Modeling Dynamics in Time-Series–Cross-Section Political Economy Data*

Nathaniel Beck[†] Jonathan N. Katz[‡]

Draft of May 7, 2009

Abstract

This paper deals with a variety of dynamic issues in the analysis of time-series–cross-section (TSCS) data. While the issues raised are more general, we focus on applications to political economy. We begin with a discussion of specification, and lay out the theoretical differences implied by the various types of time series models that have been estimated. It is 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; with slower speeds of adjustment the various models make sufficiently different predictions that they can be tested against each other. As the speed of adjustment gets slower and slower, specification (and estimation) gets more and more tricky. We then turn to a discussion of estimation. It is noted that models with both a lagged dependent variable and serially correlated errors can easily be estimated; it is only OLS that is inconsistent in this situation. We then show, via 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).

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

*For research support we thank the National Science Foundation. And earlier version was presented at the Annual Meeting of the Society for Political Methodology, Stanford University, Stanford, CA., July 29-31, 2004. We thank Geof Garrett, Evelyne Huber and John Stephens for providing data and many colleagues who have discussed TSCS issues with us and allowed us to present in various forums.

†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

there are a variety of issues related to TSCS data, a number of important ones relate to the dynamics (time series) properties of the data. Obviously many of these issues are similar to those for a single time series, but the context of comparative political economy and the relatively short lengths of the TSCS time periods make for some interesting special issues. We assume that the reader is familiar with the basics of time series; since various specification issues are covered for political scientists elsewhere (Baker, 2008; Beck, 1985, 1991; De Boef and Keele, 2008; Keele and Kelly, 2006) we go fairly quickly over the basic issues, spending more time on the issues on the interpretation of lagged dependent variables.[1] We then deal with estimation issues, concentrating on the issue of estimating models with lagged dependent variables and fixed effects. We begin with some notation. Section 3 is the heart of the paper, dealing with the interpretation of alternative dynamic specifications. This section also contains two empirical analyses. While the section deals largely with stationary data, we also briefly discuss issues related to unit roots (and why the literature on unit roots is less useful for modeling political economy in Section 5. We then turn to the performance of OLS in the presence of fixed effects and a lagged dependent variable in Section 6. We then offer some general conclusions in the final section.

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 $x_{i,t}$ be either an observation on a single independent variable or a vector of such variables; if the latter, it is assumed that the dynamics apply similarly to all the components of that vector. Where we need to differentiate dynamics, we use a second variable (or vector of variables), $z_{i,t}$. Since the constant term is typically irrelevant for what we do, we omit it from our notation.

We distinguish type of error terms; $\nu_{i,t}$ refers to an independent identically distributed ("iid") error process whereas $\varepsilon_{i,t}$ refers to a generic error process which may or may not be iid. In what follows, we superscript coefficients to indicate that they are only interpretable in the context of a specific model. (When we refer to a non-iid process, we only refer to serial correlation, that is we drop the assumption of independence; non-iid and serially correlated are used interchangeably.)

Since our interest is in dynamics, we assume that all cross-sectional issues will be dealt with separately. While cross-sectional and temporal issues are not totally separable (e.g. fixed effects to deal with cross-sectional variation has implications for how one estimates a model with lagged dependent variables), conceptually the issues are separable. In any

[1] At various points we refer to critiques of the use of lagged dependent variables in an unpublished paper of Achen (2000). While it is odd to spend time critiquing an eight year old unpublished paper, this paper has been influential (182 Google Scholar cites as of this writing). We only deal with the portions of Achen's paper relevant to issues raised here.

event, any failure to deal correctly with dynamics can only be more serious when there all also cross-sectional complications. Some ways of dealing with dynamic issues make it more difficult to deal wit cross-sectional issues; thus in evaluating some different approaches, we shall take into account which allow for simpler treatment of important cross-sectional issues.

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.

When relevant, we use L as the lag operator, so that

$$Ly_{i,t} = y_{i,t-1} \text{ if } t > 1 \tag{1}$$

and $Ly_{i,1}$ is missing. The first difference operator is then $\Delta = 1 - L$.

Stationarity

We initially, and for most of the article, assume that the data are drawn from a stationary process. A univariate process is stationary if its various moments and cross-moments do not vary with the time period. In particular, the initial sections assume that the data drawn from a "covariance stationary" process, that is

$$\mathrm{E}(y_{i,t}) = \mu \tag{2a}$$
$$\mathrm{Var}(y_{i,t}) = \sigma^2 \tag{2b}$$
$$\mathrm{E}(y_{i,t}y_{i,t-k}) = \Sigma_k \tag{2c}$$

(and similarly for any other random variables).

Stationary processes have various important features. In particular, they are mean reverting, and the best long-run forecast for a stationary process is that mean. Thus we can think of the mean as the "equilibrium" of a stationary process.

TSCS vs Panel Data

Any TSCS specification could look equally like a specification for "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 so that i averaging over time yields stable results, 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. But our interest in TSCS applications to political economy, where it is usually the case that $T > 10$ and much larger T's are common. As we shall see, issues that arise for small T disappear as T gets to the range usually seen in political economy studies, and various dynamic specification issues that are irrelevant in panel analysis are critical in TSCS analysis.

3 Dynamic Specification: Stationary Data

There are a variety of specifications for any time series model (Beck, 1991; De Boef and Keele, 2008) and all of these specifications have identical counterparts in TSCS models. Since these specifications appear in any standard econometrics text, we discuss these issues without citation, without claiming that the presentation here is of new research. We can either think of testing between these specifications or seeing if theory can tell us which one we should use, though the former more commonly works in practice. In our own prior work (Beck and Katz, 1996) we argued that a specification with a lagged dependent variable (LDV) is often adequate; since that has sparked some discussion (Achen, 2000; Keele and Kelly, 2006), we spend some time on this issue. After discussing specification, we briefly discuss estimation, postponing the issue of estimation with a lagged dependent variable and fixed effects until Section 6.

Dynamic specifications

The generic static equation[2] is

$$y_{i,t} = \beta^s x_{i,t} + \nu_{i,t}. \tag{3}$$

This equation is static because any changes in x or the errors is felt instantaneously,[3] and dissipates immediately; there are no delayed effects.[4]

There are a variety of ways to add dynamics to the static equation, Equation 3. The simplest is the "finite distributed lag model" (FDL) which assumes that the impact of x sets in over two periods, but then dissipates completely. This specification has:

$$y_{i,t} = \beta^{f1} x_{i,t} \beta^{f2} x_{i,t-1} + \nu_{i,t}. \tag{4}$$

[2]Since β is commonly used to denote the "effect" of x on y, but its meaning varies depending on the dynamics, we superscript the β to indicate what specification it refers to. We use different Greek letters to indicate the parameter on the lagged dependent variable.

[3]It may be that $x_{i,t}$ is measured with a lag, so the effect could be felt with a lag, but the model is still inflexible in that the effect is completely and only felt at the one specified year; any of the models we discuss could have $x_{i,t}$ measured a some previous time point. Since the initial lag is specified a priori, this does not make any of the specifications any more general, nor does it make Equation 3 "dynamic."

[4]We restrict ourselves in this subsection to first order dynamic processes, returning briefly to higher order processes later. Similarly, this section restricts itself to processed with either a single x or a vector $\mathbf{x}$ where all variables follow the same dynamic process. We relax this in a subsequent subsection. Neither higher ordered processes nor different dynamic processes require any fundamental rethinking.

with the obvious generalization for higher ordered lags. Obviously Equation 3 is nested inside Equation 4 so testing between the two is simple.

