The Effects of Measurement Error on Two-Stage, Least-Squares Estimates

Donald Philip Green

Abstract

Two-stage least squares (2SLS) is a statistical procedure that is used to correct for simultaneity bias and errors in variables. When applied to certain kinds of models, however, 2SLS is itself susceptible to bias as a result of random and nonrandom measurement error in the data. Using data from the 1980 Center for Political Studies panel, I show how different assumptions about measurement error produce radically different impressions about the reciprocal relationship between party identification and presidential performance evaluations.

Nonrecursive estimation techniques such as two-stage least squares (2SLS), which were relatively unknown before 1970, are now in widespread use in the social sciences. The advent of these procedures reflects a growing attentiveness to the existence of "two-way causation" in social systems, whereby regressors are both causes and consequences of the dependent variable. The development of more realistic structural models of social phenomena undoubtedly represents an advance. More questionable is the way such models have been subjected to empirical assessment.

Too often, nonrecursive analyses of great theoretical complexity have proceeded on the unrealistic assumption that the data contain no measurement error. In the political science literature, for example, many recent studies of reciprocal relationships between party identification and "short-term forces"—evaluations of presidential performance in office, positions on the major issues of the day, or evaluations of the two parties' presidential nominees—implicitly assume that the survey data being analyzed capture

---

I am grateful to Jon Cowden, Tom Hartley, and Brad Palmquist for their many helpful comments and suggestions. I would also like to thank Ann Gerken, JoAnn Dionne, and the Social Science Data Archive at Yale.

voters' loyalties and evaluations without measurement error (Page and Jones 1979; Markus and Converse 1979; Fiorina 1981; Markus 1982).

The purpose of this article is to show how measurement assumptions may influence the 2SLS estimates one obtains. For purposes of algebraic exposition, I examine the statistical properties of a standard simultaneous-equations model (see Johnston 1984, 456–57) that political scientists have used to estimate the influence of short-term forces on partisanship (see Fiorina 1981; Markus and Converse 1979; Markus 1982).[1] After discussing the biases produced by random and nonrandom measurement error, I apply this model to data from the 1980 Center for Political Studies panel survey. The estimated effect of presidential approval on party identification varies from large to small depending upon the measurement assumptions one invokes. The results suggest that measurement questions must be settled before substantive disputes may be adjudicated.

Algebraic Demonstration of the Consequences of Measurement Error

Consider the nonrecursive system of $k = 2$ equations:

$$\eta_{1i} = \beta_{12}\eta_{2i} + \gamma_{11}\xi_{1i} + \zeta_{1i}, \tag{1}$$

and

$$\eta_{2i} = \beta_{21}\eta_{1i} + \gamma_{22}\xi_{2i} + \zeta_{2i}. \tag{2}$$

In this example, η_{1i} is determined by its past value, ξ_{1i}, and the contemporaneous value of η_{2i}. Conversely, η_{2i} is affected by its past value, ξ_{2i}, and the contemporaneous value of η_{1i}. When we assume $plim\ n^{-1}\xi'_{1i}\zeta_{2i} = plim\ n^{-1}\xi'_{2i}\zeta_{1i} = 0$ and the latent variables to be measured without error, this system of equations is nothing more than a standard textbook example, and all the structural parameters may be estimated consistently using two-stage least squares, instrumental variables, or indirect least squares.[2] We shall assume, however, that the observed measures are fallible. We let

$$x_{1i} = \xi_{1i} + \delta_{1i}, \tag{3}$$

$$x_{2i} = \xi_{2i} + \delta_{2i}, \tag{4}$$

1. These studies each analyze panel data and use lagged endogenous variables as instruments. I shall do likewise in the sections that follow.

2. All three estimators are algebraicly equivalent when applied to just identified systems of equations.

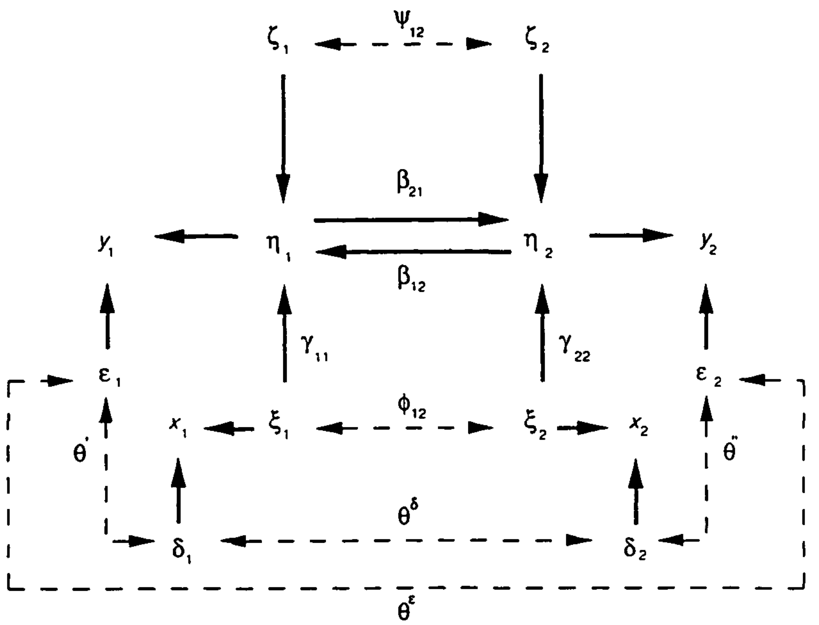

Fig. 1. Nonrecursive model with latent variables. Solid arrows represent causal paths; dashed arrows represent covariances between latent variables.

