*Instrumental and "Quasi-Instrumental" Variables**

Larry M. Bartels, *University of Rochester*

The trade-off between the efficiency of an instrumental variable and its exogeneity is widely recognized but little understood. This paper specifies the terms of that trade-off by analyzing the asymptotic mean squared errors associated with the instrumental variables estimator when the instrument may not be perfectly exogenous. The analysis shows that even seemingly minor misspecifications can play havoc with statistical inferences based on "quasi-instrumental variable" estimators. Simple rules of thumb are derived by which intuition can be applied to choices among alternative estimators based on different instrumental variables, or between instrumental variable and ordinary least squares estimators. The theoretical analysis is applied to an example drawn from Jacobson's (1990) and Green and Krasno's (1990) work on congressional campaign spending and is bolstered by Monte Carlo simulations that, for the most part, reproduce the patterns of errors predicted by the asymptotic results.

It is common in textbook treatments of the instrumental variables estimator to remark upon the difficulty of knowing or demonstrating that a potential instrument is itself really exogenous. Thus, Johnston (1972, 280–81) wrote that "the real difficulty in practice of course is actually finding variables to play the role of instruments. The true disturbance is unobservable and so it is difficult to be confident that the instruments really are uncorrelated in the limit with the disturbances." In view of this very real difficulty, it is remarkable that these same textbooks—and the econometric literature more generally—devote little or no attention to the practical consequences of using "instruments" that are really only approximately exogenous.

It is obvious that a "quasi-instrumental variable" estimator—one based on an instrumental variable that is only approximately uncorrelated with the disturbance—will not produce consistent estimates of the underlying parameters of interest. However, it should also be obvious that the use of such a quasi-instrumental variable may often be unavoidable in applied work. Moreover, even when a genuine (perfectly exogenous) instrument is available, it may be so inefficient that a quasi-instrumental variables estimator is preferable in practice (e.g., by a squared error criterion) because data are in short supply.

*This is a revised version of a paper originally presented at the 1989 Political Methodology Conference in Minneapolis and at the 1990 Political Methodology Conference in St. Louis. I am grateful to Christopher Achen, Gary King, Adrian Pagan, Douglas Rivers (twice), and David Weimer for helpful comments and suggestions and to Simon Jackman for assistance in generating the simulations described in section 9 and the figures.

American Journal of Political Science, Vol. 35, No. 3, August 1991, Pp. 777–800

The squared error criterion arises naturally in the context of the IV estimator, since the rationale for using instrumental variables is to purchase consistency at some cost in terms of precision. That trade-off is widely recognized. Kennedy (1985, 115) noted, "Because the typical case is one in which the instrumental variables are not highly correlated with the independent variable with which they are associated . . . the OLS estimator could be preferred on the MSE criterion." Hanushek and Jackson (1977, 238) even provided a Monte Carlo analysis of relative mean squared errors for OLS and IV estimators as a function of sample size and error correlation. But the trade-off is always considered under the maintained assumption that the instrument itself is perfectly exogenous. Typically, in applied work, "the choice of an instrumental variable is highly arbitrary" (Kennedy 1985, 115), and thus the assumption of perfect exogeneity is correspondingly dubious; in that case there is an additional trade-off to be made, between the efficiency of an instrument and its "relative exogeneity."

1. A Sample Squared Error Analysis

Consider the bivariate regression model

$$\mathbf{y} = \mathbf{x}\beta + \mathbf{u} \tag{1}$$

where $\mathbf{y}$ and $\mathbf{x}$ are $n \times 1$ vectors of observable data, $\mathbf{u}$ is an $n \times 1$ vector of unobserved stochastic disturbances, and β is a constant parameter to be estimated.[1]

We want to evaluate the estimator

$$b_n^{\text{IV}} = (\mathbf{z}'\mathbf{x})^{-1}\mathbf{z}'\mathbf{y} \tag{2}$$

where $\mathbf{z}$ is an $n \times 1$ vector of observations on an instrumental variable.[2] The standard textbook assumption is that the instrumental variable $\mathbf{z}$ is uncorrelated in probability with the disturbance $\mathbf{u}$—thus, that $\text{plim}(\mathbf{z}'\mathbf{u}/n) = 0$. I propose to refer to $\mathbf{z}$ as a "quasi-instrumental variable" in situations in which we entertain the possibility of misspecification of the form $\text{plim}(\mathbf{z}'\mathbf{u}/n) \neq 0$.

The squared error for the quasi-instrumental variable estimator is

$$(b_n^{\text{IV}} - \beta)^2 = [(\mathbf{z}'\mathbf{x})^{-1}\mathbf{z}'\mathbf{u}]^2 \tag{3}$$

[1]It is convenient to think of $\mathbf{x}$ and $\mathbf{y}$ as deviations from their respective means, so that no intercept is required in the model in equation (1). The presence of an intercept can be accommodated in the multiple regression framework analyzed in section 6.

[2]The instrumental variable estimator in expression (2) is often described in terms of two distinct regressions—most notably in the context of simultaneous equation models, where it is referred to as the "two-stage least squares" estimator. In the first stage, a regression of the original explanatory variable $\mathbf{x}$ on the instrumental variable $\mathbf{z}$ produces the coefficient $(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'\mathbf{x}$, and thus the fitted values $\mathbf{z}(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'\mathbf{x}$. In the second stage, the original dependent variable $\mathbf{y}$ is regressed on the fitted values from the first stage, producing the parameter estimates defined in expression (2).

Multiplying and dividing the right-hand side of this equation successively by $(\mathbf{x}'\mathbf{x})$, $(\mathbf{z}'\mathbf{z})$, and $(\mathbf{u}'\mathbf{u})$ and rearranging,

$$(b_n^{\mathrm{IV}} - \beta)^2 = [(\mathbf{u}'\mathbf{u})/(\mathbf{x}'\mathbf{x})]\ [R_{zu}^2/R_{xz}^2] \tag{4}$$

where R_{zu}^2 and R_{xz}^2 are the squared correlations between $\mathbf{z}$ and $\mathbf{u}$ and $\mathbf{z}$ and $\mathbf{x}$, respectively. Since the factor in the first brackets on the right does not depend on the choice of an instrument $\mathbf{z}$, it can be treated as a constant of proportionality for the purpose of comparing potential instruments, making the squared error for any given instrument proportional to the factor in the second brackets on the right:

$$(b_n^{\mathrm{IV}} - \beta)^2 \propto [R_{zu}^2/R_{xz}^2] \tag{5}$$

Faced with a choice among alternative quasi-instrumental variables, the squared error criterion leads to the selection of the one for which the right-side ratio in expression (5) is minimized. Obviously, this does not constitute a mechanical selection rule, since the numerator in the right-side ratio is inherently unobservable. (The denominator is, of course, the directly observable squared sample correlation between $\mathbf{z}$ and $\mathbf{x}$). Nevertheless, expression (5) does convey the general nature of the trade-off between efficiency and exogeneity in instrumental variables estimation. Given some intuition about the relative exogeneity of alternative instrumental variables, the expression suggests how to manage that trade-off intelligently.

Notice, however, that sampling variation would make R_{zu}^2 greater than zero, even if $\mathbf{z}$ and $\mathbf{u}$ were uncorrelated in the population; hence, R_{zu}^2 does not measure the "exogeneity" of $\mathbf{z}$ in the usual sense.[3] Since our intuition about population correlations is likely to be stronger than our intuition about sample correlations, it will be helpful to reformulate the error properties of quasi-IV estimators in terms of the relevant population correlations.

2. An Asymptotic Mean Squared Error Analysis

Assume for the model in equation (1) that

$$\begin{array}{ll} \operatorname{plim}(\mathbf{x}'\mathbf{x}/n) = \sigma_{xx} & \operatorname{plim}(\mathbf{z}'\mathbf{z}/n) = \sigma_{zz} \\ \operatorname{plim}(\mathbf{x}'\mathbf{z}/n) = \sigma_{xz} & \operatorname{plim}(\mathbf{z}'\mathbf{u}/n) = \sigma_{zu} \\ \operatorname{plim}(\mathbf{x}'\mathbf{u}/n) = \sigma_{xu} & \operatorname{plim}(\mathbf{u}'\mathbf{u}/n) = \sigma_{uu} \end{array} \tag{6}$$

Again, we want to evaluate the quasi-instrumental variables estimator b_n^{IV}, now using an asymptotic mean squared error criterion.

[3] Interpreted as an estimate of the corresponding population value ρ^2, any sample R^2 is biased upward by approximately $1/n$. In very small samples, it may be prudent to estimate ρ^2 (following Johnson and Kotz 1972, 244) by $[R^2(n - 1) - 1]/(n - 2)$. I ignore that complication here.

