Like shoes and shirt, one size does not fit all: Evidence on time series cross-section estimators and specifications from Monte Carlo experiments

Christopher Adolph
University of Washington
Department of Political Science

Daniel M. Butler
Stanford University
Department of Political Science

Sven E. Wilson
Brigham Young University
Department of Political Science

March 19, 2005

Abstract

Political scientists often and increasingly analyze time-series cross-sectional (TSCS) data. These data come with signficant problems, such as accounting for unobserved variation across sample units and appropriately specifying dynamics. Furthermore, even though fixed-effects (or least squares dummy variable (LSDV) models) can address unit heterogeneity, least squares (LS) estimation of modelds with fixed-effects and lagged dependent variables are known to be biased. Alternative estimators, mostly from economics, and generally designed for short panels, have been proposed to address this bias, but it is generally not well known how these estimators perform in comparison to simple methods like LS and LSDV on TSCS data. The preliminary results we illustrate here suggest that LSDV is generally as good or better than instrumental variables (IV) approaches in terms of bias and efficiency. We examine estimator performance under conditions where the importance of the unit effects and the correlation of the unit effects with the independent variables are allowed to vary and find that LSDV performs well. Unfortunatley none of the estimators, particularly LS, perform well when the dynamics of the model are mis-specified. The lesson is that new estimators do not, in general, solve the problem of mis-specifying the model's dynamics.

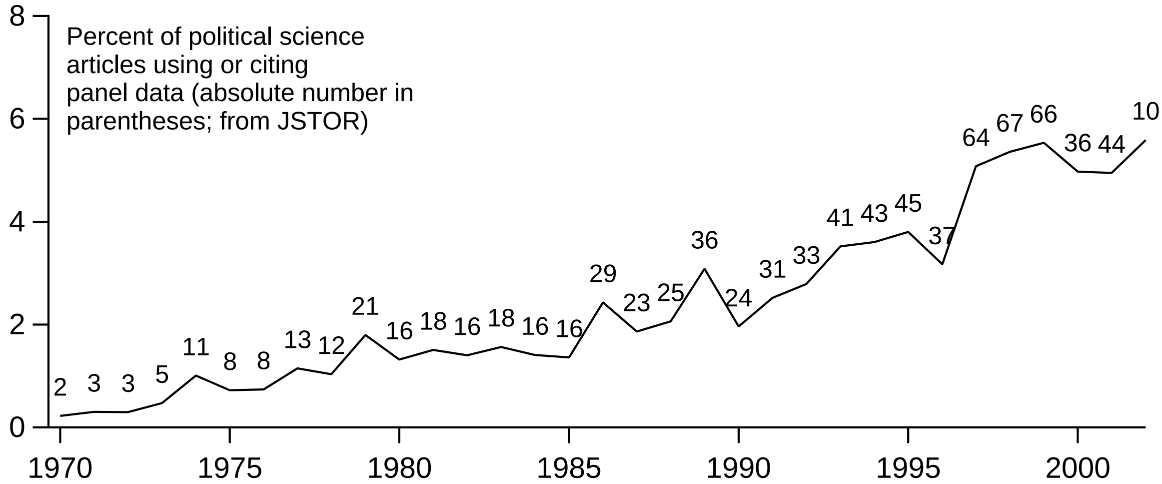

Figure 1: **Interest in TSCS data among political scientists over time.**

1 Introduction

Political scientists' interest in time-series cross-sectional (TSCS) data has increased in recent years. Figure 1 shows the number of political science articles in the JSTOR database analyzing, or at least mentioning, TSCS data, over the last three decades.[1] In the early 1970's, TSCS data was seldom used or discussed in political science. Since then, TSCS data has steadily increased in importance, and from 1996 onwards, appeared in fully 1 in every 20 political science articles indexed by JSTOR. The rise of panel data likely depends on its increased availability, the spread of easy-to-use panel data routines, and increasing quantitative sophistication in the field. TSCS data is attractive, because it allows researchers to pool time series that might be otherwise too short to analyze, to make comparisons across units, or to cope with omitted variables by sweeping out fixed unit effects.

Analysts must a pay a price to gain these advantages. TSCS data are intrinsically more challenging to model than cross-sections or time-series, and a solid understanding of the panel data econometrics and estimators is essential. In particular, there is no "one-size-fits-all" technique for TSCS data. However, researchers are not always careful in how they use TSCS data. Wilson

[1]The data in Figure 1 come from searches on JSTOR and show for each year the number of articles which includes one of the terms for panel data. The following seven terms were used in the JSTOR searches: "panel data" or "time-series cross-sectional" or "cross-sectional time-series" or "pooled cross-sectional" or "pooled cross-section" or"pooled time-series" or "repeated cross-sectional"

and Butler (2004) show that many researchers in political science treat as a "solution" to the problems of TSCS data the simple prescriptions of Beck and Katz (1995, 1996)—pool the data, include a lagged dependent variable, and calculate TSCS calculated standard errors; hereafter referred to as the BK approach. (Beck and Katz may not have meant for their recommendations to be use in this way [see Beck and Katz 2004]; however, that is how their advice has been interpreted by practitioners.) As Wilson and Butler (2004) note, treating the BK approach as a universally applicable TSCS method is problematic. In particular, it risks introduce bias if unit heterogeneity is not controlled or if the dynamics are misspecified.

But is this bias likely to be large in practice? And which, if any, of the other available techniques perform better? In this paper, we introduce a Monte Carlo (MC) framework for answering these questions, and some preliminary results from MC experiments. In general, we find that the BK approach will be seriously biased when these issues are not controlled for properly. Significantly, we also find that in the case of controlling for unit effects, the LSDV estimator outperforms the other options we consider.

To make our MC experiments as useful as possible, we took pains to choose parameters representative of the data used in political science. To this end, we surveyed a sample of political science articles analyzing TSCS data, and structured the data generating process for our experiments on the range of behavior observed in real data. Thus we begin this paper with a brief discussion of the kinds of TSCS data encountered in the field. Next, we review the estimators we compare in our experiments. Then we present results from our Monte Carlo experiments, and conclude with a discussion of their implications.

2 TSCS data as used in political science

This paper is part of a larger project to evaluate various estimators that have been proposed for TSCS data with a continuous dependent variable. Because the utility of our findings for political scientists depends on the extent to which we are able to replicate the data sets applied researchers actually face, we collected 113 recent articles using (continuous) TSCS data.[2]

[2] The sample is composed of articles from a JSTOR search in political science journals for the years 1995-2000. In the search, we used the following search terms - "panel data" or "cross-sectional time series" or "time-series cross-sectional" or "pooled cross-sectional" - and initially had 274 articles. Among these

