Improved Inference in Weakly Identified Instrumental Variables Regression

Richard Startz, Eric Zivot and Charles R. Nelson*
Department of Economics, University of Washington

December 9, 2002
This version: October 2, 2003

Abstract

It is now well known that standard asymptotic inference techniques for instrumental variable estimation perform very poorly in the presence of weak instruments. Specifically, standard asymptotic techniques give spuriously small standard errors, leading investigators to apparently tight confidence regions which may be very far from the true parameter of interest. While much research has been done on inference in models with one right-hand-side endogenous variable, not much is known about inference on individual coefficients in models with multiple right-hand-side endogenous variables. In this paper we systematically investigate inference on individual structural coefficients in instrumental variables regression models with multiple right-hand-side endogenous variables. We focus on the cases where instruments may be weak for all coefficients or only for a subset of coefficients. We introduce a new test statistic, the S-statistic, which has good properties under weak identification. We then evaluate existing techniques for performing inference on individual coefficients using Staiger and Stock's weak instrument asymptotics, and perform extensive finite sample analyses using Monte Carlo simulations.

1 Introduction

It is now well known that standard asymptotic inference techniques for instrumental variable (IV) estimation perform very poorly in the presence of weak instruments. The failure is of the worst kind — false results are accompanied by reported confidence intervals which lend an appearance of great precision. That point estimates of coefficients do a poor job of telling us the true values of those coefficients is probably irremediable, after all if an equation is poorly identified then the data do not tell us much about the parameters of the system. In this paper we uncover test statistics and related confidence intervals that work quite well in the sense that they lead to reasonably accurate inference when instruments are poor and that are essentially identical to the usual asymptotic IV test statistics and confidence intervals when the instruments are good. This sort of performance under weak and strong identification respectively is important as it discourages practitioners' natural tendency to cling to traditional methods which may give (spuriously) tight confidence bounds and erroneous inference.

We make two contributions to inference when instruments are weak. First, we introduce a new test statistic, the S-statistic and associated confidence intervals. Second,

*Corresponding author: Richard Startz. email: `startz@u.washington.edu`. Support from NSF grant SBR-9711301 is gratefully acknowledged.

we analyze the performance of the S-statistic and a number of other statistics proposed in the recent literature. Our results show that these statistics are sufficiently good and we suggest that they be produced as a routine supplement to the traditional asymptotic t-statistics based on IV estimation when there is any risk at all of there being weak instruments. No statistic dominates in terms of size stability and power in all cases, but we are able to provide some insight into when one may be particularly appropriate. In terms of size and power, the S-statistic is sometimes better and sometimes worse than competitors. We emphasize its particular advantage in having readily computable confidence intervals, since in applied work rejection or acceptance of a point hypothesis is rarely the true objective.

Most of the previous research on inference in IV regression models with weak instruments has concentrated on the simple model with a single right-hand-side, or included, endogenous variable. Unfortunately, when we consider the more general IV regression model with multiple included endogenous variables many of the results for the single included endogenous variable model do not apply for individual structural coefficients in the more general model. The fundamental issue is that when a true null is specified for the complete parameter vector then estimation under the null can give a consistent estimate of the error variance, while in contrast specification of a null on an individual coefficient does not. In this paper, we concentrate our analysis on the problem of making valid inference on individual structural coefficients in the IV regression model with multiple included endogenous variables and weak instruments. Our approach is similar in spirit to that taken by Choi and Phillips (1992), who considered finite sample and asymptotic inference in partially identified structural equations. We extend the framework of Choi and Phillips to allow for weak instruments, and we consider non-standard methods for inference on structural coefficients. We consider cases for which instruments are weak for all structural coefficients, and cases for which instruments are weak for some coefficients but not others. We also consider the case for which instruments are weak for individual coefficients but strong for a linear combination of the structural coefficients. We utilize the weak instrument asymptotic framework of Staiger and Stock (1997) to analyze the asymptotic behavior of estimators and test statistics for individual structural coefficients. We also evaluate the finite sample performance of various estimators and test statistics through an extensive set of Monte Carlo experiments.

The plan of the paper is as follows. After a review of the recent literature on estimation and inference in IV regression models with weak instruments, we present the standard IV regression model for the case of two right-hand-side endogenous variables to set notation. Next we present the standard identification conditions and establish cases of partial identification and weak instruments. The S-statistic is presented as a joint test of identification and a structural hypothesis. We then survey estimation and inference methods in IV regression, paying particular attention to estimation and inference on individual structural parameters. Following this, we summarize the asymptotic behavior of various estimators and test statistic under a variety of weak instrument cases. We then appraise the finite sample performance of various statistics through an extensive set of Monte Carlo simulations. We conclude with a brief summary, recommendations for empirical practice and suggestions for future research.

2 A Brief Review of the Literature

A series of recent papers have examined the distribution of the instrumental variable estimator under weak identification and the related issue of the performance of the traditional asymptotic tests. Papers include Bekker (1994), Blomquest and Dahlberg (1999), Bound, Jaeger, and Baker (1995), Choi and Phillips (1992), Hahn and Hausman (2002), Hahn and Inuoe (2002), Hall, Rudebusch and Wilcox (1996), Imbens and Chamberlain (2003), Kleibergen (2000, 2002), Kleibergen and Zivot (2003), Maddala and Jeong (1992), Morera (2001a,b), Nelson and Startz (1990a, b), Phillips (1989), Staiger and Stock (1997), Stock, Wright and Yogo (2002), Wang and Zivot (1998), Wong (1999) and Zivot, Startz and Nelson (1998). Dufour (1997) gives general results for obtaining correct probability levels with weak identification. In particular, Dufour shows that for a statistic of nominal size α to be valid under weak identification, the confidence intervals implied by the statistic must be unbounded at least $1-\alpha$ percent of the time.

Half a century ago, Anderson and Rubin (1949) described the Anderson-Rubin (AR) statistic, which under normality provides an exact small sample test of a hypothesis which specifies values for every element of the structural parameter vector, $\boldsymbol{\beta}$. Zivot, Startz, and Nelson (1998) show how to use the AR-statistic to construct confidence regions in the case of a single endogenous variable and provide improved statistics for maximum likelihood and generalized method of moments estimates based on degrees-of-freedom-corrected LR and LM tests. Wang and Zivot (1998) provide an asymptotic justification using the Staiger and Stock (1997) local-to-zero asymptotics for these results. Recently, Kleibergen (2002) and Moreira (2001) have proposed asymptotically similar LM tests that have better power than the AR test and the LR and LM tests studied by Wang and Zivot (1998).

The analysis in most of the above papers is limited to inference in the case of a single endogenous right hand side variable or to hypotheses specifying values for the entire vector of coefficients; here we deal with inference on individual coefficients in a model with two-right hand side variables extending the results of Choi and Phillips (1992) to the case of weak instruments. We note, however, that Dufour (1997), Wang and Zivot (1998), Dufour and Jasiak (2000) describe the use of numerical projections of joint test statistics to obtain confidence sets for individual elements of but do not study these methods in the presence of weak instruments. As a practical matter, using the projection procedure in general requires complicated numerical maximization. Recently, Taamouti (2001) Dufour and Taamouti (2003) provide a limited set of results for analytically obtaining projection-based confidence sets for individual structural coefficients based on certain types of test statistics.

Stock and Wright (2000) provides a general procedure for inference on structural parameters estimated by generalized method of moments (GMM) with weak instruments, which for the linear single equations model is based on TSLS or LIML estimates. If some endogenous variables are well identified, Stock and Wright suggest concentrating out the well identified parameters and using an Anderson-Rubin type statistic for the remaining weakly identified parameters. However, Stock and Wright point out that using their method "construction of asymptotically valid confidence intervals for subvectors ... is somewhat ... difficult," but that an asymptotically conservative confidence interval can be found by projecting out parameters as suggested in Dufour (1997). Kleibergen (2000) provides an alternative to Stock and Wright's concentrated AR statistic in the linear IV model, and a more general alternative in the GMM context is provided in Kleibergen (2002). The statistics proposed by Stock and Wright and Kleibergen require partial identification to be asymptotically valid. We evaluate these statistics in a general framework where partial identification is a special case.

3 The IV Regression Model

We begin with the classic statements about IV regression in an identified linear model, in the process defining notation for the paper.

3.1 Structure and Reduced Form

Consider the structural linear equation with k right-hand side variables[1]

$$
\begin{aligned}
\underset{(n\times1)}{\mathbf{y}} &= \underset{(n\times k)}{\mathbf{X}}\underset{(k\times1)}{\boldsymbol{\beta}} + \underset{(n\times1)}{\mathbf{u}} \\
&= \underset{(n\times1)}{\mathbf{X}_i}\underset{(1\times1)}{\beta_i} + \underset{(n\times(k-1))}{\mathbf{X}_{\sim i}}\underset{((k-1)\times1)}{\boldsymbol{\beta}_{\sim i}} + \underset{(n\times1)}{\mathbf{u}}
\end{aligned}
\tag{1}
$$

where $\mathbf{X}_i$ is the i^{th} column of $\mathbf{X}$, $\mathbf{X}_{\sim i}$ is the remainder of $\mathbf{X}$, and $\mathbf{u}$ is a random error vector. Our focus will be on making inference on the scalar parameter β_i using instrumental variables regression, when the variables in $\mathbf{X}$ are endogenous; i.e., correlated with the error term $\mathbf{u}$. The reduced form of the model consists of the population regression of $\mathbf{y}$ and each column of $\mathbf{X}$ on all q of the exogenous instruments in the matrix $\mathbf{Z}$:

$$
\underset{(n\times1)}{\mathbf{y}} = \underset{(n\times q)}{\mathbf{Z}}\underset{(q\times1)}{\boldsymbol{\theta}} + \underset{(n\times1)}{\mathbf{v}} \tag{2}
$$

$$
\underset{(n\times k)}{\mathbf{X}} = \underset{(n\times q)}{\mathbf{Z}}\underset{(q\times k)}{\boldsymbol{\Gamma}} + \underset{(n\times k)}{\mathbf{V}} \tag{3}
$$

The corresponding reduced form equations for the endogenous variables $\mathbf{X}_i$ and $\mathbf{X}_{\sim i}$ are

$$
\underset{(n\times1)}{\mathbf{X}_i} = \underset{(n\times q)}{\mathbf{Z}}\underset{(q\times1)}{\boldsymbol{\Gamma}_i} + \underset{(n\times1)}{\mathbf{V}_i} \tag{4}
$$

$$
\underset{(n\times(k-1))}{\mathbf{X}_{\sim i}} = \underset{(n\times q)}{\mathbf{Z}}\underset{(q\times(k-1))}{\boldsymbol{\Gamma}_{\sim i}} + \underset{(n\times(k-1))}{\mathbf{V}_{\sim i}} \tag{5}
$$

The model described in equations (1) - (3) is traditionally called the linear IV regression model.

3.2 Assumptions

Let $\xrightarrow{p}$ denote convergence in probability and $\xrightarrow{d}$ denote convergence in distribution. We make the following high-level assumptions that impose rather weak moment conditions on the exogenous variables and error terms:

Assumption 1

1. $\mathbf{Z}$ has full column rank q and is uncorrelated with $\mathbf{u}$, and $\mathbf{V}$.
2. $E[\mathbf{Z}_t\mathbf{Z}_t'] = \mathbf{M} > 0$, where $\mathbf{Z}_t$ denotes the t^{th} observation on $\mathbf{Z}$
3. The error terms u_t,and $\mathbf{V}_t$ are assumed to have mean zero, and to be serially uncorrelated and homoskedastic with positive definite covariance matrix
$$\boldsymbol{\Sigma} = var\begin{pmatrix} u_t \\ \mathbf{V}_t \end{pmatrix} = \begin{pmatrix} \sigma_{uu} & \boldsymbol{\Sigma}_{\mathbf{V}u}' \\ \boldsymbol{\Sigma}_{\mathbf{V}u} & \boldsymbol{\Sigma}_{\mathbf{VV}} \end{pmatrix}$$
4. $(n^{-1}\mathbf{u}'\mathbf{u}, n^{-1}\mathbf{V}'\mathbf{u}, , n^{-1}\mathbf{V}'\mathbf{V}) \xrightarrow{p} (\sigma_{uu}, \boldsymbol{\Sigma}_{\mathbf{V}u}, \boldsymbol{\Sigma}_{\mathbf{VV}})$
5. $n^{-1}\mathbf{Z}'\mathbf{Z} \xrightarrow{p} \mathbf{M}$
6. $(n^{-1/2}\mathbf{Z}'\mathbf{u}, n^{-1/2}\mathbf{Z}'\mathbf{V}) \xrightarrow{d} (\boldsymbol{\Psi}_{Zu}, \boldsymbol{\Psi}_{ZV})$, where $\boldsymbol{\Psi} = (\boldsymbol{\Psi}_{Zu}', vec(\boldsymbol{\Psi}_{ZV})')'$ is distributed as $N(0, \boldsymbol{\Sigma}\otimes\mathbf{M})$

[1]For notational brevity we omit any included exogenous variables. The model with included exogenous variables has the same form as the model without included exogenous variables by using the Frisch-Waugh-Lovell theorem and interpreting all data matrices as residuals from the projection on the included exogenous variables.

3.3 Estimation

The vector $\boldsymbol{\beta}$ is commonly estimated by the method of instrumental variables (equivalently the method of two stage least squares or generalized method of moments). The instrumental variables (IV) estimator is

$$\hat{\boldsymbol{\beta}}_{IV} = (\mathbf{X}'\mathbf{P_Z X})^{-1}\mathbf{X}'\mathbf{P_Z y} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{y} \tag{6}$$

where $\mathbf{P_Z}$=$\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ and $\hat{\mathbf{X}} = \mathbf{P_Z X}$. Using standard partitioned regression techniques, the IV estimator of β_i may be expressed as

$$\hat{\beta}_{i,IV} = (\hat{\mathbf{X}}_i'\hat{\mathbf{Q}}_{\text{-}i}\hat{\mathbf{X}}_i)^{-1}\hat{\mathbf{X}}_i'\hat{\mathbf{Q}}_{\text{-}i}\mathbf{y} \tag{7}$$

where $\hat{\mathbf{X}}_i$=$\mathbf{P}_Z\mathbf{X}_i$, $\hat{\mathbf{X}}_{\text{-}i}$=$\mathbf{P}_Z\mathbf{X}_{\text{-}i}$, and $\hat{\mathbf{Q}}_{\text{-}i}$=$\mathbf{I}_q$−$\mathbf{P}_{\hat{\mathbf{X}}_{\text{-}i}}$.

