Econometrica, Vol. 65, No. 3 (May, 1997), 557–586

INSTRUMENTAL VARIABLES REGRESSION WITH WEAK INSTRUMENTS

BY DOUGLAS STAIGER AND JAMES H. STOCK[1]

This paper develops asymptotic distribution theory for single-equation instrumental variables regression when the partial correlations between the instruments and the endogenous variables are weak, here modeled as local to zero. Asymptotic representations are provided for various statistics, including two-stage least squares (TSLS) and limited information maximum likelihood (LIML) estimators, Wald statistics, and statistics testing overidentification and endogeneity. The asymptotic distributions are found to provide good approximations to sampling distributions with 10–20 observations per instrument. The theory suggests concrete guidelines for applied work, including using nonstandard methods for construction of confidence regions. These results are used 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 estimates with fewer instruments fall between 8% and 10%, with a typical 95% confidence interval of (5%, 15%).

KEYWORDS: Two stage least squares, LIML, overidentification tests, endogeneity tests.

1. INTRODUCTION

IN EMPIRICAL WORK using instrumental variables (IV) regression, 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 first stage F statistic, which tests the hypothesis that the instruments do not enter the first stage regression, to take on a value less than 10.[2] Unfortunately, it is well known that standard asymptotic approximations to the distributions of the main instrumental variables statistics break down when the mean of this F statistic is small. Recently this has been highlighted for TSLS in quite different settings by Nelson and Startz (1990a,b) using a short sample and a single instrument and by Bound, Jaeger, and Baker (1995) using up to 180 instruments and over 329,000 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

[1] The authors thank Joshua Angrist, John Bound, Adolph Buse, Gary Chamberlain, Jean-Marie Dufour, Jerry Hausman, Guido Imbens, Tom Kane, Alan Krueger, Tom Rothenberg, Richard Startz, and three anonymous referees for helpful discussions and/or comments on earlier drafts, and Matthew Eichner and Jonathan Wright for research assistance. The research was supported in part by National Science Foundation Grant No. SES-91-22463.

[2] It is difficult to provide systematic evidence on this because first stage F statistics are often not reported. For examples, our review of articles published in the *American Economic Review* between 1988 and 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 the standard normal approximations.

conventional asymptotics, the finite-sample approach has several drawbacks which impede its use in practice, including the restrictive assumptions of fixed instruments and Gaussian errors, unwieldy expressions for distributions which can be computationally intractable, a focus on estimators rather than tests or confidence intervals, and the failure to produce clear quantitative guidelines which empirical researchers can follow.

This paper develops an alternative asymptotic framework for approximating distributions of statistics in single-equation IV regression with n endogenous regressors. Conventional asymptotics, both first-order and higher-order such as Edgeworth expansions (cf. Anderson and Sawa (1973, 1979), Morimune (1983, 1989), and Rothenberg (1984)), treat the coefficients on the instruments in the first stage as nonzero and fixed, an assumption which implies that each first stage F statistic increases to infinity with the sample size. Not surprisingly, when the means of these F statistics are small, these asymptotic approximations break down. We therefore adopt a device which, loosely speaking, holds the F statistics constant in expectation as the sample size increases. More precisely, the coefficients on the instruments in the first stage equation are modeled as being in a $T^{-1/2}$ neighborhood of zero; this will be referred to as the "weakly correlated" case. Based on this alternative framework, we derive the asymptotic representations for a number of IV estimators and test statistics, including tests of overidentifying restrictions and tests of exogeneity.

This paper makes three main contributions to the econometric theory of IV regression. First, the finite sample distribution of the TSLS estimator and of the LIMLK approximation to the LIML estimator (cf. Anderson (1977)), previously derived assuming fixed exogenous regressors and normal errors, is shown to apply asymptotically to the TSLS and LIML estimators, respectively, under general conditions (stochastic regressors and nonnormal errors) when instruments are weak. This extends the finite sample results to a much broader set of applications. Second, joint asymptotic representations, for which there are no counterparts in the finite sample literature, are obtained for many IV test statistics. Third, these representations facilitate summarizing in a few figures the relationship of estimator bias and test size to population parameters in a wide range of cases.

The paper makes three further contributions relevant to empirical work. First, the forms of exogeneity tests and overidentification tests which are asymptotically equivalent under conventional asymptotics are not equivalent with weak instruments, and the asymptotic results provide concrete guidance about which tests have greatest power and least size distortions in this case. Second, whereas conventional confidence intervals can be unreliable in the weakly correlated case, we provide several alternative methods of forming confidence intervals that are asymptotically valid with weak instruments. Finally, we provide justification for a readily computed estimator of the maximal bias of TSLS relative to OLS.

These results are used to reexamine Angrist and Krueger's (1991) important study of the returns to education. Using quarter of birth and its interactions with other covariates as instruments for education in an earnings equation, they concluded that OLS estimates are unbiased or perhaps understate the returns to education. However, in several of their specifications, the first stage F statistic is less than 5. Our asymptotic results suggest that TSLS estimates and confidence intervals are unreliable with F's this small even with more than 329,000 observations, and instead Anderson-Rubin (1949) and Bonferroni regions are more reliable. Based on our preferred statistics, we estimate returns to education which are higher, and confidence intervals which are wider, than suggested by Angrist and Krueger.

The literature on the distribution of instrumental variables estimators is large; recent contributions include Bekker (1994), Buse (1992), Choi and Phillips (1992), Hillier (1990), Magdalinos (1990, 1994), Morimune (1989), and Phillips (1989). Mariano (1982) and Phillips (1983) survey earlier contributions. The relationship of our theoretical results to this literature is discussed in Section 3. Our central innovation is the introduction and study of the weakly correlated case. The work most closely related to ours is Bekker (1994), Phillips (1989), and Choi and Phillips (1992). Bekker (1994) develops asymptotics for TSLS and LIML estimators (but not test statistics) in the fixed instruments/Gaussian case where, following Anderson (1976) and Morimune (1983), the number of instruments grows in proportion to the number of observations, whereas we keep the number of instruments fixed; Bekker's (1994) limiting approximations are normal, whereas ours are in general nonstandard. Phillips (1989) and Choi and Phillips (1992) study TSLS asymptotics with fixed parameters in the partially identified case (some linear combinations of the instruments are exactly uncorrelated while others are highly correlated; cf. Sargan (1983) for related work on nonlinear models); in contrast we consider the case that the instruments are weakly correlated, in which exactly uncorrelated instruments are a special case.

The paper is organized as follows. In Section 2, the basic ideas are set out and applied to the TSLS estimator with $n = 1$ and no other regressors. Results for estimators and tests for general n are developed in Section 3. Nonstandard interval estimators are studied in Section 4. Monte Carlo experiments which check the quality of the asymptotic approximation to the finite sample distributions are summarized in Section 5. Section 6 presents the main numerical results, including plots of asymptotic bias and coverage rates. Angrist and Krueger's (1991) data are used in Section 7 to study the returns to education. Section 8 concludes with some lessons for empirical practice.

2. THE MODEL AND AN EXAMPLE

A. The Model, Assumptions, and Notation

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 (2.1) is the structural equation of interest, y and Y are respectively a $T \times 1$ vector and a $T \times n$ matrix of T observations on the endogenous variables, (2.2) is the reduced form equation for Y, X is the $T \times K_1$ matrix of K_1 exogenous regressors, Z is the $T \times K_2$ matrix of K_2 instruments, u and V are respectively a $T \times 1$ vector and a $T \times n$ matrix of error terms, and β, γ, Π, and Φ are unknown parameters. The errors $(u_t \; V_t')'$, where u_t denotes the tth observation on u, etc., are assumed to have mean zero, to be serially uncorrelated, and to be homoskedastic with covariance matrix Σ, partitioned so that $Eu_t^2 = \sigma_{uu}$, $EV_t u_t = \Sigma_{Vu}$, and $EV_t V_t' = \Sigma_{VV}$. Let $\overline{Z} = [X \; Z]$ and let $Q = E\overline{Z}_t \overline{Z}_t'$, partitioned so that $EX_t X_t' = Q_{XX}$, $EX_t Z_t' = Q_{XZ}$, and $EZ_t Z_t' = Q_{ZZ}$. Also let $\rho = \Sigma_{VV}^{-1/2\prime}\Sigma_{Vu}\sigma_{uu}^{-1/2}$. It is assumed throughout that $E\overline{Z}_t(u_t \; V_t') = 0$ and that n, K_1, and K_2 are fixed. With the sole exception of the local power analysis of tests of overidentifying restrictions in Section 3C, (2.1) and (2.2) are assumed to hold throughout.

We are interested in inferences about β and γ when the instrument Z is weakly correlated with Y, specifically when the mean of the first stage F statistic testing $\Pi = 0$ in (2.2) is small or moderate even if T is large. If Π is modeled as fixed, this F statistic tends to infinity with T, which suggests why conventional fixed-Π asymptotics provide poor approximations with weak instruments. In contrast, if Π is modeled as local to zero, this F statistic is $O_p(1)$. We therefore make the following assumption.

ASSUMPTION L_{Π}: *$\Pi = \Pi_T = C/\sqrt{T}$, where C is a fixed $K_2 \times n$ matrix.*

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 typically differ. Let "$\Rightarrow$" denote convergence in distribution.

ASSUMPTION M: *The following limits hold jointly*:

(a) $(u'u/T, V'u/T, V'V/T) \xrightarrow{p} (\sigma_{uu}, \Sigma_{Vu}, \Sigma_{VV})$;

(b) $\overline{Z}'\overline{Z}/T \xrightarrow{p} Q$;

(c) $(T^{-1/2}X'u, T^{-1/2}Z'u, T^{-1/2}X'V, T^{-1/2}Z'V) \Rightarrow (\Psi_{Xu}, \Psi_{Zu}, \Psi_{XV}, \Psi_{ZV})$, *where* $\Psi \equiv (\Psi_{Xu}', \Psi_{Zu}', \mathrm{vec}(\Psi_{XV})', \mathrm{vec}(\Psi_{ZV})')'$ *is distributed* $N(0, \Sigma \otimes Q)$.

These conditions hold under various weak primitive assumptions. For example, if $(u_t \; V_t')'$ is a homoskedastic vector martingale difference sequence with respect to the filtration based on $\{u_{j-1}, V_{j-1}, X_j, Z_j, j \le t\}$, if u_t and V_t have four moments, and if X_t and Z_t are integrated of order zero with four moments and satisfy 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. These conditions arguably apply to linear rational expectations models such as the consumption Euler equation estimated using TSLS by Campbell and Mankiw (1989).[3]

Before proceeding we provide some additional definitions and notation. Let $P_W = W(W'W)^{-1}W'$ and $M_W = I - P_W$ where W is a general $a \times b$ matrix with $a \geq b$, and let "$^\perp$" denote the residuals from the projection on X, so $Z^\perp = M_X Z$, $Y^\perp = M_X Y$, etc. Let $\bar{X} = [Y\ X]$ and $\bar{Y} = [y\ Y]$, and let I_k denote the k-dimensional identity matrix. Define $\lambda = \Omega^{1/2} C \Sigma_{VV}^{-1/2}$, where $\Omega = Q_{ZZ} - Q_{ZX} Q_{XX}^{-1} Q_{XZ}$; $z_u = \Omega^{-1/2\prime}(\Psi_{Zu} - Q_{ZX} Q_{XX}^{-1} \Psi_{Xu}) \sigma_{uu}^{-1/2}$; and $z_V = \Omega^{-1/2\prime}(\Psi_{ZV} - Q_{ZX} Q_{XX}^{-1} \Psi_{XV}) \Sigma_{VV}^{-1/2}$. The random variable $[z_u' \ \mathrm{vec}(z_V)']'$ is distributed $N(0, \bar{\Sigma} \otimes I_{K_2})$, where $\bar{\Sigma}$ is the $(n+1) \times (n+1)$ matrix with $\bar{\Sigma}_{11} = 1$, $\bar{\Sigma}_{22} = I_n$, $\bar{\Sigma}_{12} = \rho'$, and $\bar{\Sigma}_{21} = \rho$, where $\bar{\Sigma}$ is partitioned conformably with Σ. Finally, let

(2.3a) $\quad \nu_1 = (\lambda + z_V)'(\lambda + z_V),$

(2.3b) $\quad \nu_2 = (\lambda + z_V)' z_u.$

If Σ_{Vu} is nonzero, then Y is endogenous and the OLS estimator of β, $\hat{\beta}_{\mathrm{OLS}}$, is inconsistent. Let β_0 denote the true value of β and let $\theta = \Sigma_{VV}^{-1} \Sigma_{Vu}$. Then, under Assumptions L_Π and M,

(2.4) $\quad \hat{\beta}_{\mathrm{OLS}} \xrightarrow{p} \beta_0 + \theta.$

Note that θ can alternatively be expressed in terms of the vector of correlations between the first and second stage errors, $\theta = \sigma_{uu}^{1/2} \Sigma_{VV}^{-1/2} \rho$, so that if $\rho \neq 0$ then OLS is inconsistent.

Most of the theorems in Section 3 are developed for the general k-class estimator, $\hat{\beta}(k)$, and associated tests. However, for clarity we will often use familiar subscripts, e.g. $\hat{\beta}_{\mathrm{TSLS}}$ for $\hat{\beta}(1)$.

