# **On the Stock-Yogo Tables**

Christopher L. Skeels Frank Windmeijer

Discussion Paper 16 / 679

9 November 2016 Revised 25 November 2016



Department of Economics University of Bristol Priory Road Complex Bristol BS8 1TU United Kingdom

## On the Stock-Yogo Tables<sup>\*</sup>

Christopher L. Skeels<sup>†</sup> Department of Economics The University of Melbourne

Frank Windmeijer Department of Economics and IEU University of Bristol

November 2016

#### Abstract

A standard test for weak instruments compares the first-stage F-statistic to a table of critical values obtained by Stock and Yogo (2005) using simulations. We derive a closed-form solution for the expectation that determines these critical values. Inspection of this new result provides insights not available from simulation, and will allow software implementations to be generalized and improved. Of independent interest, our analysis makes contributions to the theory of confluent hypergeometric functions and the theory of ratios of quadratic forms in normal variables. A by-product of our developments is an expression for the distribution function of the non-central chi-squared distribution that we have not been able to find elsewhere in the literature. Finally, we explore the calculation of p-values for the first-stage F-statistic weak instruments test.

*Keywords*: Weak instruments, hypothesis testing, Stock-Yogo tables, hypergeometric functions, quadratic forms, p-values

JEL codes: C12, C36, C46, C52, C65, C88

<sup>\*</sup>We would like to thank Jon Temple for extensive comments on an earlier draft of this paper and Mark Schaffer for more recent comments. The usual caveat applies. Windmeijer acknowledges funding from the Medical Research Council (MC\_UU\_12013/9). Skeels thanks the Department of Economics at the University of Bristol for their hospitality during his visit there, which is when this paper was first written.

<sup>&</sup>lt;sup>†</sup>Corresponding author. E-mail address: Chris.Skeels@unimelb.edu.au.

### 1 Introduction

The development by Stock & Yogo (2005), hereafter SY, of a quantitative definition of weak instruments, based on either IV estimator bias or Wald test size distortion, has had a significant impact on econometric practice. Their idea was to relate the first-stage *F*-statistic (or, when there are multiple endogenous regressors, the Cragg-Donald (1993) statistic) to a non-centrality parameter that, in turn, was related to the aforementioned estimator bias or test size distortion. In this way they were able to use this *F*-statistic to test whether instruments were weak. That such tests have become part of the toolkit of many practitioners is evidenced by the fact that critical values for the SY tests are available within standard computer software, such as STATA (StataCorp, 2015) when using the intrinsic **ivregress** command or the **ivreg2** package (Baum et al., 2010). The difficulty in the SY approach is that, in order to compute appropriate critical values, it is necessary to evaluate a complicated integral as an intermediate step. SY did this by Monte Carlo simulation and the tables of critical values they provided are widely used in practice.

In this paper we focus on the two-stage least squares (2SLS) bias representation of weak instruments for the single endogenous variable case. We show that for this case, the integral mentioned above need not be estimated by simulation methods as it can be solved analytically and evaluated numerically using the intrinsic functions of software such as MATLAB (MathWorks, 2016). This result, Theorem 1, is presented in the next section, in which we also provide complete details of the model in question and the problem to be addressed.

From a empirical perspective there are two important consequences of Theorem 1. First, it allows us to examine the accuracy of the SY critical values that have become so important in empirical research. For the most part these critical values concord reasonably well with those that we derive analytically, although the most substantial differences occur in regions that we would argue are of practical significance. Second, it is now straightforward to generate more extensive sets of critical values, something that we do in Table 1 (also in Section 2). In particular, we extend the SY tables to include the  $k_z = 2$  case, where  $k_z$  is the number of instruments. The SY tables do not start until  $k_z = 3$ , which is an especially stringent requirement in practice, as finding valid instruments can be extraordinarily difficult in many applications.

From a theoretical perspective, Theorem 1 provides a foundation that allows us to explore analytically certain patterns that exist in the SY tables, something that can only be alluded to on the basis of simulation results. These cases are explored in Section 3 of the paper, although much of the theoretical development underlying the discussion is relegated to Appendix C. In order to establish the results of Appendix C, it proved necessary to derive a number of properties of confluent hypergeometric functions. These results are developed in Appendix B and represent a contribution to the theory of confluent hypergeometric functions that is of independent interest. To further support the discussion of Section 3, we present in Section 4 some Monte Carlo simulation results, where we explore the sampling distributions of the *F*statistics in relation to the bias of the 2SLS estimator relative to that of the OLS estimator, for the  $k_z = 2$  and  $k_z = 3$  cases.

Key to the development of Theorem 1 is the expectation of the ratio of a bilinear form in perfectly correlated normally distributed random variables, that differ only in their means, to a quadratic form in one of these same random variables, which is of some independent interest. We note in passing that the problem could be re-cast as one involving the expectation of a ratio of quadratic forms in normal variables, although in this form the normal variables have a singular distribution and both the numerator and the denominator weighting matrices are also singular, with the numerator weighting matrix asymmetric. This observation explains the difficulty in evaluating the integral analytically, but is also the reason that the expectation ultimately has such a simple structure. This expectation is evaluated in Appendix A.

Given the recent statement on p-values issued by the American Statistical Association Board of Directors (Wasserstein & Lazar, 2016), it would be remiss of a paper such as this to be silent on the matter. In Section 5 we extend our discussion to show how p-values can be readily calculated on the basis of our earlier results.

Final remarks appear in Section 6, including some comments on extending our approach to more general situations than the case of a single endogenous regressor that is the focus of our discussion in this paper.

## 2 An Analytic Development of Stock-Yogo

Consider the simple model

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

where  $y = [y_1, \ldots, y_n]'$ ,  $x = [x_1, \ldots, x_n]'$  and  $u = [u_1, \ldots, u_n]'$  are  $n \times 1$  vectors, with n the number of observations. The regressor x is assumed endogenous, so that  $E[u|x] \neq 0$ . Other exogenous regressors in the model, including the constant, have been partialled out.

We can implicitly define a set of instruments via the following linear projection

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

where Z is an  $n \times k_z$  matrix of instruments (with full column rank),  $\pi$  a  $k_z \times 1$  vector of parameters and v is an  $n \times 1$  error vector. In this model,  $k_z - 1$  is the degree of over-identification. We assume that individual observations are independently and identically distributed, and

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} \mid z_i \sim (0, \Sigma), \quad \text{with} \quad \Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}, \quad i = 1, 2, \dots, n,$$

where  $z'_i$  denotes the *i*th row of Z. A test for  $H_0: \pi = 0$  against  $H_1: \pi \neq 0$ , is the

so-called first-stage F-statistic

$$F = \frac{\hat{\pi}' Z' Z \hat{\pi}}{k_z \hat{\sigma}_v^2} \xrightarrow[d]{H_0} \frac{\chi_{k_z}^2}{k_z},\tag{3}$$

where  $\hat{\pi} = (Z'Z)^{-1}Z'x$  and  $\hat{\sigma}_v^2 = n^{-1}x'(I_n - Z(Z'Z)^{-1}Z')x$ . Here a large value of the statistic is evidence against the null hypothesis, which is that the nominated instruments are irrelevant.

Following Staiger & Stock (1997), we consider values of  $\pi$  local to zero, as  $\pi = c/\sqrt{n}$ . We then obtain for the concentration parameter  $\mu_n^2$ ,

$$\mu_n^2 = \frac{\pi' Z' Z \pi}{\sigma_v^2} \xrightarrow{p} \frac{c' Q_{zz} c}{\sigma_v^2} \equiv \mu^2,$$

where  $Q_{zz} = E[z_i z'_i] = \text{plim}_{n\to\infty} n^{-1}Z'Z$  is positive definite by assumption. We see that  $k_z F$  is a sample analogue of  $\mu^2$ . With this formulation the testing problem previously discussed is equivalent to that of testing  $H'_0: \mu^2 = 0$  against  $H'_1: \mu^2 >$ 0. Rather than testing for the irrelevance of instruments, SY characterised weak instruments as a situation where  $\mu^2$  was greater than zero but proximate to it. Specifically, their testing problem can be thought of as  $H''_0: \mu^2 = \mu_0^2 > 0$  against  $H''_1: \mu^2 > \mu_0^2$ , for some suitably specified value of  $\mu_0^2$ . The statistic F is still a natural one in this problem although, of course, the null distribution is no longer the central distribution associated with  $\mu_0^2 = 0$ . Instead we have

$$F \xrightarrow[d]{H_0} \frac{\chi_{k_z,\mu_0^2}'^2}{k_z},\tag{4}$$

where  $\chi_{k,\delta}^{\prime 2}$  denotes a random variable following a non-central chi-squared distribution with k degrees of freedom and non-centrality parameter  $\delta \ge 0.1$  Let

$$\chi_{\alpha} = \chi_{k_z,\mu_0^2}^{\prime 2} (1-\alpha)$$

<sup>&</sup>lt;sup>1</sup>Some references specify the non-centrality parameter for a non-central chi-squared distribution as  $\delta$ , whereas others specify it as  $\delta/2$ . We have adopted the former convention here.

