Time-Series–Cross-Section Issues: Dynamics, 2004*

Nathaniel Beck† Jonathan N. Katz‡

Draft of July 24, 2004

ABSTRACT

This paper deals with a variety of dynamic issues in the analysis of time-series–cross-section (TSCS) data raised by recent papers; it also more briefly treats some cross-sectional issues. Monte Carlo analysis shows that for typical TSCS data that fixed effects with a lagged dependent variable performs about as well as the much more complicated Kiviet estimator, and better than the Anderson-Hsiao estimator (both designed for panels). It is also shown that there is nothing pernicious in using a lagged dependent variable, and all dynamic models either implicitly or explicitly have such a variable; the differences between the models relate to assumptions about the speeds of adjustment of measured and unmeasured variables. When adjustment is quick it is hard to differentiate between the models, and analysts may choose on grounds of convenience (assuming that the model passes standard econometric tests). When adjustment is slow it may be the case that the data are integrated, which means that no method developed for the stationary case is appropriate. At the cross-sectional level, it is argued that the critical issue is assessing heterogeneity; a variety of strategies for this assessment are discussed.

1. INTRODUCTION

Clearly the analysis of time-series–cross-section (TSCS) data is an important issue both to students of comparative political economy and students of political methodology. While it is hard to get a count on the number of substantive papers which use TSCS data, an appendix in Wilson and Butler (2004) lists about 150 published papers which cite our original APSR article (Beck and Katz, 1995) on the analysis of TSCS data . While this is clearly not a necessary condition for such analysis, it does set a lower bound on the number of TSCS papers (the majority of which are in comparative political economy). There have recently

*For research support we thank the National Science Foundation.

†Department of Politics; New York University; New York, NY 10003 USA; nathaniel.beck@nyu.edu

‡Division of the Humanities and Social Sciences; California Institute of Technology; Pasadena, CA 91125 USA; jkatz@caltech.edu

also been a number of published articles, unpublished manuscripts and conference papers related to TSCS methodology.[1]

This seems like a good time for us to state our position on a number of issues raised in these methodological articles. Here we primarily deal with issues related to modeling dynamics for continuous (in the sense of social science) dependent variables, since this is both of interest to substantive researchers and much discussed in the various methodological articles. We begin with some notation and nomenclature and then in Section 3 briefly discuss some non-dynamic issues. We then turn to the two major sections of this paper. Section 4 presents some Monte Carlo analyses of the performance of OLS in the presence of fixed effects and a lagged dependent variable; Section 5 discusses the modeling of dynamics in general, and includes some data analyses. We then offer some general conclusions in the final section.

It is not our intention to simply defend our written position against critics, or to engage in a textual analysis of what we actually wrote versus how that has been interpreted (a matter of little interest to anyone but us). Thus our goal is to come up with recommendations for analysis where the recommendations seem clear, and to lay out some issues where the recommendations are less clear.

2. NOTATION AND NOMENCLATURE

Notation

Let $y_{i,t}$ be a observation for unit i at time t for the time series y, where $i = 1, \ldots, N$ and $t = 1, \ldots, T$. We assume that y is measured as a continuous variable, or at least is close enough that we can take it as continuous. Since in what follows we typically do not care if we have one or more than one independent variable or variables, let $\mathbf{x}_{i,t}$ be a similar observation on a vector of independent variables. Since all our arguments hold for either a single independent variable or a vector of variables, we will typically suppress the boldface notation and talk of x as a single independent variable; this should minimize confusion.[2]

Since the paradigmatic applications are to comparative political economy, we will often refer to the time periods as years and the units as countries. To simplify notation, we are also assuming that the data set is rectangular, that is, each country is observed for the same time period. This assumption is benign; it should cause no problems if some units start or end a year or two later than do others.

We can begin with what we might call the pooled model:

$$y_{i,t} = \beta x_{i,t} + \varepsilon_{i,t}. \tag{1}$$

[1] Those we know of are, in alphabetical order, Achen (2000), Hayes (2000), Kittel (1999), Kittel and Winner (2003), Plümper, Troger and Manow (2004), Kristensen and Wawro (2003) and Wilson and Butler (2004). Substantive researchers, such as Garrett and Mitchell (2001) and Huber and Stephens (2001) also have methodological sections on the appropriate methods to analyze their own data.

[2] In addition, we suppress the constant term, which implies that the data have been centered; this is just to simplify notation for the time series models.

This implies that all observations, and particularly all observations for all units,[3] follow the same process with the same exact parameters.[4]

TSCS vs Panel Data

Equation 1 could look either like TSCS or "panel" data; our previous work has stressed that these are different, though various analysts have questioned this distinction. Early on we made the unfortunate distinction between temporally and serially dominated data sets (with $T > N$ and $N > T$). This obviously leaves out some categories. But it is also, for us, the wrong issue. The critical issue is whether T is large enough to do serious averaging over time, and also whether it is large enough to make some econometric issues disappear. While there is no magic cutoff level here, we note that "panel" studies almost invariably have single digit T's (with 3 being a common value) while the comparative politics TSCS data sets we work with commonly have T's of twenty or more.

As we shall see, the size of T tells us a great deal about which potential econometric problems might be serious ones for the data set being analyzed. This is important, since much econometric work is on panel data rather than TSCS data. There are issues that are critically important in panel data which econometricians have devoted much effort to resolving. Various analysts (Hayes, 2000; Kristensen and Wawro, 2003; Wilson and Butler, 2004) have argued that these same methods should be used for TSCS data. We return to our own analyses in Section 4, but here simply note that the issues which have important consequences for panel data estimation cause many fewer problems for TSCS data.

3. NON-DYNAMIC ISSUES

There is no panacea

The most important point made by all the various methodological articles is that there is no panacea for TSCS data, and one cannot write a computer program which just has the command "TSCS dv iv." Whatever recommendations one comes up with for analyzing TSCS data, the actual analysis of any data set must depend both on the relevant theory and a variety of informed choices which must be made in any empirical analysis. Obviously theory should guide us as much as possible, though we do note that many of the theories in comparative political economy are not very precise in providing such guidance (other than perhaps in providing a list of regressors).

Clearly researchers need to understand what the various methods do and do not do, and what problems they solve and what problems they do not solve. Researchers, much as they like, cannot proceed mechanistically, or assume that the technical details which underlay

[3] We stress homogeneity or heterogeneity of units. As Kittel and Winner (2003) note, we also might worry about heterogeneity across time. This, of course, is sensible, but introduces no new issues, and hence when we discuss heterogeneity, it is across units, not time.

[4] Equation 1 makes the very strong assumption that the effect of x on y is the same cross-sectionally and temporally. This is a very strong assumption. For progress on weakening this assumption we refer the reader to the work of Zorn (2001). Unfortunately this work does not seem to have been widely read.

any method are of no substantive importance. On this everyone agrees!

Our "panel corrected standard errors" (PCSEs) are no exception to the above. They are not a cure all for all TSCS problems.[5] They are what they say they are: they correct the OLS standard errors for two TSCS (not panel!) problems: groupwise heteroskedasticity and contemporaneous correlation of the errors. They do no more than that. We contend that for TSCS data they are superior to OLS standard errors. We will argue below as to why we often like OLS methods, but here we simply note that if one uses OLS for TSCS data, PCSEs are usually better than the OLS standard errors, and seldom are very much worse. PCSEs cure no other problems, and obviously are not a panacea (though we do contend they should replace the OLS standard errors more or less routinely for TSCS data).[6]

As various critics (Maddala, 1998) have noted, PCSEs give up on any attempt to model the process generating either groupwise heteroskedasticity or contemporaneous correlation of the errors, and simply attempt to "fix up" the standard errors. This is clearly correct. While it is obviously better to model the source of these "problems," and we have tried to do so elsewhere,[7] our only position here is that if one is going to use OLS and not model these various TSCS "problems" then one is better off using PCSEs in place of the usual OLS standard errors.

Homogeneity vs. heterogeneity

Obviously a major issue in TSCS data is the eternal question of "to pool or not to pool." Typically the two alternatives that are considered are the fully pooled model (Equation 1) or the completely unpooled variant, which has a separate β_i for each unit. Less discussed, but perhaps of more importance, is the question of whether one or a few units should be analyzed separately because they are dissimilar from the preponderance of units. These are critical questions. In other works we have discussed some graphical methods (box plots) and statistical methods (cross validation) which may aid researchers in figuring out which units are not sufficiently like the others so as to be pooled with those others.

It is obviously true that all units differ; it is equally obvious that we gain efficiency if we assume that these differences are not so great that we cannot analyze all units together. But how should we make that trade-off? In prior work (Beck and Katz, 2001a) we have found that the traditional F test for pooling too often rejects pooling (in favor of the alternative of complete non-pooling). That is that the gains from pooling offset the costs of pooling more than standard statistical theory asserts. We also found that alternative methods, all related

[5] We unfortunately used P instead of TSCS back in 1995; PCSE's are not particularly relevant for panel data.

[6] We also think they are superior to the White "heteroskedasticity consistent standard errors" that some TSCS analysts use. The White HCSEs do not take advantage of an intuition that heteroskedasticity in TSCS data is likely groupwise and they do not take account contemporaneous correlation of the errors (they were designed for the analysis of cross-sectional data). The White HCSEs should be superior to OLS standard errors, but they will not be as good as PCSEs if the errors show a TSCS structure (which seems likely although not inevitable).

[7] In work with Kristian Gleditsch, we are trying to use spatial econometric methods to better model contemporaneous correlation of the errors.

to random coefficient models, did not really seem to solve the problem of "partial pooling."

At this point we do not have anything to add to this. Our silence does not indicate that we do not think that this problem is critical. As Bartels (1996) has observed, all analysts, even those analyzing simple cross sections, must make decisions about whether each observation is generated by the same process as the others. While it is the rare analyst who worries about this in the cross-sectional context, it is clearly easier to assess non-pooling in the TSCS context (whether by standard F-tests or methods such as cross-validation). But the fact that some data gathering or political group, such as the OECD or the European Union, happens to provide data on N countries does not mean that all those countries should be analyzed as one. As Kittel (1999) notes, the question of which countries should be analyzed together is perhaps the most critical one faced by the TSCS (or any comparative) analyst. So we agree with Kittel and others that this question is critical. We trust that no one thinks our silence indicates that everyone should simply take all data and analyze it as if it were generated identically.

Fixed Effects

A number of analysts have noted that the simplest way to allow for heterogeneity is to assume that each unit has its own intercept, the "fixed effects" (FE) model. This model simply adds a dummy variable for each country, f_i to Equation 1.[8] This equation is simple to estimate by OLS.

In our first discussion of PCSEs we noted that they worked for models with or without fixed effects. Various analysts have noted how badly a model can be misspecified if fixed effects are not used when the data indicate they are needed; this is basically an omitted variable bias issue. These analysts are clearly correct, and misspecifying a model by incorrectly omitting fixed effects can have very bad consequences, that is, severe omitted variable bias.