A commonly used specification is to assume that the errors follow a first order autoregressive (AR1) process (rather than the iid process of Equation 3). If we assume that the errors follow an AR1 process, we have

$$y_{i,t} = \beta^{ar1} x_{i,t} + \frac{\nu_{i,t}}{1 - \theta L} \tag{5}$$

$$= \beta^{ar1} x_{i,t} + \theta y_{i,t-1} - \beta^{ar1} \theta x_{i,t-1} + \nu_{i,t}. \tag{6}$$

This is not the common way of writing the AR1 error model seen in elementary texts (which simply rewrites the static model but has the non-iid error process being $\varepsilon_{i,t} = \nu_{i,t} + \theta \varepsilon_{i,t-1}$; the advantage of the formulation in Equation 6 is that it makes clear the dynamics implied by the model, and also makes it easy to compare various models.]

Another commonly used model is the "lagged dependent variable" (LDV) model (with iid errors)

$$y_{i,t} = \beta^{ldv} x_{i,t} + \phi y_{t-1} + \nu_{i,t}. \tag{7a}$$

$$= \beta_l \frac{x_{i,t}}{1 - \phi L} + \frac{\nu_{i,t}}{1 - \phi L}. \tag{7b}$$

As the Equation 7b makes clear, the LDV model simply assumes that the effect of x decays geometrically (and for a vector of independent variables, all decay geometrically at the same rate. Note also that the compound error term is an infinite geometric sum (with the same decay parameter as for x); this error term is equivalent to a first order moving average (MA1) error process.

Both the AR1 and LDV specifications are special cases of the "autoregressive distributed lag" (ADL) model,

$$y_{i,t} = \beta^{adl} x_{i,t} + \mu y_{i,t-1} + \gamma x_{i,t-1} + \nu_{i,t} \tag{8}$$

where Equation 6 imposes the constraint that $\gamma = -\beta^{adl} \mu$ and Equation 7a assumes that $\gamma = 0$. The nesting of both the LDV and AR1 specifications within the ADL specification allows for testing between the various models.[5]

Interpretation

To see how the various specifications differ, we turn to unit and impulse response functions. Since x itself is stochastic, assume the process has run long enough for it to be at

[5] While it is not particularly relevant for our purposes, the ADL model can be rewritten as the single equation "error correction model" (Davidson, Hendry, Srba and Yeo, 1978),

$$\Delta y_{i,t} = \beta^{ec} \Delta x_{i,t} - \theta (y_{i,t-1} - \ gamma x_{i,t-1}) + \nu_{i,t} \tag{9}$$

which allows for the nice interpretation that short tun changes in y are a function of both short run changes in x and how much x and y are out of equilibrium, with y changing to move back towards equilibrium at at rate θ. Since the error correction model is just an algebraic rewriting of the ADL model, and given what we will say about non-stationary processes in political economy, we do not pursue this interpretation other than noting it can be most useful.

its equilibrium value (stationary implies the existence of such an equilibrium). We can then think of a one time shock in x (or ε) of one unit, with a subsequent return to equilibrium (or zero for the error) the next year; if we then plot y against this, we get an impulse response function (IRF). Alternatively, we can shock x by one unit and let it stay at the new value; the plot of y against x is a unit response functions (URF).[6]

The static specification assumes that all variables have an instantaneous and only instantaneous impact. Thus the IRF for either x or ν is a spike, associated with an instantaneous change in y, and if x or nu then returns to previous values in the next period, y immediately also returns to instantaneous values. The URF is simply a step function, with the height of the single step being β^s

The finite distributed lag model simply generalizes this, with the URF having two steps, of height β^{f1} and $\beta^{f1} + \beta^{f2}$, and the interval between the steps being one year. Thus, unlike the simple static model, if x changes, it takes two years for the full effect of the change to be felt, but the effect is fully felt in those two years. Thus it may take one year for a party to have an impact on unemployment, but it may be the case that after that year the new party in office has done all it can and will do in terms of changing unemployment. Similarly, an institutional change may not have all of its impact immediately, but the full impact may occur within the space of a year.

Note that the FDL has fallen out of favor with most time series analysts. This is because they typically work with higher frequency (quarterly or monthly or higher) data than typically used in TSCS studies of political economy. Given that economic time series are often highly correlated, it becomes very hard for the analyst to tease out the lag structure with an FDL specification.[7] However, the nature of TSCS data, it may well be that a model with one or two lags might both not overtax the data and also better fit theoretical notions. It is unlikely that interesting institutional changes have only an immediate impact, but the FDL model *might* be appropriate. It surely should be borne in mind in thinking about appropriate specifications.

The AR1 model has a different IRF for x and the error. The IRF for x is a spike, identical to that of the static model; the IRF function for the error is that of a declining geometric, with rate of decay θ. It seems odd that all the omitted variables have a declining geometric l IRF but the x we are modeling has only the immediate impact. Maybe that is correct, but it seems odd at first glance. Obviously it can be made less odd by moving to the FDL specification by adding more lags of x, but we would still have very different dynamics for x and the unobserved variables in the "error" term. One should clearly have a reason to believe that dynamics are of this form before using the AR1 specification.

The LDV model has an IRF for both x and the error has a declining geometric form; the initial response is β_l (or 1 for the error); this declines to zero geometrically at a rate ϕ. While the effect never completely dissipates, it becomes tiny fairly quickly unless ϕ is almost one. The URF starts with an effect β_{ldv} immediately, increasing to $\frac{\beta_{ldv}}{1-\phi}$. If ϕ is close to one,

[6]Obviously one function determines the other; it also makes no sense to think about a unit response function for the error process.

[7]This led to the work on polynomial distributed lags and such, but these too have fallen out of favor amongst standard time series econometricians.

the long run impact of x can be 10 or more times the immediate impact.

While the ADL specification appears to be much more flexible, it actually has IRF functions similar to the LDV specification, other than in the first year (and is identical for a shock to the error process). Initially y changes by β^{ar1} units, then the next period the change is $\beta^{ar1}\mu + \gamma$ which then dies out geometrically at a rate μ. Thus the ADL specification is only a bit more general than the LDV specification. It does allow for the maximal impact of x to occur a year later, rather than instantaneously (or, more generally, the effect of x after one period is not constrained to be the immediate impact with one year's decay). This may be important in some applications. A comparison of the various IRFs and URFs is in Figure 1.

