TECHNICAL WORKING PAPER SERIES

TESTING FOR WEAK INSTRUMENTS IN LINEAR IV REGRESSION

James H. Stock
Motohiro Yogo

Technical Working Paper 284
http://www.nber.org/papers/T0284

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2002

Prepared for the Festschrift in honor of Thomas Rothenberg. We thank Alastair Hall, Jerry Hausman, Takesi Hayakawa, George Judge, Whitney Newey, and Jonathan Wright for helpful comments and/or suggestions. This research was supported by NSF grants SBR-9730489 and SBR-0214131. The views expressed in this paper are those of the authors and not necessarily those of the National Bureau of Economic Research.

Testing for Weak Instruments in Linear IV Regression
James H. Stock and Motohiro Yogo
NBER Technical Working Paper No. 284
October 2002
JEL No. C2, C3

ABSTRACT

Weak instruments can produce biased IV estimators and hypothesis tests with large size distortions. But what, precisely, are weak instruments, and how does one detect them in practice? This paper proposes quantitative definitions of weak instruments based on the maximum IV estimator bias, or the maximum Wald test size distortion, when there are multiple endogenous regressors. We tabulate critical values that enable using the first-stage *F*-statistic (or, when there are multiple endogenous regressors, the Cragg-Donald (1993) statistic) to test whether given instruments are weak. A technical contribution is to justify sequential asymptotic approximations for IV statistics with many weak instruments.

James H. Stock
Department of Economics
Harvard University
Cambridge, MA 02138
and NBER
James_Stock@harvard.edu

Motohiro Yogo
Department of Economics
Harvard University
Cambridge, MA 02138
Yogo@fas.harvard.edu

1. Introduction

Standard treatments of instrumental variables (IV) regression stress that for instruments to be valid they must be exogenous. It is also important, however, that the second condition for a valid instrument, instrument relevance, holds, for if the instruments are only marginally relevant, or "weak," then first-order asymptotics can be a poor guide to the actual sampling distributions of conventional IV regression statistics.

At a formal level, the strength of the instruments matters because the natural measure of this strength – the so-called concentration parameter – plays a role formally akin to the sample size in IV regression statistics. Rothenberg (1984) makes this point in his survey of approximations to the distributions of estimators and test statistics. He considered the single equation IV regression model,

$$\boldsymbol{y} = \boldsymbol{Y}\beta + \boldsymbol{u}, \tag{1.1}$$

where $\boldsymbol{y}$ and $\boldsymbol{Y}$ are $T\times1$ vectors of observations on the dependent variable and endogenous regressor, respectively, and $\boldsymbol{u}$ is a $T\times1$ vector of i.i.d. $N(0,\sigma_{uu})$ errors. The reduced form equation for $\boldsymbol{Y}$ is

$$\boldsymbol{Y} = \boldsymbol{Z\Pi} + \boldsymbol{V}, \tag{1.2}$$

where $\boldsymbol{Z}$ is a $T\times K_2$ matrix of fixed, exogenous instrumental variables, $\boldsymbol{\Pi}$ is a $K_2\times1$ coefficient vector, and $\boldsymbol{V}$ is a $T\times1$ vector of i.i.d. $N(0,\sigma_{VV})$ errors, where $\text{corr}(u_t,V_t) = \rho$.

The two stage least squares (TSLS) estimator of β is $\hat{\beta}^{TSLS} = (\boldsymbol{Y}'\boldsymbol{P_Z y})/(\boldsymbol{Y}'\boldsymbol{P_Z Y})$, where $\boldsymbol{P_Z} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$. Rothenberg (1984) expresses $\hat{\beta}^{TSLS}$ as

$$\mu(\hat{\beta}^{TSLS} - \beta) = \left(\frac{\sigma_{uu}}{\sigma_{VV}}\right)^{1/2} \frac{\zeta_u + (S_{Vu}/\mu)}{1 + (2\zeta_V/\mu) + (S_{VV}/\mu^2)}, \qquad (1.3)$$

where $\zeta_u = \boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{u}/(\sigma_{uu}\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{1/2}$, $\zeta_V = \boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{V}/(\sigma_{VV}\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{1/2}$, $S_{Vu} = \boldsymbol{V}'\boldsymbol{P_Z u}/(\sigma_{uu}\sigma_{VV})^{1/2}$, $S_{VV} = \boldsymbol{V}'\boldsymbol{P_Z V}/\sigma_{VV}$, and μ is the square root of the concentration parameter, $\mu^2 = \boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi}/\sigma_{VV}$.

Under the assumptions of fixed instruments and normal errors, ζ_u and ζ_V are standard normal variables with correlation ρ, and S_{Vu} and S_{VV} are elements of a matrix with a central Wishart distribution. Because the distributions of ζ_u, ζ_V, S_{Vu}, and S_{VV} do not depend on the sample size, the sample size enters the distribution of the TSLS estimator only through the concentration parameter. In fact, the form of (1.3) makes it clear that μ^2 can be thought of as an effective sample size, in the sense that μ formally plays the role usually associated with $\sqrt{T}$. Rothenberg (1984) proceeds to discuss expansions of the distribution of the TSLS estimator in orders of μ, and he emphasizes that the quality of these approximations can be poor when μ^2 is small. This has been underscored by the dramatic numerical results of Nelson and Startz (1990a, 1990b) and Bound, Jaeger and Baker (1995).

If μ^2 is so small that inference based on some IV estimators and their conventional standard errors are potentially unreliable, then the instruments are said to be weak. But this raises two practical questions. First, precisely how small must μ^2 be for instruments to be weak? Second, because Π, and thus μ^2, is unknown, how is an applied researcher to know whether μ^2 is in fact sufficiently small that his or her instruments are weak?

This paper provides answers to these two questions. First, we develop quantitative definitions of what constitutes weak instruments. In our view, the matter of whether a group of instrumental variables is weak cannot be resolved in the abstract; rather, it depends on the inferential task to which the instruments are applied and how that inference is conducted. We therefore offer two alternative definitions of weak instruments. The first definition is that a group of instruments is weak if the bias of the IV estimator, relative to the bias of ordinary least squares (OLS), could exceed a certain threshold *b*, for example 10%. The second is that the instruments are weak if the conventional α-level Wald test based on IV statistics has an actual size that could exceed a certain threshold *r*, for example r = 10% when α = 5%. Each of these definitions yields a set of population parameters that defines weak instruments, that is, a "weak instrument set." Because different estimators (e.g., TSLS or LIML) have different properties when instruments are weak, the resulting weak instrument set depends on the estimator being used. For TSLS and other *k*-class estimators, we argue that these weak instrument sets can be characterized in terms of the minimum eigenvalue of the matrix version of μ^2/K_2.

Given this quantitative definition of weak instrument sets, we then show how to test the null hypothesis that a given group of instruments is weak against the alternative that it is strong. Our test is based on the Cragg-Donald (1993) statistic; when there is a single endogenous regressor, this statistic is simply the "first-stage *F*-statistic", the *F*-

statistic for testing the hypothesis that the instruments do not enter the first stage regression of TSLS. The critical values for the test statistic, however, are *not* Cragg and Donald's (1993): our null hypothesis is that the instruments are weak, even though the parameters might be identified, whereas Cragg and Donald (1993) test the null hypothesis of underidentification. We therefore provide tables of critical values that depend on the estimator being used, whether the researcher is concerned about bias or size distortion, and the numbers of instruments and endogenous regressors. These critical values are obtained using weak instrument asymptotic distributions (Staiger and Stock (1997)), which are more accurate than Edgeworth approximations when the concentration parameter is small.[1]

Additionally, this paper makes a separate contribution to the literature on distributions of IV estimators with weak instruments. Bekker (1994) obtained first-order distributions of various IV estimators under the assumptions that $K_2 \to \infty$, $T \to \infty$, and $K_2/T \to c$, $0 \le c < 1$, when μ^2/T is fixed and the errors are Gaussian. Chao and Swanson (2002) have explored the consistency of IV estimators with weak instruments when the number of instruments is large in the sense that K_2 is also modeled as increasing to infinity, but more slowly than T. Sargan (1975), Kunitomo (1980) and Morimune (1983) provided earlier treatments of large-K_2 asymptotics. We continue this line of work and provide conditions under which the Staiger-Stock (1997) weak instrument asymptotics hold, even if the number of instruments is increasing, as long as $K_2^2/T \to 0$. We refer to asymptotic limits taken under the sequence $K_2 \to \infty$, $T \to \infty$, such that $K_2^2/T \to 0$ and μ^2/K_2 is $O(1)$, as *many weak instrument limits*. It is shown in the appendix that these conditions justify using relatively straightforward sequential asymptotic calculations to compute limiting distributions under such sequences. Here, these many weak instrument limits are used to characterize the weak instrument sets when the number of instruments is moderate. Some of these results might be of more general interest, however; for example, Chao and Swanson (2002) show that LIML is consistent under these conditions, and we provide its $\sqrt{K_2}$ -limiting distribution.

This paper is part of a growing literature on detecting weak instruments, surveyed in Stock, Wright, and Yogo (2002). Cragg and Donald (1993) proposed a test of underidentification, which (as discussed above) is different than a test for weak instruments. Hall, Rudebusch, and Wilcox (1996), following on work by Bowden and Turkington (1984), suggested testing for underidentification using the minimum canonical correlation between the endogenous regressors and the instruments. Shea (1997) considered multiple included regressors and suggested looking at a partial R^2. Neither Hall, Rudebusch, and Wilcox (1996) nor Shea (1997) provide a formal characterization of weak instrument sets or a formal test for weak instruments, with controlled type I error, based on their respective statistics. For the case of a single endogenous regressor, Staiger and Stock (1997) suggested declaring instruments to be weak if the first-stage *F*-statistic is less than ten. Recently Hahn and Hausman (2002) suggested comparing the forward and reverse TSLS estimators and concluding that instruments are strong if the null hypothesis that these are the same cannot be rejected. Relative to this literature, the contribution of this paper is twofold. First, we provide a formal characterization of the weak instrument set for a general number of endogenous regressors. Second, we provide a test of whether given instruments fall in this set, that is, whether they are weak, where the size of the test is controlled asymptotically under the null of weak instruments.

The rest of the paper is organized as follows. The IV regression model and the proposed test statistic are presented in Section 2. The weak instrument sets are developed in Section 3. Section 4 presents the test for weak instruments and provides critical values for tests based on TSLS bias and size, Fuller-*k* bias, and LIML size. Section 5 examines the power of the test, and conclusions are presented in Section 6. Results on many weak instrument asymptotics are collected and proven in the appendix.

2. The IV Regression Model, the Proposed Test Statistic, and Initial Asymptotic Results

2.1. The IV Regression Model

We consider the linear IV regression model (1.1) and (1.2), generalized to have n included endogenous regressors $\boldsymbol{Y}$ and K_1 included exogenous regressors $\boldsymbol{X}$:

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

$$\boldsymbol{Y} = \boldsymbol{Z\Pi} + \boldsymbol{X\Phi} + \boldsymbol{V}, \tag{2.2}$$

where $\boldsymbol{Y}$ is now a $T\times n$ matrix of included endogenous variables, $\boldsymbol{X}$ is a $T\times K_1$ matrix of included exogenous variables (one column of which is 1's if (2.1) includes an intercept), and $\boldsymbol{Z}$ is a $T\times K_2$ matrix of excluded exogenous variables to be used as instruments. It is assumed throughout that $K_2 \geq n$. Let $\underline{\boldsymbol{Z}} = [\boldsymbol{X}\ \boldsymbol{Z}]$ denote the matrix of all the exogenous variables. The conformable vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ and the $\boldsymbol{\Pi}$ and $\boldsymbol{\Phi}$ are unknown parameters. Throughout this paper we exclusively consider inference about $\boldsymbol{\beta}$.

Let $\boldsymbol{X}_t = \begin{pmatrix} X_{1t} & \cdots & X_{K_1 t} \end{pmatrix}'$, $\boldsymbol{Z}_t = \begin{pmatrix} Z_{1t} & \cdots & Z_{K_2 t} \end{pmatrix}'$, and $\underline{\boldsymbol{Z}}_t = (\boldsymbol{X}_t'\ \boldsymbol{Z}_t')'$ denote the vectors of the t^{th} observations on these variables. Also let $\boldsymbol{Q}$ and $\boldsymbol{\Sigma}$ denote the population second moment matrices,

$$E\left[\begin{pmatrix} u_t \\ \boldsymbol{V}_t \end{pmatrix}\begin{pmatrix} u_t & \boldsymbol{V}_t' \end{pmatrix}\right] = \begin{bmatrix} \sigma_{uu} & \boldsymbol{\Sigma}_{uV} \\ \boldsymbol{\Sigma}_{Vu} & \boldsymbol{\Sigma}_{VV} \end{bmatrix} = \boldsymbol{\Sigma} \text{ and } E(\underline{\boldsymbol{Z}}_t\, \underline{\boldsymbol{Z}}_t') = \begin{bmatrix} \boldsymbol{Q}_{XX} & \boldsymbol{Q}_{XZ} \\ \boldsymbol{Q}_{ZX} & \boldsymbol{Q}_{ZZ} \end{bmatrix} = \boldsymbol{Q}. \quad (2.3)$$

2.2. *k*-Class Estimators and Wald Statistics

Let the superscript "$^\perp$" denote the residuals from the projection on $\boldsymbol{X}$, so for example $\boldsymbol{Y}^\perp = \boldsymbol{M_X Y}$, where $\boldsymbol{M_X} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. In this notation, the OLS estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\boldsymbol{Y}^{\perp\prime}\boldsymbol{Y}^\perp)^{-1}(\boldsymbol{Y}^{\perp\prime}\boldsymbol{y})$. The k-class estimator of $\boldsymbol{\beta}$ is,

$$\hat{\boldsymbol{\beta}}(k) = [\boldsymbol{Y}^{\perp\prime}(\boldsymbol{I} - k\boldsymbol{M}_{\boldsymbol{Z}^\perp})\boldsymbol{Y}^\perp]^{-1}[\boldsymbol{Y}^{\perp\prime}(\boldsymbol{I} - k\boldsymbol{M}_{\boldsymbol{Z}^\perp})\boldsymbol{y}^\perp]. \quad (2.4)$$

The Wald statistic testing the null hypothesis that $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, based on the k-class estimator, is

$$W(k) = \frac{[\hat{\boldsymbol{\beta}}(k) - \boldsymbol{\beta}_0]'[\boldsymbol{Y}^{\perp\prime}(I - k\boldsymbol{M}_{\boldsymbol{Z}^\perp})\boldsymbol{Y}^\perp][\hat{\boldsymbol{\beta}}(k) - \boldsymbol{\beta}_0]}{n\hat{\sigma}_{uu}(k)}, \quad (2.5)$$

where $\hat{\sigma}_{uu}(k) = \hat{\boldsymbol{u}}^{\perp}(k)'\hat{\boldsymbol{u}}^{\perp}(k)/(T - K_1 - n)$, where $\hat{\boldsymbol{u}}^{\perp}(k) = \boldsymbol{y}^{\perp} - \boldsymbol{Y}^{\perp}\hat{\boldsymbol{\beta}}(k)$.

This paper considers four specific *k*-class estimators: TSLS, the limited information maximum likelihood estimator (LIML), the family of modified LIML estimators proposed by Fuller (1977) ("Fuller-*k* estimators"), and bias-adjusted TSLS (BTSLS; Nagar (1959), Rothenberg (1984)). The values of *k* for these estimators are (cf. Donald and Newey (2001)):

TSLS: $k = 1,$ (2.6)

LIML: $k = \hat{k}_{LIML}$ is the smallest root of $\det(\underline{\boldsymbol{Y}}'\boldsymbol{M}_X\underline{\boldsymbol{Y}} - k\,\underline{\boldsymbol{Y}}'\boldsymbol{M}_{\underline{Z}}\underline{\boldsymbol{Y}}) = 0,$ (2.7)

Fuller-*k*: $k = \hat{k}_{LIML} - c/(T - K_1 - K_2)$, where c is a positive constant, (2.8)

BTSLS: $k = T/(T - K_2 + 2),$ (2.9)

where $\det(\boldsymbol{A})$ is the determinant of the matrix $\boldsymbol{A}$. If the errors are symmetrically distributed and the exogenous variables are fixed, LIML is median unbiased to second order (Rothenberg (1983)). In our numerical work, we examine the Fuller-*k* estimator with $c = 1$, which is the best unbiased estimator to second order among estimators with $k = 1 + a(\hat{k}_{LIML} - 1) - c/(T - K_1 - K_2)$ for some constants a and c (Rothenberg (1984)). For further discussion, see Donald and Newey (2001) and Stock, Wright, and Yogo (2002, Section 6.1).

2.3. The Cragg-Donald Statistic

The proposed test for weak instruments is based on the eigenvalue of the matrix analog of the *F*-statistic from the first stage regression of TSLS,

$$\boldsymbol{G}_T = \hat{\boldsymbol{\Sigma}}_{VV}^{-1/2}{}'\boldsymbol{Y}^{\perp}{}'\boldsymbol{P}_{\boldsymbol{Z}^{\perp}}\boldsymbol{Y}^{\perp}\hat{\boldsymbol{\Sigma}}_{VV}^{-1/2}/K_2, \tag{2.10}$$

where $\hat{\boldsymbol{\Sigma}}_{VV} = (\boldsymbol{Y}'\boldsymbol{M}_{\underline{\boldsymbol{Z}}}\boldsymbol{Y})/(T-K_1-K_2)$.[2]

The test statistic is the minimum eigenvalue of $\boldsymbol{G}_T$:

$$g_{\min} = \text{mineval}(\boldsymbol{G}_T). \tag{2.11}$$

This statistic was proposed by Cragg and Donald (1993) to test the null hypothesis of underidentification, which occurs when the concentration matrix is singular. Instead, we are interested in the case that the concentration matrix is nonsingular but still is sufficiently small that the instruments are weak. To obtain the limiting null distribution of the Cragg-Donald statistic (2.11) under weak instruments, we rely on weak instrument asymptotics.

2.4. Weak Instrument Asymptotics: Assumptions and Notation

We start by summarizing the elements of weak instrument asymptotics from Staiger and Stock (1997). The essential idea of weak instruments is that $\boldsymbol{Z}$ is only weakly related to $\boldsymbol{Y}$, given $\boldsymbol{X}$. Specifically, weak instrument asymptotics are developed by modeling $\boldsymbol{\Pi}$ as local to zero:

Assumption L$_{\Pi}$: $\boldsymbol{\Pi} = \boldsymbol{\Pi}_T = \boldsymbol{C}/\sqrt{T}$, where $\boldsymbol{C}$ is a fixed $K_2\times n$ matrix with bounded elements $C_{ij} \leq \bar{c}$.

Following Staiger and Stock (1997), we make the following assumption on the moments:

Assumption M. The following limits hold jointly for fixed K_2:

(a) $(T^{-1}\boldsymbol{u}'\boldsymbol{u}, T^{-1}\boldsymbol{V}'\boldsymbol{u}, T^{-1}\boldsymbol{V}'\boldsymbol{V}) \xrightarrow{p} (\sigma_{uu}, \boldsymbol{\Sigma}_{\boldsymbol{V}u}, \boldsymbol{\Sigma}_{\boldsymbol{VV}})$;

(b) $T^{-1}\underline{\boldsymbol{Z}}'\underline{\boldsymbol{Z}} \xrightarrow{p} \boldsymbol{Q}$;

(c) $(T^{-1/2}\boldsymbol{X}'\boldsymbol{u}, T^{-1/2}\boldsymbol{Z}'\boldsymbol{u}, T^{-1/2}\boldsymbol{X}'\boldsymbol{V}, T^{-1/2}\boldsymbol{Z}'\boldsymbol{V}) \xrightarrow{d} (\boldsymbol{\Psi}_{\boldsymbol{X}u}, \boldsymbol{\Psi}_{\boldsymbol{Z}u}, \boldsymbol{\Psi}_{\boldsymbol{XV}}, \boldsymbol{\Psi}_{\boldsymbol{ZV}})$, where $\boldsymbol{\Psi} \equiv$ $[\boldsymbol{\Psi}_{\boldsymbol{X}u}', \boldsymbol{\Psi}_{\boldsymbol{Z}u}', \text{vec}(\boldsymbol{\Psi}_{\boldsymbol{XV}})', \text{vec}(\boldsymbol{\Psi}_{\boldsymbol{ZV}})']'$ is distributed $N(0, \boldsymbol{\Sigma} \otimes \boldsymbol{Q})$.

Assumption M can hold for time series or cross-sectional data. Part (c) assumes that the errors are homoskedastic.

Notation and definitions. The following notation in effect transforms the variables and parameters and simplifies the asymptotic expressions. Let $\boldsymbol{\rho} = \boldsymbol{\Sigma}_{\boldsymbol{VV}}^{-1/2\,\prime}\boldsymbol{\Sigma}_{\boldsymbol{V}u}\sigma_{uu}^{-1/2}$, $\boldsymbol{\theta} = \boldsymbol{\Sigma}_{\boldsymbol{VV}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{V}u} = \sigma_{uu}^{1/2}\boldsymbol{\Sigma}_{\boldsymbol{VV}}^{-1/2}\boldsymbol{\rho}$, $\boldsymbol{\lambda} = \boldsymbol{\Omega}^{1/2}\boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{VV}}^{-1/2}$, $\boldsymbol{\Lambda} = \boldsymbol{\lambda}'\boldsymbol{\lambda}/K_2$, and $\boldsymbol{\Omega} = \boldsymbol{Q}_{\boldsymbol{ZZ}} - \boldsymbol{Q}_{\boldsymbol{ZX}}\boldsymbol{Q}_{\boldsymbol{XX}}^{-1}\boldsymbol{Q}_{\boldsymbol{XZ}}$. Note that $\boldsymbol{\rho}'\boldsymbol{\rho} \le 1$. Define the $K_2 \times 1$ and $K_2 \times n$ random variables, $\boldsymbol{z_u} = \boldsymbol{\Omega}^{-1/2\,\prime}(\boldsymbol{\Psi}_{\boldsymbol{Z}u} - \boldsymbol{Q}_{\boldsymbol{ZX}}\boldsymbol{Q}_{\boldsymbol{XX}}^{-1}\boldsymbol{\Psi}_{\boldsymbol{X}u})\sigma_{uu}^{-1/2}$ and $\boldsymbol{z_V} = \boldsymbol{\Omega}^{-1/2\,\prime}(\boldsymbol{\Psi}_{\boldsymbol{ZV}} - \boldsymbol{Q}_{\boldsymbol{ZX}}\boldsymbol{Q}_{\boldsymbol{XX}}^{-1}\boldsymbol{\Psi}_{\boldsymbol{XV}})\boldsymbol{\Sigma}_{\boldsymbol{VV}}^{-1/2}$, so

$$\begin{pmatrix} \boldsymbol{z_u} \\ \text{vec}(\boldsymbol{z_V}) \end{pmatrix} \sim N(\boldsymbol{0}, \bar{\boldsymbol{\Sigma}} \otimes \boldsymbol{I}_{K_2}), \text{ where } \bar{\boldsymbol{\Sigma}} = \begin{bmatrix} 1 & \boldsymbol{\rho}' \\ \boldsymbol{\rho} & \boldsymbol{I}_n \end{bmatrix}. \tag{2.12}$$

Also let

$$\nu_1 = (\boldsymbol{\lambda} + \boldsymbol{z_V})' (\boldsymbol{\lambda} + \boldsymbol{z_V}) \quad \text{and} \tag{2.13}$$

$$\nu_2 = (\boldsymbol{\lambda} + \boldsymbol{z_V})' \boldsymbol{z_u}. \tag{2.14}$$

2.5. Selected Weak Instrument Asymptotic Representations

For convenience, we summarize the following results from Staiger and Stock (1997), using simpler notation.