denote the  $(1 - \alpha)100$ th quantile of a non-central chi-squared distribution with  $k_z$ degrees of freedom and non-centrality parameter  $\mu_0^2$ . Then the relevant size  $\alpha$  critical region is

$$\left\{F: F > cv_{\alpha} = \frac{\chi_{\alpha}}{k_z}\right\},\tag{5}$$

where  $\chi_{\alpha}$  can be obtained, for given  $\mu_0^2$  and  $k_z$ , as the solution to either of the equations

$$1 - \alpha = e^{-\mu^2/2} \sum_{j=0}^{\infty} \frac{(\mu^2/2)^j}{2^{k_z/2+j} j! \Gamma\left(\frac{k_z}{2} + j\right)} \int_0^{\chi_\alpha} e^{-s/2} s^{k_z/2+j-1} \,\mathrm{d}s \tag{6a}$$

$$= e^{-\mu^2/2} e^{-\chi_{\alpha}/2} \left(\frac{\chi_{\alpha}}{2}\right)^{k_z/2} \sum_{j=0}^{\infty} \frac{(\chi_{\alpha}\mu^2/4)^j}{j! \Gamma\left(\frac{k_z+2}{2}+j\right)} {}_1F_1\left(1; \frac{k_z+2}{2}+j; \frac{\chi_{\alpha}}{2}\right), \quad (6b)$$

where  ${}_{1}F_{1}(\cdot;\cdot;\cdot)$  denotes a confluent hypergeometric function (Slater, 1960). Although (6a) is standard (e.g. Johnson et al., 1995, equation (29.2)), somewhat surprisingly, we have been unable to find (6b) elsewhere in the literature and so it may be a new representation for the cumulative distribution function (cdf) of a non-central chi-square distribution, one in the spirit of Venables (1971) (see Johnson et al., 1995, p.438, Equation (29.11)').<sup>2</sup> The distinguishing feature is that the representation involves confluent hypergeometric functions ( ${}_{1}F_{1}$ 's) rather than expressing the distribution in terms of modified Bessel functions, which can themselves be represented in terms of generalised hypergeometric functions (Slater, 1966) with no numerator parameters and only a single denominator parameter ( ${}_{0}F_{1}$ 's), as seen in the second member of Johnson et al. (1995, equation (29.4)).

The aspect of the SY approach that remains outstanding is the choice of  $\mu_0^2$ . Their quantitative definition of the weakness of a set of instruments is couched in terms of the impact that it has on inference. They provided two possible definitions that variously reflect the known consequences of weak instruments for (i) estimation,

<sup>&</sup>lt;sup>2</sup>To derive (6b), set t = s/2, say, and then express the incomplete gamma function as a confluent hypergeometric function using the first member of NIST (2015, Equation 8.5.1) (with their z replaced by our  $\frac{1}{2}\chi_{\alpha}$ ).

through the bias of the estimator, and (ii) hypothesis testing, through the size of the Wald test relative to its nominal size. It is the former that is in most common use and the approach of interest here.<sup>3</sup>

In particular, SY relate the bias of the 2SLS estimator of  $\beta$ ,  $\hat{\beta}_{2SLS}$ , relative to that of the ordinary least squares estimator of  $\beta$ ,  $\hat{\beta}_{OLS}$ , to the first-stage *F*-statistic by showing that they are both related to  $\mu^2$ . A value for  $\mu^2$ , denoted  $\mu_0^2$ , is then chosen to allow a certain level of relative bias. Specifically, let  $B_n$  denote the relative bias. Then, provided  $k_z \geq 2$ , so that the bias of 2SLS exists,

$$|B_n| = \left| \frac{\mathbf{E} \left[ \hat{\beta}_{2SLS} \right] - \beta}{\mathbf{E} \left[ \hat{\beta}_{OLS} \right] - \beta} \right|,$$

and

$$\lim_{n \to \infty} |B_n| = \left| \mathbf{E} \left[ \frac{(\xi - \lambda_0)'\xi}{\xi'\xi} \right] \right| \equiv |B|, \tag{7}$$

where  $\xi \sim N(\lambda_0, I_{k_z})$ . The test for weak instruments then proceeds as follows:

- 1. The practitioner chooses a value for |B|, e.g. |B| = 0.1, if an asymptotic relative bias of less than 10% is deemed acceptable.
- 2. Given  $k_z$  and |B|,  $\mu_0^2 = \lambda'_0 \lambda_0$  is obtained on solving (7).
- 3. Given  $\mu_0^2$ , critical values for F can be determined, which are proportional to those of the non-central chi-squared distribution as specified in (4).
- 4. The null of weak instruments is then rejected for sufficiently large values of the first-stage F-statistic, and we conclude that |B| is no larger than the value chosen in Step 1 above.

The difficulty in the procedure just described is that, at Step 2, there is an integral that must be evaluated as part of a search for  $\mu_0^2$ . SY do this using a 20,000 draw Monte Carlo simulation. This is unnecessary as the integral can be solved analytically. The result is summarized in the following theorem.

<sup>&</sup>lt;sup>3</sup>The exact details of these arguments can be found in SY and will not be repeated here.

**Theorem 1.** If B is as defined in equation (7) then, provided  $k_z \ge 2$ ,

$$B = {}_{1}F_{1}\left(1;\frac{k_{z}}{2};-\frac{\mu_{0}^{2}}{2}\right) > 0.$$
(8)

*Proof.* The result follows immediately from Theorem A.1, in Appendix A, which establishes the equality, and Corollary B.1.1, of Appendix B.1, which establishes the inequality.  $\Box$ 

Theorem 1 allows the use of efficiently programmed intrinsic functions in readily available software, such as MATLAB (MathWorks, 2016), at each step of a search for  $\mu_0^2$  rather than having to estimate an integral by simulation.<sup>4</sup> For the special case of  $k_z = 2$ ,

$${}_{1}F_{1}\left(1;\frac{k_{z}}{2};-\frac{\mu_{0}^{2}}{2}\right) = \exp\left\{-\frac{\mu_{0}^{2}}{2}\right\},\tag{9}$$

making evaluation of the expression especially simple.

Using our result, we provide in Table 1 an extended version of that panel of SY Table 5.1 corresponding to a single endogenous variable, which is the set of critical values most commonly used. We note that SY start their tables at  $k_z = 3$  even though, following the arguments of Kinal (1980), finite biases will exist for all  $k_z \ge 2$ . As this is a practically relevant case we include it in Table 1.

Where Table 1 overlaps with SY (Table 5.1), we are able to provide an indication of the difference made by the analytical evaluation of the expectation in (7). As shown in Table 2, the differences are typically small, with the largest differences when  $k_z$  and B are themselves small, which we would argue is the most important case in practice.<sup>5</sup>

<sup>&</sup>lt;sup>4</sup>In the absence of such intrinsic functions, computational aspects of hypergeometric functions are discussed in Johansson (2016).

<sup>&</sup>lt;sup>5</sup>We have also computed simulated critical values from 20,000 random draws as in SY, but repeating the exercise 1000 times. The resulting mean critical values are virtually identical to those in Table 1, with the maximum difference being 0.02.

$k_z \setminus B$	0.01	0.05	0.1	0.15	0.2	0.25	0.3
2	11.57	9.02	7.85	7.14	6.61	6.19	5.83
3	46.32	13.76	9.18	7.52	6.60	5.96	5.49
4	63.10	16.72	10.23	7.91	6.67	5.88	5.32
5	72.55	18.27	10.78	8.11	6.71	5.82	5.19
6	78.59	19.19	11.08	8.21	6.70	5.75	5.09
7	82.75	19.79	11.25	8.25	6.67	5.69	5.01
8	85.78	20.20	11.36	8.26	6.64	5.63	4.93
9	88.07	20.49	11.42	8.25	6.60	5.58	4.87
10	89.86	20.70	11.46	8.24	6.56	5.52	4.81
11	91.30	20.86	11.49	8.22	6.53	5.48	4.76
12	92.47	20.99	11.50	8.20	6.49	5.43	4.71
13	93.43	21.08	11.50	8.17	6.46	5.39	4.67
14	94.25	21.16	11.50	8.15	6.42	5.36	4.63
15	94.94	21.22	11.49	8.13	6.39	5.32	4.59
16	95.54	21.26	11.49	8.11	6.36	5.29	4.56
17	96.05	21.30	11.48	8.08	6.34	5.26	4.53
18	96.50	21.33	11.46	8.06	6.31	5.23	4.50
19	96.09	21.35	11.45	8.04	6.29	5.21	4.47
20	97.25	21.37	11.44	8.02	6.26	5.18	4.45
21	97.56	21.39	11.43	8.00	6.24	5.16	4.43
22	97.84	21.40	11.41	7.98	6.22	5.14	4.40
23	98.09	21.41	11.40	7.96	6.20	5.12	4.38
24	98.32	21.41	11.39	7.94	6.18	5.10	4.36
25	98.53	21.42	11.38	7.93	6.16	5.08	4.35
26	98.71	21.42	11.36	7.91	6.15	5.06	4.33
27	98.88	21.42	11.35	7.90	6.13	5.05	4.31
28	99.04	21.42	11.34	7.88	6.11	5.03	4.30
29	99.18	21.42	11.32	7.87	6.10	5.02	4.28
30	99.31	21.42	11.31	7.85	6.08	5.00	4.27

Table 1: 5% Critical values  $(cv_{0.05}^{SW})$  for single endogenous regressor, 2SLS bias

## 3 Some Further Consequences of Theorem 1

Theorem 1 allows us to prove a variety of further results that can only be speculated about on the basis of simulation results. These results are discussed below, with proofs relegated to Appendix B.