The probability limit of the estimator is

$$\begin{aligned} \text{plim}(b_n^{\text{IV}}) &= \text{plim}(\beta) + \text{plim}(\mathbf{z}'\mathbf{x}/n)^{-1}\,\text{plim}(\mathbf{z}'\mathbf{u}/n) \\ &= \beta + \sigma_{zu}/\sigma_{xz} \end{aligned} \tag{7}$$

and its asymptotic variance is[4]

$$\text{asy var}(b_n^{\text{IV}}) = (1/n)\,(\sigma_{uu}\sigma_{zz}/\sigma_{xz}^2) \tag{8}$$

Thus, its asymptotic mean squared error is

$$\begin{aligned} AMSE(b_n^{\text{IV}}) &= [\text{plim}(b_n^{\text{IV}}) - \beta]^2 + \text{asy var}(b_n^{\text{IV}}) \\ &= \sigma_{zu}^2/\sigma_{xz}^2 + (1/n)\,(\sigma_{uu}\sigma_{zz}/\sigma_{xz}^2) \end{aligned} \tag{9}$$

Multiplying and dividing the right side of this equation successively by σ_{xx}, σ_{zz}, and σ_{uu} and rearranging,

$$AMSE(b_n^{\text{IV}}) = [\sigma_{uu}/\sigma_{xx}]\,[\rho_{zu}^2 + 1/n]/\rho_{xz}^2 \tag{10}$$

where ρ_{zu}^2 ($= \sigma_{zu}^2/\sigma_{zz}\sigma_{uu}$) and ρ_{xz}^2 ($= \sigma_{xz}^2/\sigma_{xx}\sigma_{zz}$) are the squared population correlations (i.e., the probability limits of the corresponding squared sample correlations) between **z** and **u** and **z** and **x**, respectively.

Since the ratio in the first brackets on the right side of equation (10) does not depend on our choice of instrument, it will be convenient to write

$$AMSE(b_n^{\text{IV}}) \propto [\rho_{zu}^2 + 1/n]/\rho_{xz}^2 \tag{11}$$

Expression (11) is equivalent to expression (5) except that population correlations appear in place of sample correlations and an additional term $1/n$ appears in the numerator on the right side. Obviously, as the sample becomes large, these differences become correspondingly inconsequential, so that for large samples the asymptotic mean squared error is essentially proportional to the endogeneity of **z** (as measured by the squared correlation between **z** and **u**) and inversely proportional to the efficiency of **z** (as measured by the squared correlation between **z** and **x**).

3. Inferential Consequences of Misspecification

A useful way to gauge the inferential consequences of using quasi-instrumental variables is suggested by rewriting the asymptotic mean squared error of the quasi-IV parameter estimate in expression (10) as

$$AMSE(b_n^{\text{IV}}) = [\sigma_{uu}/n\sigma_{xx}]\,[1/\rho_{xz}^2]\,[1 + n\rho_{zu}^2] \tag{12}$$

[4]In the present instance, the usual formulation for the asymptotic variance of the IV estimator (e.g., Johnston 1972, 280) can be derived from the assumptions that the data are independently identically distributed and that the expectation of u_t^2 conditional upon z_t (and in section 6 upon $\mathbf{X}_{t2}$) is the constant value σ_{uu}. I am grateful to Douglas Rivers for providing the required derivation.

Expression (12) effectively partitions the asymptotic mean squared error of the quasi-IV parameter estimate into its asymptotic variance and its inconsistency (the first and second terms, respectively, in the last square brackets). Since the nominal standard error associated with the instrumental variable parameter estimate will capture only the first of these two components, the actual (asymptotic root mean squared) error will exceed that nominal standard error by a factor of $(1 + n\rho_{zu}^2)^{1/2}$.

The magnitude of the resulting exaggeration of the precision of quasi-instrumental variable parameter estimates for various sample sizes and degrees of endogeneity of the instrumental variable **z** is illustrated in Figure 1. The inferential distortions resulting from even minor violations of the assumption of perfect exogeneity of the instrument are very striking. For example, with a sample size of 250, the actual errors associated with the quasi-IV estimator are almost twice as large as the nominal standard errors when the population correlation of **z** and **u** is .10 ($\rho_{zu}^2 = .01$); with a sample size of 1,000 the same correlation produces actual errors more than 3.3 times as large as the nominal standard errors.[5] Since asymptotic correlations between the instrument and the disturbance as large as .10 are probably common in applied work, it seems likely that many of the statistical inferences based on (quasi-) instrumental variables estimates (at least from relatively large samples) are in fact wildly misleading.

4. Efficiency and Exogeneity

We can further partition the asymptotic mean squared error of the quasi-IV parameter estimate by rewriting expression (12) as

$$AMSE(b_n^{IV}) = [\sigma_{uu}/n\sigma_{xx}]\,[1 + (1 - \rho_{xz}^2)/\rho_{xz}^2 + n\rho_{zu}^2/\rho_{xz}^2] \tag{13}$$

Now the first term in the second square brackets corresponds to the asymptotic variance of the bivariate OLS estimator; the second term represents the additional asymptotic variance produced by using the instrumental variables estimator rather than the OLS estimator; and the third term represents the inconsistency produced by using a quasi-instrumental variable rather than a genuine (perfectly exogenous) instrument. This partitioning indicates that the endogeneity of the instrument will be a significant source of error (relative to error of the textbook sort arising from the inefficiency of the instrument) iff ρ_{zu}^2 is large relative to $(1 - \rho_{xz}^2)/n$. Thus, for even moderately large samples, textbook analyses of the sort described above probably address only a small fraction of the real cost of using instrumental variables in actual applications.

[5]The extent of the exaggerated precision increases with the size of the sample because the nominal standard error of the parameter estimate decreases as the sample gets larger, while the asymptotic bias attributable to misspecification does not. In absolute terms, parameter estimates based on large samples will not be more wrong than those based on small samples, but they will fall further outside the spurious confidence intervals suggested by the nominal standard errors.

Figure 1. Exaggerated Precision of "Quasi-Instrumental Variables" Parameter Estimates as a Function of Endogeneity and Sample Size

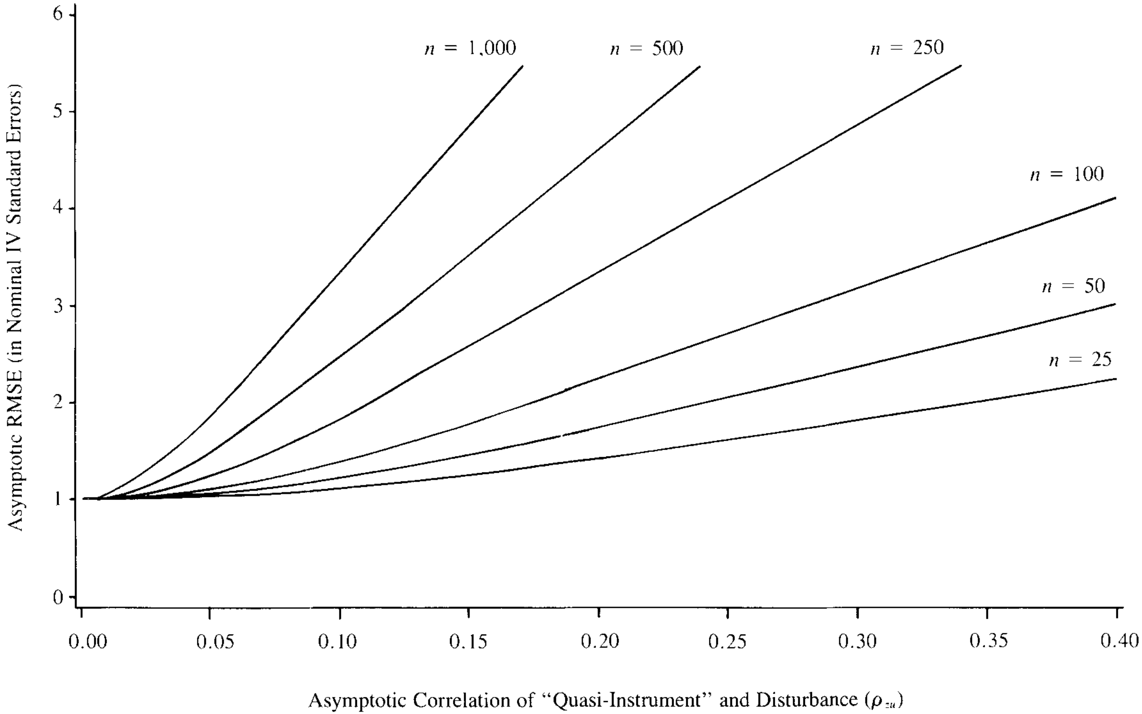


Figure 2 provides a graphical analysis of the errors associated with the quasi-instrumental variables estimator—that is, a graphical representation of the result in expression (11)—for various sample sizes and for various values of ρ^2_{xz} and ρ_{zu}.[6] The lowest curve in each panel of the figure represents the error associated with the instrumental variables estimator under the standard textbook assumption that ρ_{zu} equals zero. The additional curves in each panel show how the situation changes as the quasi-instrumental variable **z** becomes more endogenous (i.e., for increasing values of ρ_{zu} ranging from .10 to .40).