OLS estimator. Under Assumptions L_Π and M, the probability limit of the OLS estimator is $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta} + \boldsymbol{\theta}$.

k-class estimators. Suppose that $T(k - 1) \xrightarrow{d} \kappa$. Then under Assumptions L_Π and M,

$$\hat{\boldsymbol{\beta}}(k) - \boldsymbol{\beta} \xrightarrow{d} \sigma_{uu}^{1/2} \boldsymbol{\Sigma}_{VV}^{-1/2} (\nu_1 - \kappa \boldsymbol{I}_n)^{-1} (\nu_2 - \kappa \boldsymbol{\rho}) \quad \text{and} \tag{2.15}$$

$$W(k) \xrightarrow{d} \frac{(\nu_2 - \kappa\boldsymbol{\rho})'(\nu_1 - \kappa\boldsymbol{I}_n)^{-1}(\nu_2 - \kappa\boldsymbol{\rho})}{n[1 - 2\boldsymbol{\rho}'(\nu_1 - \kappa\boldsymbol{I}_n)^{-1}(\nu_2 - \kappa\boldsymbol{\rho}) + (\nu_2 - \kappa\boldsymbol{\rho})'(\nu_1 - \kappa\boldsymbol{I}_n)^{-2}(\nu_2 - \kappa\boldsymbol{\rho})]}, \tag{2.16}$$

where (2.16) holds under the null hypothesis $\boldsymbol{\beta} = \boldsymbol{\beta}_0$.

For LIML and the Fuller-k estimators, κ is a random variable, while for TSLS and BTSLS κ is nonrandom. Let $\boldsymbol{\Xi}$ be the $(n + 1) \times (n + 1)$ matrix, $\boldsymbol{\Xi} =$ $[z_u \;\; (\boldsymbol{\lambda} + \boldsymbol{z_V})]' [z_u \;\; (\boldsymbol{\lambda} + \boldsymbol{z_V})]$. Then the limits in (2.15) and (2.16) hold with:

TSLS: $\kappa = 0,$ (2.17)

LIML: $\kappa = \kappa^*$, where κ^* is the smallest root of $\det(\boldsymbol{\Xi} - \kappa \bar{\boldsymbol{\Sigma}}) = 0$, (2.18)

Fuller-k: $\kappa = \kappa^* - c$, where c is the constant in (2.8), and (2.19)

BTSLS: $\kappa = K_2 - 2.$ (2.20)

Note that the convergence in distribution of $T(\hat{k}_{LIML} - 1) \xrightarrow{d} \kappa^*$ is joint with the convergence in (2.15) and (2.16). For TSLS, the expressions in (2.15) and (2.16) simplify to

$$\hat{\boldsymbol{\beta}}^{TSLS} - \boldsymbol{\beta} \xrightarrow{d} \sigma_{uu}^{1/2} \boldsymbol{\Sigma}_{VV}^{-1/2} \boldsymbol{\nu}_1^{-1}\boldsymbol{\nu}_2 \quad \text{and} \tag{2.21}$$

$$W^{TSLS} \xrightarrow{d} \frac{\boldsymbol{\nu}_2'\boldsymbol{\nu}_1^{-1}\boldsymbol{\nu}_2}{n(1 - 2\boldsymbol{\rho}'\boldsymbol{\nu}_1^{-1}\boldsymbol{\nu}_2 + \boldsymbol{\nu}_2'\boldsymbol{\nu}_1^{-2}\boldsymbol{\nu}_2)}. \tag{2.22}$$

***Weak instrument asymptotic representations*: *the Cragg-Donald statistic*.**

Under the weak instrument asymptotic assumptions, the matrix $\boldsymbol{G}_T$ in (2.10) and the Cragg-Donald statistic (2.11) have the limiting distributions,

$$\boldsymbol{G}_T \xrightarrow{d} \boldsymbol{\nu}_1/K_2 \quad \text{and} \tag{2.23}$$

$$g_{\min} \xrightarrow{d} \text{mineval}(\boldsymbol{\nu}_1/K_2). \tag{2.24}$$

Inspection of (2.13) reveals that $\boldsymbol{\nu}_1$ has a noncentral Wishart distribution with noncentrality matrix $\boldsymbol{\lambda}'\boldsymbol{\lambda} = K_2\boldsymbol{\Lambda}$. This noncentrality matrix is the weak instrument limit of the concentration matrix:

$$\boldsymbol{\Sigma}_{VV}^{-1/2}\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi}\boldsymbol{\Sigma}_{VV}^{-1/2\,\prime} \xrightarrow{p} K_2\boldsymbol{\Lambda}. \tag{2.25}$$

Thus the weak instrument asymptotic distribution of the Cragg-Donald statistic $g_{\min}$ is that of the minimum eigenvalue of a noncentral Wishart, divided by K_2, where the noncentrality parameter is $K_2\boldsymbol{\Lambda}$. To obtain critical values for the weak instrument test based on $g_{\min}$, we characterize the weak instrument set in terms of the eigenvalues of $\boldsymbol{\Lambda}$, the task taken up in the next section.

3. Weak Instrument Sets

This section provides two general definitions of a weak instrument set, the first based on the bias of the estimator and the second based on size distortions of the associated Wald statistic. These two definitions are then specialized to TSLS, LIML, the Fuller-*k* estimator, and BTSLS, and the resulting weak instrument sets are characterized in terms of the minimum eigenvalues of the concentration matrix.

3.1. First Characterization of a Weak Instrument Set: Bias

One consequence of weak instruments is that IV estimators are in general biased, so our first definition of a weak instrument set is in terms of its maximum bias.

When there is a single endogenous regressor, it is natural to discuss bias in the units of $\boldsymbol{\beta}$, but for $n > 1$, a bias measure must scale $\boldsymbol{\beta}$ so that the bias is comparable across elements of $\boldsymbol{\beta}$. A natural way to do this is to standardize the regressors $\boldsymbol{Y}^{\perp}$ so that they have unit standard deviation and are orthogonal or, equivalently, to rotate $\boldsymbol{\beta}$ by $\boldsymbol{\Sigma}_{\boldsymbol{Y}^{\perp}\boldsymbol{Y}^{\perp}}^{1/2}$, where $\boldsymbol{\Sigma}_{\boldsymbol{Y}^{\perp}\boldsymbol{Y}^{\perp}} = E(\boldsymbol{Y}^{\perp\prime}\boldsymbol{Y}^{\perp}/T)$. In these standardized units, the squared bias of an IV estimator, which we generically denote $\hat{\boldsymbol{\beta}}^{IV}$, is $(E\hat{\boldsymbol{\beta}}^{IV} - \boldsymbol{\beta})'\boldsymbol{\Sigma}_{\boldsymbol{Y}^{\perp}\boldsymbol{Y}^{\perp}}(E\hat{\boldsymbol{\beta}}^{IV} - \boldsymbol{\beta})$. As our measure of bias, we therefore consider the relative squared bias of the candidate IV estimator $\hat{\boldsymbol{\beta}}^{IV}$, relative to the bias of the OLS estimator,

$$B_T^2 = \frac{(E\hat{\boldsymbol{\beta}}^{\mathrm{IV}} - \boldsymbol{\beta})'\boldsymbol{\Sigma}_{Y^{\perp}Y^{\perp}}(E\hat{\boldsymbol{\beta}}^{\mathrm{IV}} - \boldsymbol{\beta})}{(E\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{\Sigma}_{Y^{\perp}Y^{\perp}}(E\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}. \tag{3.1}$$

If $n = 1$, then the scaling matrix in (3.1) drops out and the expression simplifies to $B_T = |E\hat{\beta}^{IV} - \beta|/|E\hat{\beta} - \beta|$. The measure (3.1) was proposed, but not pursued, in Staiger and Stock (1997).

The asymptotic relative bias, computed under weak instrument asymptotics, is denoted by $B = \lim_{T\to\infty} B_T$. Under weak instrument asymptotics, $E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to \boldsymbol{\theta} = \sigma_{uu}^{1/2}\,\boldsymbol{\Sigma}_{\boldsymbol{VV}}^{-1/2}\boldsymbol{\rho}$ and $\boldsymbol{\Sigma}_{\boldsymbol{Y}^{\perp}\boldsymbol{Y}^{\perp}} \to \boldsymbol{\Sigma}_{\boldsymbol{VV}}$, so the denominator in (3.1) has the limit $(E\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{\Sigma}_{Y^{\perp}Y^{\perp}}(E\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to \sigma_{uu}\boldsymbol{\rho}'\boldsymbol{\rho}$. Thus the square of the asymptotic relative bias is

$$B^2 = \sigma_{uu}^{-1}\lim_{T\to\infty} \frac{(E\hat{\boldsymbol{\beta}}^{IV} - \boldsymbol{\beta})'\boldsymbol{\Sigma}_{Y^{\perp}Y^{\perp}}(E\hat{\boldsymbol{\beta}}^{IV} - \boldsymbol{\beta})}{\boldsymbol{\rho}'\boldsymbol{\rho}}. \tag{3.2}$$

We deem instruments to be strong if they lead to reliable inferences for all possible degrees of simultaneity ρ; otherwise they are weak. Applied to the relative bias measure and assuming $\rho'\rho > 0$, this leads us to consider the worst-case asymptotic bias,

$$B^{max} = \max_{\rho:\, 0 < \rho'\rho \leq 1}|B|. \tag{3.3}$$

The first definition of a weak instrument set is based on this worst-case bias. We define the weak instrument set, based on relative bias, to consist of those instruments that have the potential of leading to asymptotic relative bias greater than some value b. In population, the strength of an instrument is determined by the parameters of the reduced form equation (2.2). Accordingly, let $\mathcal{Z} = \{\boldsymbol{\Pi}, \boldsymbol{\Sigma}_{VV}, \boldsymbol{\Omega}\}$. The relative bias definition of weak instruments is

$$\mathcal{W}_{bias} = \{\mathcal{Z}\text{: } B^{max} \geq b\}. \tag{3.4}$$

Relative bias vs. absolute bias. Our motivation for normalizing the squared bias measure by the bias of the OLS estimator is that it helps to separate the two problems of endogeneity (OLS bias) and weak instrument (IV bias). For example, in an application to estimating the returns to education, based on a reading of the literature the researcher might believe that the maximum OLS bias is ten percentage points; if the relative bias measure in (3.1) is 0.1, then the maximum bias of the IV estimator is one percentage point. Thus formulating the bias measure in (3.1) as a relative bias measure allows the researcher to return to the natural units of the application using expert judgment about the possible magnitude of the OLS bias. This said, for TSLS it is possible to reinterpret the maximal relative bias measure in terms of maximal absolute bias, a point to which we return in Section 3.3.

3.2. Second Characterization of a Weak Instrument Set: Size

Our second definition of a weak instrument set is based on the maximal size of the Wald test of all the elements of β. In parallel to the approach for the bias measure, we consider an instrument strong from the perspective of the Wald test if the size of the test is close to its level for all possible configurations of the IV regression model. Let W^{IV} denote the Wald test statistic based on the candidate IV estimator $\hat{\beta}^{IV}$. For the estimators considered here, under conventional first-order asymptotics W^{IV} has a chi-squared null distribution with n degrees of freedom, divided by n. The actual rejection rate R_T under the null hypothesis is

$$R_T = \Pr_{\beta_0}[W^{IV} > \chi^2_{n;\alpha}/n], \tag{3.5}$$

where $\chi^2_{n;\alpha}$ is the α-level critical value of the chi-squared distribution with n degrees of freedom and α is the nominal level of the test.

In general, the rejection rate in (3.5) depends on ρ. As in the definitions of the bias-based weak instrument set, we consider the worst-case limiting rejection rate,

$$R^{max} = \max_{\rho:\ \rho'\rho \le 1} R, \text{ where } R = \lim_{T\to\infty} R_T. \tag{3.6}$$

The size-based weak instrument set $\mathcal{W}_{size}$ consists of instruments that can lead to a size of at least $r > \alpha$:

$$\mathcal{W}_{size} = \{\mathcal{Z}: R^{max} \geq r\}. \tag{3.7}$$

For example, if $\alpha = .05$ then a researcher might consider it acceptable if the worst case size is $r = .10$.

3.3 Weak Instrument Sets for TSLS

We now apply these general definitions of weak instrument sets to TSLS and argue that the sets can be characterized in terms of the minimum eigenvalue of $\boldsymbol{\Lambda}$.

Weak instrument set based on TSLS bias. Under weak instrument asymptotics,

$$(B_T^{TSLS})^2 \rightarrow \frac{\boldsymbol{\rho}'\boldsymbol{h}'\boldsymbol{h}\boldsymbol{\rho}}{\boldsymbol{\rho}'\boldsymbol{\rho}} \equiv (B^{TSLS})^2 \quad \text{and} \tag{3.8}$$

$$(B^{max,\, TSLS})^2 = \max_{\boldsymbol{\rho}:\, 0 < \boldsymbol{\rho}'\boldsymbol{\rho} \leq 1} \frac{\boldsymbol{\rho}'\boldsymbol{h}'\boldsymbol{h}\boldsymbol{\rho}}{\boldsymbol{\rho}'\boldsymbol{\rho}}, \tag{3.9}$$

where $\boldsymbol{h} = E[\boldsymbol{\nu}_1^{-1}(\boldsymbol{\lambda} + \boldsymbol{z_V})'\boldsymbol{z_V}]$. The asymptotic relative bias B^{TSLS} depends on $\boldsymbol{\rho}$ and $\boldsymbol{\lambda}$, which are unknown, as well as K_2 and n.

Because $\boldsymbol{h}$ depends on $\boldsymbol{\lambda}$ but not $\boldsymbol{\rho}$, by (3.8) we have that $B^{max,\, TSLS} = [\text{maxeval}(\boldsymbol{h}'\boldsymbol{h})]^{1/2}$, where maxeval($\boldsymbol{A}$) denotes the maximum eigenvalue of the matrix $\boldsymbol{A}$.

By applying the singular value decomposition to $\boldsymbol{\lambda}$ it is further possible to show that the maximum eigenvalue of $\boldsymbol{h}'\boldsymbol{h}$ depends only on K_2, n, and the eigenvalues of $\boldsymbol{\lambda}'\boldsymbol{\lambda}/K_2 = \boldsymbol{\Lambda}$. It follows that, for a given K_2 and n, the boundary b of the TSLS bias weak instrument set is a function only of the eigenvalues of $\boldsymbol{\Lambda}$.

When the number of instruments is treated as a slowly growing function of the sample size, it is further possible to show that the boundary of the weak instrument set is a decreasing function of the minimum eigenvalue of $\boldsymbol{\Lambda}$. Specifically, consider sequences of K_2 and T such that $K_2 \to \infty$ and $T \to \infty$ jointly, subject to $K_2^2/T \to 0$, where $\boldsymbol{\Lambda}$ (which in general depends on K_2) is held constant as $K_2 \to \infty$; we write this joint limit as (K_2, $T \to \infty$) and, as in the introduction, refer to it as representing "many weak instruments." It follows from (3.9) and appendix Equation (A.14) that the many weak instrument limit of B_T^{TSLS} is[3],

$$\lim_{(K_2,T\to\infty)} (B_T^{TSLS})^2 = \frac{\boldsymbol{\rho}'(\boldsymbol{\Lambda}+\boldsymbol{I})^{-2}\boldsymbol{\rho}}{\boldsymbol{\rho}'\boldsymbol{\rho}}. \tag{3.10}$$

By solving the maximization problem (3.9), we obtain the many weak instrument limit, $B^{max,\,TSLS} = (1 + \text{mineval}(\boldsymbol{\Lambda}))^{-1}$. It follows that, for many instruments, the set $\mathcal{W}_{bias,\,TSLS}$ can be characterized by the minimum eigenvalue of $\boldsymbol{\Lambda}$, and the TSLS weak instrument set $\mathcal{W}_{bias,\,TSLS}$ can be written as

$$\mathcal{W}_{bias,\,TSLS} = \{\mathcal{Z}\text{: mineval}(\boldsymbol{\Lambda}) \le \ell_{bias,\,TSLS}(b;K_2,n)\}, \tag{3.11}$$

where $\ell_{bias,\,TSLS}(b;K_2,n)$ is a decreasing function of the maximum allowable bias b.

Our formal justification for the simplification that $\mathcal{W}_{bias,\,TSLS}$ depends only on the smallest eigenvalue of $\boldsymbol{\Lambda}$, rather than on all its eigenvalues, rests on the many weak instrument asymptotic result (3.10). Numerical analysis for $n = 2$ suggests, however, that $B^{max,\,TSLS}$ is decreasing in each eigenvalue of $\boldsymbol{\Lambda}$ for all values of K_2. These numerical results suggest that the simplification in (3.11), relying only on the minimum eigenvalue, is valid for all K_2 under weak instrument asymptotics, even though we currently cannot provide a formal proof.[4]

We note that although B^{max} was defined as maximal bias relative to OLS, for TSLS this is also the maximal absolute bias in standardized units. The numerator of (3.8) is evidently maximized when $\boldsymbol{\rho}'\boldsymbol{\rho} = 1$. Thus, for TSLS, (3.2) can be restated as $(B^{max})^2 = \sigma_{uu}^{-1}\max_{\boldsymbol{\rho}:\,\boldsymbol{\rho}'\boldsymbol{\rho}=1}\lim_{T\to\infty}(E\hat{\boldsymbol{\beta}}^{TSLS}-\boldsymbol{\beta})'\boldsymbol{\Sigma}_{Y^{\perp}Y^{\perp}}(E\hat{\boldsymbol{\beta}}^{TSLS}-\boldsymbol{\beta})$. But $(E\hat{\boldsymbol{\beta}}^{TSLS}-\boldsymbol{\beta})'\boldsymbol{\Sigma}_{Y^{\perp}Y^{\perp}}(E\hat{\boldsymbol{\beta}}^{TSLS}-\boldsymbol{\beta})$ is the squared bias of $\hat{\boldsymbol{\beta}}^{TSLS}$, not relative to the bias of the OLS estimator. For TSLS, then, the relative bias measure can alternatively be reinterpreted as the maximal bias of the candidate IV estimator, in the standardized units of $\sigma_{uu}^{-1/2}\boldsymbol{\Sigma}_{\boldsymbol{Y}^{\perp}\boldsymbol{Y}^{\perp}}^{1/2}$.

Weak instrument set based on TSLS size. For TSLS, it follows from (2.22) that the worst-case asymptotic size is

$$R^{max,\,TSLS} = \max_{\boldsymbol{\rho}:\,\boldsymbol{\rho}'\boldsymbol{\rho}\le 1}\Pr\left[\frac{\boldsymbol{\nu}_2'\boldsymbol{\nu}_1^{-1}\boldsymbol{\nu}_2}{1-2\boldsymbol{\rho}'\boldsymbol{\nu}_1^{-1}\boldsymbol{\nu}_2+\boldsymbol{\nu}_2'\boldsymbol{\nu}_1^{-2}\boldsymbol{\nu}_2} > \chi^2_{n;\alpha}\right]. \qquad (3.12)$$

$R^{max,\,TSLS}$, and consequently $\mathcal{W}_{size,\,TSLS}$, depends only on the eigenvalues of $\boldsymbol{\Lambda}$ as well as n and K_2 (the reason is the same as for the similar assertion for $B^{max,\,TSLS}$).

When the number of instruments is large, the Wald statistic is maximized when $\rho'\rho = 1$ and is an increasing function of the eigenvalues of $\boldsymbol{\Lambda}$. Specifically, it is shown in the appendix (Equation (A.15)) that the many weak instrument limit of the TSLS Wald statistic, divided by K_2, is

$$W^{TSLS}/K_2 \xrightarrow{p} \frac{\rho'(\boldsymbol{\Lambda}+\boldsymbol{I}_n)^{-1}\rho}{n[1-2\rho'(\boldsymbol{\Lambda}+\boldsymbol{I}_n)^{-1}\rho+\rho'(\boldsymbol{\Lambda}+\boldsymbol{I}_n)^{-2}\rho]}. \qquad (3.13)$$

The right hand side of (3.13) is maximized when $\rho'\rho = 1$, in which case this expression can be written, $\rho'(\boldsymbol{\Lambda} + \boldsymbol{I}_n)^{-1}\rho/\rho'[\boldsymbol{I}_n - (\boldsymbol{\Lambda} + \boldsymbol{I}_n)^{-1}]^2\rho$. In turn, the maximum of this ratio over ρ depends only on the eigenvalues of $\boldsymbol{\Lambda}$ and is decreasing in those eigenvalues.

The many weak instrument limit of $R^{max,\,TSLS}$ is

$$R^{max,\,TSLS} = \max_{\rho:\,\rho'\rho\le 1} \lim_{(K_2,T\to\infty)} \Pr[W^{TSLS}/K_2 > \chi^2_{n;\alpha}/(nK_2)] = 1, \qquad (3.14)$$

where the limit follows from (3.13) and from $\chi^2_{n;\alpha}/(nK_2) \to 0$. With many weak instruments the TSLS Wald statistic W^{TSLS} increases linearly in K_2, so the boundary of the weak instrument set, in terms of the eigenvalues of $\boldsymbol{\Lambda}$, increases as a function of K_2 without bound.

For small values of K_2, numerical analysis suggests that $R^{max,\, TSLS}$ is a nonincreasing function of all the eigenvalues of $\boldsymbol{\Lambda}$, which (if so) implies that the boundary of the weak instrument set can, for small K_2, be characterized in terms of this minimum eigenvalue. The argument leading to (3.11) therefore applies here and leads to the characterization,

$$\mathcal{W}_{size,\, TSLS} = \{\mathcal{Z}\text{: mineval}(\boldsymbol{\Lambda}) \le \ell_{size,\, TSLS}(r;K_2,n,\alpha)\ \}, \qquad (3.15)$$

where $\ell_{size,\, TSLS}(r;K_2,n,\alpha)$ is decreasing in the maximal allowable size r.

3.4 Weak Instrument Sets for Other *k*-class Estimators

The general definitions of weak instrument sets given in Sections 3.1 and 3.2 also can be applied to other IV estimators. The weak instrument asymptotic distribution for general *k*-class estimators is given in Section 2.2. What remains to be shown is that the weak instrument sets, defined for specific estimators and test statistics, can be characterized in terms of the minimum eigenvalue of $\boldsymbol{\Lambda}$. As in the case of TSLS, the argument for the estimators considered here has two parts, for small K_2 and for large K_2.

For small K_2, the argument applied for the TSLS bias can be used generally for *k*-class statistics to show that, given K_2 and n, the *k*-class maximal relative bias and maximal size depend only on the eigenvalues of $\boldsymbol{\Lambda}$. In general, this dependence is complicated and we do not have theoretical results characterizing this dependence. Numerical work for $n = 1$ and $n = 2$ indicates, however, that the maximal bias and maximal size measures are decreasing in each of the eigenvalues of $\boldsymbol{\Lambda}$ in the relevant range of those eigenvalues. This in turn means that the boundary of the weak instrument set can be written in terms of the minimum eigenvalue of $\boldsymbol{\Lambda}$, although this characterization could be conservative (see footnote 4). [5]

For large K_2, we can provide theoretical results, based on many weak instrument limits, showing that the boundary of the weak instrument set depends only on mineval($\boldsymbol{\Lambda}$). These results are summarized here.

