Testing for Weak Instruments in Linear IV Regression

August 2001

James H. Stock
Kennedy School of Government, Harvard University
and the National Bureau of Economic Research

and

Motohiro Yogo*
Department of Economics, Harvard University

ABSTRACT

The quality of the asymptotic normal approximation to the distributions of instrumental variables estimators in the linear IV regression model depends on the extent to which the instruments are relevant. If the instruments are weak, so that the system is nearly unidentified for a given sample size, then the sampling distribution can be quite different from its Gaussian limit. This raises a practical problem: under what circumstances can an applied researcher be confident that identification is "good enough," that is, that the instruments are not weak? This problem has been addressed previously (informally) when there is a single endogenous regressor. This paper considers the problem of two or more included endogenous regressors. The paper has two specific contributions. First, we characterize sets of weak instruments in terms of specific population measures on quality of inference based on two stage least squares; these sets in turn depend on eigenvalues of the concentration matrix. Second, we provide a statistical procedure to test for whether the instruments at hand are weak, where the probability of a false negative (concluding that instruments are not weak, when in fact they are) is controlled asymptotically.

*Prepared for the Festschrift in honor of Thomas Rothenberg. This research was supported by NSF grant SBR-9730489.

1. Introduction

Textbook treatments of instrumental variables regression stress that for instruments to be valid they must be exogenous. In practice, however, it is equally important that the second condition for a valid instrument, instrument relevance, also holds, for if the instruments are only marginally relevant, or "weak," then the conventional first order asymptotic limits can provide poor approximations to the sampling distributions of standard instrumental variables 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. In his survey of approximations to the distributions of estimators and test statistics, Rothenberg (1984) makes this point to illustrate that asymptotic expansions need not always be performed in orders of the number of observations. He considered the single equation IV regression model,

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

where $\boldsymbol{y}$ is a $T\times 1$ vector of observations on the dependent variable, $\boldsymbol{Y}$ is the $T\times 1$ included endogenous variables, and $\boldsymbol{U}$ is $T\times 1$ a vector of i.i.d. $N(0,\sigma_{uu})$ errors (his notation is different). 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 exogenous instrumental variables, $\boldsymbol{\Pi}$ is $K_2\times 1$ coefficient vector, and $\boldsymbol{V}$ is a 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 $\hat{\beta}^{TSLS} = (\boldsymbol{Y}'\boldsymbol{P_Z}\boldsymbol{y})/(\boldsymbol{Y}'\boldsymbol{P_Z}\boldsymbol{Y})$, where $\boldsymbol{P_Z} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$, which Rothenberg (1984) expresses as,

$$c\mu(\hat{\beta}^{TSLS} - \beta) = \frac{X + (s/\mu)}{1 + (2Y/\mu) + S/\mu^2}, \tag{1.3}$$

where $\mu^2 = \boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi}/\sigma_{vv}$, $c^2 = \sigma_{vv}/\sigma_{uu}$, $X = \boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{U}/(\sigma_{uu}\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{1/2}$, $Y = \boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{U}/(\sigma_{vv}\boldsymbol{\Pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Pi})^{1/2}$, $s = \boldsymbol{V}'\boldsymbol{P_Z}\boldsymbol{U}/(\sigma_{uu}\sigma_{vv})^{1/2}$, and $S = \boldsymbol{V}'\boldsymbol{P_Z}\boldsymbol{V}/\sigma_{vv}$.

Under the assumptions of fixed instruments and normal errors, X and Y are standard normal variables with correlation ρ, and s and S are elements of a matrix with a central Wishart distribution; in particular the distributions of X, Y, s, and S do not depend on the sample size. Thus the sample size enters the distribution of the TSLS estimator only through the concentration parameter μ^2. Moreover, the form of (1.3) makes it clear that μ^2 can be thought of as an effective sample size, in sense that μ formally plays the role usually associated with $\sqrt{T}$ in (1.3). Rothenberg (1984) proceeds to describe (among other things) 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 is small.

For this reason, an important practical concern is whether μ is so small – that is, whether the instruments are sufficiently weak – that inferences based on conventional TSLS estimates and their standard errors are potentially unreliable. But μ is an unknown population parameter; how is a researcher to know in practice whether his or her instruments are sufficiently weak to jeopardize the validity of conventional TSLS inferences?

In this paper, we develop a procedure for detecting whether instruments are weak. The matter of whether a set of instruments is weak cannot be resolved in the abstract, rather it depends on the inferential task to which the instruments are applied. We therefore offer two concrete definitions of a weak instrument, both based on the performance of the most common IV method, TSLS. The first is that an instrument is weak if the bias of the TSLS estimator, relative to the bias of OLS, could exceed a certain threshold b, for example 10%. The second is that an instrument is weak if the size of the Wald test based on the TSLS estimator exceeds the level of the test by a gap r, for example 10%.

In this light, we do two things. First, we provide an asymptotic characterization of these two sets of weak instruments in terms of the minimum eigenvalue of the matrix version of μ/K_2. Next, we provide a test of the hypothesis that the instruments are weak – that is, they fall in the set of weak instruments – against the alternative that they are not. The test we propose is a simple one: compute the minimum eigenvalue of the matrix version of the F-statistic in the first stage of TSLS and compare it to a critical value, tabulated below. If the minimum eigenvalue is less than the critical value, conclude that the instruments are weak; otherwise, conclude they are not.[1]

Both the characterization of the set of weak instruments and the distribution theory for the test under the weak instrument null require asymptotic approximations to various IV statistic distributions. This requires approximating distributions that are highly accurate even when the concentration parameter is quite small, so small that the approximations discussed in Rothenberg (1984) are poor. We therefore work within the asymptotic framework developed by Staiger and Stock (1997), in which is held constant as the number of observations increases. This allows stochastic objects like *X* and *Y* in the representation (1.3) to have limiting normal distributions even if the errors are not exactly normal and the instruments are stochastic, so that (1.3); it also provides substantial simplifications in the expressions for the Wald statistic.

The tabulated boundaries of the weak instrument sets and the associated critical values for the test behave quite differently for moderate to large values of K_2. We investigate these differences by performing Nagar-type asymptotic expansions of the TSLS estimator and its Wald statistic in orders of $1/\sqrt{K_2}$ under the assumptions that the information per instrument, μ/K_2, is fixed and $K_2/T \to 0$. This is similar in spirit to Bekker's (1994) approach; he obtained a first order distribution under the sequence $K_2 \to \infty$, $T \to \infty$, $K_2/T \to c$, $0 < c < 1$, and μ/T is fixed. In Bekkers' (1994) expansion, the TSLS estimator is approximately normally distributed, although it is biased. Although Bekker's expansion might be a reasonable approximation when there are very many instruments, the assumption that $K_2/T \to c > 0$ raises questions about its suitability cases in which K_2 is moderate and the sample size is large, for example, K_2 is 15 and $T = 1000$.

The rest of the paper is organized as follows. The IV regression model and the proposed test statistic are presented in Section 2. The set of weak instruments is characterized in Section 3. Section 4 presents the limiting asymptotic distribution of the test on the boundary of the (composite) null hypothesis, and argues that the test is asymptotically unbiased. Section 5 examines the power of the test, and the asymptotic expansions are presented in Section 6. Section 7 concludes.

2. The IV Regression Model and Proposed Test Statistic

2.1. The IV Regression Model

The Population Regression Model. The general instrumental variables regression model extends (1.1) and (1.2) to have n included endogenous regressors and K_1 included exogenous regressors:

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

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

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 the regressions include 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 $\underline{\boldsymbol{Z}} = [\boldsymbol{Z}\ \ \boldsymbol{X}]$ denote the matrix of all the exogenous variables. The conformable vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ and the matrices $\boldsymbol{\Pi}$ and $\boldsymbol{\Phi}$ are unknown parameters. Throughout this paper, we shall focus on 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 $\underline{\boldsymbol{Z}}_t = (\boldsymbol{Z}_t'\ \boldsymbol{X}_t')'$ denote the vectors of the t^{th} observations on these variables. Also let $\boldsymbol{Q}$ and $\boldsymbol{\Sigma}$ denote the population second moment matrices,

$$E\begin{pmatrix} u_t \\ \boldsymbol{V}_t \end{pmatrix}\begin{pmatrix} u_t & \boldsymbol{V}_t \end{pmatrix} = \begin{bmatrix} \sigma_{uu} & \boldsymbol{\Sigma}_{uV} \\ \boldsymbol{\Sigma}_{Vu} & \boldsymbol{\Sigma}_{VV} \end{bmatrix} = \boldsymbol{\Sigma} \text{ and } E\begin{pmatrix} \boldsymbol{X}_t \\ \boldsymbol{Z}_t \end{pmatrix}\begin{pmatrix} \boldsymbol{X}_t & \boldsymbol{Z}_t \end{pmatrix} = \begin{bmatrix} \boldsymbol{Q}_{XX} & \boldsymbol{Q}_{ZX} \\ \boldsymbol{Q}_{XZ} & \boldsymbol{Q}_{ZZ} \end{bmatrix} = \boldsymbol{Q}. \qquad (2.3)$$

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

Assumption $\boldsymbol{L_{\Pi}}$: $\boldsymbol{\Pi} = \boldsymbol{\Pi}_T = \boldsymbol{C}/\sqrt{T}$, where $\boldsymbol{C}$ is a fixed $K_2 \times n$ matrix.

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

Assumption M. The following limits hold jointly:

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

(b) $T^{-1}\underline{\boldsymbol{Z}}'\underline{\boldsymbol{Z}} \xrightarrow{p} \boldsymbol{Q}$, where $\boldsymbol{Q}$ has blocks denoted $\boldsymbol{Q}_{XZ}$, etc.;

(c) $(T^{-1/2}\boldsymbol{X}'\boldsymbol{U}, T^{-1/2}\boldsymbol{Z}'\boldsymbol{U}, T^{-1/2}\boldsymbol{X}'\boldsymbol{V}, T^{-1/2}\boldsymbol{Z}'\boldsymbol{V}) \Rightarrow (\boldsymbol{\Psi}_{Xu}, \boldsymbol{\Psi}_{Zu}, \boldsymbol{\Psi}_{XV}, \boldsymbol{\Psi}_{ZV})$, where $\boldsymbol{\Psi} \equiv [\boldsymbol{\Psi}_{Xu}', \boldsymbol{\Psi}_{Zu}', \text{vec}(\boldsymbol{\Psi}_{XV})', \text{vec}(\boldsymbol{\Psi}_{ZV})']'$ 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.

The TSLS Estimator and Wald Statistic. We focus on estimation of $\boldsymbol{\beta}$ by two stage least squares. Let the superscript "$^{\perp}$" denote the residuals from the projection on $\boldsymbol{X}$; that is, $\boldsymbol{Y}^{\perp} = \boldsymbol{M_X Y}$, $\boldsymbol{y}^{\perp} = \boldsymbol{M_X y}$, and so forth, 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 TSLS estimator is,

$$\hat{\boldsymbol{\beta}}^{TSLS} = (\boldsymbol{Y}^{\perp\prime}\boldsymbol{P}_{\boldsymbol{Z}^{\perp}}\boldsymbol{Y}^{\perp})^{-1}(\boldsymbol{Y}^{\perp\prime}\boldsymbol{P}_{\boldsymbol{Z}^{\perp}}\boldsymbol{y}). \tag{2.4}$$

The Wald statistic testing the q linear restrictions that $\boldsymbol{R\beta} = \boldsymbol{r}_0$ is,

$$W_T = \frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}}^{TSLS} - \boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{Y}^{\perp\prime}\boldsymbol{P}_{\boldsymbol{Z}^{\perp}}\boldsymbol{Y}^{\perp})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}}^{TSLS} - \boldsymbol{r})}{q\hat{\sigma}_{uu}^{TSLS}} \tag{2.5}$$

where $\hat{\sigma}_{uu}^{TSLS} = \hat{\boldsymbol{U}}^{TSLS\prime}\hat{\boldsymbol{U}}^{TSLS}/(T-K_1-n)$, where $\hat{\boldsymbol{U}}^{TSLS} = \boldsymbol{y}^{\perp} - \boldsymbol{Y}^{\perp}\hat{\boldsymbol{\beta}}^{TSLS}$. We shall focus on the case $q = n$.