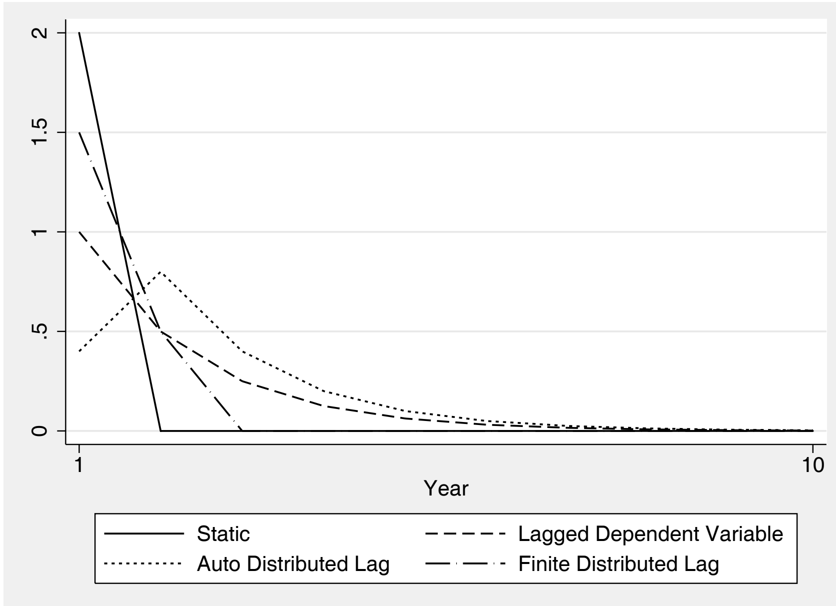
Before getting to slightly more complicated models, this analysis tells us several things. The various models differ in the assumptions they impose on the dynamics that govern how x and the errors impact y. None can be more or less right *a priori*. Later we will discuss some econometric issue, but for now we can say that various theories would suggest various specifications. The most important issue is whether we think a change in some variable is felt only immediately, or whether its impact is distributed over time. How would we expect an institutional change to affect some y of interest in terms of the timing of that effect. If only immediately or completely in one or two years, the AR1 model or the FDL model seems right; if we expect the maximal effect to be immediate, but then to continue with some monotonic decline, the LDV or ADL model seems more reason. But there is nothing "atheoretical" about the use of a lagged dependent variable, and there is nothing which should lead anyone to think that the use of a lagged dependent variable causes incorrect "harm." It may cause "correct" harm, in that it may keep us from incorrectly concluding that x has a big effect when it does not, but that cannot be a bad thing.

Similarly, the LDV or its generalization is not bad because, as claimed by Huber and Stephens (2001, 59), that these models are essentially models first differences (when ϕ or μ is close to one) , and hence not useful if we want to model levels. When these dynamic parameters are close to one, the interpretation is that the impact of a change in x sets in slowly, and the long run impact may be 10 or more times the initial impact. This may or may not be the right model, but there is nothing that *a priori* tells us that such a specification is inferior.[8] In short, one cannot prefer, *a priori*, a model which explicitly contains the lagged dependent variable over a model which does not explicitly contain such a variable.

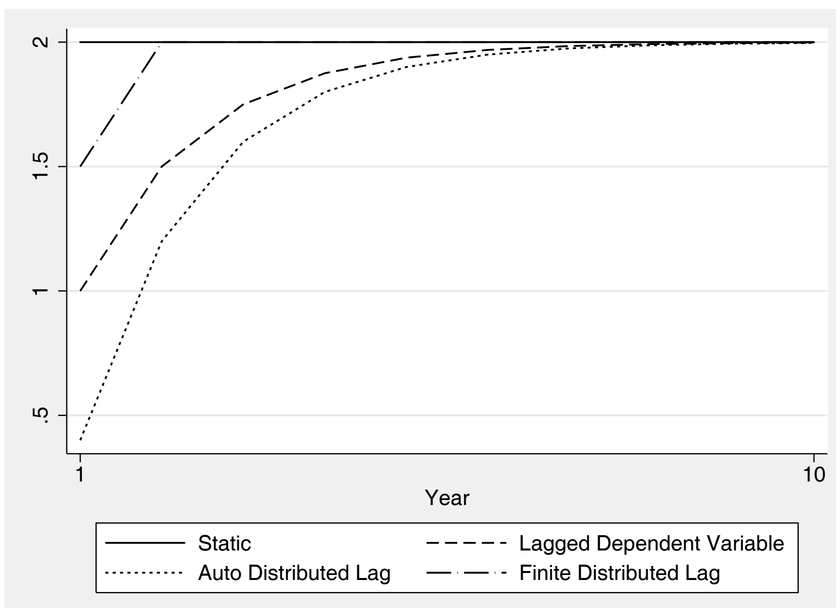
Higher order processes and other complications

We can generalize any of these models to allow for non-iid errors and we can allow for higher ordered serial correlation. Since our applications typically use annual data, it is often the case that first order error processes suffice, and it would be unusual to have more than second order processes; since, as we shall see, it is easy to test for higher order error processes,

[8]We can, of course, 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 clearly seen by the equivalence of Equations 8 and 9. While it may be easier to interpret the latter, the two forms are mathematically equivalent, so one cannot be better than the other.



(a) Impulse Response Function



(b) Unit Response Functions

Figure 1: Comparison of impulse and unit response functions for four specifications

there is no reason to simply assume that errors are iid or only follow a first order process. For notational simplicity we restrict ourselves to second order processes, but the generalization is obvious.

Consider the LDV model with AR1 errors, so that

$$y_{i,t} = \beta_{ldv} x_{i,t} + \phi y_{i,t-1} + \frac{\nu_{i,t}}{1 - \omega L}.. \tag{10}$$

After multiplying through by $(1 - \omega)$ we get a model with two lags of y, x and lagged x and some constraints on the parameters; if we generalize the ADL model similarly, we get a model with two lags of both y and x and more constraints. The interpretation of this model is similar to the model with iid errors.

If we assume that the "errors," which are usually omitted or unmeasured variables follow an MA1 process with the same dynamic parameter,
phi (which may or may not be reasonable), we then have

$$y_{i,t} = \beta^{ldv} x_{i,t} + \phi y_{t-1} + (1 - \phi L)\nu_{i,t} \tag{11a}$$

$$= \beta_{ldv} \frac{x_{i,t}}{1 - \phi L} + \nu_{i,t}. \tag{11b}$$

This is a model with a geometrically declining impact of x on y and iid errors. It is surely more likely that the "errors" (omitted variables) are correlated than that they are independent. Of course the most likely case is that the errors are neither iid nor MA1 with the same dynamics as x, so we should entertain a more general model with the errors following an unconstrained MA1 process. We return to this when we discuss econometrics.

More complicated dynamics - multiple independent variables

We typically have more than one independent variable. How much generality can we allow for? The models above easily generalize to a vector of independent variables, as long as the dynamics for each variable are identical. The generalization of this can be easily seen if we have two independent variables, x and z which are adjoined to the various models. Thus for the AR1 model, both x and z have only immediate impact, for the LDV model they have different immediate impact but the same rate of geometric decline, etc.

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

$$y_{i,t} = \beta \frac{x_{i,t}}{1 - \phi_x L} + \gamma \frac{z_t}{1 - \phi_z L} + \frac{\nu_t}{1 - \phi_\nu L}. \tag{12}$$

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 original 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 at least relatively fast (x) and those whose speed of adjustment is slower (z).[9] Since we are trying to simplify here, assume the error process shows the same slower adjustment speed as z; we can obviously build more complex models but they bring nothing additional to this discussion. We then would have

$$y_{i,t} = \beta_x x_{i,t} + \beta_z \frac{z_t}{1-\phi L} + \frac{\nu_t}{1-\phi L} \tag{13a}$$

$$= \beta_x x_{i,t} - \phi \beta_x x_{i,t-1} + \beta_z z_t + \phi y_{t-1} + \nu_t. \tag{13b}$$

Thus at the cost of one extra parameter, we can allow some variables to have only an immediate or very quick effect, while others have a longer effect, setting in geometrically. With enough years we could estimate more complex models, allowing for multiple dynamic processes, but such an opportunity is unlikely to present itself in studies of comparative political economy. We could also generalize the model by allowing for the lags of x and z to enter without constraint. It is possible to test for whether these various complications are supported by the data, or whether they simply ask too much of the data. As always, it is easy enough to ask the data and then make a decision (unless theory is so compelling that we know that impacts are immediate or set in very slowly, theory which we seldom if ever see.)