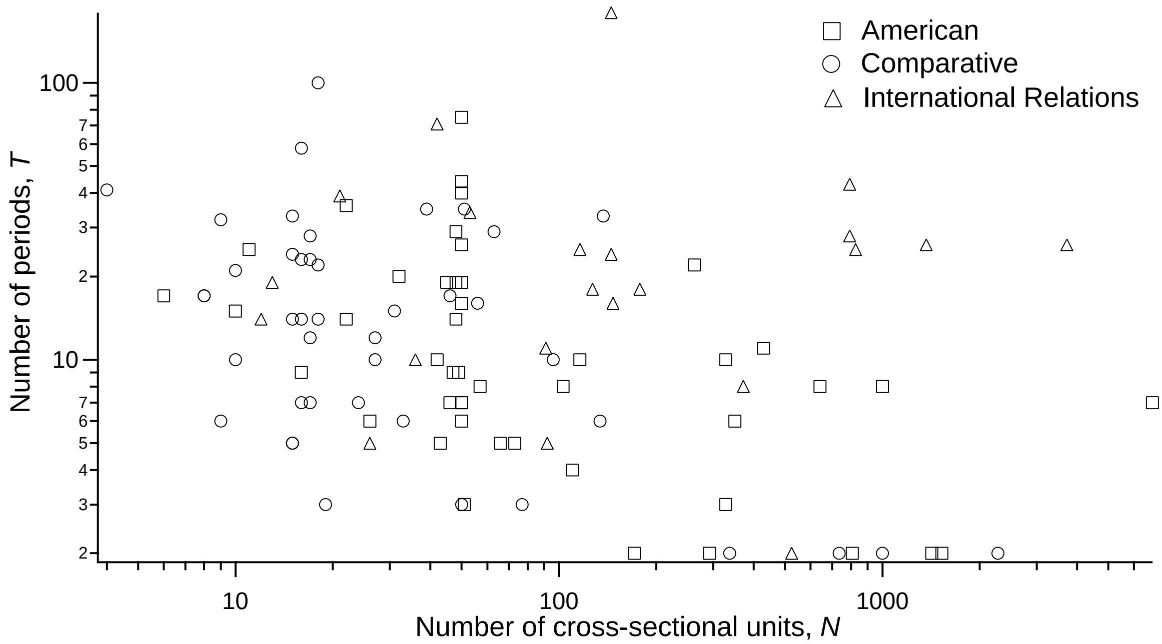

Figure 2: **Size of TSCS datasets used in political science (articles from 1995-2000).**

We start by looking at the size of the datasets used in our sample of articles. Figure 2 plots of the number of time periods T against the number of cross-sectional units, N.[3] The first thing to note is that the number of periods analyzed in political science studies is typically between 10 and 40, considerably more periods than the typical TSCS dataset examined in economics (where $T < 5$ is the norm; CITE). Most of the TSCS estimators used in political science were developed and tested for use with large N, small T datasets in economics, and may behave differently in typical political science applications.[4] Advice on the appropriateness of TSCS data estimators needs to be tailored to the evidently different problems of political scientists.

Political science panels vary by subfield, as Figure 2 illustrates. Comparativists, especially those focusing on the industrial democracies, often study between 10 and 20 units (nations) at a time. Americanists often study panels of the fifty states. International relations scholars tend to

articles there were 150 that used TSCS data sets. Because this paper only considers estimators for data with a continuous dependent variable, we excluded the 37 articles that used discrete dependent variables, leaving 113 articles.

[3]For unbalanced data sets where N was given but not T, we used the average number of time periods per unit, $\bar{T} = NT/N$. Similarly, when T was given but not N, we used $\bar{N} = NT/T$. In every case we rounded to the nearest whole number.

[4]Indeed, Beck and Katz (1995) emphasized just this distinction in recommending political scientists use LS estimators for panels with smaller N but larger T than seen in economics.

study the largest panels, sometimes treating hundreds or thousands of dyads of countries as units in a TSCS regression. It would be unsuprising if these different applications demanded different approaches.

Of course, there is more to TSCS data than its dimensions. We also wish to capture the nature of the data generating processes (DGPs) researchers face, and especially how the DGP differs from the assumptions of the standard linear regression model. To identify these issues, we looked to the concerns of applied researchers themselves, and collected a list of potential problems cited in their work. Applied researchers worry about many ways in which their data violates model assumptions, and frequently claim their data display autocorrelation, endogeneity, heteroscedasticity, complex lag structures, non-stationarity, unit heterogeneity, sample selection, and multicollinearity. While all of these issues are important and deserve attention in their own right, we focus in this paper on the issues of unit heterogeneity and lag specification.

3 Previous MC Studies of TSCS Data in Political Science

Though this is not the first Monte Carlo study of TSCS, MC evidence remains limited. Our work builds on the studies of Beck and Katz (2004), Judson and Owen (1999), and Kristensen and Wawro (2003). We share with all three papers concern for unit effects. Of the three papers however, only Kristensen and Wawro (2003) test the robustness of the pooled LS estimator (i.e. the BK approach) to unit heterogeneity. They find that as the correlation between the unit effects and the independent variable increases, the performance of the pooled LS estimator quickly deteriorates. We extend their analysis by also considering the robustness of the BK approach to changes in the size of the unit effects and the amount of trending in the independent variable.

Judson and Owen (1999) and Beck and Katz (2004) explore the potential bias that comes from including a lagged dependent variable in a LSDV, or fixed effects, model (Nickel 1981). Judson and Owen focus primarily on the bias of the coefficient of the LDV, ϕ. They note that the coefficient of the exogenous covariate, β has such small bias that there is no appreciable difference in the performance of the estimators they consider (Anderson-Hsiao, Arellano-Bond, and LSDV with the Kiviet correction). Because political scientists are generally more interested in estimates

DGP Characteristics	Judson & Owen (1999)	Kristesen & Wawro (2003)	Beck & Katz (2004)	This paper	Extensions planned
Unit effects	•	•	•	•	
Corr(x, unit effects)		•	•	•	
x is AR(1(	•		•	•	
x is AR(p)					•
ε is AR(1)		•			
ε is ARMA(p,q)					•
ε is GARCH(c,d)					•
Unit specific dynamics					•
y_{t-1} on RHS	•	•	•	•	
y_{t-2} on RHS				•	
x_{t-1} on RHS				•	
x_{t-2} on RHS				•	
Panel heteroskedasticity				•	
Endogeneity					•
Unit Roots					•
Estimation approaches					
Least Squares (LS)	•	•		•	
BK (LS with PCSE)		•			•
BK with LDV				•	
Fixed Efects (LSDV)			•	•	
Random Effects					•
LSDV-Corrected (Kiviet 1995)	•		•		•
Anderson-Hsiao (1981)	•		•	•	
Arellano-Bond GMM (1-step)	•			•	
Hsiao et al (2002)					•

Table 1: **Comparing recent Monte Carlo studies of TSCS data and estimators.**

of the effects of the exogenous independent variable, and are not as concerned about the lagged endogenous variable, Judson and Owen's results are good news. One interpretation is that so long as the researcher accounts for the unit effects, a variety of estimators give relatively unbiased results (provided, of course, N and especially T are sufficiently large). Beck and Katz (2004) concur with this basic finding, but suggest that LSDV may dominate instrumental variables methods on efficiency grounds (i.e., LSDV has lower mean squared error than the Anderson-Hsiao estimator).

Because we share many of the same interests as these previous researchers, we incorporate many of the features of their experiments in our own data generating process (DGP). Table 1 compares our approach to these earlier analyses. Our work is most closely related to Beck and Katz (2004). We build on their analysis in three ways. First, we include the least squares estimator (i.e., the original BK approach) in our experiments. Second, we explore a broader range of parameters in our MCs. In particular, we expand the dynamic nature of the model by including distributed lags of the explanatory variable, we vary the degree of correlation between the unit effects and the independent variable, and we vary the size of the unit-specific effects. Third, we examine how misspecifying the lag structure affects estimates of the parameters of interest.

As noted above, this paper is part of a larger project exploring the performance of estimators when faced with econometric issues common in political science. The last column of Table 1 shows future issues and estimators that we plan to consider. We preview some of our future efforts in the concluding section.

4 Monte Carlo design

For TSCS data, a realistic Monte Carlo experiment inevitably entails a complex data generating process. However, adding complexity to such an experiment makes it more difficult to comprehensively explore or present the full parameters space. In this section we present a DGP and MC strategy that, we hope, balances the competing demands of realism and parsimony.

4.1 Data generating process

We model the TSCS data, y_{it}, $i = 1, \ldots, N$, $t = 1, \ldots, T$ as follows:

$$\begin{aligned}
y_{it} &= \phi_1 y_{i,t-1} + \phi_2 y_{i,t-2} + \beta_1 x_{it} + \beta_{1,t-1} x_{i,t-1} + \beta_{1,t-2} x_{i,t-2} + \alpha_i + \varepsilon_{it} \\
x_{it} &= \delta x_{i,t-1} + \psi \alpha_i + \omega_{i,t} \\
\varepsilon_{it} &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \\
\omega_{it} &\sim \mathcal{N}(0, \sigma_\omega^2) \\
\alpha_i &\sim \mathcal{N}(0, \sigma_{\alpha,i}^2) \\
\sigma_{\alpha,i}^2 &= \sigma_\alpha^2 + \eta_i \\
\eta_i &\sim \mathcal{N}(0, \sigma_\eta^2)
\end{aligned} \tag{1}$$

This data generating process is similar to Beck and Katz (2004), with the notable additions of a second lag of y_{it}, distributed lags of x_{it}, and panel heteroskedasticity. This setup allows us to model most (but not all) of the features of TSCS data noted above by practitioners. In particular, it allows us to investigate the role of dynamics in x and y, as well as implications of unit heterogeneity which is correlated with the covariate of interest.

