CHAPTER 5

Testing for Weak Instruments in Linear IV Regression

James H. Stock and Motohiro Yogo

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 the given instruments are weak.

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 considers the single equation IV regression model

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

where $\mathbf{y}$ and $\mathbf{Y}$ are $T \times 1$ vectors of observations on the dependent variable and endogenous regressor, respectively, and $\mathbf{u}$ is a $T \times 1$ vector of i.i.d. $N(0, \sigma_{uu})$ errors. The reduced form equation for $\mathbf{Y}$ is

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

where $\mathbf{Z}$ is a $T \times K_2$ matrix of fixed, exogenous instrumental variables, $\mathbf{\Pi}$ is a $K_2 \times 1$ coefficient vector, and $\mathbf{V}$ is a $T \times 1$ vector of i.i.d. $N(0, \sigma_{VV})$ errors, where $\mathrm{corr}(u_t, V_t) = \rho$.

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

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

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

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 $\mathbf{\Pi}$, and thus μ^2, is unknown, how is an applied researcher to know whether μ^2 is in fact sufficiently small and that his or her instruments are weak?

This paper provides answers to these two questions. First, we develop precise, quantitative definitions of weak instruments for the general case of n endogenous regressors. 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 $\alpha = 5\%$. 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.

Second, given this quantitative definition of weak instrument sets, we 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]

This paper is part of a growing literature on detecting weak instruments, surveyed in Stock, Wright, and Yogo (2002) and Hahn and Hausman (2003). Cragg and Donald (1993) proposed a test of underidentification, which (as discussed earlier) is different from a test for weak instruments. Hall, Rudebusch, and Wilcox (1996), following the 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 et al. (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 10. 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 the 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.

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

2. THE IV REGRESSION MODEL, THE PROPOSED TEST STATISTIC, AND WEAK INSTRUMENT ASYMPTOTICS

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 $\mathbf{Y}$ and K_1 included exogenous regressors $\mathbf{X}$:

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

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

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

Let $\mathbf{X}_t = (X_{1t} \cdots X_{K_1 t})'$, $\mathbf{Z}_t = (Z_{1t} \cdots Z_{K_2 t})'$, $\mathbf{V}_t = (V_{1t} \cdots V_{nt})'$, and $\underline{\mathbf{Z}}_t = (\mathbf{X}_t'\,\mathbf{Z}_t')'$ denote the vectors of the tth observations on these variables. Also let $\boldsymbol{\Sigma}$ and $\mathbf{Q}$ denote the population second moment matrices,

$$E\left[\begin{pmatrix} u_t \\ V_{\mathrm{t}} \end{pmatrix}\begin{pmatrix} u_1 & \mathbf{V}_t' \end{pmatrix}\right] = \begin{bmatrix} \sigma_{uu} & \boldsymbol{\Sigma}_{u\mathbf{V}} \\ \boldsymbol{\Sigma}_{\mathbf{V}u} & \boldsymbol{\Sigma}_{\mathbf{VV}} \end{bmatrix} = \boldsymbol{\Sigma} \quad \text{and}$$

$$E(\underline{\mathbf{Z}}_t\underline{\mathbf{Z}}_t') = \begin{bmatrix} \mathbf{Q}_{\mathbf{XX}} & \mathbf{Q}_{\mathbf{XZ}} \\ \mathbf{Q}_{\mathbf{Z}X} & \mathbf{Q}_{\mathbf{ZZ}} \end{bmatrix} = \mathbf{Q}. \tag{2.3}$$

2.2. *k*-Class Estimators and Wald Statistics

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

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

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

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

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

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):

$$\text{TSLS:}\quad k = 1, \tag{2.6}$$

$$\text{LIML:}\quad k = \hat{k}_{\text{LIML}} \text{ is the smallest root of } \det(\underline{\mathbf{Y}}'\mathbf{M}_{\mathbf{X}}\underline{\mathbf{Y}} - k\underline{\mathbf{Y}}'\mathbf{M}_{\mathbf{Z}}\underline{\mathbf{Y}}) = 0, \tag{2.7}$$

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

$$\text{BTSLS:}\quad k = T/(T - K_2 + 2), \tag{2.9}$$

where det($\mathbf{A}$) is the determinant of the matrix $\mathbf{A}$. 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 $k = 1 + a(\hat{k}_{\text{LIML}} - 1) - c/(T - K_1 - K_2)$ for some constants a and c (Rothenberg 1984). For further discussion, see Donald and Newey (2001) and Stock et al. (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,

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

where $\hat{\mathbf{\Sigma}}_{\mathbf{VV}} = (\mathbf{Y}'\mathbf{M}_{\underline{\mathbf{Z}}}\mathbf{Y})/(T - K_1 - K_2)$.[2] The test statistic is the minimum eigenvalue of $\mathbf{G}_T$:

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

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 its eigenvalues are 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 $\mathbf{Z}$ is only weakly related to $\mathbf{Y}$, given $\mathbf{X}$. Specifically, weak instrument asymptotics are developed by modeling $\mathbf{\Pi}$ as local to zero:

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

Assumption $\mathbf{L}_{\mathbf{\Pi}}$. $\mathbf{\Pi} = \mathbf{\Pi}_T = \mathbf{C}/\sqrt{T}$, *where* $\mathbf{C}$ *is a fixed* $K_2 \times n$ *matrix.*