B. An Example: The TSLS Estimator with $n = 1$ and no X's

Consider the special case of $\hat{\beta}_{\mathrm{TSLS}}$ when $n = 1$ and $K_1 = 0$. By Assumptions L_Π and M,

$$\begin{aligned} T^{-1/2} Z'Y &= T^{-1/2} Z'(Z\Pi + V) \\ &= (T^{-1} Z'Z) C + T^{-1/2} Z'V \Rightarrow Q_{ZZ} C + \Psi_{ZV}. \end{aligned}$$

Thus

$$\begin{aligned} Y' P_Z Y &= (T^{-1/2} Y'Z)(T^{-1} Z'Z)^{-1}(T^{-1/2} Z'Y) \\ &\Rightarrow (Q_{ZZ} C + \Psi_{ZV})' Q_{ZZ}^{-1} (Q_{ZZ} C + \Psi_{ZV}) \\ &= \Sigma_{VV}^{1/2\prime} (\lambda + z_V)'(\lambda + z_V) \Sigma_{VV}^{1/2}. \end{aligned}$$

[3] These conditions could be extended to deterministic trends (by adopting a diagonal scaling matrix of the form $\mathrm{diag}(T^{1/2}, T)$) and to autocorrelated errors. Both extensions are conceptually straightforward but would complicate notation and thereby obscure the main results.

Similarly, $Y'P_Z u \Rightarrow (Q_{ZZ}C + \Psi_{ZV})'Q_{ZZ}^{-1}\Psi_{Zu} = \sigma_{uu}^{1/2}\Sigma_{VV}^{1/2\prime}(\lambda + z_V)'z_u$. These expressions and the definitions (2.3) yield

$$(2.5) \qquad \hat{\beta}_{\mathrm{TSLS}} - \beta_0 = (Y'P_Z Y)^{-1}(Y'P_Z u) \Rightarrow \sigma_{uu}^{1/2}\Sigma_{VV}^{-1/2}\nu_1^{-1}\nu_2 \equiv \beta_{\mathrm{TSLS}}^{*}.$$

Equation (2.5) expresses $\hat{\beta}_{\mathrm{TSLS}} - \beta_0$ as, asymptotically, the ratio of quadratic forms in the two $K_2 \times 1$ jointly normal random variables, z_u and z_V. This limiting distribution can be expressed as the random mixture of normals, $\int N(\beta_0 + \theta m(z_V), \mathrm{var}(z_V))\, dF(z_V)$, where $m(z_V) = (\lambda + z_V)'z_V/(\lambda + z_V)'(\lambda + z_V)$, $\mathrm{var}(z_V) = \tau^2/(\lambda + z_V)'(\lambda + z_V)$, and $\tau = [(1-\rho^2)\sigma_{uu}/\Sigma_{VV}]^{1/2}$ (this follows by rewriting $\beta_{\mathrm{TSLS}}^{*}$ in terms of the orthogonalized variable, $\eta = (z_u - z_V\rho)/\sqrt{1-\rho'\rho}$). In particular, the asymptotic bias of $\hat{\beta}_{\mathrm{TSLS}}$, relative to the asymptotic bias of OLS, is

$$(2.6) \qquad E\beta_{\mathrm{TSLS}}^{*}/\theta = Em(z_V).$$

Because the distribution of $m(z_V)$ depends only on $\lambda'\lambda/K_2$ and K_2,[4] the asymptotic bias of TSLS, relative to OLS, depends on $\lambda'\lambda/K_2$ and K_2 but not on θ or ρ.

3. ASYMPTOTIC RESULTS IN THE GENERAL CASE

A. *k-Class Estimators and Wald Statistics*

We now consider general n, K_1, and K_2. The k-class estimator of $[\beta' \ \gamma']'$ is

$$(3.1) \qquad \left[\hat{\beta}(k)' \ \hat{\gamma}(k)'\right]' = \left[\bar{X}'(I - kM_{\bar{Z}})\bar{X}\right]^{-1}\left[\bar{X}'(I - kM_{\bar{Z}})y\right].$$

By standard projection arguments, the k-class estimator of β is

$$(3.2) \qquad \hat{\beta}(k) = \left[Y^{\perp\prime}(I - kM_{Z^\perp})Y^{\perp}\right]^{-1}\left[Y^{\perp\prime}(I - kM_{Z^\perp})y^{\perp}\right].$$

Two leading cases of interest are the TSLS estimator, for which $k = 1$, and the LIML estimator. The LIML estimator is given by (3.1) (equivalently for β, (3.2)) with $k = \hat{k}_{\mathrm{LIML}}$, where $\hat{k}_{\mathrm{LIML}}$ is the smallest root of the determinantal equation $|\bar{Y}'M_X\bar{Y} - k\bar{Y}'M_{\bar{Z}}\bar{Y}| = 0$.

A standard formula for the Wald statistic testing q linear restrictions $R\beta = r$, where R is $q \times n$, is

$$(3.3) \qquad W(k) = \left[R\hat{\beta}(k) - r\right]'\left\{R\left[Y^{\perp\prime}(I - kM_{Z^\perp})Y^{\perp}\right]^{-1}R'\right\}^{-1} \times \left[R\hat{\beta}(k) - r\right]/[q\hat{\sigma}_{uu}(k)]$$

[4] This and subsequent statements about dependence of distributions on only K_2, $\lambda'\lambda/K_2$, and ρ follow from noting that these asymptotic representations are continuous functions of the random variable $\Xi_0^{*} = [z_u \ (\lambda + z_V)]'[z_u \ (\lambda + z_V)]$, which has the noncentral Wishart distribution $W_{n+1}(K_2, \bar{\Sigma}, \Lambda)$ (Muirhead (1982, pp. 441-442)), where $\Lambda = \bar{\Sigma}^{-1}\bar{\Lambda}$, where $\bar{\Lambda}_{11} = \bar{\Lambda}_{12} = \bar{\Lambda}_{21} = 0$ and $\bar{\Lambda}_{22} = \lambda'\lambda$, where $\bar{\Lambda}$ is partitioned conformably with $\bar{\Sigma}$. In particular $(\lambda + z_V)'(\lambda + z_V) \sim W_n(K_2, I_n, \lambda'\lambda)$.

where $\hat{\sigma}_{uu}(k) = \hat{u}(k)'\hat{u}(k)/(T - K_1 - n)$, where $\hat{u}(k) = y - Y\hat{\beta}(k) - X\hat{\gamma}(k) = y^{\perp} - Y^{\perp}\hat{\beta}(k)$. Similarly, a standard formula for the t ratio testing the hypothesis that a single coefficient, β_i, takes on the value β_{i0} is, $t_i(k) = [\hat{\beta}_i(k) - \beta_{i0}]/\{[Y^{\perp\prime}(I - kM_{Z^{\perp}})Y^{\perp}]^{ii}\hat{\sigma}_{uu}(k)\}^{1/2}$, where B^{ij} denotes the (i,j) element of B^{-1} for general square nonsingular matrix B (cf. Hall, Cummins, and Schake (1992, p. 145–146)).

Tests of the hypothesis that $\Pi = 0$ play an important role in this development. Accordingly, let

$$G_T = \hat{\Sigma}_{VV}^{-1/2\prime} Y^{\perp\prime} P_{Z^{\perp}} Y^{\perp} \hat{\Sigma}_{VV}^{-1/2} \tag{3.4}$$

where $\hat{\Sigma}_{VV} = Y'M_{\bar{Z}}Y/(T - K_1 - K_2)$. Note that $\operatorname{tr}(G_T)/nK_2$ is the Wald statistic testing $\Pi = 0$.

The limiting distributions of these statistics are given in Theorem 1.

THEOREM 1: *Suppose that* (2.1), (2.2), *and Assumptions* L_{Π} *and* M *hold. Also suppose that* $T(k-1) \Rightarrow \kappa$ *jointly with the limits in Appendix Lemma* A1, *where* $\kappa = O_p(1)$ (*possibly a constant*). *Then the following limits hold jointly*:

(a) $\hat{\beta}(k) - \beta_0 \Rightarrow \beta_0^*(\kappa) = \sigma_{uu}^{1/2}\Sigma_{VV}^{-1/2}\Delta_0^*(\kappa)$, *where* $\Delta_0^*(\kappa) = (\nu_1 - \kappa I_n)^{-1}(\nu_2 - \kappa\rho)$.

(b) $\hat{\sigma}_{uu}(k) \Rightarrow \sigma_{uu}^*(\kappa) = \sigma_{uu}S_1(\Delta_0^*(\kappa))$, *where* $S_1(b) = 1 - 2\rho'b + b'b$.

(c) *Under the null hypothesis* $R\beta = r$,

$$\begin{aligned} W(k) &\Rightarrow \beta_0^*(\kappa)'R'\left[R\Sigma_{VV}^{-1/2}(\nu_1 - \kappa I_n)^{-1}\Sigma_{VV}^{-1/2\prime}R'\right]^{-1} \\ &\qquad \times R\beta_0^*(\kappa)/[q\sigma_{uu}^*(\kappa)] \\ &= \Delta_0^*(\kappa)'\Sigma_{VV}^{-1/2\prime}R'\left[R\Sigma_{VV}^{-1/2}(\nu_1 - \kappa I_n)^{-1}\Sigma_{VV}^{-1/2\prime}R'\right]^{-1} \\ &\qquad \times R\Sigma_{VV}^{-1/2}\Delta_0^*(\kappa)/[qS_1(\Delta_0^*(\kappa))]. \end{aligned}$$

(d) *Under the null hypothesis* $\beta_i = \beta_{i0}$,

$$t_i(k) \Rightarrow \left\{\left[\Sigma_{VV}^{-1/2}(\nu_1 - \kappa I_n)^{-1}\Sigma_{VV}^{-1/2\prime}\right]^{ii} S_1(\Delta_0^*(\kappa))\right\}^{-1/2}\left[\Sigma_{VV}^{-1/2}\Delta_0^*(\kappa)\right]_i.$$

(e) $G_T \Rightarrow \nu_1$.

Proofs of all theorems are given in the Appendix.

Theorem 2 provides the limiting behavior of $\hat{k}_{\mathrm{LIML}}$ which, when combined with Theorem 1, gives the asymptotic distribution of the LIML estimator and test statistics.

THEOREM 2: *Suppose that* (2.1), (2.2), *and Assumptions* L_{Π} *and* M *hold. Then* $T(\hat{k}_{\mathrm{LIML}} - 1) \Rightarrow \kappa_{\mathrm{LIML}}^*$, *where* κ_{LIML}^* *is the smallest root of the determinantal equation*, $|\Xi_0^* - \kappa\bar{\Sigma}| = 0$, *where* $\Xi_0^* = [z_u \ \ (\lambda + z_V)]'[z_u \ \ (\lambda + z_V)]$, *where the convergence is joint with the limits in Appendix Lemma* A1.

When the instruments are weak, in general $\hat{\beta}(k)$ is not consistent and has a nonstandard asymptotic distribution. Moreover, $T(\hat{k}_{\mathrm{LIML}}-1)$ has a nondegenerate asymptotic distribution so $\hat{\beta}_{\mathrm{TSLS}}$ and $\hat{\beta}_{\mathrm{LIML}}$ are not equivalent under weak instrument asymptotics. The asymptotic distributions of the test statistics and $\hat{\sigma}_{uu}(k)$ are also nonstandard. The limiting representations in Theorems 1 and 2 depend on σ_{uu}, Σ_{VV}, ρ, and $\lambda'\lambda/K_2$, K_2, and n. However, K_2 and n are known and Σ_{VV} and, given ρ, σ_{uu} are consistently estimable, so $\lambda'\lambda/K_2$ and ρ are the only asymptotically unknown parameters entering the distributions.[5] In some cases, the dependence on Σ_{VV} and σ_{uu} disappears, for example, when $n=1$, the distributions of $(\hat{\beta}_{\mathrm{TSLS}}-\beta_0)/\theta$ and $(\hat{\beta}_{\mathrm{LIML}}-\beta_0)/\theta$ depend only on $\lambda'\lambda/K_2$, K_2, and $|\rho|$, and the distributions of t_{TSLS} and t_{LIML} depend only on $\lambda'\lambda/K_2$, K_2, and ρ and their pdf's are antisymmetric in ρ. The parameter $\lambda'\lambda/K_2$ has a simple interpretation: when $n=1$, the first stage F statistic G_T/K_2 converges to a noncentral $\chi^2_{K_2}$, divided by the number of instruments K_2, with noncentrality parameter $\lambda'\lambda$, and when $n>1$, $\lambda'\lambda$ is the matrix of noncentrality parameters of the limiting noncentral Wishart random variable ν_1. Although $\lambda'\lambda/K_2$ is identified, it is not consistently estimable under these asymptotics.