2.2. The Proposed Test Statistic

The proposed test is based on the eigenvalue of the matrix version of the F-statistic from the first stage regression of TSLS,

$$\boldsymbol{G}_T = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{VV}}^{-1/2\prime}\boldsymbol{Y}^{\perp\prime}\boldsymbol{P}_{\boldsymbol{Z}^{\perp}}\boldsymbol{Y}^{\perp}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{VV}}^{-1/2}/K_2, \tag{2.6}$$

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

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

$$g_{\min} = \text{mineval}(\boldsymbol{G}_T). \tag{2.7}$$

3. Characterization of the Set of Weak Instruments

3.1. Weak Instrument First Order Asymptotic Representations

We start by summarizing relevant weak instrument asymptotic results from Staiger and Stock (1997).

Notation and definitions. The following notation in effect is a transformation of variables and parameters that simplifies the asymptotic expressions. Let

$$\boldsymbol{\rho} = \boldsymbol{\Sigma}_{VV}^{-1/2\,\prime}\boldsymbol{\Sigma}_{Vu}\sigma_{uu}^{-1/2},$$

$$\boldsymbol{\theta} = \boldsymbol{\Sigma}_{VV}^{-1}\boldsymbol{\Sigma}_{Vu} = \sigma_{uu}^{1/2}\,\boldsymbol{\Sigma}_{VV}^{-1/2}\boldsymbol{\rho},$$

$$\boldsymbol{\lambda} = \boldsymbol{\Omega}^{1/2}\boldsymbol{C}\,\boldsymbol{\Sigma}_{VV}^{-1/2}, \text{ and}$$

$$\boldsymbol{\Omega} = \boldsymbol{Q}_{ZZ} - \boldsymbol{Q}_{ZX}\boldsymbol{Q}_{XX}^{-1}\boldsymbol{Q}_{XZ}.$$

Note that $\boldsymbol{\rho}'\boldsymbol{\rho} \leq 1$. Define the $K_2\times1$ and $K_2\times n$ random variables,

$$\boldsymbol{z}_u = \boldsymbol{\Omega}^{-1/2\prime}(\boldsymbol{\Psi}_{Zu} - \boldsymbol{Q}_{ZX}\boldsymbol{Q}_{XX}^{-1}\,\boldsymbol{\Psi}_{Xu})\sigma_{uu}^{-1/2}$$

$$\boldsymbol{z}_V = \boldsymbol{\Omega}^{-1/2\prime}(\boldsymbol{\Psi}_{ZV} - \boldsymbol{Q}_{ZX}\boldsymbol{Q}_{XX}^{-1}\,\boldsymbol{\Psi}_{XV})\,\boldsymbol{\Sigma}_{VV}^{-1/2}$$

so that

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

Also let

$$\nu_1 = (\boldsymbol{\lambda} + \boldsymbol{z_V})' (\boldsymbol{\lambda} + \boldsymbol{z_V}), \text{ and} \tag{3.1}$$

$$\boldsymbol{\nu}_2 = (\boldsymbol{\lambda} + \boldsymbol{z_V})' \boldsymbol{z_u}. \tag{3.2}$$

Weak Instrument Asymptotic Representations. In this notation, the probability limit of the OLS estimator is

$$\hat{\boldsymbol{\beta}} \overset{p}{\to} \boldsymbol{\beta} + \boldsymbol{\theta}. \tag{3.3}$$

The TSLS estimator and $\boldsymbol{G}_T$ have the following limits (Staiger and Stock (1997), theorem 1):

$$\hat{\boldsymbol{\beta}}^{TSLS} \Rightarrow \sigma_{uu}^{1/2} \boldsymbol{\Sigma}_{\boldsymbol{VV}}^{-1/2} \nu_1^{-1} \boldsymbol{\nu}_2, \tag{3.4}$$

$$\boldsymbol{G}_T \Rightarrow \nu_1 / K_2 \tag{3.5}$$

and, when $q = n$ and the null hypothesis $\boldsymbol{r} = \boldsymbol{r}_0$ is true,

$$W_T \Rightarrow \frac{\nu_2'\nu_1^{-1}\nu_2}{n(1-2\rho'\nu_1^{-1}\nu_2+\nu_2'\nu_1^{-2}\nu_2)} \equiv W^*, \tag{3.6}$$

where "$\Rightarrow$" denotes weak convergence. That is, (3.4) says that, under Assumptions L_Π and M, as $T \to \infty$, the distribution of $\hat{\beta}^{TSLS}$ converges to the distribution of $\sigma_{uu}^{1/2}\,\Sigma_{VV}^{-1/2}\,\nu_1^{-1}\nu_2$.

3.2. First Characterization of the Set of Weak Instruments: TSLS Bias

Our first characterizations of the set of weak instruments is based on the maximum bias of TSLS relative to OLS. The expectation of the weak instrument asymptotic representation of the TSLS estimator exists only if the degree of overidentification $K_2 - n \geq 2$, so this is assumed whenever discussing bias of TSLS.

We consider the relative bias measure B_T, which is the ratio of the bias of TSLS to the bias of OLS, where the coefficient vector has been put into standardized units by rotating by $\Sigma_{Y^\perp Y^\perp}^{1/2}$:[3]

$$B_T^2 = \frac{(E\hat{\beta}^{TSLS}-\beta)'\Sigma_{Y^\perp Y^\perp}(E\hat{\beta}^{TSLS}-\beta)}{(E\hat{\beta}-\beta)'\Sigma_{Y^\perp Y^\perp}(E\hat{\beta}-\beta)}. \tag{3.7}$$

If $n = 1$, then the scaling matrix in (3.7) drops out and the expression simplifies to $B_T = |E\hat{\beta}^{TSLS} - \beta|/|E\hat{\beta} - \beta|$.

Under weak instrument asymptotics, $\hat{\boldsymbol{\beta}}^{TSLS}$ has the limit in (3.4), and the bias measure has the asymptotic limit,

$$B_T^2 \rightarrow \frac{\boldsymbol{\rho}'\boldsymbol{h}'\boldsymbol{h}\boldsymbol{\rho}}{\boldsymbol{\rho}'\boldsymbol{\rho}} \equiv B^2, \tag{3.8}$$

where $\boldsymbol{h} = E[\boldsymbol{v}_1^{-1}(\boldsymbol{\lambda} + \boldsymbol{z_V})'\boldsymbol{z_V}]$.

The asymptotic bias measure B^2 depends on $\boldsymbol{\rho}$ and $\boldsymbol{\lambda}$, which are unknown, as well as K_2 and n.

Our approach is to consider 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 and assuming $\boldsymbol{\rho}'\boldsymbol{\rho} > 0$, this leads us to consider the worst-case asymptotic bias,

$$B^{\max} = (\max_{\boldsymbol{\rho}:\, \boldsymbol{\rho}'\boldsymbol{\rho} > 0} B^2)^{1/2} \tag{3.9}$$

This first characterization of the set of weak instruments is based on this worst-case bias. We define the set of weak instruments, 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 a weak instrument is,

$$\mathcal{W}_{bias} = \{\mathcal{Z}: B^{max} \geq b\} \tag{3.10}$$

The asymptotic results make it possible to characterize the set $\mathcal{W}_{bias}$. Because $\boldsymbol{h}$ depends on $\boldsymbol{\lambda}$ but not $\boldsymbol{\rho}$, by (3.8) we have that

$$B^{max} = (\max_{\rho} \frac{\boldsymbol{\rho}'\boldsymbol{h}'\boldsymbol{h}\boldsymbol{\rho}}{\boldsymbol{\rho}'\boldsymbol{\rho}})^{1/2} = [\text{maxeval}(\boldsymbol{h}'\boldsymbol{h})]^{1/2}, \tag{3.11}$$

where maxeval($\boldsymbol{A}$) denotes the maximum eigenvalue of the matrix $\boldsymbol{A}$. By applying the singular value decomposition to lamba, it is further possible to show that the maximum eigenvalue of $\boldsymbol{h}'\boldsymbol{h}$, and thus B^{max}, depends only on K_2, n, and the eigenvalues of $\boldsymbol{\lambda}'\boldsymbol{\lambda}/K_2$.

In Section 6, we provide an asymptotic expansion for the TSLS estimator and its Wald statistic when the number of instruments K_2 increases slowly with the sample size and in which $\boldsymbol{\lambda}'\boldsymbol{\lambda}/K_2$ is fixed, so $\boldsymbol{\lambda}'\boldsymbol{\lambda}/K_2 = \boldsymbol{\Lambda}$, a fixed matrix. Proposition 2 in that section states that the leading term in the associated expansion for B^{max} is nonincreasing in each of the eigenvalues of $\boldsymbol{\Lambda}$. It follows that, to the first order of approximation in that expansion, the set $\mathcal{W}_{bias}$ can be characterized by the minimum eigenvalue of $\boldsymbol{\lambda}'\boldsymbol{\lambda}/K_2$ for a given n and K_2. That is, $B^{max} = B^{max}(\text{mineval}(\boldsymbol{\lambda}'\boldsymbol{\lambda}/K_2);K_2,n)$, where this function is increasing in its argument and therefore invertible. This in turn implies that $B^{max}(\text{mineval}(\boldsymbol{\lambda}'\boldsymbol{\lambda}/K_2);K_2,n) \leq b$ is equivalent to $\text{mineval}(\boldsymbol{\lambda}'\boldsymbol{\lambda}/K_2) \leq \ell_{bias}(b;K_2,n)$, where

ℓ_{bias} is the inverse function of B^{max} with respect to its first argument. That is, $\mathcal{W}_{\text{bias}}$ in (3.10) can be written as,

$$\mathcal{W}_{bias} = \{\mathcal{Z}: \text{mineval}(\lambda'\lambda/K_2) \leq \ell_{\text{bias}}(b;K_2,n)\}. \tag{3.12}$$

Our formal justification for the simplification that $\mathcal{W}_{bias}$ depends only on the smallest eigenvalue of $\lambda'\lambda/K_2$, rather than on all its eigenvalues, rests on the expansion in Section 6. Numerical analysis for $n = 2$ suggests, however, that B^{max} is decreasing in all the eigenvalues of $\lambda'\lambda/K_2$ for all values of K_2. This numerical analysis suggests that the simplification in (3.12), looking only at the minimum eigenvalue, is valid for all K_2 under the weak instrument asymptotics, even though we currently cannot provide a formal justification.[4]

3.3. Second Characterization of the Set of Weak Instruments: TSLS Wald Test Size

The second characterization of the set of weak instruments is based on the maximal size of the TSLS-based Wald test of all the elements of β. Throughout we focus only on the possibility that the Wald test rejects too often under the null, which is in fact the case with weak instruments.

In parallel with the approach for the bias measure, we consider an instrument strong from the perspective of the Wald statistic if the size of the test is close to its level for all possible configurations of the IV regression model. The actual rejection rate R_T under the null hypothesis is,

$$R_T = \Pr[W_T > \chi^2_{n;\alpha}/n | \boldsymbol{r} = \boldsymbol{r}_0], \qquad (3.13)$$

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.

Under the weak instrument asymptotics the distribution of W_T has the limiting representation (3.6) under the null hypothesis. Thus,

$$R_T \rightarrow \Pr[W^* > \chi^2_{n;\alpha}/n] \equiv R, \qquad (3.14)$$

Inspection of (3.6) reveals that R depends on ρ and λ, as well as K_2, n, and α. Following our treatment of the bias, because ρ is unknown we consider the worst-case size,

$$R^{\max} = \max_\rho R = \max_\rho \Pr[\frac{\nu_2'\nu_1^{-1}\nu_2}{(1-2\rho'\nu_1^{-1}\nu_2 + \nu_2'\nu_1^{-2}\nu_2)} > \chi^2_{n;\alpha}]. \qquad (3.15)$$

It is not necessary to exclude $\rho'\rho = 0$ in the maximization in (3.15) because there is no singularity at $\rho'\rho = 0$.

The set of weak instruments, $\mathcal{W}_{\text{size}}$, based on the size of the Wald statistic, consists of instruments that can lead to a size of at least r:

$$\mathcal{W}_{size} = \{\mathcal{Z}: R^{max} \geq r\} \tag{3.16}$$

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

The maximal bias measure R^{max} depends on the eigenvalues of $\lambda'\lambda/K_2$ as well as n and K_2 (the argument is the same as for the similar assertion for B^{max}). Thus, under the weak instrument asymptotics, the weak instrument set $\mathcal{W}_{size}$ is fully characterized by $\boldsymbol{\rho}$ and the eigenvalues of $\lambda'\lambda/K_2$.