LIML and Fuller-k. It is shown in the appendix (Equations (A.19) and (A.20)) that the LIML and Fuller-*k* estimators and their Wald statistics have the many weak instrument asymptotic distributions,

$$\sqrt{K_2}\,(\hat{\boldsymbol{\beta}}^{LIML} - \boldsymbol{\beta}) \xrightarrow{d} N(0,\, \sigma_{uu}\boldsymbol{\Sigma}_{VV}^{-1/2}\boldsymbol{\Lambda}^{-1}(\boldsymbol{\Lambda} + \boldsymbol{I}_n - \boldsymbol{\rho\rho}')\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}_{VV}^{-1/2\prime}) \qquad (3.16)$$

$$W^{LIML} \xrightarrow{d} \boldsymbol{x}'(\boldsymbol{\Lambda} + \boldsymbol{I}_n - \boldsymbol{\rho\rho}')^{1/2}\boldsymbol{\Lambda}^{-1}(\boldsymbol{\Lambda} + \boldsymbol{I}_n - \boldsymbol{\rho\rho}')^{1/2\prime}\boldsymbol{x}/n, \text{ where } x \sim N(0, \boldsymbol{I}_n), \qquad (3.17)$$

where these distributions are written for LIML but also apply to Fuller-*k*.

An implication of (3.16) is that the LIML and Fuller-*k* estimators are consistent under the sequence $(K_2, T) \to \infty$, a result shown by Chao and Swanson (2002) for LIML. Thus the many weak instrument maximal relative bias for these estimators is 0.

An implication of (3.17) is that the Wald statistic is distributed as a weighted sum of n independent chi-squared random variables. When $n = 1$, it follows from (3.17) that the many weak instrument size has the simple form,

$$R^{max,\,LIML} = \max_{\boldsymbol{\rho}:\, \boldsymbol{\rho}'\boldsymbol{\rho} \le 1} \lim_{(K_2, T \to \infty)} \Pr[W^{LIML} > \chi^2_{1;\alpha}]$$

$$= \Pr[\chi_1^2 > \frac{\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}+1}\chi^2_{1;\alpha}], \qquad (3.18)$$

that is, the maximal size is the tail probability that a chi-squared distribution with one degree of freedom exceeds $(\boldsymbol{\Lambda}/(\boldsymbol{\Lambda} + 1))\ \chi^2_{1;\alpha}$. This evidently is decreasing in $\boldsymbol{\Lambda}$ and depends only on $\boldsymbol{\Lambda}$ (which trivially here is its minimum eigenvalue).

BTSLS. The many weak instrument asymptotic distributions of the BTSLS estimator and Wald statistic, derived in the appendix (Equations (A.16) and (A.17)), are

$$\sqrt{K_2}\,(\hat{\boldsymbol{\beta}}^{BTSLS} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \sigma_{uu}\boldsymbol{\Sigma}_{\boldsymbol{VV}}^{-1/2}\boldsymbol{\Lambda}^{-1}(\boldsymbol{\Lambda} + \boldsymbol{I}_n + \boldsymbol{\rho\rho}')\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{VV}}^{-1/2}{}'), \qquad (3.19)$$

$$W^{BTSLS} \xrightarrow{d} \boldsymbol{x}'(\boldsymbol{\Lambda} + \boldsymbol{I}_n + \boldsymbol{\rho\rho}')^{1/2}\boldsymbol{\Lambda}^{-1}(\boldsymbol{\Lambda} + \boldsymbol{I}_n + \boldsymbol{\rho\rho}')^{1/2}{}'\boldsymbol{x}/n, \text{ where } x \sim N(0, \boldsymbol{I}_n). \qquad (3.20)$$

It follows from (3.19) that the BTSLS estimator is consistent and that its maximal relative bias tends to zero under many weak instrument asymptotics.

For $n = 1$, the argument leading to (3.18) applies to BTSLS, except that the factor is different: the many weak instrument limit of the maximal size is

$$R^{max,\,BTSLS} = \Pr[\chi_1^2 > \frac{\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}+2}\chi^2_{1;\alpha}], \qquad (3.21)$$

which is a decreasing function of $\boldsymbol{\Lambda}$.

It is interesting to note that, according to (3.18) and (3.21), for a given value of $\boldsymbol{\Lambda}$ the maximal size distortion of LIML and Fuller-*k* tests is less than that of BTSLS when there are many weak instruments.

3.5. Numerical Results for TSLS, LIML, and Fuller-*k*

We have computed weak instrument sets based on maximum bias and size for a several *k*-class statistics. Here, we focus on TSLS bias and size, Fuller-*k* (with $c = 1$ in (2.8)) bias, and LIML size. Because LIML does not have moments in finite samples, LIML bias is not well-defined so we do not analyze it here.

The TSLS maximal relative bias was computed by Monte Carlo simulation for a grid of minimal eigenvalue of $\boldsymbol{\Lambda}$ from 0 to 30 for $K_2 = n + 2,\ldots, 100$, using 20,000 Monte Carlo draws. Computing the maximum TSLS bias entails computing $\boldsymbol{h}$ defined following (3.8) by Monte Carlo simulation, given n, K_2, then computing the maximum bias, $[\text{maxeval}(\boldsymbol{h}'\boldsymbol{h})]^{1/2}$. Computing the maximum bias of Fuller-*k* and the maximum size distortions of TSLS and LIML is more involved than computing the maximal TSLS bias because there is no simple analytic solution to the maximum problem (3.6). Numerical analysis indicates that R^{TSLS} is maximized when $\boldsymbol{\rho}'\boldsymbol{\rho} = 1$, so the maximization for $n = 2$ was done by transforming to polar coordinates and performing a grid search over the half unit circle (half because of symmetry in (2.22)). For Fuller-*k* bias and LIML size, maximization was performed over this half circle and over $0 \leq \boldsymbol{\rho}'\boldsymbol{\rho} \leq 1$. Because the bias and size measures appear to be decreasing functions of all the eigenvalues, at least in the relevant range, we set $\boldsymbol{\Lambda} = \ell \boldsymbol{I}_n$. The TSLS size calculations were performed using a grid of ℓ with $0 \leq \ell \leq 75$ (100,000 Monte Carlo draws); for Fuller-*k* bias, $0 \leq \ell \leq 12$ (50,000 Monte Carlo draws); and for LIML size, $0 \leq \ell \leq 10$ (100,000 Monte Carlo draws).

The minimal eigenvalues of $\boldsymbol{\Lambda}$ that constitute the boundaries of $\mathcal{W}_{bias,\ TSLS}$, $\mathcal{W}_{size,\ TSLS}$, $\mathcal{W}_{bias,\ Fuller\text{-}k}$, and $\mathcal{W}_{size,\ LIML}$ are plotted, respectively, in the top panels of Figures 1 – 4 for various cutoff values b and r. First consider the regions based on bias. The boundary of $\mathcal{W}_{bias,\ TSLS}$ is essentially flat in K_2 for K_2 sufficiently large; moreover, the boundaries for $n = 1$ and $n = 2$ are numerically very similar, even for small K_2. The boundary of the relative bias region for $b = .1$ (10% bias) asymptotes to approximately 8 for both $n = 1$ and $n = 2$. In contrast, the boundary of the bias region for Fuller-k tends to zero as the number of instruments increase, which agrees with the consistency of the Fuller-k estimator under many weak instrument asymptotics.

Turning to the regions based on size, the boundary of $\mathcal{W}_{size,\ TSLS}$ depends strongly on K_2 and n; as suggested by (3.14), the boundary is approximately linear in K_2 for K_2 sufficiently large. The boundary eigenvalues are very large when the degree of overidentification is large. For example, if one is willing to tolerate a maximal size of 15%, so the size distortion is 10% for the 5% level test, then with 10 instruments the minimum eigenvalue boundary is approximately 20 for $n = 1$ and approximately 16 for $n = 2$. In contrast, the boundary of $\mathcal{W}_{size,\ LIML}$ decreases with K_2 for both $n = 1$ and $n = 2$. Comparing these two plots provides a concrete assessment that tests based on LIML are far more robust to weak instruments than tests based on TSLS.

4. Test for Weak Instruments

This section provides critical values for the weak instrument test based on the Cragg-Donald (1993) statistic, $g_{\min}$. These critical values are based on the boundaries of the weak instrument sets obtained in Section 3 and on a bound on the asymptotic distribution of $g_{\min}$.

4.1 A Bound on the Asymptotic Distribution of g_{min}.

Recall that the Cragg-Donald statistic g_{min} is the minimum eigenvalue of $\boldsymbol{G}_T$, where $\boldsymbol{G}_T$ is given by (2.10). As stated in (2.23), under weak instrument asymptotics, $K_2\boldsymbol{G}_T$ is asymptotically distributed as a noncentral Wishart with dimension n, degrees of freedom K_2, identity covariance matrix, and noncentrality matrix $K_2\boldsymbol{\Lambda}$; that is,

$$\boldsymbol{G}_T \xrightarrow{d} \boldsymbol{\nu}_1/K_2 \sim \mathrm{W}_n(K_2, \boldsymbol{I}_n, K_2\boldsymbol{\Lambda})/K_2. \qquad (4.1)$$

The joint pdf for the n eigenvalues of a noncentral Wishart is known in the sense that there is an infinite series expansion for the pdf in terms of zonal polynomials (Muirhead [1978]). This joint pdf depends on all the eigenvalues of $\boldsymbol{\Lambda}$, as well as n and K_2. In principle the pdf for the minimum eigenvalue can be determined from this joint pdf for all the eigenvalues. It appears that this pdf (the "exact asymptotic" pdf of $g_{\min}$) depends on all the eigenvalues of $\boldsymbol{\Lambda}$.

This exact asymptotic distribution of $g_{\min}$ is not very useful for applications both because of the computational difficulties it poses and because of its dependence on all the eigenvalues of $\boldsymbol{\Lambda}$. This latter consideration is especially important because in practice these eigenvalues are unknown nuisance parameters, so critical values that depend on multiple eigenvalues would produce an infeasible test.

We circumvent these two problems by proposing conservative critical values based on the following bounding distribution.

***Proposition* 1**. Pr[mineval($W_n(k, \boldsymbol{I}_n, \boldsymbol{A})$) ≥ x] ≤ Pr[χ_k^2(mineval($\boldsymbol{A}$)) ≥ x], where $\chi_k^2(a)$ denotes a noncentral chi-squared random variable with noncentrality parameter a.

Proof. Let α be the eigenvector of $\boldsymbol{A}$ corresponding to its minimum eigenvalue. Then $\alpha'\boldsymbol{W}\alpha$ is distributed χ_k^2(mineval($\boldsymbol{A}$)) (Muirhead [1982, Theorem 10.3.6]). But $\alpha'\boldsymbol{W}\alpha \geq$ mineval($\boldsymbol{W}$), and the result follows.

Applying (4.1), the continuous mapping theorem, and Proposition 1, we have that

$$\Pr[g_{\min} \geq x] \rightarrow \Pr[\text{mineval}(\nu_1/K_2) \geq x] \leq \Pr\left[\frac{\chi_{K_2}^2(\text{mineval}(K_2\Lambda))}{K_2} \geq x\right]. \qquad (4.2)$$

Note that this inequality holds as an equality in the special case $n = 1$.

Conservative critical values for the test based on $g_{\min}$ are obtained as follows. First, select the desired minimal eigenvalue of Λ. Next, obtain the desired percentile, say the 95% point, of the noncentral chi-squared distribution with noncentrality parameter equal to K_2 times this selected minimum eigenvalue, and divide this percentile by K_2.[6]

4.2. The Weak Instruments Test

The bound (4.2) yields the following testing procedure to detect weak instruments. To be concrete, this is stated for a test based on the TSLS bias measure with significance level 100δ%. The null hypothesis is that the instruments are weak, and the alternative is that they are not:

$$H_0: \mathcal{Z} \in \mathcal{W}_{bias,\ TSLS} \quad \text{vs.} \quad H_1: \mathcal{Z} \notin \mathcal{W}_{bias,\ TSLS}. \tag{4.3}$$

The test procedure is

$$\text{Reject } H_0 \text{ if } g_{\min} \geq d_{bias,\ TSLS}(b;\ K_2,n,\delta), \tag{4.4}$$

where $d_{bias,\ TSLS}(b;\ K_2,n,\delta) = K_2^{-1}\ \chi^2_{K_2,1-\delta}\ (K_2\,\ell_{\ bias,\ TSLS}\ (b;K_2,n))$, where $\chi^2_{K_2,1-\delta}(m)$ is the 100(1-δ)% percentile of the noncentral chi-squared distribution with K_2 degrees of freedom and noncentrality parameter m and the function $\ell_{\ bias,\ TSLS}$ is the weak instrument boundary minimum eigenvalue of $\boldsymbol{\Lambda}$ in (3.11).

The results of Section 3 and the bound resulting from Proposition 1 imply that, asymptotically, the test (4.4) has the desired asymptotic level:

$$\lim_{T\to\infty} \Pr[g_{\min} \geq d_{bias,\ TSLS}(b;\ K_2,n,\delta) \mid \mathcal{Z} \in \mathcal{W}_{bias,\ TSLS}\ ] \leq \delta. \tag{4.5}$$

The procedure for testing whether the instruments are weak from the perspective of the size of the TSLS (or LIML) is the same, except that the critical value in (4.4) is obtained using the size-based boundary eigenvalue function, $\ell_{size,\,TSLS}(r;K_2,n,\alpha)$ (or, for LIML, $\ell_{size,\,LIML}(r;K_2,n,\alpha)$).

4.3. Critical Values

Given a minimum eigenvalue ℓ, conservative critical values for the test are percentiles of the scaled noncentral chi-squared distribution, $\chi^2_{K_2,1-\delta}(K_2\ell)/K_2$. The minimum eigenvalue ℓ is obtained from the boundary eigenvalue functions of Section 3.5.

Critical values are tabulated in Tables 1 – 4 for the weak instrument tests based on TSLS bias, TSLS size, Fuller-*k* bias and LIML size, respectively, for 1 or 2 included endogenous variables (and 3 for TSLS bias) and up to 30 instruments. These critical values are plotted in the panel below the corresponding boundaries of the weak instrument sets in Figures 1 – 4. The critical value plots are qualitatively similar to the corresponding boundary eigenvalue plots, except of course the critical values exceed the boundary eigenvalues to take into account the sampling distribution of the test statistic.

These critical value plots provide a basis for comparing the robustness to weak instruments of various procedures: the lower the critical value curve, the more robust is the procedure. For discussion and comparisons of TSLS, BTSLS, Fuller-*k*, JIVE, and LIML, see Stock, Wright, and Yogo (2002, Section 6).

***Comparison to the Staiger-Stock* (1997) *rule of thumb*.** Staiger and Stock (1997) suggested the rule of thumb that, in the $n = 1$ case, instruments be deemed weak if the first-stage F is less than ten. They motivated this suggestion based on the relative bias of

TSLS. Because the 5% critical value for the relative bias weak instrument test with b = .1 is approximately 11 for all values of K_2, the Staiger-Stock rule of thumb is approximately a 5% test that the worst case relative bias is approximately 10% or less. This provides a formal, and not unreasonable, testing interpretation of the Staiger-Stock rule of thumb.

The rule of thumb fares less well from the perspective of size distortion. When the number of instruments is one or two, the Staiger-Stock rule of thumb corresponds to a 5% level test that the maximum size is no more than 15% (so the maximum TSLS size distortion is no more than 10%). However, when the number of instruments is moderate or large, the critical value is much larger and this rule of thumb does not provide substantial assurance that the size distortion is controlled.

5. Asymptotic Properties of the Test as a Decision Rule

This section examines the asymptotic rejection rate of the weak instrument test as a function of the smallest eigenvalue of $\boldsymbol{\Lambda}$. When this eigenvalue exceeds the boundary minimum eigenvalue for the weak instrument set, the asymptotic rejection rate is the asymptotic power function.

The exact asymptotic distribution of $g_{\min}$ depends on all the eigenvalues of $\boldsymbol{\Lambda}$. It is bounded above by (4.2). Based on numerical analysis, we conjecture that this distribution is bounded below by the distribution of the minimum eigenvalue of a random matrix with the noncentral Wishart distribution $W_n(K_2, \boldsymbol{I}_n, \text{mineval}(K_2\boldsymbol{\Lambda})\boldsymbol{I}_n)/K_2$. These two bounding distributions are used to bound the distribution of g_{min} as a function of mineval($\boldsymbol{\Lambda}$).

The bounds on the asymptotic rejection rate of the test (4.4) (based on TSLS maximum relative bias) are plotted in Figure 5 for b = .1 and n = 2. The value of the horizontal axis (the minimum eigenvalue) at which the upper rejection rate curve equals 5% is $\ell_{bias}(.1;K_2,2)$. Evidently, as the minimum eigenvalue increases, so does the rejection rate. If K_2 is moderate or large, this increase is rapid and the test essentially has unit power against values of the minimum eigenvalue not much larger than the critical value. The bounding distributions give a reasonably tight range for the actual power function, which depends on all the eigenvalues of $\boldsymbol{\Lambda}$.

The analogous curves for the test based on Fuller-k bias, TSLS size, or LIML size are centered differently because the tests have different critical values but otherwise are qualitatively similar to those in Figure 5 and thus are omitted.

Interpretation as a decision rule. It is useful to think of the weak instrument test as a decision rule: if g_{min} is less than the critical value, conclude that the instruments are weak, otherwise conclude that they are strong.

Under this interpretation, the asymptotic rejection rates in Figure 5 bound the asymptotic probability of deciding that the instruments are strong. Evidently, for values of mineval($\boldsymbol{\Lambda}$) much below the weak instrument region boundary, the probability of correctly concluding that the instruments are weak is effectively one. Thus, if in fact the researcher is confronted by instruments that are quite weak, this will be detected by the weak instruments test with probability essentially one. Similarly, if the researcher has instruments with a minimum eigenvalue of $\boldsymbol{\Lambda}$ substantially above the threshold for the weak instruments set, then the probability of correctly concluding that they are strong also is essentially one.

The range of ambiguity of the decision procedure is given by the values of the minimum eigenvalue for which the asymptotic rejection rates effectively fall between zero and one. When K_2 is small, this range can be ten or more, but for K_2 large this range of potential ambiguity of the decision rule is quite narrow.

6. Conclusions

The procedure proposed here is simple: compare the minimum eigenvalue of $\boldsymbol{G}_T$, the first-stage F-statistic matrix, to a critical value. The critical value is determined by the IV estimator the researcher is using, the number of instruments K_2, the number of included endogenous regressors n, and how much relative bias or size distortion the researcher will tolerate. The test statistic is the same whether one focuses on the bias of TSLS or Fuller-k or on the size of TSLS or LIML; all that differs is the critical value.

Viewed as a test, the procedure has good power, especially when the number of instruments is large. Viewed as a decision rule, the procedure effectively discriminates between weak and strong instruments, and the region of ambiguity decreases as the number of instruments increases.

Our findings support the view that LIML is far superior to TSLS when the researcher has weak instruments, at least from the perspective of coverage rates. Actual LIML coverage rates are close to their nominal rates even for quite small values of the minimum eigenvalue, especially for moderately many instruments. Similarly, the Fuller-k estimator is more robust to weak instruments than TSLS, when viewed from the perspective of bias. Additional comparisons across estimators based on these methods are discussed in Stock, Wright, and Yogo (2002).

When there is a single included endogenous variable, this procedure provides a refinement and improvement to Staiger and Stock's (1997) rule of thumb that, in the n = 1 case, instruments be deemed "weak" if the first-stage F is less than ten. The difference between that rule of thumb and the procedure of this paper is that, instead of comparing the first-stage F to ten, it should be compared to the appropriate entry in Table 1 (TSLS bias), Table 2 (TSLS size), Table 3 (Fuller-k bias), or Table 4 (LIML size). Those critical values indicate that their rule of thumb can be interpreted as a test with approximately a 5% significance level, of the hypothesis that the maximum relative bias is at least 10%. The Staiger-Stock rule of thumb is too conservative if LIML or Fuller-k are used unless the number of instruments is very small, but it is insufficiently conservative to ensure that the TSLS Wald test has good size.

This paper has two loose ends. First, the characterization of the set of weak instruments is based on the premise that the maximum relative bias and maximum size distortion are nonincreasing in each eigenvalue of $\boldsymbol{\Lambda}$, for values of those eigenvalues in the relevant range. This was justified formally using the many weak instrument asymptotics in the appendix; although numerical analysis suggests it is true for all K_2, this remains to be proven. Second, the lower bound of the power function in Section 5 is based on the assumption that the cdf of the minimum eigenvalue of a noncentral Wishart random variable is nondecreasing in each of the eigenvalues of its noncentrality matrix. This too appears to be true based on numerical analysis but we do not have a proof nor does this result seem to be available in the literature.

Beyond this, several avenues of research remain open. First, the tests proposed here are conservative when $n > 1$ because they use critical values computed using the noncentral chi-squared bound in Proposition 1. Although the tests appear to have good power despite this, tightening the Proposition 1 bound (or constructing tests based on all the eigenvalues) could produce more powerful tests. Second, we have considered inference based on TSLS, Fuller-*k*, and LIML, but there are other estimators to explore as well. Third, the analysis here is predicated upon homoskedasticity, and it remains to extend these tests to GMM estimation of the linear IV regression model under heteroskedasticity.

Appendix

This appendix extends the fixed-K_2 weak instrument asymptotics of Staiger and Stock (1997) to the case of many weak instruments. Specifically, we suppose that the number of instruments can increase with the sample size but that $K_2^2/T \to 0$. The instruments are modeled as being weak, in the sense that the weak instrument assumption L_{Π} of Section 2.4 is assumed to hold, which in turn implies that the scaled concentration matrix Λ is finite as $T \to \infty$. Under these assumptions, plus some additional technical assumptions stated below (including i.i.d. sampling), it is shown that the limits of *k*-class IV statistics as K_2 and T jointly tend to infinity can in general be computed using sequential asymptotic limits. Under sequential asymptotics, the fixed-K_2 weak instrument limit is first obtained, then the limit of that distribution is taken as $K_2 \to \infty$. The advantage of this "first T then K_2" approach is that the sequential calculations are simpler than the calculations that arise along the joint sequence of (K_2, T).

We begin in Section A.1 by specifying the assumptions. Section A.2 justifies the sequential asymptotics by showing that, under these assumptions, a key uniform convergence result (Lemma 6 in Phillips and Moon (1999)) holds. In Section A.3, we derive the many weak instrument limits of *k*-class estimators and test statistics using sequential asymptotics. Section A.4 provides some concluding remarks.

A.1. The Model, Notation and Assumptions

To simplify the expressions, we consider the IV regression model with no included exogenous variables, that is, (2.1) and (2.2) without the $\boldsymbol{X}$ variables. Because this appendix is concerned with sequences of K_2, it is useful to indicate dependence of $\boldsymbol{\Lambda}$ on K_2. Accordingly, let $\boldsymbol{\Lambda}_{K_2} = \boldsymbol{\Sigma}_{\boldsymbol{VV}}^{-1/2}\boldsymbol{C}'\boldsymbol{Q}_{\boldsymbol{ZZ}}\boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{VV}}^{-1/2}{}'/K_2$ be the matrix $\boldsymbol{\Lambda}$ in the text, explicitly indexed by K_2; $\boldsymbol{\Lambda}_{K_2}$ is the expected value of the concentration matrix (divided by K_2) when there are K_2 instruments.