Estimation Issues

As is well known a specification with no lagged dependent variable but serially correlated errors is easy to estimate using any of several variants of feasible generalized least squares (FGLS); the Cochrane-Orcutt iterated procedure is probably the most commonly used such variant. It is also easy to estimate such a model via maximum likelihood, breaking up the full likelihood into a product of conditional likelihoods.

The LDV model with iid errors is optimally estimated by OLS. It is also well-known that OLS yields inconsistent estimates of the LDV model with serially correlated errors. Perhaps less well-known is that Cochrane-Orcutt or maximum likelihood provides consistent estimates of the LDV model with serially correlated errors non-iid errors (Hamilton, 1994, 226).[10] Thus the use of a lagged dependent variable with serially correlated errors only requires care in using a correct estimation method; it causes no other econometric problems.

It is often the case that the inclusion of a lagged dependent variable eliminates almost all serial correlation of the errors. To see this, start with the AR1 equation:

[9] We will use the words "fast" and "slow" to indicate how long it takes for a system that is perturbed either to return to equilibrium or reach a new equilibrium; for stationary models there is always some equilibrium (by definition). Thus, if x changes by one unit for one year only (an impulse change), how long will it take for y to return to its previous equilibrium value; for a unit change that persists, fast or slow refers to how quickly y approaches its new equilibrium.

[10] The Cochrane-Orcutt procedure may find a local minimum, so analysts should try various starting values. This is seldom an issue in practice, but it is clearly easy enough to try alternative starting values.

$$y_{i,t} = \beta^{adl} x_{i,t} + \varepsilon_{i,t} \tag{14}$$
$$\varepsilon_{i,t} = \nu_{i,t} + +\theta\varepsilon_{i,t-1}. \tag{15}$$

Remember, as is common, the error term is simply the error of the observer, that is, everything that determines $y_{i,t}$ that is not explained by $x_{i,t}$. If we adjoin $y_{i,t-1}$ to the specification, the error in that new specification is $\varepsilon_{i,t} - y_{i,t-1}$ where $\varepsilon_{i,t}$ is the original error in Equation 14, not some generic error term. Since the $\varepsilon_{i,t}$ are serially correlated because they contain a common omitted variable, and $y_{i,t-1}$ contains the omitted variables at time $t-1$, including $y_{i,t-1}$ will almost certainly lower the degree of serial correlation, and often will all but eliminate.

But there is no reason to simply hope this happens. We can simply estimate the LDV model assuming iid errors (so by OLS), and then test the null that the errors are iid by a Lagrange multiplier test (which only requires that OLS be consistent under the null of iid errors, which it is).[11]. If, as often happens, we do not reject the null that the remaining errors are iid, we can continue with the OLS estimates; if we reject the null of iid errors we can simply use Cochrane-Orcutt or maximum likelihood. Thus lagged dependent variables present no estimation issues. The only minor problem left is what happens if we have a lagged dependent variable and fixed effects. We turn to that issue in the next section.[12]

Discriminating between models

We can use the fact that the ADL model nests the LDV and AR1 models to allow the data to speak on which specification better fits the data. The LDV model assumes that $\gamma = 0$ in Equation 8 whereas the AR1 model assumes that $\gamma = -\mu\beta$. Thus we can estimate the full ADL model and test whether $\gamma = 0$, which would tell us that the data do not reject the LDV model or $\gamma = -\rho\beta$ which indicate the same thing for the AR1 model. If the data reject both simplifications, we can simply retain the more complicated ADL model.[13] Even in the absence of a precise test, the ADL estimates will often indicate which simplification is clearly not too costly to impose.

[11] The test is trivial to implement. Take the residuals from the OLS regression and regress them on the appropriate number of lags of those residuals and all the independent variables including the lagged dependent variable; the relevant test statistics if NTR^2 from this auxiliary regression, which is distributed χ^2 with degrees of freedom equal to the number of lags

[12] Why are we so sanguine about LDV's when Achen is so worried about them. First, since the Lagrange multiplier test allows us to assess the degree of remaining serial correlation in the LDV model, we do not simply have to worry, we can test. But, more importantly, Achen makes the mistake of thinking that the serial correlation in the errors is a fixed characteristic of the model (as we saw in the discussion near Equation 14

[13] Starting with the ADL model and then testing whether simplifications are consistent with the data is part of the idea of general to simple testing (the encompassing approach) as espoused by Hendry and his colleagues (Hendry and Mizon, 1978; Mizon, 1984). Note that this approach would start with a more complicated model with higher order specifications, but given annual data, the ADL model is often the most complicated one that need be considered. Of course if analysts believe that the simplifications of the ADL model are not consistent with the data, they are free to start with more complicated models. Nothing in our discussion precludes that, and there are no real interpretative issues raised by such a strategy.

Note that for fast dynamics (where μ is close to zero), it will be hard to distinguish between the LDV and AR1 specifications, or, alternatively, it does not make much difference which one specification we use. To see this, note that if the AR1 model is correct, but we estimate the LDV model, we are incorrectly omitting the lagged x variable, when it should be in the specification, but with the constrained coefficient, $mu\beta$. As μ goes to zero, the bias from failing to include this term goes to zero. Similarly, if we incorrectly include estimate the AR1 model when the LDV model is correct, we have incorrectly included in the specification the lagged dependent variable, with coefficient $-\mu\beta$. Again, as μ goes to zero, this bias goes to zero. Thus we might find ourselves not rejecting *either* the LDV or AR1 specifications in favor of the more general specification, but for small μ it matters little, since the impulse response models of the AR1 and LDV models, for small μ are so similar. As μ grows larger the two models diverge, and so we have a better chance of having the data speak as to which is preferred. [14]

Interestingly, this is different from the usual logic on omitted variable bias, where it is normally thought to be worse to incorrectly exclude than to incorrectly include a variable. This difference is because both models constrain the coefficient of the lagged x, and so the AR1 model "forces" the lagged x to be in the specification. But if we start with the ADL model and then test for whether simplifications are consistent with the data, we will not be misled. This testing of simplifications is easy to extent to more complicated models, such as Equation 13b, where we can test whether some variables have only the instantaneous impact implied by the AR1 logic, while others have the declining exponential impact implied by the LDV logic.

4 Examples

The growth of GDP

Our first example involves data with relatively little serial correlation. Here we use political economy explanations of the growth of GDP in 14 OECD nations observed from 1966–1990, yielding $T = 25$.[15] We note that GDP growth appears stationary, with an autoregressive coefficient of .32. Thus we expect all dynamics to be relatively fast, that is, the system will fully adjust quite quickly.

We use a model directly from Garrett, which assumes that growth in GDP is a linear additive function of political factors and economic controls. The political variables are 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 (*LEFTxCORP*); the economic and control variables are a dummy marking the relatively prosperous period through 1973 (*PER73*), overall OECD GDP growth, weighted for each

[14] If, for example, $\mu = .3$, then the LDV model has the coefficient on lagged x being 0 while the AR1 model has it as $-.3\beta$. After two periods the difference is now only $.09\beta$, and so forth, so the two specifications are substantively similar. Only as μ gets closer to one does it matter much whether low powers of μ can be assumed to be zero.

