A WEAK INSTRUMENT F-TEST IN LINEAR IV MODELS WITH MULTIPLE ENDOGENOUS VARIABLES

Eleanor Sanderson
Frank Windmeijer

Discussion Paper 15 / 644

30 June 2014
Revised 27 May 2015

Department of Economics
University of Bristol
8 Woodland Road
Bristol BS8 1TN
United Kingdom

A Weak Instrument F-Test in Linear IV Models with Multiple Endogenous Variables*

Eleanor Sanderson[a,b] and Frank Windmeijer[a,b,c]
[a]Department of Economics, University of Bristol, UK
[b]CMPO and IEU, University of Bristol, UK
[c]CEMMAP/IFS, London, UK

April 2015

Abstract

We consider testing for weak instruments in a model with multiple endogenous variables. Unlike Stock and Yogo (2005), who considered a weak instruments problem where the rank of the matrix of reduced form parameters is near zero, here we consider a weak instruments problem of a near rank reduction of one in the matrix of reduced form parameters. For example, in a two-variable model, we consider weak instrument asymptotics of the form $\pi_1 = \delta\pi_2 + c/\sqrt{n}$ where π_1 and π_2 are the parameters in the two reduced-form equations, c is a vector of constants and n is the sample size. We investigate the use of a conditional first-stage F-statistic along the lines of the proposal by Angrist and Pischke (2009) and show that, unless $\delta = 0$, the variance in the denominator of their F-statistic needs to be adjusted in order to get a correct asymptotic distribution when testing the hypothesis $H_0 : \pi_1 = \delta\pi_2$. We show that a corrected conditional F-statistic is equivalent to the Cragg and Donald (1993) minimum eigenvalue rank test statistic, and is informative about the maximum total relative bias of the 2SLS estimator and the Wald tests size distortions. When $\delta = 0$ in the two-variable model, or when there are more than two endogenous variables, further information over and above the Cragg-Donald statistic can be obtained about the nature of the weak instrument problem by computing the conditional first-stage F-statistics.

Key Words: weak instruments, multiple endogenous variables, F-test
JEL Codes: C12, C36

*This research was funded by the Economic and Social Research Council (RES-343-28-0001), the European Research Council (DEVHEALTH-269874) and the Medical Research Council (MC_UU_12013/9). Helpful comments were provided by Jörn-Steffen Pischke, Mark Schaffer, Peter Schmidt, Chris Skeels, Jonathan Temple and two anonymous referees.

1 Introduction

Following the work of Staiger and Stock (1997) and Stock and Yogo (2005), testing for weak instruments is now commonplace. For a single endogenous variable model, the standard first-stage F-statistic can be used to test for weakness of instruments, where weakness is expressed in terms of the size of the bias of the IV estimator relative to that of the OLS estimator, or in terms of the magnitude of the size distortion of the Wald test for parameter hypotheses. Stock and Yogo (2005) tabulated critical values for the standard F-statistic that have been incorporated in software packages.

For multiple endogenous variables, inspection of the individual first-stage F-statistics is no longer sufficient. The Cragg-Donald (1993) statistic can be used to evaluate the overall strength of the instruments in this case, and Stock and Yogo (2005) have tabulated critical values of the minimum eigenvalue of the Cragg-Donald statistic for testing weakness of instruments. They derive the limiting distributions under weak instrument asymptotics where the reduced form parameters are local to zero in each reduced form equation, and obtain critical values that are conservative in the sense that they are rejecting the null of weak instruments too infrequently when the null is true.

In this paper, we are interested in analysing tests for weak instruments in a model with multiple endogenous variables in a setting where the reduced form parameters are not local to zero, but where the reduced form parameter matrix is local to a rank reduction of one. In this case, the values of the F-statistics in each of the first-stage equations can be high, but the identification of (some of) the model parameters is weak. We will focus initially on a model with two endogenous variables. The weak instrument asymptotics we consider are local to a rank reduction of one, of the form

$$\pi_1 = \delta\pi_2 + c/\sqrt{n},$$

where π_1 and π_2 are the parameters in the two reduced-form equations, c is a vector of constants and n is the sample size. We call these asymptotics LRR1 weak instrument asymptotics.

We will focus solely on the properties of the 2SLS estimator. We investigate the use of a conditional first-stage F-statistic along the lines of the proposal by Angrist and Pischke (2009) and show that the variance formula in the denominator of their F-statistic needs to be adjusted in order to get a correct asymptotic distribution when testing the null hypothesis, in the two-variable model, $H_0 : \pi_1 = \delta\pi_2$. We further show that the resulting new conditional F-statistic

is equivalent to the Cragg-Donald minimum eigenvalue statistic. Using our weak instrument asymptotics we show that this conditional F-statistic cannot be used in the same way as the Stock and Yogo (2005) procedure for a single endogenous variable to assess the magnitude of the relative bias of the 2SLS estimator of an individual structural parameter. This is because the OLS bias expression contains additional terms such that the expression for the bias of the 2SLS estimator relative to that of the OLS estimator does not have the the simple expression as in the one-variable case. However, the total relative bias can be bounded as can the size distortions of Wald tests on the structural parameters.

In a two-endogenous-variable model the conditional F-statistics for each reduced form are equivalent to each other and to the Cragg-Donald minimum eigenvalue statistic under our LRR1 weak instrument asymptotics. This holds unless $\delta = 0$, in which case the local rank reduction is due to the fact that π_1 is local to zero and the first-stage F-statistic for x_1 will be small and that for x_2 will be large. In this case, both the Angrist-Pischke F-statistic and our conditional F-statistic for x_1 can be assessed against the Stock-Yogo critical value, and the 2SLS estimator for the structural parameter on x_2 is consistent. Additional information can also be obtained from our conditional F-statistics when there are more than two endogenous variables, as they will identify which variables cause the near rank reduction. For example, if in a three variable model the near rank reduction is due to the reduced form parameters on two variables only, the conditional F-statistic for the third variable will remain large giving the researcher valuable information about the nature of the problem and directions for solving it. We also show that the 2SLS estimator for the structural parameter of the third variable is consistent in that case.

The paper is organised as follows. In Section 2 we introduce the linear model with one endogenous variable and summarise the Staiger and Stock (1997) and Stock and Yogo (2005) results for testing for weak instruments. Section 3 considers weak instrument test statistics for the linear model with two endogenous explanatory variables and introduces the new conditional F-tests. Section 4 considers the relative bias and Wald test size distortions for the 2SLS estimator under the LRR1 weak instrument asymptotics as outlined above and presents some Monte Carlo results for the two-variable model. Section 4 also shows the usefulness of the conditional F-test statistics in a model with more than two endogenous variables. Finally, Section 5 concludes.

2 Weak Instrument Asymptotics in One-Variable Model

In this section we follow the basic Staiger and Stock (1997) and Stock and Yogo (2005) setup. The developments of the weak instrument setup and concepts for the one-variable model play an important role when we expand the model to multiple endogenous variables in the next section. The simple model is

$$y = x\beta + u, \tag{1}$$

where y, x, and u are $n \times 1$ vectors, with n the number of observations. There is endogeneity, such that $E(u|x) \neq 0$. The reduced form for x is

$$x = Z\pi + v, \tag{2}$$

where Z is a $n \times k_z$ matrix of instruments and v is $n \times 1$. For individual u_i and v_i we assume,

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim (0, \Sigma); \quad \Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}.$$

The 2SLS estimator is given by

$$\widehat{\beta}_{2SLS} = \frac{x' P_Z y}{x' P_Z x} = \beta_0 + \frac{x' P_Z u}{x' P_Z x},$$

where $P_Z = Z(Z'Z)^{-1} Z'$.

The concentration parameter is given by

$$CP = \frac{\pi' Z' Z \pi}{\sigma_v^2}$$

and is a measure of the strength of the instruments, see Rothenberg (1984). A small concentration parameter is associated with a bias of the 2SLS estimator and deviations from its asymptotic normal distribution.

A simple test whether the instruments are related to x is of course a Wald or F-test for the hypothesis $H_0 : \pi = 0$. The Wald test is given by

$$W_\pi = \frac{\widehat{\pi}' Z' Z \widehat{\pi}}{\widehat{\sigma}_v^2} = \frac{x' Z (Z'Z)^{-1} Z' x}{\widehat{\sigma}_v^2},$$

where $\widehat{\pi} = (Z'Z)^{-1} Z'x$ is the first-stage OLS estimator, and $\widehat{\sigma}_v^2 = x' M_Z x / n$, where $M_Z = I - P_Z$. Under the null, $W_\pi \xrightarrow{d} \chi^2_{k_z}$. The F-test is given by $F = W_\pi / k_z$. Note that we refrain from a degrees of freedom correction in the variance estimate. Also, note that the analyses here

and further below extend to a model with additional exogenous regressors, as we can replace y, x and Z everywhere by their residuals from regressions on those exogenous regressors.

Staiger and Stock (1997) introduce weak instrument asymptotics as a local to zero alternative, $\pi = c/\sqrt{n}$, which ensures that the concentration parameter does not increase with the sample size

$$CP = \frac{\pi' Z' Z \pi}{\sigma_v^2} \xrightarrow{p} \frac{c' Q_{ZZ} c}{\sigma_v^2},$$

where $Q_{ZZ} = \text{plim}\left(n^{-1} Z'Z\right)$.

Assuming that conditions are fulfilled, such that

$$\begin{pmatrix} \frac{1}{\sqrt{n}} Z'u \\ \frac{1}{\sqrt{n}} Z'v \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_{Zu} \\ \psi_{Zv} \end{pmatrix} \sim N\left(0, \Sigma \otimes Q_{ZZ}\right),$$

and $k_z \geq 3$ when assessing relative bias. Then under weak instrument asymptotics,

$$\widehat{\beta}_{2SLS} - \beta = \frac{x' Z \left(Z'Z\right)^{-1} Z'u}{x' Z \left(Z'Z\right)^{-1} Z'x} \xrightarrow{d} \frac{\sigma_u}{\sigma_v} \frac{\left(\lambda + z_v\right)' z_u}{\left(\lambda + z_v\right)' \left(\lambda + z_v\right)}.$$

where

$$\lambda = \sigma_v^{-1} Q_{ZZ}^{1/2} c; \quad z_v = \sigma_v^{-1} Q_{ZZ}^{-1/2} \psi_{Zv}; \quad z_u = \sigma_u^{-1} Q_{ZZ}^{-1/2} \psi_{Zu}.$$

The bias of the OLS estimator is given by

$$\widehat{\beta}_{OLS} - \beta = \frac{x'u}{x'x} = \frac{\left(Z\pi + v\right)' u}{\left(Z\pi + v\right)' \left(Z\pi + v\right)} = \frac{n^{-1}\left(n^{-1/2} c' Z'u + v'u\right)}{n^{-1}\left(n^{-1} c' Z'Z c + 2n^{-1/2} c' Z'v + v'v\right)}$$

$$\xrightarrow{p} \frac{\text{plim}\, n^{-1} v'u}{\text{plim}\, n^{-1} v'v} = \frac{\sigma_{uv}}{\sigma_v^2} = \frac{\sigma_u}{\sigma_v} \rho,$$

where $\rho = \frac{\sigma_{uv}}{\sigma_u \sigma_v}$.

