Regression in Space and Time: A Statistical Essay

James A. Stimson

American Journal of Political Science, Vol. 29, No. 4 (Nov., 1985), 914-947.

Stable URL:
http://links.jstor.org/sici?sici=0092-5853%28198511%2929%3A4%3C914%3ARISATA%3E2.0.CO%3B2-I

American Journal of Political Science is currently published by Midwest Political Science Association.

WORKSHOP

*Regression in Space and Time: A Statistical Essay**

James A. Stimson, *University of Houston*

Regressions on data jointly structured in space and time, commonly referred to as the pooling of cross sections of time series, can be formidable both in the strength of their design properties and in the number of special statistical problems encountered with them. This essay deals briefly with the potential applications of pooled design and more extensively with the special statistical problems commonly associated with analysis in space and time together. Four estimators—ordinary least squares, least squares with dummy variables, error components, and an adaptation of Box-Jenkins ARMA models to the pooled estimation problem—are reviewed, with an effort to suggest where each may find application in political science research. The four estimators are then illustrated by analysis of the regional dynamics in party issue polarization over issues of racial desegregation in the U.S. House of Representatives.

> *Very often a solution turns on some means of quantifying phenomena or states that have hitherto been assessed in terms of "rather more," "rather less," or "a lot of" or—sturdiest workhorse of scientific literature—"marked" ("The injection elicited a marked reaction"). Quantification as such has no merit except insofar as it helps to solve problems. To quantify is not to be a scientist, but goodness, it does help.*
>
> —*P.B. Medawar, Advice to a Young Scientist*

Political scientists deal with "space" in familiar cross-sectional designs. We deal with time in time-series regressions and with the family of techniques associated with Box and Jenkins. Although we are comfortable with each domain separately, we rarely combine them. We do comparative analyses across space, dynamic analyses over time, but almost never do we do dynamic comparison.

The quantitative subfields of political science can be roughly classified as cross-sectional (voting behavior, comparative politics), time-serial (political economy), or both (international relations, policy analysis), but in no specialty is it common to analyze both time and space at the same time. This is understandable, indeed under many conditions commendable, for introducing both space and time complicates analysis. We usually have quite

*The research reported in this article was partially supported by the National Science Foundation under grant SOC-7907543 and by the Center for Public Policy of the University of Houston. I am indebted to Sam Bookheimer, Edward Carmines, Raymond Duch, Jon Lorence, John McIver, and Gerald Wright for critical review of this paper and to Russell Dalton, Richard Eichenberg, Robert Luskin, and William Thompson for criticisms of an earlier effort.

enough complication already. But if we admit, as we must, that there are situations where space and time interact, where the one can be understood only with explicit reference to the other, then we must move on to the design of both theory and analytic strategy that accommodate that complexity. Pooling cross sections of time series is the analytic strategy I examine in this article.

Even if we have no interest in the interaction of space and time, design leverage provides a common motivation for pooling cross sections and time series. In situations where theory suggests systematic comparisons among a limited number of units—the nations of Europe, a handful of voting blocks in a legislature, or whatever—we can expect to see little more than sampling error in parameter estimates unless we can somehow increase the number of degrees of freedom in our analyses. Adding an available time dimension for each unit will sometimes be a solution. Or the case may be reversed. If theory suggests a dynamic specification, we may well have to rely on a time series far too short for reliable estimation. That case is endemic where measurement is annual, which is by far the most common situation of political time series—at least outside of political economy where finer time units of principally economic variables are commonly available.

Public policy analysts in particular often face a dilemma of stark severity. If they wish to undertake dynamic examination of the impact of particular policies and only annual measures are available, they are left to study only "old" policies (which present a time series lengthy enough for any analysis at all) to the exclusion of the new. Since the set of policies which can be studied with annual time series is precisely the opposite of the set which excites most interest, the new and experimental, any means of supplementing the time dimension with parallel time series for different units can be the solution of a particularly vexing problem.[1]

This article is a review of issues in pooled space and time analyses for political science. It is a review of issues more than published political science research, for there is little of the latter. Much of the literature on pooling models—now also commonly referred to as "panel models"[2]—is in econometrics. I shall deal with it, of necessity. It is important not to assume that any borrowed methodology is transportable to an unintended set of

[1] The difficulty doubles for election analysts who would use the biennial American national election series for longitudinal analysis. A veritable mother lode of information about the attitudes, predispositions, situations, and behaviors of groups in the American electorate, the time series is nonetheless much too short for purely time-serial analyses. That situation is in part the motivation of this article.

[2] The "panel" terminology is familiar to scholars in the survey tradition, which is in part the reason we do not employ it here, for although it has much the same connotation, it is different enough as commonly employed in the pooled regression literature to sow some confusion.

applications. I shall not. Indeed, some crucial econometric assumptions about the nature of "the" pooled data problem—particularly that it is characterized by large numbers of units and short time series—will turn out to be inappropriate for a substantial class of potential political applications.

Pooling data gathered across both units and time points can be an extraordinarily robust research design, allowing the study of causal dynamics across multiple cases, where the potential cause may even appear at different times in different cases. Many of the possible threats to valid inference are specific to either cross-sectional or time-serial design, and many of them can be jointly controlled by incorporating both space and time into the analysis. As students of research design, we must appreciate pooled designs. But as statisticians we are less enthusiastic. For pooled analyses, insofar as they are known at all, are known for special statistical problems.

Since most political scientists who study design also pay close attention to statistical theory (and vice versa), we tend to be a bit schizophrenic about combining space and time. We approve the design characteristics and are deeply suspicious of reported results. Or to put it in the reverse, we disapprove of simple cross-sectional and time-series designs because of very well known limits on valid inference, but we resort to them nonetheless from statistical conservatism.

There are means of analyzing space and time together that avoid violations of statistical assumptions. In fact there are several such means, some strikingly different than others, but all with similar names. That causes no small amount of confusion. An important first step in this statistical review is to identify the alternatives so that it will be possible to conduct a dialogue about pooled regression models where both parties have in mind the same topic.

Separating the approaches to combining space and time in regression actually is two topics, one technical and one more basic. The technical question is chiefly about error assumptions: what are the violations and how can they be remedied? But largely ignored in the statistical literature is the much larger question of the nature of the modeling enterprise when longitudinal and spatial variation are considered together. For each of the models carries assumptions about the nature of the research question (for example, which effects are tested and which merely controlled?), and, unlike the case of error assumptions, these structural assumptions invariably influence outcomes. Although that assertion is true of any statistical model, what distinguishes the pooled case is that the opportunity to be wrong is considerably enhanced when the design is two-dimensional.

Pooled models require some rethinking of the nature of the empirical enterprise. As Brunner and Liepelt (1972) have noted, the viewpoints and conventions associated with cross-sectional and longitudinal analyses may differ considerably. Analysts of cross sections, insofar as they are concerned with causation at all, typically observe covariation presumed to be produced by unobserved causal processes operating at some time before the data were gathered. Time-series analysts typically wish to model a causal process captured in the longitudinal data. And the causal "result" of cross-sectional analysis and the "process" of longitudinal analysis both appear as covariation. Clearly the simple statistical evidence used in both approaches can be seen quite differently in the two perspectives. Those differences must be confronted if we are successfully to model both process and result at the same time.

In the discussion to follow I shall take up models of space and time that are prominent in the literature and proceed in each case to identify error assumptions, structural assumptions, and appropriate applications. I then go on to further develop and illustrate one of them which seems promising for large numbers of political science applications.

Statistical models are solutions to problems, answers to questions. In lieu of problems to be solved or questions to be answered, they can come to seem far more complex than they are. Because it is handy to have concrete problems to solve and because not many pooled designs are readily available for common reference, let us presume a hypothetical research problem (which will turn out later to be a real one). For a guidepost on our brief excursion through statistical solutions, let us presume the following: (1) our problem is to explain the pattern of party polarization on racial issues in the U.S. House of Representatives under the assumptions (2) that it has a distinctly regional flavor and (3) that it is undergoing interesting changes within and between regions over time. We have for analysis the net inter-party difference on racial issues (aggregated from individual scales built from roll call votes) for, let us say, eight major regional groupings and about twenty years.

Pooling Cross Sections *and* Time Series (A Conjunction, not a Preposition)

We may begin our look at pooled data techniques by dealing with one that, except for its name, has little in common with all the others. When economists speak of pooling cross sections *and* time series, it is (sometimes) estimates, not data, that are pooled. In a fairly typical application one might have a time-serial macroeconomic model that is either underidentified or, if identified, lacks the power to produce interesting inferences (Koutsoyiannis, 1973, pp. 402–8). From altogether separate cross-sectional studies one may derive estimates of, for example, marginal propensity to save by income class. Then, with assumptions about the distribution of income, the cross-sectional estimates are incorporated into the time-series model to give it greater predictive power.