Remark 1. Implicit in Theorem 1, and a consequence of Theorem B.1 (see Corollary B.1.1), is the observation that whenever  $k_z \ge 2$ , OLS and 2SLS are always asymptotically biased in the same direction, making the absolute value function of

$k_z \setminus B$	0.05	0.10	0.20	0.30
3	0.15	-0.10	-0.14	-0.10
4	0.13	0.04	0.04	0.02
5	0.10	0.05	0.06	0.06
6	0.09	0.04	0.06	0.06
7	0.07	0.04	0.06	0.07
8	0.05	0.03	0.05	0.06
9	0.04	0.04	0.05	0.05
10	0.04	0.03	0.05	0.05
11	0.04	0.01	0.03	0.04
12	0.02	0.02	0.04	0.04
13	0.02	0.02	0.03	0.04
14	0.02	0.02	0.03	0.04
15	0.01	0.02	0.03	0.04
16	0.02	0.01	0.03	0.03
17	0.01	0.01	0.02	0.03
18	0.01	0.02	0.02	0.03
19	0.01	0.01	0.02	0.04
20	0.01	0.01	0.02	0.03
21	0.00	0.01	0.02	0.03
22	0.00	0.01	0.02	0.03
23	0.00	0.01	0.02	0.03
24	0.00	0.01	0.02	0.03
25	0.00	0.00	0.02	0.02
26	0.00	0.01	0.01	0.02
27	0.00	0.01	0.01	0.03
28	0.00	0.02	0.02	0.02
29	0.00	0.01	0.01	0.03
30	0.00	0.01	0.01	0.02

Table 2: Differences:  $cv_{0.05}^{SW} - cv_{0.05}^{SY}$ 

Note: The values of  $cv_{0.05}^{SY}$  are taken from Stock & Yogo (2005, Table 5.1).

|B| in (7) redundant.

Remark 2. The values of the relative biases of  $\hat{\beta}_{2SLS}$  and  $\hat{\beta}_{OLS}$  are explored in Figure 1 for different values of the parameters  $k_z$  and  $\mu^2/2$ . The figure illustrates that, for  $k_z \geq 2$ , the function is increasing in its argument, which is  $-\mu^2/2$ . This result is established analytically in Theorem B.2. Note also that, as  $\mu^2 \rightarrow 0$ , the information in the instruments approaches zero, and so the local-to-zero asymptotic bias of  $\hat{\beta}_{2SLS}$  approaches that of  $\hat{\beta}_{OLS}$  from below. Hence, the limit of the relative



Figure 1: Plots of  $B = {}_1F_1\left(1; \frac{k_z}{2}; -\frac{\mu^2}{2}\right)$  against  $\frac{\mu^2}{2}$  for  $k_z = 2, 3$  and 6.

asymptotic biases at  $\mu^2 = 0$  is unity, which is the value of  ${}_1F_1\left(1; \frac{k_z}{2}; 0\right)$ .

Remark 3. Certain patterns in Table 1 are readily established, as illustrated by the following result.<sup>6</sup>  $\Box$ 

**Theorem 2.** The critical values  $cv_{\alpha}$ , presented in Table 1, are decreasing functions of B for given  $k_z$ .

Proof. See Appendix C.2.

Heuristically, Theorem 2 states that the critical values will necessarily decrease as one moves from left to right across any given row of Table 1; that is, the critical values decrease as the practitioner is willing to accept increasing amounts of 2SLS bias relative to that of OLS. The intuition behind the results is as follows. An increase in *B* for fixed  $k_z$  implies, by Theorem B.2, that the argument of the confluent

<sup>&</sup>lt;sup>6</sup>Theorem 2 is similar in spirit to Das Gupta & Perlman (1974, p.180, Remark 4.1), although they only address the numerator of the ratio in equation (5). Consequently, Das Gupta & Perlman are silent on the relative magnitudes of  $\chi_{\alpha}$  and  $k_z$  which, in essence, is the content of Theorem 2.

hypergeometric function in (8) must increase, i.e. that  $\mu^2/2$  must decrease. As  $\mu^2$  approaches zero, the non-central chi-squared distribution from which critical values are drawn approaches a central chi-squared and the corresponding quantiles become smaller. Hence, as one moves across columns from left to right in Table 1, the  $cv_{\alpha}$  become smaller.

Theorem 2 explains the row behaviour of Table 1. Explaining the column behaviour is much more complicated. Observation suggests the following to be true.

**Conjecture.** For given B, the critical values  $cv_{\alpha}$ , presented in Table 1, are increasing functions of  $k_z$  up to some value, k say, whereafter they are decreasing functions of  $k_z$ . k is a decreasing function of B.

Some intuition for the Conjecture is available from the definition of  $cv_{\alpha}$ , see (5), if one considers the impact of increasing the number of instruments by one, from  $k_z$  to  $k_z + 1$ , with superscripts '0' and '1' distinguishing the two cases, respectively. For given *B* and  $\alpha$ ,

$$cv_{\alpha}^{1} - cv_{\alpha}^{0} \gtrless 0$$
 as  $\frac{\chi_{\alpha}^{1} - \chi_{\alpha}^{0}}{\chi_{\alpha}^{0}} \gtrless \frac{1}{k_{z}}$ 

k is then that value of  $k_z$  after which the  $cv_{\alpha}$  start diminishing.

Remark 4. Although *B* does not exist when  $k_z = 1$ , the confluent hypergeometric function of Theorem 1 remains well-defined. In Appendix E we analyse the properties of  ${}_1F_1\left(1; \frac{1}{2}; -\frac{\mu^2}{2}\right)$ .

### 4 Some Monte Carlo Results

We follow Sanderson & Windmejer (2016) and specify the model is as in (1) and (2), with  $\beta = 1$  and

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right).$$

	В								
	0.01		0.05		0.10		0.20		
$k_z = 3$	mean	st dev							
$\hat{eta}_{OLS}$	1.4950	0.0086	1.4989	0.0087	1.4994	0.0087	1.4997	0.0087	
$\hat{eta}_{2SLS}$	1.0054	0.0998	1.0241	0.2222	1.0506	0.3161	1.1025	0.4276	
F	34.713	6.7626	8.0336	3.1828	4.7849	2.3952	3.0948	1.8630	
rel bias	0.0	108	0.0482		0.1014		0.2052		
$\mu_0^2/k_z$	33.674		7.0445		3.7754		2.0902		
$\operatorname{cv} F$	46.316		13.765		9.1815		6.5960		
rej freq ${\cal F}$	0.0515		0.0505		0.0508		0.0511		
$k_z = 2$	mean	st dev							
$\hat{\beta}_{OLS}$	1.4996	0.0087	1.4997	0.0087	1.4997	0.0087	1.4998	0.0087	
$\hat{\beta}_{2SLS}$	1.0056	0.4398	1.0256	0.7195	1.0519	0.9651	1.0981	1.1404	
F	5.6124	3.1989	4.0004	2.6492	3.2963	2.3746	2.6011	2.0502	
rel bias	0.0111		0.0513		0.1039		0.1962		
$\mu_0^2/k_z$	4.6052		2.9957		2.3026		1.6094		
$\operatorname{cv} F$	11.572		9.0232		7.8521		6.6087		
rej freq $F$	0.0509		0.0	0.0507		0.0505		0.0498	

Table 3: Simulation results for  $k_z = 3$  and  $k_z = 2$ 

Notes: Sample size n = 10,000, number of Monte Carlo replications is 100,000.

The instruments in Z are  $k_z$  independent standard normally distributed random variables and  $\pi = (c_{Bk_z}\iota_{k_z})/\sqrt{n}$ , where  $\iota_{k_z}$  is a  $k_z$  vector of ones, and with  $c_{Bk_z}$ chosen such that the relative bias B is equal to 0.01, 0.05, 0.10 or 0.20, for values of  $k_z = 3$  and  $k_z = 2$ . The sample size n = 10,000 and the results are presented in Table 3 for 100,000 Monte Carlo replications.

For  $k_z = 3$ , the results are exactly in line with the theory: the Monte Carlo relative biases are equal to B and the rejection frequencies of the first-stage F-test are 5% at the 5% nominal level, using the critical values reported in Table 1.

The results for  $k_z = 2$  are also in line with the theory, although we see here that the standard deviations of  $\hat{\beta}_{2SLS}$  are much larger than those of the  $k_z = 3$  case at the same values of B. This is due to the fact that the information needed to obtain the same relative bias is much smaller for the  $k_z = 2$  case than for the  $k_z = 3$  case , as reflected by their smaller  $\mu_0^2/k_z$  values, but it also reflects the problem that the second moment does not exist when the degree of over-identification is equal to 1. The interquartile ranges for the 2SLS estimator when  $k_z = 2$  are 0.3296, 0.4170, 0.4811 and 0.5570 for B = 0.01, 0.05, 0.10 and 0.20, respectively. These Monte Carlo results therefore confirm our theoretical findings for the  $k_z = 2$  case. Clearly some caution should be exercised when working with 2SLS in this case because it possesses no second moment.

## 5 *p*-Values

*p*-values are readily available as a straightforward extension of our earlier analysis. Specifically, from (4) we have the limiting result

$$k_z \times F \xrightarrow[d]{H_0} \chi_{k_z,\mu_0^2}^{\prime 2}.$$
 (10)

For any particular sample value of the *F*-test, say  $\widehat{F}$ , if  $X \sim \chi_{k_z,\mu_0^2}^{\prime 2}$  then the p-value for the SY weak instruments test considered in this paper is simply  $\Pr\left(X \ge k_z \times \widehat{F}\right)$ . Of course, the problem here is the determination of  $\mu_0^2$ . Table 4 reports those values of  $\mu_0^2/k_z$  that were calculated in order to construct Table 1. For those values of *B* considered in Table 1, we now have the parameters  $k_z$  and  $\mu_0^2/k_z$ . Consequently, any computer software that can evaluate a non-central chi-squared cdf can readily calculate *p*-values for the test for weak instruments considered here.