The results for $\hat{\beta}(k)$ extend some known results in the exact distribution literature for the fixed instrument/Gaussian model. The distribution of the limiting representation of $\hat{\beta}_{\mathrm{TSLS}}$, $\beta_0^*(0)$, is the same as the exact distribution of $\hat{\beta}_{\mathrm{TSLS}}$ in the fixed instrument/Gaussian case, obtained by Richardson (1968) and Sawa (1969) for $n=1$ and by Phillips (1980) for general n. (This is most easily seen by noting that $\hat{\beta}_{\mathrm{TSLS}}$ depends only on the moments in Assumptions $M(b)$ and $M(c)$ and that, with fixed instruments and Gaussian errors, those assumptions hold as equalities). Also, the asymptotic distribution of $\hat{\beta}_{\mathrm{LIML}}$ is the same as the exact distribution of the so-called LIMLK estimator (an infeasible estimator which requires the reduced form error covariance matrix to be known) in the fixed instrument/Gaussian case (cf. Anderson (1977)). Thus Theorems 1 and 2 extend the finite sample result for estimators previously derived under the highly restrictive fixed instrument/Gaussian assumptions to the more general conditions which lead to Assumption M. While existing formulas for estimator distributions typically involve multiple infinite series expansions, the representations given here provide a simple framework for numerical evaluation of joint asymptotic distributions by Monte Carlo simulation. Finally, although LIML and LIMLK differ when the concentration parameter, of which $\lambda'\lambda$ is the probability limit in our notation, and T are finite, Anderson (1977) showed that for $n=1$ the exact fixed instrument/Gaussian LIML and LIMLK distributions converge as the concentration parameter increases to infinity (with K_2 and T fixed). Theorem 2 extends this result by implying that the LIML and LIMLK distributions converge as $T\to\infty$ for fixed $\lambda'\lambda/K_2$ and general n under Assumption L_{Π}.

The representations of the Wald and t statistics have no counterpart in the exact distribution literature, since these statistics have not yielded to finite-

[5] Given ρ, σ_{uu} can be estimated consistently by $\hat{\sigma}_{uu,\mathrm{OLS}}/(1-\rho'\rho)$, where $\hat{\sigma}_{uu,\mathrm{OLS}}=\hat{\sigma}_{uu}(0)$.

sample analysis (cf. Mariano (1982)), so these approximations are new even in the fixed instrument/Gaussian case. Because the t statistic does not have a normal asymptotic distribution, confidence intervals constructed as ± 1.96 standard errors will not in general have a 95% coverage rate, even asymptotically. Rather, the limiting representation of $t(k)$ indicates that the distribution depends on ρ in a complicated way, not just as a mean or scale shift. Thus confidence belts will depend nonlinearly on ρ. Worse, this distribution also depends on $\lambda'\lambda/K_2$, so that the confidence belts must be indexed by $\lambda'\lambda/K_2$. Because $\lambda'\lambda/K_2$ is not consistently estimable, $W(k)$ and $t(k)$ cannot be inverted directly to construct asymptotic confidence regions for β.

The limiting distributions in Theorem 1 simplify to the conventional asymptotic results when $\lambda'\lambda$ is large and K_2 is fixed. Consider the TSLS estimator for general n and ρ. If $\lambda'\lambda$ is large, then $\nu_1 = \lambda'\lambda + O_p(\|\lambda'\lambda\|^{1/2}) + O_p(1)$ (where $\|B\| = \max_{i,j}|B_{i,j}|$) and $\nu_2 = \lambda' z_u + O_p(1)$, so the distribution of $\beta_0^*(1)$ is approximately $N(0, (\Sigma_{VV}^{1/2\prime}\lambda'\lambda\Sigma_{VV}^{1/2})^{-1}\sigma_{uu})$, which is the standard fixed-Π asymptotic normal approximation for $\hat{\beta}_{\text{TSLS}}$. Similarly, $\hat{\sigma}_{uu}(1) = \sigma_{uu}[1 + O_p(\|\lambda'\lambda\|^{-1/2}) + O_p(\|\lambda'\lambda\|^{-1})]$, so the limit of $W(1)$ in Theorem 1(c) is well approximated by a χ_q^2/q, the usual result.[6]

B. *Measures of Bias of the TSLS Estimator*

Sargan (1958) reported that the bias of TSLS was of the order of the inverse of the minimum population canonical correlation between Y and Z. The sample minimum canonical correlation has also been discussed in the context of identification and testing for instrument relevance (Sargan (1958), Bowden and Turkington (1984)). Bekker (1994) interpreted it as a measure of TSLS relative bias in the fixed instrument/Gaussian model with $K_2 \to \infty$. This section provides an asymptotic interpretation of this statistic as a measure of the bias resulting from weak instruments.

Consider the squared bias of $\hat{\beta}_{\text{TSLS}}$ relative to $\hat{\beta}_{\text{OLS}}$ in OLS standard error units,

$$B^2 = \left(E\hat{\beta}_{\text{TSLS}} - \beta_0\right)' \Sigma_{Y^\perp Y^\perp}\left(E\hat{\beta}_{\text{TSLS}} - \beta_0\right) \Big/ \left(E\hat{\beta}_{\text{OLS}} - \beta_0\right)' \Sigma_{Y^\perp Y^\perp}\left(E\hat{\beta}_{\text{OLS}} - \beta_0\right) \to \rho' h' h \rho / \rho'\rho \tag{3.5}$$

where $h = E[\nu_1^{-1}(\lambda + z_V)' z_V]$ and $\Sigma_{Y^\perp Y^\perp} = \text{plim}(Y^{\perp\prime} Y^\perp / T)$ (under Assumption L_Π but not in general, $\Sigma_{Y^\perp Y^\perp} = \Sigma_{VV}$). An advantage of this measure is

[6] Another special case of interest is when $\lambda'\lambda = 0$. Then the results in Theorem 1 simplify to those obtained using fixed-Π asymptotics with $\Pi = 0$ by Phillips (1989) and Choi and Phillips (1992). With many irrelevant instruments and/or $|\rho|$ nearly one, $\hat{\beta}_{\text{TSLS}}$ tends to fall in a tight neighborhood of $\beta_0 + \theta$, and its estimated standard error is typically "too small," so tests based on t_{TSLS} incorrectly reject the null too often. Also, Theorem 2 is readily extended to modified LIML estimators. For example, Fuller (1977) proposed using $\hat{k}_F = \hat{k}_{\text{LIML}} - l/(T - K_1 - K_2)$ (cf. Morimune (1983)); for fixed K_1 and K_2, $T(\hat{k}_F - 1) \to \kappa_{\text{LIML}}^* - l$.

that it is invariant to the transformation $(Y, \beta, \Pi, \Phi, V) \Rightarrow (YA, A^{-1}\beta, \Pi A, \Phi A, VA)$, a special case of which is a change in the units of Y. For $n > 1$, numerical evaluation of B requires knowledge of ρ. Typically a candidate value of ρ is unavailable, so it is desirable to have a measure of the total relative bias which does not depend on ρ. This can be done by considering the worst case squared relative bias. Because numerical evaluation of h is somewhat cumbersome, we make the approximation that, for K_2 and/or $\lambda'\lambda/K_2$ large, $h \doteq (E\nu_1)^{-1}E[(z_V+\lambda)'z_V] = (I + \lambda'\lambda/K_2)^{-1}$. Then $B^2 \le \max_\rho B^2 = [\text{maxEval}(h'h)]^2 \doteq \{\text{minEval}(I+\lambda'\lambda/K_2)\}^{-2} \equiv B_{\max}^2$. Although $\lambda'\lambda/K_2$ cannot be consistently estimated, $\Sigma_{Y^\perp Y^\perp}^{-1/2\prime} Y^{\perp\prime} P_{Z^\perp} Y^\perp \Sigma_{Y^\perp Y^\perp}^{-1/2} \Rightarrow \nu_1$ and $E\nu_1/K_2 = I + \lambda'\lambda/K_2$. This suggests the statistic

$$(3.6) \qquad \hat{B}_{\max} = \left\{\text{minEval}\left[Y^{\perp\prime}P_{Z^\perp}Y^\perp (Y^{\perp\prime}Y^\perp/T)^{-1}/K_2\right]\right\}^{-1}$$

which is K_2/T times the inverse of the minimum squared sample canonical correlation between $Y^\perp$ and $Z^\perp$. Note that a statistic which is asymptotically equivalent under these assumptions (but not if Π is fixed) is

$$(3.7) \qquad \tilde{B}_{\max} = \{\text{minEval}(G_T/K_2)\}^{-1}.$$

When $n = 1$, $\hat{B}_{\max} = \tilde{B}_{\max} = (G_T/K_2)^{-1}$, where G_T/K_2 is the first stage F statistic testing $\Pi = 0$. Under our assumptions, the statistics (3.6) and (3.7) are asymptotically distributed as the inverse of the minimum of n noncentral independent χ^2 random variables with noncentrality parameters $\text{eig}(\lambda'\lambda/K_2)$ (cf. Anderson (1984)).

The statistics $\hat{B}_{\max}$ and $\tilde{B}_{\max}$ provide a data based measure of the worst case bias of TSLS over all ρ, relative to OLS, after the coefficients have been transformed by $\Sigma_{Y^\perp Y^\perp}^{-1/2}$. An advantage of these statistics is that they are relatively simple to compute and have a straightforward interpretation, which is simplest if $n = 1$ in which case the statistics simply measure relative bias directly (i.e. not worst case bias). For all n, it should be recognized that these are only sample measures and that bias is related to their population counterpart. Nonetheless, large values of this bias measure should alert the researcher to potential problems with instrument endogeneity.

C. Tests of Overidentifying Restrictions

This section studies the limiting behavior of two common tests of overidentifying restrictions, which are available when $K_2 > n$, under the null and under a local alternative. The test statistics are TR^2 from a regression of the IV regression residuals on the instruments and exogenous variables (here denoted ϕ_{reg}), and Basmann's (1960) test statistics (ϕ_{Bas}). The tests are analyzed for residuals from a general k-class regression, although TSLS is used most commonly in applications. Expressed in χ^2 form, the two statistics differ only in their denominators and are

(3.8a) $\phi_{\text{reg}}(k) = [\hat{u}(k)' P_{Z^\perp} \hat{u}(k)]/[\hat{u}(k)'\hat{u}(k)/T]$,

(3.8b) $\phi_{\text{Bas}}(k) = [\hat{u}(k)' P_{Z^\perp} \hat{u}(k)]/[\hat{u}(k)' M_{Z^\perp} \hat{u}(k)/(T - K_1 - K_2)]$.

Even with normal errors these statistics (divided by K_2) do not have exact F null distributions.

The power of these tests against violations of the orthogonality conditions are investigated by deriving their asymptotic representations under the local alternative:

ASSUMPTION L_ω: *$y = Y\beta + X\gamma + Z\omega + u$, where $\omega = \omega_T = T^{-1/2}d$, where d is a K_2-vector of constants.*

(This assumption is used only in this subsection.) The null hypothesis is that $\omega = 0$. The local alternative in Assumption L_ω is the natural one under standard fixed-Π asymptotics, and it also delivers a nontrivial representation for the statistics (3.8) under Assumption L_Π:

THEOREM 3: *Suppose that* (2.2) *and Assumptions M, L_Π, and L_ω hold. Let $\xi = \Omega^{1/2} d\sigma_{uu}^{-1/2}$.*

(a) *Further suppose that $T(k-1) \Rightarrow \kappa$ jointly with the limits in Appendix Lemma A1, where $\kappa = O_p(1)$ (possibly a constant). Then (i) $\hat{\beta}(k) - \beta_0 \Rightarrow \beta_\xi^*(\kappa)$, where $\beta_\xi^*(\kappa) = \sigma_{uu}^{1/2}\Sigma_{VV}^{-1/2}\Delta_\xi^*(\kappa)$, where $\Delta_\xi^*(\kappa) = (\nu_1 - \kappa I_n)^{-1}[(z_V + \lambda)'(z_u + \xi) - \kappa\rho]$; and (ii) ϕ_{Bas}, $\phi_{\text{reg}} \Rightarrow S_2(\Delta_\xi^*(\kappa), \xi)/S_1(\Delta_\xi^*(\kappa))$ and $\phi_{\text{Bas}} - \phi_{\text{reg}} \xrightarrow{p} 0$, where $S_2(b,c) = [z_u - (\lambda + z_V)b + c]'[z_u - (\lambda + z_V)b + c]$.*

(b) *$T(\hat{k}_{\text{LIML}} - 1) \Rightarrow \kappa_{\text{LIML},\xi}^*$, where $\kappa_{\text{LIML},\xi}^*$ is the smallest root of the determinantal equation, $|\Xi_\xi^* - \kappa\overline{\Sigma}| = 0$, where $\Xi_\xi^* = [(z_u + \xi)(\lambda + z_V)]'[(z_u + \xi)(\lambda + z_V)]$, where the convergence is joint with the limits in Appendix Lemma A1.*

Theorem 3(a)(i) elucidates the bias of the TSLS estimator when there are small violations of the orthogonality restrictions. If $\lambda'\lambda$ is large and $\lambda'\xi$ is small, then these small violations impart negligible bias. In the completely unidentified case, the presence of nonzero d increases the spread of the distribution but does not affect the bias.

The Basmann and regression tests are asymptotically equivalent under the null and the local alternative. Inspection of the expression for S_2 reveals that for general $\lambda'\lambda/K_2$ neither test has a χ^2 asymptotic null distribution under L_Π: although z_u is normally distributed, $\hat{\beta}_{\text{TSLS}}$ is $O_p(1)$ which makes the asymptotic distribution nonstandard.