I shall have nothing further to say of this "pooling . . . and . . ." technique. I deal with it only to resolve the considerable confusion between the "pooling . . . and . . ." technique and the "pooling . . . of . . ." methods which are the focus of this article. The "of" designs all have in common what Brown and Halaby (1982) called a "stacked" data set, where the case is a particular unit at a particular time and the arrangement of cases has times "1" through "t" for unit "i" adjacent, followed by the set of time points for unit "$i + 1$" and so on. The data themselves are cross sections of time series (or time series of cross sections).[3] Although we shall later have to discriminate between cross-sectionally and time-serially dominant species—the one with many cross sections and few time points, the other with few cross sections and many time points—of the stacked configuration (for their asymptotic properties are quite different), the "times by units" design is a given for the following discussions.

Ordinary Least Squares

Ordinary least squares (OLS) models are a natural starting point in a review of pooled designs. Sometimes appropriate (Mayor and Pearl, 1984), OLS is quite probably the most often utilized model for pooled data. Its simple assumptions illustrate the problems of combining space and time.

It is worth noting at the outset, and easy to forget in practice, that OLS simply ignores the pooled structure of the data. Each case (for our hypothetical problem, one year for one region) is treated as independent of all others, not as part of a set of related observations.

The familiar regression model:

$$\beta = [X'X]^{-1}X'Y$$

implies an expected error covariance $\sigma^2 I$, where σ^2 is the expected constant variance for all cases and I effectively incorporates the notion of uncorrelated error. One doesn't need pooled data to violate either the constant variance or uncorrelated error assumptions, but it clearly increases the likelihood of violation. Two particular violations are likely to accompany stacked pooled data.

The cases, first, are not independent along the time dimension within units. We expect serial dependence in such data. Indeed, common practice

[3] To add a bit to the semantic confusion, the clean connotations of "cross section" and "time series" disappear when analysis is of space and time together. Aside from the ambiguous "unit," there is no proper term for the unit (state, nation, gender, person, or whatever) studied at several points in time. Thus it has come to be called a "cross section," which is confusing since it is in fact a time series.

in political science (in the incremental budgeting literature, for example) is not to treat multiple readings in time as "cases" at all, but rather to deal with them as multiple variables for a single cross-sectional case.

Second, a particular form of heteroscedasticity is inherent in stacked pooled data. For a variety of reasons, often neither interesting nor important, some units are inherently more variable than others at all times. Such differential variability is usually of modest concern in unpooled data because it affects only a single case at a time. In pooled data it is likely to affect whole sets (e.g., all years for one region) and have a considerably greater potential for mischief. Simple size differences between units are one such endemic source of heterogeneity. On the reasonable assumption that variation is roughly a fixed proportion of size, analysis of units of substantially different sizes induces heteroscedasticity in any regression. But the problem can take on worrisome proportion when each cross section consists of T cases in time. The Zuk and Thompson (1982) pooled analysis of the impact of military coups on military spending, for example, had to contend with nations whose military budgets differed by a factor of more than one thousand. The size problem can be resolved by standardizing the data, for example by dividing each cross-sectional unit by a constant so that an arbitrary base year will be 100. But one pays for this procedure with artificial parameter estimates.

Violations of error assumptions are likely even in otherwise well-specified models. Some scattered and unsystematic evidence suggests that regression estimates are robust in the face of violated error assumptions, while indicators of model fit are, as usual, more suspect (but on the whole, not bad either). But "otherwise well-specified" is a tall order. Erroneous *structural* assumptions are both likely to occur with pooled data and likely to cause serious problems when they do.

We have dealt thus far with "inherent" autocorrelation and heteroscedasticity, the irreducible minimum that occurs in the data. But the autocorrelation and heteroscedasticity we observe is a function also of model specification, and common misspecification can drive both to considerably higher levels than are inherent in the data.

The misspecification that is peculiar to pooled data, and one for which other pooled models are solutions, is the assumption of homogeneity in the dependent variable (the level of the dependent variable) across units. If we assume that units are homogeneous in level and they are not (equivalently, if we assume a common intercept for all units), then the least squares estimator will be a compromise unlikely to be a good predictor of any of the units, and the apparent levels of heteroscedasticity and autocorrelation will be substantially inflated.

An illustration is useful. Assume for simplicity that we have a sample of three units, named High, Medium, and Low, each with T readings in time. Medium fluctuates over time around the pooled mean. High is always above it, Low always below. Assume further that the units are homoscedastic and nonautocorrelated. Let us assume a common intercept (B_0) and fit a regression to the pooled cases using whatever independent variables we have (which we assume will fail to account for the between-unit differences). When we analyze the result by unit (and we would never see what's going on if we did not), Medium will be well behaved. The regression will be a relatively good fit, leaving small error variance and little apparent autocorrelation of residuals. But High (and similarly, Low) will be substantially different. There we will observe a regression prediction that is always below the data points, leaving uniformly positive residuals (and therefore strong positive autocorrelation) and large (observed) error variance.

The net result in our illustration will be observed heteroscedasticity (High and Low will produce error variance which may be many times larger than Medium) and substantial autocorrelation, a result of pooling two cases where autocorrelation is always positive at all lags[4] with one case where it is near zero at all lags. Thus starting with data that were homoscedastic and nonautocorrelated, we produce a regression with observed heteroscedastic and autocorrelated error. That this misspecification is common in pooled analyses probably arises from the fact that differences in level are rarely of theoretical interest, usually a given with a "so what?" quality. But no matter how uninteresting, misspecification of the intercept term(s) can produce massive error variance and leave an estimator biased beyond recognition.

If between-unit variation is large relative to over-time variation, commonly the case, then this misspecification will seriously distort the least squares estimator β. In particular, the coefficients for independent variables that discriminate between units will be biased as they become in effect proxies for omitted intercept terms. All estimators will be biased and inconsistent (Maddala, 1971; Mundlak, 1978; Kiefer, 1980; Anderson and Hsiao, 1982), a qualitative statement. Quite probably the degree of bias will be large.

The illustration suggests one major class of solutions. If we had fit individual intercepts to High, Medium, and Low (the covariance fixed effects model), all three would have been well behaved.

[4] Further, the autocorrelation will be nonstationary and thus not amenable to any correction procedure that requires estimation of autoregressive or moving average error parameters, both of which are undefined for nonstationary series. Because nonstationarity masquerades as autoregressiveness, most conventional tests would suggest an autoregressive correction procedure, which would be quite meaningless for the problem at hand. The importance of the stationarity assumption depends upon the degree to which the pooled data are time-serially dominant.

When then can OLS be recommended for pooled analyses? OLS is reasonably robust when units are homogeneous on Y (in fact, identical to covariance and GLS error components models in the limiting case) and when heteroscedasticity and autocorrelation are not at outrageous levels. There is however a catch. What we need to know to be comfortable with OLS estimation, OLS itself will not tell us. Appropriate diagnostic information for distortions that arise from the pooled structure of the data requires explicit recognition of that structure. And OLS does not recognize a structure of N units at T times as anything more than NT independent cases ($NT = N \times T$). OLS produces no unit-specific measure of heteroscedasticity, and its standard indicators of autocorrelation (e.g., Durbin-Watson) are inappropriate (and biased against significance). There are tests for homogeneity, but performing them takes one beyond OLS. Thus OLS may well be appropriate for a particular research question, but without entertaining other models there is no satisfactory way to know that it is appropriate.

Least Squares with Dummy Variables: The Covariance Model

Significant between-unit (or between-time-point, or both) differences are the source of the OLS bias. The obvious remedy is to remove them, either by transformation (see Wallace and Hussain, 1969, pp. 60–61) or by introducing dummy variables for the effects and estimating with OLS. In either case the goal is to produce a zero expectation for residuals within and between units (or time points, or both, etc.).

The least squares with dummy variables model (LSDV), the econometric version of the covariance model, may be written:

$$y_{ij} = \Sigma\beta_r x_{rij} + \alpha_i + \tau_j + \mu_{ij} \qquad (i = 1, 2, \ldots, N; j = 1, 2, \ldots, T) \qquad (1)$$

where x_{rij} are the r covariates, α_i are unit-specific effects, τ_j are time-specific effects, and μ_{ij} are specific to both time and unit. Our error, ε_{ij}, as in equation (2):

$$\varepsilon_{ij} = \alpha_i + \tau_j + \mu_{ij} \qquad (2)$$

consists of three components, only one of which, μ_{ij}, is suitable for OLS estimation. Let us presume, as is generally the case, no time-specific effects (i.e., $E(\tau_j) = 0$ for all j) other than those to be modeled. The remaining problem then is to remove the unit-specific error from the error term by (1) transforming the equation, (2) recomputing each value as a deviation from its unit mean, or (3) fitting dummy variables to estimate a fixed effect, α_i, for each unit (either through the origin with n dummies or with an intercept and $n - 1$ dummies). Any of the three produces the same coefficients, which in turn can be proven to possess desirable asymptotic and finite properties (Wallace and Hussain, 1969, p. 62). For convenience, we adopt the third, which does differ in the potentially important matter of producing a goodness-of-fit measure that takes explanatory credit for the unit effect dummies, which might or might not be appropriate, given the particular theory in question.