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

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

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

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 $\boldsymbol{\rho} = \mathbf{\Sigma}_{\mathbf{VV}}^{-1/2\prime}\mathbf{\Sigma}_{\mathbf{V}u}\sigma_{uu}^{-1/2}$, $\boldsymbol{\theta} = \mathbf{\Sigma}_{\mathbf{VV}}^{-1}\mathbf{\Sigma}_{\mathbf{V}u} = \sigma_{uu}^{1/2}\mathbf{\Sigma}_{\mathbf{VV}}^{-1/2}\boldsymbol{\rho}$, $\boldsymbol{\lambda} = \mathbf{\Omega}^{1/2}\mathbf{C}\mathbf{\Sigma}_{\mathbf{VV}}^{-1/2}$, $\mathbf{\Lambda} = \boldsymbol{\lambda}'\boldsymbol{\lambda}/K_2$, and $\mathbf{\Omega} = \mathbf{Q}_{\mathbf{ZZ}} - \mathbf{Q}_{\mathbf{ZX}}\mathbf{Q}_{\mathbf{XX}}^{-1}\mathbf{Q}_{\mathbf{XZ}}$. Note that $\boldsymbol{\rho}'\boldsymbol{\rho} \le 1$. Define the $K_2 \times 1$ and $K_2 \times n$ random variables $\mathbf{z}_u = \mathbf{\Omega}^{-1/2\prime}(\mathbf{\Psi}_{\mathbf{Z}u} - \mathbf{Q}_{\mathbf{ZX}}\mathbf{Q}_{\mathbf{XX}}^{-1}\mathbf{\Psi}_{\mathbf{X}u})\sigma_{uu}^{-1/2}$ and $\mathbf{z}_{\mathbf{V}} = \mathbf{\Omega}^{-1/2\prime}(\mathbf{\Psi}_{\mathbf{ZV}} - \mathbf{Q}_{\mathbf{ZX}}\mathbf{Q}_{\mathbf{XX}}^{-1}\mathbf{\Psi}_{\mathbf{XV}})\mathbf{\Sigma}_{\mathbf{VV}}^{-1/2}$, so

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

Also let

$$\boldsymbol{\nu}_1 = (\boldsymbol{\lambda} + \mathbf{z_V})'(\boldsymbol{\lambda} + \mathbf{z_V}) \quad \text{and} \tag{2.13}$$

$$\boldsymbol{\nu}_2 = (\boldsymbol{\lambda} + \mathbf{z_V})'\mathbf{z_u}. \tag{2.14}$$

2.5. Selected Weak Instrument Asymptotic Representations

We first summarize some results from Staiger and Stock (1997).

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,

$$\hat{\boldsymbol{\beta}}(k) - \boldsymbol{\beta} \xrightarrow{d} \sigma_{uu}^{1/2}\mathbf{\Sigma}_{\mathbf{VV}}^{-1/2}(\boldsymbol{\nu}_1 - \kappa\mathbf{I}_n)^{-1}(\boldsymbol{\nu}_2 - \kappa\boldsymbol{\rho}) \quad \text{and} \tag{2.15}$$

$$W(k) \xrightarrow{d} \frac{(\boldsymbol{\nu}_2 - \kappa\boldsymbol{\rho})'(\boldsymbol{\nu}_1 - \kappa\mathbf{I}_n)^{-1}(\boldsymbol{\nu}_2 - \kappa\boldsymbol{\rho})}{n[1 - 2\boldsymbol{\rho}'(\boldsymbol{\nu}_1 - \kappa\mathbf{I}_n)^{-1}(\boldsymbol{\nu}_2 - \kappa\boldsymbol{\rho}) + (\boldsymbol{\nu}_2 - \kappa\boldsymbol{\rho})'(\boldsymbol{\nu}_1 - \kappa\mathbf{I}_n)^{-2}(\boldsymbol{\nu}_2 - \kappa\boldsymbol{\rho})]}, \tag{2.16}$$

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 Ξ be the $(n+1)\times(n+1)$ matrix, $\Xi = [z_u(\boldsymbol{\lambda}+\mathbf{z_V})]'[z_u(\boldsymbol{\lambda}+\mathbf{z_V})]$. Then the limits in (2.15) and (2.16) hold with

$$\text{TSLS:} \quad \kappa = 0, \tag{2.17}$$

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

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

$$\text{BTSLS:} \quad \kappa = K_2 - 2. \tag{2.20}$$

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

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

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

Weak Instrument Asymptotic Representations: The Cragg–Donald Statistic. Under the weak instrument asymptotic assumptions, the matrix $\mathbf{G}_T$ in (2.10) and the Cragg–Donald statistic (2.11) have the limiting distributions

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

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

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

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

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 $\mathbf{Y}^{\perp}$ so that they have unit standard deviation and are orthogonal or, equivalently, to rotate $\boldsymbol{\beta}$ by $\boldsymbol{\Sigma}^{1/2}_{\mathbf{Y}^{\perp}\mathbf{Y}^{\perp}}$, where $\boldsymbol{\Sigma}_{\mathbf{Y}^{\perp}\mathbf{Y}^{\perp}} = \text{plim}(\mathbf{Y}^{\perp\prime}\mathbf{Y}^{\perp}/T)$. In these standardized units, the squared bias of an IV estimator, which we generically denote by $\hat{\boldsymbol{\beta}}^{\text{IV}}$, is $(E\hat{\boldsymbol{\beta}}^{\text{IV}} - \boldsymbol{\beta})'\boldsymbol{\Sigma}_{\mathbf{Y}^{\perp}\mathbf{Y}^{\perp}}(E\hat{\boldsymbol{\beta}}^{\text{IV}} - \boldsymbol{\beta})$. As our measure of bias, we therefore consider the relative squared bias of the candidate IV estimator $\hat{\boldsymbol{\beta}}^{\text{IV}}$, relative to the squared bias of the OLS estimator

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

If $n = 1$, then the scaling matrix in (3.1) drops out and the expression simplifies to $B_T = |E\hat{\boldsymbol{\beta}}^{\text{IV}} - \boldsymbol{\beta}|/|E\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}|$. 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, $E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to \boldsymbol{\theta} = \sigma_{uu}^{1/2}\boldsymbol{\Sigma}_{\mathbf{VV}}^{-1/2}\boldsymbol{\rho}$ and $\boldsymbol{\Sigma}_{\mathbf{Y}^{\perp}\mathbf{Y}^{\perp}} \to \boldsymbol{\Sigma}_{\mathbf{VV}}$, so that 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 for $\boldsymbol{\rho}'\boldsymbol{\rho} > 0$, the square of the asymptotic relative bias is

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

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

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

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 $\mathcal{Z} = \{\mathbf{\Pi}, \mathbf{\Sigma_{VV}}, \mathbf{\Omega}\}$. The relative bias definition of weak instruments is

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

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, we will show that the maximal TSLS relative bias is also its maximal absolute bias in standardized units, so that for TSLS the maximal relative and absolute bias can be treated interchangeably. We return to this point 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 $\boldsymbol{\beta}$. 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 $\hat{\boldsymbol{\beta}}^{\text{IV}}$. 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

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

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 $\boldsymbol{\rho}$. As in the definitions of the bias-based weak instrument set, we consider the worst-case limiting rejection rate

$$R^{\max} = \max_{\rho:\rho'\rho\leq 1} R, \quad \text{where } R = \lim{}_{T\to\infty} R_T. \tag{3.6}$$

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

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

For example, if $\alpha = .05$ then a researcher might consider it acceptable if the worst-case size is $r = 0.10$.