D. Tests of Exogeneity

The Durbin-Wu-Hausman (DWH) test examines the null hypothesis that Y is exogenous (that $\rho = 0$) by checking for a statistically significant difference between the OLS and TSLS estimates of β. There are various versions of this test, three of which are

$$F_{\mathrm{DWH},i} = \left(\hat{\beta}_{\mathrm{TSLS}} - \hat{\beta}_{\mathrm{OLS}}\right)' V_i^{-1}\left(\hat{\beta}_{\mathrm{TSLS}} - \hat{\beta}_{\mathrm{OLS}}\right) \qquad (i=1,2,3), \tag{3.9}$$

where

$$V_1 = (Y^{\perp\prime}P_{Z^\perp}Y^{\perp})^{-1}\hat{\sigma}_{uu,\mathrm{TSLS}} - (Y^{\perp\prime}Y^{\perp})^{-1}\hat{\sigma}_{uu,\mathrm{OLS}},$$

$$V_2 = \left[(Y^{\perp\prime}P_{Z^\perp}Y^{\perp})^{-1} - (Y^{\perp\prime}Y^{\perp})^{-1}\right]\hat{\sigma}_{uu,\mathrm{TSLS}}, \quad \text{and}$$

$$V_3 = \left[(Y^{\perp\prime}P_{Z^\perp}Y^{\perp})^{-1} - (Y^{\perp\prime}Y^{\perp})^{-1}\right]\hat{\sigma}_{uu,\mathrm{OLS}}$$

(where $\hat{\sigma}_{uu,\mathrm{TSLS}} = \hat{\sigma}_{uu}(1)$ in the notation of (3.3)). $F_{\mathrm{DWH},2}$ was proposed by Wu (1973; his T_3 statistic) and by Hausman (1978). $F_{\mathrm{DWH},3}$ was proposed by Durbin (1954) and will be referred to as the Durbin form of the test.

From Theorem 1 and Lemma A1, these statistics have the limiting representations

$$F_{\mathrm{DWH},i} \Rightarrow [\Delta_0^*(0) - \rho]'\nu_1[\Delta_0^*(0) - \rho]/S_1(\Delta_0^*(0)) \qquad (i=1,2), \tag{3.10a}$$

$$F_{\mathrm{DWH},3} \Rightarrow [\Delta_0^*(0) - \rho]'\nu_1[\Delta_0^*(0) - \rho]/S_1(\rho). \tag{3.10b}$$

Because the limits apply for general ρ, (3.10) yields asymptotic null distributions and power functions. When $\lambda'\lambda = 0$, the limits in (3.10) do not depend on ρ so their asymptotic power equals their size. Because the tests are $O_p(1)$ for finite $\lambda'\lambda/K_2$, they are not consistent.

Under the null hypothesis $\rho = 0$, (3.10a) simplifies to $F_{\mathrm{DWH},i} \Rightarrow \zeta'\zeta/(1 + \zeta'\nu_1^{-1}\zeta)$, $i = 1, 2$, where $\zeta = \nu_1^{-1/2\prime}(\lambda + z_V)'\eta \sim N(0, I_n)$ (where $\eta = (z_u - z_V\rho)/\sqrt{1-\rho'\rho}$) and ζ and ν_1 are independent. Because $\zeta'\zeta/(1 + \zeta'\nu_1^{-1}\zeta) \leq \zeta'\zeta \sim \chi_n^2$, applying χ_n^2 critical values to $F_{\mathrm{DWH},1}$ and $F_{\mathrm{DWH},2}$ results in asymptotically conservative tests. Size adjustment of $F_{\mathrm{DWH},1}$ and $F_{\mathrm{DWH},2}$ is infeasible because their size depends on $\lambda'\lambda/K_2$. In contrast, from (3.10b), $F_{\mathrm{DWH},3}$ has an asymptotic distribution which is a mixture of noncentral χ_n^2's with random noncentrality parameter $\rho'\lambda'P_{z_V+\lambda}\lambda\rho/(1-\rho'\rho)$. Thus, under the null $\rho = 0$, $F_{\mathrm{DWH},3} \Rightarrow \chi_n^2$, and when $\lambda'\lambda > 0$ the test has power that increases with $\lambda'\lambda/K_2$ and with $|\rho|$. Under the alternative $|\rho| \neq 0$, $S_1(\Delta_0^*(0)) - S_1(\rho)$ is a nonnegative $O_p(1)$ random variable, so $F_{\mathrm{DWH},3}$ has greater asymptotic power than $F_{\mathrm{DWH},1}$ or $F_{\mathrm{DWH},2}$ (cf. Wu (1974)). This suggests using $F_{\mathrm{DWH},3}$ when instruments are weak.

E. Distribution of k-Class Estimator of the Coefficients on Exogenous Variables

The asymptotic representation of the k-class estimator of γ, $\hat{\gamma}(k)$, is examined in the case that X weakly enters (2.2). Specifically, it is assumed that Φ in (2.2) is local to zero:

ASSUMPTION L_Φ: *$\Phi = \Phi_T = T^{-1/2}H$, where H is a fixed $K_1 \times n$ matrix.*

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 unidentified. In the fixed-Φ case, the regressor moment matrix is asymptotically singular and the identified and weakly correlated linear combinations need to be treated separately, as done by Phillips (1989) in the partially identified case. In contrast, letting Φ be local to zero permits the unified treatment in the next theorem.

THEOREM 4: *Suppose that* (2.1), (2.2), *and Assumptions* M, L_Π, *and* L_Φ *hold, and that* $T(k-1) \Rightarrow \kappa$ *jointly with the limits in Appendix Lemma* A1, *where* $\kappa = O_p(1)$ (*possibly a constant*). *Then*,

$$T^{1/2}(\hat{\gamma}(k) - \gamma_0) \Rightarrow \sigma_{uu}^{1/2} Q_{XX}^{-1/2}[z_{Xu} - (\mu + z_{XV})\Delta_0^*(\kappa)], \quad \text{where}$$

$$\mu = Q_{XX}^{-1/2\prime} Q_{XZ} C \Sigma_{VV}^{-1/2} + Q_{XX}^{1/2} H \Sigma_{VV}^{-1/2},$$

$$z_{Xu} = Q_{XX}^{-1/2\prime} \Psi_{Xu} \sigma_{uu}^{-1/2}, \quad \text{and} \quad z_{XV} = Q_{XX}^{-1/2\prime} \Psi_{XV} \Sigma_{VV}^{-1/2},$$

where $[z_{Xu}' \;\; z_u' \;\; \mathrm{vec}(z_{XV})' \;\; \mathrm{vec}(z_V)']'$ *is distributed* $N(0, \bar{\Sigma} \otimes I_{K_1+K_2})$.

The representation in Theorem 4 combined with the previous results provides joint representations of $\hat{\beta}_{\mathrm{TSLS}}$, $\hat{\gamma}_{\mathrm{TSLS}}$, $\hat{\beta}_{\mathrm{LIML}}$, and $\hat{\gamma}_{\mathrm{LIML}}$. Although the expression in Theorem 4 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. In particular, $T^{1/2}(\hat{\gamma}(k) - \gamma_0)$ has an asymptotic mixture of normals distribution which depends on the local parameters H and C. This poses an additional 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 although H and C are identified, neither is consistently estimable.

4. NONSTANDARD CONFIDENCE REGIONS FOR β

A central difficulty for inference is the asymptotic dependence of the estimator and Wald statistic distributions on $\lambda'\lambda/K_2$. Because $\lambda'\lambda/K_2$ is not consistently estimable, asymptotically valid confidence regions cannot be constructed by directly inverting t statistics using the distributions from Theorem 1. This section investigates two solutions to this problem, Anderson-Rubin (1949) (AR) confidence regions and confidence regions based on Bonferroni's inequality.

A. *Anderson-Rubin Confidence Regions*

Anderson and Rubin (1949) suggested testing the null hypothesis $\beta = \beta_0$ using the statistic

$$\text{(4.1)} \qquad A_T(\beta_0) = \{(y^\perp - Y^\perp\beta_0)' P_{Z^\perp}(y^\perp - Y^\perp\beta_0)/K_2\} / \{(y^\perp - Y^\perp\beta_0)' M_{Z^\perp}(y^\perp - Y^\perp\beta_0)/(T - K_1 - K_2)\}.$$

If $(u_t \; V_t')'$ are i.i.d. $N(0, \Omega)$ and X and Z are strictly exogenous, then under the null $A_T(\beta_0)$ has an exact $F_{K_2, T-K_1-K_2}$ distribution, which has a $\chi^2_{K_2}/K_2$ limit as T gets large. This result obtains asymptotically under the more general conditions of Assumption M:

THEOREM 5: *Suppose that* (2.1), (2.2), *and Assumptions* L_Π *and M hold.*
(a) *Under the null hypothesis* $\beta = \beta_0$, $A_T(\beta_0) \Rightarrow \chi^2_{K_2}/K_2$.
(b) *Under the fixed alternative hypothesis* $\beta = \beta_1$, $A_T(\beta_0) \Rightarrow S_2(\bar{\Delta}, 0)/[K_2 S_1(\bar{\Delta})]$, *which is distributed as* K_2^{-1} *times a noncentral* $\chi^2_{K_2}$ *with noncentrality parameter* $\bar{\Delta}'\lambda'\lambda\bar{\Delta}/S_1(\bar{\Delta})$, *where* $\bar{\Delta} = \sigma_{uu}^{-1/2}\Sigma_{VV}^{1/2}(\beta_0 - \beta_1)$.

Theorem 5(a) shows that, as discussed in the fixed instrument/Gaussian case by Anderson and Rubin (1949), joint confidence regions for β can be constructed as the set of β_0 for which $A_T(\beta_0)$ fails to reject using the asymptotic $\chi^2_{K_2}/K_2$ critical values.

Theorem 5(b) implies that the probability that $A_T(\beta_0)$ rejects distant alternatives asymptotes to a value which is typically less than one. For example, for alternatives of the form $\beta_0 - \beta_1 = b\sigma_{uu}^{1/2}\Sigma_{VV}^{-1/2}\iota$ where b is a scalar and ι is the n-vector of 1's, the noncentrality parameter tends to $\iota'\lambda'\lambda\iota/n$ as $b \to \infty$. Thus tests based on the AR statistic are not consistent under weak instrument asymptotics. This accords with the failure of $\hat{\beta}(k)$ to concentrate in a decreasing region.

Variations on this approach are readily analyzed using these techniques. For example, 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 regions. One approach to this problem is to construct a "split-sample" AR statistic: run the first stage regression using the first subsample to obtain $\hat{\Pi}^{(1)}$, say, then construct (4.1) using the second subsample, where $Z^{\perp}$ is replaced by $Z^{\perp(2)}\hat{\Pi}^{(1)}$ (where $Z^{\perp(2)}$ is $Z^{\perp}$ constructed using the second subsample).[7] If the subsamples are randomly selected, if the two subsample sizes are proportional to T, and if the data are independently distributed, then the resulting statistic has a χ^2_n/n limiting distribution under the null that $\beta = \beta_0$. Like the full-sample AR statistic, the split-sample AR statistic can be inverted to construct asymptotically valid confidence regions.

The AR statistic has power against both $\beta \neq \beta_0$ and failure of the overidentifying restrictions. Thus if the overidentifying restrictions are false, the intervals could be tight and could lead a researcher to believe that β is precisely estimated, when in fact the tight interval reflects the endogeneity of an instrument. Indeed, the AR intervals can be null, as is the case in several specifications in the empirical application in Section 7.

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

B. Bonferroni Confidence Regions

The preceding remarks suggest a role for an asymptotically valid interval estimator of β that, in contrast to the AR interval estimator, imposes instrument validity as is done in the conventional application of tests based on $t(k)$. This is pursued here using an approach based on Bonferroni's inequality. Let $C_{\lambda'\lambda/K_2}(\alpha_1)$ denote a $100(1-\alpha_1)\%$ confidence region for $\lambda'\lambda/K_2$, and let $C_{\beta|\lambda'\lambda/K_2}(\alpha_2)$ denote a $100(1-\alpha_2)\%$ confidence region for β, constructed given $\lambda'\lambda/K_2$. The region $C_{\lambda'\lambda/K_2}(\alpha_1)$ can be constructed by inverting the noncentral Wishart distribution of G_T. The conditional region $C_{\beta|\lambda'\lambda/K_2}(\alpha_2)$ can be computed by inverting the Wald statistic $W(k)$ in (3.3), given $\lambda'\lambda/K_2$. Then a confidence region for β which does not depend on $\lambda'\lambda/K_2$ is

$$C_\beta^B(\alpha) = \bigcup_{\lambda'\lambda/K_2 \in C_{\lambda'\lambda/K_2}(\alpha_1)} C_{\beta|\lambda'\lambda/K_2}(\alpha_2) \tag{4.2}$$

where $\alpha = \alpha_1 + \alpha_2$. By Bonferroni's inequality, the region $C_\beta^B(\alpha)$ has confidence level of at least $100(1-\alpha)\%$. Although this approach is theoretically valid for general n, computational requirements increase sharply with n so we focus on the case $n = 1$ henceforth.

