Revisiting Dynamic Specification*

Suzanna De Boef
Department of Political Science
Pennsylvania State University
State College, PA 16802
Tele: 814-863-9402
Email: sdeboef@psu.edu

Luke Keele
Department of Politics and International Relations
Nuffield College and Oxford University
Manor Road, Oxford
OX1 3UQ UK
Tele: +44 1865 278716
Email: luke.keele@politics.ox.ac.uk

July 18, 2005

*Paper prepared for presentation at the 2005 annual meeting of the Society of Political Methodology, Florida State University. Thanks to Jim Stimson our mentor in all things time series (and more), colleague, and friend, for the long run multiplier effect his time series analysis continues to have in political science. Thanks also to Amber Boydstun for comments.

Abstract

Dramatic change in the world around us has stimulated a wealth of interest in research questions about the dynamics of political processes. At the same time we have seen increases in the number of time series data sets and the length of typical time series. Parallel advances have occurred in time series econometrics. These events have turned more political scientists into time series analysts and motivated more political methodologists to delve further into the annals of time series econometrics. But before taking the next advanced time series course, we recommend that time series analysts devote more time to issues of specification and interpretation. While advances in time series methods have helped us to change how we think about the process of political change in important ways, too often analysts have failed to recognize the wide number of general models available for stationary time series data, have estimated restricted models without testing the implied restrictions, and have done a poor job of drawing interpretations from their results. The consequences, at best, are poor connections between theory and tests and thus a limited cumulation of knowledge. More likely, the costs include biased results as well. We identify a number of general dynamic specifications, each a linear parameterization of the basic autoregressive distributed lag model and each highlighting different types of information. We then discuss the consequences of imposing restrictions on any of them. We recommend that analysts start with one or a combination of these general models and test for restrictions before adopting them. We illustrate this strategy with data on support for the Supreme Court and on presidential approval. Finally, we recommend that analysts make use of the wide array of information that can be gleaned from dynamic specifications. Such a practice will help us to better equate dynamic econometrics with dynamic theory.

The field of political science has made many advances in time series econometrics over the last two decades. Increases in the availability of time series data across all subfields have stimulated interest in research questions about the dynamics of politics. This has led to a search for methodologies appropriate for the analysis of our data; a search that has taken the discipline as far afield as hydrology and signal analysis. Work on the properties of individual time series, the consequences of temporal aggregation, methods for dealing with endogenous, nonstationary, and pooled time series, and models for time series of durations and regime switching have all played an important role in advancing how we think about temporal political processes. In spite, or perhaps because of, these advances at the frontiers of time series analysis, analysts have spent little time familiarizing themselves with the kinds of dynamic specifications that are relevant for most political data. As Neal Beck wrote in 1991, "the major problems in this area are not technical. They are instead problems of interpretation, or problems of relating econometrics and politics" (Beck 1991, 72). In short, one of the greatest shortcomings in applied time series analysis revolves around something as mundane as dynamic specification.

Dynamic specification, the inclusion of lagged variables (X or Y) on the right hand side of the model, is fundamentally about capturing temporal aspects of politics through statistical specification. By definition good dynamic specifications must (1) be consistent with the empirical properties of the data, (2) be general enough to subsume the actual process that generated the data (the DGP), and (3) allow for statistical tests of the relevant theoretical model. Achieving a specification that reliably presents the richness of political dynamics is a difficult task, requiring careful attention—both theoretical and empirical—to the properties of individual time series and the relationships among them.

Standard practice in political science is to use a highly specific form of dynamic specification, typically in the form of a lagged dependent variable model.[1] Little at-

[1]The notable exceptions to this rule occurs when an inference of cointegration guides specification or use a Box Jenkins approach to time series analysis.

tention is paid to the restrictions that a particular specification imposes on the DGP. Moreover, once a model specification is selected and estimated, inferences are typically limited to short-run effects and interpretation follows that of a static model; "a unit change in X (at $t-s$) leads to an expected change in Y (at time t)". As such, analysts frequently fail to compute and interpret quantities such as long-run impacts of exogenous variables, mean and median lag lengths of effects, equilibrium conditions, and the rate of equilibrium correction, each of which may be of theoretical interest.

This limited form of interpretation represents not just poor econometric practice, but has broader consequences. Since analysts are unfamiliar with a class of statistical models that allow for the representation of a variety of temporal causal effects, they develop theories and test hypotheses that fall far short of the detail that statistical models allow. This comes in part from a long tradition of research using cross-sectional data. Temporal dynamics are largely (necessarily) excluded from that research tradition. The imposition of statistical practices from cross-sectional research has left theories of political dynamics underdeveloped. The problems worsen when an overly simple model is chosen that is too restrictive, leading to misspecification and bias.

Contributing further to the use of a limited class of dynamic specifications for stationary data is the fact that we have come to tie the dynamic specifications associated with relatively recent time series econometric advances for unit root processes almost exclusively with the particulars of those advances. In the process we have ignored decades of work with dynamic regressions for stationary processes that are, we think, more relevant for the kinds of research questions we ask and data we use.

Here, we revisit dynamic specification by focusing on autoregressive distributed lag and error correction models and the dynamic quantities that can be derived from them. We outline a typology of dynamic specifications, so that choosing an optimal specification is a matter of writing theory that matches the complexity of available models. As such, time series analysis will be more about "relating politics to econometrics".

1 Outlining a Strategy for Choosing a Dynamic Specification

The interest in time series methods among political scientists parallels the passage of time and the subsequent increase in available time series data. Survey data on U.S. presidential approval, partisanship, and consumer sentiment, for example, have been available now for over 50 years. Quarterly and monthly data, available since at least the mid-1960s for some series, means we have time series of over 200 observations on key American political attitudes. Data on trade, economic sanctions, military expenditures, budgets, vetoes, cabinet durations and more are also available for extended periods of time. The field of comparative politics is even now seeing time series of pre and post communist behavior that are long enough to draw inferences about important research questions.

Time series analysis presents challenges to the interpretation of estimated statistical models unlike those of cross-sectional data. In cross-sectional analysis all estimated effects are necessarily contemporaneous and therefore static. Cross-sectional data do not allow us to assess whether causal effects are contemporaneous or lagged, let alone whether they have some component that is distributed over future time periods.[2] In contrast, consider the two types of effects that we might encounter in a time series model:

- An exogenous variable may have only short term effects on the outcome variable. These may occur at any lag, but the effect does not persist into the future. The reaction of economic prospections to the machinations of politicians, for example, may be quite evisceral—influencing evaluations today, but not tomorrow. Here the effect of X_t on Y_t has no memory.

[2]When lagged values of exogenous variables or indicators for observation-year are included, they are generally treated as controls and absent a lagged dependent variable are not dynamic models in the sense to which we refer.

- An exogenous variable may have both short and long term effects. In this case, the changes in X_{t-s} affect Y_t, but that effect is distributed across several future time periods. Often this occurs because two processes exist in an equilibrium and any movement by X_t will cause Y_t to adjust toward the level of X_t so that equilibrium will be maintained. Levels of democracy may affect trade between nations both contemporaneously and into the future. These effects may be distributed across only a few or perhaps many future time periods. How many time periods is an empirical question that we can answer with our data. Likewise, levels of trust in government may affect regime support in post-communist countries today and well into the future.

The ability to observe these different effects and make inferences about them are a function of the particular form of model estimated and the parameter values from that model. The power of dynamic specification is that it allows us to estimate and test for both short and long run effects and to compute a variety of other quantities that can help us better understand political processes. The difficulty is identifying the *best* specification for the task. While theory may suggest that we allow for both types of effects and consider a variety of dynamic quantities, it typically speaks little to the precise nature of these effects. Under these conditions, we need a strategy for identifying optimal dynamic specification.

We advocate using highly general dynamic models that encompass all the possible temporal effects. Having selected a general model specification, analysts should then test restrictions on that model to see if the dynamics implied in theory are consistent with the data.[3] Finally, analysts should draw inferences about short and long run dynamics as well as mean/median lag lengths, equilibrium correction rates, and the patterns of decay in effects, *regardless of the specification estimated.* Without such a strategy, much of the potential information about the dynamics of politics remains past our fingertips, just out of reach.

[3] Of course, diagnostic tests should be performed at this stage as well.

Our approach requires knowledge of how to interpret such general dynamic models in detail. The general dynamic model we consider is the auto-distributed lag (ADL) model and others that are mathematically related to it. The ADL model is the cornerstone of all dynamic regressions and a thorough understanding of the information it contains provides a foundation for our strategy. From it we define such relevant terms as long run multipliers, equilibria, and mean/median lag lengths and we demonstrate how to compute them.

After considering the ADL model, we discuss error correction models (ECMs). ECMs represent an equally general and particularly useful class of dynamic models, but one seldom exploited in political science outside the context of cointegration. Yet ECMs are isomorphic with the autoregressive distributed lag model and therefore contain the same information, fit the data the same, imply the same underlying dynamic, and are appropriate for stationary data. However, error correction models allow for more convenient testing of long run effects than does the ADL. We also discuss error correction models as a class of models to introduce some dynamic specifications that are virtually unknown in political science. We now turn to the ADL model and discuss how to interpret it.