The software we have written to conduct these experiments implements a substantially broader DGP, allowing for unit specific coefficients, ARMA-GARCH errors, and other features listed in the final column of Table 2. In further work, we hope to investigate this broader setup, but for now restrict our attention to the special case in 1.

4.2 Exploring the parameter space

Beck and Katz (2004) examined a similar DGP, and varied the number of time periods T and the first lag of the dependent variable ϕ_1, while keeping all other parameters fixed. Here, we conduct two separate experiments to extend their work.

4.2.1 Correlated unit effects

A fundamental feature of TSCS data is that key differences between different units of analysis are unobserved by the researcher. In other words, there are important unit-specific omitted variables. The presence of these unit effects is captured in equation 1 by the α_i parameter. In simple regression analysis, omitted variables do not bias the regression coefficients if they are uncorrelated with X. However, in the TSCS data context, unit effects (the α's) are not independent by construction—they are all the same value within each additional unit. Furthermore, the unit effects may be correlated with the X variables, In equation 1 this correlation results from a non-zero value of the ψ parameter.

Unit heterogeneity has recently received attention in the literature in the form of a debate over when, if at at all, fixed effects should be used (Green et al 2001, Beck and Katz 2001, Wilson and Butler 2004, Plumper et al. 2005). Of course the issue itself is broader than the debate over fixed effects. In general, unit heterogeneity refers to the circumstance where units (countries, states, individuals, etc.) differ in ways not explained by observed independent variables. The same issues are also relevant not only across units but also across time periods in the form of shocks that are common to the the time component of the data structure.

In running the experiments, we are interested in comparing estimators' performance as size of the unit effects increases in magnitude and in the degree of correlation with the independent variable. We capture these features in our DGP by parameterizing the effects as draws from a normal distribution with variance σ^2_α, which we change, and by making x_{it} depend on $\psi\alpha_i$, so that non-zero ψ induces a correlation between the unit fixed effect and the covariate. Hence, our first experiment lets ψ and σ^2_α take on any combination of values from the set $\{0, 0.5, 1, 1.5\}$. That is, we consider the effect of having no fixed effects, small fixed effects, or large fixed effects, combined with no correlation with the covariate, a small correlation, or a large correlation.

For the first experiment, we restrict the dynamics of the DGP to three special cases:

	ϕ_1	ϕ_2	$\beta_{1,t-1}$	$\beta_{1,t-2}$
No dynamics	0	0	0	0
A lagged dependent variable	0.5	0	0	0
An LDV and an LIV	0.5	0	0.5	0

In these preliminary experiments, all other parameters are held fixed, including N and T. In particular, we fix $\beta_{1,t} = 1$, $N = 10$, $T = 20$, $\sigma_\varepsilon = 1$, $\sigma_\omega = 1$, $\sigma_\eta = 1$, $\delta = 0.5$.

4.2.2 Time series dynamics

A critical issue in TSCS data analysis is choosing the appropriate lag structure. Unfortunately, applied researchers often overlook or quickly dismiss the possibility of lag misspecification, perhaps because of a lack of theoretical and methodological guidance. In particular, Wilson and Butler (2004) found that many authors using TSCS data either ignore dynamics or use the BK approach and assume a single lag of the dependent variable is sufficient to model the effect of history. This is problematic because misspecifying the lag structure introduces bias to the contemporaneous effect of the independent variable (i.e. β_1) and results in errors related to interpretation or the implications of the results for policy.

Thus, in our second experiment, we turn to the role of dynamics. Here we go beyond Beck and Katz (2004) to include in our DGP both a second lag of the dependent variable, and distributed lags of the covariate. We consider all combinations of the following parameters,

$$\begin{aligned}
\beta_{1,t-1} &= \{0, 0.2, 0.6, 0.8\} \\
\beta_{1,t-2} &= \{0, 0.3, 0.5, 0.7\} \\
\phi_1 &= \{0, 0.2, 0.6, 0.8\} \\
\phi_2 &= \{0, 0.3, 0.5, 0.7\}
\end{aligned}$$

subject to the constraint that the sum of all lag coefficients of a particular variable have no more than a unit sum.[5]

[5]In the case of y_{it}, this restriction avoids explosive series. In the case of x_{it}, it restricts attention to an *a priori* plausible subset of scenarios in which the effects of a shock to x_{it} are strongest in the initial period, and gradually die out over time.

In these experiments, we again hold fixed $\beta_{1,t} = 1$, $N = 10$, $T = 20$, $\sigma_\varepsilon = 1$, $\sigma_\omega = 0.6$, $\sigma_\eta = 1$, $\delta = 0.5$, and additionally fix $\psi = 0.5$ and $\sigma_{\alpha 0} = 1$.

5 TSCS estimators

In this paper, we consider four classes of estimators for TSCS data. The first, popularized by Beck and Katz (1995), is to estimate a pooled regression by least squares, coping with TSCS heteroskedascity by calculating "panel corrected standard errors" (PCSEs). We refer to this method as LS.

A second approach, widely used in political science and economics, is to pool the data and estimate by least squares while adding a dummy variable for each unit.[6] This method is known as the fixed effects or least-squares dummy variable (LSDV) estimator. (Note that it is easy to combine the Beck-Katz method with the fixed effects estimator; one need only adjust the standard errors).

It is well-known that including a lagged dependent variable in an LSDV model introduces bias (Nickel 1981). A third method, less widely known in political science, uses instrumental variables to overcome this bias. Introduced by Anderson and Hsiao (1981, 1982), this approach regresses the first difference in y_{it} on the first difference in x_{it} and lags of the difference in y_{it}, using the lagged level of y_{it} as an instrument. This approach retains the virtues of LSDV (controlling for unit effects, and hence potentially mitigating omitted variable bias), while avoiding Nickel bias. We refer to this estimator as AH.

As Arellano and Bond (1991) noted, not only the first lag, but all the lags of y_{it} can be used as instruments in a TSCS data model set up in differences. Accordingly, they proposed a more efficient Generalized Methods of Moments (GMM) estimator employing all available lags as instruments.[7] They offer one and two-step versions of this GMM estimator, which may included the full set of feasible instruments or some subset. We focus on the one-step estimator with all instruments, and denote this estimator as AB.

[6]Equivalently, one could "sweep out" the fixed effects by substracting the unit means from both sides of the equation.

[7]See Wawro (2002) for an introduction to GMM aimed at political scientists, and Wawro (2002) and Rodden (2004) for examples of their use.

These four classes of estimators hardly exhaust the possibilities. In future work, we will also consider random effects models as well as the bias-corrected LSDV methods of Kiviet and associates (Kiviet 1995, 1999; Bun and Kiviet 2003).

6 Results

In the following subsections, we present a few preliminary results from the MC analysis. A signficant challenge in a complex MC study is deciding which of the many parameters in the data generating process to highlight and analyze. Here, we have opted to fix (for the moment) traditional MC concerns, like N and T, to more closely examine the effects of correlated unit effects, dynamics of the dependent and independent variables, and mis-specification. Future work will explore the robustness of these results across the range of N and T shown in Figure 2; for now, we simply note that the cases examined are in many ways typical of much time series cross section work in political science.

