Pooling Disparate Observations

Larry M. Bartels

American Journal of Political Science, Vol. 40, No. 3 (Aug., 1996), 905-942.

Stable URL:
http://links.jstor.org/sici?sici=0092-5853%28199608%2940%3A3%3C905%3APDO%3E2.0.CO%3B2-I

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

*Pooling Disparate Observations**

Larry M. Bartels, *Princeton University*

Theory: Classical statistical inference takes as given the population governed by a posited statistical model and associated set of parameters. But social theories seldom include clear specifications of the populations to which they are supposed to be applicable, so data analysts frequently face difficult choices about which observations to include in their analyses.
Hypotheses: Conventional approaches to selecting relevant observations are likely either to underexploit the available data (by discarding problematic observations that could provide some information about the parameters of interest) or to overexploit the available data (by estimating alternative models and interpreting the "best" results as though they were produced in accordance with the standard assumptions of classical statistical inference).
Methods: I propose a technique, dubbed "fractional pooling," which provides a simple and coherent way either to incorporate prior beliefs about the theoretical relevance of disparate observations or to explore the implications of prior uncertainty about their relevance. The technique is easy to implement and has a plausible rationale in Bayesian statistical theory.
Results: I illustrate the potential utility of fractional pooling by applying the technique to political data originally analyzed by Ashenfelter (1994), Powell (1982), and Alesina, Londregan, and Rosenthal (1993). These examples demonstrate that conventional approaches to analyzing disparate observations can sometimes be seriously misleading, and that the approach proposed here can enrich our understanding of the inferential implications of unavoidably subjective judgments about the theoretical relevance of available data.

How to choose the set of observations to which a statistical model should be applied is one of the least understood aspects of model specification.

*This work was presented at the 1994 Political Methodology Summer Meeting in Madison and at the 1994 Annual Meeting of the American Political Science Association in New York, and to seminars at Princeton University and the University of Rochester. I am grateful to Christopher Achen, R. Michael Alvarez, Janet Box-Steffensmeier, Simon Jackman, Gary King, Renée Smith, and anonymous referees for especially helpful comments, and to G. Bingham Powell for publishing his data (Powell 1982), to Orley Ashenfelter for including his data in an unpublished report (Ashenfelter 1994), and to Howard Rosenthal for providing unpublished data analyzed by Alesina, Londregan, and Rosenthal (1993). The research reported here was originally stimulated by some comments of Nathaniel Beck's (1985). Douglas Rivers pointed out an important error in an earlier version of the analysis in Section 1.2 below when I presented it at the 1987 Political Methodology Summer Meeting in Durham. Christopher Achen introduced me to the work of Edward Leamer (1978) at an early age, and pointed out an important error in my understanding of contract curves like the one introduced in Section 2.2 below when I displayed it on a final exam. I am especially pleased to have an opportunity to acknowledge these three long-standing intellectual debts.

American Journal of Political Science, Vol. 40, No. 3, August 1996, Pp. 905–42

Problems of this sort arise both in time series analyses (for example, in deciding whether to pool observations across eras, presidential administrations, or measurement regimes) and in cross-sectional analyses (for example, in deciding whether to pool observations from different opinion surveys, households, locales, or political systems).

At the most basic level, our problem is the fundamental problem of induction: what, if anything, entitles us to make inferences about the behavior of an individual, nation, or other unit on the basis of the observed behavior of some different unit, or of the same unit at some different point in time or in some different context, or of some different unit at some different point in time or in some different context? The answer can only be, *a prior belief* in the *similarity* of the bases of behavior across units or time periods or contexts. In the case of regression analysis, our prior belief is embodied in the assumption that the relevant observations represent a single population, in the sense that the underlying regression parameters of interest apply equally to all the observations.

Our practical problem is that conventional (classical) statistical techniques are quite inflexible in representing what may be a rather complicated set of relevant prior beliefs. At some point, classical techniques force us either to act *as if* the relevant observations were governed by the same causal process, or to act *as if* the relevant observations were governed by wholly unrelated causal processes. We take the former stance when we include the relevant observations in a single regression anlaysis, and the latter stance when we estimate separate regressions in different subsets of the whole set of potentially relevant observations.[1]

My aims in this article are to trace the inferential implications of this inflexibility and to propose an alternative approach that allows for a richer range of assumptions about the theoretical relevance of the available data.[2]

[1] This point can be fudged in various ways (for example, by including a more or less complicated set of interaction terms that allow some parameters to vary across observations while others remain constant), but it cannot be avoided: in the end, whatever model is specified, each available observation must be either all the way in or all the way out of the analysis. Switching regime models (Quandt 1958; Goldfeld and Quandt 1973) and stochastic parameter regression models (Beck 1983; Newbold and Bos 1985) likewise provide additional flexibility but do not obviate the classical demand to specify categorically the set of observations for which a single stochastic mechanism will be assumed to govern the underlying parameters. Bayesian approaches to these models are presented by Swamy and Mehta (1975) and Leamer (1978, sec. 8.3-8.5), respectively.

[2] My project here shares an affinity with the superficially unrelated project of Collier and Mahon (1993), who studied how analytic categories get "stretched" to fit new contexts in comparative research. In each case, the general goal is to make appropriate use of theoretically problematic data.

The proposed technique can be thought of either as a tool for exploring through sensitivity analysis the implications of prior uncertainty about which observations "belong" in the analysis, or as a tool for Bayesian analysis incorporating prior beliefs generated from theoretically problematic data in a simple but plausible way. The basic dilemma of identifying theoretically relevant observations and its implications are described in Section 1, and my alternative approach is described and justified in Section 2. In Section 3 I use three empirical examples to further explore the issues raised in Section 1 and the approach proposed in Section 2. Section 4 concludes with some practical observations and recommendations for data analysts deciding (as all data analysts must decide)[3] whether and how to pool disparate observations.

1. The Dilemma of Disparate Observations

How do data analysts decide which observations to include in their analyses? What are the inferential implications of those decisions? I address these questions in the context of a simple regression model with two sets of available data, one unproblematic (in which the parameters of theoretical interest are "known" to apply) and the other problematic (governed by a "similar," but not necessarily identical, set of parameters). I examine the properties of the parameter estimates produced by three distinct estimation strategies: (1) analyzing the two sets of data separately, (2) pooling all of the available data in a single analysis, and (3) adopting one or the other of these approaches depending upon the results of a preliminary comparison of the results they produce. Each of these three strategies will be shown to entail significant inferential difficulties.

1.1. Model, Assumptions, and Basic Results

We begin with the dual regression model

$$\mathbf{y}_0 = \mathbf{X}_0\boldsymbol{\alpha} + \mathbf{u}_0 \qquad [1a]$$

$$\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{u}_1 \qquad [1b]$$

where $\mathbf{y}_0$ and $\mathbf{y}_1$ are $N_0 \times 1$ and $N_1 \times 1$ vectors, respectively, of observa-

[3] It should be obvious that I agree with the second half but not the first half of Beck's claim (1985, 79) that "Survey researchers usually analyze the entire sample but the time series analyst always faces a choice." Even when it is clear what "the entire sample" is (all 37,456 cases in the 1952–92 American National Election Studies cumulative data file?), good survey researchers seldom analyze it, choosing instead a smaller set of relevant observations (just as time series analysts often do) on the basis of theoretical considerations and data availability.

tions on a common dependent variable, $\mathbf{X}_0$ and $\mathbf{X}_1$ are $N_0 \times K$ and $N_1 \times K$ matrices, respectively, of observations on a common set of K explanatory variables, $\mathbf{u}_0$ and $\mathbf{u}_1$ are $N_0 \times 1$ and $N_1 \times 1$ vectors, respectively, of unobserved stochastic disturbances, and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $K \times 1$ vectors of constant parameters to be estimated.

By convention, $\boldsymbol{\beta}$ will be treated as the parameter vector of theoretical interest in the subsequent analysis: our aim will be to estimate $\boldsymbol{\beta}$ as accurately as possible, and the available data will be relevant if and only if they contribute to that aim. The N_1 observations in the "$\boldsymbol{\beta}$ regime" will be treated as having been drawn from a population in which model [1b] is assumed to hold, while the N_0 observations in the "$\boldsymbol{\alpha}$ regime" will be treated as being of uncertain theoretical relevance, in the sense that the parameter vector $\boldsymbol{\alpha}$ in model [1a] is believed *a priori* to be "similar," but not necessarily identical, to the parameter vector $\boldsymbol{\beta}$ in model [1b].

We shall assume throughout the analysis that this model is well specified, in the sense that

$$E(\mathbf{X}_0'\mathbf{u}_0) = \mathbf{0}; E(\mathbf{u}_0\mathbf{u}_0') = \sigma_0^2\mathbf{I}, \qquad [2a]$$

$$E(\mathbf{X}_1'\mathbf{u}_1) = \mathbf{0}; E(\mathbf{u}_1\mathbf{u}_1') = \sigma_1^2\mathbf{I}, \qquad [2b]$$

and

$$E(\mathbf{u}_0\mathbf{u}_1') = \mathbf{0}. \qquad [2c]$$

How should we use the available data to estimate the parameters of interest? At one extreme, we could treat $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as completely distinct parameter vectors to be estimated independently using ordinary least squares regression. The OLS estimators are

$$\mathbf{a} = (\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{y}_0 \qquad [3a]$$

and

$$\mathbf{b} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}_1. \qquad [3b]$$

Under the assumptions given in expression [2], each of these estimators is unbiased for the corresponding parameter vector, and their respective covariance matrices are

$$\text{var}(\mathbf{a}) = \sigma_0^2(\mathbf{X}_0'\mathbf{X}_0)^{-1} \qquad [4a]$$

and

$$\text{var}(\mathbf{b}) = \sigma_1^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}. \quad [4b]$$

At the other extreme, we could treat the "similar" parameter vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as indentical (and σ_0^2 and σ_1^2 as identical disturbance variances).[4] In that case, we would want to combine the two data sets and use ordinary least squares regression to estimate a single parameter vector. The pooled OLS estimator is

$$\mathbf{b}_p = (\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}(\mathbf{X}_0'\mathbf{y}_0 + \mathbf{X}_1'\mathbf{y}_1), \quad [5]$$

the expectation of $\mathbf{b}_p$ is

$$E(\mathbf{b}_p) = \boldsymbol{\beta} + (\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_0'\mathbf{X}_0(\boldsymbol{\alpha} - \boldsymbol{\beta}), \quad [6]$$

and the covariance matrix of $\mathbf{b}_p$ is

$$\text{var}(\mathbf{b}_p) = \sigma^2(\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}. \quad [7]$$

Finally, we shall build in what follows upon an important relationship between the pooled regression estimator $\mathbf{b}_p$ and the subset regression estimators **a** and **b.** Rewriting expression [5], and making use of [3a] and [3b],

$$\begin{aligned}\mathbf{b}_p &= (\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}[\mathbf{X}_0'\mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{y}_0 + \mathbf{X}_1'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}_1] \\ &= (\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}(\mathbf{X}_0'\mathbf{X}_0\mathbf{a} + \mathbf{X}_1'\mathbf{X}_1\mathbf{b}).\end{aligned} \quad [8]$$

In words, $\mathbf{b}_p$ is a matrix-weighted average of the separate OLS parameter vectors **a** and **b**, where the weight matrix associated with each parameter vector is proportional to the inverse of the covariance matrix of that parameter vector.

