This article was downloaded by:[The University of Manchester]
On: 27 May 2008
Access Details: [subscription number 773564140]
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Econometric Reviews

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713597248>

Entropy-Based Moment Selection in the Presence of Weak Identification

Alastair R. Hall ^{ab}; Atsushi Inoue ^{bc}; Changmock Shin ^d

^a University of Manchester, Manchester, UK

^b North Carolina State University, Raleigh, North Carolina, USA

^c University of British Columbia, Vancouver, Canada

^d Samsung Economic Research Institute, Seoul, S. Korea

Online Publication Date: 01 July 2008

To cite this Article: Hall, Alastair R., Inoue, Atsushi and Shin, Changmock (2008)
'Entropy-Based Moment Selection in the Presence of Weak Identification',
Econometric Reviews, 27:4, 398 — 427

To link to this article: DOI: 10.1080/07474930801960261

URL: <http://dx.doi.org/10.1080/07474930801960261>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

ENTROPY-BASED MOMENT SELECTION IN THE PRESENCE OF WEAK IDENTIFICATION

Alastair R. Hall,¹ Atsushi Inoue,² and Changmock Shin³

¹*University of Manchester, Manchester, UK and
North Carolina State University, Raleigh, North Carolina, USA*

²*University of British Columbia, Vancouver, Canada and
North Carolina State University, Raleigh, North Carolina, USA*

³*Samsung Economic Research Institute, Seoul, S. Korea*

□ *Hall et al. (2007) propose a method for moment selection based on an information criterion that is a function of the entropy of the limiting distribution of the Generalized Method of Moments (GMM) estimator. They establish the consistency of the method subject to certain conditions that include the identification of the parameter vector by at least one of the moment conditions being considered. In this article, we examine the limiting behavior of this moment selection method when the parameter vector is weakly identified by all the moment conditions being considered. It is shown that the selected moment condition is random and hence not consistent in any meaningful sense. As a result, we propose a two-step procedure for moment selection in which identification is first tested using a statistic proposed by Stock and Yogo (2003) and then only if this statistic indicates identification does the researcher proceed to the second step in which the aforementioned information criterion is used to select moments. The properties of this two-step procedure are contrasted with those of strategies based on either using all available moments or using the information criterion without the identification pre-test. The performances of these strategies are compared via an evaluation of the finite sample behavior of various methods for inference about the parameter vector. The inference methods considered are based on the Wald statistic, Anderson and Rubin's (1949) statistic, Kleibergen (2002) K statistic, and combinations thereof in which the choice is based on the outcome of the test for weak identification.*

Keywords Generalized method of moments; Inference; Moment selection; Weak identification.

JEL Classification C10; C12.

Received November 22, 2005; Accepted March 13, 2007

Address correspondence to Alastair R. Hall, Economics, School of Social Studies, University of Manchester, Manchester M13 9PL, UK; E-mail: alastair.hall@manchester.ac.uk

1. INTRODUCTION

Generalized Method of Moments (GMM) provides a method for obtaining consistent and asymptotically normally distributed estimators of the parameters of economic models based on a set of population moment conditions.¹ In most cases of interest, the researcher is actually faced with a candidate set from which to choose the moment conditions to be used in the estimation.

It was understood early on in the GMM literature that the asymptotic efficiency of the estimator is sensitive to the choice of moment condition; see Hansen (1982, 1985). Subsequently, it has been realized the quality of the asymptotic approximation to finite sample behavior can also be very sensitive to the choice of moment condition; for example, see Kocherlakota (1990).² These findings have stimulated interest in understanding the circumstances in which the asymptotic approximation may be poor. Following a seminal article by Nelson and Startz (1990) and important sequels by Staiger and Stock (1997) and Stock and Wright (2000), weak identification has emerged as an important explanation for why standard GMM asymptotic distribution theory provides a poor approximation in certain economic models of interest. This finding has stimulated the development of inference techniques that are valid irrespective of whether the parameters are identified or weakly identified parameters. Leading contributions in this area were made by Staiger and Stock (1997), Stock and Wright (2000), Kleibergen (2002), and Moreira (2003).

In a recent article, Hall et al. (2007) (HIJS hereafter) argue that the entropy of the limiting distribution of the GMM estimator provides a metric for the information about the parameters contained in a moment condition, and, as a consequence, they propose an information criterion based on this entropy that can be used to guide moment selection. This information criterion is denoted *RMSC*, which stands for relevant moment selection criterion, a terminology that is explained below. HIJS establish conditions for the consistency of the associated moment selection method in two settings of interest.³ One key aspect of these conditions relates to what is assumed about the identification of the parameter vector by the combinations of moments from the candidate set over which the selection is made. While it is standard in the moment selection literature to assume

¹See Hansen (1982).

²There have been numerous simulation studies that have made this point; see Hall (2005, Section 6.3) for a recent review.

³In this context, consistency is the property that a particular moment condition is selected with probability one in the limit as the sample size goes to infinity. The two settings are: nonlinear dynamic models estimated via GMM and linear static models estimated via instrumental variables. The difference between the analyses lies in the regularity conditions imposed; see discussion below.

that the parameter vector is identified by all the combinations of moment condition considered,⁴ it is desirable to allow for the possibility of weak identification. HIJS take a step in this direction in their analysis of *RMSC* in the static linear model estimated by Instrumental Variables (IV). They allow for the possibility that the parameter vector is only identified by some of the combinations of moments being considered and is weakly identified by the rest. In this case, they establish that their method selects a moment condition that identifies the parameter vector with probability one in the limit, and that the method is consistent provided certain other conditions hold.

In this article, we extend HIJS's analysis by considering the behavior of their moment selection method when the parameter vector is weakly identified by *all* the moment conditions being considered. It is shown that the selected moment condition is random and hence not consistent in any meaningful sense. As a result, we propose a two-step procedure for moment selection in which identification is first tested using a statistic proposed by Stock and Yogo (2003) and then only if this statistic indicates identification does the researcher proceed to the second step in which *RMSC* is used to select moments. The properties of this two-step procedure are contrasted with those of strategies based on either using all available moments or using the information criterion without the pre-test for weak identification. The inference methods considered are based on the Wald statistic, Anderson and Rubin's (1949) statistic, Kleibergen's (2002) K statistic, and combinations thereof in which the choice is based on the outcome of the test for weak identification.

The outline of the article is as follows. Section 2 describes various inference procedures based on the IV estimator, and discusses the impact of weak identification on their limiting behavior. Section 3 describes Stock and Stock and Yogo's (2003) framework for testing weak identification in this model. Section 4 describes *RMSC* and its relationship to the entropy of the limiting distribution of the IV estimator, and presents new results on the limiting behavior of *RMSC* when the parameter of interest is weakly identified by all the moments being considered. Section 5 reports results from a simulation study to investigate the properties of various strategies for instrument selection. Section 6 concludes.

2. IV ESTIMATION AND WEAK IDENTIFICATION

In this section, we briefly describe IV estimation in the linear regression model and discuss the impact of weak identification on standard IV-based inference procedures. This impact motivates the use of methods

⁴For example, see Donald and Newey (2001), Hall and Peixe (2003), and HIJS's analysis of *RMSC* in nonlinear dynamic models.

that are robust to weak identification, and so such methods are also reviewed as part of this discussion.

Consider the case in which the structural equation of interest takes the form

$$y_t = x_t' \theta_0 + u_t, \quad (1)$$

where y_t is a scalar, x_t is $p \times 1$ vector of (observable) variables, and θ_0 is an unknown parameter vector. Suppose that the reduced form for x_t is given by

$$x_t = \Pi z_t + e_t \quad (2)$$

where z_t is a $q \times 1$ vector of (observable) variables. It is assumed that $v_t = (x_t', u_t, z_t', e_t')'$ is independently and identically distributed. We further assume that the combined error vector $w_t = (u_t, e_t')'$ satisfies the following moment conditions:

$$E[w_t | z_t] = 0_{(q+1) \times 1} \quad (3)$$

$$\text{Var}[w_t | z_t] = \begin{bmatrix} \sigma_u^2 & \Sigma_{u,e} \\ \Sigma_{u,e}' & \Sigma_e \end{bmatrix} \quad (4)$$

where $0_{a \times b}$ denotes the $a \times b$ null matrix, σ_u^2 is a scalar and Σ_e is $q \times q$. For convenience below, we define $u_t(\theta) = y_t - x_t' \theta$; note that $u_t(\theta_0) = u_t$.

Equation (3) implies that θ_0 satisfies the population moment condition

$$E[z_t u_t(\theta_0)] = 0. \quad (5)$$

Under our assumptions, the GMM estimator based on (5) with the optimal weighting matrix is just the Two Stage Least Squares (2SLS) estimator,

$$\hat{\theta}_T = \{X'Z(Z'Z)^{-1}Z'X\}^{-1}X'Z(Z'Z)^{-1}Z'y, \quad (6)$$

where X is a $T \times p$ matrix with t th row x_t' , Z is a $T \times q$ matrix with t th row z_t' and y is $T \times 1$ vector with t th element y_t .⁵ Although we focus on 2SLS throughout the article, we note parenthetically that all the issues addressed, methods proposed, and limiting theory in this article apply equally to the Limited Information Maximum Likelihood estimator of θ_0 .

Within this framework, θ_0 is identified by (5) if

$$E[x_t z_t'] = M_{xz}, \quad \text{finite with } \text{rank}\{M_{xz}\} = p. \quad (7)$$

⁵Recall 2SLS is the GMM estimator based on $E[z_t u_t] = 0$ with weighting matrix $(T^{-1}Z'Z)^{-1}$.

If (7) and certain other regularity conditions hold,⁶ it can be shown

$$T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, \sigma_u^2 (M_{xz} M_{zz}^{-1} M'_{xz})^{-1}), \quad (8)$$

where $M_{zz} = E[z_t z'_t]$.

A wide variety of inference procedures have been developed for 2SLS estimators. For our purposes here, it suffices to describe just one. Suppose it is desired to test the point null hypothesis, $H_0 : \theta_0 = \bar{\theta}$. Based on (8), this hypothesis can be tested using the Wald statistic

$$W_T(\bar{\theta}) = \frac{(\hat{\theta}_T - \bar{\theta})' X' Z (Z' Z)^{-1} Z' X (\hat{\theta}_T - \bar{\theta})}{\hat{\sigma}_u^2}, \quad (9)$$

where $\hat{\sigma}_u^2 = T^{-1} \sum_{t=1}^T [u_t(\hat{\theta}_T)]^2$. Under H_0 , it can be shown that

$$W_T(\bar{\theta}) \xrightarrow{d} \chi_p^2 \quad (10)$$

and hence an asymptotically valid 5% significance level test of H_0 has the decision rule

$$\text{reject } H_0 \text{ when } W_T(\bar{\theta}) > c_p(0.95), \quad (11)$$

where $c_p(0.95)$ is the 95th percentile of the χ_p^2 distribution.

Now suppose that the parameter vector θ_0 is only weakly identified by (5). Following Staiger and Stock (1997), this situation is captured by the assumption that

$$E[x_t z'_t] = T^{-1/2} A, \quad (12)$$

where A is a finite matrix of constants. It follows from (12) that $\lim_{T \rightarrow \infty} E[x_t z'_t] = 0_{p \times q}$, and so the identification condition in (7) fails in the limit. This failure undermines the consistency of 2SLS and the use of standard inference procedures. Staiger and Stock (1997) show that under weak identification $\hat{\theta}_T - \theta_0$ converges to a nondegenerate distribution and $W_T(\bar{\theta})$ converges to a different distribution under H_0 than the standard result presented in (10). This would mean, of course, that the decision rule in (11) does not yield a 5% test if the parameters are weakly identified, and hence that the Wald test based on standard asymptotics is subject to a size distortion.