[15] The data are from Garrett (1998) and interested readers should consult his book.

country by its trade with the other OECD nations, (*DEMAND*), trade openness (*TRADE*), capital mobility (*CAPMOB*) and a measure of oil imports (*OILD*).[16] Some specifications contain lagged growth (*GDPL*). Results of estimating various specifications are in Table 1.[17]

Garrett included fixed effects in his model. Given our discussion above, we simply estimated using OLS (LSDV), not bothering to use the more complicated Kiviet estimator. For simplicity, we just mean centered each variable (centering around the unit mean); this is identical to OLS/LSDV. Given mean centering, there is no constant term in the model, and we do not report the fixed effects estimated.

We began by estimating the model by OLS (without a lagged dependent variable). Since a Lagrange multiplier test showed we could clearly reject the null of serially independent errors ($\chi^2_1 = 8.6, p < .001$); substantively, the serial correlation of the errors was small,.10. We therefore only show the results of estimating the specification with AR1 correlated errors and then a lagged dependent variable as well as the apparently incorrect static specification.

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

	OLS		AR1 Errors		LDV	
Variable	$\hat{\beta}$	SE	$\hat{\beta}$	SE	$\hat{\beta}$	SE
DEMAND	.007	.001	.007	.001	.007	.001
TRADE	−.02	.02	−.02	.02	−.02	.02
CAPMOB	−.19	.22	−.26	.23	−.24	.22
OILD	7.86	6.26	−6.69	6.66	−5.85	6.21
PER73	1.75	.32	1.76	.34	1.45	.33
CORP	.45	.57	.43	.60	.30	.57
LEFT	−.08	.18	−.08	.19	−.08	.18
LEFTxCORP	.10	.65	.10	.68	.17	.64
GDP_{-L}					.16	.05
ϕ			.10			
N	$cols2c350$		336		336	

Given the rapid speed of adjustment (the coefficient on the LDV is .15), it is not surprising that all three specification show similar estimates. Very little is significant in any of the specifications, but the two variables that show a strong impact in the static specification continue to show a strong impact in the two dynamic specifications. Clearly the simple

[16]Remember that all variables were country centered, so all specifications contained fixed effects. The variables are defined in Garrett (1998).

[17]For simplicity we report OLS standard errors. These are within a few percent of the panel corrected standard errors.

static OLS is a bit misspecified; we can reject the null hypothesis of no serial correlation at $p < .01$ even though substantively the amount of serial correlation is not great.

The similarity of the AR1 and LDV estimates is not surprising; because of the fast dynamics the two models are not really very different. After one period the various independent variables in the LDV specification have only 3% of their original impact; the long-run effects in the LDV specification are only 18% larger than the immediate impacts. Thus the two specifications are saying more or less the same things, and the estimated coefficients are quite similar. Substantively, it appears as though GDP growth in a country is largely determined by GDP growth in its trading partners, and politics appears to play little if any role.

Both specifications were tested against the full ADL specification which contained all the one year lags of the independent variables. Standard hypothesis tests do not come near to allowing rejection of the simpler AR1 or LDV models in favor of the ADL model; 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 F-statistic for all eight coefficients was about .3. In short, the data are consistent with very short run impacts, and it does not particularly matter how we exactly specify the dynamics.

Finally, in terms of the critique of the use of lagged dependent variables (Achen, 2000), there are two predictors of GDP that are strong in the AR1 model; they remain about equally strong in the LDV model. As the previous section showed, there is nothing about LDVs which "dominate a regression" or which make "real" effects disappear. Given the nature of dynamics, this will always be the case when variables adjust quickly (that is, the serial correlation of the errors or the coefficient on the lagged dependent variable is small). We now turn to a second example where variables adjust much more slowly.

Capital Taxation Rates

Our second example models the capital taxation rates in 17 OECD nations from 1961–93, using the data and specification as in Garrett and Mitchell (2001).[18] Obviously tax rates move relatively slowly over time; the autoregressive coefficient of tax rates is .77. Thus, while tax rates are clearly stationary, it will take some number of years for the system to fully adjust.

We drop a few variables from the Garrett and Mitchell specification (that were insignificant in all specifications and not particularly substantively interesting). We thus regress the capital tax rate (*CAPTAX*) on unemployment (*UNEM*), economic growth (*GDPPC*), the dependency ratio, that is the proportion of the population that is elderly (*AGED*), vulnerability of the workforce as measured by low wage imports (*LOWWAGE*, foreign direct investment (*FDI*), and two political variables, the proportion of the cabinet portfolios held by the left (*LEFT*) and the proportion held by Christian Democrats (*CDEM*). Since Garrett and Mitchell used both fixed country and year effects in all specifications, as in the previous example, we centered all variables by country (which again leads to a model with no constant term). Following Garrett and Mitchell, we also mean centered by year (that is, used year as

[18] The data set is not rectangular; some countries only report tax rates for a portion of the period under study.

well as country fixed effects). As in the previous example, we first show the AR1 and LDV results in Table 2, noting that an OLS regression will clearly show serial dependence.[19]

Table 2: Comparison of AR1, LDV and ADL estimates of Garrett and Mitchell's model of capital taxation in 17 OECD nations, 1967–1992 (country and year centered)

	OLS		AR1 Errors		LDV		ADL	
Variable	$\hat{\beta}$	SE	$\hat{\beta}$	SE	$\hat{\beta}$	SE	$\hat{\beta}$	SE
LOWWAGE	−.16	.07	−.21	.09	−.09	.05	−.28	.12
FDI	.44	.26	.52	.22	.34	.19	.59	.23
UNEM	.20	.17	−.21	.20	−.35	.13	−.68	.30
AGED	1.66	.29	1.34	.40	.34	.22	.26	.74
GDPPC	−.89	.14	−.68	.09	−.59	.10	−.80	.12
LEFT	.004	.009	.004	.009	.006	.006	.003	.01
CDEM	.023	.023	.017	.028	.014	.017	.015	.03
TAX_L					.70	.04	.76	.04
$LOWWAGE_L$							.21	.11
FDI_L							−.55	.25
$UNEM_L$							.48	.30
$AGED_L$							.24	.75
$GDPPC_L$							.29	.11
$LEFT_L$							.005	.01
$CDEM_L$							.005	.03
ϕ			.66					
N	338		338		330		322	

The simple static model is clearly wrong; either the LDV or AR1 models are strongly preferred to this static model. This is the type of model that worries Achen; the AR1 model shows a strong affect of the proportion of the population that is aged on capital taxation; the LDV model cuts the impact of this variable by a factor of three with a t-ratio barely exceeding one. Note that other variables that are important in the AR1 estimation (foreign direct investment, the growth of GDP) remain important in the LDV specification; the magnitudes of the coefficients and standard errors for these two variables is similar for both specifications. The coefficient on the proportion of low waged workers is also cut in half, though it is statistically significant in both specifications.

Interestingly, unemployment has both a larger effect, and one that is statistically significant, in the LDV specification, but not the AR1 specification. (Neither political variable has

[19] We omit checks of the various specifications. Each specification shows a small but statistically significant amount of remaining serial correlation of the errors. Estimates correcting for this are almost identical to those shown in the table.

a strong or significant impact in either specification). But whatever is going on, the LDV simply does not destroy all interesting relations between variables.

We compare both of these specifications to the full ADL model, we note that the lagged coefficients on both low waged workers, unemployment and foreign direct investment is of the opposite sign as the contemporaneous coefficient (and either statistically significant or close). Based on our prior discussion, this shows that the impact of these three variables is more or less instantaneous, with said impact almost disappearing in a year. While the lagged coefficient on the growth of GDP is significant and of the opposite sign as the contemporaneous coefficient, it is much smaller than that contemporaneous coefficient, indicating that the impact of the growth of GDP dissipates more or less exponentially, though a bit more quickly than exponentially in the first year. Neither of the two political variables shows any effect in any of the specifications.