It is evident from Figure 2 that misspecification dominates inefficiency as a source of error over much of the plausibly occupied parameter space. With 250 observations, for example, seemingly trivial endogeneity (ρ_{zu} = .10) is sufficient to make an efficient-looking "quasi-instrument" (one whose purging regression has a population R^2 of .70) no better in reality than a very weak truly exogenous instrument (one whose purging regression has a population R^2 of .20). Of course, there are other circumstances in which efficiency is crucial: with 25 observations the same "quasi-instrument" could be fairly endogenous (ρ_{zu} = .30) and still be clearly preferable to the weak "real" instrument. My argument is not that one strategy is appropriate for every circumstance but that some careful thought along the lines proposed here may help suggest the right strategy for any specific circumstance.

5. OLS as a Quasi-IV Estimator

Whatever other potential instrumental variables may be available, it is obviously always possible to use the endogenous variable **x** as an instrument for itself. The result in expression (10) applies directly with ρ^2_{xx} equal to 1, giving

$$AMSE(b_n) = [\sigma_{uu}/\sigma_{xx}]\ [\rho^2_{xu} + 1/n] \tag{14}$$

The comparison between expressions (10) and (14) provides a basis for choosing in practice between the OLS and quasi-IV estimators. The OLS estimator b_n will be superior to a quasi-IV estimator b_n^{IV} based on the instrumental variable **z**, in spite of the stipulated endogeneity of **x**, iff

$$[\rho^2_{xu} + 1/n] < [\rho^2_{zu} + 1/n]/\rho^2_{xz} \tag{15}$$

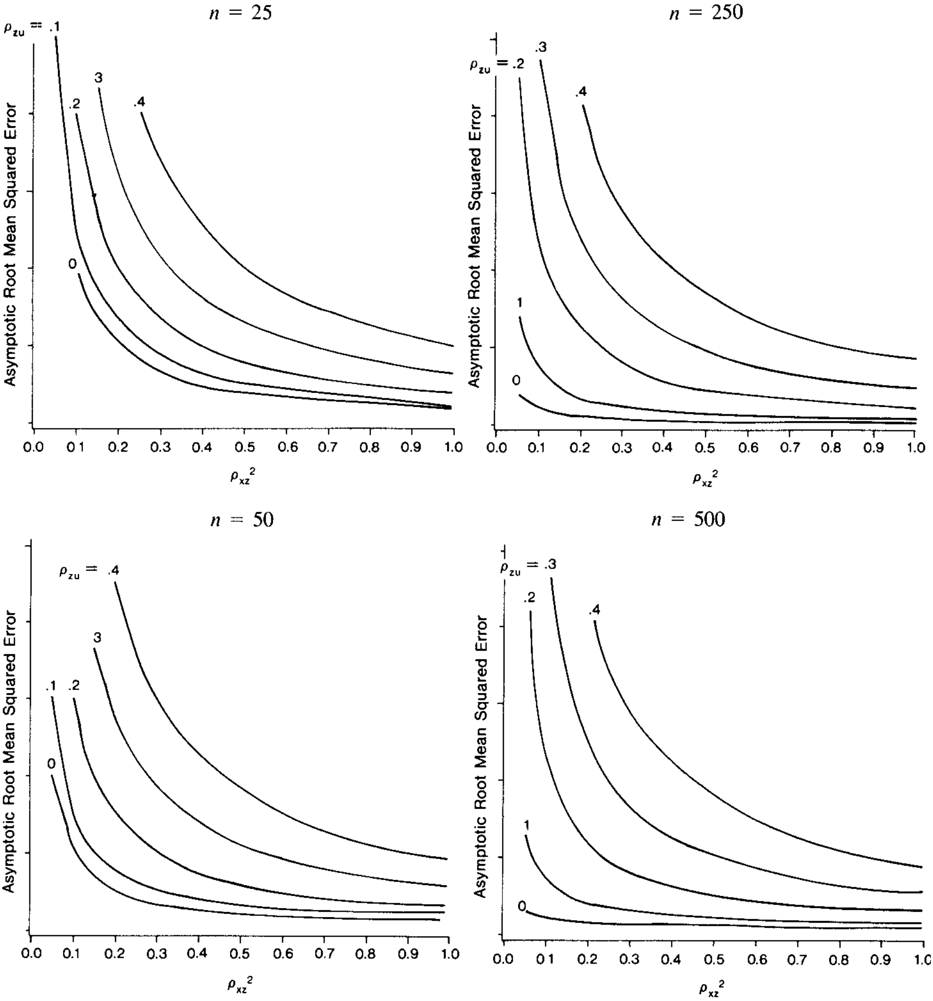
Conversely, the quasi-IV estimator based on the instrumental variable **z** will be superior to the OLS estimator by the asymptotic mean squared error criterion iff

$$\rho^2_{xz} > [\rho^2_{zu} + 1/n]/[\rho^2_{xu} + 1/n] \tag{16}$$

Of course, this comparison involves quantities that are not only unobserv-

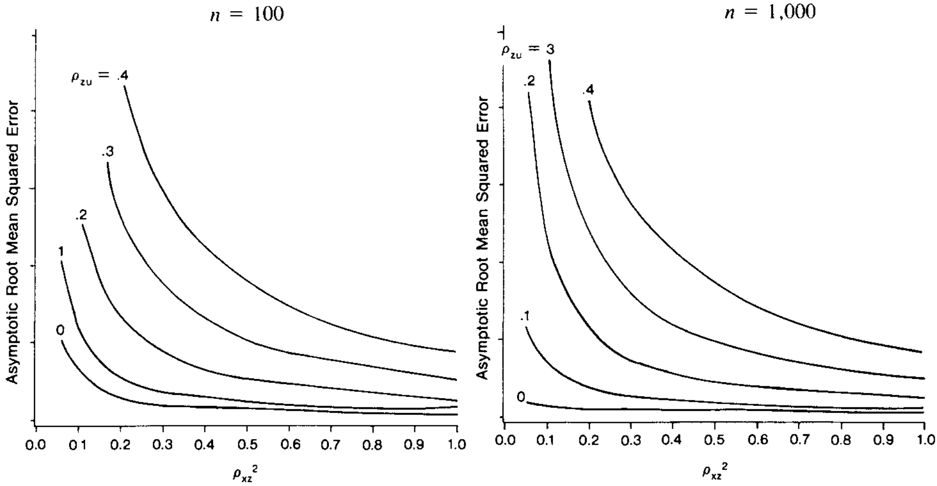
[6]More precisely, the figure shows the asymptotic root mean squared error for each combination of parameter values, ignoring the constant of proportionality in expression (10). Notice that the relationship between **x** and **z** is expressed as a squared correlation, since analysts typically think about (and report) the efficiency of an instrumental variable in terms of an R^2 statistic from a purging regression.

Figure 2. Relative Errors of "Quasi-Instrumental Variables" Parameter Estimates as a Function of Efficiency, Exogeneity, and Sample Size



able but impossible to estimate directly—population correlations that involve the disturbance term **u**. Nevertheless, it is not implausible to suppose that a good data analyst, given a direct sample estimate of the left-side squared correlation in expression (16), might have sufficient intuition about the right-hand side ratio to judge whether the corresponding quasi-IV estimator was or was not likely to improve upon OLS. In any event, such decisions must be made; when they are made on the basis of conscious consideration, they may simply be made more intelligently.

Figure 2. (*continued*)



Note: Vertical axis is proportionate to asymptotic root mean squared error, horizontal axis is ρ_{xz}^2. Lines on the graphs correspond to the level of ρ_{zu}.