2 A General Model

Most dynamic theories in political science tell us little about the specific statistical model that we should estimate. Often we are only comfortable asserting that the processes are stationary and weakly exogenous and that effects may be short and/or long term. Given that we typically start with such a general level of knowledge, we want a general dynamic model that imposes few restrictions on the data generating process. The following model fits our criteria:

$$Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{n} \sum_{i=0}^{q} \beta_{jp} X_{jt-i} + \varepsilon_t \tag{1}$$

where ε_t is white noise, $|\sum_{i=1}^{p} \alpha_i| < 1$ so that Y_t is stationary, and the processes generating X_j are weakly exogenous for the parameters of interest such that $E(\varepsilon_t, X_{js}) = 0 \forall t, s$, and j. This is an ADL$(p, q; n)$ model, where p refers to the number of lags of Y_t, q the number of lags of X_t, and n the number of exogenous regressors included in the model.[4] Given that there are no contemporaneous dependent variables on the right side, it should be consistently estimated by ordinary least squares (OLS) (Davidson & MacKinnon 1993).[5]

For simplicity, we will refer to the case when $p = q = n = 1$, but the results generalize:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t. \tag{2}$$

We can use the above model to test hypotheses about the dynamic relationship we have previously discussed. That is, we can estimate both short run effects and the long run equilibrium relationship among the variables. The short run effects of the independent variables are given directly by the estimated coefficients, β_0 and β_1, also called impact multipliers. They give the immediate effect on Y_t, of a unit change in X_t at some given t. For example, if X_t is a measure of economic expectations and our data is quarterly, β_0 tells us how levels of economic expectations in 1991Q2 affect presidential approval in that same quarter. β_1 tells us how previous levels of economic expectations in 1991Q1 affect presidential approval in the subsequent quarter.

Two time series share a long run equilibrium any time that we can express one as a function of the other. The long run equilibrium is given by the unconditional expectations or the expected value of Y_t. Let $y^* = E(Y_t)$ and $x^* = E(X_t)$ for all t. If the two processes moved together without error, in the long-run, they would converge to the following equilibrium values for the ADL(1,1;1):

$$y^* = \alpha_0 + \alpha_1 y^* + \beta_0 x^* + \beta_1 x^*. \tag{3}$$

[4] q need not be uniform for X_t, but rather gives the maximum lag length; any β_{jt-i} can be set to zero so that this notation generalizes.

[5] The proof for the consistency of OLS assumes that ε_t is IID after the lag of Y_t is included in the model. See (Keele & Kelly 2006) for a study of when this is not true.

Solving for y_t^* in terms of x_t^*, yields:

$$
\begin{aligned}
y^* &= \frac{\alpha_0}{1-\alpha_1} + \frac{\beta_0+\beta_1}{1-\alpha_1} x^* \\
&= k_0 + k_1 x^* \qquad (4)
\end{aligned}
$$

where $k_0 = \frac{\alpha_0}{(1-\alpha_1)}$ and $k_1 = \frac{(\beta_0+\beta_1)}{(1-\alpha_1)}$, and k_1 gives the long run multiplier of X_t with respect to Y_t. We can think of the long run multiplier as the total effect X_t has on Y_t distributed over future time periods. The long run equilibrium is likely to be of direct interest in ascertaining the conditions under which two series sustain an equilibrium. In fact in some cases long run equilibria and the long-run multiplier are of greater interest than the short run effects. Policymakers, for example, debate the optimal defense spending needed to generate a sustained peaceful equilibrium or the long run effects (the dynamic multiplier) of deficit spending on economic growth.

When the equilibrium relationship between two time series is disturbed, let's say economic expectations and presidential approval, then $y^* - (k_0 + k_1 x^*)0$ will not be zero. In this case, we expect a change in the level of presidential approval in the next period back toward the equilibrium. Interest in the rate of return to equilibrium, also known as error correction, is often motivated by the desire to understand just how responsive a process is. Does consumer sentiment, for example, respond quickly to good economic news or are consumers more skeptical, responding slowly, in fact willing to tolerate sentiment too low for the long-run equilibrium in the interim? The ADL also provides us with information about the speed of this error correction. The speed of adjustment is given by $(1-\alpha_1)$, as it dictates how much Y_t changes over each future period. If the long run multiplier is 5.0, and the error correction rate is 0.50, Y_t will change 2.5 points in $t+1$, and then another 1.25 point at $t+2$ and then 0.625 in $t+3$ and so on until the two series have equilibrated. Obviously increases in α_1 produce slower rates of error correction, with the reverse also being true.[6]

[6]When the maximum number of lags of Y_t in the model exceeds 1, sum of their coefficients minus one give the *cumulative adjustments*:

Other quantities also inform us about politics. In addition to knowing the magnitude of the total effect of a shock as measured by the long run multiplier, it is often useful to know how many periods it takes for some portion of the total effect of a shock to dissipate or how much of the shock has dissipated after some number of periods. Here features of the lag distribution are informative for theory. Econometric texts often give some treatment of the mean and median of the lag distribution of X_t, which provides information about the pattern of adjustment to disequilibrium. The median lag tells us the first lag, r, at which at least half of the adjustment toward long-run equilibrium has occurred following a shock to X_t, providing information about the speed with which the majority of a shock dissipates. It is calculated by listing the effect of a unit change in X_t at each successive lag, standardizing it as a proportion of the cumulative effect, and then noting at which lag the sum of these individual effects exceeds half of the long-run or equilibrium effect. A median lag of 0 tells us that half of the effect is gone in the period it has occurred. We might expect such short median lags when equilibria are very "tight", that is, lagged Y_t has a small coefficient and X_t a large one. A median effect of 4 periods would be quite long for most political processes. Given quarterly data on presidential approval, for example, theory tells us that a majority of the effects of shock in inflation will be realized (well) within a year.[7] Mean lags tell us how long it takes to adjust back to equilibrium, the average amount of time for a shock to play out. Political processes of the kind we consider here are likely to have mean lag lengths on the order of 6, perhaps less.

Median lag lengths are somewhat tedious to calculate and are typically given short shrift in political science. In econometrics, median lag lengths are the topic of textbook treatments, but the calculation generalizes to any percentage effect we might be

$$\sum_{i=1}^{p} \alpha_i - 1 \tag{5}$$

where p gives the maximum number of lags of Y_t in the model (Bannerjee, Dolado, Galbraith & Hendry 1993). The cumulative adjustment rate is thus equal to the rate of error correction when only one lag of the dependent variable is included in the model.

[7] Experience suggests to us that a median lag length of 2 is not uncommon is the study of American public opinion where there is a fair amount of inertia in y but x is often a strong predictor of y as well.

interested in. When deriving the formula for the median lag, it is useful to write the general model using lag polynomials:

$$A(L)Y_t = B(L)X_t + \varepsilon_t \tag{6}$$

where L is the lag operator: $L^i X_t = X_{t-i}$, $A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_p L^p$ and $B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \ldots + \beta_q L^q$.

We can calculate the median lag by computing m for successive values of r and recording the value of r when $m \geq .5$:

$$m = \frac{\sum_{r=0}^{R} \omega_r}{\sum_{r=0}^{\infty} \omega_r} \tag{7}$$

where ω_r:

$$\omega_r = \frac{B(L)}{A(L)} \tag{8}$$

and $\sum_{r=0}^{\infty} \omega_r = \frac{B(1)}{A(1)}$, where $A(1) = 1 - \sum_{i=1}^{p} \alpha_i$ and $B(1) = \sum_{i=0}^{q} \beta_q$. ω_r represents the effect of a shock r periods after it occurs. The summation in the denominator is across all values of time and thus provides the long-run effect: the familiar k_1. The summation in the numerator allows us to calculate the effects for up through any number of periods, R. The division in equation 7 thus normalizes the adjustment as a proportion of the total adjustment up through r periods. It is often useful to graph the standardized lag distribution to best see the patterns in affects. Graphs of lag distributions can help us answer questions such as: "what proportion of the total effect has dissipated after 4 periods (quarters)?[8] Unstandardized lag distributions or cumulative standardized lag distributions may also provide useful visuals for policy proscriptions.

[8] These graphs make the relationship between the lag distribution, the impulse response function (IRF) and the cross correlation function (CCF) readily apparent. The IRF and CCF are familiar to those acquainted with the Box-Jenkins time series framework. The former give the response of Y to an impulse shock in X in its natural units, the latter standardizes the effects. Thus the graph of the standardized lag distribution is equivalent to the CCF.

In contrast, the mean lag length tells us how long it takes to adjust back to equilibrium. The mean lag for X_j is given by:

$$\mu = \frac{\partial W(1)}{W(1)} = \frac{\beta(1)'}{\beta(1)} - \frac{\alpha(1)'}{\alpha(1)} = \frac{\sum_{r=0}^{\infty} r\omega_r}{\sum_{r=0}^{\infty} \omega_r} \tag{9}$$

where $W(1) = \frac{B(1)}{A(1)}$ and $'$ denotes the derivative with respect to L.

The mean and median lags are useful when we wish to know how many periods it takes for a process to return to equilibrium. If the level of economic expectations increases, how long will it take to see movement in trust to its new equilibrium value? Is change fast or slow? A President looking ahead to his reelection campaign may be particularly interested in the length of time it takes the public to forget recession. Where α_1 is large the mean lag length will be long.[9] The median lag length will be short when β_0 approaches one half of the long run multiplier, $.5 \times k_1$, and will equal zero when it is greater than this quantity. When both the median and mean lag length are small, adjustment is fast, when the median is short and the mean long, a large part of the disequilibrium is corrected quickly while the long-run response takes some time to complete the adjustment.[10]

The ADL is a versatile and general model with much to recommend it. From it we can draw inferences about dynamic behavior in rich detail. One might then assume that we recommend that all analysts use the ADL all the time. But, in fact, there is another class of dynamic models that are more versatile than the ADL and are equally general. This class of error correcting models (ECMs) provide estimates of short and long term effects as well as the rate of return to equilibrium in a form that is more easily interpreted, often making it the model of choice.

Unfortunately, the ECM suffers from benign neglect in political science. Because of the intimate connection between error correction and cointegration, political scientists have largely—and incorrectly—associated error correction models exclusively with unit

[9]The mean lag length will be infinite when Y is nonstationary such that $\alpha_1 = 1$.