Given that Assumption 1 and the traditional rank and order conditions hold, $\sqrt{n}(\hat{\boldsymbol{\beta}}_{IV} - \beta) \xrightarrow{d} N(\boldsymbol{\beta}, \sigma_{uu}\mathbf{H})$ where $\mathbf{H} = (\boldsymbol{\Gamma}'\mathbf{M}\boldsymbol{\Gamma})^{-1}$. A consistent estimate of the asymptotic variance $\sigma_{uu}\mathbf{H}$ is given by $n\hat{\sigma}_{uu,IV}\hat{\mathbf{H}}$, where $\hat{\sigma}_{uu,IV} = n^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{IV})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{IV})$ and $\hat{\mathbf{H}} = (\mathbf{X}'\mathbf{P_Z X})^{-1}$. Similarly, $\sqrt{n}(\hat{\beta}_{i,IV} - \beta_i) \xrightarrow{d} N(\beta_i, \sigma_{uu}H_{ii})$ where, given the partitioning in (1), $H_{ii} = \left(\boldsymbol{\Gamma}_i'\left(\mathbf{M} - \mathbf{M}\boldsymbol{\Gamma}_{\text{-}i}(\boldsymbol{\Gamma}_{\text{-}i}\mathbf{M}\boldsymbol{\Gamma}_{\text{-}i})^{-1}\boldsymbol{\Gamma}_{\text{-}i}'\mathbf{M}\right)\boldsymbol{\Gamma}_i\right)^{-1}$ is the first diagonal element of $\mathbf{H}$. A consistent estimate of the asymptotic variance $\sigma_{uu}H_{ii}$ is given by $n\hat{\sigma}_{uu,IV}\hat{H}_{ii}$, where $\hat{H}_{ii} = (\hat{\mathbf{X}}_i'\hat{\mathbf{Q}}_{\text{-}i}\hat{\mathbf{X}}_i)^{-1}$ is the first diagonal element of $\hat{\mathbf{H}}$.

The asymptotic variance of $\hat{\boldsymbol{\beta}}_{IV}$ is finite provided the rank condition for identification holds (see Davidson and MacKinnon (1993), chapter 18). Equivalently, $\mathbf{H}$ and H_{ii} exist if the rank condition holds. For our purposes, β_i is identified in the full model (1)-(3) if H_{ii} exists, unidentified if H_{ii} does not exist, and is weakly identified if H_{ii} is "nearly infinite"[2]. Since H_{ii} is a scalar, we can alternatively characterize the identifiability of β_i by examining

$$H_{ii}^{-1} = \boldsymbol{\Gamma}_i'\left(\mathbf{M} - \mathbf{M}\boldsymbol{\Gamma}_{\text{-}i}(\boldsymbol{\Gamma}_{\text{-}i}\mathbf{M}\boldsymbol{\Gamma}_{\text{-}i})^{-1}\boldsymbol{\Gamma}_{\text{-}i}'\mathbf{M}\right)\boldsymbol{\Gamma}_i$$

Notice that $H_{ii}^{-1} = 0$ if $\boldsymbol{\Gamma}_i = \mathbf{0}$ or if $\boldsymbol{\Gamma}_i = \boldsymbol{\Gamma}_{\text{-}i}\mathbf{a}$ for some non-zero $(k-1) \times 1$ vector $\mathbf{a}$. In the former case, β_i is not identified but $\boldsymbol{\beta}_{\text{-}i}$ is identified provided $\boldsymbol{\Gamma}_{\text{-}i}$ has full rank. In the latter case, both β_i and $\boldsymbol{\beta}_{\text{-}i}$ are not separately identified but the linear combination $\boldsymbol{\alpha} = \mathbf{a}\beta_i + \boldsymbol{\beta}_{\text{-}i}$ is identified[3].

The reduced form coefficients are appropriately estimated by least squares, $\hat{\boldsymbol{\theta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$, $\hat{\boldsymbol{\Gamma}}_i = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}_i$, $\hat{\boldsymbol{\Gamma}}_{\text{-}i} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}_{\text{-}i}$ and $\hat{\boldsymbol{\Gamma}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$. For the construction of the S-statistic in the next section, it is useful to note that $\hat{\boldsymbol{\beta}}_{IV}$ and $\hat{\mathbf{H}}$ can be written in terms of the estimated reduced form parameters and the instruments $\mathbf{Z}$:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{IV} &= (\hat{\boldsymbol{\Gamma}}'\mathbf{Z}'\mathbf{Z}\hat{\boldsymbol{\Gamma}})^{-1}\hat{\boldsymbol{\Gamma}}'\mathbf{Z}'\mathbf{Z}\hat{\boldsymbol{\theta}} \\ \hat{\mathbf{H}} &= (\hat{\boldsymbol{\Gamma}}'\mathbf{Z}'\mathbf{Z}\hat{\boldsymbol{\Gamma}})^{-1} \end{aligned}$$

Because we need the covariance matrix of the estimated reduced form coefficients, it is convenient to think of the reduced form as a system of seemingly unrelated regressions

$$vec(\mathbf{y}, \mathbf{X}_i, \mathbf{X}_{\text{-}i}) = (\mathbf{Z} \otimes I)vec(\boldsymbol{\theta}, \boldsymbol{\Gamma}_i, \boldsymbol{\Gamma}_{\text{-}i}) + vec(\mathbf{v}, \mathbf{V}_i, \mathbf{V}_{\text{-}i}) \tag{8}$$

[2] A formal characterization of weak identification based on Staiger and Stock's (1997) weak instrument asymptotics is given in section xxx.

[3] We note that $H_{ii}^{-1} = \hat{\mathbf{X}}_i'\hat{\mathbf{Q}}_{-i}\hat{\mathbf{X}}_i$ is closely related to Shea's partial R^2 statistic for testing instrument relevance (see Shea (1997)). Specifically, a little algebra, which is implicit in Shea's paper, shows that the numerator of Shea's partial R^2 is equal to H_{ii}^{-1}. Consequently, Shea's partial R^2 will be close to zero whenever H_{ii}^{-1} is close to zero.

Define $\boldsymbol{\lambda}$ to be the $q\cdot(k+1)\times 1$ column vector of reduced form parameters in (8) and $\hat{\boldsymbol{\lambda}}$ to be the corresponding least squares coefficients. Then, under assumption 1 $\sqrt{n}(\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda})\to N(\mathbf{0},\boldsymbol{\Sigma}_r\otimes\mathbf{M}^{-1})$ where $\boldsymbol{\Sigma}_r = var((v_t, V_{it}, \mathbf{V'}_{\tilde{}it})')$. A consistent estimate of the asymptotic covariance of $\hat{\boldsymbol{\lambda}}$ is given by $\hat{\boldsymbol{\Sigma}}_r\otimes(\mathbf{Z}'\mathbf{Z})^{-1}$ where $\hat{\boldsymbol{\Sigma}}_r = n^{-1}\sum_{t=1}^{n}(v_t, V_{it}, \mathbf{V'}_{\tilde{}it})'(v_t, V_{it}, \mathbf{V'}_{\tilde{}it})$.

4 The S-Statistic

We first introduce the S-statistic as a joint test of a hypothesis on one out of several structural coefficients and on identification. We then relate the S-statistic and the traditional IV Wald statistic to tests of hypotheses on the reduced form parameters and we present intuition for thinking of the S-statistic as providing a scaling correction to the usual IV asymptotic t-statistic (asymptotic-t).

4.1 Combining tests of the structural hypothesis and identification

First, we need an empirical measure of identification for β_i. Define

$$\hat{\Delta}_i = \sqrt{\hat{H}_{ii}^{-1}} \tag{9}$$

Computationally, $\hat{\Delta}_i$ is the square root of the reciprocal of the reported standard error of $\hat{\beta}_{i,IV}$ divided by $\sqrt{\hat{\sigma}_{uu,IV}}$, the standard error of the IV regression. Note that $\Delta_i = \sqrt{H_{ii}^{-1}} > 0$ is necessary for the rank condition to hold and for β_i to be identified in the IV regression model (1) - (3).

Suppose we wish to test $H_0 : \beta_i = \beta_i^0$. Standard practice is to use the asymptotic t-statistic

$$t_{IV}(\beta_i^0) = \frac{\hat{\beta}_{i,IV} - \beta_i^0}{SE(\hat{\beta}_{i,IV})} \tag{10}$$

where $SE(\hat{\beta}_{i,IV}) = \sqrt{\hat{\sigma}_{uu,IV}\hat{H}_{ii}}$. We augment this comparison so that the test statistic will be close to zero either if the estimated deviation is small or if the evidence for identification is weak by forming

$$\hat{\Psi}_i = \hat{\Delta}_i(\hat{\beta}_{i,IV} - \beta_i^0). \tag{11}$$

If β_i is weakly identified then $\hat{\Delta}_i$, and therefore $\hat{\Psi}_i$ will be close to zero and we will be appropriately unable to reject the hypothesized value β_i^0.

In order to derive a formal test statistic, it is useful to write $\hat{\Psi}_i$ explicitly as a function of $\mathbf{Z}$, β_i^0 and the estimated reduced form parameters $\hat{\boldsymbol{\Gamma}}$ and $\hat{\boldsymbol{\theta}}$:

$$\Psi_i(\hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{\theta}}; \mathbf{Z}; \beta_i^0) = \sqrt{(\hat{\boldsymbol{\Gamma}}'\mathbf{Z}'\mathbf{Z}\hat{\boldsymbol{\Gamma}})_{ii}^{-1}}\left(\left((\hat{\boldsymbol{\Gamma}}'\mathbf{Z}'\mathbf{Z}\hat{\boldsymbol{\Gamma}})^{-1}\hat{\boldsymbol{\Gamma}}'\mathbf{Z}'\mathbf{Z}\hat{\boldsymbol{\theta}}\right)_i - \beta_i^0\right)$$

In order to studentize $\hat{\Psi}_i$ we require an estimate of the asymptotic variance, $\hat{\sigma}^2_{\hat{\Psi}_i}$. Since the estimated reduced form parameters are asymptotically normal, we can compute $\hat{\sigma}^2_{\hat{\Psi}_i}$ by the usual Taylor series approximation (delta method):

$$\hat{\sigma}^2_{\hat{\Psi}_i} = \frac{\partial\hat{\Psi}_i}{\partial\hat{\boldsymbol{\lambda}}'}\widehat{var}(\hat{\boldsymbol{\lambda}})\frac{\partial\hat{\Psi}_i}{\partial\hat{\boldsymbol{\lambda}}} \tag{12}$$

The S-statistic for testing $H_0 : \beta_i = \beta_i^0$ is then defined as the ratio

$$S_i(\beta_i^0) \equiv \frac{\hat{\Psi}_i}{\sqrt{\hat{\sigma}^2_{\hat{\Psi}_i}}} \tag{13}$$

Note that $S_i(\beta_i^0)$ is a function of the reduced form parameters, which can be consistently estimated even when the structural parameters are not identified, the hypothesized value β_i^0 and the instruments $\mathbf{Z}$. The partial derivatives $\frac{\partial \hat{\Psi}_i}{\partial \hat{\boldsymbol{\lambda}}}$ used in computing $\hat{\sigma}^2_{\hat{\Psi}_i}$ may be conveniently calculated by the numerical delta method, and $\widehat{var}(\hat{\boldsymbol{\lambda}}) = \hat{\boldsymbol{\Sigma}}_r \otimes (\mathbf{Z}'\mathbf{Z})^{-1}$ follows immediately from the reduced form estimates[4].

Zivot, Startz and Nelson (1998) emphasize that a problem with the use of TSLS or LIML t-ratios in the presence of weak instruments is that the estimate of the structural error variance σ_{uu} is inconsistent. The S-statistic may be thought of as a t-statistic that is more robust to the presence of weak instruments by utilizing a better estimate of σ_{uu}[5]. To see this, note that (10) may be re-expressed as

$$t_{IV}(\beta_i^0) = \frac{\left(\hat{\beta}_{i,TSLS} - \beta_i^0\right)\hat{\Delta}_i}{\sqrt{\hat{\sigma}_{uu,IV}}} = \frac{\hat{\Psi}_i}{\sqrt{\hat{\sigma}_{uu,IV}}}$$

The variance estimate (12) is now seen to be an estimate of the structural error variance σ_{uu} using the reduced form estimates $\hat{\boldsymbol{\lambda}}$ and β_i^0.

5 Alternative Tests for Individual Coefficients

Several other statistics have been proposed for making inference on individual structural coefficients in the IV regression model. Some of these methods are based on the IV estimator and some are based on the limited information maximum likelihood (LIML) estimator. In this section, we briefly describe some of these statistics as they are natural competitors to the S-statistic.

5.1 LIML t and LR statistics

The LIML estimator of $\boldsymbol{\beta}$ maximizes the log-likelihood function concentrated with respect to $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$

$$L^c(\boldsymbol{\beta}) = -n\ln(2\pi) - \frac{n}{2}\ln k(\boldsymbol{\beta}) - \frac{n}{2}\ln|\mathbf{Y}\mathbf{Q}_{\mathbf{Z}}'\mathbf{Y}| \tag{14}$$

where $\mathbf{Y} = [\mathbf{y}\ \mathbf{X}]$ and

$$k(\boldsymbol{\beta}) = \frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{Q}_{\mathbf{Z}}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}.$$

The LIML estimator of $\boldsymbol{\beta}$ equivalently minimizes $k(\boldsymbol{\beta})$ and the minimized value, $k(\widehat{\boldsymbol{\beta}}_{LIML}) = \widehat{k}_{LIML}$, can be shown to be the smallest root of the determinantal equation $|\mathbf{Y}'\mathbf{Q}_{\mathbf{X}}\mathbf{Y} - k\mathbf{Y}'\mathbf{Q}_{\mathbf{Z}}\mathbf{Y}|$. The LIML estimator is usually expressed as the $k-$class estimator

$$\widehat{\boldsymbol{\beta}}_{LIML} = \left[\mathbf{X}'(\mathbf{I}_n - \widehat{k}_{LIML}\mathbf{Q}_{\mathbf{Z}})\mathbf{X}\right]^{-1}\left[\mathbf{X}'(\mathbf{I}_n - \widehat{k}_{LIML}\mathbf{Q}_{\mathbf{Z}})\mathbf{y}\right].$$

[4] Matlab code for computing the S-statistic is available from the authors upon request.