Analysts all realize that including fixed effects also has other consequences. The standard result (Greene, 2003, 287–8) is that estimating the FE model is identical to estimating Equation 1 with y and x replaced by their unit centered deviations.[9] Thus the fixed effects remove any of the average unit to unit variation from the analysis, and simply ask whether intra-unit changes in y are associated with intra-unit changes in x. This is akin to assuming that only interrupted time series quasi-experimental evidence should be taken into account, and that unit to unit differences in average x and y tell us nothing about the relationship of x to y (Green, Kim and Yoon, 2001). In general this is not the position held by most comparative political economists, who would be quite happy to explain comparative economic performance by institutional variables that either do not change much or are totally stable. Fixed effects clearly eliminates any stable variables from the analysis, but also makes it difficult for variables that change only slowly to show their impact (when their impact is by and large inter- and not intra-unit).

[8] Usually Equation 1 has a constant term, in which case one unit is taken as the reference unit and has no dummy variable associated with it.

[9] Formally, regress y^* on x^* where $y^*_{i,t} = y_{i,t} - \bar{y}_i$, and $\bar{y}_i$ is the average of the T observations on unit i and similarly for x^*.

Obviously a fixed effects analysis should not conclude anything about the inter-unit effects of the independent variables, since such effects have been removed. As several have suggested, one could then regress the estimated fixed effects on various institutional variables to assess the inter-unit effects of those variables. We leave this to others (Plümper and Troger, 2004), but clearly want to warn about the appropriate conclusions from a fixed effects analysis.

In Beck and Katz (2001b) we argued that perhaps fixed effects were not necessary in one specific analysis of international trade. There we argued that the classical F-test may be overly liberal in rejecting the null hypothesis of no effects. Since that null hypothesis is that all effects are exactly zero, we may be likely to reject when we either have a lot of data (so even small effects can be almost certainly known to be not zero) or a lot of effects, so a few are quite likely not to be zero. We have no analyses that show that the F-test rejects the null of no effects too often, or that our preferred, more conservative measure, the BIC performs better. But we do worry that when the null of no fixed effects is only marginally rejected that the overall interpretive costs of including fixed effects may be more costly than the extra bias from excluding them (if that bias is seen to be small). Thus with large T's, where we can often reject the null that even a very small estimated fixed effect is zero, we must be very careful in deciding about whether or not to include fixed effects in our model. Where the estimated effects are large and clearly significant, there is no doubt that they should be included in the model.

Of course it is possible that the F-test rejects the null of no effects because one unit shows a large effect. That could be taken as an indication that the one unit should not be analyzed with the others, particularly if there are other indications that this might be the case. Even if that approach is rejected, it still might make sense to simply include the one effect in the model, which would have many fewer pernicious consequences than the complete FE model. Or it might turn out that the effects really reflect two "clubs" of units; this again could be modelled by a single dummy variable with much less pernicious effect (particularly when the club grouping makes theoretical sense, as in perhaps a division of EU nations between the old and new EU). There are no rules for this, so analysts must think hard about how to model effects. We fully agree with other analysts that one should not automatically exclude all fixed effects because of their undesirable removal of inter-unit variability, but nor should we simply go to a full fixed effects specification based on one F-test of one particular null (nor is it necessarily the case that we need to rely on only one common test reporting a P value over or under .05).

Simple methods

A number of analysts have suggested a variety of complicated estimation methods which fix various problems in TSCS models (either of some bias or inconsistency). Obviously serious problems should be fixed, but we should also note the cost of the fixes. Sometimes the cost is statistical, as in the use of instrumental variable ("IV") estimation to deal with issues of inconsistency. While obviously consistency is a desirable property, what we really care about are the mean squared error properties of the various estimators for the sample sizes we are using, not for infinite sample sizes. In general the justification for IV estimators is purely asymptotic, and it is not obvious that the finite sample properties of these estimators is superior to that of the inconsistent (usually OLS) estimator (Bound, Jaeger and Baker, 1995).

But there are also non-statistical costs to these complicated estimators. Most applied researchers use standard statistical packages. Very often these packages require that users of complicated estimators forgo other methods that would have been available using OLS (or related least squares estimators, such as NLLS). Thus, for example, these complicated estimators make it difficult to check for non-linearities (whether parametrically via transforms like the Box-Cox transform or non-parametrically via such devices as the Generalized Additive Model). It is also almost impossible to combine these complicated estimators with various spatial methods. These are but examples; the commitment to a complicated method clearly limits most applied researchers to a subset of the analyses that could have been done in the OLS framework. This is clearly a cost, and so the move to these more complicated methods must justify this (hard to quantify) cost.

At the extreme, some recommended methods are not implemented in any popular packages and are hard to program in general. Thus, for example, Hayes (2000) recommends the Kiviet (1995) bias corrected estimator for models with a lagged dependent variable and fixed effects. As far as we know, this method is only available via some Gauss code, and is not implemented in any of the popular panel/TSCS packages such as Stata or Eviews. Moreover, even the Gauss implementation requires a fully balanced panel which is rare in applied TSCS work. Thus, in the only published Monte Carlo study (which we know of) which assesses the various dynamic panel estimators concludes that "for a sufficiently large N and T, the differences in efficiency, bias and RMSEs of the different techniques become quite small. Even so, the results ... do highlight one technique that consistently out performs the others, LSDVc [Kiviet]. Unfortunately, while LSDVc may produce superior results, it is not always practical to implement. In particular, a method of implementing LSDVc for an unbalanced panel has not yet been developed. ... Our results indicate that if LSDVc cannot be implemented that (1) when $T > 30$, LSDV [OLS with FEs] performs just as well or better than the viable alternatives...." (Judson and Owen, 1999, 13). While we think that Judson and Owen are optimistic in believing that applied researchers will use the Kiviet estimator for balanced panels, even they, for very practical reasons, recommend the simpler OLS based method for many practical situations.

Judson and Owen also rightly point out that the cost of various mistakes varies with, at least, N and T. Many of the methods considered by the various analysts were developed to deal with very real problems that arise for very small T. This is a panel, not a TSCS, type of problem. So in order to investigate the gains from departing from simple (and easily available methods), we must do our Monte Carlo (or other) studies using parameters similar to what we might find in actual research. Only then can we know the real gains from complicated methods, and then perhaps assess whether those gains are worth it. In the next section we do one such analysis for models similar to those considered by Judson and Owen.

4. FIXED EFFECTS WITH LAGGED DEPENDENT VARIABLES

In this section, we consider the problems introduced by the presence of unit effects in a dynamic model of TSCS data. We will limit our discussion to dynamic models including only the lagged dependent variable and leave a more general discussion of dynamic specification to Section 5. We also note that most of the findings presented in the section are well known in the econometrics literature.

Proposed models

Formally the model we are interested in is:

$$y_{i,t} = \phi y_{i,t-1} + \beta x_{i,t} + \alpha_i + \varepsilon_{i,t}. \tag{2}$$

This model allows for dynamics via the lagged dependent variable and unit heterogeneity in the mean via α_i. A natural first approach to estimate Equation (2) is to including a series of dummy variables for each unit in the pooled model, leading to the Least Squares Dummy Variable Estimator (LSDV).

As is well known, the LSDV estimator is equivalent to demeaning all of the variables by their individual specific means. Since α_i is constant across a unit, removing its mean removes it from the equation. There is a problem, however, since the demeaning procedure makes use of all available time periods it induces a correlation between the demeaned lagged dependent variable and the demeaned error term. That is,

$$\tilde{y}_{i,t-1} = y_{i,t-1} - \frac{1}{T_i}\sum_{t=1}^{T_i} y_{i,t-1}$$

$$\tilde{\varepsilon}_{i,t-1} = \varepsilon_{i,t} - \frac{1}{T_i}\sum_{t=1}^{T_i} \varepsilon_{i,t}$$

and, therefore,

$$\mathrm{E}[\tilde{y}_{i,t-1}\tilde{\varepsilon}_{i,t}] \neq 0.$$

The error term, $\varepsilon_{i,t-1}$ is contained with weight $1 - 1/T_i$ in $\tilde{y}_{i,t-1}$ and with weight $1/T_i$ in $\tilde{\varepsilon}$. This correlation renders the LSDV estimators of ϕ and β biased. In fact, Nickell (1981) derived the asymptotic bias (as $N \to \infty$) and showed that it was $\mathrm{O}(T^{-1})$.

Given that the LSDV is biased, there have been many alternative estimators proposed. Perhaps the most common approach is to use instrumental variables (IV) as suggested by Anderson and Hsiao (1982). The Anderson-Hsiao (AH) estimator begins by handling the unit effects by first differencing Equation (2). As with the demeaning procedure, this eliminates the unit effect but introduces correlation between the transformed errors and lagged transformed dependent variable. This correlation is then handled by using an instrumental variable that is correlated with the lagged first differenced but not the differenced error term. AH proposed using either the second lag of the dependent variable, $y_{i,t-2}$, or the second lag of the differenced lagged dependent variable. The consensus is that the level instrument works better, so we will only consider it here. A central problem with any IV estimator is that while it is unbiased it may dramatically increase mean squared error if the instrument is not highly correlated with the problematic variable. That is, the researcher needs to understand the cost of correcting the biases. We might be trading a small reduction in bias for a large decrease in efficiency.

We should also note that given that the instrument proposed by AH is weak, there have been several alternative IV estimators proposed within the general method of moments (GMM) framework. GMM estimators can handle different numbers of instruments for each observation. Therefore, Arellano and Bond (1991) suggested using all available lags at each observation as instruments. This estimator is more efficient than the AH estimator but has not seen much use in political science.

A completely different approach is taken by Kiviet (1995). While the LSDV is biased, it often has a smaller mean squared error than the proposed IV estimators (as we will see below). Therefore, if the bias of the LSDV could be estimated and used to correct the estimate, it might prove superior to either the uncorrected LSDV or the AH estimators. Kiviet (1995) derives a formula for the bias of the LSDV which has a $O(N^{-1}T^{-3/2})$ approximation error.

However, applying Kiviet's procedure is not as straight forward as it seems. First, the calculations needed to compute the bias approximation are complex.[10] Second, the formula requires knowledge of the true parameters in Equation (2), which are not known (otherwise why would we be doing estimation?). Kiviet suggests plugging in values from a consistent estimator of the model, such as AH discussed above, but this will add noise to the estimate. Third, the approximation formula implicitly assumes the data are balanced — i.e., all units are fully observed for the same number of time periods. To the best of our knowledge, the approximation has never been extended to the case of unbalanced data. Lastly, we have no direct way to calculate standard errors using this correction. The mostly likely approach to measure uncertainty in this case would be a (block) bootstrap method. However, care would need to be taken to maintain the proper dynamic structure of the data.

The motivating case for the development of all of these dynamic panels models was the case of very short panels with T's in the single digits. In that context, a bias of $O(T^{-1})$ is extremely problematic. In fact, the Monte Carlo study in Kiviet (1995) are for the cases of $T = 3$ and $T = 6$, where the alternatives to LSDV perform substantially better that it. However, in the case of TSCS, we typically see a T of greater than 10, and 20 or 30 is not uncommon. It is not clear in these cases that the proposed fixes are worth their costs, either in terms of mean square error or not allowing researchers to pursue other issues. We will evaluate them in our own Monte Carlo experiments for the more typical cases seen in TSCS data.