The argument can be generalized to both units and time simultaneously, with some complication, but we shall follow here the normal emphasis on units. As a practical matter we would seldom wish simultaneously to control both temporal and spatial disturbances. That entails both complexity and large losses of degrees of freedom. More important, it may curtail our ability to model either dimension of the phenomenon in question by assigning much of its systematic variation to dummy atheoretical variables which are collinear with explanatory variables.

Error Components Models

The transition from OLS to covariance estimators was both simple and orthodox. Between-unit variance in OLS causes problems for which the covariance estimator is the solution. The transition from covariance to error components specifications is neither simple nor orthodox.

The development of error components models can best be understood as a search for *efficiency* in pooled estimators. The covariance "within" estimator is a best linear unbiased estimator. But "best unbiased" is not necessarily "best" in an absolute sense, the estimator with minimum variance. Search for this "best" estimator is the motivation underlying the development of error components models. "Since only a few observations are available over time, but a great many observations are available for different individuals at a point in time," wrote Marc Nerlove (1971b), a central figure in these developments, "it is exceptionally important to make the most efficient use of the data across individuals . . . in order that the lesser amount of information over time can be used to best advantage". (p. 359)[5]

The error (or variance) components model arises from dissatisfaction with the covariance estimator and rests on a reconceptualization of the "problem" of pooled analyses. Maddala's (1971: 341) three point critique puts the issue well:

1. If between-unit (or between-time-point) variance is large, the covariance model attenuates total variance considerably by assigning much of it to fixed unit (or time) effects.
2. Large sets of dummy variables entail large losses of degrees of freedom.

[5] Nerlove's empirical spirit and his findings from extensive Monte Carlo experiments are anticipated by his opening quotation, ". . . the Dodo suddenly called out, 'The race is over!' and they all crowded round it, panting, and asking 'But who has won?'" from *Alice's Adventures in Wonderland.*

3. The coefficients associated with fixed unit (or time) effects are often uninterpretable.

For political science applications the second (degrees of freedom) critique can be evaluated only in the context of a particular design. But points one and three are fundamental to pooled design. If sets of dummy variables are highly collinear with explanatory variables (as they often are), how should variance be apportioned? LSDV divides explanation between the two types of variables, a process which presents the usual multicollinearity efficiency problems. If the third critique (dummies uninterpretable) is on the mark for a particular application, then loss of efficiency in estimates of the variables of interest may be a very high price to pay. If, on the other hand, there is little collinearity between explanatory and dummy variables, then OLS rather than LSDV becomes the probable estimator of choice.

The GLSE reconceptualization begins with the assertion that the estimated dummy coefficients (α_i) are not explanation, but rather summary measures of our ignorance about the causes of between-unit differences, what Maddala (1971) called "specific ignorance" as opposed to the "general ignorance" represented by the error term (μ_{ij}). If so, and if we regard our selection of units as a sample from which we wish to infer estimates (rather than a fixed group of cases to be measured), it makes sense to regard the α_i as random variables, not fixed constants—and for random variables the distribution parameters (mean and variance) are of interest, but not the individual "effects."

Error components, like OLS, then estimates a model without fixed effects (represented by dummy variables):

$$y_{ij} = \Sigma\, \beta_r x_{rij} + \varepsilon_{ij} \qquad (i = 1, 2, \ldots, N; j = 1, 2, \ldots, T) \tag{3}$$

where the error is given by:

$$\varepsilon_{ij} = \alpha_j + \tau_j + \mu_{ij} \tag{4}$$

With unit (and sometimes time) effects now collected in the error term, the GLSE looks like OLS, except that the error term is unsuitable for OLS estimation. The OLS estimate is often biased, the source of all our problems. The error components solution to the problem is to correct the bias with a GLS solution that specifies the bias.

The error components model is the GLS estimator:

$$\beta = [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}Y$$

$$\Omega = \sigma^2 \begin{vmatrix} A & 0 & \ldots & 0 \\ 0 & A & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & A \end{vmatrix}$$

A is s (T by T) partition of time dependence coefficients:

$$A = \begin{vmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \dots & \dots & \dots & \dots \\ \rho & \rho & \dots & 1 \end{vmatrix},$$

0 is a (T by T) partition of zeros; and $\rho = \sigma_u^2/\sigma^2$, the intra-class correlation coefficient).[6]

Partitioning of the error covariance matrix Ω arises from the assumption that the disturbance in any unit at any particular time affects only other time points for that unit. Each diagonal partition in the matrix represents the expected association between a particular unit and itself at various times. The 0 off-diagonal partitions reflect the assumption that the error in unit i at time t is uncorrelated with the error for other units at all times, including t. The possibility of cross-unit error covariation at t, called "spatial autocorrelation," (Doreian, 1981), raises considerably different questions than those taken up here. Because partition borders discriminate between the spatial and longitudinal dimensions, they are ubiquitous in pooled regression models.

The partitions A in the GLSE model amount to the assumption that unit effects are captured as serial correlation ρ, which is assumed to be constant at all lags.[7] The constant autocorrelation within each unit arises, as in the informal example above, from the unit effect, which by definition is time wise invariant. That is an unusual treatment of serial correlation and requires an assumption, which Nerlove (1971a, pp. 361–62) termed "fundamental," that no true (i.e., time-serial) autocorrelation is present in the error, a restriction that may be difficult to satisfy in time-dominant designs. The presence of a unit effect which is timewise invariant and the absence of time-serial autocorrelation produces the expectation of positive autocorrelation ρ at all lags.

Estimation of the error components model is either by maximum likelihood or by a two-round procedure, where least squares with dummy variables for unit effects is used to estimate ρ for the second-round GLS solution, which omits the dummies. A number of alternative procedures can be used to estimate ρ. After experimental examination of 11 of them, Maddala and Mount (1973) reached the pleasant conclusion that all lead to similar results, a conclusion supported more recently by Baltagi (1981) and

[6] Dividing the total error into its three components, (1) unit-specific, time-invariant, (2) time-specific, unit-invariant, and (3) unit- and time-specific, for the estimation of ρ is the origin of the name for the model.

[7] Recall from the previously discussed example that misspecification (by omission) of unit effects does appear as autocorrelation, which does in fact tend to be constant over time within units.

Wansbeek and Kapteyn (1982). These results are consistent with recent work (Taylor, 1980; Baltagi, 1981) showing that estimator efficiency is relatively insensitive to the quality of variance component estimates.

The error components model can be seen as a compromise between OLS, which is biased in the presence of fixed effects, and LSDV, which is inefficient. It gives generally greater weight to X (the "covariate" or "within" term) than does LSDV, but less weight than the biased OLS. At the price of imposing restrictions (covariates X and unit effects α_i uncorrelated; no time-serial autocorrelation), error components gains efficiency. But each of these restrictions may be controversial. Mundlak, taking particular aim at the assumption of no correlation between X and unit (or time) effects, caught the model on the horns of a statistical dilemma. If the assumption is true, he proved the error components (GLSE) estimator to be equivalent to the LSDV "within" estimator. "Equivalence" deprives it of any claim to greater efficiency. But if the restriction is lifted, GLSE becomes a biased estimator.

> When the model is properly specified, the GLSE is identical to the [covariance] "within" estimator. The whole literature which has been based on an imaginary difference between the two estimators, starting with Balestra and Nerlove [1966] is based on an incorrect specification which ignores the correlation between the effects and the explanatory variables. . . . the GLSE . . . is actually a restricted estimator, and when the restriction does not hold it is a biased estimator. (Mundlak, 1978, pp. 70–71)

Mundlak's demonstration does not quite settle the issue, but it has shifted the grounds of debate (and spawned yet another series of models we shall not pursue here). If the "best" estimator (GLSE) is biased (on the assumed correlation of X and error) and the consistent "within" estimator is inefficient, we still do not know which, if either, is the "preferred" estimator. That depends not upon the fact of bias or inefficiency, but upon its *degree*, under a variety of circumstances. Bias may be trivial (as for example the case of the uncorrected variance estimate from sample data in large sample applications) or it may be serious. Differences in estimator efficiency may be large or small. These issues require experimental, not mathematical, treatment. The jury is still out, for there are more specifications than can ever be examined with experimental techniques.