1.2. A Mean Squared Error Analysis

Expression [6] is sufficient to demonstrate that pooling disparate observations runs the risk of biasing our estimates of the parameters of theoretical

[4] From this point on, I shall add to the assumption in [2] the further assumption that $\sigma_0^2 = \sigma_1^2 = \sigma^2$. This assumption is not especially plausible in most applications, but simplifies the exposition significantly. More general versions of most of the subsequent results can be derived without this additional assumption, but with less clarity and no appreciable gain in insight. Heteroskedasticity provides a technical rationale for differential weighting of the data from the two regimes logically distinct from the theoretical rationale suggested below.

interest, $\boldsymbol{\beta}$. If our only aim were to avoid bias, we would always prefer the subset regression estimator $\mathbf{b}$ to the pooled regression estimator $\mathbf{b}_p$. But of course, we might then question the assumption that all of the N_1 observations used to estimate $\mathbf{b}$ really come from the same regime, and prefer an estimator based upon a subset of this subset of the data, or on a subset of a subset of a subset. Obviously, this logic leads inexorably to very small data sets, and thus to very imprecise parameter estimates. A more reasonable intuition suggests that we must weigh potential gains in the precision of parameter estimates from including theoretically problematic observations against the bias engendered when the underlying parameter values governing the problematic observations differ significantly from those governing the observations of primary theoretical interest.

A natural way to formalize this intuition is by specifying the conditions under which a more inclusive data set will be preferable by a coefficient mean squared error criterion to a less inclusive data set. The mean squared error criterion balances the competing demands of unbiasedness and precision. In the present context, the pooled regression estimator $\mathbf{b}_p$ is superior to the subset regression estimator $\mathbf{b}$ by a generalized mean squared error criterion if the mean squared error for every possible linear combination of the elements of $\mathbf{b}_p$ is less than or equal to the mean squared error for the same combination of the elements of $\mathbf{b}$ (Judge et al. 1985, 47–8). This will be true if the difference between the generalized mean squared error matrices

$$\mathrm{E}[(\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})'] - \mathrm{E}[(\mathbf{b}_p - \boldsymbol{\beta})(\mathbf{b}_p - \boldsymbol{\beta})'] = \boldsymbol{\Delta} \qquad [9]$$

is a positive definite matrix (so that $\boldsymbol{\omega}'\boldsymbol{\Delta}\boldsymbol{\omega} > 0$ for any $K \times 1$ weight vector $\boldsymbol{\omega} \neq \mathbf{0}$). Conversely, $\mathbf{b}$ is superior to $\mathbf{b}_p$ if $\boldsymbol{\Delta}$ is negative definite (so that $\boldsymbol{\omega}'\boldsymbol{\Delta}\boldsymbol{\omega} < 0$ for any $\boldsymbol{\omega} \neq \mathbf{0}$).

This generalized mean squared error matrix of $\mathbf{b}_p$ is the sum of the covariance matrix of $\mathbf{b}_p$ and the bias squared matrix,

$$\mathrm{E}[(\mathbf{b}_p - \boldsymbol{\beta})(\mathbf{b}_p - \boldsymbol{\beta})'] = \mathrm{var}(\mathbf{b}_p) + [\mathbf{E}(\mathbf{b}_p) - \boldsymbol{\beta}][\mathbf{E}(\mathbf{b}_p) - \boldsymbol{\beta}]'. \qquad [10]$$

Substituting from expressions [6] and [7] in Section 1.1,

$$\begin{aligned}\mathrm{E}[(\mathbf{b}_p - \boldsymbol{\beta})(\mathbf{b}_p - \boldsymbol{\beta})'] = {} & \sigma^2(\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1} \\ & + (\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_0'\mathbf{X}_0(\boldsymbol{\alpha} - \boldsymbol{\beta}) \\ & \times (\boldsymbol{\alpha} - \boldsymbol{\beta})'\mathbf{X}_0'\mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}.\end{aligned} \qquad [11]$$

Since the subset regression estimator $\mathbf{b}$ is unbiased $\boldsymbol{\beta}$, its generalized mean

squared error matrix is just its covariance matrix,

$$E[(\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})'] = \text{var}(\mathbf{b}). \quad [12]$$

Substituting from expression [4],

$$E[(\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})'] = \sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}. \quad [13]$$

Thus, the difference matrix for the generalized mean squared error comparison is

$$\begin{aligned}\boldsymbol{\Delta} = {} & \sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1} - \sigma^2(\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1} \\ & - (\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_0'\mathbf{X}_0(\boldsymbol{\alpha} - \boldsymbol{\beta}) \\ & \times(\boldsymbol{\alpha} - \boldsymbol{\beta})'\mathbf{X}_0'\mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}.\end{aligned} \quad [14]$$

By some straightforward matrix algebra,

$$\begin{aligned}\boldsymbol{\Delta} = {} & (\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_0'\mathbf{X}_0[\sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1} + \sigma^2(\mathbf{X}_0'\mathbf{X}_0)^{-1} \\ & - (\boldsymbol{\alpha} - \boldsymbol{\beta})(\boldsymbol{\alpha} - \boldsymbol{\beta})']\mathbf{X}_0'\mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}.\end{aligned} \quad [15]$$

This difference matrix will be positive definite if the data matrices $\mathbf{X}_0'\mathbf{X}_0$, $\mathbf{X}_1'\mathbf{X}_1$, and $(\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)$ have full rank K[5] and the matrix in square brackets is positive definite (Judge and Bock 1978, 316, Theorem A.3.7). Roughly speaking, the latter condition will be satisified if $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are sufficiently similar; if each element of $\boldsymbol{\alpha}$ exactly equals the corresponding element of $\boldsymbol{\beta}$, the matrix $(\boldsymbol{\alpha} - \boldsymbol{\beta})(\boldsymbol{\alpha} - \boldsymbol{\beta})'$ is a matrix of zeroes, and the condition is obviously satisfied for any $\sigma^2 > 0$.

The mean squared error comparison becomes much simpler in the special case of a bivariate regression model (with mean-deviated data so that no intercept is required), since in that case $(\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}$, $\mathbf{X}_0'\mathbf{X}_0$, $\sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}$, $\sigma^2(\mathbf{X}_0'\mathbf{X}_0)^{-1}$, and $(\boldsymbol{\alpha} - \boldsymbol{\beta})(\boldsymbol{\alpha} - \boldsymbol{\beta})'$ are all positive scalars. Then the difference matrix $\boldsymbol{\Delta}$ is positive definite if

$$\sigma^2/\Sigma x_1^2 + \sigma^2/\Sigma x_0^2 - (\alpha - \beta)^2 > 0, \quad [16]$$

and negative definite if the inequality is reversed. The first two terms on the left side of this inequality are the variances of the least squares parameter

[5] This condition simply ensures that the vectors of parameter estimates **a**, **b**, and $\mathbf{b}_p$ referred to above can in fact be computed given the available data.

estimates a and b, respectively, and their sum is the variance of the difference between a and b (since the variance of the difference between two variables is equal to the sum of their variances minus twice the covariance, and the covariance here is equal to zero). Thus, we can rewrite condition [16] as

$$\text{var}(a - b) - (\alpha - \beta)^2 > 0, \qquad [17]$$

which can be rearranged to produce

$$-1 < (\alpha - \beta)/[\text{var}(a - b)]^{1/2} < 1. \qquad [18]$$

The middle term in expression [18] is the population value corresponding to a *t*-statistic for the difference between the separate subset parameter values a and b. The pooled regression coefficient b_p will be superior to the subset regression coefficient b by a mean squared error criterion if and only if this population value is less than one in absolute value (Wallace 1964; Feldstein 1973). Unfortunately, this population value is unknown. Thus, an analyst attempting to minimize the mean squared error of his or her parameter estimate must either judge the probability that condition [18] is satisfied on *a priori* grounds, or else rely on the corresponding sample *t*-statistic to make an inference about the population value of interest.[6] Even in the simple bivariate case, each of these approaches has its pitfalls. Prior beliefs may or may not be easily translated into bets about *t*-statistics, and using sample data to substitute for prior beliefs courts the inferential problems of pretest estimation outlined in Section 1.3. In any case, neither approach generalizes readily to the more usual situation in which there is more than one explanatory variable. Thus, although the mean squared error criterion is a useful guide in principle, it fails to provide a practical solution to the dilemma of disparate observations.

1.3. Pretest Estimation

It should be clear by now that, roughly speaking, it makes sense to pool disparate observations if the underlying parameters governing those observations are sufficiently similar, but not otherwise. A common practice among data analysts faced with situations like the one outlined here is to test the hypothesis that the parameter vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are *identical* by

[6] In most cases, the easiest way to generate the relevant sample *t*-statistic will be to regress $[\mathbf{y}_0'\mathbf{y}_1']'$ on $[\mathbf{x}_0'\mathbf{x}_1']'$ and $[\mathbf{x}_0'\mathbf{0}']'$. The parameters in this regression are β and $\alpha - \beta$, and the relevant *t*-statistic can be calculated simply by dividing the second parameter estimate by its standard error.

comparing the total sum of squared residuals from the two subset regressions (with N_0 and N_1 observations) with the sum of squared residuals from the pooled regression (with $N_0 + N_1$ observations). If the improvement in fit from estimating two sets of regression coefficients rather than one is sufficiently large, the null hypothesis of parameter equality is rejected and inference proceeds on the basis of the separate subset regression results. If the improvement in fit is not sufficiently large to reject the null hypothesis of parameter equality, inference proceeds on the basis of the pooled regression results. The test statistic

$$\{[SSR_p - (SSR_a + SSR_b)]/K\}/\{(SSR_a + SSR_b)/(N_0 + N_1 - 2K)\}$$

has an F distribution with K and $N_0 + N_1 - 2K$ degrees of freedom, where SSR_p is the sum of squared residuals from the pooled regression and SSR_a and SSR_b are the sums of squared residuals from the subset regressions in the α and β regimes, respectively.

In the simple case of a bivariate regression model, this *F*-statistic is just the *t*-statistic for $(\alpha - \beta)$ described at the end of Section 1.2. In the time-series setting, the usual *F*-test for structural change is commonly referred to as the "Chow test" (Chow 1960).[7] Other common variable selection criteria, such as Theil's adjusted R^2 statistic and the Mallows C_p statistic, are also based upon the sum of squared residuals, and can be interpreted as *F*-tests with appropriately chosen critical values (Edwards 1969; Amemiya 1980; Judge et al. 1985, chap. 21). One important disadvantage of the F-test is that it is only valid under the assumption that $\sigma_0^2 = \sigma_1^2 = \sigma^2$. There are more flexible alternative tests, however, including one that is asymptotically valid under heteroskedasticity of arbitrary form, based upon an auxiliary regression (Davidson and MacKinnon 1993, sec. 11.6).