Throughout this appendix, it is assumed that the random variables are i.i.d. with four moments, the instruments are not multicollinear, and the errors are homoskedastic. Specifically, we assume,

***Assumption A*1**.

(a) There exists a constant $D_1 > 0$ such that mineval$(\boldsymbol{Z}'\boldsymbol{Z}/T) \geq D_1$ a.s. for all K_2 and for all T greater than some T_0.

(b) $\boldsymbol{Z}_t$ is i.i.d. with $E\boldsymbol{Z}_t\boldsymbol{Z}_t' = \boldsymbol{Q}_{\boldsymbol{ZZ}}$, where $\boldsymbol{Q}_{\boldsymbol{ZZ}}$ is positive definite, and $EZ_{it}^4 \leq D_2 < \infty$, $i = 1,\ldots, K_2$.

(c) $\boldsymbol{\eta}_t = [u_t \;\; \boldsymbol{V}_t']'$ is i.i.d. with $E(\boldsymbol{\eta}_t \,|\, \boldsymbol{Z}_t) = 0$, $E(\boldsymbol{\eta}_t\boldsymbol{\eta}_t' \,|\, \boldsymbol{Z}_t) = \boldsymbol{\Sigma}$, and $E(|\eta_{it}\eta_{jt}\eta_{kt}\eta_{lt}| \,|\, \boldsymbol{Z}_t) = E(|\eta_{it}\eta_{jt}\eta_{kt}\eta_{lt}|) \leq D_2 < \infty$, $i, j, k, l = 1,\ldots, n + 1$.

Our analysis focuses on sequences of K_2 that, if they increase, do so slower than $\sqrt{T}$. Specifically, we assume,

***Assumption A*2**. $K_2^2/T \to 0$ as $T \to \infty$.

Note that Assumption $A2$ does not require K_2 to increase, but it limits the rate at which it can increase.

A.2. Uniform Convergence Result

This section provides the uniform convergence result (Theorem A.1) that justifies the use of sequential asymptotics to compute the many weak instrument limiting representations. We adopt Phillips and Moon's (1999) notation in which $(T, K_2 \to \infty)_{\text{seq}}$ denotes the sequential limit in which first $T \to \infty$, then $K_2 \to \infty$; the notation $(K_2, T \to \infty)$ denotes the joint limit in which K_2 is implicitly indexed by T.

Lemma 6 of Phillips and Moon (1999) provides general conditions under which sequential convergence implies joint convergence.

***Phillips and Moon* (1999) *Lemma 6*.**

(a) Suppose there exist random vectors X_K and X on the same probability space as $X_{K,T}$ satisfying, for all K, $X_{K,T} \xrightarrow{p} X_K$ as $T \to \infty$ and $X_K \xrightarrow{p} X$ as $K \to \infty$. Then, $X_{K,T} \xrightarrow{p} X$ as $(K, T \to \infty)$ if and only if,

$$\limsup_{K,\,T} \Pr[\ \|X_{K,T} - X_K\| > \varepsilon] = 0 \text{ for all } \varepsilon > 0. \tag{A.1}$$

(b) Suppose there exist random vectors X_K such that, for any fixed K, $X_{K,T} \xrightarrow{d} X_K$ as $T \to \infty$ and $X_K \xrightarrow{d} X$ as $K \to \infty$. Then $X_{K,T} \xrightarrow{d} X$ as $(K, T \to \infty)$ if and only if, for all bounded continuous functions f,

$$\lim \sup_{K,T} \mid E[f(X_{K,T})] - E[f(X_K)] \mid = 0. \qquad \text{(A.2)}$$

Note that condition (A.2) is equivalent to the requirement that

$$\lim \sup_{K,T} \sup_x \mid F_{X_{K,T}}(x) - F_{X_K}(x) \mid = 0, \qquad \text{(A.3)}$$

where $F_{X_{K,T}}$ is the cdf of $X_{K,T}$ and F_{X_K} is the cdf of X_K.

The rest of this section is devoted to showing that the conditions of this lemma, that is, (A.1) and (A.3), hold under assumptions L_Π, A1, and A2 for the statistics that enter the *k*-class estimators and tests. To do so, we use the following Berry-Esseen bound proven by Götze (1991, equation (1.5)):

***Berry-Esseen Bound* (*Götze* (1991))**. Let $\{X_1,\ldots, X_T\}$ be an i.i.d. sequence in $\mathbb{R}^K$ with zero means, a nonsingular second moment matrix, and finite absolute third moments. Let P_T be the probability measure associated with $T^{-1/2}\sum_{t=1}^{T} X_t$, and let P be the limiting Gaussian measure. Then for each T,

$$\sup_{A\in C^K} |P_T(A) - P(A)| \le const \times (K/T)^{1/2}, \qquad \text{(A.4)}$$

where C^K is the class of all measurable convex sets in $\mathbb{R}^K$.

We now turn to k-class statistics. First note that, for fixed K_2, under Assumptions $L_{\boldsymbol{\Pi}}$ and A1, the weak law of large numbers and the central limit theorem imply that the following limits hold jointly for fixed K_2:

$$(T^{-1}\boldsymbol{u}'\boldsymbol{u},\ T^{-1}\boldsymbol{V}'\boldsymbol{u},\ T^{-1}\boldsymbol{V}'\boldsymbol{V}) \xrightarrow{p} (\sigma_{uu},\ \boldsymbol{\Sigma}_{Vu},\ \boldsymbol{\Sigma}_{VV}), \tag{A.5}$$

$$\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi} \xrightarrow{p} \boldsymbol{C}'\boldsymbol{Q}_{ZZ}\boldsymbol{C}, \tag{A.6}$$

$$(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{u},\ \boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{V}) \xrightarrow{d} (\boldsymbol{C}'\boldsymbol{\Psi}_{Zu},\ \boldsymbol{C}'\boldsymbol{\Psi}_{ZV}), \tag{A.7}$$

$$(\boldsymbol{u}'\boldsymbol{P}_Z\boldsymbol{u},\ \boldsymbol{V}'\boldsymbol{P}_Z\boldsymbol{u},\ \boldsymbol{V}'\boldsymbol{P}_Z\boldsymbol{V}) \xrightarrow{d} (\boldsymbol{\Psi}_{Zu}'\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{\Psi}_{Zu},\ \boldsymbol{\Psi}_{ZV}'\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{\Psi}_{Zu},\ \boldsymbol{\Psi}_{ZV}'\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{\Psi}_{ZV}), \tag{A.8}$$

where $\boldsymbol{\Psi} \equiv [\boldsymbol{\Psi}_{Zu}',\ \text{vec}(\boldsymbol{\Psi}_{ZV})']'$ (defined in Assumption M) is distributed $N(0, \boldsymbol{\Sigma} \otimes \boldsymbol{Q}_{ZZ})$.

The following theorem shows that, in fact, the limits in (A.5) – (A.8) and related limits hold uniformly in K_2 under the weak instrument assumption (Assumption $L_{\boldsymbol{\Pi}}$), the sampling assumption (Assumption A1), and the rate condition (Assumption A2). As in (A.3), let F_X denote the cdf of the random variable X (etc.).

Theorem A.1. Under Assumptions $L_{\boldsymbol{\Pi}}$, A1, and A2,

(a) $\limsup_{K_2,T} \Pr[\,|\,(\boldsymbol{u}'\boldsymbol{u}/T,\ \boldsymbol{V}'\boldsymbol{u}/T,\ \boldsymbol{V}'\boldsymbol{V}/T) - (\sigma_{uu},\ \boldsymbol{\Sigma}_{Vu},\ \boldsymbol{\Sigma}_{VV})\,| > \varepsilon] = 0 \ \forall\ \varepsilon > 0$,

(b) $\limsup_{K_2,T} \Pr[\,|\,\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi}/K_2 - \boldsymbol{C}'\boldsymbol{Q}_{ZZ}\boldsymbol{C}/K_2\,| > \varepsilon] = 0 \ \forall\ \varepsilon > 0$,

(c) $\limsup_{K_2,T} \sup_{\boldsymbol{x}} |\,F_{\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{u}}(\boldsymbol{x}) - F_{\boldsymbol{C}'\boldsymbol{\Psi}_{Zu}}(\boldsymbol{x})\,| = 0$,

(d) $\limsup_{K_2,T} \sup_{\boldsymbol{x}} |\,F_{\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{V}}(\boldsymbol{x}) - F_{\boldsymbol{C}'\boldsymbol{\Psi}_{ZV}}(x)\,| = 0$,

(e) $\limsup_{K_2,T} \sup_x | F_{\boldsymbol{u}'P_Z\boldsymbol{u}}(x) - F_{\boldsymbol{\Psi}_{Zu}'\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{\Psi}_{Zu}}(x) | = 0$,

(f) $\limsup_{K_2,T} \sup_{\boldsymbol{x}} | F_{\boldsymbol{V}'P_Z\boldsymbol{u}}(\boldsymbol{x}) - F_{\boldsymbol{\Psi}_{ZV}'\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{\Psi}_{Zu}}(\boldsymbol{x}) | = 0$,

(g) $\limsup_{K_2,T} \sup_{\boldsymbol{x}} | F_{\boldsymbol{V}'P_Z\boldsymbol{V}}(\boldsymbol{x}) - F_{\boldsymbol{\Psi}_{ZV}'\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{\Psi}_{ZV}}(\boldsymbol{x}) | = 0$.

Theorem A.1 verifies the conditions (A.1) and (A.3) of Phillips and Moon's (1999) Lemma 6 for statistics that enter the *k*-class estimator and Wald statistic. Some of these objects converge in probability uniformly under the stated assumptions (parts (a) and (b)), while others converge in distribution uniformly (parts (c) – (g)). It follows from the continuous mapping theorem that continuous functions of these objects also converge in probability (and/or distribution) uniformly under the stated assumptions. Because the *k*-class estimator $\hat{\boldsymbol{\beta}}(k)$ and Wald statistic $W(k)$ are continuous functions of these statistics (after centering and scaling as needed), it follows that the $(K_2, T \to \infty)$ joint limit of these *k*-class statistics can be computed as the sequential limit, $(T, K_2 \to \infty)_{seq}$.

The proof of Theorem A.1 uses the following lemma.

Lemma A.2. Let $\Delta_T = (\boldsymbol{Z}'\boldsymbol{Z}/T)^{-1} - \boldsymbol{Q}_{ZZ}^{-1}$. Under Assumptions A1 and A2,

(a) $\limsup_{K_2,T} \Pr[| T^{-1}\boldsymbol{u}'\boldsymbol{Z}\Delta_T \boldsymbol{Z}'\boldsymbol{u} | > \varepsilon] = 0 \ \forall \ \varepsilon > 0$,

(b) $\limsup_{K_2,T} \Pr[| T^{-1}\boldsymbol{V}'\boldsymbol{Z}\Delta_T \boldsymbol{Z}'\boldsymbol{u} | > \varepsilon] = 0 \ \forall \ \varepsilon > 0$,

(c) $\limsup_{K_2,T} \Pr[| T^{-1}\boldsymbol{V}'\boldsymbol{Z}\Delta_T \boldsymbol{Z}'\boldsymbol{V} | > \varepsilon] = 0 \ \forall \ \varepsilon > 0$.

Proof of Lemma A.2.

(a) The strategy is to show that $T^{-1}\boldsymbol{u}'\mathbf{Z}\Delta_T\,\mathbf{Z}'\boldsymbol{u}$ has expected mean square that is bounded by $const \times (K_2^2/T)$, then to apply Chebychev's inequality. The expected square of $T^{-1}\boldsymbol{u}'\mathbf{Z}\Delta_T\,\mathbf{Z}'\boldsymbol{u}$ is

$$E(T^{-1}\boldsymbol{u}'\mathbf{Z}\Delta_T\,\mathbf{Z}'\boldsymbol{u})^2 = \frac{1}{T^2}E\sum_{q=1}^{T}\sum_{r=1}^{T}\sum_{s=1}^{T}\sum_{t=1}^{T}u_q u_r u_s u_t \mathbf{Z}_q{}'\Delta_T\mathbf{Z}_r\mathbf{Z}_s{}'\Delta_T\mathbf{Z}_t$$

$$= \frac{1}{T^2}E\sum_{t=1}^{T}u_t^4(\mathbf{Z}_t{}'\Delta_T\mathbf{Z}_t)^2 + \frac{1}{T^2}E\sum_{s=1}^{T}\sum_{t=1,t\neq s}^{T}u_s^2u_t^2(\mathbf{Z}_t{}'\Delta_T\mathbf{Z}_t)(\mathbf{Z}_s{}'\Delta_T\mathbf{Z}_s)$$

$$+\frac{2}{T^2}E\sum_{s=1}^{T}\sum_{t=1,t\neq s}^{T}u_s^2u_t^2(\mathbf{Z}_t{}'\Delta_T\mathbf{Z}_s)^2\,. \qquad \text{(A.9)}$$

The first term on the right hand side of (A.9) is

$$\frac{1}{T^2}E\sum_{t=1}^{T}u_t^4(\mathbf{Z}_t{}'\Delta_T\mathbf{Z}_t)^2 = \frac{1}{T^2}(Eu_t^4)\sum_{t=1}^{T}E\{\mathbf{Z}_t{}'[(\mathbf{Z}'\mathbf{Z}/T)-\mathbf{Q}_{\mathbf{ZZ}}^{-1}]\mathbf{Z}_t\}^2$$

$$= \frac{1}{T}(Eu_t^4)E[\mathbf{Z}_t{}'(\mathbf{Z}'\mathbf{Z}/T)\mathbf{Z}_t - \mathbf{Z}_t{}'\mathbf{Q}_{\mathbf{ZZ}}^{-1}\mathbf{Z}_t]^2$$

$$= \frac{1}{T}(Eu_t^4)\{E[\mathbf{Z}_t{}'(\mathbf{Z}'\mathbf{Z}/T)^{-1}\mathbf{Z}_t]^2$$

$$-2E[(\mathbf{Z}_t{}'(\mathbf{Z}'\mathbf{Z}/T)^{-1}\mathbf{Z}_t)(\mathbf{Z}_t{}'\mathbf{Q}_{\mathbf{ZZ}}^{-1}\mathbf{Z}_t)]+E[\mathbf{Z}_t{}'\mathbf{Q}_{\mathbf{ZZ}}^{-1}\mathbf{Z}_t]^2\}\,, \qquad \text{(A.10)}$$

where the second line follows because $\mathbf{Z}_t$ is i.i.d. Now $E[\mathbf{Z}_t'(\mathbf{Z}'\mathbf{Z}/T)^{-1}\mathbf{Z}_t]^2 \leq E[\mathbf{Z}_t'\mathbf{Z}_t \text{maxeval}[(\mathbf{Z}'\mathbf{Z}/T)^{-1}]]^2 = E[\mathbf{Z}_t'\mathbf{Z}_t/\text{mineval}(\mathbf{Z}'\mathbf{Z}/T)]^2 \leq D_1^{-2}\,E(\mathbf{Z}_t'\mathbf{Z}_t)^2 = const \times K_2^2$, where the second inequality follows from Assumption A.1. Similar calculations show that $E[\boldsymbol{Z}_t'(\boldsymbol{Z}'\boldsymbol{Z}/T)^{-1}\boldsymbol{Z}_t(\boldsymbol{Z}_t'\boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1}\boldsymbol{Z}_t)] \leq const \times K_2^2$ and $E[\boldsymbol{Z}_t'\boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1}\boldsymbol{Z}_t]^2 \leq const \times K_2^2$, the second two terms after the final equality in (A.10) are $\leq const \times K_2^2$. It follows from (A.10) that

$$T^{-2}E\sum_{t=1}^{T}u_t^4(\boldsymbol{Z}_t'\varDelta_T\boldsymbol{Z}_t)^2 \leq const \times K_2^2/T. \tag{A.11}$$

The second term on the right hand side of (A.9) is

$$\begin{aligned}
\frac{1}{T^2}E\sum_{s=1}^{T}\sum_{t=1,t\neq s}^{T}u_s^2u_t^2(\boldsymbol{Z}_t'\varDelta_T\boldsymbol{Z}_t)(\boldsymbol{Z}_s'\varDelta_T\boldsymbol{Z}_s) &= \frac{\sigma_{uu}^2}{T^2}E\sum_{s=1}^{T}\sum_{t=1,t\neq s}^{T}(\boldsymbol{Z}_t'\varDelta_T\boldsymbol{Z}_t)(\boldsymbol{Z}_s'\varDelta_T\boldsymbol{Z}_s) \\
&\leq \frac{\sigma_{uu}^2}{T^2}E\sum_{s=1}^{T}\sum_{t=1}^{T}(\boldsymbol{Z}_t'\varDelta_T\boldsymbol{Z}_t)(\boldsymbol{Z}_s'\varDelta_T\boldsymbol{Z}_s) \\
&= \sigma_{uu}^2E\left[\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{Z}_t'\varDelta_T\boldsymbol{Z}_t)\right]^2 = \sigma_{uu}^2E\left\{tr\left[\varDelta_T\left(\frac{\boldsymbol{Z}'\boldsymbol{Z}}{T}\right)\right]\right\}^2 \\
&= \sigma_{uu}^2E\left\{tr\left[\boldsymbol{I}-\boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1}\left(\frac{\boldsymbol{Z}'\boldsymbol{Z}}{T}\right)\right]\right\}^2 = \sigma_{uu}^2\left(E\left\{tr\left[\boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1}\left(\frac{\boldsymbol{Z}'\boldsymbol{Z}}{T}\right)\right]\right\}^2 - K_2^2\right) \\
&= \sigma_{uu}^2\left[E\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Z}_t'\boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1}\boldsymbol{Z}_t\right)^2 - K_2^2\right] \\
&= \sigma_{uu}^2\left[\frac{1}{T^2}\sum_{t=1}^{T}E(\boldsymbol{Z}_t'\boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1}\boldsymbol{Z}_t)^2 + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}E(\boldsymbol{Z}_t'\boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1}\boldsymbol{Z}_t)(\boldsymbol{Z}_s'\boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1}\boldsymbol{Z}_s) - K_2^2\right] \\
&= \sigma_{uu}^2\left[const\times\frac{TK_2^2}{T^2} + \frac{T(T-1)K_2^2}{T^2} - K_2^2\right] = const\times\sigma_{uu}^2\frac{K_2^2}{T},
\end{aligned} \tag{A.12}$$

where the penultimate equality follows because $E(\boldsymbol{Z}_t' \boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1} \boldsymbol{Z}_t) = K_2$ and $E(\boldsymbol{Z}_t' \boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1} \boldsymbol{Z}_t)^2 =$ $const \times K_2^2$.

The third term on the right hand side of (A.9) is

$$\frac{2}{T^2} E \sum_{s=1}^{T} \sum_{t=1, t \neq s}^{T} u_s^2 u_t^2 (\boldsymbol{Z}_t ' \boldsymbol{\Delta}_T \boldsymbol{Z}_s)^2 = 2\sigma_{uu}^2 E \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t=1, t \neq s}^{T} (\boldsymbol{Z}_t ' \boldsymbol{\Delta}_T \boldsymbol{Z}_s)^2$$

$$\leq 2\sigma_{uu}^2 E \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t=1}^{T} (\boldsymbol{Z}_t ' \boldsymbol{\Delta}_T \boldsymbol{Z}_s)^2$$

$$= 2\sigma_{uu}^2 E \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t=1}^{T} tr(\boldsymbol{\Delta}_T \boldsymbol{Z}_s \boldsymbol{Z}_s ' \boldsymbol{\Delta}_T \boldsymbol{Z}_t \boldsymbol{Z}_t ')$$

$$= 2\sigma_{uu}^2 Etr\left[\boldsymbol{\Delta}_T \left(\frac{\boldsymbol{Z}'\boldsymbol{Z}}{T}\right) \boldsymbol{\Delta}_T \left(\frac{\boldsymbol{Z}'\boldsymbol{Z}}{T}\right)\right] = 2\sigma_{uu}^2 E\left\{tr\left[\boldsymbol{I} - \boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1}\left(\frac{\boldsymbol{Z}'\boldsymbol{Z}}{T}\right)\right]^2\right\}$$

$$= 2\sigma_{uu}^2 Etr\left\{\boldsymbol{I} - 2\boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1}\left(\frac{\boldsymbol{Z}'\boldsymbol{Z}}{T}\right) + \left[\boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1}\left(\frac{\boldsymbol{Z}'\boldsymbol{Z}}{T}\right)\right]^2\right\}$$

$$= 2\sigma_{uu}^2 \left[E\left(\frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \boldsymbol{Z}_t ' \boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1} \boldsymbol{Z}_s \boldsymbol{Z}_s ' \boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1} \boldsymbol{Z}_t\right) - K_2\right]$$

$$= 2\sigma_{uu}^2 \left[\frac{1}{T^2} \sum_{t=1}^{T} E(\boldsymbol{Z}_t ' \boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1} \boldsymbol{Z}_t)^2 + \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1, s \neq t}^{T} Etr(\boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1} \boldsymbol{Z}_s \boldsymbol{Z}_s ' \boldsymbol{Q}_{\boldsymbol{ZZ}}^{-1} \boldsymbol{Z}_t \boldsymbol{Z}_t ') - K_2\right]$$

$$= 2\sigma_{uu}^2 \left[const \times \frac{TK_2^2}{T^2} + \frac{T(T-1)K_2}{T^2} - K_2\right]$$

$$= 2\sigma_{uu}^2 \left(const \times \frac{K_2^2}{T} - \frac{K_2}{T}\right), \qquad \text{(A.13)}$$

where the penultimate equality uses $Etr(\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{Z}_s\boldsymbol{Z}_s'\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{Z}_t\boldsymbol{Z}_t') = \mathrm{tr}(\boldsymbol{I}) = K_2$ for $s \neq t$ and $E(\boldsymbol{Z}_t'\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{Z}_t)^2 = const \times K_2^2$.

The application of (A.11), (A.12), and (A.13) to (A.9) implies that $E(T^{-1}\boldsymbol{u}'\boldsymbol{Z}\boldsymbol{\Delta}_T\boldsymbol{Z}'\boldsymbol{u})^2 \leq const \times \sigma_{uu}^2 K_2^2/T$, and the desired result (a) follows by Chebychev's inequality.

The proofs of (b) and (c) are analogous to the proof of (a).

Proof of Theorem A.1.

(a) This follows from the weak law of large numbers because ($\boldsymbol{u}'\boldsymbol{u}/T$, $\boldsymbol{V}'\boldsymbol{u}/T$, $\boldsymbol{V}'\boldsymbol{V}/T$) do not depend on K_2.