As a measure of relative bias, Stock and Yogo (2005) propose

$$B_n^2 = \left(\frac{E\left[\widehat{\beta}_{2SLS}\right] - \beta}{E\left[\widehat{\beta}_{OLS}\right] - \beta} \right)^2.$$

From the derivations above, and as $E\left[z_u | z_v\right] = \rho z_v$, it follows that

$$B_n^2 = \left(E\left[\frac{\left(\lambda + z_v\right)' z_v}{\left(\lambda + z_v\right)' \left(\lambda + z_v\right)} \right] \right)^2,$$

or

$$B_n = \left| E\left[\frac{\left(\lambda + z_v\right)' z_v}{\left(\lambda + z_v\right)' \left(\lambda + z_v\right)} \right] \right|.$$

Using weak instrument asymptotics, Stock and Yogo (2005) are therefore able to assess the size of the relative bias in relation to the first-stage F-statistic. As $z_v \sim N(0, I_{k_z})$, B_n is determined by the values of λ and k_z. Let

$$l = \lambda'\lambda/k_z = \frac{1}{k_z}\frac{c'Q_{ZZ}c}{\sigma_v^2},$$

then using Monte Carlo simulation, i.e. draws of $z_v \sim N(0, I_{k_z})$, Stock and Yogo (2005) find the values of l such that B_n is a certain value, say 0.1, for different values of k_z. For example, when $k_z = 4$ and using 100,000 Monte Carlo draws, we obtain a relative expected bias $E\left[\frac{(\lambda+z_v)'z_v}{(\lambda+z_v)'(\lambda+z_v)}\right] = 0.1$ for $l = 4.98$. When $k_z = 8$, we find $l = 7$, again for $B_n = 0.1$.

Using weak instrument asymptotics, Staiger and Stock (1997) derive the asymptotic distribution for the first-stage F-statistic, which is given by

$$F \xrightarrow{d} \chi^2_{k_z}(k_z l)/k_z,$$

where $\chi^2_{k_z}(a)$ is the non-central chi-squared distribution with non-centrality parameter a. The F-test statistic can therefore be used to test the hypothesis

$$H_0 : CP/k_z \leq l_b \quad \text{vs} \quad H_1 : CP/k_z > l_b,$$

where l_b is the value for l determined above such that the $B_n = b$. For $b = 0.1$, we find from the scaled non-central chi-squared distribution a critical values of 10.20 when $k_z = 4$ and 11.38 when $k_z = 8$. In comparison, Stock and Yogo (2005), henceforth SY, find very similar critical values of 10.27 and 11.39 for these two cases respectively.

As an illustration, we performed a small simulation. The model is as in (1) and (2), with $\beta = 1$;

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right);$$

the instruments in Z are four independent standard normally distributed random variables and $\pi = \begin{pmatrix} c & c & c & c \end{pmatrix}'/\sqrt{n}$, with c chosen such that the relative bias B_n for $n \to \infty$ is equal to 0.1, or 10%. We set the sample size $n = 10,000$ and show the results in Table 1 for $10,000$ Monte Carlo replications. The results are clearly in line with the theory. The observed relative bias is just over 10% and the rejection frequency of the F-test using the SY weak instrument critical value is 5% at the 5% nominal level.

Table 1. Estimation and relative bias results for one-variable model

	mean	st dev	rel bias	SY rej freq
$\widehat{\beta}_{OLS}$	1.4989	0.0086		
$\widehat{\beta}_{2SLS}$	1.0529	0.2173	0.1060	
F	5.97	2.36		0.0502

Notes: sample size 10,000; 10,000 MC replications; $\beta = 1$;
F is the first-stage F-statistic for x; rel bias is the relative bias of the 2SLS estimator, relative to that of the OLS estimator; SY rej freq uses the 5% Stock-Yogo critical value for the F-test for a 10% relative bias.

The Wald test for testing the restriction $H_0 : \beta = \beta_0$ is given by

$$W = \frac{\left(\widehat{\beta}_{2SLS} - \beta_0\right)^2 (x' P_Z x)}{\widehat{\sigma}_u^2},$$

where $\widehat{\sigma}_u^2 = \left(y - x\widehat{\beta}_{2SLS}\right)' \left(y - x\widehat{\beta}_{2SLS}\right)/n$. Staiger and Stock (1997) show that, under weak instrument asymptotics,

$$W \xrightarrow{d} \frac{\nu_2^2/\nu_1}{1 - 2\rho\nu_2/\nu_1 + (\nu_2/\nu_1)^2},$$

where

$$\nu_1 = (\lambda + z_v)' (\lambda + z_v); \quad \nu_2 = (\lambda + z_v)' z_u.$$

The Wald size distortion is maximised for $\rho = 1$, and SY find the critical values for the F-test such that the maximal size of the Wald test is a certain value, say 10%, at a nominal 5% level. For the Monte Carlo example above, we set $\rho = 1$ and choose c such that the maximal size distortion of the Wald test is 10%, in which case the value of l is given by 16.415. The SY critical value in this case is given by 24.58. The results are given in Table 2, and confirm that the size of the Wald test is 10% and the rejection frequency of the F-test using the SY critical values is indeed 5%.

Table 2. Estimation and Wald test results for one-variable model

	mean	st dev	rej freq	SY rej freq
$\widehat{\beta}_{OLS}$	1.9935	0.0008		
$\widehat{\beta}_{2SLS}$	1.0318	0.1184		
W	1.42	2.52	0.0994	
F	17.45	4.11		0.0501

Notes: sample size 10,000; 10,000 MC replications; $\beta = 1, \rho = 1$;
W is the Wald test for testing H_0: $\beta = 1$; rej freq uses 5% critical value of χ_1^2; SY rej freq uses the 5% Stock-Yogo critical value for the F-test, for a maximal 10% size of W.

3 Two Variable Model

Following the exposition in Angrist and Pischke (2009), we first consider the following two-variable model

$$\begin{aligned} y &= x_1\beta_1 + x_2\beta_2 + u \\ x_1 &= Z\pi_1 + v_1 \\ x_2 &= Z\pi_2 + v_2 \end{aligned} \tag{3}$$

where y, x_1, x_2, u, v_1 and v_2 are $n \times 1$ vectors, with n the number of observations. Z is an $n \times k_z$ matrix of instruments, with $k_z \geq 2$ ($k_z \geq 4$ when assessing relative bias), and π_1 and π_2 are $k_z \times 1$ vectors. For an individual observation i,

$$\begin{pmatrix} u_i \\ v_{1i} \\ v_{2i} \end{pmatrix} \sim \left(0, \begin{pmatrix} \sigma_u^2 & \sigma_{Vu}' \\ \sigma_{Vu} & \Sigma_V \end{pmatrix}\right); \quad \Sigma_V = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

Equivalently, we can write

$$\begin{aligned} y &= X\beta + u \\ X &= Z\Pi + V \end{aligned}$$

where $\beta = (\beta_1, \beta_2)'$; $X = \begin{pmatrix} x_1 & x_2 \end{pmatrix}$; $\Pi = \begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix}$ and $V = \begin{pmatrix} v_1 & v_2 \end{pmatrix}$. Further, let $x = \text{vec}(X)$, $\pi = \text{vec}(\Pi)$ and $v = \text{vec}(V)$.

The OLS estimates for π_j are denoted $\widehat{\pi}_j = (Z'Z)^{-1} Z'x_j$, $j = 1, 2$, and the estimated variances are given by

$$\widehat{\Sigma}_V = \begin{pmatrix} \widehat{\sigma}_1^2 & \widehat{\sigma}_{12} \\ \widehat{\sigma}_{12} & \widehat{\sigma}_2^2 \end{pmatrix} = \frac{1}{n}\widehat{V}'\widehat{V} = \frac{1}{n}\begin{pmatrix} \widehat{v}_1'\widehat{v}_1 & \widehat{v}_1'\widehat{v}_2 \\ \widehat{v}_1'\widehat{v}_2 & \widehat{v}_2'\widehat{v}_2 \end{pmatrix},$$

where $\widehat{V} = X - Z\widehat{\Pi}$.

The first-stage F-statistics are given by

$$F_j = \frac{\widehat{\pi}_j' Z'Z \widehat{\pi}_j}{k_z \widehat{\sigma}_j^2} = \frac{x_j' Z (Z'Z)^{-1} Z' x_j}{k_z \widehat{\sigma}_j^2}; \; j = 1, 2,$$

and $k_z F_j$ converges in distribution to a $\chi^2_{k_z}$ distribution under the null $H_0 : \pi_j = 0$. Significant first-stage F-statistics are clearly necessary, but not sufficient, for identification of β. For example, if $\pi_1 = \delta\pi_2 \neq 0$, both first-stage F-statistics will reject their null in large samples, but the model is clearly underidentified.

Staiger and Stock (1997) and Stock and Yogo (2005) consider weak instrument asymptotics where all reduced form parameters are local to zero, i.e. $\Pi = C/\sqrt{n}$. The Wald test for $H_0 : \pi = 0$ is given by

$$W_\pi = \widehat{\pi}' \left(\widehat{\Sigma}_V^{-1} \otimes Z'Z \right) \widehat{\pi},$$

which is identical to the trace of the Cragg-Donald (1993) statistic

$$CD = \widehat{\Sigma}_V^{-1/2} \widehat{\Pi}' Z' Z \widehat{\Pi} \widehat{\Sigma}_V^{-1/2}.$$

However, this Wald test statistic on the reduced form cannot be used in an equivalent way to assess relative bias and 2SLS Wald test size distortions as in the one-variable model above, because these are determined largely by the minimum eigenvalue of CD, $\tau_{\min}$. In other words, relative bias and Wald size distortions can be large if $\mathrm{tr}\,(CD)$ is large but $\tau_{\min}$ is small. In a general setting with g endogenous explanatory variables, $W_\pi = \mathrm{tr}\,(CD)$ is a test for $H_0 : \mathrm{rank}\,(\Pi) = 0$, whereas $\tau_{\min}$ is a test for $H_0 : \mathrm{rank}\,(\Pi) = g - 1$. SY derive critical values for $\tau_{\min}/k_z$ under the local to zero weak instrument asymptotics for maximal total relative bias and Wald test distortions, where the total relative bias is given by

$$B^2 = \frac{\left(E\left[\widehat{\beta}_{2SLS}\right] - \beta \right)' \Sigma_X \left(E\left[\widehat{\beta}_{2SLS}\right] - \beta \right)}{\left(E\left[\widehat{\beta}_{OLS}\right] - \beta \right)' \Sigma_X \left(E\left[\widehat{\beta}_{OLS}\right] - \beta \right)}, \tag{4}$$

with $\Sigma_X = \mathrm{plim}\left(n^{-1} X'X\right)$. In this case, as $\tau_{\min}$ is not the test statistic for $H_0 : \pi = 0$, unlike in the case of one endogenous variable, the correspondence is not exact and use of the SY critical values results in a conservative test in the sense that the null of weak instruments is rejected too infrequently when the null is true. This is not altogether an undesirable feature of the test, as a researcher can be quite confident that instruments are not weak when $\tau_{\min}/k_z$ is larger than the SY critical value.

