TECHNICAL WORKING PAPER SERIES

INSTRUMENT RELEVANCE IN MULTIVARIATE LINEAR MODELS: A SIMPLE MEASURE

John Shea

Technical Working Paper 193

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 1996

Research support from the National Science Foundation is acknowledged. I thank Ken West, Jim Stock, Harry Kelejian and a referee for helpful discussions and suggestions. This paper is forthcoming in the *Review of Economics and Statistics*, 1997. This paper is part of NBER's research program in Monetary Economics. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Technical Working Paper 193
March 1996

INSTRUMENT RELEVANCE IN MULTIVARIATE LINEAR MODELS: A SIMPLE MEASURE

ABSTRACT

The correlation between instruments and explanatory variables is a key determinant of the performance of the instrumental variables estimator. The R-squared from regressing the explanatory variable on the instrument vector is a useful measure of relevance in univariate models, but can be misleading when there are multiple endogenous variables. This paper proposes a computationally simple partial R-squared measure of instrument relevance for multivariate models.

John Shea
Department of Economics
University of Wisconsin
1180 Observatory Drive
Madison, WI 53711
and NBER

I. INTRODUCTION

The method of instrumental variables (IV) is one of the most powerful tools of econometrics, because it allows consistent parameter estimation in the presence of correlation between explanatory variables and disturbances. Econometricians have long realized that the performance of the IV estimator depends crucially on the degree of instrument relevance --the correlation between instruments and explanatory variables. Low relevance increases the inconsistency of IV estimates whenever instruments are not perfectly exogenous. Even when instruments are perfectly exogenous, low relevance increases asymptotic standard errors and therefore reduces the power of hypothesis tests. Moreover, low relevance can cause the finite sample distribution of IV estimates to depart considerably from the asymptotic normal distribution; depending on the data generating process, the resulting problems can include finite sample bias, fat tails and missized hypothesis tests (see e.g. Buse (1992), Stock and Staiger (1993), Nelson and Startz (1990) and Hall, Rudebusch and Wilcox (1994)).

In models with one explanatory variable, the R-squared from regresssing the endogenous variable on the instrument vector is a useful measure of relevance. In multivariate models, however, one cannot measure relevance by simply regressing each explanatory variable on the instrument vector in turn. If instruments are highly collinear, for instance, IV can work poorly even when the R-squared is high for each explanatory variable. This paper proposes a simple way to measure relevance in multivariate models. For a given explanatory variable X_i, I suggest computing the squared correlation between the component of X_i orthogonal to the other explanatory variables, and the component of X_i's projection on the instruments orthogonal to the projection of the other explanatory variables on the instruments. This partial R-squared measure can be computed using a series of simple OLS regressions.

The rest of the paper proceeds as follows. Section II motivates the partial R-squared measure of relevance by examining the consistency and precision of IV estimates in multivariate models. Section III presents Monte Carlo evidence on the finite sample behavior of multivariate IV, while Section IV presents a brief empirical example. Section V concludes with a brief discussion of how relevance measures might be used by practitioners.

II. RELEVANCE IN MULTIVARIATE MODELS: ASYMPTOTIC THEORY

Consider the following setup. Suppose Y is a T-by-1 vector of observations of a dependent variable, X is a T-by-k matrix of explanatory variables, and ε is an unobservable mean-zero T-by-1 disturbance correlated with some elements of X. Suppose one wants to estimate

$$Y = X\beta + \varepsilon \tag{1}$$

with two-stage least squares (2SLS), using a T-by-n matrix Z to instrument for X, where $n \geq k$. How should one measure instrument relevance in this case? At first glance, the answer seems obvious: regress each element of X on Z in turn and compute the standard R-squared from each regression. This procedure is in fact common in applied work (see, for instance, Miron and Zeldes (1988), Campbell and Mankiw (1990), Caballero and Lyons (1992) and Attanasio and Weber (1995)). Unfortunately, this procedure may be misleading, as pointed out by Nelson and Startz (1990). For instance, suppose X and Z both have rank two. Suppose that Z_1 is highly correlated with X_1 and X_2, but Z_2 is uncorrelated with X. Then a regression of X_1 or X_2 on Z will produce a high R-squared, even though β is unidentified for practical purposes.