(b) Note that $E[\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi}/K_2 - \boldsymbol{C}'\boldsymbol{Q}_{ZZ}\boldsymbol{C}/K_2] = 0$. Under Assumption $L_{\boldsymbol{\Pi}}$, the (1, 1) element of this matrix is $(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi} - \boldsymbol{C}'\boldsymbol{Q}_{ZZ}\boldsymbol{C})_{1,1}/K_2 = (TK_2)^{-1}\sum_{i=1}^{K_2}\sum_{j=1}^{K_2} C_{i1}C_{j1}(Z_{it}Z_{kt} - q_{ij})$, where q_{ij} is the (i, j) element of $\boldsymbol{Q}_{ZZ}$. Because $\boldsymbol{Z}_t$ is i.i.d. (Assumption A1(b)) and the elements of $\boldsymbol{C}$ are bounded (Assumption $L_{\boldsymbol{\Pi}}$), the expected value of the square of this element is

$$E\{[(\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi} - \boldsymbol{C}'\boldsymbol{Q}_{ZZ}\boldsymbol{C})_{1,1}/K_2]^2\} = \frac{1}{T}E\left[\frac{1}{K_2}\sum_{i=1}^{K_2}\sum_{j=1}^{K_2} C_{i1}C_{j1}(Z_{it}Z_{kt} - q_{ij})\right]^2$$

$$= \frac{1}{T}\frac{1}{K_2^2}\sum_{i=1}^{K_2}\sum_{j=1}^{K_2}\sum_{k=1}^{K_2}\sum_{l=1}^{K_2} C_{i1}C_{j1}C_{k1}C_{l1}E[(Z_{it}Z_{jt} - q_{ij})(Z_{kt}Z_{lt} - q_{kl})] \leq const \times \frac{K_2^2}{T}.$$

By the same argument applied to the (1,1) element, the remaining elements of $\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi}/K_2 - \boldsymbol{C}'\boldsymbol{Q}_{ZZ}\boldsymbol{C}/K_2$ are also bounded in mean square by $const \times (K_2^2/T)$. The matrix $\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi}/K_2$ is $n \times n$ so the number of elements does not depend on K_2 and the result (b) follows by Chebychev's inequality and noting that, under Assumption A2, $K_2^2/T \to 0$.

(c) Under Assumption $L_{\boldsymbol{\Pi}}$, $\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{u} = T^{-1/2}\boldsymbol{C}'\boldsymbol{Z}'\boldsymbol{u} = \boldsymbol{C}'(T^{-1/2}\sum_{t=1}^{T}\boldsymbol{Z}_t u_t)$. Let P_T denote the probability measure associated with $T^{-1/2}\boldsymbol{Z}'\boldsymbol{u}$ and let P denote the limiting probability measure associated with $\boldsymbol{\Psi}_{Zu}$. Define the convex set $A(\boldsymbol{x}) = \{y \in \mathbb{R}^{K_2} : \boldsymbol{C}'\boldsymbol{y} \le \boldsymbol{x}\}$, so that $P_T(A(x)) = F_{\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{u}}(\boldsymbol{x})$ and $P(A(x)) = F_{\boldsymbol{C}'\boldsymbol{\Psi}_{Zu}}(\boldsymbol{x})$. By Assumption A.1, $\boldsymbol{Z}_t u_t$ is an i.i.d., mean zero K_2-dimensional random variable with finite third moments, so Götze's (1991) Berry-Esseen bound (A.4) applies and $\sup_{\boldsymbol{x}} | F_{\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{u}}(\boldsymbol{x}) - F_{\boldsymbol{C}'\boldsymbol{\Psi}_{Zu}}(\boldsymbol{x}) | \le const \times \sqrt{K_2/T}$. The result (c) follows from $K_2/T \to 0$ by Assumption A2. We note that this line of argument is used in Jensen and Mayer (1975).

(d) The proof is the same as for (c).

(e) Write, $\boldsymbol{u}'\boldsymbol{P_Z u} = (T^{-1/2}\boldsymbol{u}'\boldsymbol{Z})(T^{-1}\boldsymbol{Z}'\boldsymbol{Z})(T^{-1/2}\boldsymbol{Z}'\boldsymbol{u}) = \xi_1 + \xi_2$, where $\xi_1 = (T^{-1/2}\boldsymbol{u}'\boldsymbol{Z})\boldsymbol{Q}_{ZZ}^{-1}(T^{-1/2}\boldsymbol{Z}'\boldsymbol{u})$, and $\xi_2 = (T^{-1/2}\boldsymbol{u}'\boldsymbol{Z})\Delta_T(T^{-1/2}\boldsymbol{Z}'\boldsymbol{u})$. As in the proof of (c), let P_T denote the probability measure associated with $T^{-1/2}\boldsymbol{Z}'\boldsymbol{u}$ and let P denote the limiting probability measure of $\boldsymbol{\Psi}_{Zu}$. Let $B(x)$ be the convex set, $B(x) = \{y \in \mathbb{R}^{K_2} : \boldsymbol{y}'\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{y} \le x\}$, so that $P_T(B(x)) = F_{\xi_1}(x)$ and $P(B(x)) = F_{\boldsymbol{\Psi}_{Zu}'\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{\Psi}_{Zu}}(x)$. It follows from (A.4) that $\sup_x | F_{\xi_1}(x) - F_{\boldsymbol{\Psi}_{Zu}'\boldsymbol{Q}_{ZZ}^{-1}\boldsymbol{\Psi}_{Zu}}(x) | \le const \times \sqrt{K_2/T}$. By Lemma A.2(a), $\xi_2 \xrightarrow{p} 0$ uniformly as $(K_2, T \to \infty)$, and the result (e) follows.

(f) and (g). The dimensions of $\boldsymbol{V}'\boldsymbol{P_Z u}$ and $\boldsymbol{V}'\boldsymbol{P_Z V}$ do not depend on K_2, and the proofs of (f) and (g) are similar to that of (e).

A.3. Many Weak Instrument Asymptotic Limits

This section collects calculations of the many weak instrument asymptotic limits of *k*-class estimators and Wald statistics. Because of Theorem A.1, these calculations are performed using sequential asymptotics, in which the fixed-K_2 weak instrument asymptotic limits in Section 2.4 are analyzed as $K_2 \to \infty$. The limiting distributions differ depending on the limiting behavior of *k*. The main results are collected in Theorem A.3.

***Theorem* A.3**. Suppose that Assumptions $L_{\boldsymbol{\Pi}}$ and A1 hold; that $\boldsymbol{\Lambda}_{K_2} \to \boldsymbol{\Lambda}_\infty$, where maxeval($\boldsymbol{\Lambda}_\infty$) < ∞; and that $K_2 \to \infty$ and $T \to \infty$ subject to $K_2^2/T \to 0$. Let $\boldsymbol{x}$ be a *n*-dimensional standard normal random variable. Then the following limits hold as (K_2, $T \to \infty$):

(a) (TSLS) If $T(k-1)/K_2 \to 0$, then

$$\hat{\boldsymbol{\beta}}(k) - \boldsymbol{\beta} \xrightarrow{p} \sigma_{uu}^{1/2}\, \boldsymbol{\Sigma}_{VV}^{-1/2} (\boldsymbol{\Lambda}_\infty + \boldsymbol{I}_n)^{-1} \boldsymbol{\rho} \text{ and} \tag{A.14}$$

$$W(k)/K_2 \xrightarrow{p} \frac{\boldsymbol{\rho}'(\boldsymbol{\Lambda}_\infty + \boldsymbol{I}_n)^{-1}\boldsymbol{\rho}}{n[1 - 2\boldsymbol{\rho}'(\boldsymbol{\Lambda}_\infty + \boldsymbol{I}_n)^{-1}\boldsymbol{\rho} + \boldsymbol{\rho}'(\boldsymbol{\Lambda}_\infty + \boldsymbol{I}_n)^{-2}\boldsymbol{\rho}]}. \tag{A.15}$$

(b) (BTSLS) If $\sqrt{K_2}\,[T(k-1)/K_2 - 1] \to 0$ and mineval($\boldsymbol{\Lambda}_\infty$) > 0 , then

$$\sqrt{K_2}\,(\hat{\boldsymbol{\beta}}(k) - \boldsymbol{\beta}) \xrightarrow{d} N(0, \sigma_{uu} \boldsymbol{\Sigma}_{VV}^{-1/2}\, \boldsymbol{\Lambda}_\infty^{-1} (\boldsymbol{\Lambda}_\infty + \boldsymbol{I}_n + \boldsymbol{\rho}\boldsymbol{\rho}')\, \boldsymbol{\Lambda}_\infty^{-1}\, \boldsymbol{\Sigma}_{VV}^{-1/2\prime}) \text{ and} \tag{A.16}$$

$$W(k) \xrightarrow{d} \boldsymbol{x}'(\boldsymbol{\Lambda}_\infty + \boldsymbol{I}_n + \boldsymbol{\rho\rho}')^{1/2} \boldsymbol{\Lambda}_\infty^{-1} (\boldsymbol{\Lambda}_\infty + \boldsymbol{I}_n + \boldsymbol{\rho\rho}')^{1/2\prime}\boldsymbol{x}/n. \tag{A.17}$$

(c) (LIML, Fuller-k). If $T(k - k_{LIML})/\sqrt{K_2} \to 0$ and mineval($\boldsymbol{\Lambda}_\infty$) > 0, then

$$\sqrt{K_2}\,[T(k-1)/K_2 - 1] \xrightarrow{d} N(0, 2), \tag{A.18}$$

$$\sqrt{K_2}\,(\hat{\boldsymbol{\beta}}(k) - \boldsymbol{\beta}) \xrightarrow{d} N(0, \sigma_{uu} \boldsymbol{\Sigma}_{VV}^{-1/2} \boldsymbol{\Lambda}_\infty^{-1} (\boldsymbol{\Lambda}_\infty + \boldsymbol{I}_n - \boldsymbol{\rho\rho}') \boldsymbol{\Lambda}_\infty^{-1} \boldsymbol{\Sigma}_{VV}^{-1/2\prime}) \text{ and} \tag{A.19}$$

$$W(k) \xrightarrow{d} \boldsymbol{x}'(\boldsymbol{\Lambda}_\infty + \boldsymbol{I}_n - \boldsymbol{\rho\rho}')^{1/2} \boldsymbol{\Lambda}_\infty^{-1} (\boldsymbol{\Lambda}_\infty + \boldsymbol{I}_n - \boldsymbol{\rho\rho}')^{1/2\prime}\boldsymbol{x}/n. \tag{A.20}$$

Before proving Theorem A.3, we first state some limiting properties of random variables that appear in the weak instrument representations. Let $\boldsymbol{z}_u$, $\boldsymbol{z}_V$, and $\boldsymbol{\rho}$ be as defined in Section 2.4, and let $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ be as defined in (2.13) and (2.14). Then the following limits hold jointly as $K_2 \to \infty$:

$$\boldsymbol{\nu}_1/K_2 \xrightarrow{p} \boldsymbol{\Lambda}_\infty + \boldsymbol{I}_n, \tag{A.21}$$

$$\boldsymbol{\nu}_2/K_2 \xrightarrow{p} \boldsymbol{\rho}, \tag{A.22}$$

$$\begin{pmatrix} \dfrac{\boldsymbol{z}_u'\boldsymbol{z}_u - K_2}{\sqrt{K_2}} \\ \dfrac{\boldsymbol{\lambda}'\boldsymbol{z}_u}{\sqrt{K_2}} \\ \dfrac{\boldsymbol{z}_V'\boldsymbol{z}_u - K_2\boldsymbol{\rho}}{\sqrt{K_2}} \end{pmatrix} \xrightarrow{d} N(0, B), \text{ where } B = \begin{bmatrix} 2 & 0 & 2\boldsymbol{\rho}' \\ 0 & \boldsymbol{\Lambda}_\infty & 0 \\ 2\boldsymbol{\rho} & 0 & \boldsymbol{I}_n + \boldsymbol{\rho\rho}' \end{bmatrix}, \tag{A.23}$$

$$(\boldsymbol{\nu}_2 - K_2\boldsymbol{\rho})/\sqrt{K_2} \to N(0, \boldsymbol{\Lambda}_\infty + \boldsymbol{I}_n + \boldsymbol{\rho}\boldsymbol{\rho}'). \tag{A.24}$$

The results (A.21) – (A.24) follow by straightforward calculations using the central limit theorem, the weak law of large numbers, and the joint normal distribution of $\boldsymbol{z}_u$ and $\boldsymbol{z}_V$ in (2.12).

Proof of Theorem A.3.

(a) From (2.21), the fixed-K_2 weak instrument approximation to the distribution of the TSLS estimator is $\hat{\boldsymbol{\beta}}^{TSLS} - \boldsymbol{\beta} \sim \sigma_{uu}^{1/2}\,\boldsymbol{\Sigma}_{VV}^{-1/2}\,\boldsymbol{\nu}_1^{-1}\boldsymbol{\nu}_2 = \sigma_{uu}^{1/2}\,\boldsymbol{\Sigma}_{VV}^{-1/2}\,(\boldsymbol{\nu}_1/K_2)^{-1}(\boldsymbol{\nu}_2/K_2)$. The limit stated in the theorem for the estimator follows by substituting (A.21) and (A.22) into this expression. The many weak instrument limit for the Wald statistic follows by modifying (2.22) to be

$$W^{TSLS}/K_2 \sim \frac{(\boldsymbol{\nu}_2/K_2)'(\boldsymbol{\nu}_1/K_2)^{-1}(\boldsymbol{\nu}_2/K_2)}{n[1-2\boldsymbol{\rho}'(\boldsymbol{\nu}_1/K_2)^{-1}(\boldsymbol{\nu}_2/K_2)+(\boldsymbol{\nu}_2/K_2)'(\boldsymbol{\nu}_1/K_2)^{-2}(\boldsymbol{\nu}_2/K_2)]}$$

and applying (A.21) and (A.22).

(b) The fixed-K_2 weak instrument approximation to the distribution of a k-class estimator, given in (2.15), in general can be written,

$$\sqrt{K_2}\,[\hat{\boldsymbol{\beta}}(k) - \boldsymbol{\beta}] \sim \sigma_{uu}^{1/2}\boldsymbol{\Sigma}_{VV}^{-1/2}\left[\frac{\boldsymbol{\nu}_1 - K_2\boldsymbol{I}_n}{K_2} - \frac{1}{\sqrt{K_2}}\left(\frac{\kappa - K_2}{\sqrt{K_2}}\right)\boldsymbol{I}_n\right]^{-1}$$

$$\times\left[\frac{\nu_2 - K_2\rho}{\sqrt{K_2}} - \left(\frac{\kappa - K_2}{\sqrt{K_2}}\right)\rho\right], \tag{A.25}$$

where $T(k-1) \xrightarrow{d} \kappa$ for fixed K_2. The assumption $\sqrt{K_2}\,[T(k-1)/K_2 - 1] \to 0$ implies that $(\kappa - K_2)/\sqrt{K_2} \to 0$, so by (A.21) and (A.24) we have, as $K_2 \to \infty$,

$$\frac{\nu_1 - K_2 \boldsymbol{I}_n}{K_2} - \frac{1}{\sqrt{K_2}}\left(\frac{\kappa - K_2}{\sqrt{K_2}}\right)\boldsymbol{I}_n \xrightarrow{p} \boldsymbol{\Lambda}_\infty \text{ and}$$

$$\frac{\nu_2 - K_2\rho}{\sqrt{K_2}} - \left(\frac{\kappa - K_2}{\sqrt{K_2}}\right)\rho \xrightarrow{d} \mathrm{N}(0, \boldsymbol{\Lambda}_\infty + \boldsymbol{I}_n + \rho\rho'),$$

and the result (A.16) follows. The assumption mineval($\boldsymbol{\Lambda}_\infty$) > 0 is used to ensure the invertibility of $\boldsymbol{\Lambda}_\infty$. The distribution of the Wald statistic follows.

(c) For fixed K_2, $T(k_{LIML} - 1) \xrightarrow{d} \kappa^*$. We show below that, as $K_2 \to \infty$,

$$\frac{\kappa^* - K_2}{\sqrt{K_2}} = \frac{\mathbf{z}_u'\mathbf{z}_u - K_2}{\sqrt{K_2}} + o_p(1). \tag{A.26}$$

The result (A.18) follows from (A.26) and (A.23). Moreover, applying (A.21), (A.23), (A.24), and (A.26) yields,

$$\frac{\boldsymbol{\nu}_1 - K_2\boldsymbol{I}_n}{K_2} - \frac{1}{\sqrt{K_2}}\left(\frac{\kappa^* - K_2}{\sqrt{K_2}}\right)\boldsymbol{I}_n \xrightarrow{p} \boldsymbol{\Lambda}_\infty \text{ and}$$

$$\frac{\boldsymbol{\nu}_2 - K_2\boldsymbol{\rho}}{\sqrt{K_2}} - \left(\frac{\kappa^* - K_2}{\sqrt{K_2}}\right)\boldsymbol{\rho} = \frac{\boldsymbol{\lambda}'\boldsymbol{z}_u}{\sqrt{K_2}} + \frac{\boldsymbol{z}_V{}'\boldsymbol{z}_u - K_2\boldsymbol{\rho}}{\sqrt{K_2}} - \left(\frac{\boldsymbol{z}_u{}'\boldsymbol{z}_u - K_2}{\sqrt{K_2}}\right)\boldsymbol{\rho} + o_p(1)$$

$$\xrightarrow{d} \mathrm{N}(0, \boldsymbol{\Lambda}_\infty + \boldsymbol{I}_n - \boldsymbol{\rho}\boldsymbol{\rho}'),$$

where $\boldsymbol{\Lambda}_\infty$ is invertible by the assumption mineval($\boldsymbol{\Lambda}_\infty$) > 0. The result (A.19) follows, as does the distribution of the Wald statistic.

It remains to show (A.26). From (2.18), κ^* is the smallest root of,

$$0 = \det\left[\begin{pmatrix} \boldsymbol{z}_u{}'\boldsymbol{z}_u & \boldsymbol{\nu}_2{}' \\ \boldsymbol{\nu}_2 & \boldsymbol{\nu}_1 \end{pmatrix} - \kappa^*\begin{pmatrix} 1 & \rho' \\ \rho & \boldsymbol{I}_n \end{pmatrix}\right]. \tag{A.27}$$

Let $\phi = (\kappa^* - K_2)/\sqrt{K_2}$, $a = (\boldsymbol{z}_u{}'\boldsymbol{z}_u - K_2)/\sqrt{K_2}$, $\boldsymbol{b} = (\boldsymbol{\nu}_2 - K_2\boldsymbol{\rho})/\sqrt{K_2}$, and $\boldsymbol{L} = (\boldsymbol{\nu}_1 - K_2\boldsymbol{I}_n)/K_2$. Then (A.27) can be rewritten so that ϕ is the smallest root of

$$0 = \det\begin{bmatrix} a - \phi & (\boldsymbol{b} - \phi\boldsymbol{\rho})' \\ \boldsymbol{b} - \phi\boldsymbol{\rho} & \sqrt{K_2}\boldsymbol{L} - \phi\boldsymbol{I}_n \end{bmatrix}. \tag{A.28}$$

We first show that $K_2^{-1/4}\phi \xrightarrow{p} 0$. Let $\tilde{\phi} = K_2^{-1/4}\phi$. By (A.21), (A.23), and (A.24), $K_2^{-1/4}a \xrightarrow{p} 0$, $K_2^{-1/4}\boldsymbol{b} \xrightarrow{p} 0$, and $L \xrightarrow{p} \boldsymbol{\Lambda}_\infty$. By the continuity of the determinant, it follows that in the limit $K_2 \to \infty$, $\tilde{\phi}$ is the smallest root of the equation,

$$0 = \det\begin{bmatrix} \tilde{\phi} & \tilde{\phi}\boldsymbol{\rho}' \\ \tilde{\phi}\boldsymbol{\rho} & \tilde{\phi}\boldsymbol{I}_n + O_p(K_2^{1/4}) \end{bmatrix}, \tag{A.29}$$

from which it follows that $\tilde{\phi} = K_2^{-1/4}\phi \xrightarrow{p} 0$.

To obtain (A.26), write the determinantal equation (A.28) as

$$\begin{aligned} 0 &= [(a - \phi) - (\boldsymbol{b} - \phi\boldsymbol{\rho})'(K_2^{1/2}\boldsymbol{L} - \phi\boldsymbol{I}_n)^{-1}(\boldsymbol{b} - \phi\boldsymbol{\rho})]\det(K_2^{1/2}\boldsymbol{L} - \phi\boldsymbol{I}_n) \\ &= K_2^{n/2}\{(a - \phi) - [K_2^{-1/4}(\boldsymbol{b} - \phi\boldsymbol{\rho})]'(\boldsymbol{L} - K_2^{-1/2}\phi\boldsymbol{I}_n)^{-1}[K_2^{-1/4}(\boldsymbol{b} - \phi\boldsymbol{\rho})\}\det(\boldsymbol{L} - K_2^{-1/2}\phi\boldsymbol{I}_n) \\ &= K_2^{n/2}\{[(a - \phi)]\det(\boldsymbol{\Lambda}_\infty) + o_p(1)\}, \end{aligned} \tag{A.30}$$

where the final equality follows from $K_2^{-1/4}\boldsymbol{b} \xrightarrow{p} 0$, $L \xrightarrow{p} \boldsymbol{\Lambda}_\infty$, $K_2^{-1/4}\phi \xrightarrow{p} 0$, and $\det(\boldsymbol{\Lambda}_\infty) > 0$. By the continuity of the solution to (A.28), it follows that $\phi = a + o_p(1)$ which, in the original notation, is (A.26).

A.4. Remarks

1. To simplify the calculations, we have assumed i.i.d. sampling. Götze (1991) provides a Berry-Esseen bound for i.n.i.d. sampling in which the rate is $K_2^2/T \to 0$ rather than the rate of $K_2/T \to 0$, which appears in his result quoted in Section A.2. Because we already assume that $K_2^2/T \to 0$ (this rate is used in the consistency calculations in Theorem A.1(b) and Lemma A.2), the results in Section A.2 extend to the case where

the errors and instruments are independently but not necessarily identically distributed.

2. The many weak instrument representations in Theorem A.3 for BTSLS, LIML and the Fuller-*k* estimator rule out the partially identified and unidentified cases, for which mineval($\boldsymbol{\Lambda}_\infty$) = 0. This suggests that the approximations in Theorem A.3(b) and (c) might become inaccurate as $\boldsymbol{\Lambda}_{K_2}$ becomes nearly singular. The behavior of the many weak instrument approximations in partially identified and unidentified cases remain to be explored.

Endnotes

[1] See Rothenberg (1984, pp. 921) for a discussion of the quality of the Edgeworth approximation as a function of μ^2 and K_2.

[2] The definition of $\boldsymbol{G}_T$ in (2.10) is $\boldsymbol{G}_T$ in Staiger and Stock (1997, eq. (3.4)), divided by K_2 to put it in F-statistic form.

[3] In the appendix, the assumption that $\boldsymbol{\Lambda}$ is constant is generalized slightly to consider sequences of $\boldsymbol{\Lambda}$, indexed by K_2, that have a finite limit $\boldsymbol{\Lambda}_\infty$ as $K_2 \rightarrow \infty$.

[4] Because in general the maximal bias depends on all the eigenvalues, the maximal bias when all the eigenvalues are equal to some value ℓ_0 might be greater than the maximal bias when one eigenvalue is slightly less than ℓ_0 but the others are large. For this reason the set $\mathcal{W}_{bias}$ is potentially conservative when K_2 is small. This comment applies to size-based sets as well.

[5] It appears that there is some non-monotonicity in the dependence on the eigenvalues for Fuller-k bias when the minimum eigenvalue is very small, but for such small eigenvalues the bias is sufficiently large that this non-monotonicity does not affect the boundary eigenvalues.