[10]Complexity is always difficult to accurately measure, but the Gauss code that implements the calculation takes more than 30 lines.

Monte Carlo Experiments

The Monte Carlo experiments we ran are based on those from Kiviet (1995) but with T and N chosen to match TSCS data as seen in typical political science. We are going to explore how the LSDV, AH, and Kiviet Correction (KC) perform in finite samples where the data generating process is very clean. In fact, the experiments are similar to those presented by Judson and Owen (1999) with similar conclusions.

The data were generated according to Equation (2) with the following additional assumptions:

$$\varepsilon \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$
$$\alpha_i \overset{iid}{\sim} N(0, \sigma_\alpha^2)$$
$$\sigma_\alpha = \mu(1-\phi)\sigma_\varepsilon$$
$$x_{i,t} = \delta x_{i,t-1} + \gamma(1-\delta)\alpha_i + \omega_{i,t}$$
$$\omega_{i,t} \overset{iid}{\sim} N(0, \sigma_\omega^2)$$
$$i = 1, 2, \ldots, N$$
$$t = 1, 2, \ldots, T.$$

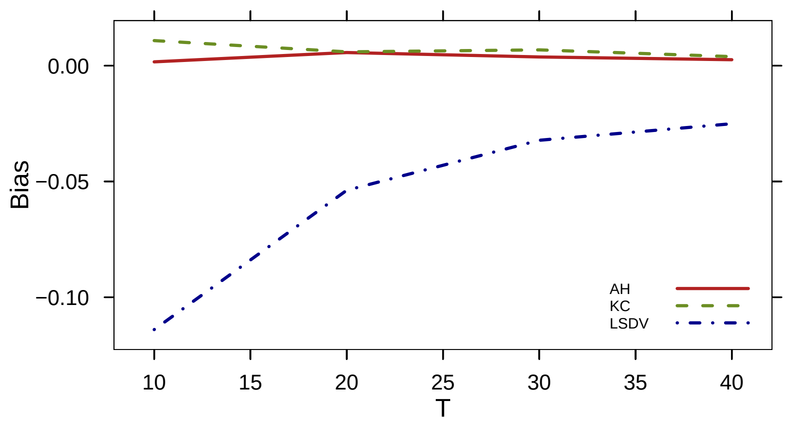
These assumptions are fairly standard in the literature. The parameterization of σ_α lets μ give the relative importance of the unit effects to the idiosyncratic errors in a straight forward manner. Further, the inclusion of $\gamma(1-\delta)\alpha_i$ induces a correlation between the unit effects and the exogenous variable, $x_{i,t}$. This is not crucial in these experiments, since all of the estimators can handle correlation between the unit effects and the regressors (unlike the random effects estimator), but we think that such correlation is common in actual data. We are particularly interested in how the estimators perform as both T and ϕ vary. The other parameters were fixed at a single value for the experiments, since they did not qualitatively change the findings.

We are interested in two criteria for evaluating the proposed estimators: bias and root mean square error. However, root mean square error is more important since it incorporates both bias and estimation variability. That is, we might be willing to use a slightly biased estimator that had dramatically smaller sampling variance.

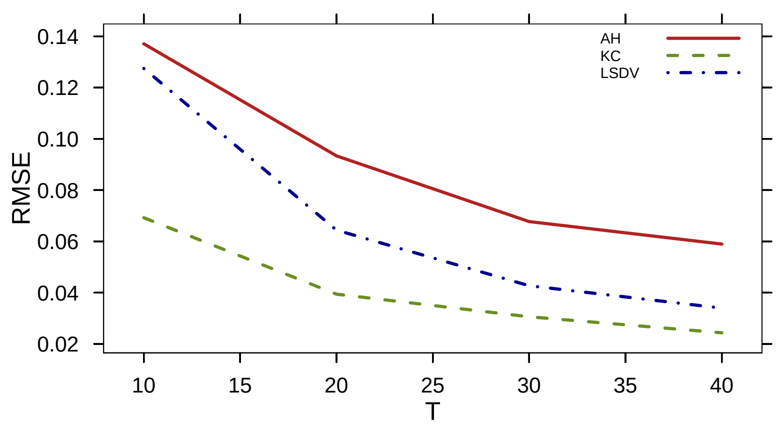
The experiments proceed by drawing the error terms and constructing the autoregressive series, $x_{i,t}$ and $y_{i,t}$. Since we do not want to worry about the impact of initial conditions, we actually let the process run for $T + 50$ periods and discard the first 50 observations. Then given the data we estimate the model using the three proposed estimators. This is repeated 1000 times. The average bias and root mean squared error is calculated for the estimates of the two parameters, ϕ and β, by the three estimators.

A complete set of results for the Monte Carlo may be found in the tables. We will look at some graphs of selected results to get a feel for what is occurring. We will first examine what happens as T varies. We fixed ϕ at an intermediate level of 0.6. In Figure 1 we see the results for the estimates of ϕ. We see from the results on bias, that as expected both the AH IV and the KC estimators are essentially unbiased, but there is substantial bias in the LSDV, particularly for small T. The picture changes somewhat when we look at RMSE. Here the KC estimator continues to dominate, but the AH estimator pays a high cost in terms of sampling variability to get unbiasedness. In fact, in terms of RMSE, the LSDV is superior to the AH.

The picture for LSDV continues to improve if we look at the results for β, typically the parameter of interest in most analysis. Figure 2 graphs out bias and RMSE for the estimates of β as a function of T. Here again, the AH estimate is unbiased, but is clearly dominated in terms of RMSE by both the KC and LSDV, even though both are slightly biased. The



(a) Bias



(b) RMSE

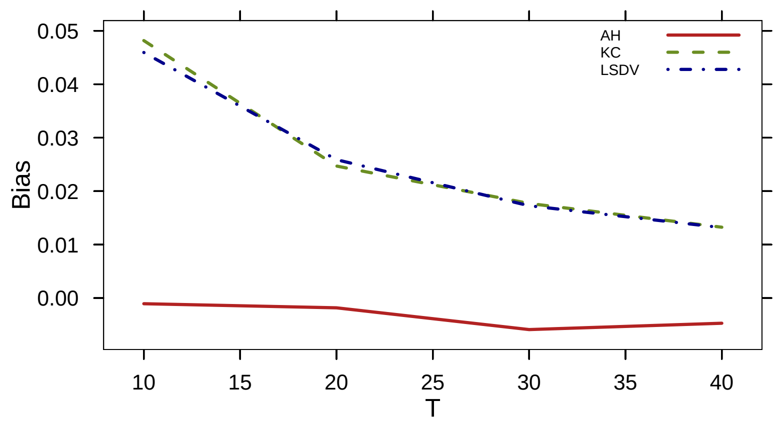
Figure 1: Monte Carlo Results for estimates of ϕ as a function of T from LSDV, Kiviet Correction, and Anderson-Hsiao estimators. Simulation parameters: $N = 20$, $\beta = 1$, $\phi = 0.6$, $\delta = 0.5$, $\sigma_\omega = 0.6$, $\mu = 1$, $\gamma = 0.3$, and $\sigma_\varepsilon = 1$

.

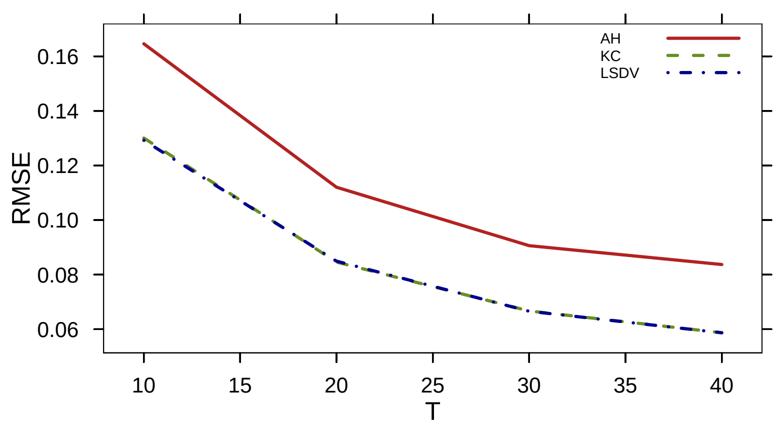
RMSE of both the LSDV and KC estimators are virtually identical, so the simplicity of LSDV clearly gives it the nod over KC for estimating β.

In addition, if one looks at the estimated long run impact of the x ($\beta/(1-\phi)$), which is often the quantity of interest, the graph would look very similar to Figure 2, with LSDV doing as well as the much more complicated KC estimate. These results may be found in Table A.3.

We have also examined how the estimators perform as the serial dependence in the



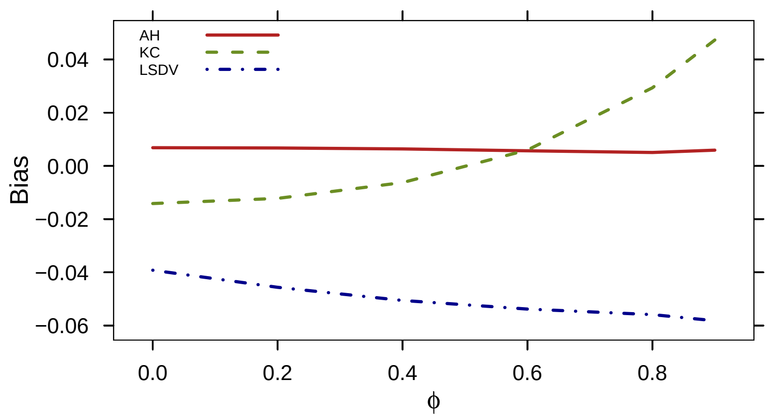
(a) Bias



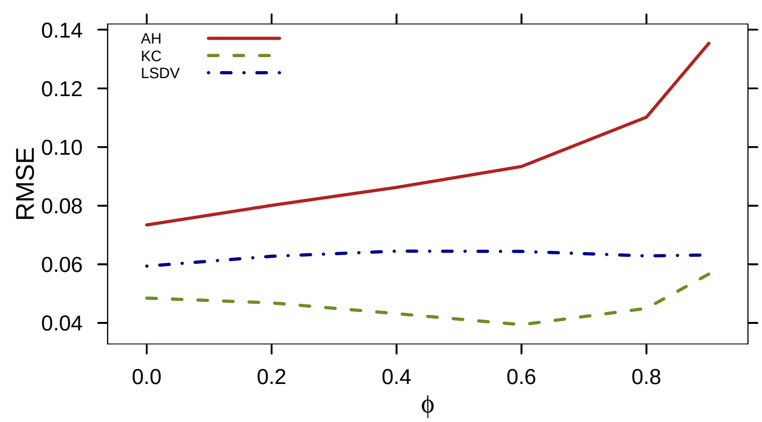
(b) RMSE

Figure 2: Monte Carlo Results for estimates of β as a function of T from LSDV, Kiviet Correction, and Anderson-Hsiao estimators. Simulation parameters: $N = 20$, $\beta = 1$, $\phi = 0.6$, $\delta = 0.5$, $\sigma_\omega = 0.6$, $\mu = 1$, $\gamma = 0.3$, and $\sigma_\varepsilon = 1$. Note that in the RMSE graph, the Kiviet Correction and LSDV have practically the same value so cannot be seen separately on the graph.

dependent variable change. Figure 3 graphs the results for the estimates of ϕ as ϕ varies from 0 to 0.9. We fixed T at 20, but got similar results for other values of T. As we have seen before, the AH estimator is practically unbiased across the entire range. Similarly, the level of the bias of the LSDV remains relatively constant, perhaps with a slight deterioration as ϕ increases. The KC estimator performance, however, in terms of bias, is quite variable. However, for RMSE we see that performance of the KC estimator, while variable, always out performs the LSDV and AH estimators. But, in terms of RMSE, the difference between the LSDV and KC estimators is not large, with both out performing the AH estimator.