Fortunately, there is a simple way to measure instrument relevance in multivariate models. To motivate this measure, rewrite (1) as

$$(2) \qquad Y = X_1\beta_1 + X_2\beta_2 + \varepsilon,$$

where X_1 is T-by-1 and X_2 is T-by-(k-1). Define $\tilde{X}_1 = X_1 - X_2(X_2'X_2)^{-1}(X_2'X_1)$ as the component of X_1 orthogonal to X_2. Let $\hat{X}_1$ and $\hat{X}_2$ denote the projections of X_1 and X_2 on Z, and let $\bar{X}_1 = \hat{X}_1 - \hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}(\hat{X}_2'\hat{X}_1)$ denote the component of X_1's projection on Z orthogonal to X_2's projection on Z. Now suppose we estimate (2) using 2SLS, using Z to instrument for X. Then the usual partialling-out arguments imply

$$(3) \qquad \beta_1^{2SLS} = (\bar{X}_1'\bar{X}_1)^{-1}(\bar{X}_1'Y),$$

which in turn implies

$$(4) \qquad \beta_1^{2SLS} = \beta_1 + (\bar{X}_1'\bar{X}_1)^{-1}(\bar{X}_1'\varepsilon).$$

From (4), the probability limit of $(\beta_1^{2SLS} - \beta)$ can be written as a function of $\text{plim}(\beta_1^{OLS} - \beta)$, the covariance between ε and $\tilde{X}_1$, the covariance between ε and $\bar{X}_1$, and the population squared correlation between $\bar{X}_1$ and $\tilde{X}_1$, denoted R_p^2:

$$(5) \quad \text{plim}(\beta_1^{2SLS} - \beta) = \text{plim}(\beta_1^{OLS} - \beta) * [\text{Cov}(\bar{X}_1,\varepsilon)/\text{Cov}(\tilde{X}_1,\varepsilon)] * (R_p^2)^{-1}$$

From (5), β_1^{2SLS} is consistent if Z is perfectly exogenous, so that $\text{Cov}(\bar{X}_1,\varepsilon)$ is zero. If the instruments are not exactly exogenous, however, the degree of inconsistency depends on relevance, where in a multivariate context relevance requires that Z have components important to X_1 that are lineraly independent of those important to X_2. Note that if the partial R-squared is low enough, the degree of inconsistency may be larger using 2SLS than using OLS, even if the degree of instrument endogeneity is relatively small.

Provided that β_1^{2SLS} is consistent, its asymptotic standard error is given by

(6) $$ASE(\beta_1^{2SLS}) = \sigma_\varepsilon^2 (\bar{X}_1'\bar{X}_1)^{-1},$$

which conditional on knowing the true σ_ε^2 can be rewritten as

(7) $$ASE(\beta_1^{2SLS}) = SE(\beta_1^{OLS}) * (R_p^2)^{-1},$$

so that even if instruments are exogenous, a low partial R-squared reduces precision and thus reduces the power of hypothesis tests.

The above discussion suggests that practitioners estimating multivariate models may want to compute the sample partial R-squared statistic for each endogenous explanatory variable. For a given X_1, this statistic can be computed as follows:

STEP ONE: Regress X on Z. Save the fitted values $\hat{X}$.

STEP TWO: Regress X_1 on the remaining X. Save the residuals $\tilde{X}_1$.

STEP THREE: Regress $\hat{X}_1$ on the remaining $\hat{X}$. Save the residuals $\bar{X}_1$.

STEP FOUR: Compute the sample squared correlation between $\bar{X}_1$ and $\tilde{X}_1$.

Notice that for scalar X, partial R-squared reduces to the standard R-squared from regressing X on Z. Notice too that if X contains only one endogenous variable, but at least one exogenous variable, the statistic proposed in this paper equals the squared correlation between the components of X_1 and $\hat{X}_1$ orthogonal to X_2, a statistic sometimes reported in the previous literature under the name "partial R-squared" (e.g. Bound, Jaeger and Baker (1995)). Third, as stated above, partial R-squared automatically increases with the number of (possibly irrelevant) overidentifying instruments. Asymptotically, of course, adding irrelevant instruments to Z does no harm; however, Buse (1992) finds that adding irrelevant overidentifying instruments increases the finite sample bias of IV. Practitioners may therefore want to correct partial R-squared for degrees of freedom, as follows:

$$(8) \qquad \bar{R}_p^2 = 1 - (T-1)/(T-n)*(1 - R_p^2),$$

where $\bar{R}_p^2$ denotes the corrected partial R-squared, R_p^2 denotes uncorrected partial R-squared, T is sample size, and n is the number of instruments (including, of course, any exogenous variables that are part of both X and Z). Note that corrected and uncorrected partial R-squared are identical in the case of a scalar X and Z (in which case, of course, partial R-squared would be identical to standard R-squared), and that, like a corrected standard R-squared statistic, corrected partial R-squared falls relative to uncorrected partial R-squared as both the dimension of X and the number of overidentifying instruments rise.

Finally, it is worth comparing partial R-squared to canonical correlations. Bowden and Turkington (1984, pp. 29-32) show that the IV estimator and its covariance matrix can be rewritten in terms of Z and X's canonical correlations and canonical factor loadings; in particular, they show that the estimated standard error of β_i^{2SLS} can be rewritten as the square root of

$$(9) \qquad \sigma_\varepsilon^2 * \sum_{j=1}^{k} (a_{ij}/r_j)^2,$$

where k is the dimension of X, r_j is the jth canonical correlation (where the r are arranged in descending order), and a_{ij} is the loading of canonical variable j on X_i. From (9), a low r_j can cause imprecise estimates of one or more elements of β, so it is natural to think of a low r_j as a sign of instrument irrelevance. Indeed, Hall, Rudebusch and Wilcox (1994) suggest assessing the relevance of Z for X by testing the null hypothesis that the kth canonical correlation is zero; if the null is not rejected, then the effective rank of $\hat{X}$ is less than k, and 2SLS is likely to perform poorly.

Like partial R-squared, canonical correlations "partial out" correlation among instruments, and are thus not vulnerable to the Nelson-Startz critique of standard R-squared. Furthermore, Hall *et al*'s approach utilizes a well-developed distribution theory for testing zero canonical correlations. On the other hand, partial R-squared has the advantage of assigning a relevance measure to each X_i, allowing the researcher to pinpoint variables needing better instruments. Canonical correlations, meanwhile, do not map readily into particular X variables. More importantly, canonical correlations do not distinguish problems due to instrument irrelevance from those due to poor conditioning of X. From (7), a high standard error for β_1^{2SLS} can result either from a low partial R-squared or a high OLS standard error. In turn, the latter can result either from a high variance of the underlying disturbance (a high σ_ε^2), a low variance of X_1, or a high degree of multicollinearity between X_1 and X_2 (the latter two of which would reduce $\tilde{X}_1'\tilde{X}_1$). Equation (9) thus implies that canonical correlations depend on the the variance and multicollinearity of X as well as on instrument relevance. Partial R-squared, meanwhile, measures instrument relevance alone. The distinction is important in practice, since irrelevance can sometimes be cured by finding better instruments, while low variance or multicollinearity of X is presumably incurable.[1]

III. FINITE SAMPLE EVIDENCE

The above discussion motivates partial R-squared by examining the consistency and precision of IV in multivariate models. Aside from these asymptotic problems, a recent literature has examined the effect of relevance on the finite sample behavior of the IV estimator. In general, this literature has found that under low relevance the finite sample distribution of IV can depart dramatically from the asymptotic normal distribution. Buse (1992), for example, approximates the exact finite sample distribution of IV and shows that IV is biased in the direction of OLS, with the bias increasing as instruments grow less relevant. Stock and Staiger (1993) derive the asymptotic distribution of IV in a model where the coefficients from projecting X on Z decline as the sample size grows, so that the F-statistic from projecting X on Z does not automatically go to infinity; they too find that low relevance increases the bias of IV estimates. Nelson and Startz (1990) simulate a just-identified univariate model; they find that low relevance can cause fat tails or, in extreme cases, concentration of IV estimates away from true values with low estimated standard errors. Hall, Rudebusch and Wilcox (1994) extend Nelson and Startz' results, and find that low relevance causes oversized t-tests primarily when the correlation between the disturbance and the explanatory variable is extremely high--that is, when instruments have low relevance, IV performs worst exactly when it is needed the most.