[10]If any of the lag weights are negative, then the mathematics used do not apply and mean lag lengths cannot be computed. In such cases, it is likely that the model is misspecified so that negative lag weights can be a good diagnostic for dynamic specification (Hendry 1995, 216).

root time series and cointegration. This association has meant that often (we believe) contrived arguments about the nature of political time series are offered to justify use of an ECM. More problematic, however, the potential advantages of ECMs have gone largely ignored, as analysts assume the an ECM cannot be used since their data is not integrated. This is a problem we seek to rectify since careful consideration of ECMs is an important part of choosing the best dynamic specification.

3 ECMs–An Alternative General Model

The term error correction model applies to any parameterization of the ADL that directly estimates the rate at which Y_t changes to correct perturbations from the long run equilibrium. There are, in fact, many parameterizations for which this occurs, all of which are equivalent in the sense that they contain the same information. The equivalence of the ADL and various ECM models is well known in econometric circles. Yet there is debate in political science on this point (Beck 1993, Williams 1993, Durr 1993*a*, Durr 1993*b*, Smith 1993). We prove the equivalence between the ADL and two dynamic specifications that fall in the class of ECMs. Parts of these proofs can be seen elsewhere (Davidson & MacKinnon 1993, Bannerjee et al. 1993). First, consider once again the ADL(1,1;1) model:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t. \tag{10}$$

First, we take the first difference of Y_t to produce:

$$\Delta Y_t = \alpha_0 + (\alpha_1 - 1) Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t. \tag{11}$$

Then we add and subtract $\beta_0 X_{t-1}$ from the right hand side:

$$\Delta Y_t = \alpha_0 + (\alpha_1 - 1) Y_{t-1} + \beta_0 \Delta X_t + (\beta_0 + \beta_1) X_{t-1} + \varepsilon_t. \tag{12}$$

Finally, we add and subtract $(\alpha_1 - 1)X_{t-1}$ from the right hand side and collect terms to form the Generalized Error Correction Model (GECM):

$$\Delta Y_t = \alpha_0 + \gamma(Y_{t-1} - X_{t-1}) + \lambda_1 \Delta X_t + \lambda_2 X_{t-1} + \varepsilon_t \tag{13}$$

where $\gamma = (\alpha_1 - 1)$, $\lambda_1 = \beta_0$, and $\lambda_2 = \beta_1 + \beta_0 + \alpha_1 - 1$.

It is immediately apparent that the estimated parameters, with the exception of λ_1, which equals β_0, are different than those of the ADL. The information provided *by* the estimates themselves differs across the two models. However, because the GECM is a linear reparameterization, the information that can be drawn *from* the estimates is identical. Most notably, the GECM, unlike the ADL model, provides a direct estimate of the error correction rate and its standard error. Denoted here by γ, the coefficient on lagged Y_t, the error correction rate tells us the speed with which Y_t adjusts whenever it is not equal to $(k_0 + k_1 x_t)$. One implication of error correction is that even in the absence of change in X for many periods, we expect Y to continue to change as long as the series are not in equilibrium. Only when $y_t = (k_0 + k_1 x_t)$ do we expect $\Delta Y = 0$.

Note that it appears that k_1, the long-run multiplier, is one and the equilibrium relationship is one-to-one. However, an additional X_{t-1} term is included in the GECM to "break homogeneity" or to allow the long-run relationship to generalize beyond the restrictive one-to-one case implied by the disequilibrium term: $(Y_{t-1} - X_{t-1})$. The long run multiplier is again given by the sum of the short run effects divided by 1 minus the sum of the coefficients on the lagged Y. Retaining the notation $B(1)$ for the sum of the short run coefficients and $A(1)$ for 1 minus the sum of the coefficients on the lagged Y:[11]

$$k_1 = \frac{B(1)}{A(1)} = \frac{\lambda_1 + \lambda_2 - \lambda_1 - \gamma}{1 - (1 + \gamma)} = -\frac{(\lambda_2 - \gamma)}{\gamma} \tag{14}$$

[11]Rewrite equation 13 in levels to show that $B(1) = \lambda_2 - \gamma$ and $A(1) = -\gamma$.

$$\begin{aligned} Y_t &= \alpha_0 + Y_{t-1} + \gamma Y_{t-1} - \gamma X_{t-1} + \lambda_1 X_t - \lambda_1 X_{t-1} + \lambda_2 X_{t-1} + \varepsilon_t \\ &= \alpha_0 + (1+\gamma) Y_{t-1} + (-\gamma - \lambda_1 + \lambda_2) X_{t-1} + \lambda_1 X_t + \varepsilon_t \end{aligned}$$

Because $B(1)$ is equal to the sum of the coefficients on the X_{t-i}, we have $B(1) = \lambda_1 + \lambda_2 - \lambda_1 - \gamma = \lambda_2 - \gamma$. Because $A(1)$ is 1 minus the sum of the coefficients on lagged Y, we have $A(1) = 1 - (1 + \gamma) = -\gamma$.

which by substitution produces the same estimate as that calculated from the ADL:

$$k_1 = -\frac{(\lambda_2 - \gamma)}{\gamma} = -\frac{(\beta_1 + \beta_0 + \alpha - 1 - (\alpha_1 - 1))}{(\alpha_1 - 1)} = \frac{(\beta_1 + \beta_0)}{(1 - \alpha_1)} \tag{15}$$

such that the GECM produces the same value for k_1 as the ADL does.[12]

The GECM also produces the same estimates of the short-run effects as the ADL model. In the GECM, the short-run effects of X_t and X_{t-1} are given by $\lambda_1 = \beta_0$ and $\lambda_2 - \lambda_1 - \gamma = \beta_1 + \beta_0 + \alpha_1 - 1 - \beta_0 - (\alpha_1 - 1) = \beta_1$. So while the same short-run effects estimated in the ADL model may be retrieved from the GECM, we will only have a standard error for λ_1. Since $\gamma = (\alpha_1 - 1)$, and α_1 is estimated in the ADL representation, the error correction rate is also available to us in either specification. The standard error for γ and α_1 are the same, so the inferences about error correction can be drawn from either specification.

While the GECM is useful for demonstrating the equivalence between the ADL and ECMs, it can be cumbersome when used in applied contexts and is not usually estimated (But for an example see De Boef & Kellstedt 2004). There are, however, two other additional transformations of the ADL, one due to Bewley (1979) and one to Bardsen (1989), that tend to be more useful with applied data. We first describe the Bardsen transformation, which is perhaps the most useful form of the error correction model. We again start with the ADL (1,1;1):

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t. \tag{16}$$

To parameterize the model as a Bardsen ECM, first take the first difference of Y_t:

$$\Delta Y_t = \alpha_0 + (\alpha_1 - 1) Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t. \tag{17}$$

Now add and subtract $\beta_0 X_{t-1}$ from the right hand side:

[12]Even though the disequilibrium term itself is restricted, γ remains a consistent estimator of the error correction rate in this case (For a proof see Bannerjee et al. 1993, 61).

$$\Delta Y_t = \alpha_0 + (\alpha_1 - 1)Y_{t-1} + \beta_0 \Delta X_t + (\beta_0 + \beta_1)X_{t-1} + \varepsilon_t. \tag{18}$$

Regrouping terms leaves us with the following equation:

$$\Delta Y_t = \alpha_0 + \alpha_1^* Y_{t-1} + \beta_1^* \Delta X_t + \beta_2^* X_{t-1} + \varepsilon_t \tag{19}$$

Again, by substitution, we can see the equivalence between the ADL, GECM, and the Bardsen error correction model: $\alpha_1^* = (\alpha_1 - 1)$, $\beta_1^* = \beta_0 = \lambda_1$, and $\beta_2^* = \beta_1 + \beta_0$.

Once again the parameters estimated are different but contain the same information. We notice, too, that the regressors are more "natural" than with the GECM; there is no need to include the homogenous disequilibrium term (the $(Y_{t-1} - X_{t-1})$ term). Using levels and first differences of the time series we can directly estimate the error correction rate, α_1^* and its standard error and the short run effect of X_t and its standard error. The long run multiplier, k_1, is more readily calculated here than from the GECM or the ADL:

$$k_1 = \frac{\beta_2^*}{\alpha_1^*} = \frac{(\beta_1 + \beta_0)}{(\alpha_1 - 1)} \tag{20}$$

A simple example demonstrates the interpretation of all three coefficients in the model. Let's say we regress the first difference of presidential approval on one lag of presidential approval, one lag of economic expectations, and the first difference of economic expectations as in Equation 19. The estimated coefficients are $\hat{\beta}_1^* = 0.5$, $\hat{\alpha}_1^* = -0.5$, and $\hat{\beta}_2^* = 1.0$. If economic expectations were to increase five points, how will that affect presidential approval in the context of the error correction model? First, presidential approval will increase 2.5 points immediately (5 * 0.5, the coefficient of $\hat{\beta}_1^*$). But the error correction model implies that presidential approval and economic expectations also have an equilibrium relationship, where this increase in economic expectations disturbs the equilibrium, causing presidential approval to be too low. As a result, presidential approval will increase another five points (5 * 1.0, the coefficient for $\hat{\beta_2^*}$). But the increase in presidential approval (or re-equilibration, in error correction parlance) is not immediate, occurring over future time periods at a rate dictated by $\hat{\alpha_1}$. The largest portion of the movement in presidential approval will occur in the next time period, when 50% of the shift will occur. In the following time period ($t+1$), presidential approval will increase 2.5 points, increasing 1.25 points at $t+2$ and .63 points in $t+3$ and so on, until presidential approval has increased five points. Thus, the economy has two effects on presidential approval: one that occurs immediately and another impact dispersed across future time periods.

So one distinct advantage of the Bardsen ECM is that the long term multiplier is easily calculated in this model. But, as with the GECM and the ADL, the standard error for the long run multiplier is not directly estimated. But since the long run multiplier is the ratio of two coefficients ($\frac{\beta_2^*}{\alpha_1^*}$) the standard error is easier to derive in this form than in either the GECM or ADL form. In particular, we know that the variance for the long run multiplier is given by the formula for the approximation of the variance of a ratio of coefficients with known variances in this case:

$$\text{Var}(a/b) = (1/b^2)\text{Var}(a) + (a^2/b^4)\text{Var}(b) - 2(a/b^3)\text{Cov}(a,b) \tag{21}$$

So with the above formula, we can calculate the standard error for the long run multiplier. However, we can directly estimate the long-run multiplier and its standard error using a transformation first proposed by Bewley (1979).

