Inference on a Structural Parameter in Instrumental Variables Regression with Weak Instruments

Jiahui Wang and Eric Zivot*
Dept. of Economics, Box 353330
University of Washington
Seattle, WA 98195-3330

February 25, 1996
This Revision: July 12, 1996

*The authors can be also be contacted by email at ezivot@u.washington.edu or wang1@u.washington.edu. Comments are welcome, but please do not quote without the authors' permission.

Abstract

In this paper we consider the problem of making inference on a structural parameter in instrumental variables regression when the instruments are only weakly correlated with the endogenous explanatory variables. Adopting a local-to-zero assumption as in Staiger and Stock (1994) on the coefficients of the instruments in the first stage equation, the asymptotic distributions of various test statistics are derived under a limited information framework. We show that Wald-type test statistics are not pivotal, thus $(1-\alpha)\cdot 100\%$ confidence intervals implied by those test statistics can have zero coverage probability if the standard asymptotic distribution theory is used. In contrast, the likelihood type test statistics are pivotal when the model is just-identified, thus providing valid confidence intervals. Even when the model is over-identified, we show that the distributions of the likelihood type test statistics are bounded from above by a χ^2 distribution with degrees of freedom given by the number of instruments. Hence, we can always invert the likelihood type test statistics to obtain valid, although conservative, confidence intervals. The confidence intervals obtained by using this bounding distribution are compared with those obtained by using the standard $\chi^2(1)$ asymptotic distribution and an alternative bounding distribution, a transformation of the distribution of the Wilks statistic, suggested by Dufour (1994). Confidence intervals based on our χ^2 bounding distribution are shown to be tighter than those based on the Wilks bounding distribution by Monte Carlo experiments.

1 Introduction

Instrumental variables (IV) estimation with weak instruments in finite samples has recently recaptured the attention of econometricians. There are two problems associated with the inference on structural parameters when instrumental variables estimation is used with weakly correlated instruments. First, Nelson and Startz (1990a), Nelson and Startz (1990b), Maddala and Jeong (1992), and Staiger and Stock (1994) show that in finite samples the IV estimates are strongly biased in the same direction as OLS estimates and may lead to incorrect inference when the instruments used are weakly correlated with the endogenous explanatory variables. Although exact small sample distributions have been developed in this context,[2] they usually involve expansion methods and are very hard to apply. Second, Dufour (1994) shows that in limited information simultaneous equations and instrumental variables regression models, the standard asymptotically justified "estimate $\pm$ 2 $\cdot$ asymptotic standard error" confidence intervals for a locally almost unidentified (LAU) structural parameter have zero coverage probability, and even expansion methods and bootstrap techniques cannot handle this problem.

To understand the first problem, Staiger and Stock (1994), hereinafter SS, suggest some alternative asymptotic distributions for TSLS (two-stage-least-squares) and LIML (limited-information-maximum-likelihood) estimators of a structural parameter, using a local-to-zero assumption for the coefficients of the instruments in the reduced form equation. If the coefficients on the instruments in the first stage equation are modeled as nonzero and fixed, then the first stage F statistic for testing the quality of the instruments tends to infinity asymptotically. In contrast, a local-to-zero assumption yields a first stage F statistic which converges to a random variable. SS's Monte Carlo experiments show that the new asymptotic distributions work well with just 20 observations per instrument. However, their asymptotic distributions of t-statistics from TSLS and LIML depend on unknown nuisance parameters in a complicated way, which makes hypothesis testing on the structural parameter very difficult. Also, since the t-statistics are always bounded, they can have zero coverage probability for a LAU structural parameter.

Adopting a local-to-zero assumption as in SS, in this paper we derive the asymptotic distributions of likelihood ratio (LR), Lagrange multiplier (LM) and Wald statistics for testing hypothesis on a structural parameter in a

[2] See Phillips (1983) for a review.

limited information framework, using either generalized method of moments (GMM) or maximum likelihood estimation. Whereas Wald type test statistics are asymptotically nonpivotal, we show that likelihood type test statistics are asymptotically pivotal when the model is just-identified. When the model is over-identified, the asymptotic distributions of the likelihood type test statistics are bounded from above by a χ^2 distribution, which does not depend on unknown nuisance parameters, with degree of freedom equal to the number of instruments. Similar results hold for the GMM statistics if the variance of the structural equation is estimated under the null model considered.

Based on the asymptotic (bounding) distributions of the likelihood type test statistics, we follow Dufour (1994) and consider the construction of valid $(1-\alpha)\cdot 100\%$ confidence sets for a structural parameter. Dufour (1994) proves that under general regularity conditions, any valid confidence set with coverage probability $(1-\alpha)$ for a LAU parameter should be unbounded with probability close to $(1-\alpha)$. We note that valid confidence intervals can be obtained by inverting test statistics whose distributions or bounding distributions do not depend on nuisance parameters. We verify Dufour's results about valid confidence sets being unbounded with probability $(1-\alpha)$ in nearly non-identified models using the goodness-of-fit statistics from the first-stage regression. When the model is over-identified, our Monte Carlo experiments show that the confidence intervals obtained by using the χ^2 bounding distribution are too conservative if the first stage F statistic is somewhat large. Thus, as in Nelson, Startz, and Zivot (1996), we suggest a switching confidence interval when the model is over-identified, based on the goodness-of-fit statistics from the first-stage regression. The confidence intervals obtained by using the χ^2 bounding distribution and the switching confidence intervals are compared with the confidence intervals obtained by inverting the likelihood ratio statistic and using another bounding distribution suggested by Dufour (1994), a transformation of the Wilks statistic in the context of multivariate regression models. Monte Carlo experiments show that the switching confidence intervals and the confidence intervals based on the χ^2 bounding distribution are tighter.