Given these results, it seems prudent to conjecture that a low partial R-squared in a multivariate setting may cause the finite sample distribution of β_1^{2SLS} to differ considerably from the asymptotic distribution. This section presents simulation evidence on the finite sample behavior of IV in a multivariate model. I consider the following data generating process:

(10a) $$Y = \beta_1 X_1 + \beta_2 X_2 + \lambda u_1 + (1-\lambda) u_2$$

(10b) $$X_1 = \gamma u_1 + (1-\gamma) e_1$$

(10c) $$X_2 = \gamma u_2 + (1-\gamma) e_2$$

(10d) $$Z_1 = \delta e_1 + (1-\delta) e_2 + \phi v_1$$

(10e) $$Z_2 = (1-\delta) e_1 + \delta e_2 + \phi v_2,$$

where u_1, u_2, e_1, e_2, v_1 and v_2 are unobserved disturbances, assumed to be standard normal and joint orthogonal; and where Y, X_1, X_2, Z_1 and Z_2 are observable variables. Equation (10a) is the structural equation of interest. From (10b) and (10c), OLS estimation is inappropriate, since X_1 and X_2 are correlated with u_1 and u_2 respectively. From (10d) and (10e), the Z's are correlated with the X's but uncorrelated with the u's, so that 2SLS estimation of (10a) is warranted. The parameter δ governs correlation among the Z's; prior reasoning suggests that 2SLS should be poorly behaved as δ approaches 0.5, since in the limit Z_1 and Z_2 are identical up to disturbances irrelevant to X. The parameters λ and γ govern the correlation between the X's and the disturbance to (10a), with increases in γ raising the endogeneity of both X_1 and X_2, and increases in λ raising the endogeneity of X_1 relative to that of X_2. Prior research suggests that the performance of β_1^{2SLS} may deteriorate as γ or λ increases, particularly if relevance is weak. The parameter ϕ, finally, governs the amount of variation in the Z's that is unrelated to the exogenous components of the X's; I set ϕ to be nonzero so that 2SLS estimation of (10a) is still mechanically possible in the limiting case of $\delta = 0.5$.

Table 1 presents results from a series of experiments investigating the empirical distribution of β_1^{2SLS} generated by (10a)-(10e). In all cases, I set β_1 and β_2 equal to zero, λ equal to 0.9, and ϕ equal to 0.1; δ and γ vary across experiments. Each experiment consists of 10,000 trials. For each trial, I draw 100 observations and estimate β_1^{2SLS} and its t-statistic. For each experiment, I report the population correlation between X_1 and the disturbance to (10a), denoted $\sigma_{x\varepsilon}$; the population standard R-squared (denoted R_s^2), equal to the squared correlation between X_1 and the projection of X_1 on Z; the population partial R-squared (denoted R_p^2), equal to the squared correlation between the part of X_1 orthogonal to X_2 and the part of X_1's projection on Z orthogonal to X_2's projection on Z; the theoretical asymptotic standard error for β_1^{2SLS} (denoted ASE), defined as in (7); the median of the empirical distribution of β_1^{2SLS}; the empirical size, defined as the empirical fraction of estimated t-statistics for β_1^{2SLS} greater than 1.96 in absolute value, where t-statistics are computed using estimated rather than theoretical asymptotic standard errors; and (one minus) the coverage rates of the 95 and 99 percent confidence intervals, defined as the empirical fraction of estimates lying more than 1.96 or 2.326 theoretical asymptotic standard errors away from zero.

Reading down the rows of Table 1, I find that increases in γ increase the correlation between X_1 and the disturbance to (10a), as expected; the set-up of my data generating process implies that increases in γ also reduce the correlation between X_1 and the instruments, as reflected in the results for standard R-squared. For a given γ, both standard and partial R-squared decline as δ approaches