$$y_{1i} = \eta_{1i} + \epsilon_{1i}, \tag{5}$$

and

$$y_{2i} = \eta_{2i} + \epsilon_{2i}, \tag{6}$$

where ϵ_{ki} and δ_{ki} represent errors of measurement (see fig. 1). Assume further that as $n \rightarrow \infty$, the structural disturbance term and the errors of measurement take on the following properties:

$$plim\ n^{-1}\xi'_{ki}\zeta_{ki} = 0 \qquad plim\ n^{-1}\delta'_{ki}\zeta_{ki} = 0 \qquad plim\ n^{-1}\xi'_{ki}\delta_{ki} = 0$$

$$plim\ n^{-1}\delta'_{1i}\delta_{2i} = 0 \qquad plim\ n^{-1}\xi'_{ki}\epsilon_{ki} = 0 \qquad plim\ n^{-1}\delta'_{ki}\epsilon_{ki} = 0$$

$$plim\ n^{-1}\zeta'_{ki}\epsilon_{ki} = 0 \qquad plim\ n^{-1}\delta'_{1i}\delta_{1i} = \theta_{11} \qquad plim\ n^{-1}\delta'_{2i}\delta_{2i} = \theta_{22}$$

$$plim\ n^{-1}\zeta'_{1i}\zeta_{2i} = \psi_{12} \qquad plim\ n^{-1}\zeta'_{1i}\zeta_{1i} = \psi_{11} \qquad plim\ n^{-1}\zeta'_{2i}\zeta_{2i} = \psi_{22}.$$

In essence, the errors of measurement associated with a given variable are uncorrelated with the latent traits, the structural disturbance term, and the errors of measurement associated with other variables. We shall refer to measurement error of this kind as "random" error.

For notational convenience we shall assume that as the sample size increases, the regressors take on the following properties:

$$plim\ n^{-1}\xi'_{1i}\xi_{1i} = \phi_{11}$$

$$plim\ n^{-1}\xi'_{1i}\xi_{2i} = \phi_{12}$$

$$plim\ n^{-1}\xi'_{2i}\xi_{2i} = \phi_{22}.$$

From the foregoing assumptions, we may deduce the expected population covariance matrix for the observed variables, x_{ki} and y_{ki}. Looking ahead to the algebra that follows, we find the terms of most interest to be $cov(y_{ki},x_{ki})$, $var(x_{ki})$, and $cov(x_{1i},x_{2i})$. Taking *plims* gives

$$plim\ n^{-1}x'_{1i}x_{1i} = \phi_{11} + \theta_{11}$$

$$plim\ n^{-1}x'_{2i}x_{2i} = \phi_{22} + \theta_{22}$$

$$plim\ n^{-1}x'_{1i}x_{2i} = \phi_{12}$$

$$plim\ n^{-1}x'_{1i}y_{1i} = D(\gamma_{11}\phi_{11} + \beta_{12}\gamma_{22}\phi_{12})$$

$$plim\ n^{-1}x'_{2i}y_{1i} = D(\gamma_{11}\phi_{12} + \beta_{12}\gamma_{22}\phi_{22})$$

$$plim\ n^{-1}x'_{1i}y_{2i} = D(\gamma_{22}\phi_{12} + \beta_{21}\gamma_{11}\phi_{11})$$

$$plim\ n^{-1}x'_{2i}y_{2i} = D(\gamma_{22}\phi_{22} + \beta_{21}\gamma_{11}\phi_{12})$$

where

$$D = \frac{1}{1 - \beta_{12}\beta_{21}}.$$

What happens when we apply two-stage least squares to this system of equations? Let us first consider γ^*_{11}, the first element in a (2×1) vector of 2SLS estimates from the formula

$$(Z'X)^{-1}Z'y_1 = \begin{pmatrix} \gamma^*_{11} \\ \beta^*_{12} \end{pmatrix}$$

where the $(n \times 2)$ instrument matrix Z is composed of the two vectors, x_{1i} and x_{2i}, and where the $(n \times 2)$ regressor matrix X is composed of the two vectors, x_{1i} and y_{2i}. If we assume (without loss of generality) that the x_{ki} have mean zero, we may express the 2SLS estimator as

$$\gamma^{*}_{11} = \frac{cov(x_{2i}, y_{2i}) \cdot cov(x_{1i}, y_{1i}) - cov(x_{1i}, y_{2i}) \cdot cov(x_{2i}, y_{1i})}{var(x_{1i}) \cdot cov(x_{2i}, y_{2i}) - cov(x_{1i}, x_{2i}) \cdot cov(x_{1i}, y_{2i})}. \quad (7)$$

Taking *plim*s of the previous expression gives

$$\frac{D^2(\gamma_{22}\phi_{22} + \beta_{21}\gamma_{11}\phi_{12})(\gamma_{11}\phi_{11} + \beta_{12}\gamma_{22}\phi_{12})}{(\phi_{11} + \theta_{11})D(\gamma_{22}\phi_{22} + \beta_{21}\gamma_{11}\phi_{12})}$$

$$\frac{-\,D^2(\gamma_{22}\phi_{12} + \beta_{21}\gamma_{11}\phi_{11})(\gamma_{11}\phi_{12} + \beta_{12}\gamma_{22}\phi_{11})}{-\,(\phi_{12})D(\gamma_{22}\phi_{12} + \beta_{21}\gamma_{11}\phi_{11})}.$$