This paper is organized as follows. Section 2 lays down the Limited Information Simultaneous Equations System (LISEM), and introduces the local-to-zero assumptions made by SS. In section 3 we derive the asymptotic distributions of GMM type statistics for testing hypothesis on the structural parameter. In section 4 we derive the asymptotic distributions of the trio of test statistics in the LIML framework. The relationship between GMM and

LIML test statistics are also analyzed. In section 5 we derive the bounding distributions of the LR and LM test statistics and consider the construction of valid confidence intervals using these test statistics. In section 6 we evaluate the finite sample performance of the various test statistics and derived confidence sets using the Monte Carlo design from SS. Section 7 concludes the paper. All the proofs in the paper are relegated to the appendix.

2 The Local-to-Zero LISEM Framework

The LISEM model we consider is:

$$\mathbf{y} = \mathbf{x}\beta + \mathbf{Z}_1\boldsymbol{\gamma} + \mathbf{u} \tag{1}$$

$$\mathbf{x} = \mathbf{Z}_1\boldsymbol{\phi} + \mathbf{Z}_2\boldsymbol{\pi} + \mathbf{v} \tag{2}$$

where $\mathbf{y}$ is the $T \times 1$ vector of observations on the variable we are interested in explaining, $\mathbf{x}$ is the $T \times 1$ vector of observations on the single included endogenous variable, $\mathbf{Z}_1$ is the $T \times k_1$ matrix of k_1 included exogenous variables, and $\mathbf{Z}_2$ is the $T \times k_2$ matrix of k_2 excluded exogenous variables, which are usually called the instruments. Let $k = k_1 + k_2$. $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are assumed to be uncorrelated with $\mathbf{u}$ and $\mathbf{v}$. The disturbance terms $\mathbf{u}$ and $\mathbf{v}$ are only contemporaneously correlated and have mean zero and positive semi-definite variance-covariance matrix given by

$$\boldsymbol{\Omega} = \text{var}\begin{bmatrix} u_i \\ v_i \end{bmatrix} = \begin{bmatrix} \sigma_{uu} & \sigma_{uv} \\ \sigma_{vu} & \sigma_{vv} \end{bmatrix}.$$

Although we assume that there is only one variable in $\mathbf{x}$, the asymptotic distributions of the test statistics considered in this paper can be easily generalized for multivariate $\mathbf{x}$.

It is now widely known that when the instruments $\mathbf{Z}_2$ are weakly correlated with the endogenous variables $\mathbf{x}$ (e.g., $\boldsymbol{\pi} \approx 0$), the TSLS estimator of β and the inference based on it might be misleading; for example, see Nelson and Startz (1990a), Nelson and Startz (1990b). As SS point out, if $\boldsymbol{\pi}$ is modeled as fixed, as in the standard asymptotic analysis in this context, the first-stage F statistic testing $\boldsymbol{\pi} = 0$ in (2) goes to infinity as T increases. Therefore, the standard fixed-$\boldsymbol{\pi}$ asymptotics are inappropriate in the weak instrument case and could lead to incorrect inference. To circumvent this problem, SS suggest using a local-to-zero assumption on $\boldsymbol{\pi}$ so that the first-stage F statistic is $O_p(1)$. Using this local-to-zero assumption, they are able to derive a new asymptotic distribution theory which works well even in very small samples. For completeness, we restate Staiger and Stock's assumptions here. Let "$\Rightarrow$" denote convergence in distribution.

Assumption 1 $\boldsymbol{\pi} = \boldsymbol{\pi}_T = T^{-\frac{1}{2}}\mathbf{g}$, *where* $\mathbf{g}$ *is* $k_2 \times 1$ *vector.*

Assumption 2 *The following limits hold jointly:*

(a) $(T^{-1}\mathbf{u}'\mathbf{u}, T^{-1}\mathbf{u}'\mathbf{v}, T^{-1}\mathbf{v}'\mathbf{v}) \xrightarrow{p} (\sigma_{uu}, \sigma_{uv}, \sigma_{vv})$.

(b) $(T^{-1}\mathbf{Z}_1{}'\mathbf{Z}_1, T^{-1}\mathbf{Z}_1{}'\mathbf{Z}_2, T^{-1}\mathbf{Z}_2{}'\mathbf{Z}_2) \xrightarrow{p} (\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{12}, \boldsymbol{\Sigma}_{22})$.

(c) $(T^{-\frac{1}{2}}\mathbf{Z}_1{}'\mathbf{u}, T^{-\frac{1}{2}}\mathbf{Z}_2{}'\mathbf{u}, T^{-\frac{1}{2}}\mathbf{Z}_1{}'\mathbf{v}, T^{-\frac{1}{2}}\mathbf{Z}_2{}'\mathbf{v}) \Rightarrow (\boldsymbol{\Psi}_{1u}, \boldsymbol{\Psi}_{2u}, \boldsymbol{\Psi}_{1v}, \boldsymbol{\Psi}_{2v})$.

where $\boldsymbol{\Psi} \equiv (\Psi_{1u}{}', \Psi_{2u}{}', \Psi_{1v}{}', \Psi_{2v}{}')'$ *is normally distributed with mean zero and variance* $\boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}$*, and*

$$\boldsymbol{\Sigma} = \mathrm{E}([\mathbf{Z}_1\, \mathbf{Z}_2]'[\mathbf{Z}_1\, \mathbf{Z}_2]) = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

