TECHNICAL WORKING PAPER SERIES

INSTRUMENTAL VARIABLES REGRESSION WITH WEAK INSTRUMENTS

Douglas Staiger
James H. Stock

Technical Working Paper No. 151

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 1994

The authors thank Joshua Angrist, John Bound, Adolph Buse, Gary Chamberlain, Jean-Marie Dufour, Jerry Hausman, Guido Imbens, Tom Kane, Alan Krueger, Tom Rothenberg, and Richard Startz for helpful discussions and/or comments on an earlier draft, and Matthew Eichner for research assistance. The research was supported in part by National Science Foundation grant no. SES-91-22463. This paper is part of NBER's research program in Asset Pricing. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #151
January 1994

INSTRUMENTAL VARIABLES
REGRESSION WITH WEAK INSTRUMENTS

ABSTRACT

This paper develops asymptotic distribution theory for instrumental variable regression when the partial correlation between the instruments and a single included endogenous variable is weak, here modeled as local to zero. Asymptotic representations are provided for various instrumental variable statistics, including the two-stage least squares (TSLS) and limited information maximum likelihood (LIML) estimators and their t-statistics. The asymptotic distributions are found to provide good approximations to sampling distributions with just 20 observations per instrument. Even in large samples, TSLS can be badly biased, but LIML is, in many cases, approximately median unbiased. The theory suggests concrete quantitative guidelines for applied work. These guidelines help to interpret Angrist and Krueger's (1991) estimates of the returns to education: whereas TSLS estimates with many instruments approach the OLS estimate of 6%, the more reliable LIML and TSLS estimates with fewer instruments fall between 8% and 10%, with a typical confidence interval of (6%, 14%).

Douglas Staiger
Kennedy School of Government
Harvard University
Cambridge, MA 02138
and NBER

James H. Stock
Kennedy School of Government
Harvard University
Cambridge, MA 02138
and NBER

1. Introduction

In empirical applications of instrumental variable techniques, very often the partial correlation between the instruments and the included endogenous variable is low, that is, the instruments are weak. It is our impression that, in applications of two-stage least squares (TSLS), it is common for the F-statistic, testing the hypothesis that the instruments do not enter the first-stage regression, to take on a value less than 10.[1] Unfortunately, standard asymptotic approximations to the distributions of the main instrumental variables statistics break down in this case. Recently this has been highlighted for TSLS in quite different settings by Nelson and Startz (1990a,b) using only a few observations and a single instrument and by Bound, Jaeger and Baker (1993) using up to 180 instruments and over 300,000 observations. Both Nelson and Startz and Bound, Jaeger and Baker find that the TSLS estimator is biased in the direction of the ordinary least squares (OLS) estimator, and that the TSLS standard error is small relative to the bias. While a large literature on finite-sample distribution theory has tackled these departures from conventional asymptotics, the finite-sample approach has several drawbacks which impede its use in practice, including the assumption of Gaussian errors and fixed instruments, unwieldy expressions for distributions which can be computationally intractable, and most importantly the failure to produce clear quantitative guidelines which empirical researchers can follow.

This paper develops an alternative framework for analyzing the distributions of statistics arising in instrumental variables regression in a single equation with a single endogenous variable included as a regressor. Modern applications of instrumental variables techniques typically use many observations, so our approximations are asymptotic in the sample size N. Conventional asymptotics, both first-order and higher-order such as Edgeworth expansions (see for example Anderson and Sawa (1973, 1979) and for a review see Rothenberg (1984)), treat the coefficient on the instruments in the first stage equation as nonzero and fixed, an assumption which implies that the first-stage F statistic increases to infinity with the sample size. Not surprisingly, when this F is small, these asymptotic approximations break down. Because our objective is to inform inference when the first-stage F-statistic is small, the asymptotics developed here employ a device which, loosely speaking, holds this F-statistic constant as the sample size increases. More precisely, the coefficients on the instruments in the first-stage equation are modeled as being in a $N^{-1/2}$ neighborhood of zero. We refer to this as the "nearly unidentified" case. Under this assumption, the first-stage F statistic is asymptotically proportional to a noncentral χ^2 random variable.

These local-to-zero asymptotics are used to develop alternative asymptotic approximations to the distribution of various statistics arising in instrumental variables regression, including the TSLS and limited information maximum likelihood (LIML) estimators and their t-ratios; k-class estimators of the coefficients on the included exogenous variables; tests of overidentifying restrictions; and tests for endogeneity. Some of these distributions are closely related to ones obtained in the finite-sample/Gaussian literature (see the review article by Phillips (1983) and, for the exactly unidentified case, Phillips (1984, 1985)), but the results here hold under weaker conditions and have simpler derivations and computations[2]. For example, the errors can be martingale difference sequences and the instruments can be stochastic and merely predetermined, in contrast to the assumptions of normal errors and fixed instruments in the finite-sample literature. The local-to-zero asymptotics bridge the exact (Gaussian) distribution theory and fixed-parameter asymptotics, including as special cases both the usual normal asymptotics and the nonstandard asymptotics obtained by Phillips' (1989) and Choi and Phillips' (1992) in the "partially identified" model, who treat the first-stage coefficients as fixed but permit some to be zero.

The paper has four main methodological conclusions. First, the resulting asymptotic distributions are very close to the finite-sample distributions, even with as few as 20 observations in the extreme bimodal cases considered by Nelson and Startz (1990a) and Maddala and Jeong (1992). Second, the nonnormal local-to-zero asymptotic distribution of the usual TSLS t-statistic implies that conventionally constructed confidence intervals will fail to have the desired coverage rates, even in large samples. However, an alternative approach to the construction of confidence intervals, based on a statistic proposed by Anderson and Rubin (1949), is asymptotically justified for general error distributions. Third, LIML point estimates are found to be approximately median-unbiased and LIML confidence intervals have approximately their nominal coverage rates, even for weak instruments. Fourth, the theoretical results suggest that key features of the distributions, such as bias and coverage rates, can be summarized in a series of simple plots which are applicable to a wide range of models and number of instruments. These plots can be used to provide concrete quantitative guidelines for empirical applications of instrumental variables regressions.

We use these results to interpret Angrist and Krueger's (1991) important and innovative study of the returns to education, in which they used the quarter of birth and its interactions with other covariates as instruments for education in an earnings equation. Labor economists have long hypothesized that OLS estimates of the returns to education are biased upward due to a positive correlation between innate ability and years of education. Strikingly, Angrist and Krueger's TSLS estimates suggested that OLS estimates are unbiased or are biased slightly downward, perhaps due to measurement error. However, in several of their specifications, the first-stage F statistic is less than 5. The local-to-zero asymptotics suggests that the TSLS estimates and confidence intervals based on many instruments are unreliable, despite having more than 300,000 observations. In particular the theory explains the observed movement of TSLS towards OLS as the number of instruments increases. In contrast, LIML estimates and especially Anderson-Rubin (1949) (AR) confidence intervals are arguably more reliable here. LIML and AR confidence intervals are similar to the conventional TSLS confidence intervals when the first stage F is large and the number of instruments is small, but for many instruments and low first-stage F's the LIML and AR confidence intervals are wider than the conventional but unreliable TSLS confidence intervals. Based on our preferred statistics, we estimate returns to education which are higher, but confidence intervals which are wider, than suggested by Angrist and Krueger.

The paper is organized as follows. The asymptotic framework is developed in section 2, where asymptotic representations are presented for various statistics arising in instrumental variable regression including TSLS and LIML. Monte Carlo experiments which check the quality of the asymptotic approximation to the finite sample distributions are summarized in section 3. Section 4 suggests guidelines for applying these results in empirical work. Angrist and Krueger's (1991) data are used in Section 5 to study the returns to education. Section 6 concludes.

2. Asymptotics with Weakly Correlated Instruments

A. The model, assumptions, and an example

In matrix notation, the model considered is,

$$y = Y\beta + X\gamma + u \tag{2.1}$$

$$Y = Z\pi + X\phi + v \tag{2.2}$$

where y is the $N \times 1$ vector of observations on the endogenous variable in the equation of interest, Y is the $N \times 1$ vector of observations on the single included endogenous variable, X is the $N \times K_1$ matrix of K_1 exogenous regressors, Z is the $N \times K_2$ matrix of K_2 instruments, v and u are each $N \times 1$ error terms, and β, γ, π, and ϕ are unknown parameters. The structural equation for y is (2.1) and the reduced form equation for Y is (2.2). The errors (u_i, v_i) are assumed to be serially uncorrelated and to be homoskedastic with covariance matrix Ω, so that $\Omega_{11} = \sigma_{uu}$, $\Omega_{12} = \Omega_{21} = \sigma_{uv}$, and $\Omega_{22} = \sigma_{vv}$. Let $\rho = \sigma_{uv}/(\sigma_{uu}\sigma_{vv})^{1/2}$. Throughout, it is assumed that X_i and Z_i are uncorrelated with u_i and v_i, where X_i denotes the i-th observation on X, etc. With the sole exception of the local power analysis of tests of overidentifying restrictions in section 2C, (2.1) and (2.2) are assumed to hold throughout.

We are interested in statistical inferences about β and γ when the instrument Z is only weakly related to Y, that is, when the F-statistic testing $\pi = 0$ in (2.2) is small or moderate even though the number of observations N might be large. If π is modeled as fixed, then this first-stage F statistic tends to infinity as N increases, suggesting that conventional fixed-π asymptotics is inappropriate in the situation at hand. If, however, π is local to zero, this F statistic is $O_p(1)$. We therefore assume,

Assumption L_π. $\pi = \pi_N = N^{-1/2}g$, where g is a fixed $K_2 \times 1$ vector.

Rather than make primitive assumptions on the errors and exogenous variables, we instead assume moment conditions which they must satisfy. This permits the subsequent application of the results in either time series or cross sectional settings, where the primitive assumptions on the variables typically differ. Let "=>" denote convergence in distribution.

Assumption M. The following limits hold jointly:

(a) $(u'u/N, v'u/N, v'v/N) \xrightarrow{p} (\sigma_{uu}, \sigma_{uv}, \sigma_{vv})$;

(b) $(X'X/N, X'Z/N, Z'Z/N) \xrightarrow{p} (\Sigma_{XX}, \Sigma_{XZ}, \Sigma_{ZZ})$;

(c) $(N^{-1/2}X'u, N^{-1/2}Z'u, N^{-1/2}X'v, N^{-1/2}Z'v) => (\Psi_{Xu}, \Psi_{Zu}, \Psi_{Xv}, \Psi_{Zv})$,

where $\Psi = (\Psi_{Xu}', \Psi_{Zu}', \Psi_{Xv}', \Psi_{Zv}')'$ is distributed $N(0, \Omega \otimes \Sigma)$

and $\Sigma = E(X_i' \ Z_i')'(X_i' \ Z_i')$.

These conditions hold under the standard assumptions in the literature on finite sample distributions of IV estimators, namely X and Z fixed but having nonstochastic limits analogous to assumption M(b). They also hold under much weaker assumptions, such as those found in time series settings. For example, if (u_i, v_i) is a homoskedastic vector martingale difference sequence with respect to the filtration based on $\{u_{j-1}, v_{j-1}, X_j, Z_j, j \leq i\}$, if u_i and v_i have 4 moments, and if X_i and Z_i are integrated of order zero with four moments and satisfy additional weak conditions limiting dependence, then (a) and (b) follow from the weak law of large numbers and (c) follows from the central limit theorem for martingale difference sequences.

If σ_{uv} is nonzero, then Y is endogenous and the OLS estimator of β, $\hat{\beta}_{OLS}$, is inconsistent. Let β_0 denote the true value of β and let $\theta = \sigma_{uv}/\sigma_{vv}$. Then, under assumptions L_π and M, (2.3) $$\hat{\beta}_{OLS} \xrightarrow{p} \beta_0 + \theta.$$

Before turning to the general results, we illustrate the approach by sketching the asymptotics for the TSLS estimator $\hat{\beta}_{TSLS}$ in the special case that there are no exogenous regressors in (2.1). Then $\hat{\beta}_{TSLS} = (\hat{Y}'\hat{Y})^{-1}(\hat{Y}'y)$ where $\hat{Y} = Z(Z'Z)^{-1}Z'Y$, so

(2.4) $$\hat{\beta}_{TSLS} - \beta_0 = [(N^{-\frac{1}{2}}Y'Z)(N^{-1}Z'Z)^{-1}(N^{-\frac{1}{2}}Z'Y)]^{-1}[(N^{-\frac{1}{2}}Y'Z)(N^{-1}Z'Z)^{-1}(N^{-\frac{1}{2}}Z'u)].$$

By assumption M, $N^{-1}Z'Z \xrightarrow{p} \Sigma_{ZZ}$ and $(N^{-\frac{1}{2}}Z'u, N^{-\frac{1}{2}}Z'v) \Rightarrow (\Psi_{Zu}, \Psi_{Zv})$. Applying assumptions L_π and M to $N^{-\frac{1}{2}}Z'Y$ therefore yields,

(2.5) $$N^{-\frac{1}{2}}Z'Y = N^{-\frac{1}{2}}Z'(Z\pi+v) = (N^{-1}Z'Z)g + N^{-\frac{1}{2}}Z'v \Rightarrow \Sigma_{ZZ}g + \Psi_{Zv}.$$

It is useful to define,

(2.6) $$(z_u, z_v) = (\Sigma_{ZZ}^{-\frac{1}{2}}\Psi_{Zu}/\sqrt{\sigma_{uu}},\ \Sigma_{ZZ}^{-\frac{1}{2}}\Psi_{Zv}/\sqrt{\sigma_{vv}}),$$

so that $(z_u', z_v')'$ is distributed $N(0,\bar{\Omega}\otimes I_{K_2})$, where $\bar{\Omega}$ is the 2×2 matrix with $\bar{\Omega}_{11}=\bar{\Omega}_{22}=1$ and $\bar{\Omega}_{12}=\bar{\Omega}_{21}=\rho$ and I_{K_2} is the $K_2\times K_2$ identity matrix. Define $\lambda = V_{ZZ}^{\frac{1}{2}}g/\sqrt{\sigma_{vv}}$, where $V_{ZZ} = \Sigma_{ZZ} - \Sigma_{ZX}\Sigma_{XX}^{-1}\Sigma_{XZ}$ (in this example there are no X's so $V_{ZZ} = \Sigma_{ZZ}$). Applying (2.5) to (2.4) and using the definition of λ and (2.6) yields,

(2.7) $$\hat{\beta}_{TSLS} - \beta_0 \Rightarrow (\sigma_{uu}/\sigma_{vv})^{\frac{1}{2}}(\lambda+z_v)'z_u/(\lambda+z_v)'(\lambda+z_v).$$

Thus $\hat{\beta}_{TSLS}-\beta_0$ is asymptotically the ratio of quadratic forms in the K_2-dimensional jointly normal random variables, z_v and z_u.

Because u and v are correlated the quadratic form in (2.7) does not have an elementary interpretation. It is therefore useful to rewrite z_u as its expectation given z_v, plus an independent residual, specifically, $z_u = \rho z_v + (1-\rho^2)^{1/2}\eta$, where η is an K_2-dimensional standard normal random variable which is independent of z_v. Because $\rho(\sigma_{uu}/\sigma_{vv})^{1/2} = \sigma_{uv}/\sigma_{vv} = \theta$, applying this factorization to (2.7) yields the alternative asymptotic representation,

$$(2.8) \qquad \hat{\beta}_{TSLS} - \beta_0 => \theta\{(\lambda+z_v)'z_v/(\lambda+z_v)'(\lambda+z_v)\} + \kappa\{(\lambda+z_v)'\eta/(\lambda+z_v)'(\lambda+z_v)\}$$

where $\kappa = [(1-\rho^2)\sigma_{uu}/\sigma_{vv}]^{1/2}$. Because $E(\eta|z_v) = 0$, the bias of $\hat{\beta}_{TSLS}$ is $\theta E[(\lambda+z_v)'z_v/(\lambda+z_v)'(\lambda+z_v)]$. Moreover, it is evident that $\hat{\beta}_{TSLS}$ is not consistent but rather has the asymptotic mixture-of-normals distribution, $\int N(\beta_0+m(z_v),v(z_v))dF(z_v)$, where $m(z_v) = \theta\{(\lambda+z_v)'z_v/(\lambda+z_v)'(\lambda+z_v)\}$ and $v(z_v) = \kappa^2/(\lambda+z_v)'(\lambda+z_v)$.

B. Asymptotics for TSLS

We now turn to a more precise statement of this result, extended to include exogenous regressors X. Let the superscript "$\perp$" denote the projection orthogonal to X, so $y^\perp = M_X y$, $Y^\perp = M_X Y$ and $Z^\perp = M_X Z$ where $M_X = I_N - X(X'X)^{-1}X'$, and let $\hat{Y}^\perp = P_{Z^\perp} Y^\perp$ where $P_{Z^\perp} = Z^\perp(Z^{\perp\prime}Z^\perp)^{-1}Z^{\perp\prime}$. By standard projection arguments, $\hat{\beta}_{TSLS}$, its t-statistic, t_{TSLS}, and the squared standard error of the regression are,

$$(2.9a) \qquad \hat{\beta}_{TSLS} = (\hat{Y}^{\perp\prime}\hat{Y}^\perp)^{-1}(\hat{Y}^{\perp\prime}y^\perp)$$

$$(2.9b) \qquad t_{TSLS} = (\hat{Y}^{\perp\prime}\hat{Y}^\perp/\hat{\sigma}_{uu})^{1/2}(\hat{\beta}_{TSLS} - \beta_0)$$

$$(2.10) \qquad \hat{\sigma}_{uu} = (y-Y\hat{\beta}_{TSLS}-X\hat{\gamma}_{TSLS})'(y-Y\hat{\beta}_{TSLS}-X\hat{\gamma}_{TSLS})/(N-K_1-1)$$
$$= (y^\perp-Y^\perp\hat{\beta}_{TSLS})'(y^\perp-Y^\perp\hat{\beta}_{TSLS})/(N-K_1-1).$$

Let F_N denote the usual Wald statistic testing the joint significance of the instruments in the first stage regression, that is, testing the hypothesis that $\pi=0$ against the alternative $\pi\neq0$. The main results on the limiting distribution of the TSLS estimator and its related statistics are given in theorem 1.

Theorem 1. Suppose that assumptions L_π and M hold. Then:

(a) $\hat{\beta}_{TSLS} - \beta_0 => \theta\nu_1/\nu_3 + \kappa\nu_2/\nu_3 \equiv \beta^*$,

(b) $\hat{\sigma}_{uu} => \sigma_{uu}[1-2\rho(\sigma_{vv}/\sigma_{uu})^{1/2}\beta^* + (\sigma_{vv}/\sigma_{uu})\beta^{*2}]$,

(c) Under the null hypothesis $\beta = \beta_0$, $t_{TSLS} => \nu_4/\{\nu_3[1-2\rho(\sigma_{vv}/\sigma_{uu})^{1/2}\beta^* + (\sigma_{vv}/\sigma_{uu})\beta^{*2}]\}^{1/2}$, and

(d) $F_N => \nu_3/K_2$,

where $\nu_1 = (\lambda+z_v)'z_v$, $\nu_2 = (\lambda+z_v)'\eta$, $\nu_3 = (\lambda+z_v)'(\lambda+z_v)$, $\nu_4 = (\lambda+z_v)'z_u$, $\theta = \sigma_{uv}/\sigma_{vv}$, $\kappa = [(\sigma_{uu}/\sigma_{vv})(1-\rho^2)]^{1/2}$, $\lambda = V_{ZZ}^{1/2}g/\sqrt{\sigma_{vv}}$, $z_u = \rho z_v + (1-\rho^2)^{1/2}\eta$, and $(z_v' \;\; \eta')'$ is a $N(0, I_{2K_2})$ vector of random variables.

Proofs are given in the appendix.

A comparison of theorem 1(a) and (2.8) reveals that the inclusion of exogenous variables in (2.1) does not change the limiting distribution of $\hat{\beta}_{TSLS}$, and that in the general case $\hat{\beta}_{TSLS}$ has the asymptotic mixture-of-normals distribution given following (2.8).

The representations in theorem 1 depend on only three unknown parameters: $\lambda'\lambda/K_2 = g'V_{ZZ}g/(K_2\sigma_{vv})$, where g is the local-to-zero parameter from the first stage regression; the ratio of the error variances, σ_{uu}/σ_{vv}; and ρ or equivalently θ, which determines the magnitude of the limiting inconsistency of the OLS estimator. The parameter $\lambda'\lambda$ has a simple interpretation: the first stage F-statistic converges to a noncentral $\chi^2_{K_2}$, divided by the number of instruments K_2, with noncentrality parameter $\lambda'\lambda$. This noncentrality parameter is the asymptotic analog of the so-called concentration parameter which appears in finite-sample treatments of the distribution of $\hat{\beta}_{TSLS}$.[3]

In some cases, the distributions do not depend on the ratio σ_{uu}/σ_{vv}. One important example is the ratio of the TSLS bias to the OLS bias, which from theorem 1 has the limiting representation, $(\hat{\beta}_{TSLS}-\beta_0)/\theta => \nu_1/\nu_3 + \rho^{-1}(1-\rho^2)^{1/2}\nu_2/\nu_3$, where the joint distribution of (ν_1, ν_2,

ν_3) depends only on $\lambda'\lambda/K_2$, K_2, and ρ. A second example is the TSLS t-statistic, which from theorem 1(c) has the representation, $t_{TSLS} => \nu_4/\{\nu_3[1-2\rho\tilde{\beta}^* + \tilde{\beta}^{*2}]\}^{1/2}$, where $\tilde{\beta}^* = \rho(\nu_1/\nu_3) + (1-\rho^2)^{1/2}(\nu_2/\nu_3)$, the distribution of which also only depends on $\lambda'\lambda/K_2$, K_2, and ρ. Moreover, the limiting distributions of $(\hat{\beta}_{TSLS}-\beta_0)/\theta$ depends only on $|\rho|$, while the limiting distribution of t_{TSLS} is symmetric in ρ.[4] In particular, the limiting relative bias, $E\beta^*/\theta$, and the coverage rates of symmetric two-sided confidence intervals based on t_{TSLS} depend only on $|\rho|$, $\lambda'\lambda/K_2$, and K_2, a simplification which shall be exploited in sections 3 - 5.

A case of special interest is when the instruments are uncorrelated with the included endogenous variable so $\lambda'\lambda = 0$, the so-called leading case. A direct calculation using theorem 1 reveals that,

(2.11) $$\hat{\beta}_{TSLS} => \beta_0 + \theta + \kappa t_{K_2}/K_2^{1/2},$$

(2.12) $$t_{TSLS} => (z_v'z_u)/\{(z_v'z_v)[1-2\rho(\sigma_{vv}/\sigma_{uu})^{1/2}\beta^* + (\sigma_{vv}/\sigma_{uu})\beta^{*2}]\}^{1/2}$$

where t_{K_2} denotes a t distribution with K_2 degrees of freedom. These results were previously obtained as exact results for the Gaussian/strictly exogenous regressor model by Phillips (1984, 1985), and Phillips (1989) and Choi and Phillips (1992) showed that they hold asymptotically under weak assumptions on the errors and instruments such as those here. (Phillips (1989) and Choi and Phillips (1992) consider multiple included endogenous variables but their results are restricted to the fixed-π case.) If $K_2 = 1$ then $\hat{\beta}_{TSLS}$ has no moments asymptotically, although it does if $K_2 > 1$. The distribution of $\hat{\beta}_{TSLS}$ is centered around the probability limit of the OLS estimator, $\beta_0 + \theta$. When the number of instruments is large, the normal approximation to the t distribution can be used and $\hat{\beta}_{TSLS}$ approximately has the distribution, $N(\beta_0+\theta, \kappa^2/K_2)$. Clearly, with many irrelevant instruments and/or $|\rho|$ nearly one (so κ is nearly zero) this distribution can be tightly concentrated around $\beta_0+\theta$.

If $\lambda'\lambda/K_2 = 0$, tests based on t_{TSLS} can reject the true value β_0 with high probability. To see this, consider the extreme case of $|\rho| = 1$. Then $z_u=z_v$ and $\rho(\sigma_{vv}/\sigma_{uu})^{1/2}\beta^* = 1$, so from (2.12)

t_{TSLS} is asymptotically the square root of $z_v'z_v$ (a $\chi^2_{K_2}$ random variable), divided by 0 and asymptotically the t statistic is infinite. For $|\rho|$ nearly one, the t-statistic will be very large under the null and typically will reject when tests are based on conventional critical values. Similar problems occur for K_2 large even if $|\rho|$ is moderate. To see this, suppose K_2 is sufficiently large that $\kappa t_{K_2}/K_2^{1/2}$ is negligible so $\beta^* \dot{=} \theta$. Then (2.12) becomes, $t_{TSLS} => \rho(1-\rho^2)^{-1/2}(z_v'z_v)^{1/2} + (z_v'\eta)/(z_v'z_v)^{1/2} = \rho(1-\rho^2)^{-1/2}\zeta_1 + \zeta_2$, where (ζ_1, ζ_2) are independent $(\chi^2_{K_2})^{1/2}$ and N(0,1) random variables, respectively. Thus, for large K_2 the t-statistic will incorrectly reject $\beta = \beta_0$ with high probability.