When $n = 1$, $C_{\lambda'\lambda/K_2}(\alpha_1)$ can be constructed by inverting the noncentral chi-squared statistic G_T. The construction of $C_{\beta|\lambda'\lambda/K_2}(\alpha_2)$ requires obtaining asymptotic critical values of $W(k)$ or, since $n = 1$, the k-class t statistic, which depend on $\lambda'\lambda/K_2$ and ρ. Because $\theta = \sigma_{uu}^{1/2}\Sigma_{VV}^{-1/2}\rho$ and $\sigma_{uu}^{1/2}\Sigma_{VV}^{-1/2}$ is consistently estimable given β_0, this requires using a data-dependent mapping from β to ρ to obtain critical values. The relations $\hat{\sigma}_{uu,\mathrm{OLS}} \xrightarrow{p} \sigma_{uu}(1-\rho^2)$ and $\hat{\beta}_{\mathrm{OLS}} \xrightarrow{p} \beta_0 + \theta$ suggest using

$$\hat{\rho}(\beta_0) = \left(\hat{\Sigma}_{VV}/\hat{\sigma}_{uu,\mathrm{OLS}}\right)^{1/2}\left(\hat{\beta}_{\mathrm{OLS}} - \beta_0\right)/ \left[1 + \left(\hat{\Sigma}_{VV}/\hat{\sigma}_{uu,\mathrm{OLS}}\right)\left(\hat{\beta}_{\mathrm{OLS}} - \beta_0\right)^2\right]^{1/2}. \tag{4.3}$$

Because $\hat{\rho}(\beta_0) \xrightarrow{p} \rho(\beta_0) = (\Sigma_{VV}/\sigma_{uu})^{1/2}\theta(\beta_0)$ uniformly in $\beta_0, C_{\beta|\lambda'\lambda/K_2}(\alpha_2)$, constructed as the acceptance region of t_{TSLS} given $\lambda'\lambda/K_2$, has asymptotic confidence level $100(1-\alpha_2)\%$. Thus the Bonferroni region (4.2) will have coverage rate of at least $100(1-\alpha)\%$.

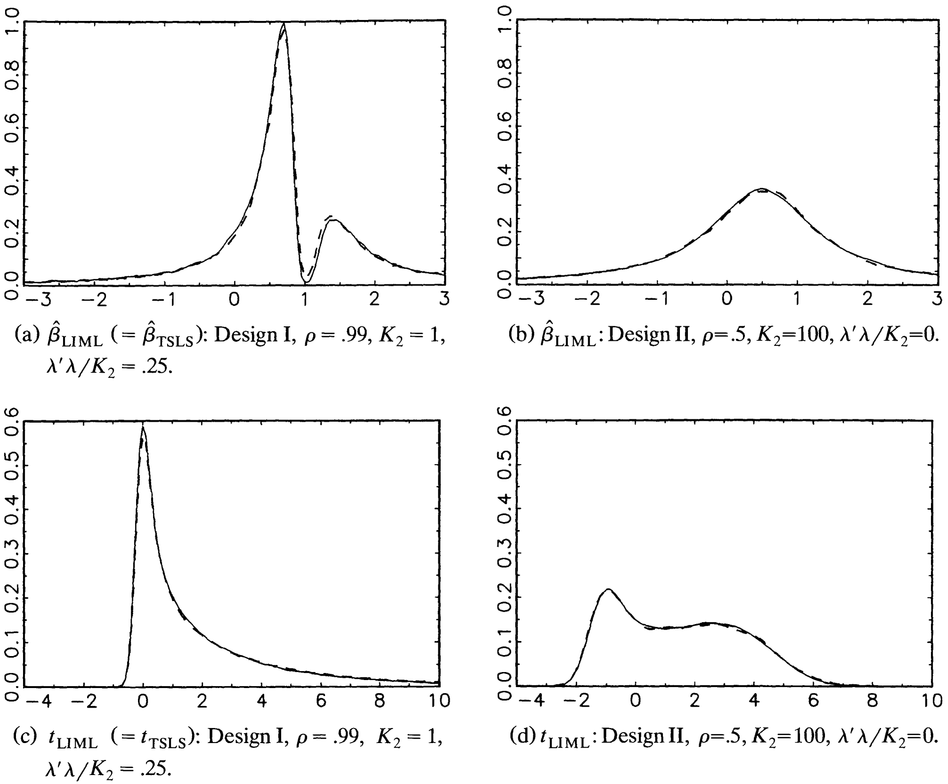
In the numerical work below, we consider two alternative methods for constructing (4.2), in which $C_{\beta|\lambda'\lambda/K_2}(\alpha_2)$ is alternatively based on the TSLS and LIML t statistics; the resulting confidence regions are respectively called TSLS and LIML Bonferroni regions. For both, the first stage confidence interval for $\lambda'\lambda/K_2$ and the second stage confidence interval for β_0 are equal-tailed, and $\alpha_1 = \alpha_2$. Using unequal-tailed intervals or letting α_1 and α_2 differ might improve performance, but these extensions are left to future work.

5. MONTE CARLO COMPARISON OF ASYMPTOTIC AND FINITE-SAMPLE DISTRIBUTIONS WHEN $n = 1$

Monte Carlo experiments were performed to examine the quality of the preceding asymptotic approximations to the finite-sample distributions of $\hat{\beta}_{\mathrm{TSLS}}$, t_{TSLS}, $\hat{\beta}_{\mathrm{LIML}}$, and t_{LIML} when $n = 1$. Because our asymptotic distributions are exact for $\hat{\beta}_{\mathrm{TSLS}}$ in the fixed instrument/Gaussian case, two designs that focus on stochastic instruments and nonnormal errors were considered. The first (design I) reflects time series applications where the number of instruments is small and the instruments are stochastic. The errors and instruments were drawn according to Z_t i.i.d. $N(0, I_{K_2})$ and $(u_t V_t)'$ i.i.d. $N(0, \overline{\Sigma})$. The second design (design II) is motivated by cross-sectional applications with a large number of fixed binary instruments, as in Angrist and Krueger (1991), and nonnormal errors. In this design, the instruments are $Z_t = \mathbf{1}_{jt}$, where $\mathbf{1}_{jt}$ is an indicator variable which equals one if observation t is in cell j, where $j = 1, \ldots, K_2 + 1$ and the final cell was omitted, and $(u_t, V_t)' = ((\xi_{1t}^2 - 1)/\sqrt{2}, (\xi_{2t}^2 - 1)/\sqrt{2})'$, where $(\xi_{1t}\ \xi_{2t})' \sim N(0, \tilde{\Sigma})$, where $\tilde{\Sigma}_{11} = \tilde{\Sigma}_{22} = 1$, and $\tilde{\Sigma}_{12} = \tilde{\Sigma}_{21} = \sqrt{\rho}$, $\rho \geq 0$. Thus in design II, $(u_t, V_t)'$ are scaled centered diagonal elements of a Wishart random variable. An equal number of observations were drawn from each cell (up to integer constraints). 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 $\hat{\beta} - \beta_0$. The data are generated according to (2.1) and (2.2) with $\sigma_{uu} = \Sigma_{VV} = 1$ and with $X_t = 1$. Finite sample distributions were computed using 20,000 Monte Carlo replications. Asymptotic distributions were computed using 100,000 draws of the random variates appearing in the limiting representations.

Selected asymptotic and finite-sample pdf's of $\hat{\beta}_{\mathrm{LIML}}$ (top panel) and t_{LIML} (bottom panel) are plotted in Figure 1 for two cases of interest. The finite sample results are for $T = 20K_2$. The design I case (Figure 1(a), (c)) is similar to one of the cases examined by Nelson and Startz (1990a,b) and Maddala and Jeong (1992) and both the asymptotic and finite-sample estimator pdf's are bimodal; because $K_2 = 1$ in this case, TSLS and LIML are equivalent. The design II case (Figure 1(b), (d)) is similar to a case simulated by Bound et al. (1995) and estimated by Angrist and Krueger (1991), except that T/K_2 is much larger in their cases. In each case the asymptotics provide a good approximation to the finite-sample distributions, with the differences often nearly indistinguishable at the scale of the plot.

The maximum absolute difference between the finite sample and asymptotic cumulative distributions of $\hat{\beta}_{\mathrm{TSLS}}$, t_{TSLS}, $\hat{\beta}_{\mathrm{LIML}}$, and t_{LIML} are given in Table I for various parameter values. Even for as few as 5 observations per instrument, the asymptotic distributions provide good approximations to the sampling distributions for $\hat{\beta}_{\mathrm{TSLS}}$ and $\hat{\beta}_{\mathrm{LIML}}$: over all cases in Table I, the largest differences between the two estimator cdf's are .111 for $T/K_2 = 5$, .050 for $T/K_2 = 10$, and .042 for $T/K_2 = 20$. The asymptotic approximations to the distribution of the t



(a) $\hat{\beta}_{\text{LIML}}$ $(=\hat{\beta}_{\text{TSLS}})$: Design I, $\rho = .99$, $K_2 = 1$, $\lambda'\lambda/K_2 = .25$.

(b) $\hat{\beta}_{\text{LIML}}$: Design II, ρ=.5, K_2=100, $\lambda'\lambda/K_2$=0.

(c) t_{LIML} $(= t_{\text{TSLS}})$: Design I, $\rho = .99$, $K_2 = 1$, $\lambda'\lambda/K_2 = .25$.

(d) t_{LIML}: Design II, ρ=.5, K_2=100, $\lambda'\lambda/K_2$=0.

FIGURE 1—Asymptotic (solid line) and finite-sample (dashed line) pdf's of the LIML estimator and t statistic. True $\beta_0 = 0$; plim$(\hat{\beta}_{\text{OLS}}) = \rho$.

TABLE I

MAXIMUM ABSOLUTE DIFFERENCE BETWEEN FINITE SAMPLE CDF AND ASYMPTOTIC CDF[a]

Parameters			$T/K_2=5$				$T/K_2=10$				$T/K_2=20$			
K_2	ρ	$\lambda'\lambda/K_2$	$\hat\beta_{TSLS}$	t_{TSLS}	$\hat\beta_{LIML}$	t_{LIML}	$\hat\beta_{TSLS}$	t_{TSLS}	$\hat\beta_{LIML}$	t_{LIML}	$\hat\beta_{TSLS}$	t_{TSLS}	$\hat\beta_{LIML}$	t_{LIML}
A. *Design I*														
1	.99	0	0.007	0.075	0.007	0.075	0.008	0.029	0.008	0.029	0.004	0.012	0.004	0.012
1	.99	.25	0.089	0.093	0.089	0.093	0.047	0.046	0.047	0.046	0.025	0.022	0.025	0.022
1	.99	1	0.111	0.087	0.111	0.087	0.050	0.038	0.050	0.038	0.026	0.026	0.026	0.026
1	.99	10	0.081	0.090	0.081	0.090	0.037	0.059	0.037	0.059	0.021	0.045	0.021	0.045
4	.99	0	0.007	0.052	0.008	0.025	0.006	0.024	0.005	0.013	0.005	0.008	0.005	0.006
4	.99	.25	0.041	0.135	0.032	0.027	0.023	0.085	0.016	0.016	0.012	0.048	0.015	0.015
4	.99	1	0.063	0.058	0.030	0.051	0.031	0.041	0.012	0.034	0.021	0.024	0.009	0.024
4	.99	10	0.033	0.015	0.015	0.012	0.022	0.010	0.010	0.008	0.010	0.006	0.011	0.009
B. *Design II*														
4	.5	0	0.041	0.080	0.028	0.038	0.027	0.044	0.017	0.021	0.014	0.023	0.007	0.014
4	.5	.25	0.051	0.061	0.038	0.046	0.025	0.039	0.020	0.024	0.015	0.018	0.013	0.018
4	.5	1	0.059	0.039	0.037	0.038	0.034	0.024	0.024	0.023	0.015	0.014	0.013	0.016
4	.5	10	0.067	0.051	0.058	0.060	0.041	0.031	0.041	0.043	0.025	0.019	0.025	0.026
100	.5	0	0.050	0.048	0.011	0.019	0.029	0.026	0.012	0.011	0.020	0.015	0.004	0.009
100	.5	.25	0.052	0.045	0.021	0.020	0.031	0.028	0.012	0.013	0.016	0.016	0.013	0.013
100	.5	1	0.037	0.034	0.034	0.034	0.027	0.028	0.025	0.026	0.021	0.020	0.022	0.022
100	.5	10	0.067	0.069	0.077	0.077	0.048	0.047	0.050	0.050	0.037	0.038	0.042	0.042

[a] Entries are the Kolmogorov-Smirnov statistics testing the equality of the two distributions; specifically, $\sup_x |F_{asy}(x) - F_{exact}(x)|$, where F_{asy} and F_{exact} are the Monte Carlo estimates of the asymptotic and exact finite sample distributions. Quantiles of this statistic, under the hypothesis that the two population distributions are identical, are: 50%, .0064; 95%, .0105. Asymptotic distributions were computed using 100,000 replications of the representation in Theorems 1 and 2. Finite sample distributions were computed using 20,000 replications.

statistic for TSLS and LIML are somewhat less good, but for $T/K_2 = 10$ they are typically within .03 in both designs.

Finite-sample coverage rates of 95% Bonferroni (both TSLS and LIML, with $\alpha_1 = \alpha_2 = .025$) and AR confidence intervals were checked for the models in Table I. For Bonferroni intervals, the lowest coverage rate for $T/K_2 = 10$ is 93.4%; for $T/K_2 = 20$ coverage is at least 95% and typically is between 96% and 99%. The AR interval coverage rates are between 93% and 95% for $T/K_2 = 20$.[8]

These results suggest that the weak instrument asymptotics provide good approximations to the finite sample distributions of the estimators and t statistics when exogenous regressors are stochastic and errors are nonnormal, for a wide range of parameter values including cases previously studied by Nelson and Startz (1990a,b) and Bound, Jaeger, and Baker (1995). The approximations are

[8] Details of these and other unreported results are available from the authors upon request.

typically good with only ten observations per instrument, and improve as this ratio increases.