A Time-Serial GLS Model

Econometricians who write about cross sections of time series pay notably little attention to autocorrelated error, particularly in contrast to the regular (i.e., time-serial) econometric literature. Nerlove (1971b) quickly noted that autocorrelation is assumed not to exist in the error components model (as it must be: the ρ parameter in the GLSE model is used to specify cross-sectional effects). Others simply do not mention it.

Three explanations may be advanced for the omission. All have to do with typical econometric *applications,* rather than more theoretical issues. One of the most important applications for pooled designs in economics, indeed an explanation for the considerable attention econometricians pay to pooled models, is the microeconomic study of the behavior of firms over time. Maddala (1971b) even used the term "firm" to designate the cross-sectional unit in his general model. The focus on firms, with presumed important individual characteristics, immediately forces attention to the issue of between unit effects, the predominant concern of econometric treatments of "the problem" of pooled designs.

The numerical dominance of the cross-sectional over the time-serial dimensions of applications suggests a second explanation for lack of concern over autocorrelated disturbances. Bòth in the typical "firm" application and in the Balestra and Nerlove (1966) "demand for natural gas" study that had much to do with the origins of the error components formulation, it has always been assumed that the number of units is considerably larger than the number of time points. Nerlove (1971a, 1971b) was most explicit about this assumption. Cross-sectional dominance simultaneously minimizes the threat of autocorrelated error (only $N(T-1)$ cases may be serially dependent, which can be a relatively small proportion of NT when N is large and T is small) and maximizes the possibility of bias (or inefficiency) from the specification of unit effects.

Economists probably show little interest in autocorrelation, last, because GLS methods of dealing with it require an unbiased estimate of the time-serial disturbance, ρ. Such estimation frequently is very difficult because of the econometric popularity of lagged endogenous model specifications, which preclude unbiased estimation of autocorrelation.

Both because autocorrelation is frequently a small problem relative to unit effects and because it is frequently impossible to deal with it in any case, given the applications in question, econometric treatments have tended to "assume" it to the sidelines. Models that deal with it, such as the one I shall take up here, have enjoyed little popularity among the econometric fraternity.

GLS-ARMA

The approach I describe here is more traditional than the error components model. Like the normal time-series case, it treats the pooled estimation problem by estimating and then specifying the time dependence process in the residuals (see Kmenta, 1971, for development of the AR(1) case). It also incorporates a conventional approach to (between-unit) heteroscedasticity by a WLS weighting procedure. A GLS specification for cross sections of time series, it reduces to the normal (time-serial) GLS model in the limiting case of only one unit.

The model is similar in form to GLSE, identical except for the specification of Ω. It may be written:

$$\Omega = \begin{vmatrix} \sigma_1^2 A & 0 & \dots & 0 \\ 0 & \sigma_2^2 A & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 A \end{vmatrix}$$

where the A partitions specify time dependence processes. In the common first order autoregressive case:

$$A = \begin{vmatrix} 1 & \rho & \rho^2 & \dots & \rho^{t-1} \\ \rho & 1 & \rho & \dots & \rho^{t-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{t-1} & \dots & \rho^2 & \dots & 1 \end{vmatrix}$$

The GLS-ARMA estimator is similar in appearance to GLSE. Its covariance matrix is partitioned on the same logic (within-unit time dependence, no cross-sectional correlated error) discussed above. Its critical difference is that it specifies the within-unit over-time partitions A as ARMA processes rather than the "fixed at all times" specification of GLSE. ARMA processes (Box and Jenkins, 1976) may take on many appearances—the first-order autoregressive process presented is only an illustration of one prominent possibility. But what distinguishes all of them from the constant ρ of GLSE is that they decay over time. The ρ of GLSE is constant over time because it is a cross-sectional effect. The decay inherent in ARMA processes is a time-serial effect; it represents the assumption that successively more distant points in time are expected to show declining association with one another. The choice between the two models thus comes down to whether "the problem" of pooled designs is unspecified unit effects (GLSE) or time-serial effects (GLS-ARMA). GLS-ARMA becomes relatively desirable when design is time dominant and when explanatory variables are dynamic. If either of these conditions is not present, autocorrelation may be more a nuisance than a threat, and other considerations would dominate the choice of estimator.

As illustrated earlier, where there are significant between-unit effects, error will be nonstationary and GLS-ARMA inappropriate unless modified by the addition of dummy variables. The econometric unpopularity of the model may now be restated: it deals with a problem of little concern in economic application (autocorrelation) and fails to contend with a problem (unit effects) of great concern.

GLSE is a compromise between the efficient OLS and the consistent LSDV estimators. It eliminates dummy variables and compensates with the specification of "ρ." GLS-ARMA, as proposed here, is also a compromise. It deals specifically with two of the defects of LSDV (lost degrees of freedom and uninterpretable fixed-effect coefficients) by an iterative specification-diagnosis-respecification modeling procedure that fits as many unit dummy variables as are necessary to produce homogeneous behavior, but no more. It is well adapted to pooled analyses where time is dominant (i.e., $T > N$), both because autocorrelation becomes more threatening with lengthier time series and because it is easier to model unit effects when dealing with a relative handful of units. When N is small (eight for the illustration used here), the number of dummies is manageable and ready interpretability is probable.[8]

The threat of bias in this estimator comes from underspecification of unit effects, from not providing dummy variables where they are needed. (It gains efficiency by not providing more than are needed.) Reaching the appropriate specification requires sensitive diagnostic information which addresses the particular question of which unit dummies to include. If appropriately specified, the residuals will display three properties: (1) the summed unit means will be approximately zero, (2) the unit residual variances will be approximately equal,[9] and (3) the pattern of autocorrelation within each unit will be stationary. Underspecification where unit effects are significant for unit "i" produces the opposite properties; systematic under- or overprediction (at all times) produces observed mean residuals considerably different from zero, error variance substantially higher than for other units, and nonstationary autocorrelation for the unit. The first two conditions could be misleading in the hypothetical situation of all units equally badly specified. But nonstationary autocorrelation of residuals (i.e., a pattern that shows no tendency to decay to zero at progressively longer lags) is always a tipoff to underspecification. I shall illustrate the diagnostic procedure in the analysis below.

I conclude the review of alternative models with Table 1, an informal guide to appropriate estimation techniques under interactions of three pooled design characteristics: (1) whether the design is cross-sectionally or time-serially dominant, (2) whether or not between-unit effects are present (and, if so, whether they are fixed constants to be estimated or random variation to be controlled), and (3) whether or not significant time-serial correlation of error is present. In the same spirit as this essay, what the table

[8] GLS-ARMA presumes dynamic models of some sort for timewise structural effects and specifies time-serial error with ρ. It therefore makes no sense to introduce dummy variables for time effects. If such dummies were necessary with this time-dominant specification, we would face the same problems (too many lost degrees of freedom and uninterpretable coefficients) that accompany unit dummies in cross-sectionally dominant designs.

[9] "Approximately equal" means in this case same order of magnitude. The variances may be different both from sampling fluctuation and heteroscedastic error in the data.

TABLE 1

Design Characteristics and Pooled Estimators: An Informal Guide (Cell Entries Are Appropriate Estimators)

	No Timewise Autocorrelation in Error	Timewise Autocorrelation Present
Cross-sectional dominance ($N > T$)		
No between-unit effects	Ordinary Least Squares (OLS)	*
Between-unit effects present		*
(Fixed)	Least Squares with Dummy Variables (LSDV)	
(Random)	Error Components (GLSE)	
Time-serial dominance ($T > N$)		
No between-unit effects	Ordinary Least Squares (OLS)	GLS-ARMA
Between-unit effects present	Least Squares with Dummy Variables (LSDV)	GLS-ARMA (with Dummy Variables)

*No estimator developed specifically for this case.

shows is not "the best pooled estimator," but the conditionally best under various design variables and with typical assumptions of the proponents of various models about what is "the problem" of pooled analyses.

Now let us turn to the real world of American politics for illustration of a problem requiring pooled estimation, and then go on to some solutions.

Party Polarization on Racial Desegregation in the U.S. House, 1957–80: An Example Analysis in Time and Space

The racial positions of the two American political parties have changed dramatically over the last three decades. The phenomenon is so uniform that it can be observed by indicators as diverse as party platforms, citizen perceptions of what the parties stand for, the issue positions of party identifiers, and, most strikingly, in partisan divisions over the issues of racial desegregation in both houses of Congress (Carmines, Renten, and Stimson, 1984; Carmines and Stimson, 1981, 1984, forthcoming [1985].

Starting from relatively centrist positions, with the Republicans on balance more liberal, the parties first switched positions on the issues and then, having done so, began an evolutionary polarization which at the end of the period leaves neither party centrist and the Democrats strikingly more liberal. For this article I take all that as given.