As in other applications, this pretest estimation strategy has two main deficiencies. First, the choice of a significance level for the test of the constraints embodied in the pooled regression is essentially arbitrary, since there is no clear specification of the inferential costs of Type I and Type II errors. The mean squared error analysis in Section 1.2 suggests using whatever significance level produces a critical value for the test statistic of 1.0, at least in the simple bivariate case; but other considerations argue in

[7] The Chow test is based on the assumption that the potential structural change occurs at a known point in time. More general diagnostic tests for structural change at any unspecified point in the time series are also available (for example, Brown, Durbin, and Evans 1975), although it seems odd in many applications to maintain the assumption of a discrete structural shift while professing ignorance about the timing of that shift.

favor of a critical value closer to 2.0 (Judge et al. 1985, 77).[8] Choosing a .05 significance level, or any other conventional level, has no obvious justification. Indeed, as the empirical examples below will demonstrate, an F-test with a significance level much looser than .05 may fail to reject the null hypothesis even when the parameter estimates we care about differ substantially, while an F-test with a significance level of .001 may reject the null hypothesis even when the parameter estimates we care about are similar.

Second, and more importantly, standard statistical inferences based upon either the separate or pooled regression results will be misleading because they fail to reflect the specification uncertainty reflected in the first (model selection) phase of the pretest estimation strategy. Except in very simple cases, the real statistical properties of the pretest estimator will be unknown (Judge and Bock 1978). Although this is a substantial theoretical embarrassment, it is usually ignored in practice. As Leamer (1978, 130) put it,

> few researchers are willing to accept "peculiar" estimates, and the standard operating procedure is to search for constraints that yield "acceptable" estimates. The fact that the resulting estimator is neither unbiased, linear, nor "best" is no large deterrent to a person whose research project would be dubbed "fruitless" if it were summarized in a nonsensical estimate.

2. Fractional Pooling

We have seen in Section 1 that each of the usual approaches to the dilemma of disparate observations has significant flaws. If we simply discard problematic observations and base our analysis on data "known" to represent the regime of primary theoretical interest, we will avoid bias but our parameter estimates may be very imprecise. If we simply pool disparate observations, our parameter estimates will be more precise but possibly biased. If we allow the data to dictate our handling of disparate observations by adopting a pretest estimation strategy, we will suffer one or the other of these same unhappy fates while deceiving ourselves about how much we have actually learned from our data.

Intuitively, it would be desirable to have an estimation strategy that

[8] In any case, a thoroughgoing application of the mean squared error criterion argues against any pretest estimator of the sort considered here, since the pretest estimator is a discontinuous function of the data (at some point, a small change in the test statistic produces a discrete jump between the subset parameter estimate $\mathbf{b}$ and the pooled parameter estimate $\mathbf{b}_p$) and therefore inadmissible (Leamer 1978, 135). Feldstein (1973) and others have suggested continuous mixing schemes in which $\mathbf{b}$ and $\mathbf{b}_p$ are averaged, with weights that are a continuous function of the test statistic for the constraints embodied in the pooled regression.

relied heavily upon the observations "known" to represent the regime of primary theoretical interest, while discounting but not discarding completely the problematic data. The simplest way to do this is just to weight the problematic "α data" $\mathbf{X}_0$ and $\mathbf{y}_0$ less heavily than the "β data" $\mathbf{X}_1$ and $\mathbf{y}_1$ in our regression analysis. I propose here an estimator with exactly that feature, the fractionally pooled regression estimator

$$\mathbf{b}_\lambda = (\lambda^2\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}(\lambda^2\mathbf{X}_0'\mathbf{y}_0 + \mathbf{X}_1'\mathbf{y}_1), \qquad [19]$$

with λ a suitably chosen constant satisfying $0 \leq \lambda \leq 1$.

The fractionally pooled regression estimator $\mathbf{b}_\lambda$ is simply a weighted least squares estimator of the sort introduced in every textbook discussion of heteroskedasticity. But the motivation for weighting the available data in the present case is not that the disturbances associated with the various observations are heteroskedastic, but that the parameters governing some observations are only approximately the parameters of theoretical interest. In the former case we choose to discount data that are especially noisy; in the latter case we choose to discount data that are of problematic theoretical relevance.

The estimator $\mathbf{b}_\lambda$, like the pooled regression estimator $\mathbf{b}_p$, is a matrix-weighted average of the separate OLS parameter vectors $\mathbf{a}$ and $\mathbf{b}$:

$$\begin{aligned}\mathbf{b}_\lambda &= (\lambda^2\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1} \\ &\quad \times [\lambda^2\mathbf{X}_0'\mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{y}_0 + \mathbf{X}_1'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}_1] \qquad [20] \\ &= (\lambda^2\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}(\lambda^2\mathbf{X}_0'\mathbf{X}_0\mathbf{a} + \mathbf{X}_1'\mathbf{X}_1\mathbf{b}).\end{aligned}$$

But here the relative weight attached to the vector $\mathbf{b}$ estimated from the data of primary theoretical interest is greater than in the pooled regression estimator $\mathbf{b}_p$, while the relative weight attached to the vector $\mathbf{a}$ estimated from the theoretically problematic data is correspondingly smaller. By appropriate choice of the pooling fraction λ, we can represent any desired degree of confidence in the α data, from treating them at face value to ignoring them completely. It is easy to see, by comparing expression [20] with expression [8], that when $\lambda = 1$, the estimator $\mathbf{b}_\lambda$ reduces to the pooled regression estimator $\mathbf{b}_p$. At the other extreme, it is clear that when $\lambda = 0$, the estimator $\mathbf{b}_\lambda$ reduces to the subset regression estimator $\mathbf{b}$.

2.1. Implementation

Fractional pooling is easy to implement, requiring simply a weighted least squares regression:

(1) Multiply the N_0 observations in the $\boldsymbol{\alpha}$ subset for the dependent variable and each of the K explanatory variables (including the constant, if any) by the pooling fraction λ, where $0 \leqslant \lambda \leqslant 1$.
(2) Run a regression using the N_0 weighted observations from (1) together with the N_1 unweighted observations in the $\boldsymbol{\beta}$ subset.
(3) The coefficients produced by the regression in (2) are the parameter estimates $\mathbf{b}_\lambda$. However, the nominal standard errors of these parameter estimates (and the nominal standard error of the regression) are artificially small; the correct standard errors (and the correct standard error of the regression) can be recovered by multiplying the printed standard errors (and the printed standard error of the regression) by $[(N_0 + N_1 - K)/(\lambda N_0 + N_1 - K)]^{1/2}$.[9]

This approach can be generalized in the obvious way to deal with more complicated situations in which there are more than two subsets of available data. Indeed, there is no reason in principle why each observation cannot have its own associated weight, λ_n, reflecting the subjective theoretical relevance of that observation. However, my analysis here will continue to focus on the simple case considered so far, in which the relevant data can be categorized into two subsets, one of which we wish to discount by the factor λ.

2.2. *The Locus of Fractionally Pooled Parameter Estimates*

It is illuminating to observe how the parameter estimates produced by fractional pooling vary with the choice of the pooling fraction λ. As I have already indicated, when λ equals zero the parameter estimates are identical to those produced by the subset regression using only the observations from the $\boldsymbol{\beta}$ regime (the coefficient vector $\mathbf{b}$ in Section 1.1 above). At the opposite extreme, when λ equals one the parameter estimates are identical to those from the pooled regression using the observations from both the $\boldsymbol{\alpha}$ regime and the $\boldsymbol{\beta}$ regime (the coefficient vector $\mathbf{b}_p$ in Section 1.1). Intermediate values of λ produce a continuum of parameter estimates that can be represented as a curve connecting the points $\mathbf{b}$ and $\mathbf{b}_p$ in the K-dimensional parameter space.[10]

[9] The printed R^2 statistic will also be too large, but the required adjustment in this case is somewhat more complicated.

[10] Mathematically, this curve is equivalent to a portion of the "curve décolletage" of Dickey (1975), which Leamer (1978, sec. 3.3, 5.6) referred to as the "information contract curve." In their case the curve is a locus of possible compromises between Bayesian prior and sample information; here it is a locus of possible compromises between data from the disparate regimes. The analogy reflects the Bayesian rationale for fractional pooling developed in Section 2.3.

It is worth noting that the continuum of points along this curve will not, in general, lie "between" the endpoints in any given dimension.[11] Thus, it will not, in general, be sufficient to estimate the subset coefficient vector $\mathbf{b}$ and the pooled coefficient vector $\mathbf{b}_p$ and presume that "the truth must lie somewhere in between." Even when the curve approximates a straight line, moreover, distances along this line will not, in general, be proportional to the differences in values of λ that produce them. For example, the point corresponding to $\lambda = .5$ may lie near $\mathbf{b}$, near $\mathbf{b}_p$, or midway between these endpoints.[12] On the other hand, it is often easy enough to rerun the regression several times with alternative values of λ in order to explore the shape of the curve, and to report results corresponding to alternative values of λ in order to convey to readers the sensitivity of the analysis to the choice of λ.

2.3. Bayesian Rationale

My analysis so far has emphasized that intelligent decisions about how to treat disparate observations must be based, in one way or another, upon prior beliefs about the theoretical relevance of the available data. The main conceptual attraction of the fractional pooling approach proposed here is that it relies upon prior beliefs in a simple and natural way: the subjective relevance of the problematic data is summarized in a single fraction, λ. In this section I show that this way of incorporating prior beliefs is consistent with the more general precepts of Bayesian statistical theory.[13]

For the simple case of a linear regression model with normally distributed errors and a "natural conjugate" normal-gamma prior distribution for $(\boldsymbol{\beta}, \sigma^2)$, the Bayesian formula for combining prior and sample data is directly analogous to the classical formula for combining data from two separate samples (Leamer 1978, 76–9). For example, if the prior distribution for $\boldsymbol{\beta}$ given σ^2 in Equation [1b] above is normal with mean vector $\hat{\boldsymbol{\beta}}^*$ and covariance matrix $\sigma^2(\boldsymbol{\Omega}^*)^{-1}$ and the prior distribution for σ^2 is gamma with parameters $\hat{\sigma}^{*2}$ and ν^*, then the posterior distribution for $\boldsymbol{\beta}$ given σ^2 is normal with mean vector $\hat{\boldsymbol{\beta}}^{**}$ and covariance matrix $\sigma^2(\Omega^{**})^{-1}$ and the

[11] This point is nicely illustrated by the curve in Figure 2 in Section 3.2.

[12] For example, in Figure 3 in Section 3.3, the apparent effect of election year GNP growth on presidential election outcomes is almost invariant for values of λ greater than .5, but varies considerably for values of λ less than .3.