Beginning with the ADL(1,1;1) model, we first subtract $\alpha_1 Y_t$ from both sides:

$$Y_t(1-\alpha_1) = \alpha_0 - \alpha_1 \Delta Y_t + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t.$$

Let $\eta = (1-\alpha_1)^{-1}$ and add and subtract $\beta_1 X_t$ to get:

$$Y_t = \eta\alpha_0 - \eta\alpha_1 \Delta Y_t + \eta(\beta_0 + \beta_1)X_t - \eta\beta_1 \Delta X_t + \eta\varepsilon_t. \tag{22}$$

Substituting and rewriting, we estimate the following regression:

$$Y_t = \phi_0 - \phi_1 \Delta Y_t + \psi_0 X_t - \psi_1 \Delta X_t + \mu_t \tag{23}$$

where $\phi_0 = \eta\alpha_0$, $\phi_1 = \eta\alpha_1$, $\psi_0 = \eta(\beta_0 + \beta_1)$, $\psi_1 = \eta\beta_1$, and $\mu = \eta\varepsilon_t$.

Estimation of this model is slightly less straightforward, and it is not strictly speaking an error correction model, although it is still an equally general transformation of the ADL and GECM. The complications arise from the fact that inclusion of ΔY_t means that a contemporaneous value of Y_t is on the right hand side of the model and the estimates will be biased if OLS is used to estimate the model. This specification requires analysts to use instrumental variables for consistent estimation (Bewley 1979). Instruments of a constant X_t, X_{t-1}, and Y_{t-1} should be used to estimate the model. In spite of this added step, the Bewley transformation is appealing because the long-run multiplier is estimated directly as the coefficient on X_t in this specification: $\psi_0 = \frac{B(1)}{A(1)} = \eta(\beta_0 + \beta_1) = k_1$ as $\eta = A^{-1}(1)$. Further, because it is estimated directly, we are provided with an estimate of the variance associated with the long-run multiplier. Of the ADL, the two other error correction specifications, and this transformation, this is the only specification that gives the variance associated with the long-run multiplier, k_1, directly. It is this direct estimate of the long-run multiplier and its variance that makes the Bewley transformation particularly useful.[13]

Two things should be noted about this variance estimate. First, it is an approximation to the variance since instruments are used to estimate the model. But, second, it can be shown that this estimate is equivalent to that in equation 21 (Hendry 1995). Analysts interested in long run behavior thus have a way to estimate not only the total long run effect but also the precision of that estimate.

To better demonstrate the equivalence among the ADL, the two forms of ECMs, and the Bewley transformation, and to show how each of the quantities discussed above can be calculated from each specification, we next use simulated data to estimate each

[13]The individual short-run effects of a unit change in X_t are more tedious to calculate. They are given by $\eta^{-1}(\psi_1 - \psi_0) = \beta_1 - \beta_0 + \beta_1 = \beta_0$ and $\eta^{-1}\psi_1 = \beta_1$.

model and calculate the short and long run effects of each. The exercise underscores how each model recovers the same estimates with stationary data.

3.1 An Example of Model Equivalence

While we have demonstrated analytically that the ADL model, GECM, Bardsen ECM and Bewley transformation all estimate the same quantities in different forms, an example brings this point into sharper focus. For the example, we use simulated data. The data generation process (DGP) for Y_t is the ADL(1,1) model and the DGP for X_t is a simple autoregressive process:

$$\begin{aligned} Y_t &= \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_{1t} \\ X_t &= \rho X_{t-1} + \varepsilon_{2t} \end{aligned} \tag{24}$$

where ρ is 0.75 making the model clearly stationary. We set the parameter values for the Y_t DGP as follows: $\alpha_0 = 0$, $\alpha_1 = 0.75$, $\beta_0 = 0.50$, and $\beta_1 = 0.25$. We, then, estimated an ADL model, a GECM, the Bardsen ECM, and a Bewley model. The results appear in Table 1.

First, let us examine the short run effects of X_t on Y_t. For the ADL model in column 1, these are given explicitly by the estimated coefficients β_0 and β_1—the coefficients on X_t and X_{t-1}. These are estimated as 0.53 and 0.25 respectively. For the GECM and ECM, λ_1 and η_1—the coefficients on ΔX_t for each model—are both equal to β_0, which is what we find as both of these estimated coefficients are 0.53 (notice that the estimates of the standard errors are also the same). For the second short run effect, the comparison is less obvious. In the ADL, $\hat{\beta}_1 = 0.25$. To calculate this effect in the GECM, we use $\lambda_2 - \lambda_1 - \gamma$ (the coefficients on $X_{t-1} - \Delta X_t - Y_{t-1} = 0.52$ - 0.53 - (-0.25)), which equals 0.24. Finally, for the ECM we just use $\beta_2^* - \beta_1^*$ ($X_{t-1} - \Delta X_t = 0.77 - 0.53$), again this is 0.24. The only difference in these estimates is due to rounding.

Table 1: ADL, GECM, Bardsen ECM, and Bewley Transformation Estimates

	ADL Model	GECM	Bardsen ECM	Bewley Transformation
Y_{t-1}	0.75	–	-0.25	–
	(0.02)		(0.02)	
Error Correction Term	–	-0.25		–
		(0.02)		
X_t	0.53	–	–	3.06
	(0.06)			(0.18)
X_{t-1}	0.25	0.52	0.77	–
	(0.07)	(0.05)	(0.07)	
ΔX_t	–	0.53	0.53	-0.97
		(0.06)	(0.06)	(0.25)
ΔY_t	–	–	–	-2.96
				(0.35)
N	249	249	249	249
Adj. R^2	.94	.45	.45	.18

Note: Simulated Data.

We can also calculate the same values for the long-run multiplier across all the models. Recall that for the ADL model, the long-run multiplier, k_1, is $\frac{(\beta_1+\beta_0)}{(1-\alpha_1)}$. The true value for the generated data is $\frac{0.50+0.25}{0.25} = 3.00$. Using the values from the estimated model, we find that $\hat{k_1}$ is $\frac{0.53+0.25}{0.75} = 3.12$ for the ADL. In the GECM, the long-run multiplier is given by $\frac{(\lambda_2-\gamma)}{-\gamma}$ or $\frac{0.52-(-0.25)}{0.25} = 3.08$. And this value is easily calculated in the ECM where it is simply $\frac{\beta_2^*}{-\alpha_1^*} = \frac{0.77}{-(-0.25)} = 3.08$. Again, there are minor differences due to rounding. The difference between the estimated and true long run multipliers is caused by sampling error in the simulation. Finally, we see the estimate for ψ_0, from the Bewley model, is 3.06, but now the estimate has a standard error.

We can also calculate the mean and median lag length from each specification. The median lag length in the generated data is 2. To calculate this quantity recall the formula for the median lag length:

$$m = \frac{\sum_{r=0}^{R} \omega_r}{\sum_{r=0}^{\infty} \omega_r}. \tag{25}$$

We've already calculated the denominator, the total or long run effect above: $k_1 = \frac{(\beta_1+\beta_0)}{(1-\alpha_1)} \frac{.50+.25}{.25} = 3.0$. Now we need to calculate the numerator for r from 0 to R, stopping at the value of R for which we have accounted for half of the total effect. The easiest way to think of this problem is by substituting into the ADL and assuming equilibrium conditions except for a single unit shock in X_t so that in the first period the response in Y_t is given by:

$$Y_t = 0.53X_t. \tag{26}$$

Normalizing as a proportion of the total or long run effect we have 0.53/3.0 or .167 = 16.7% of the effect for $R = 0$. Continuing, the effect of that same shock one period later is given by:

$$\begin{aligned} Y_{t+1} &= 0.75Y_t + 0.25X_t && (27) \\ &= 0.75(0.503) + 0.25 && (28) \\ &= 0.625. && (29) \end{aligned}$$

We again divide by k_1 to normalize: $0.625/3.0 = 0.208 = 20.8\%$. Now adding this effect, the cumulative portion of the effect expended is 0.208 + 0.167 or 37.5% of the effect after $R = 1$ lags. Continuing, the formula is simpler as no additional short term effects of X_t enter the model.

$$\begin{aligned} Y_{t+2} &= 0.75Y_{t+1} && (30) \\ &= 0.75(0.625) && (31) \\ &= 0.469 && (32) \end{aligned}$$

$0.469/3.0 = 0.156$. Adding, we exceed one half: $0.156+0.208+0.167 = 53.1\%$.[14]

[14] We can write this sequence of lag values in terms of the ω_r as well:

$$\begin{aligned} \omega_0 &= \beta_0 X_0 = .50 \\ \omega_1 &= \beta_1 X_0 + \alpha Y_0 = .25 + .75(.50) = .625 \\ \omega_2 &= \alpha Y_1 = .75(.625) = .46875 \\ &\vdots \end{aligned}$$

A couple of additional points are worth making about the median lag length. First, it is almost always the case the median lags, and other quantities like it, will need to be calculated by hand. As we noted above, it is not always the case that the median (or mean) will be of theoretical interest, but it is often the case that we care about some aspect of the distribution of the decay of effects and characterizing that distribution in some form is worth doing. While there are other ways to calculate medians or other features of the distribution, substitution is often the most intuitive way to understand the patterns of decay. Second, we can begin with the sum of the short run effects from any of the other general models and and forms of the long run multipliers from these models as well and we will produce the same estimated median lag length. Third, repeating the above procedure for the estimated parameter values for the simulated data produces the same estimated median lag length as from the theoretical DGP.