In the discussion that follows, we evaluate the performance of each estimator using bias and efficiency as criteria. In the Monte Carlo context, bias ($\beta_1 - E(\hat{\beta}_1)$) is the difference between the true value, β_1, and the mean of its estimator. The most prominent efficiency criteria is mean squared error (MSE). As is well known, the MSE can be expressed as the sum of the variance of the estimator plus the square of the bias. Thus, increasing bias tends to be associated with increasing MSE. However, significant gains in MSE may be obtained by employing a slightly biased estimator that has low variance. In the graphs that follow, we present results for bias and MSE separately, and the value for each is displayed by using a gray scale, with darker shades representing higher values. Lower values (lighter shading) are, of course, preferred.

6.1 Unit Effects

In this analysis we highlight the effects of two important aspects of unit heterogeneity: 1) the importance of the unit effects, which is captured by increasing the variance of α and 2) the extent to which the unit effects are correlated with the independent variable, which is captured by the ψ parameter. The working paper by Kristensen and Waro (2003) suggests that the LS framework performs particularly poorly as the second effect becomes stronger. Ignoring unit effects as LS does will obvioulsy result in biased estimates, but it is not clear how the available estimators will perform under different specifications of the data generating process.

In order to isolate the effects of the unit effects, we assume in the analysis that follows that the researcher has correctly determined the dynamic structure of the data. Here we consider the following three cases: 1) the simple case where no dynamics exist; 2) the LDV model (which has become the default standard in political science since Beck and Katx (1995) promoted its use; and 3) the case where lagged independent variables (LIVs) are important as well. A natural extension of our analysis is to evaluate estimator performance when researchers fail to appropriately specify the dynamic structure of the model.

Figures 3 and 4 present the results of our simulations for the four estimation approaches and three dynamic specificaitons discussed above. Not surprisingly, the LS estimator performs poorly except in the base case where no dynamics are involved and where σ_α and ψ are both zero (in other words, when the assumptions of the classical model hold, LS is the best approach). Figure 3 shows, however, some additional insight into how unit heterogeneity interacts with the dynamics of the model. In the LDV case, we find the same thing as Kristensen and Waro, namely that increasing the correlation between the unit effects and the indpendent variables worsens the performance of the LS estimator. For the LSDV estimator, increasing σ_α actually lowers the bias that results from including an LDV in the specification. This is because as unit effects become more important, they explain more (relative to X) of the variance in Y (though there are no corresponding gains in effeciency, as Figure 4 shows).

The findings of Kristsen and Waro are reversed, however, when the specification includes an LIV in addition to an LDV. In this case, it is the importance of the unit effects (σ_α) that is the primary culprit. In fact, increasing ψ actually reduces the bias associated with the LS estimator in this case. This is likely because neglecting the LIV is capturing some of the effect of the unit effects, an effect that grows stronger as the ψ parameter gets larger. Furthermore, Figure 4 reveals that in the case where no dynamics are involved, the efficiency of the LS model is primarily a function of σ_α, not ψ. Thus the Kristensen and Waro results hold for the LDV case but not for the no dynamics case or the case where an LIV is also present. Thus, their paper should not be interpreted to mean that, in general, LS is appropriate as long as the unit effects are not correlated with the regressors. It can easily be the case that the overall importance of the unit effects, even when uncorrelated with X, will be the dominant factor in increasing both the bias and MSE of the LS estimator.

One argument for the simple LDV approach of Beck and Katz is that including an LDV does much the same thing as including fixed effects in the model, since it accounts for the overall level of the dependent variable, much as fixed effects do. However this seems to be true only when the importance of the fixed effects is low and when the model does not include an LIV. An LDV may partially mitigate the exclusion of fixed effects, but it is hardly a general solution.

Finally, a central motivating question in this analysis is whether the IV estimators proposed by Anderson and Hsiao and Arellano and Bond outperform the simple LSDV estimator in the types of data sets typically used by political scientists. With respect to the analyses illustrated in Figures 3 and 4, the answer seems to be that they do not. In particular, the Anderson-Hsiao estimator is strictly dominated by the LSDV and Arellano-Bond estimators. Furthermore, the LSDV estimator performs at least as well as and sometimes better than Arellano-Bond (this is particularly the case when there are no dynamics in the model, though in that case, GMM approaches are not recommended in the first place, because the LSDV estimator is unbiased when dynamics are not present).

6.2 Time series dynamics

In this section we switch our focus from unit heterogeneity to misspecification of the dynamic structure, though the data used in the analysis still includes unit heterogeneity. In short, we examine the effects of failing to identify the correct number of lagged variables in the data. A review of the applied literature on TSCS data in political science by Wilson and Butler (2004) show relatively little attention paid to the question of dynamics except for the ad hoc inclusion of an LDV in many studies. Any good time series textbook discusses issues involved in correctly specifying the lag structure of the data and those lessons apply just as well to the TSCS case.

We illustrate in the following analys the effects of neglecting lagged values in both the dependent and independent variables. For the sake of parsimony in presentation, we treat the LDVs separtely from the LIVs, though they could just as easily be considered together (and we hae

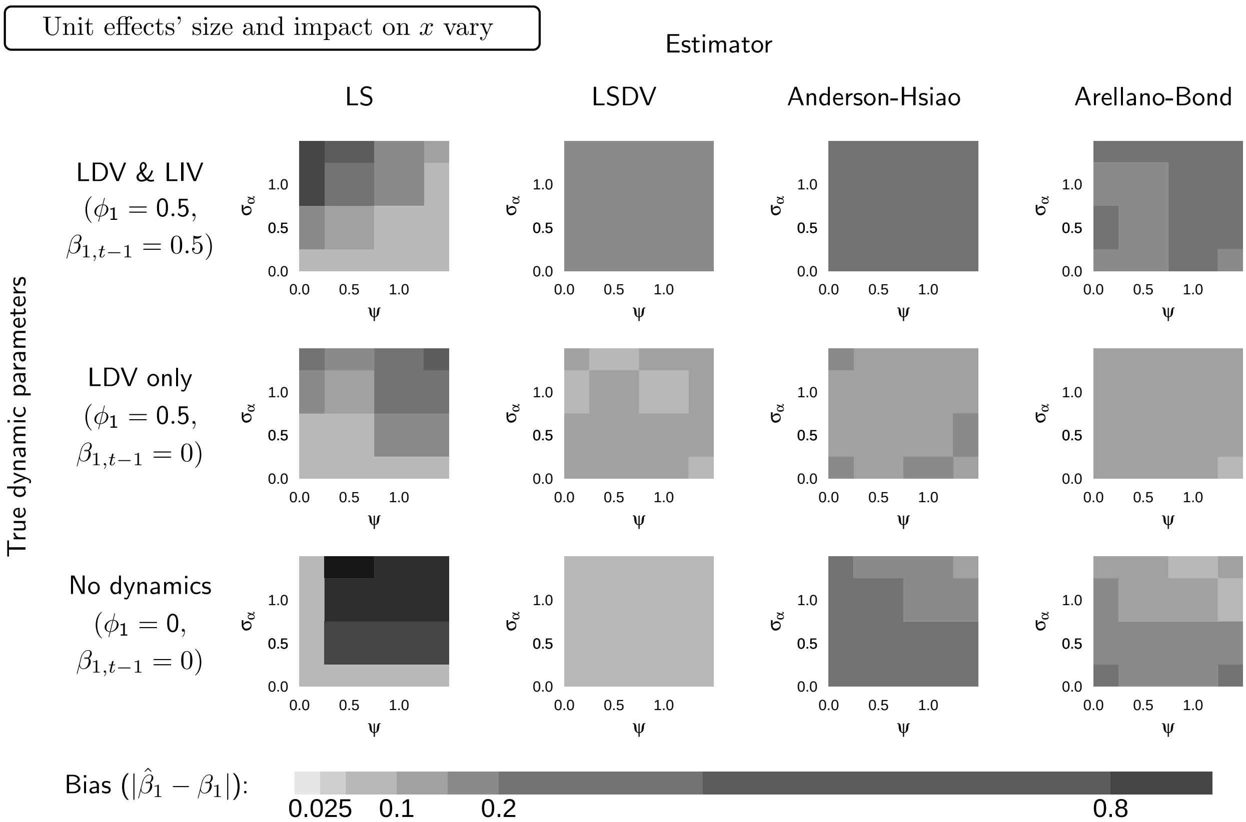

Figure 3: **Bias ($|\hat{\beta}_1 - \beta_1|$) in various time series cross-section estimators (columns) under different dynamic parameters (rows), different impacts of unit effects on x (horizontal axes of plots), and different sizes of unit effects (vertical axes).** Each cell in each plot shows the bias in a different scenario for the given combination of estimator and dynamics (all models are correctly specified). For all scenarios, 1000 simulations were drawn from a TSCS process with $N = 10$, $T = 20$, $\beta_{1,t} = 1$, $\sigma_\varepsilon = 1$, $\delta = 0.5$, $\sigma_\omega = 0.6$, and $\sigma_\eta = 1$. Each of the N time series begins with a burn-in period of 50 observations, which is discarded.