### 6 Final Remarks

The main contribution of this paper has been to resolve analytically an integral as a special function, obviating the need to resolve it by simulation. This integral is of independent interest in the theory of ratios of quadratic forms in normal variables. Here it is of primary interest because it provides a functional relationship between the bias in the 2SLS estimator and the limiting sampling distribution of a test statistic that SY proposed for testing the presence of weak instruments, when the

$k_z \setminus B$	0.01	0.05	0.1	0.15	0.2	0.25	0.3
2	4.605	2.996	2.303	1.897	1.609	1.386	1.204
3	33.674	7.045	3.775	2.677	2.090	1.706	1.426
4	50.000	10.000	5.000	3.329	2.483	1.960	1.599
5	59.799	11.793	5.784	3.774	2.761	2.144	1.724
6	66.332	12.991	6.315	4.081	2.958	2.277	1.816
7	70.998	13.848	6.696	4.304	3.102	2.375	1.885
8	74.498	14.491	6.982	4.472	3.212	2.450	1.938
9	77.221	14.992	7.205	4.604	3.298	2.510	1.980
10	79.398	15.392	7.384	4.709	3.367	2.558	2.014
11	81.180	15.720	7.531	4.796	3.424	2.597	2.043
12	82.665	15.993	7.653	4.868	3.471	2.630	2.066
13	83.922	16.224	7.756	4.929	3.511	2.658	2.086
14	84.999	16.423	7.845	4.981	3.546	2.682	2.104
15	85.932	16.594	7.922	5.027	3.576	2.703	2.119
16	86.749	16.745	7.989	5.067	3.602	2.721	2.132
17	87.470	16.877	8.048	5.102	3.626	2.738	2.144
18	88.110	16.995	8.101	5.133	3.646	2.752	2.154
19	88.683	17.101	8.148	5.161	3.665	2.765	2.163
20	89.199	17.196	8.191	5.186	3.681	2.777	2.172
21	89.666	17.281	8.229	5.209	3.697	2.787	2.179
22	90.090	17.360	8.264	5.230	3.710	2.797	2.186
23	90.477	17.431	8.296	5.249	3.723	2.806	2.193
24	90.833	17.496	8.326	5.266	3.734	2.814	2.198
25	91.159	17.556	8.353	5.282	3.745	2.821	2.204
26	91.461	17.612	8.377	5.297	3.755	2.828	2.209
27	91.740	17.663	8.400	5.311	3.764	2.834	2.213
28	91.999	17.711	8.422	5.323	3.772	2.840	2.217
29	92.241	17.755	8.442	5.335	3.780	2.846	2.221
30	92.466	17.797	8.460	5.346	3.787	2.851	2.225

Table 4: Values for  $\mu_0^2/k_z$  corresponding to Table 1

null of weak instruments is true. Analysis of this special function provides theoretical foundations for the remarks of Section 3, which explore patterns observed in Table 1 as the parameters B and  $k_z$  vary. This analysis required the derivation of certain results that are of independent interest in the theory of confluent hypergeometric functions. A final contribution of the paper is to explore the problem of p-values of the aforementioned test for weak instruments, on the basis of our earlier theoretical developments. We provide information such that any computer software than can evaluate a non-central chi-squared cdf can readily compute *p*-values in essentially

all practical circumstances.

Relative to what SY provided, there are some things that we have not done and they merit further comment. First, one aspect of the SY tables that we have not addressed relates to those tables based on size distortions of a Wald statistic. This is a much more difficult analytical problem than has been addressed here and it is not clear that there is much benefit in tackling it as, in our estimation, the bias tables are in much more frequent use, making them of greater practical relevance.

Second, in the results presented here, we have restricted attention to the case of a single endogenous regressor. It is possible to extend our results to the more general case, however, the outcome involves invariant polynomials with multiple matrix arguments, sometimes referred to as Davis polynomials, which are computationally troublesome.<sup>7</sup> Also, the SY results for multiple endogenous variable cases are only approximate and provide upper bounds on critical values for the Cragg-Donald minimum eigenvalue test. This is due to the fact that the SY analysis is for the full matrix of first-stage parameters  $\Pi$  being local to zero. In contrast, Sanderson & Windmejer (2016) define weak identification as the rank of  $\Pi$  being local to a rank reduction of one. For these asymptotics, the single endogenous variable results apply, only the degrees of freedom need to be adjusted for the number of endogenous variables, see Sanderson & Windmejer (2016) for details.

Finally, in support of the results presented in the paper we provide two MATLAB programs on an 'as is' basis. The first of these, Table1.m, provides the body of Table 1. The second program, entitled sypval.m, provides *p*-values. Appendix D provides some discussion on the contents of these programs. The programs are available at https://sites.google.com/site/skeelscv/.

<sup>&</sup>lt;sup>7</sup>Some progress towards addressing the computational aspects of these polynomials has been made by Hillier et al. (2009, 2014).

## References

- Abramowitz, M., Stegun, I. A., 1964. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables (Dover, New York), ninth Dover printing.
- Ancarani, L. U., Gasaneo, G., 2008. Derivatives of any order of the confluent hypergeometric function  ${}_{1}F_{1}(a;b;z)$  with respect to the parameter a or b, Journal of Mathematical Physics 49.
- Angrist, J. D., Pischke, J.-S., 2009. Mostly Harmless Econometrics. An Empiricist's Companion (Princeton University Press, Princeton and Oxford).
- Baum, C. F., Schaffer, M. E., Stillman, S., 2010. ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML, and k-class regression, boston College Department of Economics, Statistical Software Components S425401. Downloadable from http://ideas.repec.org/c/boc/bocode/ s425401.html.
- Cohen, J. D., 1988. Noncentral chi-square: Some observations on recurrence, The American Statistician 42, 120–122.
- Cragg, J. G., Donald, S. G., 1993. Testing identifiability and specification in instrument variable models, Econometric Theory 9, 222–240.
- Das Gupta, S., Perlman, M. D., 1974. Power of the noncentral f-test: Effect of additional variates on Hotelling's  $t^2$ -test, Journal of the American Statistical Association 69, 174–180.
- Herz, C. S., 1955. Bessel functions of matrix argument, Annals of Mathematics 61, 474–523.
- Hillier, G. H., Kan, R., Wang, X., 2009. Computationally efficient recursions for

top-order invariant polynomials with applications, Econometric Theory 25, 211–242.

- Hillier, G. H., Kan, R., Wang, X., 2014. Generating functions and short recursions, with applications to the moments of quadratic forms in noncentral normal vectors, Econometric Theory 30, 436–473.
- Hillier, G. H., Kinal, T. W., Srivastava, V. K., 1984. On the moments of Ordinary Least Squares and Instrumental Variables estimators in a general structural equation, Econometrica 52, 185–202.
- Johansson, F., 2016. Computing hypergeometric functions rigorously, https://hal. inria.fr/hal-01336266v2. Downloaded 15 September, 2016.
- Johnson, N. L., Kotz, S., Balakrishnan, N., 1995. Continuous Univariate Distributions, volume 2, 2nd edition (John Wiley & Sons, Inc., New York).
- Kinal, T. W., 1980. The existence of moments of k-class estimators, Econometrica 49, 241–249.
- MathWorks, 2016. MATLAB and Statistics Toolbox Release 2016b, The MathWorks Inc., Natick, Massachusetts, United States.
- NIST, 2015. Digital Library of Mathematical Functions, National Institute of Standards and Technology, Release 1.0.10 of 2015-08-07. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, and B. V. Saunders, eds. Online companion to Olver et al. (2010). http://dlmf.nist.gov/.
- Olver, F. W. J., Lozier, D. W., Boisvert, R. F., Clark, C. W., eds., 2010. NIST Handbook of Mathematical Functions (Cambridge University Press, New York, NY), print companion to NIST (2015).

- Piegorsch, W. W., Casella, G., 1985. The existence of the first negative moment, The American Statistician 39, 60–62.
- Sanderson, E., Windmejer, F., 2016. A weak instrument F-test in linear IV models with multiple endogenous variables, Journal of Econometrics 190, 212–221.
- Slater, L. J., 1960. Confluent Hypergeometric Functions (Cambridge University Press, Cambridge).
- Slater, L. J., 1966. Generalized Hypergeometric Functions (Cambridge University Press, Cambridge).
- Staiger, D., Stock, J. H., 1997. Instrumental variables regression with weak instruments, Econometrica 65, 557–586.
- StataCorp, 2015. Stata Statistical Software: Release 14, StataCorp LP, College Station, TX.
- Stock, J. H., Yogo, M., 2005. Testing for weak instruments in linear IV regression, in: Andrews, D. W. K., Stock, J. H., eds., Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, chapter 5, 80–108 (Cambridge University Press, Cambridge, UK).
- Venables, W. N., 1971. Inference Problems Based on Non-Central Distributions, Ph.D. thesis, Department of Statistics, University of Adelaide, Australia.
- Wasserstein, R. L., Lazar, N. A., 2016. The ASA's statement on p-values: Context, process, and purpose, The American Statistician 70, 129–133.