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 $\mathbf{\Lambda}$.

3.3.1. Weak Instrument Set Based on TSLS bias

Under weak instrument asymptotics,

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

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

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

Because $\mathbf{h}$ depends on $\boldsymbol{\lambda}$ but not on $\boldsymbol{\rho}$, by (3.8) we have $B^{\max,\text{TSLS}} = [\text{maxeval}(\mathbf{h}'\mathbf{h})]^{1/2}$, where maxeval($\mathbf{A}$) denotes the maximum eigenvalue of the matrix $\mathbf{A}$. By applying the singular value decomposition to $\boldsymbol{\lambda}$, it is further possible to show that the maximum eigenvalue of $\mathbf{h}'\mathbf{h}$ depends only on K_2, n, and the eigenvalues of $\boldsymbol{\lambda}'\boldsymbol{\lambda}/K_2 = \mathbf{\Lambda}$. It follows that, for a given K_2 and n, the maximum TSLS asymptotic bias is a function only of the eigenvalues of $\mathbf{\Lambda}$.

When the number of instruments is large, it is possible to show further that the maximum TSLS asymptotic bias is a decreasing function of the minimum eigenvalue of $\mathbf{\Lambda}$. Specifically, consider sequences of K_2 and T such that $K_2 \to \infty$ and $T \to \infty$ jointly, subject to $K_2^4/T \to 0$, where $\mathbf{\Lambda}$ (which in general depends on K_2) is held constant as $K_2 \to \infty$.[3] We write this joint limit as (K_2, $T \to \infty$) and, following Stock and Yogo (2005), we refer to it as representing "many weak instruments." It follows from (3.9) and Theorem 4.1(a) of Stock and Yogo (2003) that the many weak instrument limit of B_T^{TSLS} is

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

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

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

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

Our formal justification for the simplification that $\mathcal{W}_{\text{bias,TSLS}}$ depends only on the smallest eigenvalue of $\mathbf{\Lambda}$, rather than on all its eigenvalues, rests on

[3] In Stock and Yogo (2005), the assumption that $\mathbf{\Lambda}$ is constant is generalized to consider sequences of $\mathbf{\Lambda}$, indexed by K_2, that have a finite limit $\mathbf{\Lambda}_\infty$ as $K_2 \to \infty$.

the many weak instrument asymptotic result (3.10). Numerical analysis for $n = 2$ suggests, however, that $B^{\max,\mathrm{TSLS}}$ 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]

3.3.2. *Reinterpretation in Terms of Absolute Bias*

Although $B^{\max}$ was defined as maximal bias relative to OLS, for TSLS $B^{\max}$ is also the maximal absolute bias in standardized units. The numerator of (3.9) is evidently maximized when $\boldsymbol{\rho}'\boldsymbol{\rho} = 1$. Thus, for TSLS, (3.2) can be restated as $(B^{\max})^2 = \sigma_{uu}^{-1} \max_{\rho:\rho'\rho=1} \lim_{T\to\infty}(E\hat{\boldsymbol{\beta}}^{\mathrm{TSLS}} - \boldsymbol{\beta})'\Sigma_{Y^\perp Y^\perp}(E\hat{\boldsymbol{\beta}}^{\mathrm{TSLS}} - \boldsymbol{\beta})$. But $(E\hat{\boldsymbol{\beta}}^{\mathrm{TSLS}} - \boldsymbol{\beta})'\Sigma_{Y^\perp Y^\perp}(E\hat{\boldsymbol{\beta}}^{\mathrm{TSLS}} - \boldsymbol{\beta})$ is the squared bias of $\hat{\boldsymbol{\beta}}^{\mathrm{TSLS}}$, 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 $\sigma_{uu}^{-1/2}\boldsymbol{\Sigma}_{\mathbf{Y}^\perp\mathbf{Y}^\perp}^{1/2}$.

3.3.3. *Weak Instrument Set Based on TSLS Size*

For TSLS, it follows from (2.22) that the worst-case asymptotic size is

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

$R^{\max,\mathrm{TSLS}}$, and consequently $\mathcal{W}_{\text{size,TSLS}}$, 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 $B^{\max,\mathrm{TSLS}}$).

When the number of instruments is large, the Wald statistic is maximized when $\boldsymbol{\rho}'\boldsymbol{\rho} = 1$ and is an increasing function of the eigenvalues of $\boldsymbol{\Lambda}$. Specifically, it is shown in Stock and Yogo (2005), Theorem 4.1(a), that the many weak instrument limit of the TSLS Wald statistic, divided by K_2, is

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

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

[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 $\mathcal{W}_{\text{bias}}$ is potentially conservative when K_2 is small. This comment applies to size-based sets as well.

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

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

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} is $O_p(K_2)$, so that 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 $R^{\max,\mathrm{TSLS}}$ 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

$$\mathcal{W}_{\mathrm{size,TSLS}} = \{\mathcal{Z}\colon \mathrm{mineval}(\boldsymbol{\Lambda}) \leq \ell_{\mathrm{size,TSLS}}(r; K_2, n, \alpha)\}, \tag{3.15}$$

where $\ell_{\mathrm{size,TSLS}}(r; K_2, n, \alpha)$ 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 can also 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 $n = 2$, however, indicates 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.[5] 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).