The mean lag length for the simulated data can also be calculated readily from the ADL representation. Recalling that the general formula for the mean lag is given by the partial derivative of ω with respect to L:

$$\mu = \frac{\omega'}{\omega} = \frac{B(1)'}{B(1)} - \frac{A(1)'}{A(1)}, \tag{33}$$

for the ADL(1,1;1) case we consider, we can write:

$$\mu = \frac{\beta_1}{\beta_0 + \beta_1} - \frac{-\alpha}{1-\alpha} = \frac{.25}{.50 + .25} + \frac{.75}{1 - .75} = 3\frac{1}{3}. \tag{34}$$

It is particularly easy to see the effect of changes in the dynamic parameter α on the mean lag length; bigger α, produce bigger mean lag lengths. The mean lag length is relatively short in our example, although not overly so.

Most analysts seldom draw inferences about either the equilibrium relationship, the long-run multiplier, or the rate of error correction, or about features of the lag distribution, yet often these quantities are not only of interest, but of central importance for understanding politics. Each of the general models we've discussed will provide estimates of these quantities and it is important to consider each specification when selecting a dynamic model. In spite of their ready availability, however, these quantities are typically underdeveloped and are often omitted altogether from discussions of results. In addition, however, analysts also estimate models that place restrictions on these quantities, restrictions that are theoretically and empirically invalid.

3.2 Restricting ADL and EC Models

Analysts often estimate and present models that are restricted versions of the ADL and ECM's. Instead of starting with a general model and then imposing restrictions, the starting point is a specific model that imposes a distinct functional form on the dynamics. Here, we consider some examples of this practice and examine the consequences.

A commonly estimated model is the lagged dependent variable model known in economics as the partial adjustment (PA) model. In this model, β_1 from the ADL(1,1:1) is assumed to be zero. The theoretical justification for the model is typically that individuals or governments pursue target values of Y given current values of X but that changing Y is costly so that immediate adjustment—change—in Y is slow or partial.[15] The appropriateness of this theoretical story aside, this statistical model is encompassed by the ADL and ECM. But since a more general model is rarely presented in conjunction with the PA model, one does not know whether the restriction has been imposed *a priori* or whether it has been tested and supported by the data. Similar issues arise when analysts estimate finite distributed lag models or include lagged X but omit contemporaneous X terms.[16] Too little attention has been given to the consequences of this practice. We focus on the restriction most commonly estimated

[15]Note that such logic itself implies that with a long enough time horizon and given certain patterns in X, imbalance in X and Y is virtually guaranteed so that the model will make no sense empirically or theoretically. This is easily seen when the model is written in error correction form.

[16]Beginning analysis with the ECM gives the analyst an early indication on testing the appropriateness of the FDL restriction on the general model. If one estimates an ECM and finds $\gamma = -1$ then dropping lagged Y from the general model in ADL form and estimating an FDL is appropriate.

in applied work, that $\beta_1 = 0$ in the ADL model.[17]. We ask: what are the consequences if the partial adjustment model is used when the ADL(1,1;n) actually encompasses the DGP?

We begin by showing the consequences for the estimated short run effect in the simple bivariate case. Because OLS is used to estimate both models, the analytical derivations are relatively simple. The model that actually encompasses the DGP is:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t. \tag{35}$$

but we estimate:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + e_t. \tag{36}$$

now the estimate for β_0 is $\mathrm{E}(\beta_0) = \beta_0 + \mathrm{b}_1\beta_1$ where b_1 equals:

$$\mathrm{b}_1 = \frac{\sum(X_t - \bar{X}_t)(X_{t-1} - \bar{X}_{t-1})}{\sum(X_t - \bar{X}_t)^2} \tag{37}$$

The degree of bias obviously depends on the correlation between X_t and X_{t-1}, which will typically be quite large for a time series. There will be the same bias in the estimate of α_1 depending on the covariance between X_{t-1} and Y_{t-1}. The partial adjustment model implies the following long run multiplier: $\frac{\beta_1}{(1-\alpha_1)}$. This long run multiplier will be biased as well, with the size of the bias depending both on $\beta_1 - \mathrm{b}_1$ and the bias in α_1. The partial adjustment model also constrains the mean and median lag lengths to values different from those in the ADL. Unfortunately performing diagnostic tests in this case can also be misleading.

There will be similar consequences if any of a wide number of more restricted models are estimated when the more general model describes the DGP. For example, if we estimate a static model in its ADL form we regress Y_t on X_t. In terms of the

[17]In the ECM the restriction is $\lambda_2 - \lambda_1 - \gamma = 0$

ADL, the hypothesis under the restriction is that the joint restriction $\alpha_1 = \beta_1 = 0$ is valid.[18]

The consequences of estimating the static model when the ADL is the true model are well known: β_0 will neither capture the long nor the short-run multiplier and calculations of the long-run equilibrium will be biased downward, unless X_t is a unit root process or $\beta_0\alpha_1 + \beta_1 = 0$. The static model constrains the median and mean lag lengths to be zero so that the bias in β_0 will be greater when the true mean and median lag lengths are longer. The errors of the equation will be autocorrelated. If this autocorrelation is positive, standard errors will be too small. Finally, testing for the constancy of the parameters of the static model results in rejection of the null of constant parameters too often, providing misleading information about how to improve the specification. One might object that few would be foolish enough to estimate such a model with time series data. But the static model is estimated when GLS estimators such as Prais-Winsten are used or OLS with Newey-West standard errors.

To give the reader a sense of how using a restricted model affects our understanding, we plot the standardized lag distribution for four models in Figure 3.2. The first lag distribution is that of the ADL from section 3.1 where $\beta_0 = .50$, $\beta_1 = .25$, and $\alpha = .75$. The next plot is the lag distribution estimating a static regression with the same data. The third plot is the lag distribution for a partial adjustment model, again when the true data generating process follows equation 24. The final plot is for an ADL model, where we increased the coefficient on Y_{t-1} to 0.90. The fourth plot underscores how much more of the effect feeds into the future as the autoregressive nature of Y_t changes. The static model produces the most distorted lag distribution, while we see the pattern of geometric decay that is imposed under the partial adjustment model. The distortion we observe in the lag distribution due to using an overly restrictive model will be felt across most any other estimate from the models.

[18] In terms of the GECM, the hypothesis for the static model is the joint hypothesis that $\gamma = -1$ and $\lambda_2 = \lambda_1 - 1$ when estimated in the familiar static regression in ADL form, $y_t = \alpha_0 + \beta_0 X_t + \varepsilon_t$, and in the less familiar, but equivalent ECM form, $\Delta Y_t = \alpha_0 - \gamma Y_{t-1} + \lambda_1 \Delta X_t + \lambda_2 X_{t-1} + \varepsilon_t$.

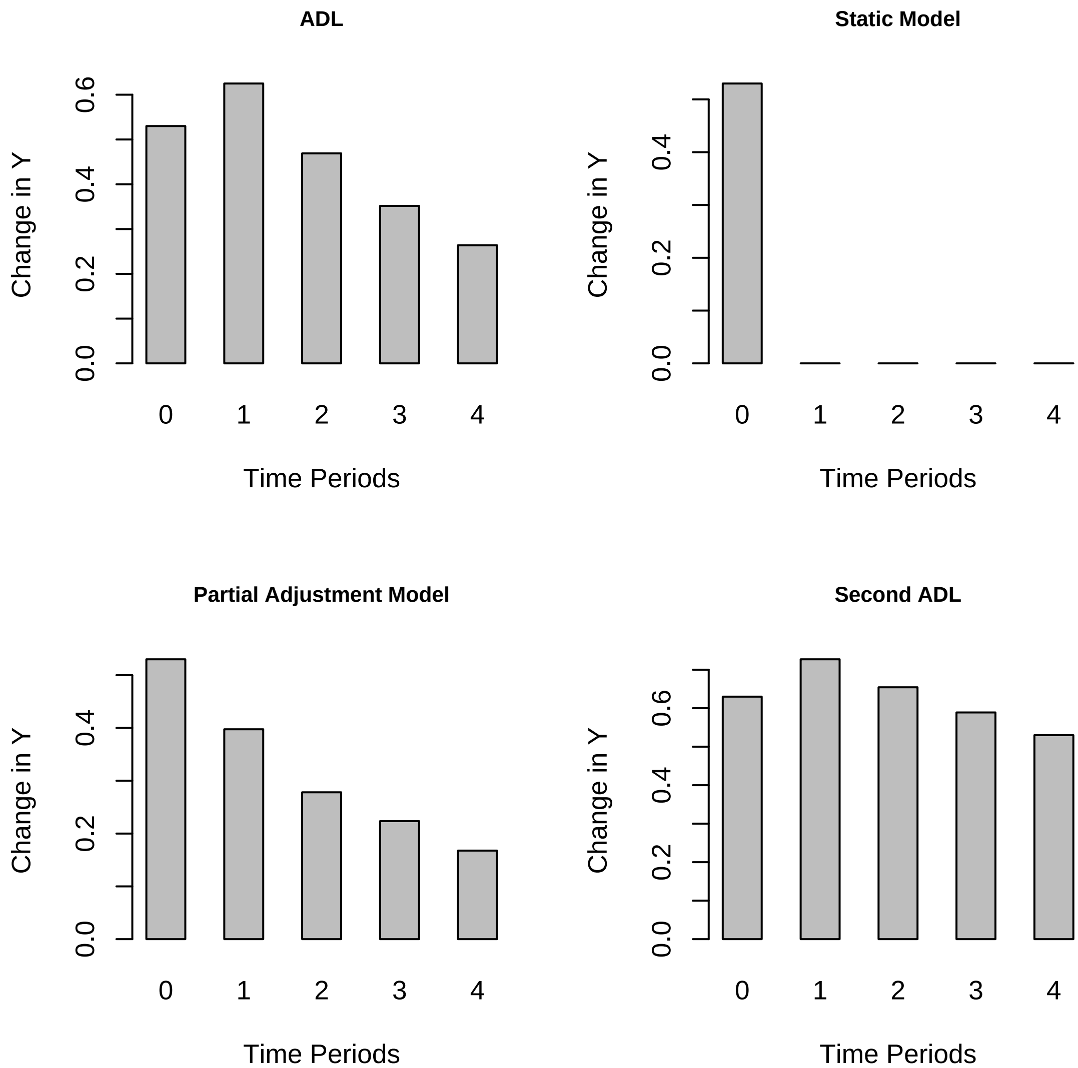

Figure 1: Simulated Lag Distributions: The Effect of Restrictions

We could continue in this vein for other time series specifications and examine the consequences of estimating each. Instead in Table 2, we list a number of restrictive specifications and show what they imply for both the ADL and ECM. This should serve as a useful references for applied analysts. If they feel one of the more restrictive models listed here is the true model, they can estimate an unrestricted ADL or ECM and then test the appropriate restriction instead of starting with the more restrictive specification, keeping in mind the features of the restricted forms of those models.