[5] We experimented with other versions of the IV $t-$statistic that made use of estimates of the structural error variance under the restriction that $\beta_i = \beta_i^0$ and found that the $S-$statistic had the best performance in finite samples.

For testing $\beta_i = \beta_i^0$, the LIML $t - ratio$ is

$$t_{LIML}(\beta_i^0) = \frac{\hat{\beta}_{i,LIML} - \beta_i^0}{\widehat{SE}(\hat{\beta}_{i,LIML})} \tag{15}$$

where $\widehat{SE}(\hat{\beta}_{LIML}) = \sqrt{\widehat{var}(\hat{\beta}_{i,LIML})} = \sqrt{\hat{\sigma}_{uu,LIML} \cdot \left[\mathbf{X}'(\mathbf{I}_n - \widehat{k}_{LIML}\mathbf{Q}_{\mathbf{Z}})\mathbf{X}\right]_{ii}^{-1}}$. The LR statistic is

$$LR_{LIML}(\beta_i^0) = n\ln(\widetilde{k}_{LIML}(\beta_i^0)) - n\ln(\widehat{k}_{LIML}), \tag{16}$$

where $\widetilde{k}_{LIML}(\beta_i^0)$ is computed from the concentrated log-likelihood function imposing the restriction $\beta_i = \beta_i^0$, and $\widehat{k}_{LIML}$ is computed from the unconstrained log-likelihood function (14). The restricted LIML estimator of $\beta_{\sim i}$, $\tilde{\boldsymbol{\beta}}_{LIML,\sim i}(\beta_i^0)$, minimizes the restricted variance ratio

$$k(\boldsymbol{\beta}_{\sim i}) = \frac{(\mathbf{y} - \mathbf{X}_i\beta_i^0 - \mathbf{X}_{\sim i}\boldsymbol{\beta}_{\sim i})'(\mathbf{y} - \mathbf{X}_i\beta_i^0 - \mathbf{X}_{\sim i}\boldsymbol{\beta}_{\sim i})}{(\mathbf{y} - \mathbf{X}_i\beta_i^0 - \mathbf{X}_{\sim i}\boldsymbol{\beta}_{\sim i})'\mathbf{Q}_{\mathbf{Z}}(\mathbf{y} - \mathbf{X}_i\beta_i^0 - \mathbf{X}_{\sim i}\boldsymbol{\beta}_{\sim i})} \tag{17}$$

5.2 Concentrated AR statistic

Stock and Wright (2000) consider a concentrated Anderson-Rubin type statistic for testing $H_0 : \beta_i = \beta_i^0$ in a GMM framework. In the linear IV regression, this statistic has the form

$$AR(\beta_i^0) = \frac{(\mathbf{y} - \mathbf{X}_i\beta_i^0 - \mathbf{X}_{\sim i}\tilde{\beta}_{\sim i}(\beta_i^0))'\mathbf{P}_{\mathbf{Z}}(\mathbf{y} - \mathbf{X}_i\beta_i^0 - \mathbf{X}_{\sim i}\tilde{\beta}_{\sim i}(\beta_i^0))}{(\mathbf{y} - \mathbf{X}_{\sim i}\beta_i^0 - \mathbf{X}_{\sim i}\tilde{\beta}_{\sim i}(\beta_i^0))'\mathbf{Q}_{\mathbf{Z}}(\mathbf{y} - \mathbf{X}_i\beta_i^0 - \mathbf{X}_{\sim i}\tilde{\beta}_{\sim i}(\beta_i^0))/(n-k)} \tag{18}$$

where $\tilde{\beta}_{\sim i}(\beta_i^0)$ denotes either the IV or LIML estimate of $\beta_{\sim i}$ imposing $\beta_i = \beta_{\sim i}^0$. The restricted LIML estimate minimizes (17), whereas the restricted IV estimate has the analytic form

$$\tilde{\beta}_{\sim i,IV}(\beta_i^0) = (\mathbf{X}_{\sim i}'\mathbf{P}_{\mathbf{Z}}\mathbf{X}_{\sim i})^{-1}\mathbf{X}_{\sim i}'\mathbf{P}_{\mathbf{Z}}(\mathbf{y} - \mathbf{X}_i\beta_i^0)$$

When $\tilde{\beta}_{\sim i}(\beta_i^0) = \tilde{\beta}_{\sim i,IV}(\beta_i^0)$, we use $AR_{IV}(\beta_i^0)$; when $\tilde{\beta}_{\sim i}(\beta_i^0) = \tilde{\beta}_{\sim i,LIML}(\beta_i^0)$, we use $AR_{LIML}\left(\beta_i^0\right)$. Under $H_0 : \beta_i = \beta_i^0$ and the assumption that $\beta_{\sim i}$ is well identified, Stock and Wright's results imply that $AR(\beta_i^0)$ is asymptotically distributed $\chi^2(q - k + 1)$.

5.3 Concentrated K statistic

Kleibergen (2000) proposes a concentrated version of his joint K-statistic (see Kleibergen (2002) for details) for testing the individual hypothesis $H_0 : \beta_i = \beta_i^0$, which has the form

$$K(\beta_i^0) = \frac{(\mathbf{y} - \mathbf{X}_i\beta_i^0 - \mathbf{X}_{\sim i}\tilde{\beta}_{\sim i,LIML}(\beta_i^0))'\mathbf{P}_{\mathbf{W}(\beta_i^0)}(\mathbf{y} - \mathbf{X}_i\beta_i^0 - \mathbf{X}_{\sim i}\tilde{\beta}_{\sim i,LIML}(\beta_i^0))}{(\mathbf{y} - \mathbf{X}_i\beta_i^0 - \mathbf{X}_{\sim i}\tilde{\beta}_{\sim i,LIML}(\beta_i^0))'\mathbf{Q}_{\mathbf{Z}}(\mathbf{y} - \mathbf{X}_i\beta_i^0 - \mathbf{X}_{\sim i}\tilde{\beta}_{\sim i,LIML}(\beta_i^0))/(n-k)} \tag{19}$$

where $\tilde{\beta}_{\sim i,LIML}(\beta_i^0)$ is the LIML estimate of $\beta_{\sim i}$ imposing $\beta_i = \beta_i^0$,

$$\begin{aligned}
\mathbf{W}(\beta_i^0) &= [\mathbf{Q}_{\mathbf{Z}\widetilde{\Gamma}_{LIML,\sim i}(\beta_i^0)} - \mathbf{Q}_Z](\mathbf{X}_i - \mathbf{H}(\beta_i^0)\mathbf{S}_{22}(\beta_i^0)^{-1}\mathbf{S}_{21}(\beta_i^0)) \\
\mathbf{H}(\beta_i^0) &= \begin{pmatrix} \mathbf{X}_2 & \mathbf{y} - \mathbf{X}_i\beta_i^0 \end{pmatrix} \\
\mathbf{S}_{22}(\beta_i^0) &= \frac{1}{n-q}\mathbf{H}(\beta_i^0)'\mathbf{Q}_Z\mathbf{H}(\beta_i^0) \\
\mathbf{S}_{21}(\beta_i^0) &= \frac{1}{n-q}\mathbf{H}(\beta_i^0)'\mathbf{Q}_Z\mathbf{X}_1
\end{aligned}$$

and $\tilde{\mathbf{\Gamma}}_{LIML,\text{-}i}(\beta_i^0)$ is the LIML estimate of $\mathbf{\Gamma}_{\text{-}i}$ imposing $\beta_i = \beta_i^0$. Under $H_0 : \beta_i = \beta_i^0$ and the assumption that $\beta_{\text{-}i}$ is well identified, Kleibergen shows that $K(\beta_i^0)$ is asymptotically distributed $\chi^2(1)$.

5.4 Projected AR statistic

Let $\boldsymbol{\beta} = (\beta_i, \boldsymbol{\beta}'_{\text{-}i})'$ and consider testing the hypotheses

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^0$$

using the Anderson-Rubin statistic

$$AR(\boldsymbol{\beta}^0) = \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0)' P_{\mathbf{Z}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0)/k}{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0)' Q_{\mathbf{Z}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0)/(n-k)} \tag{20}$$

Under the null, Staiger and Stock (1997) show that $AR(\boldsymbol{\beta}^0)$ is asymptotically $\chi^2(q)$ regardless of the quality of the instruments. A confidence set for $\boldsymbol{\beta}$ with level $1-\alpha$ can be obtained by inverting $AR(\boldsymbol{\beta}^0)$ giving

$$C_\beta(\alpha) = \{\boldsymbol{\beta}^0 : AR(\boldsymbol{\beta}^0) \leq \chi^2_\alpha(m)\} \tag{21}$$

where $\chi^2_\alpha(m)$ is the $1-\alpha$ quantile of the chi-square distribution with m degrees of freedom.

If we are interested in making inference on β_i or some function of $\boldsymbol{\beta}$, say $\eta = g(\boldsymbol{\beta})$, then a Scheffe-type projection method as described in Dufour (1997), Wang and Zivot (1998) and Dufour and Jasiak (2001) can be employed to make valid inference. They show that a confidence set defined by

$$C_\eta(\alpha) = \{\eta_0 : \eta_0 = g(\boldsymbol{\beta}) \text{ for some } \boldsymbol{\beta} \in C_\beta(\alpha)\}$$

has asymptotic coverage probability at least $1-\alpha$. If $g(\boldsymbol{\beta}) = \beta_i$ the set $C_\eta(\alpha)$ is simply the projection of $C_\beta(\alpha)$ on the β_i axis. Dufour and Taamouti (2003) give an analytic formula for computing projection-based confidence sets for linear functions $g(\boldsymbol{\beta}) = \mathbf{w}'\boldsymbol{\beta}$ based on (21).

6 Asymptotic Properties Under Weak Instruments

In this section, we evaluate the asymptotic properties under weak instruments of a subset of the competing statistics for making inference on individual coefficients in the IV regression. To simplify the asymptotic analysis, we restrict our attention to the IV regression model (1) - (3) with two right-hand side endogenous variables so that $\boldsymbol{\beta} = (\beta_1, \beta_2)'$.

6.1 Weak Instrument Cases

We follow Staiger and Stock (1997) and Wang and Zivot (1998) and characterize weak instruments using a local-to-zero framework. With multiple endogenous variables, the characterization of weak instruments becomes a bit complicated because the instruments $\mathbf{Z}$ may be weak for the coefficients on all of the endogenous variables, or for only a subset of the coefficients. Therefore, we consider the following weak instrument (WI) cases:

1. Weak Instrument Case I: $\mathbf{\Gamma} = [\mathbf{\Gamma}_1, \mathbf{\Gamma}_2] = n^{-1/2}\mathbf{G}$, where $\mathbf{G}$ is a fixed $q \times 2$ matrix of full rank. This case is considered by Staiger and Stock (1997) and Wang and Zivot (1998), and specifies that instruments are weak for both structural coefficients.

2. Weak Instrument Case II: $\mathbf{\Gamma}_1 = n^{-1/2}\mathbf{g}_1$, where $\mathbf{g}_1$ is a fixed $q \times 1$ vector, and $\mathbf{\Gamma}_2$ is a non-zero fixed $q \times 1$ vector linearly independent of $\mathbf{\Gamma}_1$. This case specifies that instruments are weak for β_1 but not for β_2.

3. Weak Instrument Case III: $\mathbf{\Gamma}_1 = a\mathbf{\Gamma}_2 + n^{-1/2}\mathbf{g}_1$, where a is a non-zero scalar, $\mathbf{g}_1$ is a fixed $q \times 1$ vector, and $\mathbf{\Gamma}_2$ is a non-zero fixed $q \times 1$ vector linearly independent of $\mathbf{g}_1$. This case specifies that instruments are weak for both structural coefficients except for the linear combination $\alpha = a\beta_1 + \beta_2$.

6.2 Standardized Variables

The asymptotic distributions of various estimators and test statistics under the weak instrument cases defined above depend on nuisance parameters measuring the degree of endogeneity of $\mathbf{X}_1$ and $\mathbf{X}_2$, standardized multivariate normal random vectors, and standardized measures of the quality of the instruments $\mathbf{Z}$. Endogeneity is measured using the simple correlation coefficients

$$\rho_{u1} = \sigma_{u1}/(\sigma_{uu}\sigma_{11})^{1/2},\ \rho_{u2} = \sigma_{u2}/(\sigma_{uu}\sigma_{22})^{1/2}, \rho_{12} = \sigma_{12}/(\sigma_{11}\sigma_{22})^{1/2}$$

The weak instrument asymptotic distributions are functions of the standardized random vectors

$$\begin{aligned}\mathbf{z}_u &= \mathbf{M}^{-1/2}\mathbf{\Psi}_{\mathbf{Z}u}/\sigma_{uu}^{1/2}\\ \mathbf{z}_1 &= \mathbf{M}^{-1/2}\mathbf{\Psi}_{\mathbf{Z}1}/\sigma_{11}^{1/2}\\ \mathbf{z}_2 &= \mathbf{M}^{-1/2}\mathbf{\Psi}_{\mathbf{Z}2}/\sigma_{22}^{1/2}\end{aligned}$$

with $\mathbf{Z}_{\mathbf{V}}= [\mathbf{z}_1, \mathbf{z}_2]$ such that[6]

$$\begin{pmatrix} \mathbf{z}_u \\ vec(\mathbf{Z}_V) \end{pmatrix} \sim N(0, \mathbf{R} \otimes \mathbf{I}_q),\ \mathbf{R} = \begin{pmatrix} 1 & \rho_{u1} & \rho_{u2} \\ \rho_{u1} & 1 & \rho_{12} \\ \rho_{u2} & \rho_{12} & 1 \end{pmatrix}$$

Staiger and Stock (1997) use an alternative standardization such that $\rho_{12} = 0$.