Given these findings, it is clearly desirable to base inference on statistics whose properties are robust to the potential presence of weak identification. Within the framework considered here, Staiger

⁶See *inter alia* Hall (2005, Section 2.3).

and Stock (1997) show that such inference can be performed using the statistic proposed by Anderson and Rubin (1949) (referred to as the AR statistic hereafter), that is,⁷

$$AR_T(\bar{\theta}) = \left\{ \frac{u(\bar{\theta})' Z(Z'Z)^{-1} Z' u(\bar{\theta})}{u(\bar{\theta})'(I_T - Z(Z'Z)^{-1} Z') u(\bar{\theta})} \right\} \left\{ \frac{T - q}{q} \right\}. \quad (13)$$

Staiger and Stock (1997) show that H_0 , $AR_T(\bar{\theta}) \xrightarrow{d} \chi_q^2/q$ irrespective of whether the parameter is identified or weakly identified. However, an unattractive feature of the AR statistic is that the degrees of freedom of its limiting distribution equal the number of instruments. Therefore, when there are many more instruments than parameters, then this approach to inference leads to tests with low power or equivalently very wide confidence intervals; see Kleibergen (2002). Therefore, while the advantage of the AR statistic over the Wald is clear if the parameters are weakly identified, it does not carry over to models in which the parameters are, in fact, identified. Kleibergen (2002) and Moreira (2003) propose statistics that have the advantages of the AR statistic but without this weakness. Here we focus on Kleibergen's (2002) K statistic which is given by

$$K_T(\bar{\theta}) = (T - q) \left\{ \frac{u(\bar{\theta})' M(\bar{\theta})^{-1} u(\bar{\theta})}{u(\bar{\theta})'(I_T - Z(Z'Z)^{-1} Z') u(\bar{\theta})} \right\}, \quad (14)$$

where

$$\begin{aligned} M(\theta) &= N(\theta)[N(\theta)'N(\theta)]^{-1}N(\theta)' \\ N(\theta) &= Z\tilde{\Pi}(\theta) \\ \tilde{\Pi}(\theta) &= (Z'Z)^{-1}Z' \left[y - u(\theta) \frac{s_{ue}(\theta)}{s_{uu}(\theta)} \right] \\ s_{uu}(\theta) &= (T - q)^{-1} u(\theta)'(I_T - Z(Z'Z)^{-1} Z') u(\theta) \\ s_{ue}(\theta) &= (T - q)^{-1} u(\theta)'(I_T - Z(Z'Z)^{-1} Z') X. \end{aligned}$$

Kleibergen (2002) shows that under $H_0 : \theta_0 = \bar{\theta}$, $K_T(\bar{\theta}) \xrightarrow{d} \chi_p^2$ irrespective of whether the parameters are identified or weakly identified. Notice that the degrees of freedom equal the number of parameters and hence the K statistic does not suffer from the aforementioned weakness of the AR statistic.

Kleibergen (2002) evaluates the performance of the Wald, AR, and K statistics via a simulation study. He reports that the AR and K statistics exhibit empirical size very close to the nominal size irrespective of

⁷This statistic is specialized here to the model in (1)–(2).

whether the parameters are identified or weakly identified and irrespective of the number of instruments. In contrast, he finds that the Wald statistic exhibits severe size distortions in two settings: the first is where the parameters are weakly identified, the second is where the parameters are identified and the degree of overidentification is large. He also reports a power comparison of the AR and K tests that illustrate the K test dominates the AR in models where there are more instruments than parameters irrespective of the quality of the identification.

This simulation evidence would appear to recommend the adoption of the K statistic as the basis for inference about the parameters in linear models estimated via IV. However, there is one aspect of Kleibergen's (2002) simulation design that merits further scrutiny before this recommendation is adopted. In the case where multiple instruments are used, all instruments beyond the first are redundant.⁸ HJIS report evidence that the inclusion of redundant instruments leads to a severe distortions in the empirical size of the Wald test in finite samples; further they find that these size distortions are ameliorated by the use of *RMSC* to select instruments.⁹

Given practitioners preference for inferences based on Wald tests, for example, through the ubiquitous reporting of standard errors and *t*-statistics, it is interesting to compare the properties of inference based on the K test with strategies that use *RMSC* to select instruments and then base inference on the Wald statistic. However, given the preceding discussion, it is clear that inferences should not be based on the Wald statistic if the parameters are weakly identified. Therefore, the Wald test can only be used in combination with some method for determining whether or not the parameters are identified.

Therefore in the next section, we outline Stock and Yogo's (2003) test for weak identification. This is followed by a discussion of *RMSC* and an evaluation of its properties if the parameter vector is weakly identified by all combinations of instruments considered. Collectively, Sections 2–4 highlight the key statistical results that motivate the various strategies for inference considered in the simulation study reported in Section 5.

3. TESTING FOR WEAK IDENTIFICATION

Early examinations of the use of tests for identification drew rather pessimistic conclusions about their efficacy; see Hall et al. (1996) and Zivot et al. (1998). However, this pessimism seems to be in part attributable to how the pre-test was performed and the assumed action if weak identification is diagnosed. Recent work by Stock and Yogo (2003) suggests

⁸See Kleibergen (2002, p. 1791).

⁹Also see Hall and Peixe (2003).

that this approach may be more fruitful than at first thought. So we now elaborate on Stock and Yogo (2003) test for weak identification, and then return to the issue of its use as a pre-test as part of an inference strategy that is robust to weak identification.

Stock and Yogo (2003) propose drawing inference about the quality of the identification using the following statistic originally proposed by Cragg and Donald (1993),¹⁰

$$g_{min} = \text{mineval}\{G_T\}, \quad (15)$$

where $\text{mineval}\{\cdot\}$ denotes the minimum eigenvalue of the matrix within the curly brackets,

$$G_T = \widehat{\Sigma}_e^{-1/2} X' Z (Z' Z)^{-1} Z' X \widehat{\Sigma}_e^{-1/2'} / q \quad (16)$$

and $\widehat{\Sigma}_e = (T - q)^{-1} X' (I_T - Z(Z' Z)^{-1} Z') X$. Originally, this statistic was designed to test under-identification, but Stock and Yogo (2003) propose using it to test the null hypothesis of weak identification in the sense of (12). Under the latter null, Stock and Yogo (2003) show that

$$G_T \xrightarrow{d} v_1 / q \quad (17)$$

where $v_1 \sim W_p(q, I_p, \lambda' \lambda)$, a noncentral Wishart distribution, with $\lambda = M_{zz}^{1/2} A \Sigma_e^{-1}$.¹¹ A key contribution of Stock and Yogo's (2003) work is the recognition that the definition of weak identification varies by estimator and also by the type of inference desired, and hence that the critical value for a test of weak identification should also depend on these factors. To illustrate, suppose inference is to be performed about θ_0 using the decision rule in (11). Stock and Yogo (2003) show that under weak instrument asymptotics, the rejection frequency of the Wald statistic is determined by p , q , $\rho = \Sigma_e^{-1/2} \Sigma'_{u,e} \sigma_u^{-1}$, and $\text{mineval}\{\Lambda\}$, where $\Lambda = \lambda' \lambda / q$. They propose that, in this context, the null and alternative hypothesis of weak identification should reflect the maximal size distortion in the size of the Wald test. So, for instance, for the decision rule in (11), the null and alternative are, respectively,¹²

$$H_0^{wi} : \text{not } H_A^{wi}$$

H_A^{wi} : the size distortion of the decision rule in (11) is not greater than $\alpha\%$ for any choice of ρ or $\text{mineval}(\Lambda)$,

¹⁰This statistic is specialized here to the model in (1)–(2).

¹¹See Rao (1973, p. 534).

¹²Note that p and q are fixed.

where α is some level set by the researcher. For given values of p , q , and α , Stock and Yogo (2003) calculate numerically the value of the minimum eigenvalue of Λ such that H_A^{wi} holds for all ρ ; let this value of the minimum eigenvalue of Λ be denoted $\text{mineval}\{\Lambda(\alpha, p, q)\}$. Stock and Yogo (2003) use $\text{mineval}\{\Lambda(\alpha, p, q)\}$ to tabulate critical points for g_T based on the distribution in (17) that can be used to test H_0^{wi} against H_A^{wi} .

In Section 5, we explore the performance of various inference strategies that use Stock and Yogo (2003) statistic to determine which statistic inference is based upon. Another aspect of these inference strategies is the use of *RMSC* to select the instruments used. However, before presenting these inference strategies, it is necessary to consider the limiting behavior of *RMSC* in the presence of weak identification; this is the topic of the next section.

4. ENTROPY BASED MOMENT SELECTION

In most cases of interest, the researcher is actually faced with a candidate set of instruments from which to choose in order to construct the moment condition used in the estimation. The inference strategies described in the previous section involve using all available instruments. Assuming the parameter vector to be identified, there is no cost asymptotically to the use of all available instruments but there is evidence that there may be finite sample gains to using only a subset of the instruments in certain settings. This evidence has stimulated interest in developing methods for instrument selection. In this section, we focus on one such method, the *RMSC* proposed by HIJS.¹³ We survey results presented by HIJS on the behavior of *RMSC* if the parameter vector is identified, and also present a new result that shows the limiting behavior of *RMSC* if the parameter vector is weakly identified by all choices of instruments considered. Together these results yield useful insights into the behavior of *RMSC* that inform the design of the inference strategies evaluated in Section 5.

RMSC can be used to select moments in the more general setting of GMM estimation, but we specialize our discussion of the method here to its application to the problem of instrument selection in the linear model described in the previous section. To this end, we now assume that the candidate set of moment conditions is given by (5). In this case, the only difference between various choices of moments lies in the chosen instrument vector and so without loss of generality, we frame the problem in terms of instrument selection. Following Andrews (1999), we use a $q \times 1$ selection vector c to denote which elements of the instrument vector z_i are

¹³Other methods for instrument selection are proposed by Donald and Newey (2001) and Hall and Peixe (2003).

included in a particular moment condition: if $c_j = 1$ then the j th element of z_t is included; if $c_j = 0$ implies then j th element of z_t is excluded. The case in which all instruments are used is denoted by $c = \iota_q$, where ι_q is $q \times 1$ vector of ones. The moments associated with c are written as

$$E[z_t(c)u_t(\theta_0)] = 0, \quad (18)$$

where

$$z_t(c) = S(c)z_t$$

and $S(c)$ is a selection matrix that picks out the elements of z_t indicated by c . Note that $|c| = c'c$ equals the number of elements in $z_t(c)$. The set of all possible selection vectors is denoted by $\mathcal{C}$, that is,

$$\mathcal{C} = \{c \in \mathbb{R}^q; c_j = 0, 1, \text{ for } j = 1, 2, \dots, q, \text{ and } c = (c_1, \dots, c_q)', |c| \geq p\}.$$

For brevity, statistics of interest are now indexed by c and so, for example, $\hat{\theta}_T(c)$ denotes the 2SLS estimator based on (18).