[13] The best surveys of Bayesian approaches to regression analysis and related econometric techniques are by Zellner (1971) and Leamer (1978). Western and Jackman (1994) recently argued for the utility of a Bayesian approach to cross-national comparative political research. The technique of "mixed estimation" proposed by Theil and Goldberger (1961) is practically similar, but less general and without the compelling theoretical justification of Bayesian methods.

posterior distribution for σ^2 is gamma with parameters $\hat{\sigma}^{**2}$ and ν^{**}, where

$$\hat{\boldsymbol{\beta}}^{**} = (\boldsymbol{\Omega}^* + \mathbf{X}_1'\mathbf{X}_1)^{-1}(\boldsymbol{\Omega}^*\hat{\boldsymbol{\beta}}^* + \mathbf{X}_1'\mathbf{X}_1\mathbf{b})$$

$$\boldsymbol{\Omega}^{**} = \boldsymbol{\Omega}^* + \mathbf{X}_1'\mathbf{X}_1$$

$$\hat{\sigma}^{**2} = [\nu^*\hat{\sigma}^{*2} + \mathrm{SSR}_b + (\mathbf{b} - \hat{\boldsymbol{\beta}}^*)' \times \boldsymbol{\Omega}^*(\boldsymbol{\Omega}^* + \mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_1(\mathbf{b} - \hat{\boldsymbol{\beta}}^*)]/(\nu^* + N_1)$$

$$\nu^{**} = \nu^* + N_1.$$

The marginal posterior distribution of $\boldsymbol{\beta}$ is multivariate Student with mean vector $\hat{\boldsymbol{\beta}}^{**}$, covariance matrix $\hat{\sigma}^{**2}(\boldsymbol{\Omega}^{**})^{-1}$, and degrees of freedom ν^{**}. The posterior mean, $\hat{\boldsymbol{\beta}}^{**}$, is a matrix-weighted average of the prior mean $\hat{\boldsymbol{\beta}}^*$ and the sample estimate $\mathbf{b}$, where the weight matrices are proportional to the inverted covariance matrices of $\hat{\boldsymbol{\beta}}^*$ and $\mathbf{b}$, respectively. Thus, the Bayesian analyst in this case treats the prior information as being equivalent to a previous sample of size ν^* of the same process that generated the sample data.[14]

If we substitute $\mathbf{a}$ for $\hat{\boldsymbol{\beta}}^*$, $\mathbf{X}_0'\mathbf{X}_0$ for $\boldsymbol{\Omega}^*$, and N_0 for ν^* in these expressions, the marginal posterior mean vector and covariance matrix of $\boldsymbol{\beta}$ are

$$\hat{\boldsymbol{\beta}}^{**} = (\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}(\mathbf{X}_0'\mathbf{y}_0 + \mathbf{X}_1'\mathbf{y}_1) \qquad [21]$$

and

$$\mathrm{var}(\hat{\boldsymbol{\beta}}^{**}) = \hat{\sigma}^{**2}(\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}, \qquad [22]$$

which exactly parallel the formulas for $\mathbf{b}_p$ and $\mathrm{var}(\mathbf{b}_p)$ in expressions [5] and [7] in Section 1.1. If instead we substitute $\mathbf{a}$ for $\hat{\boldsymbol{\beta}}^*$, $\lambda^2\mathbf{X}_0'\mathbf{X}_0$ for $\boldsymbol{\Omega}^*$, and λN_0 for ν^*, the marginal posterior mean vector and covariance matrix of $\boldsymbol{\beta}$ are

$$\hat{\boldsymbol{\beta}}^{**} = (\lambda^2\mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}(\lambda^2\mathbf{X}_0'\mathbf{y}_0 + \mathbf{X}_1'\mathbf{y}_1) \qquad [23]$$

[14] The degrees of freedom parameter in the denominator of $\hat{\sigma}^{**2}$ is $\nu^* + N_1$ rather than $\nu^* + N_1 - K$; this result parallels the maximum-likelihood estimate of σ^2 but not the more usual unbiased estimate. In cases where the equivalence of prior and sample information seems inappropriate, there are altenatives to the normal-gamma family of prior distribution in which, for example, a marked discrepancy between prior beliefs and sample evidence may produce a bimodal posterior distribution with modes corresponding to the prior and sample means (Dickey 1975; Leamer 1978, 79–81).

and

$$\text{var}(\hat{\boldsymbol{\beta}}^{**}) = \hat{\sigma}^{**2}(\lambda^2 \mathbf{X}_0'\mathbf{X}_0 + \mathbf{X}_1'\mathbf{X}_1)^{-1}, \qquad [24]$$

which exactly parallel the analogous formulas for $\mathbf{b}_\lambda$ and $\text{var}(\mathbf{b}_\lambda)$. Thus, fractional pooling amounts to a Bayesian regression analysis of the data from the $\boldsymbol{\beta}$ regime of primary interest, with a prior mean vector equal to the ordinary least squares coefficient vector $\mathbf{a}$ from the $\boldsymbol{\alpha}$ regime of uncertain theoretical relevance and a prior covariance matrix equal to the covariance matrix of $\mathbf{a}$ inflated by the scalar value $1/\lambda^2$.

At first glance, it may seem odd that our prior beliefs about $\boldsymbol{\beta}$ depend upon all the features of the $\boldsymbol{\alpha}$ data that affect the estimated covariance matrix of $\mathbf{a}$. But this dependence is natural if we recall that the point of fractional pooling is to learn what we can about $\boldsymbol{\beta}$ from the $\boldsymbol{\alpha}$ data. If the $\boldsymbol{\alpha}$ data are few in number or nearly colinear or simply exhibit little variation, they can tell us little about $\boldsymbol{\alpha}$ and thus even less about $\boldsymbol{\beta}$; in that case we must approach the $\boldsymbol{\beta}$ data themselves with more uncertainty than we otherwise would. The important point to bear in mind here is that the mean vector $\mathbf{a}$ and covariance matrix $\hat{\sigma}^2\ (\lambda^2 \mathbf{X}_0'\mathbf{X}_0)^{-1}$ are the mean vector and covariance matrix of a "prior" distribution only from the perspective of the $\boldsymbol{\beta}$ data, since they are actually estimated from the $\boldsymbol{\alpha}$ data.[15]

Under what circumstances would this specification for the prior mean vector and covariance matrix make sense? Whenever we want to discount the data from the $\boldsymbol{\alpha}$ regime to some extent, but otherwise treat them as if they represented additional data from the $\boldsymbol{\beta}$ regime of primary theoretical

[15] In this respect the approach proposed here is akin to "empirical Bayes" techniques, which use relevant data to estimate "prior" distributions (Maritz 1970; Lindley and Smith 1972; Rubin 1980). I wish, however, to emphasize the subjectivity of the maintained assumptions used to transform relevant data into a "prior" distribution (in particular, the choice of a pooling fraction λ), whereas "empirical Bayes" techniques typically emphasize the role of the data and deemphasize the role of the maintained assumptions used to transform the data into a "prior" distribution. Rubin's (1980) application is especially relevant, since his problem was to pool admissions data from dozens of different law schools to estimate the relative importance of college grades and standardized test scores in predicting success in each law school. (Analyzing the data from each law school separately produced rather imprecise and temporally unstable parameter estimates.) Rubin's approach was to assume that the relevant parameter for each law school was drawn from a distribution whose hyperparameters he estimated from the data for all law schools. It is interesting to note that Rubin (and Efron and Morris 1977 in a similar application) used thoroughly conventional (and *ad hoc*) exploratory techniques to assess the possibility that the parameters governing different subsets of the data (for example, more versus less selective law schools) were drawn from different distributions, rather than relying upon theory to specify the relevance of each subset of the data from the perspective of the others, as I propose here.

interest. Of course, this specification will not be appropriate for every occasion. For example, data analysts with specific prior beliefs about the relative magnitudes of the parameters in the two regimes would not want to simply treat the coefficient vector **a** from the α regime as their prior mean vector when analyzing the data from the β regime; in that case, a more complicated approach is necessary to appropriately represent the impact of the α data on our beliefs about β. Alternatively, data analysts might prefer different weights, and hence different values of λ, to represent the subjective import of the problematic data for their beliefs about different parameters; again, a more complicated specification is necessary to capture the relevant prior beliefs. But despite these potential complications, the approach proposed here makes it possible to represent one interesting family of prior beliefs in a way that accords with Bayesian theory, while avoiding much of the complexity of a full-blown Bayesian analysis. The fact that fractional pooling has this natural Bayesian interpretation seems to me to add significantly to its attractiveness.

The problem remains of how to choose an appropriate value of λ. The explicit subjectivity of the Bayesian approach highlights the fact that reasonable people will often disagree about such matters. Indeed, despite the fact that Bayesian analysis is often hindered by the difficulty of specifying precise prior beliefs in a meaningful way, one of the main virtues of Bayesian techniques is that they can help to monitor and clarify the implications of alternative subjective judgments about appropriate model specification. Often this is accomplished by calculating ranges of posterior estimates corresponding to meaningful ranges of prior beliefs. For example, a very general result of Chamberlain and Leamer (1976) provides an ellipsoid bound for a posterior mean vector of regression coefficients given only the prior mean vector, the sample coefficient vector and covariance matrix; every point in this ellipsoid corresponds to a possible choice of prior covariance matrix, and no choice of prior covariance matrix produces a posterior mean outside the ellipsoid.

The λ-family of pooled estimates described in Section 2.2 can be thought of as representing an intermediate approach between a fully specified Bayesian analysis and the partially specified analysis of Chamberlain and Leamer (1976). Simply specifying that our prior distribution for β is centered at the least squares point **a** estimated from the problematic data is sufficient to reduce the range of possible posterior mean vectors from the entire parameter space to Chamberlain and Leamer's ellipsoid, whose boundary includes the prior and sample mean vectors (here, the subset regression estimates **a** and **b**). Specifying in addition that our prior covariance matrix for β must be proportional to the covariance matrix of **a** further reduces the family of possible posterior mean vectors to a curve (here, for

constants of proportionality $1/\lambda^2 \geq 1$ as implied by $0 \leq \lambda \leq 1$, the locus of λ-estimates described in Section 2.2). Finally, specifying in addition a precise value of λ reduces the family of possible posterior mean vectors to a single point on the curve, the parameter vector $\mathbf{b}_\lambda$.

Each of these levels of specificity may be illuminating in some circumstances, depending upon how much we are willing to assume about the relevance of the $\boldsymbol{\alpha}$ data for beliefs about $\boldsymbol{\beta}$. The important point is that fractional pooling provides a simple and flexible way to explore the implications of an interesting class of prior beliefs about the relevance of problematic data. On the one hand the approach can be used to estimate a single parameter vector $\mathbf{b}_\lambda$ based on a specific value of λ; on the other hand it can be used to produce a *family* of fractionally pooled estimates reflecting a whole range of alternative values of λ, as described in Section 2.2. This flexibility seems to me to add significantly to the attractiveness of fractional pooling as a technique for analyzing disparate observations.

3. Empirical Examples

My aim in this section is to provide concrete illustrations of the issues addressed in Section 1 and the approach proposed in Section 2. The illustrations are drawn from real data analyses in which the problem of pooling disparate observations seems to arise. For the sake of convenience, all of the relevant data sets are sufficiently compact to be reproduced in the Appendix. The issues arising in the analysis of these data sets are equally relevant in analyses with much larger data sets, however. For example, analysts of large opinion surveys with both panel and fresh cross-section components, including the Census Bureau's Current Population Survey, the Survey of Income and Program Participation, and some recent American National Election Studies, might legitimately wonder whether to pool data from new and old respondents; analysts of primary season campaign dynamics might wonder whether or not to include out-party identifiers in their analyses (Brady and Johnston 1987; Bartels 1988); and analysts of congressional behavior might wonder whether to pool data from Senate sessions in which the Republicans were the majority party with data from sessions in which the Democrats were the majority party (Schiller 1995). In these cases and many others, the problem arises of how to make valid inferences on the basis of disparate observations.