Alternatively, consider the case that $\lambda'\lambda/K_2$ is large. Then from appendix lemma A1(i), $\hat{\nu}^{\perp\prime}\hat{\nu}^{\perp}/\lambda'\lambda = \sigma_{vv} + O_p((\lambda'\lambda)^{-1/2}) + O_p((\lambda'\lambda/K_2)^{-1})$. When $\lambda'\lambda/K_2$ is large enough for the remainder terms to be negligible, the result in theorem 1(a) simplifies to,

$$\hat{\beta}_{TSLS} - \beta_0 \dot{=}> (\sigma_{uu}/\sigma_{vv})^{1/2}\lambda'z_u/\lambda'\lambda \sim N(0, \sigma_{uu}/(\sigma_{vv}\lambda'\lambda)). \tag{2.13}$$

This is the usual fixed-π asymptotic normal approximation for $\hat{\beta}_{TSLS}$. Similarly, for $\lambda'\lambda/K_2$ large, the distribution of t_{TSLS} in theorem 1(c) is well approximated by a standard normal, the usual result.

For intermediate values of the concentration parameter, $\hat{\beta}_{TSLS}$ is biased with the magnitude of the bias depending on θ, $\lambda'\lambda/K_2$, K_2, and σ_{uu}/σ_{vv}. Similarly, t_{TSLS} does not have a standard t or normal distribution. The nonstandard distribution for t_{TSLS} comes from two sources: the nonstandard distribution of $\hat{\beta}_{TSLS}$ and the nonstandard distribution of the estimator of the variance, $\hat{\sigma}^2_{uu}$. This latter contribution arises because $\hat{\beta}_{TSLS}$ is not consistent for β but rather has a limiting distribution, so $\hat{\sigma}^2_{uu}$ does not estimate σ^2_{uu} consistently.

Unless $\lambda'\lambda/K_2$ is large, the nonstandard distribution of the t-statistic impedes the construction of asymptotic confidence intervals in this problem. Because the t-statistic does not have an asymptotic normal distribution, confidence intervals contructed as ± 1.96 standard errors will not in general have a 95% coverage rate, even asymptotically. Rather, the limiting representation of

t_{TSLS} indicates that the distribution depends on ρ in a complicated way, not just as a mean or scale shift. Thus critical values for t_{TSLS} depend on the null value being tested, so confidence intervals must be constructed using confidence belts which depend on ρ. Worse, the distribution of t_{TSLS} also depends on $\lambda'\lambda/K_2$, so that the confidence belts must be indexed by $\lambda'\lambda/K_2$. However, $\lambda'\lambda/K_2$ is not consistently estimable and is unknown in applications. Thus, without resorting to conservative and cumbersome methods such as Scheffe or Bonferroni intervals, t_{TSLS} does not provide a suitable statistic for the construction of asymptotic confidence intervals.

C. Tests of Overidentifying Restrictions and Endogeneity

Two tests commonly used in instrumental variables regression are tests of overidentifying restrictions and the Durbin-Wu-Hausman (DWH) test for endogeneity. This section provides the limiting behavior of these tests in the local-to-zero setting under the null and alternative hypotheses.

We first consider two standard tests of overidentifying restrictions applicable if $K_2 > 1$: NR^2 from a regression of the residuals on the instruments and exogenous variables, here denoted χ^2_{reg}, and Basman's (1960) test, here denoted F_{Bas}. The two statistics are respectively,

$$\chi^2_{reg} = \hat{u}'P_{Z^\perp}\hat{u}/(\hat{u}'\hat{u}/N) \tag{2.14}$$

$$F_{Bas} = \{\hat{u}'P_{Z^\perp}\hat{u}/(K_2-1)\}/\{\hat{u}'M_{Z^\perp}\hat{u}/(N-K_1-K_2)\} \tag{2.15}$$

where $\hat{u} = y^\perp - Y^\perp\hat{\beta}_{TSLS}$ is the residual from the second-stage regression. Because $\hat{u}$ depends on $\hat{\beta}_{TSLS}$, which in turn involves the errors v, even under normality these statistics do not have exact χ^2 or F distributions under the null. Relevant questions therefore are whether these tests have standard null distributions asymptotically in the nearly unidentified case, and what their properties are when there is a small violation of the identification orthogonality restrictions.

The properties of these tests under small violations of the orthogonality conditions are investigated by deriving their asymptotic representations under a local alternative in which (2.1) fails to hold and instead the instruments weakly enter the equation of interest:

$$(2.16) \qquad y = Y\beta + X\gamma + Z\omega + u, \text{ where } \omega = \omega_N = N^{-\frac{1}{2}}d.$$

The limiting representation of the regression and Basman tests of overidentifying restrictions is given in the next theorem.

Theorem 2. Suppose that assumptions L_π and M hold and that (2.1) is replaced by (2.16).

(a) $\hat{\beta}_{TSLS}-\beta_0 => \beta_d^*$, where $\beta_d^* = \beta^* + \nu_3^{-1}(\lambda+z_v)'V_{ZZ}^{\frac{1}{2}}d/\sqrt{\sigma_{vv}}$ where β^* is defined in theorem 1.

(b) $F_{Bas} => Q(\beta_d^*)/(K_2-1)$, $\chi^2_{reg} => Q(\beta_d^*)$, and $F_{Bas} - \chi^2_{reg}/(K_2-1) \xrightarrow{p} 0$, where

$$Q(b) = \{1-2\rho(\sigma_{vv}/\sigma_{uu})^{\frac{1}{2}}b+(\sigma_{vv}/\sigma_{uu})b^2\}^{-1}[z_u-(\sigma_{vv}/\sigma_{uu})^{\frac{1}{2}}(\lambda+z_v)b]'[z_u-(\sigma_{vv}/\sigma_{uu})^{\frac{1}{2}}(\lambda+z_v)b].$$

The two tests are asymptotically equivalent under the null and the local alternative. Under the null, $d=0$ and $\beta_d^* = \beta^*$. Inspection of the expression for Q reveals that for general $\lambda'\lambda/K_2$ neither the regression nor Basmann overidentifying tests have their standard χ^2 asymptotic null distributions: although z_u is normally distributed, $\hat{\beta}_{TSLS}$ is $O_p(1)$ which makes the asymptotic distribution nonstandard. However, the null distributions of the tests approach their standard limits when $\lambda'\lambda/K_2$ is large. To see this, note that because $d=0$ the numerator of $Q(\beta^*)$ can be written as $z_u'M_{\lambda+z_v}z_u$, where $M_{\lambda+z_v} = I_{K_2} - (\lambda+z_v)[(\lambda+z_v)'(\lambda+z_v)]^{-1}(\lambda+z_v)'$. When $\lambda'\lambda/K_2$ is large, $M_{\lambda+z_v} \doteq M_\lambda$ and $\beta_d^* = \beta^* \doteq 0$, so the denominator of $Q(\beta^*)$ approaches one and $Q(\beta^*) \doteq z_u'M_\lambda z_u$, which has a $\chi^2_{K_2-1}$ distribution since z_u is an K_2-dimensional standard normal.

Under the local alternative, only the limiting distribution of $\hat{\beta}_{TSLS}$ changes, namely $\hat{\beta}_{TSLS} - \beta => \beta_d^*$. The expression for the limit of $\hat{\beta}_{TSLS}$ in theorem 2(a) elucidates the bias of the TSLS estimator when there are small violations of the orthogonality restrictions. If $\lambda'\lambda = g'V_{ZZ}g/\sigma_{vv}$ is large and $g'V_{ZZ}d/\sigma_{vv}$ is small, then these small violations impart negligible bias although they increase the spread of the distribution (because of the term $z_v'V_{ZZ}^{\frac{1}{2}}d$). In the completely unidentified case, the presence of nonzero d neither reduces nor increases the bias, but it does increase the spread of the distribution.

We next turn to the behavior of the DWH test for endogeneity. There is a large literature on this test, its computation and interpretation; see Davidson and MacKinnon (1993, ch. 7.9) for a discussion and references. In the case at hand of a single included endogenous variable, the test statistic is,

$$t_{DWH} = (\hat{\beta}_{TSLS}-\hat{\beta}_{OLS})/[\hat{var}(\hat{\beta}_{TSLS})-\hat{var}(\hat{\beta}_{OLS})]^{1/2} \tag{2.17}$$

where $\hat{var}(\hat{\beta}_{TSLS})$ and $\hat{var}(\hat{\beta}_{OLS})$ are respectively the standard estimators of the variances of $\hat{\beta}_{TSLS}$ and $\hat{\beta}_{OLS}$. The DWH statistic tests the null that $\rho=0$, that is, that Y is exogenous, against the alternative that $\rho\neq0$.

The limiting representation of t_{DWH} for general ρ obtains as a consequence of theorem 1(a) and (b) and appendix lemma A1(I). Under assumptions M and L_π, $\hat{\beta}_{OLS} \xrightarrow{p} \beta_0+\theta$ and $\hat{var}(\hat{\beta}_{OLS}) \xrightarrow{p} 0$, so $t_{DWH} = (\hat{\beta}_{TSLS}-\beta_0-\theta)/\hat{var}(\hat{\beta}_{TSLS})^{1/2} + o_p(1)$. Thus,

$$t_{DWH} \Rightarrow (\lambda+z_v)'[(1-\rho^2)^{1/2}\eta-\rho\lambda]/\{\nu_3[1-2\rho(\sigma_{vv}/\sigma_{uu})^{1/2}\beta^*+(\sigma_{vv}/\sigma_{uu})\beta^{*2}]\}^{1/2}. \tag{2.18}$$

Under the null hypothesis $\rho=0$, this simplifies to,

$$t_{DWH} \Rightarrow \nu_2/\{\nu_3[1+(\nu_2/\nu_3)^2]\}^{1/2}. \tag{2.19}$$

If $\lambda'\lambda/K_2$ is large, then ν_2/ν_3 is approximately $N(0, 1/\lambda'\lambda)$ and $\nu_3/\lambda'\lambda \doteq 1$, so t_{DWH} is approximately N(0,1) under the null. In general, however, the null distribution of t_{DWH} is nonnormal and depends on $\lambda'\lambda/K_2$. This is perhaps unsurprising in light of the nonnormality of $\hat{\beta}_{TSLS}$ and its t-statistic and in any event implies that conventional critical values are inappropriate for the DWH test unless $\lambda'\lambda/K_2$ is large.

The limit (2.18) applies under nonlocal alternatives ρ and thus can be used to obtain asymptotic power functions of the test. Because t_{DWH} is $O_p(1)$ for general ρ, the test is not consistent. For general ρ and $\lambda'\lambda/K_2$, the power function based on (2.18) must be computed numerically and this is not undertaken here. However, some intuition can be gained by examining the leading case of $\lambda=0$. Substituting $\lambda=0$ into (2.18), one finds that for general ρ, $t_{DWH} \Rightarrow z_v'\eta/\{(z_v'z_v)[1+(z_v'\eta/z_v'z_v)^2]\}^{1/2}$. This does not depend on ρ and is nonnormal. Thus, in the leading case the DWH test has power equal to its size against all ρ, that is, the DWH test is uninformative about endogeneity, but its size is not equal to its level if standard normal critical values are used.

D. *Asymptotics for alternative estimators and test statistics.*

The twin problems of biased estimation and the invalidity of conventionally constructed confidence intervals for TSLS suggest exploring alternative methods for inference. This subsection explores two alternatives which might fruitfully be applied with weak instruments.

The first alternative concerns estimation. Various estimators of β with reduced bias have been previously proposed; see for example Sawa (1973a, 1973b), Morimune (1978), and for a review, see Phillips (1983, section 3.10). The asymptotic representation in theorem 1 suggests another estimator which is a linear combination of the OLS and TSLS estimator. Inspection of the limiting representation for $\hat\beta_{TSLS}$ indicates that the bias in the $\hat\beta_{TSLS}$ estimator, conditional on z_v, is $\theta\nu_1/\nu_3$, that is, the bias is proportional to the bias in the OLS estimator. This suggests that using a linear combination of the OLS and TSLS estimators might reduce bias. From (2.3) and the expression in theorem 1(a), we have,

$$\hat\beta_{TSLS} - (\nu_1/\nu_3)\hat\beta_{OLS} = [1-(\nu_1/\nu_3)]\beta_0 + \kappa\nu_2/\nu_3 + o_p(1)\,. \tag{2.20}$$

Rearranging (2.20) and defining $\tilde b_N = 1/(1-\nu_1/\nu_3)$ yields,

$$\tilde{b}_N\hat{\beta}_{TSLS} + (1-\tilde{b}_N)\hat{\beta}_{OLS} = \beta_0 + \kappa\tilde{b}_N^{-1}\nu_2/\nu_3 + o_p(1). \tag{2.21}$$

Conditional on z_v, $E\tilde{b}_N^{-1}\nu_2/\nu_3 = 0$, so the right hand side of (2.21) has asymptotic mean β_0.

Although $\nu_3/K_2 = F_N + o_p(1)$, ν_1 is unobservable so (2.21) cannot be used as an estimator. However, when K_2 is large and λ is fixed, $\nu_1/K_2 = z_v'z_v/K_2 + \lambda'z_v/K_2 \doteq 1 + o_p(K_2^{-1/2})$. For $\lambda'\lambda/K_2$ moderate and K_2 large, the stochastic error in this approximation is likely to be small. This suggests the approximation, $\tilde{b}_T \doteq 1/(1 - 1/F_N)$, which leads to the modified estimator,

$$\hat{\beta}_m = b_T\hat{\beta}_{TSLS} + (1-b_T)\hat{\beta}_{OLS}, \text{ where } b_T = F_N/(F_N - 1). \tag{2.22}$$

If $F_N > 1$ then $b_T > 1$ and $\hat{\beta}_m$ is not a convex combination of TSLS and OLS but rather is further from OLS than TSLS. This estimator has the advantage of involving only standard regression output. The limiting representation of $\hat{\beta}_m$ is readily obtained from theorem 1:

$$\hat{\beta}_m - \beta_0 => \theta(\nu_1-K_2)/(\nu_3-K_2) + \kappa\nu_2/(\nu_3-K_2). \tag{2.23}$$

The presence of ν_3-K_2 in the denominator of the second term in (2.23), rather than ν_3 as appears in the limiting representation for $\hat{\beta}_{TSLS}$, suggests that the variance of $\hat{\beta}_m$ will exceed that of $\hat{\beta}_{TSLS}$, particularly for first-stage F statistics near 1. However, for large K_2, in this case the representation (2.23) suggests that there will be a bias reduction. Numerical evaluation of (2.23) is needed to ascertain whether the bias reduction offsets the variance increase and yields a reduction in mean squared error.

We next turn to the construction of asymptotically valid confidence intervals. Anderson and Rubin (1949) suggested testing the null hypothesis $\beta = \beta_0$ using the statistic,

$$A_N(\beta_0) = \{(y^{\perp}-Y^{\perp}\beta_0)'P_{Z^{\perp}}(y^{\perp}-Y^{\perp}\beta_0)/K_2\}/\{(y^{\perp}-Y^{\perp}\beta_0)'M_{Z^{\perp}}(y^{\perp}-Y^{\perp}\beta_0)/(N-K_1-K_2)\} \tag{2.24}$$

It should be emphasized that it is maintained that Z satisfies the instrument orthogonality conditions, that is, $\omega=0$ in (2.16). If (u_i, v_i) are i.i.d. $N(0, \Omega)$ and Z is strictly exogenous, then under the null $A_N(\beta_0)$ has an exact $F_{K_2,N-K_1-K_2}$ distribution, which has a $\chi^2_{K_2}/K_2$ limit as N gets large. With nonnormal errors and/or instruments which are valid but not strictly exogenous, this result obtains asymptotically, as stated in theorem 3:

Theorem 3. Suppose that assumptions L_π and M hold.

(a) Under the null hypothesis $\beta = \beta_0$, $A_N(\beta_0) => \chi^2_{K_2}/K_2$.

(b) Under the fixed alternative hypothesis $\beta = \beta_1$, $A_N(\beta_0) => Q(\beta_0-\beta_1)/K_2$, where $Q(\bullet)$ is defined in theorem 2.

Theorem 3(a) shows that, as pointed out in the finite sample case by Anderson and Rubin (1949) and discussed by Phillips (1983), confidence intervals can be constructed as the set of points $\{\beta_0\}$ for which $A_N(\beta_0)$ fails to reject, using the asymptotic $\chi^2_{K_2}/K_2$ critical values. (In practice, it might be desirable to use the more conservative $F_{K_2,N-K_1-K_2}$ critical values, so that the resulting confidence set will have a somewhat higher coverage rate.) However, this raises conceptual issues which are discussed in section 5.

Under the fixed alternative, the AR statistic has a noncentral $\chi^2_{K_2}/K_2$ distribution. To see this, let $\upsilon = [z_u-(\sigma_{vv}/\sigma_{uu})^{1/2}(\beta_0-\beta_1)z_v]/[1-2(\sigma_{vv}/\sigma_{uu})^{1/2}\rho(\beta_0-\beta_1)+(\sigma_{vv}/\sigma_{uu})(\beta_0-\beta_1)^2]^{1/2}$, so that υ is an K_2-dimensional standard normal variable. Then $Q(\beta_0-\beta_1) = (\upsilon-a)'(\upsilon-a)$ which has a noncentral $\chi^2_{K_2}$ distribution where $a'a = \{(\sigma_{vv}/\sigma_{uu})(\beta_0-\beta_1)^2/[1-2(\sigma_{vv}/\sigma_{uu})^{1/2}\rho(\beta_0-\beta_1)+(\sigma_{vv}/\sigma_{uu})(\beta_0-\beta_1)^2]\}(\lambda'\lambda)$ is the noncentrality parameter. Note that $\lim_{|\beta_0-\beta_1|\to\infty} a'a = \lambda'\lambda$, so that the probability of rejecting very distant alternatives depends not on the alternative or the sample size but rather only on $\lambda'\lambda$. An implication is that tests based on the AR statistic are not consistent. This accords with the failure of $\hat{\beta}_{TSLS}$ to concentrate on a decreasing region.

When the number of instruments is large, the AR statistic involves projections onto a high-dimensional subspace which could result in reduced power and thus wide confidence intervals (but see the Monte Carlo study of Maddala (1974)). One approach to this problem is therefore to construct a "split-sample" AR statistic, in which the projection is on a smaller subspace formed by a linear combination estimated from the first-stage regression run using a separate subsample[5]. Let subscripts "1" and "2" denote terms or data from a first and second randomly chosen subsample of sizes N_1 and N_2, respectively, where $N_1+N_2=N$ and N_1/N tends to a constant limit in (0,1). The split-sample AR statistic is,

$$(2.25)\qquad A_{N_1,N_2}(\beta_0) = \{(y_2^{\perp}-Y_2^{\perp}\beta_0)'P_{\tilde{Y}_2^{\perp}}(y_2^{\perp}-Y_2^{\perp}\beta_0)/K_2\}/\{(y_2^{\perp}-Y_2^{\perp}\beta_0)'M_{\tilde{Y}_2^{\perp}}(y_2^{\perp}-Y_2^{\perp}\beta_0)/(N_2-K_1-K_2)\},$$

where $P_{\tilde{Y}_2^{\perp}} = \tilde{Y}_2^{\perp}(\tilde{Y}_2^{\perp\prime}\tilde{Y}_2^{\perp})^{-1}\tilde{Y}_2^{\perp\prime}$ and $M_{\tilde{Y}_2^{\perp}} = I-P_{\tilde{Y}_2^{\perp}}$, where $\tilde{Y}_2^{\perp} = Z_2^{\perp}\hat{\pi}_1$ and $\hat{\pi}_1$ is the OLS estimator of π from the first subsample. If the data are independently distributed, then straightforward modifications of the proof of theorem 3a show that, conditional on the first subsample, $A_{N_1,N_2}(\beta_0) \Rightarrow \chi_1^2$ under the null that $\beta=\beta_0$; this limit does not depend on first-subsample data and thus is the unconditional distribution as well. It follows that the split-sample AR statistic (2.25) can be inverted to construct confidence intervals for β_0 in the same way as the AR statistic (2.24).

E. Asymptotic Distribution of LIML

The k-class estimator of the vector of parameters in the equation of interest, $\Theta = (\beta\ \gamma')'$, is,

$$(2.26)\qquad \hat{\Theta}(k) = \{W'(I-kM_{Z^*})W\}^{-1}\{W'(I-kM_{Z^*})y\}$$

where $W = (Y\ \ X)$ and $Z^* = (X\ \ Z)$. When $k=1$, this is of course the TSLS estimator. The LIML estimator is given by (2.26) with $k = k_{LIML} = 1 + \hat{\zeta}_N$, where $\hat{\zeta}_N$ is the smallest root of $|G - \zeta C| = 0$, where $G = V'(P_{Z^*} - P_X)V$ and $C = V'M_{Z^*}V$, where $V = (y\ \ Y)$. By projection arguments, the k-class estimator of β and a standard formula for its t-statistic, t(k), are

(2.27a) $$\hat{\beta}(k) = \{Y^{\perp\prime}(I-kM_{Z^\perp})Y^\perp\}^{-1}\{Y^{\perp\prime}(I-kM_{Z^\perp})y^\perp\}$$

(2.27b) $$t(k) = [\hat{\beta}(k)-\beta_0]/\{[Y^{\perp\prime}(I-kM_{Z^\perp})Y^\perp]^{-1}\hat{\sigma}_{uu}(k)\}^{1/2}$$

where $\hat{\sigma}_{uu}(k) = \hat{u}(k)'\hat{u}(k)/(N-K_1-1)$, where $\hat{u}(k) = y^\perp - Y^\perp\hat{\beta}(k)$. The LIML estimator, $\hat{\beta}_{LIML}$, is $\hat{\beta}(k_{LIML})$. For $k=1$, $t(k) = t_{TSLS}$ as defined in (2.9b). For LIML, the formula (2.27b) can be justified using fixed-π asymptotics and is used in applications (for example, it is implemented in TSP version 4.2 (Hall, Cummins and Schake (1992, p. 145-6)).

The limiting representation of $\hat{\beta}_{LIML}$, given in the next theorem, is obtained by expressing k_{LIML} as $k_{LIML} = 1+\zeta_N/N$, so $\zeta_N = N\hat{\zeta}_N$.

Theorem 4. Suppose that assumptions L_π and M hold. Then:

(a) $\zeta_N \Rightarrow \zeta^*$, where ζ^* is the smallest root of $|G^* - \zeta\bar{\Omega}| = 0$, where $G^* = (z_u\ \ (\lambda+z_v))'(z_u\ \ (\lambda+z_v))$ and $\bar{\Omega}$ is the 2×2 matrix with $\bar{\Omega}_{11}=\bar{\Omega}_{22}=1$ and $\bar{\Omega}_{12}=\bar{\Omega}_{21}=\rho$.

(b) $\hat{\beta}_{LIML} - \beta_0 \Rightarrow \Delta^*(\zeta^*)$, where $\Delta^*(\zeta) = (\theta\nu_1+\kappa\nu_2-\theta\zeta)/(\nu_3-\zeta)$

(c) $t_{LIML} \Rightarrow (\nu_4-\rho\zeta^*)/\{(\nu_3-\zeta^*)[1-2\rho(\sigma_{vv}/\sigma_{uu})^{1/2}\Delta^*(\zeta^*)+(\sigma_{vv}/\sigma_{uu})\Delta^*(\zeta^*)^2]\}^{1/2}$,

where ν_1, ν_2, ν_3, and ν_4 are defined in theorem 1.