The Staiger and Stock (1997) and Stock and Yogo (2005) results for the F-test and Cragg-Donald statistic in the one-variable and multiple-variable model respectively in relation to the relative bias and Wald test size distortions hold under the stated assumptions of the model and the reduced form equations for the endogenous variables. When the variances in the reduced forms are conditionally heteroskedastic, then one can compute robust F-statistics and the Kleibergen-Paap (2006) robust version of the Cragg-Donald statistic. These test statistics are then valid tests for underidentification as they have correct size under the null that the instruments are not informative, i.e. for testing that $\mathrm{rank}\,(\Pi) = 0$. But the documented

relationship of the weak-instrument critical values and the sizes of the relative bias and Wald-test size distortion no longer holds, see for example Bun and De Haan (2010). This limits the exact use of the weak-instrument tests, as for example binary endogenous explanatory variables automatically produce a conditionally heteroskedastic reduced form. Also, this relationship brakes down in simple panel data models, when there is serial correlation in the reduced form errors, or indeed in simple time-series models with serial correlation. In our development of the conditional F-statistics for models with multiple endogenous variables, we maintain the same assumptions as Staiger and Stock (1007) and Stock and Yogo (2005), and hence the same limitations. Olea and Pflueger (2013) have recently proposed an alternative robust F-test type procedure for weak instruments, but thus far it can only be applied to the one-endogenous variable model.

3.1 Conditional F-test

Angrist and Pischke (2009) propose an alternative conditional first-stage F-statistic for the case of multiple endogenous variables by reformulating the estimation problem to a one-variable model after replacing the other endogenous variables with their reduced form predictions. For instance, for the two-variable model, the 2SLS estimator for β_1 is obtained by 2SLS in the model

$$y = x_1\beta_1 + \widehat{x}_2\beta_2 + u^*, \tag{5}$$

where $\widehat{x}_2 = Z\widehat{\pi}_2 = P_Z x_2$, using Z as the instruments, and hence

$$\widehat{\beta}_1 = \left(x_1' M_{\widehat{x}_2} P_Z M_{\widehat{x}_2} x_1\right)^{-1} x_1' M_{\widehat{x}_2} P_Z y.$$

Therefore, $\widehat{\beta}_1$ can be seen as the 2SLS estimator in the one-variable model

$$y = M_{\widehat{x}_2} x_1 \beta_1 + \xi, \tag{6}$$

where the residual $M_{\widehat{x}_2} x_1 = x_1 - \widehat{x}_2\widetilde{\delta}$, with $\widetilde{\delta} = \left(\widehat{x}_2'\widehat{x}_2\right)^{-1}\widehat{x}_2' x_1$, is instrumented by Z. The reduced form is then

$$M_{\widehat{x}_2} x_1 = Z\kappa + \varepsilon \tag{7}$$

and the Angrist-Pischke F-statistic is testing the hypothesis $H_0 : \kappa = 0$, given by

$$F_{AP} = \frac{\widehat{\kappa}' Z' Z \widehat{\kappa}}{(k_z - 1)\,\widehat{\sigma}_\varepsilon^2} = \frac{x_1' M_{\widehat{x}_2} P_Z M_{\widehat{x}_2} x_1}{(k_z - 1)\,\widehat{\sigma}_\varepsilon^2}, \tag{8}$$

where $\widehat{\kappa}$ is the OLS estimator of κ,

$$\widehat{\kappa} = \left(Z'Z\right)^{-1} Z' M_{\widehat{x}_2} x_1 = \left(Z'Z\right)^{-1} Z' \left(x_1 - \widehat{x}_2 \widetilde{\delta}\right) = \widehat{\pi}_1 - \widehat{\pi}_2 \widetilde{\delta};$$

and $\widehat{\sigma}_\varepsilon^2 = \widehat{\varepsilon}'\widehat{\varepsilon}/n$, with $\widehat{\varepsilon} = M_{\widehat{x}_2} x_1 - Z\widehat{\kappa}$. The degrees of freedom correction follows because $\widehat{x}_2$ has been predicted using the same instruments Z. If we partition $Z = \left[\begin{array}{cc} z_1 & Z_2 \end{array}\right]$ with Z_2 a $(k_z - 1) \times n$ matrix, then the instrument set for (5) could equivalently be written as $\left[\begin{array}{cc} \widehat{x}_2 & Z_2 \end{array}\right]$.

As the problem seems to have been reduced to a one-endogenous variable model, F_{AP} has been proposed to determine instrument strength for identification of individual structural parameters, like β_1 in the above derivation, and Stock and Yogo (2005) weak instrument critical values used to determine maximum relative bias of the IV estimator, relative to the OLS estimator for the single parameter. There are some issues with this, however, that seem to invalidate such an approach. Under the null that $\kappa = 0$, $(k_z - 1) F_{AP}$ does not follow an asymptotic $\chi^2_{k_z - 1}$ distribution, unless $\pi_1 = 0$. An alternative F-statistic is easily derived that corrects for this, but the relative bias results as described in the previous section for the one-variable model do not carry over to the individual parameters in this multiple endogenous variables model.

To consider the asymptotic distribution, for any given value of δ we have that

$$\begin{aligned} x_1 - \widehat{x}_2 \delta &= x_1 - x_2\delta + \left(x_2 - \widehat{x}_2\right)\delta = Z\left(\pi_1 - \delta\pi_2\right) + v_1 - \delta v_2 + \delta M_Z v_2 \\ &= Z\left(\pi_1 - \delta\pi_2\right) + v_1 - \delta P_Z v_2. \end{aligned}$$

Clearly, the OLS estimator for κ_δ in the model

$$x_1 - \widehat{x}_2\delta = Z\kappa_\delta + \varepsilon^* \tag{9}$$

is given by

$$\begin{aligned} \widehat{\kappa}_\delta &= \left(Z'Z\right)^{-1} Z'\left(x_1 - \widehat{x}_2\delta\right) = \left(Z'Z\right)^{-1} Z'\left(x_1 - x_2\delta\right) \\ &= \widehat{\pi}_1 - \delta\widehat{\pi}_2 = \pi_1 - \delta\pi_2 + \left(Z'Z\right)^{-1} Z'\left(v_1 - \delta v_2\right) \end{aligned}$$

and hence the variance of the OLS estimator is given by

$$Var\left(\widehat{\kappa}_\delta\right) = \left(\sigma_1^2 - 2\delta\sigma_{12} + \delta^2\sigma_2^2\right)\left(Z'Z\right)^{-1}. \tag{10}$$

The F-statistic for testing $H_0 : \kappa_\delta = 0$ in (9) is

$$F_\delta = \frac{\widehat{\kappa}' Z' Z \widehat{\kappa}}{k_z\left(\widehat{\sigma}_1^2 - 2\delta\widehat{\sigma}_{12} + \delta^2\widehat{\sigma}_2^2\right)},$$

and $k_z F_\delta$ converges in distribution to a $\chi^2_{k_z}$ distribution under the null that $\kappa_\delta = 0$, or $\pi_1 = \delta\pi_2$. However, computing the standard F-test statistic in (9) as

$$F_s = \frac{\widehat{\kappa}' Z' Z \widehat{\kappa}}{k_z \widehat{\sigma}^2_{\varepsilon^*}}$$

does not result in F_δ as

$$\widehat{\varepsilon}^{*\prime}\widehat{\varepsilon}^* = (x_1 - \widehat{x}_2\delta)' M_Z (x_1 - \widehat{x}_2\delta) = x_1' M_Z x_1 = \widehat{v}_1'\widehat{v}_1$$

and hence

$$F_s = \frac{\widehat{\kappa}' Z' Z \widehat{\kappa}}{k_z \widehat{\sigma}^2_1}.$$

Therefore the denominator of F_s does not estimate the variance as in (10) correctly and $k_z F_s$ does not converge to a $\chi^2_{k_z}$ distribution under the null, unless $\delta = 0$. The correct F-statistic would be obtained by the standard F-test if the dependent variable in (9) was $x_1 - \delta x_2$ instead of $x_1 - \delta\widehat{x}_2$.

The Angrist-Pischke approach does replace δ by an estimate $\widetilde{\delta}$. By developing a formal testing framework we show that the same issues arise and that $(k_z - 1) F_{AP}$ does not have an asymptotic $\chi^2_{k_z-1}$ distribution under the null that $\pi_1 = \delta\pi_2$, unless $\delta = 0$.

Partition $Z = \left[\begin{array}{cc} z_1 & Z_2 \end{array} \right]$. We can write the reduced from for x_1 as

$$\begin{aligned} x_1 &= Z\pi_1 + v_1 = Z\pi_2 + Z(\pi_1 - \pi_2) + v_1 \\ &= Z\pi_2\delta + Z_2(\pi_{12} - \pi_{22}\delta) + v_1 = x_2\delta + Z_2(\pi_{12} - \pi_{22}\delta) + v_1 - \delta v_2 \end{aligned} \tag{11}$$

where π_1 and π_2 are partitioned as $\left[\begin{array}{cc} \pi_{11} & \pi_{12}' \end{array} \right]'$ and $\left[\begin{array}{cc} \pi_{21} & \pi_{22}' \end{array} \right]'$ respectively; $\delta = \pi_{11}/\pi_{21}$, implicitly assuming that $\pi_{21} \neq 0$. Hence a test for underidentification is a test for $H_0 : \gamma = 0$, in the model

$$x_1 = x_2\delta + Z_2\gamma + v^*, \tag{12}$$

where $v^* = v_1 - \delta v_2$. Clearly, x_2 is an endogenous variable in (12), but we can estimate the parameters δ and γ by IV, using Z as instruments. The 2SLS estimators for δ and γ are given by

$$\widehat{\delta} = \left(\widehat{x}_2' M_{Z_2} \widehat{x}_2\right)^{-1} \widehat{x}_2' M_{Z_2} x_1; \quad \widehat{\gamma} = \left(Z_2' M_{\widehat{x}_2} Z_2\right)^{-1} Z_2' M_{\widehat{x}_2} x_1$$

and

$$Var(\widehat{\gamma}) = \sigma^2_{v^*} \left(Z_2' M_{\widehat{x}_2} Z_2\right)^{-1},$$

with $\sigma^2_{v^*} = \sigma^2_1 - 2\delta\sigma_{12} + \delta^2\sigma^2_2$. The F-test statistic for testing $H_0 : \gamma = 0$ is therefore given by

$$F_\gamma = \frac{x_1' M_{\widehat{x}_2} Z_2 \left(Z_2' M_{\widehat{x}_2} Z_2\right)^{-1} Z_2' M_{\widehat{x}_2} x_1}{(k_z - 1)\left(\widehat{v}^{*\prime}\widehat{v}^*/n\right)}$$

with

$$\widehat{v}^* = x_1 - x_2\widehat{\delta} - Z_2\widehat{\gamma} = Z\widehat{\pi}_1 + \widehat{v}_1 - Z\widehat{\pi}_2\widehat{\delta} - \widehat{\delta}\widehat{v}_2 - Z_2\widehat{\gamma} = \widehat{v}_1 - \widehat{\delta}\widehat{v}_2,$$

as the IV estimates are given by

$$\widehat{\delta} = \frac{\widehat{\pi}_{11}}{\widehat{\pi}_{21}};\ \widehat{\gamma} = \widehat{\pi}_{12} - \widehat{\pi}_{22}\widehat{\delta}.$$

Hence,

$$\widehat{\sigma}^2_{v^*} = \frac{1}{n}\widehat{v}^{*\prime}\widehat{v}^* = \widehat{\sigma}^2_1 - 2\widehat{\delta}\widehat{\sigma}_{12} + \widehat{\delta}^2\widehat{\sigma}^2_2$$

is a consistent estimator of $\sigma^2_{v^*}$.