3.1. Election Fraud in Philadelphia

My first example is based on Ashenfelter's (1994) analysis of alleged election fraud in the casting of absentee ballots in a special election in the 2nd Senate District of Pennsylvania in 1993. The Democratic candidate trailed his Republican opponent by 564 votes in the tally of ballots cast by

voting machine, but recorded a plurality of 1,025 absentee votes to win the election by 461 votes. The Republican candidate challenged the election result in court, alleging that many of the absentee ballots were cast illegally and should be voided. In order to assess the plausibility of the Democratic margin in absentee votes, Ashenfelter analyzed the relationship between absentee vote margins and machine vote margins in 21 previous Pennsylvania Senate elections in the Philadelphia area in the preceding decade.

Ashenfelter's regression analysis indicated a significant positive relationship between machine vote margins and absentee vote margins, as well as a small (126 vote) and statistically insignificant pro-Republican bias in absentee votes. From these results, Ashenfelter calculated the expected Democratic absentee vote margin in the 2nd District in 1993 as −133, with a standard error of 345. Assuming a well-specified regression with normally distributed errors, Ashenfelter concluded that the probability of observing a Democratic absentee vote margin as far from the expected margin as the one reported was less than 1%, and that the probability of observing a Democratic absentee vote margin as far from the expected margin as 565 (the minimum absentee vote margin required to offset the Republican machine vote margin of 564 votes) was about 6%. Judge Clarence Newcomer cited Ashenfelter's analysis in support of his decision to overturn the contested election result, awarding the 2nd District seat (and, as it happens, partisan control of the Pennsylvania State Senate as a whole) to the Republicans.

My aim here is to examine the sensitivity of Ashenfelter's conclusion to his choice of relevant prior election results. While the spatial and temporal delimitations embodied in his analysis are not unreasonable, other choices seem equally reasonable. Why not use the previous twenty years' results rather than ten? Why not use results from all of Pennsylvania's Senate districts, or from all urban districts, rather than just those from Philadelphia? On the other hand, why not limit the analysis to previous results in the district where the disputed election actually occurred, rather than including results from other Philadelphia districts with rather different social and political characteristics?

My reanalysis of Ashenfelter's data focuses upon the last of these issues. With only three previous elections in the 2nd Senate District since it was last redistricted, it might seem hopeless to limit the analysis to that district alone. Nevertheless, it seems prudent to recognize that previous results from other districts are less obviously relevant than previous results from the 2nd District in evaluating the allegation of vote fraud in the 2nd District in 1993. Fractional pooling seems to provide an attractive approach in this situation, since it allows us to make use of data from Philadelphia districts (and, in principle, from other relevant times or places) without

Table 1. Regression Analysis of Absentee Vote Margins in Philadelphia

	All Philadelphia Districts	District 2 Only	Other Districts Only
Intercept	−125.9	153.5	−200.9
	(114.3)	(22.5)	(133.4)
Machine Vote Margin	.01270	.00770	.01394
	(.00298)	(.00127)	(.00327)
Std error of regression	324.8	23.2	338.0
Adjusted R^2	.46	.95	.50
N	21	3	18

(Ordinary least squares parameter estimates with standard errors in parentheses.)

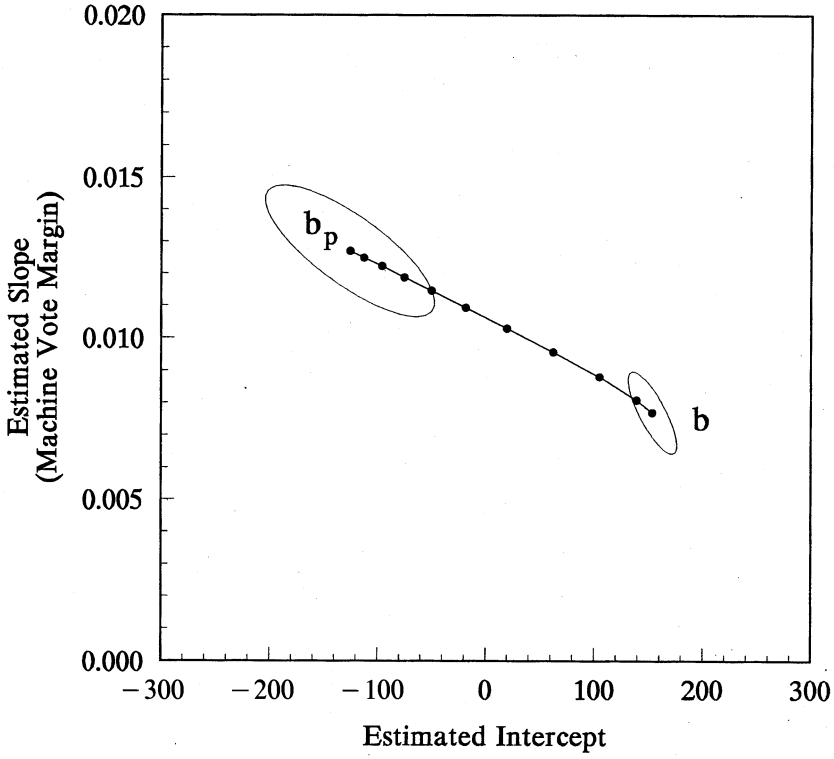
having to pretend that the relationship between machine votes and absentee votes is known in advance to be indentical in every district.

The results of Ashenfelter's (1994) analysis based on 21 previous elections in seven Philadelphia districts from 1982 to 1992 are replicated in the first column of Table 1. The second column of Table 1 repeats Ashenfelter's analysis using only the data from the three previous elections in the 2nd Senate District. Rather remarkably, in view of the fact that there is only one degree of freedom in this version of the model, the parameter estimates are quite precise, the standard error of the regression is less than one-tenth as large as in the pooled analysis in the first column, and the adjusted R^2 statistic is .95. The absentee vote bias is pro-Democratic rather than pro-Republican, and the slope of the relationship between machine vote margins and absentee vote margins is considerably smaller than in the pooled analysis in the first column of Table 1.

The third column of Table 1 shows the corresponding regression results for the 18 elections in districts other than the 2nd. The estimated slope of the relationship between machine vote margins and absentee vote margins is 80% larger in these districts than in District 2. The t-statistic for this difference is 1.8, while the t-statistic for the difference in intercepts is −2.6. There are so few observations from District 2, however, that an F-test comes nowhere near rejecting the null hypothesis of parameter equality embodied in the pooled regression at conventional significance levels.[16]

[16] The F-statistic is .818; with 2 and 17 degrees of freedom, $p = .458$. Note, however, that the dramatic disparity between the standard errors of the regressions in the second and third column of Table 1 casts grave doubt upon the assumption underlying the F-test that σ_0^2 equals σ_1^2. My unease about relying on F-tests to determine whether to pool disparate observations stems in part from my suspicion that this assumption will often be implausible

Figure 1. Philadelphia Vote Fraud Analysis



This is a case where the *F*-test seems not to be answering the question that ought to be crucial in most decisions about whether or not to pool disparate observations: are the parameter values governing the problematic observations similar to those governing the observations of primary theoretical interest?

The parameter estimates in the first and second columns of Table 1 are represented graphically in Figure 1, with the intercept estimates shown on the horizontal dimension and the slope estimates shown on the vertical dimension. The point $\mathbf{b}_p$ at the upper left of the figure represents Ashenfelter's parameter estimates (in the first column of Table 1) based upon 21 previous elections in all seven Philadelphia districts; the point **b** at the lower

in applied work; the present and subsequent examples reinforce that suspicion and illustrate some of its implications.

Table 2. Statistical Tests of Absentee Vote Margin as a Function of λ

λ	$\hat{y}_{(-564)}$	Standard error of $\hat{y}$	*p*-value ($\hat{y} > 564$)
1.0	−133	345	.029
0.9	−120	332	.027
0.8	−103	319	.027
0.7	−83	305	.026
0.6	−57	290	.027
0.5	−25	273	.028
0.4	14	253	.030
0.3	57	227	.032
0.2	100	190	.031
0.1	134	132	.026
0.0	149	33	.025

right of the figure represents the parameter estimates (in the second column of Table 1) based upon the three previous elections in District 2 only. The ellipse surrounding each of these points is the boundary of a two-dimensional 50% joint confidence region. The distance between these two ellipses in the parameter space indicates that our inferences about where the true parameter values probably lie are quite sensitive to our choice of relevant observations.

The locus of parameter estimates produced by fractional pooling for Ashenfelter's analysis is also shown in Figure 1. Intermediate points between the endpoints $\mathbf{b}$ and $\mathbf{b}_p$ correspond to intermediate values of λ at .1 intervals between zero and one. All of these values imply smaller slopes and more Democratic intercepts than in Ashenfelter's analysis. For values of λ less than .6, these estimates are outside the two-dimensional 50% confidence region for Ashenfelter's estimates (the larger of the two ellipses shown in Figure 1). Nevertheless, for plausible values of λ—say, in a range from .4 to .8—the parameter estimates implied by fractional pooling are only modestly different from appropriately weighted simple averages of the subset and pooled regression coefficients $\mathbf{b}$ and $\mathbf{b}_p$. In this sense, at least, the locus of fractionally pooled parameter estimates seems to provide relatively little information beyond that contained in the subset and pooled regressions.

Finally, since Ashenfelter's main interest in the regression analysis was not in the parameter estimates themselves but in a comparison between the disputed 1993 absentee vote margin and the predicted absentee vote margin implied by the regression analysis, it is worth observing how this comparison depends upon the pooling fraction λ. For the 11 values of λ illustrated in Figure 1, Table 2 shows the predicted absentee vote margin correspond-

ing to a machine vote margin of −564, the standard error of this predicted absentee vote margin, and the probability of observing a Democratic absentee vote margin large enough to offset the observed Republican margin in machine votes.[17] The range of *p*-values in the last column of Table 2 is only from .025 to .032, with the lowest *p*-value occurring when λ equals zero (that is, when elections from districts other than the 2nd are assigned no weight, the assumption at the opposite extreme from the one adopted by Ashenfelter). Thus, although the parameter estimates in Ashenfelter's analysis are fairly sensitive to how data from other districts are handled, his conclusion that the disputed absentee vote margin was inconsistent with previous experience is not.

3.2. Political Violence

My second example is derived from Powell's (1982) analysis of regime performance in contemporary democracies. Powell related various measures of political participation, stability, and violence in 26 democracies to a variety of social and institutional factors, including economic development, ethnic cleavages, and constitutional design. My reanalysis of Powell's data focuses on the effect of population size, economic development (measured by the natural logarithm of GNP per capita), ethnic fractionalization (an index measuring "the probability that two randomly drawn citizens will be of different ethnic or linguistic groups"), and constitutional design (a measure of "representativeness," with presidential systems at the low end and parliamentary systems with at least five representatives per district at the high end) upon political violence, measured by the number of deaths per year from political violence in the period 1967–76.[18]

It seems plausible to suppose that the primary aim of a cross-national analysis of political violence in contemporary democracies would be to shed light upon the causes of violence in the minority of "developing" or "Third World" countries that have managed to create and maintain democratic political institutions. For one thing, the magnitude of the problem of political violence is clearly greater in "traditional" than in "modern" democracies. For example, 83% of all the deaths from political violence

[17] The *p*-values in the last column of Table 2 are based upon one-tailed *t*-tests; Ashenfelter (1994) reported the probability value for a two-tailed *t*-test. Thus, the *p*-value of .029 reported for $\lambda = 1$ in the last column of Table 2 corresponds with Ashenfelter's claim (1994, 3–4) that "we would have expected the Democratic candidate to win the election (based on the sum of the machine and absentee votes) in fewer than 5.8 in 100 cases with the observed configuration of the facts."