The design of RMSC is motivated by the entropy of the standard limiting distribution of the IV estimator. Given (8), the entropy of the limiting distribution of $\hat{\theta}_T(c)$ is¹⁴

$$\text{ent}_\theta(c) = 0.5p[1 + \ln(2\pi)] - 0.5 \ln[|M_{xx}(c)M_{zz}(c)^{-1}M_{xz}(c)'|], \quad (19)$$

where $M_{xx}(c) = M_{xx}S(c)'$ and $M_{zz}(c) = S(c)M_{zz}S(c)'$. HIJS argue that $\text{ent}_\theta(c)$ provides a metric for the information about θ_0 in the moment condition (18). Specifically, they show that $\text{ent}_\theta(c)$ is minimized by the asymptotically efficient choice of instruments and is infinite in the presence of weak identification. In view of these properties, HIJS argue that this entropy provides a potentially useful basis for instrument (moment) selection. Accordingly, they propose the following rule for instrument selection: select $z_t(\hat{c}_T)$, where

$$\hat{c}_T = \underset{c \in \mathcal{C}}{\text{argmin}} RMSC(c), \quad (20)$$

where

$$RMSC(c) = \ln[|\hat{V}_{\theta,T}(c)|] + \kappa(|c|, T), \quad (21)$$

$$\hat{V}_{\theta,T}(c) = \hat{\sigma}_u^2(c) \left[\frac{1}{T} \sum_{t=1}^T x_t z_t(c)' \left(\frac{1}{T} \sum_{t=1}^T z_t(c) z_t(c)' \right)^{-1} \frac{1}{T} \sum_{t=1}^T z_t(c) x_t' \right]^{-1}, \quad (22)$$

¹⁴The entropy is defined to be the expectation of the log of the probability density function of the distribution. Ahmed and Gokhale (1989) derive the entropy of the normal distribution.

$$\hat{\sigma}_u^2(c) = (1/T) \sum_{t=1}^T [u_t(\hat{\theta}_T(c))]^2, \quad (23)$$

and $\kappa(|c|, T)$ is a deterministic penalty term such as¹⁵

$$\kappa(|c|, T) = (|c| - p) \ln(T^{1/2})/T^{1/2} \quad (24)$$

As can be seen, the entropy of the normal distribution is a function of the variance of the distribution and so instrument selection is driven by its impact on the limiting variance of the IV estimator. The penalty term plays an important role because it enables the criterion to distinguish between choices of instrument that yield the same asymptotic variance. The use of RMSC, therefore, is designed to select from the candidate set the instrument vector of minimal length that achieves the minimum variance possible (given this candidate set).¹⁶

In their analysis of *RMSC*, HIJS consider the case where the candidate set consists of three types of moment conditions: (i) those that weakly identify θ_0 ; (ii) those that identify a subset of θ_0 and weakly identify the remaining elements; (iii) and those that identify θ_0 . Below, we will establish the limiting behavior of *RMSC* when θ_0 is weakly identified by all the moment conditions in the candidate set. However, for purposes of comparison, it is useful to first reproduce HIJS's consistency result. For brevity, we specialize this consistency result here to the case when the moments are only of the types given in (i) and (iii) above. To do so, we must first introduce some additional notation. Let the reduced form for x_t in terms of $z_t(c)$ be

$$x_t = \Pi(c)z_t(c) + e_t(c) \quad (25)$$

It is assumed that (25) evaluated at $c = \iota_q$ yields (2), that is, $\Pi(\iota_q) = \Pi$, $z_t(\iota_q) = z_t$ and $e_t(\iota_q) = e_t$. The appropriate subsets of $\mathcal{C}$ are defined as follows: the subset associated with weak identification is

$$\mathcal{C}_{wi} = \{c \text{ such that}$$

$$\Pi(c) = T^{-1/2}A(c), \text{ for some matrices of constants } A(c); c \in \mathcal{C}\} \quad (26)$$

¹⁵HIJS place certain generic restrictions on the form of the deterministic penalty term in order to establish the consistency of *RMSC*; see below.

¹⁶This approach can be viewed as minimizing a penalized version of the entropy of the limiting distribution of the IV estimator as in HIJS or it can be viewed purely in terms of minimizing a penalized measure of the variance. Either way, further theoretical work is needed to justify the use of BIC type penalty term. To date, the choice of the BIC penalty has only been justified through simulation evidence.

and that associated with identification is

$$\mathcal{C}_{id} = \{c \text{ such that } \Pi(c) \text{ is a matrix of constants with } \text{rank}\{\Pi(c)\} = p; c \in \mathcal{C}\}. \quad (27)$$

As mentioned above, it suffices here to consider the case in which moment selection is over $\mathcal{C}_{wi} \cup \mathcal{C}_{id}$, and so we define

$$\tilde{c}_T = \underset{c \in \mathcal{C}_{wi} \cup \mathcal{C}_{id}}{\operatorname{argmin}} RMSC(c). \quad (28)$$

To establish the consistency of the moment selection method, certain regularity conditions need to be imposed. These conditions place restrictions on both the limiting behavior of certain sample moments and the deterministic penalty term, and also involve an identification condition that serves to tie down the probability limit of $\tilde{c}_T$. These first of these two conditions are as follows.

Assumption 1. (i) $\text{rank}\{M_{zz}\} = q$; (ii) $T^{-1}Z'Z \xrightarrow{p} M_{z,z}$; (iii) $T^{-1/2}Z'u \xrightarrow{d} N(0, \sigma_u^2 M_{z,z})$; (iv) $T^{-1/2} \sum_{t=1}^T z_t \otimes e_t \xrightarrow{d} N(0, \Sigma_1)$; (v) $T^{-1}u'u \xrightarrow{p} \sigma_u^2$; (vi) $T^{-1/2} \sum_{t=1}^T (e_t u_t - \Sigma_{e,u}) \xrightarrow{d} N(0, \Sigma_2)$.

Assumption 2. For any $\tilde{c}, \bar{c} \in C$ such that $|\bar{c}| > |\tilde{c}|$, $T^{1/2}[\kappa(|\bar{c}|, T) - \kappa(|\tilde{c}|, T)] \rightarrow +\infty$ as $T \rightarrow \infty$, and $\kappa(|c|, T) = o(1)$ for every $c \in \mathcal{C}$.

To present the identification condition, we must introduce some additional notation. We first define $V_\theta(c) = \sigma_u^2 [M_{xz}(c)M_{zz}(c)^{-1}M_{xz}(c)']^{-1}$. Noting that within the framework here, the minimum value for $V_\theta(c)$ is $V(l_q)$, we define the set of selection vectors that attain this minimum to be $\mathcal{C}_{eff}$, that is,

$$\mathcal{C}_{eff} = \{c; V_\theta(l_q) = V_\theta(c), c \in \mathcal{C}\}.$$

We need also to refer to the subset of $\mathcal{C}_{eff}$ that contains the selection vectors of minimum length, that is,

$$\mathcal{C}_{min} = \{c; c \in \mathcal{C}_{eff}, |c| \leq |\bar{c}| \text{ for all } \bar{c} \in \mathcal{C}_{eff}\}.$$

Notice that $\mathcal{C}_{min}$ contains the set of selection vectors that are both asymptotically efficient relative to the candidate set and also include no redundant moment conditions. This combination of efficiency and nonredundancy is termed *relevance* by HIJS and is the source of the name of their information criterion. The identification condition is as follows.

Assumption 3. There is a selection vector c_r such that $c_r \in \mathcal{C}_{id}$ and $\mathcal{C}_{min} = \{c_r\}$.

The following lemma gives the limiting behavior of $\tilde{c}_T$ when subsets of the candidate set provide only weak identification.

Lemma 1 (HIJS, Theorem 6). *Let the data be generated by (1) and (2) and Assumptions 1–3 hold, then $\tilde{c}_T \xrightarrow{p} c_r$.*

This consistency result depends crucially on the assumption that θ_0 is identified by at least one combination of moment conditions considered. We now turn to the case in where this assumption fails.

Now consider the case in which θ_0 is weakly identified by the candidate set of moments and so $\mathcal{C} = \mathcal{C}_{wi}$. For this part of the analysis, it is convenient to refine the conditions imposed on the variables driving the data generation process as follows.

Assumption 4.

$$(a) \quad \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} u_t \\ e_t(c) \end{pmatrix} \begin{pmatrix} u_t \\ e_t(c) \end{pmatrix}' \xrightarrow{p} \begin{bmatrix} \sigma_u^2 & \Sigma_{u,e}(c) \\ \Sigma_{e,u}(c) & \Sigma_e(c) \end{bmatrix},$$

where $\Sigma_{u,e}(c) = \Sigma_{e,u}(c)'$.

$$(b) \quad \frac{1}{T^{1/2}} \sum_{t=1}^T \begin{pmatrix} z_t(c) \otimes u_t \\ z_t(c) \otimes e_t(c) \end{pmatrix} \xrightarrow{d} N \left(0, \begin{bmatrix} \sigma_u^2 M_{z,z}(c) & \Sigma_{zu,ez}(c) \\ \Sigma_{zu,ez}(c)' & \Sigma_{ze,ez}(c) \end{bmatrix} \right)$$

$$\equiv \begin{bmatrix} \zeta_u(c) \\ \zeta_e(c) \end{bmatrix}, \quad (29)$$

where the asymptotic covariance matrix is positive definite for all $c \in \mathcal{C}$.

The following theorem characterizes the limiting behavior of RMSC when the parameter vector is weakly identified by the candidate set.

Theorem 1. *Let $\mathcal{C} = \mathcal{C}_{wi}$ and Assumption 4 hold, then we have*

$$\begin{aligned} \hat{c}_T &\equiv \underset{c \in \mathcal{C}}{\operatorname{argmin}} RMSC(c) \\ &\Rightarrow \underset{c \in \mathcal{C}}{\operatorname{argmin}} \{ \ln(\sigma_u^2 - 2\Sigma_{e,u}\zeta(c) + \zeta(c)'\Sigma_e\zeta(c)) \\ &\quad - \ln |(M_{z,z}(c)A'(c) + \zeta_e(c))'M_{z,z}(c)^{-1}(M_{z,z}(c)A'(c) + \zeta_e(c))| \}, \end{aligned} \quad (30)$$

where

$$\begin{aligned} \zeta(c) &= [(M_{z,z}(c)A'(c) + \zeta_e(c))'M_{z,z}(c)^{-1}(M_{z,z}(c)A'(c) + \zeta_e(c))]^{-1} \\ &\quad \times (M_{z,z}(c)A'(c) + \zeta_e(c))'M_{z,z}(c)^{-1}\zeta_u(c). \end{aligned}$$

Proof. Under Assumption 4, we have

$$\begin{aligned} \ln \left| \sum_{t=1}^T x_t z_t(c)' \left(\sum_{t=1}^T z_t(c) z_t(c)' \right)^{-1} \sum_{t=1}^T z_t(c) x_t' \right| - p \ln(T) \\ \xrightarrow{d} (M_{z,z}(c) A'(c) + \zeta_e(c))' M_{z,z}(c)^{-1} (\Sigma_{zz}(c) A'(c) + \zeta_e(c)), \end{aligned} \quad (31)$$

where the convergence is uniform in $c \in \mathcal{C}$. Since

$$\begin{aligned} \hat{\theta}(c) - \theta &= [(M_{z,z}(c) A'(c) + \zeta_e(c))' M_{z,z}(c)^{-1} (M_{z,z}(c) A'(c) + \zeta_e(c))]^{-1} \\ &\quad \times (M_{z,z}(c) A'(c) + \zeta_e(c))' M_{z,z}(c)^{-1} \zeta_u \\ &\equiv \xi(c), \end{aligned}$$

it follows that

$$\hat{\sigma}_u^2(c) \xrightarrow{d} \sigma_0^2 + 2 \Sigma'_{e,u} \xi(c) + \xi(c)' \Sigma_e \xi(c) \quad (32)$$

Thus, the desired result follows from (31) and (32). $\square$

Theorem 31 indicates that the selected moment condition is random when the parameter vector is weakly identified by the candidate set. Clearly, this is in marked contrast to the case where identification is achieved by at least one set of moment conditions considered for which $\hat{c}_T$ converges in probability to the constant c_r . This difference has important consequences for subsequent inferences about θ_0 . If $\hat{c}_T$ converges in probability to the constant c_r , then $T^{1/2}[\hat{\theta}_T(\hat{c}_T) - \theta_0]$ has the same limiting distribution as $T^{1/2}[\hat{\theta}_T(c_r) - \theta_0]$, that is (8) evaluated at $z_t = z_t(c_r)$.¹⁷ In other words, if θ_0 is identified by at least one set of moment conditions,¹⁸ then inference can proceed using standard methods. In such cases, data based moment selection can be beneficial for the types of reason discussed above. However, if $\hat{c}_T$ converges to a random variable, then this dependence on the data has the potential to contaminate subsequent inference procedures. For example, if θ_0 is weakly identified by the candidate set, then there is no guarantee that the AR statistic evaluated at $z_t = z_t(\hat{c}_T)$ converges to a χ_q^2/q distribution even if the null hypothesis is correct. In such cases, data-based moment selection may not be beneficial because its use may undermine subsequent inference procedures. These considerations suggest that, unless the quality of the identification is not in question, it may be advisable to employ a pre-test for weak identification when selecting moments via RMSC. In the next section, we explore the properties of this strategy along with various alternatives.