Thus the proportion of the population that is aged is the only variable that seems like it "ought" to determine tax rates, that appears to strongly determine tax rates but either the LDV or ADL specifications fail to find a significant effect of this variable. It may be noted that while *AGED* perhaps "ought" to effect tax rates, its coefficient in the AR1 specification "seems" a bit large; would a one point increase in the aged population be expected to lead to over a one point increase in capital taxation rates? Thus perhaps it is not so simple to discuss which results make "sense."

Note that *AGED* is itself highly trending (its autoregression has a coefficient of .93 with a standard error of .01). While we can reject the null that *AGED* has a unit root, it, like the capital tax rate, changes very slowly. Thus we might suspect that the simple contemporaneous relationship between the two variables is spurious (in the sense of Granger and Newbold (1974). Of course we cannot know the "truth" here, but it is not obvious that the ADL (or LDV) results on the impact of *AGED* are somehow foolish. Note that the ADL specification seems "sensible" for all the other variables. Here is one garden variety model where the use of a lagged dependent variable, in what appears to be a correct specification, is consistent with perfectly reasonable results, and subject to a perfectly clear interpretation.

5 No-stationarity in political economy TSCS data

"For the case of first-order autoregression, it is known from such calculations the [estimating of the autoregression coefficient] is downward-biased in small samples, with the bias becoming more severe as [the autoregressive coefficient] approaches unity." (Hamilton, 1994, 217)

During the last two decades, with the pioneering work of Engle and Granger (1987), time series econometrics has been dominated by the study of non-stationary series. While there are many ways to violate the assumptions of stationarity (Equation 2. the new work has all dealt with the issue of unit roots, that is, how do we estimate models where the data are integrated (we restrict ourselves to integration of order one with no loss of generality). Such data, denoted I(1), are not stationary but their first difference is stationary. The simplest example of such a process is a random walk, where

$$y_{i,t} = y_{i,t-1} + \nu_{i,t}. \tag{16}$$

. Such data look very different from data generated by a stationary process. Of most importance for us is that they do not have simple equilibrium, in that the expected time for recrossing the mean is unbounded, so there is no tendency to return to the mean, even in the long run, and the best prediction of a series many periods ahead is the current value of that series.

There is a huge literature on estimating models with integrated data. Such methods must take into account that standard asymptotic theory does not apply, and also that

$$\lim_{t \to \infty} \text{VAR}(yit) = \infty. \tag{17}$$

Thus if we wait long enough, any integrated series will wander "infinitely" far from its mean. Much work on both diagnosing and estimating models with inter grated series builds heavily on both these issues. Our interest is not in the estimation of single time series, but rather TSCS political economy data.[20]

Political economy data is typically observed annually for relatively short periods of time (often 20-40 years). Of most relevance, during that time period, we often observe very few cycles. Thus, while the series are very persistent, we have no idea if a longer time period wold show the series to be stationary (though with a slow speed of adjustment) or non-stationary. But annual observations on GDP or left political control of the economy are very different than the daily observations we may have on various financial rates. So while it may appear from an autoregression that political economy series have unit roots, is this the right characterization of these series. Thus, for example, using Huber and Stephens' (2001) data, an autoregression of social security on its lag yileds a point estimate of the autoregressive coefficient of 1.003 with a standard error of .009; a similar autoregression of Christian Democratic party cabinet participation of 1.03 with a standard error of .001. It does not take heavy duty testing to see we cannot reject the null that the autoregressive coefficient is one in favor of alternative that it is less than one (so the series is stationary). But does this mean that we think the series might be I(1).?

Making an argument similar to that of Alvcarez and Katz (2000), if these series had unit roots, there would be tendency for them to return to their means and the variance of the observations would grow larger and larger over time. But by definition social security spending and the Christian Democratic cabinet participation are between zero and one hundred per cent). If either series were I(1), then we would be equally likely to see an increase or decrease in either variable regardless of its present value; do we really believe that there is no tendency for social security spending to be more likely to rise when it is low and to fall when high, or similarly for Christian Democratic cabinet strength? Note that in the data set, social security spending only ranges between 3% and 33% with similar numbers for Chrsitian Democratic cabinet strength being zero and 34%, So these series are very persistent, but

[20]There is a literature on panel unit roots (Im, Pesaran and Shin, 2003; Levin, Lin and Chu, 2002), but at this point the literature is still largely about testing for unit roots.

they simply cannot be I(1). So the impressive apparatus built over the last two decades to estimate models with I(1) series does not provide the tools needed for many if not most political science TSCS datasets.

Fortunately, the issue is not really the univariate properties of any series, but the properties of the stochastic process which generate the y's conditional on the covariates, which is the same as the error process. Even with data similar to Huber and Stephens', the errors may appear stationary and so the methods of the previous section can be used. At that point all the lessons of that section apply.

6 Fixed effects with lagged dependent variables

In this section, we consider the problems introduced by the presence of fixed unit effects in a dynamic model of TSCS data. Thus we consider the autoregression

$$y_{i,t} = \phi y_{i,t-1} + \alpha_i + \varepsilon_{i,t} \tag{18}$$

which generalizes easily to a specification including other exogenous variables. The obvious way to estimate Equation 18 is with OLS, which is exactly equivalent to first centering each $y_{i,t}$ around its unit mean and then regressing the centered variables on the lagged centered variables. In the context of fixed effects, OLS is thus known as least squares dummy variables (LSDV) and we will use that terminology here, but remembering that LSDV is exactly OLS.

But this induces a correlation between the demeaned lagged y and the demeaned error term. This point was made long ago for a single time series by Hurwicz (1950), and was generalized to TSCS (or panel) data by Nickell (1981). The algebra is trivial. Let $\tilde{y}$ and $\tilde{\varepsilon}$ be the centered y and the error,

$$\tilde{y}_{i,t-1} = y_{i,t-1} - \frac{1}{T_i}\sum_{t=1}^{T_i} y_{i,t-1} \tag{19}$$

$$\tilde{\varepsilon}_{i,t-1} = \varepsilon_{i,t} - \frac{1}{T_i}\sum_{t=1}^{T_i} \varepsilon_{i,t}. \tag{20}$$

But then clearly $\mathrm{E}[\tilde{y}_{i,t-1}\tilde{\varepsilon}_{i,t}] \neq 0$. The error term, $\varepsilon_{i,t-1}$ is contained with weight $1 - \frac{1}{T_i}$ in $\tilde{y}_{i,t-1}$ and with weight $\frac{1}{T_i}$ in $\tilde{\varepsilon}$. This correlation renders the LSDV estimators of ϕ and β biased. Nickell (1981) derived the asymptotic bias (as $N \to \infty$) and showed that it was $\mathrm{O}(T^{-1})$. Note that the bias gets smaller as we increase T, that is move from the "panel" world to the TSCS world. Clearly the bias term is huge for two or three wave panels; is it similarly an issue for typical political economy data? And is the cure that is commonly used in panel situations worse than the disease in TSCS situations.

Many cures to the problem have been 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 (18). 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 $\text{O}(N^{-1}T^{-3/2})$ 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. Second, the formula requires knowledge of the true parameters in Equation (18), 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 $\text{O}(T^{-1})$ 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. But with TSCS data we have much larger T's.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.