[18] Powell reported a parallel analysis of violence in the period 1958–67, but since most of his explanatory variables are measured circa 1965 I limit my attention to the later time period.

in Powell's 26 countries occurred in the ten countries he classified as "predominantly traditional" or "mixed traditional and modern." (More than half of the deaths in "modern" countries occurred in the United Kingdom, mostly in Northern Ireland.)

Even an analyst who attached equal importance to understanding political violence in traditional and modern democracies would probably concede that somewhat different social and institutional factors might be associated with political violence in the two types of regimes. Thus, while it might well be reasonable to attach *some* weight to the experience of traditional democracies in attempting to understand the causes of political violence in modern democracies, and vice versa, each subset of the data must presumably be considered problematic from the theoretical viewpoint of the other. For the purposes of this analysis my focus is on the causes of political violence in Powell's traditional democracies, and my aim is to explore how evidence from the modern democracies might shed light on those causes.

The parameter estimates in the first column of Table 3 replicate Powell's analysis (1982, 156) using data from his 26 contemporary democracies. My analysis is of logged deaths from political violence, rather than Powell's truncated version of the unlogged variable.[19] This difference may account for the one notable disparity between the results reported by Powell and those reported in the first column of Table 3: whereas Powell reported a small but significant positive impact of ethnic fractionalization on political violence (a standardized regression coefficient of +.11, significant at the .10 level), the corresponding estimate in Table 3 is negative and smaller than its standard error. In other respects, the two sets of results are similar.

The second column of Table 3 shows the parameter estimates produced by applying the same regression model to the subset of the complete data set consisting of ten of the 11 countries classified by Powell as "predominantly traditional" or "mixed traditional and modern." I omit Costa Rica from this subset of "traditional" countries because it is a glaring outlier, having experienced much less political violence than other countries with similar institutions and levels of economic development.[20] Limiting the analysis to

[19] Powell (1982, 235) reported that "For riots, deaths, and protests extreme cases are truncated to ninetieth percentile to prevent bias; log transformation yields similar results." Since Powell's sources record no deaths from political violence in some of his countries between 1967 and 1976, I add .05 (half the lowest positive recorded value of deaths per year) to each observation before taking natural logarithms.

[20] Including Costa Rica with the other ten "traditional" countries reduces the adjusted R^2 statistic from .84 in the second column of Table 2 to .25; the standard errors of the parameter estimates are almost three times as large as those reported in the second column of Table 2, and none of the parameter estimates remains statistically different from zero at the .20 significance level. It is important to note that my treatment of Coast Rica here is

these ten traditional democracies produces some substantial differences in the estimated effects of the explanatory variables. The coefficient for population, which had a *t*-statistic of 3.4 in the complete data set, is almost exactly zero in the subset of traditional democracies, perhaps suggesting that regimes rather than citizens are the primary focus of political violence in traditional democracies. The coefficient for ethnic fractionalization is almost exactly zero (though it was smaller than its standard error even in the complete data set), while the intercept level of violence is substantially larger.

The third columm of Table 3 shows the parameter estimates produced by applying Powell's regression model to the complementary subset of modern countries.[21] In this subset of the data the estimated effect of population is almost twice as large as in the complete data set, the estimated effect of GNP is of roughly the same magnitude but with a much larger standard error, and the estimated effect of having a representational constitution is essentially zero. The goodness-of-fit statistics are less impressive than for the corresponding analysis of traditional countries only, but more impressive than for the pooled analysis.

The *F*-statistic for a test of the constraints embodied in the pooled analysis is 7.57 (with 5 and 16 degrees of freedom), sufficiently large to reject the constraints at the .001 significance level.[22] It seems clear from this result, and from the differences in some of the parameter estimates that appear in Table 3, that data from traditional and modern democracies should not simply be pooled for the purposes of Powell's analysis. It does not follow, however, that the two data sets should be treated as though they were entirely unrelated, as they presumably would be from this point on by an analyst adopting the usual pretest estimation strategy. The intuition that the experience of modern democracies can tell us *something* about the corre-

---

prompted by a loud message from the data, and not by *a priori* theoretical considerations; thus, it cannot be justified by the approach to disparate observations proposed in this paper, or, I fear, by any compelling statistical theory.

[21] For the sake of simplicity I include Costa Rica among the modern countries in the third column of Table 3. Using only the 15 countries classified by Powell as "modern with traditional sector" or "predominantly modern" produces similar results; none of the parameter estimates changes by more than .15 standard errors from those reported in the third column of Table 3. The parameter estimate for a dummy variable for Costa Rica added to the model in the third column of Table 3 gets a *t*-statistic of .07; the parameter estimate for the same dummy variable added to a model like the one in the second column of Table 3 but including Costa Rica gets a *t*-statistic of −7.5.

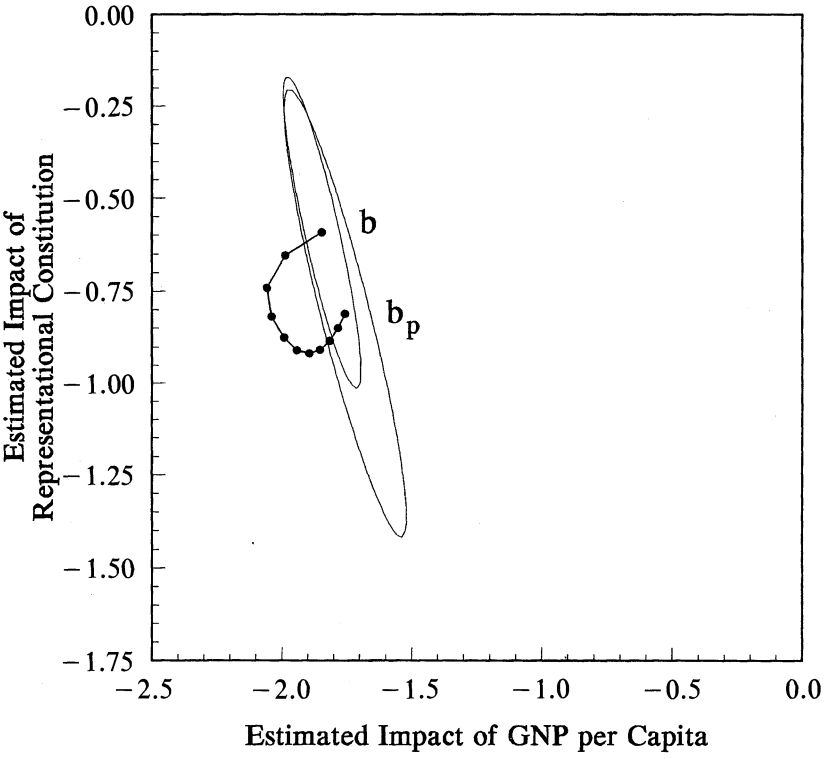
[22] The standard error of the regression for modern democracies in the third column of Table 3 is exactly twice as large as the standard error of the regression for traditional democracies in the second column. Although not as dramatic as the corresponding disparity in Table 1, this difference is sufficiently large to warrant considerable skepticism about the appropriateness of the *F*-test in this setting.

Table 3. Regression Analysis of Deaths from Political Violence

	All Contemporary Democracies	"Traditional" Only (minus Costa Rica)	"Modern" Only (plus Costa Rica)
Intercept	−.96	16.54	−19.24
	(5.54)	(5.61)	(6.70)
Population (logged)	1.015	−.027	1.855
	(.297)	(.215)	(.318)
GNP per capita (logged)	−1.757	−1.846	−1.594
	(.432)	(.835)	(1.008)
Ethnic Fractionalization	−1.598	.041	.836
	(1.908)	(2.735)	(2.118)
Representational Constitution	−.811	−.592	−.019
	(.342)	(.301)	(.351)
Std error of regression	1.93	.69	1.38
Adjusted R_2	.63	.84	.72
N	26	10	16

(Ordinary least squares parameter estimates with standard errors in parentheses.)

Figure 2. Analysis of Political Violence



sponding political processes in traditional democracies seems sufficiently compelling to warrant some investigation of its inferential implications.

Figure 2 shows the least squares parameter estimate **b** and the pooled least squares estimate $\mathbf{b}_p$ for two of Powell's explanatory variables. The horizontal dimension of the figure represents the estimated impact of GNP per capita, and the vertical dimension represents the estimated impact of a representational constitution. Figure 2 also shows the two-dimensional 50% confidence region associated with each estimator and the locus of fractionally pooled parameter estimates connecting the endpoints **b** (corresponding to $\lambda = 0$) and $\mathbf{b}_p$ (corresponding to $\lambda = 1$).

Although the apparent restraining effect of a representational constitution is smaller in the traditional democracies than in the whole data set, each estimate is well within the 50% confidence region of the other, and indeed virtually the entire two-dimensional confidence region for **b** lies

within the corresponding confidence region for $\mathbf{b}_p$. Thus, for these explanatory variables, the inferential implications of the choice between $\mathbf{b}$ and $\mathbf{b}_p$ turn out to be relatively minor, notwithstanding the emphatic rejection of the null hypothesis of parameter equality prompted by the F-statistic of 7.57 cited earlier.

The locus of fractionally pooled parameter estimates connecting the points $\mathbf{b}$ and $\mathbf{b}_p$, however, tells a somewhat more complicated story. The shape of the λ-curve illustrates the fact, noted in Section 2.2, that the parameter estimates produced by fractional pooling need not lie between the least squares and pooled least squares estimates in any single dimension. Indeed, in this instance, *any* value of λ greater than zero and less than one produces parameter estimates outside the range from $\mathbf{b}$ to $\mathbf{b}_p$ for one of the two variables, and values of λ between about .3 and .7 produce parameter estimates outside the range from $\mathbf{b}$ to $\mathbf{b}_p$ for *both* variables.

Furthermore, a wide range of plausible values of λ produce parameter estimates for these explanatory variables outside the 50% confidence regions implied by either of the endpoint estimates $\mathbf{b}$ or $\mathbf{b}_p$. In that sense, our inferences about the likely effects of economic development and constitutional design in traditional democracies are fairly sensitive to subjective judgments about how much we can learn from the distinct—but not completely unrelated—experience of modern democracies.

3.3. Economic Conditions and Presidential Election Outcomes

My third example is based on Alesina, Londregan, and Rosenthal's (1993) analysis of presidential election outcomes, part of a larger empirical analysis based upon "a model of the political economy of the United States" encompassing presidential election outcomes, midterm and presidential year congressional election outcomes, and GNP growth.

Alesina, Londregan, and Rosenthal's analysis was based on election data from 1916 to 1988. Most other analysts of economic voting in United States presidential elections (Tufte 1978; Markus 1988; Erikson 1989) have begun their analyses with either the 1948 or 1952 election. Either of these starting points minimizes potential complications associated with World War II and the aftershocks of the New Deal realignment, while limiting the analysis to a historical period in which the federal government was clearly assigned significant responsibility for macroeconomic management.[23] In addition, the 1952 election is the first for which survey data on

[23] The latter historical observation might be taken to imply an *a priori* expectation that the impact of economic conditions on presidential election outcomes would be greater in the post-World War II period than earlier. I ignore that complication here, except to note that the parameter estimates in Table 4 are consistent with such an expectation.