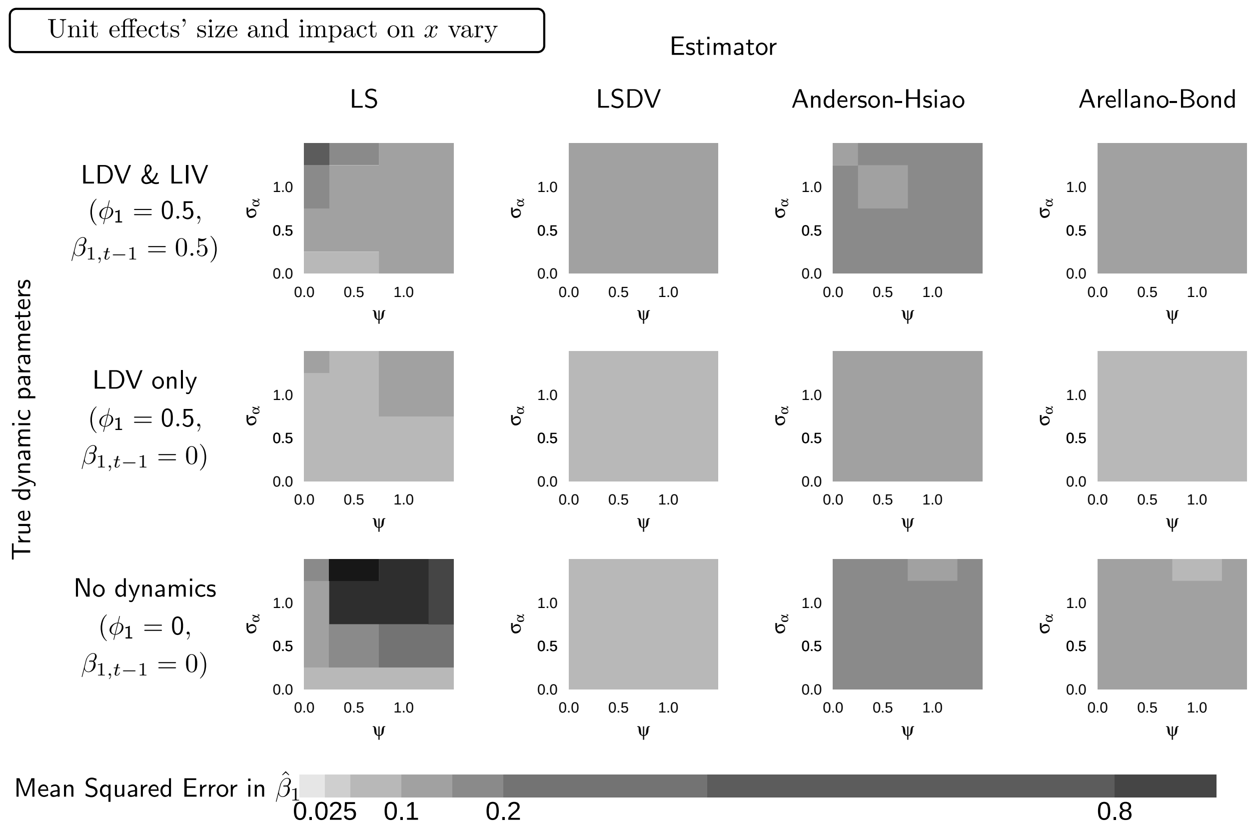

Figure 4: **Efficiency (MSE) of various time series cross-section estimators (columns) under different dynamic parameters (rows), different impacts of unit effects on x (horizontal axes of plots), and different sizes of unit effects (vertical axes).** Each cell in each plot shows the mean squared error for a different scenario for the given combination of estimator and dynamics (all models are correctly specified). For all scenarios, 1000 simulations were drawn from a TSCS process with $N = 10$, $T = 20$, $\beta_{1,t} = 1$, $\sigma_\varepsilon = 1$, $\delta = 0.5$, $\sigma_\omega = 0.6$, and $\sigma_\eta = 1$. Each of the N time series begins with a burn-in period of 50 observations, which is discarded.

done so in other MC experiments, not shown). While many researchers include an LDV, very few studies have considered LIVs, even though there is a long literature on properly identifying the appropriate number of lags. The Anderson-Hsiao and Arrellano-Bond estimators include lagged values as instruments but not as direct effects. For this and other reasons, it is possible that the performance of these IV estimators will differ from the LS and LSDV estimators.

We examine first the case of LDVs, where the results are shown in Figures 5 and 6. Each of the twelve plots shows the performance under various values of the true parameters. The rows of the figures correspond to different specifications. In the bottom row, no lags are included, in the middle row, only a first-order lag is included, while the top row is correctly specified. As before, we consider the performance of the four estimators in terms of bias and MSE. Note that because the choice of row (specification) and column (estimator) is up the researcher, but the location of the data within the plot is in practice unknown, a good estimator/specification combination would need to have globally good performance across the plotted space.

The abysmal performance of the LS estimator is not unexpected, but the profound bias that results even when the coefficients on the LDVs are very small illustrates the importance of correctly specifying the equations. Indeed, this lesson holds as well for the other estimators. Compared with a correct specification, the choice between the LSDV and the IV estimators is a second-order concern. We note that when using the simple LS specification, including an LDV is definitely superior to the simple case of excluding it all together. Thus TSCS studies that include no lagged values are particularly suspect. The advice by Beck and Katz (1995) to include an LDV when doing pooled-OLS turns out to be quite helpful in this regard, though note that for cases where the ϕ_1 is smaller than ϕ_2 the simple LDV does not do well (in practice, this might occur when adjustment processes are slow or periodic in some fashion).

For all the estimators, mis-specification (in terms of leaving out a signficant second order lag) has a strong effect on the bias estiamates, though there is less effect for MSE. The choice of estimator does little to alleviate this. When we turn to LIVs (Figures 7 and 8), much the same story is evident. Mis-specification has very negative consequences, particularly in the LS case. In contrast, a correctly specified LSDV is clearly the superior estimator. Interstingly, the one case where the LSDV variable seems to underperform the IV estimators is when the second lag is important and incorrectly omitted. This is the one case where the Anderson-Hsiao estimator is actually superior to the others. The LSDV performs worse than the other two in this case in terms of both bias inefficiency. In contrast, the LSDV does better when both lags are incorrectly omitted.

In the particular scenario we have chosen to illustrate, there is no advantage to using the IV estimators when the equation is correctly specfied. The LSDV is simpler to implement and interpret and it outperforms the IV approaches. But we caution the reader that we have not fully examined variations in other parameters that may undermine this general result. For instance, much more work needs to be done in exploring different combinations of N and T. Judson and Owen find that the IV approaches are somewhat better than LSDV when T is very small.