Table 2: Restrictions Of The General Model

Type	ADL Model	Restriction	Features
General	$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t$	None	None
PA[a]	$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \varepsilon_t$	$\beta_1 = 0$	None
Static	$Y_t = \alpha_0 + \beta_0 X_t + \varepsilon_t$	$\alpha_1 = \beta_1 = 0$	$k_1 = \beta_0$
FDL[b]	$Y_t = \alpha_0 + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t$	$\alpha_1 = 0$	$k_1 = \sum_{j=1}^{n} \sum_{i=0}^{q-1} \beta_{ji}$
Differences	$\Delta Y_t = \alpha_0 + \beta_0 \Delta X_t + \varepsilon_t$	$\alpha_1 = 1$, $\beta_0 = -\beta_1$	Infinite mean lag length
Dead Start	$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_1 X_{t-1} + \varepsilon_t$	$\beta_0 = 0$	None
Common Factor	$Y_t = \beta_0 X_t + \varepsilon_t$, $\varepsilon_t = \beta_1 \varepsilon_{t-1} + u_t$	$\beta_1 = -\beta_0 \alpha_1$	$k_1 = \beta_0$, $\mu = 0$, ec rate 100%
Type	GECM Model	Restriction	Features
General	$\Delta Y_t = \alpha_0 + \gamma(Y_{t-1} - X_{t-1}) + \lambda_1 \Delta X_t + \lambda_2 X_{t-1} + \varepsilon_t$	None	None
PA[a]	$\Delta Y_t = \alpha_0 + \gamma(Y_{t-1} - X_{t-1}) + \beta_0 X_t + \varepsilon_t$	$\lambda_2 - \lambda_1 = \gamma = 0$	None
Static	$\Delta Y_t = \alpha_0 + \gamma Y_{t-1} + \lambda_1 \Delta X_t + \lambda_2 x_{t-1} + \varepsilon_t$	$\gamma = -1, \lambda_2 = \lambda_1 - 1$	$k_1 = \lambda_1$
FDL[b]	$\Delta Y_t = \alpha_0 + \gamma Y_{t-1} + \lambda_1 \Delta X_t + \lambda_2 x X t - 2 + \varepsilon_t$	$\gamma = -1$	$k_1 = \lambda_2 + 1$, EC rate 100%
Differences	$\Delta Y_t = \alpha_0 + \lambda_1 \Delta X_t + \varepsilon_t$	$\gamma = 0$, $\lambda_2 = 0$	Infinite mean lag length
Dead Start	$\Delta Y_t = \alpha_0 + \gamma(Y_{t-1} - X_{t-1}) + \lambda_2 X_{t-1} + \varepsilon_t$	$\lambda_1 = 0$	None
Common Factor	$\Delta Y_t = \alpha_0 + \lambda_1 \Delta X_t + \varepsilon_t$, $\varepsilon_t = \lambda_2 \varepsilon_{t-1} + u_t$	$\lambda_2 = 2\gamma + 1$	$\mu = 0$, EC rate 100%

[a] Partial adjustment

[b] Finite distributed lag

Regardless which restrictive model is estimated, what is important is that restricting the ADL has consequences that extend beyond the restricted coefficients. If the restrictions are invalid this will lead to specification errors and bias in *virtually all the dynamic quantities that we care about.* That is, the mean and median lag lengths, error correction rate, and long run multipliers as well as any other estimated or derived quantities may be biased.

Unfortunately, post-estimation diagnostic tests will tend to be of little help in constructive model building when analysts begin with models that impose untenable restrictions on the data. But failing a specification test does not mean we've found *the*, or even *a*, problem with a given model specification. Consider an example. We can't be certain that a finding of autocorrelation in the residuals of an estimated partial adjustment model means that the underlying data generating process is a PA model containing autocorrelation. In fact using any single diagnostic test to build up a dynamic specification from a restricted (here the PA model) to a general model (the ADL or ECM model) is dangerous. Diagnostic information is symptomatic of some ill only, but not necessarily of the particular ill or its cause. The autocorrelation may mask a functional form misspecification rather than serial correlation in the DGP, for exmample. Even an absence of assumption violations may merely mask other problems. In short, starting with a restricted model like the PA model can lead us down a road in which rejections of the null can mean many things—misspecification of any number of kinds or simply data generated under the alternative hypothesis, but we cannot know because the tests are conditional on the model.

The dangers associated with adopting restricted dynamic specifications further emphasize the need for analysts to consider a variety of general models, to begin analysis with one of these models, to test restrictions on it, and to interpret, interpret, interpret, the abundant amount of information that is available to us as analysts. Given the ease of using a general model—be it the ADL, the ECM, or other linear forms—there is really no reason to ever start with a more restrictive specification. Moreover given the dangers of estimating a restricted model, analysts should present the results from the general model even if later they settle on a more restrictive specification.

3.3 Selecting a Model: Advantages and Disadvantages of Alternatives

One might, at this point, ask is there any reason to chose the ADL over the ECM (or the reverse)? Both are equally general and contain the same information. Both fit the data equally well, provided that the model encapsulates the process generating the data. And both are easily estimated with most any statistical software. The only distinction presented thus far is that differing quantities are directly estimated in each model. In the ADL, for example, both short term effects are directly estimated, while in the ECM only one is, but the error correction rate is directly estimated. Additionally, however, each of the models has its own unique advantages, which we now review.

The ADL's greatest advantage may be the familiarity analysts tend to have with the model given how close it is to the partial adjustment model. Additionally, the ADL provides estimates of short-run effects (and their standard errors) in a transparent fashion. This gives us an explicit indicator of the "stickiness" of the process we care about.

ECMs, on the other hand, allow for a tighter link between theory and model. The behavioral story of error correction is broadly appealing: two (or more) processes are tied together in the long run such that when one process increases (or decreases), the other must adjust to maintain this long run equilibrium relationship. This error correction process is one that would appear to apply to many political relationships and EC models estimate the rate of error correction (and their standard errors) directly.

ECMs offer other advantages. In contrast to the ADL where the regressors are often highly colinear, the regressors in the ECMs—ΔX_t and X_{t-1} and $(Y_{t-1} - X_{t-1})$—are unlikely to be highly correlated. As a result, ECMs tend to give more precise estimates of the model's parameters. More precise estimates in turn provide for greater confidence about inferences. This low correlation among regressors in the ECM also means that the diagnosis of misspecification that may arise when testing restrictions on the general model in its ADL form can be avoided using an ECM. Some forms of the general model can be understood as restrictions on the ECM more easily than on the ADL, also aiding in testing restrictions on dynamic behavior. Finally, two of the EC models parameterize the dependent variable as a change in Y_t—both Bardsen and the GECM—helping us to avoid spurious regressions due to strongly autoregressive (or near-integrated data) or unit roots in cases where that concern arises (De Boef & Granato 1997, De Boef 2001).

If the long-run multiplier is of particular interest, the Bewley transformation is particularly attractive, providing a direct estimate and an associated standard error, although it does require instrumental variables regression. While OLS is certainly simpler to implement, the ready availability of instruments makes the cost usually associated with IV estimation low, and worth paying if knowledge of the precision of the long run multiplier is integral to theory or the central hypothesis of the analysis. On the downside, however, short-run effects are more tedious to derive from this specification and come without standard errors.

At this point it should be clear that each of these specifications has something to recommend it. Which model is the right one? The answer depends on the specificity of the theory. In situations where the primary concern is about short run behavior, then the ADL is most appropriate. But when one is concerned about both short and long run effects, or theory dictates equilibrium behavior then the ECM is best. And when we are particularly interested in the long run multiplier, then the Bewley model is an obvious choice. And while we don't recommend fishing expeditions for models, when theory is general it may be best to estimate a combination of models.

In the next section, we re-estimate a model of Supreme Court approval based on (Durr, Martin & Wolbrecht 1993). The original research uses a partial adjustment model that has imposed some restrictions. While the original research was estimated with standard techniques for stationary data that are technically correct, we apply our strategy for dynamic specification and show how consideration of different dynamic specifications can help us to gain a richer understanding of politics.

4 An Example (or two)

We consider a time series model from the applied literature on the Supreme Court to demonstrate how using an error correction model changes our understanding of political dynamics over and above using standard techniques for stationary data. The model is from (Durr, Martin & Wolbrecht 1993). The authors argue that public support for the Supreme Court over time is a function of three factors: (1) how far the ideological position of the Court diverges from the ideological position of the public, (2) countermajoritarian behavior by the Court, and (3) general evaluations of government. Their dependent variable is a semi-annual time series measure of public support for the Supreme Court built from a variety of survey items. They find in their tests that the series is stationary, as we would expect.[19]

The authors contend that support for the Supreme Court is a dynamic process, where attitudes toward the Supreme Court are a function of past attitudes toward the Supreme Court as modified by new information on ideological divergence, countermajoritarian behavior, and general evaluations of government. As such, they expect past shocks to feed forward into the future, decaying at an exponential rate, such that shocks from the last period will matter half as much in the present time period, and so on. To capture these dynamics, the authors use a partial adjustment model. So like most analysts, they start with a restrictive rather than a general specification.