[6]The critical values based on Proposition 1 can be quite conservative when all the eigenvalues of $\boldsymbol{\Lambda}$ are small. For example, the boundary of the TSLS bias-based weak instrument set with b = 0.1, n = 2, and K_2 = 4 is mineval($\boldsymbol{\Lambda}$) = 3.08, and the critical value for a 5% test with b = .10 based on Proposition 1 is 7.56. If the second eigenvalue in fact equals the first, the correct critical value should be 4.63, and the rejection probability under the null is only 0.1%. (Of course, it is infeasible to use this critical value because

the second eigenvalue of Λ is unknown.) If the second eigenvalue is 10, then the rejection rate is approximately 2%. On the other hand, if the second eigenvalue is large, the Proposition 1 bound is tighter. For example, for values of K_2 from 4 to 34 and $n = 2$, if the second eigenvalue exceeds 20 the rejection probability under the null range from 3.3% to 4.1% for the nominal 5% weak instrument test based on TSLS bias with $b = .10$.

References

Anderson, T.W. (1974): "An Asymptotic Expansion of the Distribution of the Limited Information Maximum Likelihood Estimate of a Coefficient in a Simultaneous Equation System," *Journal of the American Statistical Association*, 69, 565 – 573.

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

Bound, J., D.A. Jaeger, and R.M. 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. and D. Turkington (1984): *Instrumental Variables*. Cambridge: Cambridge University Press.

Chao, J.C. and N.R. Swanson (2002): "Consistent Estimation with a Large Number of Weak Instruments," manuscript, University of Maryland.

Cragg, J.G. and S.G. Donald (1993): "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 9, 222 – 240.

Donald, S.G. and W.K. Newey (2001): "Choosing the Number of Instruments," *Econometrica*, 69, 1161 – 1191.

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

Götze, F. (1991): "On the Rate of Convergence in the Multivariate CLT," *Annals of Probability*, 19, 724 – 739.

Hall, A., G.D. Rudebusch and D. Wilcox (1996): "Judging Instrument Relevance in Instrumental Variables Estimation," *International Economic Review* 37, 283 – 298.

Hahn, J. and Hausman, J. (2002): "A New Specification Test for the Validity of Instrumental Variables," *Econometrica*, 70, 163 – 189.

Jensen, D.R. and L.S. Mayer (1975): "Normal-Theory Approximations to Tests of Linear Hypotheses," *Annals of Statistics*, 3, 429 – 444.

Kunitomo, N. (1980): "Asymptotic Expansions of the Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations," *Journal of the American Statistical Association*, 75, 693-700.

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.

Muirhead, R.J. (1978): "Latent Roots and Matrix Variates: A Review of Some Asymptotic Results," *Annals of Statistics*, 6, 5-33.

——— (1982): *Aspects of Multivariate Statistical Theory*. New York: Wiley.

Nagar, A.L. (1959): "The Bias and Moment Matrix of the General *k*-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27, 575 – 595.

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

——— (1990b): "The Distribution of the Instrumental Variables Estimator and its *t*-ratio when the Instrument is a Poor One," *Journal of Business* 63, 5125 – 5140.

Rothenberg, T.J. (1983): "Asymptotic Properties of Some Estimators in Structural Models," in S. Karlin, T. Amemiya, and L. Goodman, eds., *Studies in Econometrics, Time Series, and Multivariate Statistics*, Academic Press.

———— (1984): "Approximating the Distributions of Econometric Estimators and Test Statistics," Ch. 15 in *Handbook of Econometrics, Vol. II*, ed. by Z. Griliches and M.D. Intriligator. Amsterdam: North Holland, 881 – 935.

Sargan, D. (1975): "Asymptotic Theory and Large Models," *International Economic Review*, 16, 75 – 91.

Shea, J. (1997): "Instrument Relevance in Multivariate Linear Models: A Simple Measure," *Review of Economics and Statistics* 79, 348 – 352.

Staiger, D. and J.H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557 – 586.

Stock, J.H., J.H. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20, 518 – 529.

Table 1.
Critical Values for the Weak Instrument Test Based on TSLS Bias
Significance level is 5%

	***n* = 1, *b* =**				***n* = 2, *b* =**				***n* = 3, *b* =**			
K_2	**0.05**	**0.10**	**0.20**	**0.30**	**0.05**	**0.10**	**0.20**	**0.30**	**0.05**	**0.10**	**0.20**	**0.30**
3	13.91	9.08	6.46	5.39	.	.	.	.	.	.	.	.
4	16.85	10.27	6.71	5.34	11.04	7.56	5.57	4.73	.	.	.	.
5	18.37	10.83	6.77	5.25	13.97	8.78	5.91	4.79	9.53	6.61	4.99	4.30
6	19.28	11.12	6.76	5.15	15.72	9.48	6.08	4.78	12.20	7.77	5.35	4.40
7	19.86	11.29	6.73	5.07	16.88	9.92	6.16	4.76	13.95	8.50	5.56	4.44
8	20.25	11.39	6.69	4.99	17.70	10.22	6.20	4.73	15.18	9.01	5.69	4.46
9	20.53	11.46	6.65	4.92	18.30	10.43	6.22	4.69	16.10	9.37	5.78	4.46
10	20.74	11.49	6.61	4.86	18.76	10.58	6.23	4.66	16.80	9.64	5.83	4.45
11	20.90	11.51	6.56	4.80	19.12	10.69	6.23	4.62	17.35	9.85	5.87	4.44
12	21.01	11.52	6.53	4.75	19.40	10.78	6.22	4.59	17.80	10.01	5.90	4.42
13	21.10	11.52	6.49	4.71	19.64	10.84	6.21	4.56	18.17	10.14	5.92	4.41
14	21.18	11.52	6.45	4.67	19.83	10.89	6.20	4.53	18.47	10.25	5.93	4.39
15	21.23	11.51	6.42	4.63	19.98	10.93	6.19	4.50	18.73	10.33	5.94	4.37
16	21.28	11.50	6.39	4.59	20.12	10.96	6.17	4.48	18.94	10.41	5.94	4.36
17	21.31	11.49	6.36	4.56	20.23	10.99	6.16	4.45	19.13	10.47	5.94	4.34
18	21.34	11.48	6.33	4.53	20.33	11.00	6.14	4.43	19.29	10.52	5.94	4.32
19	21.36	11.46	6.31	4.51	20.41	11.02	6.13	4.41	19.44	10.56	5.94	4.31
20	21.38	11.45	6.28	4.48	20.48	11.03	6.11	4.39	19.56	10.60	5.93	4.29
21	21.39	11.44	6.26	4.46	20.54	11.04	6.10	4.37	19.67	10.63	5.93	4.28
22	21.40	11.42	6.24	4.43	20.60	11.05	6.08	4.35	19.77	10.65	5.92	4.27
23	21.41	11.41	6.22	4.41	20.65	11.05	6.07	4.33	19.86	10.68	5.92	4.25
24	21.41	11.40	6.20	4.39	20.69	11.05	6.06	4.32	19.94	10.70	5.91	4.24
25	21.42	11.38	6.18	4.37	20.73	11.06	6.05	4.30	20.01	10.71	5.90	4.23
26	21.42	11.37	6.16	4.35	20.76	11.06	6.03	4.29	20.07	10.73	5.90	4.21
27	21.42	11.36	6.14	4.34	20.79	11.06	6.02	4.27	20.13	10.74	5.89	4.20
28	21.42	11.34	6.13	4.32	20.82	11.05	6.01	4.26	20.18	10.75	5.88	4.19
29	21.42	11.33	6.11	4.31	20.84	11.05	6.00	4.24	20.23	10.76	5.88	4.18
30	21.42	11.32	6.09	4.29	20.86	11.05	5.99	4.23	20.27	10.77	5.87	4.17

Notes: The test rejects if g_{min} exceeds the critical value. The critical value is a function of the number of included endogenous regressors (n), the number of instrumental variables (K_2), and the desired maximal bias of the IV estimator relative to OLS (b).

Table 2.
Critical Values for the Weak Instrument Test Based on TSLS Size
Significance level is 5%

	***n* = 1, *r* =**				***n* = 2, *r* =**			
***K*₂**	**0.10**	**0.15**	**0.20**	**0.25**	**0.10**	**0.15**	**0.20**	**0.25**
1	16.38	8.96	6.66	5.53	.	.	.	.
2	19.93	11.59	8.75	7.25	7.03	4.58	3.95	3.63
3	22.30	12.83	9.54	7.80	13.43	8.18	6.40	5.45
4	24.58	13.96	10.26	8.31	16.87	9.93	7.54	6.28
5	26.87	15.09	10.98	8.84	19.45	11.22	8.38	6.89
6	29.18	16.23	11.72	9.38	21.68	12.33	9.10	7.42
7	31.50	17.38	12.48	9.93	23.72	13.34	9.77	7.91
8	33.84	18.54	13.24	10.50	25.64	14.31	10.41	8.39
9	36.19	19.71	14.01	11.07	27.51	15.24	11.03	8.85
10	38.54	20.88	14.78	11.65	29.32	16.16	11.65	9.31
11	40.90	22.06	15.56	12.23	31.11	17.06	12.25	9.77
12	43.27	23.24	16.35	12.82	32.88	17.95	12.86	10.22
13	45.64	24.42	17.14	13.41	34.62	18.84	13.45	10.68
14	48.01	25.61	17.93	14.00	36.36	19.72	14.05	11.13
15	50.39	26.80	18.72	14.60	38.08	20.60	14.65	11.58
16	52.77	27.99	19.51	15.19	39.80	21.48	15.24	12.03
17	55.15	29.19	20.31	15.79	41.51	22.35	15.83	12.49
18	57.53	30.38	21.10	16.39	43.22	23.22	16.42	12.94
19	59.92	31.58	21.90	16.99	44.92	24.09	17.02	13.39
20	62.30	32.77	22.70	17.60	46.62	24.96	17.61	13.84
21	64.69	33.97	23.50	18.20	48.31	25.82	18.20	14.29
22	67.07	35.17	24.30	18.80	50.01	26.69	18.79	14.74
23	69.46	36.37	25.10	19.41	51.70	27.56	19.38	15.19
24	71.85	37.57	25.90	20.01	53.39	28.42	19.97	15.64
25	74.24	38.77	26.71	20.61	55.07	29.29	20.56	16.10
26	76.62	39.97	27.51	21.22	56.76	30.15	21.15	16.55
27	79.01	41.17	28.31	21.83	58.45	31.02	21.74	17.00
28	81.40	42.37	29.12	22.43	60.13	31.88	22.33	17.45
29	83.79	43.57	29.92	23.04	61.82	32.74	22.92	17.90
30	86.17	44.78	30.72	23.65	63.51	33.61	23.51	18.35

Notes: The test rejects if g_{min} exceeds the critical value. The critical value is a function of the number of included endogenous regressors (n), the number of instrumental variables (K_2), and the desired maximal size (r) of a 5% Wald test of $\beta = \beta_0$.

Table 3.
Critical Values for the Weak Instrument Test Based on Fuller-*k* Bias
Significance level is 5%

	***n* = 1, *b* =**				***n* = 1, *b* =**			
K_2	**0.05**	**0.10**	**0.20**	**0.30**	**0.05**	**0.10**	**0.20**	**0.30**
1	23.63	19.35	15.42	12.86	.	.	.	.
2	15.60	12.38	7.93	6.62	14.14	11.94	9.50	8.11
3	12.04	9.59	6.15	5.13	11.62	9.21	6.57	5.70
4	10.09	8.10	5.36	4.46	9.96	7.80	5.43	4.70
5	8.85	7.16	4.89	4.07	8.84	6.94	4.84	4.16
6	7.99	6.51	4.58	3.82	8.02	6.34	4.47	3.82
7	7.35	6.02	4.35	3.63	7.41	5.90	4.22	3.58
8	6.86	5.65	4.17	3.48	6.93	5.56	4.03	3.41
9	6.47	5.35	4.02	3.36	6.54	5.29	3.89	3.27
10	6.14	5.11	3.90	3.27	6.22	5.06	3.77	3.16
11	5.87	4.90	3.79	3.18	5.94	4.87	3.66	3.07
12	5.64	4.72	3.70	3.11	5.71	4.71	3.58	3.00
13	5.43	4.57	3.62	3.05	5.50	4.57	3.50	2.93
14	5.26	4.43	3.54	2.99	5.33	4.44	3.43	2.87
15	5.10	4.31	3.48	2.94	5.17	4.33	3.37	2.82
16	4.95	4.20	3.41	2.90	5.02	4.23	3.32	2.78
17	4.83	4.10	3.36	2.86	4.89	4.13	3.27	2.74
18	4.71	4.01	3.30	2.82	4.77	4.05	3.22	2.70
19	4.60	3.93	3.25	2.78	4.67	3.97	3.18	2.67
20	4.50	3.85	3.21	2.75	4.56	3.90	3.13	2.64
21	4.41	3.78	3.16	2.72	4.47	3.83	3.10	2.61
22	4.32	3.71	3.12	2.69	4.39	3.76	3.06	2.59
23	4.24	3.65	3.08	2.66	4.31	3.70	3.02	2.56
24	4.17	3.59	3.04	2.63	4.23	3.65	2.99	2.54
25	4.09	3.54	3.01	2.61	4.16	3.59	2.96	2.52
26	4.03	3.48	2.97	2.59	4.09	3.54	2.93	2.50
27	3.96	3.43	2.94	2.56	4.03	3.49	2.90	2.48
28	3.90	3.39	2.91	2.54	3.97	3.45	2.87	2.47
29	3.85	3.34	2.88	2.52	3.91	3.40	2.85	2.45
30	3.79	3.30	2.85	2.50	3.86	3.36	2.82	2.43

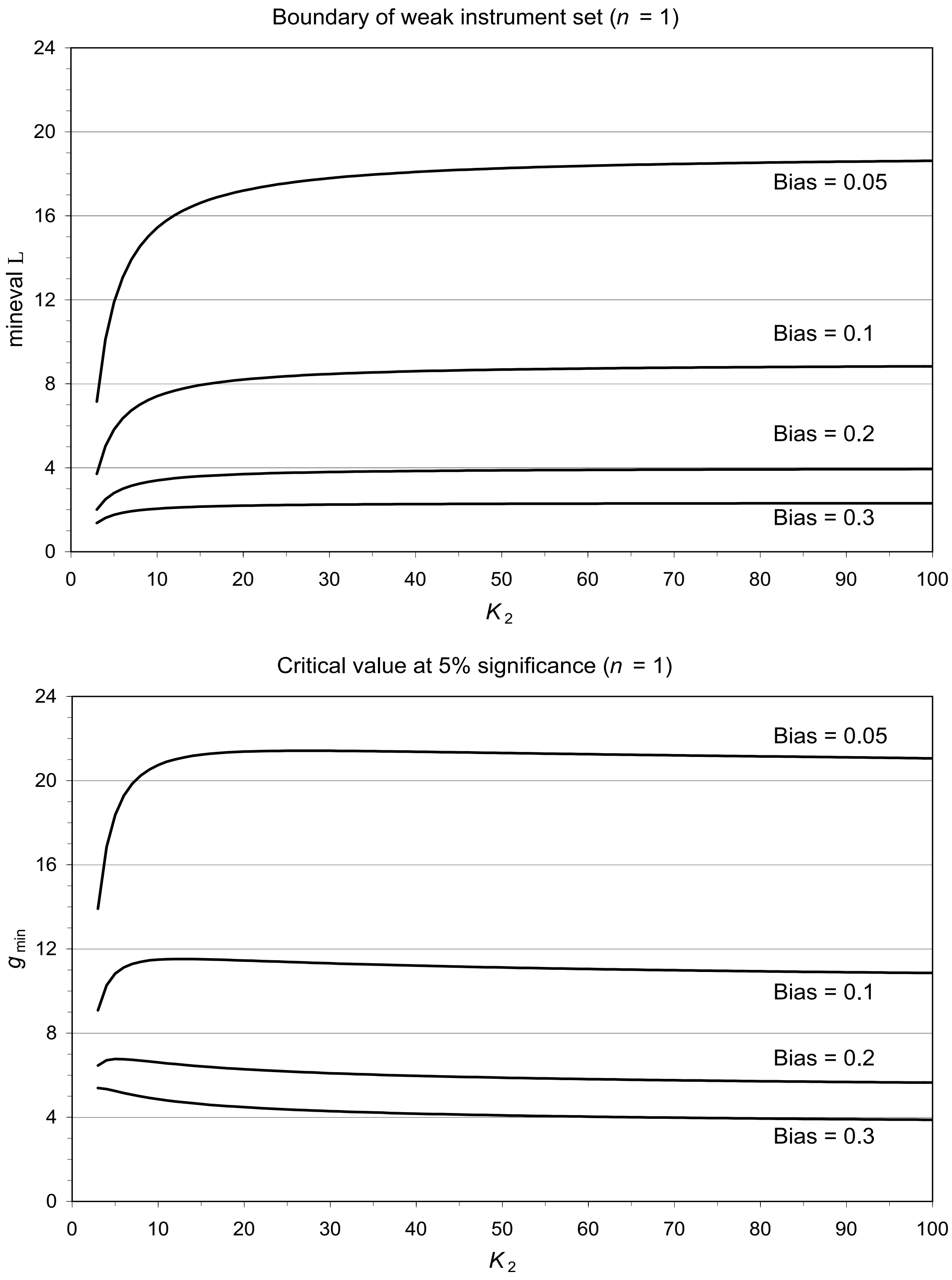
See the notes to Table 1.

Table 4.
Critical Values for the Weak Instrument Test Based on LIML Size
Significance level is 5%

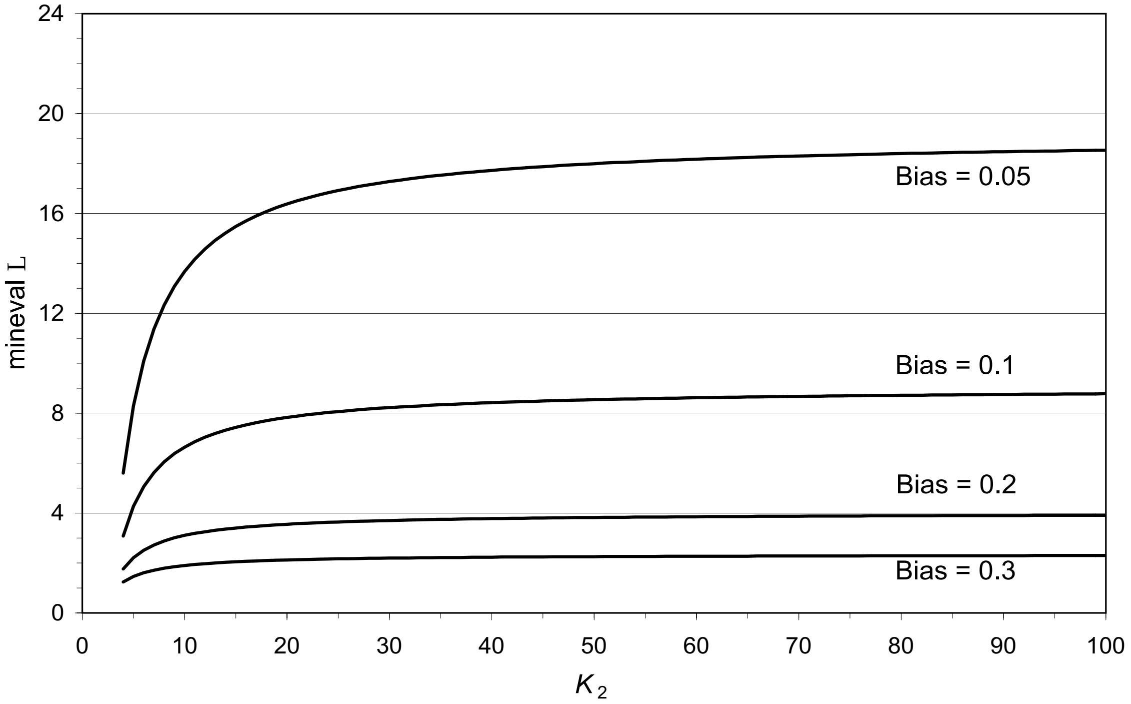
	***n* = 1, *r* =**				***n* = 1, *r* =**			
K_2	**0.10**	**0.15**	**0.20**	**0.25**	**0.10**	**0.15**	**0.20**	**0.25**
1	16.38	8.96	6.66	5.53	.	.	.	.
2	8.68	5.33	4.42	3.92	7.03	4.58	3.95	3.63
3	6.46	4.36	3.69	3.32	5.44	3.81	3.32	3.09
4	5.44	3.87	3.30	2.98	4.72	3.39	2.99	2.79
5	4.84	3.56	3.05	2.77	4.32	3.13	2.78	2.60
6	4.45	3.34	2.87	2.61	4.06	2.95	2.63	2.46
7	4.18	3.18	2.73	2.49	3.90	2.83	2.52	2.35
8	3.97	3.04	2.63	2.39	3.78	2.73	2.43	2.27
9	3.81	2.93	2.54	2.32	3.70	2.66	2.36	2.20
10	3.68	2.84	2.46	2.25	3.64	2.60	2.30	2.14
11	3.58	2.76	2.40	2.19	3.60	2.55	2.25	2.09
12	3.50	2.69	2.34	2.14	3.58	2.52	2.21	2.05
13	3.42	2.63	2.29	2.10	3.56	2.48	2.17	2.02
14	3.36	2.57	2.25	2.06	3.55	2.46	2.14	1.99
15	3.31	2.52	2.21	2.03	3.54	2.44	2.11	1.96
16	3.27	2.48	2.18	2.00	3.55	2.42	2.09	1.93
17	3.24	2.44	2.14	1.97	3.55	2.41	2.07	1.91
18	3.20	2.41	2.11	1.94	3.56	2.40	2.05	1.89
19	3.18	2.37	2.09	1.92	3.57	2.39	2.03	1.87
20	3.21	2.34	2.06	1.90	3.58	2.38	2.02	1.86
21	3.39	2.32	2.04	1.88	3.59	2.38	2.01	1.84
22	3.57	2.29	2.02	1.86	3.60	2.37	1.99	1.83
23	3.68	2.27	2.00	1.84	3.62	2.37	1.98	1.81
24	3.75	2.25	1.98	1.83	3.64	2.37	1.98	1.80
25	3.79	2.24	1.96	1.81	3.65	2.37	1.97	1.79
26	3.82	2.22	1.95	1.80	3.67	2.38	1.96	1.78
27	3.85	2.21	1.93	1.78	3.74	2.38	1.96	1.77
28	3.86	2.20	1.92	1.77	3.87	2.38	1.95	1.77
29	3.87	2.19	1.90	1.76	4.02	2.39	1.95	1.76
30	3.88	2.18	1.89	1.75	4.12	2.39	1.95	1.75

See the notes to Table 2.