(a) Bias



(b) RMSE

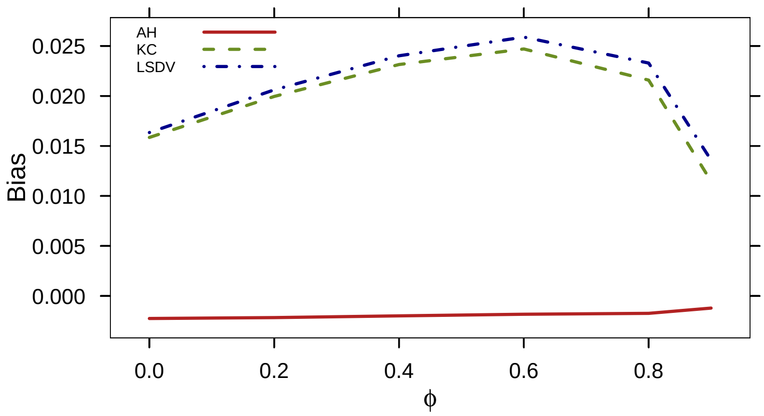
Figure 3: Monte Carlo Results for estimates of ϕ as a function of ϕ from LSDV, Kiviet Correction, and Anderson-Hsiao estimators. Simulation parameters: $N = 20$, $T = 20$, $\beta = 1$, $\delta = 0.5$, $\sigma_\omega = 0.6$, $\mu = 1$, $\gamma = 0.3$, and $\sigma_\varepsilon = 1$

.

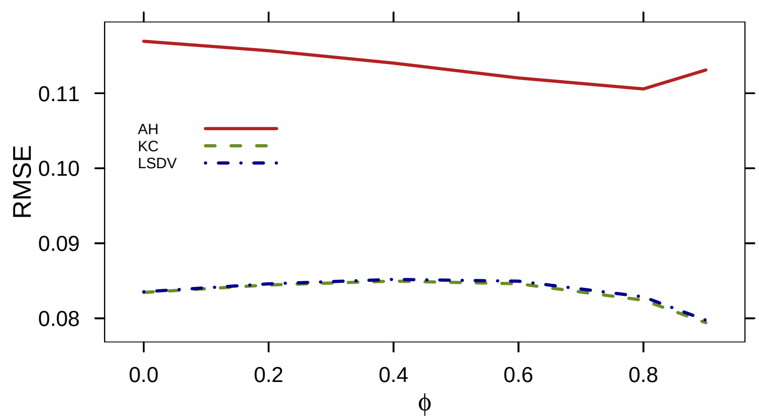
The results for the estimates of β as ϕ varies are found in Figure 4. As in Figure 2, the performance of the KC and LSDV are hard to distinguish but they both clearly outperform the AH estimator in terms of RMSE.

Advice

Given the results from these simulations the AH estimator should not be used for TSCS data. While it is clearly unbiased, the cost for this is very high. The picture with regard to



(a) Bias



(b) RMSE

Figure 4: Monte Carlo Results for estimates of β as a function of ρ from LSDV, Kiviet Correction, and Anderson-Hsiao estimators. Simulation parameters: $N = 20$, $T = 20$, $\beta = 1$, $\delta = 0.5$, $\sigma_\omega = 0.6$, $\mu = 1$, $\gamma = 0.3$, and $\sigma_\varepsilon = 1$

.

the Kiviet correction versus the simpler LSDV estimator is less straightforward. It is clear for our results, and those of others, that the Kiviet correction works well to lower the bias, particularly of the estimate of ϕ, with little cost in terms of RMSE.

That said, as discussed above, there are real costs in using the Kiviet correction, not the least of which is that it will not currently work with unbalanced data and standard errors will need to be calculated by some sort of block bootstrap. Given these costs and relatively good performance of LSDV for longer TSCS data that we typically see in applications, we see little reason, *in general*, not to prefer LSDV over the Kiviet estimator when T is twenty or more. The LSDV performs relatively well and is flexible enough to allow other estimation and/or specification problems to be dealt with.[11]

5. TIME SERIES ISSUES

In our previous work, we recommended that including a lagged dependent variable (LDV) is a convenient way of handling dynamics. We note four critiques of this: a) one should always think about dynamic specifications, which should be based on theory, and which will differ based on theoretical considerations; b) LDVs are atheoretical (we should not explain current y by lagged y); c) using LDVs is similar to modeling first differences, and we should not model first differences if we want to model lags; and d) LDVs cause serious econometric problems which lead LDVs to obscure the actual relationship between independent variables of interest and the dependent variable (they incorrectly "dominate" the regression). Implicit in the latter three objections is the notion that using a GLS correction of the errors (Cochrane-Orcutt or the like) causes fewer problems than the use of an LDV and is a better way to proceed. (We have discussed in Section 4 another critique which has to do with LDVs in the presence of fixed effects and do not reconsider that issue here.) The issues of this section are identical to those that would be made in a single time series context. In this section we assume that all processes are stationary, postponing the (brief) discussion of modeling non-stationary data to the next section.

Thinking about specification is always good

The first critique is impossible to disagree with. Researchers should always think about their specification, and should insofar as possible base this specification on well developed theory. We are perhaps more skeptical than some about whether theory is sufficiently well developed as to commonly be a guide, but where it is, it should be used. We also note that many assumptions can be relaxed, but at the cost of estimating many extra parameters. Thus, for example, it has been suggested that lag structures may differ by unit. This is surely plausible, but the consequence of this is that we must estimate separate dynamic parameters for each unit. Given that our time series are fairly short (perhaps 30 or 40 years at most), it is asking a lot of the data to estimate separate dynamic parameters for each unit. Thus, as always, there is a trade-off between parsimony and verisimilitude. The situation is

[11] It is likely that even for slightly smaller T 's (but still in the TSCS, not panel, range) that the costs of the Kiviet estimator outweigh the gains, but the case is not quite so clear cut.

not any different for time series. So of course researchers should think about their dynamic specifications, and start with fairly general ones, but if they start with too general ones, the data will be unable to say much about them.

Time series specifications

To deal with the other issues we must proceed a bit more formally. Everything we say is well known and in any good time series econometrics text, but the lessons sometime seem to be forgotten. Since our discussion is about the time series properties of the data, we can suppress the unit subscript, i, without confusion. All the discussion below holds equally for a single time series or for TSCS data.

Let y_t be an observation at time t for the time series y, where $t = 1, \ldots, T$ and let x_t be a similar observation on an independent variable(s), x. As before, we denote the independent variable as x and talk of it as scalar, with nothing changing if we have a vector of independent variables. We begin with the simplest model,[12] the "static" specification:

$$y_t = \beta_s x_t + \epsilon_t. \tag{3}$$

If we then assume that the errors follow an AR1 process, letting L be the lag operator, with ϕ the AR parameter[13] and, letting ν represent an iid error process (independent of all other processes), we can write the model with AR1 errors as

$$y_t = \beta_{ar1} x_t + \frac{\nu_t}{1 - \phi L} \tag{4a}$$

$$= \beta_{ar1} x_t + \epsilon_t + \phi \epsilon_{t-1} \tag{4b}$$

$$= \beta_{ar1} x_t + \phi y_{t-1} - \beta_{ar1} \phi x_{t-1} + \nu_t \tag{4c}$$

where

$$\epsilon_t = \phi \epsilon_{t-1} + \nu_t. \tag{4d}$$

This is a special case of the "autoregressive distributed lag" (ADL) model,

$$y_t = \beta_{adl} x_t + \phi y_{t-1} + \gamma_{adl} x_{t-1} + \nu_t \tag{5}$$

where Equation 4c imposes the constraint that $\gamma_{adl} = -\beta_{adl}\phi$. This setup is equivalent to the single equation DHSY "error correction model" (Davidson, Hendry, Srba and Yeo, 1978),

$$\Delta y_t = \beta_{adl} \Delta x_t - \phi_{dhsy}(y_{t-1} - \gamma_{dhsy} x_{t-1}) + \nu_t \tag{6}$$

where $\phi_{dhsy} = 1 - \phi_{adl}$ and $\gamma_{dhsy} = \frac{\gamma_{adl} + \beta_{adl}}{1 - \phi_{adl}}$.

[12] Again, we suppress the constant term to simplify notation; this implies that all equilibria are at zero.

[13] The alert reader will note that we subscript the β's in the different specifications to highlight they are conceptually different, but use a generic ϕ in each specification, since conceptually these speed of adjustment parameters are similar across all the specifications. While, of course, the estimates of ϕ will vary with choice of dynamic specification, typically this variation is small, at least for stationary data.

We can also impose the constraint $\gamma = 0$ yielding the LDV model (with iid errors)

$$y_t = \beta_{ldv} \frac{x_t}{1 - \phi L} + \frac{\nu_t}{1 - \phi L} \tag{7a}$$

$$= \beta_{ldv} x_t + \phi y_{t-1} + \nu_t. \tag{7b}$$

We can generalize the LDV model to allow the errors to be non-iid; we denote these errors by ω to remind the reader they are not the same as the ϵ in Equation 3. The LDV model with arbitrary errors is

$$y_t = \beta_{ldv} x_t + \phi y_{t-1} + \omega_t. \tag{8}$$

Why go through all this tedious notation? What do we learn from it? (We make no claim that any of this learning is less than a decade old.)

LDVs vs. AR1 errors

First, we see that if one has a model with AR1 errors, and one wants to deal with those errors, one is going to put a lagged dependent variable in the equation. Thus, even though the coefficient of a lagged dependent variable does not show up on the table of regression coefficients in Stata output, estimation of Equation 4a by some GLS-based method has a lagged dependent variable in the specification. We could, of course, estimate Equation 4c by constrained least squares (as in, say, Eviews), or by maximum likelihood, in which case we would see the lagged dependent variable in the specification. But these are issues about the best way to estimate a model, not the fundamental specification of the model. For that fundamental question, there is literally no difference in whether one uses the AR1 error model or the LDV model *in terms of whether the model has a lagged dependent variable.*

This has enormous implications. Thus, if one is worried about including the lagged dependent variable in a specification (Huber and Stephens, 2001, 59–60), one cannot avoid that problem by assuming that the errors are serially correlated and then choosing a GLS estimation method. As we can see, one can choose amongst the models based on ideas about the speed of adjustment of y to the x's, but all the models typically considered share a lagged dependent variable, either explicitly or implicitly.