Table 4. Regression Analysis of Presidential Election Outcomes

	1916–88	1948–88	1916–44
Intercept	7.58	−12.22	31.75
	(11.47)	(35.15)	(19.09)
Partisan Balance	.739	1.083	.200
	(.221)	(.663)	(.398)
Republican Incumbent	10.11	9.84	17.08
	(2.18)	(4.89)	(4.43)
GNP Growth	1.239	2.350	1.445
	(.202)	(.963)	(.213)
Std error of regression	4.24	4.92	2.88
Adjusted R_2	.70	.37	.92
Durbin-Watson	2.86	2.91	2.56
N	19	11	8

(Ordinary least squares parameter estimates with standard errors in parentheses.)

voters' perceptions of the presidential candidates are available from the American National Election Studies.

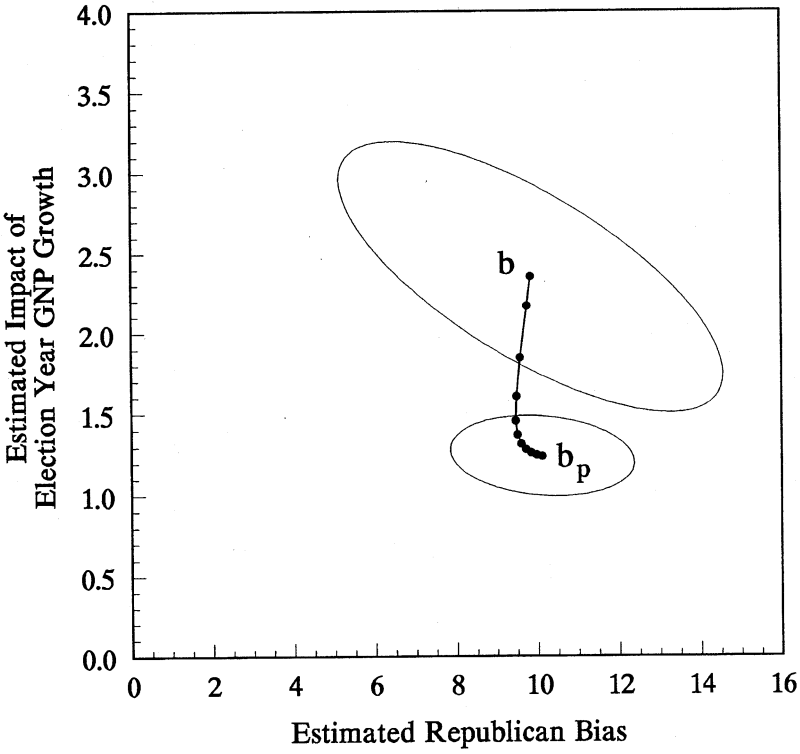
The obvious disadvantage of limiting the anlaysis to the post-World War II era is that the number of post-war presidential elections available for analysis is unhappily small—11 at the time of Alesina, Londregan, and Rosenthal's analysis, now 12. A less obvious disadvantage is that the range of observed variation in election year GNP growth has been considerably narrower after World War II than before, making it correspondingly more difficult to estimate the impact of GNP growth on election outcomes.[24]

Alesina, Londregan, and Rosenthal estimated separate effects for expected election-year GNP growth (based upon lagged growth, military mobilization, and partisan effects) and current growth shocks, but found no significant difference between the two estimates. They interpreted the similarity of the two growth effects as evidence of naive retrospection on the part of voters. Here I simplify the analysis by estimating a single effect for election year GNP growth; as a result, the parameter estimates for the 1916–88 data set reported in the first column of Table 4 do not exactly match those reported by Alesina, Londregan, Rosenthal (1993, Table 3), although the differences are minor.[25]

[24] Election year GNP growth in the 11 elections between 1948 and 1988 ranged from −.2% to 6.6%; the corresponding range in the eight elections between 1916 and 1944 was from −14.4% to 13.2%.

[25] Alesina, Londregan, and Rosenthal's "equation restricted" parameter estimates (and the corresponding ordinary least squares estimates from my Table 4) are 5.981 for the intercept (7.584), .739 for lagged House vote (.739), 10.362 for the Republican bias (10.11), and 1.590 and 1.140 for expected GNP growth and the current growth shock, respectively (1.239).

Figure 3. Presidential Election Analysis



The dependent variable in the regression analysis in Table 4 is the two-party presidential vote percentage won by the candidate of the incumbent party. The explanatory variables, in addition to election year GNP growth, are a constant, partisan balance (measured by the incumbent party's share of the national congressional vote in the previous midterm election), and a dummy variable indicating when the Republicans were the incumbent party. My focus here is on the GNP growth effect and the Republican bias. Parameter estimates representing those effects are shown on the vertical and horizontal axes, respectively, of Figure 3.

Rather surprisingly, the estimated Republican bias is essentially insensitive to the choice of time periods. Other things being equal, Republican incumbents did about ten percentage points better than Democrats whether we focus on the post-World War II period only or on the entire period from 1916 to 1988. On the other hand, the estimated effect of election-year GNP

growth appears to be much more sensitive to the choice of sample period. The estimated effect in the post-World War II period (in the second column of Table 4) is almost twice as large as the estimated effect for the entire period from 1916 to 1988 (in the first column of Table 4); as a result, the two-dimensional 50% confidence regions for $\mathbf{b}$ and $\mathbf{b}_p$ in Figure 3 do not overlap at all. Thus, a simple comparison of the pooled and subset regression estimates and their associated confidence regions seems to cast considerable doubt upon the wisdom of Alesina, Londregan, and Rosenthal's decision to pool data from the entire period from 1916 to 1988, despite the fact that an F-test fails to reject the hypothesis that the coefficient vectors $\mathbf{b}$ and $\mathbf{b}_p$ are equal.[26]

A closer examination of the locus of fractionally pooled parameter estimates in Figure 3 indicates that most of the variation in the apparent impact of economic growth occurs in the range $0 < \lambda < .3$; all of the fractionally pooled estimates in the range $.4 < \lambda < 1$ are within 20% of the fully pooled estimate $\mathbf{b}_p$, and within the two-dimensional 50% confidence region for that estimate shown in Figure 3. Although different analysts would naturally assign somewhat different relevance to data from elections before 1948, most would, I think, agree that these elections are sufficiently relevant to the contemporary period to warrant weights somewhere between .4 and 1. Thus, contrary to the implication of the pooled and subset regressions alone, fractional pooling demonstrates that Alesina, Londregan, and Rosenthal's estimates provide a good representation of the available evidence about the impact of GNP growth on presidential election outcomes for a wide range of weights we might plausibly want to attach to the 1916–44 data.

4. Conclusion

Beck (1985, 79) argued a decade ago that "It ought to be common practice to test all time series results for sensitivity to choice of sample period." It is not. Nor is it common practice to test cross-sectional results for sensitivity to equally subjective choices about the range of observations to which a given statistical model is applied. It should be.

Having said that, I must add that the analysis and examples presented in this article suggest that even moderately conscientious sensitivity testing of the sort advocated by Beck (1985), Bartels (1990), and others may not be enough to avoid quite misleading inferences. In particular, two aspects of moderately conscientious common practice seems to me to be dubious.

[26] The F-statistic is .916; with 4 and 11 degrees of freedom, $p = .488$. Once again, however, the disparity between the standard errors of the regressions in the second and third columns of Table 4 casts doubt upon the assumption underlying the F-test that σ_0^2 equals σ_1^2.

First, using F-tests to decide whether or not to pool disparate observations seems unlikely to result (except by chance) in intelligent treatment of problematic data, given the mismatch between what F-tests can do and what analysts need done. F-tests invite formulaic *post hoc* revisions in model specifications, whereas analysts want (or should want) guidance about the implications of their own judgments about issues of model specification. F-tests are about goodness of fit, whereas analysts focus (or should focus) primarily on parameter values. And F-tests are tests of exact equality, whereas analysts care (or should care) about magnitudes of inequality. As a result, F-tests can resoundingly reject the null hypothesis of parameter equality even when the pooled and subset parameter estimates of interest are, in fact, quite similar (as in the example in Section 3.2), and can resoundingly fail to reject the null hypothesis of parameter equality even when the pooled and subset regression results are quite different (as in the examples in Sections 3.1 and 3.3).

The second (and probably less common) conventional approach, actually reporting pooled and subset regression results, is significantly better than simply using those results to compute an F-statistic. But even a direct comparison of pooled and subset regression results may or may not provide a good indication of the inferential import of problematic data. When the alternative sets of parameter estimates produced by fractional pooling are relatively evenly spaced along a relatively straight line connecting the pooled regression estimate $\mathbf{b}_p$ and the subset regression estimate $\mathbf{b}$, as in Figure 1, it is easy enough (and safe enough) to interpolate intermediate results between these endpoints. But when the λ-curve is highly non-linear, as in Figure 2, or when points along this curve corresponding to equally spaced values of λ are very unevenly spaced, as in Figure 3, casually splitting the difference between the two extremes (or feeling relieved when there is not much difference between the two extremes) can lead analysts and readers significantly astray.

The fundamental contradiction undermining both these conventional practices, in my view, is that they pretend to maintain the rigid classical dichotomy between observations that belong in the analysis and those that do not, while smuggling more complex, uncertain prior beliefs about the theoretical relevance of the available data in by a back door. The alternative approach proposed here, fractional pooling, replaces the classical dichotomy with a more realistic continuum of assumptions about the relevance of the available data, purchasing coherence, flexibility, and realism at the cost of modest additional complexity.

Fractional pooling is easy to implement and has a plausible theoretical justification in Bayesian statistical theory. Those are attractive features in any statistical technique. Fractional pooling will not be an appropriate solution

to every problem of pooling disparate observations, however. Analysts with specific expectations about the relative magnitudes of parameters in different subsets of the complete data set may want to incorporate those expectations in a more elaborate Bayesian analysis. Data that are more relevant for estimating some parameters than others likewise invite a more detailed, and correspondingly complex, Bayesian specification. The approach proposed here has both the virtue and the limitation of simplicity, representing the theoretical relevance of each observation by a single fraction.

One other apparent limitation of the approach proposed here can easily be circumvented. Even if the theoretical relevance of each observation can be represented by a single fraction, there is no guarantee that the available data will fall naturally into exactly two subsets, one with weight 1.0 and the other with weight λ. Fortunately, there is no reason in principle why fractional pooling should not be extended to associate different values of λ with each of several categories of observations, or even with each observation separately. Of course, more complicated weighting schemes will be harder to describe and justify; but that in itself should not be sufficient reason to eschew them if they more accurately reflect considered judgments about the theoretical relevance of the available data.

Finally, it will seldom be the case that any single value of λ (or, for that matter, any single division of the available data into "clearly relevant" and "theoretically problematic" subsets) is so compelling on *a priori* grounds that it alone warrants examination and reporting. But that is not a serious hindrance to fractional pooling, since it is easy enough to compute and report parameter estimates corresponding to a variety of alternative values of λ (or to a variety of alternative divisions of the available data).