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.

[5] It appears that there is some nonmonotonicity 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 so that this nonmonotonicity does not affect the boundary eigenvalues.

3.4.1. *LIML and Fuller-k*

As shown in Stock and Yogo (2003), Theorem 2(c), the LIML and Fuller-k estimators and their Wald statistics have the many weak instrument asymptotic distributions

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

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

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 zero.

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

$$\begin{aligned} R^{\max,\text{LIML}} &= \max_{\rho:\rho'\rho\le 1} \lim_{(K_2,T\to\infty)} \Pr\left[W^{\text{LIML}} > \chi^2_{1;\alpha}\right] \\ &= Pr\left[\chi_1^2 > \frac{\boldsymbol{\Lambda}}{\boldsymbol{\Lambda}+1}\chi^2_{1;\alpha}\right], \end{aligned} \quad (3.18)$$

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

3.4.2. *BTSLS*

The many weak instrument asymptotic distributions of the BTSLS estimator and Wald statistic are (Stock and Yogo 2003, Theorem 2(b))

$$\sqrt{K_2}(\hat{\boldsymbol{\beta}}^{\text{BTSLS}} - \boldsymbol{\beta}) \xrightarrow{d} N\left[0, \sigma_{uu}\boldsymbol{\Sigma}_{\mathbf{VV}}^{-1/2}\boldsymbol{\Lambda}^{-1}(\boldsymbol{\Lambda} + \mathbf{I}_n + \boldsymbol{\rho}\boldsymbol{\rho}')\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}_{\mathbf{VV}}^{-1/2\prime}\right], \quad (3.19)$$

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

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

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

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 $\mathbf{\Lambda}$ 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 several k-class statistics. Here, we focus on TSLS bias and size, Fuller-k (with $c = 1$ in (2.8)) bias, and LIML size. Additional results are reported in Stock et al. (2002). 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 $\mathbf{\Lambda}$ from 0 to 30 for $K_2 = n + 2, \ldots, 100$, using 20,000 Monte Carlo draws. Computing the maximum TSLS bias entails computing $\mathbf{h}$, defined following (3.8), by Monte Carlo simulation, given n, K_2, and then computing the maximum bias $[\text{maxeval}(\mathbf{h}'\mathbf{h})]^{1/2}$. 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 $\boldsymbol{\rho}'\boldsymbol{\rho} = 1$, and 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 $\mathbf{\Lambda} = \ell\mathbf{I}_n$. 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 $\mathbf{\Lambda}$ that constitute the boundaries of $\mathcal{W}_{\text{bias,TSLS}}$, $\mathcal{W}_{\text{size,TSLS}}$, $\mathcal{W}_{\text{bias,Fuller-k}}$, and $\mathcal{W}_{\text{size,LIML}}$ are plotted, respectively, in the top panels of Figures 5.1–5.4 for various cutoff values b and r. The figures show the boundary eigenvalues for $n = 1$; the corresponding plots of boundary eigenvalues for $n = 2$ are qualitatively, and in many cases quantitatively, similar. First consider the regions based on bias. The boundary of $\mathcal{W}_{\text{bias,TSLS}}$ is essentially flat in K_2 for K_2 sufficiently large. The boundary of the relative bias region for $b = 0.1$ (10% bias) asymptotes to approximately 8. In contrast, the boundary of the bias region for Fuller-k tends to zero as the number of instruments increases, 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 $\mathcal{W}_{\text{size,TSLS}}$ depends strongly on K_2; 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

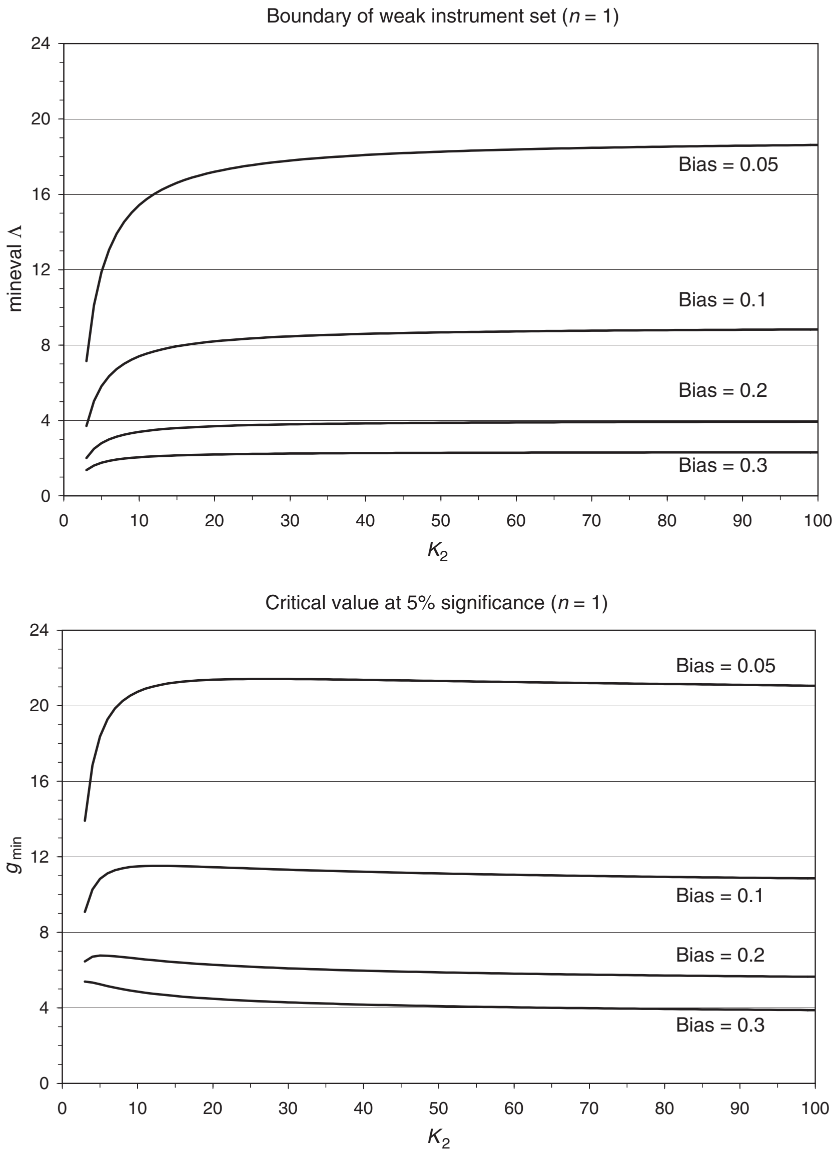

Figure 5.1. Weak instrument sets and critical values based on bias of TSLS relative to OLS.