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 (18) with the following additional assumptions:

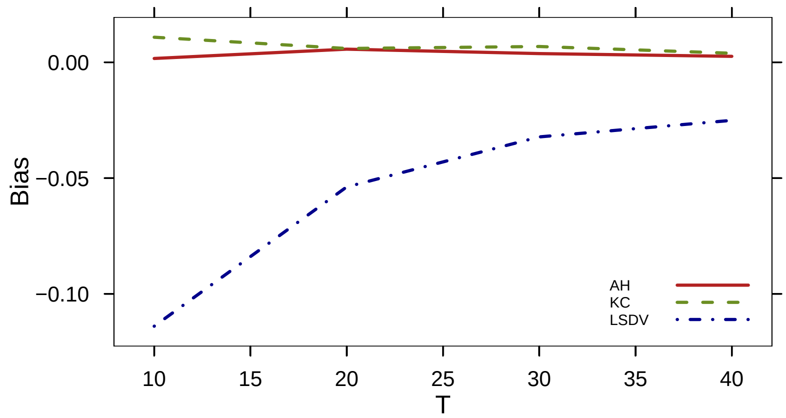
$$\varepsilon \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$$
$$\alpha_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$$
$$\sigma_\alpha = \mu(1-\phi)\sigma_\varepsilon$$
$$x_{i,t} = \delta x_{i,t-1} + \gamma(1-\delta)\alpha_i + \omega_{i,t}$$
$$\omega_{i,t} \stackrel{iid}{\sim} N(0, \sigma_\omega^2)$$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

These assumptions are fairly standard in the literature. The parametrization 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.

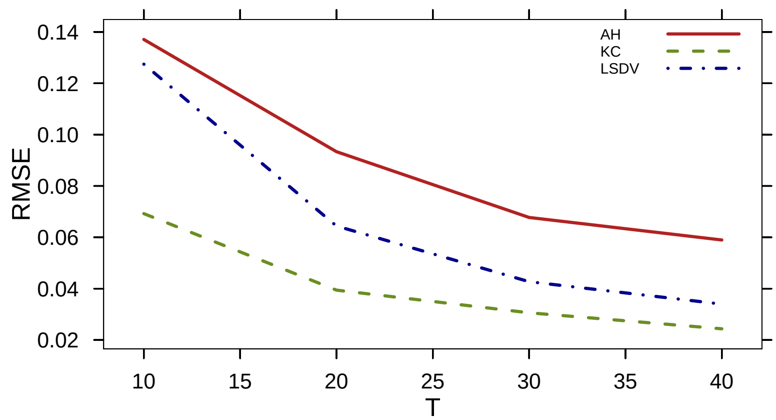
We are interested in two criteria for evaluating the proposed estimators: bias and root mean square error. However, root mean square error is more important since it incorporates both bias and estimation variability. That is, we might be willing to use a slightly biased estimator that had dramatically 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 on-line appendix. 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 2 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 when we look at RMSE. Here



(a) Bias

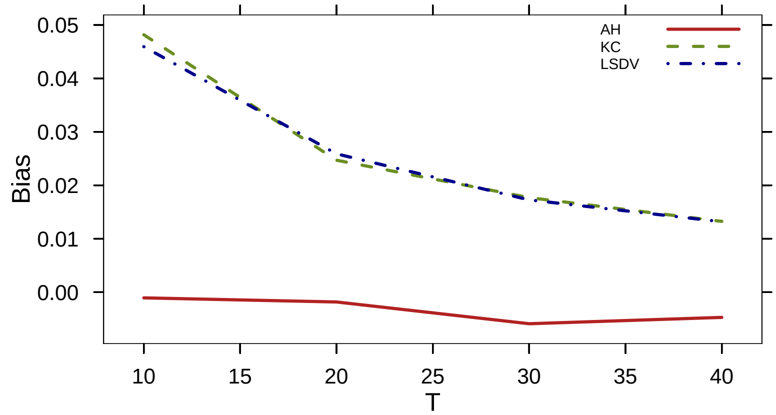


(b) RMSE

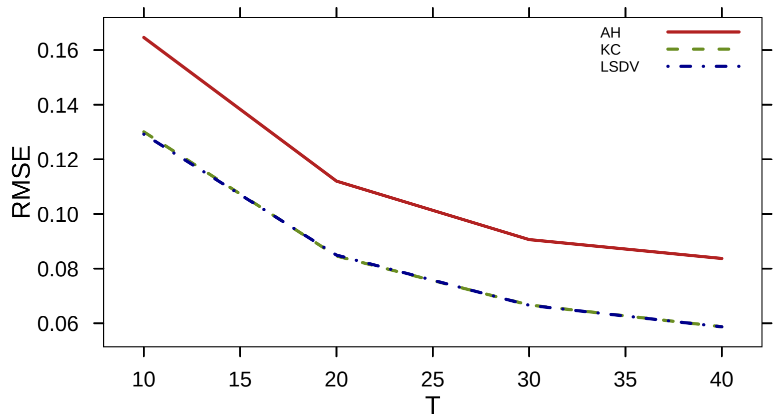
Figure 2: Monte Carlo Results for estimates of ϕ as a function of T from LSDV, Kiviet Correction, and Anderson-Hsiao estimators. Simulation parameters: $N = 20$, $\beta = 1$, $\phi = 0.6$, $\delta = 0.5$, $\sigma_\omega = 0.6$, $\mu = 1$, $\gamma = 0.3$, and $\sigma_\varepsilon = 1$

the KC estimator continues to dominate, but the AH estimator pays a high cost in terms of sampling variability to get unbiasedness. In terms of RMSE, the LSDV is superior to the AH. The advantage of the Kiviet estimator over LSDV declines as T gets larger, though even for $T = 40$ the advantage of Kiviet is discernible albeit far from enormous.

The picture for LSDV continues to improve if we look at the results for β, typically the parameter of interest in most analysis. Figure 3 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 RMSE of both the LSDV and KC estimators are virtually identical.



(a) Bias



(b) RMSE

Figure 3: 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.

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 3, with LSDV doing as well as the much more complicated KC estimate.[21]

[21]These results are in the on-line appendix which also contains a figure showing similar results as the autocorrelation in the dependent variable changes.

Which method to use

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 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.[22]

7 Conclusion

There is no cookbook for dealing with modeling the dynamics of TSCS models; examination of the specifications, and what they entail substantively, can allow TSCS analysts to think about and model these dynamics. Well known econometric tests help in this process, and validated methods make it easy to estimate the appropriate dynamic model. Modeling decisions are less critical where variables equilibrate quickly; as the adjustment process slows, the various models imply more and more different characteristics of the data. Analysts should take advantage of this to choose the appropriate model. Analysts should be very cautious about using ideas about integrated series unless it is plausible that the data could have properties consistent with integrated processes.

Being more specific, we have provided evidence that, unlike the claim made by Achen, there is nothing pernicious in the use of a model with a lagged dependent variable. Obviously attention to issues of testing and specification are as important here as anywhere, but there is nothing about lagged dependent variables that make them generically harmful. As we have seen, there are a variety of generic dynamic specifications, and researchers should choose amongst them using the same general methodology they use in other cases.

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.

The brief conclusion of this paper is that issues of dynamics are substantive issues, and that researchers should so think of them. Thus, the dynamic issues confronting political

[22] The Kiviet estimator has been almost exclusively used in the panel situation, and so our advice here is nothing new. Similar advice is given by Judson and Owen (1999).

economists using TSCS data are not technically formidable, but they do require matching the poliitcal economy models to the choice of dynamic specification.