Indeed, it seems likely that the most fruitful application of the approach proposed here will be as a tool for sensitivity testing. Whenever there is uncertainty about the appropriate specification of a statistical model, as there almost always is in nonexperimental work, it is potentially illuminating to be able to explore the implications of alternative assumptions. Fractional pooling provides a rich new range of alternative assumptions whose implications may seem worth exploring, rather than a method for choosing one assumption or another. In any case, it is worth bearing in mind that the inadequacy of an analysis based upon any single value of λ applies not only to values between the endpoints of zero and one, but also to the endpoints themselves. The fact that analysts are used to choosing one endpoint or the other casually or implicitly does not make their choices any less subjective or any less problematic.

In the end, my plea is simply for more self-conscious realism about the theoretical relevance of available data. While availability is itself an inherently practical limitation, the fact that a given observation does or

does not appear in a table or on a data tape should not be the data analyst's first and last consideration. On one hand, maximizing sample size by taking every available observation at face value will usually produce confusion or overconfidence or both. On the other hand, good social science data are in sufficiently short supply that data analysts should not hesitate to make whatever honest use they can of observations whose theoretical pedigree is less than perfect. The technique proposed here offers one way to steer a reasonable course between the unreasonable extremes of voracious inclusiveness and fastidious exclusiveness.

Manuscript submitted 25 May 1995.
Final manuscript received 26 October 1995.

APPENDIX

This appendix provides raw data for the three extended empirical examples considered in Section 3. Ashenfelter's (1994) data on absentee and machine voting in Philadelphia are reproduced in Table A1, Powell's (1982) data on regime characteristics and political violence are reproduced in Table A2, and Alesina, Londregan, and Rosenthal's (1993) data on economic growth and election outcomes in the United States are reproduced in Table A3. Readers are referred to these sources for further information.

Table A1. Ashenfelter Data on Machine and Absentee Vote Margins in Philadelphia

District	Year	Machine Margin (D)	Absentee Margin (D)
2	1982	26,427	346
4	1982	15,904	282
8	1982	42,448	223
1	1984	19,444	593
3	1984	71,797	572
5	1984	−1,017	−229
7	1984	63,406	671
2	1986	15,671	293
4	1986	36,276	360
8	1986	36,710	306
1	1988	21,848	401
3	1988	65,862	378
5	1988	−13,194	−829
7	1988	56,100	394
2	1990	700	151
4	1990	11,529	−349
8	1990	26,047	160
1	1992	44,425	1,329
3	1992	45,512	368
5	1992	−5,700	−434
7	1992	51,206	391
2	1993	−564	1,025

Data on election returns from Pennsylvania Senate districts in Philadelphia area, 1982–93, from Ashenfelter (1994). **Machine Margin:** Democratic margin in votes cast by machine ballot. **Absentee Margin:** Democratic margin in votes cast by absentee ballot.

Table A2. Powell Data on Social, Economic, and Constitutional Determinants of Political Violence

Country	Pop	GNP	Ethnic	Rep	Deaths
Australia	11.3	2,694	.32	2	0
Austria	7.3	1,732	.13	4	0
Belgium	9.5	2,427	.55	4	.1
Canada	19.6	3,327	.75	2	.3
Ceylon	11.2	193	.47	2	121.1
Chile	8.6	760	.14	1	87.3
Costa Rica	1.4	556	.07	1	0
Denmark	4.8	2,853	.05	4	0
Finland	4.6	2,354	.16	4	0
France	48.9	2,589	.26	1	.7
West Germany	59.0	2,558	.03	3	2.7
India	486.7	136	.89	2	328.4
Ireland	2.9	1,319	.04	3	4.0
Italy	51.6	1,485	.04	4	8.9
Jamaica	1.8	669	.04	2	17.9
Japan	98.0	1,159	.02	3	2.7
Netherlands	12.3	2,091	.10	4	.2
New Zealand	2.6	2,664	.37	2	0
Norway	3.7	2,543	.04	4	0
Philippines	32.3	215	.74	1	332.7
Sweden	7.7	3,441	.08	4	0
Turkey	31.1	379	.25	4	18.1
United Kingdom	54.6	2,446	.37	2	160.0
United States	194.6	4,810	.50	1	20.7
Uruguay	2.7	771	.20	1	18.2
Venezuela	8.7	1,187	.11	1	5.9

Data on contemporary democracies from Powell (1982). Bold faced countries categorized as "predominantly traditional" or "mixed traditional and modern" (Table 3.3). **Pop:** population in millions, 1965 (Table 3.1). **GNP:** GNP per capita in $US, 1965 (Table 3.2). **Ethnic:** ethnic fractionalization index (Table 3.4). **Rep:** representational constitution (presidential = 1; majoritarian parliamentary = 2; mixed parliamentary = 3; representational parliamentary = 4) (Table 4.1, 234). **Deaths:** deaths per year from political violence, 1967–76 (Table A.1). Switzerland omitted (missing data on **Rep**).

Table A3. Alesina, Londregan, and Rosenthal Data on GNP Growth and Presidential Election Outcomes

Year	Republican Incumbent	Incumbent Party's Midterm House Vote (%)	Election Year GNP Growth (%)	Incumbent Party's Presidential Vote (%)
1916	0	50.338	7.279	51.626
1920	0	45.096	−1.146	36.190
1924	1	53.600	2.919	65.259
1928	1	58.428	1.191	58.788
1932	1	54.129	−14.406	40.825
1936	0	56.184	13.219	62.487
1940	0	50.815	7.563	54.975
1944	0	47.662	7.863	53.776
1948	0	45.272	3.862	52.326
1952	0	50.041	3.826	44.623
1956	1	47.272	2.033	57.746
1960	1	43.603	2.198	49.899
1964	0	52.327	5.201	61.345
1968	0	51.327	4.064	49.593
1972	1	45.775	4.858	61.813
1976	1	41.323	4.771	48.930
1980	0	54.322	−.166	44.711
1984	1	43.782	6.559	59.155
1988	1	45.005	3.803	53.939

Unpublished data analyzed by Alesina, Londregan, and Rosenthal (1993) provided by Howard Rosenthal.

REFERENCES

Alesina, Alberto, John Londregan, and Howard Rosenthal. 1993. "A Model of the Political Economy of the United States." *American Political Science Review* 87:12–33.

Amemiya, Takeshi. 1980. "Selection of Regressors." *International Economic Review* 21: 331–54.

Ashenfelter, Orley. 1994. "Report on Expected Absentee Ballots." Revised March 29, 1994. Department of Economics, Princeton University. Typescript.

Bartels, Larry M. 1988. *Presidential Primaries and the Dynamics of Public Choice.* Princeton: Princeton University Press.

Bartels, Larry M. 1990. "Five Approaches to Model Specification." *The Political Methodologist* 3:2–6.

Beck, Nathaniel. 1983. "Time-varying Parameter Regression Models." *American Journal of Political Science* 27:557–600.

Beck, Nathaniel. 1985. "Estimating Dynamic Models is Not Merely a Matter of Technique." *Political Methodology* 11:71–89.

Brady, Henry E., and Richard Johnston. 1987. "What's the Primary Message: Horse Race or Issue Journalism?" In *Media and Momentum: The New Hampshire Primary and Nomination Politics,* ed. Gary R. Orren and Nelson W. Polsby. Chatham, NJ: Chatham House.

Brown, R. L., J. Durbin, and J. M. Evans. 1975. "Techniques for Testing the Constancy of Regression Relationships Over Time" (with discussion). *Journal of the Royal Statistical Society, Series B* 37:149–92.

Chamberlain, Gary, and Edward E. Leamer. 1976. "Matrix Weighted Averages and Posterior Bounds." *Journal of the Royal Statistical Society, Series B* 38:73–84.

Chow, Gregory C. 1960. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." *Econometrica* 28:591–605.

Collier, David, and James E. Mahon, Jr. 1993. "Conceptual 'Stretching' Revisited: Adapting Categories in Comparative Analysis." *American Political Science Review* 87:845–55.

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

Dickey, J. M. 1975. "Bayesian Alternatives to the *F*-test and Least-Squares Estimates in the Normal Linear Model." In *Bayesian Studies in Econometrics and Statistics,* ed. S. E. Fienberg and A. Zellner. Amsterdam: North-Holland.

Edwards, John B. 1969. "The Relation between the *F*-test and $\bar{R}^2$." *The American Statistician* 23:No. 5, 28.

Efron, Bradley, and Carl Morris. 1977. "Stein's Paradox in Statistics." *Scientific American* 236:119–27.

Erikson, Robert S. 1989. "Economic Conditions and the Presidential Vote." *American Political Science Review* 83:567–73.

Feldstein, Martin. 1973. "Multicollinearity and the Mean Square Error Criterion." *Econometrica* 41:337–46.

Goldfeld, Stephen, and Richard Quandt. 1973. "The Estimation of Structural Shifts by Switching Regressions." *Annals of Economic and Social Measurement* 2:475–85.

Judge, George G., and M. E. Bock. 1978. *The Statistical Implications of Pre-Test and Stein Rule Estimators in Econometrics.* New York: North-Holland.

Judge, George G., W. E. Griffiths, R. Carter Hill, Helmut Lütkepohl and Tsoung-Chao Lee. 1985. *The Theory and Practice of Econometrics.* 2nd ed. New York: John Wiley and Sons.

Leamer, Edward E. 1978. *Specification Searches: Ad Hoc Inference with Nonexperimental Data.* New York: John Wiley & Sons.

Lindley, D. V., and A. F. M. Smith. 1972. "Bayes Estimates for the Linear Model" (with discussion). *Journal of the Royal Statistical Society, Series B* 34:1–41.

Maritz, J. S. 1970. *Empirical Bayes Methods.* London: Methuen.

Markus, Gregory B. 1988. "The Impact of Personal and National Economic Conditions on the Presidential Vote: A Pooled Cross-Sectional Analysis." *American Journal of Political Science* 32:137–54.

Newbold, Paul, and Theodore Bos. 1985. *Stochastic Parameter Regression Models.* Quantitative Applications in the Social Sciences, No. 51. Newbury Park: Sage Publications.

Powell, G. Bingham, Jr. 1982. *Contemporary Democracies: Participation, Stability, and Violence.* Cambridge: MA: Harvard University Press.

Quandt, Richard. 1958. "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes." *Journal of the American Statistical Association* 53: 873–80.

Rubin, Donald B. 1980. "Using Empirical Bayes Techniques in the Law School Validity Studies." *Journal of the American Statistical Association* 75:801–16.

Schiller, Wendy. 1995. "Senators as Political Entrepreneurs: Using Bill Sponsorship to Shape Legislative Agendas." *American Journal of Political Science* 39:186–203.

Swamy, P. A. V. B., and J. S. Mehta. 1975. "Bayesian and Non-Bayesian Analysis of Switching Regressions and of Random Coefficient Regression Models." *Journal of the American Statistical Association* 70:593–602.

Theil, Henry, and Arthur S. Goldberger. 1961. "On Pure and Mixed Statistical Estimation in Economics." *International Economic Review* 2:65–78.

Tufte, Edward R. 1978. *Political Control of the Economy.* Princeton: Princeton University Press.

Wallace, T. D. 1964. "Efficiencies for Stepwise Regression." *Journal of the American Statistical Association* 59:1179–82.

Western, Bruce, and Simon Jackman. 1994. "Bayesian Inference for Comparative Research." *American Political Science Review* 88:412–23.

Zellner, Arnold. 1971. *An Introduction to Bayesian Inference in Econometrics.* New York: John Wiley & Sons.