6. NUMERICAL EVALUATION OF ASYMPTOTIC DISTRIBUTIONS AND TEST POWER FUNCTIONS WHEN $n = 1$

The asymptotic representations can be used to study numerically the properties of various inferential procedures. This section focuses on four issues: bias of TSLS and LIML point estimates; coverage rates of conventional TSLS and LIML confidence intervals; size distortions of tests of overidentifying restrictions; and the power of the AR and Bonferroni tests of $\beta = \beta_0$. As in Section 5, attention is restricted to the case of a single included endogenous variable ($n = 1$).

A. Estimator Bias

The ratio of the asymptotic TSLS bias to the OLS bias ($E\beta^*_{\text{TSLS}}/\theta$) and the ratio of the median LIML bias to the OLS bias (median$[\Delta^*_0(\kappa^*_{\text{LIML}})/\theta]$) are plotted in Figure 2 for $2 \leq K_2 \leq 100$ and $0 \leq \lambda'\lambda/K_2 \leq 20$ for $|\rho| = .2$, .5, and .99. The relative bias of TSLS does not depend on ρ and evidently depends more strongly on $\lambda'\lambda/K_2$ than on K_2. The population counterpart of $\hat{B}_{\max}$ is $(1 + \lambda'\lambda/K_2)^{-1}$, and this provides a good approximation to the TSLS bias for all but very small values of K_2: if $K_2 \geq 5, \max_{0 \leq \lambda'\lambda/K_2 \leq 20}|E\beta^*_{\text{TSLS}}/\theta - (1 + \lambda'\lambda/K_2)^{-1}| = .07$, while if $K_2 \geq 10$, this maximal approximation error drops to .03. In contrast to TSLS, $\hat{\beta}_{\text{LIML}}$ rapidly becomes median unbiased as $\lambda'\lambda/K_2$ increases, particularly for large values of $|\rho|$. For the cases in Figure 2 with $\lambda'\lambda/K_2 \geq 2$, the maximal relative median bias of LIML is 10% for $K_2 \geq 2$ and is 1% for $K_2 \geq 8$. Anderson (1982), Hillier (1990) and others have noted the relative lack of bias of LIML in the fixed instrument/Gaussian model; the results here extend their conclusions to more general conditions on the instruments and errors and to a more comprehensive set of cases.

B. TSLS and LIML Confidence Interval Coverage Rates

Coverage rates for conventional 95% TSLS and LIML confidence intervals are plotted in Figure 3 for the same parameter values as in Figure 2. The TSLS coverage rate is quite sensitive to K_2 and $|\rho|$ and generally falls as K_2 increases, $\lambda'\lambda/K_2$ decreases, and $|\rho|$ increases. For example, when $|\rho| = .2$, coverage rates are near 95% for all K_2 once $\lambda'\lambda/K_2$ is greater than 10, but when $|\rho| = .99$, coverage rates exceed 90% only if both $\lambda'\lambda/K_2$ is large and K_2 is small. Thus TSLS confidence intervals can fail dramatically for moderate and, depending on K_2 and ρ, large $\lambda'\lambda/K_2$. In contrast, coverage rates for LIML

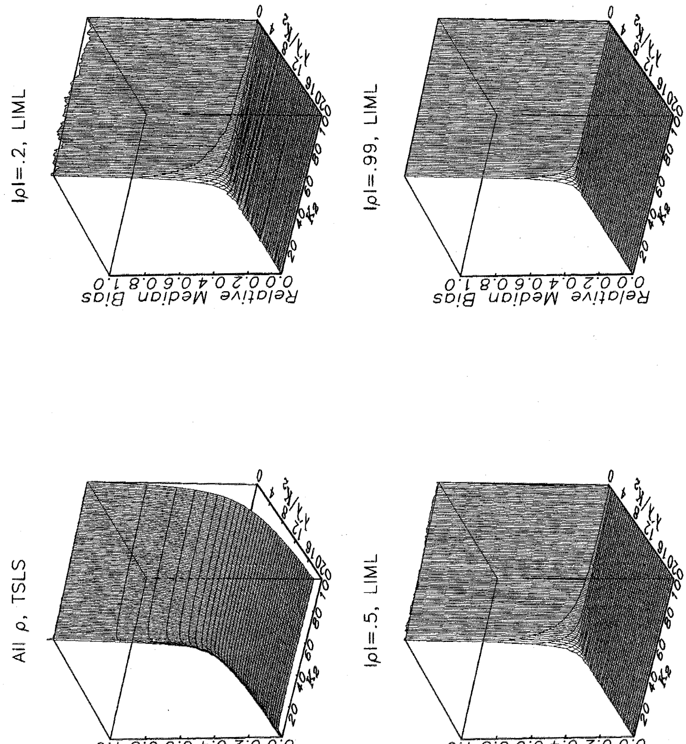


FIGURE 2.—Asymptotic bias of β_{TSLS} and asymptotic median bias of β_{LIML}, as a fraction of the bias of $\hat{\beta}_{\mathrm{OLS}}$.

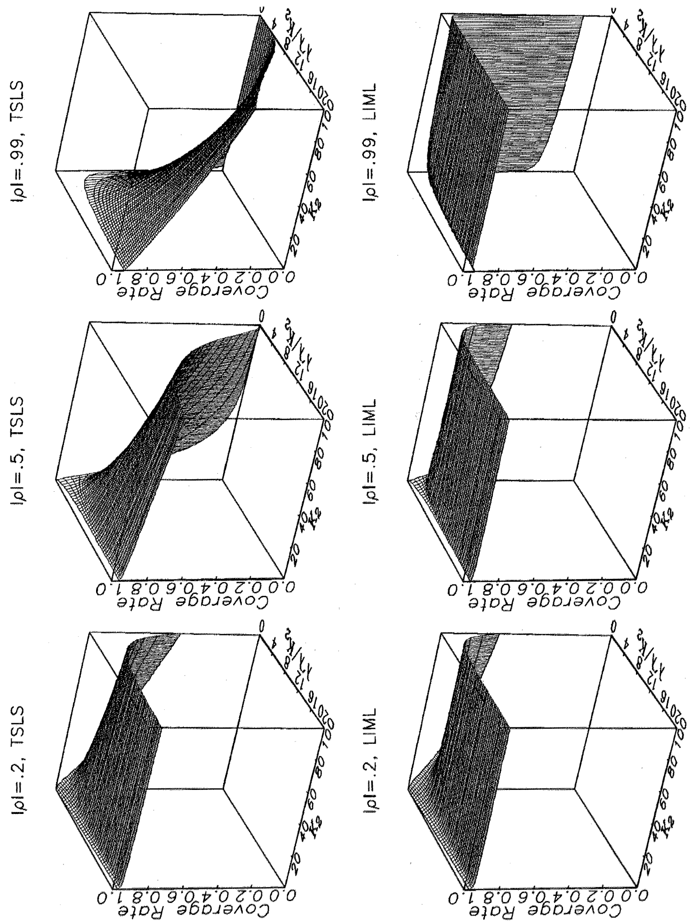


FIGURE 3.—Asymptotic coverage rates of conventional 95% TSLS and LIML confidence intervals.

confidence intervals are less sensitive to K_2 and $|\rho|$ and in an absolute sense they can be considered fairly good, as long as $\lambda'\lambda/K_2 \geq 1$: for $1 \leq K_2 < 100$, $\lambda'\lambda/K_2 \geq 1$, and $|\rho| = .2$, .5, and .99, the asymptotic coverage rates lie between 81.6% and 99.8%. For $\lambda'\lambda/K_2 \geq 10$, the asymptotic coverage rates lie between 91.6% and 98.1%.[9]

C. Size of Tests of Overidentifying Restrictions

The asymptotic size (the rejection rate under the null of instrument exogeneity) of the 5% Basmann test of the overidentifying restrictions based on TSLS ($\phi_{\text{Bas}}(1)$ in (3.8b)) were computed using the representations in Theorem 3; to save space, no figures or tables are provided but the findings are summarized. Rejection rates under the null are generally close to 5% for $|\rho|$ small, but for large $|\rho|$ and large K_2 the size distortions can be dramatic. For example, with $K_2 = 100$ and $\lambda'\lambda/K_2 = 1$, the rejection rate is 47% when $|\rho| = .75$ and is 97% when $|\rho| = .99$. The 5% test based on LIML, $\phi_{\text{Bas}}(\hat{k}_{\text{LIML}})$, has much better size than its TSLS counterpart. Over the parameter values $|\rho| = (.2, .5, .75, .99)$, $0 \leq \lambda'\lambda/K_2 \leq 20$, $2 \leq K_2 \leq 100$, the size is between .001 and .052; if $1 \leq \lambda'\lambda/K_2 < 20$, the size is between .012 and .052. Thus the Basmann TSLS overrejections under the null are essentially absent for its LIML counterpart, although $\phi_{\text{Bas}}(\hat{k}_{\text{LIML}})$ is asymptotically conservative for $\lambda'\lambda/K_2$ small and K_2 small. This suggests using $\phi_{\text{Bas}}(\hat{k}_{\text{LIML}})$ in practice.

D. Power of AR and Bonferroni Tests

One way to compare the accuracy of the AR and Bonferroni confidence regions is to compare the asymptotic power of AR and Bonferroni tests of the hypothesis $\beta = \beta_0$ against the alternative $\beta = \beta_0 + (\sigma_{uu}/\Sigma_{VV})^{1/2}\,\overline{\Delta}$. When $K_2 = 1$, because the Bonferroni tests are conservative the AR test is uniformly (in $\overline{\Delta}$) more powerful for all ρ and $\lambda'\lambda/K_2$. When $K_2 > 1$, no test dominates the other so the asymptotic power of the three tests were compared numerically and are briefly summarized. When K_2 and/or $\lambda'\lambda/K_2$ is large, AR has the lowest power against most alternatives and LIML Bonferroni tends to be more powerful than TSLS Bonferroni, particularly for large $|\rho|$ and when K_2 is large and $\lambda'\lambda/K_2$ is small. When both K_2 and $\lambda'\lambda/K_2$ are small, AR is more powerful than either Bonferroni test. This suggests using the LIML Bonferroni confidence regions if K_2 is large and/or $\lambda'\lambda/K_2$ is suspected to be large, and using the AR regions otherwise.

[9] These results accord with Morimune's (1989, Sec. 3) Monte Carlo finding of greater size distortions for t_{TSLS} than t_{LIML} in selected models in the fixed instrument/Gaussian case. In Morimune's (1989) designs, $K_2 \leq 11$, $\lambda'\lambda/K_2 \geq 2.6$, and $\rho \leq .9$, so Morimune's results understate the distortions found here for more instruments and smaller $\lambda'\lambda/K_2$.

7. APPLICATIONS TO THE RETURNS TO EDUCATION

This section reexamines Angrist and Krueger's (1991) estimates of the returns to education in light of the foregoing results. Angrist and Krueger's insight was that quarter of birth, and quarter of birth interacted with other covariates, can serve as instruments for education in an earnings equation: quarter of birth is arguably randomly distributed across the population, yet it 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.[10] 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.

Table II summarizes the results of regressions of log weekly earnings onto years of education and additional control variables (listed at the bottom of the table). 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 estimated return to education, that is, the coefficient on years of education, estimated by OLS, TSLS, and LIML in four basic specifications. Subsequent rows report 95% Bonferroni and AR confidence intervals for the returns to education, the first stage F statistic, the Durbin endogeneity test statistic ($F_{\mathrm{DWH},3}$), and Basmann's over-identification test statistic based on the LIML estimator ($\phi_{\mathrm{Bas}}(\hat{k}_{\mathrm{LIML}})$). The regression specifications are taken from Angrist and Krueger (1991, Tables 5–7) and Bound, Jaeger, and Baker (1995, Tables 1 and 2). Three quarter-of-birth dummies are used as instruments in column I. Columns II and III add quarter-of-birth × year-of-birth interactions to the instrument list, for a total of 30 instruments in column II and 28 instruments in column III (due to the inclusion of age and age^2). Column IV adds quarter-of-birth × state-of-birth interactions to the instrument list for a total of 178 instruments.

The asymptotic theory helps to interpret these empirical results. In order to apply some of the asymptotic results, *a-priori* reasoning is used to obtain a range in which ρ might plausibly fall. To do this we posit that the return to education lies between 0 and .18. In specification I of Table II, $\hat{\beta}_{\mathrm{OLS}} = .063$ and $(\hat{\Sigma}_{VV}/\hat{\Sigma}_{uu})^{1/2} = 5.05$; using (4.3), this yields $-.51 \leq \hat{\rho}(\beta_0) \leq .30$. Because coverage rates decrease and bias increases as $|\rho|$ increases, it therefore suffices to consider $|\rho| = .5$. (The other specifications in Table II yield similar ranges for ρ.)

[10] For details of construction, see Appendix 1 of Angrist and Krueger (1991). We thank David Jaeger for providing these data.