7 Discussion

Many analysts have reached a concensus that unit effects are an important consideration in any TSCS analysis. Even Beck and Katz, who ignored the issue in their highly influential paper of 1995, now point out the importance of accounting for unit effects (2004). The results here confirm, under a fairly general data generating process, the poor performance of the LS estimator when the estimating equation is not correctly specified and the classical assumptions do not hold. We have shown that adding an LDV helps, but, in general, the original BK prescriptions (pooled-OLS with LDV and PCSEs) lead to seriously biased and inefficient estimation.

This concensus on fixed effects does not necessarily mean that we know enough to approach TSCS analysis with any degree of confidence. Analytical results that have been well-known for decades imply that least squares estimation of panel data models are biased for finite T, even as as N goes to infinity. This bias becomes worse the smaller T gets.

The good news is that evidence is beginning to accumulate (not only this paper, but previously Monte Carlo work) that the LSDV model generally performs—when the dyanmics are correctly specified—as well or better than more complicated IV estimators. A clear result from our experiments is the superior performance of the LSDV model in handling correlation between the unit effects and the independent variable (i.e. Figures 3 and 4), a situation that is likely to occur often in practice. Ours is the first analysis to show this result. The strong performance of

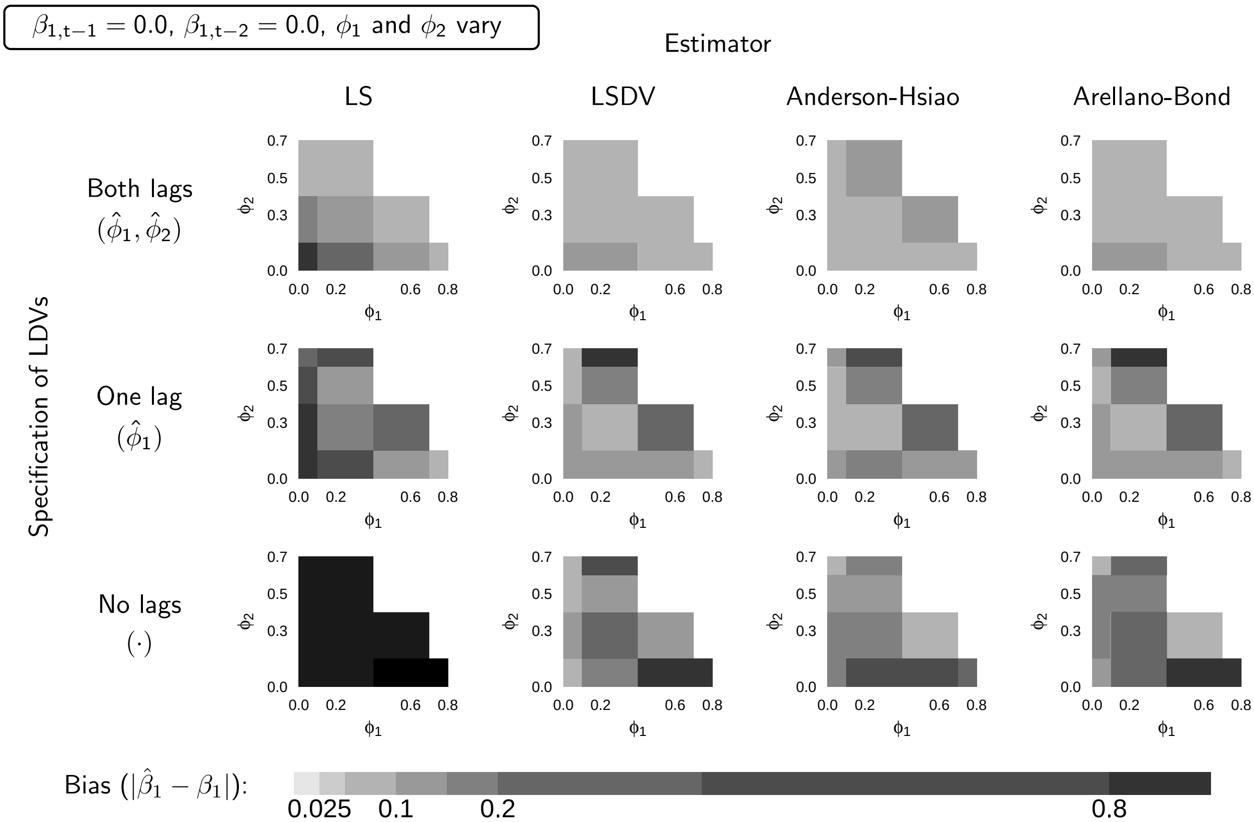

Figure 5: **Bias ($|\hat{\beta}_1 - \beta_1|$) in various time series cross-section estimators (columns) under different dynamic specifications (rows) and different true coefficients on first and second lags of the dependent variable (horizontal and verical axes of plots).** Each cell in each plot shows the bias in a different scenario for the given combination of estimator and specification. For all scenarios, 1000 simulations were drawn from a TSCS process with $N = 10$, $T = 20$, $\beta_{1,t} = 1$, $\beta_{1,t-1} = 0.0$, $\beta_{1,t-2} = 0.0$, $\sigma_\varepsilon = 1$, $\delta = 0.5$, $\psi = 0.5$, $\sigma_\omega = 0.6$, $\sigma_\alpha = 1$, and $\sigma_\eta = 1$. Each of the N time series begins with a burn-in period of 50 observations, which is discarded. For these runs, the specification also includes no lags of the independent variable, which happens to be the correct specification.