# A The Expectation of a Particular Function of Normal Random Variables

**Theorem A.1.** Suppose that  $\xi \sim N(\lambda, I_k)$ . Then,

$$\mathbf{E}\left[\frac{(\xi-\lambda)'\xi}{\xi'\xi}\right] = \begin{cases} {}_1F_1\left(1;\frac{k}{2};-\frac{\mu^2}{2}\right), & k \ge 2, \\ \\ diverges, & k = 1, \end{cases}$$

where  $\mu^2 = \lambda' \lambda$ .

*Proof.* The proof is in two parts. First, we establish conditions under which the expectation exists, which proves to be when k > 1. Second, we evaluate the expectation in this case.

#### (i) Existence

Let  $\mathcal{E}(\xi; \lambda) = (\xi - \lambda)' \xi / \xi' \xi$ . The existence of  $\mathbb{E}[\mathcal{E}(\xi; \lambda)]$  requires the existence of some finite constant  $\mathcal{C}$  such that  $\mathbb{E}[|\mathcal{E}(\xi; \lambda)|] \leq \mathcal{C}$ . Observe that

$$|\mathcal{E}(\xi;\lambda)| = \left|1 - \frac{\lambda'\xi}{\xi'\xi}\right| \le 1 + \frac{|\lambda'\xi|}{\xi'\xi} \le 1 + (\mu^2)^{1/2} \frac{(\xi'\xi)^{1/2}}{\xi'\xi} = 1 + \left(\frac{\mu^2}{\xi'\xi}\right)^{1/2}$$

Hence,

$$E[|\mathcal{E}(\xi;\lambda)|] \le 1 + (\mu^2)^{1/2} E[(\xi'\xi)^{-1/2}].$$

Next, write

$$\mathbf{E}\left[(\xi'\xi)^{-1/2}\right] = \frac{\exp\{-\mu^2/2\}}{(2\pi)^{k/2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{\xi'\xi}{2}\right\} \, (\xi'\xi)^{-1/2} \, \exp\{\lambda'\xi\} \, \mathrm{d}\xi,$$

where  $\int_{-\infty}^{\infty} f(\xi) d\xi$  denotes the k-fold integral,  $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(\xi_1, \dots, x_k) \prod_{i=1}^k d\xi_i$ with  $\xi_i$  the *i*th element of  $\xi$ . In accord with Herz (1955, Lemma 1.4), we can decompose almost all k-vectors  $\xi$  into  $\xi = hr^{1/2}$ , where  $h = \xi(\xi'\xi)^{-1/2}$ , so that h'h = 1, and  $r = \xi'\xi > 0$ , with volume elements

$$d\xi = 2^{-1} r^{(k-2)/2} dh dr.$$
 (A.1)

This is essentially a transformation to polar coordinates. The resulting expression is

$$\mathbf{E}\left[(\xi'\xi)^{-1/2}\right] = \frac{\exp\{-\mu^2/2\}}{2(2\pi)^{k/2}} \int_{r>0} \int_{h'h=1} \exp\left\{-\frac{r}{2}\right\} r^{(k-3)/2} \exp\left\{r^{1/2}\lambda'h\right\} \,\mathrm{d}h \,\mathrm{d}r,$$

almost everywhere. The integral with respect to h is readily evaluated using Hillier et al. (1984, Equation (6)):

$$\int_{h'h=1} \exp\left\{r^{1/2}\lambda'h\right\} \,\mathrm{d}h = \frac{2\pi^{k/2}}{\Gamma\left(\frac{k}{2}\right)} {}_0F_1\left(\frac{k}{2};\frac{\mu^2 r}{4}\right). \tag{A.2}$$

Making this substitution we obtain

$$\mathbf{E}\left[(\xi'\xi)^{-1/2}\right] = \frac{\exp\{-\mu^2/2\}}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} \int_{r>0} \exp\left\{-\frac{r}{2}\right\} r^{(k-3)/2} \,_{0}F_1\left(\frac{k}{2};\frac{\mu^2 r}{4}\right) \,\mathrm{d}r,$$

where the integral a variant of the well-known Laplace transform of a generalised hypergeometric function; see for example (NIST, 2015, Equation 16.5.3).<sup>8</sup> The Laplace transform will be convergent provided  $(k-1)/2 > 0 \Rightarrow k > 1$ , in which case

$$\begin{split} \frac{1}{\Gamma\left(\frac{k}{2}\right)} \int_{r>0} \exp\left\{-\frac{r}{2}\right\} \, r^{(k-3)/2} \, _{0}F_{1}\left(\frac{k}{2};\frac{\mu^{2}r}{4}\right) \, \mathrm{d}r \\ &= 2^{(k-1)/2} \, _{1}F_{1}\left(\frac{k-1}{2};\frac{k}{2};\frac{\mu^{2}}{2}\right). \quad (A.3) \end{split}$$

<sup>&</sup>lt;sup>8</sup>All the references to NIST (2015) are also available online at http://dlmf.nist.gov by adding the equation number with the final digit prepended by 'E' instead of a decimal point. Thus, NIST (2015, Equation 16.5.3) is available at http://dlmf.nist.gov/16.5E3, and so on. Similarly, http://dlmf.nist.gov/16.2 and http://dlmf.nist.gov/16.2.ii are the URL's of Section 16.2 and Section 16.2(ii), respectively.

Consequently, we have established that, for k > 1,

$$\begin{split} \mathbf{E}\left[\left|\mathcal{E}(\xi;\lambda)\right|\right] &\leq 1 + \exp\{-\mu^2/2\}(\mu^2/2)^{1/2} \ _1F_1\left(\frac{k-1}{2};\frac{k}{2};\frac{\mu^2}{2}\right) \\ &= 1 + (\mu^2/2)^{1/2} \ _1F_1\left(\frac{1}{2};\frac{k}{2};-\frac{\mu^2}{2}\right), \end{split}$$

where the final equality is an application of Kummer's transformation (NIST, 2015, Equation 13.2.39), and so  $E[\mathcal{E}(\xi; \lambda)]$  exists as the confluent hypergeometric function is known to be convergent for all finite values of its parameters and argument (Slater, 1966, p.45).

In the event that k = 1, we see that  $E[\mathcal{E}(\xi; \lambda)]$  reduces to  $1 - \lambda E[\xi^{-1}]$ . But the expectation of the inverse of a Normal random variable is known not to exist, see for example Piegorsch & Casella (1985, Example 2.2), which completes our existence results.

#### (ii) Expectation when k > 1

The development of the expectation is very similar to the process of determining its existence. Given the normality assumption on  $\xi$  and assuming that  $k \ge 2$ , we can write

$$\operatorname{E}\left[\frac{(\xi-\lambda)'\xi}{\xi'\xi}\right] = \frac{\exp\{-\lambda'\lambda/2\}}{(2\pi)^{k/2}} \int_{-\infty}^{\infty} \left[1 - \frac{\lambda'\xi}{\xi'\xi}\right] \exp\left\{-\frac{\xi'\xi}{2}\right\} \exp\{\lambda'\xi\} \,\mathrm{d}\xi = \mathcal{I} \text{ (say).}$$

Make the transformation to polar coordinates in accord with (A.1) to obtain

$$\begin{aligned} \mathcal{I} &= \frac{\exp\{-\lambda'\lambda/2\}}{2(2\pi)^{k/2}} \int_{r>0} \exp\{-r/2\} r^{(k-2)/2} \\ &\times \left\{ \int_{h'h=1} \exp\{\lambda' h r^{1/2}\} \,\mathrm{d}h - r^{-1} \int_{h'h=1} \lambda' h r^{1/2} \exp\{\lambda' h r^{1/2}\} \,\mathrm{d}h \right\} \,\mathrm{d}r. \end{aligned}$$

Next, write

$$\lambda' h r^{1/2} \exp\{\lambda' h r^{1/2}\} = \left. \frac{\mathrm{d} \exp\{(1+t)\lambda' h r^{1/2}\}}{\mathrm{d}t} \right|_{t=0}$$

and evaluate the integrals over h'h = 1 using (A.2). This yields, on replacing  $\lambda'\lambda$ 

by  $\mu^2$ ,

$$\begin{split} \mathcal{I} &= \frac{\exp\{-\mu^2/2\}}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} \, \int_{r>0} \, \exp\left\{-\frac{r}{2}\right\} \, r^{(k-2)/2} \\ & \times \left\{{}_0F_1\left(\frac{k}{2};\frac{\mu^2 r}{4}\right) - r^{-1}\left[\frac{\mathrm{d}}{\mathrm{d}t} \, {}_0F_1\left(\frac{k}{2};\frac{(1+t)^2\mu^2 r}{4}\right)\right]_{t=0}\right\} \, \mathrm{d}r, \end{split}$$

where  ${}_{p}F_{q}(a_{1}, \ldots, a_{p}; b_{1}, \ldots, b_{q}; \xi)$  denotes a generalized hypergeometric function . Finally, differentiating with respect to t, using say NIST (2015, Equation 16.3.1), evaluating the derivative at t = 0, and then resolving the resulting Laplace transforms using (A.3) yields

$$\begin{split} \mathcal{I} &= \exp\{-\mu^2/2\} \left[ {}_1F_1\left(\frac{k}{2};\frac{k}{2};\frac{\mu^2}{2}\right) - \frac{\mu^2}{k} {}_1F_1\left(\frac{k}{2};\frac{k+2}{2};\frac{\mu^2}{2}\right) \right] \\ &= \exp\{-\mu^2/2\} {}_1F_1\left(\frac{k-2}{2};\frac{k}{2};\frac{\mu^2}{2}\right) \\ &= {}_1F_1\left(1;\frac{k}{2};-\frac{\mu^2}{2}\right), \end{split}$$
(A.4)

where the second last equality exploits one of the relationships for contiguous confluent hypergeometric functions (NIST, 2015, Equation 13.3.4) and the final equality is another application of Kummer's transformation. This completes the proof.  $\Box$ 

# B Some Useful Properties of Confluent Hypergeometric Functions

In this appendix we establish some properties of the confluent hypergeometric function that are used in the body of the paper. Specifically, we develop two results that apply in quite general settings to the response of confluent hypergeometric functions to changes in their argument when the numerator and denominator parameters are held fixed. A third result is presented that applies to a confluent hypergeometric function with specific parameter values that is of particular interest to this paper, the proof of which exploits the previously developed results.