¹⁷See Pötscher (1991) or the discussion in Hall (2005, Section 7.1).

¹⁸The other regularity conditions in Lemma 1 must also be satisfied, of course.

5. EVALUATION OF VARIOUS INFERENCE STRATEGIES

In this section, we evaluate the impact of various instrument selection strategies on the finite sample properties of inference procedures based on 2SLS. These inference strategies involve the use of the Wald, AR, and K statistics both with and without the use of *RMSC* and the Stock–Yogo pre-test for weak identification. The design allows for different degrees of identification and thus provides insights into the performance of these methods in the variety of settings that are encountered in practice.

Data are generated from the following model:

$$y_t = \theta_0 x_t + u_t, \quad (33)$$

$$x_t = \pi' z_t + e_t, \quad (34)$$

where y_t and x_t are scalars; z_t is a $q_{\max} \times 1$ vector for $q_{\max} = 8$; $[u_t, e_t, z_t']' \sim NID(0_{(q_{\max}+2) \times 1}, \Sigma)$ with Σ equal to a matrix whose diagonal elements are all one and whose only nonzero off diagonal elements are the (1, 2) and (2, 1) entries which are both σ_{ue} for $\Sigma_{ue} = \sigma_{ue} \in \{0.1, 0.5, 0.9\}$. The parameter vector π is chosen to fix the population concentration parameter, $\lambda' \lambda / q$, via

$$\frac{\lambda' \lambda}{q} = \frac{\pi' E[Z' Z] \pi}{q_{\max}} = \frac{T}{q_{\max}} \sum_{i=1}^{q_{\max}} \pi_i^2 \quad (35)$$

Since π is a vector, (35) does not yield a unique solution. We consider three solutions:

1. **ID _{π} = I**: In this specification, only the first instrument is relevant and so only the first element of π is nonzero, that is,

$$\pi = (\gamma_1, 0'_{(q_{\max}-1) \times 1})'. \quad (36)$$

Note that in this case, $c_r = (1, 0'_{(q_{\max}-1) \times 1})'$;

2. **ID _{π} = II**: In this specification, all the instruments are equally important and so all the elements of π are equal, that is,

$$\pi = (\gamma_2, \gamma_2, \dots, \gamma_2)'. \quad (37)$$

In this case, $c_r = \iota_{q_{\max} \times 1}$.

3. **ID _{π} = III**: In this specification, the instruments are of diminishing marginal importance with $\pi_1 \geq \pi_2 \geq \dots \geq \pi_{q_{\max}}$. Following Donald and Newey (2001), the i th element of π is determined via

$$\pi_i = \gamma_3 \left(1 - \frac{i}{q_{\max} + 1} \right)^4 \quad \text{for } i = 1, \dots, q_{\max}. \quad (38)$$

In this case, $c_r = \iota_{q_{\max} \times 1}$.

The values of constants γ_1 , γ_2 , and γ_3 are chosen so that (35) holds for the set $\lambda'/\lambda/q \in \{1, 10, 25, 50\}$.¹⁹ The simulation is based on 10,000 Monte Carlo draws for sample size $T = 100$.

In strategies involving *RMSC*, this criterion is calculated using the formula in (21)–(23) with the *BIC* type penalty in (24). As noted above, *RMSC* can be employed by performing the minimization over all possible instruments or some subset thereof. We consider two approaches. First, we consider the sequential approach in which $RMSC(c)$ is minimized over choices of c : $c = (1'_{q \times 1}, 0'_{(q_{\max} - q) \times 1})'$, $q = 1, 2, \dots, q_{\max}$, where $1_{n \times 1}$ is a $n \times 1$ vector of ones and $0_{n \times 1}$ is the $n \times 1$ null vector. This represents the case where the researcher may have some a priori information about the relative importance of the instruments; notice that this information is correct in models $\mathbf{ID}_\pi = \mathbf{I}$ and $\mathbf{ID}_\pi = \mathbf{III}$. Second, we consider the case in which the minimization is over all possible choices of instrument.²⁰ For expositional brevity, we only explicitly report in the tables the results for the *RMSC*-based strategies for the case in which the instruments are selected sequentially. The results for the case in which the selection is over all possible choices are summarized in the text and the tables upon which the discussion is based can be obtained from the authors upon request.

Our evaluation of the inference procedures is based on a number aspects of the finite sample behavior of the 2SLS estimator. We report: (i) the median bias; (ii) the coverage probability of the 95% confidence interval based on the Wald, AR, and K statistics; (iii) the median width of the aforementioned intervals based on the Wald and AR statistics;²¹ (iv) the power of the Wald, AR, and K tests of $H_0: \theta_0 = 0$. For (i)–(iii), the parameter in the structural equation is set to zero, that is, $\theta_0 = 0$; for (iv) the true value of θ_0 is set to 0.05, 0.1, and 0.5.

Within our design, the AR statistic has an exact F distribution with (8, 92) degrees of freedom. Therefore, at the suggestion of a referee we use percentiles from this distribution in our calculations so that any distortions from the nominal coverage probabilities can be attributed purely to the other causes of interest such as the use of the Stock–Yogo test for weak identification or the use of *RMSC* to select instruments.

For expositional convenience, the various inference strategies are discussed in four natural groupings. The results are discussed collectively at the end of the section.

¹⁹The concentration parameter is widely accepted as a unit less measure of the “quality” of the identification; the lower the value, the weaker the identification.

²⁰Since we must have $|c| \geq p$, the total number of choices is 255 because $p = 1$.

²¹Zivot et al. (1998) provide closed form expressions for the interval based on the AR statistic for the case here in which $p = 1$. The median width of the interval based on the K statistic is not reported because its calculation requires solving a fourth order polynomial inequality in θ to which there is no closed form solution. Furthermore, since the interval can be of infinite length, it is infeasible to find the interval numerically in Monte Carlo experiments.

(a) *All instruments are used:*

- *Strategy 1:* 2SLS is calculated using all available instruments and inference is based on the Wald statistic in (9) and the decision rule in (11).
- *Strategy 2:* Inference is based on the AR statistic in (13) with all instruments.
- *Strategy 3:* Inference is based on the K statistic in (14) with all instruments.

Table 1 reports the median bias of the 2SLS estimator along with the coverage probabilities and median widths of 95% confidence intervals for θ_0 . It can be seen that (as expected) the AR and K tests maintain the correct coverage probabilities irrespective of the quality of the identification. In contrast, the Wald based interval can exhibit severe coverage probability distortions if the parameter is weakly identified: for low levels of endogeneity (i.e., $\sigma_{ue} = 0.1$), the coverage is accurate for values of the concentration parameter above 10, but for high levels of endogeneity (i.e., $\sigma_{ue} = 0.9$), the coverage is only accurate once the concentration parameter is as high as 50. A comparison of the median widths of the intervals shows that the Wald intervals are far narrower than those based on the AR, and so it is clearly advantageous to base inference on the Wald statistic rather than the AR statistic provided, of course, that the coverage probability of the former is accurate.

From Table 2, it can be seen that the K test is more powerful than the AR test (confirming Kleibergen's 2002 findings). It can also be seen that for models with the concentration parameter equal to 50—the case in which the confidence interval based on the Wald statistic has approximately the correct coverage probability—the Wald and K tests have similar power.

(b) *RMSC is used to select the instruments:*

- *Strategy 4:* RMSC is used to select the instrument vector from the candidate set and then inference is based on the Wald statistic in (9) and the decision rule in (11).
- *Strategy 5:* RMSC is used to select the instrument vector from the candidate set and then inference is based on the AR statistic in (13).
- *Strategy 6:* Inference is based on the K statistic in (14) with instruments selected using *RMSC*.

Table 3 reports the median bias of the post-selection 2SLS estimator along with the coverage probabilities and median widths for the 95% confidence intervals based on these three strategies. Table 4 reports the corresponding power calculations. Note that these intervals are based

TABLE 1 All 8 instruments are used and no pretest is conducted (Strategies 1–3) – median bias, coverage probabilities and median width of 95% confidence intervals

$\frac{\lambda'\lambda}{q}$	σ_{ue}	ID_{π}^a	Med bias	Wald	Width	AR	Width	K
1	0.1	I	0.048	0.967	1.023	0.950	∞	0.941
1	0.1	II	0.054	0.966	1.022	0.953	∞	0.946
1	0.1	III	0.051	0.968	1.021	0.950	∞	0.945
1	0.5	I	0.245	0.789	0.926	0.949	∞	0.945
1	0.5	II	0.241	0.789	0.921	0.952	∞	0.946
1	0.5	III	0.245	0.786	0.926	0.949	∞	0.945
1	0.9	I	0.436	0.336	0.649	0.949	∞	0.942
1	0.9	II	0.435	0.342	0.648	0.952	∞	0.942
1	0.9	III	0.435	0.338	0.649	0.949	∞	0.945
10	0.1	I	0.007	0.950	0.419	0.951	0.742	0.949
10	0.1	II	0.009	0.951	0.419	0.955	0.744	0.949
10	0.1	III	0.009	0.951	0.420	0.950	0.744	0.946
10	0.5	I	0.041	0.919	0.411	0.952	0.748	0.947
10	0.5	II	0.042	0.921	0.411	0.951	0.747	0.946
10	0.5	III	0.041	0.920	0.411	0.951	0.746	0.950
10	0.9	I	0.074	0.836	0.391	0.948	0.769	0.947
10	0.9	II	0.074	0.851	0.389	0.950	0.762	0.949
10	0.9	III	0.073	0.848	0.392	0.953	0.767	0.948
25	0.1	I	0.003	0.949	0.271	0.949	0.450	0.947
25	0.1	II	0.004	0.946	0.271	0.948	0.449	0.945
25	0.1	III	0.003	0.944	0.271	0.949	0.451	0.944
25	0.5	I	0.017	0.934	0.270	0.953	0.455	0.943
25	0.5	II	0.017	0.936	0.269	0.950	0.453	0.946
25	0.5	III	0.016	0.932	0.270	0.949	0.454	0.942
25	0.9	I	0.032	0.906	0.263	0.953	0.455	0.949
25	0.9	II	0.031	0.904	0.264	0.944	0.453	0.948
25	0.9	III	0.031	0.905	0.263	0.951	0.454	0.950
50	0.1	I	0.001	0.944	0.194	0.951	0.316	0.945
50	0.1	II	0.001	0.947	0.194	0.949	0.315	0.948
50	0.1	III	0.002	0.948	0.193	0.948	0.314	0.948
50	0.5	I	0.008	0.940	0.192	0.948	0.316	0.947
50	0.5	II	0.009	0.940	0.192	0.944	0.315	0.944
50	0.5	III	0.009	0.938	0.193	0.947	0.314	0.945
50	0.9	I	0.016	0.922	0.191	0.948	0.318	0.941
50	0.9	II	0.017	0.924	0.190	0.948	0.314	0.950
50	0.9	III	0.014	0.925	0.190	0.951	0.318	0.946