The only model we have looked at without at least an implicit lagged dependent variable is the simple Equation 3. Why not estimate this by OLS and correct the standard errors by the Newey-West procedure; this would keep the lagged dependent variable out of the specification? But, if the AR1 error specification (or the LDV specification or the ADL) is correct, this procedure is subject to severe omitted variable bias. And the bias must be severe, else the lagged dependent variable would not be seen as troublesome.

So, while it is not clear (at least to us) what it means for the lagged dependent variable to dominate the regression, it is clear that estimating a model with either AR1 errors or an LDV has the lagged dependent variable in the specification. The LDV cannot be benign in one case and malignant in the second. This is not to say that the specifications in Equations 4c and 7b are in any other way equivalent; they are not. We return to this issue below when we return to our discussion of specification.

Is the use of an LDV "atheoretical?"

Is the lagged dependent variable "atheoretical?" As can be seen by examining Equation 7a, the past state of y is not doing any explaining; it is simply in Equation 7b because of a convenient algebraic manipulation. If the effect of both the measured x and the variables contained in the error term (the error term simply consists of all the determinants of y that we chose not to, or could not, put in the vector of independent variables) impacts geometrically (or a one time pulse change in either of these dies out exponentially), then it simplifies estimation to use the LDV setup, even though there is literally no claim of the lagged dependent variable having any causal status. Obviously those who estimate LDV models must remember not to interpret the ϕ coefficient in Equation 7b causally, that is, not to conclude that a unit change in last year's y causes (whatever that means) a ϕ unit change in current y. But this should not prevent us from using algebraic transformations to simplify estimation.

LDVs and differencing

A related question raised by some (Huber and Stephens, 2001, 59) is whether the LDV (with ϕ close to one) is essentially a model of first differences, and hence not useful if we want to model levels? This issue is nearly one of semantics. Clearly y is changing slowly if ϕ is near one, but this is simply what the data are saying about the process which generate the y_t. So we are still modeling levels.

Note that any method we use looks somewhat like differencing. From an algebraic perspective, we can always subtract y_{t-1} from both sides of a specification to turn a model of levels into one of differences; alternatively, one can add a lagged y to turn a model of differences into one of levels. This is most clearly seen by comparing the ADL model (Equation 5) with its DHSY equivalent (Equation 6). While we think it more convenient to interpret the DHSY model in terms of what affects the first difference of y, obviously this is identical to modeling the level of y.

To beat a dead horse but one more time, note that the GLS procedure for dealing with AR1 errors is pseudo-differences, that is, it estimates a model where each observation is transformed by subtracting off ϕ times the previous observation (in Equation 4b) and then doing OLS on the pseudo-differenced observations. When the estimated ϕ is near one (say .8 or above), as it is in almost all of the Huber and Stephens estimations, then the GLS correction for serially correlated errors is essentially first differencing. The choice between modeling levels or first differences (or nearly first differences) is econometric, not substantive.

Do LDVs cause econometric harm which destroys relationships?

In a recent very influential conference paper, Achen (2000) argued that LDVs "dominate" the regression, that is, they incorrectly mask relationships between the substantive x and y. As Achen clearly states, the results are well known, but the implications have been incorrectly ignored. As clearly shown by Achen the problem arises when both the errors show strong serial correlation and x also shows strong trend.[14] There is no doubt that OLS is inconsistent in the presence of serially correlated errors and an LDV. In our own work, where we have advocated the LDV specification, we have said that researchers should test for serially correlated errors using a Lagrange multiplier test. The advantage of this test is that it only requires estimation under the null, and hence can be done using only OLS residuals. Thus it is quite easy to implement.

Note that if the LDV model has serially correlated errors, it is only the case that OLS is inconsistent. We can still estimate Equation 8 by either maximum likelihood, non-linear least squares, or Cochrane-Orcutt.[15] Thus if we prefer the LDV specification, but have reason to believe that the errors are still serially correlated, all we need do is switch estimation techniques; either of the two mentioned alternatives are available in many canned packages.

Obviously we should test for remaining serial correlation. We should then take the appropriate action if the null of independent errors can be rejected so long as *there is enough remaining serial correlation to make this worthwhile*. Before getting to the trade-off issue, we note that it will commonly be the case that the LDV model will show little if any serial correlation of the errors. Why do we not see a serious problem where Achen sees a potential disaster?

The reason is that the errors are moving targets. Clearly time series analysis of the static model often shows huge serial correlation. But note that the errors in this model are $y_t - \beta x_t$. When we add an LDV, the errors change, which is why we have denoted them with a different Greek letter. Using Equation 4c, we see that the error from assuming the static model is

$$\epsilon_t = \phi y_{t-1} - \beta_{ar1}\phi x_{t-1} + \nu_t. \tag{9}$$

But when we add the LDV to the static specification, the error (from Equation 8 assuming serially correlated errors) is

$$\omega = \beta_{ar1}\phi x_{t-1} + \nu_t. \tag{10}$$

Note the difference between the two error terms: ϕy_{t-1}. With long memoried y, this variable surely accounts for a large portion of the serial correlation of the static model errors. Obviously the remaining $\beta_{ar1}\phi x_{t-1}$ still will induce serial correlation, but the remaining serial correlation in the LDV model must be smaller than in the static model, and most often it is very much smaller.

So of course one must test for serially correlated errors in the LDV model, but they will often not be there or be trivially small. (TSCS data sets are often large, so a small amount of remaining serial correlation of the errors may be statistically significant.) Even with a small, but statistically significant, serial correlation of the errors, we still might prefer OLS to a full maximum likelihood estimator or a GLS estimator. Why would we knowingly make such a mistake?

[14] That is, x is generated by a relatively long memoried autoregressive process.

[15] For Cochrane-Orcutt, we must take care to rule out local minima (Hamilton, 1994, 226). For maximum likelihood, we must use the usual time series "trick" of breaking up the density of the sample into the product of conditional densities.

Why stay with OLS?

As we discussed in the Section 3, the advantage of staying within the OLS framework is that it allows us to do many other things, which are often hard once we leave that framework.[16] Given the costs of abandoning OLS, what are the benefits here? We know from decades of time series experience that one can freely ignore a small amount of serial correlation at almost no cost. Clearly one cannot set a hard and fast limit and say "ignore serial correlation until ϕ exceeds some threshold." But experience does tell us that for small ϕ (say less than 0.1) there is little difference between OLS and GLS.[17] Researchers nervous about ignoring even small amounts of remaining serial correlation can, of course, use either a maximum likelihood or GLS routine to see if the OLS results would change when taking remaining serial correlation into account. The costs of this precaution are minimal, so it is hard to argue against it. By using this check, researchers can reassure themselves that the OLS and GLS results are similar. If this is the case, there are good reasons to eschew the small gains from leaving the OLS framework insofar as such a move imposes non-trivial costs (albeit these costs are difficult to assess in a statistical framework).

Back to specification

So, econometrically, we are not unhappy with the LDV model, given that suitable precautions are taken. But, of course, no specification should ever be used without thought, and either guesses about how the world operates, or theory, should inform the choice of specification. We have already seen that LDV's do not incorrectly "dominate" a regression, and that the AR1 serially correlated error model implicitly has an LDV in the specification. But there are important differences in the specification of the LDV, AR1 and ADL models. For simplicity, we assume that both the LDV and ADL errors are either serially independent or almost so, so we can take those errors as iid with little loss.

To see how the specifications differ, it is easiest to look at the impact of a permanent one unit level change in x. Since the models are linear, we can assume that $x = 0$ until some point, and then move to $x = 1$ thereafter. What are the differences in the behavior of y in the three models?

For the AR1 model, y instantaneously adjusts to the change by increasing by β_{ar1}. For the LDV model, y adjusts to the change in x geometrically; the initial impact of the change is β_{ldv}, with steady-state impact $\frac{\beta_{ldv}}{1-\phi}$, the latter attained geometrically with parameter ϕ.

The ADL model is more complicated; initially y responds to the level shift in x by increasing β_{adl} units, with the long run change in y being $\frac{\beta_{adl}+\gamma_{adl}}{1-\phi}$. This can also be seen in the DHSY restatement of the ADL model, where a one unit change in the long run value of x is associated with a γ_{dhsy} increase in y; as we see from the line following Equation 6, this

[16] Hard because these other things are not programmed in the statistical packages once we move beyond the OLS commands, not because they are inherently hard, but clearly most researchers limit themselves to what they can do with canned statistical packages.

[17] Since GLS proceeds by pseudo-differencing, OLS is just ignoring that we should have subtracted a small number times the previous observation from the current observation. As long as this number is a small number, this should not have an enormous impact.

is the same $\frac{\beta_{adl}+\gamma_{adl}}{1-\phi}$ as in the ADL formulation. Note that if $\gamma_{adl} = -\beta_{adl}\phi$ this reduces to a one time increase in y of β_{adl} units, which is also the long run steady state increase. If γ_{adl} is negative, the γ term moderates the long run effect of a unit change in x; the initial short run effect is still β_{adl}, but the long run effect is less than what we would see for the LDV model.

Of course the ADL model nests the LDV and AR1 error models, and one can start with the more general model and then test the null hypotheses that one of the two specialized models is adequate. This is clearly a reasonable way to proceed. But we must remember that we typically have several independent variables, not one, and that these variables and their lags are often highly collinear, so the ADL model may be asking too much of the data. If that is the case, we will not reject the null that the LDV model is correct. Our procedure of starting with the LDV and then testing for serially correlated errors will lead to the same model being selected, since if there is serious serial correlation of the errors in the LDV model, we will simply add lagged x's to the specification (either explicitly, or implicitly through the AR1 error term).

If one had to choose between the AR1 and LDV models on *a priori* grounds, it is not the presence of the LDV that matters, but rather what one thinks the speed of adjustment to a level change in x looks like. Is it immediate, leading to the AR1 error specification, or does it take place slowly, with the final long term effect being much larger than the immediate effect? Note that the error term just consists of variables we either chose not to or could not measure; these other unmeasured variables have the same geometric impact structure as in the LDV model (as can be seen in Equation 7a) . So if x is similar to the variables that make up the error process, one might expect the LDV model to be suitable. If x represents a change in regime that we expect to have an immediate one-time impact, the AR1 formulation seems plausible. If the data are willing to speak, the ADL model is a good compromise between the two. At least one of us has pushed the DHSY model at great length (Beck, 1991); the DHSY and ADL models are identical. So it would be hypocritical to like the DHSY model for time series but not like the ADL model for TSCS data. We think it will often be the case that we cannot reject the hypothesis that the LDV model is adequate, but we should let the data speak on this. We note that this position is rather close to that taken by Wilson and Butler (2004).