### **B.1** Some Consequences for ${}_1F_1(a;b;s)$ as s Changes

**Theorem B.1.** If a > 0, b > 0, and  $s \ge 0$  then  ${}_{1}F_{1}(a;b;s) \ge 1$ .

*Proof.* Consider the series expansion of the confluent hypergeometric function (NIST, 2015, Equation 13.2.2):

$$_{1}F_{1}(a;b;s) = \sum_{j=0}^{\infty} \frac{(a)_{j}}{(b)_{j}} \frac{s^{j}}{j!},$$
(B.1)

where the Pochhammer symbol  $(\alpha)_j$  denotes the rising factorial function, a polynomial of order j in a,<sup>9</sup>

$$\alpha(\alpha+1)(\alpha+2)\dots(\alpha+n-1) = \Gamma(\alpha+j)/\Gamma(\alpha), \tag{B.2}$$

and where  $\Gamma(\alpha)$  denotes the usual Gamma function. For  $\alpha > 0$ ,  $(\alpha)_n > 0$  follows immediately from the left-hand side of (B.2).

From the positivity of  $(\alpha)_n$  for all  $0 < \alpha < \infty$ , it follows that if a > 0, b > 0, and s > 0 then  ${}_1F_1(a;b;s)$  is an infinite sum of positive terms, with the first term in the sum being equal to unity. Given that the series converges absolutely to some finite number (see, for example, Slater, 1966, p.45), it must be that the value of the series is some number greater than unity.

If s = 0 then the series on the right-hand side of (B.1) terminates at n = 0, whereupon  ${}_{1}F_{1}(a;b;0) = 1$ , which completes the proof.

**Corollary B.1.1.** If  $b \ge a > 0$  but s < 0 then  $0 < {}_1F_1(a; b; s) \le 1$ .

*Proof.* Positivity: If s < 0, Kummer's transformation yields  $e^{s} {}_{1}F_{1}(b-a;b;-s)$ . Then  $e^{s} > 0$  for all  $-\infty < s < \infty$  and, if b > a,  ${}_{1}F_{1}(b-a;b;-s) > 0$  by Theorem B.1, whereas, if b = a, the series terminates after the first term and so  ${}_{1}F_{1}(0;b;-s) = 1 > 0$ .

<sup>&</sup>lt;sup>9</sup>Important special cases of this function are  $(a)_0 = 1$ , including  $(0)_0 = 1$ , and  $(1)_j = j!$ . A useful collection of results on  $(a)_j$  can be found in Slater (1966, Appendix I).

Upper Bound: If  $b \ge a > 0$  then  $(b-a)_j \le (b)_j$  for all j = 1, 2, 3, ... Consequently, for s < 0,

$$_{1}F_{1}\left(b-a;b;-s\right) = |_{1}F_{1}\left(b-a;b;-s\right)| \leq \sum_{j=0}^{\infty} \left|\frac{(b-a)_{j}}{(b)_{j}}\frac{s^{j}}{j!}\right| \leq \sum_{j=0}^{\infty} \left|\frac{s^{j}}{j!}\right| = e^{|s|},$$

where the first equality follows from the positivity of the confluent hypergeometric function and the second inequality follows because  $b \ge a > 0 \Rightarrow (b-a)_j \le (b)_j$ . Hence,  $e^s {}_1F_1 (b-a;b;-s) \le 1$ .

**Theorem B.2.** If  $b \ge a > 0$  then  ${}_{1}F_{1}(a;b;s)$  is an increasing function of s.

*Proof.* We need to establish that the derivative  ${}_{1}F_{1}(a;b;s)$  with respect to s is everywhere positive. It is well-known (NIST, 2015, Equation 16.3.1) that

$$\frac{\mathrm{d}}{\mathrm{d}s} {}_{1}F_{1}(a;b;s) = \frac{a}{b} {}_{1}F_{1}(a+1;b+1;s).$$
(B.3)

Clearly, a/b > 0, as  $b \ge a > 0$  by assumption. Similarly, by Theorem B.1,  ${}_1F_1(a+1;b+1;s) \ge 1$  for  $s \ge 0$ . Consequently,  ${}_1F_1(a;b;s)$  is an increasing function of s under these conditions.

Now suppose that s < 0, which we will represent as s = -w, w = |s|. By Kummer's transformation (NIST, 2015, Equation 13.2.39)

$${}_{1}F_{1}(a+1;b+1;-w) = e^{-w} {}_{1}F_{1}(b-a;b+1;w).$$
 (B.4)

The exponential function is positive for all  $-\infty < w < \infty$  and, by Theorem B.1,  ${}_{1}F_{1}(b-a; b+1; w) > 0$ , as  $b-a \ge 0$  by assumption. That is,  $d_{1}F_{1}(a; b; s) / ds > 0$ when s < 0. This completes the proof.

**Theorem B.3.** The function  ${}_{1}F_{1}\left(1;\frac{1}{2};-s\right)$  is decreasing for  $0 < s < s_{0}$ , increasing for  $s > s_{0}$ , and is neither increasing nor decreasing for  $s = s_{0}$ , where  $s_{0} \approx 2.2559$ .

*Proof.* Differentiating the function with respect to s and then applying Kummer's

transformation yields

$$\frac{\mathrm{d}}{\mathrm{d}s} {}_{1}F_{1}\left(1;\frac{1}{2};-s\right) = -2 {}_{1}F_{1}\left(2;\frac{3}{2};-s\right) = -2e^{-s} {}_{1}F_{1}\left(-\frac{1}{2};\frac{3}{2};s\right).$$

Hence,

$$\begin{split} \frac{\mathrm{d}}{\mathrm{d}s} \,_{1}F_{1}\left(1;\frac{1}{2};-s\right) &= -2e^{-s}\sum_{j=0}^{\infty} \frac{\left(-\frac{1}{2}\right)_{j}}{\left(\frac{3}{2}\right)_{j}} \frac{s^{j}}{j!} \\ &= -2e^{-s}\left\{1-\frac{s}{3}\sum_{j=1}^{\infty} \frac{\left(\frac{1}{2}\right)_{j-1}\left(1\right)_{j-1}}{\left(2\right)_{j-1}\left(\frac{5}{2}\right)_{j-1}} \frac{s^{j-1}}{(j-1)!}\right\} \\ &= -2e^{-s} + \frac{2se^{-s}}{3} \,_{2}F_{2}\left(\frac{1}{2},1;2,\frac{5}{2};s\right), \end{split}$$

where the second equality has used the identity  $j! = (1)_j = (2)_{j-1}$ . In particular,

The zero of the derivative is  $s_0 \approx 2.2559$ ;  ${}_1F_1\left(1; \frac{1}{2}; -s_0\right) \approx -0.2847$ . For s > 0, the hypergeometric function is an increasing function of s, whereas 3/s is a decreasing function of s, hence there can be at most one such zero; see Figure B.1. It follows that  ${}_1F_1\left(1; \frac{1}{2}; -s\right)$  is decreasing for  $0 < s < s_0$  and increasing for  $s > s_0$ .

## **B.2** Some Consequences for $_1F_1(a;b;s)$ as b Changes

Let us now turn to the behaviour of confluent hypergeometric functions as the denominator parameter changes.

**Theorem B.4.** If s > 0 then, for b > m,  $m = 1, 2, 3, \ldots, {}_{1}F_{1}(m; b; -s)$  is an increasing function of b.