^a ID_{π} denotes the specification how the first stage parameter vector is constructed. ‘I’ means only the first instrument is relevant, ‘II’ means all the instruments are equally important, and ‘III’ means the instruments are in order of their importance.

on the distributional results that are only valid for a fixed choice of instrument. Table 3 reveals that the use of the moment selection technique only has a slight distorting effect on the coverage probabilities of the AR and K statistic-based intervals. With the regard to the Wald statistic, an interesting pattern emerges. The use of *RMSC* is efficacious for models $ID_{\pi} = \mathbf{I}$ and $ID_{\pi} = \mathbf{III}$: the coverage probabilities are now accurate at concentration parameter values of 10 and 25, respectively, for the model

TABLE 2 All 8 instruments are used and no pretest is conducted (Strategies 1–3) – power

$\frac{\lambda'\lambda}{q}$	σ_{ue}	ID_{π}^a	Wald			AR			K		
			0.05	0.1	0.5	0.05	0.1	0.5	0.05	0.1	0.5
1	0.1	I	0.05	0.08	0.56	0.05	0.05	0.10	0.06	0.06	0.15
1	0.1	II	0.05	0.07	0.57	0.05	0.05	0.10	0.06	0.06	0.14
1	0.1	III	0.05	0.08	0.57	0.05	0.05	0.10	0.05	0.06	0.15
1	0.5	I	0.28	0.35	0.80	0.05	0.05	0.09	0.06	0.06	0.14
1	0.5	II	0.28	0.35	0.81	0.05	0.05	0.09	0.05	0.06	0.14
1	0.5	III	0.28	0.34	0.80	0.05	0.05	0.09	0.06	0.06	0.13
1	0.9	I	0.72	0.77	0.95	0.05	0.05	0.08	0.06	0.06	0.15
1	0.9	II	0.71	0.76	0.94	0.05	0.05	0.09	0.06	0.06	0.16
1	0.9	III	0.72	0.77	0.95	0.05	0.05	0.08	0.06	0.06	0.15
10	0.1	I	0.09	0.18	0.99	0.06	0.08	0.74	0.07	0.14	0.95
10	0.1	II	0.09	0.19	0.99	0.05	0.08	0.75	0.08	0.14	0.95
10	0.1	III	0.09	0.18	0.99	0.06	0.08	0.75	0.07	0.14	0.94
10	0.5	I	0.17	0.30	0.98	0.05	0.07	0.61	0.07	0.13	0.89
10	0.5	II	0.17	0.30	0.99	0.05	0.07	0.61	0.07	0.13	0.89
10	0.5	III	0.17	0.30	0.99	0.05	0.07	0.61	0.07	0.13	0.89
10	0.9	I	0.29	0.44	0.98	0.06	0.07	0.51	0.08	0.14	0.84
10	0.9	II	0.28	0.44	0.99	0.05	0.07	0.51	0.07	0.13	0.85
10	0.9	III	0.28	0.43	0.98	0.05	0.07	0.50	0.07	0.13	0.85
25	0.1	I	0.13	0.33	1.00	0.07	0.12	0.99	0.11	0.28	1.00
25	0.1	II	0.13	0.33	1.00	0.07	0.12	0.99	0.11	0.29	1.00
25	0.1	III	0.13	0.33	1.00	0.07	0.12	0.99	0.11	0.28	1.00
25	0.5	I	0.18	0.41	1.00	0.06	0.11	0.97	0.11	0.26	1.00
25	0.5	II	0.18	0.41	1.00	0.07	0.12	0.97	0.10	0.26	1.00
25	0.5	III	0.18	0.41	1.00	0.06	0.11	0.96	0.11	0.26	1.00
25	0.9	I	0.26	0.50	1.00	0.06	0.11	0.93	0.10	0.25	1.00
25	0.9	II	0.27	0.50	1.00	0.07	0.12	0.93	0.11	0.26	1.00
25	0.9	III	0.26	0.50	1.00	0.06	0.11	0.93	0.11	0.26	1.00
50	0.1	I	0.19	0.54	1.00	0.08	0.21	1.00	0.17	0.49	1.00
50	0.1	II	0.19	0.54	1.00	0.09	0.21	1.00	0.18	0.49	1.00
50	0.1	III	0.19	0.54	1.00	0.09	0.22	1.00	0.17	0.50	1.00
50	0.5	I	0.24	0.58	1.00	0.08	0.20	1.00	0.17	0.47	1.00
50	0.5	II	0.24	0.59	1.00	0.09	0.20	1.00	0.17	0.48	1.00
50	0.5	III	0.25	0.59	1.00	0.09	0.20	1.00	0.17	0.47	1.00
50	0.9	I	0.30	0.63	1.00	0.09	0.19	1.00	0.17	0.45	1.00
50	0.9	II	0.31	0.64	1.00	0.09	0.20	1.00	0.16	0.46	1.00
50	0.9	III	0.29	0.63	1.00	0.08	0.18	1.00	0.16	0.44	1.00

^aIn each case, the power is of a 5% significance level test of $H_0: \theta_0 = 0$ when the true value of this parameter is given at the head of each column.

^bFor other definitions, see the notes to Table 1.

with high endogeneity ($\sigma_{ue} = 0.9$) whereas this is only the case for a concentration parameter value of 50 in the case where all instruments are used; however, its use increases the median interval width compared to the case where all instruments are used. In contrast, the use of *RMSC* has a slightly adverse effect on the coverage probabilities for model $ID_{\pi} = \text{II}$. If selection is over all possible choices of instrument, then a similar pattern of results holds for models $ID_{\pi} = \text{I}$ and $ID_{\pi} = \text{III}$. However, the adverse

TABLE 3 Instruments are selected by RMSC and no pre-test is conducted (Strategies 4–6) – median bias, coverage probabilities and median width of 95% confidence intervals

$\frac{\lambda' \lambda}{q}$	σ_{ue}	ID_{π}^a	Med bias	Wald	Width	AR	Width	K
Sequential approach								
1	0.1	I	0.028	0.984	1.293	0.956	2.093	0.961
1	0.1	II	0.059	0.981	1.242	0.956	10.993	0.967
1	0.1	III	0.036	0.984	1.263	0.955	2.528	0.962
1	0.5	I	0.148	0.914	1.169	0.937	2.049	0.957
1	0.5	II	0.283	0.827	1.119	0.939	11.714	0.947
1	0.5	III	0.184	0.901	1.150	0.941	2.517	0.953
1	0.9	I	0.325	0.592	0.815	0.888	2.218	0.943
1	0.9	II	0.504	0.331	0.761	0.900	24.263	0.918
1	0.9	III	0.347	0.563	0.807	0.895	2.595	0.934
10	0.1	I	−0.000	0.953	0.438	0.952	0.457	0.949
10	0.1	II	0.016	0.966	0.539	0.958	0.847	0.961
10	0.1	III	0.007	0.956	0.466	0.952	0.548	0.951
10	0.5	I	0.000	0.950	0.437	0.952	0.459	0.949
10	0.5	II	0.076	0.907	0.525	0.947	0.834	0.946
10	0.5	III	0.024	0.945	0.459	0.951	0.544	0.950
10	0.9	I	0.001	0.943	0.434	0.940	0.461	0.944
10	0.9	II	0.129	0.785	0.487	0.931	0.811	0.930
10	0.9	III	0.038	0.917	0.447	0.944	0.538	0.948
25	0.1	I	−0.000	0.949	0.276	0.951	0.284	0.948
25	0.1	II	0.010	0.959	0.354	0.953	0.526	0.956
25	0.1	III	0.002	0.949	0.295	0.949	0.342	0.948
25	0.5	I	−0.000	0.945	0.277	0.945	0.285	0.943
25	0.5	II	0.038	0.932	0.350	0.948	0.523	0.952
25	0.5	III	0.010	0.943	0.295	0.950	0.343	0.944
25	0.9	I	0.002	0.948	0.275	0.950	0.285	0.948
25	0.9	II	0.066	0.864	0.338	0.938	0.511	0.935
25	0.9	III	0.018	0.933	0.291	0.946	0.338	0.947
50	0.1	I	−0.000	0.946	0.195	0.949	0.200	0.947
50	0.1	II	0.003	0.956	0.256	0.950	0.373	0.954
50	0.1	III	0.001	0.949	0.208	0.947	0.241	0.948
50	0.5	I	−0.001	0.946	0.195	0.948	0.199	0.946
50	0.5	II	0.021	0.940	0.254	0.949	0.370	0.949
50	0.5	III	0.006	0.942	0.209	0.951	0.242	0.944
50	0.9	I	0.000	0.941	0.195	0.945	0.200	0.941
50	0.9	II	0.037	0.898	0.247	0.940	0.364	0.939
50	0.9	III	0.009	0.937	0.208	0.950	0.241	0.944

^aFor definitions see the notes to Table 1.

effect for $ID_{\pi} = \text{II}$ is much stronger and the coverage probabilities of the post-selection confidence interval are not accurate for models with high endogeneity even with a concentration parameter of 50.

Combining the results from Tables 1 and 3, we obtain some interesting insights into the problems with Wald based inference. It was found that if all available instruments are used, then the coverage probability is distorted for concentration parameter values of 1, 10, and 25. However, this appears to be due to a combination of effects in models $ID_{\pi} = \text{I}$ and

TABLE 4 Instruments are selected by RMSC and no pre-test is conducted (Strategies 4–6) – power