We may simplify the above expression to the following:

$$plim\,\gamma^{*}_{11} =$$

$$\frac{D^2[\gamma_{11}\gamma_{22}\phi_{11}\phi_{22} + \beta_{12}\beta_{21}\gamma_{11}\gamma_{22}\phi^2_{12} - \gamma_{11}\gamma_{22}\phi^2_{12} - \beta_{12}\beta_{21}\gamma_{11}\gamma_{22}\phi_{11}\phi_{22}]}{D[\gamma_{22}\phi_{11}\phi_{22} + \theta_{11}\gamma_{22}\phi_{22} + \theta_{11}\beta_{21}\gamma_{11}\phi_{12} - \gamma_{22}\phi^2_{12}]}$$

$$= \frac{\gamma_{11}[\phi_{11}\phi_{22} - \phi^2_{12}]}{\phi_{11}\phi_{22} - \phi^2_{12} + \theta_{11}\phi_{22} + \dfrac{\theta_{11}\beta_{21}\gamma_{11}\phi_{12}}{\gamma_{22}}}. \quad (8)$$

Provided $\gamma_{11} > 0$ and the denominator in (8) is positive, γ^{*}_{11} will be an overestimate when

$$\theta_{11}\left(\phi_{22} + \frac{\beta_{21}\gamma_{11}\phi_{12}}{\gamma_{22}}\right) < 0. \quad (9)$$

If $\theta_{11} = 0$, then γ^{*}_{11} will be a consistent estimator of γ_{11}. If ξ_1 is measured with error, however, the estimate of γ_{11} may be seriously biased. In the special case where ξ_{1i} and ξ_{2i} are orthogonal (implying that $\phi_{12} = 0$) but x_{1i} contains random error, γ^{*}_{11} will underestimate γ_{11}. More generally, when x_{1i} contains error and the ξ_{ki} are correlated, γ_{11} will be underestimated as long as all of the parameters take on positive values (which is generally the case in nonrecursive models of partisanship). In the more complex case in which all of the parameters are positive except β_{21}, γ^{*}_{11} may be biased in either direction. Finally, note that parameter θ_{22} nowhere appears in the result, indicating that the presence of measurement error in x_{2i} has no bearing on the estimate of γ_{11}.

Turning now to the estimate of β_{12}, we find that the effects of measurement error are asymmetrical in that error in x_{1i} leads to bias, while error in x_{2i} does not. The 2SLS estimator of β_{12} may be written

$$\beta_{12}^{*} = \frac{var(x_{1i}) \cdot cov(x_{2i},y_{1i}) - cov(x_{1i},x_{2i}) \cdot cov(x_{1i},y_{1i})}{var(x_{1i}) \cdot cov(x_{2i},y_{2i}) - cov(x_{1i},x_{2i}) \cdot cov(x_{1i},y_{2i})}. \tag{10}$$

Taking *plims* gives

$$plim\ \beta_{12}^{*} =$$

$$\frac{(\phi_{11} + \theta_{11})D(\gamma_{11}\phi_{12} + \beta_{12}\gamma_{22}\phi_{22}) - (\phi_{12})D(\gamma_{11}\phi_{11} + \beta_{12}\gamma_{22}\phi_{12})}{(\phi_{11} + \theta_{11})D(\gamma_{22}\phi_{22} + \beta_{21}\gamma_{11}\phi_{12}) - (\phi_{12})D(\gamma_{22}\phi_{12} + \beta_{21}\gamma_{11}\phi_{11})},$$

which simplifies to

$$plim\ \beta_{12}^{*} = \frac{\beta_{12}\left(\phi_{11}\phi_{22} - \phi_{12}^{2} + \theta_{11}\phi_{22} + \dfrac{\theta_{11}\gamma_{11}\phi_{12}}{\beta_{12}\gamma_{22}}\right)}{\phi_{11}\phi_{22} - \phi_{12}^{2} + \theta_{11}\phi_{22} + \dfrac{\theta_{11}\beta_{21}\gamma_{11}\phi_{12}}{\gamma_{22}}}. \tag{11}$$

Assuming the numerator and denominator of (11) to be positive, we see that $\beta_{12}^{*} > \beta_{12}$ for $\beta_{12} > 0$ when

$$\frac{\theta_{11}\gamma_{11}\phi_{12}}{\gamma_{22}}\left(\frac{1}{\beta_{12}} - \beta_{21}\right) > 0. \tag{12}$$

As before, we find that if $\theta_{11} = 0$, β_{12} will be a consistent estimator. Similarly, β_{12}^{*} will be consistent if $\gamma_{11} = 0$, since x_{1i} would represent an irrelevant variable that may be dropped from the estimation equation. When $\phi_{12} = 0$, the regressors are orthogonal, and x_{1i} may be dropped from the estimating equation without changing the results. If $\gamma_{22} = 0$, the equation is simply unidentified.

In the more interesting case where θ_{11}, γ_{11}, γ_{22}, and ϕ_{12} all exceed zero, the bias to β_{12}^{*} depends on the relative magnitudes of β_{12} and β_{21}. Although one might conceive of situations in which reciprocal causation is so strong that the inverse of β_{12} is less than β_{21}, studies of voting behavior characteristically indicate that $1/\beta_{12} > \beta_{21}$. Thus, when ξ_1 is measured with error (and hence $\theta_{11} > 0$), the result is typically an overestimate of β_{12}. Finally, note that the error component of x_{2i}, θ_{22}, has no bearing on the final estimate. In sum, the assumptions we choose to make about the presence or absence of random error in x_{2i} have no impact on the estimates we derive for γ_{11} and β_{12}. For that matter, neither do the assumptions about random error in y_{ki}. On the other hand, assumptions concerning θ_{11} may have substantial ramifications for the estimates of these two parameters.