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.

Alvcarez, R.Michael and Jonathan N. Katz. 2000. "Aggregation and Dynamics of Survey Responses: The Case of Presidential Approval." Social Scicnce Working Paper 1103, Division of the Humanities and Social Science, California Institute of Technology,.

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.

Baker, Regina. 2008. "Lagged Depemdent Variables and Reality: Did you specify that autocorrelation *a priori*." Unpublished paper, Department of Politcal Science, Universit of Oregon.

Beck, Nathaniel. 1985. "Estimating Dynamic Models is Not Merely a Matter of Technique." *Political Methodology* 11:71–90.

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

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

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.

De Boef, Suzanna and Luke Keele. 2008. "Taking Time Seriously." *American Journal of Political Science* 52:184–200.

Engle, Robert and C. W. J. Granger. 1987. "Co-Integration and Error Correction: Representation, Estimation and Testing." *Econometrica* 55:251–76.

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.

Granger, Clive W. and Paul Newbold. 1974. "Spurious Regressions in Econometrics." *Journal of Econometrics* 2:111–20.

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

Hendry, David and Graham Mizon. 1978. "Serial Correlation as a Convenient Simplification, Not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England." *Economic Journal* 88:549–563.

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

Hurwicz, L. 1950. "Least-Squares Bias in Time Series." In *Statistical Inference in Dynamic Economic Models*, ed. T. Koopmans. New York: Wiley pp. 365–83.

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.

Keele, Luke and Nathan J. Kelly. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analsis* 14:186–205.

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

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.

Mizon, Graham. 1984. "The Encompassing Approach in Econometrics." In *Econometrics and Quantitative Economics*, ed. David Hendry and Kenneth Wallis. Oxford: Basic Blackwell pp. 135–172.

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

A On Line Appendix: 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

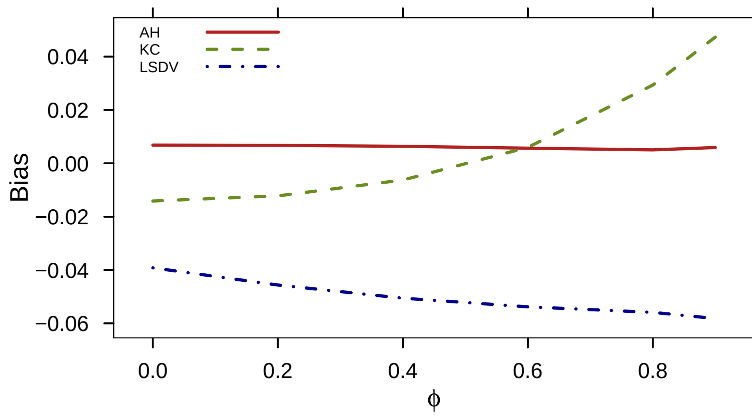
		LSDV		Anderson-Hsiao		Kiviet	
T	ϕ	RMSE	Bias	RMSE	Bias	RMSE	Bias
4	0.00	0.287	0.056	0.467	−0.010	0.285	0.044
4	0.20	0.284	0.073	0.601	−0.011	0.281	0.060
4	0.40	0.279	0.082	1.426	−0.008	0.275	0.069
4	0.60	0.269	0.074	3.262	−0.061	0.267	0.062
4	0.80	0.252	0.031	10.827	0.209	0.252	0.018
4	0.90	0.246	−0.018	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

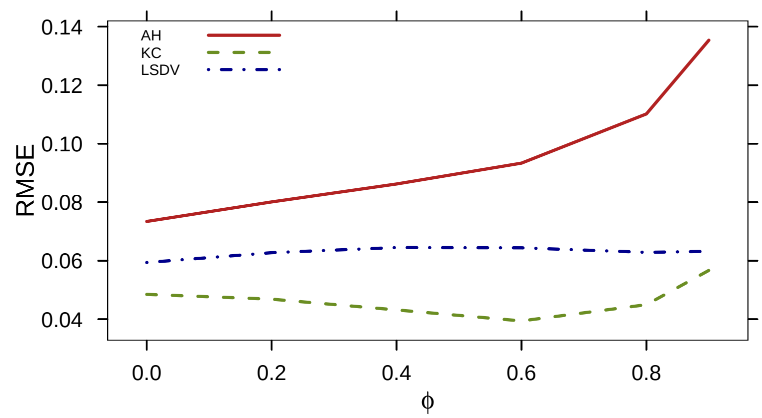
		LSDV		Anderson-Hsiao		Kiviet	
T	ϕ	RMSE	Bias	RMSE	Bias	RMSE	Bias
4	0.00	0.303	−0.275	0.765	0.012	0.216	0.162
4	0.20	0.347	−0.321	2.021	0.076	0.194	0.119
4	0.40	0.381	−0.355	2.927	0.013	0.185	0.089
4	0.60	0.403	−0.379	22.710	−0.146	0.185	0.072
4	0.80	0.424	−0.401	25.756	0.902	0.181	0.054
4	0.90	0.426	−0.404	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

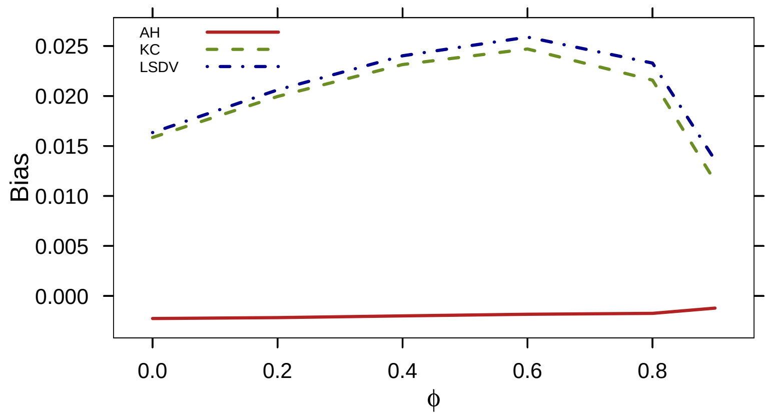


(a) Bias

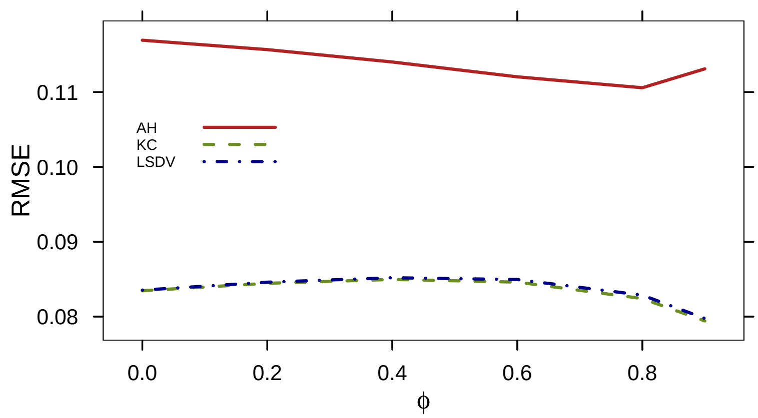


(b) RMSE

Figure A.1: 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 $\sigma_\varepsilon = 1$



(a) Bias



(b) RMSE

Figure A.2: 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 $\sigma_\varepsilon = 1$