$\frac{\lambda'\lambda}{q}$	σ_{ue}	ID_{π}^a	Wald			AR			K		
			0.05	0.1	0.5	0.05	0.1	0.5	0.05	0.1	0.5
Sequential approach											
1	0.1	I	0.02	0.04	0.37	0.05	0.05	0.23	0.04	0.05	0.24
1	0.1	II	0.03	0.04	0.44	0.05	0.05	0.12	0.04	0.04	0.14
1	0.1	III	0.03	0.04	0.40	0.05	0.06	0.20	0.04	0.05	0.22
1	0.5	I	0.12	0.16	0.57	0.07	0.08	0.23	0.05	0.07	0.22
1	0.5	II	0.23	0.29	0.74	0.07	0.07	0.15	0.06	0.07	0.18
1	0.5	III	0.14	0.19	0.62	0.07	0.08	0.21	0.05	0.07	0.22
1	0.9	I	0.47	0.52	0.84	0.12	0.13	0.24	0.06	0.07	0.18
1	0.9	II	0.72	0.76	0.94	0.11	0.11	0.17	0.09	0.10	0.21
1	0.9	III	0.50	0.56	0.86	0.11	0.12	0.22	0.07	0.08	0.20
10	0.1	I	0.08	0.15	0.98	0.07	0.14	0.96	0.08	0.14	0.97
10	0.1	II	0.06	0.13	0.94	0.05	0.07	0.65	0.06	0.10	0.83
10	0.1	III	0.07	0.15	0.97	0.07	0.11	0.91	0.07	0.14	0.94
10	0.5	I	0.10	0.18	0.95	0.07	0.14	0.91	0.08	0.14	0.92
10	0.5	II	0.17	0.28	0.96	0.07	0.10	0.61	0.09	0.14	0.81
10	0.5	III	0.12	0.21	0.96	0.08	0.13	0.86	0.08	0.15	0.91
10	0.9	I	0.12	0.21	0.92	0.08	0.14	0.85	0.08	0.14	0.85
10	0.9	II	0.34	0.47	0.98	0.09	0.12	0.57	0.11	0.18	0.79
10	0.9	III	0.16	0.27	0.95	0.08	0.14	0.79	0.09	0.15	0.85
25	0.1	I	0.12	0.30	1.00	0.10	0.28	1.00	0.11	0.29	1.00
25	0.1	II	0.09	0.22	1.00	0.07	0.12	0.96	0.08	0.19	0.99
25	0.1	III	0.11	0.29	1.00	0.09	0.21	1.00	0.11	0.27	1.00
25	0.5	I	0.13	0.32	1.00	0.11	0.26	1.00	0.11	0.27	1.00
25	0.5	II	0.17	0.36	1.00	0.08	0.14	0.93	0.11	0.23	0.99
25	0.5	III	0.15	0.33	1.00	0.10	0.22	1.00	0.11	0.27	1.00
25	0.9	I	0.14	0.33	1.00	0.10	0.25	1.00	0.10	0.26	1.00
25	0.9	II	0.29	0.49	1.00	0.10	0.17	0.91	0.14	0.27	0.98
25	0.9	III	0.18	0.38	1.00	0.10	0.23	0.99	0.12	0.27	1.00
50	0.1	I	0.18	0.52	1.00	0.17	0.49	1.00	0.17	0.50	1.00
50	0.1	II	0.13	0.36	1.00	0.08	0.19	1.00	0.12	0.32	1.00
50	0.1	III	0.16	0.48	1.00	0.13	0.37	1.00	0.15	0.46	1.00
50	0.5	I	0.19	0.51	1.00	0.16	0.46	1.00	0.17	0.47	1.00
50	0.5	II	0.20	0.47	1.00	0.09	0.20	1.00	0.14	0.35	1.00
50	0.5	III	0.20	0.51	1.00	0.14	0.38	1.00	0.18	0.46	1.00
50	0.9	I	0.21	0.52	1.00	0.16	0.44	1.00	0.17	0.45	1.00
50	0.9	II	0.30	0.57	1.00	0.11	0.24	1.00	0.18	0.39	1.00
50	0.9	III	0.23	0.53	1.00	0.14	0.37	1.00	0.18	0.45	1.00

^aSee Tables 1 and 2 for definitions.

$ID_{\pi} = \text{III}$: for lower values of the concentration parameter, the distortions are clearly due to weak identification, but for more moderate values of the concentration parameter the distortions are due to the inclusion of redundant or nearly redundant instruments.

Comparing the results in Tables 2 and 4, it can be seen the use of RMSC does not affect the power comparisons observed in the case where all instruments are used. Similar findings are also obtained if the selection is over all possible choices.

(c) *All instruments are used but the choice of statistic is based on the Stock–Yogo test:*

- *Strategy 7:* The Stock–Yogo statistic in (15)–(16) is used to test the null hypothesis of weak identification against the alternative that the maximal relative bias of the 2SLS estimator is 5%. If the null is not rejected, then all available instruments are used and inference is based on the AR statistic in (13). If the null is rejected, then inference is based on the Wald statistic. In either case, all instruments are used.
- *Strategy 8:* The Stock–Yogo statistic in (15)–(16) is used to test the null hypothesis of weak identification against the alternative that the maximal relative bias of the 2SLS estimator is 5%. If the null is not rejected, then all available instruments are used and inference is based on the K statistic in (14). If the null is rejected, then inference is based on the Wald statistic. In either case, all instruments are used.
- *Strategy 9:* The Stock–Yogo statistic in (15)–(16) is used to test the null hypothesis of weak identification against the alternative that the maximal size distortion of 10% for the Wald test. If the null is not rejected, then all available instruments are used and inference is based on the Anderson Rubin statistic in (13). If the null is rejected, then inference is based on the Wald statistic. In either case, all instruments are used.
- *Strategy 10:* The Stock–Yogo statistic in (15)–(16) is used to test the null hypothesis of weak identification against the alternative that the maximal size distortion of 10% for the Wald statistic. If the null is not rejected, then all available instruments are used and inference is based on the K statistic in (14). If the null is rejected, then inference is based on the Wald statistic. In either case, all instruments are used.

Table 5 reports the coverage probabilities and median widths for the 95% confidence intervals based on these four strategies along with the *nonrejection* frequencies of the Stock–Yogo test. Table 6 reports the corresponding power calculations. First consider the performance of the Stock–Yogo test. It can be seen that the version based on the size of 2SLS t -statistic is more conservative than the version based on the bias of 2SLS. This difference has minimal impact for concentration parameter values less than or equal to 10 as both versions fail to reject weak identification all or nearly all of the time. However, the two versions yield differing conclusions when the concentration parameter is 25: in this case, the version based on the bias rejects weak identification approximately 85% of the time but the version based on the size rejects weak identification only 15%. Not surprisingly, it is the version of the test based on the size of the 2SLS t -statistic that yields the strategies that are closest to the nominal level. Furthermore, with these latter strategies, the coverage probabilities are very close to the nominal level at all values of the concentration

TABLE 5 All 8 instruments are used and statistics are selected by pre-tests (Strategies 6–9) – median bias, coverage probabilities and median width of 95% confidence intervals

$\frac{\lambda' \lambda}{q}$	σ_{ue}	ID_{π}^a	Pretest based on 2SLS bias ($b = 0.05$)				Pretest based on 2SLS size ($r = 0.10$)			
			Weak ^b	AR/Wald	Width	K/Wald	Weak ^c	AR/Wald	Width	K/Wald
1	0.1	I	1.000	0.950	∞	0.941	1.000	0.950	∞	0.941
1	0.1	II	1.000	0.953	∞	0.946	1.000	0.953	∞	0.946
1	0.1	III	1.000	0.950	∞	0.945	1.000	0.950	∞	0.945
1	0.5	I	1.000	0.949	∞	0.945	1.000	0.949	∞	0.945
1	0.5	II	1.000	0.952	∞	0.946	1.000	0.952	∞	0.946
1	0.5	III	1.000	0.949	∞	0.945	1.000	0.949	∞	0.945
1	0.9	I	1.000	0.949	∞	0.942	1.000	0.949	∞	0.942
1	0.9	II	1.000	0.952	∞	0.942	1.000	0.952	∞	0.942
1	0.9	III	1.000	0.949	∞	0.945	1.000	0.949	∞	0.945
10	0.1	I	0.988	0.950	0.742	0.949	1.000	0.951	0.742	0.949
10	0.1	II	0.988	0.955	0.744	0.949	1.000	0.955	0.744	0.949
10	0.1	III	0.990	0.950	0.744	0.946	1.000	0.950	0.744	0.946
10	0.5	I	0.990	0.951	0.748	0.946	1.000	0.952	0.748	0.947
10	0.5	II	0.990	0.951	0.747	0.946	1.000	0.951	0.747	0.946
10	0.5	III	0.987	0.949	0.746	0.949	1.000	0.951	0.746	0.950
10	0.9	I	0.989	0.942	0.769	0.942	1.000	0.948	0.769	0.947
10	0.9	II	0.987	0.945	0.762	0.944	1.000	0.950	0.762	0.949
10	0.9	III	0.989	0.947	0.767	0.944	1.000	0.953	0.767	0.948
25	0.1	I	0.165	0.948	0.272	0.949	0.867	0.949	0.439	0.947
25	0.1	II	0.160	0.947	0.272	0.946	0.860	0.947	0.437	0.944
25	0.1	III	0.163	0.946	0.272	0.943	0.871	0.948	0.441	0.944
25	0.5	I	0.173	0.934	0.270	0.930	0.866	0.943	0.446	0.937
25	0.5	II	0.165	0.935	0.269	0.931	0.866	0.943	0.443	0.940
25	0.5	III	0.171	0.930	0.270	0.926	0.863	0.940	0.445	0.936
25	0.9	I	0.162	0.904	0.263	0.896	0.864	0.929	0.452	0.925
25	0.9	II	0.165	0.900	0.263	0.894	0.868	0.922	0.450	0.927
25	0.9	III	0.166	0.902	0.263	0.895	0.869	0.929	0.452	0.927
50	0.1	I	0.000	0.944	0.194	0.944	0.036	0.945	0.194	0.944
50	0.1	II	0.000	0.947	0.194	0.947	0.039	0.947	0.193	0.946
50	0.1	III	0.000	0.948	0.193	0.948	0.043	0.948	0.193	0.948
50	0.5	I	0.000	0.940	0.192	0.940	0.038	0.940	0.192	0.938
50	0.5	II	0.000	0.940	0.192	0.940	0.039	0.940	0.192	0.938
50	0.5	III	0.000	0.938	0.193	0.938	0.041	0.939	0.193	0.937
50	0.9	I	0.000	0.922	0.191	0.922	0.040	0.921	0.191	0.919
50	0.9	II	0.000	0.924	0.190	0.924	0.036	0.924	0.190	0.922
50	0.9	III	0.000	0.925	0.190	0.925	0.038	0.925	0.190	0.922

^aEmpirical frequencies that the Stock and Yogo test based on 2SLS bias ($b = 0.05$) cannot reject H_0 that instruments are weak, $c = 18$.

^bEmpirical frequencies that the Stock and Yogo test based on 2SLS size distortion ($r = 0.10$) cannot reject H_0 that instruments are weak, $c = 18$.

^cFor other definitions, see the notes to Table 1.

parameter. In models with low or moderate endogeneity, the combined AR/Wald (strategy 9) dominates both the strategies based on just one statistic (strategies 1 and 2) as it yields accurate coverage probabilities at all concentration parameter values combined with the narrower median

TABLE 6 All 8 instruments are used and statistics are selected by pre-tests (Strategies 6-9) – power