The Consequences of Nonrandom Error

The previous section assumed the errors of measurement (δ_{ki} and ϵ_{ki}) to be mutually uncorrelated. In social science, this assumption may prove to be false for two reasons. First, the measurement errors of different traits at a given time may be correlated if both are measured within the same environment or by the same measuring procedure (Campbell and Fiske 1959; Werts, Joreskog, and Linn 1976). For example, if survey respondents present themselves as more liberal than they really are when questioned by a black interviewer, and more conservative when interviewed by a white Southerner, a liberal or conservative bias may influence responses throughout the interview. If this is the case, it may be unrealistic to assume that $cov(\delta_{1i},\delta_{2i}) = cov(\epsilon_{1i},\epsilon_{2i}) = 0$. Second, the same method is often used to measure a trait at two times. To the extent that the errors of measurement at time one are repeated at time two, the assumption that $cov(\delta_{1i},\epsilon_{1i}) = cov(\delta_{2i},\epsilon_{2i}) = 0$ will prove to be false (Werts, Joreskog, and Linn 1976; Wiley and Wiley 1974). We shall refer to the former type of nonrandom error as within-wave correlated error and the latter as across-wave correlated error.

First consider the case of within-wave correlated error. We amend our earlier assumptions by defining the following *plims*:

$$plim\ n^{-1}\delta_{1i}'\delta_{2i} = \theta^{\delta},$$

and

$$plim\ n^{-1}\epsilon_{1i}'\epsilon_{2i} = \theta^{\epsilon}.$$

These assumptions imply the following:

$$plim\ \beta_{12}^{*} = \frac{(\phi_{11} + \theta_{11})D(\gamma_{11}\phi_{12} + \beta_{12}\gamma_{22}\phi_{22}) - (\phi_{12} + \theta^{\delta})D(\gamma_{11}\phi_{11} + \beta_{12}\gamma_{22}\phi_{12})}{(\phi_{11} + \theta_{11})D(\gamma_{22}\phi_{22} + \beta_{21}\gamma_{11}\phi_{12}) - (\phi_{12} + \theta^{\delta})D(\gamma_{22}\phi_{12} + \beta_{21}\gamma_{11}\phi_{11})},$$

which simplifies to

$$plim\ \beta_{12}^{*} = \frac{\beta_{12}\left[\phi_{11}\phi_{22} - \phi_{12}^{2} + \theta_{11}\phi_{22} + \dfrac{\theta_{11}\gamma_{11}\phi_{12}}{\beta_{12}\gamma_{22}} - \theta^{\delta}\left(\phi_{12} + \dfrac{\gamma_{11}\phi_{11}}{\beta_{12}\gamma_{22}}\right)\right]}{\phi_{11}\phi_{22} - \phi_{12}^{2} + \theta_{11}\phi_{22} + \dfrac{\theta_{11}\beta_{21}\gamma_{11}\phi_{12}}{\gamma_{22}} - \theta^{\delta}\left(\phi_{12} + \dfrac{\beta_{21}\gamma_{11}\phi_{11}}{\gamma_{22}}\right)}. \quad (13)$$

Again, assuming that β_{12} and (13) are positive, β^*_{12} will be an overestimate when

$$\left(\frac{\gamma_{11}}{\gamma_{22}}\right)\left(\frac{1}{\beta_{12}} - \beta_{21}\right)(\theta_{11}\phi_{12} - \theta^{\delta}\phi_{11}) > 0. \tag{14}$$

The first term in the equation will invariably be positive for models such as ours that use lagged endogenous variables as instruments. And as we mentioned earlier, the middle term in equation 14 is likely to be positive. Under these conditions, β^*_{12} will be an overestimate when

$$\frac{\phi_{12}}{\phi_{11}} > \frac{\theta^{\delta}}{\theta_{11}}.$$

This inequality may be interpreted as a comparison of two regression coefficients. On the left is what one would obtain were one to regress ξ_{2i} on ξ_{1i}; on the right, what one would obtain from a regression of δ_{2i} on δ_{1i}. In essence, if there is a stronger relationship between the error component of the observed regressors than between the trait component, β^*_{12} will be an underestimate of β_{12}.

Turning to the case of across-wave correlated error, we assume that $plim\ n^{-1}\delta'_{1i}\delta_{2i} = plim\ n^{-1}\epsilon'_{1i}\epsilon_{2i} = 0$, but allow the following *plim*s to take on nonzero values:

$$cov(\delta_{1i},\epsilon_{1i}) = \theta';$$

$$cov(\delta_{2i},\epsilon_{2i}) = \theta''.$$

Under these assumptions

$plim\ \beta^*_{12} =$

$$\frac{(\phi_{11} + \theta_{11})D(\gamma_{11}\phi_{12} + \beta_{12}\gamma_{22}\phi_{22}) - (\phi_{12})[D(\gamma_{11}\phi_{11} + \beta_{12}\gamma_{22}\phi_{12}) + \theta']}{(\phi_{11} + \theta_{11})[D(\gamma_{22}\phi_{22} + \beta_{21}\gamma_{11}\phi_{12}) + \theta''] - (\phi_{12})D(\gamma_{22}\phi_{12} + \beta_{21}\gamma_{11}\phi_{11})}.$$

Rearranging terms gives

$plim\ \beta^*_{12} =$

$$\frac{\beta_{12}\left(\phi_{11}\phi_{22} - \phi_{12}^2 + \theta_{11}\phi_{22} + \dfrac{\theta_{11}\gamma_{11}\phi_{12}}{\beta_{12}\gamma_{22}} - \dfrac{\phi_{12}\theta'}{\beta_{12}\gamma_{22}D}\right)}{\phi_{11}\phi_{22} - \phi_{12}^2 + \theta_{11}\phi_{22} + \dfrac{\theta_{11}\beta_{21}\gamma_{11}\phi_{12}}{\gamma_{22}} + \dfrac{\theta''(\phi_{11} + \theta_{11})}{\gamma_{22}D}}. \tag{15}$$