In the expansion of W^* in Section 6, the leading term is nonincreasing in the eigenvalues of $\lambda'\lambda/K_2$ (Proposition 3). This implies that we can characterize $\mathcal{W}_{size}$ in terms of the minimum eigenvalue of $\lambda'\lambda/K_2$, as well as K_2 and n. The argument leading to (3.12) therefore applies here and leads to the characterization,

$$\mathcal{W}_{size} = \{\mathcal{Z}: \text{mineval}(\lambda'\lambda/K_2) \leq \ell_{size}(r;K_2,n,\alpha)\}. \tag{3.17}$$

where $\ell_{size}(r;K_2,n,\alpha)$ is the inverse function of $R^{max}(\text{mineval}(\lambda'\lambda/K_2); K_2,n,\alpha)$ with respect to its first argument.

As is the case for $\mathcal{W}_{bias}$, the justification for the simplification (3.17) in which $\mathcal{W}_{size}$ depends on only the smallest eigenvalue of $\lambda'\lambda/K_2$ formally is based on the expansion in Section 6 for weak instruments and moderate K_2. However, numerical analysis of the $n = 2$ case suggests that this property holds for small K_2 as well.

3.4. Numerical Results

The worst case bias and the maximal size distortion were computed by Monte Carlo simulation for a grid of minimal eigenvalue of $\lambda'\lambda/K_2$ from 0 to 50, for $K_2 = n + 2,\ldots,100$ for the relative bias, and for $K_2 = n,\ldots, 100$ for the size distortions. All results are based on 20,000 Monte Carlo draws. Computing the maximum bias entails first computing $\boldsymbol{h}$ defined following (3.8) by Monte Carlo simulation, given n, K_2, and the minimum eigenvalue; as discussed above, for these computations we set $\lambda'\lambda/K_2 = \ell \boldsymbol{I}_n$, where ℓ is the minimum eigenvalue. The maximum relative bias is then $[\text{maxeval}(\boldsymbol{h}'\boldsymbol{h})]^{1/2}$. Computing the maximum size distortion is more involved because there is no simple analytic solution to the maximum problem (3.15). Numerical analysis indicates that R 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 the expression for W^*). As in the relative bias computations and as discussed above, in these computations we set $\lambda'\lambda/K_2 = \ell \boldsymbol{I}_n$, where ℓ is the minimum eigenvalue.

The relative bias is plotted for n =1 as a function of K_2 and $\lambda'\lambda/K_2$ in Figure 1, and the maximal relative bias is plotted for $n = 2$ as a function of K_2 and the minimum eigenvalue of $\lambda'\lambda/K_2$ in Figure 2 (when $n = 1$, the relative bias B does not depend on $\boldsymbol{\rho}$, so $B^{max} = B$).[5] Evidently, for K_2 sufficiently large, the maximal relative bias depends only weakly on K_2. The maximal relative bias seems to decline as approximately the inverse of the minimum eigenvalue. Also, the contours for $n = 1$ in Figure 1 and $n = 2$ in Figure 2 are quite similar.

The worst case size, R^{max}, is plotted in Figure 3 for $n = 1$ and in Figure 4 for $n = 2$ for a Wald test of level 5%. In contrast to the plots of the worst case bias, the worst case size strongly depends on K_2 over the full range considered. The worst case size approaches 5% much more rapidly as a function of the minimum eigenvalue for small K_2 than when K_2 is large. Also, the surfaces for $n = 1$ and $n = 2$ are qualitatively similar but quantitatively different; curiously, given a value of K_2 and the minimum eigenvalue, the size distortion when $n = 1$ is worse than when $n = 2$.

These surfaces can be used to compute boundaries of the weak instrument regions $\mathcal{W}_{bias}$ and $\mathcal{W}_{size}$, where the boundaries are determined by the minimum eigenvalue of $\lambda'\lambda/K_2$, as well as by n and K_2. These boundaries are plotted in Figure 5: the relative bias region boundary is plotted in the top two panels for selected values of the relative bias cutoff b, and the size region boundary is plotted in the bottom two panels for selected values of the size cutoff r.

First consider the boundary of $\mathcal{W}_{bias}$. These are 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 $\mathcal{W}_{size}$ depends strongly on K_2 and n. The boundary is approximately linear in K_2 for K_2 sufficiently large, for all the size distortion cutoffs r considered. Note also that the cutoffs are numerically quite 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 20 instruments the minimum eigenvalue boundary is approximately 25 for $n = 1$ and approximately 20 for $n = 2$.

We return to some of these features of the weak instrument region boundary functions when in the discussion of the asymptotic expansion in Section 6.

4. Asymptotic Distribution of the Test Statistic

This section provides the weak instrument asymptotic representation of the test statistic $g_{\min}$ and a bound on its distribution. This bound provides conservative critical values for the test based on $g_{\min}$.

Distribution of $g_{\min}$. Recall that the statistic g_{min} is the minimum eigenvalue of $\boldsymbol{G}_T$, where $\boldsymbol{G}_T$ is given by (2.6). 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 $\lambda'\lambda$:

$$\boldsymbol{G}_T \Rightarrow \boldsymbol{G}^* \sim \mathrm{W}_n(K_2, \boldsymbol{I}_n, \lambda'\lambda)/K_2. \tag{4.1}$$

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 $\lambda'\lambda$, as well as n and K_2. In principal 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 $\lambda'\lambda$.

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 $\lambda'\lambda$. This latter consideration is especially important because in practice these eigenvalues are unknown nuisance parameters, so any critical values based on these eigenvalues would produce an infeasible test.

We circumvent these two problems by performing the test using conservative critical values. These conservative critical values are based on a noncentral chi-squared bounding distribution, given in the following proposition.

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

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

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

$$\Pr[g_{\min} \geq x] \to \Pr[\text{mineval}(\boldsymbol{G}^*) \geq x] \leq \Pr\left[\frac{\chi^2_{K_2}(\text{mineval}(\lambda'\lambda)/K_2)}{K_2} \geq x\right]. \quad (4.2)$$

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

Weak instruments test. This yields the following testing procedure to detect for weak instruments. To be concrete, this is stated for a test based on the bias measure. The null hypothesis is that the instruments are weak, and the alternative is that they are not:

$$H_0: \mathcal{Z} \in \mathcal{W}_{\text{bias}} \text{ v. } H_1: \mathcal{Z} \notin \mathcal{W}_{\text{bias}}. \quad (4.3)$$

The test procedure is,

$$\text{Reject } H_0 \text{ if } g_{\min} \geq \chi^2_{K_2,\delta}(\ell_{\text{bias}}(b;K_2,n)), \quad (4.4)$$

where $\chi^2_{K_2,\delta}(\ell)$ is the 100δ% critical value ($100(1-\delta)$% percentile) of the noncentral chi-squared distribution with K_2 degrees of freedom and noncentrality parameter ℓ.

The critical value chosen by the researcher depends on the attitude of the researcher towards the maximal bias. If the researcher is satisfied if the maximal relative bias is no more than 10%, then she would set $b = .10$.

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

$$\lim_{T\to\infty} \Pr[g_{\min} \geq \chi^2_{K_2,\delta}(\ell_{\text{bias}}(b;K_2,n)) \mid \mathcal{Z} \in \mathcal{W}_{\text{bias}}] \leq \delta. \tag{4.5}$$

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

Critical Values. Given a minimum eigenvalue ℓ, conservative critical values for the test are percentiles of the noncentral chi-squared distribution, $\chi^2_{K_2,\delta}(\ell)$. The minimum eigenvalue ℓ is obtained from the boundary eigenvalue functions, numerical values of which were reported in Section 3.4.

Critical values for the relative bias version of the test are tabulated in Tables 1, 2, and 3 for $n = 1$, 2 and 3, respectively, for various K_2 and relative bias tolerances b. Critical values for the size version of the test are reported in Tables 4 and 5.

The critical values reflect the features of the boundary eigenvalue functions discussed in Section 3.4. For example, the critical values for the relative bias at first increase then, for moderate K_2, essentially do not depend on K_2. In contrast, the critical values for the size version increase approximately linearly with K_2.

The critical values are plotted in Figure 6. These critical value plots are qualitatively similar to the boundary eigenvalue plots in Figure 5, except of course the critical values exceed the boundary eigenvalues to take into account the sampling distribution of the test statistic.

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 can be viewed as 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 Staiger-Stock 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% (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 discusses the asymptotic properties of the weak instruments test as a function of the smallest eigenvalue of $\lambda'\lambda/K_2$. This entails providing and studying the asymptotic rejection rate of the test as a function of the smallest eigenvalue of $\lambda'\lambda/K_2$.

When this eigenvalue exceeds the cutoff population eigenvalue $\ell_{bias}(b;K_2,n)$ for a given b, this asymptotic rejection rate is the asymptotic power function.

The asymptotic distribution of g_{min} depends on all the eigenvalues of $\lambda'\lambda/K_2$. It is bounded above by (4.2). We conjecture that this distribution is bounded below by the distribution of the minimum eigenvalue of a random matrix with the noncentral Wishart distribution $W_n(K_2, \boldsymbol{I}_n, \text{mineval}(\lambda'\lambda)\boldsymbol{I}_n)/K_2$.[6] We will use these two bounding distributions to bound the distribution of g_{min} as a function of mineval$(\lambda'\lambda/K_2)$.

The resulting bounds on the asymptotic rejection rate of the test (4.4) (based on relative bias) are plotted in Figures 7 for $b = .10$ and $n = 2$. The value of the horizontal axis (the minimum eigenvalue) at which the upper rejection rate curve equals 5% is $\ell_{bias}(.10;K_2,2)$. Evidently, as the minimum eigenvalue increases, the rejection rate increases. For large K_2, this increase is rapid, and the test effectively has unit power against values of the minimum eigenvalue not much larger than one. The bounding distributions give a reasonably tight range for the actual power function, which depends on all the eigenvalues of $\lambda'\lambda/K_2$.

The analogous bounds for the test based on the size of the Wald statistic are plotted in Figure 8. The curves have a similar shape, but they are centered at much larger values of the minimal eigenvalue for the reasons discussed in Sections 4 and 5. Otherwise, the qualitative conclusions for the bias-based test apply to the size-based test.

Interpretation as a decision rule. The test in (4.4) can, of course, be interpreted as a decision rule: if g_{min} is less than the critical value, conclude that the instruments are weak; if it exceeds the critical value, conclude that they are strong.

Under this interpretation, the asymptotic rejection rates plotted in Figures 7 and 8 bound the asymptotic probability of deciding that the instruments are strong. Evidently, for small values of mineval($\lambda'\lambda/K_2$), the probability of correctly concluding that the instruments are weak is effectively one. Thus, if in fact one is confronted by instruments that are quite weak, then the probability of not recognizing this using the testing procedure (4.4) is vanishingly small. Similarly, if one has instruments for which the minimum eigenvalue of $\lambda'\lambda/K_2$ is substantially above the threshold for the weak instruments set, then the probability of correctly concluding that they are strong also is effectively 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 is fairly large. For K_2 large, this range of potential ambiguity of the decision rule is quite small.

From a practical perspective, it matters whether the researcher is concerned about bias or size. Concern about relative bias leads to much smaller weak instrument sets than concern about size distortions.

6. Asymptotic Expansion for Many Weak Instruments

This section reports the results of, and conclusions based on, Nagar-type asymptotic expansions of the TSLS estimator $\hat{\beta}^{TSLS}$ and its Wald statistic W. These expansions have two purposes. First, they provides a formal justification for the characterizations of the sets $\mathcal{W}_{\text{bias}}$ and $\mathcal{W}_{\text{size}}$ solely in terms of the minimum eigenvalue of

$\lambda'\lambda/K_2$; this is summarized in Propositions 2 and 3 below. Second, they explain the peculiar feature, noted in Section 4, that the boundary of the weak instrument region $\mathcal{W}_{\text{bias}}$ is essentially constant for K_2 is sufficiently large (say, $K_2 - n \geq 10$), whereas the boundary of the region $\mathcal{W}_{\text{size}}$ is essentially linear in K_2 for K_2 sufficiently large.