$\frac{\lambda'\lambda}{q}$	σ_{ue}	ID_{π}^a	Pretest based on 2SLS bias ($b = 0.05$)						Pretest based on 2SLS size ($r = 0.10$)					
			AR/Wald			K/Wald			AR/Wald			K/Wald		
			0.05	0.1	0.5	0.05	0.1	0.5	0.05	0.1	0.5	0.05	0.1	0.5
1	0.1	I	0.05	0.05	0.10	0.06	0.06	0.15	0.05	0.05	0.10	0.06	0.06	0.15
1	0.1	II	0.05	0.05	0.10	0.06	0.06	0.14	0.05	0.05	0.10	0.06	0.06	0.14
1	0.1	III	0.05	0.05	0.10	0.05	0.06	0.15	0.05	0.05	0.10	0.05	0.06	0.15
1	0.5	I	0.05	0.05	0.09	0.06	0.06	0.14	0.05	0.05	0.09	0.06	0.06	0.14
1	0.5	II	0.05	0.05	0.09	0.05	0.06	0.14	0.05	0.05	0.09	0.05	0.06	0.14
1	0.5	III	0.05	0.05	0.09	0.06	0.06	0.13	0.05	0.05	0.09	0.06	0.06	0.13
1	0.9	I	0.05	0.05	0.08	0.06	0.06	0.15	0.05	0.05	0.08	0.06	0.06	0.15
1	0.9	II	0.05	0.05	0.09	0.06	0.06	0.16	0.05	0.05	0.09	0.06	0.06	0.16
1	0.9	III	0.05	0.05	0.08	0.06	0.06	0.15	0.05	0.05	0.08	0.06	0.06	0.15
10	0.1	I	0.06	0.08	0.74	0.07	0.14	0.95	0.06	0.08	0.74	0.07	0.14	0.95
10	0.1	II	0.06	0.08	0.75	0.08	0.14	0.95	0.05	0.08	0.75	0.08	0.14	0.95
10	0.1	III	0.06	0.08	0.75	0.08	0.14	0.94	0.06	0.08	0.75	0.07	0.14	0.94
10	0.5	I	0.06	0.08	0.61	0.08	0.13	0.89	0.05	0.07	0.61	0.07	0.13	0.89
10	0.5	II	0.06	0.08	0.61	0.07	0.13	0.89	0.05	0.07	0.61	0.07	0.13	0.89
10	0.5	III	0.06	0.08	0.61	0.08	0.13	0.89	0.05	0.07	0.61	0.07	0.13	0.89
10	0.9	I	0.06	0.08	0.51	0.08	0.14	0.84	0.06	0.07	0.51	0.08	0.14	0.84
10	0.9	II	0.06	0.07	0.51	0.07	0.13	0.85	0.05	0.07	0.51	0.07	0.13	0.85
10	0.9	III	0.06	0.07	0.50	0.07	0.13	0.85	0.05	0.07	0.50	0.07	0.13	0.85
25	0.1	I	0.12	0.31	0.99	0.12	0.32	1.00	0.08	0.16	0.99	0.11	0.29	1.00
25	0.1	II	0.12	0.31	1.00	0.13	0.33	1.00	0.08	0.16	0.99	0.11	0.29	1.00
25	0.1	III	0.12	0.31	1.00	0.13	0.32	1.00	0.08	0.16	0.99	0.11	0.29	1.00
25	0.5	I	0.18	0.38	0.98	0.18	0.39	1.00	0.10	0.17	0.97	0.12	0.28	1.00
25	0.5	II	0.18	0.39	0.98	0.18	0.39	1.00	0.10	0.17	0.97	0.12	0.28	1.00
25	0.5	III	0.18	0.38	0.97	0.18	0.39	1.00	0.09	0.17	0.97	0.12	0.28	1.00
25	0.9	I	0.26	0.48	0.94	0.26	0.48	1.00	0.12	0.18	0.93	0.14	0.28	1.00
25	0.9	II	0.26	0.48	0.94	0.26	0.48	1.00	0.13	0.18	0.93	0.14	0.28	1.00
25	0.9	III	0.26	0.48	0.94	0.26	0.48	1.00	0.12	0.18	0.93	0.14	0.29	1.00
50	0.1	I	0.19	0.54	1.00	0.19	0.54	1.00	0.19	0.53	1.00	0.19	0.54	1.00
50	0.1	II	0.19	0.54	1.00	0.19	0.54	1.00	0.19	0.53	1.00	0.19	0.54	1.00
50	0.1	III	0.19	0.54	1.00	0.19	0.54	1.00	0.18	0.53	1.00	0.19	0.54	1.00
50	0.5	I	0.24	0.58	1.00	0.24	0.58	1.00	0.24	0.57	1.00	0.24	0.58	1.00
50	0.5	II	0.24	0.59	1.00	0.24	0.59	1.00	0.24	0.59	1.00	0.24	0.59	1.00
50	0.5	III	0.25	0.59	1.00	0.25	0.59	1.00	0.24	0.58	1.00	0.24	0.58	1.00
50	0.9	I	0.30	0.63	1.00	0.30	0.63	1.00	0.30	0.63	1.00	0.30	0.63	1.00
50	0.9	II	0.31	0.64	1.00	0.31	0.64	1.00	0.31	0.64	1.00	0.31	0.64	1.00
50	0.9	III	0.29	0.63	1.00	0.29	0.63	1.00	0.29	0.62	1.00	0.29	0.62	1.00

^aSee Tables 1, 2 and 5 for definitions.

widths of the Wald based interval for high concentration parameter values. However, in the case of high endogeneity, the Wald coverage probability is slightly too low even with a concentration parameter equal to 50 and this is inherited by the AR/Wald and K/Wald strategies.

In terms of power, it can be seen from Table 6 that the combined strategies are slightly more powerful than using K statistic on its own but this finding must be treated with some caution in view of the comments at

the end of the previous paragraph because as the statistics are compared to their nominal 5% critical values.

(d) *Statistics are selected via the Stock–Yogo test and instruments are selected by RMSC if weak identification is rejected:*

- *Strategy 11:* The Stock–Yogo statistic in (15)–(16) is used to test the null hypothesis of weak identification against the alternative that the maximal relative bias of the 2SLS estimator is 5%. If the null is not rejected, then all available instruments are used and inference is based on the Anderson Rubin statistic in (13). If the null is rejected, then the instrument is chosen using *RMSC* and inference is based on the Wald statistic.
- *Strategy 12:* The Stock–Yogo statistic in (15)–(16) is used to test the null hypothesis of weak identification against the alternative that the maximal relative bias of the 2SLS estimator is 5%. If the null is not rejected, then all available instruments are used and inference is based on the K statistic in (14). If the null is rejected, then the instrument is chosen using *RMSC* and inference is based on the Wald statistic.
- *Strategy 13:* The Stock–Yogo statistic in (15)–(16) is used to test the null hypothesis of weak identification against the alternative that the maximal size distortion of 10% for the Wald statistic. If the null is not rejected, then all available instruments are used and inference is based on the Anderson Rubin statistic in (13). If the null is rejected, then the instrument is chosen using *RMSC* and inference is based on the Wald statistic.
- *Strategy 14:* The Stock–Yogo statistic in (15)–(16) is used to test the null hypothesis of weak identification against the alternative that the maximal size distortion of 10% for the Wald statistic. If the null is not rejected, then all available instruments are used and inference is based on the K statistic in (14). If the null is rejected, then the instrument is chosen using *RMSC* and inference is based on the Wald statistic.

Table 7 reports the coverage probabilities and median widths for the 95% confidence intervals based on these four strategies. Table 8 contains the corresponding power calculations. Due to the properties of the Stock–Yogo test reported in (c), the use of *RMSC* only impacts the performance of the intervals in models with concentration parameters of 25 or 50. For these settings, the use of *RMSC* yields more accurate coverage probabilities for models $\mathbf{ID}_\pi = \mathbf{I}$ and $\mathbf{ID}_\pi = \mathbf{III}$ but less accurate probabilities for model $\mathbf{ID}_\pi = \mathbf{II}$. This echoes the finding reported in (b) above. A similar pattern also emerges if the selection is over all possible instrument choices.

TABLE 7 Statistics are selected by pre-tests and instruments are selected by RMSC only when a pretest rejects the null of weak instruments (Strategies 11–14) – median bias, coverage probabilities and median width of 95% confidence intervals

$\frac{\lambda' \lambda}{q}$	σ_{ue}	ID_{π}^a	Pretest based on 2SLS bias				Pretest based on 2SLS size			
			Med bias	AR/Wald	Width	K/Wald	Med bias	AR/Wald	Width	K/Wald
			Sequential approach							
1	0.1	I	0.048	0.950	∞	0.941	0.045	0.950	∞	0.941
1	0.1	II	0.054	0.953	∞	0.946	0.052	0.953	∞	0.946
1	0.1	III	0.051	0.950	∞	0.945	0.048	0.950	∞	0.945
1	0.5	I	0.245	0.949	∞	0.945	0.233	0.949	∞	0.945
1	0.5	II	0.241	0.952	∞	0.946	0.234	0.952	∞	0.946
1	0.5	III	0.245	0.949	∞	0.945	0.239	0.949	∞	0.945
1	0.9	I	0.436	0.949	∞	0.942	0.425	0.949	∞	0.942
1	0.9	II	0.435	0.952	∞	0.942	0.421	0.952	∞	0.942
1	0.9	III	0.435	0.949	∞	0.945	0.423	0.949	∞	0.945
10	0.1	I	0.007	0.950	0.742	0.949	0.008	0.951	0.742	0.949
10	0.1	II	0.009	0.955	0.744	0.949	0.008	0.955	0.744	0.949
10	0.1	III	0.009	0.950	0.744	0.946	0.007	0.950	0.744	0.946
10	0.5	I	0.041	0.952	0.748	0.947	0.037	0.952	0.748	0.947
10	0.5	II	0.042	0.951	0.747	0.945	0.036	0.951	0.747	0.946
10	0.5	III	0.041	0.949	0.746	0.949	0.037	0.951	0.746	0.950
10	0.9	I	0.074	0.944	0.769	0.945	0.065	0.948	0.769	0.947
10	0.9	II	0.074	0.946	0.762	0.945	0.066	0.950	0.762	0.949
10	0.9	III	0.073	0.949	0.767	0.946	0.064	0.953	0.767	0.948
25	0.1	I	0.000	0.948	0.277	0.949	0.002	0.949	0.439	0.947
25	0.1	II	0.008	0.956	0.357	0.955	0.005	0.949	0.438	0.946
25	0.1	III	0.002	0.950	0.296	0.948	0.003	0.949	0.441	0.945
25	0.5	I	0.004	0.946	0.277	0.942	0.013	0.948	0.446	0.941
25	0.5	II	0.032	0.930	0.352	0.926	0.017	0.944	0.444	0.941
25	0.5	III	0.010	0.941	0.295	0.937	0.013	0.943	0.445	0.939
25	0.9	I	0.006	0.950	0.275	0.941	0.024	0.947	0.452	0.943
25	0.9	II	0.057	0.863	0.340	0.857	0.031	0.918	0.450	0.923
25	0.9	III	0.019	0.931	0.291	0.923	0.026	0.940	0.452	0.938
50	0.1	I	−0.000	0.946	0.195	0.946	−0.000	0.947	0.195	0.946
50	0.1	II	0.003	0.956	0.256	0.956	0.004	0.955	0.256	0.954
50	0.1	III	0.001	0.949	0.208	0.949	0.000	0.949	0.208	0.949
50	0.5	I	−0.001	0.946	0.195	0.946	−0.001	0.947	0.195	0.945
50	0.5	II	0.021	0.940	0.254	0.940	0.020	0.940	0.254	0.938
50	0.5	III	0.006	0.942	0.209	0.942	0.005	0.943	0.209	0.942
50	0.9	I	0.000	0.941	0.195	0.941	−0.001	0.942	0.195	0.940
50	0.9	II	0.037	0.898	0.247	0.898	0.035	0.897	0.247	0.895
50	0.9	III	0.009	0.937	0.208	0.937	0.007	0.937	0.208	0.934

^aSee Tables 1 and 5 for definitions.

In terms of power, consider first the cases of models $ID_\pi = \mathbf{I}$ and $ID_\pi = \mathbf{III}$ for which the coverage probabilities are approximately accurate for these strategies. In these models, *K/Wald* strategy (strategy 14) appears the best. A comparison with Tables 2, 4, and 6 indicates that this strategy dominates those that use either the Wald or K tests all the time no matter whether all instruments are used or the instruments are selected via *RMSC*.