If β_{12} and (15) are positive, β^*_{12} will be an overestimate when

$$\frac{1}{\gamma_{22}}\left[\left(\frac{1}{\beta_{12}}-\beta_{21}\right)(\theta_{11}\gamma_{11}\phi_{12})-\frac{\phi_{12}\theta'}{\beta_{12}D}-\frac{\theta''(\phi_{11}+\theta_{11})}{D}\right]>0. \qquad (16)$$

While this expression does not lend itself to intuitive interpretation, a number of informative conclusions might be drawn from it. Let us, for the sake of simplicity, limit our discussion to the typical case in which γ_{22} and D are positive and ϕ_{22} is nonzero. In contrast to earlier results based on the assumption of random error, the measurement properties of x_{2i} come into play when measurement errors are correlated. We find, for example, that the greater the covariance between δ_{2i} and ϵ_{2i}, the smaller the estimate of β^*_{12}. Positive covariance between errors in x_{1i} and y_{1i} also decreases the estimate of β^*_{12} as long as the latent exogenous variables are not orthogonal. Indeed, if across-wave error covariance is of sufficient magnitude, β^*_{12} will prove to be an underestimate when 2SLS is run without allowance for measurement error.[3]

Notice that whether we assume random or nonrandom error, parameter θ_{22} nowhere enters the formula for β^*_{12}. Only the covariances between δ_{2i} and the other measurement errors affect the estimate of β_{12}. This is significant because research on measurement error in social science data focuses primarily on estimating item reliability and devotes relatively little attention to nonrandom error. As a result, researchers seeking to take measurement error into account when analyzing simultaneous equations may often find that relatively little is known about the nonrandom error structure of the data.

Finally, we point out that the variances of ϵ_{1i} and ϵ_{2i}, as well as the covariance between them, have no effect on the estimates of β_{12} or γ_{11}. Measurement error in the endogenous variables, y_{1i} and y_{2i}, is problematic only when correlated with the errors in the predetermined variables.

A Substantive Illustration of the Effects of Measurement Error for Two-Stage, Least-Squares Analysis

In the previous section, we demonstrated that 2SLS results may be profoundly affected by the presence of measurement error in certain variables. In this section, we illustrate this proposition by examining the interplay between partisanship and presidential performance evaluations. Fiorina (1981) has suggested that an individual's party identification is both a cause and consequence of the way he or she evaluates the incumbent administration's perfor-

3. Both within-wave and across-wave error covariance affect the estimate of γ_{11} as well. The algebraic expressions are tedious, however, and do not lend themselves to easy substantive interpretation.

mance in office. Thus, all things being equal, individuals who approve of the incumbent Republican administration become more Republican in their party loyalties; those who disapprove become more Democratic. In this way, the performance of the economy, scandals, and foreign affairs may influence the balance of party loyalties among the electorate.

Data

In order to test this proposition, we analyze data from the 1980 National Election Panel Study, which conducted interviews with a random sample of adults in January and June of that year. This period was marked by the Iran hostage crisis as well as a sharp decline in the nation's economy, and President Carter's approval rating fell precipitously (Brody and Rothenberg 1988; Green and Palmquist 1990). The question is: Did increasing disapproval of Carter's performance in office produce a shift in party loyalties?

Measures

The measure of party loyalty is the standard Michigan party identification scale. Respondents are asked whether they think of themselves as Democrats, Republicans, or Independents. Those who think of themselves as Independents are asked whether they feel closer to the Democratic or Republican parties. Those who, in the initial question, think of themselves as Democrats (Republicans) are asked whether they think of themselves as "strong" Democrats (Republicans). Piecing the responses together creates a seven-point continuum ranging from Strong Democrat to Strong Republican, with Independents who do not lean toward either party in the middle.

The measure of perceived presidential performance in office is a branched question which asks the respondent whether he or she approves or disapproves of the way Carter has handled his job as president, and then asks whether he or she approves (disapproves) strongly or not strongly. The exact wording of each question may be found in the appendix.

It seems clear that neither measure is immune to measurement error. This conclusion has been demonstrated empirically in the case of the Michigan party identification scale. Test-retest reliability analyses based on data from 1956–60, 1972–76, as well as the 1980 Panel suggest that the measurement error variance is approximately .4 (Asher 1974; Converse and Markus 1979; Green and Palmquist 1990).[4] Although this degree of reliability is good by survey research standards (see Achen 1975), the presence of even this small

4. These findings imply that the reliability of the party identification measure is approximately .9.

TABLE 1. Covariance Matrix: Party Identification and Presidential Approval

Party Identification (January)	Party Identification (June)	Presidential Approval (January)	Presidential Approval (June)
3.9597			
3.4984	4.1897		
0.8615	0.8807	2.3841	
0.9549	1.0040	1.3066	2.1887

Source: 1980 National Election Study Panel Study.

Notes: Missing data have been deleted in listwise fashion. See the appendix for missing data codes and frequencies.

amount of measurement error turns out to create serious problems for two-stage, least-squares analysis.

Results

Let party identification (PID) at time one be ξ_{1i} and let PID at time two be η_{1i}. Similarly, let presidential performance evaluations at time one be ξ_{2i}; at time two, η_{2i}. From the previous discussion, it follows that only the error properties of x_{1i}, party identification at time one, will be relevant to the estimation of β_{12}, the effect of presidential approval on party identification. By the same token, only the error properties of x_{1i} will have any bearing on the estimate of γ_{11}, the over-time stability of party identification. To be sure, measurement error in x_{2i} will tend to bias β_{21} and γ_{22}, but these parameters do not concern us here.