Of particular interest here is the question of *how* this polarization occurred. The basic hypothesis, already tested at the national level, is that the polarization is a case of "issue evolution." Issue evolutions, in the simplist case, begin with a "critical moment," an event large enough to be noticeable, but far less dramatic than the critical elections of the realignment literature. Following the critical moment is an evolutionary adaptation to new issue positions which in the long run far overshadows the event which set it in motion. The evolutionary adaptation, a positive feedback process, is the nearly inevitable combination of a perceived change accompanied by inertial resistance to the change. Neither citizens nor politicians readily change attitudes or parties. But both sorts of actors tend to line up "correctly" when they newly enter the system after the critical moment. Changing issue alignments thus follow a dynamic growth pattern which is bounded by an eventual equilibrium state which occurs only when the level of issue polarization of "old" citizens and politicians grows to match the level of new entrants. This sort of dynamic is well suited to the Box-Tiao (1975) intervention model:[10]

$$Y_t = \frac{\omega_0}{(1 - \delta_1 B)} I_t + N_t$$

where

Y_t is the dependent issue polarization at t;

I_t is a binary intervention variable coded zero before the intervention and one after;

B is the backshift operator such that $BX_t = X_{t-1}$;

N_t is a noise model for the residuals; and

$\omega_0/(1 - \delta_1 B)$ models the dynamics.

The dynamic term, $\delta_0/(1 - \delta_1 B)$, is a close fit to the issue evolution theory: ω_0 is the initial impact at the critical moment and δ expresses the growth rate. A fraction less than 1.0, δ implies that the evolutionary growth never matches the change at the critical moment in any later year, but the

[10] McCleary and Hay (1980) wrote the basic first-order model as:

$$Y_t^* = \delta_1 Y_{t-1}^* + \omega_0 I_t + a_y$$

where Y^* is the dependent variable after its noise process N_t has been filtered out. This form, easier to interpret, is deceptively similar to a lagged dependent variable formulation. It is different in the critically important respect that Y_{t-1}^* is (by definition) uncorrelated with error a_t. The measured variable Y_{t-1} never appears on the right-hand side.

cumulation of dynamic growth over a period of years may far exceed the original change.[11] The eventual equilibrium level is given by $\omega_0/(1 - \delta_1)$. I shall say nothing further about intervention modeling, which is both well documented in the literature and a distraction from the task at hand in this essay.

For an operational version of the issue evolution notion let us postulate a critical moment based upon the decisive issue contrasts (and outcome) of the 1964 Johnson-Goldwater contest (which affects the series under consideration in 1965 when the Congress elected in 1964 is seated). The data are individual issue scales based upon annual factor analyses of roll call votes. Scored with a standard metric (50, 25), high scores indicate racial liberalism. When aggregated by party (or party within region), Republican scores are subtracted from Democratic scores, making positive values connote relative Democratic liberalism. Thus we would expect ω_0 always to be positive, reflecting the growing alignment of liberal with Democratic, and δ should also be positive, reflecting the positive feedback process implied by the issue evolution theory.

Evolution and issue evolution are inherently dynamic concepts. But racial desegregation is an issue framed in a regional context. Levels of popular support for desegregation vary across regions, and have as long as the issue has existed. And there is every reason to believe that similar variation is to be found in the voting patterns of regionally defined subgroups of the House of Representatives. I ascribe that variation to regional political culture, a term I wish to use as an empty vessel to capture the multitude of factors, social, historical, economic, and so forth, that impinge upon the typical outlooks, beliefs, and behaviors of geographically defined populations. These "between" differences are not the central focus of my interest, but as is often the case in pooled designs, one must come to terms with them to understand the dynamics of the composite national parties.

American political parties are not and have never been national parties. They are organized at various levels, from precinct to national committee, but their character seems but modestly associated with any of their levels of formal organization. They are diverse, but that point may easily be overstated. Each party is composed of 50 state units, but it is not the case that each of those subunits has a wholly distinctive character, a political world of its own. The parties are more cohesive than that. The lines of division are fewer; they are drawn by the distinctions between shared and not shared

[11] The dynamic parameter δ_1 is estimated by program POOL used for these analyses. The program provides a nonlinear maximum likelihood solution. A number of alternative model specifications for the estimate, corresponding to the variety of models of structure and error processes entertained below, all produced essentially the same estimate, with maximum variation on the order of .01. Because δ is insensitive to model specification in this instance, it is sometimes treated below as a fixed constant rather than a parameter to be estimated.

interests, shared and not shared experiences, shared and not shared values, shared and not shared speech. The lines of division are largely regional.

"Region" is a stepchild in American political science. A critical concept in the formal study of other nations, it has never achieved much prominence in the United States. That is true for a number of reasons. A discipline with a legalistic heritage, political science has divided the world of American politics into a tripartite national, state, and municipal scheme which recognizes governing units which have constitutional standing at one or another level. Regions lack both governing units and constitutional standing.[12] The political behavior tradition, the most prominent challenge to the legalistic heritage, has equally little room for a regional focus. It divides the world into two parts, government and individual citizen, relegating all geographical subdivisions to the residual category "context." Its typical tool, the national sample survey, is capable of regional analyses, but best suited to the study of the individual citizen/voter. As Beck (forthcoming [1985]) has noted, that preoccupation has distracted attention from a host of alternative ways of conceptualizing the political world.

Region, the generic concept, suffers probably most of all in the American context from the fact that one particular region, the South, is so distinctive that it draws attention away from the more subtle distinctions along other regional lines. That distinction is so large and so durable that we commonly treat explanations and analyses as two components, one each for "South" and "Nonsouth." It is widely noted that the South is not a homogeneous region, and that is true. It is more remarkable that the "Nonsouth" is far less homogeneous. The process of controlling for the residue of the Civil War has blurred all other distinctions. If "Nonsouth" really were homogeneous, if for example, we could be comfortable with the notion that the political outlooks and behaviors of New Englanders and Californians were essentially similar, that would not be troublesome.

Region is a stepchild not because it is rejected—few would argue the converse, that American regions respond uniformly to political choices—but because it is ignored. But it is not always ignored. Sundquist's (1973) masterful treatment of realignment made a strong case that the phenomenon requires a regional explanation, that realignments follow a different course, occur at different times, in different ways, and for different reasons between regions.

Journalistic commentators on American politics do not ignore regional considerations. Indeed, it may even be fair to say it is their typical stock in trade. Lacking the rich data of the political scientist, they frequently look at

[12] It is worthy of note that two deservedly prominent exceptions to this rule, Key's *Southern Politics in State and Nation* (1949) and Lockard's *New England State Politics* (1959), both found it necessary to justify their considerable regional insights by title reference to state politics.

the simple issue of *where* votes come from. If one looks at this question, one cannot fail to see substantial regional differentiation. And thus journalists have long noted a substantial east/west polarization of the American political parties, a pattern rarely noticed by political scientists (Abramson, Aldridge, and Rohde, 1982, is an exception to this point). And the definitive work on the regional basis of American politics, Garreau's *The Nine Nations of North America* (1981) is not academic scholarship at all, but a popular book written by a working journalist.

Four Estimators Compared

I am ready now to model the effect of a covariate, the intervention variable, on issue polarization within region. Rather than picking the one best pooled estimator, I shall use the illustration with four estimators, OLS, LSDV, GLSE, and GLS-ARMA and examine assumptions and outcomes in each case.

I. Ordinary Least Squares

Ordinary least squares is not appropriate for this application. Its statistical model and assumptions are violated. More seriously, it is misfit to this particular research; it doesn't allow us to address important questions of interest. It serves nonetheless as a useful baseline from which we can calibrate the refinements of other estimators.

To employ OLS estimation we need to assume homogeneity of units and time-serial independence, both derivative of the standard OLS uncorrelated error assumption. We will discover—alas, only after the fact—that neither assumption is tenable for this case.

With 24 annual readings (1957 through 1980) for eight regions,[13] our design is time-dominant. The mere fact that there are more readings along the time dimension than across the spatial one alerts us to the probability that time-serial problems have ample opportunity to produce serious distortions. But since OLS does not recognize the two-dimensional data structure, being alerted is the best we can do in either dimension.

The first estimates are presented in Table 2. The OLS model consists of but three parameters, a constant which may be interpreted as the pre-intervention level of polarization (implicitly for all regions equally), a

[13] I employ the ICPSR standard region scheme which divides the states into New England, Middle Atlantic, East North Central, West North Central, South, Border, Mountain, and Pacific (including Alaska and Hawaii) regions. The ICPSR scheme has its problems, as do all alternatives. Two fundamental difficulties limit the prospect for improvement; (1) the cultural borders of regions are not constrained by the legal limits of state boundaries, and (2) regional definitions are not fixed; states may wander back and forth across regional boundaries over time.