*Proof.* Observe that

$$\frac{\mathrm{d}}{\mathrm{d}b} \,_{1}F_{1}\left(m;b;-s\right) = e^{-s} \frac{\mathrm{d}}{\mathrm{d}b} \,_{1}F_{1}\left(b-m;b;s\right) = e^{-s} \sum_{j=0}^{\infty} \, \frac{s^{j}}{j!} \frac{\mathrm{d}}{\mathrm{d}b} \frac{(b-m)_{j}}{(b)_{j}},$$



Figure B.1: Plots of  $y = {}_2F_2\left(\frac{1}{2}, 1; 2, \frac{5}{2}; s\right)$  and  $y = \frac{3}{s}$ .

where the first equality follows from Kummer's transformation and the second from (B.1). Noting that (Ancarani & Gasaneo, 2008, Equation (4))

$$\frac{\mathrm{d}\left(\tau\right)_{j}}{\mathrm{d}\tau} = \left(\tau\right)_{j} \left[\Psi(\tau+j) - \Psi(\tau)\right],$$

where  $\Psi(\tau)\Gamma(\tau) = d\Gamma(\tau)/d\tau$  defines the psi function, and then using the recursion (Abramowitz & Stegun, 1964, Equation 6.3.6)

$$\Psi(\tau + j) = \Psi(\tau) + \sum_{p=0}^{j-1} \frac{1}{\tau + p},$$

we find that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}b} \frac{(b-m)_j}{(b)_j} &= \left[ \frac{\mathrm{d} (b-m)_j}{\mathrm{d}b} (b)_j - (b-m)_j \frac{\mathrm{d} (b)_j}{\mathrm{d}b} \right] \left[ (b)_j \right]^{-2} \\ &= \frac{(b-m)_j}{(b)_j} \left\{ \left[ \Psi(b-m+j) - \Psi(b-m) \right] - \left[ \Psi(b+j) - \Psi(b) \right] \right\} \\ &= \frac{(b-m)_j}{(b)_j} \sum_{p=0}^{m-1} \left[ \frac{1}{b-m+p} - \frac{1}{b-m+j+p} \right] . \\ &= \frac{(b-m)_j}{(b)_j} \sum_{p=0}^{m-1} \frac{1}{b-m+p} \left[ 1 - \frac{(b-m+p)_j}{(b-m+p+1)_j} \right]. \end{aligned}$$

Thus,

$$\frac{\mathrm{d}}{\mathrm{d}b} {}_{1}F_{1}(m;b;-s) = \sum_{p=0}^{m-1} \frac{e^{-s}}{b-m+p} \left[ {}_{1}F_{1}(b-m;b;s) - {}_{2}F_{2}(b-m,b-m+p;b,b-m+p+1;s) \right].$$
(B.5)

This representation of the derivative depends crucially upon the assumption that m is integer valued. Alternate representations, based on  ${}_{1}F_{1}(m;b;-s)$  directly, can be obtained from Ancarani & Gasaneo (2008, Equations (6b) and (16)), with (B.5) closest in spirit to the former.

The derivative can be signed by first noting that

$$_{1}F_{1}\left( b-m;b;s\right) = _{2}F_{2}\left( b-m,b-m+p+1;b,b-m+p+1;s\right)$$

whence it follows, on writing  $\nu_j(b-m,p) = (b-m)_j / \left[ (b)_j (b-m+p+1)_j \right]$ , that

$$\begin{split} {}_{1}F_{1}\left(b-m;b;s\right) &= {}_{2}F_{2}\left(b-m,b-m+p;b,b-m+p+1;s\right) \\ &= \sum_{j=0}^{\infty} \frac{\nu_{j}(b-m,p)s^{j}}{j!} \left[ (b-m+p+1)_{j} - (b-m+p)_{j} \right] \\ &= \sum_{j=0}^{\infty} \frac{\nu_{j}(b-m,p)s^{j}\left(b-m+p+1\right)_{j-1}j}{j!} \\ &= \frac{(b-1)s}{b(b-m+p+1)} \\ &\times \sum_{j=1}^{\infty} \frac{(b-m+1)_{j-1}\left(b-m+p+1\right)_{j-1}}{(b)_{j-1}\left(b-m+p+2\right)_{j-1}} \frac{s^{j-1}}{(j-1)!} \\ &= \frac{(b-1)s}{b(b-m+p+1)} \\ &\times {}_{2}F_{2}\left(b-m+1,b-m+p+1;b+1,b-m+p+2;s\right). \end{split}$$

Given the assumptions on the parameters, the hypergeometric function is larger

than unity. Hence

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}b} \,_{1}F_{1}\left(m;b;-s\right) &= \sum_{p=0}^{m-1} \frac{(b-1)se^{-s}}{b(b-m+p)(b-m+p+1)} \\ &\times \,_{2}F_{2}\left(b-m+1,b-m+p+1;b+1,b-m+p+2;s\right) \quad (\mathrm{B.6}) \\ &> \sum_{p=0}^{m-1} \frac{(b-1)se^{-s}}{b(b-m+p)(b-m+p+1)} > 0, \end{aligned}$$

which completes the proof.

C Analysis of Table 1

### C.1 Preliminaries

In this appendix we analyse how the critical values  $cv_{\alpha}$ , presented in Table 1, change in response to changes in one of either  $k_z$  or B when the other is held fixed; that is, as one either moves down columns of the table or across rows, from left to right, respectively. From equation (5),

$$cv_{\alpha} = \frac{\chi_{\alpha}}{k_z},$$
 (C.1)

with  $\chi_{\alpha}$  the solution to the equation

$$1 - \alpha = \int_0^{\chi_\alpha} f(s \mid k_z, \mu^2) \, \mathrm{d}s, \qquad (C.2)$$

where  $f(s \mid k_z, \mu^2)$  denotes the density function of a non-central chi-squared random variable; specifically

$$f(s \mid k_z, \mu^2) = e^{-\mu^2/2} \sum_{j=0}^{\infty} \frac{(\mu^2/2)^j}{j! 2^{(k_z+2j)/2} \Gamma\left(\frac{k_z+2j}{2}\right)} e^{-s/2} s^{(k_z+2j-2)/2}$$
$$= \sum_{j=0}^{\infty} \kappa_j(k_z, \mu^2) e^{-s/2} s^{(k_z+2j-2)/2}, \tag{C.3}$$

where

$$\kappa_j(k_z, \mu^2) = \frac{e^{-\mu^2/2} (\mu^2/2)^j}{j! 2^{(k_z+2j)/2} \Gamma\left(\frac{k_z+2j}{2}\right)}.$$
 (C.4)

The parameter  $\mu^2$  is chosen to satisfy

$$B = {}_{1}F_{1}\left(1;\frac{k_{z}}{2};-\frac{\mu^{2}}{2}\right).$$
(C.5)

By Corollary B.1.1, the absolute values can be ignored as the confluent hypergeometric function is positive for all  $\mu^2 \ge 0$  whenever  $k_z \ge 2$ , which shall be assumed for the rest of this appendix unless indicated otherwise.

### C.2 The Consequence of Varying B for Fixed $k_z \ge 2$

With  $k_z$  held fixed we have, from (C.1),

$$\frac{\mathrm{d}cv_{\alpha}}{\mathrm{d}B} = \frac{1}{k_z} \left. \frac{\mathrm{d}\chi_{\alpha}}{\mathrm{d}\mu^2} \right/ \frac{\mathrm{d}B}{\mathrm{d}\mu^2} \tag{C.6}$$

First, from (B.3) and Theorem B.4 (with m = 2),

$$\frac{\mathrm{d}B}{\mathrm{d}\mu^2} = \frac{\mathrm{d}B}{\mathrm{d}(-\mu^2/2)} \frac{\mathrm{d}(-\mu^2/2)}{\mathrm{d}\mu^2} = -\frac{1}{k_z} {}_1F_1\left(2;\frac{k_z+2}{2};-\frac{\mu^2}{2}\right) < 0, \tag{C.7}$$

for all  $\mu^2$  and  $k_z$ . Second, using Leibniz's Rule for the differentiation of integrals, we can differentiate both sides of (C.2) with respect to  $\mu^2$  to obtain,

$$0 = \int_{0}^{\chi_{\alpha}} \frac{\partial f(s \mid k_{z}, \mu^{2})}{\partial \mu^{2}} \, \mathrm{d}s + f(\chi_{\alpha} \mid k_{z}, \mu^{2}) \frac{\mathrm{d}\chi_{\alpha}}{\mathrm{d}\mu^{2}}$$
$$\Rightarrow \frac{\mathrm{d}\chi_{\alpha}}{\mathrm{d}\mu^{2}} = -\frac{1}{f(\chi_{\alpha} \mid k_{z}, \mu^{2})} \int_{0}^{\chi_{\alpha}} \frac{\partial f(s \mid k_{z}, \mu^{2})}{\partial \mu^{2}} \, \mathrm{d}s. \tag{C.8}$$

Note that (C.8) implicitly assumes  $0 < \chi_{\alpha} < \infty$ , so that  $0 < \alpha < 1$ . In the event that either  $\chi_{\alpha} = 0$  or  $\chi_{\alpha}$  is infinite, then  $f(s \mid k_z, \mu^2) = 0$ , as does its derivative with respect to  $\mu^2$ , making the representation (C.8) invalid. Indeed, as these cases

are on the boundaries of support of a non-central chi-squared random variable, the ordinary derivative is not well-defined and so the approach taken above would require modification. For this reason, hereafter, we shall assume that  $0 < \alpha < 1$ .