The limiting marginal distribution of $\zeta_N = N(k_{LIML} - 1)$ depends on only the number of instruments and $\lambda'\lambda/K_2$. To see this, note that the zeros of $|G^* - \zeta\bar{\Omega}|$ are the same as the zeros of $|\bar{\Omega}^{-1/2}G^*\bar{\Omega}^{-1/2\prime}-\zeta I|$. Now $(z_u'\ \ z_v')'$ is distributed $N(0, \bar{\Omega}\otimes I_{K_2})$, so $\bar{\Omega}^{-1/2}(z_u\ \ (\lambda+z_v))'$ has the same distribution as $(\eta\ \ (\lambda+z_v))'$, where η and z_v are independent $N(0, I_{K_2})$ random variables. It follows that the distribution of ζ^* is the distribution of the smallest eigenvalue of $\tilde{G}^* = (\eta\ \ (\lambda+z_v))'(\eta\ \ (\lambda+z_v))$, the distribution of which depends only on $\lambda'\lambda/K_2$ and K_2. Also, by arguments similar to those given for TSLS following theorem 1, $(\hat{\beta}_{LIML}-\beta_0)/\theta$ and t_{LIML} have distributions which depend on only $\lambda'\lambda/K_2$ and K_2, and ρ: the limiting distribution of $(\hat{\beta}_{LIML}-\beta_0)/\theta$ depends on $|\rho|$, while the limiting distribution of t_{LIML} is symmetric in ρ. However, the distribution of $\hat{\beta}_{LIML}-\beta_0$ depends on the ratio σ_{uu}/σ_{vv} as well.

This representation provides a further approximation to $\hat{\beta}_{LIML}$ when the number of instruments is large. In this case, from the representation in the previous paragraph, $\bar{G}^*/K_2 = \mathrm{diag}(1, 1+\lambda'\lambda/K_2) + O_p(K_2^{-1/2})$, so the minimum eigenvalue ζ^* is approximately $\zeta^*/K_2 = 1 + O_p(K_2^{-1/2})$ so $\zeta^* \doteq K_2$. Evidently, for more than one instrument, the LIML and TSLS estimators are not asymptotically equivalent in this setting, and the approximately linear dependence of ζ^* on K_2 suggests that the divergence between LIML and TSLS increases with the number of instruments, holding $\lambda'\lambda$ fixed.

The results in theorem 4 are readily extended to modified LIML estimators. As an example, consider Fuller's (1977) estimator, which is given by $\hat{\beta}(k_F)$ where $k_F = 1 + \hat{\zeta} - \ell/(N-K_1-K_2) = 1 + \zeta_{F,N}/N$, where $\zeta_{F,N} = \zeta_N - \ell N/(N-K_1-K_2)$ (see the discussion in Morimune (1983)). For fixed K_1 and K_2, this has the limit, $\zeta_{F,N} => \zeta^* - \ell$, so $\hat{\beta}(k_F)-\beta_0 => \Delta^*(\zeta^*-\ell)$. Under conventional asymptotic assumptions, this estimator with $\ell = 1$ is mean unbiased to order $O(T^{-1})$. However, these assumptions do not hold here, and it is an open question whether this estimator improves upon LIML under the current assumptions.

F. Distribution of k-class estimator of the coefficients on exogenous variables

The coefficients γ on the exogenous regressors are often of as great interest as β. This subsection provides an asymptotic representation of the k-class estimator of γ, $\hat{\gamma}(k)$, when $k = 1 + \zeta/N$, which specializes to representations for the TSLS and LIML estimators of γ.

The case studied here is when X enters the reduced form equation for Y weakly; specifically, it is assumed that ϕ in (2.2) is local to zero:

Assumption L_ϕ: $\phi = \phi_N = N^{-1/2}f$, where f is a fixed $K_1 \times 1$ vector.

One motivation for this assumption is that if ϕ is fixed but π is local to zero, then asymptotically $\hat{Y}$ and X are multicollinear and γ is nearly underidentified. In the fixed-ϕ case, the regressor moment matrix is asymptotically singular and the identified and nearly unidentified linear combinations need to be treated separately, as done by Phillips (1989) in the exactly underidentified ($\pi = 0$) case. In contrast, letting ϕ be local to zero permits a single treatment of the various linear combinations.

The k-class estimator of γ is, from the matrix inversion formula and (2.26),

$$\text{(2.28)}\qquad \hat{\gamma}(k) = \{X'X - X'Y[Y'(I-kM_{Z*})Y]^{-1}Y'X\}^{-1}\{X'y - X'Y[Y'(I-kM_{Z*})Y]^{-1}Y'(I-kM_{Z*})y\}.$$

The asymptotic representation of $\hat{\gamma}(k)$ is given in the next theorem:

Theorem 5. Suppose that assumptions L_π, L_ϕ, and M hold, and that $k = 1+\zeta/N$. Then,

$$N^{1/2}(\hat{\gamma}(k)-\gamma) \Rightarrow \sigma_{uu}^{1/2}\Sigma_{XX}^{-1/2}(z_{Xu} - Hz_{Zu} + \rho J\zeta)$$

where

$$H = C^{-1}(\mu_X+z_{Xv})(\mu_Z+z_{Zv})'/(D-\zeta)$$
$$J = C^{-1}(\mu_X+z_{Xv})/(D-\zeta)$$
$$C = I - (\mu_X+z_{Xv})(\mu_X+z_{Xv})'/(D-\zeta)$$
$$D = (\mu_X+z_{Xv})'(\mu_X+z_{Xv}) + (\mu_Z+z_{Zv})'(\mu_Z+z_{Zv})$$
$$\mu_X = \Sigma_{XX}^{-1/2}\Sigma_{XZ}g/\sqrt{\sigma_{vv}} + \Sigma_{XX}^{1/2}f/\sqrt{\sigma_{vv}}$$
$$\mu_Z = V_{ZZ}^{1/2}g/\sqrt{\sigma_{vv}}$$

and $(z_{Xu}'\ z_u'\ z_{Xv}'\ z_v')'$ is distributed $N(0, \bar{\Omega}\otimes I_{K_1+K_2})$.

If estimation is by TSLS, $\zeta = 0$, while if estimation is by LIML, then $\zeta \Rightarrow \zeta^*$, the representation of which is given in theorem 4(a). The limiting random variables z_u and z_v are the same as in the previous subsections, so the representation in theorem 5 combined with the previous results provides joint representations of $\hat{\beta}_{TSLS}$, $\hat{\gamma}_{TSLS}$, $\hat{\beta}_{LIML}$, and $\hat{\gamma}_{LIML}$.

Although the expression in theorem 5 is complicated, some general observations can be made. Most importantly, under these assumptions both the TSLS and LIML estimators of γ are consistent but their asymptotic distributions are nonstandard. As in the previous subsections, write $z_u = \rho z_v + (1-\rho^2)^{1/2}\eta$ and additionally let $z_{Xu} = \rho z_{Xv} + (1-\rho^2)^{1/2}\eta_X$, so that $(z_{Xv}'\ z_v'\ \eta_X'\ \eta')'$ is distributed $N(0, I_{2(K_1+K_2)})$. Then the representation in theorem 5 can be rewritten,

$$(2.29) \qquad N^{1/2}(\hat{\gamma}(k)-\gamma) => \rho\sigma_{uu}^{1/2}\Sigma_{XX}^{-1/2}(z_{Xv} - Hz_v + \rho J\zeta) + (1-\rho^2)^{1/2}\sigma_{uu}^{1/2}\Sigma_{XX}^{-1/2}(\eta_X - H\eta).$$

Because (η, η_X) is independent of (z_v, z_{Xv}), $N^{1/2}(\hat{\gamma}(k)-\gamma)$ is, conditionally on (z_v, z_{Xv}), asymptotically normal with mean $\rho\sigma_{uu}^{1/2}\Sigma_{XX}^{-1/2}(z_{Xv} - Hz_v + \rho J\zeta)$ and variance $(1-\rho^2)\sigma_{uu}(\Sigma_{XX}^{-1} + \Sigma_{XX}^{-1/2}H\Sigma_{XX}^{-1/2}{}')$, so that unconditionally $\hat{\gamma}(k)$ has an asymptotic mixed normal distribution. Phillips (1989) and Choi and Phillips (1992) obtained mixed normal limiting distributions for $\hat{\gamma}_{TSLS}$ when the parameters are fixed rather than local-to-zero and when some coefficients are exactly unidentified while others are identified; in our notation, this corresponds to $\pi=0$ and ϕ fixed and nonzero. Their mixed normal distributions differ from that here, however; in particular the distribution here depends on the local parameters f and g. This dependence on the local parameters f and g poses a problem for inference: the distribution of $\hat{\gamma}$ depends on the extent to which *both* Z and X enter the reduced form equation for Y, and since neither f nor g are consistently estimable, the distribution in theorem 5 cannot be consistently estimated using sample statistics.

The representation in theorem 5 simplifies in some special cases. As an example, suppose g is large. Then $C \doteq I - \mu_X\mu_X'/(\mu_X'\mu_X+\mu_Z'\mu_Z)$ which is nonrandom and $\hat{\gamma}$ has the standard asymptotic normal distribution found in the usual $\pi\neq0$, $\phi\neq0$ case. However, if f is large but g is not, then $C \doteq I - \mu_X\mu_X'/\mu_X'\mu_X$, which is singular and the limiting distribution in (2.29) is degenerate, in particular is a mixture-of-normals distribution with a conditional covariance matrix which has rank equal to K_1-1. This corresponds to the case in which one linear combination of X, namely $X\phi$, is asymptotically perfectly multicollinear with $\hat{Y}$, so that the linear combination of coefficients associated with $X\phi$ is poorly estimated.

3. Numerical Results

A. Monte Carlo Comparison of Asymptotic and Finite-Sample Distributions

Several Monte Carlo experiments were performed to examine the quality of the preceding asymptotic approximations to the finite-sample distributions of $\hat{\beta}_{TSLS}$, t_{TSLS}, $\hat{\beta}_{LIML}$, t_{LIML}, and $A_N(\beta_0)$. Two designs were considered. The first reflects time series applications where the number of instruments is small and the instruments are stochastic. The second design is motivated by cross-sectional applications with a large number of binary instruments as in Angrist and Krueger (1991). The second design also uses nonnormal errors. In both designs, the data are generated according to (2.1) and (2.2) with $\sigma_{uu}=\sigma_{vv}=1$ and with X being a vector of ones, so that the only included exogenous variable is a constant.

In the first design the errors and instruments were drawn according to,

(3.1) Design I: Z_i i.i.d. $N(0, I_{K_2})$, (u_i, v_i) i.i.d. $N(0, \bar{\Omega})$.

Results are reported for $K_2=1$ and $K_2=4$ and, for each K_2, for $\rho=.5$ and $\rho=.99$.

In the second design,

(3.2) Design II: $Z_i = 1_{ji}$, $(u_i, v_i) = (\xi_{1i}, \rho\xi_{1i}+(1-\rho^2)^{1/2}\xi_{2i})$, where (ξ_{1i}, ξ_{2i}) are i.i.d. and mutually independent with distribution $(\chi_1^2-1)/\sqrt{2}$,

where 1_{ji} is an indicator variable if observation j is in cell i, where $j = 1, \ldots, K_2+1$ and the final cell was omitted. An equal number of observations were drawn from each cell (up to integer constraints), so each cell has approximately $N/(K_2+1)$ observations. Results are reported for $K_2=4$ and $K_2=100$ and, for each K_2, for $\rho=.2$ and $\rho=.5$.

In both designs, the true value of β is taken to be zero, which is done without loss of generality by interpreting the results as pertaining to $\hat{\beta}-\beta_0$. All results are based on 20,000

Monte Carlo replications. The asymptotic distributions were computed using 20,000 draws of the random variates appearing in the limiting representations in theorems 1 and 4.[6]

The results for the first design are summarized in figures 1 - 3 and tables 1 - 5. In the just-identified ($K_2=1$) case in which LIML and TSLS are equivalent, the asymptotic and finite-sample pdf's of $\hat{\beta}_{TSLS}$, computed using a kernel density estimator (Gaussian kernel), are plotted in figure 1 for $\rho=.5$ and $\rho=.99$, in each case for $\lambda'\lambda/K_2=.25$ and 1.0, for a sample size of 20. The case $\rho = .99$ and $\lambda'\lambda/K_2=.25$ is close to one of the cases examined by Jeong and Maddala (1992) (they used $K_2=1$, $N=20$, $\rho=.99$, and $\lambda'\lambda/K_2=.2$), and the pdf's in figure 1(c) exhibit the same bimodality as theirs. As the concentration parameter increases, the upper mode gets smaller and moves to the right, and for smaller correlations it disappears. In each of the cases examined, the asymptotic distribution provides a good approximation to the finite-sample distributions: the asymptotic distributions are bimodal when the finite sample distributions are, and often the differences between the two distributions are almost indistinguishable at the level of detail of the plot.

The approximation continues to be good for the case of 4 instruments and 80 observations (20 observations/instrument) plotted in figure 2 for TSLS and in figure 3 for LIML. When $\lambda'\lambda/K_2 \neq 0$, the LIML distribution is better centered (LIML is approximately median unbiased) but is more dispersed than TSLS. This feature of LIML has been noted in the finite-sample literature (e.g. Anderson (1982) and Mariano (1982)). The asymptotic distribution continues to be bimodal for LIML but not TSLS in the overidentified case with $\rho=.99$.

Quantitative comparisons of the asymptotic and finite-sample distributions for this design are given in tables 1 and 2 for $\hat{\beta}_{TSLS}$ and t_{TSLS} and tables 3 and 4 for $\hat{\beta}_{LIML}$ and t_{LIML}, respectively. The entries are the finite-sample cdf, evaluated at selected quantiles of the asymptotic distribution. In addition to the cases in figures 1 - 3, the tables report results for smaller sample sizes and other values of $\lambda'\lambda/K_2$. The results indicate that even for as few as 5 observations per instrument the asymptotic theory provides a good approximation to the sampling distribution for $\hat{\beta}_{TSLS}$ and $\hat{\beta}_{LIML}$: the maximum absolute difference between the asymptotic cdf and the finite-sample cdf at the evalution points in tables 1 and 3 is .11 for $N/K_2=5$, and is .03 for $N/K_2=20$. The asymptotic approximations to the distribution of the t-statistic for TSLS and LIML are typically less good for $N/K_2=5$, but are typically within .02 for $N/K_2=20$.

Table 5 summarizes Monte Carlo coverage rates for conventional 95% TSLS and LIML confidence intervals, constructed as the estimate ± 1.96 standard errors. When $\lambda'\lambda/K_2$ is small or when ρ is near one, the t-statistic has a skewed and heavy-tailed distribution, and the standard confidence intervals have coverage rates which can differ substantially from their purported levels, confirming the results of Nelson and Startz (1992b). For example, when $\rho = .99$, $\lambda'\lambda/K_2 = 0$, and $K_2 = 4$, only 1% of conventional "95%" TSLS confidence intervals contain the true value of β. LIML confidence intervals have much better coverage rates than TSLS confidence intervals for the cases with $K_2=4$ and $\lambda'\lambda/K_2 \neq 0$. Although when $\lambda'\lambda/K_2$ is small confidence intervals based on the TSLS or LIML t-statistic have incorrect coverage rates, both in finite samples and asymptotically (in the sense of section 2) the AR statistic has an exact $F_{K_2,N-K_2-1}$ distribution in this design. Thus confidence intervals formed by inverting the AR statistic will have coverage rates equal to the stated confidence level both in finite samples and asymptotically. In the simulations here, AR confidence intervals are constructed using the asymptotic $\chi^2_{K_2}/K_2$ distribution of $A_N(\beta_0)$ rather than its exact F distribution, so the actual coverage rates are less than 95%.

Results for design II are summarized in figures 4 - 7 for $K_2 = 100$ and $N/K_2 = 20$, and in tables 6 - 10 for additional values of K_2, N/K_2, and $\lambda'\lambda/K_2$. As in design I, the asymptotic approximations to the distribution of the estimators is good, both capturing the qualitative features of the pdf (figures 4 and 6) and having quantiles which are close to the finite-sample quantiles (tables 6 and 8). As the theoretical expressions suggest, the finite-sample distributions and their asymptotic approximations concentrate near the probability limit of the OLS estimator for $\lambda'\lambda/K_2$ small. As $\lambda'\lambda/K_2$ increases, the distribution of $\hat{\beta}_{TSLS}$ moves towards its true value, 0. As found in design I, when $\lambda'\lambda/K_2 \neq 0$ the distribution of the LIML estimator is more centrally located around its true value than is TSLS.

As in design I, the local-to-zero asymptotics also provide good approximations to the pdf (figures 5 and 7) and quantiles (tables 7 and 9) of the t-statistics, at least for $N/K_2 = 20$. The extent of the nonnormality of the distribution of t_{TSLS} is evident in figure 5 for $\lambda'\lambda/K_2$ ranging between 0 and 10, with $K_2 = 100$ and $\rho = .5$: the distribution is shifted and slightly skewed, even for $\lambda'\lambda/K_2 = 10$. Not surprisingly, this bias and nonnormality results in poor coverage rates for conventional TSLS "95%" confidence intervals, as documented in table 10. For example, with $K_2 = 100$, $\lambda'\lambda/K_2 = .25$, and $\rho = .5$, the coverage rate of the usual confidence interval is zero. In contrast, the distribution of t_{LIML} is typically more centered around zero and conventional LIML "95%" confidence intervals have better coverage rates than TSLS (table 10), even though t_{LIML} clearly can have a nonnormal distribution (figure 7). In this design, because the errors are drawn from a χ^2 distribution the AR statistic does not have an exact F distribution but has an asymptotic $\chi^2_{K_2}$ as shown in theorem 3. Nonetheless, in contrast to the conventional confidence intervals, confidence intervals constructed by inverting the AR statistic have coverage rates which are close to their asymptotic coverage level.

These results suggest that the asymptotic results in section 2 provide a good approximation to the finite-sample distributions in cases of interest, even with very few observations per instrument. As the number of observations per instrument increases, the quality of the approximation improves. In particular, the asymptotics provide good approximations to the cases highlighted by Nelson and Startz (1990a,b) and deliver the same qualitative implications. The asymptotics provide a good approximation to the many-instrument case studied with $\lambda'\lambda/K_2 = 0$ by Bound, Jaeger and Baker (1993), and more generally for positive values of $\lambda'\lambda/K_2$ as well.

B. Quantitative Summaries of the Asymptotic Distributions

Figures 8 - 11 plot summary measures of the asymptotic TSLS and LIML distributions and constitute the main numerical results of the paper. Figure 8 plots the ratio of the asymptotic TSLS bias to the OLS bias, $E\beta^*/\theta$, for values of K_2 ranging from 2 to 100 and values of $\lambda'\lambda/K_2$ ranging from 0 to 20. Figure 9 plots the coverage rates for TSLS 95% confidence intervals. (The case $K_2 = 1$ is dropped from figure 8, but not figures 9-11, because $\hat{\beta}_{TSLS}$ has no first moment for $K_2 = 1$.) The ratio of the median LIML bias to the OLS bias, median($\Delta^*(\zeta^*)/\theta$), and the coverage rates of conventional 95% LIML confidence intervals are respectively shown in figures 10 and 11. As discussed following theorems 1 and 4, these summary measures depend only on $|\rho|$, $\lambda'\lambda/K_2$, and K_2. Each figure includes separate graphs for $|\rho| = .2$, $|\rho| = .5$, $|\rho| = .75$ and $|\rho| = .99$.

From figure 8 it is apparent that the relative bias of TSLS is largely a function of $\lambda'\lambda/K_2$. A useful rule of thumb, related to the modified estimator proposed in equation (2.22), is that the ratio of the TSLS bias to the OLS bias is approximately $1/(1+\lambda'\lambda/K_2)$. Thus for $\lambda'\lambda/K_2 = 4$, roughly 20% of the OLS bias remains. It is also apparent from figure 10 that LIML is approximately median unbiased for nearly all cases in which $\lambda'\lambda/K_2 \geq .5$. Anderson (1982) and others have noted this feature of LIML in the Gaussian/fixed instrument case through numerical evaluation of finite-sample distributions. The ability of LIML to produce median-unbiased estimates even with very weak instruments ($\lambda'\lambda/K_2 \geq .5$) contrasts sharply with TSLS.

Coverage rates for TSLS confidence intervals are more sensitive to K_2 and $|\rho|$ as can be seen from figure 9. In particular, coverage rates generally fall as K_2 increases or $\lambda'\lambda/K_2$ decreases, and the rate at which the coverage rates fall increase with $|\rho|$. For example, when $|\rho| = .2$, coverage rates are near 95% for all K_2 once $\lambda'\lambda/K_2$ is greater than 10. In contrast, when $|\rho| = .99$, coverage rates approach 95% only for combinations of relatively large $\lambda'\lambda/K_2$ and small K_2. Thus the TSLS coverage rate is quite sensitive to the parameters of the model and low coverage rates can persist even when $\lambda'\lambda/K_2$ is large. Coverage rates for LIML confidence intervals are much less sensitive than TSLS to K_2 and $|\rho|$ (figure 11) and in an absolute sense the coverage rates can be considered fairly good: for $1 \leq K_2 \leq 100$, $\lambda'\lambda/K_2 \geq 1$, and $|\rho| = .2, .5, .75$, and .99, the asymptotic coverage rates lie between 81.4% and 99.8%. For $\lambda'\lambda/K_2 \geq 10$, the asymptotic coverage rates lie between 91.5% and 98%.

These results suggest five conclusions. First, the asymptotics of section 2 provide good approximations to the sampling distributions of the LIML and TSLS estimators and their t-statistics in a wide range of designs, which suggests that useful lessons for practice can be based on these asymptotic distributions. Second, TSLS can exhibit large relative bias, which is well approximated by $(1+\lambda'\lambda/K_2)^{-1}$. Third, coverage rates of TSLS confidence intervals deteriorate dramatically as $\lambda'\lambda/K_2$ decreases, especially for K_2 or $|\rho|$ large. Fourth, LIML is approximately median unbiased, even for $\lambda'\lambda/K_2$ as small as .5. Fifth, coverage rates of LIML confidence intervals are reasonably accurate and deteriorate seriously only for $\lambda'\lambda/K_2 < 1$.

4. Use of the Results in Empirical Applications

This work leads to some concrete quantitative guidelines for applications of instrumental variable regression with a single dependent variable in samples of the size typically found in modern econometric research, say with at least 20 observations per instrument. As discussed following theorems 1 and 4, the asymptotic distributions of the relative error of the TSLS and LIML estimators (respectively $(\hat{\beta}_{TSLS}-\beta_0)/\theta$ and $(\hat{\beta}_{LIML}-\beta_0)/\theta$) and of t_{TSLS} and t_{LIML} depend on only three parameters: the number of instruments K_2, $\lambda'\lambda/K_2$, and the correlation ρ. In a given application K_2 is of course known but to apply the distribution theory and figures 8 - 11 requires estimates of $\lambda'\lambda/K_2$ and ρ.

Empirical evidence on $\lambda'\lambda/K_2$ can be obtained from the F-statistic (F_N) from the first stage regression, testing the hypothesis that the coefficients on the instruments are zero. From theorem 1(d), F_N-1 is an asymptotically unbiased estimator of $\lambda'\lambda/K_2$. The researcher could go further and construct a confidence interval for $\lambda'\lambda/K_2$ by inverting the asymptotic noncentral $\chi^2_{K_2}$ distribution of K_2F_N. This might be warranted if K_2 is small (so the spread of the distribution of F_N is large) or if the figures suggest that conclusions about bias and coverage rate are sensitive to small changes in $\lambda'\lambda/K_2$ in the case at hand.

The correlation ρ is of course unknown but we suspect that, in many applications, a "ballpark" estimate of ρ can be obtained which will be suitable for applying the asymptotics. Hypothetical values of ρ can be deduced from $\hat{\beta}_{OLS}$ and various hypothetical values of β_0. Because $\hat{\beta}_{OLS}-\beta_0$

[... is] $\theta = \sigma_{uv}/\sigma_{vv}$, $\rho = \text{plim}(\hat{\beta}_{OLS}-\beta_0)(\sigma_{vv}/\sigma_{uu})^{1/2} = \theta(\sigma_{vv}/\sigma_{uu})^{1/2}$. The reduced-form error variance σ_{vv} is consistently estimated by $\hat{\sigma}_{vv}$, the squared standard error of the first stage regression. Given β_0, σ_{uu} is estimated consistently by the standard error of the regression of $y-Y\beta_0$ on X. This permits estimation of ρ as a function of β_0. Of course, β_0 is the main object of interest and is unknown, so this procedure does not provide an estimator of ρ in the usual sense; rather, the hypothetical values of β_0 typically will rely on *a-priori* economic reasoning. However, the purpose here is only to ascertain whether the standard distributions are reliable rather than to estimate ρ consistently. This argues for using a range of β_0 and thus a range of ρ to reach conclusions about distributions in the problem at hand.