with 10 instruments the minimum eigenvalue boundary is approximately 20 for $n = 1$ (it is approximately 16 for $n = 2$). In contrast, the boundary of $\mathcal{W}_{\text{size,LIML}}$ decreases with K_2 for both $n = 1$ and $n = 2$. Comparing these two plots shows that tests based on LIML are far more robust to weak instruments than tests based on TSLS.

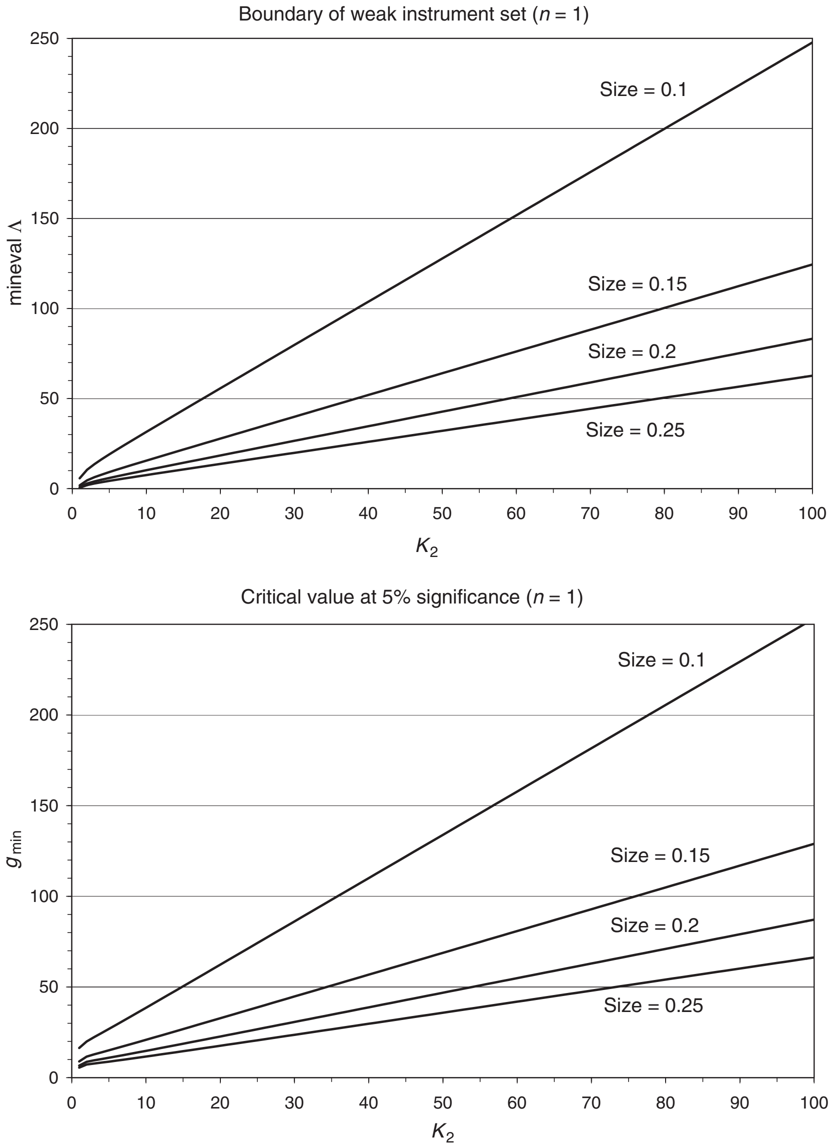

Figure 5.2. Weak instrument sets and critical values based on size of TSLS Wald test.

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 $g_{\min}$.