Additionally, define

$$\underset{(2\times 2)}{\underline{\mathbf{\Lambda}}} = \mathbf{M}^{1/2}\mathbf{G}\mathbf{\Sigma}_{\mathbf{VV}}^{-1/2} = [\underline{\boldsymbol{\lambda}_1}, \underline{\boldsymbol{\lambda}_2}]$$

and for $i = 1, 2$

$$\begin{aligned}\boldsymbol{\lambda}_i &= \mathbf{M}^{1/2}\mathbf{g}_i/\sigma_{ii}^{1/2}\\ \eta_i &= (\boldsymbol{\lambda}_i + \mathbf{z}_i)'(\boldsymbol{\lambda}_i + \mathbf{z}_i)\\ \xi_i &= (\boldsymbol{\lambda}_i + \mathbf{z}_i)'\mathbf{z}_u\end{aligned}$$

The matrix $\underline{\mathbf{\Lambda}}'\underline{\mathbf{\Lambda}}/q$ is related to the noncentrality parameter of the limiting chi-square distribution of the Wald statistic for testing $\mathbf{\Gamma} = \mathbf{0}$ in (3), and measures the global quality of the instruments. The scalars $\boldsymbol{\lambda}_i'\boldsymbol{\lambda}_i/q$ $(i = 1, 2)$ are related to the noncentrality parameter of the limiting chi-square distribution of the Wald statistics for testing $\mathbf{\Gamma}_i = \mathbf{0}$ in (4), and measure the quality of the instruments for β_i.

[6] The correlation coefficients in the matrix $\mathbf{R}$ are not unrestricted. An analysis of the Choleski decomposition of $\mathbf{R}$ will spell out the necessary restrictions.

6.3 Asymptotics Under Weak Instrument Case I

The appendix gives the convergence results for sample moments under weak instrument case I that are used in deriving the following results. In the following, "$\Rightarrow$" denotes convergence under the Staiger-Stock weak instrument asymptotics.

Theorem 1 *Under Assumption 1 and weak instrument case I, the following results hold jointly as $n \to \infty$*

1. $$\hat{\beta}_{i,IV} - \beta_i \Rightarrow \left(\frac{\sigma_{uu}}{\sigma_{ii}}\right)^{1/2} \frac{(\boldsymbol{\lambda}_i + \mathbf{z}_i)'\mathbf{Q}_{\lambda_{-i}+z_{-i}}\mathbf{z}_u}{(\boldsymbol{\lambda}_i + \mathbf{z}_i)'\mathbf{Q}_{\lambda_{-i}+z_{-i}}(\boldsymbol{\lambda}_i + \mathbf{z}_i)} = \beta_i^*,\ i = 1, 2$$
   where $\mathbf{Q}_{\lambda_{-i}+z_{-i}} = \mathbf{I}_q - \mathbf{P}_{\lambda_{-i}+z_{-i}}$ is a random idempotent matrix of rank $q-1$ w.p.1.

2. Under the null hypothesis $H_0 : \beta_i = \beta_i^0$ $(i = 1, 2)$

$$\begin{aligned}
\tilde{\beta}_{-i,IV}(\beta_i^0) - \beta_{-i} &\Rightarrow \left(\frac{\sigma_{uu}}{\sigma_{-i-i}}\right)^{1/2} \eta_{-i}^{-1}\xi_{-i} = \beta_{-i}^*(\beta_i^0) \\
\tilde{\sigma}_{uu}(\beta_i^0) &\Rightarrow \sigma_{uu}\left\{1 + \omega_{-i}^*(\beta_i^0)\right\}
\end{aligned}$$

   where $\omega_{-i}^*(\beta_i^0) = \beta_{-i}^*(\beta_i^0)^2\left(\frac{\sigma_{-i-i}}{\sigma_{uu}}\right) - 2\beta_{-i}^*(\beta_i^0)\left(\frac{\sigma_{u-i}}{\sigma_{-i-i}}\right)$.

Proof. The proof follows directly from the results of Lemma 1 in the Appendix. ■

Result 1, first derived by Staiger and Stock (1997), shows that the IV estimate of β_i is inconsistent and converges to a random variable whose distribution depends on nuisance parameters that cannot be consistently estimated. Result 2 shows that the restricted IV estimate, $\tilde{\beta}_{-i,IV}(\beta_i^0)$, is inconsistent and converges to the random variable $\beta_{-i}^*(\beta_i^0)$ that depends on unknown nuisance parameters. As a result, the restricted residual variance estimate, $\tilde{\sigma}_{uu}(\beta_i^0)$, is also inconsistent and converges to a random variable[7]. Since the denominators of the asymptotic-t, S, AR and K statistics depend on restricted estimates of σ_{uu}, the inconsistency of $\tilde{\sigma}_{uu}(\beta_i^0)$ introduces a random variable like $\omega_{-i}^*(\beta_i^0)$ into the limiting distributions of these test statistics. For example, using Lemma 1 from the Appendix, it is straightforward to show that

$$AR(\beta_i^0) \Rightarrow \frac{\mathbf{z}_u'\mathbf{Q}_{\lambda_{-i}+z_{-i}}\mathbf{z}_u}{1 + \omega_{-i}^*(\beta_i^0)}$$

The results of Theorem 1 indicate that, if instruments are weak for all structural coefficients, asymptotically valid inference cannot be made using any of the proposed test statistics. However, asymptotically valid, but conservative, confidence sets for individual coefficients may be computed using the Dufour-Taamouti projection-AR sets. If instruments are very weak, these sets will be unbounded with probability close to the stated coverage probability.

6.4 Asymptotics Under Weak Instrument Case II

Most of the asymptotic results for estimators and test statistics to date have been based on WI case I. In this section we provide some asymptotic results for a subset of the estimators and test statistics under WI case II.

[7] Similar results may be shown to hold for the LIML estimates.

Theorem 2 *Under Assumption 1 and weak instrument case II, the following results hold as $n \to \infty$*

Part 1.

(a) $\hat{\beta}_{1,IV} - \beta_1 \Rightarrow \left(\frac{\sigma_{uu}}{\sigma_{11}}\right)^{1/2} \frac{(\boldsymbol{\lambda}_1+\mathbf{z}_1)'\mathbf{Q}_{\mathbf{M}^{1/2}\boldsymbol{\Gamma}_2}\mathbf{z}_u}{(\boldsymbol{\lambda}_1+\mathbf{z}_1)'\mathbf{Q}_{\mathbf{M}^{1/2}\boldsymbol{\Gamma}_2}(\boldsymbol{\lambda}_1+\mathbf{z}_1)} = \beta_1^{\dagger}$

(b) $\sqrt{n}(\hat{\beta}_{2,IV} - \beta_2) \Rightarrow \sigma_{uu}^{1/2} \cdot \frac{\boldsymbol{\Gamma}_2'\mathbf{M}^{1/2\prime}\mathbf{Q}_{\lambda_1+z_1}\mathbf{z}_u}{\boldsymbol{\Gamma}_2'\mathbf{M}^{1/2\prime}\mathbf{Q}_{\boldsymbol{\lambda}_1+\mathbf{z}_1}\mathbf{M}^{1/2}\boldsymbol{\Gamma}_2}$

(c) $\hat{\sigma}_{uu,IV} \Rightarrow \sigma_{uu}\{1+\omega_1^{\dagger}\}$

where $\mathbf{Q}_{\mathbf{M}^{1/2}\boldsymbol{\Gamma}_2}$ is an idempotent matrix with rank $q-1$, $\mathbf{Q}_{\boldsymbol{\lambda}_1+\mathbf{z}_1}$ is a random idempotent matrix of rank $q-1$ w.p.1, and $\omega_1^{\dagger} = \left(\beta_1^{\dagger}\right)^2 \left(\frac{\sigma_{11}}{\sigma_{uu}}\right) - 2\beta_1^{\dagger}\left(\frac{\sigma_{u1}}{\sigma_{uu}}\right)$.
Part 2. Under the null hypothesis $H_0 : \beta_1 = \beta_1^0$

(d) $t_{IV}(\beta_1^0)^2 \Rightarrow (1+\omega_1^{\dagger})^{-1}\mathbf{z}_u'\mathbf{P}_{\mathbf{A}}\mathbf{z}_u$

(e) $\tilde{\beta}_{2,i}(\beta_1^0) \Rightarrow \beta_2$, $i = IV, LIML$

(f) $\sqrt{n}(\tilde{\beta}_{2,i}(\beta_1^0) - \beta_2) \Rightarrow N(0, \sigma_{uu}(\boldsymbol{\Gamma}_2'\mathbf{M}\boldsymbol{\Gamma}_2)^{-1})$, $i = IV, LIML$

(g) $\tilde{\sigma}_{uu,i}(\beta_1^0) \Rightarrow \sigma_{uu}$, $i = IV, LIML$

(h) $LR_{LIML}(\beta_1^0) \Rightarrow \mathbf{z}_u'\mathbf{Q}_{\mathbf{M}^{1/2}\boldsymbol{\Gamma}_2}\mathbf{z}_u \sim \chi^2(1)$ for $q = 2$; $LR_{LIML}(\beta_1^0) \Rightarrow \mathbf{z}_u'\mathbf{Q}_{\mathbf{M}^{1/2}\boldsymbol{\Gamma}_2}\mathbf{z}_u - k_{LIML}^* < \chi^2(q-1)$ for $q > 2$

(i) $S^2(\beta_1^0) \Rightarrow \mathbf{z}_u'\mathbf{P}_{\mathbf{A}}\mathbf{z}_u < \mathbf{z}_u'\mathbf{Q}_{\mathbf{M}^{1/2}\boldsymbol{\Gamma}_2}\mathbf{z}_u \sim \chi^2(q-1)$,

(j) $AR_i(\beta_1^0) \Rightarrow \mathbf{z}_u'\mathbf{Q}_{\mathbf{M}^{1/2}\boldsymbol{\Gamma}_2}\mathbf{z}_u \sim \chi^2(q-1)$, $i = IV, LIML$

(k) $K(\beta_1^0) \Rightarrow \chi^2(1)$

where $\mathbf{A} = \mathbf{Q}_{\mathbf{M}^{1/2}\boldsymbol{\Gamma}_2}(\boldsymbol{\lambda}_1+\mathbf{z}_1)$.
Part 3. Under the null hypothesis $H_0 : \beta_2 = \beta_2^0$

(l) $t_{IV}(\beta_2^0)^2 \Rightarrow (1+\omega_1^{\dagger})^{-1}\mathbf{z}_u'\mathbf{P}_{\mathbf{B}}\mathbf{z}_u$

(m) $\tilde{\beta}_{1,IV}(\beta_2^0) - \beta_1 \Rightarrow \left(\frac{\sigma_{uu}}{\sigma_{11}}\right)^{1/2} \eta_1^{-1}\xi_1 = \beta_1^*(\beta_2^0)$

(n) $\tilde{\sigma}_{uu}(\beta_2^0) \Rightarrow \sigma_{uu}\left\{1+\omega_1^*(\beta_2^0)\right\}$

where $\mathbf{B} = \mathbf{Q}_{\boldsymbol{\lambda}_1+\mathbf{z}_1}\mathbf{M}^{1/2}\boldsymbol{\Gamma}_2$, and $\omega_1^*(\beta_2^0) = \left(\beta_1^*(\beta_2^0)\right)^2\left(\frac{\sigma_{11}}{\sigma_{uu}}\right) - 2\beta_1^*(\beta_2^0)\left(\frac{\sigma_{u1}}{\sigma_{uu}}\right)$.
Proof. The proof follows directly from the results of Lemma 2 in the Appendix. ■

Part 1 of the theorem shows that if instruments are weak for β_1 but not for β_2, then $\hat{\beta}_{1,IV}$ is not consistent for β_1 *but* $\hat{\beta}_{2,IV}$ is consistent for β_2. This corresponds with the result from Choi and Phillips (1992) for the partially identified model. Due to the inconsistency of $\hat{\beta}_{1,IV}$, the IV estimate of the residual error variance is inconsistent and converges to a random variable. The asymptotic distribution of $\hat{\beta}_{1,IV}$ is a ratio of quadratic forms in correlated normal random vectors and is similar to the result established by Staiger and Stock (1997). When β_1 is totally unidentified, the limiting distribution reduces to the expression given in part (b) of Corollary 3.1 of Choi and Phillips (1992). The limiting distribution of $\hat{\beta}_{2,IV}$ is not normal, but may be expressed as a mixture-normal distribution using arguments from Staiger and Stock (1997). Conditional on $\mathbf{z}_1$, the asymptotic distribution of $\hat{\beta}_{2,IV}$ is normal with mean zero and variance

$$\sigma_{uu}\left(\mathbf{\Gamma}_2'\mathbf{M}^{1/2\prime}\mathbf{Q}_{\boldsymbol{\lambda}_1+\mathbf{z}_1}\mathbf{M}^{1/2}\mathbf{\Gamma}_2\right)^{-1}$$

When β_1 totally unidentified the limiting distribution reduces to the expression given in part (a) of Corollary 3.1 of Choi and Phillips (1992).