Given K_2 and these estimates of ρ and $\lambda'\lambda/K_2$, the bias and coverage rates of the various LIML and TSLS statistics can be deduced by inspection of figures 8 - 11 or by simulation of the relevant limiting representations in section 2. If $\lambda'\lambda/K_2$ is sufficiently large, both TSLS and LIML bias and coverage rates might be judged satisfactory and inference can proceed in the conventional way. In other cases, TSLS inference might be unreliable but the figures will suggest that the LIML estimator and confidence intervals are more reliable; the figures indicate that, for $\lambda'\lambda/K_2 \geq 5$ and $.2 \leq |\rho| \leq .99$, LIML is approximately median unbiased and conventional 95% LIML confidence intervals have coverage rates of at least 90%. In the event that both TSLS and LIML estimators and confidence intervals have unacceptable bias and coverage rates, a theoretically valid alternative is to construct Anderson-Rubin (1949) confidence intervals, perhaps using the split-sample approach discussed in section 2D.

5. Application to the Returns to Education

This section interprets Angrist and Krueger's (1991) estimates of the returns to education in light of the foregoing distribution theory. The analysis builds on Angrist and Krueger's insight that the quarter of birth, and quarter of birth interacted with other covariates, can serve as instruments for education in an earnings equation. Quarter of birth may be a useful instrument, they reasoned, because it is randomly distributed across the population, yet affects educational attainment through a combination of the age at which a person begins school and the compulsory schooling laws in a person's state. However, in many cases their first-stage F-statistics are low, raising the possibility that inference based on standard asymptotics might be unreliable here.

We use Angrist and Krueger's (1991) data, which is drawn from the 5% Public Use Micro Sample of the 1980 U.S. Census. For details of construction, see Appendix 1 of Angrist and Krueger (1991).[7] The sample includes men born between 1930 and 1949 with positive earnings in 1979 and no missing data on any of the relevant variables. As in Angrist and Krueger, the sample is split into two ten year birth cohorts.

A. Results and General Discussion

Regression results are reported in table 11. The top panel contains results for men born in 1930-39, and the bottom panel contains results for the 1940-49 cohort. The first rows of each panel contain the results of estimating the effect of education on log weekly earnings by OLS and TSLS in four basic specifications. All specifications use dummy variables to control for race, SMSA, marital status, region, and year of birth. We further control for age (measured in quarter years) and age^2 in column three, and for state of birth in column four. Three quarter-of-birth dummies are used as instruments in column one. Columns two and three add (quarter-of-birth)×(year-of-birth) interactions to the instrument list (27 additional instruments), for a total of 30 instruments in column 2 and 28 instruments in column 3 (due to the inclusion of age and age^2). Finally, column four adds (quarter-of-birth)×(state-of-birth) interactions to the instrument list, for a total of 178 instruments.

The asymptotic results help to interpret several features of table 11. Consistent with the theoretical predictions, as the first-stage F falls and the number of instruments increases (specifications II, III, and IV), the TSLS point estimates move towards the OLS point estimates and the TSLS standard errors (spuriously) decrease. For the 1930-39 cohort, the LIML and the combined OLS-TSLS estimates are fairly stable across specifications I, II and IV, ranging from .084 to .102, and the estimates do not approach the OLS estimate as more instruments are added and F_N falls. The differences between the AR and LIML confidence intervals and those from TSLS are striking. When the first-stage F is large, as for specification I, the LIML and AR confidence intervals coincide closely with the usual asymptotic confidence interval from TSLS. However for the specifications in which the first-stage F is small (specifications II - IV), the LIML and especially the AR and split-sample AR confidence intervals are noticably larger than those given by TSLS.

The second panel of Table 11 contains estimation results for the 1940-49 cohort. Although the TSLS estimates are less stable across specifications for this cohort, they also approach the OLS estimate as the first-stage F falls. The LIML and combined OLS-TSLS estimator are also less stable. LIML confidence intervals remain larger than TSLS intervals. AR confidence intervals are null for specifications I, II and III, that is, there is no β that yields an $A_N(\beta)$ statistic which lies below the 95% critical value. This arises because the AR statistic also tests, and in this case rejects, the over-identifying restrictions. Basmann's (1960) test of the overidentifying restrictions, reported in the last row of each panel of Table 2, also rejects at the 5% level for these specifications in the 1940-49 cohort. (Note however that the Basmann test rejections, except perhaps for specification I, must be interpreted cautiously in light of theorem 2 and the subsequent discussion.) In contrast, the split-sample AR confidence intervals, which do not test the overidentifying restrictions, are nonempty and substantially larger than the TSLS confidence intervals for specifications II, III and IV. These results point to the instruments being invalid for this cohort.

This suggests that care must be taken in interpreting AR confidence intervals. These confidence intervals are the set of β for which neither $\beta = \beta_0$ nor the overidentifying restrictions are rejected. If the overidentifying restrictions are false, then spuriously short (or empty) confidence intervals on β arise, in the sense that their shortness arises not from precision in estimation of β but from the rejection of the overidentifying restrictions. To illustrate these issues, $A_N(\beta)$ is plotted as a function of β for each of the specifications given in Table 11 in figure 12 for the 1930-39 cohort and in figure 13 for the 1940-49 cohort; 95% critical values are marked with dotted lines in each figure. Note that only for specification I is $A_N(\beta)$ tightly concentrated. For the other specifications $A_N(\beta)$ is less concentrated about its minimum, and flattens out dramatically away from the minimum. Because $A_N(\beta)$ is a ratio of quadratics in β_0, it always has a maximum and asymptotes as $\beta \to \pm\infty$. This raises the possibility of disjoint confidence sets, which we found in Monte Carlo simulations but not in this empirical application.

Figure 13 illustrates the problems with the AR confidence intervals when the model is poorly specified. In specifications I, II and III, failure of the overidentifying restrictions increases $A_N(\beta)$ to the point where it always lies above the critical value. Thus we can reject the joint hypothesis that $\beta = \beta_0$ *and* the overidentifying restrictions hold. This leads to empty confidence intervals, that is, there is no value of β for which the joint hypothesis is not rejected. Were a confidence level more nearly equal to unity used, the AR intervals could be nonempty and tight. However, this would be misleading, since the tightness of the AR intervals in this case, and their emptiness at the 95% confidence level in fact reflects the rejection of one or more of the overidentifying restrictions.

B. Choice of Preferred Specification

The strategy of section 4 is now used to suggest which of the results in table 11 are most reliable. This entails making an educated guess about ρ (or a range of ρ), estimating $\lambda'\lambda/K_2$, and then using figures 8 - 11 to ascertain which if any of the various statistics are to be preferred. The rejection of the overidentifying restrictions and null AR confidence intervals for the 1940-49 cohort leads us to focus on the results for the 1930-39 cohort.

The value $\rho=-.2$ was chosen as a plausible value and was computed as follows. Recall that $\rho = \theta(\sigma_{vv}/\sigma_{uu})^{1/2}$. In specification I of table 11, the effect of a year of education on earnings is estimated as .063 (OLS) and .099 (TSLS). Taking the TSLS estimate as the true value, we have $\theta=-.036$. Combining this with the standard error of the first stage regression ($\hat{\sigma}^{1/2}_{vv}=3.18$) and of the second stage regression ($\hat{\sigma}^{1/2}_{uu}$=.63) produces an estimate of ρ=-.2. A more extreme assumption is that the true coefficient is .15; this would yield a value of ρ=-.5. Our conclusions about the reliability or unreliability of the estimators or t-statistics, and about the choice of preferred specification, are insensitive to the use of ρ=-.5. Alternatively, economic reasoning about omitted variable bias (innate ability is unobserved and omitted but is positively correlated with both education and earnings) suggests that $\theta>0$ and thus $\rho>0$. Because $\rho=.2$ and $\rho=.5$ correspond approximately to β_0 = .03 and β_0 = -.03 and because figures 8-11 depend only on $|\rho|$, the analysis here applies equally for these small or even negative values of β_0.

In specification I, F_N=30.5; according to figures 8 - 11 with K_2=3, $\lambda'\lambda/K_2$=29.5 and $|\rho|$=.2 or $|\rho|$=.5, both TSLS and LIML are effectively unbiased and their confidence intervals have coverage rates close to 95%. In specification II, for K_2=30, $\lambda'\lambda/K_2$=3.7, and $|\rho|$=.2 the LIML estimate is median unbiased, the TSLS bias is approximately 20% of the OLS bias and coverage rates for TSLS and LIML confidence intervals are 92% and 93% for $|\rho|$=.2, but less for TSLS if $|\rho|$=.5. We therefore rely on the LIML or AR results in specification II and put less weight on the TSLS results. In specification III, for K_2=28, $\lambda'\lambda/K_2$=.6, and $|\rho|$=.2, LIML remains approximately median unbiased, the TSLS bias grows to 63% of the OLS bias, and coverage rates of TSLS and LIML confidence intervals fall to 85% and 86%, although these coverage rates fall off sharply as $\lambda'\lambda/K_2$ drops and smaller values of $\lambda'\lambda/K_2$ are also consistent with the observed F of 1.6. Here, both TSLS and LIML confidence intervals are unreliable and AR intervals are preferred. Finally, for specification IV, with K_2=178, $\lambda'\lambda/K_2$=.9, and $|\rho|$=.2, we use the case in figures 8-11 with K_2=100. Here, LIML is median unbiased, TSLS bias is 50% of the OLS bias, and coverage rates for TSLS and LIML are 70% and 85%. Again TSLS is unreliable but LIML is arguably satisfactorily behaved, albeit with a confidence interval which is somewhat too tight. In specifications II, III and IV, values of $|\rho|\geq.5$ would accentuate the preference of AR over LIML and especially LIML over TSLS.

Using the estimators supported by the asymptotics, our estimates of the returns to education are reasonably stable across specifications, ranging from .084 to .010, with the exception of specification III in which the LIML estimator is very imprecise and TSLS is arguably badly biased. Among the AR confidence intervals, the tightest occurs in specification I with 3 instruments and is (.05, .15). Among those TSLS and LIML confidence intervals that we suspect to have at least 90% coverage rates, the tightest occurs for LIML in specification II with 30 instruments and is (.05, .12). Overall, this analysis confirms the main conclusion of Angrist and Krueger that OLS estimates are if anything biased downward. While our preferred point estimates of the returns to education are higher than theirs, our confidence intervals are wider than their unreliable TSLS intervals.

6. Conclusions

The approach developed in this paper reduces the analysis of distributions of instrumental variables statistics to a straightforward calculation followed by Monte Carlo simulation of the limiting representations, which involve few nuisance parameters and low (K_2) dimensional random variables. The resulting approximations seem to work well in moderately large sample sizes, say 20 observations per instrument. The numerical results confirm conclusions from the finite-sample literature such as those summarized by Mariano (1982), in particular the dependence of the bias of TSLS on the first-stage F and the relatively lower bias of LIML, and extend them to nonGaussian errors with stochastic instruments. Moreover, the results demonstrate that, because of the nonstandard distribution of the TSLS estimator, many statistics involving this estimator, such as the Basman test of overidentifying restrictions and the DWH test for endogeneity, have large-sample null distributions which differ from their conventional normal or χ^2 approximations. When the first stage F is small, not just the point estimates but almost all the statistics conventionally used in instrumental variables analysis have distributions which are suspect. However, figures 8 - 11 can be used in practice to ascertain when these problems are likely to arise.

Although the applications discussed here involve a single included endogenous variable, the approach can be extended to multiple included endogenous variables. The split-sample AR intervals (2.25) might result in shorter intervals than the full-sample AR intervals (2.24), but the extent to which gains are possible in the local-to-zero setting has not been investigated here. Finally, the joint asymptotic distribution of (say) (F_N, t_{TSLS}) depends only on $(\lambda'\lambda/K_2, \rho)$, given K_2; this raises the possibility of constructing bounds-type intervals for ρ either by Scheffe or Bonferroni methods. In any event, while this paper has suggested guidelines for recognizing when conventional inference might be misleading, the best alternative econometric strategy in that circumstance remains unresolved. These and related issues are the topic of ongoing research.

Appendix

Before proving theorems 1 and 2, we state and prove a lemma which collects various results about sample moments. Recall that $Z^{\perp} = M_X Z$, $Y^{\perp} = M_X Y$, etc., and $\hat{Y}^{\perp} = P_{Z^{\perp}} Y^{\perp}$.

Lemma A1. Suppose that (2.1), (2.2), and assumptions L_{π} and M hold. Then the following hold jointly:

(a) $u^{\perp\prime} u^{\perp}/N \xrightarrow{p} \sigma_{uu}$

(b) $Y^{\perp\prime} u^{\perp}/N \xrightarrow{p} \sigma_{uv}$

(c) $Y^{\perp\prime} Y^{\perp}/N \xrightarrow{p} \sigma_{vv}$

(d) $Z^{\perp\prime} Z^{\perp}/N \xrightarrow{p} \Sigma_{ZZ} - \Sigma_{ZX}\Sigma_{XX}^{-1}\Sigma_{XZ} = V_{ZZ}$

(e) $(Z^{\perp\prime} Z^{\perp})^{-1/2}(Z^{\perp\prime} u^{\perp}) \Rightarrow \sigma_{uu}^{1/2} z_u$

(f) $(Z^{\perp\prime} Z^{\perp})^{-1/2}(Z^{\perp\prime} v^{\perp}) \Rightarrow \sigma_{vv}^{1/2} z_v$

(g) $(Z^{\perp\prime} Z^{\perp})^{-1/2}(Z^{\perp\prime} Y^{\perp}) \Rightarrow \sigma_{vv}^{1/2}(\lambda + z_v)$

(h) $\hat{Y}^{\perp\prime} u^{\perp} \Rightarrow (\sigma_{vv}\sigma_{uu})^{1/2}(\lambda + z_v)' z_u$

(i) $\hat{Y}^{\perp\prime} \hat{Y}^{\perp} \Rightarrow \sigma_{vv}(\lambda + z_v)'(\lambda + z_v) = \sigma_{vv}\nu_3$,

where $(z_v' \;\; z_u')'$ is distributed $N(0, \bar{\Omega} \otimes I_{K_2})$, where $\bar{\Omega}$ is defined following (2.6).

<u>Proof</u>. All limits in this proof invoke assumption M and those involving π invoke assumption L_{π}.

(a) $u^{\perp\prime} u^{\perp}/N = u'u/N - (u'X/N)(X'X/N)^{-1}(X'u/N) = u'u/N + o_p(1) \xrightarrow{p} \sigma_{uu}$.

(b) $Y^{\perp\prime} u^{\perp}/N = Y' M_X u/N = \pi' Z' M_X u/N + v' M_X u/N = v'u/N + o_p(1) \xrightarrow{p} \sigma_{uv}$.

(c) $Y^{\perp\prime} Y^{\perp}/N = \pi' Z^{\perp\prime} Z^{\perp} \pi/N + 2\pi' Z^{\perp\prime} v^{\perp}/N + v^{\perp\prime} v^{\perp}/N$. Now $\pi' Z^{\perp\prime} Z^{\perp} \pi/N \xrightarrow{p} 0$. Also, because $\pi \to 0$, $X'v/N \xrightarrow{p} 0$, and $Z'v/N \xrightarrow{p} 0$, $\pi' Z^{\perp\prime} v^{\perp}/N = \pi'[Z'v/N - (Z'X/N)(X'X/N)^{-1}(X'v/N)] \xrightarrow{p} 0$ and $v^{\perp\prime} v^{\perp}/N \xrightarrow{p} \sigma_{vv}$, so $Y^{\perp\prime} Y^{\perp}/N \xrightarrow{p} \sigma_{vv}$.

(d) This follows directly from assumption M.

(e), (f) Let

$$\text{(A.1a)} \qquad z_u = \sigma_{uu}^{-1/2} V_{ZZ}^{-1/2}(\Psi_{Zu} - \Sigma_{ZX}\Sigma_{XX}^{-1}\Psi_{Xu})$$

$$\text{(A.1b)} \qquad z_v = \sigma_{vv}^{-1/2} V_{ZZ}^{-1/2}(\Psi_{Zv} - \Sigma_{ZX}\Sigma_{XX}^{-1}\Psi_{Xv})$$

$$\text{(A.1c)} \qquad \eta = (1-\rho^2)^{-1/2}(z_u - \rho z_v)$$

where $V_{ZZ} = \Sigma_{ZZ} - \Sigma_{ZX}\Sigma_{XX}^{-1}\Sigma_{XZ}$. Because $(\Psi_{Zu}'\ \Psi_{Xu}'\ \Psi_{Zv}'\ \Psi_{Xv}')' \sim N(0, \Omega\otimes\Sigma)$, direct calculation confirms that $(z_v', z_u') \sim N(0, \bar{\Omega}\otimes I_{K_2})$. Now,

$$N^{-1/2}Z^{\perp\prime}u^{\perp} = N^{-1/2}Z'u - (Z'X/N)(X'X/N)^{-1}(N^{-1/2}X'u) \Rightarrow \Psi_{Zu} - \Sigma_{ZX}\Sigma_{XX}^{-1}\Psi_{Xu}$$

$$N^{-1/2}Z^{\perp\prime}v^{\perp} = N^{-1/2}Z'v - (Z'X/N)(X'X/N)^{-1}(N^{-1/2}X'v) \Rightarrow \Psi_{Zv} - \Sigma_{ZX}\Sigma_{XX}^{-1}\Psi_{Xv}.$$

so

$$(Z^{\perp\prime}Z^{\perp}/N)^{-1/2}(N^{-1/2}Z^{\perp\prime}u^{\perp}) \Rightarrow V_{ZZ}^{-1/2}(\Psi_{Zu} - \Sigma_{ZX}\Sigma_{XX}^{-1}\Psi_{Xu}) = \sigma_{uu}^{1/2}z_u$$

$$(Z^{\perp\prime}Z^{\perp}/N)^{-1/2}(N^{-1/2}Z^{\perp\prime}v^{\perp}) \Rightarrow V_{ZZ}^{-1/2}(\Psi_{Zv} - \Sigma_{ZX}\Sigma_{XX}^{-1}\Psi_{Xv}) = \sigma_{vv}^{1/2}z_v.$$

(g) Set $\lambda = V_{ZZ}^{1/2}g/\sigma_{vv}^{1/2}$. Then by result (f) of this lemma,

$$(Z^{\perp\prime}Z^{\perp})^{-1/2}(Z^{\perp\prime}Y^{\perp}) = (Z^{\perp\prime}Z^{\perp}/N)^{1/2}(N^{1/2}\pi) + (Z^{\perp\prime}Z^{\perp}/N)^{-1/2}(N^{-1/2}Z^{\perp\prime}v^{\perp})$$

$$\Rightarrow V_{ZZ}^{1/2}g + \sigma_{vv}^{1/2}z_v = \sigma_{vv}^{1/2}(\lambda + z_v).$$

(h) Results (e) and (g) above yield,

$$\hat{Y}^{\perp\prime}u^{\perp} = (Y^{\perp\prime}Z^{\perp})(Z^{\perp\prime}Z^{\perp})^{-1}(Z^{\perp\prime}u^{\perp}) \Rightarrow (\sigma_{vv}\sigma_{uu})^{1/2}(\lambda+z_v)'z_u.$$

(i) Result (g) above yields,

$$\hat{Y}^{\perp\prime}\hat{Y}^{\perp} = (Y^{\perp\prime}Z^{\perp})(Z^{\perp\prime}Z^{\perp})^{-1}(Z^{\perp\prime}Y^{\perp}) \Rightarrow \sigma_{vv}(\lambda+z_v)'(\lambda+z_v) = \sigma_{vv}\nu_3. \quad \square$$

Proof of theorem 1

(a) From the definition of $\hat{\beta}_{TSLS}$, $\hat{\beta}_{TSLS} - \beta_0 = (\hat{Y}^{\perp\prime}\hat{Y}^{\perp})^{-1}(\hat{Y}^{\perp\prime}u^{\perp})$. By lemma A1, $\hat{Y}^{\perp\prime}u^{\perp} \Rightarrow (\sigma_{vv}\sigma_{uu})^{1/2}(\lambda+z_v)'z_u$. Let η be as defined in (A.1c). Then,

$$\begin{aligned}\hat{\beta}_{TSLS} - \beta_0 &\Rightarrow (\sigma_{uu}/\sigma_{vv})^{1/2}(\lambda+z_v)'[\rho z_v + (1-\rho^2)^{1/2}\eta]/\nu_3 \\ &= \theta\nu_1/\nu_3 + \kappa\nu_2/\nu_3\end{aligned}$$

since $\rho(\sigma_{uu}/\sigma_{vv})^{1/2} = \sigma_{uv}/\sigma_{vv} = \theta$ and $[\sigma_{uu}(1-\rho^2)/\sigma_{vv}]^{1/2} = \kappa$.

(b) Consider $\hat{\sigma}_{uu}/\sigma_{uu}$. From lemma A1 and part (a) of this theorem, we have:

$$\begin{aligned}\hat{\sigma}_{uu}/\sigma_{uu} &= (y^{\perp} - Y^{\perp}\hat{\beta}_{TSLS})'(y^{\perp} - Y^{\perp}\hat{\beta}_{TSLS})/(N-K_1-1)\sigma_{uu} \\ &= [u^{\perp} - Y^{\perp}(\hat{\beta}_{TSLS}-\beta_0)]'[u^{\perp} - Y^{\perp}(\hat{\beta}_{TSLS}-\beta_0)]/(N-K_1-1)\sigma_{uu} \\ &= [u^{\perp\prime}u^{\perp} - 2u^{\perp\prime}Y^{\perp}(\hat{\beta}_{TSLS}-\beta_0) + Y^{\perp\prime}Y^{\perp}(\hat{\beta}_{TSLS}-\beta_0)^2]/(N-K_1-1)\sigma_{uu} \\ &\Rightarrow (\sigma_{uu} - 2\sigma_{uv}\beta^* + \sigma_{vv}\beta^{*2})/\sigma_{uu} = 1 - 2\rho(\sigma_{vv}/\sigma_{uu})^{1/2}\beta^* + (\sigma_{vv}/\sigma_{uu})\beta^{*2}.\end{aligned}$$

(c) Rewrite t_{TSLS} in (2.9) as $t_{TSLS} = (\hat{\sigma}_{uu}/\sigma_{uu})^{-1/2}[(\hat{\beta}_{TSLS}-\beta_0)/(\sigma_{uu}/\hat{Y}^{\perp\prime}\hat{Y}^{\perp})^{1/2}]$. From lemma A1, $\hat{Y}^{\perp\prime}\hat{Y}^{\perp} \Rightarrow \sigma_{vv}\nu_3$, so from part (a) of this theorem, we have

$$\begin{aligned}(\hat{\beta}_{TSLS}-\beta_0)/(\sigma_{uu}/\hat{Y}^{\perp\prime}\hat{Y}^{\perp})^{1/2} &\Rightarrow (\sigma_{uu}/\sigma_{vv})^{-1/2}(\theta\nu_1/\nu_3 + \kappa\nu_2/\nu_3)/\nu_3^{1/2} \\ &= \rho\nu_1/\nu_3^{1/2} + (1-\rho^2)^{1/2}\nu_2/\nu_3^{1/2} = \nu_4/\nu_3^{1/2}.\end{aligned}$$

Combining this expression with the limit of $\hat{\sigma}_{uu}/\sigma_{uu}$ in part (b) yields the result in the theorem.