TABLE 2

Two Intervention Models of Party Issue Polarization in the U.S. House of Representatives

	Ordinary Least Squares	Least Squares Dummy Variables
1964 election intervention (ω_0)	4.82	4.71
	(0.36)[a]	(0.28)
Dynamic growth parameter (δ_1)	0.83	0.83
	(0.08)	(0.08)
Constant (pre-intervention mean)	3.42	−5.69
	(1.28)	(2.01)
Middle Atlantic		8.53
		(2.61)
East North Central		13.20
		(2.61)
West North Central		10.79
		(2.61)
South		−0.21
		(2.61)
Border		4.84
		(2.61)
Mountain		16.55
		(2.61)
Pacific		19.58
		(2.61)
R^2 (adjusted)	.485	.665
ρ[b]	.55	.31
Autocorrelation pattern	Nonstationary	Stationary
Estimated equilibrium level $\omega_0/(1-\delta_1)$	28.35	27.71

[a]Standard errors are indicated in parentheses.

[b]The autocorrelation estimate ρ is a weighted composite of within-region residual autocorrelation.

linear intervention coefficient ω_0 (the intervention impact at its onset), and δ_1, an indicator of the rate of dynamic growth to a new equilibrium level.

The OLS estimates of Table 2 are in many regards well behaved. The intervention model is a close approximation both to the raw data and to similar models fit to a single national polarization time series. It describes a situation of slight Democratic liberalism before 1964 followed by a dramatic shift after the 1964 election, an impact that approaches an eventual equilibrium level some 28 points higher than the starting value. But two problems are notable. On the statistical side, both pooled and unit autocorrelations indicate nonstationary autocorrelated error. The pooled autocorrelation for the first five lags (.55, .36, .36, .44, .47) fails to display the rapid decay toward zero that would indicate stationarity. Within-unit autocorrelations (not presented) show at least three regions (New England, South, and Pacific) with essentially constant autocorrelation at all lags (the Pacific region, for example: .63, .58, .62, .62, .66). Flat autocorrelation, as we noted earlier, indicates underspecified unit effects.

On the substantive side, the positive constant term is misleading. The Democrats were on balance slightly more conservative than the Republicans on racial matters prior to 1964, which should produce a negative or nonsignificant constant term. That is obscured in part by aggregation problems (southern Democrats were about a third of the party in the House, but all regions are weighted equally in the regression) and in part by failure to include Region effects. Inferences from the OLS estimates would take us only slightly astray, but the estimator is clearly not optimal for the situation.[14] Our diagnosis indicates underspecified unit effects. The next logical step is to turn to a model that specifies unit effects.

II. Least Squares with Dummy Variables

The LSDV specification in the second column of Table 2 is in all regards similar to the OLS specification except for the inclusion of seven region dummy variables. (The eighth, New England, is an omitted reference category.) As applied here, the LSDV model specifies a dynamic process that is uniform across all regions, but shifted upward or downward at all times by a constant (the dummy variable coefficient) in each. Although this model does not specify space/time interactions, it begins to allow specification of the problem into its spatial and temporal dimensions for a cleaner look at each.

[14] The reader may rightly question how OLS problems can be diagnosed with confidence after the earlier assertion that diagnosis was problematic with OLS. The assertion should actually be drawn a bit narrower; it is OLS software that is problematic. The estimation and diagnosis here were performed with a pooled estimation routine constrained to OLS assumptions.

LSDV improves upon the OLS specification of this problem in several ways. Adjusted R^2 is substantially boosted, from .49 to .67. Autocorrelation is nearly halved, from .55 to .31. More importantly, the nonstationarity of the OLS autocorrelation is entirely removed; of the autocorrelations at the first five lags (.31, -.01, -.02, .10, .15) only the first is significant. This clearly confirms the suspicion that the OLS autocorrelation results mainly from underspecified unit effects.

Interpreting the LSDV estimates is a less satisfactory matter, as Maddala warned. Setting aside the intervention parameters, nearly identical to those estimated in OLS, it is unclear what sense can be made of the coefficients. Ignoring what we know about American politics, a researcher looking at these estimates would likely interpret them as showing the South relatively normal and six other regions as significantly more polarized, which would do justice to the data and at the same time be a substantial misunderstanding.

What we would like to see in dummy parameters is how each category deviates from some meaningful norm. What we actually see is how each category differs from the reference category. New England, the reference category here, is itself significantly less polarized (which is not apparent from these estimates) than the national norm. Choice of reference category is arbitrary; New England was chosen by the particular software employed because it was the first category. (The other reasonable arbitrary choice, the last category, would have been even more deviant, but in the opposite direction.) The problem in essence is that there are as many equally good LSDV solutions—mathematically equivalent, but not equally interpretable—as there are categories. Knowing that the estimates are misleading, there are ways we might adjust them to produce a more reasonable interpretation. But in the general case we wouldn't know that our interpretation was misleading.

A second aspect of this same problem is that normal tests of significance (i.e., for coefficient difference from zero) are also misleading, for lack of a meaningful zero point. The problem works in both directions; as the reference point shifts from the reference category to some meaningful norm, some significant coefficients become nonsignificant and some nonsignificant coefficients become significant. Significance tests, without great caution, are likely to obscure more than aid the interpretative process. This awkwardness, along with fears of less-than-optimal efficiency, underlies the development of the error components alternative.

III. Error Components (GLSE)

The error components approach to pooled estimation begins with the presumption that we are interested in unit effects only insofar as they confound estimation of the covariates. It is as if we replaced the names of our categories (New England, Middle Atlantic, and so forth) with the numbers 1 through 8, each of no inherent interest (Maddala's term "specific ignorance" is apt) except that failing to control it might lead to biased estimation of the truly interesting covariate. The error components advocates have noted all along that this presumption is a substantive, not statistical, question. The error components solution cannot be justified if it is inappropriate. It *is* inappropriate here, because we believe the regional units are part of a meaningful explanation of issue polarization, not mere residual categories that must be dealt with to avoid bias. If not for the illustrative purpose of this essay, it would be inappropriate to proceed further.

On design and statistical criteria the error components estimator gets a mixed review for this case. The positive side is that this is one of the very rare cases where it is tenable to assume the independence of the covariates (here purely time-serial) and unit effects, the assumption Mundlak highlighted as necessary for unbiased estimation. The misfit of estimator to illustration here is that error components is designed to cope with short (in time) and fat (in space) design problems. This design is long and thin, presenting a different set of problems than error components was designed to control.

Application of error components here is by a two-round procedure, where the first round is devoted to estimating ρ, which is then incorporated into the second-round GLS solution. For the first round we need to estimate the error components:[15]

$$\sigma^2 = \sigma_u^2 + \sigma_e^2$$

a simple apportionment of error into total, between, and within categories. Total variance σ^2 may be estimated from the residual variance of the OLS solution. Within variance σ_e^2 may be similarly estimated from the residual variance of the LSDV solution, leaving $\sigma_u^2 = \sigma^2 - \sigma_e^2$. The estimate of ρ is then the ratio σ_u^2/σ^2. From the regressions already reported, $\sigma^2 = 125.91$, $\sigma_e^2 = 81.85$, $\sigma_u^2 = 44.06$; thus:

$$\hat{\rho} = \sigma_u^2/\sigma^2 \approx .35.$$

The second-round GLS solution of Table 3 suggests that, *for this illustration,* the controversial GLSE assumptions produce estimates only slightly different from OLS. The GLS solution differs in producing an estimate of fit[16] somewhat higher than OLS and an associated reduction in standard error of one of the coefficients. The R^2, as advertised, is a compromise between OLS and LSDV. In effect the error components model has

[15] This is the common cross-section-only version of error components, of which the full model may be extended to both space and time.

[16] The R^2 used with both GLS estimators is the R^2 estimate developed in Buse (1973).

TABLE 3

An Error Components (GLSE) Model of Party Issue Polarization in the U.S. House of Representatives

	Estimates and Standard Errors
1964 election intervention (ω_o)	4.80
	(0.28)
Dynamic growth parameter (δ_1)	0.83
	(0.08)
Constant (pre-intervention mean)	3.45
	(0.49)
R^2 (adjusted)	.602
$\hat{\rho}$ [a]	.35
Estimated equilibrium level	
$\omega_o/(1-\delta_1)$	28.24

[a] $\hat{\rho} = \sigma_u^2/\sigma^2$.

capitalized upon our knowledge of the magnitude of "between" effects—the specific ignorance—to produce the same parameter estimates as OLS, but smaller standard errors and a 25 percent higher estimate of explained variance.

IV. GLS-ARMA

The estimators examined thus far may be distinguished by the a priori assumptions regarding between-unit effects they require. OLS requires us to assume none. LSDV is almost the converse; it presumes that unit effects are present, without making a priori claims about their size or direction. Because the units have individual histories, they are expected to differ in ways that might be explained. GLSE takes that loose expectation a step further in its presumption that unit effects are random—or at least may be treated as such. GLS-ARMA differs from the others in presuming that unit effects are partially specified. With a small number of units and greater relative knowledge of each, the analyst in this case should be prepared to postulate which units would be expected to differ from the norm and in what direction. But that specification is only partial; we expect sometimes to be wrong, and when we are right we have not so much tested a postulated model as simply made the search for a satisfactory model more constrained and efficient.