TABLE 8 Statistics are selected by pre-tests and instruments are selected by *RMSC* only when a pretest rejects the null of weak instruments (Strategies 11–14) – power

$\frac{\lambda'\lambda}{q}$	σ_{ue}	ID_{π}^a	Pretest based on 2SLS bias ($b = 0.05$)						Pretest based on 2SLS size ($r = 0.10$)					
			AR/Wald			K/Wald			AR/Wald			K/Wald		
			0.05	0.1	0.5	0.05	0.1	0.5	0.05	0.1	0.5	0.05	0.1	0.5
Sequential approach														
1	0.1	I	0.05	0.05	0.10	0.06	0.06	0.15	0.05	0.05	0.10	0.06	0.06	0.15
1	0.1	II	0.05	0.05	0.10	0.06	0.06	0.14	0.05	0.05	0.10	0.06	0.06	0.14
1	0.1	III	0.05	0.05	0.10	0.05	0.06	0.15	0.05	0.05	0.10	0.05	0.06	0.15
1	0.5	I	0.05	0.05	0.09	0.06	0.06	0.14	0.05	0.05	0.09	0.06	0.06	0.14
1	0.5	II	0.05	0.05	0.09	0.05	0.06	0.14	0.05	0.05	0.09	0.05	0.06	0.14
1	0.5	III	0.05	0.05	0.09	0.06	0.06	0.13	0.05	0.05	0.09	0.06	0.06	0.13
1	0.9	I	0.05	0.05	0.08	0.06	0.06	0.15	0.05	0.05	0.08	0.06	0.06	0.15
1	0.9	II	0.05	0.05	0.09	0.06	0.06	0.16	0.05	0.05	0.09	0.06	0.06	0.16
1	0.9	III	0.05	0.05	0.08	0.06	0.06	0.15	0.05	0.05	0.08	0.06	0.06	0.15
10	0.1	I	0.06	0.08	0.74	0.07	0.14	0.95	0.06	0.08	0.74	0.07	0.14	0.95
10	0.1	II	0.05	0.08	0.75	0.08	0.14	0.95	0.05	0.08	0.75	0.08	0.14	0.95
10	0.1	III	0.06	0.08	0.75	0.07	0.14	0.94	0.06	0.08	0.75	0.07	0.14	0.94
10	0.5	I	0.05	0.07	0.61	0.08	0.13	0.89	0.05	0.07	0.61	0.07	0.13	0.89
10	0.5	II	0.06	0.08	0.61	0.07	0.13	0.89	0.05	0.07	0.61	0.07	0.13	0.89
10	0.5	III	0.06	0.08	0.61	0.07	0.13	0.89	0.05	0.07	0.61	0.07	0.13	0.89
10	0.9	I	0.06	0.08	0.51	0.08	0.14	0.84	0.06	0.07	0.51	0.08	0.14	0.84
10	0.9	II	0.06	0.07	0.51	0.07	0.13	0.85	0.05	0.07	0.51	0.07	0.13	0.85
10	0.9	III	0.06	0.07	0.50	0.07	0.13	0.85	0.05	0.07	0.50	0.07	0.13	0.85
25	0.1	I	0.11	0.28	0.99	0.11	0.30	1.00	0.08	0.15	0.99	0.11	0.28	1.00
25	0.1	II	0.09	0.21	0.99	0.09	0.22	1.00	0.07	0.14	0.99	0.11	0.27	1.00
25	0.1	III	0.11	0.27	1.00	0.11	0.28	1.00	0.07	0.15	0.99	0.11	0.28	1.00
25	0.5	I	0.13	0.30	0.98	0.13	0.31	1.00	0.09	0.16	0.97	0.11	0.27	1.00
25	0.5	II	0.16	0.33	0.98	0.16	0.33	1.00	0.09	0.15	0.97	0.11	0.27	1.00
25	0.5	III	0.14	0.31	0.97	0.14	0.32	1.00	0.08	0.15	0.97	0.11	0.27	1.00
25	0.9	I	0.14	0.33	0.94	0.15	0.33	1.00	0.09	0.16	0.93	0.11	0.26	1.00
25	0.9	II	0.28	0.46	0.94	0.28	0.46	1.00	0.12	0.17	0.93	0.13	0.27	1.00
25	0.9	III	0.18	0.37	0.94	0.18	0.37	1.00	0.10	0.16	0.93	0.12	0.27	1.00
50	0.1	I	0.18	0.52	1.00	0.18	0.52	1.00	0.18	0.51	1.00	0.18	0.51	1.00
50	0.1	II	0.13	0.36	1.00	0.13	0.36	1.00	0.13	0.36	1.00	0.13	0.36	1.00
50	0.1	III	0.16	0.48	1.00	0.16	0.48	1.00	0.16	0.47	1.00	0.16	0.48	1.00
50	0.5	I	0.19	0.51	1.00	0.19	0.51	1.00	0.19	0.50	1.00	0.19	0.51	1.00
50	0.5	II	0.20	0.47	1.00	0.20	0.47	1.00	0.20	0.46	1.00	0.20	0.46	1.00
50	0.5	III	0.20	0.51	1.00	0.20	0.51	1.00	0.20	0.51	1.00	0.20	0.51	1.00
50	0.9	I	0.21	0.52	1.00	0.21	0.52	1.00	0.21	0.51	1.00	0.21	0.52	1.00
50	0.9	II	0.30	0.57	1.00	0.30	0.57	1.00	0.30	0.57	1.00	0.30	0.57	1.00
50	0.9	III	0.23	0.53	1.00	0.23	0.53	1.00	0.23	0.53	1.00	0.23	0.53	1.00

^aSee Tables 1, 2, and 5 for definitions.

Now consider model $ID_{\pi} = II$. In this case, as noted above, none of this group of strategies yields accurate coverage probabilities and so power comparisons are of little interest.

As a final comment, we note that there is a way in which the use of the Stock–Yogo test undermines some of the gains from *RMSC* that are

noted in (b) above. Recall that for models $\mathbf{ID}_\pi = \mathbf{I}$ and $\mathbf{ID}_\pi = \mathbf{III}$ with the concentration parameter equal to 10 or 25, the use of *RMSC* yielded considerable improvement in the coverage probabilities of the Wald based inferences relative to the case in which all instruments are used. However, due to the conservative nature of the Stock–Yogo test, these settings are diagnosed (misdiagnosed?) as weak identification and so *RMSC* is not used in those models when inference is based on strategies 11–14.

Collective Evaluation of the Strategies in (a)–(d)

Pulling all these results together, it would appear that strategy 14 dominates for models $\mathbf{ID}_\pi = \mathbf{I}$ and $\mathbf{ID}_\pi = \mathbf{III}$. Therefore, in these models, it is beneficial to proceed as follows: use the Stock–Yogo test based on the size of *t*-statistic to determine the quality of the identification; if weak identification is not rejected inference is based on the K statistic using all the instruments; if weak identification is rejected then inference is based on the Wald statistic using the instruments selected via *RMSC*. However, while this strategy yields accurate coverage probabilities in these two models, it does not do so in model $\mathbf{ID}_\pi = \mathbf{II}$. In the latter case, the dominant strategy is to base inference on the K statistic calculated using all the instruments (strategy 3) irrespective of the quality of the identification.

6. CONCLUDING REMARKS

Hall et al. (2007) propose a method for moment selection based on an information criterion that is a function of the entropy of the limiting distribution of the GMM estimator. They establish the consistency of the method subject to certain conditions that include the identification of the parameter vector by at least one of the moment conditions being considered. In this article, we examine the limiting behavior of this moment selection method when the parameter vector is weakly identified by all the moment conditions being considered. It is shown that the selected moment condition is random and hence not consistent in any meaningful sense. As a result, we propose a two-step procedure for moment selection in which identification is first tested using a statistic proposed by Stock and Yogo (2003) and then only if this statistic indicates identification does the researcher proceed to the second step in which the aforementioned information criterion is used to select moments. The properties of this two-step procedure are contrasted with those of strategies based on either using all available moments or using the information criterion without the identification pre-test. The inference methods considered are based on the Wald statistic, Anderson and Rubin's (1949) statistic, Kleibergen's (2002) K statistic, and combinations thereof in which

the choice is based on the outcome of the test for weak identification. Our simulation results reveal some interesting findings. First and foremost, the performance of the various inference strategies depends on not only the collective explanatory power of the instruments but also on the pattern of coefficients in the first-step regression. Of the three models considered, two involve patterns in which the first-step regression coefficients decline and hence the instruments are of differing degrees of importance in the explanation of the endogenous regressor. In these cases, there appears an advantage to the use of the pre-test for weak identification and to the use of *RMSC* to select instruments if weak identification is rejected. However, in the other model, the first-step regression coefficients are all equal, and in this case the dominant strategy involves using the *K* statistic calculated with all the instruments irrespective of the quality of the identification. Therefore, on the basis of this evidence, it appears that if a single strategy is to be selected then it should be to base inference on the *K* statistic calculated using all instruments.

Our findings also raise some interesting questions: Is the use of *K* statistic with all instruments the dominant strategy in other simulation designs? Is there a way to diagnose the dominant inference strategy for the data/model in hand? These questions are beyond the scope of this article and so are left to future research.

ACKNOWLEDGMENTS

Article prepared for “A Conference in Honor of George Judge: Recent Developments in the Theory, Method, and Application of Information and Entropy Econometrics,” American University, Washington DC, September 23–25, 2005. We are very grateful to a referee and the editors for valuable comments on an earlier draft of the article.

REFERENCES

- Ahmed, N. A., Gokhale, D. V. (1989). Entropy expressions and their estimators for multivariate distributions. *IEEE Transactions on Information Theory* 35:688–692.
- Anderson, T. W., Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 20:46–63.
- Andrews, D. W. K. (1999). Consistent moment selection procedures for generalized method of moments estimation. *Econometrica* 67:543–564.
- Cragg, J. G., Donald, S. G. (1993). Testing identifiability and specification in instrumental variables estimation. *Econometric Theory* 9:222–240.
- Donald, S. G., Newey, W. K. (2001). Choosing the number of instruments. *Econometrica* 69:1161–1192.
- Hall, A. R. (2005). *Generalized Method of Moments*. Oxford, UK: Oxford University Press.
- Hall, A. R., Peixe, F. P. M. (2003). A consistent method for the selection of relevant instruments in linear models. *Econometric Reviews* 22:269–288.
- Hall, A. R., Rudebusch, G., Wilcox, D. (1996). Judging instrument relevance in instrumental variables estimation. *International Economic Review* 37:283–298.

- Hall, A. R., Inoue, A., Jana, K., Shin, C. (2007). Information in generalized method of moments estimation and entropy based moment selection. *Journal of Econometrics* 138:488–512.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50:1029–1054.
- Hansen, L. P. (1985). A method of calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *Journal of Econometrics* 30:203–238.
- Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumentatl variables regression. *Econometrica* 70:1781–1803.
- Kocherlakota, N. R. (1990). On tests of representative consumer asset pricing models. *Journal of Monetary Economics* 26:285–304.
- Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica* 71:1027–1048.
- Nelson, C. R., Startz, R. (1990). The distribution of the instrumental varaibles estimator and its t ratio when the instrument is a poor one. *Journal of Business* 63:S125–S140.
- Pötscher, B. M. (1991). Effects of model selection on inference. *Econometric Theory* 7:163–185.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. 2nd ed. New York, NY: Wiley.
- Staiger, D., Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica* 65:557–586.
- Stock, J. H., Wright, J. H. (2000). Gmm with weak identification. *Econometrica* 68:1055–1096.
- Stock, J. H., Yogo, M. (2003). Testing for weak instruments in linear IV regression. Discussion paper, Kennedy School of Government, Harvard University, Cambridge, MA.
- Zivot, E., Startz, R., Nelson, C. R. (1998). Valid confidence intervals and inference in the presence of weak instruments. *International Economic Review* 39:1119–1144.