The Angrist and Pischke (2009) F-statistic as described above is related to F_γ, as

$$F_{AP} = \frac{x_1' M_{\widehat{x}_2} Z \left(Z'Z\right)^{-1} Z' M_{\widehat{x}_2} x_1}{(k_z - 1)\left(\widehat{\varepsilon}'\widehat{\varepsilon}/n\right)} = \frac{x_1' M_{\widehat{x}_2} Z_2 \left(Z_2' M_{\widehat{x}_2} Z_2\right)^{-1} Z_2' M_{\widehat{x}_2} x_1}{(k_z - 1)\,\widehat{\sigma}^2_1},$$

because

$$\begin{aligned} x_1' M_{\widehat{x}_2} P_Z M_{\widehat{x}_2} x_1 &= x_1' P_Z M_{\widehat{x}_2} P_Z x_1 = \widehat{x}_1' M_{\widehat{x}_2} \widehat{x}_1 \\ &= \widehat{\gamma} Z_2' M_{\widehat{x}_2} Z_2 \widehat{\gamma} = x_1' M_{\widehat{x}_2} Z_2 \left(Z_2' M_{\widehat{x}_2} Z_2\right)^{-1} Z_2' M_{\widehat{x}_2} x_1, \end{aligned}$$

and the sum of squared residuals is given by

$$\widehat{\varepsilon}'\widehat{\varepsilon} = x_1' M_{\widehat{x}_2} M_Z M_{\widehat{x}_2} x_1 = x_1' M_Z x_1 = \widehat{v}_1'\widehat{v}_1$$

and hence $\widehat{\varepsilon}'\widehat{\varepsilon}/n = \widehat{\sigma}^2_1$. Therefore, whilst the numerators are the same in F_{AP} and F_γ, the denominators are different. $(k_z - 1)\, F_{AP}$ is therefore not asymptotically $\chi^2_{k_z - 1}$ distributed under the null, $H_0 : \pi_1 = \delta\pi_2$, unless $\delta = 0$ and hence $\pi_1 = 0$.

Clearly, $\widetilde{\delta} = \left(\widehat{x}_2'\widehat{x}_2\right)^{-1}\widehat{x}_2' x_1$ is an estimate of δ under the null that $\pi_1 = \delta\pi_2$ and hence $\gamma = 0$. Let $\widetilde{v}^* = x_1 - x_2\widetilde{\delta}$ be the residual under the null, then the LM test for the null $H_0 : \gamma = 0$ is given by

$$LM = \frac{\widetilde{v}^{*\prime} Z \left(Z'Z\right)^{-1} Z'\widetilde{v}^*}{\widetilde{v}^{*\prime}\widetilde{v}^*/n}$$

which converges to a $\chi^2_{k_z - 1}$ distribution under the null. LM is equal to nR^2 in the model

$$x_1 - x_2\widetilde{\delta} = Z\kappa + \xi. \tag{13}$$

The F-test in (13), with appropriate degrees of freedom correction, is given by

$$\begin{aligned} F_{1|2} &= \frac{\widehat{\kappa}' Z' Z \widehat{\kappa}}{(k_z - 1)\left(\widehat{\xi}'\widehat{\xi}/n\right)} = \frac{\left(\widehat{\pi}_1 - \widetilde{\delta}\widehat{\pi}_2\right)' Z' Z \left(\widehat{\pi}_1 - \widetilde{\delta}\widehat{\pi}_2\right)}{(k_z - 1)\left(\widehat{\sigma}_1^2 + \widetilde{\delta}^2\widehat{\sigma}_2^2 - 2\widetilde{\delta}\widehat{\sigma}_{12}\right)} \\ &= \frac{x_1' M_{\widehat{x}_2} Z_2 \left(Z_2' M_{\widehat{x}_2} Z_2\right)^{-1} Z_2' M_{\widehat{x}_2} x_1}{(k_z - 1)\left(\widehat{\sigma}_1^2 + \widetilde{\delta}^2\widehat{\sigma}_2^2 - 2\widetilde{\delta}\widehat{\sigma}_{12}\right)}, \end{aligned} \tag{14}$$

which is only different from F_γ through the estimate of δ in the denominator. In $F_{1|2}$ this is invariant to which instrument has been excluded from Z in forming Z_2, making it therefore preferable to F_γ. Clearly, $F_{1|2}$ differs from F_{AP} by using the IV residual $x_1 - x_2\widetilde{\delta}$ in (13) instead of the second stage residual $x_1 - \widehat{x}_2\widetilde{\delta}$ for F_{AP} in (7).

Analogous to (11), we can write for x_2

$$x_2 = x_1\delta^* + Z_2\left(\pi_{22} - \pi_{21}\delta^*\right) + v_2 - \delta^* v_1 = x_1\delta^* + Z_2\gamma^* + v^{**}$$

where $\delta^* = \pi_{12}/\pi_{22} = \delta^{-1}$, $\gamma^* = -\gamma/\delta$ and $v^{**} = v^*/\delta$. Clearly

$$F_{2|1} = \frac{\left(\widehat{\pi}_2 - \widetilde{\delta}^*\widehat{\pi}_1\right)' Z' Z \left(\widehat{\pi}_2 - \widetilde{\delta}^*\widehat{\pi}_1\right)}{(k_z - 1)\left(\widehat{\sigma}_2^2 + \widetilde{\delta}^{*2}\widehat{\sigma}_1^2 - 2\widetilde{\delta}^*\widehat{\sigma}_{12}\right)},$$

where $\widetilde{\delta}^* = \left(\widehat{x}_1'\widehat{x}_1\right)^{-1}\widehat{x}_1' x_2$, has the same asymptotic properties as $F_{1|2}$ under $H_0 : \pi_1 = \delta\pi_2$, but it is not identical to $F_{1|2}$ as $\widetilde{\delta}^* \neq \widetilde{\delta}^{-1}$.

3.2 Relationship with Cragg-Donald Statistic

With g endogenous variables, the minimum eigenvalue of the Cragg-Donald statistic, $\tau_{\min}$, is a test for $H_0 : \text{rank}\,(\Pi) = g - 1$ against the alternative $H_1 : \text{rank}\,(\Pi) = g$. For the two-variable model, this null is of course equivalent to $H_0 : \pi_1 = \delta\pi_2$. The Cragg-Donald test is based on the restricted estimates under the null, using the minimum-distance criterion,

$$\left(\overline{\delta}, \overline{\pi}_2\right) = \underset{\delta, \pi_2}{\arg\min} H\left(\delta, \pi_2\right),$$

with

$$H\left(\delta, \pi_2\right) = \left(\begin{pmatrix} \widehat{\pi}_1 \\ \widehat{\pi}_2 \end{pmatrix} - \begin{pmatrix} \delta\pi_2 \\ \pi_2 \end{pmatrix}\right)' \left(\widehat{\Sigma}^{-1} \otimes Z'Z\right) \left(\begin{pmatrix} \widehat{\pi}_1 \\ \widehat{\pi}_2 \end{pmatrix} - \begin{pmatrix} \delta\pi_2 \\ \pi_2 \end{pmatrix}\right).$$

The Cragg-Donald test statistic is then

$$\tau_{\min} = H\left(\overline{\delta}, \overline{\pi}_2\right) \xrightarrow{d} \chi^2_{k_z - 1}$$

under the null. We show in the Appendix that

$$H\left(\overline{\delta},\overline{\pi}_2\right)=\frac{\left(\widehat{\pi}_1-\overline{\delta}\widehat{\pi}_2\right)'Z'Z\left(\widehat{\pi}_1-\overline{\delta}\widehat{\pi}_2\right)}{\widehat{\sigma}_1^2+\overline{\delta}^2\widehat{\sigma}_2^2-2\overline{\delta}\widehat{\sigma}_{12}}$$

and hence the only difference between $F_{1|2}$, $F_{2|1}$ and $\tau_{\min}/\left(k_z-1\right)$ is the estimate for δ. Clearly, unlike the F-statistics, $\tau_{\min}$ is invariant to normalisation, as $H\left(\overline{\delta}^*,\overline{\pi}_1\right)=H\left(\overline{\delta},\overline{\pi}_2\right)$. Because of this, computation of both $F_{1|2}$ and $F_{2|1}$ can provide further information about the nature of the weakness of the instruments, as their values can indicate whether the rank reduction is due to e.g. $\pi_1=0$ ($\delta=0$), which $\tau_{\min}$ cannot distinguish, as we will show below. We will also present a three-endogenous variables example in Section 4.3 which further highlights the additional information about instrument strength revealed by the three conditional F-statistics relative to that of the Cragg-Donald statistic.

4 Local to Rank One Weak Instrument Asymptotics in the Two-Variable Model

In the previous section, we have shown that $\left(k_z-1\right)F_\gamma$ has a limiting $\chi^2_{k_z-1}$ distribution under the null that $\gamma=0$ in (12). We next investigate whether F_γ can be used to assess whether instruments are weak for individual parameters as described in Section 2. We focus in the derivation below on F_γ as the setup for this test is easier to use with our weak instruments asymptotics, but results of course carry over directly to $F_{1|2}$, $F_{2|1}$ and $\tau_{\min}$.

We are interested in the case that the instruments are not weak for each equation, but where the rank of Π approaches a rank reduction of one. We specify LRR1 weak instrument asymptotics as $\gamma=c/\sqrt{n}$, or

$$\pi_{12}=\delta\pi_{22}+c/\sqrt{n}.$$

We can then write the reduced form of x_1 as

$$x_1=Z\pi_2\delta+Z_2\left(\pi_{12}-\delta\pi_{22}\right)+v_1=\widehat{x}_2\delta+Z_2c/\sqrt{n}+\left(v_1-\delta P_Z v_2\right).$$

The IV estimator for β_1 is given by

$$\widehat{\beta}_{1,2SLS}=\frac{x_1'M_{\widehat{x}_2}Z_2\left(Z_2'M_{\widehat{x}_2}Z_2\right)^{-1}Z_2'M_{\widehat{x}_2}y}{x_1'M_{\widehat{x}_2}Z_2\left(Z_2'M_{\widehat{x}_2}Z_2\right)^{-1}Z_2'M_{\widehat{x}_2}x_1}$$

and it follows that

$$\widehat{\beta}_{1,2SLS}-\beta_1=\frac{x_1'M_{\widehat{x}_2}Z_2\left(Z_2'M_{\widehat{x}_2}Z_2\right)^{-1}Z_2'M_{\widehat{x}_2}u}{x_1'M_{\widehat{x}_2}Z_2\left(Z_2'M_{\widehat{x}_2}Z_2\right)^{-1}Z_2'M_{\widehat{x}_2}x_1}$$

as $M_{\widehat{x}_2}\widehat{x}_2 = 0$, $M_{\widehat{x}_2}M_Z = \left(I - P_Z x_2 \left(x_2' P_Z x_2\right)^{-1} x_2' P_Z\right) M_Z = M_Z$, and hence $Z_2' M_{\widehat{x}_2} M_Z v_2 = Z_2' M_Z v_2 = 0$.