(d) The Wald, LM and LR statistics are asymptotically equivalent under the local alternative L. Consider the LM statistic, $F_N = [\hat{Y}^{\perp\prime}\hat{Y}^{\perp}/K_2]/[Y^{\perp\prime}Y^{\perp}/(N-K_1)]$. The result stated in the theorem follows directly from lemma A1 (c) and (i). □

Proof of theorem 2

(a) Under the local alternative (2.16),

$$\hat{\beta}_{TSLS}-\beta_0 = (\hat{Y}^{\perp\prime}\hat{Y}^{\perp})^{-1}(\hat{Y}^{\perp\prime}u^{\perp}) + (\hat{Y}^{\perp\prime}\hat{Y}^{\perp})^{-1}(N^{-1/2}\hat{Y}^{\perp\prime}Z^{\perp})d.$$

From theorem 1(a), $(\hat{Y}^{\perp\prime}\hat{Y}^{\perp})^{-1}(\hat{Y}^{\perp\prime}u^{\perp}) => \beta^*$. From Lemma A1(d), (g), and (i), $(\hat{Y}^{\perp\prime}\hat{Y}^{\perp})^{-1}(N^{-1/2}\hat{Y}^{\perp\prime}Z^{\perp}) => (\lambda+z_v)'V^{1/2}_{ZZ}/(\sigma^{1/2}_{vv}\nu_3)$. Thus,

$$\hat{\beta}_{TSLS}-\beta_0 => \beta^* + \nu_3^{-1}(\lambda+z_v)'V^{1/2}_{ZZ}d/\sigma^{1/2}_{vv} = \beta^*_d .$$

(b) Consider $\hat{u}'P_{Z^\perp}\hat{u}$. Now $\hat{u} = y^{\perp}$ [illegible] $= u^{\perp} - Y^{\perp}(\hat{\beta}_{TSLS}-\beta_0)$, so

$$\hat{u}'P_{Z^\perp}\hat{u} = \{(Z^{\perp\prime}Z^{\perp})^{-1/2}(Z^{\perp\prime}u^{\perp}) - (\text{[illegible]})^{-1/2}(Z^{\perp\prime}Y^{\perp})(\hat{\beta}_{TSLS}-\beta)\}'$$
$$\{(Z^{\perp\prime}Z^{\perp})^{-1/2}(Z^{\perp\prime}u^{\perp}) - (\text{[illegible]})^{-1/2}(Z^{\perp\prime}Y^{\perp})(\hat{\beta}_{TSLS}-\beta)\}.$$

Applying lemma A1(e) and (g) and part (a) of this theorem yields,

$$\hat{u}'P_{Z^\perp}\hat{u} => \sigma_{uu}\{z_u - (\sigma_{vv}/\sigma_{uu})^{1/2}(\lambda+z_v)\beta^*_d\}'\{z_u - (\sigma_{vv}/\sigma_{uu})^{1/2}(\lambda+z_v)\beta^*_d\}.$$

Next consider $\hat{u}'\hat{u}/N$:

$$\hat{u}'\hat{u}/N = u^{\perp\prime}u^{\perp}/N - 2(u^{\perp\prime}Y^{\perp}/N)(\hat{\beta}_{TSLS}-\beta) + (Y^{\perp\prime}Y^{\perp}/N)(\hat{\beta}_{TSLS}-\beta)^2$$
$$=> \sigma_{uu} - 2\sigma_{uv}\beta^*_d + \sigma_{vv}\beta^{*2}_d.$$

so

$$\chi^2_{reg} = \hat{u}'M_{Z^\perp}\hat{u}/(\hat{u}'\hat{u}/N)$$
$$=> (\sigma_{uu}-2\sigma_{uv}\beta_d^0+\sigma_{vv}\beta_d^{0\,2})^{-1}\sigma_{uu}[z_u-(\sigma_{vv}/\sigma_{uu})^{1/2}(\lambda+z_v)\beta_d^0]'[z_u-(\sigma_{vv}/\sigma_{uu})^{1/2}(\lambda+z_v)\beta_d^0].$$

which yields the result in the theorem.

The asymptotic equivalence of $\chi^2_{reg}/(K_2-1)$ and F_{Bas} under the null and local alternative ($d \neq 0$) follows from noting that their numerators are identical (up to the factor K_2-1) and their denominators are asymptotically equivalent: $\hat{u}'\hat{u}/N - \hat{u}'M_{Z^\perp}\hat{u}/N = \hat{u}'P_{Z^\perp}\hat{u}/N \xrightarrow{p} 0$. □

<u>Proof of theorem 3</u>

(a) Under the null $\beta=\beta_0$, from the definition of A_N and lemma A1, we have,

$$A_N(\beta_0) = \{(u^{\perp\prime}Z^\perp)(Z^{\perp\prime}Z^\perp)^{-1}(Z^{\perp\prime}u^\perp)/K_2\}/\{[u^{\perp\prime}u^\perp - (u^{\perp\prime}Z^\perp)(Z^{\perp\prime}Z^\perp)^{-1}(Z^{\perp\prime}u^\perp)]/(N-K_1)$$
$$=> z_u'z_u/K_2 - \chi^2_{K_2}/K_2.$$

(b) Under the fixed alternative $\beta = \beta_1$, $y^\perp - Y^\perp\beta_0 = u^\perp - Y^\perp(\beta_0-\beta_1)$. By lemma A1,

$$(Z^{\perp\prime}Z^\perp)^{-1/2}Z^{\perp\prime}[u^\perp - Y^\perp(\beta_0-\beta_1)] => \sigma_{uu}^{1/2}z_u - \sigma_{vv}^{1/2}(\beta_0-\beta_1)(\lambda+z_v)$$

so

$$(y^\perp - Y^\perp\beta_0)'Z^\perp(Z^{\perp\prime}Z^\perp)^{-1}Z^{\perp\prime}(y^\perp - Y^\perp\beta_0) =>$$
$$\sigma_{uu}[z_u-(\sigma_{vv}/\sigma_{uu})^{1/2}(\beta_0-\beta_1)(\lambda+z_v)]'[z_u-(\sigma_{vv}/\sigma_{uu})^{1/2}(\beta_0-\beta_1)(\lambda+z_v)].$$

Also, $(y^\perp - Y^\perp\beta_0)'(y^\perp - Y^\perp\beta_0) => \sigma_{uu} - 2\sigma_{uv}(\beta_0-\beta_1) + \sigma_{vv}(\beta_0-\beta_1)^2$. Substitution of this and the previous expression into the definition of $A_N(\beta_0)$ yields the result in the theorem. □

<u>Proof of theorem 4</u>

(a) Let $f_N(\zeta_N) = |G - \zeta_N C/N|$, so that ζ_N corresponds to N times the value of ζ in the determinantal equation, $|G - \zeta C| = 0$. By standard projection algebra, $G = V'(P_{Z^*} - P_X)V = V^{\perp\prime} P_{Z^\perp} V^{\perp}$, where $V^{\perp} = (y^{\perp}\ \ Y^{\perp})$. Similarly, $C = V'M_{Z^*}V = V^{\perp\prime} M_{Z^\perp} V^{\perp}$. Define the 2×2 matrix R by $R_{11} = R_{22} = 1$, $R_{12} = 0$, and $R_{21} = -\beta$. Because $|R| = 1$, $f_N(\zeta_N) = |R'GR - \zeta_N R'CR/N|$. Now,

$$R'GR = R'V^{\perp\prime} P_{Z^\perp} V^{\perp} R = (u^{\perp}\ \ Y^{\perp})' P_{Z^\perp} (u^{\perp}\ \ Y^{\perp}) \Rightarrow G^{\perp *},$$
$$R'CR/N = (u^{\perp}\ \ Y^{\perp})'(u^{\perp}\ \ Y^{\perp})/N - R'GR/N \xrightarrow{p} \Omega$$

where the convergence is joint and follows from lemma A1, and where $G^{\perp *}$ is the 2×2 matrix, $G^{\perp *} = (\sigma_{uu}^{1/2} z_u\ \ \sigma_{vv}^{1/2}(\lambda + z_v))'(\sigma_{uu}^{1/2} z_u\ \ \sigma_{vv}^{1/2}(\lambda + z_v))$.

We first show that the smallest root of $|G - \zeta C|$ is $O_p(N^{-1})$. Let c_N be any sequence increasing to infinity with N, and consider the equation, $c_N^{-1} f_N(\zeta_N) = |c_N^{-1/2} R'GR - (c_N^{-1/2}\zeta_N) R'CR/N|$. Now $c_N^{-1/2} R'GR$ converges in probability to the 2×2 zero matrix, so it follows that the roots of $c_N^{-1} f_N(\zeta_N)$ converge in probability to zero. Thus $c_N^{-1/2}\zeta_N \xrightarrow{p} 0$ for all sequences c_N tending to infinity with N, hence ζ_N is $O_p(1)$ and the smallest root of $|G - \zeta C|$ is $O_p(N^{-1})$.

Because the solution to a determinantal equation is continuous in its elements, from the continuous mapping theorem it follows that the limiting distribution of the solution to $f_N(\zeta_N) = 0$ has the distribution of the solution to the limiting determinantal equation, $f^*(\zeta^*) = 0$, where $f^*(\zeta) = |G^{\perp *} - \zeta\Omega|$. Let $\bar{R} = \mathrm{diag}(\sigma_{uu}^{1/2}, \sigma_{vv}^{1/2})$, so $\Omega = \bar{R}\,\tilde{\Omega}\bar{R}$ and $G^{\perp *} = \bar{R}G^*\bar{R}$, where G^* is defined in theorem 4. Because the roots of $f^*(\zeta) = 0$ are the same as the roots of $|\bar{R}^{-1}|f^*(\zeta)|\bar{R}^{-1}| = 0$, this provides the representation of ζ^*, namely as the smallest root of $|G^* - \zeta\tilde{\Omega}| = 0$.

(b) Let $k = 1 + \zeta/N$ and $\Delta(\zeta) = \hat{\beta}(1 + \zeta/N) - \beta_0$. Then $\Delta_N(\zeta) = a_N(\zeta)^{-1} b_N(\zeta)$, where

$$a_N(\zeta) = Y^{\perp\prime}Y^{\perp} - kY^{\perp\prime}(I - P_{Z^\perp})Y^{\perp} = -\zeta Y^{\perp\prime}Y^{\perp}/N + (1 + \zeta/N)Y^{\perp\prime}P_{Z^\perp}Y^{\perp}$$
$$b_N(\zeta) = Y^{\perp\prime}u^{\perp}/N - kY^{\perp\prime}(I - P_{Z^\perp})u^{\perp} = -\zeta Y^{\perp\prime}u^{\perp}/N + (1 + \zeta/N)Y^{\perp\prime}P_{Z^\perp}u^{\perp}.$$

From lemma A1(b), (c), (e), and (g), a_N and b_N have the limits,

$$(a_N(\zeta),\ b_N(\zeta)) \Rightarrow (\sigma_{vv}\{(\lambda+z_v)'(\lambda+z_v) - \zeta\},\ (\sigma_{uu}\sigma_{vv})^{1/2}\{(\lambda+z_v)'z_u - \rho\zeta\})$$

where the convergence is uniform in ζ over compact sets. Thus,

$$\Delta_N(\zeta) \Rightarrow (\sigma_{uu}/\sigma_{vv})^{1/2}\{(\lambda+z_v)'(\lambda+z_v) - \zeta\}^{-1}\{(\lambda+z_v)'z_u - \rho\zeta\}.$$

The result $\hat{\beta}_{LIML} - \beta_0 \Rightarrow \Delta^*(\zeta^*)$ follows by writing $(\sigma_{uu}/\sigma_{vv})^{1/2}z_u = \theta z_v + \kappa\eta$ and $(\sigma_{uu}/\sigma_{vv})^{1/2}\rho\zeta = \theta\zeta$, by invoking the continuous mapping theorem and by noting that the limiting representations are joint as a consequence of lemma A1. □

(c) By arguments which parallel those used for $\hat{\sigma}_{uu}$ in the proof of theorem 1(b),

$$\hat{\sigma}_{uu}(k) \Rightarrow \sigma_{uu}\{1 - 2\rho(\sigma_{vv}/\sigma_{uu})^{1/2}\Delta^*(\zeta) + (\sigma_{vv}/\sigma_{uu})\Delta^*(\zeta)^2\}.$$

Combining this expression, the limits given in the proof of part (b) of this theorem for $a_N(\zeta)$ and $b_N(\zeta)$, and the limit for $\hat{\beta}(k)-\beta_0$ yields,

$$t(k) \Rightarrow \{(\lambda+z_v)'z_u - \rho\zeta\}/\{[(\lambda+z_v)'(\lambda+z_v)-\zeta][1-2\rho(\sigma_{vv}/\sigma_{uu})^{1/2}\Delta^*(\zeta)+(\sigma_{vv}/\sigma_{uu})\Delta^*(\zeta)^2]\}^{1/2}.$$

The result in the theorem for LIML follows by setting $\zeta = \zeta^*$.

<u>Proof of theorem 5</u>

Write $N^{1/2}(\hat{\gamma}(k)-\gamma) = C_N(\zeta)^{-1}B_N(\zeta)$, where

$$C_N(\zeta) = N^{-1}X'X - (N^{-1/2}X'Y)[-\zeta N^{-1}Y'Y + kY'P_{Z*}Y]^{-1}(N^{-1/2}Y'X)$$
$$B_N(\zeta) = N^{-1/2}X'u - (N^{-1/2}X'Y)[-\zeta N^{-1}Y'Y + kY'P_{Z*}Y]^{-1}(-\zeta N^{-1}Y'u + kY'P_{Z*}u).$$

Now $Y'P_{Z^*}Y = Y'P_{Z^{\perp *}}Y = (N^{-1/2}Y'X \;\; N^{-1/2}Y'Z^{\perp})(N^{-1}Z^{\perp *\prime}Z^{\perp *})^{-1}(N^{-1/2}Y'X \;\; N^{-1/2}Y'Z^{\perp})'$, where $Z^{\perp *} = (X \;\; Z^{\perp})$, and similarly for $Y'P_{Z^*}u$. The various sample moments have limits which are either obtained from lemma A1 or can be calculated directly:

$$
\begin{aligned}
&N^{-1/2}X'Y = N^{-1/2}X'(Z\pi+X\phi+v) \Rightarrow \Sigma_{ZX}g + \Sigma_{XX}f + \Psi_{Xv}\\
&N^{-1/2}Z^{\perp\prime}Y = N^{-1/2}Z'M_X(Z\pi+X\phi+v) \Rightarrow V_{ZZ}g + (\Psi_{Zv}-\Sigma_{ZX}\Sigma_{XX}^{-1}\Psi_{Xv})\\
&N^{-1/2}Z^{\perp\prime}u \Rightarrow \Psi_{Zu}-\Sigma_{ZX}\Sigma_{XX}^{-1}\Psi_{Xu}\\
&N^{-1}Z^{\perp *\prime}Z^{\perp *} \xrightarrow{p} \mathrm{diag}(\Sigma_{XX}, V_{ZZ})
\end{aligned}
$$

These limits imply,

$$
\begin{aligned}
&C_N(\Gamma) \Rightarrow \Sigma_{XX} - (\Sigma_{ZX}g+\Sigma_{XX}f+\Psi_{Xv})(\Sigma_{ZX}g+\Sigma_{XX}f+\Psi_{Xv})'/(\tilde{D}_1-\sigma_{vv}\Gamma)\\
&B_N(\Gamma) \Rightarrow \Psi_{Xu} - (\Sigma_{ZX}g+\Sigma_{XX}f+\Psi_{Xv})(\tilde{D}_2-\sigma_{vu}\Gamma)/(\tilde{D}_1-\sigma_{vv}\Gamma)
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{D}_1 = {} & (\Sigma_{ZX}g+\Sigma_{XX}f+\Psi_{Xv})'\Sigma_{XX}^{-1}(\Sigma_{ZX}g+\Sigma_{XX}f+\Psi_{Xv})\\
& + (V_{ZZ}g+\Psi_{Zv}-\Sigma_{ZX}\Sigma_{XX}^{-1}\Psi_{Xv})'V_{ZZ}^{-1}(V_{ZZ}g+\Psi_{Zv}-\Sigma_{ZX}\Sigma_{XX}^{-1}\Psi_{Xv})\\
\tilde{D}_2 = {} & (\Sigma_{ZX}g+\Sigma_{XX}f+\Psi_{Xv})'\Sigma_{XX}^{-1}\Psi_{Xu} + (V_{ZZ}g+\Psi_{Zv}-\Sigma_{ZX}\Sigma_{XX}^{-1}\Psi_{Xv})'V_{ZZ}^{-1}(\Psi_{Zu}-\Sigma_{ZX}\Sigma_{XX}^{-1}\Psi_{Xu}).
\end{aligned}
$$

In addition to the definitions of z_u, z_v, and η in (A.1), further define:

(A.2a) $\quad z_{Xu} = \sigma_{uu}^{-1/2}\Sigma_{XX}^{1/2}\Psi_{Xu}$

(A.2b) $\quad z_{Xv} = \sigma_{vv}^{-1/2}\Sigma_{XX}^{1/2}\Psi_{Xv}$

(A.2c) $\quad \eta_X = (1-\rho^2)^{-1/2}(z_{Xu} - \rho z_{Xv})$.

Direct calculation confirms that $(z_{Xu}' \;\; z_u' \;\; z_{Xv}' \;\; z_v')'$ is distributed $N(0, \tilde{\Omega}\otimes I_{(K_1+K_2)})$. The result in the theorem follows from substituting these definitions into the preceding limits for $C_N(\Gamma)$ and $B_N(\Gamma)$, rearranging, and collecting terms. □

Footnotes

1. It is difficult to provide systematic evidence on this because first-stage F statistics are often not reported. For example, a review of articles published in the *American Economic Review* between 1988 to 1992 found 18 which used TSLS but none reported first-stage F's or partial R^2's. In each of the 18 articles, econometric inference was performed using conventional asymptotic normal approximations.

2. Although there is some recent work which has relaxed the normality assumption (e.g. Buse (1992)), far less is known about the nonnormal than the normal case.

3. In the finite-sample literature with Z nonrandom, $\pi'Z'Z\pi/\sigma_{vv}$ is referred to as the concentration parameter. For Z stochastic, under assumptions L_π and M this has the probability limit $\lambda'\lambda$.

4. This follows by first noting that the distribution of (z_v, z_u) for $\rho=\rho_0$ is the same as the distribution of $(z_v, -z_u)$ for $\rho=-\rho_0$. Making the appropriate substitutions into (2.7) (divided by θ) and into the representation for t_{TSLS} in theorem 1(c) respectively shows that the limiting representations of $(\hat{\beta}_{TSLS}-\beta_0)/\theta$ and t_{TSLS} for $\rho=\rho_0$ are the same as those of $(\hat{\beta}_{TSLS}-\beta_0)/\theta$ and $-t_{TSLS}$ for $\rho=-\rho_0$.

5. We thank Jean-Marie Dufour for suggesting to us the split-sample Anderson-Rubin test.

6. The distributions depend only on $\lambda'\lambda = \bar{h}$, say, so without loss of generality the data were generated with $\lambda = (\bar{h}^{1/2}, 0, \ldots, 0)'$.

7. We thank David Jaeger for providing these data.

References

Anderson, T.W. (1982), "Some Recent Developments on the Distributions of Single-Equation Estimators," in W. Hildebrand (ed.), *Advances in Econometrics*. Amsterdam: North-Holland, 109 - 122.

Anderson, T.W. and H. Rubin (1949), "Estimation of the Parameters of a Single Equation in a Complete Systems of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46-63.

Anderson, T.W. and T. Sawa (1973), "Distributions of Estimates of Coefficients of a Single Equation in a Simultaneous Systems and their Asymptototic Expansions," *Econometrica*, 41, 683-714.

Anderson, T.W. and T. Sawa (1979), "Evaluation of the Distrubition Function of the Two Stage Least Squares Estimate," *Econometrica*, 47, 163-182.

Angrist, Joshua D. and Alan B. Krueger (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics*, 106, 979-1014.

Basmann, R.L. (1960), "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics," *Journal of the American Statistical Association*, 55, 650-659.

Bound, John, David A. Jaeger, and Regina Baker (1993), "The Cure Can Be Worse than the Disease: A Cautionary Tale Regarding Instrumental Variables," NBER Technical Working Paper no. 137 (June 1993).

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

Choi, I. and P.C.B. Phillips (1992), "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations," *Journal of Econometrics*, 51, 113-150.

Davidson, R. and J. MacKinnon (1993), *Estimation and Inference in Econometrics*. Oxford University Press, New York.

Fuller, W. (1977), "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45, 939-953.

Hall, B.H., C. Cummins, and R. Schnake (1992), *Time Series Processor Version 4.2 Reference Manual*, TSP International, Palo Alto.

Maddala, G.S. (1974), "Some Small Sample Evidence on Tests of Singificance in Simultaneous Equations Models," *Econometrica*, 42, 841-51.

Maddala, G.S. and J. Jeong (1992), "On the Exact Small Sample Distribution of the Instrumental Variable Estimator," *Econometrica*, 60, 181-183.

Mariano, R.S. (1982), "Analytical Small-Sample Distribution Theory in Econometrics: The Simultaneous Equations Case," *International Economic Review*, 23, 503-34.

Morimune, K. (1978), "Improving the Limited Information Maximum Likelihood Estimator when the Disturbances are Small," *Journal of the American Statistical Association*, 75, 230-237.

Morimune, K. (1983), "Approximate Distributions of k-Class Estimators When the Degree of Overidentifiability is Large Compared with the Sample Size," *Econometrica*, 51, 821-841.

Nelson, C.R. and R. Startz (1990a), "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967-976.

Nelson, C.R. and R. Startz (1990b), "The Distribution of the Instrumental Variable Estimator and its t-ratio When the Instrument is a Poor One," *Journal of Business*, 63, 5125-5140.

Phillips, P.C.B. (1983), "Exact Small Sample Theory in the Simultaneous Equations Model," in Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics, Vol. I*, North-Holland, .

Phillips, P.C.B. (1984), "The Exact Distribution of LIML: I," *International Economic Review*, 25 249-261.

Phillips, P.C.B. (1985), "The Exact Distribution of LIML: II," *International Economic Review*, 26 21-36.

Phillips, P.C.B. (1989), "Partially Identified Econometric Models," *Econometric Theory*, 5, 181-240.

Rothenberg, T.J. (1984), "Approximating the Distributions of Econometric Estimators and Test Statistics," in Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics, Vol. II*, North-Holland, 881-936.

Sawa, T. (1973a), "Almost Unbiased Estimator in Simultaneous Equations Systmes," *International Economic Review*, 14, 97-106.

Sawa, T. (1973b), "The Mean Square Error of a Combined Estimator and Numerical Comparison with the TSLS Estimator," *Journal of Econometrics*, 1, 115-132.

TABLE 1

Finite Sample CDF of $\hat{\delta}_{TSLS}-\delta_0$ for Design I, Evaluated At Selected Quantiles of the Asymptotic Distribution

Values of finite sample CDF evaluated at:

Parameters			$N/K_2 = 5$					$N/K_2 = 20$				
ρ	$\lambda'\lambda/K_2$	K_2	5%	20%	50%	80%	95%	5%	20%	50%	80%	95%
0.50	0	1	0.05	0.20	0.50	0.80	0.95	0.05	0.20	0.49	0.80	0.95
0.50	0.25	1	0.05	0.20	0.49	0.79	0.95	0.05	0.20	0.49	0.80	0.95
0.50	1	1	0.05	0.20	0.48	0.77	0.94	0.05	0.20	0.49	0.79	0.95
0.50	10	1	0.09	0.23	0.49	0.74	0.88	0.06	0.21	0.50	0.78	0.94
0.50	0	4	0.05	0.20	0.50	0.81	0.95	0.05	0.21	0.51	0.80	0.95
0.50	0.25	4	0.05	0.20	0.50	0.79	0.95	0.04	0.20	0.50	0.80	0.95
0.50	1	4	0.05	0.19	0.48	0.78	0.94	0.05	0.20	0.49	0.80	0.95
0.50	10	4	0.06	0.20	0.50	0.79	0.93	0.05	0.20	0.50	0.80	0.95
0.99	0	1	0.05	0.20	0.50	0.79	0.95	0.05	0.19	0.50	0.80	0.95
0.99	0.25	1	0.05	0.17	0.42	0.82	0.96	0.05	0.20	0.48	0.81	0.95
0.99	1	1	0.04	0.17	0.42	0.69	0.95	0.05	0.20	0.48	0.77	0.95
0.99	10	1	0.08	0.23	0.48	0.73	0.87	0.06	0.21	0.51	0.79	0.94
0.99	0	4	0.05	0.20	0.49	0.79	0.95	0.05	0.21	0.50	0.80	0.95
0.99	0.25	4	0.05	0.19	0.46	0.78	0.95	0.05	0.20	0.49	0.80	0.95
0.99	1	4	0.05	0.18	0.45	0.74	0.93	0.05	0.20	0.48	0.78	0.94
0.99	10	4	0.05	0.20	0.49	0.77	0.92	0.05	0.20	0.51	0.80	0.94

• Finite Sample CDF's based on 20,000 Monte Carlo draws.