The expansion here holds the noncentrality matrix, $\lambda'\lambda/K_2$, constant. We let $K_2 \to \infty$ and $T \to \infty$ but, in contrast to Bekker (1994), $K_2/T \to 0$. The expansion is carried out only to order $o_p(1/\sqrt{K_2})$, insufficient to calculate an Edgeworth approximation but sufficient to address the matters of the previous paragraph.

The expansion is performed for the IV regression model (1.1) and (1.2) with no included exogenous variables, fixed $\boldsymbol{Z}$, and i.i.d. Gaussian errors, except that we allow for n included endogenous variables. Let

$$\tilde{\lambda} = (\mathbf{Z}'\mathbf{Z}/T)^{-1/2}(\sqrt{T}\,\Pi_{\boldsymbol{T}})\boldsymbol{\Sigma}_{\boldsymbol{VV}}^{-1/2},$$

$$\tilde{\boldsymbol{z}}_{\boldsymbol{V}} = (\mathbf{Z}'\mathbf{Z}/T)^{-1/2}(\mathbf{Z}'\mathbf{V}/\sqrt{T}\,)\boldsymbol{\Sigma}_{\boldsymbol{VV}}^{-1/2},$$

$$\tilde{\boldsymbol{z}}_{u} = (\mathbf{Z}'\mathbf{Z}/T)^{-1/2}(\mathbf{Z}'\mathbf{U}/\sqrt{T}\,)\sigma_{uu}^{-1/2},$$

$$\tilde{\boldsymbol{v}}_1 = (\tilde{\lambda} + \tilde{\boldsymbol{z}}_{\boldsymbol{V}})'(\tilde{\lambda} + \tilde{\boldsymbol{z}}_{\boldsymbol{V}}), \text{ and}$$

$$\tilde{\boldsymbol{v}}_2 = (\tilde{\lambda} + \tilde{\boldsymbol{z}}_{\boldsymbol{V}})'\tilde{\boldsymbol{z}}_u.$$

These definitions parallel the corresponding definitions of λ, etc. in the weak instrument asymptotic representation in Section 3, except that these are defined in the original sample space.

With these definitions, we have the exact relations,

$$\hat{\boldsymbol{\beta}}^{TSLS} = \boldsymbol{\beta} + \sigma_{uu}^{1/2}\, \boldsymbol{\Sigma}_{VV}^{-1/2}\, \tilde{\boldsymbol{\nu}}_1^{-1}\, \tilde{\boldsymbol{\nu}}_2, \tag{6.1}$$

$$W = \tilde{\boldsymbol{\nu}}_2{}'\tilde{\boldsymbol{\nu}}_1^{-1}\, \tilde{\boldsymbol{\nu}}_2/(ns_T), \tag{6.2}$$

where the expression for W holds under the null hypothesis and

$$s_T = (\boldsymbol{U}'\boldsymbol{U}/T)/\sigma_{uu} - 2[\,\tilde{\boldsymbol{z}}_u{}'\tilde{\boldsymbol{\lambda}} + \boldsymbol{\rho}'\boldsymbol{\Sigma}_{VV}^{-1/2}(\boldsymbol{V}'\boldsymbol{V}/T)\,\boldsymbol{\Sigma}_{VV}^{-1/2}{}' + \sigma_{uu}^{-1/2}(\boldsymbol{U}'\boldsymbol{V}/T)\,\boldsymbol{\Sigma}_{VV}^{-1/2}\,]\tilde{\boldsymbol{\nu}}_1^{-1}\tilde{\boldsymbol{\nu}}_2$$

$$+ \tilde{\boldsymbol{\nu}}_2{}'\tilde{\boldsymbol{\nu}}_1^{-1}[(\tilde{\boldsymbol{\lambda}}'\tilde{\boldsymbol{\lambda}})/T + (\tilde{\boldsymbol{\lambda}}'\tilde{\boldsymbol{z}}_V)/T + (\tilde{\boldsymbol{z}}_V{}'\tilde{\boldsymbol{\lambda}})/T + \boldsymbol{\Sigma}_{VV}^{-1/2}(\boldsymbol{V}'\boldsymbol{V}/T)\,\boldsymbol{\Sigma}_{VV}^{-1/2}{}']\,\tilde{\boldsymbol{\nu}}_1^{-1}\tilde{\boldsymbol{\nu}}_2.$$

We make the following assumption:

Assumption E. $K_2 \to \infty$, $T \to \infty$, $K_2/T \to 0$, and $\tilde{\boldsymbol{\lambda}}'\tilde{\boldsymbol{\lambda}}/K_2 = \boldsymbol{\Lambda}$ is fixed.

Assumption E replaces assumption L_Π. Under this assumption, $\boldsymbol{\Pi}$ is implicitly modeled as $\boldsymbol{\Pi}_T = \sqrt{K_2/T}\,\boldsymbol{C}$, that is, as being in a $\sqrt{K_2/T}$ neighborhood of zero. This captures the notion that the instruments are weak while permitting an analysis of the statistics of interest for the case that the number of instruments is moderate, relative to the sample size.

Under Assumption E, the assumption of homoskedastic Gaussian errors, and the assumption that $\boldsymbol{Z}'\boldsymbol{Z}/T \xrightarrow{p} \boldsymbol{Q}_{ZZ}$, we find that,

$$\tilde{\boldsymbol{v}}_1/K_2 = \boldsymbol{\Lambda} + \boldsymbol{I}_n + \frac{\tilde{\boldsymbol{z}}_1 + \tilde{\boldsymbol{z}}_1' + \tilde{\boldsymbol{z}}_2}{\sqrt{K_2}} + o_p(\frac{1}{\sqrt{K_2}}), \text{ and} \tag{6.3}$$

$$\tilde{\boldsymbol{v}}_2/K_2 = \boldsymbol{I}_n\boldsymbol{\rho} + \frac{\tilde{\boldsymbol{z}}_1 + \tilde{\boldsymbol{z}}_2\boldsymbol{\rho} + \tilde{\boldsymbol{z}}_3 + \tilde{\boldsymbol{z}}_4}{\sqrt{K_2}} + o_p(\frac{1}{\sqrt{K_2}}) \tag{6.4}$$

where $\tilde{\boldsymbol{z}}_1$, $\tilde{\boldsymbol{z}}_2$, $\tilde{\boldsymbol{z}}_3$, and $\tilde{\boldsymbol{z}}_4$ are independent and vec($\tilde{\boldsymbol{z}}_1$) $\sim N(\mathbf{0},\boldsymbol{I}_n\otimes\boldsymbol{\Lambda})$, vec($\tilde{\boldsymbol{z}}_2$) $\sim N(\mathbf{0},\boldsymbol{H})$,

$\tilde{\boldsymbol{z}}_3 \sim N(\mathbf{0}, (1-\boldsymbol{\rho}'\boldsymbol{\rho})\boldsymbol{\Lambda})$, and $\tilde{\boldsymbol{z}}_4 \sim N(\mathbf{0}, (1-\boldsymbol{\rho}'\boldsymbol{\rho})\boldsymbol{I}_n)$, and

$$\boldsymbol{H} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

Bias. Applying these results to the relative bias formula yields the leading term,

$$B^2 = \frac{\boldsymbol{\rho}'(\boldsymbol{\Lambda}+\boldsymbol{I})^{-2}\boldsymbol{\rho}}{\boldsymbol{\rho}'\boldsymbol{\rho}} + O_p(\frac{1}{\sqrt{K_2}}). \tag{6.5}$$

This allows us to characterize the weak instrument set $\mathcal{W}_{\text{bias}}$ to the order of approximation in (6.5).

***Proposition* 2**. For any $\boldsymbol{\rho}$ such that $\boldsymbol{\rho}'\boldsymbol{\rho} > 0$, the leading term in the expansion (6.5) for the relative bias B^2 is nonincreasing in each eigenvalue of $\boldsymbol{\Lambda}$. Moreover, up to this leading term in the expansion, $B^{max} = [\text{ mineval}(\boldsymbol{\Lambda}) + 1]^{-1}$.

Proof. For any $\boldsymbol{\rho}$, the leading term in (6.5) can be written as $\sum_{i=1}^{n} a_i^2(\lambda_i + 1)^{-2}$, where λ_i is the i^{th} eigenvalue of $\boldsymbol{\Lambda}$, $a_1, \ldots, a_n$ do not depend on the eigenvalues of $\boldsymbol{\Lambda}$ and $\sum_{i=1}^{n} a_i^2 = 1$. The result follows.

Size. The Wald statistic has the expansion,

$$W/K_2 = \frac{(\bar{v}_2/K_2)'(\bar{v}_1/K_2)^{-1}(\bar{v}_2/K_2)}{n[1-2\boldsymbol{\rho}'(\bar{v}_1/K_2)^{-1}(\bar{v}_2/K_2)+(\bar{v}_2/K_2)'(\bar{v}_1/K_2)^{-2}(\bar{v}_2/K_2)]} + o_p\left(\frac{1}{\sqrt{K_2}}\right). \qquad (6.6)$$

where $\bar{v}_1$ and $\bar{v}_2$ denote the terms up to order $O_p(1/\sqrt{K_2})$ in (6.3) and (6.4), respectively.

The leading term in this expansion is,

$$W/K_2 = \frac{\boldsymbol{\rho}'(\boldsymbol{\Lambda}+\boldsymbol{I})^{-1}\boldsymbol{\rho}}{n[1-2\boldsymbol{\rho}'(\boldsymbol{\Lambda}+\boldsymbol{I})^{-1}\boldsymbol{\rho}+\boldsymbol{\rho}'(\boldsymbol{\Lambda}+\boldsymbol{I})^{-2}\boldsymbol{\rho}]} + O_p\left(\frac{1}{\sqrt{K_2}}\right). \qquad (6.7)$$

This allows us to characterize the weak instrument set $\mathcal{W}_{size}$ to the order of approximation in (6.7):

***Proposition* 3**. For any $\boldsymbol{\rho}$, the leading term in the expansion for W/K_2 in (6.7) is nonincreasing in each eigenvalue of $\boldsymbol{\Lambda}$. Moreover, this leading term is maximized when $\boldsymbol{\rho}'\boldsymbol{\rho} = 1$.

Proof. Write $(\boldsymbol{\Lambda} + \boldsymbol{I})^{-1} = \boldsymbol{JHJ}'$, where $\boldsymbol{H}$ is the diagonal matrix with diagonal elements $1/(1+\lambda_i)$, where λ_i is the i^{th} eigenvector of $\boldsymbol{\Lambda}$ and $\boldsymbol{J}$ is the matrix of the associated eigenvectors of $\boldsymbol{\Lambda}$. Then the leading term in (6.7) can be written, $h = \boldsymbol{a}'\boldsymbol{Ha}/\{n[\boldsymbol{a}'(\boldsymbol{I} - \boldsymbol{H})^2\boldsymbol{a} + (1-\boldsymbol{a}'\boldsymbol{a})]\}$, where $\boldsymbol{a} = \boldsymbol{H\rho}$. Differentiating h with respect to λ_j shows that $\partial h/\partial\lambda_j \le 0$, $j = 1,\ldots, n$, proving the first statement. Because $\boldsymbol{a}'\boldsymbol{a} \le 1$, $\partial h/\partial a_j^2 \ge 0$, $j = 1,\ldots, n$, for any $\boldsymbol{a}$, h can be increased by increasing $\boldsymbol{a}$ proportionately, or equivalently by increasing $\boldsymbol{\rho}$ proportionately until $\boldsymbol{\rho}'\boldsymbol{\rho} = 1$. It follows that h is maximized when $\boldsymbol{\rho}'\boldsymbol{\rho} = 1$.

The leading terms in the expansions (6.5) and (6.7) explain the curious feature that the boundary of $\mathcal{W}_{\text{bias}}$ is constant for large K_2 while it essentially increases linearly in K_2 for $\mathcal{W}_{\text{size}}$. First consider the expression (6.5) for B^2. The leading term is constant for a given $\boldsymbol{\rho}$ and $\boldsymbol{\Lambda}$, and by Proposition 2, $B^{\max}$ depends only on the minimum eigenvalue of $\boldsymbol{\Lambda}$, up to the order of this leading term. But $\boldsymbol{\Lambda}$, and thus $B^{\max}$, does not depend on K_2. Similarly, the minimum eigenvalue of $\boldsymbol{\Lambda}$ does not depend on n. Thus, for K_2 sufficiently large, $B^{\max}$ depends only on the minimum eigenvalue of $\boldsymbol{\Lambda}$, not on K_2 or n.