6. Extension to Multiple Regression

The expression corresponding to expression (10) for a multiple regression model with endogenous explanatory variable $\mathbf{x}_1$ and additional (exogenous) explanatory variables $\mathbf{X}_2$ is a straightforward generalization. The model is

$$\mathbf{y} = \mathbf{x}_1\beta_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u} \tag{17}$$

and the (quasi-) IV estimator for β_1 is

$$b_{1n}^{\mathrm{IV}} = (\mathbf{z}'\mathbf{M}_2\mathbf{x}_1)^{-1}\,\mathbf{z}'\mathbf{M}_2\mathbf{y} \tag{18}$$

where $\mathbf{M}_2 = \mathbf{I} - \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'$ is the "least squares residualizing matrix" based on the exogenous explanatory variables in $\mathbf{X}_2$.[7] We add to the assumptions in expression (6) that

$$\begin{aligned}
\operatorname{plim}(\mathbf{X}_2'\mathbf{X}_2/n) &= \boldsymbol{\Sigma}_{22} \quad \text{(a positive definite matrix)}\\
\operatorname{plim}(\mathbf{X}_2'\mathbf{x}_1/n) &= \boldsymbol{\sigma}_{2x}\\
\operatorname{plim}(\mathbf{X}_2'\mathbf{z}/n) &= \boldsymbol{\sigma}_{2z}\\
\operatorname{plim}(\mathbf{X}_2'\mathbf{u}/n) &= \mathbf{0} \quad \text{(since } \mathbf{X}_2 \text{ is exogenous)}
\end{aligned} \tag{19}$$

[7] The terminology reflects the fact that multiplying any vector by $\mathbf{M}_2$ produces the least squares residual vector from a regression of the original vector on $\mathbf{X}_2$. For example, $\mathbf{z}'\mathbf{M}_2$ is the (transposed) least squares residual vector obtained by regressing $\mathbf{z}$ on $\mathbf{X}_2$, and $\mathbf{M}_2\mathbf{X}_1$ is the least squares residual vector obtained by regressing $\mathbf{x}_1$ on $\mathbf{X}_2$. The analysis below makes use of the facts that $\mathbf{M}_2' = \mathbf{M}_2$, $\mathbf{M}_2\mathbf{M}_2 = \mathbf{M}_2$, and $\mathbf{M}_2\mathbf{X}_2 = \mathbf{0}$, which can be verified from the definition of $\mathbf{M}_2$ in the text by performing the required multiplications.

Then the probability limit of the quasi-IV estimator in expression (18) is

$$\begin{aligned}
\text{plim}(b_{1n}^{\text{IV}}) &= \beta_1 + \text{plim}(\mathbf{z}'\mathbf{M}_2\mathbf{x}_1/n)^{-1}\,\text{plim}(\mathbf{z}'\mathbf{M}_2\mathbf{u}/n) \\
&\qquad (\text{since } \mathbf{M}_2\mathbf{X}_2 = \mathbf{0} \text{ exactly}) \\
&= \beta_1 + \text{plim}(\mathbf{z}'\mathbf{M}_2\mathbf{x}_1/n)^{-1}\,\text{plim}(\mathbf{z}'\mathbf{u}/n) + 0 \\
&\qquad (\text{since plim}(\mathbf{X}_2'\mathbf{u}/n) = \mathbf{0} \text{ by assumption}) \\
&= \beta_1 + \sigma_{zu}/(\sigma_{xz} - \boldsymbol{\sigma}_{2x}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2z})
\end{aligned} \tag{20}$$

its asymptotic variance is

$$\text{asy var}(b_{1n}^{\text{IV}}) = (1/n)\,\sigma_{uu}(\sigma_{zz} - \boldsymbol{\sigma}_{2z}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2z})/(\sigma_{xz} - \boldsymbol{\sigma}_{2x}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2z})^2 \tag{21}$$

and its asymptotic mean squared error is

$$\begin{aligned}
AMSE(b_{1n}^{\text{IV}}) &= \sigma_{zu}^2/(\sigma_{xz} - \boldsymbol{\sigma}_{2x}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2z})^2 + \\
&\quad (1/n)\,\sigma_{uu} \times (\sigma_{zz} - \boldsymbol{\sigma}_{2z}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2z})/(\sigma_{xz} - \boldsymbol{\sigma}_{2x}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2z})^2
\end{aligned} \tag{22}$$

Multiplying and dividing the right-hand side of expression (22) successively by $(\sigma_{xx} - \boldsymbol{\sigma}_{2x}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2x})$, $(\sigma_{zz} - \boldsymbol{\sigma}_{2z}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2z})$, and σ_{uu} and rearranging,

$$\begin{aligned}
AMSE\,(b_{1n}^{\text{IV}}) &= [\sigma_{uu}/(\sigma_{xx} - \boldsymbol{\sigma}_{2x}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2x})] \\
&\quad [\sigma_{zu}^2/\sigma_{uu}(\sigma_{zz} - \boldsymbol{\sigma}_{2z}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2z}) + 1/n] \\
&\quad [(\sigma_{xx} - \boldsymbol{\sigma}_{2x}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2x})\,(\sigma_{zz} - \boldsymbol{\sigma}_{2z}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2z})/(\sigma_{xz} - \boldsymbol{\sigma}_{2x}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2z})^2] \\
&= [\sigma_{uu}/(\sigma_{xx} - \boldsymbol{\sigma}_{2x}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2x})]\,[\rho_{zu|2}^2 + 1/n]/\rho_{xz|2}^2
\end{aligned} \tag{23}$$

where $\rho_{zu|2}$ and $\rho_{xz|2}$ are the population correlations between $\mathbf{z}$ and $\mathbf{u}$ and $\mathbf{z}$ and $\mathbf{x}_1$, respectively, *holding constant* $\mathbf{X}_2$.

Since the ratio in the first square brackets does not vary with our choice of instrument,[8] we can write

$$AMSE(b_{1n}^{\text{IV}}) \propto [\rho_{zu|2}^2 + 1/n]/\rho_{xz|2}^2 \tag{24}$$

which is identical to expression (11) except that the relevant quantities are now partial correlations holding constant the exogenous variables $\mathbf{X}_2$. Thus, all of the analysis in sections 3, 4, and 5 (including the graphical analysis in Figures 1 and 2) applies to the multiple regression model as well, if direct correlations are replaced by partial correlations holding constant $\mathbf{X}_2$ (i.e., if $\rho_{xz|2}^2$, $\rho_{zu|2}^2$, and $\rho_{xu|2}^2$ are substituted for ρ_{xz}^2, ρ_{zu}^2, and ρ_{xu}^2, respectively).

Some intuition regarding the likely magnitudes of the partial correlation in the numerator of expression (24) may follow from noting that

$$\rho_{zu|2}^2 = \rho_{zu}^2/(1 - \rho_{z2}^2) \tag{25}$$

[8]The constant of proportionality in expression (22) will be greater by a factor of $1/(1 - \rho_{x2}^2)$ than the corresponding constant in expression (10); the difference may be thought of as a penalty for collinearity in the multiple regression model.

where ρ_{z2} is the (directly estimable) population multiple correlation between the instrument and the exogenous explanatory variables. Thus, it is possible to translate judgments about the direct correlation between **z** and **u** straightforwardly into judgments about the partial correlation in expression (24) between **z** and **u** holding constant $\mathbf{X}_2$.

Similarly, in thinking about the partial correlation in the denominator of expression (24), it may be enlightening to recall (from Theil 1971, 175) that

$$\rho^2_{xz|2} = (\rho^2_{xp} - \rho^2_{x2})/(1 - \rho^2_{x2}) \tag{26}$$

which is simply the population analog of the proportion of previously unexplained variance explained by adding **z** to the regression of $\mathbf{x}_1$ on $\mathbf{X}_2$. (I use ρ^2_{xp} to denote the squared multiple population correlation from the purging regression of $\mathbf{x}_1$ on **z** and $\mathbf{X}_2$, and ρ^2_{x2} is the corresponding squared multiple population correlation from a purging regression of $\mathbf{x}_1$ on $\mathbf{X}_2$ only.)

It is common in textbook treatments of the instrumental variables estimator to note that "the higher the correlation between Z_k and X_k (so long as Z_k remains uncorrelated with U), the lower the variance of the estimated coefficients" (Hanushek and Jackson 1977, 234). While that statement is true, the appearance of $\rho^2_{xz|2}$ rather than ρ^2_{xz} in the denominator of expression (24) makes it clear that the statement is true in the multiple regression framework only to the extent that the higher correlation is with that portion of $\mathbf{x}_1$ uncorrelated with $\mathbf{X}_2$. Certainly the implication that an instrument strongly correlated with $\mathbf{x}_1$ is superior to one less strongly correlated with $\mathbf{x}_1$ can be a misleading rule of thumb for evaluating alternative instruments, since **z** may be a relatively inefficient instrument even if it is strongly correlated with $\mathbf{x}_1$, if it happens to be correlated with that portion of $\mathbf{x}_1$ that is collinear with $\mathbf{X}_2$. (In that case it will add little to a purging regression in which $\mathbf{X}_2$ already appears, making the right-hand side ratio in expression (26)—and hence the denominator in expression (24)—relatively small.)

So far I have examined only the error associated with the parameter for the single endogenous variable $\mathbf{x}_1$ in the multiple regression model. Might not some estimator do poorly with respect to this one parameter but still be preferable because it produces better estimates of the other parameters in the model? In a word, no. We can see this by writing the vector of IV estimates for the other parameters (using a standard result on inverses of partitioned matrices from Theil 1971, 18) as[9]

$$\begin{aligned}
\mathbf{b}^{\mathrm{IV}}_{2n} &= (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{y} - (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{x}_1\,(\mathbf{z}'\mathbf{M}_2\mathbf{x}_1)^{-1}\mathbf{z}'\mathbf{M}_2\mathbf{y} \\
&= (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{y} - (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{x}_1\, b^{\mathrm{IV}}_{1n} \qquad (27) \\
&= \boldsymbol{\beta}_2 + (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{u} - (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{x}_1\,(b^{\mathrm{IV}}_{1n} - \beta_1)
\end{aligned}$$

[9]The second line of expression (27) follows from a direct substitution of expression (18); the third line from a substitution of expression (17) and rearrangement.