For justification of these assumptions, see SS. Since we are usually interested in making inference only on the parameter β, it is convenient to transform the LISEM in (1) and (2) using the Frisch-Waugh-Lovell theorem[3] as follows:

$$\mathbf{y}^{\perp} = \mathbf{x}^{\perp}\beta + \mathbf{u}^{\perp} \tag{3}$$

$$\mathbf{x}^{\perp} = \mathbf{Z}_2{}^{\perp}\boldsymbol{\pi} + \mathbf{v}^{\perp} \tag{4}$$

where $\mathbf{A}^{\perp} = \mathbf{M}_{z_1}\mathbf{A}$ for any conformable matrix $\mathbf{A}$, and

$$\mathbf{M}_{z_1} = \mathbf{I} - \mathbf{Z}_1(\mathbf{Z}_1{}'\mathbf{Z}_1)^{-1}\mathbf{Z}_1{}'$$

is the matrix that projects onto the space orthogonal to that spanned by $\mathbf{Z}_1$.

As a preparation for the asymptotics of the estimators and test statistics, it is necessary to introduce a lemma. We will offer some intuitive explanations, but for the proof the reader is referred to SS. Before we introduce the lemma, let's define some notations used by SS:

$\theta = \sigma_{uv}/\sigma_{vv}$,

$\rho = \sigma_{uv}/(\sigma_{uu}\sigma_{vv})^{\frac{1}{2}}$,

[3] For example, see Davidson and MacKinnon (1993).

$\kappa = [(\sigma_{uu}/\sigma_{vv})(1-\rho^2)]^{\frac{1}{2}}$,

$\mathbf{V}_{22} \equiv \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}{}^{-1}\mathbf{\Sigma}_{12}$,

$\mathbf{z}_u \equiv \sigma_{uu}{}^{-\frac{1}{2}}\mathbf{V}_{22}{}^{-\frac{1}{2}}(\mathbf{\Psi}_{2u} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}{}^{-1}\mathbf{\Psi}_{1u})$ is a standard normal random vector,

$\mathbf{z}_v \equiv \sigma_{vv}{}^{-\frac{1}{2}}\mathbf{V}_{22}{}^{-\frac{1}{2}}(\mathbf{\Psi}_{2v} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}{}^{-1}\mathbf{\Psi}_{1v})$ is a standard normal random vector,

$\mathbf{z}_u \equiv \rho\mathbf{z}_v + (1-\rho^2)^{\frac{1}{2}}\boldsymbol{\eta}$, and $\boldsymbol{\eta}$ is a k_2-dimensional normal random vector independent of $\mathbf{z}_v$,

$\boldsymbol{\lambda} \equiv \sigma_{vv}{}^{-\frac{1}{2}}\mathbf{V}_{22}{}^{\frac{1}{2}}\mathbf{g}$ (can be interpreted as the concentration parameter),

$\nu_1 \equiv (\boldsymbol{\lambda} + \mathbf{z}_v)'\mathbf{z}_v$,

$\nu_2 \equiv (\boldsymbol{\lambda} + \mathbf{z}_v)'\boldsymbol{\eta}$,

$\nu_3 \equiv (\boldsymbol{\lambda} + \mathbf{z}_v)'(\boldsymbol{\lambda} + \mathbf{z}_v)$,

$\nu_4 \equiv (\boldsymbol{\lambda} + \mathbf{z}_v)'\mathbf{z}_u$,

$\hat{\mathbf{x}^{\perp}} \equiv \mathbf{P}_{z_2^{\perp}}\mathbf{x}^{\perp}$ is the matrix that projects $\mathbf{x}^{\perp}$ onto $\mathbf{Z}_2^{\perp}$.

Lemma 1 (Staiger and Stock, 1994) *Suppose that Assumption 1 and Assumption 2 hold, then:*

(a) $T^{-1}\mathbf{u}^{\perp\prime}\mathbf{u}^{\perp} \xrightarrow{p} \sigma_{uu}$.

(b) $T^{-1}\mathbf{x}^{\perp\prime}\mathbf{u}^{\perp} \xrightarrow{p} \sigma_{uv}$.

(c) $T^{-1}\mathbf{x}^{\perp\prime}\mathbf{x}^{\perp} \xrightarrow{p} \sigma_{vv}$.

(d) $T^{-1}\mathbf{Z}_2{}^{\perp\prime}\mathbf{Z}_2{}^{\perp} \xrightarrow{p} \mathbf{V}_{22}$.

(e) $(\mathbf{Z}_2{}^{\perp\prime}\mathbf{Z}_2{}^{\perp})^{-\frac{1}{2}}\mathbf{Z}_2{}^{\perp\prime}\mathbf{u}^{\perp} \Rightarrow \sigma_{uu}{}^{\frac{1}{2}}\mathbf{z}_u$.

(f) $(\mathbf{Z}_2{}^{\perp\prime}\mathbf{Z}_2{}^{\perp})^{-\frac{1}{2}}\mathbf{Z}_2{}^{\perp\prime}\mathbf{v}^{\perp} \Rightarrow \sigma_{vv}{}^{\frac{1}{2}}\mathbf{z}_v$.

(g) $(\mathbf{Z}_2{}^{\perp\prime}\mathbf{Z}_2{}^{\perp})^{-\frac{1}{2}}\mathbf{Z}_2{}^{\perp\prime}\mathbf{x}^{\perp} \Rightarrow \sigma_{vv}{}^{\frac{1}{2}}(\boldsymbol{\lambda} + \mathbf{z}_v)$.