Next turn to the expression (6.7) for W/K_2. Calculations reveal that this leading term, maximized over ρ, depends on n although it does not depend on K_2. Because W is divided by K_2 in this expression, (6.7) alternatively implies that $W = cK_2 + O_p(\sqrt{K_2})$, where c depends on ρ, the eigenvalues of Λ, and n. This explains the effectively linear increase in the boundary of $\mathcal{W}_{\text{size}}$ and its dependence on n.

7. Conclusions

The procedure proposed here is simple to implement: it entails comparing the minimum eigenvalue of the first stage F-statistic matrix – the multivariate generalization of the first stage F-statistic – to a critical value. The critical value is determined by the number of instruments K_2, the number of included endogenous regressors n, and the researcher's willingness to tolerate relative bias or size distortions. The test statistic is the same whether one focuses on size or relative bias; all that differs is the critical value.

Viewed as a test, the procedure has good power (especially so when the number of instruments is large). Viewed as a decision rule, the procedure discriminates between weak and strong instruments quite effectively, with the region of ambiguity decreasing with the number of instruments.

When there is a single included endogenous variable, this procedure provides a refinement and improvement to Staiger and Stock's (1997) suggested 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 (bias) or 4 (size). Those critical values indicate that their rule of thumb can be interpreted as a 5% test of the hypothesis that the maximum relative bias is (approximately) 10%. However, their rule of thumb does not ensure that TSLS-based Wald statistics will have good size.

The results in this paper have 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 $\lambda'\lambda/K_2$. This was justified for moderate K_2 using the expansion in Section 6 and numerical analysis suggests it is true for all K_2, but this remains to be proven. Second, the lower bound of the power function in Section 6 is based on the similar 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 testing procedure described here focuses on TSLS. One extension of this research would be to consider other estimators and other Wald statistics, particularly the LIML Wald statistic. LIML and its Wald statistic have better-behaved weak instrument sampling distributions that TSLS and its Wald statistic, so LIML-based inference might produce tighter boundaries of the weak instrument region.

Second, 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.

Third, the asymptotic expansion in Section 6 could be extended in several ways. The expansion was for the classical fixed instrument/normal error model. Relaxing this assumption, as Rothenberg (1983) did for the Π fixed case, would complicate the expressions only somewhat to order $o_p(1/\sqrt{K_2})$ and would generalize the result considerably. It also remains to explore the quality of the resulting distributional approximations to the finite sample distribution of the TSLS estimator.

Endnotes

[1] Cragg and Donald (1993) propose a test of the null hypothesis that the coefficients of interest are underidentified, against the alternative that they are identified. Our purpose is quite different: just because the parameters are identified does not mean that an applied researcher can reliably use use the conventional first order asymptotic approximations to the limiting distributions of TSLS statistics.

[2] The definition of $\boldsymbol{G}_T$ in (2.6) is $\boldsymbol{G}_T$ in Staiger and Stock (1997, eq. (3.4)), divided by K_2 to put it in F-statistic form.

[3] Note that $\boldsymbol{\Sigma}_{\boldsymbol{Y}^\perp\boldsymbol{Y}^\perp} = \boldsymbol{\Sigma}_{VV}$ under Assumption L_Π.

[4] The assumption that $B^{max}(\ell;K_2,n)$ is nonincreasing in ℓ, needed to invert the function, has the effect of potentially including in (3.12) some instruments that are not weak. Because 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 very large. For this reason the set $\mathcal{W}_{bias}$ is potentially conservative. This comment applies to the size-based set developed in Section 3.3 as well.

[5] Figure 1 previously appeared in Staiger and Stock (1997) and is included here for completeness.

[6] We have not located a formal statement of this in the literature but numerical investigations suggest that it is true.

References

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

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

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.

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.

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

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

Table 1: Critical values of minEval G_T/K_2 at the 5% significance level ($n = 1$)

	Relative bias			
K_2	0.05	0.10	0.20	0.30
3	13.94	9.11	6.49	5.41
4	16.87	10.29	6.72	5.35
5	18.39	10.84	6.78	5.25
6	19.28	11.13	6.77	5.16
7	19.86	11.30	6.73	5.07
8	20.25	11.40	6.69	4.99
9	20.53	11.46	6.65	4.92
10	20.74	11.49	6.61	4.86
11	20.89	11.51	6.57	4.80
12	21.01	11.52	6.53	4.75
13	21.10	11.52	6.49	4.71
14	21.17	11.51	6.45	4.67
15	21.23	11.51	6.42	4.63
16	21.27	11.50	6.39	4.59
17	21.30	11.49	6.36	4.56
18	21.33	11.47	6.33	4.53
19	21.35	11.46	6.31	4.50
20	21.37	11.45	6.28	4.48
21	21.38	11.43	6.26	4.45
22	21.39	11.42	6.24	4.43
23	21.40	11.40	6.21	4.41
24	21.41	11.39	6.19	4.39
25	21.41	11.38	6.17	4.37
26	21.41	11.36	6.16	4.35
27	21.41	11.35	6.14	4.33
28	21.41	11.34	6.12	4.32
29	21.41	11.32	6.11	4.30
30	21.41	11.31	6.09	4.29
31	21.40	11.30	6.08	4.27
32	21.40	11.29	6.06	4.26
33	21.40	11.27	6.05	4.25
34	21.39	11.26	6.04	4.23
35	21.39	11.25	6.02	4.22
36	21.38	11.24	6.01	4.21
37	21.38	11.23	6.00	4.20
38	21.37	11.22	5.99	4.19
39	21.37	11.21	5.98	4.18
40	21.36	11.20	5.97	4.17
41	21.36	11.19	5.96	4.16
42	21.35	11.18	5.95	4.15
43	21.34	11.17	5.94	4.14
44	21.34	11.16	5.93	4.13
45	21.33	11.15	5.92	4.13
46	21.33	11.14	5.91	4.12
47	21.32	11.14	5.90	4.11
48	21.31	11.13	5.89	4.10
49	21.31	11.12	5.89	4.09
50	21.30	11.11	5.88	4.09
51	21.30	11.10	5.87	4.08
52	21.29	11.10	5.86	4.07
53	21.28	11.09	5.86	4.07
54	21.28	11.08	5.85	4.06
55	21.27	11.07	5.84	4.05
56	21.27	11.07	5.83	4.05
57	21.26	11.06	5.83	4.04
58	21.26	11.05	5.82	4.04
59	21.25	11.05	5.82	4.03
60	21.24	11.04	5.81	4.03
61	21.24	11.03	5.80	4.02
62	21.23	11.03	5.80	4.01
63	21.23	11.02	5.79	4.01
64	21.22	11.01	5.79	4.00
65	21.22	11.01	5.78	4.00
66	21.21	11.00	5.78	3.99
67	21.21	11.00	5.77	3.99
68	21.20	10.99	5.77	3.99
69	21.20	10.98	5.76	3.98
70	21.19	10.98	5.76	3.98
71	21.19	10.97	5.75	3.97
72	21.18	10.97	5.75	3.97
73	21.18	10.96	5.74	3.96
74	21.17	10.96	5.74	3.96
75	21.17	10.95	5.73	3.96

Table 2: Critical values of minEval G_T/K_2 at the 5% significance level ($n = 2$)

	Relative bias			
K_2	0.05	0.10	0.20	0.30
4	10.99	7.57	5.60	4.75
5	13.93	8.79	5.93	4.80
6	15.70	9.49	6.09	4.79
7	16.87	9.93	6.17	4.77
8	17.69	10.22	6.21	4.73
9	18.29	10.43	6.23	4.70
10	18.75	10.58	6.24	4.66
11	19.11	10.69	6.23	4.62
12	19.40	10.78	6.22	4.59
13	19.63	10.84	6.21	4.56
14	19.82	10.89	6.20	4.53
15	19.98	10.93	6.19	4.50
16	20.12	10.96	6.17	4.48
17	20.23	10.99	6.16	4.45
18	20.33	11.00	6.14	4.43
19	20.41	11.02	6.13	4.41
20	20.48	11.03	6.11	4.39
21	20.55	11.04	6.10	4.37
22	20.60	11.05	6.08	4.35
23	20.65	11.05	6.07	4.33
24	20.69	11.05	6.06	4.31
25	20.73	11.05	6.04	4.30
26	20.77	11.05	6.03	4.28
27	20.80	11.05	6.02	4.27
28	20.82	11.05	6.01	4.25
29	20.85	11.05	5.99	4.24
30	20.87	11.05	5.98	4.23
31	20.89	11.05	5.97	4.22
32	20.91	11.04	5.96	4.20
33	20.92	11.04	5.95	4.19
34	20.94	11.04	5.94	4.18
35	20.95	11.03	5.93	4.17
36	20.96	11.03	5.92	4.16
37	20.97	11.03	5.91	4.15
38	20.98	11.02	5.91	4.14
39	20.99	11.02	5.90	4.13
40	20.99	11.01	5.89	4.13
41	21.00	11.01	5.88	4.12
42	21.01	11.00	5.87	4.11
43	21.01	11.00	5.87	4.10
44	21.02	11.00	5.86	4.09
45	21.02	10.99	5.85	4.09
46	21.03	10.99	5.84	4.08
47	21.03	10.98	5.84	4.07
48	21.03	10.98	5.83	4.07
49	21.03	10.97	5.82	4.06
50	21.04	10.97	5.82	4.05
51	21.04	10.96	5.81	4.05
52	21.04	10.96	5.81	4.04
53	21.04	10.96	5.80	4.03
54	21.04	10.95	5.79	4.03
55	21.04	10.95	5.79	4.02
56	21.04	10.94	5.78	4.02
57	21.05	10.94	5.78	4.01
58	21.05	10.93	5.77	4.01
59	21.05	10.93	5.77	4.00
60	21.05	10.93	5.76	4.00
61	21.05	10.92	5.76	3.99
62	21.05	10.92	5.75	3.99
63	21.05	10.91	5.75	3.98
64	21.05	10.91	5.74	3.98
65	21.04	10.91	5.74	3.97
66	21.04	10.90	5.73	3.97
67	21.04	10.90	5.73	3.96
68	21.04	10.89	5.72	3.96
69	21.04	10.89	5.72	3.96
70	21.04	10.89	5.71	3.95
71	21.04	10.88	5.71	3.95
72	21.04	10.88	5.71	3.94
73	21.04	10.88	5.70	3.94
74	21.04	10.87	5.70	3.94
75	21.04	10.87	5.69	3.93

Table 3: Critical values of minEval G_T/K_2 at the 5% significance level ($n = 3$)