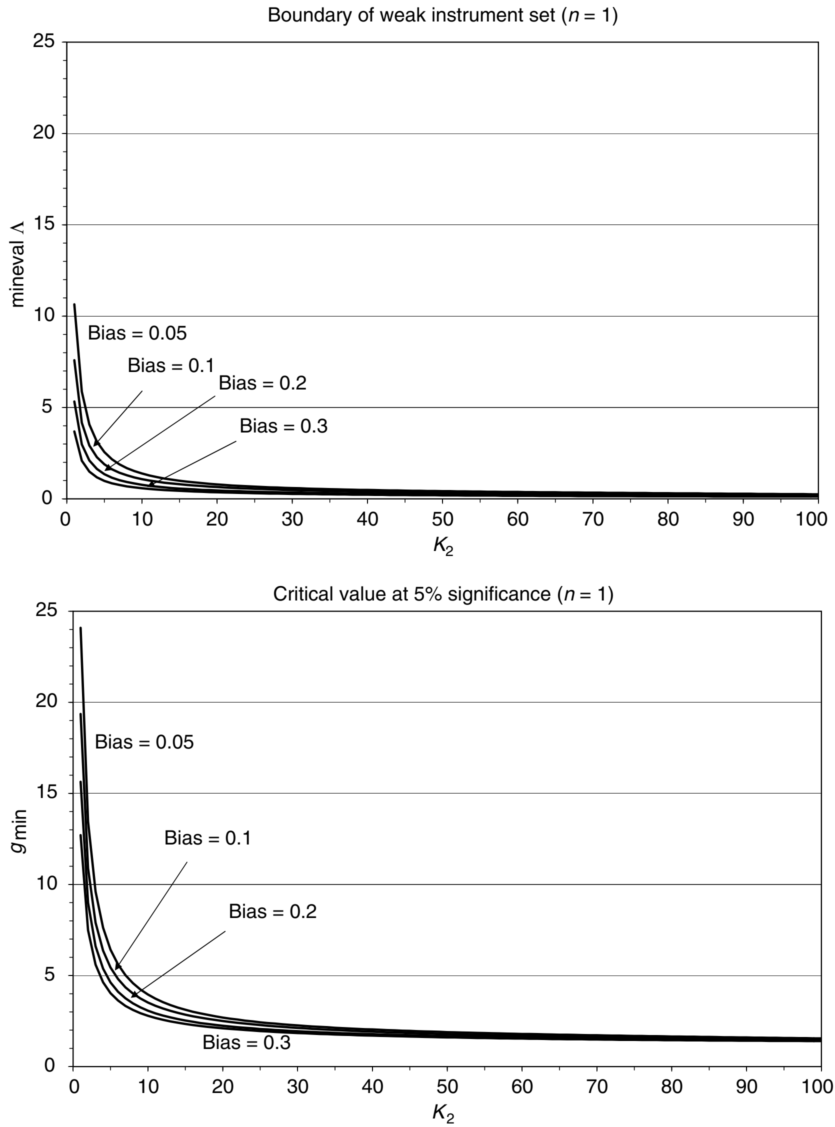

Figure 5.3. Weak instrument sets and critical values based on bias of Fuller-k relative to OLS.

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

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

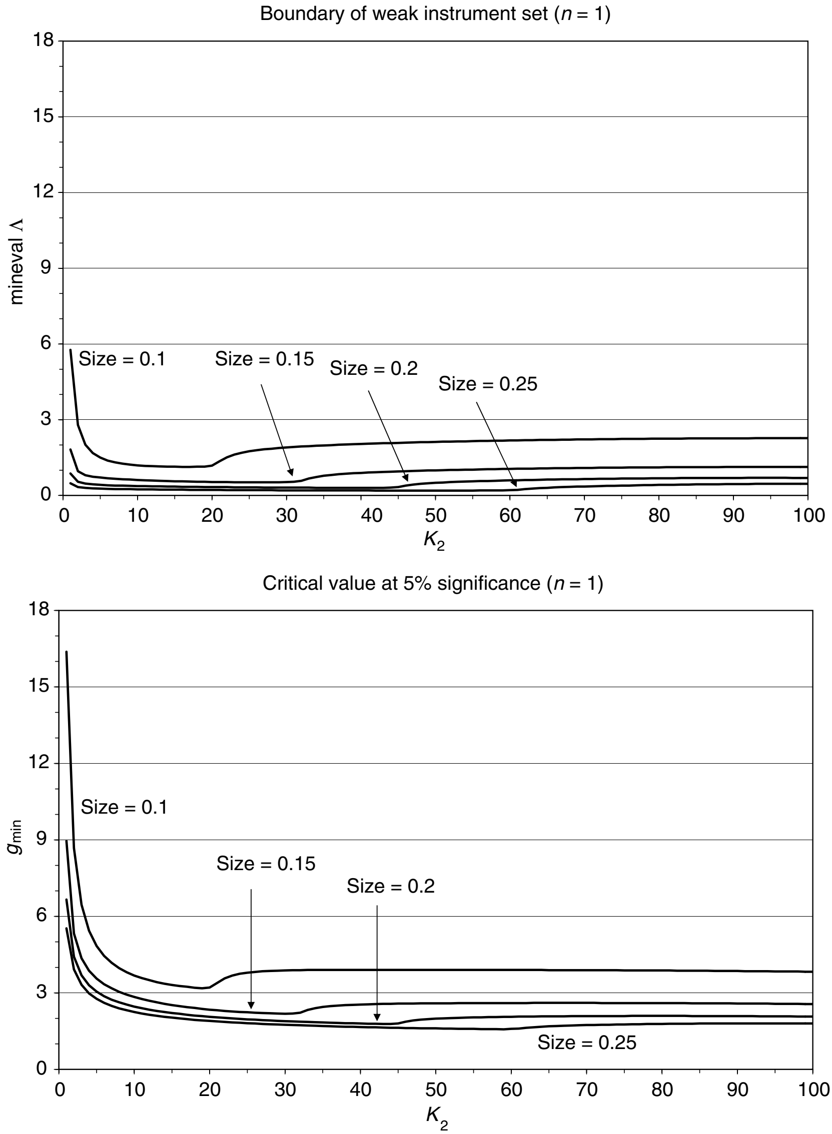

Figure 5.4. Weak instrument sets and critical values based on size of LIML Wald test.

and noncentrality matrix $K_2\mathbf{\Lambda}$; that is,

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

The joint pdf for the n eigenvalues of a noncentral Wishart has an infinite series expansion in terms of zonal polynomials (Muirhead 1978). This joint pdf depends on all the eigenvalues of $\mathbf{\Lambda}$, 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 $\mathbf{\Lambda}$.

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 $\mathbf{\Lambda}$. This latter consideration is especially important because in practice these eigenvalues are unknown nuisance parameters, and 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 4.1. $\Pr[\textit{mineval}(\mathrm{W}_n(k, \mathbf{I}_n, \mathbf{A})) \geq x] \leq \Pr[\chi_k^2(\textit{mineval}(\mathbf{A})) \geq x]$, *where* $\chi_k^2(a)$ *denotes a noncentral chi-squared random variable with noncentrality parameter* a.

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

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

$$\begin{aligned} \Pr[g_{\min} \geq x] &\to \Pr[\text{mineval}(\boldsymbol{\nu}_1/K_2) \geq x] \\ &\leq \Pr\left[\frac{\chi_{K_2}^2(\text{mineval}(K_2\mathbf{\Lambda}))}{K_2} \geq x\right]. \end{aligned} \tag{4.2}$$

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 $\mathbf{\Lambda}$. 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]

[6] The critical values based on Proposition 4.1 can be quite conservative when all the eigenvalues of $\mathbf{\Lambda}$ 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($\mathbf{\Lambda}$) $= 3.08$, and the critical value for a 5% test with $b = 0.1$ 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 $\mathbf{\Lambda}$ 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 ranges from 3.3% to 4.1% for the nominal 5% weak instrument test based on TSLS bias with $b = 0.1$.