TABLE 2

Finite Sample CDF of t_{TSLS} for Design I, Evaluated At Selected Quantiles of the Asymptotic Distribution

Values of finite sample CDF evaluated at:

Parameters			$N/K_2 = 5$					$N/K_2 = 20$				
ρ	$\lambda'\lambda/K_2$	K_2	5%	20%	50%	80%	95%	5%	20%	50%	80%	95%
0.50	0	1	0.08	0.21	0.49	0.75	0.89	0.06	0.20	0.50	0.79	0.94
0.50	0.25	1	0.08	0.21	0.48	0.76	0.89	0.05	0.20	0.49	0.79	0.94
0.50	1	1	0.09	0.20	0.48	0.76	0.90	0.06	0.19	0.49	0.79	0.94
0.50	10	1	0.09	0.20	0.49	0.77	0.91	0.06	0.20	0.50	0.79	0.95
0.50	0	4	0.06	0.20	0.49	0.77	0.92	0.05	0.21	0.50	0.80	0.94
0.50	0.25	4	0.06	0.20	0.50	0.78	0.92	0.05	0.20	0.50	0.79	0.94
0.50	1	4	0.06	0.19	0.48	0.78	0.93	0.05	0.20	0.49	0.80	0.95
0.50	10	4	0.06	0.20	0.50	0.80	0.94	0.05	0.20	0.50	0.80	0.95
0.99	0	1	0.05	0.19	0.47	0.73	0.88	0.05	0.20	0.49	0.79	0.94
0.99	0.25	1	0.06	0.17	0.43	0.71	0.88	0.05	0.20	0.49	0.78	0.94
0.99	1	1	0.09	0.15	0.43	0.71	0.87	0.07	0.19	0.48	0.78	0.94
0.99	10	1	0.11	0.17	0.48	0.76	0.90	0.07	0.17	0.51	0.80	0.94
0.99	0	4	0.04	0.19	0.48	0.76	0.91	0.05	0.19	0.50	0.80	0.94
0.99	0.25	4	0.04	0.19	0.57	0.92	0.99	0.05	0.20	0.51	0.85	0.98
0.99	1	4	0.05	0.18	0.48	0.85	0.99	0.05	0.20	0.49	0.81	0.97
0.99	10	4	0.05	0.20	0.49	0.79	0.94	0.05	0.20	0.51	0.80	0.95

* Finite Sample CDF's based on 20,000 Monte Carlo draws.

TABLE 3

Finite Sample CDF of $\hat{\delta}_{LIML}-\delta$, for Design I, Evaluated At Selected Quantiles of the Asymptotic Distribution

Values of finite sample CDF evaluated at:

Parameters			$N/K_2 = 5$					$N/K_2 = 20$				
ρ	$\lambda'\lambda/K_2$	K_2	5%	20%	50%	80%	95%	5%	20%	50%	80%	95%
0.50	0	1	0.05	0.20	0.50	0.80	0.95	0.05	0.20	0.50	0.80	0.95
0.50	0.25	1	0.05	0.20	0.49	0.80	0.95	0.05	0.20	0.50	0.80	0.95
0.50	1	1	0.05	0.20	0.47	0.76	0.94	0.05	0.20	0.50	0.80	0.95
0.50	10	1	0.09	0.23	0.49	0.74	0.88	0.06	0.21	0.50	0.79	0.94
0.50	0	4	0.05	0.20	0.51	0.80	0.95	0.05	0.20	0.51	0.80	0.94
0.50	0.25	4	0.05	0.20	0.48	0.79	0.95	0.05	0.20	0.49	0.79	0.95
0.50	1	4	0.06	0.20	0.49	0.77	0.94	0.05	0.20	0.50	0.80	0.95
0.50	10	4	0.07	0.22	0.50	0.78	0.93	0.05	0.20	0.50	0.80	0.95
0.99	0	1	0.05	0.20	0.50	0.80	0.95	0.06	0.20	0.50	0.80	0.95
0.99	0.25	1	0.04	0.17	0.43	0.83	0.96	0.05	0.20	0.50	0.82	0.95
0.99	1	1	0.05	0.17	0.42	0.69	0.95	0.05	0.20	0.49	0.78	0.95
0.99	10	1	0.08	0.22	0.47	0.72	0.87	0.06	0.21	0.49	0.78	0.93
0.99	0	4	0.04	0.20	0.49	0.79	0.95	0.05	0.20	0.49	0.80	0.95
0.99	0.25	4	0.05	0.19	0.48	0.76	0.95	0.05	0.20	0.49	0.79	0.95
0.99	1	4	0.06	0.20	0.49	0.77	0.92	0.05	0.20	0.49	0.79	0.94
0.99	10	4	0.06	0.21	0.50	0.79	0.94	0.05	0.21	0.51	0.80	0.95

• Finite Sample CDF's based on 20,000 Monte Carlo draws.

pl

TABLE 4

Finite Sample CDF of t_{LmL} for Design I, Evaluated At Selected Quantiles of the Asymptotic Distribution

Values of finite sample CDF evaluated at:

Parameters			N/K₂ = 5					N/K₂ = 20				
ρ	$\lambda'\lambda/K_2$	K_2	5%	20%	50%	80%	95%	5%	20%	50%	80%	95%
0.50	0	1	0.08	0.22	0.49	0.75	0.89	0.06	0.21	0.50	0.79	0.94
0.50	0.25	1	0.08	0.21	0.48	0.75	0.89	0.05	0.20	0.49	0.79	0.94
0.50	1	1	0.09	0.21	0.48	0.77	0.90	0.06	0.20	0.50	0.80	0.94
0.50	10	1	0.10	0.20	0.49	0.77	0.90	0.06	0.20	0.50	0.80	0.94
0.50	0	4	0.06	0.21	0.51	0.80	0.94	0.05	0.20	0.51	0.80	0.95
0.50	0.25	4	0.06	0.20	0.49	0.79	0.93	0.05	0.20	0.50	0.80	0.95
0.50	1	4	0.06	0.19	0.49	0.79	0.94	0.06	0.20	0.50	0.80	0.95
0.50	10	4	0.06	0.21	0.50	0.79	0.94	0.05	0.20	0.50	0.80	0.95
0.99	0	1	0.05	0.19	0.47	0.74	0.88	0.05	0.20	0.49	0.79	0.94
0.99	0.25	1	0.05	0.17	0.43	0.70	0.87	0.05	0.20	0.49	0.78	0.94
0.99	1	1	0.09	0.16	0.44	0.71	0.88	0.07	0.18	0.50	0.78	0.94
0.99	10	1	0.11	0.16	0.47	0.74	0.89	0.07	0.17	0.49	0.78	0.94
0.99	0	4	0.05	0.20	0.50	0.79	0.93	0.05	0.20	0.49	0.80	0.94
0.99	0.25	4	0.07	0.18	0.48	0.77	0.93	0.05	0.20	0.50	0.79	0.94
0.99	1	4	0.08	0.17	0.49	0.78	0.93	0.07	0.18	0.49	0.79	0.95
0.99	10	4	0.06	0.20	0.50	0.79	0.94	0.05	0.20	0.51	0.80	0.95

* Finite Sample CDF's based on 20,000 Monte Carlo draws.

TABLE 5

Finite Sample Coverage Rates For Design I Tests at the 95% Confidence Level

Parameters			Coverage Rates For					
			$N/K_2=5$			$N/K_2=20$		
ρ	$\lambda'\lambda/K_2$	K_2	t_{TSLS}	t_{LIML}	A-R	t_{TSLS}	t_{LIML}	A-R
0.50	0	1	0.98	0.98	0.86	0.98	0.98	0.93
0.50	0.25	1	0.98	0.98	0.85	0.98	0.98	0.93
0.50	1	1	0.97	0.97	0.85	0.98	0.98	0.93
0.50	10	1	0.96	0.96	0.86	0.96	0.96	0.94
0.50	0	4	0.87	0.93	0.90	0.87	0.92	0.94
0.50	0.25	4	0.88	0.92	0.90	0.86	0.92	0.94
0.50	1	4	0.89	0.93	0.90	0.88	0.93	0.94
0.50	10	4	0.94	0.95	0.90	0.94	0.95	0.94
0.99	0	1	0.47	0.47	0.85	0.39	0.39	0.93
0.99	0.25	1	0.70	0.70	0.86	0.68	0.68	0.94
0.99	1	1	0.82	0.82	0.85	0.81	0.82	0.93
0.99	10	1	0.93	0.93	0.86	0.92	0.91	0.93
0.99	0	4	[illegible]	[illegible].26	0.90	0.01	0.26	0.94
0.99	0.25	4	[illegible]	[illegible]30	0.90	0.15	0.80	0.94
0.99	1	[illegible]	[illegible]	0.89	0.90	0.45	0.89	0.94
0.99	10	[illegible]	[illegible].38	0.94	0.90	0.88	0.94	0.94

* Finite Sample Coverage rates based on 20,000 Monte Carlo Draws.

TABLE 6

Finite Sample CDF of $\hat{\beta}_{TSLS}-\beta_0$ for Design II, Evaluated At Selected Quantiles of the Asymptotic Distribution

Values of finite sample CDF evaluated at:

Parameters			$N/K_2 = 5$					$N/K_2 = 20$				
ρ	$\lambda'\lambda/K_2$	K_2	5%	20%	50%	80%	95%	5%	20%	50%	80%	95%
0.50	0	4	0.03	0.21	0.50	0.74	0.92	0.04	0.20	0.50	0.78	0.94
0.50	0.25	4	0.04	0.21	0.53	0.78	0.94	0.04	0.20	0.51	0.80	0.95
0.50	1	4	0.03	0.20	0.55	0.82	0.96	0.04	0.20	0.51	0.81	0.96
0.50	10	4	0.02	0.17	0.57	0.84	0.95	0.03	0.19	0.53	0.82	0.95
0.50	0	100	0.10	0.26	0.50	0.73	0.88	0.06	0.22	0.51	0.77	0.93
0.50	0.25	100	0.08	0.25	0.51	0.75	0.90	0.06	0.22	0.51	0.79	0.94
0.50	1	100	0.06	0.24	0.53	0.79	0.92	0.05	0.21	0.52	0.80	0.95
0.50	10	100	0.01	0.22	0.59	0.83	0.93	0.03	0.21	0.55	0.81	0.94
0.20	0	4	0.04	0.17	0.50	0.76	0.91	0.05	0.19	0.49	0.78	0.94
0.20	0.25	4	0.04	0.18	0.52	0.80	0.94	0.05	0.19	0.51	0.80	0.94
0.20	1	4	0.04	0.18	0.53	0.82	0.95	0.05	0.20	0.52	0.82	0.95
0.20	10	4	0.02	0.18	0.57	0.84	0.94	0.04	0.20	0.53	0.82	0.95
0.20	0	100	0.05	0.21	0.51	0.78	0.92	0.05	0.20	0.50	0.80	0.94
0.20	0.25	100	0.05	0.21	0.51	0.79	0.93	0.04	0.20	0.51	0.79	0.94
0.20	1	100	0.04	0.21	0.53	0.81	0.94	0.04	0.21	0.51	0.80	0.94
0.20	10	100	0	0.20	0.57	0.83	0.94	0.03	0.21	0.54	0.82	0.94

* Finite Sample CDF's based on 20,000 Monte Carlo draws.

TABLE 7

Finite Sample CDF of t_{TMS} for Design II, Evaluated At Selected Quantiles of the Asymptotic Distribution

Values of finite sample CDF evaluated at:

Parameters			N/K₂ = 5					N/K₂ = 20				
ρ	$\lambda'\lambda/K_2$	K_2	5%	20%	50%	80%	95%	5%	20%	50%	80%	95%
0.50	0	4	0.07	0.20	0.45	0.71	0.85	0.05	0.19	0.47	0.77	0.92
0.50	0.25	4	0.07	0.21	0.48	0.73	0.88	0.05	0.20	0.49	0.78	0.93
0.50	1	4	0.07	0.22	0.51	0.79	0.92	0.05	0.20	0.50	0.79	0.94
0.50	10	4	0.04	0.22	0.55	0.81	0.93	0.04	0.20	0.53	0.81	0.94
0.50	0	100	0.06	0.23	0.50	0.76	0.90	0.06	0.21	0.51	0.79	0.93
0.50	0.25	100	0.06	0.22	0.51	0.78	0.93	0.05	0.21	0.50	0.80	0.95
0.50	1	100	0.05	0.21	0.54	0.81	0.94	0.05	0.20	0.52	0.81	0.95
0.50	10	100	0.01	0.22	0.59	0.83	0.94	0.03	0.21	0.54	0.81	0.94
0.20	0	4	0.06	0.21	0.48	0.74	0.89	0.05	0.20	0.48	0.78	0.94
0.20	0.25	4	0.07	0.22	0.50	0.76	0.90	0.05	0.20	0.50	0.79	0.94
0.20	1	4	0.06	0.23	0.51	0.78	0.91	0.06	0.21	0.51	0.79	0.94
0.20	10	4	0.03	0.22	0.56	0.80	0.92	0.04	0.21	0.53	0.81	0.94
0.20	0	100	0.05	0.21	0.50	0.79	0.94	0.05	0.20	0.50	0.80	0.95
0.20	0.25	100	0.05	0.20	0.51	0.80	0.94	0.04	0.20	0.50	0.80	0.94
0.20	1	100	0.04	0.21	0.53	0.81	0.94	0.04	0.21	0.51	0.80	0.95
0.20	10	100	0.01	0.20	0.57	0.83	0.93	0.03	0.21	0.54	0.82	0.94

• Finite Sample CDF's based on 20,000 Monte Carlo draws.

TABLE 8

Finite Sample CDF of $\hat{\beta}_{LIML}-\beta_0$ for Design II, Evaluated At Selected Quantiles of the Asymptotic Distribution

Values of finite sample CDF evaluated at:

Parameters			$N/K_2 = 5$					$N/K_2 = 20$				
ρ	$\lambda'\lambda/K_2$	K_2	5%	20%	50%	80%	95%	5%	20%	50%	80%	95%
0.50	0	4	0.05	0.21	0.51	0.79	0.94	0.05	0.21	0.51	0.80	0.95
0.50	0.25	4	0.05	0.22	0.53	0.83	0.95	0.05	0.21	0.51	0.81	0.95
0.50	1	4	0.05	0.20	0.55	0.84	0.97	0.05	0.21	0.52	0.83	0.96
0.50	10	4	0.03	0.19	0.56	0.83	0.94	0.04	0.21	0.52	0.81	0.94
0.50	0	100	0.05	0.20	0.50	0.80	0.95	0.05	0.20	0.50	0.80	0.95
0.50	0.25	100	0.06	0.22	0.51	0.78	0.94	0.05	0.21	0.50	0.80	0.95
0.50	1	100	0.04	0.21	0.53	0.80	0.93	0.04	0.20	0.52	0.81	0.94
0.50	10	100	0.00	0.19	0.57	0.82	0.93	0.03	0.20	0.54	0.81	0.94
0.20	0	4	0.05	0.21	0.52	0.80	0.95	0.05	0.20	0.50	0.81	0.95
0.20	0.25	4	0.05	0.20	0.53	0.83	0.96	0.05	0.20	0.51	0.81	0.95
0.20	1	4	0.03	0.19	0.53	0.84	0.97	0.04	0.20	0.52	0.82	0.96
0.20	10	4	0.03	0.19	0.55	0.82	0.94	0.04	0.21	0.52	0.80	0.94
0.20	0	100	0.05	0.21	0.51	0.81	0.95	0.05	0.20	0.50	0.80	0.95
0.20	0.25	100	0.06	0.22	0.52	0.79	0.95	0.05	0.21	0.52	0.80	0.95
0.20	1	100	0.04	0.20	0.51	0.79	0.94	0.04	0.20	0.51	0.80	0.95
0.20	10	100	0.01	0.19	0.57	0.82	0.93	0.03	0.20	0.53	0.80	0.94

* Finite Sample CDF's based on 20,000 Monte Carlo draws.

TABLE 9

Finite Sample CDF of t_{LM} for Design II, Evaluated At Selected Quantiles of the Asymptotic Distribution

Values of finite sample CDF evaluated at:

Parameters			$N/K_2 = 5$					$N/K_2 = 20$				
ρ	$\lambda'\lambda/K_2$	K_2	5%	20%	50%	80%	95%	5%	20%	50%	80%	95%
0.50	0	4	0.06	0.22	0.51	0.77	0.90	0.05	0.21	0.51	0.79	0.94
0.50	0.25	4	0.07	0.24	0.53	0.80	0.93	0.06	0.22	0.51	0.79	0.94
0.50	1	4	0.08	0.24	0.54	0.81	0.94	0.06	0.22	0.51	0.81	0.95
0.50	10	4	0.03	0.22	0.56	0.80	0.92	0.04	0.21	0.52	0.80	0.94
0.50	0	100	0.06	0.21	0.49	0.78	0.93	0.05	0.20	0.50	0.80	0.94
0.50	0.25	100	0.06	0.22	0.51	0.78	0.93	0.05	0.21	0.50	0.80	0.94
0.50	1	100	0.04	0.21	0.53	0.80	0.93	0.04	0.20	0.52	0.81	0.94
0.50	10	100	0.00	0.20	0.57	0.82	0.93	0.03	0.20	0.54	0.81	0.94
0.20	0	4	0.06	0.23	0.51	0.77	0.92	0.05	0.21	0.50	0.79	0.94
0.20	0.25	4	0.07	0.24	0.52	0.79	0.93	0.05	0.21	0.50	0.80	0.94
0.20	1	4	0.06	0.24	0.53	0.79	0.93	0.06	0.22	0.52	0.80	0.95
0.20	10	4	0.03	0.22	0.55	0.79	0.92	0.04	0.22	0.52	0.79	0.94
0.20	0	100	0.06	0.22	0.51	0.79	0.94	0.05	0.20	0.50	0.79	0.94
0.20	0.25	100	0.06	0.22	0.52	0.79	0.94	0.05	0.21	0.51	0.80	0.95
0.20	1	100	0.04	0.21	0.51	0.79	0.93	0.04	0.20	0.51	0.80	0.94
0.20	10	100	0.01	0.20	0.57	0.82	0.93	0.03	0.21	0.53	0.80	0.94

• Finite Sample CDF's based on 20,000 Monte Carlo draws.

TABLE 10

Finite Sample Coverage Rates For Design II Tests at the 95% Confidence Level

Parameters			Coverage Rates For					
			$N/K_2=5$			$N/K_2=20$		
ρ	$\lambda'\lambda/K_2$	K_2	t_{TSLS}	t_{LIML}	A-R	t_{TSLS}	t_{LIML}	A-R
0.50	0	4	0.79	0.89	0.92	0.84	0.91	0.95
0.50	0.25	4	0.83	0.92	0.93	0.85	0.91	0.95
0.50	1	4	0.88	0.93	0.93	0.88	0.93	0.95
0.50	10	4	0.93	0.94	0.92	0.94	0.95	0.95
0.50	0	100	0	0.58	0.92	0	0.59	0.94
0.50	0.25	100	0	0.74	0.92	0	0.76	0.94
0.50	1	100	0.04	0.86	0.92	0.04	0.87	0.94
0.50	10	100	0.72	0.94	0.92	0.64	0.95	0.94
0.20	0	4	0.96	0.98	0.92	0.98	0.99	0.95
0.20	0.25	4	0.96	0.98	0.93	0.98	0.98	0.95
0.20	1	4	0.96	0.97	0.93	0.97	0.97	0.95
0.20	10	4	0.95	0.95	0.92	0.95	0.95	0.95
0.20	0	100	0.47	0.65	0.92	0.47	0.66	0.94
0.20	0.25	100	0.56	0.68	0.92	0.56	0.70	0.94
0.20	1	100	0.72	0.84	0.92	0.71	0.85	0.94
0.20	10	100	0.90	0.94	0.92	0.91	0.94	0.94

* Finite Sample Coverage rates based on 20,000 Monte Carlo Draws.

Table 11

Estimated Effects of Years of Education on Log Weekly Earnings in the 1980 Census

	I	II	III	IV
A. Men Born 1930-39 (n=329,509)				
OLS	.0632	.0632	.0632	.0628
(S.E.)	(.0003)	(.0003)	(.0003)	(.0003)
TSLS	.0990	.0806	.0600	.0811
(S.E.)	(.0207)	(.0164)	(.0290)	(.0109)
LIML	.0999	.0838	.0574	.0982
(S.E.)	(.0210)	(.0179)	(.0385)	(.0153)
Combined	.1002	.0852	.0546	.1021
A-R Confidence Interval	[.052,.153]	[-.003,.179]	[-.441,.490]	[-.015,.240]
A-R Split-Sample Confidence Interval	[.054,.168]	[-.002,.134]	[-∞,+∞]	[-.073,.131]
F (first stage)	30.53	4.747	1.613	1.869
(P-Value)	(.000)	(.000)	(.021)	(.000)
F (over-ID)	1.160	0.775	0.725	0.916
(P-Value)	(.313)	(.800)	(.849)	(.781)
B. Men Born 1940-49 (n=486,926)				
OLS	.0520	.0520	.0521	.0516
(S.E)	(.0003)	(.0003)	(.0003)	(.0003)
TSLS	-.0734	.0393	.0779	.0666
(S.E.)	(.0273)	(.0145)	(.0239)	(.0113)
LIML	-.0902	.0286	.1243	.0878
(S.E.)	(.0301)	(.0177)	(.0420)	(.0178)
Combined	-.0783	.0371	.0928	.0829
A-R Confidence Interval	[ϕ]	[ϕ]	[ϕ]	[.033,.148]
A-R Split-Sample Confidence Interval	[-.135,.010]	[0,.111]	[.028,.245]	[-.035,.162]
F(first stage)	26.32	6.849	2.736	1.929
(P-Value)	(.000)	(.000)	(.000)	(.000)
F(over-ID)	4.847	3.226	1.873	1.140
(P-Value)	(.008)	(.000)	(.004)	(.098)
<u>Controls</u>				
Race, SMSA, married, region, year of birth dummies	Yes	Yes	Yes	Yes
Age, Age^2	No	No	Yes	Yes
State of Birth	No	No	No	Yes
<u>Instruments</u>				
Quarter of Birth	Yes	Yes	Yes	Yes
(Quarter of birth) x (year of birth)	No	Yes	Yes	Yes
(Quarter of birth) X (state of birth)	No	No	No	Yes
# Instruments	3	30	28	178

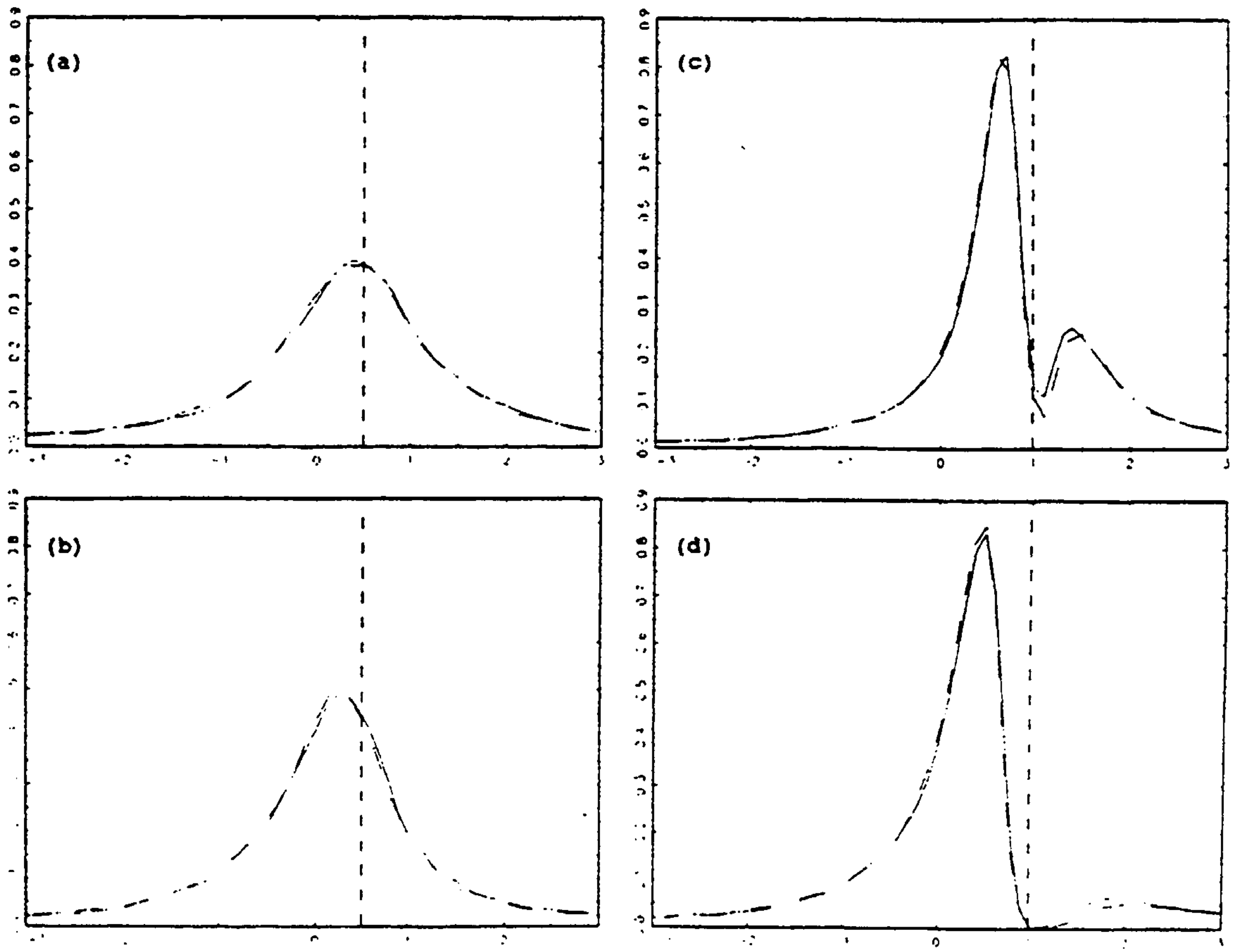

Figure 1. Finite-sample (solid line) and asymptotic (dashed line) pdfs of the TSLS/LIML estimator with $K_2=1$ in design I with: (a) $\rho=.5$, $\lambda'\lambda/K_2=.25$; (b) $\rho=.5$, $\lambda'\lambda/K_2=1$; (c) $\rho=.99$, $\lambda'\lambda/K_2=.25$; (d) $\rho=.99$, $\lambda'\lambda/K_2=1$. Vertical line indicates OLS plim; true beta=0.