TABLE II

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)
A-R Confidence Interval	[.052, .153]	[−.003, .179]	[−.441, .490]	[−.015, .240]
TSLS Bonferroni Confidence Interval	[.052, .152]	[.038, .137]	$[-\infty, +\infty]$	[.048, .172]
LIML Bonferroni Confidence Interval	[.052, .153]	[.036, .134]	$[-\infty, +\infty]$	[.043, .158]
F (first stage)	30.53	4.747	1.613	1.869
{p-value}	{.000}	{.000}	{.021}	{.000}
Durbin test, TSLS	3.087	1.126	0.013	2.853
{p-value}	{0.079}	{.289}	{.910}	{.091}
Basmann test, LIML	2.318	22.45	19.55	161.1
{p-value}	{.314}	{.801}	{.849}	{.800}
B. *Men Born* 1940–49 (n = 486, 926)				
OLS	.0520	.0520	.0520	.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)	(.0197)	(.0420)	(.0178)
A-R Confidence Interval	[∅]	[∅]	[∅]	[.033, .148]
TSLS Bonferroni Confidence Interval	[−.155, −.018]	[−.004, .076]	[.000, .219]	[.027, .150]
LIML Bonferroni Confidence Interval	[−.174, −.028]	[−.023, .079]	[−.009, .290]	[.023, .156]
F (first stage)	26.32	6.849	2.736	1.929
{p-value}	{.000}	{.000}	{.000}	{.000}
Durbin test, TSLS	28.90	0.780	1.188	1.780
{p-value}	{.000}	{.377}	{.276}	{.182}
Basmann test, LIML	9.356	93.29	49.22	200.36
{p-value}	{.009}	{.000}	{.006}	{.110}
Controls				
Race, Standard Metropolitan Statistical Area, Married, Region, Year of Birth Dummies	yes	yes	yes	yes
Age, Age^2	no	no	yes	yes
State of Birth	no	no	no	yes
Instruments				
Quarter of birth	yes	yes	yes	yes
Quarter of birth *(year of birth)	no	yes	yes	yes
Quarter of birth *(state of birth)	no	no	no	yes
# Instruments	3	30	28	178

First consider the results for the 1930–39 cohort. In specification I the first stage F statistic is large, implying that the expected bias of the TSLS estimator is negligible. Inverting the first stage F statistic yields a 97.5% conference interval for $\lambda'\lambda/K_2$ of (17.3, 45.8). Over this range of $\lambda'\lambda/K_2$ with $K_2 = 3$ and $|\rho| \leq < .5$, the asymptotic theory suggests that standard TSLS and LIML statistics are reliable. However, in specifications II–IV the first stage F statistic falls into a range in which some of the TSLS and LIML results become unreliable. In specification II the relative bias measure $\hat{B}_{\max} = .21$, and the 97.5% confidence interval for $\lambda'\lambda/K_2$ is (2.26, 5.64). Based on Figures 2 and 3, for $\lambda'\lambda/K_2 \geq 2$, $K_2 = 30$, and $|\rho| \leq .5$, LIML is approximately median unbiased but the TSLS relative bias is as high as 33%; coverage rates for LIML and TSLS confidence intervals may be as low as 90% and 60%, respectively. Also, Bonferroni intervals, particularly LIML Bonferroni, are generally more accurate than AR confidence intervals for K_2 large as discussed in Section 6D. For specification II, this suggests focusing on the LIML estimates and either conventional or Bonferroni LIML confidence intervals. In specification III the 97.5% interval for $\lambda'\lambda/K_2$ includes $\lambda'\lambda/K_2 = 0$, so none of the the TSLS or LIML estimates or confidence intervals are reliable. Bonferroni and AR tests and the Durbin endogeneity test have correct size for this specification but could have negligible power. Finally, for specification IV the 97.5% confidence interval for $\lambda'\lambda/K_2$ is (0.53, 1.32). Figures 2 and 3 do not go as high as $K_2 = 178$, but for $K_2 = 100$, $\lambda'\lambda/K_2 \geq .5$, and $|\rho| = .5$, LIML remains approximately median unbiased but TSLS relative bias is as high as 67%; coverage rates for LIML and TSLS confidence intervals could be as low as 77% and 1%; and Bonferroni confidence intervals are generally tighter than AR confidence intervals. This suggests focusing on the LIML point estimates and LIML Bonferroni confidence intervals for specification IV.

For the 1940–49 cohort, rejection of the over-identifying restrictions using the Basmann-LIML test suggests that the results from specifications I–III are unreliable, particularly since the asymptotics imply that if anything this test is undersized. Note that AR confidence intervals are empty for these specifications as a result of rejecting the overidentifying restrictions. The Basmann test does not reject in specification IV but the first stage F statistic is 1.9. Using reasoning similar to that given for specification IV of the 1930–39 cohort, this suggests focussing on the LIML point estimate and LIML Bonferroni confidence intervals for this specification.

Using the estimators supported by the asymptotics, the point estimates are reasonably stable across specifications and cohorts, ranging from .084 to .100. The shortest AR interval is (.05, .15) in specification I (1930–39 cohort), and Bonferroni intervals from specifications II (1930–39 cohort) and IV (both cohorts) are similarly short. Among the TSLS and LIML confidence intervals which we suspect to have at least 90% coverage rates, the tightest is (.05, .12) for LIML in specification II (1930–39 cohort). Importantly, the Durbin endogeneity test rejects the hypothesis that OLS and TSLS estimands are the same at the 10% level in specifications I and IV for the 1930–39 cohort. Overall, this analysis confirms the main conclusion of Angrist and Krueger that OLS estimates are if anything biased downward. However, our preferred estimates of the returns to education are higher than theirs, implying roughly twice as much downward bias in OLS estimates, and our preferred confidence intervals are much wider than the unreliable TSLS intervals.

8. CONCLUSIONS AND LESSONS FOR EMPIRICAL PRACTICE

When the instruments are weakly correlated with the endogenous regressors, conventional asymptotic results fail even if the sample size is large. In particular, TSLS can be badly biased and can produce confidence intervals with severely distorted coverage rates even if $\lambda'\lambda/K_2$ is moderate or, if K_2 is large, if $\lambda'\lambda/K_2$ is large. More generally, Figures 2 and 3 summarize the circumstances in which TSLS and LIML will be unbiased and will form reliable confidence intervals. Using conventional asymptotics after pretesting for instrument significance is an unsatisfactory solution because the pretest will have power against small values of $\lambda'\lambda/K_2$, for which TSLS and LIML statistics can be ill behaved. For example, if $\lambda'\lambda/K_2 = 3$ and $K_2 = 10$, the power of a 5% pretest using the first stage F statistic exceeds 99%, but the TSLS bias is fully one-fourth the OLS bias.

The results have some constructive implications for empirical practice. At a minimum, first stage F statistics (or, when $n > 1$, G_T and/or the bias measures in Section 3b) should be reported. Although some forms of the DWH test are conservative, the Durbin form ($F_{\mathrm{DWH},3}$) was found to have correct asymptotic size and to have power against differences between the TSLS and OLS estimands, even for small $\lambda'\lambda/K_2$, recommending its use. While tests of overidentifying restrictions have size distortions, under the null the TSLS version of the Basmann test tends to overreject while the LIML version tends to underreject. This suggests relying on the Basmann-LIML test but recognizing that, for some parameter values, it will have low power against small violations of instrument orthogonality.

When $n = 1$, these results have two additional constructive implications. First, estimator bias is less of a problem for LIML than TSLS, particularly when $\lambda'\lambda/K_2 \geq 2$, which suggests using LIML rather than TSLS point estimates. Second, given the difficulties with conventional IV confidence intervals, these results strongly suggest using nonstandard methods for interval estimation. Of the asymptotically valid methods analyzed here, none is uniformly more accurate than the others: LIML Bonferroni tests tend to have greatest power for large $\lambda'\lambda/K_2$ and/or large K_2, but AR tests are relatively more powerful for $\lambda'\lambda/K_2$ and K_2 both small. In the empirical application to the returns to education, both procedures produce plausible and comparable confidence intervals in the cases in which the overidentifying restrictions were not rejected, even when the first stage F statistic is quite small (less than two).

John F. Kennedy School of Government, Harvard University, 79 John F. Kennedy St., Cambridge, MA 02138, U.S.A.

Manuscript received October, 1993; final revision received May, 1996.

APPENDIX

Before proving the theorems we state a lemma which collects various results about sample moments. The proof makes repeated use of Assumptions L_{Π} and M and is omitted. Adopt the notational conventions, $B = B^{1/2\prime}B^{1/2}$ and $B^{-1} = B^{-1/2}B^{-1/2\prime}$ where B is a nonsingular symmetric matrix, and let $P_W^{1/2} = (W'W)^{-1/2\prime}W'$ for general $a \times b$ matrix W with $a \geq b$.

LEMMA A1: *Suppose that* (2.1), (2.2), *and Assumptions* L_{Π} *and M hold. Then the following hold jointly*:

(a) $(u^{\perp\prime}u^{\perp}/T, Y^{\perp\prime}u^{\perp}/T, Y^{\perp\prime}Y^{\perp}/T) \xrightarrow{p} (\sigma_{uu}, \Sigma_{Vu}, \Sigma_{VV})$.

(b) $Z^{\perp\prime}Z^{\perp}/T \xrightarrow{p} \Omega$, *where* $\Omega = Q_{ZZ} - Q_{ZX}Q_{XX}^{-1}Q_{XZ}$.

(c) $\{P_{Z^{\perp}}^{1/2}u^{\perp}, P_{Z^{\perp}}^{1/2}V^{\perp}\} \Rightarrow \{z_u\sigma_{uu}^{1/2}, z_V\Sigma_{VV}^{1/2}\}$, *where* $(z_u' \operatorname{vec}(z_V)')'$ *is distributed* $N(0, \overline{\Sigma} \otimes I_{K_2})$.

(d) $P_{Z^{\perp}}^{1/2}Y^{\perp} \Rightarrow (\lambda + z_V)\Sigma_{VV}^{1/2}$.

(e) $(Y^{\perp\prime}P_{Z^{\perp}}u^{\perp}, Y^{\perp\prime}P_{Z^{\perp}}Y^{\perp}, u^{\perp\prime}P_{Z^{\perp}}u^{\perp}) \Rightarrow (\Sigma_{VV}^{1/2\prime}\nu_2\sigma_{uu}^{1/2}, \Sigma_{VV}^{1/2\prime}\nu_1\Sigma_{VV}^{1/2}, \sigma_{uu}z_u'z_u)$.

PROOF OF THEOREM 1: (a) Define $\kappa_T = T(k-1)$ and write $\hat{\beta}(k) - \beta_0 = L_T(\kappa_T)^{-1}N_T(\kappa_T)$, where $L_T(\kappa) = Y^{\perp\prime}[I - (1 + \kappa/T)M_{Z^{\perp}}]Y^{\perp}$ and $N_T(\kappa) = Y^{\perp\prime}[I - (1 + \kappa/T)M_{Z^{\perp}}]u^{\perp}$. Using Lemma A1 (a) and (e), L_T and N_T can be shown to have the limits $\{L_T(\kappa), N_T(\kappa)\} \Rightarrow \{\Sigma_{VV}^{1/2\prime}(\nu_1 - \kappa I_n)\Sigma_{VV}^{1/2}, \sigma_{uu}^{1/2}\Sigma_{VV}^{1/2\prime}(\lambda + z_V)'z_u - \kappa\Sigma_{Vu}\}$, where the convergence is uniform in κ over compact sets. By assumption the convergence of the moments in Lemma A1, and thus the convergence of $\{L_T(\kappa), N_T(\kappa)\}$, is joint with $\kappa_T \Rightarrow \kappa$. Recall that $\rho = \Sigma_{VV}^{-1/2\prime}\Sigma_{Vu}\sigma_{uu}^{-1/2}$, so after some algebra one has $\hat{\beta}(k) - \beta_0 \Rightarrow \sigma_{uu}^{1/2}\Sigma_{VV}^{-1/2}(\nu_1 - \kappa I_n)^{-1}(\nu_2 - \kappa\rho) = \sigma_{uu}^{1/2}\Sigma_{VV}^{-1/2}\Delta_0^*(\kappa) \equiv \beta_0^*(\kappa)$.

(b) Note that $\hat{u}(k) = y^{\perp} - Y^{\perp}\hat{\beta}(k) = u^{\perp} - Y^{\perp}(\hat{\beta}(k) - \beta_0)$, so by Lemma A1 and part (a), $\hat{\sigma}_{uu}(k) \Rightarrow \sigma_{uu} - 2\Sigma_{Vu}'\beta_0^*(\kappa) + \beta_0^*(\kappa)'\Sigma_{VV}\beta_0^*(\kappa) = \sigma_{uu}S_1(\Delta_0^*(\kappa))$.

(c) Using $R\hat{\beta}(k) - r = R(\hat{\beta}(k) - \beta_0)$, the definition of $L_T(\kappa)$, and the previous results, we have

$$W(k) \Rightarrow \beta_0^*(\kappa)'R'\left[R\Sigma_{VV}^{-1/2}(\nu_1 - \kappa I_n)^{-1}\Sigma_{VV}^{-1/2\prime}R'\right]^{-1}R\beta_0^*(\kappa)/[q\sigma_{uu}S_1(\Delta_0^*(\kappa))].$$

The representation in the theorem follows using $\beta_0^*(\kappa) = \sigma_{uu}^{1/2}\Sigma_{VV}^{-1/2}\Delta_0^*(\kappa)$.