We assume that

$$\begin{pmatrix} \frac{1}{\sqrt{n}} Z_2' M_{\widehat{x}_2} u \\ \frac{1}{\sqrt{n}} Z_2' M_{\widehat{x}_2} (v_1 - \delta v_2) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_{Z_2^* u} \\ \psi_{Z_2^*(v_1 - \delta v_2)} \end{pmatrix} = N(0, \Omega \otimes Q),$$

where

$$\Omega = \begin{pmatrix} \sigma_u^2 & \sigma_{u1} - \delta\sigma_{u2} \\ \sigma_{u1} - \delta\sigma_{u2} & \sigma_1^2 + \delta^2\sigma_2^2 - 2\delta\sigma_{12} \end{pmatrix}; \quad Z_2^* = M_{\widehat{x}_2} Z_2; \quad Q = \operatorname{plim}\left(n^{-1} Z_2^{*\prime} Z_2^*\right).$$

It is then easily shown that

$$x_1' M_{\widehat{x}_2} Z_2 \left(Z_2' M_{\widehat{x}_2} Z_2\right)^{-1} Z_2' M_{\widehat{x}_2} x_1 \xrightarrow{d} \sigma^2_{v_1 - \delta v_2} \left(\widetilde{\lambda} + \widetilde{z}_v\right)' \left(\widetilde{\lambda} + \widetilde{z}_v\right)$$

and

$$x_1' M_{\widehat{x}_2} Z_2 \left(Z_2' M_{\widehat{x}_2} Z_2\right)^{-1} Z_2' M_{\widehat{x}_2} u \xrightarrow{d} \sigma_u \sigma_{v_1 - \delta v_2} \left(\widetilde{\lambda} + \widetilde{z}_v\right)' \widetilde{z}_u$$

where

$$\sigma_{v_1 - \delta v_2} = \sqrt{\sigma_1^2 + \delta^2\sigma_2^2 - 2\delta\sigma_{12}};$$

$$\widetilde{\lambda} = \sigma^{-1}_{v_1 - \delta v_2} Q^{1/2} c; \quad \widetilde{z}_v = \sigma^{-1}_{v_1 - \delta v_2} Q^{-1/2} \psi_{Z_2^*(v_1 - \delta v_2)}; \quad \widetilde{z}_u = \sigma_u^{-1} Q^{-1/2} \psi_{Z_2^* u}.$$

We are therefore in the same setup as Staiger and Stock (1997) and Stock and Yogo (2005), and the distribution of the bias of $\widehat{\beta}_{1,2SLS}$ is given by

$$\widehat{\beta}_{1,2SLS} - \beta_1 = \frac{x_1' M_{\widehat{x}_2} Z_2 \left(Z_2' M_{\widehat{x}_2} Z_2\right)^{-1} Z_2' M_{\widehat{x}_2} u}{x_1' M_{\widehat{x}_2} Z_2 \left(Z_2' M_{\widehat{x}_2} Z_2\right)^{-1} Z_2' M_{\widehat{x}_2} x_1} \xrightarrow{d} \frac{\sigma_u}{\sigma_{v_1 - \delta v_2}} \frac{\left(\widetilde{\lambda} + \widetilde{z}_v\right)' \widetilde{z}_u}{\left(\widetilde{\lambda} + \widetilde{z}_v\right)' \left(\widetilde{\lambda} + \widetilde{z}_v\right)},$$

and

$$E\left(\widehat{\beta}_{1,2SLS}\right) - \beta_1 \longrightarrow \frac{\sigma_{u1} - \delta\sigma_{u2}}{\sigma_1^2 + \delta^2\sigma_2^2 - 2\delta\sigma_{12}} E\left(\frac{\left(\widetilde{\lambda} + \widetilde{z}_v\right)' \widetilde{z}_v}{\left(\widetilde{\lambda} + \widetilde{z}_v\right)' \left(\widetilde{\lambda} + \widetilde{z}_v\right)}\right).$$

One would therefore think that one could proceed as in the one-variable model as specified above, with

$$\widetilde{l} = \widetilde{\lambda}'\widetilde{\lambda} / (k_z - 1) = \frac{1}{k_z - 1} \frac{c' Q c}{\sigma_1^2 + \delta^2\sigma_2^2 - 2\delta\sigma_{12}}$$

and the critical values from the non-central chi-squared distribution applied to

$$F_\gamma = \frac{x_1' M_{\widehat{x}_2} Z_2 \left(Z_2' M_{\widehat{x}_2} Z_2\right)^{-1} Z_2' M_{\widehat{x}_2} x_1}{(k_z - 1)\left(\widehat{\sigma}_1^2 + \widehat{\delta}^2 \widehat{\sigma}_2^2 - 2\widehat{\delta}\widehat{\sigma}_{12}\right)}.$$

However, in this case the bias of the OLS estimator of β_1 in the model

$$y = x_1\beta_1 + x_2\beta_2 + u$$

is given by

$$\widehat{\beta}_{1,OLS} - \beta_1 = \frac{x_1' M_{x_2} u}{x_1' M_{x_2} x_1}.$$

As

$$x_1 = x_2\delta + Z_2 c/\sqrt{n} + (v_1 - \delta v_2),$$

we get that

$$\operatorname{plim}\left(n^{-1}\left(x_1' M_{x_2} u\right)\right) = \operatorname{plim}\left(n^{-1}\left(\frac{c}{\sqrt{n}} Z_2' M_{x_2} u + (v_1 - \delta v_2)' M_{x_2} u\right)\right).$$

Further,

$$\begin{aligned} & \operatorname{plim}\left(n^{-1}\left(x_1' M_{x_2} x_1\right)\right) \\ = \; & \operatorname{plim}\left(n^{-1}\left(\frac{c}{\sqrt{n}} Z_2' M_{x_2} Z_2 \frac{c}{\sqrt{n}} + 2\frac{c'}{\sqrt{n}} Z_2' M_{x_2}(v_1 - \delta v_2) + (v_1 - \delta v_2)' M_{x_2}(v_1 - \delta v_2)\right)\right). \end{aligned}$$

From these results we find that the bias of the OLS estimator converges to

$$\operatorname{plim}\left(\widehat{\beta}_{1,OLS} - \beta_1\right) = \frac{\operatorname{plim}\left(n^{-1}(v_1 - \delta v_2)' M_{x_2} u\right)}{\operatorname{plim}\left(n^{-1}(v_1 - \delta v_2)' M_{x_2}(v_1 - \delta v_2)\right)} = \frac{\sigma_{u1} - \delta\sigma_{u2} - \frac{(\sigma_{12} - \delta\sigma_2^2)\sigma_{u2}}{\pi_2' Q_{ZZ}\pi_2 + \sigma_2^2}}{\sigma_1^2 + \delta^2\sigma_2^2 - 2\delta\sigma_{12} - \frac{(\sigma_{12} - \delta\sigma_2^2)^2}{\pi_2' Q_{ZZ}\pi_2 + \sigma_2^2}}$$

and therefore, we now have that

$$B_{n,1} = \frac{\left|E\left[\widehat{\beta}_{1,2SLS}\right] - \beta_1\right|}{\left|E\left[\widehat{\beta}_{1,OLS}\right] - \beta_1\right|} \neq \left|E\left[\frac{\left(\widetilde{\lambda} + \widetilde{z}_v\right)' \widetilde{z}_v}{\left(\widetilde{\lambda} + \widetilde{z}_v\right)' \left(\widetilde{\lambda} + \widetilde{z}_v\right)}\right]\right|$$

and so the direct relationship between the relative bias of the individual parameter and the value of the concentration parameter does not hold in this setting.[1]

[1] The one-variable model as described above was $y = M_{\widehat{x}_2} x_1 \beta_1 + \xi$.and so one could ask the question whether the weak instrument relative bias could apply to the OLS estimator in this model instead. The OLS estimator is given by

$$\widetilde{\beta}_{1,OLS} = \frac{x_1' M_{\widehat{x}_2} y}{x_1' M_{\widehat{x}_2} x_1} = \beta_1 + \frac{\beta_2 x_1' M_{\widehat{x}_2} x_2 + x_1' M_{\widehat{x}_2} u}{x_1' M_{\widehat{x}_2} x_1}$$

and therefore

$$\operatorname{plim}\widetilde{\beta}_{1,OLS} - \beta_1 = \frac{\beta_2\sigma_{12} + \sigma_{u1}}{\sigma_1^2}$$

and hence, again

$$\widetilde{B}_{n,1} = \frac{\left|E\left[\widehat{\beta}_{1,2SLS}\right] - \beta_1\right|}{\left|E\left[\widetilde{\beta}_{1,OLS}\right] - \beta_1\right|} \neq \left|E\left[\frac{\left(\widetilde{\lambda} + \widetilde{z}_v\right)' \widetilde{z}_v}{\left(\widetilde{\lambda} + \widetilde{z}_v\right)' \left(\widetilde{\lambda} + \widetilde{z}_v\right)}\right]\right|.$$

However, we can get a result for the total relative bias. First of all, we show in the Appendix that for the 2SLS estimator for β_2 the following holds,

$$\widehat{\beta}_{2,2SLS} - \beta_2 \xrightarrow{d} -\delta \frac{\sigma_u}{\sigma_{v_1 - \delta v_2}} \frac{\left(\widetilde{\lambda} + \widetilde{z}_v\right)' \widetilde{z}_u}{\left(\widetilde{\lambda} + \widetilde{z}_v\right)' \left(\widetilde{\lambda} + \widetilde{z}_v\right)},$$

and hence, asymptotically,

$$E\left[\widehat{\beta}_{2,2SLS}\right] - \beta_2 = -\delta \left(E\left[\widehat{\beta}_{1,2SLS}\right] - \beta_1\right).$$

From this it follows that $\widehat{\beta}_{2,2SLS}$ is consistent when $\delta = 0$, that is in the situation where the instruments are strong for x_2, but weak for x_1 in the sense that π_1 is local to zero.

We then show in the Appendix that

$$B^2 = \frac{\left(E\left[\widehat{\beta}_{2SLS}\right] - \beta\right)' \Sigma_X \left(E\left[\widehat{\beta}_{2SLS}\right] - \beta\right)}{\left(E\left[\widehat{\beta}_{OLS}\right] - \beta\right)' \Sigma_X \left(E\left[\widehat{\beta}_{OLS}\right] - \beta\right)} \leq b^2,$$

where

$$b = E\left(\frac{\left(\widetilde{\lambda} + \widetilde{z}_v\right)' \widetilde{z}_v}{\left(\widetilde{\lambda} + \widetilde{z}_v\right)' \left(\widetilde{\lambda} + \widetilde{z}_v\right)}\right).$$

From this it follows that we can use the SY critical values for $\tau_{\min}/(k_z - 1)$, $F_{1|2}$ and $F_{2|1}$ to assess LRR1 weak instrument maximal total relative bias. These are the critical values tabulated for the one-endogenous variable case with $k_z - 1$ instruments.