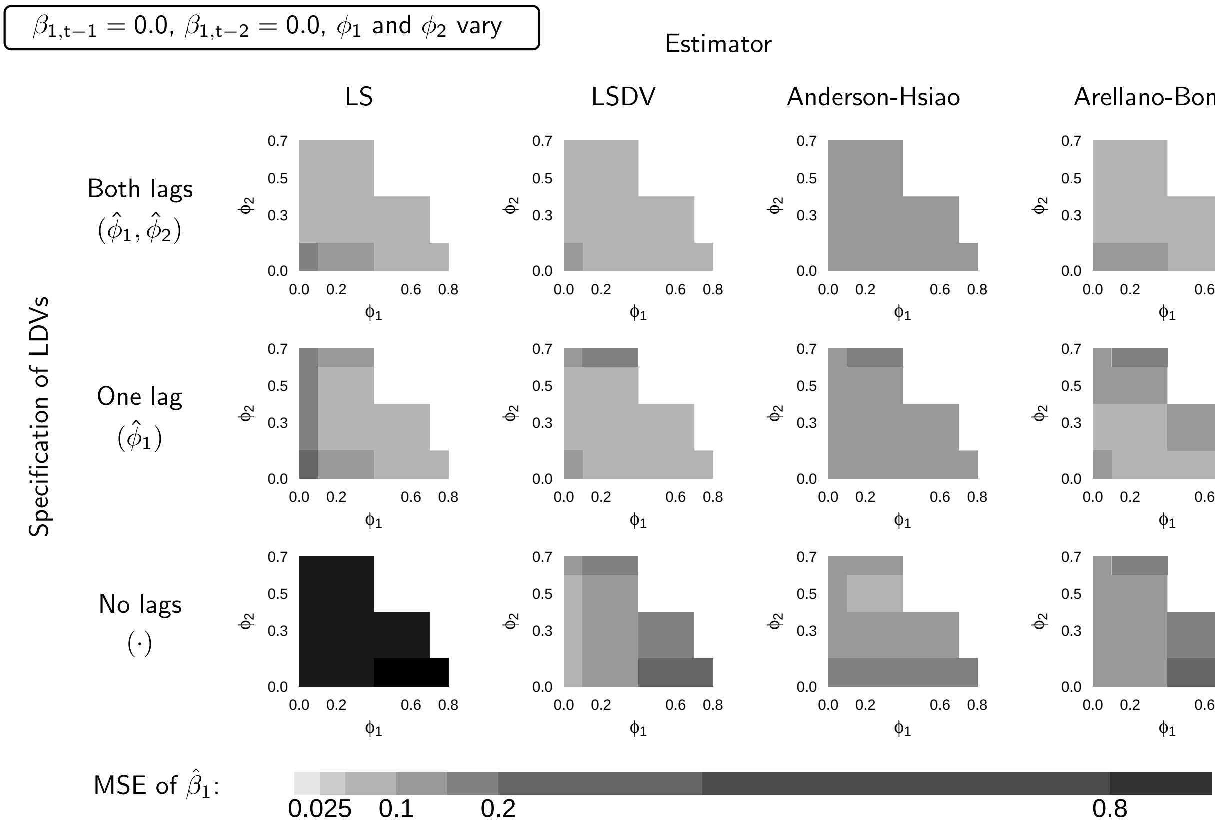

Figure 6: **Efficiency (MSE) of various time series cross-section estimators (columns) under different dynamic specifications (rows) and different true coefficients on first and second lags of the dependent variable (horizontal and verical axes of plots).** Each cell in each plot shows the mean squared error for a different scenario for the given combination of estimator and specification. For all scenarios, 1000 simulations were drawn from a TSCS process with $N = 10$, $T = 20$, $\beta_{1,t} = 1$, $\beta_{1,t-1} = 0.0$, $\beta_{1,t-2} = 0.0$, $\sigma_\varepsilon = 1$, $\delta = 0.5$, $\psi = 0.5$, $\sigma_\omega = 0.6$, $\sigma_\alpha = 1$, and $\sigma_\eta = 1$. Each of the N time series begins with a burn-in period of 50 observations, which is discarded. For these runs, the specification also includes no lags of the independent variable, which happens to be the correct specification.

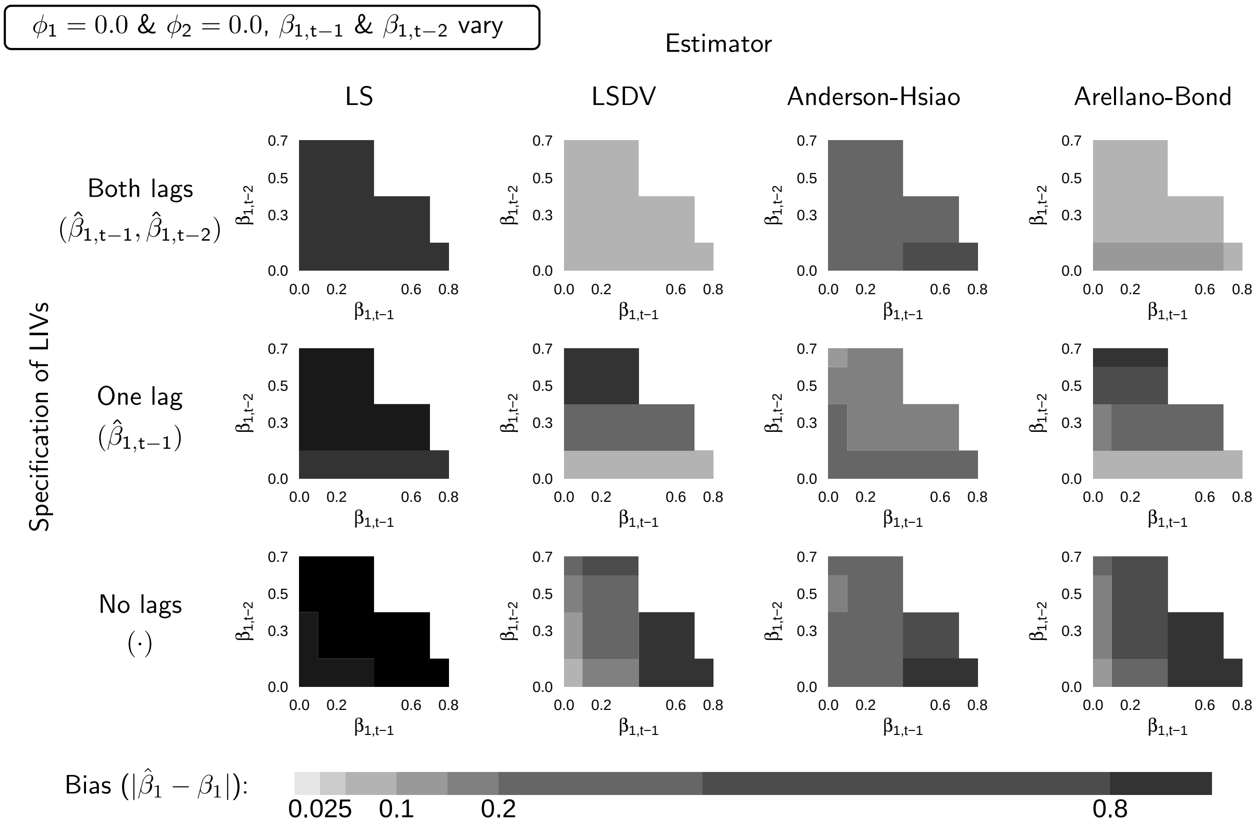

Figure 7: **Bias ($|\hat{\beta}_1 - \beta_1|$) in various time series cross-section estimators (columns) under different dynamic specifications (rows) and different true coefficients on first and second distributed lags of the covariate (horizontal and vertical axes of plots).** Each cell in each plot shows the bias in a different scenario for the given combination of estimator and specification. For all scenarios, 1000 simulations were drawn from a TSCS process with $N = 10$, $T = 20$, $\beta_{1,t} = 1$, $\phi_1 = 0.0$, $\phi_2 = 0.0$, $\sigma_\varepsilon = 1$, $\delta = 0.5$, $\psi = 0.5$, $\sigma_\omega = 0.6$, $\sigma_\alpha = 1$, and $\sigma_\eta = 1$. Each of the N time series begins with a burn-in period of 50 observations, which is discarded. For these runs, the specification also includes no lags of the dependent variable, which happens to be the correct specification.