0.5. However, as expected, partial R-squared declines much more rapidly than standard R-squared as δ approaches 0.5, and asymptotic standard errors rise. I find that the finite sample distribution of 2SLS is similar to the asymptotic distribution when $\delta = 1$ and γ is low: the empirical median is zero, the t-test is correctly sized, and the coverage rates of the theoretical asymptotic confidence intervals are accurate. However, the finite sample performance of 2SLS deteriorates as instrument relevance declines and as the correlation between X_1 and the disturbance increases. As δ approaches 0.5, the median of the empirical distribution of β_1^{2SLS} departs from zero, the more so the higher is γ; this result is consistent with Buse (1992) and Stock and Staiger (1993). The empirical distribution initially develops fat tails relative to the theoretical asymptotic distribution as δ falls; the fat tails eventually subside as δ nears 0.5, which is not surprising given that the theoretical asymptotic standard error becomes infinity in the limit. Increases in γ, meanwhile, cause the median of β_1^{2SLS} to diverge farther from zero, and cause the t-test to become oversized, consistent with Nelson and Startz (1990) and Hall *et al* (1994). Interestingly, the effect of instrument relevance on size appears to depend on γ; size declines as δ approaches 0.5 for γ less than 0.8, but rises as δ approaches 0.5 for $\gamma = 0.9$. Note, finally, that the standard R-squared is often quite high despite the presence of median inconsistency and fat tails for δ near 0.5; not surprisingly, standard R-squared can give misleading information about relevance when instruments are highly correlated among themselves.

IV. AN EMPIRICAL EXAMPLE

This section presents a brief empirical example, inspired by the recent macroeconomic literature on returns to scale.[2] My estimating equation is as follows:

$$(11) \qquad dx_t = \alpha + \beta\, dy_t + \gamma\, doil_t + \varepsilon_t,$$

where dx_t is the growth rate of a composite labor and capital input for the US manufacturing sector at time t, dy_t is real manufacturing value added growth at t, $doil_t$ is the growth rate of the real price of oil at t, and ε_t is a disturbance term that represents a technology disturbance. Equations similar to (11) have been estimated by Hall (1990), Caballero and Lyons (1992), and many others. The key parameter in (11) is β, the elasticity of input use with respect to output growth; under constant returns, β should equal one, while under increasing returns β should be less than one. Following Caballero and Lyons (1992), I include the growth of oil prices as an additional regressor in order to control for potential inconsistencies arising from using double-deflated NIPA value added data rather than gross output data (see Bruno (1978) and Basu and Fernald (1995)).

I estimate (11) using US manufacturing data from 1948 through 1987. I measure dx_t as a weighted average of labor and capital growth, where labor is measured as NIPA total hours of all employees in manufacturing, capital is measured as BEA fixed reproducible tangible wealth in manufacturing, and dx_t is weighted using labor's average share of nominal manufacturing output between t and t-1. I measure the real price of oil as the nominal BLS producer price index for crude oil divided by the GNP deflator. Since ε is a technology shock and since technology shocks are likely to affect output, I must instrument for dy_t using demand-shift variables that are relevant for output but uncorrelated with technology. Following Ramey (1989) and Hall (1990), my instruments are the constant term, the growth rate of the real oil price, the growth rate of real US military spending in terms of the GNP deflator, and a dummy variable equal to one when the year-end US President is a Democrat. The 2SLS results are as follows:

$$\text{(12)} \qquad dx_t = \underset{(0.009)}{-0.013} + \underset{(0.244)}{0.818}\, dy_t + \underset{(0.032)}{0.068}\, doil_t + \varepsilon_t$$

Inputs thus increase less than one-for-one with output, consistent with increasing returns, although β is not significantly different from one. The high standard error on β suggests that the instruments might not be very relevant for output growth. The uncorrected standard R-squared from regressing dy_t on the vector of instruments is 0.10. The uncorrected partial R-squared, however, is only 0.05.[3] The discrepancy arises because the most relevant instrument for output growth over my sample period is the oil price, which is an included variable in (11). Partial R-squared recognizes the fact that relevance due to doil cannot help identify β in (11), and thus correctly indicates that the instrument set is not very relevant for output growth in my sample.

V. CONCLUSION

I conclude by discussing the possible uses of partial R-squared in applied work. Given the importance of instrument relevance for the performance of 2SLS, it is tempting for practitioners to use relevance measures such as partial R-squared as a screening device, narrowing down a list of potential instruments by discarding those displaying insufficient relevance.[4] As Hall, Rudebusch and Wilcox (1994) demonstrate, however, such pretesting can be dangerous, because high measured correlation between Z and X *in the sample* can be due to a high correlation between Z and the endogenous part of X *in the sample*, even if Z is exogenous in population. Screening instruments for relevance *ex ante* can thus increase the chance of inconsistent estimates *ex post*.