Table 1 presents the covariance matrix for party identification and presidential approval, based on the 637 respondents who were interviewed in both January and June, 1980. Table 2 presents two-stage, least-squares results under various measurement assumptions. The first is that each of the observed indicators contains no measurement error. The estimates based on this common, if blithe, assumption suggest that retrospective performance evaluations have a profound influence on party loyalty. The 2SLS estimate of β_{12} (.11) suggests that a four-unit change from strong approval of Carter to strong disapproval causes partisanship to shift almost a half-point in the Republican direction. This, in conjunction with the fact that the estimate of γ_{11} (.86) falls well below unity, confirms the notion that partisanship is relatively labile over short periods of time.

When the party identification measure at time one is assumed to have a measurement error variance (θ_{11}) of .40, however, the picture that emerges is quite different. The middle column of table 2 shows partisanship to be highly

TABLE 2. Apparent Effects of Performance Evaluations on Party Identification Under Different Assumptions about Random Measurement Error

	Assume No Error		Assume θ_{11} = .4		Assume an Error Variance .4 for all Variables	
	Estimate	SE	Estimate	SE	Estimate	SE
β_{12}	0.109	.056	0.032	.058	0.032	.058
β_{21}	0.152	.028	0.152	.028	0.124	.029
γ_{11}	0.857	.025	0.974	.029	0.974	.029
γ_{22}	0.492	.032	0.492	.032	0.603	.040
ϕ_{11}	3.960	.222	3.560	.222	3.560	.222
ϕ_{22}	2.384	.134	2.384	.134	1.984	.134
ϕ_{12}	0.862	.127	0.862	.122	0.862	.127
ψ_{11}	1.087	.061	0.749	.063	0.349	.063
ψ_{22}	1.403	.079	1.403	.079	0.875	.080
ψ_{12}	−0.217	.098	−0.109	.100	−0.026	.079

Source: 1980 National Election Study Panel Study.

Notes: All entries are 2SLS estimates and maximum likelihood standard errors. N = 637. Party identification is a seven-point scale, ranging from 0 to 6. Presidential approval is a four-point scale, ranging from 1 to 5. For question wording and scoring procedures, see the appendix.

stable (γ^*_{11} = .97) and virtually unaffected by performance evaluations (β^*_{12} = .03). As the algebraic demonstration presented above would suggest, failure to take measurement error in x_{1i} into account results in an overestimate of β_{12} and an underestimate of γ_{11}.

Finally, when we assume that *both* lagged party identification and lagged retrospective performance variables contain measurement error ($\theta_{11} = \theta_{22}$ = .4), we turn up identical results for β^*_{12} and γ^*_{11}. Nor do these estimates or their standard errors change when we also assume that the variances of the contemporaneous measurement errors (ϵ_{1i} and ϵ_{2i}) equal .4 (see the right-hand column of table 2). In short, when measurement error is assumed to be random, the only measurement assumptions that matter are those pertaining to θ_{11}.[5]

Results Assuming Nonrandom Error

Within-Wave Correlated Error

By failing to take random error into account, nonrecursive models may tend to overstate the influence of short-term forces on party identification. Survey

5. Thus, the question of whether .4 is an accurate estimate for θ_{22} is moot.

TABLE 3. Apparent Effects of Performance Evaluations on Party Identification Under Different Assumptions about Within-Wave Nonrandom Measurement Error

	Assume $\theta^\delta = \theta^\epsilon = .1$		Assume $\theta^\delta = \theta^\epsilon = .2$		Assume $\theta^\delta = \theta^\epsilon = .3$	
	Estimate	SE	Estimate	SE	Estimate	SE
β_{12}	0.120	.057	0.204	.056	0.285	.056
β_{21}	0.144	.029	0.162	.028	0.180	.028
γ_{11}	0.951	.028	0.928	.028	0.907	.028
γ_{22}	0.595	.040	0.587	.039	0.579	.038
ϕ_{11}	3.560	.222	3.560	.222	3.560	.222
ϕ_{22}	1.984	.134	1.984	.134	1.984	.134
ϕ_{12}	0.762	.127	0.662	.127	0.562	.127
ψ_{11}	0.382	.068	0.470	.078	0.609	.091
ψ_{22}	0.905	.081	0.945	.082	0.994	.083
ψ_{12}	−0.269	.078	−0.509	.078	−0.746	.080

Source: 1980 National Election Study Panel Study.

Notes: All entries are 2SLS estimates and maximum likelihood standard errors. $N = 637$. All variables are assumed to have a measurement error variance of .4. Party identification is a seven-point scale, ranging from 0 to 6. Presidential approval is a four-point scale, ranging from 1 to 5. For question wording and scoring procedures, see the appendix.

data, however, may contain nonrandom as well as random measurement error. Earlier we derived a result that suggested that β^*_{12} would be an underestimate if the error component of x_{ki} is more strongly interrelated than the trait component. Thus, the question of whether 2SLS results overstate or understate the influence of short-term forces hinges on the value of θ^δ.

As it turns out, θ^δ is a difficult parameter to estimate in general, but especially difficult given simultaneous causality. Suppose for the purposes of illustration, however, that the measurement error variance for each of the indicators is .4, Table 3 shows how the parameter estimates change as we stipulate values for θ^δ ranging from .1 to .3.[6] In each case, the estimates of β_{12} strongly support Fiorina's theory of retrospective adjustment of partisanship. Indeed, when θ^δ is set to .2, β^*_{12} proves to be twice as large as when the data were assumed to contain no measurement error (cf. table 2). The contrast between these results and those obtained previously illustrates the importance of measurement assumptions and the need for further research on the measurement properties of survey data.