We can also use the equivalent SY critical values for assessing the maximal size of the individual 2SLS Wald tests. We get for the Wald test for the simple null $H_0 : \beta_1 = \beta_1^0$

$$W_1 = \frac{\left(\widehat{\beta}_{1,2SLS} - \beta_1^0\right)^2 \left(x_1' M_{\widehat{x}_2} Z_2 \left(Z_2' M_{\widehat{x}_2} Z_2\right)^{-1} Z_2' M_{\widehat{x}_2} x_1\right)}{\widehat{\sigma}_u^2} = \frac{\sigma_u^2}{\widehat{\sigma}_u^2} \frac{\left(\left(\widetilde{\lambda} + \widetilde{z}_v\right)' \widetilde{z}_u\right)^2}{\left(\widetilde{\lambda} + \widetilde{z}_v\right)' \left(\widetilde{\lambda} + \widetilde{z}_v\right)}$$

where

$$\widehat{\sigma}_u^2 = \left(y - x_1 \widehat{\beta}_{1,2SLS} - x_2 \widehat{\beta}_{2,2SLS}\right)' \left(y - x_1 \widehat{\beta}_{1,2SLS} - x_2 \widehat{\beta}_{2,2SLS}\right)/n.$$

We find that

$$\widehat{\sigma}_u^2 \xrightarrow{d} \sigma_u^2 \left(1 - 2 \frac{\sigma_{u1} - \delta \sigma_{u2}}{\sigma_u \sigma_{v_1 - \delta v_2}} \frac{\widetilde{\nu}_2}{\widetilde{\nu}_1} + \left(\frac{\widetilde{\nu}_2}{\widetilde{\nu}_1}\right)^2\right),$$

where

$$\widetilde{\nu}_1 = \left(\widetilde{\lambda} + \widetilde{z}_v\right)' \left(\widetilde{\lambda} + \widetilde{z}_v\right); \quad \widetilde{\nu}_2 = \left(\widetilde{\lambda} + \widetilde{z}_v\right)' \widetilde{z}_u.$$

The Wald test is then, as in Staiger and Stock (1997) and Stock and Yogo (2005), equal to

$$W_1 = \frac{\widetilde{v}_2^2/\widetilde{v}_1}{1 - 2\widetilde{\rho}\widetilde{\nu}_2/\widetilde{\nu}_1 + (\widetilde{\nu}_2/\widetilde{\nu}_1)^2}$$

where $\widetilde{\rho} = \frac{\sigma_{u1} - \delta\sigma_{u2}}{\sigma_u \sigma_{v_1 - \delta v_2}}$, and so we can again use the SY critical values for the F-statistic for maximal size of the Wald-test, achieved when $\widetilde{\rho} = 1$. Clearly, we get the same results for W_2, the Wald test for $H_0 : \beta_2 = \beta_2^0$.

4.1 Monte Carlo Illustration

To illustrate, we generate data from the model as specified above, with

$$\begin{pmatrix} u_i \\ v_{1i} \\ v_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{u1} & \sigma_{u2} \\ \sigma_{u1} & \sigma_1^2 & \sigma_{12} \\ \sigma_{u2} & \sigma_{12} & \sigma_2^2 \end{pmatrix} \right).$$

The instruments are drawn independently from the standard normal distribution, with $k_z = 4$, and hence $Q_{ZZ} = I_4$. We set $\pi_2 = (-0.5, 0.5, -0.5, 0.5)'$ and $\pi_1 = \delta\pi_2 + (0, c, c, c)'/\sqrt{n}$. We have

$$Q = \text{plim}\, \frac{1}{n} Z_2' M_{\widehat{x}_2} Z_2 = I_{k_z - 1} - \frac{\pi_{22}\pi_{22}'}{\pi_2'\pi_2},$$

where $\pi_2 = \begin{bmatrix} \pi_{21} & \pi_{22} \end{bmatrix}'$ is partitioned commensurate with $Z = \begin{bmatrix} z_1 & Z_2 \end{bmatrix}$.

The limit of the concentration parameter for this specific configuration is given by

$$CPl = \frac{c'Qc}{\sigma_1^2 + \delta^2\sigma_2^2 - 2\delta\sigma_{12}} = \frac{3c^2 - \frac{c^2}{4}}{\sigma_1^2 + \delta^2\sigma_2^2 - 2\delta\sigma_{12}}.$$

We choose c such that the concentration parameter has the value for which the IV estimator for β has a maximal total relative bias of 10%. We have further set the parameters as follows: $\beta_1 = 0.5;\ \beta_2 = -0.3;\ \sigma_u^2 = \sigma_1^2 = \sigma_2^2 = 1;\ \sigma_{u1} = 0.1;\ \sigma_{u2} = -0.7;\ \sigma_{12} = -0.7$ and $\delta = 0.7$. This design is such that the additional terms in the OLS bias are important, with

$$\frac{\sigma_{u1} - \delta\sigma_{u2}}{\sigma_1^2 + \delta^2\sigma_2^2 - 2\delta\sigma_{12}} \frac{\sigma_1^2 + \delta^2\sigma_2^2 - 2\delta\sigma_{12} - \frac{(\sigma_{12} - \delta\sigma_2^2)^2}{\pi_2' Q_{ZZ}\pi_2 + \sigma_2^2}}{\sigma_{u1} - \delta\sigma_{u2} - \frac{(\sigma_{12} - \delta\sigma_2^2)\sigma_{u2}}{\pi_2' Q_{ZZ}\pi_2 + \sigma_2^2}} = 3.5591.$$

i.e. the OLS bias for β_1 is much smaller than $\frac{\sigma_{u1} - \delta\sigma_{u2}}{\sigma_1^2 + \delta^2\sigma_2^2 - 2\delta\sigma_{12}}$. The results are given in Table 3 for a sample size of $10,000$ observations. The individual standard F-statistics are very large. As expected, the IV estimator of β_1 has a large relative bias of 0.3441, approximately equal to $3.56 * 0.1$, but the relative bias of β_2 is much smaller at 0.0498. The distributions of $F_{1|2}$, $F_{2|1}$ and $\tau_{\min}/(k_z - 1)$ are virtually identical, each with a mean of 4.7 and rejection frequency

of 4.6% at the 5% nominal level using the weak instrument critical value. In comparison, the AP F-statistics are much larger in this case with the mean of $F_{AP,1}$ equal to 11.82, and that of $F_{AP,2}$ equal to 22.93.

Table 3. Estimation results and relative bias for two-variable model

	mean	st dev	rel bias	SY rej freq
$\widehat{\beta}_{1,OLS}$	0.5695	0.0070		
$\widehat{\beta}_{2,OLS}$	-0.6506	0.0062		
$\widehat{\beta}_{1,2SLS}$	0.5239	0.1979	0.3441	
$\widehat{\beta}_{2,2SLS}$	-0.3174	0.1419	0.0498	
F_1	1290	44		
F_2	2503	71		
$F_{AP,1}$	11.82	5.91		0.6256
$F_{AP,2}$	22.93	11.46		0.9082
$F_{1\|2}$	4.70	2.35		0.0460
$F_{2\|1}$	4.71	2.36		0.0464
$\tau_{\min}/(k_z-1)$	4.70	2.35		0.0457
$\tau_{\min}/k_z$	3.52	1.76		0.0267

Notes: sample size 10,000; 10,000 MC replications; $\beta_1 = 0.5$; $\beta_2 = -0.3$; F_j is the first-stage F-statistic for $x_j, j = 1, 2$; $F_{AP,j}$ is the Angrist-Pischke F-statistic and $F_{1|2}$ and $F_{2|1}$ are the conditional F-statistics as in (14); $\tau_{\min}$ is the Cragg-Donald minimum eigenvalue statistic; rel bias is the relative bias of the 2SLS estimator, relative to that of the OLS estimator; SY rej freq uses the 5% Stock-Yogo critical values for a maximum 10% total relative bias

The total relative bias in this design is found to be equal to 7.6%, which is less than 10%, as predicted by the theory above. The SY test for weak instruments for Π local to 0 is conservative and has a rejection frequency of 2.6%. This test is given by $\tau_{\min}/k_z$ and the weak instrument critical value is derived for two endogenous variables with k_z instruments. In contrast, the weak instrument critical values for $F_{1|2}$, $F_{2|1}$ and $\tau_{\min}/(k_z-1)$ are those for one endogenous variable with $k_z - 1$ instruments. From Table 1 in SY, it is easily established that when $\tau_{\min}/k_z$ is larger than its associated tabulated critical value, then $\tau_{\min}/(k_z-1)$ is also larger than its weak instrument critical value, so we would always reject LRR1 weak instrument problems whenever we reject rank zero weak instrument problems.

In Table 4 we present results for the Wald test statistics in a design with $\widetilde{\rho} = 1$, by changing the variance parameters to $\sigma_{u1} = 0.755$, $\sigma_{u2} = 0.35$ and $\sigma_{12} = -0.35$, again choosing c such that the size of the Wald tests is 10% at the 5% level. The simulations confirm the analytical results.

The rejection frequencies of the Wald tests are just over 10% and the rejection frequencies of $F_{1|2}$, $F_{2|1}$ and $\tau_{\min}/(k_z-1)$ just over 5%. In this case, the SY weak instrument test $\tau_{\min}/k_z$ using the tabulated critical value for two endogenous variables and four instruments is also just over 5%.

Table 4. Estimation and Wald tests results for two-variable model

	mean	st dev	rej freq	SY rej freq
$\widehat{\beta}_{1,OLS}$	1.4990	0.0007		
$\widehat{\beta}_{2,OLS}$	0.3899	0.0006		
$\widehat{\beta}_{1,2SLS}$	0.5257	0.1565		
$\widehat{\beta}_{2,2SLS}$	-0.2827	0.1071		
W_1	1.47	2.86	0.1016	
W_2	1.46	2.87	0.1017	
W_{12}	2.61	3.58	0.1080	
$F_{1\|2}$	14.85	4.40		0.0548
$F_{2\|1}$	14.93	4.45		0.0585
$\tau_{\min}/(k_z-1)$	14.84	4.40		0.0517
$\tau_{\min}/k_z$	11.13	3.30		0.0545

Notes: sample size 10,000; 10,000 MC replications; $\beta_1 = 0.5$; $\beta_2 = -0.3$
W_j is the Wald test for H_0: $\beta_j = \beta_{0j}$; W_{12} is joint Wald test;
$F_{1|2}$ and $F_{2|1}$ are the conditional F-statistics as in (14); $\tau_{\min}$ is the Cragg-Donald minimum eigenvalue statistic; rej freq for Wald tests uses 5% critical value of χ^2 distribution; SY rej freq uses the 5% Stock-Yogo critical values for a maximal 10% size of Wald tests.