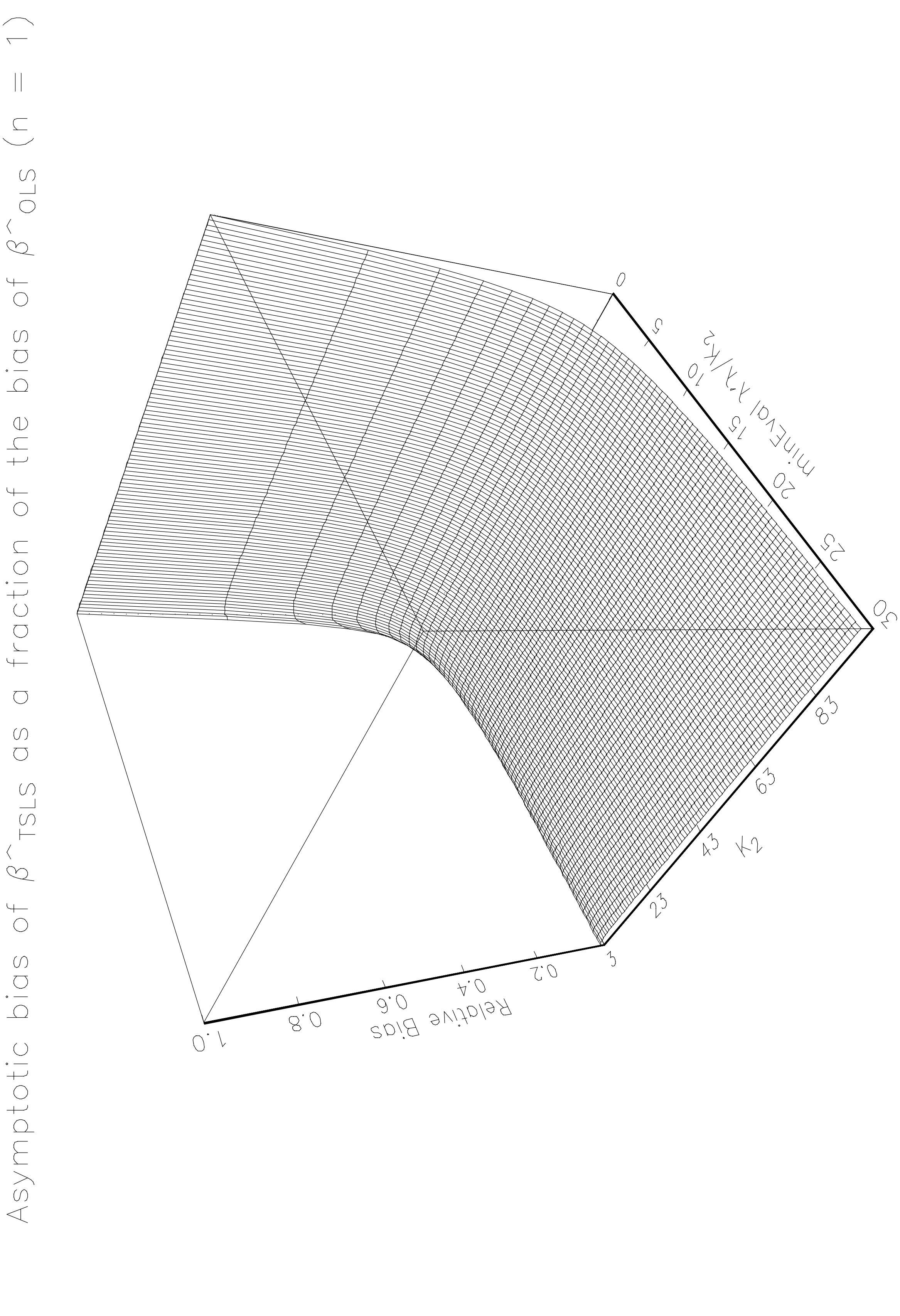
	Relative bias			
K_2	0.05	0.10	0.20	0.30
5	9.46	6.63	5.01	4.32
6	12.15	7.78	5.36	4.42
7	13.91	8.51	5.57	4.45
8	15.16	9.01	5.70	4.47
9	16.08	9.38	5.78	4.46
10	16.79	9.65	5.84	4.45
11	17.34	9.85	5.88	4.44
12	17.79	10.01	5.90	4.42
13	18.16	10.14	5.92	4.41
14	18.47	10.25	5.93	4.39
15	18.73	10.34	5.94	4.37
16	18.95	10.41	5.94	4.35
17	19.14	10.47	5.94	4.34
18	19.30	10.52	5.94	4.32
19	19.44	10.56	5.93	4.31
20	19.57	10.60	5.93	4.29
21	19.68	10.63	5.92	4.28
22	19.78	10.65	5.92	4.26
23	19.87	10.68	5.91	4.25
24	19.95	10.70	5.91	4.24
25	20.02	10.71	5.90	4.22
26	20.08	10.73	5.89	4.21
27	20.14	10.74	5.89	4.20
28	20.19	10.75	5.88	4.19
29	20.24	10.76	5.87	4.18
30	20.29	10.77	5.87	4.17
31	20.33	10.78	5.86	4.16
32	20.36	10.78	5.85	4.15
33	20.40	10.79	5.85	4.14
34	20.43	10.79	5.84	4.13
35	20.46	10.80	5.83	4.12
36	20.49	10.80	5.83	4.11
37	20.51	10.80	5.82	4.10
38	20.53	10.81	5.82	4.10
39	20.55	10.81	5.81	4.09
40	20.57	10.81	5.80	4.08
41	20.59	10.81	5.80	4.07
42	20.61	10.81	5.79	4.07
43	20.63	10.81	5.79	4.06
44	20.64	10.81	5.78	4.05
45	20.66	10.81	5.78	4.05
46	20.67	10.81	5.77	4.04
47	20.68	10.81	5.77	4.03
48	20.69	10.81	5.76	4.03
49	20.70	10.81	5.76	4.02
50	20.71	10.81	5.75	4.02
51	20.72	10.81	5.75	4.01
52	20.73	10.81	5.74	4.01
53	20.74	10.80	5.74	4.00
54	20.75	10.80	5.73	4.00
55	20.75	10.80	5.73	3.99
56	20.76	10.80	5.72	3.99
57	20.77	10.80	5.72	3.98
58	20.77	10.80	5.71	3.98
59	20.78	10.80	5.71	3.97
60	20.79	10.79	5.71	3.97
61	20.79	10.79	5.70	3.96
62	20.80	10.79	5.70	3.96
63	20.80	10.79	5.69	3.95
64	20.80	10.79	5.69	3.95
65	20.81	10.79	5.69	3.95
66	20.81	10.78	5.68	3.94
67	20.82	10.78	5.68	3.94
68	20.82	10.78	5.67	3.93
69	20.82	10.78	5.67	3.93
70	20.83	10.78	5.67	3.93
71	20.83	10.77	5.66	3.92
72	20.83	10.77	5.66	3.92
73	20.83	10.77	5.66	3.92
74	20.84	10.77	5.65	3.91
75	20.84	10.77	5.65	3.91

Table 4: Critical values of minEval G_T/K_2 at the 5% significance level ($n = 1$)

	Size distortion			
K_2	0.10	0.15	0.20	0.25
1	16.52	8.88	6.96	6.78
2	19.84	11.60	8.75	7.41
3	22.18	12.86	9.50	7.79
4	24.46	14.00	10.20	8.23
5	26.77	15.13	10.92	8.72
6	29.10	16.28	11.67	9.25
7	31.45	17.43	12.42	9.80
8	33.81	18.59	13.19	10.37
9	36.19	19.76	13.96	10.94
10	38.57	20.94	14.75	11.53
11	40.96	22.12	15.53	12.12
12	43.35	23.30	16.32	12.71
13	45.75	24.49	17.12	13.31
14	48.15	25.67	17.92	13.91
15	50.55	26.87	18.71	14.52
16	52.96	28.06	19.52	15.12
17	55.37	29.25	20.32	15.73
18	57.78	30.45	21.12	16.34
19	60.19	31.65	21.93	16.95
20	62.61	32.85	22.73	17.56
21	65.02	34.04	23.54	18.17
22	67.44	35.24	24.35	18.79
23	—	36.44	25.16	19.40
24	—	37.65	25.97	20.01
25	—	38.85	26.78	20.63
26	—	40.05	27.59	21.24
27	—	41.25	28.40	21.86
28	—	42.45	29.21	22.48
29	—	43.66	30.02	23.09
30	—	44.86	30.83	23.71
31	—	46.07	31.64	24.33
32	—	47.27	32.46	24.94
33	—	—	33.27	25.56
34	—	—	34.08	26.18
35	—	—	34.89	26.80
36	—	—	35.71	27.42
37	—	—	36.52	28.04
38	—	—	37.33	28.65
39	—	—	38.15	29.27
40	—	—	38.96	29.89

Table 5: Critical values of minEval G_T/K_2 at the 5% significance level ($n = 2$)

	Size distortion			
K_2	0.10	0.15	0.20	0.25
2	6.96	5.32	5.17	5.02
3	13.34	8.21	6.60	5.75
4	16.78	9.76	7.45	6.26
5	19.38	10.98	8.16	6.73
6	21.63	12.05	8.83	7.20
7	23.69	13.06	9.48	7.66
8	25.64	14.03	10.11	8.13
9	27.53	14.98	10.74	8.60
10	29.37	15.92	11.37	9.07
11	31.19	16.84	12.00	9.54
12	32.98	17.77	12.63	10.02
13	34.75	18.68	13.25	10.49
14	36.52	19.60	13.88	10.97
15	38.27	20.51	14.50	11.45
16	40.01	21.41	15.13	11.93
17	41.75	22.32	15.76	12.41
18	43.49	23.23	16.38	12.89
19	45.22	24.13	17.01	13.37
20	46.94	25.04	17.63	13.86
21	48.67	25.94	18.26	14.34
22	50.39	26.84	18.89	14.82
23	52.11	27.75	19.51	15.31
24	53.82	28.65	20.14	15.79
25	55.54	29.55	20.77	16.27
26	57.25	30.45	21.40	16.76
27	—	31.35	22.02	17.24
28	—	32.26	22.65	17.73
29	—	33.16	23.28	18.21
30	—	34.06	23.91	18.70
31	—	34.96	24.53	19.18
32	—	35.86	25.16	19.67
33	—	36.76	25.79	20.15
34	—	37.66	26.42	20.64
35	—	38.56	27.04	21.13
36	—	39.46	27.67	21.61
37	—	40.36	28.30	22.10
38	—	—	28.93	22.58
39	—	—	29.56	23.07
40	—	—	30.19	23.56



Asymptotic bias of β^TSLS as a fraction of the bias of β^OLS (n = 1)
Relative Bias
1.0
0.8
0.6
0.4
0.2
3
23
43
63
83
K2
minEval λ'λ/K2
0
5
10
15
20
25
30

Figure 1:

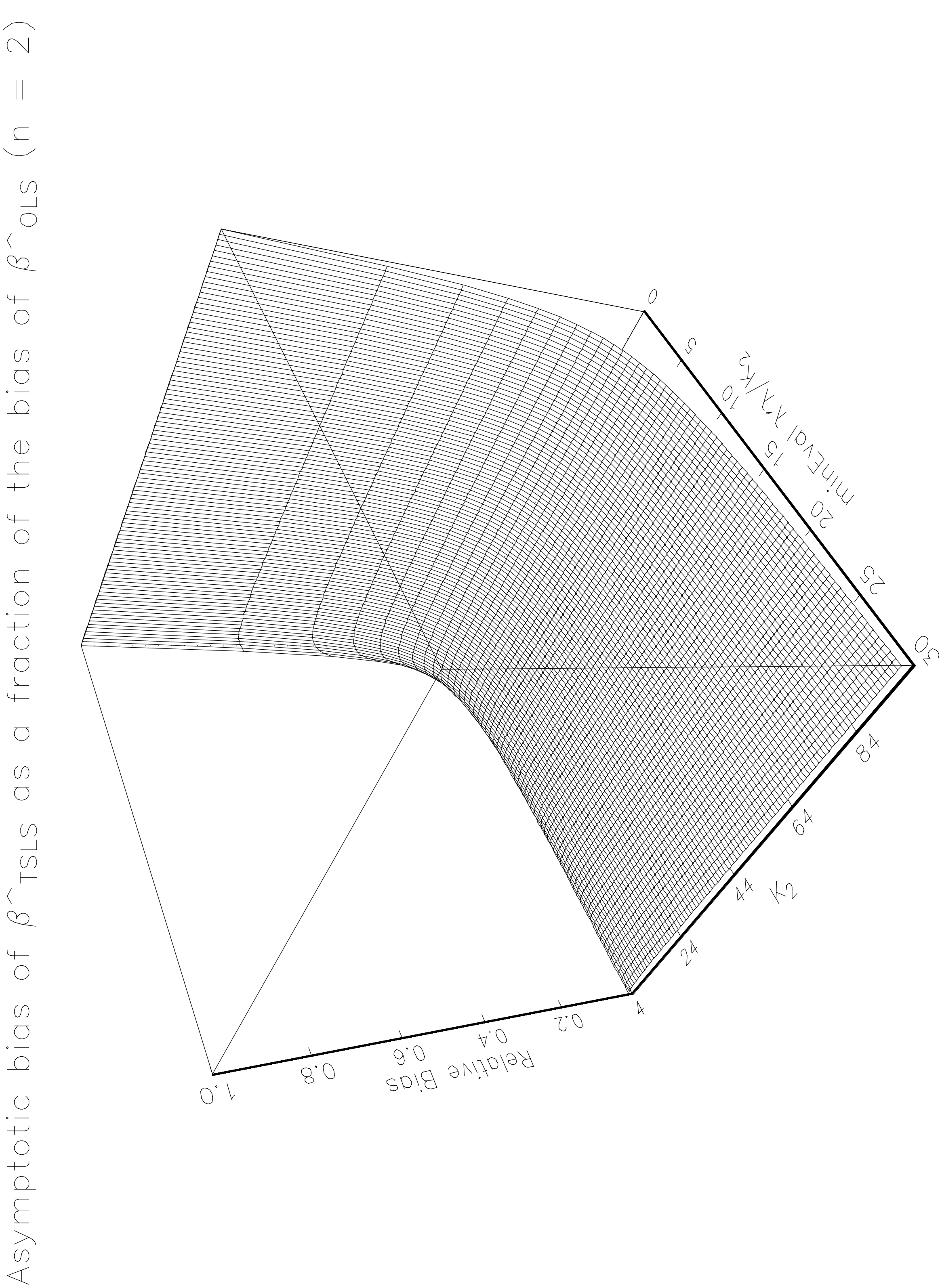

Asymptotic bias of β^TSLS as a fraction of the bias of β^OLS (n = 2)
Relative Bias
1.0
0.8
0.6
0.4
0.2
4
24
44
64
84
K2
minEval λ'λ/K2
0
5
10
15
20
25
30