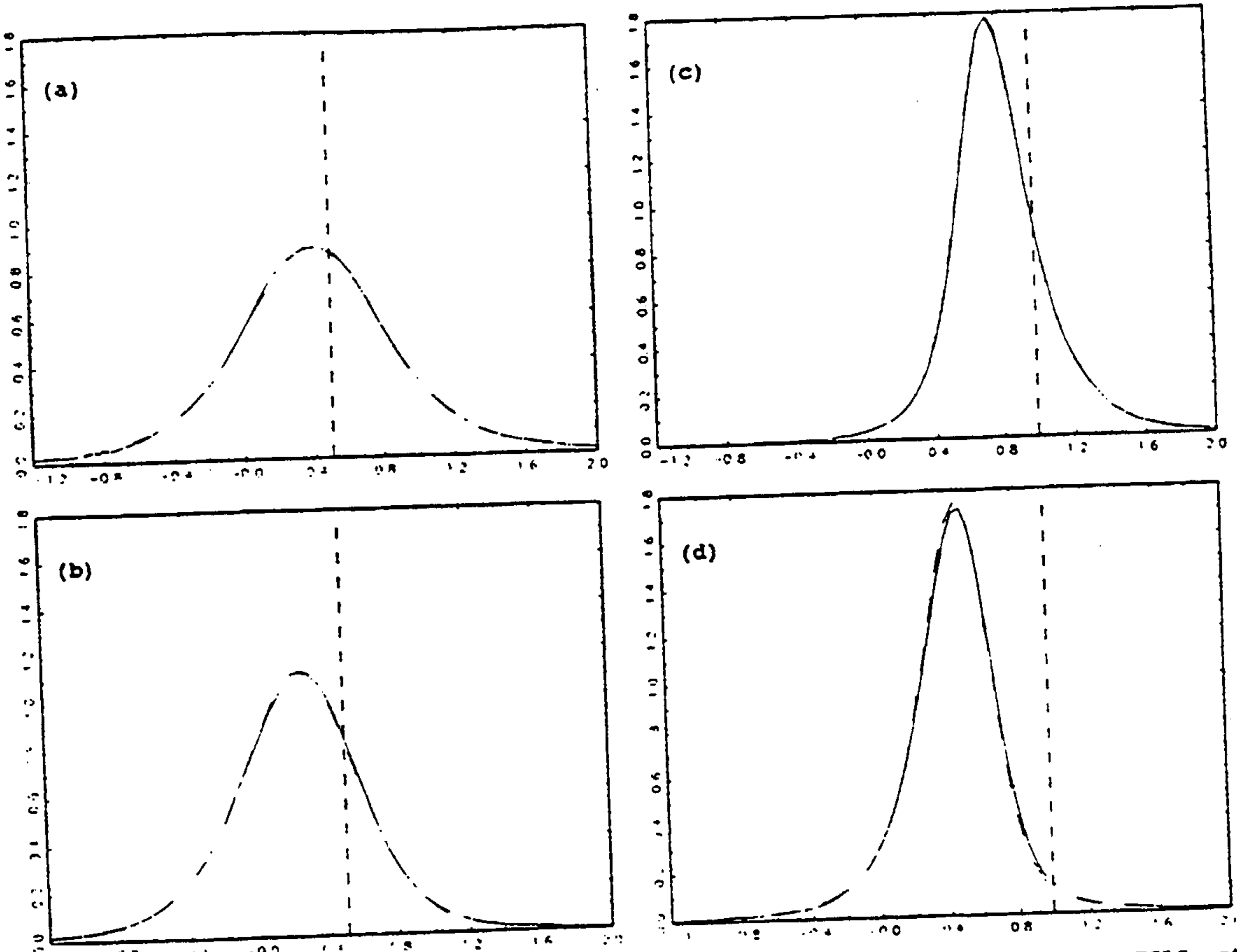

Figure 2. Finite-sample (solid line) and asymptotic (dashed line) pdfs of the TSLS estimator with $K_2=4$ in design I with: (a) $\rho=.5$, $\lambda'\lambda/K_2=.25$; (b) $\rho=.5$, $\lambda'\lambda/K_2=1$; (c) $\rho=.99$, $\lambda'\lambda/K_2=.25$; (d) $\rho=.99$, $\lambda'\lambda/K_2=1$. Vertical line indicates OLS plim; true beta=0.

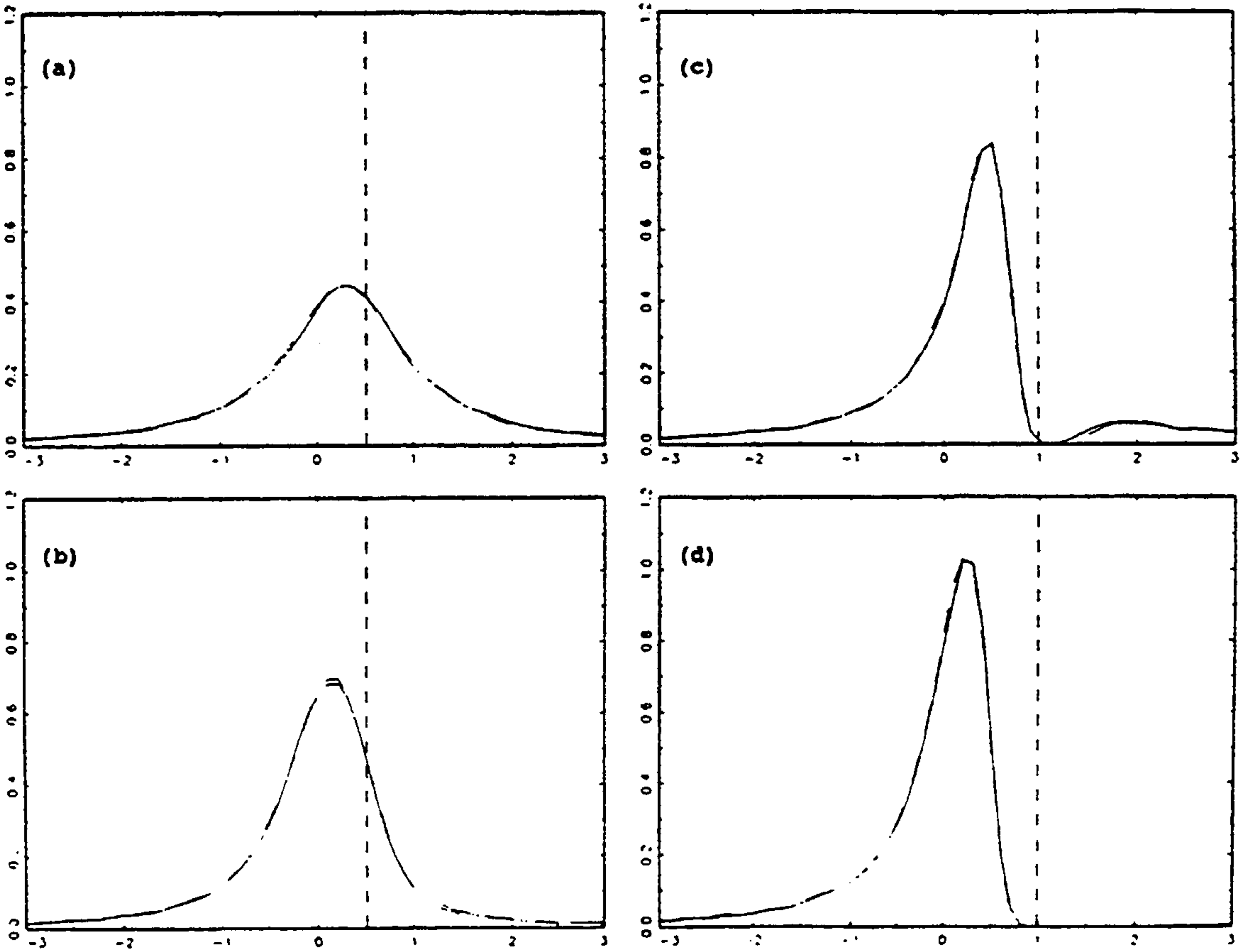

Figure 3. Finite-sample (solid line) and asymptotic (dashed line) pdfs of the LIML estimator with K_2=4 in design I with: (a) ρ=.5, $\lambda'\lambda/K_2$=.25; (b) ρ=.5, $\lambda'\lambda/K_2$=1; (c) ρ=.99, $\lambda'\lambda/K_2$=.25;

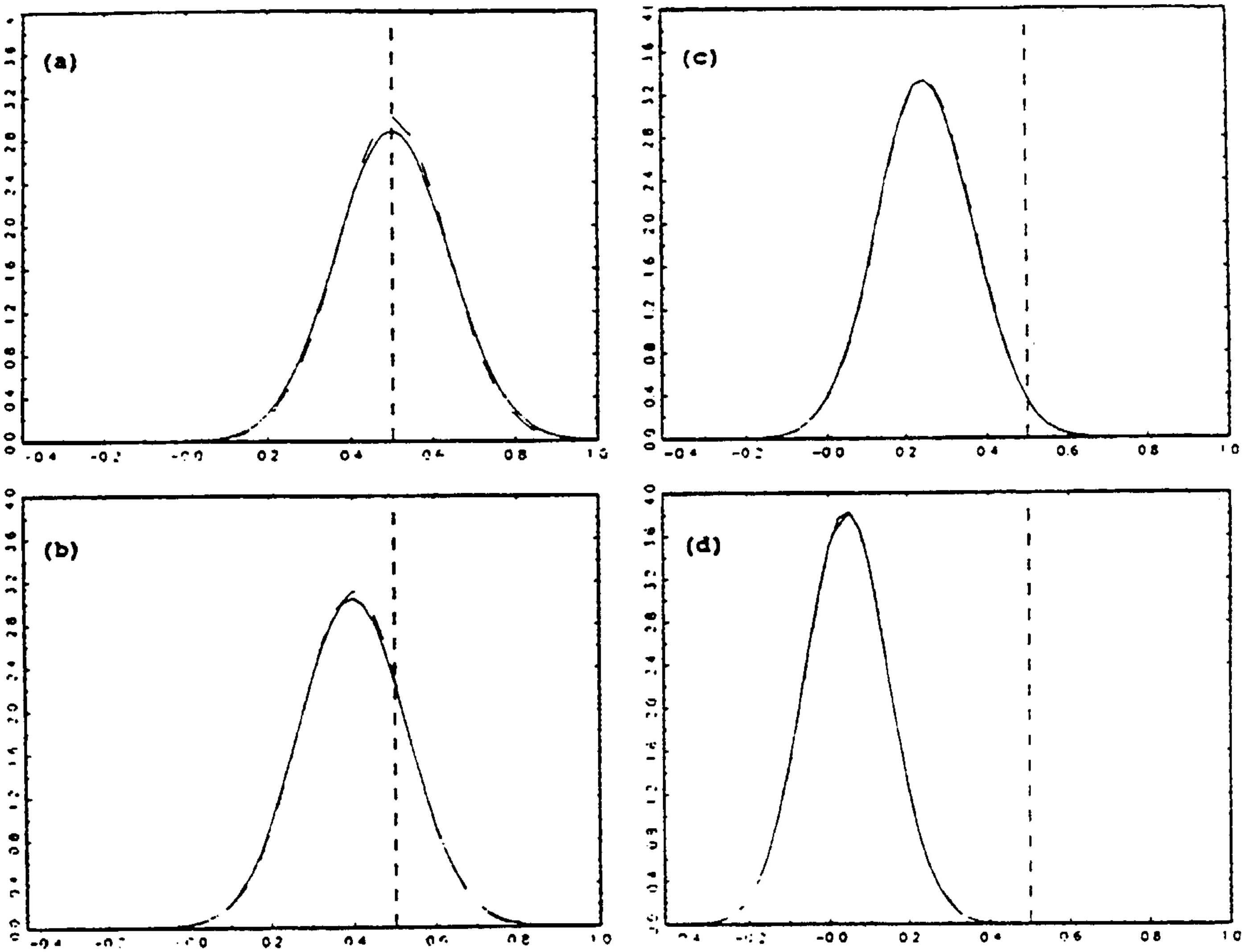

Figure 4. Finite-sample (solid line) and asymptotic (dashed line) pdfs of the TSLS estimator with K_2=100 in design II with: (a) ρ=.5, $\lambda'\lambda/K_2$=0; (b) ρ=.5, $\lambda'\lambda/K_2$=.25; (c) ρ=.5, $\lambda'\lambda/K_2$=1; (d) ρ=.5, $\lambda'\lambda/K_2$=10. Vertical line indicates OLS plim; true beta=0.

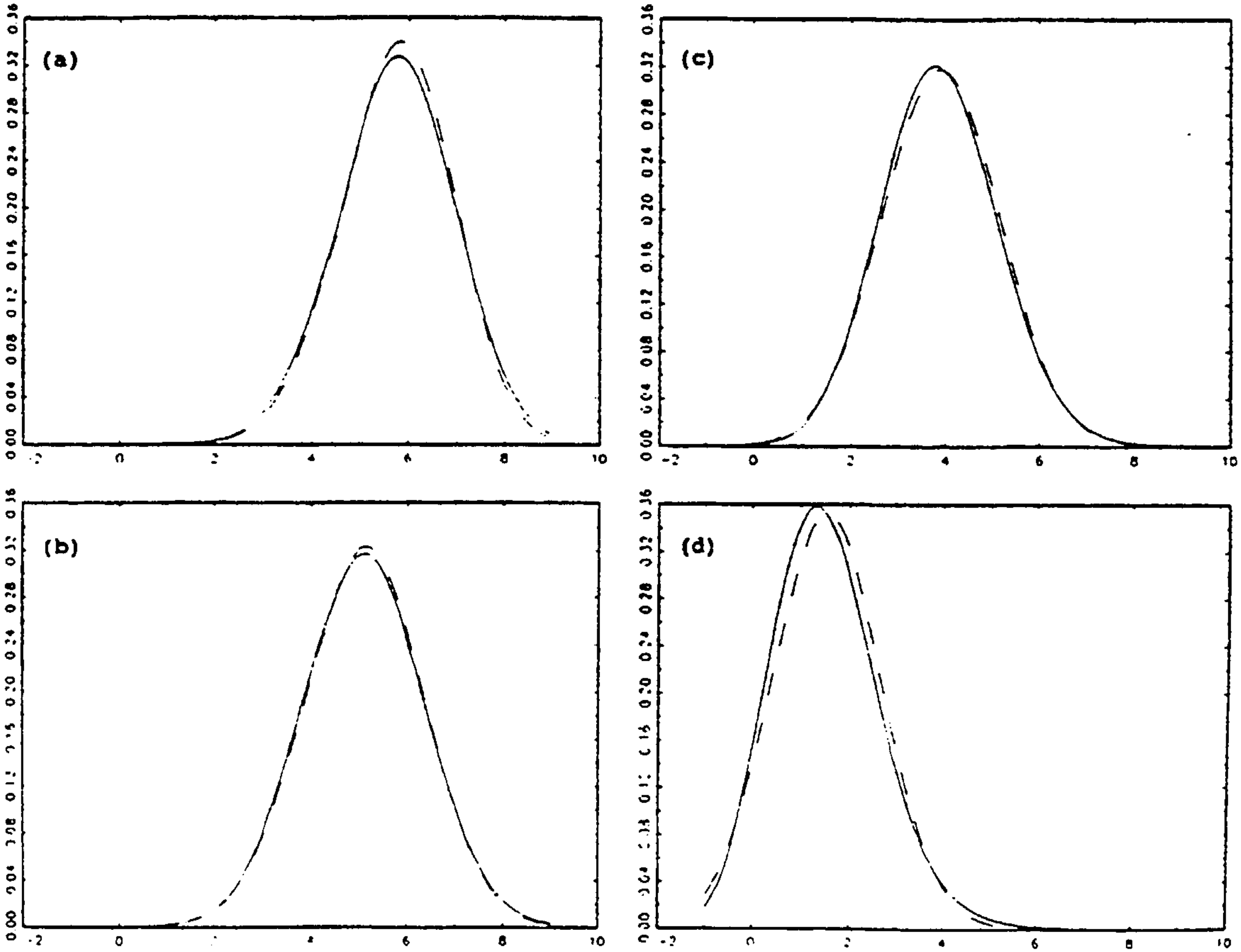

Figure 5. Finite-sample (solid line) and asymptotic (dashed line) pdfs of the TSLS t-statistics with K_2=100 in design II with: (a) ρ=.5, $\lambda'\lambda/K_2$=0; (b) ρ=.5, $\lambda'\lambda/K_2$=.25; (c) ρ=.5, $\lambda'\lambda/K_2$=1; (d) ρ=.5, $\lambda'\lambda/K_2$=10.

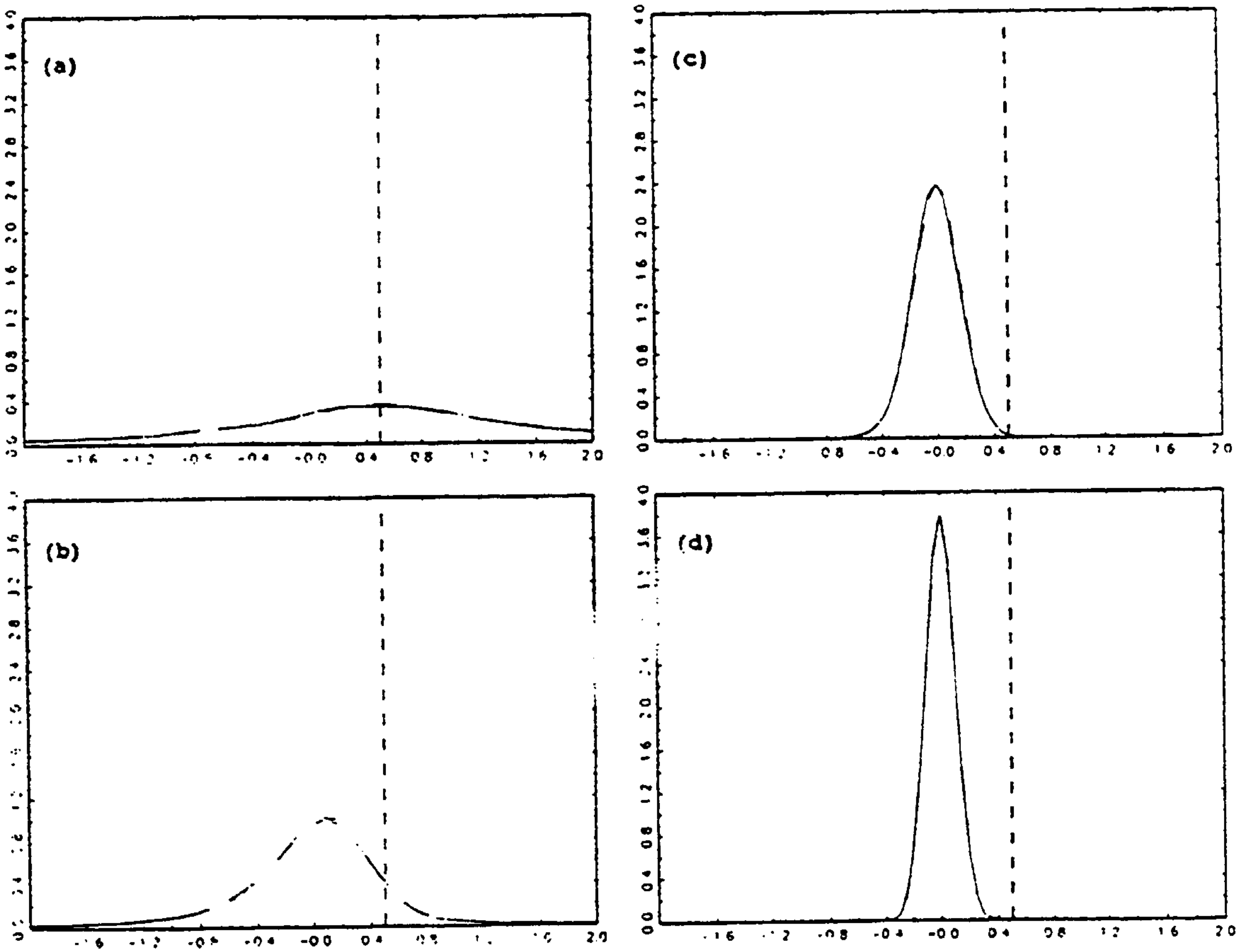

Figure 6. Finite-sample (solid line) and asymptotic (dashed line) pdfs of the LIML estimator with K_2=100 in design II with: (a) ρ=.5, $\lambda'\lambda/K_2$=0; (b) ρ=.5, $\lambda'\lambda/K_2$=.25; (c) ρ=.5, $\lambda'\lambda/K_2$=1; (d) ρ=.5, $\lambda'\lambda/K_2$=10. Vertical line indicates OLS plim; true beta=0.

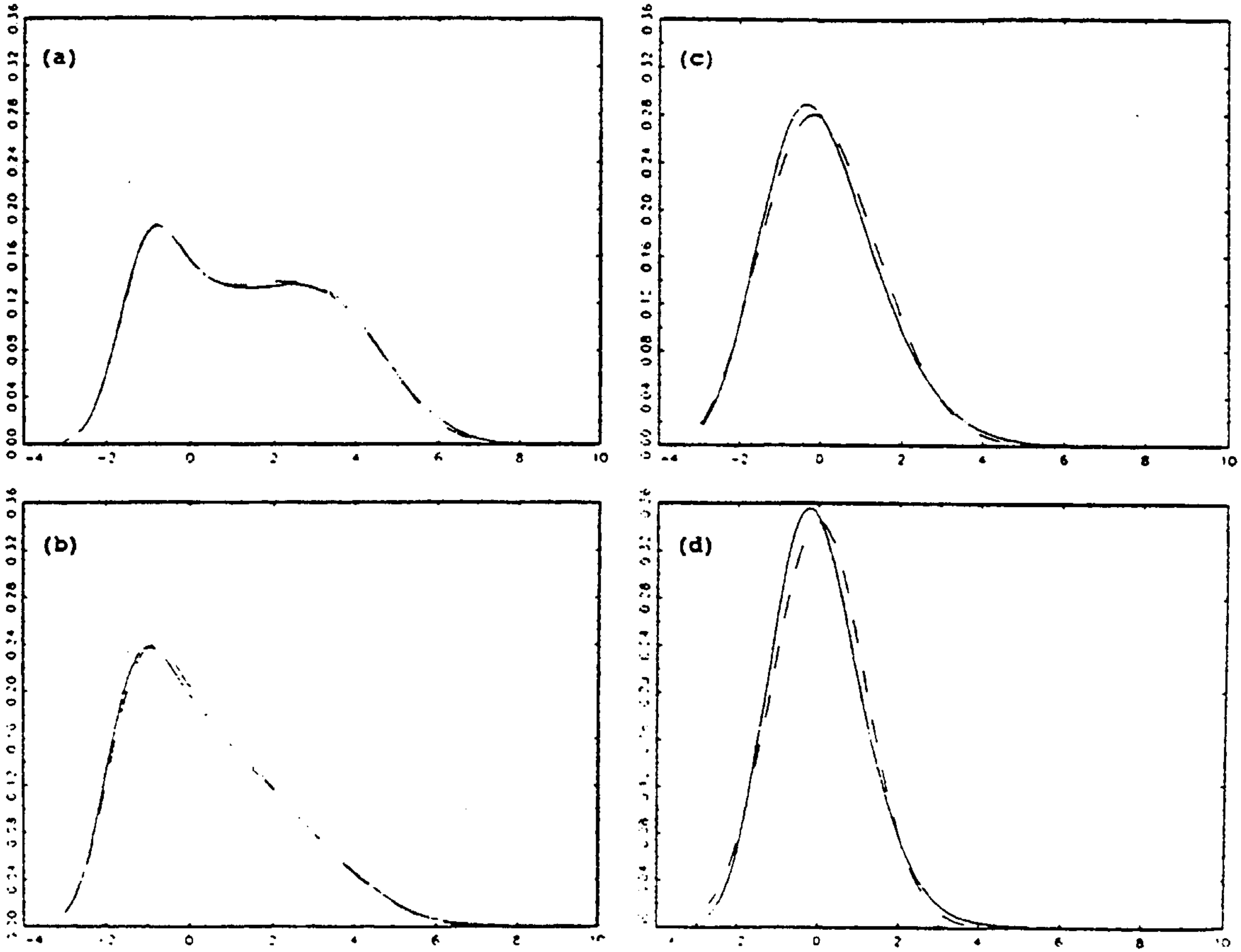

Figure 7. Finite-sample (solid line) and asymptotic (dashed line) pdfs of the LIML t-statistics with K_2=100 in design II with: (a) ρ=.5, $\lambda'\lambda/K_2$=0; (b) ρ=.5, $\lambda'\lambda/K_2$=.25; (c) ρ=.5, $\lambda'\lambda/K_2$=1; (d) ρ=.5, $\lambda'\lambda/K_2$=10.