We estimate a more parsimonious version of the model, dropping presidential approval and the measures of countermajoritarian behavior, which were insignificant in all specifications. The omission of these variables has no effect on the results of either

[19]It seems unlikely that the Dred Scott case still influences public opinion about the Supreme Court, but it is not unreasonable to think that *Roe v. Wade* still matters, implying the presence of long-term (but not permanent) memory in support for the Supreme Court. The properties of this time series are similar to many in applied work: surely not integrated, but likely to be strongly autoregressive.

Table 3: Model Comparison of Public Support for the Supreme Court

	Supreme Court Support_t PA	Supreme Court Support_t ADL	ΔSupreme Court Support_t ECM	Supreme Court Support_t Bewley
Supreme Court Support_{t-1}	0.24 (0.14)	0.28 (0.16)	$-0.72**$ (0.16)	–
Δ Supreme Court Support_t	–	–	–	-0.38 (0.30)
Ideological Divergence_t	$-5.48*$ (2.58)	-5.97 (2.77)	–	-5.48 (2.58)
Congressional Approval_t	$0.47**$ (0.13)	$0.46*$ (0.21)	–	$0.62**$ (0.17)
ΔIdeological Divergence_t	–	–	$-5.97*$ (2.77)	-2.49 (4.29)
Ideological Divergence_{t-1}	–	1.80 (2.95)	-4.17 (3.39)	–
ΔCongressional Approval_t	–	–	$0.46*$ (0.21)	0.20 (0.32)
Congressional Approval_{t-1}	–	-0.14 (0.23)	$0.44**$ (0.15)	–
Constant	$41.63**$ (11.74)	$40.25*$ (12.24)	$40.25**$ (12.24)	$55.70**$ (11.05)
N	41	41	41	41
Adjusted R^2	0.47	0.40	0.43	–
Box-Ljung Q Test	8.20	9.54	9.54	9.54
χ^2 p-value	0.97	0.95	0.95	0.95

Note: OLS Estimates. Standard Errors in Parentheses.
Data are semi-annual, 1973:1 to 1993:2
Two tailed tests.
* $p < .05$
** $p < .01$

the original model or any of the new models. We estimate the partial adjustment model, the ADL, a Bardsen ECM, and a Bewley model. In the first column of Table 3, we include a model with the original dynamic specification as estimated by the authors (the PA model). Again, we see that both ideological divergence and Congressional approval are significant predictors, while past support for the Supreme Court is not.

We see one effect of using a restrictive model in the calculation of the median lags. In the original model, 49% of the total change in Y_t occurs immediately. Using the ADL to calculate the median lag, we find that 75% of the change in Y_t occurs immediately. So the estimated dynamic change in Y_t is much faster in the unrestricted model than in the restricted model.

In the fourth column, we estimated the model with the Bewley transformation. The parameters immediately highlighted here are the coefficients for ideological divergence and congressional approval. These values are the long run multipliers and give the total effect of an impulse change. Only the coefficient for congressional approval is statistically different from 0; we interpret it to mean that a unit change in congressional approval has an expected total long run effect on Supreme Court approval of near two-thirds of a percentage point. As the reader can see, the Bewley transformation allows the analyst to easily calculate the long run multiplier and test whether it is statistically different from zero.

With the error correction model, we find a larger effect for congressional approval. A one point increase in congressional approval moves support for the Supreme Court up 0.46 points immediately, with another 0.62 point increase over future time periods for a total effect of 1.08 points.In the authors' model, the total effect of a one point increase in Congressional approval increases support for the Supreme Court about 0.62, a difference of nearly 0.50 points. The reason for this is that under the partial adjustment model, the short run effect gets lumped into the long run multiplier, so we can't distinguish between the two types of effects. The effect, then, is over 1/3 larger than before. Moreover, the ECM is attractive for as a behavioral story. It is entirely reasonable to assume that approval for the Supreme Court and Congressional approval, as indicators of general government approval have an equilibrium. Overall, we see from this example that imposing a more restrictive specification causes us to miss information that from the data that is easily extracted with a few extra steps.

4.1 Dynamic Specifications of Approval

Presidential approval was one of the first areas of study in which political scientist utilized time series data and its development parallels the advances in time series econometrics. Early work was conducted before the dangers of autocorrelated errors had reached political science, with attention to autocorrelated errors came GLS estimation, often in combination with distributed lags (Frey & Schneider 1978, Golden & Poterba 1980, Chappell 1990). Early on Hibbs (1982) paid detailed attention to theoretical dynamics, specifying approval as a function of cumulative and exponentially weighted conditions under a given party. Since then Ostrom and Smith (1992) have argued that approval is a unit root process cointegrated with economic conditions. They estimated the long-run equilibrium with a (2 step) error correction model. Norpoth and Yantek based their modeling decisions solely on empirical evidence using Box Jenkins methodology (Norpoth & Yantek 1983). Rounding out the class of time series regressions, static specifications (in the context of data pooled by administration) (Brace & Hinckley 1992) and specifications in first differences (Lanoue 1987) have been estimated as well. Beck (1991) compared a variety of these alternative specifications for approval including static relationships, first differences, exponentially distributed lags, partial adjustment, error correction, and transfer functions. Wood (2000) allows for time varying relationships in the context of flexible least squares.

Controversy continues to characterize the time series research on presidential approval. Mackuen, Erikson, and Stimson (1992) took the research on presidential approval in a different direction by arguing that the theory of rational expectations predicts that sociotropic prospective evaluations of the economy should be the only aspect of economic evaluations that matter. They found such evidence with a partial adjustment model. Norpoth (1996) used Box Jenkins methodology and found no effect for prospective evaluations, while Clarke and Stewart (1994) used an error correction model to find a mix of effects across retrospective and prospective sociotropic evaluations. Again much of the controversy revolved around the appropriate model.

While we would agree with some aspects of the modeling approach adopted by Clarke and Stewart (1994), they argue, as most analysts in political science feel compelled to do when using an ECM, that presidential approval and sociotropic evaluations are cointegrated. Of course, they admit that presidential approval is not integrated, but do find some limited evidence for cointegration between presidential approval and one of the economic time series. It is our position that it is absurd to argue that presidential approval is cointegrated with economic evaluations no matter what a test may suggest. How can a series that is made up of ratings for different presidents have a permanent memory? Can we really argue that the ratings of John F. Kennedy still fully affect those of George W. Bush? Moreover, by limiting themselves to only modeling cointegration between presidential approval and one of the economic time series, they use an overly restrictive specification. They allow only one of the economic variables to have a long run effect, a decision that makes little sense theoretically.

Instead, as we have demonstrated, an analyst can readily adopt the behavioral story of error correction without the baggage of cointegration. Error correction seems particularly appropriate in the context here. That is we should fully expect that economic evaluations, be they retrospective or prospective, to have both short and long run effects. As new economic information becomes available, it should immediately translate into a new evaluation of the president. But it would also be reasonable to assume that the economic performance and presidential approval have an equilibrium and the two will move together in the long run.

We re-estimate the Clarke and Stewart model, testing among retrospective, prospective egocentric and sociotropic economic evaluations as explanations for presidential approval evaluations. However, we allow all of the economic variables to have long run effects. We control as they do for events, the honeymoon periods, the Vietnam conflict, and for changes in political administrations for the period from 1960 to 2003.

Since we make no claim of cointegration we need not restrict only some of the economic variables to have both short and long-run effects. We also use a Bewley model to test whether any of the long run effects are bounded away from zero. Table 4 presents the results for both an ECM and a Bewley model.

The evidence, here, supports (MacKuen, Erikson & Stimson 1992), as sociotropic prospective evaluations are the only economic indicator that has a significant effect on presidential approval. As we might expect, such evaluations have both short and long run effects. A five point increase in forward looking economic optimism will increase presidential approval 0.60 points. The long run effect is much larger, however. This five point increase will increase presidential approval nearly three points. Over 90% of the reequilibration will take place over six quarters.

Without having to rely on evidence of cointegration, we can fully test the short and long run effects between all aspects of economic evaluations. Moreover, since we use a general model, we need not worry about imposing a model that is too restrictive as MacKuen, Erikson and Stimson and Clarke and Stewart did.

5 Conclusion

Many times political time series analysts are attracted to the behavioral story associated with the error correction model, but assuming ECMs were appropriate only when time series were cointegrated and either of the opinion that it is quite rare to find true unit roots in political science or fearful of reviewers who are suspicious of inferences that unit roots characterize our data, they dismissed it as a potential dynamic specification. Indeed such fears are well founded. We are of the opinion that while political time series may often look like unit roots, given their theoretical properties, it is highly unlikely they are unit roots. Analysts often forget that the presence of a unit root in economic data is more than just an empirical property; it also implies a theory that is rarely applicable in political science (Beck 1992).