Part 2 shows that standard methods of inference are not valid for β_1. Valid inference may be performed on β_1 using $S^2(\beta_1^0), LR(\beta_1^0), AR_{IV}(\beta_1^0), AR_{LIML}(\beta_1^0)$ or $K(\beta_1^0)$. The reason for this is that the residual error variance σ_{uu} may be consistently estimated when $\beta_1 = \beta_1^0$ and instruments are strong for β_2. The limiting distributions of $AR_{IV}(\beta_1^0), AR_{LIML}(\beta_1^0)$ are $\chi^2(q-1)$, whereas the limiting distributions of $S^2(\beta_1^0)$ and $LR(\beta_1^0)$ are bounded from above by $\chi^2(q-1)$. Consequently, conservative inference based on $S^2(\beta_1^0)$ and $LR(\beta_1^0)$ may be performed using $\chi^2(q-1)$ critical values. Since the limiting distribution of Kleibergen's $K(\beta_1^0)$ statistic is $\chi^2(1)$, it has an apparent power advantage when q is large over the other statistics whose limiting distributions are $\chi^2(q-1)$ or are bounded by $\chi^2(q-1)$. The Monte Carlo results in section 8, however, show that no statistic dominates in terms of power.

Part 3 of the Theorem implies, surprisingly, that no method, except for the projection AR confidence sets, provides valid inference for β_2. This is due to fact that, under the null $\beta_2 = \beta_2^0$, we cannot remove the effects of weak instruments for β_1. For example, using straightforward calculations, it can be shown that the denominator of $AR_{IV}(\beta_2^0)$ converge to $\sigma_{uu}\left\{1+\omega_1^*(\beta_2^0)\right\}$ instead of σ_{uu}. In effect, since we cannot consistently estimate σ_{uu} when $\beta_2 = \beta_2^0$ and instruments are weak for β_1, we cannot get asymptotically pivotal statistics for testing $\beta_2 = \beta_2^0$. If $\hat{\sigma}_{uu}$ is a consistent estimator of σ_{uu}, then straightforward conditioning arguments may be used to show that $t_{IV}(\beta_2^0) \Rightarrow N(0,1)$ so that valid inference on β_2 may be performed using standard methods.

7 Confidence Regions

An asymptotically valid confidence set for the scalar β_i with level $1-\alpha$ based on inverting the statistic $T(\beta_i^0)$ is defined by

$$C_{\beta_i}(\alpha) = \{\beta_i^0 : T(\beta_i^0) \leq c_\alpha\} \tag{22}$$

where c_α is the $1-\alpha$ quantile of the limiting distribution of $T(\beta_i^0)$. Computing the set (22) requires finding the values of β_i^0 such that $T(\beta_i^0) < c_\alpha$. In general, this process requires a numerical search. However, utilizing the insights of Dufour (1997), Zivot, Startz and Nelson (1998) and Dufour and Jasiak (2001), if the inequality $T(\beta_i^0) \leq c_\alpha$ may be re-written as a quadratic inequality

$$a\left(\beta_i^0\right)^2 + b\beta_i^0 + c \leq 0, \tag{23}$$

where values of a, b, and c depend on the data and c_α, then the confidence regions defined by (22) have convenient closed form expressions and may take one of four shapes: a familiar connected interval of the form (β_i^L, β_i^H); the union of two rays $(-\infty, \beta_i^L) \cup (\beta_i^H, \infty)$; the entire real line; or the empty set. In the appendix, we show that the inequality $T(\beta_i^0) \leq c_\alpha$ may be expressed in the form (23) for $S(\beta_i^0)^2$ and $AR_{IV}(\beta_i^0)$, but not for $AR_{LIML}(\beta_i^0), LR(\beta_i^0)$ and $K(\beta_i^0)$. Consequently, easy to compute analytic formulas for the confidence sets based on $S(\beta_i)^2$ and $AR_{IV}(\beta_i^2)$ are readily available.

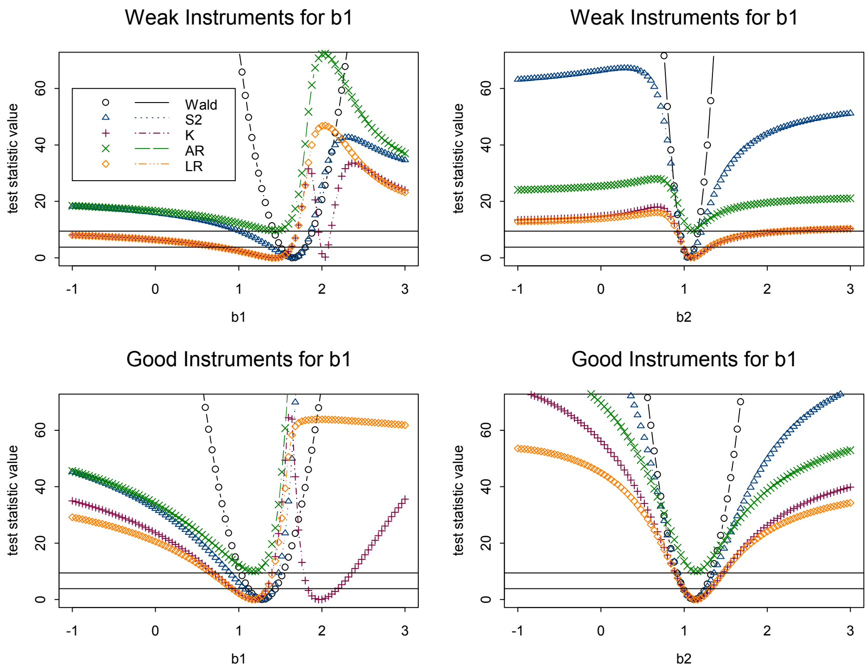

Figure 1: Test statistic values as a function of β_i^0 from Monte Carlo realization under weak instrument case II.

Figure 1 illustrates the shapes of typical confidence sets based on inverting the statistics $t_{IV}(\beta_i^0)^2$, $S(\beta_i^0)^2$, $AR_{LIML}(\beta_i^0)$, $LR(\beta_i^0)$ and $K(\beta_i^0)$ in the presence of weak and good instruments. The plot of $AR_{IV}(\beta_i^0)$ is almost identical to $AR_{LIML}(\beta_i^0)$ and so is omitted. The top panel of Figure 1 shows plots of these statistics, as functions of β_i^0, for a particular Monte Carlo realization of the IV regression model (1) - (3) with $k = 2$ and $q = 5$ in which $\beta_1 = 1$ is weakly identified and highly endogenous and $\beta_2 = 1$ is well identified and minimally endogenous. The bottom panel of Figure 1 shows plots of the statistics for a Monte Carlo realization in which both β_1 and β_2 are well identified[8]. Ninety five percent confidence intervals are the values of β_i^0 such that the test statistic in question lies below the appropriate critical value. These confidence intervals are summarized in Table 1, along with the projection-based confidence set determined from the joint AR statistic (20). Based on the asymptotic results in the previous section, $c_{.05} = \chi^2_{.05}(1) = 3.84$ is the appropriate critical value for $K(\beta_1)$ and, potentially for $S^2(\beta_1^0)$ and $LR(\beta_1^0)$ as well. This critical value is the lower horizontal line on the plots. The critical value $c_{.05}(4) = \chi^2_{.05}(4) = 9.481$ is appropriate for $AR_{LIML}(\beta_1^0)$ and serves as a conservative critical value for $S^2(\beta_1^0)$ and $LR(\beta_1^0)$. This value is upper horizontal line on the plots. There is no valid critical value for $t_{IV}(\beta_1^0)^2$. For the statistics testing $\beta_2 = 1$, the chi-square critical values are not correct[9]. However, for illustrative purposes, these critical values are used when forming the confidence sets.

[8]In particular, for the top panel, the design is as described in section 8 for weak instrument case II with $\gamma_{11} = 0.2236$ and $\gamma_{22} = 1.1180$. For the bottom panel, $\gamma_{11} = \gamma_{22} = 1.1180$. For both panels, $\rho_{u1} = 0.99$ and $\rho_{u2} = 0.1$

[9]The Monte Carlo results in the next section indicate that the chi-square critical values often produce reasonable results.

	Weak Instruments		Good Instruments	
	β_1	β_2	β_1	β_2
t_{IV}^2	$[1.52, 1.80]$	$[1.00, 1.12]$	$[1.12, 1.44]$	$[1.00, 1.24]$
S_1^2	$[1.39, 1.80]$	$[0.99, 1, 14]$	$[1.08, 1.42]$	$[0.99, 1.26]$
S_4^2	$[1.00, 1.87]$	$[0.95, 1.20]$	$[0.89, 1.48]$	$[0.92, 1.36]$
LR_1	$[0.76, 1.62]$	$[1.00, 1.30]$	$[0.92, 1.32]$	$[1.00, 1.32]$
LR_4	$[-\infty, 1.70]$	$[0.92, 2.22]$	$[0.68, 1.40]$	$[0.90, 1.46]$
K	$[0.76, 1.62] \cup [2.00, 2.06]$	$[1.00, 1.30]$	$[0.94, 1.32] \cup [1.84, 2.18]$	$[1.00, 1.32]$
AR_{LIML}	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
AR_{proj}	$[1.17, 1.56]$	$[1.03, 1.18]$	$[1.07, 1.27]$	$[1.07, 1.23]$
Notes: S_1^2 and LR_1 are computed using $\chi^2_{.05}(1)$; S_4^2 and LR_4 are computed using $\chi^2_{.05}(4)$				

Table 1: 95 Percent Confidence Sets from Simulated Data

Consider first the plots of the statistics as functions of β_1 for the weak instrument case given in the upper left panel of Figure 1. The $t_{IV}(\beta_1^0)^2$ statistic plots as a parabola and the incorrect 95 percent confidence region is a reasonably tight closed interval around $\hat{\beta}_{1,IV} = 1.66$, $(1.52, 1.80)$, a region which excludes the true value. The shapes of the other statistics are generally similar to each other but display some important differences. All statistics eventually asymptote to a finite value. $AR_{LIML}(\beta_1)$ always lies above the critical value 9.481 and so produces an empty confidence set. This results from the data rejecting the overidentifying restrictions when one imposes $\beta_1 = 1$. Unlike the $AR_{LIML}(\beta_1)$ set, the projected-AR confidence set is a closed interval but excludes $\beta_1 = 1$. The statistics $LR(\beta_1^0)$ and $K(\beta_1^0)$ follow each other very closely except for a region around $\beta_1^0 = 2$. In fact $K(\beta_1^0)$ drops to zero at the point where $AR_{LIML}(\beta_1)$ attains it maximum value, a phenomenon noted by Kleibergen (2001). As a result, the 95 percent confidence set based on $K(\beta_1^0)$ consists of two disconnected intervals with the true value $\beta_1 = 1$ contained in the first interval. The statistics $LR(\beta_1^0)$ and $S^2(\beta_1^0)$ behave similarly. Both statistics produce closed confidence sets using the $\chi^2(1)$ critical value, with $\beta_1 = 1$ covered by the LR set but not by the S^2 set. When the conservative $\chi^2(4)$ critical value is used, the lower limit of the LR set becomes unbounded whereas the S^2 remains closed. With the $\chi^2(4)$ critical value, the S^2 set contains $\beta_1 = 1$.

Next, consider the statistics for testing $\beta_1 = 1$ when instruments are good for both coefficients. All confidence sets except for those based on $AR_{LIML}(\beta_1^0)$ and $K(\beta_1^0)$ are closed, and the right end-points of the sets are similar. Due to the high degree of endogeneity of X_1, the $t_{IV}(\beta_1^0)^2$ confidence set still excludes $\beta_1 = 1$. However, $\beta_1 = 1$ is covered by the LR and K confidence sets using the $\chi^2(1)$ critical value, and is covered by the S^2 confidence set using the $\chi^2(4)$ critical value.

Now consider the test statistics as functions of β_2. For both weak and good instruments for β_1, all confidence sets, except those based on $AR_{LIML}(\beta_2^0)$, are closed. The AR_{LIML} set for β_2^0 is empty, and the projection AR set does not cover the value $\beta_2 = 1$. The remaining sets are very similar to each other and cover the true value $\beta_2 = 1$. The sets based on $t_{IV}(\beta_1^0)^2$ and $S^2(\beta_2^0)$ using the $\chi^2(1)$ critical value are very similar and have the smallest width. The sets based on $LR(\beta_2^0)$ and $K(\beta_2^0)$ using the $\chi^2(1)$ critical value are essentially identical.

8 Finite Sample Properties Under Weak Instruments

In this section, we evaluate the finite sample properties of the competing statistics for making inference on individual structural coefficients using a comprehensive set of Monte

Carlo experiments. Several authors have considered Monte Carlo designs for IV regressions with weak instruments. The main papers are Choi and Phillips (1992), Hall, Rudebusch and Wilcox (1997), Shea (1997), Staiger and Stock (1997), Zivot, Startz and Nelson (1998), Dufour and Khalaf (1998), Blomquist and Dahlberg (1999), Flores-Lagunes (2000), Kleibergen (2000, 2002), Taamouti (2001) and Hahn and Inuoe (2002). Most of the Monte Carlo studies that have focussed on the performance of estimation and inference methods in the presence of weak instruments are based on designs with a single right-hand-side endogenous variable. Choi and Phillips (1992), Flores-Lagunes (2000) and Kleibergen (2000) have considered designs with two right-hand-side endogenous variables, and the results from these papers indicate that it may be misleading to extrapolate the results from the one variable case to the multiple variable case. Much more work is needed in the multiple right-hand-side variable case and we provide the most comprehensive study to date.

8.1 Monte Carlo Designs for Multiple Endogenous Variables

The Staiger-Stock weak instrument asymptotics show that the distributions of IV estimators and test statistics depend on three key nuisance parameters: (1) the degree of endogeneity as measured by the correlation coefficients ρ_{u1}, ρ_{u2} and ρ_{12}; (2) the number of instruments, q; and (3) the relevance of the instruments as measured by $\underline{\mathbf{\Lambda}}'\underline{\mathbf{\Lambda}}/q$. Instruments are irrelevant when $\underline{\mathbf{\Lambda}}'\underline{\mathbf{\Lambda}}/q = \mathbf{0}$. For one right-hand side endogenous variable, Staiger and Stock's simulation experiments reveal that instruments are essentially weak when $0 < \underline{\mathbf{\Lambda}}'\underline{\mathbf{\Lambda}}/q < 10$. Instruments become pretty good when $\underline{\mathbf{\Lambda}}'\underline{\mathbf{\Lambda}}/q > 10$. In the multiple right-hand-side endogenous variable case, weak instruments are characterized by the minimum eigenvalue of $\underline{\mathbf{\Lambda}}'\underline{\mathbf{\Lambda}}/q$. In addition, Staiger and Stock show that the performance of standard inference methods is the worst in models with many irrelevant instruments (large value of q and $\mathbf{\Gamma} \approx \mathbf{0}$) and very high degrees of endogeneity.