(h) $\hat{\mathbf{x}^{\perp}}{}'\mathbf{u}^{\perp} \Rightarrow (\sigma_{uu}\sigma_{vv})^{\frac{1}{2}}\nu_4$.

(i) $\hat{\mathbf{x}^{\perp}}{}'\hat{\mathbf{x}^{\perp}} \Rightarrow \sigma_{vv}\nu_3$.

Since we assume that $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are uncorrelated with $\mathbf{u}$ and $\mathbf{v}$, the projection of $\mathbf{u}$ off $\mathbf{Z}_1$ will be just itself, which gives us the result in (a). $\mathbf{x}$ is given by (2), so if we project $\mathbf{x}$ off $\mathbf{Z}_1$, by the local-to-zero assumption on $\boldsymbol{\pi}$, the projection will be just $\mathbf{v}$, from which (b) and (c) follow naturally. If we project $\mathbf{Z}_2$ off $\mathbf{Z}_1$, the variance of the projection will be equal to the difference between the variance of $\mathbf{Z}_2$ and the variance of the projection of $\mathbf{Z}_2$ on $\mathbf{Z}_1$, which is what (d) implies. To understand (e) and (f), it suffices to note that

$$T^{-\frac{1}{2}}\mathbf{Z}_2^{\perp\prime}\mathbf{u}^{\perp} \Rightarrow \mathbf{\Psi}_{2u} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}{}^{-1}\mathbf{\Psi}_{1u},$$

$$T^{-\frac{1}{2}}\mathbf{Z}_2^{\perp\prime}\mathbf{v}^{\perp} \Rightarrow \mathbf{\Psi}_{2v} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}{}^{-1}\mathbf{\Psi}_{1v},$$

and $\mathbf{z}_u$ and $\mathbf{z}_v$ are just standardized normal random vectors. (g) follows naturally by noting that $\mathbf{x}^\perp$ is given by equation 4. Direct calculations can show (h) and (i) using (g).

3 Asymptotic Distributions of GMM-Type Test Statistics

To estimate β using GMM, it is convenient to use the transformed model (3) and (4) instead of the original model (1) and (2). The sample moment condition is given by:

$$m_T(\beta) = \mathbf{Z}_2^{\perp\prime}(\mathbf{y}^\perp - \mathbf{x}^\perp\beta),$$

the asymptotic variance of which in the good instrument case is

$$\mathbf{W} = \text{asy.var}\left[m_T(\beta)\right] = \sigma_{uu}\mathbf{V}_{22},$$

so an efficient GMM estimator of β is given by:

$$\hat{\beta}_{GMM} = \underset{\beta}{\text{argmin}}\, J_T(\beta) = \underset{\beta}{\text{argmin}}\, m_T(\beta)'\mathbf{W}_T^{-1}m_T(\beta),$$

where $\mathbf{W}_T$ is a consistent estimator of $\mathbf{W}$. Using

$$\mathbf{W}_T = \hat{\sigma}_{uu}\mathbf{Z}_2^{\perp\prime}\mathbf{Z}_2^\perp/T$$

as the weight matrix gives:

$$\hat{\beta}_{GMM} = (\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp)^{-1}\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{y}^\perp,$$

which is just the TSLS estimator of β. The results of SS for TSLS thus apply in the GMM context provided

$$\hat{\sigma}_{uu} = (\mathbf{y}^\perp - \mathbf{x}^\perp\hat{\beta}_{GMM})'(\mathbf{y}^\perp - \mathbf{x}^\perp\hat{\beta}_{GMM})/T.$$

Theorem 1 (Staiger and Stock, 1994) *Suppose that Assumption 1 and Assumption 2 hold, then:*

(a) $\hat{\beta}_{GMM} - \beta_0 \Rightarrow \theta\nu_1/\nu_3 + \kappa\nu_2/\nu_3 \equiv \beta^*$.

(b) $\hat{\sigma}_{uu} \Rightarrow \sigma_{uu}[1 - 2\rho(\sigma_{vv}/\sigma_{uu})^{\frac{1}{2}}\beta^* + (\sigma_{vv}/\sigma_{uu})\beta^{*2}]$.

(c) Under the null hypothesis $\beta = \beta_0$,
$t_{GMM} \Rightarrow \nu_4/\{\nu_3[1 - 2\rho(\sigma_{vv}/\sigma_{uu})^{\frac{1}{2}}\beta^* + (\sigma_{vv}/\sigma_{uu})\beta^{*2}]\}^{\frac{1}{2}}$.

(d) The first-stage F statisitc $F_T \Rightarrow \nu_3/k_2$.