Hall *et al*'s results do not mean, however, that partial R-squared and other relevance measures are useless. In the first place, one can use relevance measures as *ex post* diagnostic tools, perhaps to identify the cause of high standard errors or to alert the researcher to possible inconsistency if one suspects that the instruments are not perfectly exogenous (Bound, Jaeger and Baker (1995) provide an example of the latter). Second, one may be able to avoid pretesting problems by using a split sample technique; specifically, one might set aside part of the sample for screening instruments for relevance, and then conduct estimation using the rest of the sample.[5] As long as the instruments are exogenous in population, such a screening procedure should not select instruments that are especially likely to be correlated with the disturbance in the estimation sample. I plan to explore the efficacy of such an instrument selection procedure in future work. Finally, one can use the *concept* of instrument relevance as a guiding principle when designing instrument selection strategies *ex ante*, even if one does not want to pretest instruments for relevance. For instance, to estimate equation (11) controlling for energy prices, the concept of partial R-squared suggests that one must find plausible demand-shift instruments *besides the oil price* that are correlated with output. At worst, Hall *et al*'s results do not mean that relevance should be abandoned as a criterion for instrument selection; they merely suggest that researchers should rely primarily on prior reasoning rather than pretests to determine whether an instrument is likely to be relevant or not.

REFERENCES

Angrist, Joshua, and Alan Krueger. "Split-Sample Instrumental Variables Estimates of the Return to Schooling," Journal of Business and Economic Statistics, Volume 13, 1995, 225-235.

Attanasio, Orazio P. and Guglielmo Weber. "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey," Journal of Political Economy, Volume 103 (6), December 1995, 1121-1157.

Basu, Susanto, and John Fernald (1995), "Are Apparent Productive Spillovers in US Manufacturing a Figment of Specification Error?" Journal of Monetary Economics, Volume 36 (1), August 1995, 165-188.

Bruno, Michael. "Duality, Intermediate Inputs and Value-Added," in Melvyn Fuss and Daniel McFadden, eds. Production Economics: A Dual Approach to Theory and Applications, Volume 2. Amsterdam: North-Holland, 1978.

Bound, John; Jaeger, David; and Regina Baker. "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak," Journal of the American Statistical Association, Volume 90, 1995, 443-450.

Bowden, Roger and Darrell Turkington. Instrumental Variables. Cambridge UK: Cambridge University Press, 1984.

Burnside, Craig. "What Do Production Function Regressions Tell Us About Increasing Returns to Scale and Externalities?" mimeo, 1995.

Buse, A. "The Bias of Instrumental Variable Estimators," Econometrica, Volume 60 (1), January 1992, 173-180.

Caballero, Ricardo and Richard Lyons. "External Effects in US Procyclical Productivity," Journal of Monetary Economics, Volume 29, 1992, 209-225.

Campbell, John and Greg Mankiw, "Permanent Income, Current Income and Consumption," Journal of Business and Economic Statistics, Volume 8, 1990, 265-279.

Fuhrer, Jeffery; Moore, George; and Scott Schuh. "Estimating the Linear-Quadratic Inventory Model: Maximum Likelihood versus Generalized Method of Moments," Journal of Monetary Economics, Volume 35 (1), February 1995, 115-157.

Hall, Alastair; Rudebusch, Glenn; and David Wilcox. "Judging Instrument Relevance in Instrumental Variables Estimation," forthcoming International Economic Review, 1994.

Hall, Robert. "Invariance Properties of Solow's Productivity Residual," in Peter Diamond, ed. Growth/Productivity/Employment: Essays in Honor of Robert Solow, Cambridge US: MIT Press, 1990.

Miron, Jeffrey and Steven Zeldes. "Seasonality, Cost Shocks, and the Production Smoothing Model of Inventories," Econometrica, Volume 56, 1988, 877-908.

Nelson, Charles and Richard Startz. "The Distribution of the Instrumental Variables Estimator and its T-Ratio when the Instrument is a Poor One," Journal of Business, Volume 58, 1990, S125-S140.