4.2 The case $\delta = 0$

When $\delta = 0$, we have in the process above that π_1 is local to zero, and hence the instruments for x_1 are weak, but not for x_2. As shown above, $\widehat{\beta}_{2,2SLS}$ is in this case consistent for β_2, but $\widehat{\beta}_{1,2SLS}$ will suffer from a weak instrument bias. This situation may actually be of interest if the main research focus is on the effect of x_2 on y. If the instruments used are then strong for x_2 but weakly or not informative for x_1, the IV estimator for β_2 will be well behaved. In Table 5, we show the results for the bias of the 2SLS estimates, for when $\delta = 0$ and where we have further set $\sigma_{u1} = 0.8$. All other parameters remain the same as for the results presented in Table 3, and we have set the value of c again such that the maximum total relative bias is 10%. As can be seen from the table, the results are as expected. The value of the first-stage F-statistic for x_1, F_1 is now small, whilst that of F_2 is large. The behaviour of $F_{AP,1}$ is now the same as that of $F_{1|2}$, both rejecting the null of weak instruments 5% of the time using the

SY critical values for $k_z - 1$ instruments. $\widehat{\beta}_{2,2SLS}$ is consistent, but the total relative bias is at 9.7% only just below 10%.

Table 5. Estimation results and relative bias for two-variable model, $\delta = 0$

	mean	st dev	rel bias	SY rej freq
$\widehat{\beta}_{1,OLS}$	1.2317	0.0067		
$\widehat{\beta}_{2,OLS}$	-0.3976	0.0047		
$\widehat{\beta}_{1,2SLS}$	0.5776	0.3001	0.0776	
$\widehat{\beta}_{2,2SLS}$	-0.3010	0.0103	-0.0010	
F_1	4.08	1.88		0.0044
F_2	2503	70		1.0000
$F_{AP,1}$	4.79	2.39		0.0515
$F_{AP,2}$	2922	502		1.0000
$F_{1\|2}$	4.72	2.36		0.0474
$F_{2\|1}$	462	1184		0.8811
$\tau_{\min}/(k_z - 1)$	4.72	2.36		0.0470
$\tau_{\min}/k_z$	3.54	1.77		0.0259

Notes: sample size 10,000; 10,000 MC replications; $\beta_1 = 0.5$; $\beta_2 = -0.3$
F_j is the first-stage reduced form F-statistic for $x_j, j = 1, 2$;
$F_{AP,j}$ is the Angrist-Pischke F-statistic and $F_{1|2}$ and $F_{2|1}$ are the conditional F-statistics as in (14); $\tau_{\min}$ is the Cragg-Donald minimum eigenvalue statistic; rel bias is the relative bias of the 2SLS estimator, relative to that of the OLS estimator; SY rej freq uses the 5% Stock-Yogo critical values for a maximum 10% total relative bias

4.3 More than Two Endogenous Variables

As is clear from the analyses above for the two-variable model, the use of $F_{1|2}$ and $F_{2|1}$ under our LRR1 weak instrument asymptotics do not reveal more information than the Cragg-Donald statistic $\tau_{\min}/(k_z - 1)$, unless $\delta = 0$ and hence π_1 is local to zero. The derivations for the two-variable model easily extend to the general case of several endogenous variables. The computation of the individual conditional F-statistics could then reveal further interesting patterns that the Cragg-Donald statistic will not be able to. For example, consider a three-variable model, which has a local rank reduction of one, of the form

$$\pi_1 = \delta_2 \pi_2 + \delta_3 \pi_3 + c/\sqrt{n}$$

but with $\delta_3 = 0$. The conditional F-statistics are in this case computed from

$$x_j - X_{-j}\widetilde{\delta} = Z\kappa + \xi,$$

where X_{-j} is the matrix of endogenous variables with x_j excluded and $\widetilde{\delta} = \left(\widehat{X}'_{-j}\widehat{X}_{-j}\right)^{-1}\widehat{X}'_{-j}x_j$. The conditional F-statistics are then

$$F_{x_j|X_{-j}} = \frac{\widehat{\kappa}'Z'Z\widehat{\kappa}}{(k_z-2)\left(\widehat{\xi}'\widehat{\xi}/n\right)}, \tag{15}$$

see the Appendix for simple Stata code to calculate $F_{x_j|X_{-j}}$.

Table 6 presents some simulation results for this particular case for the following design

$$\begin{pmatrix} u_i \\ v_{1i} \\ v_{2i} \\ v_{3i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.8 & 0.3 & 0.6 \\ 0.8 & 1 & 0.3 & 0.5 \\ 0.3 & 0.3 & 1 & 0.4 \\ 0.6 & 0.5 & 0.4 & 1 \end{pmatrix}\right),$$

$\delta_2 = 0.5$; $\delta_3 = 0$; $\beta_1 = 0.5$; $\beta_2 = -0.3$; $\beta_3 = 0.7$. The instruments are again drawn independently form the standard normal distribution, with $k_z = 5$, and c is again chosen such that the total relative bias is less than 10%.

Table 6. Estimation results and relative bias for three-variable model

	mean	st dev	rel bias	SY rej freq
$\widehat{\beta}_{1,OLS}$	1.1337	0.0068		
$\widehat{\beta}_{2,OLS}$	-0.4581	0.0050		
$\widehat{\beta}_{3,OLS}$	0.9526	0.0055		
$\widehat{\beta}_{1,2SLS}$	0.5709	0.3086	0.1120	
$\widehat{\beta}_{2,2SLS}$	-0.3361	0.1575	0.2285	
$\widehat{\beta}_{3,2SLS}$	0.6990	0.0161	-0.0040	
F_1	650	26		
F_2	2504	67		
F_3	902	32		
$F_{1\|2,3}$	4.82	2.38		0.0514
$F_{2\|1,3}$	4.84	2.41		0.0531
$F_{3\|1,2}$	198.21	329.06		0.8779
$\tau_{\min}/(k_z-2)$	4.82	2.38		0.0513
$\tau_{\min}/k_z$	2.89	1.43		0.0156

Notes: sample size 10,000; 10,000 MC replications; β_1= 0.5; β_2= −0.3; β_3= 0.7
F_j is the first-stage reduced form F-statistic for x_j, j =1, 2, 3;
$F_{1|2,3}$, $F_{2|1,3}$ and $F_{3|1,2}$ are the conditional F-statistics as in (15);
$\tau_{\min}$ is the Cragg-Donald minimum eigenvalue statistic; rel bias is the relative bias of the 2SLS estimator, relative to that of the OLS estimator; SY rej freq uses the 5% Stock-Yogo critical values for a maximum 10% total relative bias

It is clear from the conditional F-statistics that the near rank reduction is due to parameters in the reduced form equations for x_1 and x_2. From a straightforward extension of the analytical

results for the two-variable case in the Appendix we get that $\widehat{\beta}_{3,2SLS}$ is consistent as $\delta_3 = 0$. This is confirmed by the simulation results. The total relative bias in this case is equal to 8.8%, which is less than 10%. It is clear that the conditional F-statistics now provide important additional information to that provided by the Cragg-Donald statistic.

5 Conclusions

We have shown that a conditional first-stage F-test statistic can be informative about the information that instruments provide for models with multiple endogenous variables. The conditional F-test is similar to the one proposed by Angrist and Pischke (2009), but takes the variance of the multiple equations into account for testing a rank reduction of one of the matrix of reduced from parameters. Our weak instrument asymptotics is defined as local to a rank reduction of one of this matrix. We find that the conditional F-statistics in a two-endogenous variables model provide the same information as the Cragg-Donald test statistic for testing a rank reduction of one, unless the rank reduction is due to the fact that the instruments are uninformative for one of the endogenous variables. The conditional F-statistics are informative for total relative bias and Wald test size distortions for individual structural parameters. With more than two endogenous variables, the conditional F-statistics can provide additional information regarding the strength of the instruments for the different reduced forms. We therefore recommend in applied work that researchers report standard first-stage F-statistics, the Cragg-Donald statistic and the conditional F-statistics in order to gauge the nature of the weak instrument problem, if any. The Stock and Yogo (2005) weak instrument critical values can be used for the Cragg-Donald and conditional F-statistics. When reduced form errors are conditionally heteroskedastic and/or serially correlated, robust conditional F-statistics can be computed and used as tests for underidentification. However, the exact link of the Stock and Yogo (2005) critical values with the magnitude of the relative bias and Wald-test size distortions no longer holds for the robust statistics and is therefore an important avenue for future research.

References

[1] Angrist J. and J.-S. Pischke, 2009. Mostly Harmless Econometrics: An Empiricist's Companion, Princeton University Press, Princeton.

[2] Bun, M. and M. de Haan, 2010, Weak Instruments and the First Stage F-Statistic in IV Models with a Nonscalar Error Covariance Structure, Discussion Paper 2010/02, UvA Econometrics, University of Amsterdam.

[3] Cragg, J.G. and S.G. Donald, 1993. Testing Identifiability and Specification in Instrumental Variables Models, *Econometric Theory* 9, 222-240.

[4] Kleibergen, F. and R. Paap, 2006. Generalized reduced rank tests using the singular value decomposition, *Journal of Econometrics* 133, 97-126.

[5] Olea, J.L.M. and C. Pflueger, 2013, A Robust Test for Weak Instruments, *Journal of Business and Economic Statistics* 31, 358-368.

[6] Rothenberg, T.J., 1984, Approximating the Distributions of Econometric Estimators and Test Statistics. In Z. Griliches and M.D. Intriligator (Eds.), *Handbook of Econometrics, Volume 2*, 881-935. Amsterdam: North Holland.

[7] Staiger, D. and J.H. Stock, 1997. Instrumental Variables Regression with Weak Instruments, *Econometrica* 65, 557-586.

[8] Stock, J.H. and M. Yogo, 2005. Testing for weak instruments in linear IV regression. In D.W.K. Andrews and J.H. Stock (Eds.), *Identification and Inference for Econometric Models, Essays in Honor of Thomas Rothenberg*, 80-108. New York: Cambridge University Press.