Newey and West (1987) show that the trio of test statistics — Wald, Likelihood Ratio[4], Lagrange Multiplier — applies in the context of GMM. For the hypothesis $H_0 : \beta = \beta_0$ vs. $H_1 : \beta \neq \beta_0$, it is trivial to show that:

$$\begin{aligned} D_{GMM} &= [(\mathbf{y}^{\perp} - \mathbf{x}^{\perp}\beta_0)'\mathbf{P}_{z_2^{\perp}}(\mathbf{y}^{\perp} - \mathbf{x}^{\perp}\beta_0) \\ &\quad - (\mathbf{y}^{\perp} - \mathbf{x}^{\perp}\hat{\beta}_{GMM})'\mathbf{P}_{z_2^{\perp}}(\mathbf{y}^{\perp} - \mathbf{x}^{\perp}\hat{\beta}_{GMM})]/\hat{\sigma}_{uu}, \\ LM_{GMM} &= (\mathbf{y}^{\perp} - \mathbf{x}^{\perp}\beta_0)'\mathbf{P}_{\hat{\mathbf{x}}^{\perp}}(\mathbf{y}^{\perp} - \mathbf{x}^{\perp}\beta_0)/\hat{\sigma}_{uu}, \\ Wald_{GMM} &= (\beta_0 - \hat{\beta}_{GMM})^2\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^{\perp}}\mathbf{x}^{\perp}/\hat{\sigma}_{uu}, \end{aligned}$$

where $\mathbf{P}_{\hat{x}^{\perp}}$ is the matrix that projects onto $\hat{\mathbf{x}^{\perp}}$. We remark in passing that these three test statistics are numerically identical.

Proposition 1 *For the hypothesis $H_0 : \beta = \beta_0$ vs. $H_1 : \beta \neq \beta_0$, $Wald_{GMM}$, D_{GMM}, and LM_{GMM} statistics are numerically identical as long as the same $\hat{\sigma}_{uu}$ is used.*

Since the three test statistics are numerically identical as long as the same value of $\hat{\sigma}_{uu}$ is used, we will use $\mathcal{GMM}$ to denote them. However, the choice of $\hat{\sigma}_{uu}$ does make a crucial difference, as the following theorem shows. When $\hat{\sigma}_{uu}$ is chosen to be $(\mathbf{y}^{\perp} - \mathbf{x}^{\perp}\beta_0)'(\mathbf{y}^{\perp} - \mathbf{x}^{\perp}\beta_0)/T$, which is a consistent estimator of σ_{uu} under the null hypothesis regardless of the value of $\boldsymbol{\pi}$, we will refer to the test statistic as $\mathcal{GMM}_0$. Instead, when $\hat{\sigma}_{uu}$ is chosen to be $(\mathbf{y}^{\perp} - \mathbf{x}^{\perp}\hat{\beta}_{GMM})'(\mathbf{y}^{\perp} - \mathbf{x}^{\perp}\hat{\beta}_{GMM})/T$, we will refer to the statistic as $\mathcal{GMM}_1$.

Theorem 2 *Suppose Assumption 1 and Assumption 2 hold. For testing the hypothesis $H_0 : \beta = \beta_0$ vs. $H_1 : \beta \neq \beta_0$:*

(a) If the model is over-identified, then $\mathcal{GMM}_0 \Rightarrow \nu_4{}^2/\nu_3$; if the model is just-identified, then $\mathcal{GMM}_0 \Rightarrow \mathbf{z}_u^2 \equiv \chi^2(1)$.

(b) If the model is over-identified,
$\mathcal{GMM}_1 \Rightarrow \frac{\nu_4{}^2}{\nu_3}[1 - 2\rho(\sigma_{vv}/\sigma_{uu})^{\frac{1}{2}}\beta^* + (\sigma_{vv}/\sigma_{uu})\beta^{*2}]^{-1}$;
if the model is just-identified, then
$\mathcal{GMM}_1 \Rightarrow \mathbf{z}_u^2[1 - 2\rho(\sigma_{vv}/\sigma_{uu})^{\frac{1}{2}}\beta^* + (\sigma_{vv}/\sigma_{uu})\beta^{*2}]^{-1}$.

$\mathcal{GMM}_1$ always follows a nonstandard asymptotic distribution, because the asymptotic distribution of $\hat{\sigma}_{uu}$ is a mixed one dependent on the nuisance

[4] The likelihood ratio type statistic in the context of GMM is equal to the difference in the GMM objective function, and is often referred to as the "D" statistic.

parameters ρ and $\boldsymbol{\lambda}'\boldsymbol{\lambda}/k_2$. Consequently the weight matrix, $\mathbf{W}_T$, converges to a random variable and not a constant. Also, the asymptotic distribution is the same as the asymptotic distribution of the square of t_{TSLS}, which is derived in SS. However, $\mathcal{GMM}_0$ follows a $\chi^2(1)$ distribution asymptotically when the model is just-identified, and later we will show that even when the model is over-identified the asymptotical distribution of $\mathcal{GMM}_0$ can be bounded by a $\chi^2(k_2)$ distribution. This result depends crucially on the fact that $\hat{\sigma_{uu}} \xrightarrow{p} \sigma_{uu}$ and so the weight matrix, $\mathbf{W}_T$, converges to a constant. Therefore, Theorem 2 suggests that $\mathcal{GMM}_0$ should be preferred to $\mathcal{GMM}_1$ for testing the hypothesis $H_0 : \beta = \beta_0$.