The data generating process (DGP) for our experiments is similar to the designs in Flores-Lagunes (2000) and has the form

$$\begin{aligned}
\mathbf{y} &= \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \mathbf{u} \\
\mathbf{X}_1 &= \mathbf{Z}_1\gamma_{11} + \mathbf{Z}_2\gamma_{12} + \mathbf{v}_1 \\
\mathbf{X}_2 &= \mathbf{Z}_1\gamma_{21} + \mathbf{Z}_2\gamma_{22} + \mathbf{v}_2
\end{aligned}$$

where $\beta_1 = \beta_2 = 1$, and the covariates are generated following[10]:

$$\begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix} \sim \text{iid } N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

$$\begin{bmatrix} u_t \\ v_{1t} \\ v_{2t} \end{bmatrix} \sim \text{iid } N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{u1} & \rho_{u2} \\ \rho_{u1} & 1 & 0 \\ \rho_{u2} & 0 & 1 \end{pmatrix} \right)$$

The main difference from the Flores-Lagunes designs is that the instruments are mutually uncorrelated with unit variances. In this design,

$$\begin{aligned}
\mathbf{\Sigma_{VV}} &= \mathbf{I}_2, \mathbf{Q} = \mathbf{I}_q, \boldsymbol{\lambda} = \mathbf{C} = \mathbf{\Gamma}/\sqrt{n} \\
\boldsymbol{\rho} &= \mathbf{\Sigma}_{Vu} = (\rho_{u1}, \rho_{u2})' \\
\underline{\mathbf{\Lambda}}'\underline{\mathbf{\Lambda}} &= \mathbf{C}'\mathbf{C} = \mathbf{n}\mathbf{\Gamma}'\mathbf{\Gamma}
\end{aligned}$$

We set $n = 100$ and consider designs for which $\boldsymbol{\rho} = (0.5, 0.5)', (-0.5, 0.5)', (0.1, 0.99)', (0.99, 0.1)'$. We consider the following designs for Weak Instrument Cases I, II and III:

[10] The correlation coefficients must satisfy $\rho_{ue_1}^2 + \rho_{ue_2}^2 < 1$ for the error covariance matrix to be positive definite.

- Weak Instrument Case I (Staiger-Stock): $\gamma_{12} = \gamma_{21} = 0$. Set γ_{11} and γ_{22} such that

$$\underline{\mathbf{\Lambda}}'\underline{\mathbf{\Lambda}}/q = \left(\frac{n}{q}\right)\begin{pmatrix} \gamma_{11}^2 & 0 \\ 0 & \gamma_{22}^2 \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix}, \ \alpha = 0, 1, 10$$

  This implies that

$$\gamma_{11} = \gamma_{22} = \gamma = \left(\frac{\alpha \cdot q}{n}\right)^{1/2}$$

  Here instruments are weak for both β_1 and β_2. The following table summarizes the values of γ for $q = 2, 5, 20$ and $n = 100$

q/α	0	0.25	1
2	0	0.0707	0.1414
5	0	0.1118	0.2236
20	0	0.2236	0.4472

- Weak Instrument Case II: $\gamma_{12} = \gamma_{21} = 0$. Set γ_{11} and γ_{22} such that

$$\underline{\mathbf{\Lambda}}'\underline{\mathbf{\Lambda}}/q = \left(\frac{n}{q}\right)\begin{pmatrix} \gamma_{11}^2 & 0 \\ 0 & \gamma_{22}^2 \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ 0 & 25 \end{pmatrix}, \ \alpha = 0, 1, 10$$

  This implies that

$$\gamma_{11} = \left(\frac{\alpha \cdot q}{n}\right)^{1/2}, \ \gamma_{22} = \left(\frac{25 \cdot q}{n}\right)^{1/2}$$

  Here instruments are weak for β_1 but not for β_2. The following table summarizes the values of γ_{11} for $q = 2, 5, 20$ and $n = 100$

q/α	0	1	10
2	0	0.1414	0.4472
5	0	0.2236	0.7071
20	0	0.4472	1.4142

The values of γ_{22}, for $q = 2, 5$ and 20, are 0.7071, 1.1180 and 2.2361, respectively.

- Weak Instrument Case III: $\gamma_{21} = 0, \gamma_{11} = c_{11}/\sqrt{n}, \gamma_{12} = \gamma_{22}$ so that

$$\mathbf{\Gamma}_1 = \begin{pmatrix} c_{11}/\sqrt{n} \\ \gamma_{22} \end{pmatrix}, \ \mathbf{\Gamma}_2 = \begin{pmatrix} 0 \\ \gamma_{22} \end{pmatrix}$$

  and

$$\underline{\mathbf{\Lambda}}'\underline{\mathbf{\Lambda}}/q = \left(\frac{n}{q}\right)\begin{pmatrix} \gamma_{22}^2 + c_{11}^2/n & \gamma_{22}^2 \\ \gamma_{22}^2 & \gamma_{22}^2 \end{pmatrix} = \begin{pmatrix} a+b & b \\ b & b \end{pmatrix}, \ b = 25, a = 0, 2, 10.$$

  where $a = c_{11}^2/n$ and $b = \gamma_{22}^2$. Here $c_{11} = \sqrt{q \cdot a}$. As $a \to 0$, $rank(\boldsymbol{\lambda}'\boldsymbol{\lambda}/q) \to 1$ provided $b \neq 0$. The weak instrument asymptotics will be influenced by the minimum eigenvalue of $\boldsymbol{\lambda}'\boldsymbol{\lambda}/q$.

8.2 Results

There are 36 different designs for each of the weak instrument cases. For each design, 10,000 simulations are performed and the Monte Carlo experiments for each design use the same random numbers to eliminate simulation noise between experiments. We compute the unrestricted IV and LIML estimates as well as the corresponding estimates that impose the restriction $\beta_i = 1$. We also compute the IV and LIML t-statistics, the concentrated AR statistics using the restricted IV and LIML estimates (18), the concentrated K statistic (19), the projection-based confidence sets based on the joint AR statistic, and the S statistic (13) for testing the individual hypothesis

$$H_0 : \beta_i = 1, \ i = 1, 2$$

The results of a subset of the experiments are summarized in Tables 2 - 4 and described below. Power results are only reported for a subset of Weak Instrument Case II designs.

8.2.1 Weak Instrument Case I

In this design, the instruments are weak for both β_1 and β_2 in a symmetric way. The asymptotic results indicate that none of the tests considered are asymptotically pivotal. The empirical sizes of the tests for individual coefficients are summarized in Table 2. For the just identified models, the IV and LIML t-statistics for β_i $(i = 1, 2)$ have size distortions that increase with ρ_i. The LR, AR_{LIML} and K statistics for β_i are well behaved and appear to be very close numerically. This result was also noted by Kleibergen (2000). The S and AR_{IV} statistics for β_i are nearly identical and are slightly oversized when $\rho_{\tilde{}i}$ is large. The tests for β_i based on the projected AR confidence sets are very conservative. Most of the confidence sets are unbounded, even for moderately strong instruments.

For the overidentified models, the IV t-statistics and the S-statistics can be severely size distorted, especially for highly overidentified models with very weak instruments and $\rho_i \approx 1$. The LIML-t, LR and AR_{IV} statistics are also size distorted but to a lesser degree. Interestingly, the size distortions of $AR_{IV}(\beta_i^0)$ and $LR(\beta_i^0)$ are more sensitive to the degree of endogeneity of $X_{\tilde{}i}$ than the degree of endogeneity of X_i. In contrast, there is surprisingly little size distortion in the $AR_{LIML}$and K statistics for β_i $(i = 1, 2)$. As with the AR_{IV} and LR statistics, the size distortions of $AR_{LIML}(\beta_i^0)$ and $K(\beta_i^0)$ are larger for higher values of $\rho_{\tilde{}i}$. The tests for β_i based on the projected AR confidence sets are again very conservative when instruments are very weak, but are nearly correctly sized when instruments become stronger. For instruments of moderate strength, roughly thirty five percent of the confidence sets are closed.

8.2.2 Weak Instrument Case II: Size

In this design, the instruments are weak for β_1 but not for β_2. The asymptotic results indicate that the S, LR, AR and K statistics have asymptotically pivotal or bounded pivotal distributions in this case. The empirical sizes for the individual coefficient tests are summarized in Table 3. For the just identified models, the IV and LIML t-statistics are sized distorted for β_1 when ρ_1 is high. The S, LR, AR and K statistics for β_1 are properly sized, as predicted by theory. Interestingly, when instruments are very weak, most of the test statistics for β_2 are undersized. However, $S(\beta_2^0)$ and $AR_{IV}(\beta_2^0)$ are slightly over-sized when $\rho_1 \approx 1$ and β_1.

For the overidentified models, the IV and LIML t-statistics, S-statistic and LR statistic for β_1 can be severely size distorted using the $\chi^2(1)$ critical value. The S and LR statistics become size controlled when using the $\chi^2(q-1)$ critical value as predicted by theory. The AR and K statistics for β_1 have stable sizes, as predicted by theory. However, $AR_{IV}(\beta_2^0)$ is over sized when $\rho_1 \approx 1$ and β_1 is weakly identified.

The tests for β_i $(i = 1, 2)$ based on the projected AR confidence sets are conservative, with the tests for β_2 much more conservative than the tests for β_1. The empirical size of the tests for β_1 does not vary with the quality of the instruments whereas the empirical size of the tests for β_2 becomes closer to the nominal size as the instruments for β_1

improve. When instruments are weak for β_1, most of the confidence sets for β_1 and β_2 are unbounded. When instruments become moderately strong for β_1, the percentage of closed confidence sets approaches the nominal coverage rates.

8.2.3 Weak Instrument Case II: Power

Three sets of power experiments are run to evaluate the tests for individual coefficients. In the first two sets, five percent tests for the null hypothesis $\beta_1 = 1$ against the alternatives $\beta_1 = 1 + \delta$, for selected values of δ, are computed for designs in which instruments are moderately strong for β_1 and for designs for which instruments are weak for β_1. In the third set, five percent tests for the null hypothesis $\beta_2 = 1$ against the alternatives $\beta_2 = 1 + \delta$ are computed in designs for which instruments are weak for β_1. In all cases size-adjusted power is computed based on $10,000$ simulations for each value of δ.

Figure 2 shows the power results for tests on β_1 for the designs in which instruments are good ($\alpha = 10$) and endogeneity is moderate for both variables ($\rho_1 = \rho_2 = 0.5$). In these designs all of the test statistics have size close to nominal size[11]. For $q = 2$, the power of the IV and LIML t-statistics are almost identical, and the power of the LR, AR_{LIML}, K and S statistics are almost identical. The IV and LIML t-statistics have higher power for $\delta > 0$ and the LR, AR_{LIML}, K and S statistics have slightly higher power for $\delta < 0$. The power of the projected AR confidence set is uniformly below the power of the other test statistics. For $q > 2$, the power of the IV and LIML t-statistics differ slightly, and the LR, K and S statistics have very similar power. Due to the larger critical values, the AR_{LIML} statistic loses power relative to the other statistics and is close to the power of the projected AR confidence set. Interestingly, the power of the K statistic is not monotonic for $\delta < 0$. The IV t-statistic has the best power for $\delta > 0$ and the S statistic has the best overall power for $\delta < 0$.

Figure 3 gives the power results for the tests on β_1 for designs in which instruments are weak ($\alpha = 1$) and endogeneity is moderate for both variables. In these designs, the IV and LIML t-statistics and LR statistics are moderately over sized whereas the other statistics are size controlled. In general, the size adjusted power of the IV and LIML t-statistics dominate the power of the other statistics. However, this is somewhat misleading since size adjustment is not possible in empirical applications. When $q = 2$, the power of the AR_{LIML}, K, LR and S statistics is identical and never exceeds 0.4. When $q > 2$ the powers of the AR_{LIML}, K, LR and S statistics diverge and the S statistic generally dominates in terms of power. The shapes of the LR, AR_{LIML} and projected AR power curves are very similar with the ranking $LR > AR_{LIML} >$ projected AR. Contrary to the results of Kleibergen (2000), the K statistic often has lower power than the AR_{LIML} statistic and the projected AR statistic.

Figure 4 gives the power results for the tests on β_2 for designs in which instruments are weak for β_1 ($\alpha = 1$) and endogeneity is moderate for both variables ($\rho_1 = \rho_2 = 0.5$). In these designs, all of the tests for β_2 are generally undersized. When $q = 2$, the power of the IV and LIML t-statistics are similar and dominate the power of the other statistics. The power curves for AR_{LIML}, K, LR and S are similar in shape and flatten out for large values of $|\delta|$, suggesting that they are inconsistent tests. The power of the projected AR confidence set is uniformly the lowest, and never rises above about 17%. For $q > 2$, all statistics exhibit higher power than when $q = 2$. The IV and LIML t-statistics and the S statistic are very similar and exhibit the highest power, followed by the LR, AR_{LIML} and projected AR statistics. For highly overidentified models, the power of the K statistic is non-monotonic in $|\delta|$ and is lower than the power of the projected AR confidence set.

[11] For all plots, sized adjusted power based on the $\chi^2(1)$ critical value is used for the S and LR statistics.