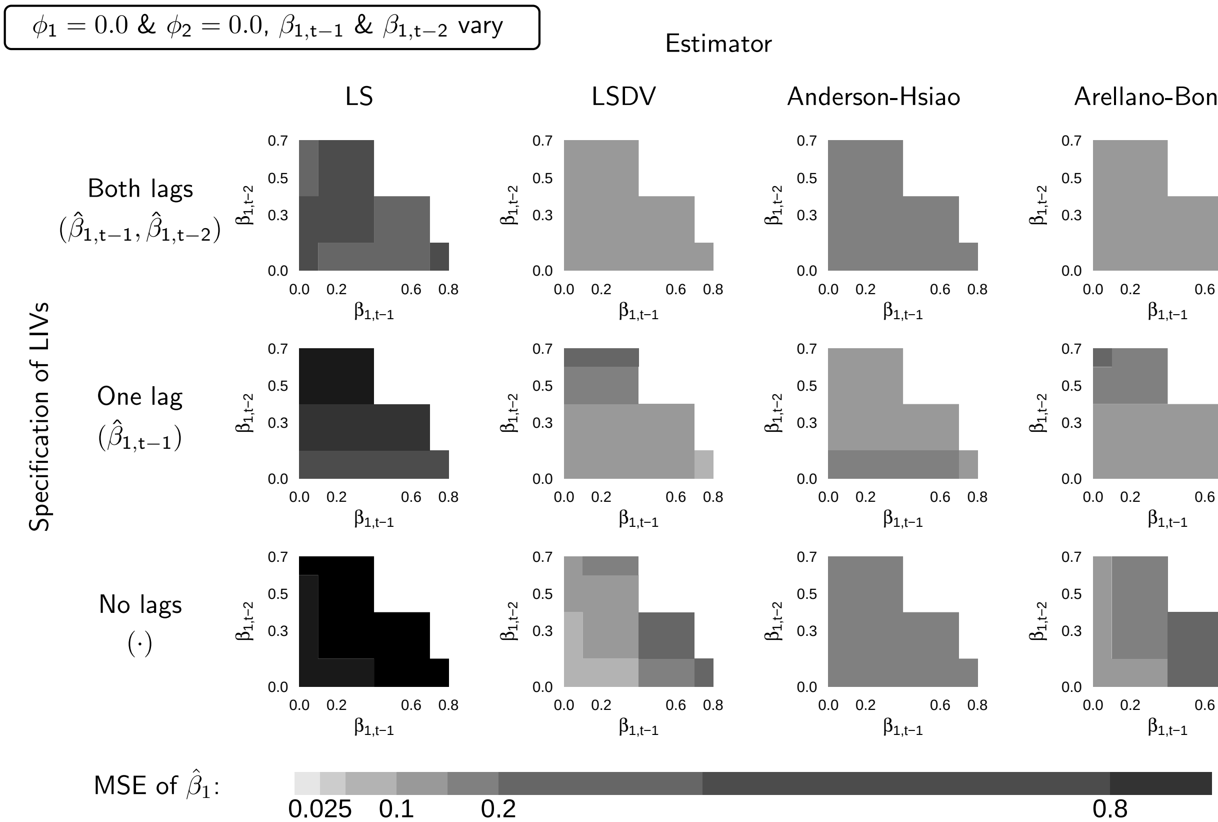

Figure 8: **Efficiency (MSE) of various time series cross-section estimators (columns) under different dynamic specifications (rows) and different true coefficients on first and second distributed lags of the covariate (horizontal and vertical axes of plots).** Each cell in each plot shows the mean squared error for a different scenario for the given combination of estimator and specification. For all scenarios, 1000 simulations were drawn from a TSCS process with $N = 10$, $T = 20$, $\beta_{1,t} = 1$, $\phi_1 = 0.0$, $\phi_2 = 0.0$, $\sigma_\varepsilon = 1$, $\delta = 0.5$, $\psi = 0.5$, $\sigma_\omega = 0.6$, $\sigma_\alpha = 1$, and $\sigma_\eta = 1$. Each of the N time series begins with a burn-in period of 50 observations, which is discarded. For these runs, the specification also includes no lags of the dependent variable, which happens to be the correct specification.

the LSDV also shows up in previous MC work we cite elsewhere. We suspect that the LSDV will be shown to be the estimator of choice for a wide range of settings. Furthermore, the bias that does exist in dynamic panel data models is mostly an effect on the coefficient of the LDV, not the coefficient on the independent variables. In political science, the LDV is usually included for purposes of control, not because it is of central importance. Even though estimates of β_1 may be very reliable, researchers should be cautioned that ϕ_1 may still have significant bias. (Note this also biases estimates of the long-run effects of the independent variables, which may often be the quantities of real interest.)

Given the preliminary nature of this study, a caveat is in order. Although Judson and Owen find in favor of LSDV when $T > 20$, they show that IV approaches are slighlty better when T is 10 or less, though their data generating process is not as general as the one employed here. They also note the strong performance of Kiviet's (1995) corrected-LSDV estimator (available only for balanced panels). The results here are very preliminary and different scenarios may serve to illustrate cases where the LSDV is not optimal. Data sets where T is very small will pose signficant challenges under any estimation technique, and fixed-effects, in particular, are less useful the smaller T becomes.

Not all news is good, however. Mis-specification of the lag structure of the model both in terms of LDVs and LIVs results in serious bias and inefficiency for all the parameters of the model, including β_1. This holds for all estimators, particularly LS. Unfortunately, though accurate specification of lag structures in time-series research is not a new problem, it remains a challenging one. For instance, deciding what to do with a group of variables that are individually insignificantly but are a) jointly significant and b)influential on the parameters of interest is not an easy question. Given that models usually have multiple independent variables, adding lags can result in a very cumbersome presentation and interpretation of results. The hassle involved in accounting for LIVs is surely a major reason that almost no one in applied political science research actually does it (evidence is given in Wilson and Butler (2004)).

Unfortunately, it appears that political scientists will need to face the problem of lagged variables. As we have shown, the errors are simply too profound to sweep this problem under the rug. (Beck and Katz (2004) are certainly in agreement on the point that dynamics need to be carefully addressed.) Nor is the solution likely to come in the form of a better estimation technique. We might hope that IV approaches including past values of both the dependent and independent variables would yield adequate solutions when the estimating equations themselves are mis-specified. They don't.

Another serious analytical problem the profession has to come to grips with is that there may not be an appropriate panel data approach for dealing with datasets where the variable of interest is either time-invariant (which makes FE impossible) or slowly-moving (which causes standard errors to be very high). Beck and Katz (2004) continue to advise researchers that it may be the case that fixed-effects should not be used if the variables of most interest are invariant or slowly moving. We question the utility of using seriously biased coefficients just because they have a strong theoretical or intuitive appeal. Perhaps alternative estimation schemes will be developed that can yield unbiased (or even consistent) estimates of these effects, but we are not there yet.

References

Adolph, Christopher. 'Visual Interpretation and Presentation of Monte Carlo Results.' *The Political Methodologist*, 2003, 11.2: 31–35.

Adolph, Christopher. 2003. *SeeMC*. R and Gauss software for visualization of Monte Carlo results, available at `http://www.chrisadolph.com`.

Anderson, T.W. and C. Hsiao. 1982. "Formulation and Estimation of Dynamic Models Using Panel Data." *Journal of Econometrics* 18:47-82.

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *The Review of Economic Studies* 58(April): 277-297.

Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to do (and not to do) with Time-Series Cross-Section Data." *American Political Science Review* 89(September):634-64

Beck, Nathaniel, and Jonathan N. Katz. 1996. "Nuisance v. Substance: Specifying and Estimating Time-Series-Cross-Section Models." *Political Analysis* 6:1-36.

Beck, Nathaniel, and Jonathan N. Katz. 2001. "Throwing out the Baby with the Bath Water: A Comment on Green, Kim, and Yoon." *International Organization* 55(spring): 487-495.

Beck, Nathaniel, and Jonathan N. Katz. 2004. "Time-Series-Cross-Section Issues: Dynamics, 2004." Annual Meeting of the Society for Political Methodology, July, Stanford, CA.

Bun, M. J. G., and J. F. Kiviet. 2003. "On the diminishing returns of higher order terms in asymptotic expansions of bias." *Economics Letters* 79: 145-152.

Green, Donald P., Soo Yeon Kim, and David H. Yoon. 2001. "Dirty Pool." *International Organization* 55(spring): 441-468

Judson, Ruth A., and Ann L. Owen. 1999. "Estimating Dynamic Panel Data Models: A Guide for Macroeconomists." *Economic Letters* 6: 9-15.

Kiviet, Jan F. 1995. "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models." *Journal of Econometrics* 68(July): 53-78.

Kiviet, Jan F. 1999. "Expectation of expansions for estimators in a dynamic panel data model: some results for weakly exogenous regressors." In Hsiao, C., K. Lahiri, L-F Lee, and M. H. Pesaran. eds. *Analysis of Panel Data and Limited Dependent Variables.* Cambridge: Cambridge University Press.

Kristensen, Ida P. and Gregory Wawro (2003). "Lagging the Dog? The Robustness of Panel Corrected Standard Errors in the Presence of Serial Correlatoin and Observation Specific Effects." Manuscript, Columbia University.

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49(November): 1417-1426

Plumper, Thomas, Vera E. Troeger, and Philip Manow. 2005. "Panel Data Analysis in Comparative Politics: Linking Method to Theory." *European Journal of Political Research* 44: 327-354.

Rodden, Jonathan. 2002. “The Dilemma of Fiscal Federalism: Grants and Fiscal Performance around the World.” *American Journal of Political Science* 46 (July): 670-687.

Wawro, Gregory. 2002. “Estimating Dynamic Panel Models in Political Science.” *Political Analysis* 10(1): 25-48.

Wilson, Sven E., and Daniel M. Butler. 2004. “A Lot More to Do: The Promise and Peril of Panel Data in Political Science.” Manuscript, Department of Political Science, BYU.