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 $100\delta\%$. The null hypothesis is that the instruments are weak, and the alternative is that they are not:

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

The test procedure is

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

where $d_{\text{bias,TSLS}}(b; K_2, n, \delta) = K_2^{-1}\chi^2_{K_2,1-\delta}(K_2\ell_{\text{bias,TSLS}}(b; K_2, n))$, where $\chi^2_{K_2,1-\delta}(m)$ is the $100(1-\delta)\%$ percentile of the noncentral chi-squared distribution with K_2 degrees of freedom and noncentrality parameter m, and the function $\ell_{\text{bias,TSLS}}$ 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:

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

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, $\ell_{\text{size,TSLS}}(r; K_2, n, \alpha)$ (or, for LIML, $\ell_{\text{size,LIML}}(r; K_2, n, \alpha)$).

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 in Section 3.5.

Critical values are tabulated in Tables 5.1–5.4 for the weak instrument tests based on TSLS bias,TSLS size, Fuller-*k* bias, and LIML size, respectively, for one and two included endogenous variables (and three 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 5.1–5.4. The critical value plots are qualitatively similar to the corresponding boundary eigenvalue plots, except of course that 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

Table 5.1. *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 5.2. *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 $\boldsymbol{\beta} = \boldsymbol{\beta}_0$.

more robust is the procedure. For discussion and comparisons of TSLS, BTSLS, Fuller-k, JIVE, and LIML, see Stock et al. (2002, Section 6).

4.3.1. *Comparison to the Staiger–Stock 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 10. They motivated

Table 5.3. *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	24.09	19.36	15.64	12.71				
2	13.46	10.89	9.00	7.49	15.50	12.55	9.72	8.03
3	9.61	7.90	6.61	5.60	10.83	8.96	7.18	6.15
4	7.63	6.37	5.38	4.63	8.53	7.15	5.85	5.10
5	6.42	5.44	4.62	4.03	7.16	6.07	5.04	4.44
6	5.61	4.81	4.11	3.63	6.24	5.34	4.48	3.98
7	5.02	4.35	3.75	3.33	5.59	4.82	4.08	3.65
8	4.58	4.01	3.47	3.11	5.10	4.43	3.77	3.39
9	4.23	3.74	3.25	2.93	4.71	4.12	3.53	3.19
10	3.96	3.52	3.07	2.79	4.41	3.87	3.33	3.02
11	3.73	3.34	2.92	2.67	4.15	3.67	3.17	2.88
12	3.54	3.19	2.80	2.57	3.94	3.49	3.04	2.77
13	3.38	3.06	2.70	2.48	3.76	3.35	2.92	2.67
14	3.24	2.95	2.61	2.41	3.60	3.22	2.82	2.58
15	3.12	2.85	2.53	2.34	3.47	3.11	2.73	2.51
16	3.01	2.76	2.46	2.28	3.35	3.01	2.65	2.44
17	2.92	2.69	2.39	2.23	3.24	2.92	2.58	2.38
18	2.84	2.62	2.34	2.18	3.15	2.84	2.52	2.33
19	2.76	2.56	2.29	2.14	3.06	2.77	2.46	2.28
20	2.69	2.50	2.24	2.10	2.98	2.71	2.41	2.23
21	2.63	2.45	2.20	2.07	2.91	2.65	2.36	2.19
22	2.58	2.40	2.16	2.04	2.85	2.60	2.32	2.16
23	2.52	2.36	2.13	2.01	2.79	2.55	2.28	2.12
24	2.48	2.32	2.10	1.98	2.73	2.50	2.24	2.09
25	2.43	2.28	2.06	1.95	2.68	2.46	2.21	2.06
26	2.39	2.24	2.04	1.93	2.63	2.42	2.18	2.03
27	2.36	2.21	2.01	1.90	2.59	2.38	2.15	2.01
28	2.32	2.18	1.99	1.88	2.55	2.35	2.12	1.98
29	2.29	2.15	1.96	1.86	2.51	2.31	2.09	1.96
30	2.26	2.12	1.94	1.84	2.47	2.28	2.07	1.94

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).

this suggestion based on the relative bias of TSLS. Because the 5% critical value for the relative bias weak instrument test with $b = 0.1$ 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.

Table 5.4. *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

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 $\boldsymbol{\beta} = \boldsymbol{\beta}_0$.

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 that 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 the 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 $\mathbf{\Lambda}$. 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 $\mathbf{\Lambda}$. It is bounded above by (4.2). On the basis of 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 $\mathrm{W}_n(K_2, \mathbf{I}_n, \text{mineval}(K_2\mathbf{\Lambda})\mathbf{I}_n)/K_2$. These two bounding distributions are used to bound the distribution of $g_{\min}$ as a function of mineval($\mathbf{\Lambda}$).

The bounds on the asymptotic rejection rate of the test (4.4) (based on TSLS maximum relative bias) are plotted in Figure 5.5 for $b = 0.1$ and $n = 2$. The value of the horizontal axis (the minimum eigenvalue) at which the upper rejection rate curve equals 5% is $\ell_{\text{bias}}(.1; K_2, 2)$. Evidently, as the minimum eigenvalue increases, so does the rejection rate. The rejection curve becomes steeper as K_2 increases. The bounding distributions give a fairly tight range for the actual power function, which depends on all the eigenvalues of $\mathbf{\Lambda}$.

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 5.5 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 5.5 bound the asymptotic probability of deciding that the instruments are strong. Evidently, for values of mineval($\mathbf{\Lambda}$) much below the weak instrument region boundary, the probability of correctly concluding that the instruments are weak is effectively equal to 1. 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 equal to 1. Similarly, if the researcher has instruments with a minimum eigenvalue of $\mathbf{\Lambda}$ substantially above the threshold for the weak instruments set, then the probability of correctly concluding that they are strong also is essentially equal to 1.