6 Appendix

6.1 Cragg-Donald Statistic

The Cragg-Donald statistic in the two-variable model is obtained as

$$\tau_{\min} = \min H(\delta, \pi_2) = \left(\begin{pmatrix} \widehat{\pi}_1 \\ \widehat{\pi}_2 \end{pmatrix} - \begin{pmatrix} \delta\pi_2 \\ \pi_2 \end{pmatrix} \right)' \left(\widehat{\Sigma}^{-1} \otimes Z'Z \right) \left(\begin{pmatrix} \widehat{\pi}_1 \\ \widehat{\pi}_2 \end{pmatrix} - \begin{pmatrix} \delta\pi_2 \\ \pi_2 \end{pmatrix} \right).$$

The first-order condition is given by, writing $\widehat{\Sigma}^{-1} = \begin{pmatrix} \widehat{\sigma}^{11} & \widehat{\sigma}^{12} \\ \widehat{\sigma}^{12} & \widehat{\sigma}^{22} \end{pmatrix}$,

$$-\frac{1}{2}\frac{\partial H(\delta, \pi_2)}{\partial \pi_2} = \left(\begin{pmatrix} \delta\widehat{\sigma}^{11} + \widehat{\sigma}^{12} & \delta\widehat{\sigma}^{12} + \widehat{\sigma}^{22} \end{pmatrix} \otimes Z'Z \right) \left(\begin{pmatrix} \widehat{\pi}_1 \\ \widehat{\pi}_2 \end{pmatrix} - \begin{pmatrix} \delta\pi_2 \\ \pi_2 \end{pmatrix} \right) = 0,$$

resulting in

$$\left(\delta\widehat{\sigma}^{11}+\widehat{\sigma}^{12}\right)Z'Z\widehat{\pi}_1+\left(\delta\widehat{\sigma}^{12}+\widehat{\sigma}^{22}\right)Z'Z\widehat{\pi}_2=\delta\left(\delta\widehat{\sigma}^{11}+\widehat{\sigma}^{12}\right)Z'Z\pi_2+\left(\delta\widehat{\sigma}^{12}+\widehat{\sigma}^{22}\right)Z'Z\pi_2.$$

Hence,

$$\overline{\pi}_2=\frac{\left(\overline{\delta}\widehat{\sigma}^{11}+\widehat{\sigma}^{12}\right)\widehat{\pi}_1+\left(\overline{\delta}\widehat{\sigma}^{12}+\widehat{\sigma}^{22}\right)\widehat{\pi}_2}{\overline{\delta}\left(\overline{\delta}\widehat{\sigma}^{11}+\widehat{\sigma}^{12}\right)+\left(\overline{\delta}\widehat{\sigma}^{12}+\widehat{\sigma}^{22}\right)}=\frac{\left(\overline{\delta}\widehat{\sigma}_2^2-\widehat{\sigma}_{12}\right)\widehat{\pi}_1+\left(\widehat{\sigma}_1^2-\overline{\delta}\widehat{\sigma}_{12}\right)\widehat{\pi}_2}{\widehat{\sigma}_1^2+\overline{\delta}^2\widehat{\sigma}_2^2-2\overline{\delta}\widehat{\sigma}_{12}},$$

and

$$\begin{aligned}\widehat{\pi}_2-\overline{\pi}_2 &= -\frac{\left(\overline{\delta}\widehat{\sigma}_2^2-\widehat{\sigma}_{12}\right)}{\widehat{\sigma}_1^2+\overline{\delta}^2\widehat{\sigma}_2^2-2\overline{\delta}\widehat{\sigma}_{12}}\left(\widehat{\pi}_1-\overline{\delta}\widehat{\pi}_2\right);\\ \widehat{\pi}_1-\overline{\delta}\overline{\pi}_2 &= \frac{\left(\widehat{\sigma}_1^2-\overline{\delta}\widehat{\sigma}_{12}\right)}{\widehat{\sigma}_1^2+\overline{\delta}^2\widehat{\sigma}_2^2-2\overline{\delta}\widehat{\sigma}_{12}}\left(\widehat{\pi}_1-\overline{\delta}\widehat{\pi}_2\right).\end{aligned}$$

As

$$\begin{pmatrix}\left(\widehat{\sigma}_1^2-\overline{\delta}\widehat{\sigma}_{12}\right)\\ -\left(\overline{\delta}\widehat{\sigma}_2^2-\widehat{\sigma}_{12}\right)\end{pmatrix}'\widehat{\Sigma}^{-1}\begin{pmatrix}\left(\widehat{\sigma}_1^2-\overline{\delta}\widehat{\sigma}_{12}\right)\\ -\left(\overline{\delta}\widehat{\sigma}_2^2-\widehat{\sigma}_{12}\right)\end{pmatrix}=\begin{pmatrix}1 & -\overline{\delta}\end{pmatrix}\widehat{\Sigma}\begin{pmatrix}1\\ -\overline{\delta}\end{pmatrix}=\widehat{\sigma}_1^2+\overline{\delta}^2\widehat{\sigma}_2^2-2\overline{\delta}\widehat{\sigma}_{12},$$

it follows that

$$H\left(\overline{\delta},\overline{\pi}_2\right)=\frac{\left(\widehat{\pi}_1-\overline{\delta}\widehat{\pi}_2\right)'Z'Z\left(\widehat{\pi}_1-\overline{\delta}\widehat{\pi}_2\right)}{\widehat{\sigma}_1^2+\overline{\delta}^2\widehat{\sigma}_2^2-2\overline{\delta}\widehat{\sigma}_{12}}.$$

6.2 Total Relative Bias

Equivalently to (11) we can write

$$x_2=x_1\delta^*+Z_2\left(\pi_{22}-\pi_{12}\delta^*\right)+v_2-\delta^*v_1$$

where $\delta^*=\pi_{21}/\pi_{11}=\delta^{-1}$. Hence, under LRR1 weak instrument asymptotics, we have

$$x_2=x_1\delta^*-Z_2c\delta^*/\sqrt{n}+v_2-\delta^*v_1.$$

As

$$\begin{pmatrix}\frac{1}{\sqrt{n}}Z_2'M_{\widehat{x}_1}u\\ \frac{1}{\sqrt{n}}Z_2'M_{\widehat{x}_1}\left(v_2-\delta^*v_1\right)\end{pmatrix}\xrightarrow{d}\begin{pmatrix}\psi_{Z_2^{**}u}\\ \psi_{Z_2^{**}(v_2-\delta^*v_1)}\end{pmatrix}=N\left(0,\Omega^*\otimes Q^*\right),$$

$$\Omega^*=\begin{pmatrix}\sigma_u^2 & \sigma_{u2}-\delta^*\sigma_{u1}\\ \sigma_{u2}-\delta^*\sigma_{u1} & \sigma_2^2+\delta^{*2}\sigma_1^2-2\delta^*\sigma_{12}\end{pmatrix};\quad Z_2^{**}=M_{\widehat{x}_1}Z_2;\quad Q^*=\operatorname{plim}\left(n^{-1}Z_2^{**\prime}Z_2^{**}\right)$$

$$Q^*=\operatorname{plim}\frac{1}{n}Z_2'M_{\widehat{x}_1}Z_2=\operatorname{plim}\left(\frac{1}{n}\left(Z_2'Z_2-Z_2'\delta x_2\left(\delta x_2'Z\left(Z'Z\right)Z'\delta x_2\right)\delta x_2'Z_2\right)\right)=Q$$

It follows that $\psi_{Z_2^{**}u}=\psi_{Z_2^*u}$ and $\psi_{Z_2^{**}(v_2-\delta^*v_1)}=\psi_{Z_2^*(v_2-\delta^*v_1)}=-\frac{1}{\delta}\psi_{Z_2^*(v_1-\delta v_2)}$, as e.g.

$$Z_2'M_{\widehat{x}_1}u=Z_2'u-Z_2'x_1\left(x_1'P_Zx_1\right)^{-1}x_1'P_Zu=Z_2'u-Z_2'x_1\left(x_1'Z\left(Z'Z\right)^{-1}Z'x_1\right)x_1'Z\left(Z'Z\right)^{-1}Z'u.$$

Further,

$$\operatorname{plim}\left(n^{-1}Z'x_1\right) = \delta \operatorname{plim}\left(Z'x_2\right),$$

$$\widetilde{\lambda}^* = \sigma^{-1}_{v_2-\delta^* v_1} Qc\delta^* = \left(\delta^* \sigma_{v_1-\delta v_2}\right)^{-1} Qc\delta^* = \widetilde{\lambda},$$

so we get that

$$\widehat{\beta}_{2,SLS} - \beta_2 \xrightarrow{d} -\delta \frac{\sigma_u}{\sigma_{v_1-\delta v_2}} \frac{\left(\widetilde{\lambda}+\widetilde{z}_v\right)' \widetilde{z}_u}{\left(\widetilde{\lambda}+\widetilde{z}_v\right)' \left(\widetilde{\lambda}+\widetilde{z}_v\right)}$$

and hence, asymptotically,

$$E\left[\widehat{\beta}_{2,SLS}\right] - \beta_2 = -\delta E\left[\widehat{\beta}_{1,2SLS}\right] - \beta_1.$$

Using this, we can express the total relative bias (4) as

$$B^2 = b^2 \frac{\Sigma'_{Xu}\Sigma_X^{-1/2}\Sigma_X^{1/2} D\Sigma_X D\Sigma_X^{1/2}\Sigma_X^{-1/2}\Sigma_{Xu}}{\Sigma'_{Xu}\Sigma_X^{-1}\Sigma_{Xu}}$$

where

$$b = E\left[\frac{\left(\widetilde{\lambda}+\widetilde{z}_v\right)' \widetilde{z}_v}{\left(\widetilde{\lambda}+\widetilde{z}_v\right)' \left(\widetilde{\lambda}+\widetilde{z}_v\right)}\right]; \quad \Sigma_{Xu} = \begin{pmatrix} \sigma_{u1} \\ \sigma_{u2} \end{pmatrix}; \quad D = \frac{1}{\sigma_1^2 + \delta^2\sigma_2^2 - 2\delta\sigma_{12}} \begin{pmatrix} 1 & -\delta \\ -\delta & \delta^2 \end{pmatrix}.$$

Hence

$$B^2 \leq b^2 \max \operatorname{eval}\left(\Sigma_X^{1/2} D\Sigma_X D\Sigma_X^{1/2}\right).$$

As $\Sigma_X^{1/2} D\Sigma_X D\Sigma_X^{1/2}$ is a symmetric idempotent matrix, we get that $B^2 \leq b^2$. To show this, note that

$$\Sigma_X = \operatorname{plim} \frac{1}{n} \begin{pmatrix} x_1'x_1 & x_1'x_2 \\ x_1'x_2 & x_2'x_2 \end{pmatrix} = \pi_2' Q_{ZZ} \pi_2 \begin{pmatrix} \delta^2 & \delta \\ \delta & 1 \end{pmatrix} + \Sigma_V$$

and hence $\Sigma_X D = \Sigma_V D$. Let

$$d = \frac{1}{\sqrt{\sigma_1^2 + \delta^2\sigma_2^2 - 2\delta\sigma_{12}}} \begin{pmatrix} 1 \\ -\delta \end{pmatrix}$$

so that $D = dd'$, then $d'\Sigma_V d = 1$ and hence

$$\Sigma_X D\Sigma_X D = \Sigma_V D\Sigma_V D = \Sigma_V dd'\Sigma_V dd' = \Sigma_V dd' = \Sigma_V D = \Sigma_X D$$

and therefore

$$\Sigma_X^{1/2} D\Sigma_X D\Sigma_X^{1/2}\Sigma_X^{1/2} D\Sigma_X D\Sigma_X^{1/2} = \Sigma_X^{1/2} D\Sigma_X D\Sigma_X D\Sigma_X D\Sigma_X^{1/2} = \Sigma_X^{1/2} D\Sigma_X D\Sigma_X^{1/2}.$$

6.3 Stata Code

Simple Stata code to calculate the conditional F-statistic $F_{1|2,3}$ for the case of 3 endogenous variables, x1, x2 and x3, with 5 instruments z1,...,z5 and 2 other exogenous variables w1 and w2 is as follows:

```
ivregress 2sls x1 (x2 x3 = z1 z2 z3 z4 z5) w1 w2
predict res123, r
reg res123 z1 z2 z3 z4 z5 w1 w2
test z1 z2 z3 z4 z5
scalar Fsw = r(F)*r(df)/(r(df)-2)
di Fsw
```