The squared inconsistencies in these parameter estimates are the main diagonal elements of

$$[\text{plim}(\mathbf{b}_{2n}^{IV}) - \boldsymbol{\beta}_2][\text{plim}(\mathbf{b}_{2n}^{IV}) - \boldsymbol{\beta}_2]' = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2x}\boldsymbol{\sigma}_{2x}'\boldsymbol{\Sigma}_{22}^{-1} \times [\text{plim}(b_{1n}^{IV}) - \beta_1]^2 \tag{28}$$

(because plim $(\mathbf{X}_2'\mathbf{u}/n) = \mathbf{0}$ by assumption). The asymptotic variances are the main diagonal elements of

$$\text{asy var}(\mathbf{b}_{2n}^{IV}) = (1/n)\,\sigma_{uu}\,\boldsymbol{\Sigma}_{22}^{-1} + \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2x}\boldsymbol{\sigma}_{2x}'\boldsymbol{\Sigma}_{22}^{-1}\,[\text{asy var}(b_{1n}^{IV})] \tag{29}$$

Thus, the asymptotic mean squared errors for the parameter vector $\mathbf{b}_{2n}^{IV}$ are the main diagonal elements of

$$AMSE(\mathbf{b}_{2n}^{IV}) = (1/n)\,\sigma_{uu}\,\boldsymbol{\Sigma}_{22}^{-1} + \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2x}\boldsymbol{\sigma}_{2x}'\boldsymbol{\Sigma}_{22}^{-1}\,[AMSE(b_{1n}^{IV})] \tag{30}$$

The first term of expression (30) is invariant with respect to **z**, and the second term has main diagonal elements that are increasing functions of the asymptotic mean squared error of b_{1n}^{IV} (since the main diagonal elements of the matrix $\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{2x}\boldsymbol{\sigma}_{2x}'\boldsymbol{\Sigma}_{22}^{-1}$ must be nonnegative). It follows that whatever quasi-instrument seems likely to produce the best estimate of β_1 will be equally likely to produce the best estimate of each element of $\boldsymbol{\beta}_2$ as well.

7. Two-Stage Least Squares

Suppose we regress the endogenous variable $\mathbf{x}_1$ on the exogenous variables $\mathbf{X}_2$ plus some additional variables $\mathbf{X}_0$ excluded from the equation of interest and use the vector of fitted values

$$\hat{\mathbf{x}}_1 = \mathbf{M}_2\mathbf{X}_0(\mathbf{X}_0'\mathbf{M}_2\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{M}_2\mathbf{x}_1 + \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{x}_1 \tag{31}$$

as our instrumental variable in the original equation. The resulting IV estimator is the familiar two-stage least squares estimator.

Just as textbooks cling to the assumption that instrumental variables are perfectly exogenous, they cling to the analogous assumption that excluded variables used to construct an instrument in the 2SLS framework are perfectly exogenous. But it is evident that, when this assumption is false, the 2SLS estimator may be inferior to other quasi-IV estimators, including (possibly) the OLS estimator. Given the highly optimistic nature of the exclusion restrictions in much applied work, this is more than a theoretical possibility.[10]

The form of the 2SLS estimator for β_1 is

$$\begin{aligned} b_{1n}^{2SLS} &= (\hat{\mathbf{x}}_1'\mathbf{M}_2\mathbf{x}_1)^{-1}\hat{\mathbf{x}}_1'\mathbf{M}_2\mathbf{y} \\ &= [\mathbf{x}_1'\mathbf{M}_2\mathbf{X}_0(\mathbf{X}_0'\mathbf{M}_2\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{M}_2\mathbf{x}_1]^{-1} \\ &\quad \times \mathbf{x}_1'\mathbf{M}_2\mathbf{X}_0(\mathbf{X}_0'\mathbf{M}_2\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{M}_2\mathbf{y} \end{aligned} \tag{32}$$

[10]Comparisons of the sort considered here among misspecified simultaneous-equations models

which is equivalent to the IV estimator b_{1n}^{IV} of expression (18) based on the instrumental variable $\mathbf{z} = \mathbf{X}_0(\mathbf{X}_0'\mathbf{M}_2\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{M}_2\mathbf{x}_1$. Thus, if there is only one excluded variable in the simultaneous-equations model, the effective instrument $\mathbf{z}$ is proportional to it; in that case the analysis in section 6 applies directly with $\mathbf{x}_0$ in place of $\mathbf{z}$. If there are two or more excluded variables, the effective instrument $\mathbf{z}$ is a weighted average of the excluded variables, with the weight for each being the corresponding parameter estimate from the purging regression. In either case, the apparent complication introduced by including the right-hand side exogenous variables $\mathbf{X}_2$ in the purging regression turns out to be illusory, since the relevant characteristics of $\mathbf{z}$ in expression (23) are its partial correlations with $\mathbf{x}_1$ and with $\mathbf{u}$ *holding constant* $\mathbf{x}_2$.

8. Empirical Illustration

How might we apply the results of the preceding theoretical analysis to the real problems faced by practicing data analysts? In this section I illustrate the implications of my analysis for one ongoing empirical controversy. The question at issue is how campaign spending by incumbent representatives affects congressional election outcomes.

The pioneering work on congressional campaign spending by Jacobson (1980) seemed to demonstrate that spending by congressional challengers has a strong impact on election outcomes but that spending by incumbents has little or no effect. This striking asymmetry—and the associated puzzle of why incumbents go to such trouble to raise and spend money if doing so makes no appreciable difference to their electoral prospects—prompted revisionist analyses by Green and Krasno (1988) and others. Green and Krasno's analysis is of particular interest here because one of their chief departures from Jacobson's original model involved the addition of a new instrumental variable to the set utilized by Jacobson. Subsequent debate (Jacobson 1990; Green and Krasno 1990) has focused in significant part on the appropriateness of this additional instrumental variable—precisely the sort of question to which the theoretical analysis presented here should be able to contribute some insight.

Green and Krasno proposed treating incumbent spending in the current election cycle as an endogenous explanatory variable in a model that accounts for the challenger's vote percentage. They further proposed using an incumbent's spending in the prior election cycle as an instrument for spending in the current cycle; they argued that the prior incumbent spending level reflects an incumbent's "propensity to spend money given his or her abilities and tastes for fundraising" (1988, 897). This specification produced estimated effects of incumbent spending much larger than those produced by Jacobson.

Jacobson criticized Green and Krasno's use of prior incumbent spending as

---

were pursued in a somewhat more general framework, albeit from the standpoint of an asymptotic bias criterion only, by Bartels (1985).

an instrument for current incumbent spending on several grounds (1990, 337), most notably because prior incumbent spending is strongly correlated with the other explanatory variables in the challenger vote model. Green and Krasno dismissed this aspect of Jacobson's criticism as "simply rhetorical fluff, as collinearity biases neither the TSLS estimates nor the standard errors" (1990, 364). If our only goal is to draw unbiased inferences from the data, then Green and Krasno were right to dismiss collinearity as a significant issue. However, if we also care about drawing the most *accurate* possible inferences from the data, then the theoretical analysis in section 6 above should make clear that Jacobson's criticism is quite relevant.[11]

Early versions of Jacobson's model (1980, 1985) included incumbent seniority and dummy variables for primary competition in each party as potential instrumental variables. More recently, Jacobson has expressed substantial skepticism about the appropriateness of these and other readily available instrumental variables, arguing that the findings produced by his own and other similar analyses "are by no means conclusive because it is not clear that any of the models is really identified" (1990, 341). Thus, having criticized Green and Krasno's proposed instrumental variable as well, he was left to conclude that "TSLS models, where interpretable, merely repeat the ordinary least squares findings (Jacobson 1985), implying that simultaneity bias is small and that the OLS model is adequate after all" (1990, 341).[12]

The analysis presented here is a somewhat stylized representation of this controversy. I use Jacobson's data from one election year, 1986, to illustrate the implications of the theoretical results reported above for two alternative models, one treating incumbent spending as exogenous (in more or less the way advocated by Jacobson 1990) and estimated using ordinary least squares, and the other treating incumbent spending as endogenous (in more or less the way advocated by Green and Krasno 1990) and estimated using two-stage least squares.[13]

The first column of Table 1 shows the results of the two-stage least squares purging regression of incumbent spending on prior incumbent spending and the

[11] Curiously, Green and Krasno earlier seemed to recognize this sort of criticism as more than "rhetorical fluff," having criticized Jacobson's original instrumental variables on similar grounds (1988, 896).