Obviously we normally have a vector of independent variables. How much generality can we allow for? The point can be made with two independent variables, x and z. These can be adjoined to the AR1, LDV and ADL model in the obvious way. In all three models, both variables must show the same relationship between the initial short run impact on y and the long term total impact. For the AR1 model, all the impact of both x and z occurs immediately; for the LDV model, the impact sets in geometrically, but at the same rate for both variables; the error correction representation of the ADL model shows that the speed of adjustment to the new equilibrium y has only one parameter, ϕ. What is the cost of generalizing?

The more general model, with separate speeds of adjustment for both independent variables (and the errors) is

$$y_t = \beta \frac{x_t}{1 - \phi_x L} + \gamma \frac{z_t}{1 - \phi_z L} + \frac{\nu_t}{1 - \phi_e L}. \tag{11}$$

Obviously each new variable now requires us to estimate two additional parameters. Also, on multiplying out the lag structures, we see that with three separate speeds of adjustment we have a third-order lag polynomial multiplying y, which means that we will have the first three lags of y on the right hand side of the specification (and two lags of both x and z) and the iid error (so the errors would be MA2). While there are of course many constraints on the parameters of this model, the need for 3 lags of y costs us 3 years worth of observations (assuming the original data set contained as many observations as were available). With k independent variables we would lose $k + 1$ years of data; for a typical problem where T is perhaps 30 and k is perhaps 5, this is non-trivial. Thus we are unlikely to ever be able to (or want to) estimate a model where each variable has its own speed of adjustment.

But we might get some leverage by allowing for two kinds of independent variables; those where adjustment is instantaneous or nearly so (changes in institutions?) and those whose speed of adjustment is similar to the speed of adjustment of the error process (that is, variables that could freely move from the measured to unmeasured category and vice-versa). Let x (scalar) be the first type of variable and z (scalar) be of the second type. We then would have

$$y_t = \beta x_t + \gamma \frac{z_t}{1 - \phi L} + \frac{\nu_t}{1 - \phi L} \tag{12a}$$

$$= \beta x_t - \phi \beta x_{t-1} + \gamma z_t + \phi y_{t-1} + \nu_t. \tag{12b}$$

As before, this could be estimated by constrained least squares, or we could allow the coefficient of the lagged x to be free and simply estimate the unconstrained version of Equation 12b by OLS (and then test the non-linear constraints implied by Equation 12b). The nice thing about this formulation is that we could make **z** into a vector of variables without losing any observations, at the usual cost of one parameter per added variable. If we have only one (or very few) variables that only have instantaneous impact, the unconstrained version of Equation 12b might be a good compromise between estimability and verisimilitude. Thus we would get some of the gains of the ADL model without paying all of the costs. If many of the variables in the vector **z** are controls which could just as easily be in the error term, this could be a very useful approach that still leaves us in the OLS world. As always, one would have to see how this plays out for any model. But there is no reason that we either have to assume that all variables (including the error) have the same speed of adjustment, or that all measured variables only have an immediate impact. How much flexibility we can allow in practice is clearly a function of the data set being analyzed.

When does all this matter?

If ϕ, in any of the specifications, is relatively small, then shocks quickly die out (or we can say there is a quick return to equilibrium). Suppose, for example, that $\phi = .20$. In a model with AR1 errors, the impact of any of the independent variables is felt only immediately. But in the LDV model, where the long run impact of any given x is $1.33\beta_x$, over 80% of this long run effect is seen immediately, and almost all of the long run effect is seen within one year. In this situation, the LDV and AR1 error specification will appear quite similar. As ϕ gets larger, the difference between the two models get larger. As ϕ gets towards one neither approach is correct, since we move into the world of non-stationary time series. Before looking at that world, we present one example with a small ϕ.

The growth of GDP

This example uses data from Garrett (1998). He uses data on 14 OECD nations observed from 1966–1990, yielding $T = 25$. Our first analysis is of the growth of GDP (*GDP*), which is stationary (with autoregressive representation $GDP_t = .4GDP_{t-1} + 2$, in a model with no fixed effects). Given that at this point we do not want to get into a controversy about fixed effects, and given that standard F tests clearly indicate that the model for GDP wants fixed effects, we include such in the model. These have the effect of making several political economy variables that do not change much over time statistically insignificant. For simplicity, these variables are just eliminated from our discussion.[18]

We thus have a model which assumes that growth in GDP is a linear function of the country dummies (fixed effects), a dummy marking the relatively prosperous period through 1973, lagged growth (*GDPL*), overall OECD GDP growth, weighted for each country by its trade with the other OECD nations, (*DEMAND*), the proportion of cabinet posts occupied by left parties (*LEFT*), the degree of centralized labor bargaining as a measure of corporatism (*CORP*) and the product of the latter two variables (*LEFT*x*CORP*). Obviously the model is sparse on the economic end, with the country dummies, lagged growth and growth in trading partners proxying for whatever economic variables actually determine growth. For our purposes this is not a problem, since we are interested in the effect of the three political variables on growth. The underlying political economic theory being tested is the "social democratic corporatist" model of economic performance, which argues that nations with congruent political and labor bargaining perform best; nations with politically powerful left parties and centralized labor bargaining or nations with politically powerful right parties and decentralized bargaining have congruent institutions. (Both the theory and data are spelled out in Garrett's book.)

To get some feel for the data, the growth in GDP has a mean of about 3.3%, with a standard deviation of 2.4% and a range of -4.3% to 12.8%. The labor organization variable ranges, in principal, from 0 to 5, with a mean of 3 and a standard deviation of 1; the *LEFT* variable ranges from 0 to 3.5, with a mean of 2 and a standard deviation of 1. Both of these variables were standardized by Garrett. Results of an LDV estimation are in the left columns of Table 1, with the AR1 error estimation in the middle columns.[19] The more general ADL

[18] The issue of fixed effects in LDV models is discussed in Section 4, and is not further discussed here. Given that discussion, we do not see a problem in our approach here.

[19] The LDV specification was checked for remaining serial correlation of the errors with a Lagrange multiplier test, which indicates (very strongly) that we cannot reject the null hypothesis that the errors are

model is in the two rightmost columns. The errors reported are PCSE's.

Table 1: Comparison of LDV and AR1 error estimates of Garrett's model of economic growth in 14 OECD nations, 1966–1990 (with fixed effects)

	LDV		AR1 Errors		AR1	
Variable	$\hat{\beta}$	PCSE	$\hat{\beta}$	PCSE	$\hat{\beta}$	PCSE
GDP_{-1}	.16	.07			.15	.08
DEMAND	.72	.16	.70	.18	.70	.17
CORP	−.72	.60	−.78	.70	−.92	1.16
LEFT	−.77	.34	−.88	.38	−.63	.53
LEFTxCORP	.27	.14	.31	.15	.19	.20
PER6673	1.64	.37	1.98	.41	1.65	.37
CONSTANT	2.76	1.77	3.42	2.08	2.45	1.82
$DEMAND_{-1}$					.07	.18
$CORP_{-1}$					.23	1.09
$LEFT_{-1}$					−.23	.53
$LEFTxCORP_{-1}$					.14	.20
ϕ			.15	.08		
N	336		336		336	
BIC	4.3846		4.3945		4.4504	
SSE	1116.131		1127.226		1112.210	

Given the rapid speed of adjustment (the coefficient on the LDV is .15), it is not surprising that the AR1 and LDV estimates are quite similar. Each of these specifications were tested against the full ADL specification; the usual tests show (very decisively) that we cannot reject either the LDV or AR1 error model in favor of the full ADL model. The BIC also indicates that either the LDV or AR1 error model is preferred to the full ADL model, and that the LDV and AR1 models perform very similarly, with a slight nod to the LDV model.

In this case, with rapid speeds of adjustment (or very short memoried processes), it simply does not matter whether one uses the LDV or AR1 error specification. Nor would it be worth it to see if some more complicated model would allow for some variables to adjust more quickly than others. Given our previous arguments that the LDV specification allows us to easily worry about other features of the data, we would stick with the LDV/OLS model. But if one is going to do nothing more than present the results in Table 1, it clearly does not matter which of the two specifications is used. It is also clear that the full ADL specification is too general; because of multicolinearity, this generality is harmful, making

serially independent. If for some reason that test did not allay all fears, we estimated the LDV specification allowing for serially correlated errors (using NLLS in Eviews); the estimated serial correlation is -.01 with a standard error of about .06.

it impossible to conclude anything. Since the various tests indicate that for these data this generality adds little, we can stay with the simpler LDV (or AR1 error) models. Both support the social democratic corporatist position that growth is lower in the presence of either left governments or corporatist labor arrangements *in the absence of the other*, and that left governments and corporatist labor arrangements in combination produces growth.

6. MODELING NON-STATIONARY TSCS DATA

There is much recent work on non-stationary models (Im, Pesaran and Shin, 2003; Levin, Lin and Chu, 2002). That literature deals with testing for unit roots in TSCS and panel data, and whether TSCS or panel unit root test statistics have the same non-standard distributions as their single time series cousins do.

While estimation of TSCS models with unit root data is just beginning to be studied, our experience from single time series analysis tells us that we cannot simply use stationary methods to analyze such data. As an example here, we consider the Huber and Stephens (2001) analysis of the determinants of social security spending in the OECD countries in the post-World War II period (16 OECD countries observed 1960–85). Huber and Stephens (HS for short) fit a model with an AR1 error term and a variety of political variables and economic controls to explain social security spending (*SSBEN*).[20] Obviously one would expect *SSBEN* to be very smooth (that is, to be well predicted by its past). In fact, if we fit a simple autoregression to *SSBEN* (with a constant), we find the point estimate of the autoregressive coefficient is 1.003 with a standard error of .008. Thus we do not have to worry whether statistically we can or cannot reject the null of a unit root; it is pretty clear that *SSBEN* can be characterized as a random walk (with an upward drift of about .3% per year). Thus standard stationary methods (whether of an AR1 error or LDV or ADL variant) are likely to be very misleading for understanding the determinants of *SSBEN*.[21]

[20] HS have a series of analyses on a variety of measures of the size of the welfare state. Since this is just an example, we focus on only a subset of the variables used in their analysis in their Table 3.2, specification 1, which has a dependent variable *SSBEN*, defined to be social security benefit expenditures as a proportion of GDP. We work with a subset of their independent variables, those that were statistically significant in their analysis. While this section is not intended as a critique of HS, we do note that most of what is seen in our one analysis holds for the various HS analyses in their econometric chapter (3). We would also expect that what is said here would hold for other analyses of government spending and similar variables, such as that of Garrett and Mitchell (2001).