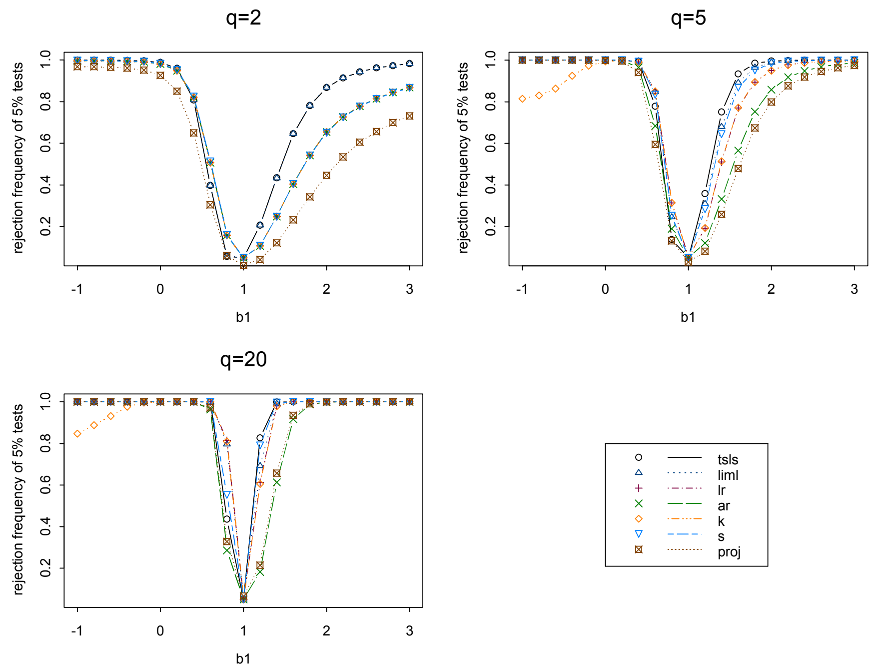

Figure 2: Size adjusted power of various test statistics. Weak instrument case II, moderate instruments for X_1.

8.2.4 Weak Instrument Case III

In this design, the instruments are such that the reduced form matrix Γ is close to being of reduced rank, so that only the linear combination $\beta_1 + \beta_2$ is identified in the extreme case of rank failure.

The tests on the individual coefficients are summarized in Table 4. When $q = 2$, the IV and LIML t-statistics and S statistics for β_1 and β_2 are undersized only when $\rho_1 = \rho_2 = 0.5$. The LR, AR_{LIML} and K statistics for β_1 and β_2 have stable size. When $q > 2$, the TSLS and LIML t-statistics, S statistics and LR statistics for β_1 and β_2 are size distorted, whereas the AR_{LIML} and K statistics do not show much size distortion. In particular, the K statistics for β_1 and β_2 are very well behaved. For all values of q, the tests for β_1 and β_2 based on the projected AR confidence sets are conservative and corresponding confidence sets are unbounded for near rank failure models.

9 Conclusion

For inference on individual structural coefficients in IV regression models with weak instruments we make the following observations. Valid inference using the statistics described in the paper is available for weakly identified coefficients as long as the remaining coefficients are well identified. The asymptotic results show that valid inference on well identified coefficients in the presence of some weakly identified coefficients is problematic. The only asymptotically valid tests for individual coefficients regardless of the quality of the instruments are based on projections of asymptotically valid tests for all coefficients.

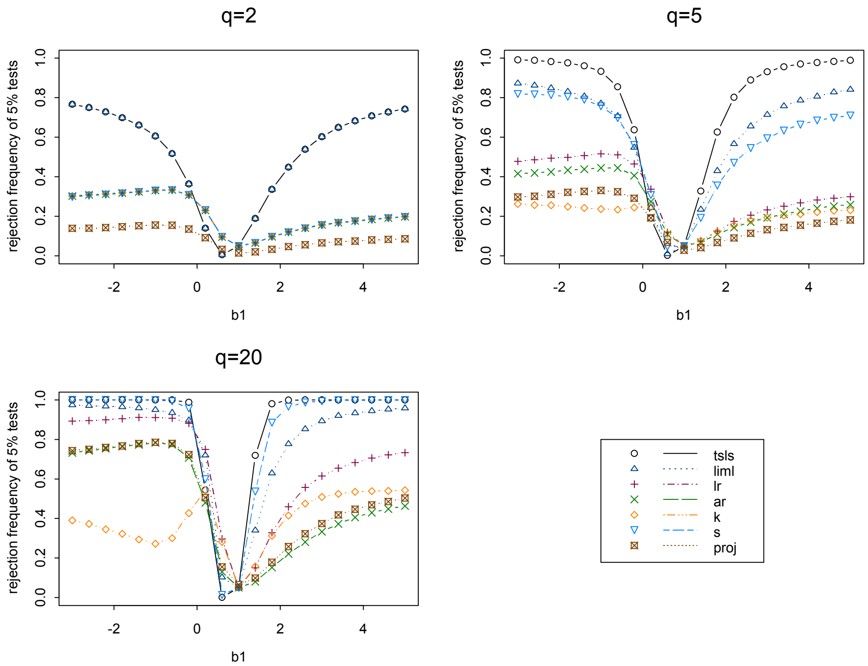

Figure 3: Size adjusted power for various test statistics. Weak instrument case II, weak instruments for X_1.

The S-statistic and some of its competitors are about as good as the IV t-statistic when instruments are good and generally much better than the IV t-statistic when instruments are weak. We suggest that these statistics ought to routinely supplement the IV t-statistic when the results of IV estimation are reported in practice. In particular confidence intervals based on the S-statistic are easily computed and ought to be routinely supplied. None of the newer statistics dominates in terms of size and power. Unfortunately, the best statistic to use depends on the characteristics of the data generating process which are unobservable in real data. In situations in which the researcher has priors about the data generating process our results in section 8 of the paper provide some guidance in picking a preferred statistic. Since our primary focus is the on the S-statistic, we would like to point out one situation where the S-statistic performs noticeably poorly; when instruments are very weak for all structural coefficients and endogeneity is very severe, then the size of the S-statistic is noticeably too large.

The current practice in empirical software is that estimated coefficients are accompanied by the estimated asymptotic standard error. This allows for trivial computation of confidence intervals and tests of point hypotheses. We recommend that this should be supplemented by the analytic confidence intervals from the S-statistic, AR_{IV} statistic and the projected joint AR statistic for a standard nominal size such as five percent. Auxiliary procedures should then be provided to test point hypotheses.

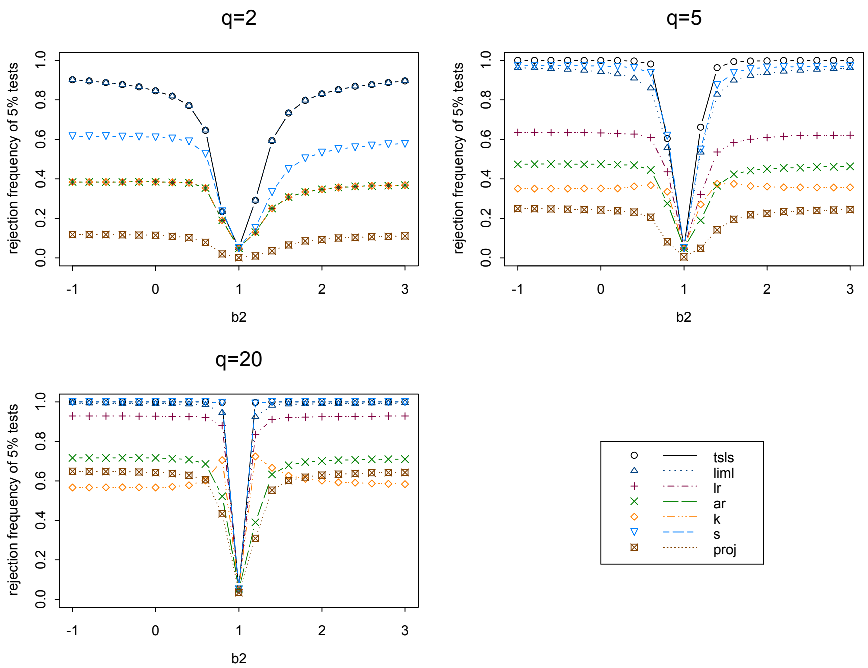

Figure 4: Size adjusted power for various test statistics. Weak instrument case II. Weak instruments for X_1.

10 Appendix A

The proofs of Theorems 1 and 2 follow from straightforward manipulations of the results in the following lemmas.

Lemma 1. *Under Assumption 1 and weak instrument case I, the following results hold jointly as* $n \to \infty$ *for* $i = 1, 2:$

1. $n^{-1}\mathbf{Z}'\mathbf{X}_i \xrightarrow{p} \mathbf{0}$
2. $n^{-1/2}\mathbf{Z}'\mathbf{X}_i \Rightarrow \sigma_{ii}^{1/2}\mathbf{M}^{1/2}(\boldsymbol{\lambda}_i + \mathbf{z}_i)$
3. $\mathbf{X}_i'\mathbf{P}_Z\mathbf{X}_i = \hat{\mathbf{\Gamma}}_i'\mathbf{Z}'\mathbf{Z}\hat{\mathbf{\Gamma}}_i \Rightarrow \sigma_{ii}(\boldsymbol{\lambda}_i + \mathbf{z}_i)'(\boldsymbol{\lambda}_i + \mathbf{z}_i) = \sigma_{ii}\eta_i$
4. $\mathbf{X}_i'\mathbf{P}_Z\mathbf{X}_{-i} = \hat{\mathbf{\Gamma}}_i'\mathbf{Z}'\mathbf{Z}\hat{\mathbf{\Gamma}}_{-i} \Rightarrow (\sigma_{ii}\sigma_{-i-i})^{1/2}(\boldsymbol{\lambda}_i + \mathbf{z}_i)'(\boldsymbol{\lambda}_{-i} + \mathbf{z}_{-i})$
5. $\mathbf{X}_i'\mathbf{P}_Z\mathbf{u} = \hat{\mathbf{\Gamma}}_i'\mathbf{Z}'\mathbf{u} \Rightarrow (\sigma_{ii}\sigma_{uu})^{1/2}(\boldsymbol{\lambda}_i + \mathbf{z}_i)'\mathbf{z}_u = (\sigma_{ii}\sigma_{uu})^{1/2}\xi_i$

Lemma 2. *Under Assumption 1 and weak instrument case II, the following results hold jointly as* $n \to \infty$

1. $n^{-1}\mathbf{X}_1'\mathbf{X}_1 \Rightarrow \sigma_{11}$
2. $n^{-1}\mathbf{X}_1'\mathbf{u} \Rightarrow \sigma_{u1}$
3. $n^{-1}\mathbf{Z}'\mathbf{X}_1 \xrightarrow{p} \mathbf{0}$
4. $n^{-1/2}\mathbf{Z}'\mathbf{X}_1 \Rightarrow \sigma_{11}^{1/2}\mathbf{M}^{1/2}(\boldsymbol{\lambda}_1 + \mathbf{z}_1)$

5. $n^{-1}\mathbf{Z}'\mathbf{X}_2 \xrightarrow{p} \mathbf{M}\boldsymbol{\Gamma}_2$
6. $\mathbf{X}_1'\mathbf{P}_Z\mathbf{X}_1 = \hat{\boldsymbol{\Gamma}}_1'\mathbf{Z}'\mathbf{Z}\hat{\boldsymbol{\Gamma}}_1 \Rightarrow \sigma_{11}(\boldsymbol{\lambda}_1 + \mathbf{z}_1)'(\boldsymbol{\lambda}_1 + \mathbf{z}_1)$
7. $n^{-1/2}\mathbf{X}_1'\mathbf{P}_Z\mathbf{X}_2 = n^{-1/2}\hat{\boldsymbol{\Gamma}}_1'\mathbf{Z}'\mathbf{Z}\hat{\boldsymbol{\Gamma}}_2 \Rightarrow \left(\sigma_{11}^{1/2}\mathbf{M}^{1/2}(\boldsymbol{\lambda}_1 + \mathbf{z}_1)\right)'\boldsymbol{\Gamma}_2$
8. $n^{-1}\mathbf{X}_2\mathbf{P}_Z\mathbf{X}_2 = n^{-1}\hat{\boldsymbol{\Gamma}}_2'\mathbf{Z}'\mathbf{Z}\hat{\boldsymbol{\Gamma}}_2 \xrightarrow{p} \boldsymbol{\Gamma}_2'\mathbf{M}\boldsymbol{\Gamma}_2$
9. $\mathbf{X}_1'\mathbf{P}_Z\mathbf{u} = \hat{\boldsymbol{\Gamma}}_1'\mathbf{Z}'\mathbf{u} \Rightarrow (\sigma_{11}\sigma_{uu})^{1/2}(\boldsymbol{\lambda}_1 + \mathbf{z}_1)'\mathbf{z}_u$
10. $n^{-1/2}\mathbf{X}_2'\mathbf{P}_Z\mathbf{u} = n^{-1/2}\hat{\boldsymbol{\Gamma}}_2'\mathbf{Z}'\mathbf{u} \Rightarrow {\sigma_{uu}}^{1/2}\boldsymbol{\Gamma}_2'\mathbf{M}^{1/2}\mathbf{z}_u$

11 Appendix B

In this appendix, we follow Zivot, Startz and Nelson (1998) and show how to compute AR_{IV} and S confidence regions in closed form, that are asymptotically valid under weak instrument case II. As stated in section 7, analytic confidence intervals may be obtained if the non-rejection region of the test statistic in question may be expressed as the quadratic inequality (23). For the statistic $AR_{IV}(\beta_i)$, the $(1-\alpha)\cdot 100\%$ confidence region (30) is determined by those values of β_1^0 such that

$$\frac{(\mathbf{y} - \mathbf{X}_i\beta_1^0 - \mathbf{X}_{\text{-}i}\tilde{\beta}_{\text{-}i,IV}(\beta_i^0))'\mathbf{P}_{\mathbf{Z}}(\mathbf{y} - \mathbf{X}_i\beta_i^0 - \mathbf{X}_{\text{-}i}\tilde{\beta}_{\text{-}i,IV}(\beta_i^0))}{(\mathbf{y} - \mathbf{X}_{\text{-}i}\beta_i^0 - \mathbf{X}_{\text{-}i}\tilde{\beta}_{\text{-}i,IV}(\beta_i^0))'\mathbf{Q}_{\mathbf{Z}}(\mathbf{y} - \mathbf{X}_i\beta_i^0 - \mathbf{X}_{\text{-}i}\tilde{\beta}_{\text{-}i,IV}(\beta_i^0))/(n-k)} \leq \chi^2_\alpha(q-1) \quad (24)$$

where $\chi^2_\alpha(q-1)$ denotes the right tail α percent quantile of the chi-square distribution with $q-1$ degrees of freedom. Since $\tilde{\boldsymbol{\beta}}_{\text{-}i,IV}(\beta_i^0) = (\mathbf{X}'_{\text{-}i}\mathbf{P}_{\mathbf{Z}}\mathbf{X}_{\text{-}i})^{-1}\mathbf{X}'_{\text{-}i}\mathbf{P}_{\mathbf{Z}}(\mathbf{y} - \mathbf{X}_i\beta_i^0)$, (24) may be re-written as

$$(\mathbf{y} - \mathbf{X}_i\beta_i^0)'\mathbf{D}(\mathbf{y} - \mathbf{X}_i\beta_i^0) \leq 0 \quad (25)$$

where $\mathbf{D} = \mathbf{B}'\mathbf{A}\mathbf{B}$ with $\mathbf{B} = \mathbf{I}_n - \mathbf{X}_{\text{-}i}(\mathbf{X}'_{\text{-}i}\mathbf{P}_{\mathbf{Z}}\mathbf{X}_{\text{-}i})^{-1}\mathbf{X}'_{\text{-}i}\mathbf{P}_{\mathbf{Z}}$ and $\mathbf{A} = \mathbf{P}_{\mathbf{Z}} - \left(\frac{\chi^2_\alpha(q-1)}{n-k}\right)\mathbf{Q}_{\mathbf{Z}}$. The inequality (25) may be put in the form (23) with $a = \mathbf{y}'\mathbf{D}\mathbf{y}$, $b = -2\mathbf{y}'\mathbf{D}\mathbf{X}_i$ and $c = \mathbf{X}_i'\mathbf{D}\mathbf{X}_i$.