[12] Having concluded that "simultaneity bias is small," Jacobson went on to treat both incumbent spending and challenger spending as exogenous in a logit analysis of individual vote choices using panel data (1990, 342–56). I share Green and Krasno's (1990, 370–71) skepticism regarding the adequacy of this new analysis.

[13] I am grateful to Stephen Ansolabahere—and indirectly to Gary Jacobson—for providing these data. For the sake of simplicity, I follow both Jacobson (1990) and Green and Krasno (1990) in treating challenger spending as exogenous, though both had earlier made strong arguments for its endogeneity (Jacobson 1980, 1985; Green and Krasno 1988). Similarly, I follow the lead of the principals in excluding open seats, first-term incumbents, and districts without major party challengers in the current or immediately preceding election cycle, as well as Jack Kemp (R-NY) and Claude

Table 1. Purging and Auxiliary Regressions for Incumbent Spending in 1986 Congressional Elections

	Purging Regression (Including Prior Incumbent Spending)	Auxiliary Regression (Excluding Prior Incumbent Spending)
Intercept	2.523	9.760
	(.280)[a]	(.277)
Republican challenger	.0363	−.0460
	(.0494)	(.0678)
Challenger party strength, 1984	−.00575	.01981
	(.00369)	(.00452)
Challenger prior office	.0724	.2001
	(.0676)	(.0927)
Challenger spending	.1472	.1455
	(.0229)	(.0317)
Incumbent spending, 1984	.6816	0
	(.0455)	(—)
R^2	.6428	.3140
Standard error of regression	.366	.507
Number of observations	250	250

[a] Standard errors of parameter estimates are in parentheses.

exogenous explanatory variables. For purposes of comparison, the second column of Table 1 shows the results of the auxiliary regression of incumbent spending on the exogenous explanatory variables only. (The R^2 statistic from this auxiliary regression is a sample estimate of the squared population correlation ρ^2_{x2}, which appears in expression 26 above.) It is clear from these results that current incumbent spending is strongly related to prior incumbent spending and that the use of prior incumbent spending as an exogenous variable dramatically improves the goodness of fit of the purging regression.

The implications of these results for the choice between two-stage least squares and ordinary least squares as an estimator for the challenger vote model are laid out in Table 2. For each of the alternative estimators, Table 2 shows the observed sample correlations corresponding to the population correlations appearing in the expression for the asymptotic mean squared error of the incumbent spending parameter estimate (expression 24 above). These include the R^2 statistics from the purging and auxiliary regressions in Table 1, the R^2 statistic from

Pepper (D-FL), who spent heavily in 1986 in pursuit of national rather than local ambitions. These exclusions leave 250 observations. (Jacobson's 255 observations apparently included five first-term incumbents due to miscoding.)

Table 2. Observed Correlations between Alternative Instruments and Explanatory Variables in 1986 Congressional Elections

Exp.	2SLS	OLS
R^2_{x2} ($\mathbf{x}_1$ on $\mathbf{X}_2$)	.3140	.3140
R^2_{xp} ($\mathbf{x}_1$ on $\mathbf{X}_2$ and $\mathbf{z}$)	.6428	1.0000
R^2_{z2} ($\mathbf{z}$ on $\mathbf{X}_2$)	.4885	.3140
(26) $R^2_{xz\mid 2} = (R^2_{xp} - R^2_{x2}) \,/\, (1 - R^2_{x2})$	.4793	1.0000
(25) $R^2_{zu\mid 2} = R^2_{zu} \,/\, (1 - R^2_{z2})$	1.955 R^2_{zu}	1.458 R^2_{zu}
(24) $\dfrac{[R^2_{zu\mid 2} + 1/n]}{R^2_{xz\mid 2}}$	4.079 R^2_{zu} +.0083	1.458 R^2_{zu} +.0040

an auxiliary regression of the 2SLS instrumental variable **z** (derived from the purging regression in Table 1) on the exogenous explanatory variables in the challenger vote model, and the partial correlations derived from these observed correlations via expressions (25) and (26).

By expression (24) the asymptotic mean squared error of the parameter estimate for incumbent spending in each specification is proportional to the population equivalent of the corresponding expression in the last row of Table 2. Which of these two values is likely to be smaller? Here we must draw upon some intuition regarding the likely correlation between each of the alternative instrumental variables and the unobserved disturbance term in the challenger vote equation. Ignoring the terms that do not depend on R^2_{zu} (in each case these will be relatively inconsequential if R^2_{zu} exceeds .01 or so), the estimated asymptotic mean squared error from the OLS estimation will exceed that from the 2SLS estimation when the correlation between the disturbance in the challenger vote equation and current incumbent spending exceeds the correlation between the disturbance in the challenger vote equation and the instrument based on the purging regression in Table 1 by about 67%.

Given the imperfections of the available measures of incumbents' vulnerability and challengers' electoral potential—challenger party strength is measured by the previous challenger's vote in the district and challenger quality by a dichotomous variable for prior experience in elective office[14]—it seems likely that incumbents and their potential contributors have considerably more information about their true electoral prospects than is captured in the challenger vote model. If contributions are heavily conditioned on this information (and the work of Jacobson and others suggests that they are), then it seems likely that the en-

[14] I do not pursue here another innovation in Green and Krasno's (1988) analysis, the use of a more elaborate eight-point challenger quality scale.

Table 3. OLS and 2SLS Parameter Estimates for Challenger Vote Percentage in 1986 Congressional Elections

	OLS Estimates	2SLS Estimates
Intercept	−11.82	4.47
	(6.14)[a]	(8.63)
Republican challenger	−4.63	−4.70
	(.61)	(.62)
Challenger party strength, 1984	.485	.518
	(.042)	(.045)
Challenger prior office	2.00	2.34
	(.84)	(.87)
Challenger spending	3.057	3.300
	(.298)	(.315)
Incumbent spending	−.185	−1.855[b]
	(.575)	(.845)
R^2	.724	.714
Standard error of regression	4.56	4.64
Number of observations	250	250

[a]Standard errors of parameter estimates are in parentheses.

[b]Endogenous explanatory variable replaced by instrument based on purging regression in Table 1 (including prior incumbent spending).

dogeneity of current incumbent spending is fairly severe. By contrast, prior incumbent spending seems relatively immune to this endogeneity, since the prior challenger's realized vote already appears in the model as an explanatory variable.[15] Thus, on balance, it seems likely that the relative exogeneity of the instrumental variable constructed from the purging regression in Table 1 is sufficient to warrant its cost in terms of inefficiency, relative to the readily available alternative of estimating the challenger vote model using ordinary least squares.

The implication of this judgment for the estimated impact of incumbent spending on congressional election outcomes is illustrated in Table 3. The OLS parameter estimates in the first column of Table 3 essentially recapitulate Jacobson's story: each one-unit increase in challenger spending (e.g., from \$30,000 to about \$90,000 or from \$100,000 to about \$280,000) apparently increased the

[15]Some evidence for the relative exogeneity of prior incumbent spending is provided by the fact that prior incumbent spending contributes little or nothing to an auxiliary regression of current challenger spending on the other exogenous variables. It seems implausible that prior incumbent spending could be strongly related to unmeasured aspects of the district's political situation, yet entirely unrelated to factors recognized by the current challenger and his or her potential contributors.

challenger's vote by three percentage points, while incumbent spending had virtually no effect. The 2SLS parameter estimates in the second column of Table 3 tell a quite different story: the estimated effect of challenger spending increases slightly, while the estimated effect of incumbent spending increases by an order of magnitude, to a level somewhat more than half that of challenger spending.

9. Monte Carlo Analysis

The analysis in sections 2 through 7 addresses the asymptotic behavior of the instrumental variables estimator under misspecification. By abstracting from the vagaries of sampling variation, the analysis highlights the inherent inferential cost of misspecification (characterized by nonzero population correlations between quasi-instrumental variables and disturbances), even in an ideal world where the data at hand are always nicely representative of the underlying population from which they are sampled. Nevertheless, it is important as well to have some idea of how the vagaries of real sampling variation may modify, magnify, or obscure this inherent cost, especially in very small samples.

The Monte Carlo simulation is a standard tool for analyzing the behavior of statistical estimators in finite samples.[16] Given a well-specified statistical model, a random number generator can produce a large number of artificial data sets that mimic the posited statistical process. The distribution of parameter estimates produced by applying the statistical estimator of interest to these artificial data sets then provides some insight into the range of likely outcomes that might be expected when the same estimator is applied to a single real data set generated by the statistical process represented in the simulation.