Table 4: Error Correction Models for Presidential Approval

	Δ Presidential Approval$_t$	Presidential Approval$_t$
Pres. Approval$_{t-1}$	$-0.30**$	–
	(0.05)	
ΔPres. Approval$_t$	–	$-2.32**$
		(0.59)
Sociotropic Prospective$_t$	–	$0.59**$
		(0.19)
ΔSociotropic Prospective$_t$	$0.12*$	-0.19
	(0.06)	(0.19)
Sociotropic Prospective$_{t-1}$	$0.18**$	–
	(0.05)	
Egocentric Prospective$_t$	–	-0.30
		(0.43)
ΔEgocentric Prospective$_t$	0.07	0.54
	(0.10)	(0.35)
Egocentric Prospective$_{t-1}$	-0.09	–
	(0.13)	
Sociotropic Retrospective$_t$	–	-0.08
		(0.07)
ΔSociotropic Retrospective$_t$	0.01	0.11
	(0.03)	(0.10)
Sociotropic Retrospective$_{t-1}$	-0.02	–
	(0.02)	
Egocentric Retrospective$_t$	–	0.08
		(0.25)
ΔEgocentric Retrospective$_t$	-0.07	-0.31
	(0.08)	(0.24)
Egocentric Retrospective$_{t-1}$	0.03	–
	(0.08)	
N	41	41
Adjusted R^2	0.33	–
LM Test	34.73	34.73
χ^2 p-value	0.71	0.71

Note: OLS Estimates. Standard Errors in Parentheses.
Data are quarterly, 1960:1 to 2003:2
Controls omitted from table.
Two tailed tests.
* $p < .05$
** $p < .01$

While we do not often work with unit roots, we need not enter debates about unit roots and cointegration to discuss long run equilibria and rates of re-equilibration. As we have shown, these dynamic quantities are implied by virtually all dynamic regressions involving stationary data, although some specifications impose restrictions on these quantities. General models like the ADL, the GECM, the Bardsen ECM, and the Bewley transformation fit the behavioral story that attracted analysts to cointegration methods. We should be estimating these models. Each of these models tells the behavioral story with a different tune and beat. We should select the optimal dynamic specification from among one of these general models depending on our theory and the needs of the audience.

We recommend the Bardsen ECM, but as we discussed previously, theory must play a role. As we noted, the choice will depend on whether the theory suggests an emphasis on short versus long term. When the burden of proof is on the short run, the ADL will provide the estimates of most relevance. When the theory focuses on the long run, the Bewley model will estimate the quantity of interest. Of course, as is often the case, our theories are not that specific. And it for this reason that the Bardsen ECM is so attractive. With it one has both the short and long run estimates most readily available. The results from this model then may lead one to focus on either the short or the long run.

The more general point we wish to make is that applied analysts should not report the results from restricted models without presenting the general model and demonstrating that those restrictions are valid. There is no excuse for not estimating the more general model since they require no special software and use OLS. Moreover, whatever general is estimated, the analyst should fully interpret the model. By fully interpret, we mean discussing short and long run effects, mean and median lag lengths, and rates of error correction.

We often deny the role that statistical models play in the development of theory. But often, analyses that use linear models have theories that predict linear effects. While the functional form of a statistical model should not dictate how we formulate theory, it is inevitable that some such influence will occur. That seems to have been the case thus far in time series analysis. The influence has limited the development of theory. By expanding our repertoire of models, we will build better dynamic theories. Often we bemoan the fact that our models aren't sophisticated enough to match the real world. The new goal should be making our theories sophisticated enough to match our models.

References

Bannerjee, Anindya, Juan Dolado, John W. Galbraith & David F. Hendry. 1993. *Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data.* Oxford: Oxford University Press.

Bardsen, G. 1989. "The Estimation of Long-Run Coefficients from Error Correction Models." *Oxford Bulletin of Economics and Statistics* 54:225–55.

Beck, Nathaniel. 1992. "Comparing Dynamic Specifications: The Case of Presidential Approval." *Political Analysis* 3:27–50.

Beck, Nathaniel. 1993. "The Methodology of Cointegration." *Political Analysis* 4:237–248.

Beck, Neal. 1991. "Model Selection: Are Time Series Techniques Useful in Cross-Sectional Problems?" 999:Midwest Meeting.

Bewley, R. A. 1979. "The Direct Estimation of the Equilibrium Response in a Linear Model." *Economic Letters* 3:357–61.

Brace, Paul & Barbara Hinckley. 1992. "Follow the Leader: Opinion Polls and the Modern Presidents." *Journal of Politics* 1991:993–1017.

Chappell, Henry W. 1990. "Economic Performance, Voting, and Political Support." *Review of Economics and Statistics* 72:313–320.

Clarke, Harold D. & Marianne C. Stewart. 1994. "Prospections,Retrospections and Rationality: The 'Bankers' Model of Presidential Approval Reconsidered." *American Journal of Political Science* 38:1104–1123.

Davidson, Russell & James G. MacKinnon. 1993. *Estimation and Inference in Econometrics.* New York: Oxford University Press.

De Boef, Suzanna. 2001. "Testing for Cointegrating Relationships with Near-integrated Data." *Political Analysis* 9:78–94.

De Boef, Suzanna & Jim Granato. 1997. "Near-integrated Data and the Analysis of Political Relationship." *American Journal of Political Science* 41:619–640.

De Boef, Suzanna & Paul Kellstedt. 2004. "The Political (And Economic) Origins of Consumer Confidence." *American Journal of Political Science* Forthcoming.

Durr, Robert H. 1993*a*. "An Essay on Cointegration and Error Correction Models." *Political Analysis* 4:185–228.

Durr, Robert H. 1993*b*. "Of Forests and Trees." *Political Analysis* 4:255–258.

Durr, Robert H., Andrew D. Martin & Christina Wolbrecht. 1993. "Ideological Divergence and Public Support for the Supreme Court." *American Journal of Political Science* 44:768–776.

Frey, Bruno & Frederich Schneider. 1978. "An Empirical Study of Politico-Economic Interaction." *Review of Economics and Statistics* 60:174–183.

Golden, David G. & James M. Poterba. 1980. "The Price of Popularity: The Political Business Cycle Reexamined." *American Journal of Political Science* 24:696–714.

Hendry, David F. 1995. *Dynamic Econometrics.* Oxford: Oxford University Press.

Hibbs, Douglas A., with the assistance of Douglas Rivers & NicholasVasilatos. 1982. "The Dynamics of Political Support for American PresidentsAmong Occupational and Partisan Groups." *American Journal of Political Science* 26:312–323.

Keele, Luke J. & Nathan J. Kelly. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis* Forthcoming.

Lanoue, David J. 1987. "Economic Prosperity and Presidential Popularity: Sorting Out The Effects." *The Western Political Quarterly* 40:237–245.

MacKuen, Michael B., Robert S. Erikson & James A. Stimson. 1992. "Peasants or Bankers: The American Electorate and the U.S. Economy." *American Political Science Review* 86:597–611.

Norpoth, Helmut. 1996. "Politics and the Prospective Voter." *Journal of Politics* 58:776–792.

Norpoth, Helmut & Thom Yantek. 1983. "Macroeconomic Conditions and theFluctuations of Presidential Popularity: The Question of Lagged Effects." *American Journal of Political Science* 27:785–807.

Ostrom, Charles & Renee Smith. 1992. "Error Correction, Attitude Persistence, and Executive Rewards and Punishments: A Behavioral Theory of Presidential Approval." *Political Analysis* 4:127–83.

Smith, Renee. 1993. "Error Correction, Attractors, and Cointegration: Substantive and Methodological Issues." *Political Analysis* 4:249–254.

Williams, John T. 1993. "What Goes Around Comes Around: Unit Root Tests and Cointegration." *Political Analysis* 4:229–236.

Wood, B. Dan. 2000. "Weak Theories and Parameter Instability: Using Flexible Least Squares to Take Time Varying Relationships Seriously." *American Journal of Political Science* 44:603–618.

Appendix

A.1 The General EC Model

It is often useful to write the general model using lag polynomials:

$$A(L)Y_t = B(L)X_t + \varepsilon_t \tag{A.1}$$

where L is the lag operator: $L^i X_t = X_{t-i}$, $A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_p L^p$ and $B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \ldots + \beta_q L^q$.

More generally, the long-run equilibrium for the general model is given by:

$$y^* = \left(1 - \sum_{i=1}^{p} \alpha_i\right)^{-1} \left(\alpha_0 + \sum_{j=1}^{n}\sum_{i=0}^{q} \beta_{ij} x_j^*\right) \tag{A.2}$$

$$= \frac{\alpha_0}{(1 - \sum_{i=1}^{p} \alpha_i)} + \frac{\sum_{j=1}^{n}\sum_{i=0}^{q} \beta_{ij} x_j^*}{(1 - \sum_{i=1}^{p} \alpha_i)} \tag{A.3}$$

$$= k_0 + k_1 x_j^* \tag{A.4}$$

and the long-run multiplier for any given j is given by:

$$\frac{B_j(1)}{A(1)} \tag{A.5}$$

where $L = 1$ refers to the case where 1 replaces the lag operator so that the long-run multiplier is given by the sum of the coefficients on X_j, the short-run effects, divided by 1 minus the sum of the coefficients on the lagged Y_t.

A.2 The General Bardsen Model

The Bardsen transformation for the general ADL$(p, q; n)$ is given by:

$$\Delta y_t = \alpha_0 + \sum_{i=1}^{p-1} \alpha_i^* \Delta y_{t-i} + \alpha_p^* y_{t-p} + \sum_{i=1}^{j}\sum_{i=1}^{q-1} \beta_{ji}^* \Delta x_{jt-i} + \sum_{i=1}^{j} \beta_{jq}^* x_{jt-q} + \varepsilon_t \tag{A.6}$$

where $\alpha_i^* = -A(1) = \sum_{l=1}^{i} \alpha_l - 1 = \sum_{i=1}^{l} \gamma_l$ and $\beta_{ji}^* = B(1) = \sum_{i=0}^{l} \beta_{jl}$ (Bannerjee et al. 1993, 54).

Each successive α_i^* is thus the sum of the error correction coefficients for all lags of Y_t up to and including p in the DGP. The sum is appropriately thought of as the cumulative impact of a unit change in X_t on Y_t up through $t-i$ lags. The sum of the short-run effects is given by β_{ji}^* such that we can compute the long-run multiplier as $\frac{\beta_{ji}^*}{-\alpha_p^*}$ for any X_j. In the case where $p = q = n = 1$, the second term in the equation drops out and we have only one error correction term, that given by the coefficient on y_{t-1}: $\alpha_p^* = \alpha_1 - 1$ leading to the more familiar and easily estimated error correction model.