Figure 1: Weak Instrument Sets and Critical Values based on Bias of TSLS Relative to OLS



Boundary of weak instrument set (n = 2)



Critical value at 5% significance (n = 2)

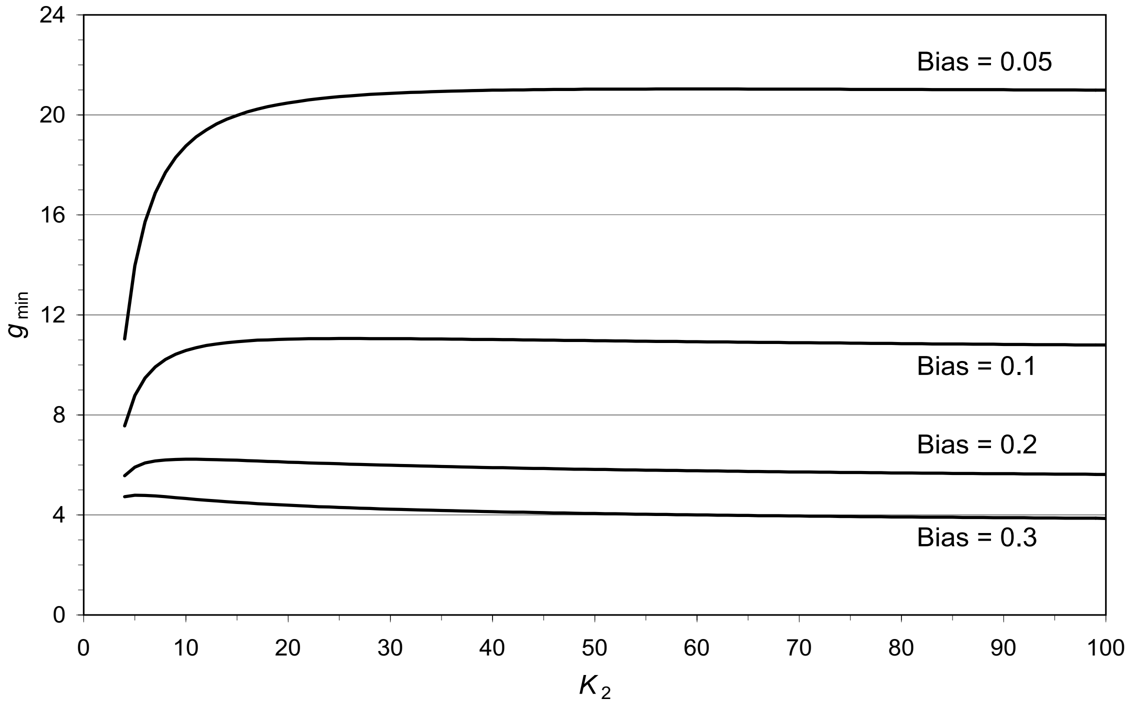
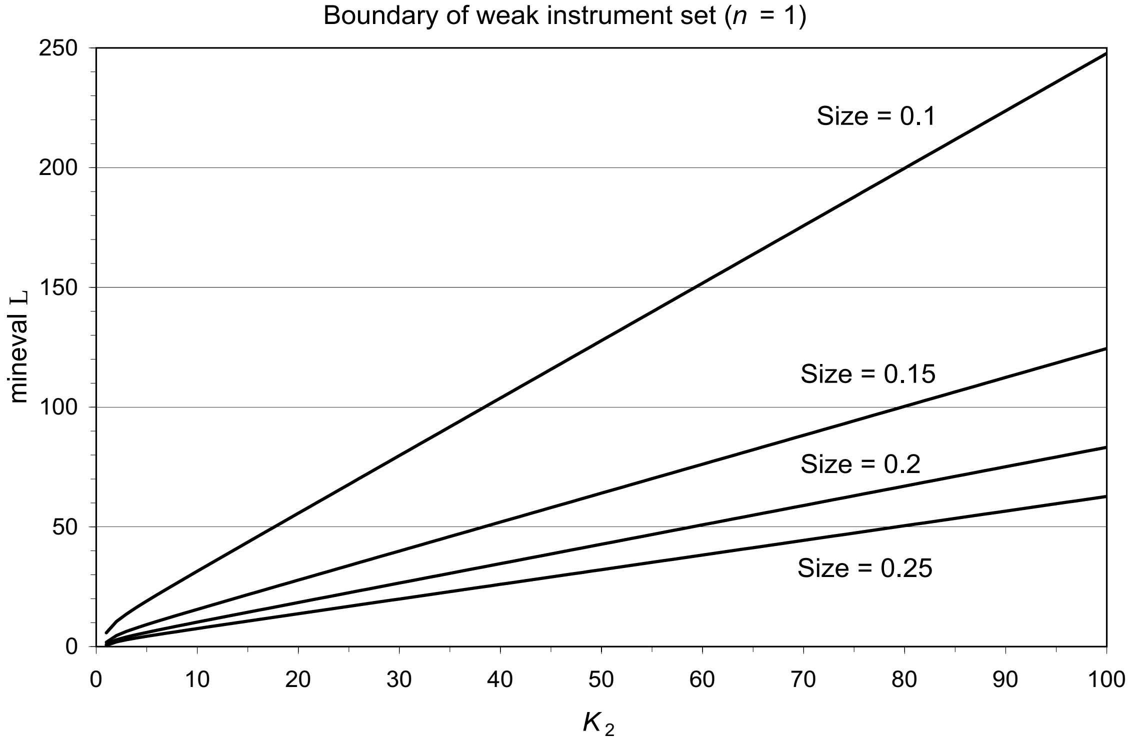
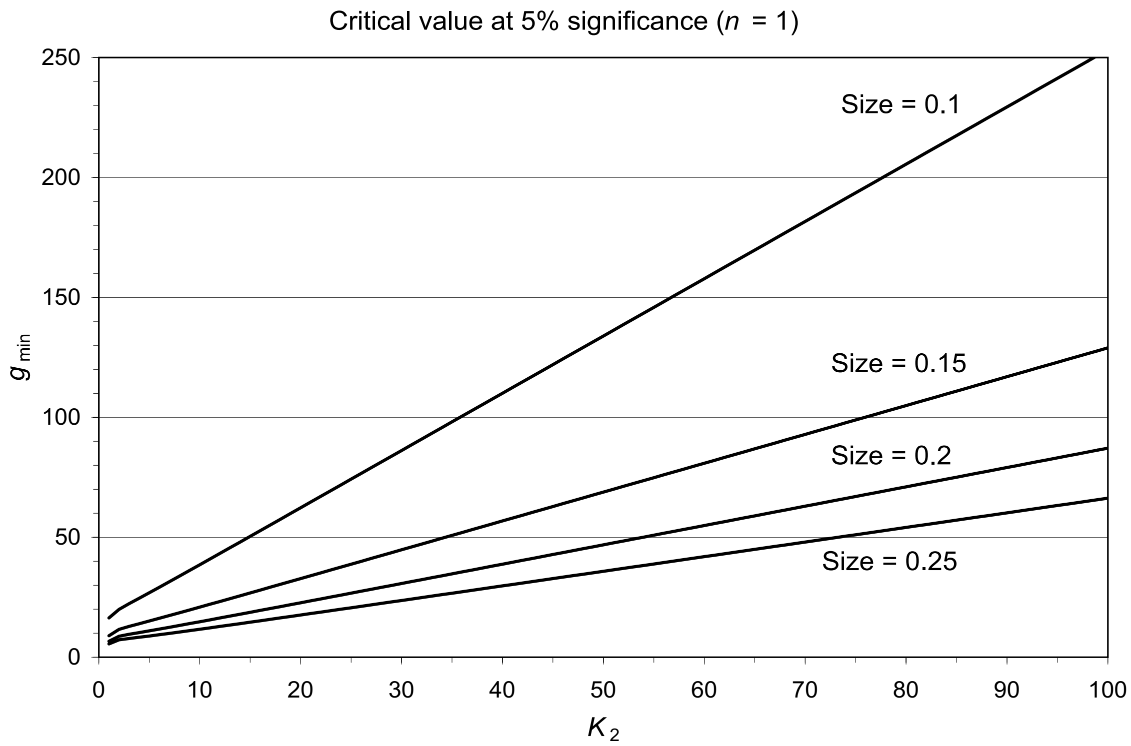
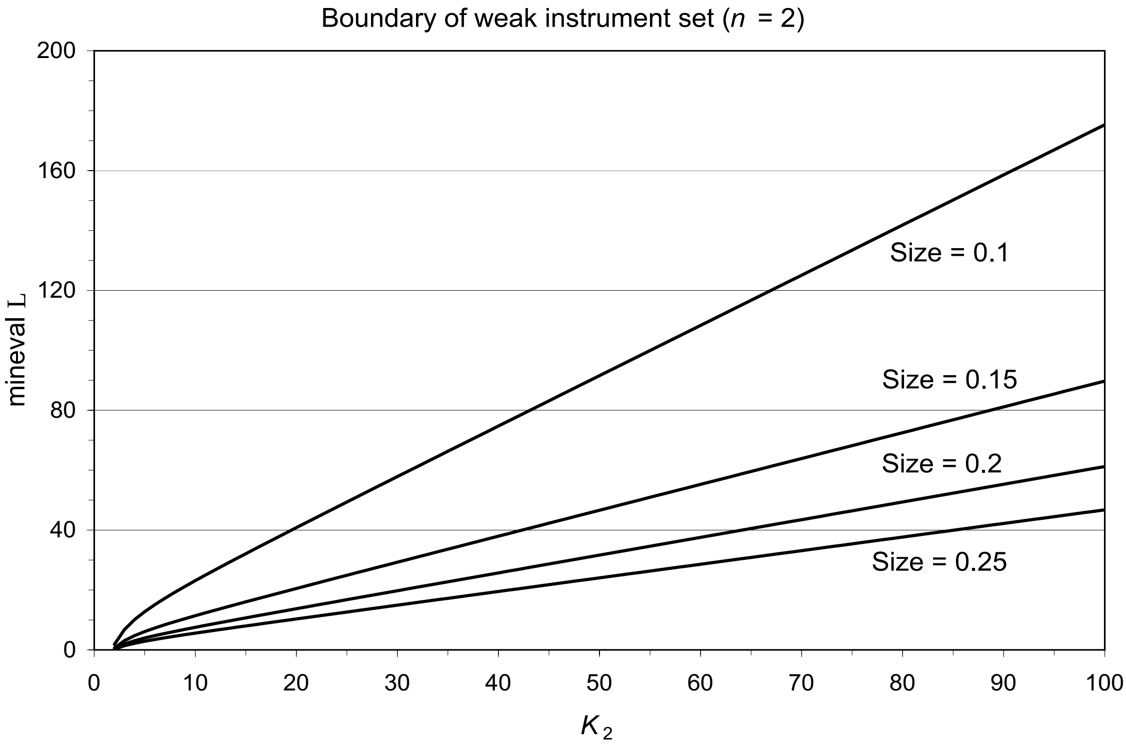


Figure 2: Weak Instrument Sets and Critical Values based on Size of TSLS Wald Test






Boundary of weak instrument set (n = 2)
mineval L
K 2
Size = 0.1
Size = 0.15
Size = 0.2
Size = 0.25
200
160
120
80
40
0
0
10
20
30
40
50
60
70
80
90
100

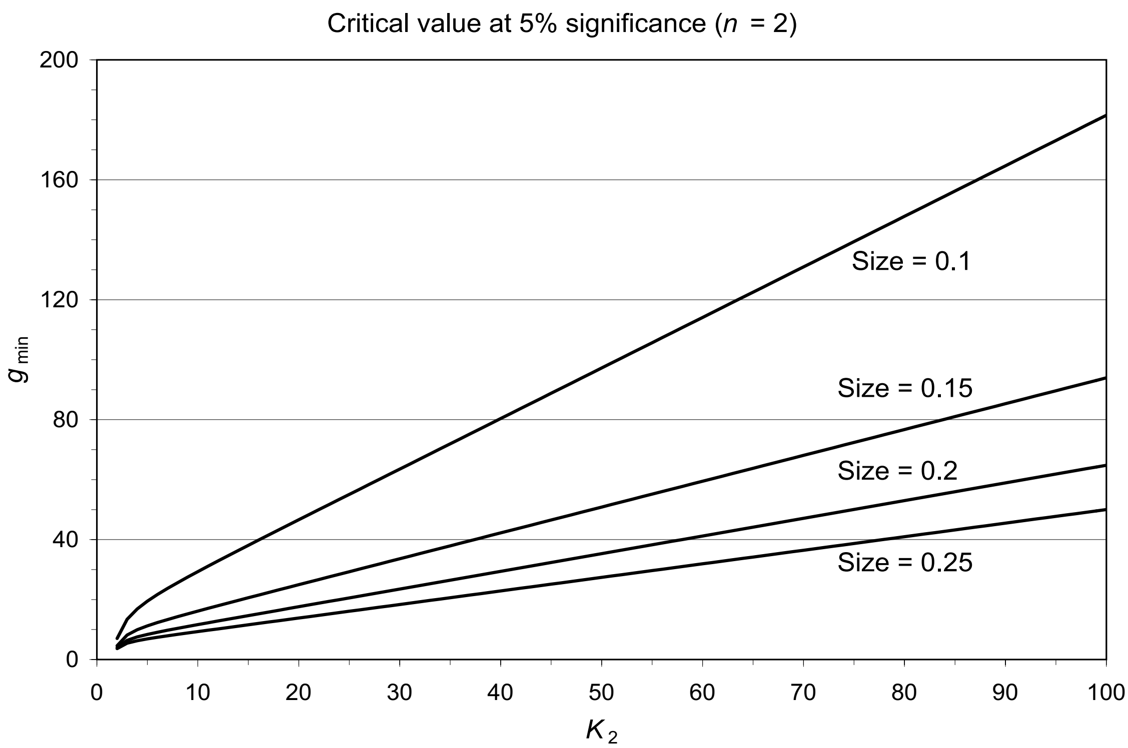
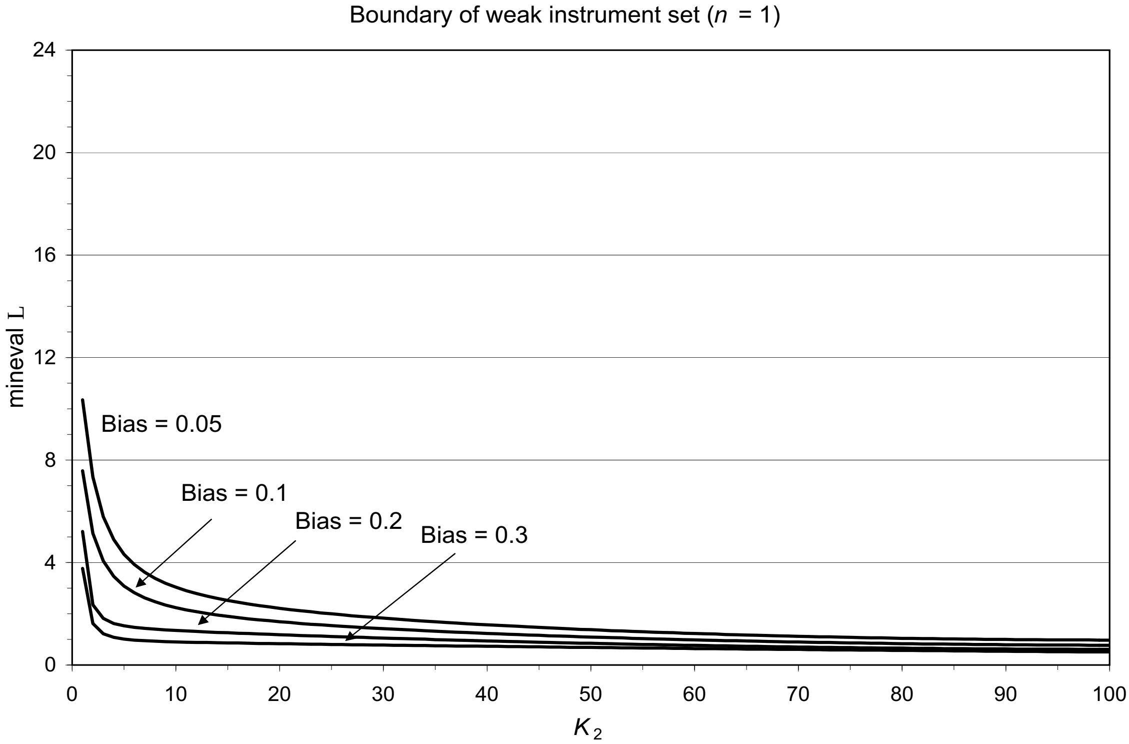
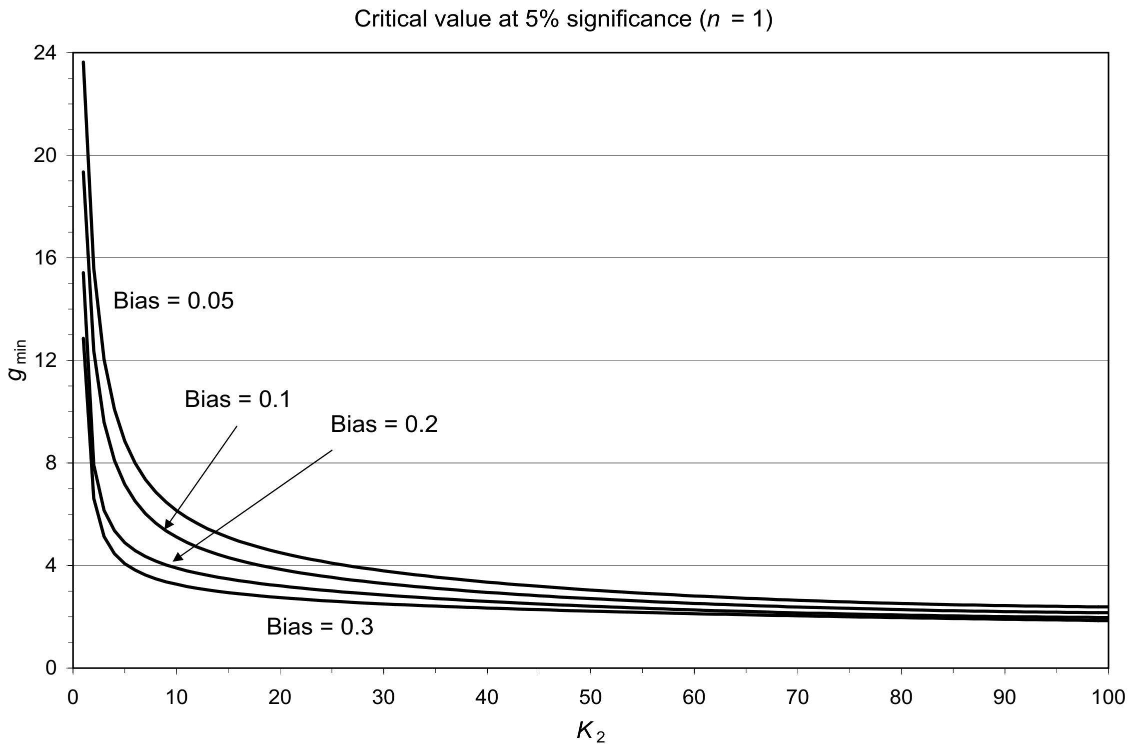
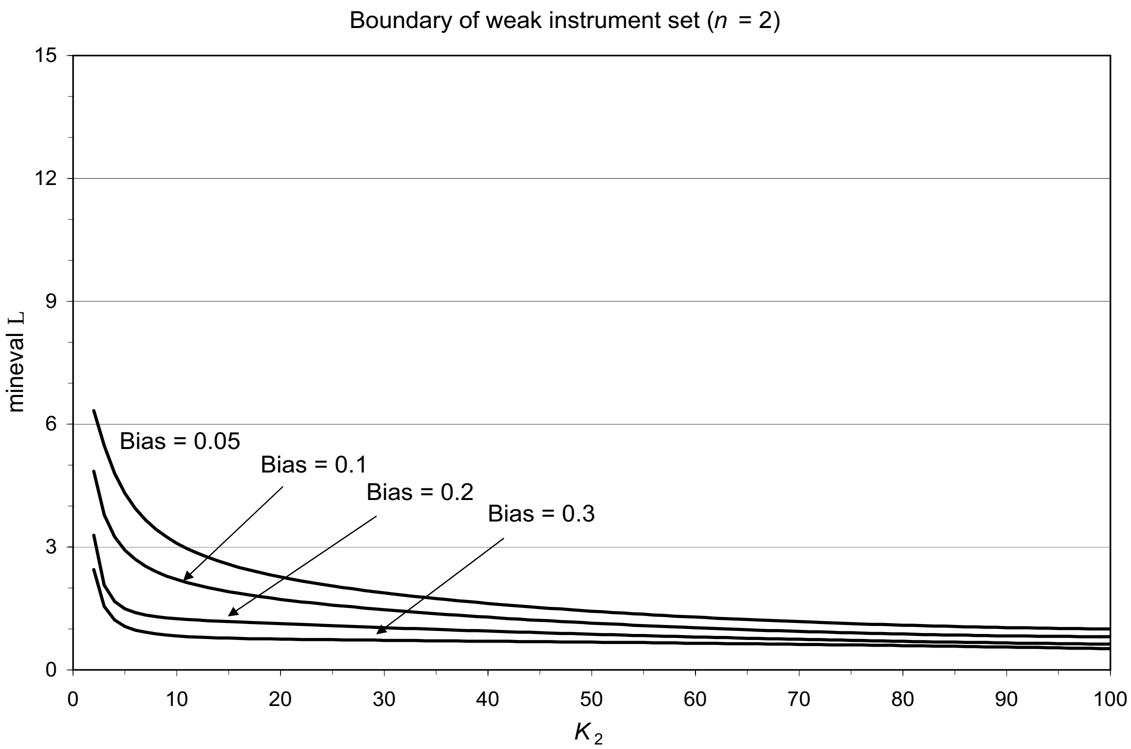

Critical value at 5% significance (n = 2)
g min
K 2
Size = 0.1
Size = 0.15
Size = 0.2
Size = 0.25
200
160
120
80
40
0
0
10
20
30
40
50
60
70
80
90
100

Figure 3: Weak Instrument Sets and Critical Values based on Bias of Fuller-k Relative to OLS






Boundary of weak instrument set (n = 2)
mineval L
Bias = 0.05
Bias = 0.1
Bias = 0.2
Bias = 0.3
K 2
0
3
6
9
12
15
0
10
20
30
40
50
60
70
80
90
100