The crucial characteristics of the statistical models in the theoretical analyses above are the endogeneity of a proposed instrumental variable (represented by the population correlation between the quasi-instrument and the disturbance in the regression equation of interest), its efficiency (represented by the population correlation between the quasi-instrument and the original endogenous variable), and the size of the sample on which the parameter estimates are based. (The inferential implications of these three characteristics are highlighted in Figure 2 above.)

For purposes of illustration, the simulations reported here take as given the exogenous variables in Jacobson's 1986 congressional spending data and the 2SLS parameter estimates in Tables 1 and 3 above. The simulated endogeneity of the quasi-instrumental variable and the sample size are both varied systematically, with 500 artificial data sets generated from each of the 30 combinations of six predetermined sample sizes (N = 25, 50, 100, 250, 500, and 1,000) and

[16] Hanushek and Jackson (1977, 60–65ff.) provide a useful description and applications of the Monte Carlo technique.

Figure 3. Monte Carlo and Asymptotic Results

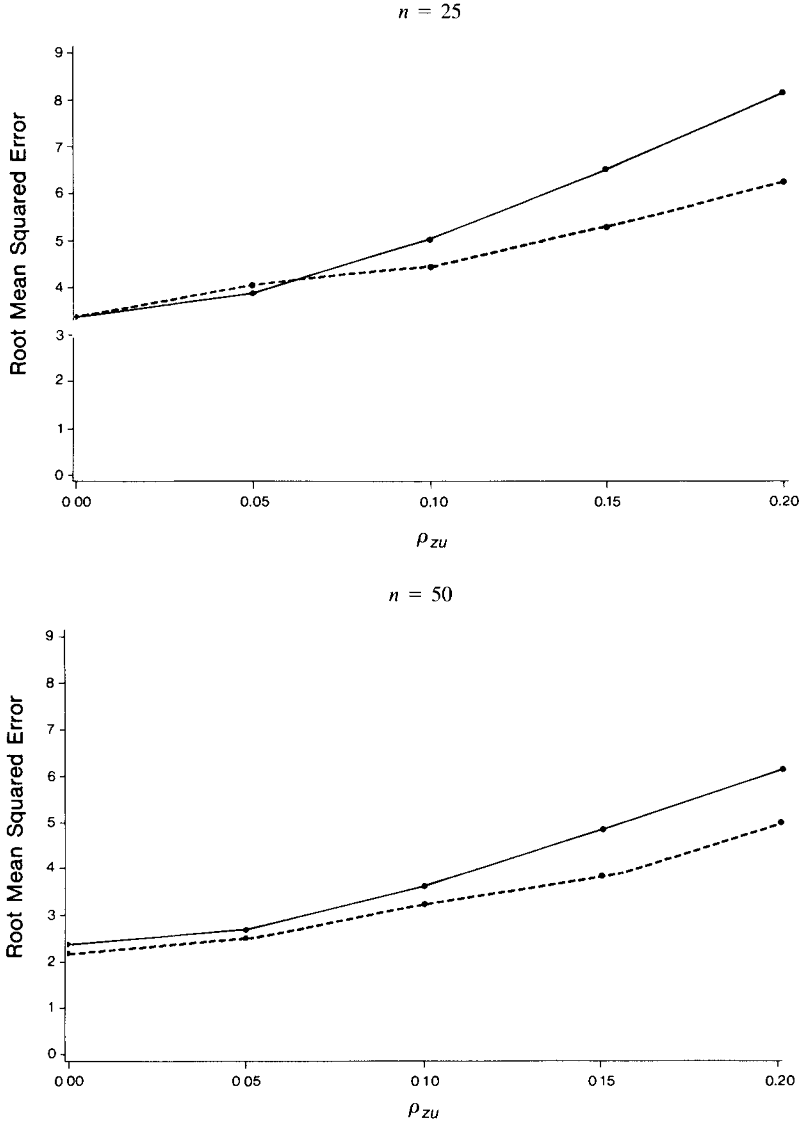
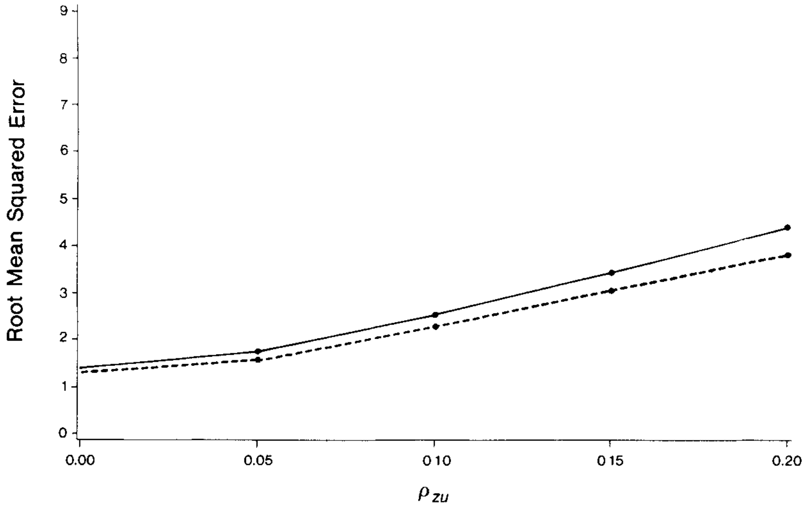


Figure 3. (*continued*)

$n = 100$



$n = 250$

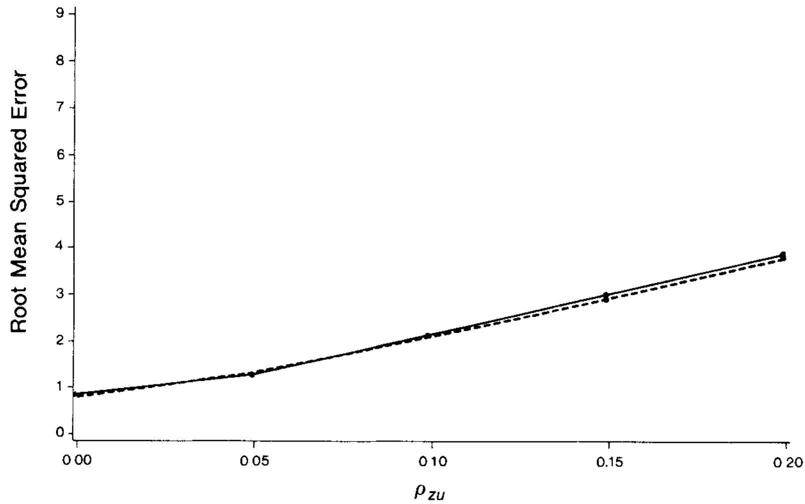
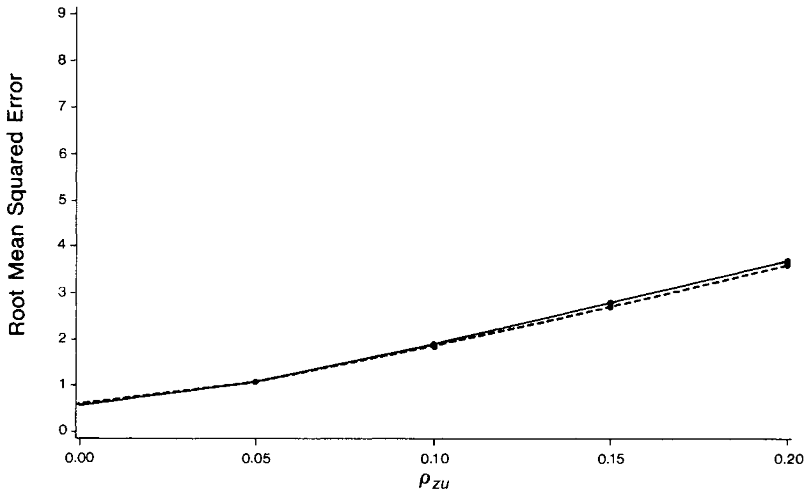
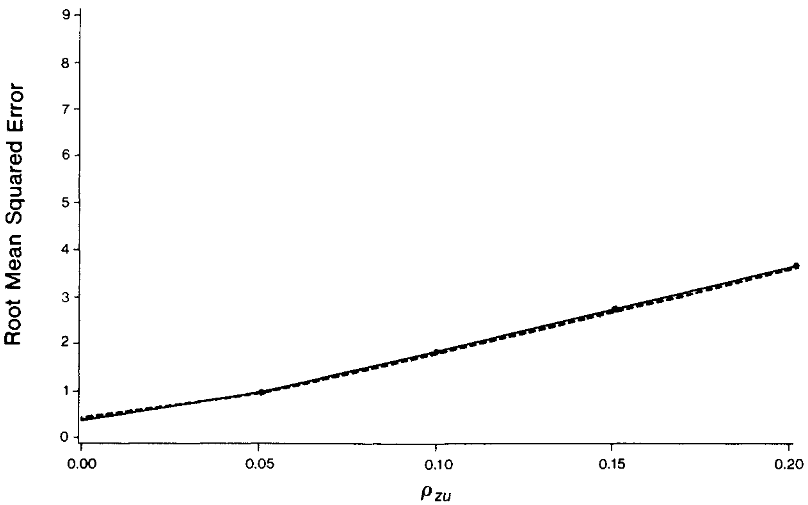