If one were to hazard a guess as to which set of assumptions is more plausible in this particular case, the random error model might win out over the within-wave covariance model. In the first place, the error covariance

6. In order to be consistent, we assume that $\theta^\delta = \theta^\epsilon$.

assumptions just invoked produce a highly questionable estimate of ψ_{12}, the covariance between the structural disturbances. While intuition suggests that the unobserved factors that cause one to change one's partisanship in the Republican direction would be *positively* correlated with the unobserved factors that cause one to evaluate Carter more harshly, the results in table 3 indicate that this correlation is $-.76$ when θ^{δ} is set to .20. This implausible result may well indicate that incorrect measurement assumptions have been imposed. There are also substantive reasons for thinking that within-wave error covariance is not a serious problem here. Interviewer effects arising from social desirability tend to be mild for questions on nonracial topics, such as presidential performance ratings and partisanship (Cotter, Cohen, and Coulter 1982; Hatchett and Schuman 1975). The two questions were posed at different points in the January interview, separated by a considerable number of questions; thus, memory and priming effects do not seem to be a problem. Finally, the two items do not share biases due to common response format (Green 1988) or respondent acquiescence (Couch and Keniston 1960). Thus, the case of party identification and presidential approval is not one which leads us to suspect the presence of within-wave error covariance. But in the absence of direct evidence on this point, the question of within-wave covariance remains open.

Across-wave correlated error

The fact that partisanship and presidential approval are each measured in the same fashion at both wave one and wave two raises the possibility of across-wave correlated error. Estimating across-wave covariance is difficult, however, requiring either multiple indicators or at least four waves of panel data (Palmquist and Green 1989). Not surprisingly, the across-wave error covariance for approval (θ'') has not been studied. On the other hand, two analyses of the error covariance for party identification (θ') have been undertaken, using the 1956–60 and 1980 four-wave CPS panel studies. In both cases, the across-wave error covariance seems to be negligible (Palmquist and Green 1989).

In view of this evidence, we construct three versions of the simultaneous equations model. The first stipulates an error covariance of .20 for both θ' and θ'' (again assuming that $\theta_{11} = \theta_{22} = .40$). The second assumes $\theta' = .10$, but $\theta'' = .20$. The third sets θ' to zero and θ'' to .20. The results appear in table 4. We find that when a substantial degree of across-wave error covariance is assumed for both variables, the results are similar to what we obtained when we assumed no error in the data. But when error covariance is stipulated only for the approval measures, the estimates closely resemble the results of the random error model, and support for Fiorina's position again evaporates.

TABLE 4. Apparent Effects of Performance Evaluations on Party Identification Under Different Assumptions about Across-Wave Nonrandom Measurement Error

	Assume $\theta' = \theta'' = .2$		Assume $\theta'' = .2,\ \theta' = .1$		Assume $\theta'' = 0.2,\ \theta' = 0.0$	
	Estimate	SE	Estimate	SE	Estimate	SE
β_{12}	0.094	.070	0.067	.071	0.039	.071
β_{21}	0.163	.031	0.157	.030	0.152	.029
γ_{11}	0.901	.030	0.937	.031	0.972	.031
γ_{22}	0.486	.040	0.488	.040	0.490	.039
ϕ_{11}	3.560	.222	3.560	.222	3.560	.222
ϕ_{22}	1.984	.134	1.984	.134	1.984	.134
ϕ_{12}	0.862	.127	0.862	.127	0.862	.127
ψ_{11}	0.725	.061	0.540	.061	0.349	.063
ψ_{22}	1.094	.079	1.094	.079	1.094	.079
ψ_{12}	−0.143	.096	−0.094	.095	−0.047	.095

Source: 1980 National Election Study Panel Study.

Notes: All entries are 2SLS estimates and maximum likelihood standard errors. $N = 637$. All variables are assumed to have measurement error variances of .4. Party identification is a seven-point scale, ranging from 0 to 6. Presidential approval is a four-point scale, ranging from 1 to 5. For question wording and scoring procedures, see the appendix.

Thus, depending on the measurement assumptions we make, our results vary from strong endorsement to strong rejection of the Fiorina thesis.

Conclusion

Much of the debate surrounding 2SLS models of partisanship and voting behavior focuses on the *structural* assumptions that underlie these models. One scholar proposes a set of variables to be excluded from the second-stage equation (e.g., Jackson 1975); another replies that these instruments may be systematically related to omitted determinants of the dependent variable (e.g., Markus 1982). The issue of measurement error is absent from the discussion of whether a proposed exogenous variable is truly exogenous. It may be that biases stemming from reciprocal causation are perceived to be much more serious than those that arise from mismeasurement. This perception is not unfounded, for even slight specification errors can be devastating (Bartels 1989), whereas certain nonrecursive models are entirely immune to biases due to measurement error. But as we have seen, measurement error can, under certain conditions, play havoc with 2SLS estimates. In particular, when instrumental variables in one equation serve as regressors in another, 2SLS corrects for random measurement error in the y_{ki} variable, but allows the remaining random and nonrandom error to introduce serious biases.