[21] As one example of how problematic things get with integrated series, we reran the HS AR1 model using 6 different methods of calculating ρ, the AR1 error coefficient, that are available in Stata. These are all known to be asymptotically equivalent, and analysts usually use whatever method their program defaults to. Here, the five full GLS methods (with different ways of estimating ρ) produce estimates of ρ varying from .80 to .97. Stata does not produce standard errors for these, but the full NLLS procedure in Eviews, which yields the estimate of ρ of .97, shows a standard error .01; clearly the variability of the different estimates of ρ well exceeds this standard error. The impact of these different estimates of ρ is non-trivial. For example, if we want to know how much social welfare spending increases when a Christian democratic government comes to power, we get full GLS estimates of from .24% to .42% (with standard errors under .20). If we lived and died by statistical significance, we would find that three of the arbitrary choices lead to a conclusion that Christian democracy does not have a statistically significant effect on social welfare spending, while the other two analyses lead to the opposite conclusion. If we go further, and drop the first observation

The independent variables used by HS[22] are also integrated (though it is not obvious what this means for a dummy variable). For single time series we would then ask if the variables are co-integrated. Issues of co-integration are just being worked out for TSCS data. However, we can regress *SSBEN* on the independent variables used by HS and test the residuals for stationarity since our interest is whether *SSBEN* equilibrates back to a value specified by the other independent variables, not whether some of the independent variables are in an equilibrium relationship with each other. An autoregression of the residuals from the "co-integrating regression" yield an autoregressive coefficient of about .95 with a standard error of about .018. The various tests for panel unit roots indicate that these residuals are sufficiently likely to have a unit root that we would not conclude that *SSBEN* and the various independent variables are in a long run equilibrium relationship.[23]

In the absence of co-integration, we can only explain short run changes in *SSBEN* by short run changes in the independent variable. Since the number of veto gates does not change over time in almost all countries, we entered that in the regression as a level, asking if the number of veto gates increases or slows the growth of social welfare spending. Results are in Table 2. (The residuals from this regression appear independent.)

The only variables that have a significant impact on welfare spending are economic (the growth of GDP and the unemployment rate) and the level of constitutional veto gates (which slightly slows the growth of the welfare state). Changes in the more conventional party control variables, and other social variables, seem to have no statistically (or substantively) significant impact on the growth of welfare spending. This is rather different from the findings of HS about the impacts of the levels of these variables on the level of social welfare spending (using the improper AR1 error model).

7. CONCLUSION

Our goal has been to see what advice we would give TSCS analysts at the current moment. Some of that advice agrees with that of the other TSCS analysts, some differs. Much is consistent with what we have always said, but some differs. Given our interests, we do not distinguish between these, nor do we attempt in any way to defend what we no longer wish to defend.

---

in each country, as in the traditional Cochrane-Orcutt GLS-like procedure, we would find that a Christian democracy slightly *decreases* social welfare spending, by about $-.005\%$ (with a standard error of .16, so one would not conclude that one could determine the sign of the impact of Christian democracy). The point of all this is that integrated series are very different from stationary ones, and our usual rules of thumb for stationary series do not hold for integrated series.

[22] The variables in our analysis are: left political power, Christian democratic political power, veto gates due to constitutional structure, percent of women in the labor force, the interaction of that variable and left political power, the proportion of the population over age 65, real gross domestic product (in dollars) per capita and the unemployment rate.

[23] This is consistent with estimates of the error correction model. The coefficient on the lag of *SSBEN* is $-.047$ with a standard error of .022. The test statistic for the null that this coefficient is zero (that is, there is no tendency for *SSBEN* to return to equilibrium) is -2.10. Under the null this statistic has a non-standard distribution; had it been generated by a single time series, it would not be close to be being significant.

Table 2: First differences of *SSBEN*

	$\hat{\beta}$	PCSE
Δ*AGED*	2.98	2.35
Δ*GDP*	−0.97	0.24
Δ*FEMALE LABOR*	0.22	0.15
Δ*CHRISTIAN DEMOCRACY*	−0.10	0.21
Δ*LEFT*	0.05	0.14
Δ*UNEM*	0.21	0.07
Δ*FEMALE LABOR*x*LEFT*	0.00	0.01
VETOGATES	−0.05	0.02
CONSTANT	0.43	0.16

General pieties

- There is no mechanical solution to all interesting issues for dealing with TSCS data; analysts must think about how to model their particular data. No one could possibly even imagine disagreeing with this.

- Where we have theory, we should use it in specifying our models. Again, no one can disagree with this. We are perhaps more skeptical than other analysts about the state of theory in comparative political economy.

- While TSCS and panel data may share a common notation, they differ. Researchers should be aware of what statistical fixes are designed for panel data, and whether those fixes are relevant for TSCS data (and whether the problem those fixes remedy are even issues for TSCS data).

- There is no simple technical fix for all the issues of TSCS data. PCSEs improve on OLS standard errors, but they do not fix any other problems, and only fix the OLS standard errors with respect to panel heteroskedasticity and contemporaneous correlation of the errors. The cost of PCSEs is low, both in terms of allowing researchers to do many other things and also in that if in any situation they are inferior to the usual standard errors, they are not very much inferior. Any other issues relating to TSCS are not fixed by PCSEs. So the only solid claim is that OLS with PCSEs is superior to OLS with the usual standard errors

- We have a preference for methods which do not solve one particular problem at the expense of all others. Thus we prefer least squares approaches, which tend to be compatible with many extensions, so long as the least squares approaches are not "too bad." Least squares approaches are often as good or better than more complicated ones, so there is no issue about choosing the simpler approaches. Where least squares is only slightly inferior there is a hard trade-off. While in theory least squares may be much inferior to other estimators, this occurs less often in practice than econometrics books might lead one to believe. This is not to say this will always be the case. There will often be tests to indicate this problem, and researchers should obviously use such tests. But we should be careful not to discard a simpler method because a test indicates a small violation.

- While consistency is a nice property, it is not the be all and end all of criteria. Lots of estimators are consistent but some are better than others. And it is quite possible for an inconsistent estimator to have better small sample mean square error properties than a consistent estimator; we have seen this with the Anderson-Hsiao estimator. While instrumental variable estimators have great asymptotic properties, they may often be inferior to inconsistent estimators in the situations commonly faced by researchers.

Fixed effects/heterogeneity

- If the data clearly indicate that fixed effects are needed, they should be included in the specification, otherwise one runs the risk of severe omitted variable bias.

- This conclusion is only for the continuous dependent variable case, not the binary "event history" case.

- If the test for fixed effects is close, researchers should weigh the costs and benefits of including fixed effects. This is particularly true if rejecting the null hypothesis that no fixed effects are needed does not lead to the conclusion that all unit intercepts are different.

- If fixed effects are used, researchers should be cautious about interpreting non-findings. Knowing that the temporal variation in some variable seems unrelated to the temporal variation in the dependent variable does not mean that there may not be a strong cross-sectional relationship. In particular, institutional variables may have little impact over time (and likely do not even change much over time), but may have a big cross-sectional impact.

- Fixed effects are not the only type of heterogeneity that should be considered. Researchers should always ask whether the various units can be seen to follow the same model with similar parameters. There are real gains from pooling if the units follow similar (not identical) processes, but costs to pooling highly dissimilar units. The fact that some IGO contains a set of countries, or that some data collection organization collected data on a set of countries, is far from a sufficient condition for using a fully pooled analysis.

- Given the gains from pooling, researchers should be careful about moving to a fully heterogeneous model based on a close test result. This is particularly true if rejecting the null hypothesis that all unit coefficients are the same does not lead to the conclusion that all of them differ.

- Both data analysis and testing may show that the relevant alternatives are neither complete pooling nor complete heterogeneity. It may well be the case that one or a few units appear different; at that point one should analyze the remaining units with a pooled model. Of course there is a danger of obtaining a model that appears to fit well by throwing out the units that do not fit well. Thus it would be nice to have some ideas as to why units might or might not be similar.

- While it might appear that the random coefficient (or hierarchical) model might solve the heterogeneity, previous research indicates that it does not. Given the typical time span for TSCS data, the random coefficient model does not appear to do better than choosing between complete pooling and a fully heterogeneous model.

Dynamics

- Consistent with our first bullet point under pieties, obviously one should think about the dynamic model that one wants to use; no one dynamic model, no matter how generally attractive it may or may not be, will always be right.

- There is no evidence that the use of lagged dependent variables (LDVs) is pernicious, other than in conditions where tests indicate they clearly should not be used. Researchers using lagged dependent variables should test (using a Lagrange multiplier test) for remaining serial correlation of the errors. While this will rarely be a problem in practice, researchers should clearly test rather than rely on even good rules of thumb. As always, the question of what do in the presence of close test results requires much judgement.

- For typical comparative TSCS data, it does not appear that OLS with fixed effects and a lagged dependent variable (LSDV) is problematic. It is clearly better than the instrumental variable alternatives proposed. The Kiviet correction to LSDV might be considered if estimating the dynamics is crucial for the application, but using this estimator makes it very difficult to treat other complications of the model, and is likely infeasible for many researchers given the present state of software.

- There are a variety of well known time series specifications that can be used for TSCS data. The same principles that apply to time series apply to TSCS data. But typically the time series component of TSCS data is a lot shorter than for single time series data, and usually TSCS data is observed at a higher (annual) periodicity than is single time series data. Thus it is unlikely that TSCS models will require the complexity often necessary for single time series models.

- We should be clear on how the various dynamic specifications differ. In particular, the LDV and AR1 error model do NOT differ in whether they include the lagged dependent variable in the specification; they do differ on the speed of adjustment for the modeled independent variables. Most of the various specifications differ as to which lags of the independent variables are included, an issue which is well suited to standard statistical adjudication.

- Where the dynamics indicate that adjustments are quick, the differences between the various time series specifications will be minimal. Thus the advantages of the simpler lagged dependent variables may show best here. As the speed of adjustment slows, the various specifications become more different.

- Researchers must take particular care to not use stationary methods for non-stationary data. The results from stationary analyses can be extremely misleading if the data are not stationary. Work on non-stationary TSCS data is in its infancy; researchers should continue to monitor these developments.

REFERENCES

Achen, Christopher. 2000. "Why Lagged Dependent Variables Can Supress the Explanatory Power of Other Independent Variables." Presented at the Annual Meeting of the Society for Political Methodology, UCLA.

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

Arellano, M. and S.R. Bond. 1991. "Some Tests of Specification for Panel Dat: Monte Carlo Ecidence and an Application to Employment Equations." *Review of Economic Studies* 58:277–297.

Bartels, Larry M. 1996. "Pooling Disparate Observations." *American Journal of Political Science* 40:905–42.

Beck, Nathaniel. 1991. "Comparing Dynamic Specifications: The Case of Presidential Approval." *Political Analysis* 3:51–87.

Beck, Nathaniel and Jonathan N. Katz. 1995. "What To Do (and Not To Do) with Time-Series Cross-Section Data." *American Political Science Review* 89:634–47.

Beck, Nathaniel and Jonathan N. Katz. 2001a. "Random Coefficient Models for Time-Series–Cross-Section Data: The 2001 Version." Paper presented at the Annual Meeting of the Society for Political Methodology, Emory University.

Beck, Nathaniel and Jonathan N. Katz. 2001b. "Throwing Out the Baby With the Bath Water: A Comment on Green, Kim and Yoon." *International Organizations* 55:487–95.