4 Asymptotic Distributions of LIML-Type Test Statistics

If we estimate the parameter β using LIML, the hypothesis $H_0 : \beta = \beta_0$ vs. $H_1 : \beta \neq \beta_0$ can tested using the trio of test statistics — Wald, Likelihood Ratio, Lagrange Multiplier. The LIML estimator of β can be derived by minimizing the concentrated log likelihood function:

$$\mathcal{L}^c(\beta) = -T\ln(2\pi) - \frac{T}{2}\ln k(\beta) - \frac{T}{2}\log|\mathbf{Y}'\mathbf{M}_{z_2^\perp}\mathbf{Y}|, \tag{5}$$

where $\mathbf{Y} = [\mathbf{y}^\perp\ \mathbf{x}^\perp]$ and

$$k(\beta) = \frac{(\mathbf{y}^\perp - \mathbf{x}^\perp\beta)'(\mathbf{y}^\perp - \mathbf{x}^\perp\beta)}{(\mathbf{y}^\perp - \mathbf{x}^\perp\beta)'\mathbf{M}_{z_2^\perp}(\mathbf{y}^\perp - \mathbf{x}^\perp\beta)}. \tag{6}$$

It's obvious that minimizing the concentrated log likelihood function is equivalent to minimizing $k(\beta)$, and it can be shown that the minimized $\hat{k}$ is the smallest eigenvalue of $(\mathbf{x}'\mathbf{M}_z\mathbf{x})^{-\frac{1}{2}}(\mathbf{x}'\mathbf{M}_{z_1}\mathbf{x})(\mathbf{x}'\mathbf{M}_z\mathbf{x})^{-\frac{1}{2}}$, or equivalently, $\bar{\zeta}_T \equiv \hat{k} - 1$ is the smallest eigenvalue of $\mathbf{Y}'\mathbf{P}_{z_2^\perp}\mathbf{Y}(\mathbf{Y}'\mathbf{M}_{z_2^\perp}\mathbf{Y})^{-1}$, and the LIML estimator of β is given as a k-class estimator:

$$\hat{\beta}_{LIML} = [\mathbf{x}^{\perp\prime}(\mathbf{I} - \hat{k}\mathbf{M}_{z_2^\perp})\mathbf{x}^\perp]^{-1}\mathbf{x}^{\perp\prime}(\mathbf{I} - \hat{k}\mathbf{M}_{z_2^\perp})\mathbf{y}^\perp. \tag{7}$$

Note that when the model is just-identified, $\hat{k} = 1$, so $\hat{\beta}_{GMM} = \hat{\beta}_{LIML} = \hat{\beta}_{TSLS}$. Defining $\zeta_T = T\bar{\zeta}_T$, SS show that:

Theorem 3 (Staiger and Stock, 1994) *Suppose that Assumption 1 and Assumption 2 hold, then:*

(a) $\zeta_T \Rightarrow \zeta^*$*, where* ζ^* *is the smallest root of* $|G^* - \zeta\bar{\boldsymbol{\Omega}}| = 0$*, where* $G^* = \begin{bmatrix} \mathbf{z}_u \\ \boldsymbol{\lambda} + \mathbf{z}_v \end{bmatrix}' \begin{bmatrix} \mathbf{z}_u & \boldsymbol{\lambda} + \mathbf{z}_v \end{bmatrix}$ *and* $\bar{\boldsymbol{\Omega}} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$.

(b) $\hat{\beta}_{LIML} - \beta_0 \Rightarrow \Delta(\zeta^*)$*, where* $\Delta(\zeta) = (\theta\nu_1 + \kappa\nu_2 - \theta\zeta)/(\nu_3 - \zeta)$.

(c) $\hat{\sigma_{uu}}(k) \Rightarrow \sigma_{uu}[1 - 2\rho(\sigma_{vv}/\sigma_{uu})^{-\frac{1}{2}}\Delta(\zeta^*) + (\sigma_{vv}/\sigma_{uu})\Delta^2(\zeta^*)]$*, where* $\hat{\sigma_{uu}}(k) = (\mathbf{y}^\perp - \mathbf{x}^\perp\hat{\beta}_{LIML})'(\mathbf{y}^\perp - \mathbf{x}^\perp\hat{\beta}_{LIML})/T$.

Using the concentrated log likelihood function (5), we can easily show:

$$\begin{aligned} LR_{LIML} &= T\ln k(\beta_0) - T\ln \hat{k}, \\ LM_{LIML} &= \mathcal{G}(\beta_0)'\mathcal{I}(\beta_0)^{-1}\mathcal{G}(\beta_0)/T, \\ Wald_{LIML} &= (\hat{\beta}_{LIML} - \beta_0)^2 \mathbf{x}^{\perp\prime}(\mathbf{I} - \hat{k}\mathbf{M}_{z_2^\perp})\mathbf{x}^\perp/\hat{\sigma_{uu}}(k), \end{aligned}$$

where

$$\begin{aligned} \mathcal{G}(\beta_0) &= T\left[\frac{\mathbf{x}^{\perp\prime}\mathbf{u}^\perp}{\mathbf{u}^{\perp\prime}\mathbf{u}^\perp} - \frac{\mathbf{x}^{\perp\prime}\mathbf{M}_{z_2^\perp}\mathbf{u}^\perp}{\mathbf{u}^{\perp\prime}\mathbf{M}_{z_2^\perp}\mathbf{u}^\perp}\right], \\ \mathcal{I}(\beta_0) &= T\left(\frac{\mathbf{x}^{\perp\prime}\mathbf{x}^\perp}{\mathbf{u}^{\perp\prime}\mathbf{u}^\perp} - \frac{\mathbf{x}^{\perp\prime}\mathbf{M}_{z_2^\perp}\mathbf{x}^\perp}{\mathbf{u}^{\perp\prime}\mathbf{M}_{z_2^\perp}\mathbf{u}^\perp} - 2\left(\frac{\mathbf{x}^{\perp\prime}\mathbf{u}^\perp}{\mathbf{u}^{\perp\prime}\mathbf{u}^\perp}\right)^2 + 2\left(\frac{\mathbf{x}^{\perp\prime}\mathbf{M}_{z_2^\perp}\mathbf{u}^\perp}{\mathbf{u}^{\perp\prime}\mathbf{M}_{z_2^\perp}}\right)^2\right), \end{aligned}$$

and

$$\mathbf{u}^\perp = \mathbf{y}^\perp - \mathbf{x}^\perp\beta_0.$$

Unlike in the context of GMM, the Wald, Likelihood Ratio and Lagrange Multiplier statistics are not identical in the context of the LIML model. As such their asymptotic distributions differ as the following theorem shows:

Theorem 4 *Suppose that Assumption 1 and Assumption 2 hold, then:*

(a) If the model is just-identified, $LR_{LIML} \Rightarrow \mathbf{z}_u^2 \equiv \chi^2(1)$*; if the model is over-identified,* $LR_{LIML} \Rightarrow \mathbf{z}_u{}'\mathbf{z}_u - \zeta^*$*, where* ζ^* *is defined in Theorem 3.*

(b) If the model is just-identified, $LM_{LIML} \Rightarrow \mathbf{z}_u^2 \equiv \chi^2(1)$*; if the model is over-identified,* $LM_{LIML} \Rightarrow {\nu_4}^2/\nu_3$.

(c) If the model is just-identified, then
$Wald_{LIML} \Rightarrow \mathbf{z}_u^2[1 - 2\rho(\sigma_{vv}/\sigma_{uu})^{\frac{1}{2}}\beta^* + (\sigma_{vv}/\sigma_{uu})\beta^{*2}]^{-1}$;
if the model is over-identified, then
$Wald_{LIML} \Rightarrow \frac{(\nu_4 - \rho\zeta^*)^2}{\nu_3 - \zeta^*}[1 - 2\rho(\sigma_{vv}/\sigma_{uu})^{\frac{1}{2}}\Delta(\zeta^*) + (\sigma_{vv}/\sigma_{uu})\Delta^2(\zeta^*)]^{-1}$.

In the just-identified case, the asymptotic distributions of LR_{LIML} and LM_{LIML} are $\chi^2(1)$ and thus free of nuisance parameters. In the next section we will show that even in the over-identified case, they are bounded from above by a $\chi^2(k_2)$ distribution. Interestingly, the asymptotic distribution for LM_{LIML} is identical to that for $\mathcal{GMM}_0$. The LR statistic, due to its invariance to nonlinear transformations, can be used to test hypothesis about nonlinear functions of β. Also note that the distribution of $Wald_{LIML}$ is just the square of the distribution of LIML t-ratio given by SS, and in the just-identified case, is the same as $Wald_{GMM}$ ($\mathcal{GMM}_1$).

5 Construction of Valid Confidence Intervals

The last two sections of this paper describe the asymptotic distributions of various test statistics based on SS's local-to-zero asymptotics. In principle we can invert the test statistics to obtain confidence intervals for β_0. Generically, given a test statistic $\Psi(\beta_0)$ for testing the hypothesis $H_0 : \beta = \beta_0$ at significance level α, the $(1-\alpha)\cdot 100\%$ (conservative) confidence set for β_0 is the set:

$$C_\psi(\beta_0; 1-\alpha) = \{\beta_0 : \Psi(\beta_0) \leq \mathcal{F}_{1-\alpha}\},$$

where $\mathcal{F}_{1-\alpha}$ is the $(1-\alpha)$ quantile from the asymptotic (bounding) distribution $\mathcal{F}$ of the test statistic. However, as we can see from Theorem 2 and Theorem 4, the asymptotic distributions of Wald type statistics depend on nuisance parameters, so those test statistics are not asymptotically pivotal. For a nonpivotal test statistic, Dufour (1994) shows that the corresponding confidence interval for β_0 has zero coverage probability when the model is locally non-identified.

Obtaining valid confidence intervals for β_0 in the presence of weak instruments requires finding a test statistic that is asymptotically pivotal. One such statistic, originally proposed by Anderson and Rubin (1949), for testing $H_0 : \beta = \beta_0$ is

$$AR = \frac{(\mathbf{y}^{\perp} - \mathbf{x}^{\perp}\beta_0)'\mathbf{P}_{z_2^{\perp}}(\mathbf{y}^{\perp} - \mathbf{x}^{\perp}\beta_0)/k_2}{(\mathbf{y}^{\perp} - \mathbf{x}^{\perp}\beta_0)'\mathbf{M}_{z_2^{\perp}}(\mathbf{y}^{\perp} - \mathbf{x}^{\perp}\beta_0)/(T - k_1 - k_2)},$$

which follows an exact F distribution under the assumption of normality, which is independent of nuisance parameters. Under weaker conditions, SS show that $k_2 AR \Rightarrow \chi^2(k_2)$, which does not depend on nuisance parameters either. Therefore, in a more general set-up the AR statistic is asymptotically pivotal, and valid confidence intervals for β_0 can be obtained by inverting the AR statistic, as suggested by Dufour (1994).

We further note that from Theorem 2 and Theorem 4, when the model is just-identified, the test statistics LR_{LIML}, LM_{LIML} and $\mathcal{GMM}_0$ are asymptotically pivotal, so they can also provide asymptotically valid confidence intervals. When the model is over-identified, we can show that LR_{LIML}, LM_{LIML}, $\mathcal{GMM}_0$ are bounded from above by a $\chi^2(k_2)$ distribution as the next theorem demonstrates.