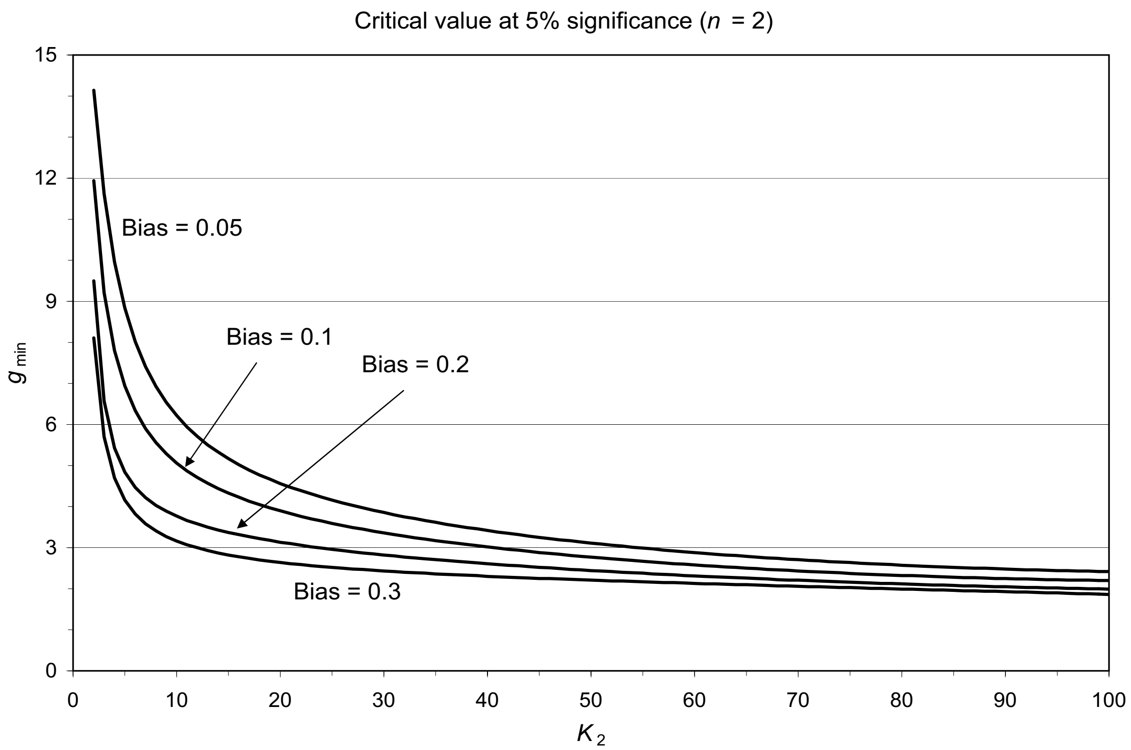
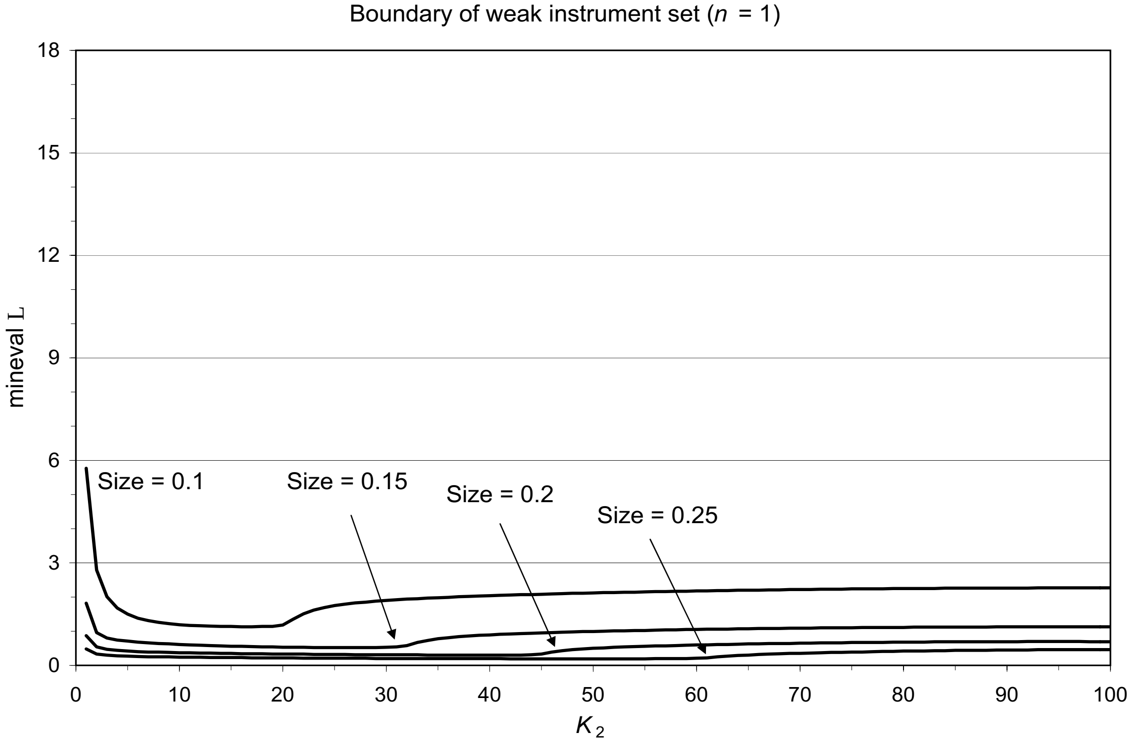
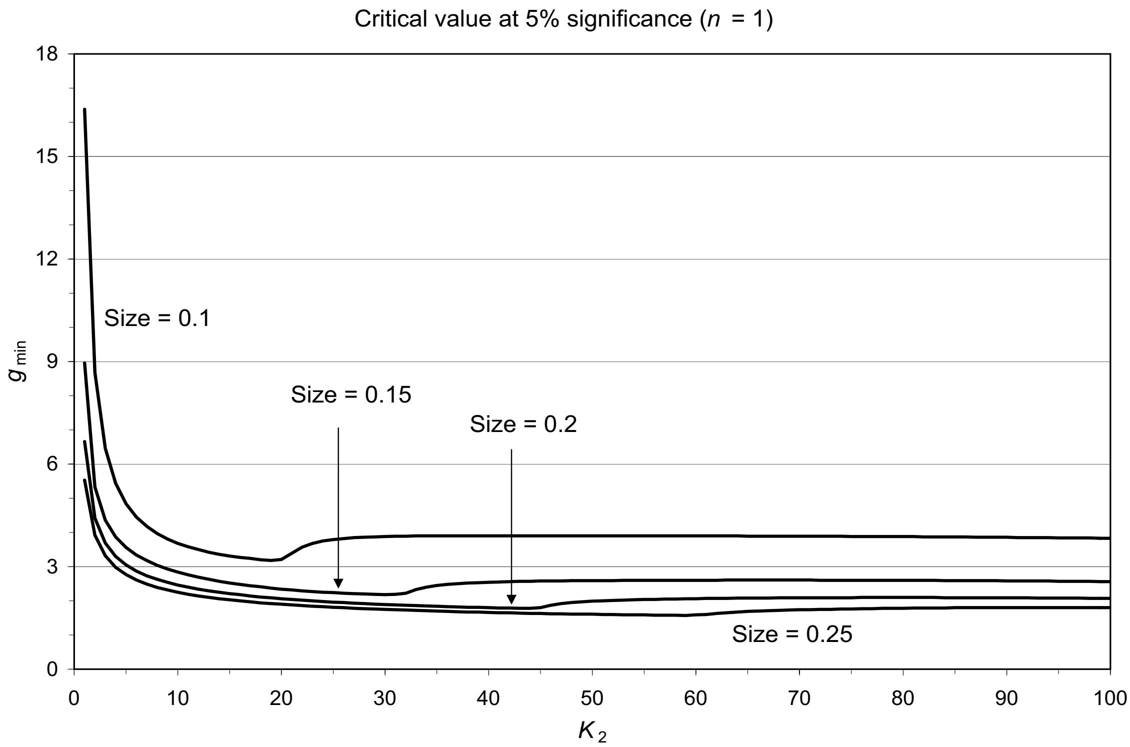
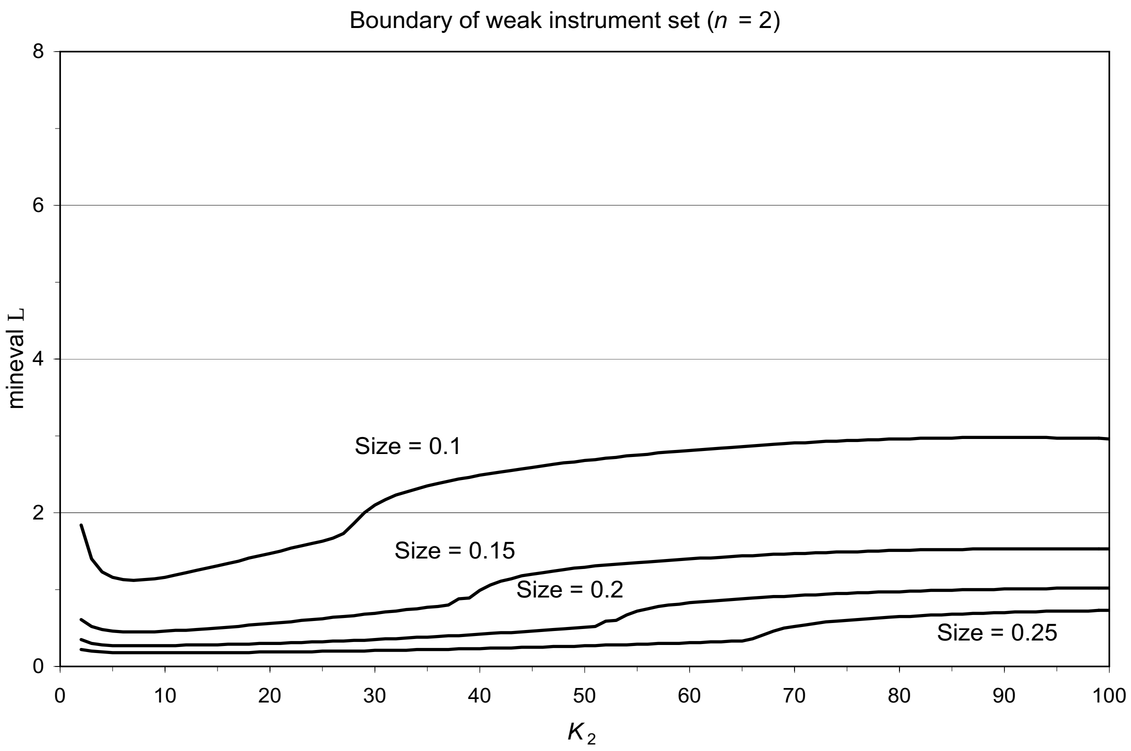

Critical value at 5% significance (n = 2)
g min
Bias = 0.05
Bias = 0.1
Bias = 0.2
Bias = 0.3
K 2
0
3
6
9
12
15
0
10
20
30
40
50
60
70
80
90
100

Figure 4: Weak Instrument Sets and Critical Values based on Size of LIML Wald Test






Boundary of weak instrument set (n = 2)
mineval L
K 2
Size = 0.1
Size = 0.15
Size = 0.2
Size = 0.25
0
2
4
6
8
0
10
20
30
40
50
60
70
80
90
100

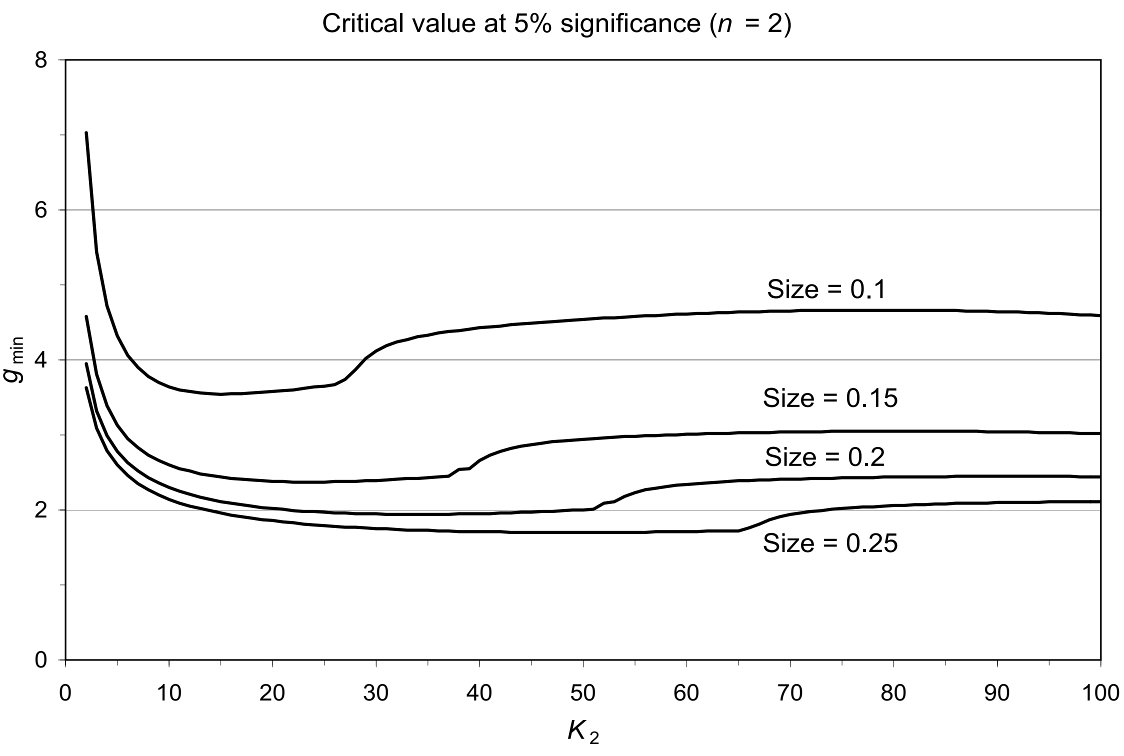
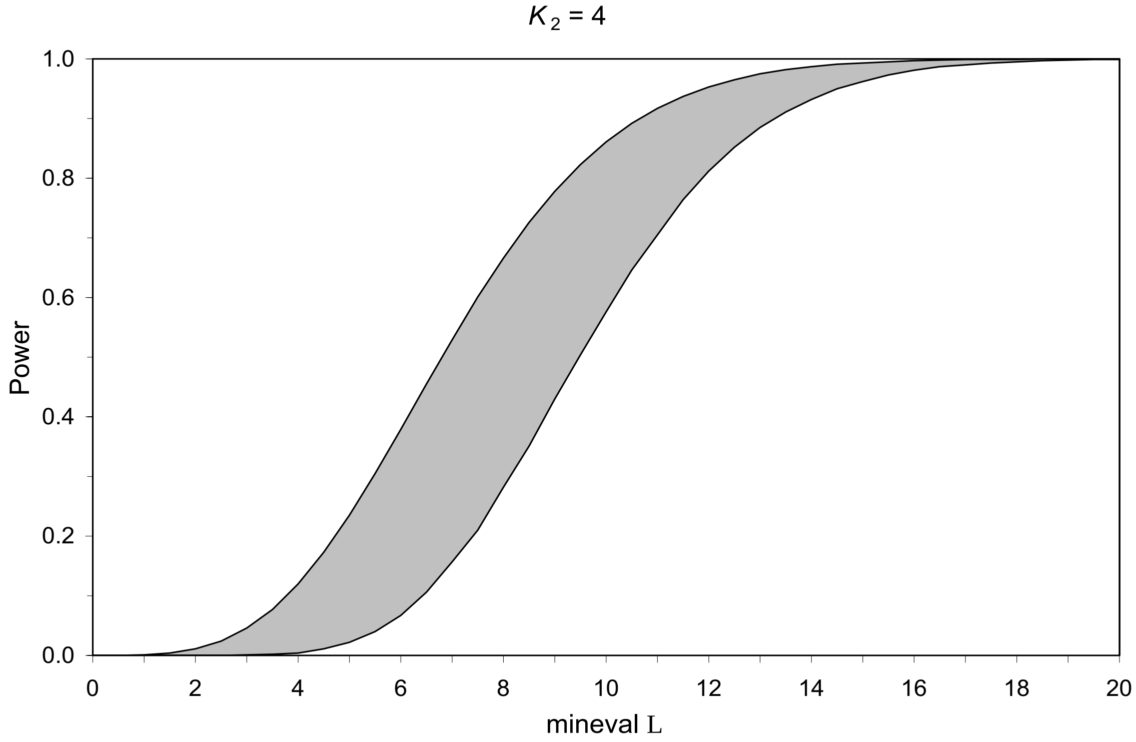
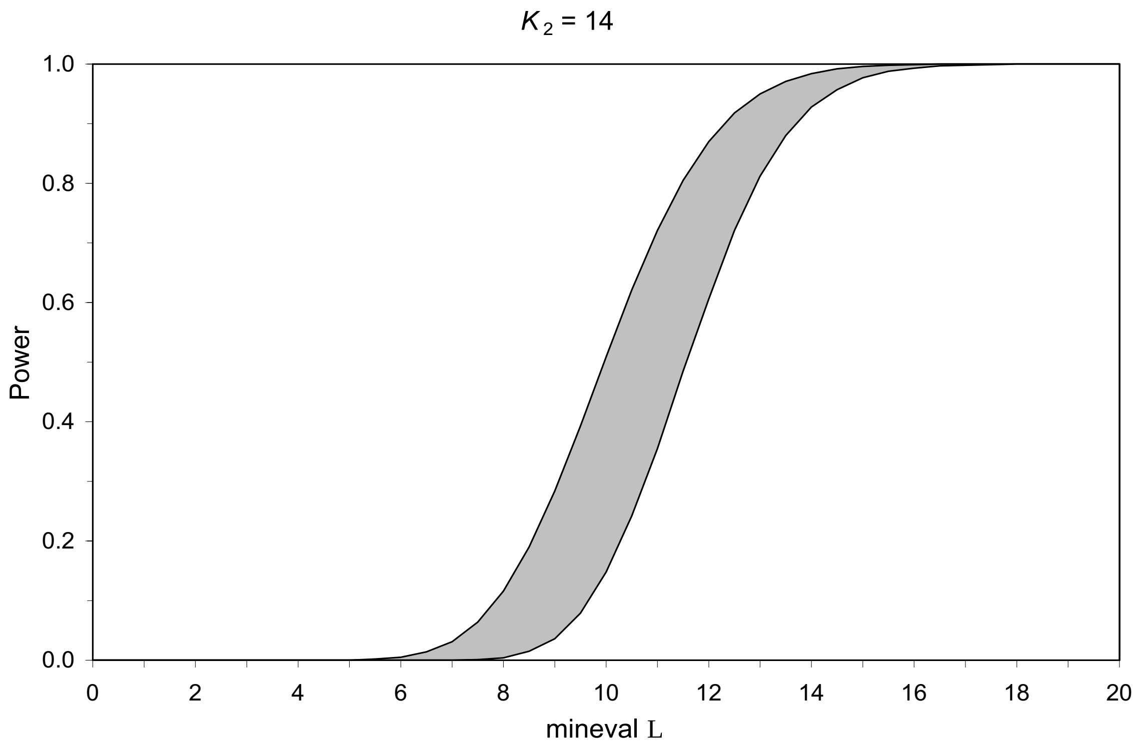
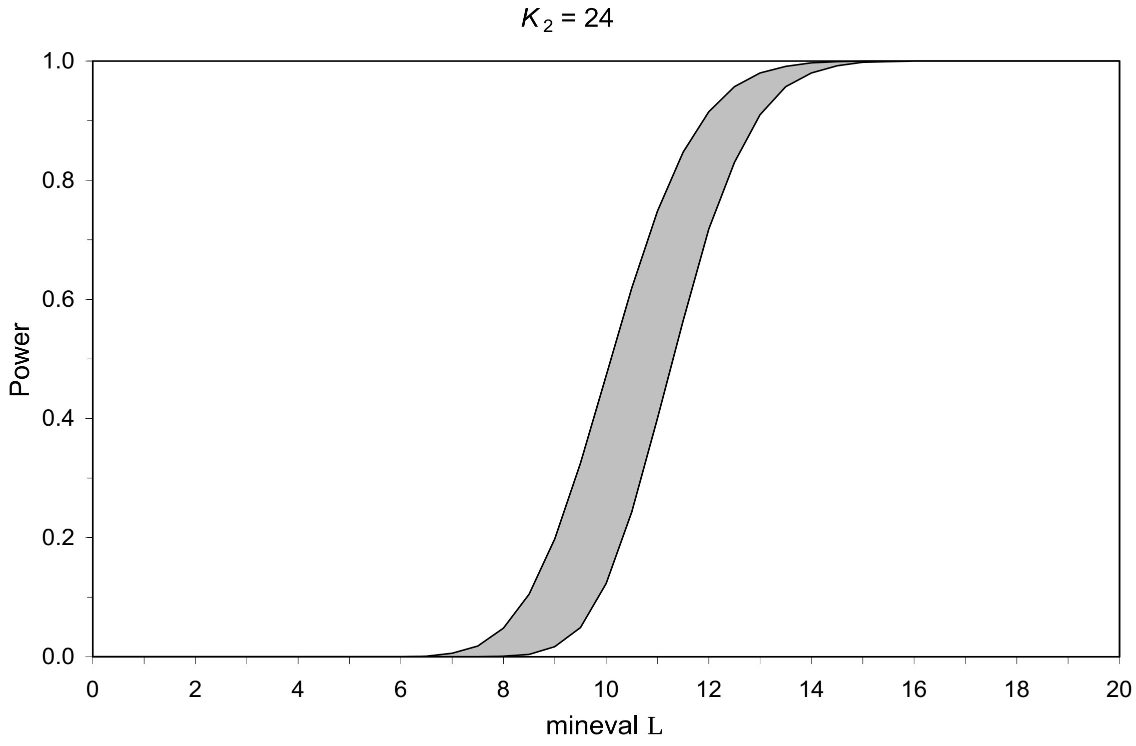

Critical value at 5% significance (n = 2)
g min
K 2
Size = 0.1
Size = 0.15
Size = 0.2
Size = 0.25
0
2
4
6
8
0
10
20
30
40
50
60
70
80
90
100

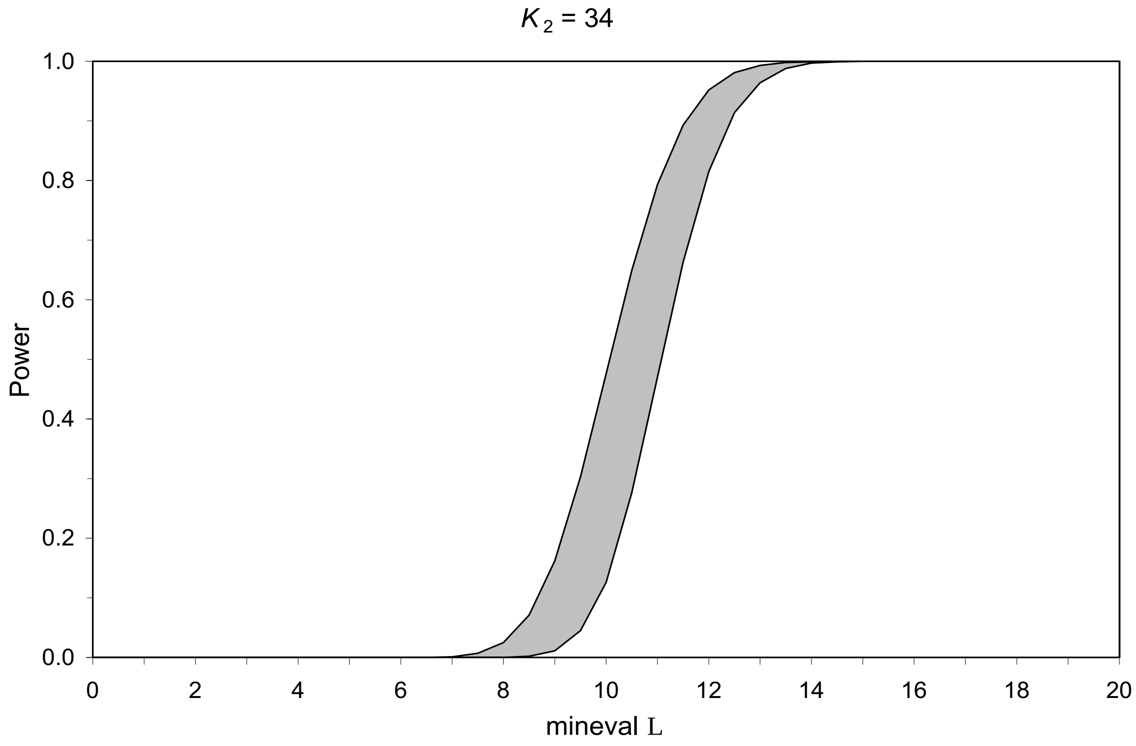
Figure 5: Power Function for TSLS Bias Test (Relative Bias = 0.1, $n = 2$)







K_2 = 24
Power
mineval L
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
12
14
16
18
20



K_2 = 34
Power
mineval L
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
12
14
16
18
20