Bound, J., D.A. Jaeger and R. M. Baker. 1995. "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endodgenous Explanatory Variable is Weak." *Journal of the American Statistical Association* 90:443–50.

Davidson, J., D. Hendry, F. Srba and S. Yeo. 1978. "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditures and Income in the United Kingdom." *Economic Journal* 88:661–92.

Garrett, G. and D. Mitchell. 2001. "Globalization, Government Spending and Taxation in the OECD." *European Journal of Political Research* 39:144–77.

Garrett, Geoffrey. 1998. *Partisan Politics in the Global Economy.* New York: Cambridge University Press.

Green, Donald, Soo Yeon Kim and David Yoon. 2001. "Dirty Pool." *International Organizations* 55:441–68.

Greene, William. 2003. *Econometric Analysis.* Fifth ed. Upper Saddle River, N.J.: Prentice Hall.

Hamilton, J. 1994. *Time Series Analysis*. Princeton: Princeton University Press.

Hayes, Jude. 2000. Globalization and the Crisis of Embedded Liberalism: The Role of Domestic Political Institutions PhD thesis University of Minnesota. Ph.D. Dissertation, Department of Political Science.

Huber, Evelyne and John D. Stephens. 2001. "Development and Crisis of the Welfare State.".

Im, K.S., M. H. Pesaran and Y. Shin. 2003. "Testing for Unit Roots in Heterogeneous Panels." *Journal of Econometrics* 115:53–74.

Judson, Katherine A. and Anne L. Owen. 1999. "Estimating Dynamic Panel Data Models: A Guide for Macroeconomists." *Economics Letters* 65:9–15.

Kittel, B. 1999. "Sense and Sensitivity in Pooled Analysis of Political Data." *European Journal of Political Research* 38:225–54.

Kittel, B. and H. Winner. 2003. "How Reliable is Pooled Analysis in Political Economy. The Globalization—Welfare State Nexus Revisited." Institut für Gesellschaftsforschung Discussion Paper.

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

Kristensen, Ida P. and Gregory Wawro. 2003. "Lagging the Dog? The Robustness of Panel Corrected Standard Errors in the Presence of Serial Correlation and Observation Specific Effects." Presented at the Annual Meeting of the Society for Political Methodology, University of Minnesota.

Levin, Andrew, C.-F Lin and C-S. J. Chu. 2002. "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties." *Journal of Econometrics* 108:1–24.

Maddala, G.S. 1998. "Recent Developments in Dynamic Econometric Modeling: A Personal Viewpoint." *Political Analysis* 7:59–87.

Nickell, S. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49:1417–26.

Plümper, Thomas and Vera Troger. 2004. "The Estimation of Time-Invariant Variables in Panel Analyses with Unit Fixed Effects." Presented at the 2004 Annual Meeting of the Society for Political Methodology, Stanford, July.

Plümper, Thomas, Vera Troger and Philip Manow. 2004. "Panel Data Analysis in Comparative Politics: Linking Method to Theory." *European Journal of Political Research*.

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, Brigham Young University.

Zorn, Christopher. 2001. "Estimating Between- and Within-Cluster Covariate Effects, with an Application to Models of International Disputes." *International Interactions*.

A. COMPLETE MONTE CARLO RESULTS

This appendix presents the complete results for the Monte Carlo experiments. Simulation parameters: $N = 20$, $\beta = 1$, $\delta = 0.5$, $\sigma_\omega = 0.6$, $\mu = 1$, $\gamma = 0.3$, and $\sigma_\varepsilon = 1$.

Table A.1: Monte Carlo Results for β

		LSDV		Anderson-Hsiao		Kiviet	
T	ϕ	RMSE	Bias	RMSE	Bias	RMSE	Bias
4	0.00	0.287	0.056	0.467	−0.010	0.285	0.044
4	0.20	0.284	0.073	0.601	−0.011	0.281	0.060
4	0.40	0.279	0.082	1.426	−0.008	0.275	0.069
4	0.60	0.269	0.074	3.262	−0.061	0.267	0.062
4	0.80	0.252	0.031	10.827	0.209	0.252	0.018
4	0.90	0.246	−0.018	16.843	−0.107	0.268	−0.034
10	0.00	0.128	0.040	0.172	0.000	0.129	0.041
10	0.20	0.131	0.047	0.170	−0.000	0.131	0.048
10	0.40	0.132	0.051	0.168	−0.001	0.132	0.052
10	0.60	0.129	0.046	0.165	−0.001	0.130	0.048
10	0.80	0.122	0.017	0.163	−0.002	0.122	0.020
10	0.90	0.124	−0.026	0.168	−0.000	0.123	−0.023
20	0.00	0.084	0.016	0.117	−0.002	0.083	0.016
20	0.20	0.085	0.021	0.116	−0.002	0.084	0.020
20	0.40	0.085	0.024	0.114	−0.002	0.085	0.023
20	0.60	0.085	0.026	0.112	−0.002	0.085	0.025
20	0.80	0.083	0.023	0.111	−0.002	0.082	0.022
20	0.90	0.080	0.014	0.113	−0.001	0.079	0.011
30	0.00	0.066	0.009	0.093	−0.007	0.066	0.009
30	0.20	0.066	0.012	0.093	−0.006	0.066	0.012
30	0.40	0.067	0.015	0.092	−0.006	0.067	0.015
30	0.60	0.067	0.017	0.091	−0.006	0.067	0.018
30	0.80	0.066	0.018	0.089	−0.005	0.066	0.018
30	0.90	0.064	0.012	0.089	−0.005	0.064	0.013
40	0.00	0.058	0.006	0.086	−0.005	0.058	0.006
40	0.20	0.058	0.009	0.086	−0.005	0.058	0.009
40	0.40	0.059	0.011	0.085	−0.005	0.059	0.011
40	0.60	0.059	0.013	0.084	−0.005	0.059	0.013
40	0.80	0.058	0.014	0.082	−0.005	0.058	0.015
40	0.90	0.057	0.011	0.082	−0.005	0.057	0.011

Table A.2: Monte Carlo Results for ϕ

		LSDV		Anderson-Hsiao		Kiviet	
T	ϕ	RMSE	Bias	RMSE	Bias	RMSE	Bias
4	0.00	0.303	−0.275	0.765	0.012	0.216	0.162
4	0.20	0.347	−0.321	2.021	0.076	0.194	0.119
4	0.40	0.381	−0.355	2.927	0.013	0.185	0.089
4	0.60	0.403	−0.379	22.710	−0.146	0.185	0.072
4	0.80	0.424	−0.401	25.756	0.902	0.181	0.054
4	0.90	0.426	−0.404	39.220	−1.528	0.180	0.056
10	0.00	0.105	−0.083	0.109	0.008	0.075	−0.027
10	0.20	0.115	−0.095	0.119	0.006	0.074	−0.022
10	0.40	0.121	−0.105	0.128	0.004	0.070	−0.010
10	0.60	0.127	−0.114	0.137	0.002	0.069	0.011
10	0.80	0.141	−0.131	0.154	−0.001	0.076	0.033
10	0.90	0.157	−0.148	0.191	0.004	0.078	0.037
20	0.00	0.059	−0.039	0.073	0.007	0.048	−0.014
20	0.20	0.063	−0.046	0.080	0.007	0.047	−0.012
20	0.40	0.064	−0.051	0.086	0.006	0.043	−0.006
20	0.60	0.064	−0.054	0.093	0.006	0.039	0.006
20	0.80	0.063	−0.056	0.110	0.005	0.045	0.029
20	0.90	0.063	−0.058	0.135	0.006	0.057	0.047
30	0.00	0.044	−0.024	0.055	0.002	0.039	−0.008
30	0.20	0.045	−0.028	0.059	0.003	0.037	−0.006
30	0.40	0.045	−0.030	0.064	0.003	0.034	−0.002
30	0.60	0.043	−0.032	0.068	0.004	0.031	0.007
30	0.80	0.039	−0.033	0.074	0.005	0.033	0.023
30	0.90	0.038	−0.034	0.084	0.006	0.043	0.038
40	0.00	0.035	−0.018	0.048	0.003	0.031	−0.006
40	0.20	0.036	−0.021	0.052	0.003	0.030	−0.005
40	0.40	0.035	−0.024	0.056	0.003	0.027	−0.002
40	0.60	0.034	−0.025	0.059	0.003	0.024	0.004
40	0.80	0.031	−0.026	0.064	0.003	0.025	0.016
40	0.90	0.030	−0.026	0.070	0.002	0.031	0.027

Table A.3: Monte Carlo Results for $\beta/(1-\rho)$

		LSDV		Anderson-Hsiao		Kiviet	
T	ϕ	RMSE	Bias	RMSE	Bias	RMSE	Bias
4	0.00	0.287	0.056	0.467	−0.010	0.285	0.044
4	0.20	0.355	0.091	0.751	−0.013	0.351	0.075
4	0.40	0.465	0.136	2.377	−0.013	0.459	0.115
4	0.60	0.672	0.186	8.156	−0.152	0.668	0.154
4	0.80	1.259	0.157	54.134	1.045	1.261	0.091
4	0.90	2.458	−0.185	168.433	−1.073	2.683	−0.338
10	0.00	0.128	0.040	0.172	0.000	0.129	0.041
10	0.20	0.164	0.059	0.213	−0.000	0.164	0.061
10	0.40	0.220	0.084	0.280	−0.001	0.221	0.087
10	0.60	0.323	0.115	0.412	−0.003	0.325	0.120
10	0.80	0.608	0.084	0.813	−0.009	0.610	0.098
10	0.90	1.240	−0.263	1.684	−0.000	1.235	−0.230
20	0.00	0.084	0.016	0.117	−0.002	0.083	0.016
20	0.20	0.106	0.026	0.145	−0.003	0.106	0.025
20	0.40	0.142	0.040	0.190	−0.003	0.142	0.039
20	0.60	0.212	0.065	0.280	−0.005	0.212	0.062
20	0.80	0.414	0.116	0.553	−0.009	0.412	0.108
20	0.90	0.798	0.136	1.131	−0.012	0.794	0.115
30	0.00	0.066	0.009	0.093	−0.007	0.066	0.009
30	0.20	0.083	0.015	0.116	−0.008	0.083	0.015
30	0.40	0.111	0.025	0.153	−0.010	0.111	0.025
30	0.60	0.166	0.043	0.227	−0.015	0.167	0.044
30	0.80	0.329	0.088	0.447	−0.026	0.330	0.091
30	0.90	0.640	0.123	0.892	−0.047	0.641	0.131
40	0.00	0.058	0.006	0.086	−0.005	0.058	0.006
40	0.20	0.073	0.011	0.107	−0.006	0.073	0.011
40	0.40	0.098	0.018	0.141	−0.008	0.098	0.018
40	0.60	0.147	0.033	0.209	−0.012	0.147	0.033
40	0.80	0.292	0.072	0.411	−0.023	0.292	0.073
40	0.90	0.573	0.113	0.817	−0.047	0.573	0.115