From (C.3), the integrand in (C.8) is (Cohen, 1988, Equation (2))

$$\frac{\partial f(s \mid k_z, \mu^2)}{\partial \mu^2} = \frac{1}{2} \left[ f(s \mid k_z + 2, \mu^2) - f(s \mid k_z, \mu^2) \right].$$

Integrating by parts allows us to write

$$\int_0^{\chi_\alpha} e^{-s/2} s^{(k_z+2j)/2} \, \mathrm{d}s = \left[ -2e^{-s/2} s^{(k_z+2j)/2} \right]_0^{\chi_\alpha} + (k_z+2j) \int_0^{\chi_\alpha} e^{-s/2} s^{(k_z+2j-2)/2} \, \mathrm{d}s$$
$$= -2e^{-\chi_\alpha/2} \chi_\alpha^{(k_z+2j)/2} + (k_z+2j) \int_0^{\chi_\alpha} e^{-s/2} s^{(k_z+2j-2)/2} \, \mathrm{d}s,$$

and so (C.8) becomes

$$\frac{\mathrm{d}\chi_{\alpha}}{\mathrm{d}\mu^{2}} = -\frac{1}{2f(\chi_{\alpha} \mid k_{z}, \mu^{2})} \sum_{j=0}^{\infty} \left\{ \kappa_{j}(k_{z}+2, \mu^{2}) \left[ -2e^{-\chi_{\alpha}/2}\chi_{\alpha}^{(k_{z}+2j)/2} + (k_{z}+2j) \int_{0}^{\chi_{\alpha}} e^{-s/2} s^{(k_{z}+2j-2)/2} \, \mathrm{d}s \right] - \kappa_{j}(k_{z}, \mu^{2}) \int_{0}^{\chi_{\alpha}} e^{-s/2} s^{(k_{z}+2j-2)/2} \, \mathrm{d}s \right\}$$

$$= \frac{f(\chi_{\alpha} \mid k_{z}+2, \mu^{2})}{f(\chi_{\alpha} \mid k_{z}, \mu^{2})}$$

$$- \sum_{j=0}^{\infty} \frac{\kappa_{j}(k_{z}+2, \mu^{2})(k_{z}+2j) - \kappa_{j}(k_{z}, \mu^{2})}{2f(\chi_{\alpha} \mid k_{z}, \mu^{2})} \int_{0}^{\chi_{\alpha}} e^{-s/2} s^{(k_{z}+2j-2)/2} \, \mathrm{d}s$$

$$= \frac{f(\chi_{\alpha} \mid k_{z}+2, \mu^{2})}{f(\chi_{\alpha} \mid k_{z}, \mu^{2})} > 0,$$
(C.9)

as  $\kappa_j(k_z+2,\mu^2)(k_z+2j) - \kappa_j(k_z,\mu^2) = 0$ . The positivity of the ratio follows because each of the functions f are values of non-central chi-squared density functions which differ only in their degrees of freedom,  $k_z$  versus  $k_z + 2$  respectively, and so are both everywhere positive for all  $0 < \chi_\alpha < \infty$ , as is assumed above. As an aside, we know that as degrees of freedom increase for given  $\mu^2$  these functions cross, which means that sometimes  $f(\chi_\alpha \mid k_z, \mu^2) > f(\chi_\alpha \mid k_z + 2, \mu^2)$  and sometimes the converse is true. That is, we are unable to bound  $d\chi_\alpha/d\mu^2$  from above. Combining (C.6), (C.7), and (C.9), we find that

$$\frac{\mathrm{d}cv_{\alpha}}{\mathrm{d}B} = -\frac{f(\chi_{\alpha} \mid k_z + 2, \mu^2)}{f(\chi_{\alpha} \mid k_z, \mu^2) \, {}_1F_1\left(2; \frac{k_z + 2}{2}; -\frac{\mu^2}{2}\right)} < 0, \tag{C.10}$$

which confirms the behaviour observed in Table 1. That is, for given values of  $k_z$ , the critical values  $cv_{\alpha}$  are decreasing functions of the asymptotic bias B.

### D Some Remarks on Computational Aspects

For the most part, both the programs Table1.m and sypval.m rely on intrinsic MAT-LAB functions. Once the relevant inputs are available then the structure of the programs is immediately apparent. Specifically, for given values of  $k_z$  and B, it is necessary to obtain the corresponding value for  $\mu_0^2$  from the non-linear equation (8). We adopt a fairly simple-minded approach to this, by iterating from a starting value to the correct solution using a bisection algorithm.

Our starting values are chosen as follows. When  $k_z = 2$ , we know from (9) that the values of  $\mu_0^2$  can be calculated exactly as  $\mu_0^2 = -2 \ln B$  and so no search is required. When  $k_z > 2$  we exploit an approximation asymptotic in  $\mu_0^2$  (Slater, 1960, equation (4.1.8)) that reduces to  $\mu_0^2 \approx (k_z - 2)/B$ . As expected, the performance of the approximation improves as B decreases which, for fixed  $k_z$ , corresponds to increasing  $\mu_0^2$  (see Appendix C.2). Nevertheless, for all cases where  $k_z > 2$ , this approximation provides much better starting values in the search for  $\mu_0^2$  than do naive alternatives, such as starting the search from zero (say). Moreover, this approximation performs best under exactly the same circumstances that naive methods are at their slowest, affording considerable computational time savings. As the values of  $\mu_0^2/k_z$  are much more stable for any given B than are the  $\mu_0^2$ , as can be deduced from Table 4, we use this parameterisation in our search algorithm.

### **E** Some Remarks on the $k_z = 1$ Case

The SY approach is not available if  $k_z = 1$  because the bias of 2SLS does not exist, hence |B| is undefined.<sup>10</sup> Nevertheless, given the difficulties often encountered in finding appropriate instruments, the exactly identified model is one of considerable practical relevance. As the confluent hypergeometric function of Theorem 1 remains well-defined when  $k_z = 1$ , one might ask if it could provide an ad hoc basis for a test for weak instruments in this case, based on F, in the spirit of the SY approach.

The function  ${}_{1}F_{1}\left(1;\frac{1}{2};-\frac{\mu^{2}}{2}\right)$  displays behaviours that are quite different to what was observed in over-identified models. These behaviours are displayed in Figure E.1 where we plot both the confluent hypergeometric function and its absolute value against  $\mu^{2}/2$ , when  $k_{z} = 1$ . Note that in the figure we use the symbol *B* to represent the confluent hypergeometric function  ${}_{1}F_{1}\left(1;\frac{1}{2};-\frac{\mu^{2}}{2}\right)$ , rather than the expectation  $\mathrm{E}\left[(\xi - \lambda)'\xi/\xi'\xi\right]$ , with the latter unbounded when  $k_{z} = 1$ .

In Figure E.1 we see that neither *B* nor |B| are monotonic in  $\mu^2/2$  when  $k_z = 1$ , in contrast to the over-identified cases. This lack of monotonicity is established in Theorem B.3. That the confluent hypergeometric function can take negative values when  $k_z = 1$  means that this case is the only one considered where taking the absolute value of the hypergeometric function has any material impact on observed behaviour. We can establish numerically that *B*, and hence |B|, both have a zero at  $\mu^2/2 \approx 0.8540$ . As this is in the region where the hypergeometric function is a decreasing value of its argument (Theorem B.3), we see that as *B* moves through its zero to the right, so that  $\mu^2$  is increasing, it becomes negative and appears to stay that way, with a minimum of approximately -0.2847 occurring at  $\mu^2/2 \approx$ 2.2559. Clearly |B| cannot become negative and so, at  $\mu^2/2 \approx 2.2559$ , it has a local maximum of approximately 0.2847. Consequently, there are three values of  $\mu^2$ that yield the same value of |B| for 0 < |B| < 0.2847, there are two values of  $\mu^2$ 

<sup>&</sup>lt;sup>10</sup>Similarly, in the proof of Theorem A.1, we established that  $E[(\xi - \lambda_0)'\xi/\xi'\xi]$  was unbounded when  $k_z = 1$ .



Figure E.1: Plots of  $B = {}_{1}F_{1}\left(1; \frac{1}{2}; -\frac{\mu^{2}}{2}\right)$  and its absolute value against  $\frac{\mu^{2}}{2}$ .

corresponding to |B| = 0.2847, and for  $|B| \in \{0\} \cap (0.2847, 1]$  there is a one-to-one mapping between |B| and  $\mu^2 < \infty$ .

Observe in Figure E.1 that there are 3 values of  $\mu^2$  corresponding to |B| = 0.1. Setting  $\mu_0^2 = 13.83$ , the largest of these numbers, we find a critical value for the first-stage *F*-test of 28.77. At this level of information, the 2SLS estimator appears well-behaved. This is shown in Table E.1, which shows the estimation result of a Monte Carlo analysis as in Section 4 for  $k_z = 1$ . Even though it has no moments as the model is just-identified, we find that the Monte Carlo relative bias is indeed 10% with the rejection frequency of the *F*-test again 5%. The same holds at the smaller values of |B| of 0.05 and 0.01, for which the largest implied values of  $\mu_0^2$  are 23.41 and 103.06, with the estimation results very similar to those for  $k_z = 3$ . However, when we consider the |B| = 0.20 case, for which  $\mu_0^2$  is 8.198, the lack of moments of the 2SLS estimator becomes apparent, with the standard deviation now very large at 6.05. These results suggest that the approximation might be useful for

	В							
	-0.01		-0.05		-0.10		-0.20	
	mean	st dev						
$\hat{eta}_{OLS}$	1.4949	0.0086	1.4988	0.0086	1.4993	0.0086	1.4996	0.0087
$\hat{\beta}_{2SLS}$	0.9954	0.1003	0.9753	0.2327	0.9495	0.4389	0.8936	6.0492
F	104.11	20.411	24.429	9.8352	14.821	7.5732	9.1757	5.8671
relbias	-0.0092		-0.0496		-0.1011		-0.2130	
$\mu_0^2$	103.06		23.412		13.830		8.198	
$\operatorname{cv} F$	139.17		42.035		28.769		20.323	
rej freq ${\cal F}$	0.0496		0.0507		0.0485		0.0489	

Table E.1: Simulation results for  $k_z = 1$ 

Notes: Sample size n = 10,000, number of Monte Carlo replications is 100,000.

the smaller values of |B|, if one works with the largest implied values of  $\mu_0^2$ , even though the 2SLS estimator does not possess any moments in this case.

Confirming the approximate median unbiasedness of the just-identified 2SLS estimator, see for example the discussion in Angrist & Pischke (2009, p.209), we find that the median biases, not reported in the table, are very close to 0 at all values of  $\mu^2$ .