Figure 2:

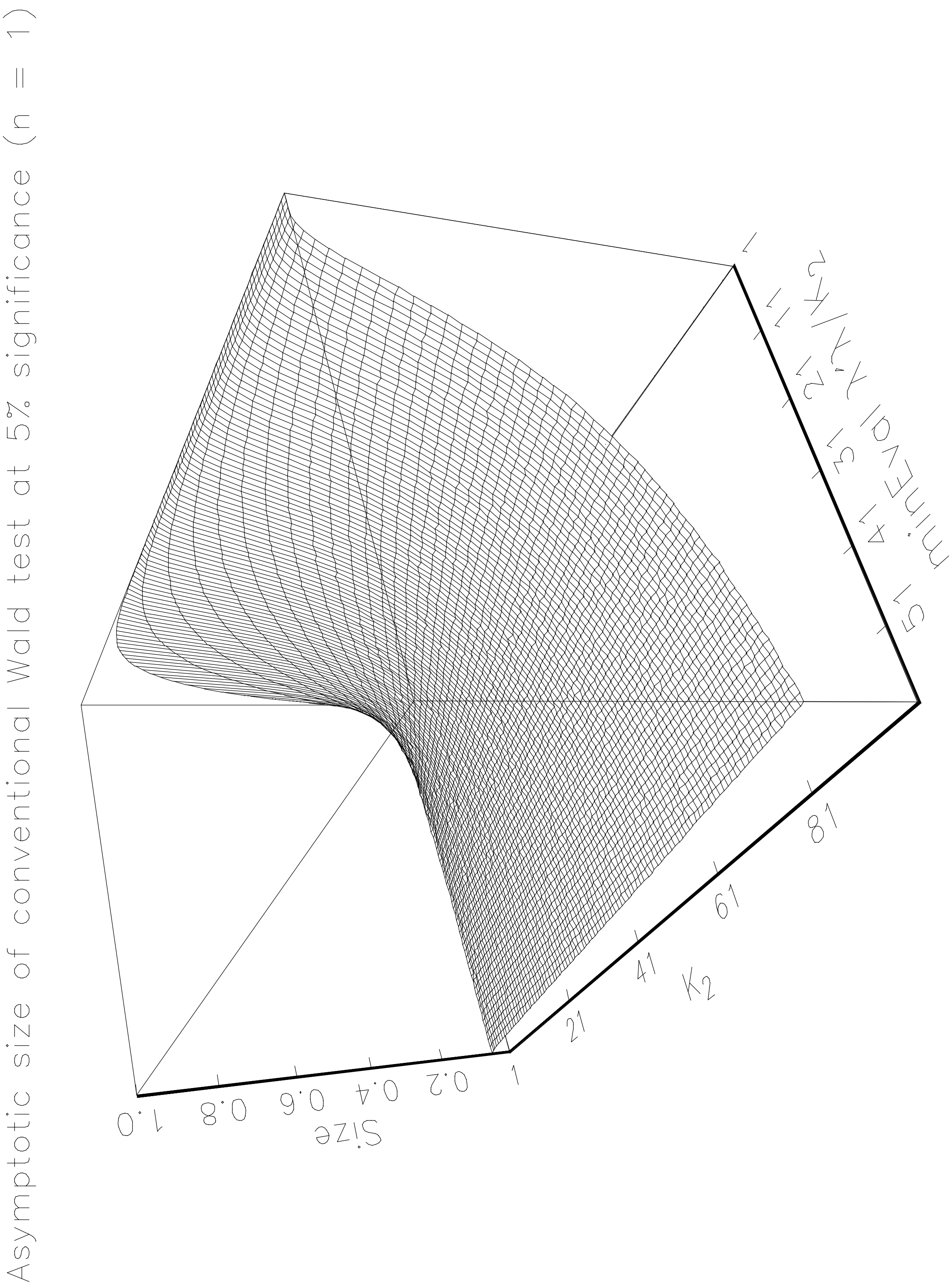

Asymptotic size of conventional Wald test at 5% significance (n = 1)
Size
0.2 0.4 0.6 0.8 1.0
K2
1 21 41 61 81
minEval λ'λ/K2
1 11 21 31 41 51

Figure 3:

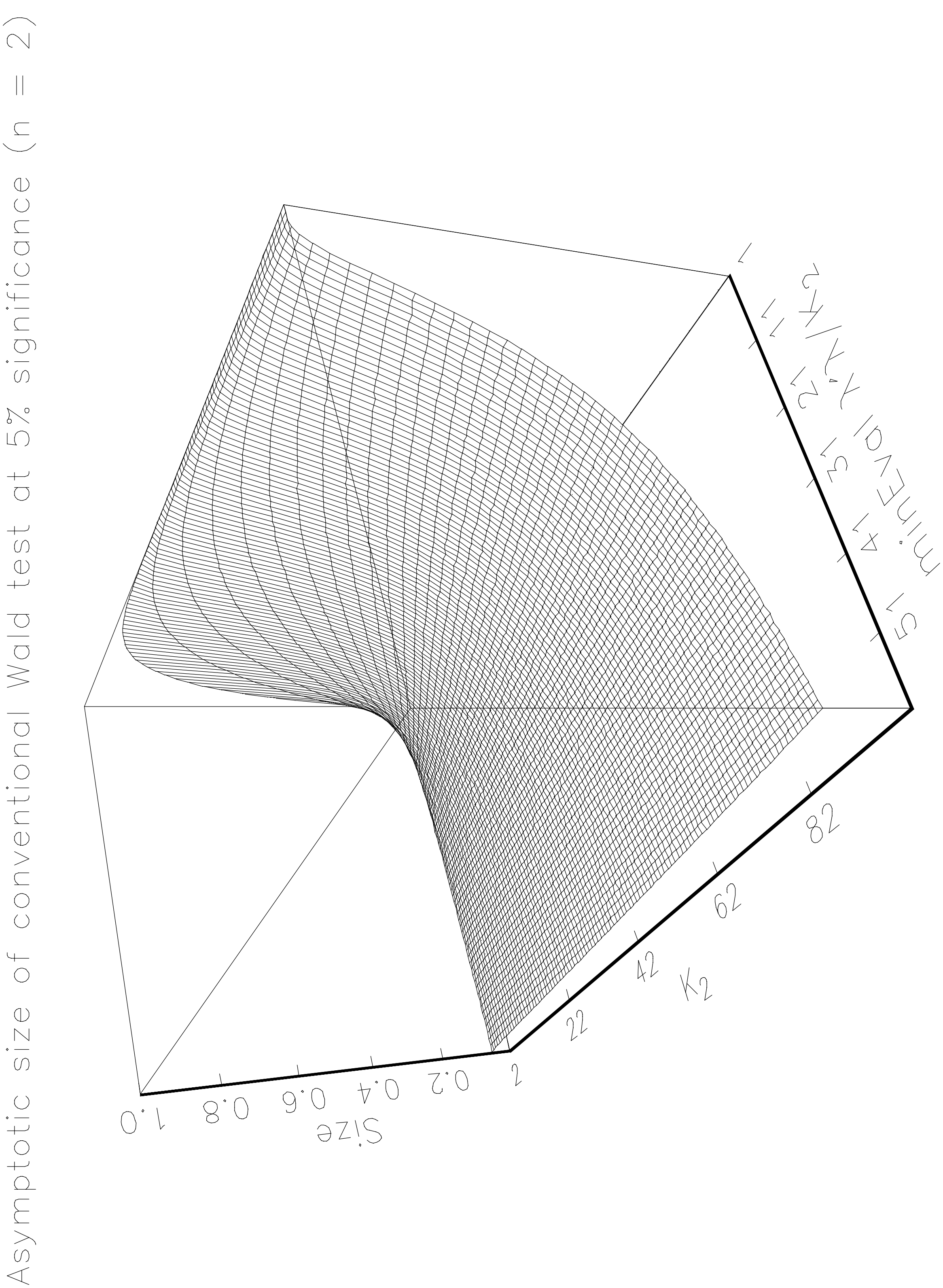

Asymptotic size of conventional Wald test at 5% significance (n = 2)
Size
1.0
0.8
0.6
0.4
0.2
K2
2
22
42
62
82
minEval λ'λ/K2
1
11
21
31
41
51

Figure 4:

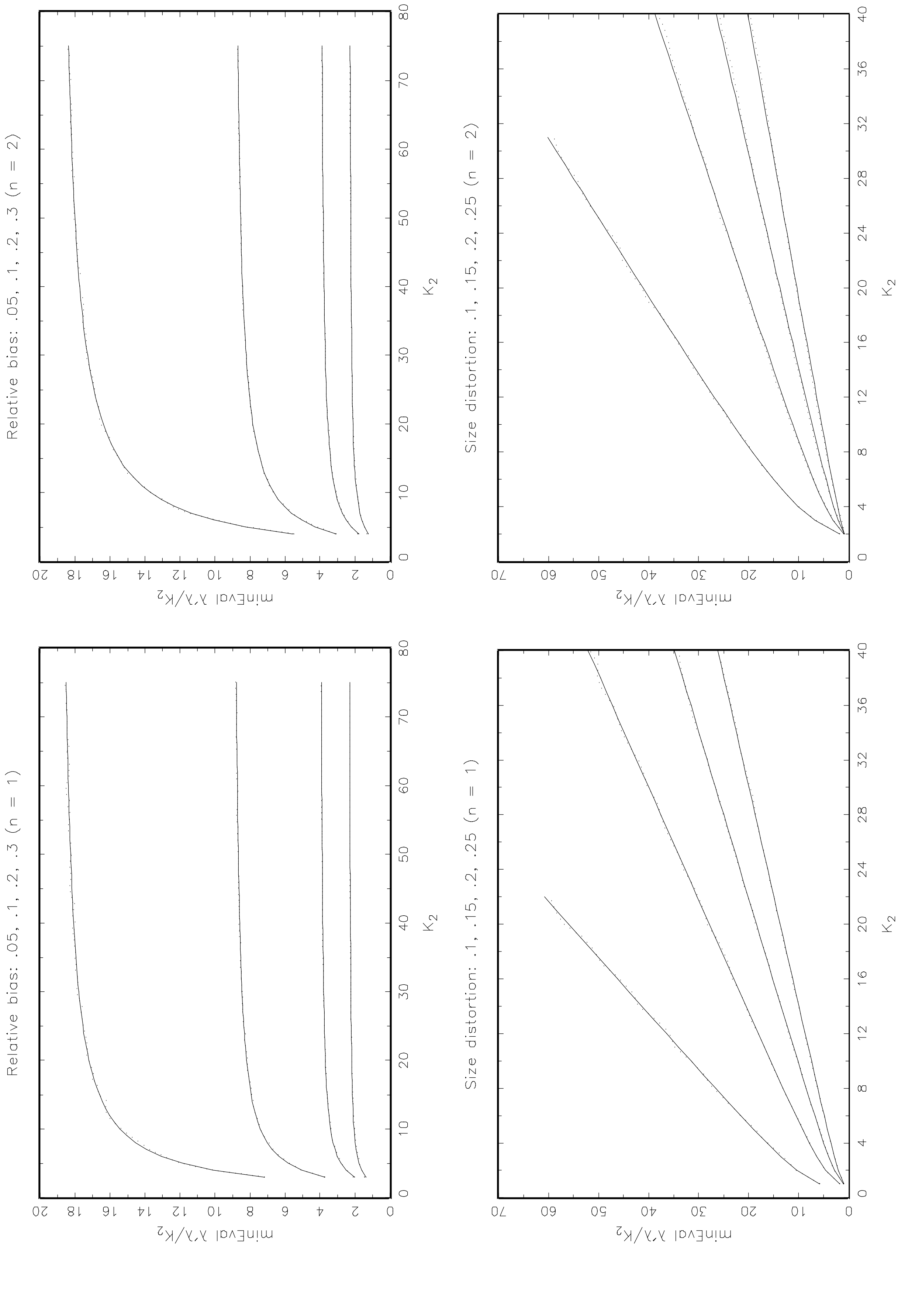


Figure 5: Boundary of the weak instrument set

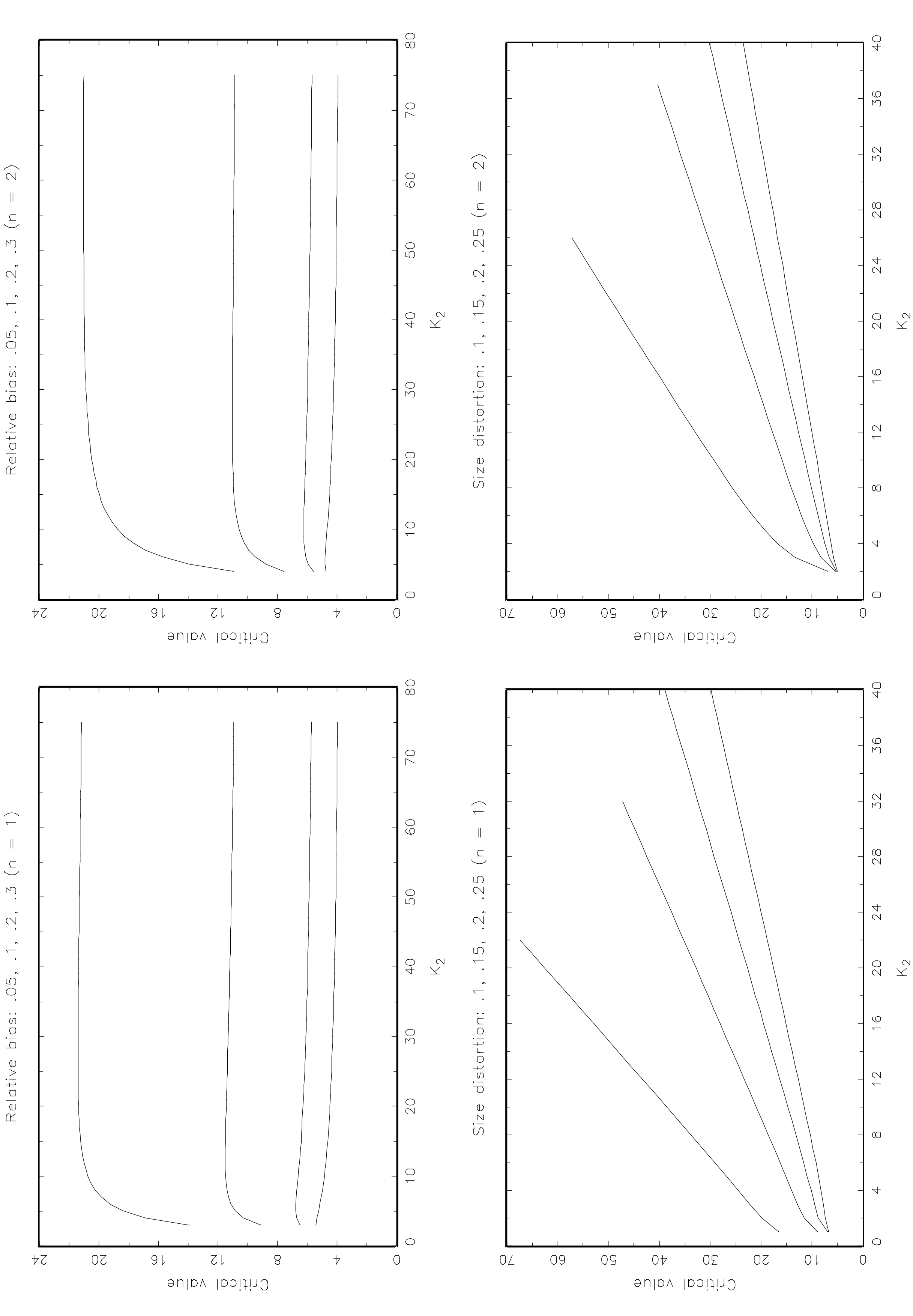


Figure 6: Critical values of minEval G_T/K_2 at the 5% significance level

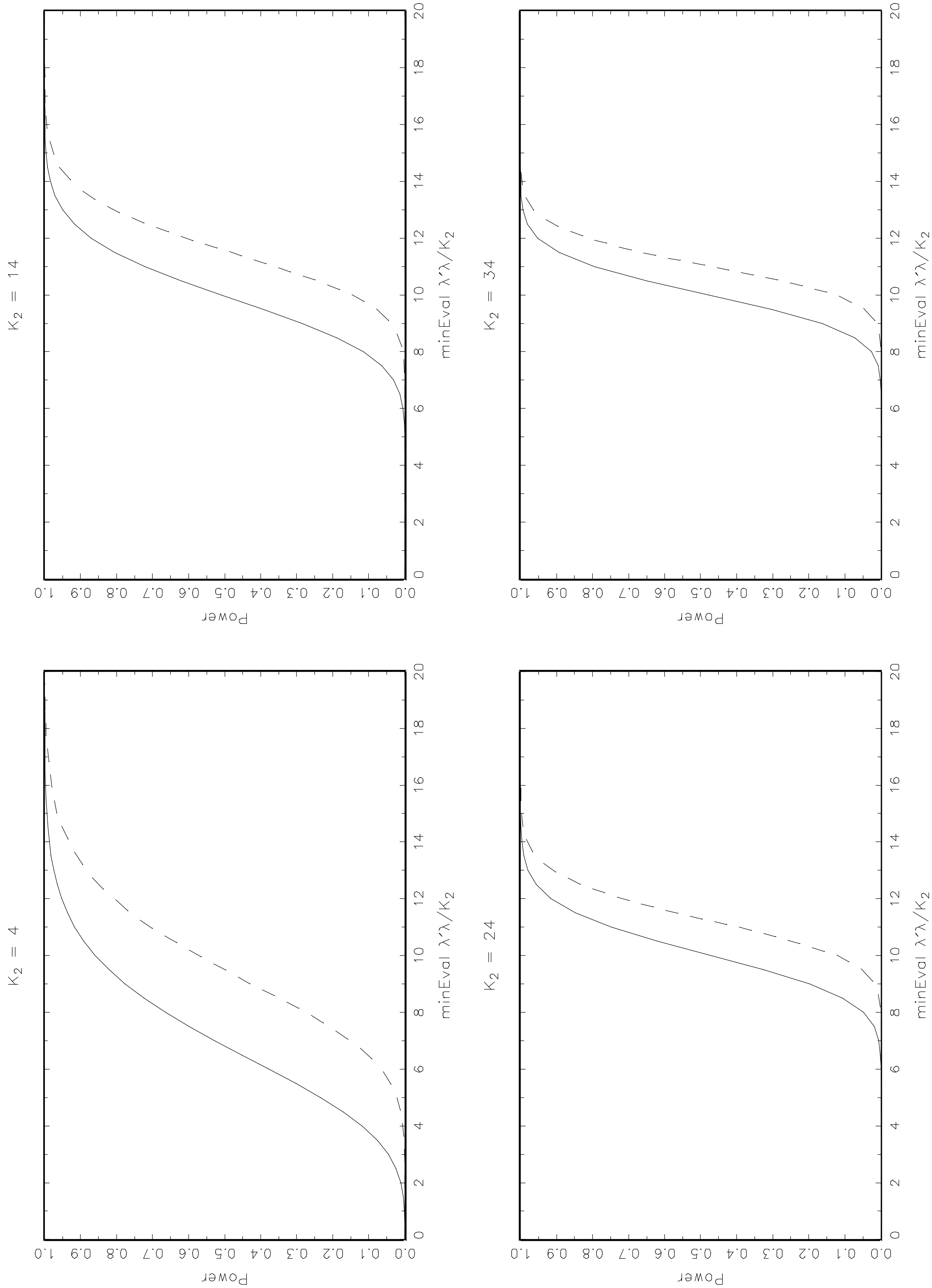


Figure 7: Bounds of the power function when relative bias is 0.1 (n = 2)

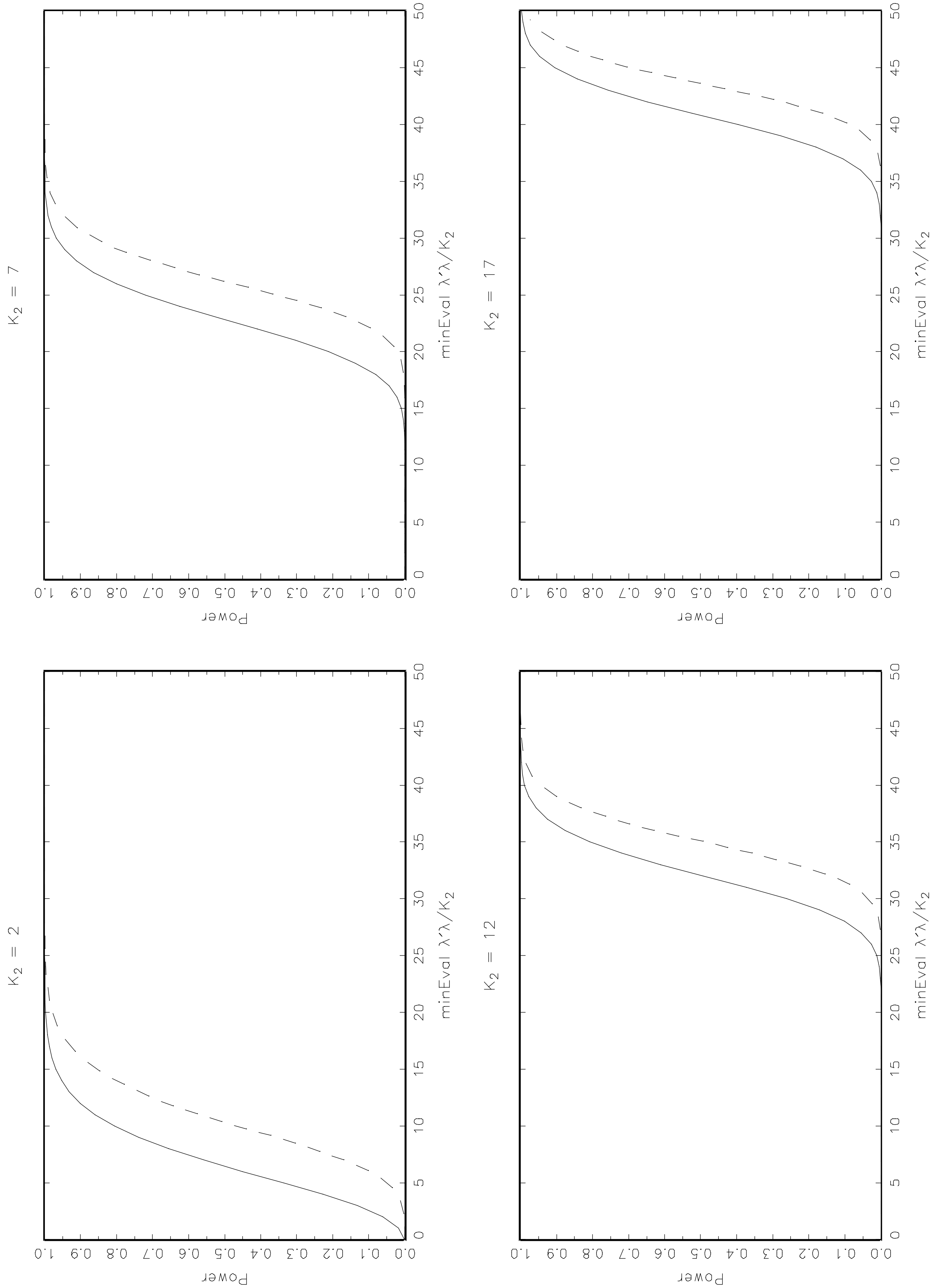


Figure 8: Bounds of the power function when size distortion is 0.1 (n = 2)