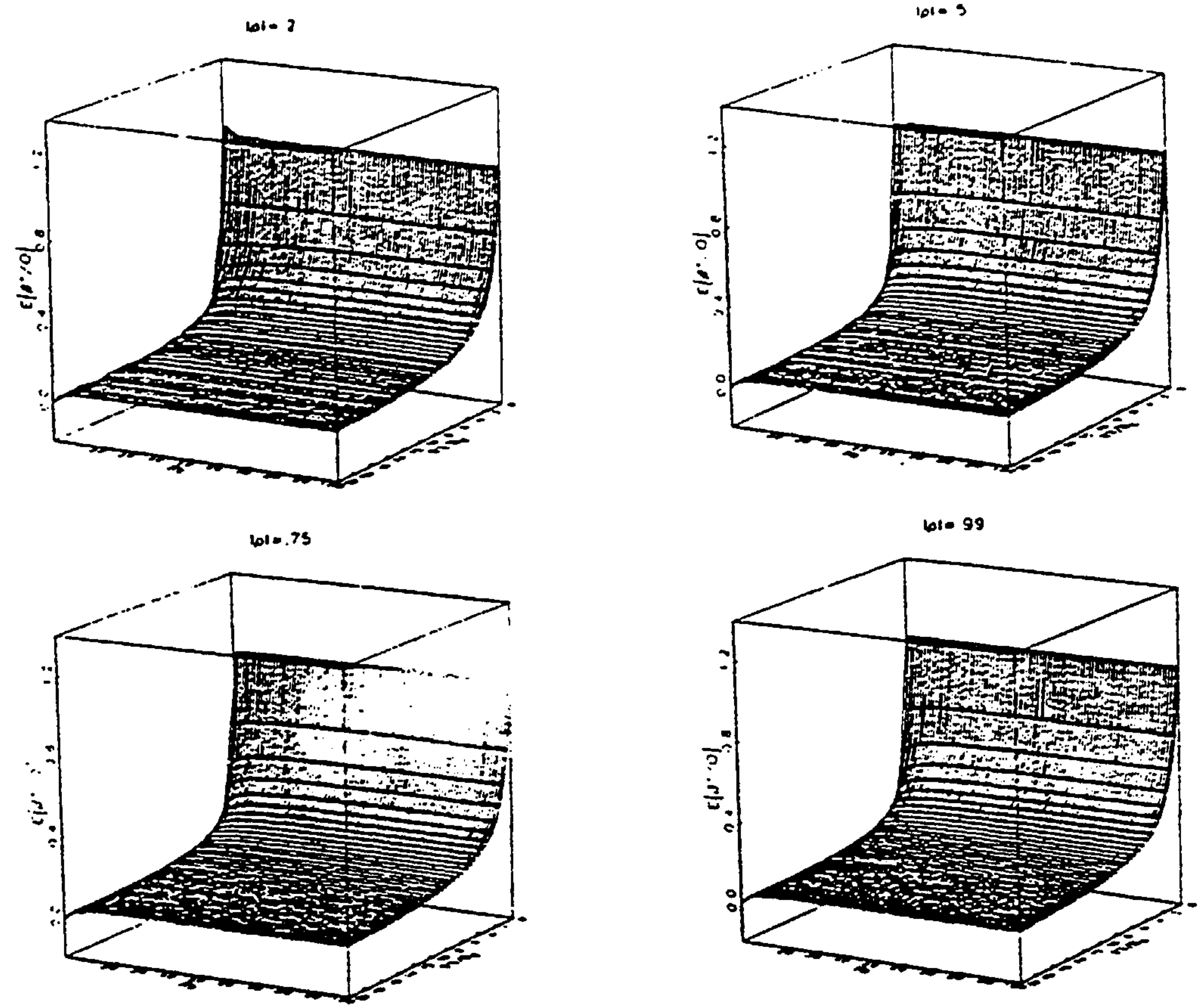

Figure 8. Asymptotic bias of $\hat{\beta}_{TMS}$ as a fraction of the bias of $\hat{\beta}_{OLS}$

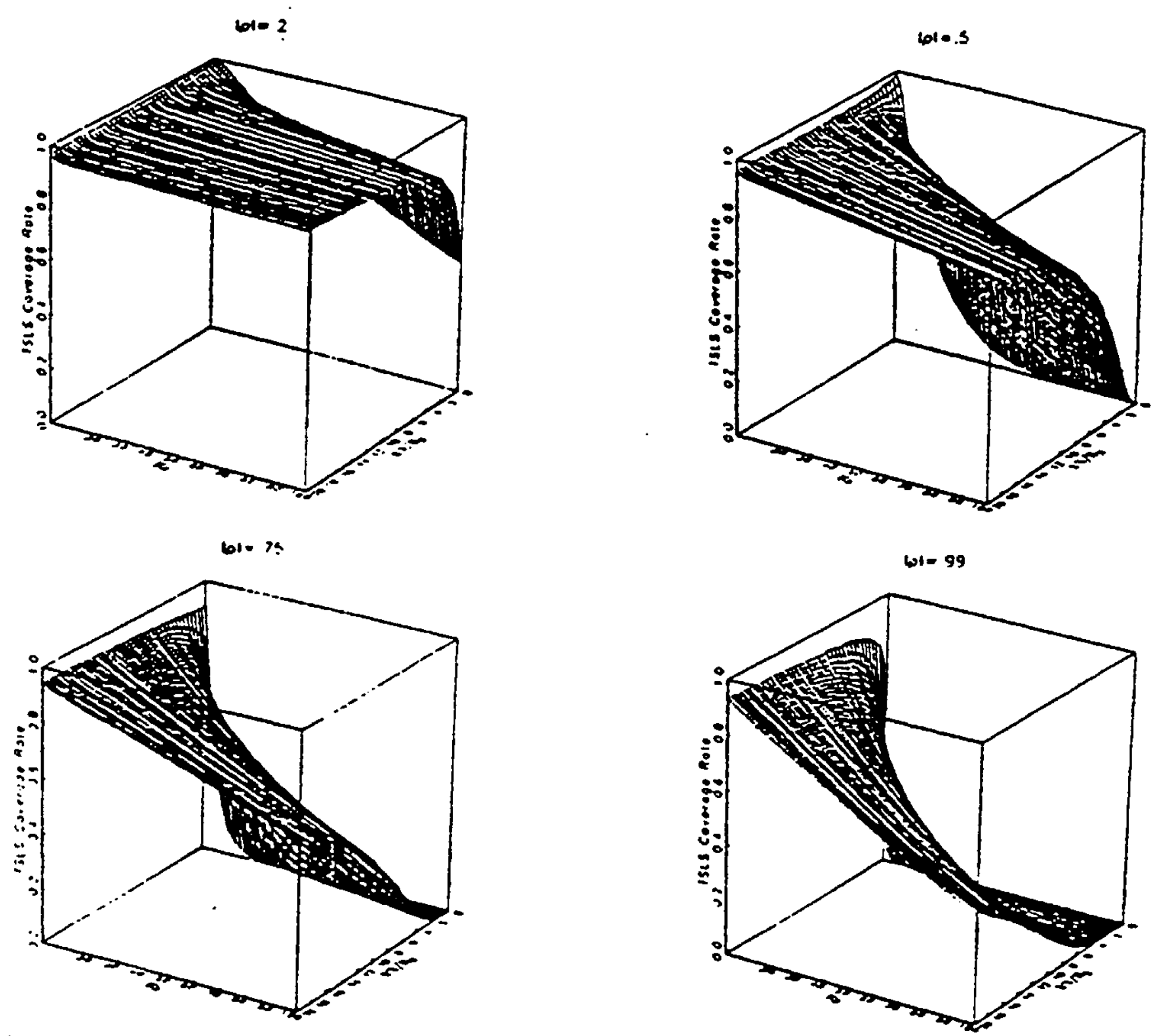

Figure 9. Asymptotic coverage rates for conventional TSLS 95% confidence intervals

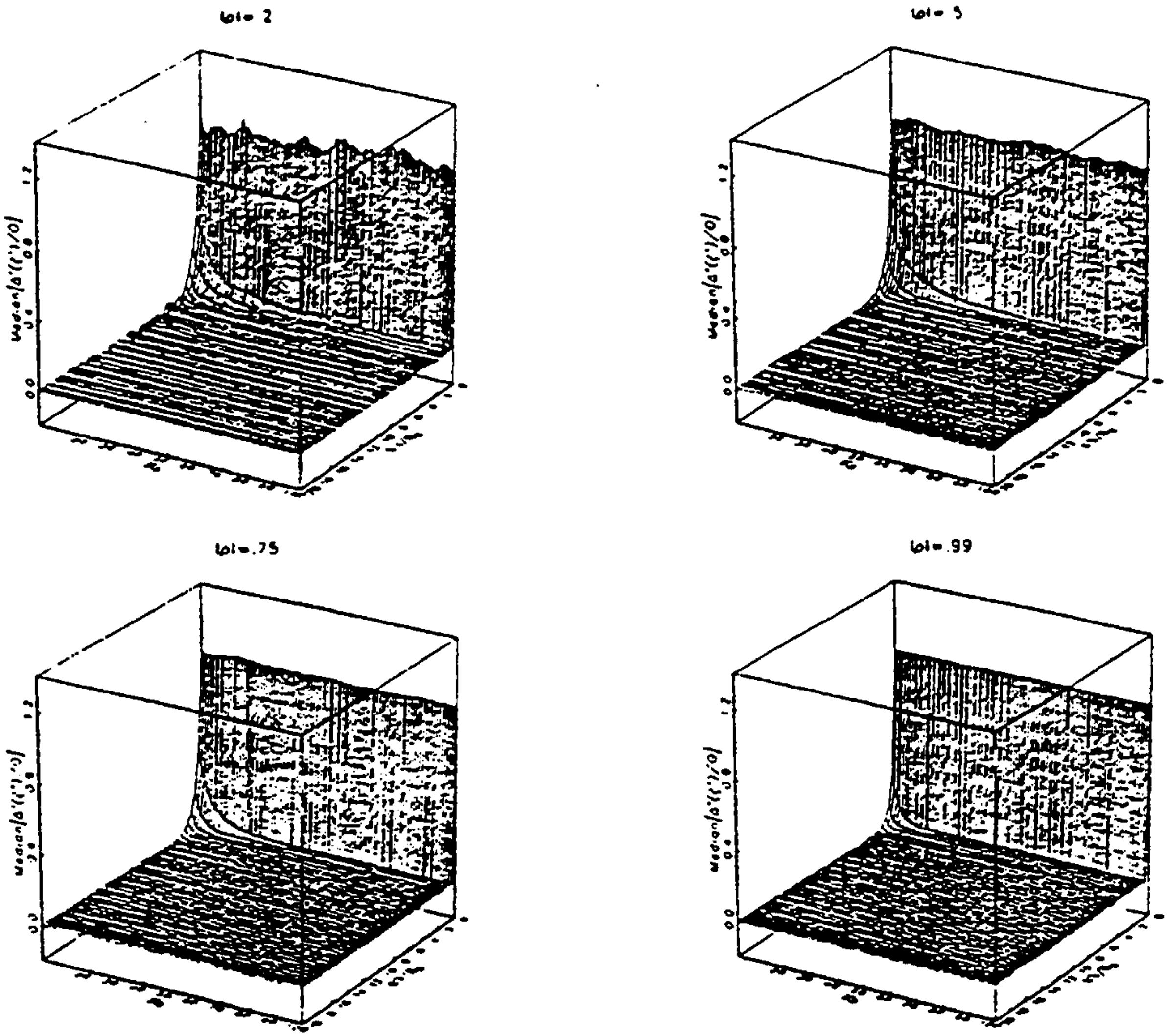

Figure 10. Asymptotic median bias of $\hat{\beta}_{LIML}$ as a fraction of the bias of $\hat{\beta}_{OLS}$

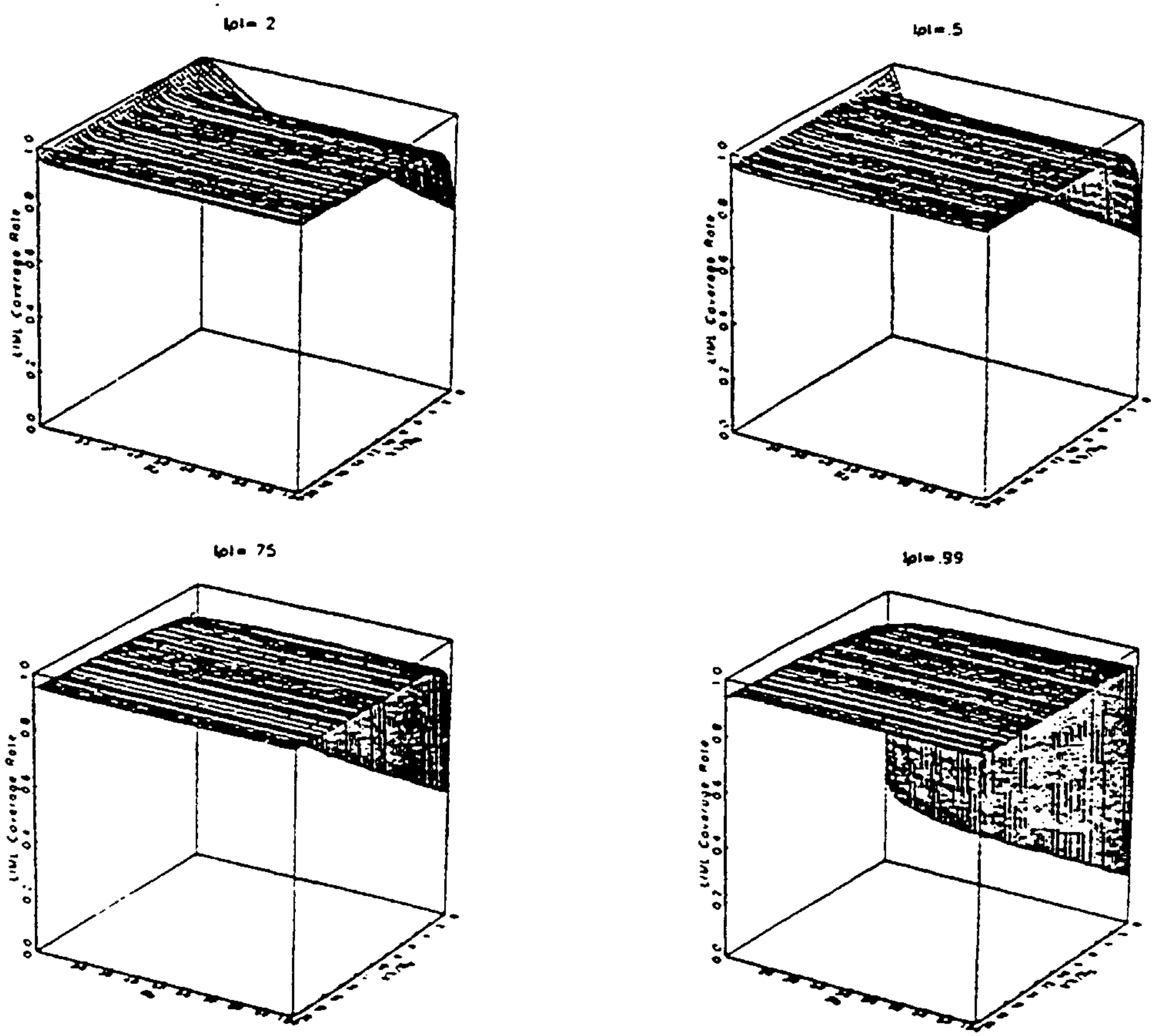

Figure 11. Asymptotic coverage rates for conventional LIML 95% confidence intervals

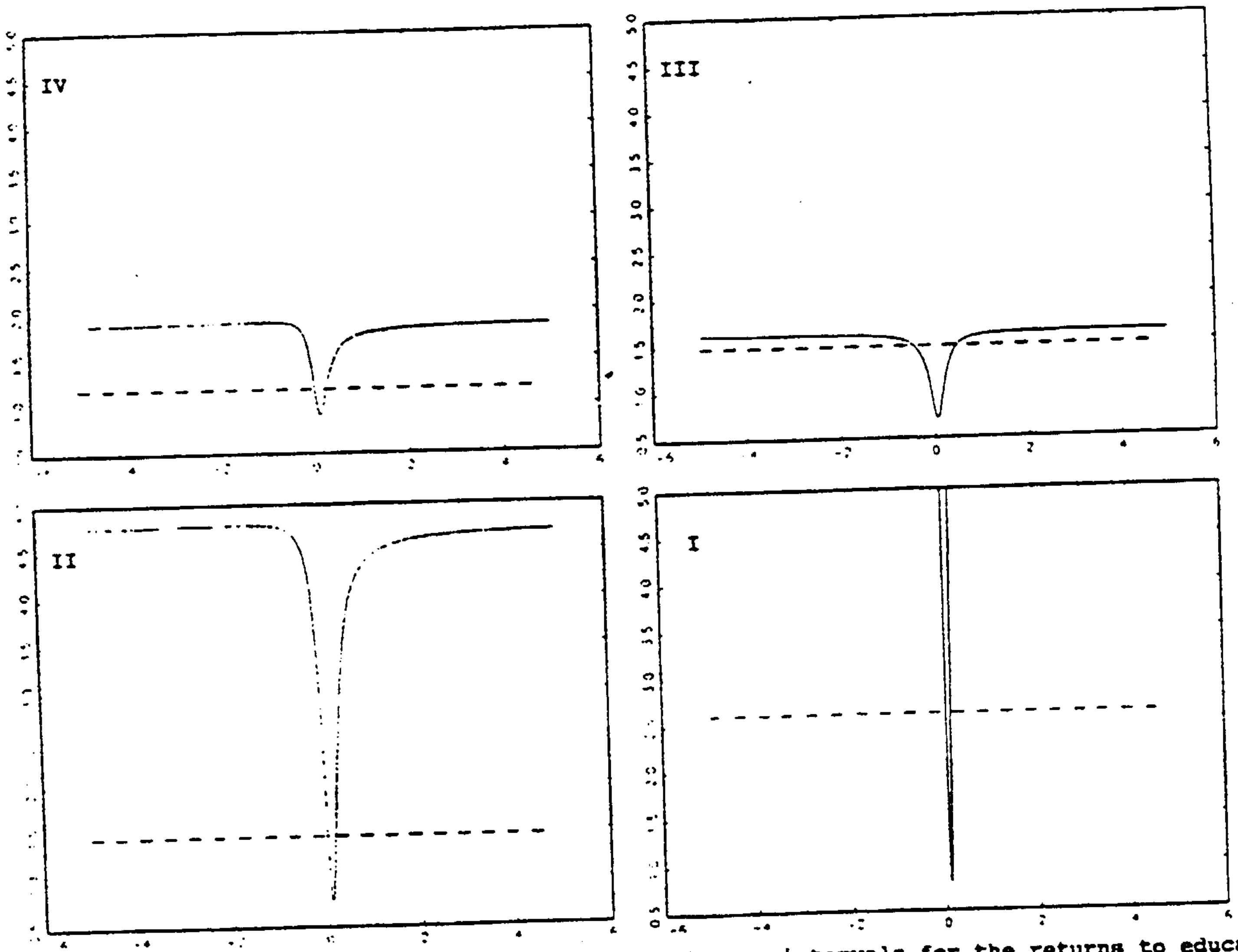

Figure 12. Construction of Anderson-Rubin confidence intervals for the returns to education, 1930-39 birth year cohort, for specifications I, II, III, and IV from table 11. Lines are AR test statistic and 5% asymptotic critical value (horizontal line).

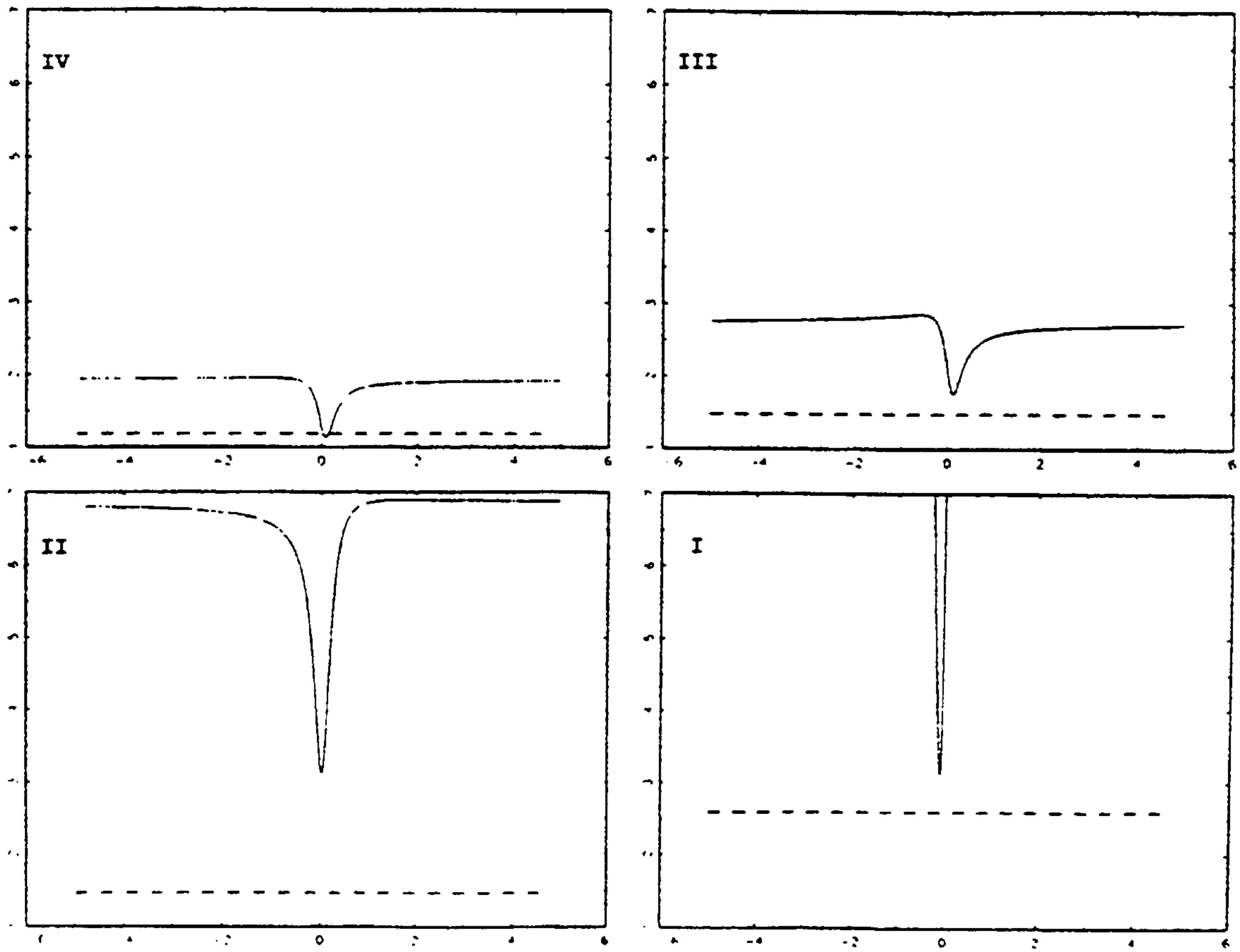

Figure 13. Construction of Anderson-Rubin confidence intervals for the returns to education, 1940-49 birth year cohort, for specifications I, II, III, and IV from table 11. Lines are AR test statistic and 5% asymptotic critical value (horizontal line).