Examining the direction of these biases is one of the objectives of this article. While the algebraic results presented here are not mathematically innovative, neither are they wholly intuitive or widely known. Intuition suggests that to correct for random measurement error in party identification but not presidential approval unfairly stacks the deck in favor of partisan stability. This turns out not to be the case; when random error is assumed, the reliability of the approval measures has no effect on the estimates for β_{12} or γ_{11}. The results concerning nonrandom error are no more intuitive. One might suppose that positive within-wave error covariance would inflate the correlation between party identification and presidential approval and thus increase the apparent endogeneity of party identification, but the opposite turns out to be the case. And in the case of across-wave correlated error, we find that positive values of *both* θ' and θ'' may lead to underestimates of β_{12}. Thus, the algebraic section of this paper provides several interesting insights into the underlying assumptions of 2SLS.[7]

In sum, the results presented here have important implications for the practice of quantitative political science. For econometric methods to be employed successfully, it is essential that political scientists take a greater interest in the measurement assumptions implicit in their statistical analyses. Our empirical analysis suggests that, for certain nonrecursive models, the choice of measurement assumptions may have a profound influence on the results. Only when researchers develop a better understanding of the error structure of their data will they be able to adjudicate between competing theories using 2SLS.

REFERENCES

Achen, Christopher H. 1975. "Mass Political Attitudes and the Survey Response." *American Political Science Review* 69:1218–31.

Asher, Herbert B. 1974. "Some Consequences of Measurement Error in Survey Data." *American Journal of Political Science* 18:469–85.

Bartels, Larry M. 1989. "Instrumental and Quasi-Instrumental Variables." Presented at the annual Political Methodology Conference, Minneapolis.

Brody, Richard A., and Lawrence S. Rothenberg. 1988. "The Instability of Partisanship: An Analysis of the 1980 Presidential Election." *British Journal of Political Science* 18:445–65.

7. Admittedly, we have examined just one relatively simple simultaneous equations model. Our results are not directly generalizable to more elaborate simultaneous models. When large systems of equations are just identified, expressions for the biased parameter estimates tend to be complex. And when systems of equations are overidentified, unique closed-form expressions for the 2SLS estimates cannot be obtained. Nevertheless, our results and the procedure used to derive them are applicable to the fairly small systems of equations one typically encounters in political science.

Campbell, Donald and Donald Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin* 56:81–105.

Converse, Philip E. and Gregory B. Markus. 1979. "Plus ca Change. . . : The New CPS Election Study Panel." *American Political Science Review* 73:32–49.

Couch, Arthur, and Kenneth Keniston. 1960. "Yeasayers and Naysayers: Agreeing Response set as a Personality Variable." *Journal of Abnormal and Social Psychology* 60:151–74.

Cotter, Patrick R., Jeffrey Cohen, and Philip B. Coulter. 1982. "Race of Interviewer Effects on Telephone Interviews." *Public Opinion Quarterly* 46:278–84.

Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections.* New Haven: Yale University Press.

Green, Donald P. 1988. "On the Dimensionality of Public Sentiment Toward Partisan and Ideological Groups." *American Journal of Political Science* 32:758–80.

Green, Donald P., and Bradley L. Palmquist. 1990. "Of Artifacts and Partisan Instability." *American Journal of Political Science* 34:872–902.

Hatchett, Shirley, and Howard Schuman. 1975. "White Respondents and Race of Interviewer Effects." *Public Opinion Quarterly* 39:523–28.

Jackson, John. 1975. "Issues and Party Alignment." In *The Future of Political Parties,* ed. Louis Maisel and Paul Sacks. Beverly Hills: Sage.

Johnston, J. 1984. *Econometric Methods.* 3d ed. New York: McGraw-Hill.

Markus, Gregory B. 1982. "Political Attitudes during an Election Year: A Report on the 1980 NES Panel Study." *American Political Science Review* 76:538–60.

Markus, Gregory B., and Philip E. Converse. 1979. "A Dynamic Simultaneous Equation Model of Electoral Choice." *American Political Science Review* 73:1055–70.

Page, Benjamin I., and Calvin C. Jones. 1979. "Reciprocal Effects of Policy Preferences, Party Loyalties, and the Vote." *American Political Science Review* 73:1071–89.

Palmquist, Bradley L., and Donald P. Green. 1989. "Assessing Nonrandom Measurement Error using Panel Data and Single-Indicator Models." Presented at the annual meeting of the American Political Science Association, Atlanta.

Werts, Charles E., and Robert L. Linn. 1972. "A Multitrait-Multimethod Model for Studying Growth." *Educational and Psychological Measurement* 32:655–78.

Werts, Charles E., Karl G. Joreskog, and Robert L. Linn. 1976. "Analyzing Ratings with Correlated Intrajudge Measurement Errors." *Educational and Psychological Measurement* 36:319–28.

Wiley, James A., and Mary Glenn Wiley. 1974. "A Note on Correlated Errors in Repeated Measurements." *Sociological Methods and Research* 3:172–88.

APPENDIX

Question Wording and Scoring Conventions

Party Identification. The initial question is, "Generally speaking, do you think of yourself as a Republican, Democrat, Independent, or what?" Those who classify themselves as Republicans or Democrats are asked, "Would you call yourself a strong (Republican/Democrat) or a not very strong (Republican/Democrat)?" Those who classify themselves as Independents are asked, "Do you think of yourself as closer to the Republican party or the Democratic party?" Responses to these questions form a seven-point scale ranging from strong Democrat (coded 0) to strong Republican (coded 6). All but fourteen of the 1,008 people interviewed in January, 1980, provided valid responses; eleven of the fourteen described themselves as "apolitical."

Presidential Performance Evaluations. Respondents are first asked "Do you approve or disapprove of the way Jimmy Carter is handling his job as President?": Those who respond "approve" are asked, "Do you approve strongly or not strongly?" Those who respond "disapprove" are asked, "Do you disapprove strongly or not strongly?" Responses form a four-category scale: strongly approve is coded 1; weak approval, 2; weak disapproval, 4; and strong disapproval, 5. Of the 1,008 respondents interviewed in January, 1980, 81 responded "Don't Know," and another 15 failed to provide an answer.