Partial specification entails incremental modeling. We expect the initial specification to take us part of the way to a final model, but weak a priori specification in a complex world is unlikely to produce an optimal model. And with pooled data a less than optimal fit in either dimension threatens to bias the other. Thus an incremental process—specify, estimate, diagnose, respecify, reestimate, rediagnose, . . . —takes advantage of what we think we know to move toward a specification beyond our initial theory. The strategy requires sensitive diagnoses and, for this particular problem, pooled estimation provides three clean indicators. If a particular unit is correctly specified, (1) its summed residuals (over time) should approximate zero, (2) its residual variance should be in reasonable proportion to the other units, and (3) the pattern of autocorrelated error should be stationary, indicating stochastic fluctuation around the regression prediction rather than deterministic errors arising from systematic under-or overprediction.

From extensive experience with a variety of independent data, most particularly parallel data for the U.S. Senate, I might be prepared to specify probable region effects quite completely. But to do so in this case would eliminate the illustration of the iterative model building. Instead, let us presume atypically little knowledge of the unit effects to illustrate what can be learned from pooled diagnoses. Table 4 is a baseline for the procedure; it presents regional breakdowns of patterns of error from the "no region effects" OLS model of Table 2.

The relative robustness of estimators thus far observed has not prepared us for the evidence of Table 4, which shows a fit which is very poor within regions. Summed residuals do not approximate zero; just how far off they are can be best appreciated in contrast to the better fits to come. The residual variance ratio, a ratio of actual to expected variation in each unit, shows a relatively uniform fit, deceptive because it is uniformly poor. Most important, residual autocorrelation is both high and nonstationary. Both within several regions and in the pooled sample estimate, the pattern of the autocorrelation function is flat; it shows no tendency to decay toward zero after several lags. The evidence, in sum, demonstrates cross-sectional underspecification.

That the estimator is relatively robust in the face of serious underspecification is *not* generalizable. It occurs here because we have a special case of purely time-serial covariates. Were any cross-sectional variables present in the model, they would become proxies for the omitted region effects, inducing very serious distortions.

Both a priori knowledge and diagnostic evidence suggest similar respecifications. We would expect polarization (nothing more than between-party issue difference) to be low in the South (because Democrats there are relatively conservative) and in New England (because Republicans there are relatively liberal) and high in the West, where other data have indicated a

TABLE 4

Diagnostic Information for the OLS Specification of Table 2

	Summed Residuals[a]	Residual Variance Ratio[b]	Autocorrelations Lag 1	Autocorrelations Pattern
New England	−223.	1.48	.71	flat
Middle Atlantic	−18.	0.30	.27	decay
East North Central	94.	0.55	.26	decay
West North Central	36.	0.58	.22	decay
South	−228.	1.23	.76	flat
Border	−107.	0.95	.45	decay
Mountain	174.	1.24	.40	flat
Pacific	247.	1.65	.63	flat
ρ			.55	flat

[a] $\sum_{t=1}^{24} u_{it}$.

[b] Residual variance ratio is the ratio of the actual residual variance for the ith cross section (σ_i^2) divided by the expected variance (σ^2/n).

strong issue polarization that precedes the national salience of the issue. Thus for our first GLS-ARMA specification dummy variables are introduced for three regions, New England, South, and Pacific.

Table 5 displays both the estimates for specification 1 and diagnostic evidence for the next step. Fitting the three region dummy variables clearly improves model specification a great deal, most notably in reducing residual autocorrelation (from .55 to .35) and eliminating nonstationarity. Comparing R^2 is less clean because the GLS estimate in the ARMA case is consistently less than the OLS R^2, (.630 for this model). That is exactly the opposite pattern we just observed with GLSE, where the GLS estimate ran consistently higher than its OLS counterpart. The GLS R^2 estimate must be interpreted with caution.

The success of this initial specification is most evident in residual diagnostics. Two regions not yet modeled, the Border and Mountain states, show evidence of continuing underspecification, but the others are considerably improved. The addition of dummy variables produces approximately zero summed residuals for our three cases and reduces the autocorrelation of error. The exception is New England, where autocorrelation is reduced

TABLE 5

Party Issue Polarization in the U.S. House of Representatives
GLS-ARMA Specification 1: Dummies for
New England, South, and Pacific

	Estimates and Standard Errors	Cross-Section Diagnostics			
		Summed Residuals	Residual Variance Ratio	Autocorrelations Lag 1	Pattern
1964 election intervention (ω_0)	4.42 (0.37)				
Dynamic growth parameter (δ_1)	0.84 (0.08)				
Constant	5.44 (1.54)				
New England	−10.68 (3.16)	2.5	1.07	.47	flat
Middle Atlantic		−49.0	0.47	.35	decay
East North Central		62.9	0.69	.15	none
West North Central		5.2	0.80	.19	none
South	−10.70 (2.55)	−1.83	0.69	.35	decay
Border		−137.6	1.53	.49	decay
Mountain		143.5	1.58	.31	decay
Pacific	8.91 (3.10)	2.3	1.18	.16	none
R^2 (adjusted)	.468				
ρ	.352				
Estimated equilibrium level	27.63				

but still significant and still nonstationary even though the model is correctly predicting its central tendency.

In specification 2 (Table 6) additional dummy variables are introduced for the Border and Mountain regions. The Border region shows the same pattern of overprediction (i.e., negative residuals) as the adjoining South, while the Mountain states resemble the adjoining Pacific region in higher-than-normal polarization. The pattern is weaker in both cases than for adjoining regions and stands out cleanly only when the stronger adjoining

TABLE 6

Party Issue Polarization in the U.S. House of Representatives
GLS-ARMA Specification 2: Dummies for
New England, South, Pacific, Border, and Mountain

	Estimates and Standard Errors	Cross-Section Diagnostics			
		Summed Residuals	Residual Variance Ratio	Autocorrelations Lag 1	Pattern
1964 election intervention (ω_0)	4.70				
	(0.38)				
Dynamic growth parameter (δ_1)	0.83				
	(0.08)				
Constant	5.17				
	(1.64)				
New England	−10.68	−211.8	1.95	.47	flat
	(3.16)				
Middle Atlantic		−49.7	0.46	.35	decay
East North Central		62.2	0.67	.14	none
West North Central		4.5	0.79	.19	none
South	−10.74	−1.69	0.70	.36	decay
	(2.64)				
Border	−5.91	3.6	1.12	.43	decay
	(3.14)				
Mountain	5.87	1.8	1.15	.08	none
	(3.19)				
Pacific	8.90	1.7	1.14	.16	none
	(3.15)				
R^2 (adjusted)	.488				
ρ	.308				
Estimated equilibrium level	27.65				

patterns have been modeled. Addition of the two dummy variables produces a modest improvement in model specification seen, for example, in a further reduction of autocorrelation.

The diagnostic evidence of Table 6 suggests that with one exception the model may now be said to predict uniformly across regions. The three industrial heartland regions (Middle Atlantic, East North Central, and West

North Central) are well predicted without necessity of dummy variables as are four of the five regions which have been dummied. The exception is New England, which was suspicious after the first specification and decidedly worse in the second. The "problem" with New England is not that its intercept is lower than the norm—the dummy variable is capable of modeling that—but that the within-region pattern interacts with the intervention component of the model. Not lower than the norm in issue polarization at all times, New England *becomes* lower when the norm shifts. Before 1964 Republicans from all regions except the South were moderate to liberal on racial issues. The big majorities on early civil rights legislation in the House were formed when the Republicans could be counted upon to deliver nearly all their votes in alliance with northern Democrats. What appears to be the case in New England is that its Republicans became deviant after 1964 by sticking to their traditional liberalism while the party elsewhere moved strongly to the racial right. To deal with this space/time interaction, New England is modeled as a step function (zero through 1964, one after) in place of the time constant dummy variable we employed elsewhere.

The modest revision of the specification for New England has substantial consquences, as seen in Table 7. Not only does it clean up the specification problem for the region, it has some notable effects upon the whole model. The intervention component grows stronger; evidently the New England polarization in the wrong direction suppressed the national pattern. Improvement in fit (from .49 to .55) is notable for such a small change as is the continuing reduction in autocorrelation (from .31 to .27).

Specification 3 of Table 7 would seem to be a satisfactory specification. That is a judgment call. It is always possible to continue to estimate more parameters, add new wrinkles, and so forth, but such tailoring to fit the data is not conducive to parsimonious explanation.[17] The diagnostics point to no obvious specification problems. Indeed, autocorrelation and heteroscedasticity (seen in the residual variance ratio) are both reduced to such a level that GLS estimation is probably unnecessary. The specification of Table 7 can be estimated with ordinary least squares assumptions with very similar parameter estimates, but higher R^2 (.68) and reduced standard errors.