Shea, John. "Do Supply Curves Slope Up?" Quarterly Journal of Economics, Volume 108 (1), February 1993, 1-32.

Shea, John. "Union Contracts and the Life Cycle-Permanent Income Hypothesis," American Economic Review, Volume 85 (1), March 1995, 186-200.

Staiger, Douglas and James Stock. "Asymptotics for Instrumental Variables Regressions with Weakly Correlated Instruments," mimeo, July 1993.

FOOTNOTES

[1]For instance, Campbell and Mankiw (1990) and Shea (1995) acheive powerful tests of the life cycle-permanent income hypothesis by finding instruments that are stronger predictors of income growth than those considered in previous literature.

[2]Other examples from recent literature using partial R-squared include Fuhrer, Moore and Schuh (1995) and Burnside (1995).

[3]The degrees-of-freedom corrected standard R-squared is 0.022; the degrees-of-freedom corrected partial R-squared is -0.031.

[4]Regrettably, Shea (1993) is an example of this practice.

[5]Angrist and Krueger (1995) suggest the use of split samples to avoid the usual finite sample bias in 2SLS caused by "overfitting" in the first stage regression. Their proposed procedure is to set aside part of the sample to estimate the coefficients of the first-stage projection of X on Z, and then to estimate the second-stage relationship using the other part of the sample, using the coefficients from the first sample to form fitted X. Angrist and Krueger do not discuss pretesting instruments for relevance, however.

TABLE 1

Monte Carlo simulation of (10a)-(10e)

δ	γ	$\sigma_{x\varepsilon}$	R_s^2	R_p^2	ASE	Median	Size	Coverage Rates 95	Coverage Rates 99
1	0.3	0.39	0.84	0.84	0.13	-0.00	0.05	0.06	0.01
0.53	0.3	0.39	0.53	0.36	0.20	-0.00	0.05	0.07	0.02
0.52	0.3	0.39	0.48	0.20	0.26	0.00	0.03	0.10	0.05
0.51	0.3	0.39	0.43	0.06	0.48	0.02	0.01	0.16	0.11
0.50	0.3	0.39	0.42	0	∞	0.18	0.00	---	---
1	0.5	0.70	0.50	0.50	0.18	-0.00	0.05	0.06	0.03
0.53	0.5	0.70	0.31	0.21	0.28	0.00	0.05	0.09	0.04
0.52	0.5	0.70	0.28	0.12	0.37	0.00	0.05	0.11	0.07
0.51	0.5	0.70	0.26	0.04	0.67	0.09	0.03	0.18	0.14
0.50	0.5	0.70	0.25	0	∞	0.40	0.02	---	---
1	0.7	0.91	0.15	0.15	0.30	0.00	0.07	0.10	0.06
0.53	0.7	0.91	0.10	0.06	0.47	0.03	0.07	0.15	0.12
0.52	0.7	0.91	0.09	0.04	0.61	0.09	0.07	0.19	0.15
0.51	0.7	0.91	0.08	0.01	1.11	0.30	0.06	0.17	0.13
0.50	0.7	0.91	0.08	0	∞	0.47	0.05	---	---
1	0.8	0.96	0.06	0.06	0.46	0.02	0.09	0.17	0.14
0.53	0.8	0.96	0.04	0.02	0.70	0.16	0.08	0.21	0.17
0.52	0.8	0.96	0.03	0.01	0.92	0.26	0.09	0.20	0.16
0.51	0.8	0.96	0.03	0.004	1.66	0.41	0.09	0.12	0.09
0.50	0.8	0.96	0.03	0	∞	0.51	0.08	---	---
1	0.9	0.99	0.01	0.01	0.91	0.42	0.15	0.23	0.18
0.53	0.9	0.99	0.01	0.005	1.40	0.58	0.17	0.15	0.11
0.52	0.9	0.99	0.01	0.003	1.84	0.62	0.18	0.10	0.08
0.51	0.9	0.99	0.01	0.001	3.33	0.67	0.19	0.05	0.04
0.50	0.9	0.99	0.01	0	∞	0.69	0.20	---	---

NOTES: This table presents results from simulating equations (10a)-(10e) in the text, for varying δ and γ. See the text for details.