To derive analytic expressions for the S-confidence region, it is more convenient to find those values of β_i^0 that satisfy $S_i^2 \leq c_\alpha^2$, which implies $\Psi_i^2 \leq \hat{\sigma}^2_{\hat{\Psi}_i}c_\alpha^2$. For a conservative confidence set, use $c_\alpha^2 = \chi^2_\alpha(q-1)$. However, as illustrated in section 8, the critical value $c_\alpha^2 = \chi^2_\alpha(1)$ is often appropriate. In order to compute analytic confidence intervals, it is useful to define $\hat{\phi}_i \equiv \hat{\Delta}_i\hat{\beta}_{i,IV}$ and then note[12]

$$\begin{aligned}
\hat{\Psi}_i^2 &= \hat{\Delta}_i^2\beta_i^{0^2} - 2\hat{\Delta}_i\hat{\phi}_i\beta_i^0 + \hat{\phi}_i^2 && (26)\\
\hat{\sigma}^2_{\hat{\Psi}_i} &= var(\hat{\Psi}_i) = \beta_i^{0^2}var(\hat{\Delta}_i) - 2\beta_i^0 cov(\hat{\Delta}_i, \hat{\phi}_i) + var(\hat{\phi}_i) && (27)
\end{aligned}$$

From equations (26) and (27), the condition $\Psi_i^2 < \hat{\sigma}^2_{\hat{\Psi}_i}c_\alpha^2$ gives the confidence region defined by the quadratic inequality

$$\left(\hat{\Delta}_i^2 - c_\alpha^2 var(\hat{\Delta}_i)\right)\beta_i^{0^2} + 2\left(-\hat{\Delta}_i\hat{\phi}_i + c_\alpha^2 cov(\hat{\Delta}_i, \hat{\phi}_i)\right)\beta_i^0 + \left(\hat{\phi}_i^2 - c_\alpha^2 var(\hat{\phi}_i)\right) \leq 0 \quad (28)$$

which is in the form (23) with $a = \hat{\Delta}_i^2 - c_\alpha^2 var(\hat{\Delta}_i)$, $b = 2\left(-\hat{\Delta}_i\hat{\phi}_i + c_\alpha^2 cov(\hat{\Delta}_i, \hat{\phi}_i)\right)$ and $c = \hat{\phi}_i^2 - c_\alpha^2 var(\hat{\phi}_i)$.

[12] Computationally we estimate $var(\hat{\Delta}_i) = \frac{\partial\hat{\Delta}_i}{\partial\hat{\lambda}'}var(\hat{\lambda})\frac{\partial\hat{\Delta}_i}{\partial\hat{\lambda}}$ and $cov(\hat{\Delta}_i, \hat{\phi}_i) = \frac{\partial\hat{\Delta}_i}{\partial\hat{\lambda}'}var(\hat{\lambda})\frac{\partial\hat{\phi}_i}{\partial\hat{\lambda}}$, where the vector of partial derivatives are computed numerically.

The non-rejection regions based on the AR_{LIML}, LR and K statistics cannot be expressed as the simple quadratic inequality (23), because the restricted LIML estimate $\tilde{\boldsymbol{\beta}}_{\sim i,LIML}(\beta_i^0)$ is a complicated nonlinear function of β_i^0.

The $S-$confidence region is defined by the roots of equation (28). Let

$$R = \sqrt{\left(-\hat{\Delta}_i\hat{\phi}_i + c_\alpha^2 cov(\hat{\Delta}_i,\hat{\phi}_i)\right)^2 - \left(\hat{\Delta}_i^2 - c_\alpha^2 var(\hat{\Delta}_i)\right)\left(\hat{\phi}_i^2 - c_\alpha^2 var(\hat{\phi}_i)\right)} \quad (29)$$

$$\{\beta_i^L,\beta_i^H\} = \frac{\left(\hat{\Delta}_i\hat{\phi}_i - c_\alpha^2 cov(\hat{\Delta}_i,\hat{\phi}_i)\right) \pm R}{\hat{\Delta}_i^2 - c_\alpha^2 var(\hat{\Delta}_i)} \quad (30)$$

If $\hat{\Delta}_i^2/var(\hat{\Delta}_i^2) > c_\alpha^2$, then the confidence region from inverting S_i^2 is the connected interval (β_i^L,β_i^H). For higher critical values, if $\hat{\Delta}_i^2/var(\hat{\Delta}_i^2) < c_\alpha^2$, the S-region is the union of two rays defined by $(-\infty,\beta_i^L)\cup(\beta_i^H,\infty)$ if R in (29) is real, and the entire real line otherwise. The corresponding confidence region is the entire real line when the argument to the root in (29) is negative, which occurs for critical values above c^*, where

$$c^* = \sqrt{\frac{\hat{\Delta}_i^2 var(\hat{\phi}_i) + \hat{\phi}_i^2 var(\hat{\Delta}_i) - 2\hat{\Delta}_i\hat{\phi}_i cov(\hat{\Delta}_i,\hat{\phi}_i)}{var(\hat{\Delta}_i)var(\hat{\phi}_i) - cov(\hat{\Delta}_i,\hat{\phi}_i)^2}}$$

In sufficiently well identified models the uncertainty about $\hat{\Delta}_i$ is negligible, so $var(\hat{\Delta}_i) << \hat{\Delta}_i^2$, $cov(\hat{\Delta}_i,\hat{\phi}_i) << \hat{\Delta}_i\hat{\phi}_i$, and $var(\hat{\phi}_i) \approx \hat{\Delta}_i^2 var(\hat{\beta}_{i,IV})$. Equations (28) through (30) reduce to

$$\left(\hat{\Delta}_i^2\right)\beta_i^{0^2} + 2\left(-\hat{\Delta}_i\hat{\beta}_{i,IV}\right)\beta_i^0 + \hat{\Delta}_i^2\left(\hat{\beta}_{i,IV}^2 - c_\alpha^2 var(\hat{\beta}_{i,IV})\right) < 0$$

$$R = \sqrt{\left(-\hat{\Delta}_i\hat{\beta}_{i,IV}\right)^2 - \left(\hat{\Delta}_i^2\right)^2\left(\hat{\beta}_{i,IV}^2 - c_\alpha^2 var(\hat{\beta}_{i,IV})\right)} = \hat{\Delta}_i^2 c_\alpha^2\sqrt{var(\hat{\beta}_{i,IV})}$$

$$\{\beta_i^L,\beta_i^H\} = \hat{\beta}_{i,IV} \pm c_{\alpha/2}\sqrt{var(\hat{\beta}_{i,IV})}$$

This establishes that S-regions approach the intervals based on the usual IV t-statistic as identification becomes certain. Evaluating the ratio of $\hat{\Psi}_i^2$ to $\hat{\sigma}_{\hat{\Psi}_i}^2$ as $\beta_i^0 \to \pm\infty$, it follows immediately that $\lim_{\beta_i^0\to\pm\infty} S_i^2 = \hat{\Delta}_i^2/var(\hat{\Delta}_i^2)$. Therefore, the S-region is unbounded iff the "Wald statistic" for identification of β_i is not significantly different from zero, as in this case S_i^2 asymptotes to a value less than c_α^2. Said differently, if $\hat{\Delta}_i^2/var(\hat{\Delta}_i^2) < c_\alpha^2$ then $S_i^2 < c_\alpha^2$ for large values of $|\beta_i^0|$ so extreme value of β_i^0 are not rejected. It follows that the S-region asymptotically satisfies Dufour's (1997) condition requiring, in the case of a near non-identification, that a $(1-\alpha)\cdot 100\%$ confidence region be unbounded at least $1-\alpha$ percent of the time for the regions to attain coverage probability $1-\alpha$.

References

[1] Anderson, T.W. (1950). "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *The Annals of Mathematical Statistics*, 20, 46-63.

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

[3] Blomquist, S. and Dahlberg (1999). "Small Sample Properties of LIML and Jackknife IV Estimators: Experiments with Weak Instruments," *Journal of Applied Econometrics*, 14(1), 69-88.

[4] Bound, J., Jaeger, D.A., and Baker , R.M. (1995). "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90, 443-450.

[5] Chamberlain, G. and Imbens, G. (2003). "Random Effects Estimators with Many Instrumental Variables," unpublished manuscript, Department of Economics, Harvard University.

[6] Choi, I., and Phillips, P.C.B. (1992). "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations," *Journal of Econometrics*, 51, 113-150.

[7] Davidson, R., and MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics*, New York: Oxford University Press.

[8] Dufour, J-M (1997). "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365-88.

[9] Dufour, J-M, and Jasiak, J. (2001). "Finite Sample Limited Information Inference Methods for Structural Equations and Models with Unobserved and Generated Regressors," *International Economic Review*, forthcoming.

[10] Dufour, J-M, and Khalaf, L. (1997). "Simulation Based Finite and Large Sample Inference Methods in Simultaneous Equations," working paper, Université de Montréal.

[11] Dufour, J-M, and Taamouti, M. (2003). "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," Unpublished manuscript, Department of Economics, Université de Montréal.

[12] Flores-Lagunes, A. (2000). "Estimation Methods Robust to Weak Instruments," Unpublished manuscript, Department of Economics, Ohio State University.

[13] Hall, A. R., Rudebusch, G. D., and Wilcox, D. W. (1996). "Judging Instrument Relevance in Instrumental Variables Estimation," *International Economic Review*, 37, 283-289.

[14] Hahn, J. and Hausman, J.A. (2002), "A New Specification Test for the Validity of Instrumental Variables," *Econometrica*, 70, 163-189.

[15] Hahn, J. and Inoue, A. (2002). "A Monte Carlo Comparison of Various Asymptotic Approximations to th Distribution," Econometric Reviews, 21, 309-336.

[16] Kleibergen, F. (2000). "Pivotal Statistics for Testing Subsets of Structural Parameters in the IV Regression Model," Tinbergen Institute Discussion Paper 2000-88/4, University of Amsterdam.

[17] Kleibergen, F. (2002). "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70, 1781-1803.

[18] Maddala, G.S. (1974). "Some Small Sample Evidence on Tests of Significance in Simultaneous Equations Models," *Econometrica*, 42, 841-851.

[19] Maddala, G. S. and Jeong, J. (1992). "On the Exact Small Sample Distribution of the IV Estimator," *Econometrica*, 60, pp. 181-83.

[20] Moreira, M. (2001). "Tests with Correct Size when Instruments can be Arbitrarily Weak," unpublished manuscript, Department of Economics, University of California Berkeley.

[21] Nelson, C. R. and Startz, R. (1990a). "Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator," *Econometrica*, 58, 967-976.

[22] Nelson, C. R. and Startz, R. (1990b): "The Distribution of the Instrumental Variables Estimator and Its t-ratio when the Instrument is a Poor One," *Journal of Business*, 63, S125-S140.

[23] Phillips, P.C.B. (1989). "Partially Identified Econometric Models," *Econometric Theory*, 5, 181-240.

[24] Shea, J., (1997): "Instrument Relevance in Multivariate Linear Models: A Simple Measure," *The Review of Economics and Statistics*, Vol. LXXIX, No. 2, 348-352.

[25] Staiger, D. and Stock, J. H. (1997): "Instrumental Variables Regressions with Weak Instruments," *Econometrica*, 65, 557-586.

[26] Stock, J. H. and Wright, J. (2000), "GMM with Weak Identification," *Econometrica*, 68, 1055-1096.

[27] Stock, J. H. and M. Yogo (2001). "Testing for Weak Instruments," unpublished manuscript, Department of Economics, Harvard University.

[28] Taamouti, M. (2001). "Techniques d'inférence exact dans les modèles structurels avec applications macroéconomiques," PhD thesis, Departemente de sciences économiques, Université de Montréal.

[29] Wang, J. and Zivot, E. (1998): "Inference on Structural Parameters in Instrumental Variables Regressions with Weak Instruments," *Econometrica* , Vol. 66, No. 6, 1389-1404.

[30] Wong, K.-F. (1999). "A Simulation Comparison of Inference for Instrumental Variable Estimators," Unpublished manuscript, Department of Economics, Chinese University of Hong Kong.

[31] Zivot, E., Startz, R., and Nelson, C.R. (1998): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, Vol. 39, No. 4, pp. 1119-1144.