Figure 3. (*continued*)

$n = 500$



$n = 1{,}000$



Note: ρ_{zu} on horizontal axis; root mean squared error on vertical axis; 500 simulations. Dotted lines are Monte Carlo results; continuous lines are asymptotic results.

five predetermined levels of endogeneity of the instrumental variable based on the purging regression in Table 1 (ρ_{zu} = 0, .05, .10, .15, and .20).[17]

Figure 3 provides a graphic comparison of the observed root mean squared errors of the 500 parameter estimates for incumbent spending for each sample size and level of endogeneity (the dotted lines in the figure) and the theoretically expected (asymptotic) root mean squared errors derived for the 1986 Jacobson data from the results in section 6 (the solid lines in the figure). For the samples of 250, 500, and 1,000 observations the match between the observed errors and the theoretically expected (asymptotic) errors is virtually exact. For smaller sample sizes, the observed errors are consistently smaller than the asymptotic results would lead us to expect, with discrepancies as large as 20% for the smallest samples and most endogenous instruments.

Table 4 compares the means and standard deviations of the observed distributions of parameter estimates for the 30 simulations and the corresponding probability limits and asymptotic standard deviations. It is evident from the table that the discrepancies between the observed and asymptotic errors for the smaller sample sizes in Figure 3 are attributable to significantly smaller biases in the simulated data than the asymptotic theory suggests. The observed biases are about 25% smaller than the asymptotic biases for the simulation with 25 observations, 20% smaller for the simulation with 50 observations, and 15% smaller for the simulation with 100 observations.

These discrepancies suggest that the asymptotic analysis in sections 2 through 7 should be applied with some caution to very small samples.[18] Nevertheless, the overall pattern of observed errors across sample sizes and levels of endogeneity is sufficiently similar to the pattern predicted by the asymptotic analysis to suggest that the theoretical results derived above do constitute useful rules of thumb for working data analysts concerned with the actual behavior in finite samples of the quasi-instrumental variable estimator.

10. Conclusions

The preceding analysis has at least three practical implications for data analysts who employ instrumental variables estimators (and for consumers of data analysis based on such estimators).

First, the crucial role of the squared population correlation $\rho^2_{xz|2}$ in the analysis in section 6 suggests that the corresponding squared sample correlation $R^2_{xz|2}$

[17] At the maximum level of simulated endogeneity, ρ_{zu} = .20, the implied independent effect of prior incumbent spending on current votes is 2.5, an effect about halfway between those attributed to current incumbent and challenger spending, respectively, in Table 3.

[18] Results on the exact small sample distribution of the (properly specified) instrumental variable estimator suggest that, even without the additional complication of misspecification, "the asymptotic distribution is a poor approximation to the true distribution where the instruments are poor, in the sense of not being highly correlated with the regressor, and when the number of observations is small" (Nelson and Startz 1990, 967).

Table 4. Observed (Simulated) and Expected (Asymptotic) Means and Standard Deviations of Parameter Estimate for Incumbent Spending

	n = 25		n = 50		n = 100		n = 250		n = 500		n = 1,000	
Endogeneity	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev
$\rho_{zu} = 0$	**−2.02**[a]	**3.44**	**−1.87**	**2.28**	**−1.87**	**1.24**	**−1.84**	**.81**	**−1.85**	**.62**	**−1.86**	**.42**
	(.15)[b]		(.10)		(.06)		(.04)		(.03)		(.02)	
	−1.86	3.49	−1.86	2.45	−1.86	1.36	−1.86	.84	−1.86	.60	−1.86	.42
$\rho_{zu} = .05$	**−.10**	**3.73**	**−.63**	**2.45**	**−.98**	**1.27**	**−.92**	**.92**	**−.94**	**.61**	**−.95**	**.41**
	(.17)		(.11)		(.06)		(.04)		(.03)		(.02)	
	.00	3.49	−.30	2.45	−.82	1.36	−.92	.84	−.92	.60	−.92	.42
$\rho_{zu} = .10$	**.80**	**3.64**	**.69**	**2.51**	**−.16**	**1.38**	**.01**	**.85**	**−.03**	**.61**	**−.02**	**.44**
	(.16)		(.11)		(.06)		(.04)		(.03)		(.02)	
	1.86	3.49	1.26	2.45	.21	1.36	.01	.84	.01	.60	.01	.42
$\rho_{zu} = .15$	**2.09**	**3.59**	**1.79**	**2.28**	**.82**	**1.33**	**.86**	**.80**	**.86**	**.61**	**.93**	**.44**
	(.16)		(.10)		(.06)		(.04)		(.03)		(.02)	
	3.72	3.49	2.82	2.45	1.24	1.36	.94	.84	.94	.60	.94	.42
$\rho_{zu} = .20$	**3.19**	**3.75**	**3.10**	**2.53**	**1.69**	**1.27**	**1.80**	**.85**	**1.79**	**.59**	**1.84**	**.42**
	(.17)		(.11)		(.06)		(.04)		(.03)		(.02)	
	5.58	3.49	4.38	2.45	2.28	1.36	1.87	.84	1.87	.60	1.87	.42

[a] Observed means and standard deviations from simulations are in bold; corresponding asymptotic means and standard deviations below.

[b] Standard errors of observed means are in parentheses.

(or the R^2 statistics from the purging regression of $\mathbf{x}_1$ on $\mathbf{z}$ and $\mathbf{X}_2$ and from an auxiliary regression of $\mathbf{x}_1$ on $\mathbf{X}_2$ only, from which $R^2_{xz|2}$ can be calculated using expression 26) should be computed (and reported) as a matter of course. The usual practice of reporting only the R^2 statistic from a purging regression of $\mathbf{x}_1$ on both $\mathbf{z}$ and $\mathbf{X}_2$ tends to exaggerate the real efficiency of the instrumental variable, since the total variance in $\mathbf{x}_1$ accounted for by both $\mathbf{z}$ and $\mathbf{X}_2$ may greatly exceed the incremental variance accounted for by $\mathbf{z}$ after controlling for $\mathbf{X}_2$.

Similarly, the appearance of the squared population correlation ρ^2_{z2} in expression (25) suggests that the corresponding squared sample correlation R^2_{z2} (the R^2 statistic from an auxiliary regression of $\mathbf{z}$ on $\mathbf{X}_2$) should be computed (and reported) as a matter of course in order to facilitate intuition regarding the likely magnitude of the partial correlation between $\mathbf{z}$ and $\mathbf{u}$ holding constant $\mathbf{X}_2$.

Finally, and most obviously, analysts and consumers alike should bear in mind that, under some circumstances, even seemingly minor misspecifications will seriously bias statistical inferences based on instrumental variables estimators. Until the nature and magnitude of these biases become fully familiar, considerable care will be necessary to make appropriate adjustments in the inferential weight that we attach to parameter estimates based on what are in fact "quasi-instrumental variable" estimators.

Manuscript submitted 9 August 1989
Final manuscript received 20 August 1990

REFERENCES

Bartels, Larry M. 1985. "Alternative Misspecifications in Simultaneous-Equation Models." *Political Methodology* 11:181–99.

Green, Donald Philip, and Jonathan S. Krasno. 1988. "Salvation for the Spendthrift Incumbent: Reestimating the Effects of Campaign Spending in House Elections." *American Journal of Political Science* 32:884–907.

———. 1990. "Rebuttal to Jacobson's 'New Evidence for Old Arguments.'" *American Journal of Political Science* 34:363–72.

Hanushek, Eric A., and John E. Jackson. 1977. *Statistical Methods for Social Scientists*. New York: Academic Press.

Jacobson, Gary C. 1980. *Money in Congressional Elections*. New Haven: Yale University Press.

———. 1985. "Money and Votes Reconsidered: Congressional Elections, 1972–1982." *Public Choice* 7:7–62.

———. 1990. "The Effects of Campaign Spending in House Elections: New Evidence for Old Arguments." *American Journal of Political Science* 34:334–62.

Johnson, N. L., and S. Kotz. 1972. *Distributions in Statistics: Continuous Multivariate Distributions*. New York: Wiley.

Johnston, J. 1972. *Econometric Methods*. 2d ed. New York: McGraw-Hill.

Kennedy, Peter. 1985. *A Guide to Econometrics*. 2d ed. Cambridge: MIT Press.

Nelson, Charles R., and Richard Startz. 1990. "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator." *Econometrica* 58:967–76.

Theil. Henri. 1971. *Principles of Econometrics*. New York: Wiley.