(d) The result follows from the definition of $t_\iota(k)$ and calculations similar to those in part (c).

(e) Write $\hat{\Sigma}_{VV} = V'M_{\overline{Z}}V/(T - K_1 - K_2)$. The result follows from Assumption M and Lemma A1(e). *Q.E.D.*

PROOF OF THEOREM 2: Note that for any nonsingular $(n+1) \times (n+1)$ matrix J, the roots of $|\overline{Y}'M_X\overline{Y} - k\overline{Y}'M_{\overline{Z}}\overline{Y}| = 0$ are the same as the roots of $|J'\overline{Y}'M_X\overline{Y}J - kJ'\overline{Y}'M_{\overline{Z}}\overline{Y}J| = 0$. In particular choose J, partitioned conformably with $\overline{Y}$, to be $J_{11} = 1$, $J_{21} = -\beta$, $J_{12} = 0$, and $J_{22} = I_n$. Let $D_T(\kappa) = J'\overline{Y}'M_X\overline{Y}J - (1 + \kappa/T)J'\overline{Y}'M_{\overline{Z}}\overline{Y}J$. Now use $M_{\overline{Z}} = M_XM_{Z^{\perp}}M_X$ to rewrite $D_T(\kappa)$ as $D_T(\kappa) = J'\overline{Y}^{\perp\prime}P_{Z^{\perp}}\overline{Y}^{\perp}J - \kappa J'\overline{Y}^{\perp\prime}M_{Z^{\perp}}\overline{Y}^{\perp}J/T$. Because $y^{\perp} = Y^{\perp}\beta + u^{\perp}$, $\overline{Y}^{\perp}J = [u^{\perp}\ Y^{\perp}]$, by Lemma A1(e), $J'\overline{Y}^{\perp\prime}P_{Z^{\perp}}\overline{Y}^{\perp}J \Rightarrow \Upsilon'\Xi_0^*\Upsilon$, where $\Upsilon = \operatorname{diag}(\sigma_{uu}^{1/2}, \Sigma_{VV}^{1/2})$ and Ξ_0^* is defined in the statement of the theorem. Also, $J'\overline{Y}^{\perp\prime}M_{Z^{\perp}}\overline{Y}^{\perp}J/T \xrightarrow{p} \Sigma = \Upsilon'\overline{\Sigma}\Upsilon$. Thus $D_T(\kappa) \Rightarrow \Upsilon'(\Xi_0^* - \kappa\overline{\Sigma})\Upsilon$ uniformly in κ over compact sets. The solutions to $|D_T(\kappa)| = 0$ therefore converge to those of $|\Xi_0^* - \kappa\overline{\Sigma}| = 0$. Thus $\hat{\kappa}_{\mathrm{LIML}} = T(\hat{k}_{\mathrm{LIML}} - 1) \Rightarrow \kappa_{\mathrm{LIML}}^*$, where κ_{LIML}^* is the smallest root of $|\Xi_0^* - \kappa\overline{\Sigma}| = 0$. *Q.E.D.*

PROOF OF THEOREM 3: (a)(i) Use the definition of L_T and N_T in Theorem 1 to write $\hat{\beta}(k) - \beta_0$ under L_ω as

$$\hat{\beta}(k) - \beta_0 = L_T(\kappa_T)^{-1}N_T(\kappa_T) + L_T(\kappa_T)^{-1}[Y^{\perp\prime}(I - \kappa_TM_{Z^{\perp}})Z^{\perp}\omega].$$

The limits in Theorem 1 for $L_T(\kappa_T)$ and $N_T(\kappa_T)$ continue to apply under L_ω. Also, $Y^{\perp\prime}(I - \kappa_TM_{Z^{\perp}})Z^{\perp}\omega = Y^{\perp\prime}Z^{\perp}\omega \Rightarrow \sigma_{uu}^{1/2}\Sigma_{VV}^{1/2\prime}(\lambda + z_V)'\xi$. Thus $\hat{\beta}(k) - \beta_0 \Rightarrow \beta_0^*(\kappa) + \sigma_{uu}^{1/2}\Sigma_{VV}^{-1/2}(\nu_1 - \kappa I_n)^{-1}(z_V + \lambda)'\xi = \sigma_{uu}^{1/2}\Sigma_{VV}^{-1/2}\Delta_\xi^*(\kappa)$.

(ii) Assumption L_ω implies $\hat{u} = u^\perp + Z^\perp \omega - Y^\perp(\hat{\beta}(k) - \beta)$, so by Lemma A1 and part (a)(i),

$$P_{Z^\perp}^{1/2}\hat{u}^\perp \Rightarrow \sigma_{uu}^{1/2} z_u + \Omega^{1/2} d - (\lambda + z_V)\Sigma_{VV}^{1/2}\beta_\xi^*(\kappa) = \sigma_{uu}^{1/2}[z_u + \xi - (\lambda + z_V)\Delta_\xi^*(\kappa)].$$

Thus $\hat{u}'P_{Z^\perp}\hat{u} \Rightarrow \sigma_{uu} S_2(\Delta_\xi^*(\kappa), \xi)$. Under Assumption L_ω, similar calculations show that $\hat{u}'\hat{u}/T = [u^\perp - Y^\perp(\hat{\beta}(k) - \beta)]'[u^\perp - Y^\perp(\hat{\beta}(k) - \beta)]/T + o_p(1) \Rightarrow \sigma_{uu} S_1(\Delta_\xi^*(\kappa))$. Combining these results we have $\phi_{\text{reg}}(k) \Rightarrow S_2(\Delta_\xi^*(\kappa), \xi)/S_1(\Delta_\xi^*(\kappa))$. The result $\phi_{\text{reg}} - \phi_{\text{Bas}} \xrightarrow{p} 0$ (and consequently the limiting distribution of ϕ_{Bas}) follows from $\hat{u}'M_{Z^\perp}\hat{u}/(T - K_1 - K_2) - \hat{u}'\hat{u}/T \xrightarrow{p} 0$.

(b) The proof is a modification of the proof of Theorem 2. Under Assumption L_ω,

$$P_{Z^\perp}^{1/2}\overline{Y}^\perp J = [P_{Z^\perp}^{1/2}(u^\perp + Z^\perp\omega) P_{Z^\perp}^{1/2} Y^\perp] \Rightarrow [\sigma_{uu}^{\frac{1}{2}}(z_u + \xi) \qquad (z_V + \lambda)\Sigma_{VV}^{1/2}].$$

Thus $J'\overline{Y}^{\perp\prime}P_{Z^\perp}\overline{Y}^\perp J \Rightarrow \Upsilon'\Xi_\xi^*\Upsilon$ and $J'\overline{Y}^{\perp\prime}M_{Z^\perp}\overline{Y}^\perp J/T = J'\overline{Y}^{\perp\prime}\overline{Y}^\perp J/T + o_p(1)$. Under L_ω, $J'\overline{Y}^{\perp\prime}\overline{Y}^\perp J/T \xrightarrow{p} \Sigma = \Upsilon'\overline{\Sigma}\Upsilon$ because, for example, $\omega' Z^{\perp\prime} P_{Z^\perp} Z^\perp \omega = T^{-1}d'(Z^{\perp\prime}Z^\perp/T)d \xrightarrow{p} 0$. Thus $D_T(\kappa)$ in the proof of Theorem 2 has the limit, $D_T(\kappa) \Rightarrow \Upsilon'(\Xi_\xi^* + \kappa\overline{\Sigma})\Upsilon$ uniformly in κ on compact sets. The result follows using the arguments in the proof of Theorem 2. *Q.E.D.*

Proof of Theorem 4: Note that $\hat{\gamma}(k) = (X'X)^{-1}X'[y - Y\hat{\beta}(k)]$, so $T^{1/2}(\hat{\gamma}(k) - \gamma_0) = (T^{-1}X'X)^{-1}[T^{-1/2}X'u - (T^{-1/2}X'Y)(\hat{\beta}(k) - \beta_0)]$. The result follows from calculations which invoke Assumption L_Φ and Lemma A1. *Q.E.D.*

Proof of Theorem 5: (a) Under the null $\beta = \beta_0$, from the definition of A_T and Lemma A1,

$$A_T(\beta_0) = (u^{\perp\prime}P_{Z^\perp}u^\perp/K_2)/[(u^{\perp\prime}u^\perp - u^{\perp\prime}P_{Z^\perp}u^\perp)/(T - K_1 - K_2)]$$
$$\Rightarrow z_u' z_u/K_2 \sim \chi_{K_2}^2/K_2.$$

(b) When $\beta = \beta_1$, $y^\perp - Y^\perp\beta_0 = u^\perp - Y^\perp(\beta_0 - \beta_1)$, so by Lemma A1,

$$(y^\perp - Y^\perp\beta_0)'P_{Z^\perp}(y^\perp - Y^\perp\beta_0) \Rightarrow \sigma_{uu}[z_u - (\lambda + z_V)\overline{\Delta}]'[z_u - (\lambda + z_v)\overline{\Delta}]$$
$$= \sigma_{uu}S_2(\overline{\Delta}, 0).$$

Also, $(y^\perp - Y^\perp\beta_0)'(y^\perp - Y^\perp\beta_0)/(T - K_1 - K_2) \xrightarrow{p} \sigma_{uu}S_1(\overline{\Delta})$. Substitution of these limits into the definition of $A_T(\beta_0)$ yields the expression in the theorem. The fact that the distribution is noncentral $\chi_{K_2}^2$ follows by some algebra after observing that the K_2-vector $z_u - z_V\overline{\Delta}$ is distributed $N(0, S_1(\overline{\Delta}))$, where $S_1(\overline{\Delta})$ is nonrandom. *Q.E.D.*

REFERENCES

Anderson, T. W. (1976): "Estimation of Linear Functional Relationships: Approximation Distributions and Connections with Simultaneous Equations in Econometrics," *Journal of the Royal Statistical Society*, Series B, 38, 1–36.

——— (1977): "Asymptotic Expansions of the Distributions of Estimates in Simultaneous Equations for Alternative Parameter Sequences," *Econometrica*, 45, 509–518.

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

——— (1984): *An Introduction to Multivariate Statistical Analysis*, Third Edition. New York: Wiley.

Anderson, T. W., and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System 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 System and their Asymptotic Expansions," *Econometrica*, 41, 683–714.

——— (1979): "Evaluation of the Distribution Function of the Two Stage Least Squares Estimate," *Econometrica*, 47, 163–182.

ANGRIST, J. D., AND A. 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.

BEKKER, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62, 657–681.

BOUND, J., D. A. JAEGER, AND R. BAKER (1995): "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90, 443–450.

BOWDEN, R. J., AND D. A. TURKINGTON (1984): *Instrumental Variables*. Cambridge, U.K.: Cambridge University Press.

BUSE, A. (1992): "The Bias of Instrumental Variable Estimators," *Econometrica*, 60, 173–180.

CAMPBELL, J. Y., AND N. G. MANKIW (1989): "Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence," *NBER Macroeconomics Annual 1989*, 185–215.

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.

DURBIN, J. (1954): "Errors in Variables," *Review of the International Statistical Institute*, 22, 23–32.

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*. Palo Alto: TSP International.

HAUSMAN, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1271.

HILLIER, G. H. (1990): "On the Normalization of Structural Equations: Properties of Direction Estimators," *Econometrica*, 58, 1181–1194.

MADDALA, G. S., AND J. JEONG (1992): "On the Exact Small Sample Distribution of the Instrumental Variable Estimator," *Econometrica*, 60, 181–183.

MAGDALINOS, M. A. (1990): "The Classical Principles of Testing Using Instrumental Variables Estimates," *Journal of Econometrics*, 44, 241–279.

——— (1994): "Testing Instrument Admissibility: Some Refined Asymptotic Results," *Econometrica*, 62, 373–404.

MARIANO, R. S. (1982): "Analytical Small-Sample Distribution Theory in Econometrics: The Simultaneous Equations Case," *International Economic Review*, 23, 503–534.

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.

——— (1989): "t Test in a Structural Equation," *Econometrica*, 57, 1341–1360.

MUIRHEAD, R. J. (1982): *Aspects of Multivariate Statistical Theory*. New York: Wiley.

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.

——— (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. (1980): "The Exact Finite Sample Density of Instrumental Variables Estimators in an Equation With $n+1$ Endogenous Variables," *Econometrica*, 48, 861–878.

——— (1983): "Exact Small Sample Theory in the Simultaneous Equations Model," in *Handbook of Econometrics, Vol. I*, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland, 449–516.

——— (1989): "Partially Identified Econometric Models," *Econometric Theory*, 5, 181–240.

RICHARDSON, D. H. (1968): "The Exact Distribution of a Structural Coefficient Estimator," *Journal of the American Statistical Association*, 63, 1214–1226.

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

SARGAN, J. D. (1958): "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393–415.

——— (1983): "Identification and Lack of Identification," *Econometrica*, 51, 1605–1633.

SAWA, T. (1969): "The Exact Sampling Distribution of Ordinary Least Squares and Two Stage Least Squares Estimators," *Journal of the American Statistical Association*, 64, 923–936.

WU, D.-M. (1973): "Alternative Tests of Independence Between Stochastic Regressors and Disturbances," *Econometrica*, 41, 733–750.

——— (1974): "Alternative Tests of Independence Between Stochastic Regressors and Disturbances: Finite Sample Results," *Econometrica*, 42, 529–546.