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 0 and 1. When K_2 is small this range can be 10 or more, but for K_2 large this range of potential ambiguity of the decision rule is quite narrow.

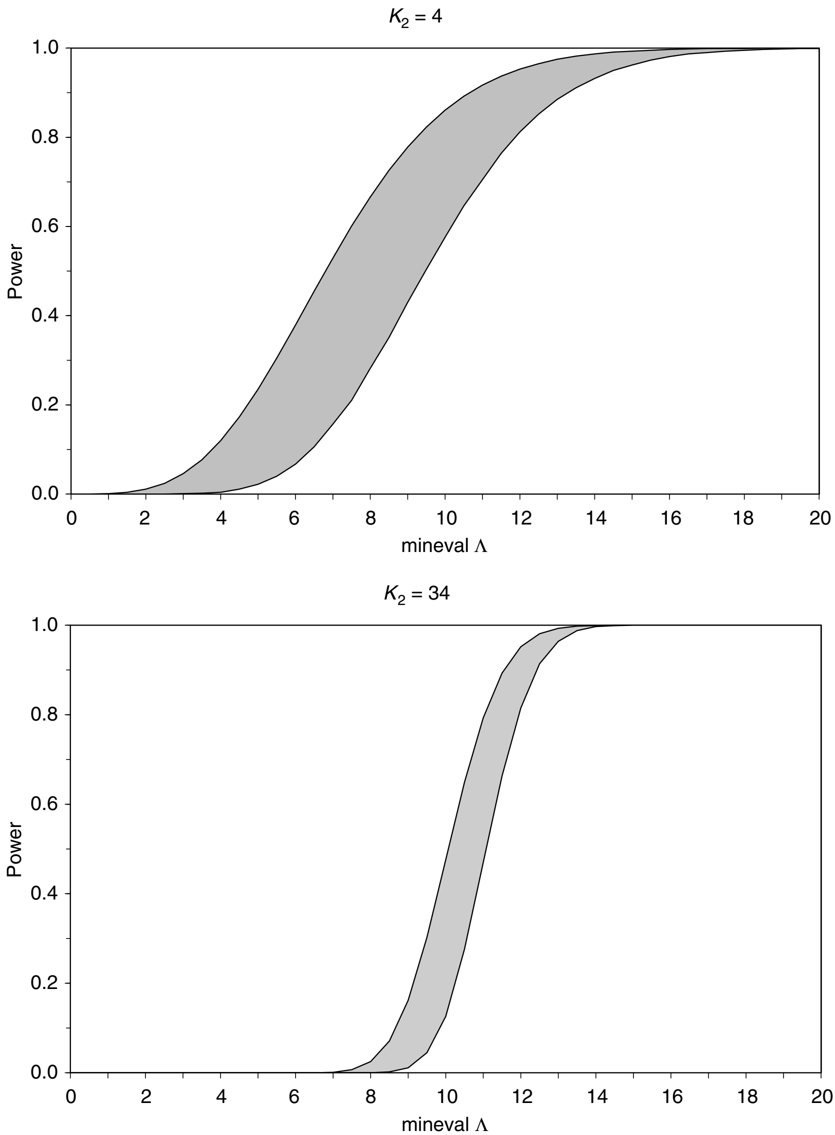

Figure 5.5. Power function for TSLS bias test (Relative bias $= 0.1, n = 2$).

6. CONCLUSIONS

The procedure proposed here is simple: compare the minimum eigenvalue of $\mathbf{G}_T$, 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 bias or size distortion the researcher is willing to 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 concentration parameter, 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 are made in Stock et al. (2002).

When there is a single included endogenous variable, this procedure provides a refinement and improvement to the Staiger–Stock (1997) rule of thumb that instruments be deemed "weak" if the first-stage F is less than 10. The difference between that rule of thumb and the procedure of this paper is that, instead of comparing the first-stage F to 10, it should be compared to the appropriate entry in Table 5.1 (TSLS bias), Table 5.2 (TSLS size), Table 5.3 (Fuller-k bias), or Table 5.4 (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 the maximum size distortion are nonincreasing in each eigenvalue of $\mathbf{\Lambda}$, for values of those eigenvalues in the relevant range. This was justified formally using the many weak instrument asymptotics of Stock and Yogo (2003); 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 on the basis of 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 4.1. Although the tests appear to have good power despite this, tightening the Proposition 4.1 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.

ACKNOWLEDGMENTS

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.

References

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, 443–50.

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," unpublished manuscript, University of Maryland.

Cragg, J. G., and S. G. Donald (1993), "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 9, 222–40.

Donald, S. G., and W. K. Newey (2001), "Choosing the Number of Instruments," *Econometrica*, 69, 1161–91.

Fuller, W. A. (1977): "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45, 939–53.

Hall, A., G. D. Rudebusch, and D. Wilcox (1996), "Judging Instrument Relevance in Instrumental Variables Estimation," *International Economic Review*, 37, 283–98.

Hahn, J., and J. Hausman (2002), "A New Specification Test for the Validity of Instrumental Variables," *Econometrica*, 70, 163–89.

Hahn, J., and J. Hausman (2003), "Weak Instruments: Diagnosis and Cures in Empirical Econometrics," *American Economic Review, Papers and Proceedings*, 93, 118–25.

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

Muirhead, R. J. (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, 575–95.

Nelson, C. R., and R. Startz (1990a), "Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator," *Econometrica*, 58, 967–76.

Nelson, C. R., and R. Startz (1990b), "The Distribution of the Instrumental Variables Estimator and Its t-Ratio when the Instrument is a Poor One," *Journal of Business*, 63, 5125–40.

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

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

Shea, J. (1997), "Instrument Relevance in Multivariate Linear Models: A Simple Measure," *Review of Economics and Statistics*, 79, 348–52.

Staiger, D., and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–86.

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, 518–29.

Stock, J. H., and M. Yogo (2005), "Asymptotic Distributions of Instrumental Variables Statistics with Many Weak Instruments," Chapter 6 in this volume.