GLS-ARMA improves a bit on the dummy variable interpretation problem that we encountered with LSDV. The improvement is that the iterative modeling procedure, by focusing on successive elimination of outliers, tends to produce a natural reference category (or set of categories). In this case the three-region industrial heartland becomes the reference

[17] The most substantial improvements to be made in model specification at this point are in the time dimension. Providing for a "backlash" effect during the 90th Congress, for example, produces a substantial improvement in fit. But that is yet another topic in an already lengthy paper.

TABLE 7

Party Issue Polarization in the U.S. House of Representatives
GLS-ARMA Specification 3: Dummies for
South, Pacific, Border, and Mountain Plus New England Interaction

	Estimates and Standard Errors	Cross-Section Diagnostics			
		Summed Residuals	Residual Variance Ratio	Autocorrelations Lag 1	Pattern
1964 election intervention (ω_0)	4.97				
	(0.36)				
Dynamic growth parameter (δ_1)	0.83				
	(0.08)				
Constant	4.17				
	(1.46)				
New England interaction	−14.31	−19.5	1.07	.27	decay
	(3.30)				
Middle Atlantic		−43.7	0.54	.36	decay
East North Central		68.2	0.80	.17	none
West North Central		10.4	0.92	.20	none
South	−10.50	−1.4	0.86	.38	decay
	(2.51)				
Border	−5.64	3.0	1.24	.42	decay
	(2.85)				
Mountain	6.13	1.58	1.32	.06	none
	(2.94)				
Pacific	9.16	1.42	1.24	.10	none
	(2.84)				
R^2 (adjusted)	.549				
ρ	.270				
Estimated equilibrium level	29.23				

category because it was so close to the national norm that it required no special attention. That makes the signs of the modeled cross sections interpretable as the direction of deviation. Tests of significance for the dummy coefficients still may be misleading because of the influence of reference categories, but in contrast to LSDV, where they can easily lead to flatly false inferences, the degree of interpretative caution required here is small.

"The Race Is Over! . . . But Who Has Won?"[18]

In a race with four contestants it is tempting to assess the order of finish, which estimators won, which lost, and how far apart did they run. That can be done for our single illustration. But given the immense variety of combinations of design, data, and hypotheses, no single illustration can support much generalization. For that, we may refer to Table 1.

Perhaps the single most interesting "finding" is that the contestants did not finish far apart. How far depends in large measure upon how much importance is attached to the region variables vis-à-vis the covariate intervention. If attention is restricted to the latter, then it is notable that four different estimators would each have produced essentially the same inferences about the dynamics of issue polarization. The estimates are notably similar. But if our focus shifts to the units, then the four estimators differ substantially. And as noted earlier, the robustness of covariate estimates is highly dependent upon model specification. Other reasonable specifications would not be similarly robust.

The answer to "who has won?" is, of course, all of the above. Each was running a different race. There are design and situational contexts where each is the estimator of choice. Where units are homogeneous and time-serial disturbances not serious, OLS has desirable properties and is the estimator of choice. Where unit effects are present and concern for estimator efficiency is not paramount (e.g., when estimating from very large samples), LSDV is a reasonable choice. With random unit effects, the error components estimator is the efficient choice, the price of efficiency being possible bias from the correlation of covariates and omitted unit effects. And where design is time dominant—and particularly if it is also dynamic—GLS-ARMA is the estimator that can deal with time-serial threats to inference.

Ignorance is sometimes a blessing. To deal with the complications of pooled design in detail makes us painfully aware of a plethora of potential problems. All of them can easily be resolved with the expedient of sticking to straight cross-sectional or time-serial design. But dealing with space and time together carries with it the possibility of insights into the political world—and into the political world as seen through one-dimensional designs—that make it sometimes worth its price.

Manuscript submitted 10 October 1984
Final manuscript received 15 April 1985

[18] The subtitle is borrowed from Nerlove and, in turn, from *Alice's Adventures in Wonderland.* See note 4.

REFERENCES

Abramson, P., J. Aldridge, and D. Rohde. 1982. *Change and continuity in the 1980 elections*. Washington: Congressional Quarterly Press.

Anderson, T. W., and Cheng Hsiao. 1982. Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18 (January):47–82.

Balestra, P., and M. Nerlove. 1966. Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica*, 34 (July):585–612.

Baltagi, Badi H. 1981. Pooling: An experimental study of alternative testing and estimation procedures in a two-way error component model. *Journal of Econometrics*, 17 (September):21–49.

Beck, P. Forthcoming, 1984. Choice context and consequences: Beaten and unbeaten paths toward a science of politics. In H. Weisberg, ed., *Political science: The science of politics*. New York: Agathon.

Box, G. E. P., and G. M. Jenkins. 1976. *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Box, G. E. P., and G. C. Tiao. 1975. Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70 (March):70–79.

Brown, M. C., and C. N. Halaby. 1982. Pooling time series and cross-sectional data: estimating the "cost" of political machines. Paper presented at the 1982 annual meeting of the American Political Science Association, Denver, September 1982.

Brunner, Ronald D., and Klaus Liepelt. 1972. Data analysis, process analysis, and system change. *American Journal of Political Science*, 17 (November):538–69.

Buse, A. 1973. Goodness of fit in generalized least squares estimation. *American Statistician*, 27 (June):106–8.

Carmines, Edward G., Steven Renten, and James A. Stimson. 1984. Events and alignments: The party image link. In R. Niemi and H. Weisberg, eds., *Controversies in voting behavior*, 2d ed. Washington, D.C.: Congressional Quarterly Press: pp. 545–60.

Carmines, Edward G., and James A. Stimson. 1981. Issue evolution, population replacement, and normal partisan change. *American Political Science Review*, 75 (March):107–18.

———. 1984. The dynamics of issue evolution: The United States. In R. Dalton, S. Flanagan, and P. Beck, eds., *Electoral change in advanced industrial democracies*. Princeton: Princeton University Press: pp. 134–58.

———. Forthcoming, 1985. The politics and policy of race in Congress. In G. Wright, L. Rieselbach, and L. Dodd, eds., *Congress and policy change* (New York: Agathon Press).

Doreian, Patrick. 1981. Estimating linear models with spatially distributed data. In S. Leinhardt, ed., *Sociological methodology 1981*. San Francisco: Jossey-Bass: pp. 359–88.

Garreau, Joel. 1981. *The nine nations of North America*. Boston: Houghton Mifflin.

Key, V. O. 1949. *Southern politics in state and nation*. New York: Knopf.

Kiefer, Nicholas M. 1980. Estimation of fixed effect models for time series of cross-sections with arbitrary intertemporal covariance. *Journal of Econometrics*, 14 (October):195–202.

Kmenta, Jan. 1971. *Elements of econometrics*. New York: Macmillan.

Koutsoyiannis, A. 1973. *Theory of econometrics*. London: Macmillan.

Lockard, Duane. 1959. *New England state politics*. Princeton: Princeton University Press.

McCleary, Richard, and Richard A. Hay, Jr. 1980. *Applied time series analysis*. Beverly Hills: Sage.

Maddala, G. S. 1971. The use of variance components models in pooling cross section and time series data. *Econometrica*, 39 (March):341–58.

Maddala, G. S., and T. D. Mount. 1973. A comparative study of alternative estimators for variance components models used in econometric applications. *Journal of the American Statistical Association*, 68 (June):324–28.

Mayor, T. H., and L. R. Pearl. 1984. Life-cycle effects, structural change and long-run movements in the velocity of money. *Journal of Money, Credit, and Banking*, 16 (May):175–84.

Mundlak, Yair. 1978. On the pooling of time series and cross section data. *Econometrica*, 46 (January):69–85.

Nerlove, Marc. 1971a. A note on error components models. *Econometrica*, 39 (March):383–96.

———. 1971b. Further evidence on the estimation of dynamic economic relations from a time series of cross sections. *Econometrica*, 39 (March):359–82.

Sundquist, James L. 1973. *Dynamics of the party system*. Washington, D.C.: Brookings Institution.

Taylor, William E. 1980. Small sample considerations in estimations from panel data. *Journal of Econometrics*, 13 (June):203–23.

Wallace, T. D., and A. Hussain. 1969. The use of error components models in combining cross section with time series data. *Econometrica*, 37 (January):55–72.

Wansbeek, Tom, and Arie Kapteyn. 1982. A class of decompositions of the variance-covariance matrix of a generalized error components model. *Econometrica*, 50 (May):713–24.

Zuk, G., and W. R. Thompson. 1982. The post-coup military spending question: A pooled cross-sectional time series analysis. *American Political Science Review*, 76 (March):60–74.