Theorem 5 *Suppose that Assumption 1 and Assumption 2 hold. When the model given in (1) and (2) is over-identified, the asymptotic distributions of LR_{LIML}, LM_{LIML} and $\mathcal{GMM}_0$ are bounded from above by $\mathbf{z}_u{}'\mathbf{z}_u$, which follows a $\chi^2(k_2)$ distribution.*

Therefore, in over-identified models valid confidence intervals for β_0 can also be constructed by inverting the LR_{LIML}, LM_{LIML} or $\mathcal{GMM}_0$ statistics, though they will be "conservative", i.e., have coverage probability at least $(1-\alpha)$, using Dufour (1990)'s terminology.

Dufour (1994) shows that valid $(1-\alpha)\cdot 100\%$ confidence sets for a LAU parameter must be unbounded with probability $(1-\alpha)$. We now demonstrate this result for confidence sets formed by inverting the AR, LR_{LIML}, LM_{LIML} and $\mathcal{GMM}_0$ statistics for testing $H_0 : \beta = \beta_0$. As shown in Nelson, Startz, and Zivot (1996), inverting the AR, LR_{LIML}, LM_{LIML}, and $\mathcal{GMM}_0$ statistics requires solving a quadratic inequality of the form

$$a\beta_0^2 + b\beta_0 + c \leq 0,$$

from which it follows that the confidence sets will be unbounded when the coefficient a is negative, cover the entire real line when $b^2 - 4ac < 0$ as well, and they can even be empty when $a > 0$ and $b^2 - 4ac < 0$. Following this line of argument, we give the conditions under which the confidence intervals inverted from the AR, LR_{LIML}, LM_{LIML} and $\mathcal{GMM}_0$ statistics will be unbounded:

Proposition 2 *In the LISEM model given by (1) and (2),*[5]

(a) The $(1-\alpha)\cdot 100\%$ *confidence interval* $C_{AR}(\beta_0; 1-\alpha)$ *obtained by inverting the AR statistic will be unbounded when the first stage F statistic for testing* $H_0 : \boldsymbol{\pi} = 0$ *is insignificant at level* α*, i.e.,* $F_{\boldsymbol{\pi}=0} \leq F_{1-\alpha}(k_2, T-k)$.

(b) When the model is nearly non-identified, the $(1-\alpha)\cdot 100\%$ *confidence interval* $C_{AR}(\beta_0; 1-\alpha)$ *will be unbounded with probability* $(1-\alpha)$.

(c) The $(1-\alpha)\cdot 100\%$ *(conservative) confidence interval* $C_{LR}(\beta_0; 1-\alpha)$ *obtained by inverting the* LR_{LIML} *statistic will be unbounded when the first stage F statistic for testing* $H_0 : \boldsymbol{\pi} = 0$ *is insignificant at level* α*, i.e.,* $F_{\boldsymbol{\pi}=0} \leq F_{1-\alpha}(k_2, T-k)$.

(d) The $(1-\alpha)\cdot 100\%$ *(conservative) confidence interval* $C_{LM}(\beta_0; 1-\alpha)$ *obtained by inverting the* LM_{LIML} *and* $\mathcal{GMM}_0$ *statistics will be unbounded when the first stage LM statistic for testing* $H_0 : \boldsymbol{\pi} = 0$ *is insignificant at level* α*, i.e.,* $LM_{\boldsymbol{\pi}=0} \leq \chi^2_{1-\alpha}(k_2)$.

(e) When the model is nearly non-identified, the $(1-\alpha)\cdot 100\%$ *(conservative) confidence intervals* $C_{LR}(\beta_0; 1-\alpha)$ *and* $C_{LM}(\beta_0; 1-\alpha)$ *will aymptotically be unbounded with probability close to* $(1-\alpha)$.

Result (a) is consistent with the common wisdom that the first stage F statistic provides a pretest for the relevance of the instruments. Results (b) and (e) confirm Dufour (1994)'s statement that valid confidence sets for β_0 must be unbounded with probability close to $(1-\alpha)$ in nearly non-identified models. Results (c) and (d) show that either the first stage F statistic or the first stage LM statistic (the TR^2 statistic) can be used as a pretest when LR or LM type test statistics are used to test the hypothesis $\boldsymbol{\pi} = 0$. If these first stage test statistics are insignificant, which implies that the instruments are "weak", the confidence intervals for β_0 will typically be unbounded. Results from Monte Carlo experiments verifying these statements can be found in Nelson, Startz, and Zivot (1996).

As pointed out by Dufour (1994), in the presence of normality, the likelihood ratio criterion[6] for testing $\beta = \beta_0$ is bounded by the Wilks statistic, a statistic commonly used in the analysis of the multivariate linear model.

[5] When inverting the test statistics, the $\chi^2(k_2)$ bounding distribution is used when necessary, and the corresponding confidence intervals will be conservative as a result.

[6] The likelihood ratio statistics used by Dufour (1994) are defined as the ratio of two maximized likelihoods, but the likelihood ratio statistics we have described so far are defined as -2 times the logarithm of the ratio. To avoid confusion, we will refer to the likelihood ratio statistics used by Dufour (1994) as the likelihood ratio criterion (LRC) hereinafter.

Hence, valid confidence intervals can also be obtained by inverting the likelihood ratio criterion for the LIML model. To see the point, consider the unrestricted reduced-form of the limited information model:

$$\mathbf{Y} = \mathbf{Z}\mathbf{B} + \boldsymbol{\epsilon},$$

where $\mathbf{Y} = [\mathbf{y}\ \mathbf{x}], \mathbf{Z} = [\mathbf{Z}_1\ \mathbf{Z}_2]$, $\mathbf{B}$ is the matrix of the reduced-form coefficients, and $\boldsymbol{\epsilon}$ is the vector of reduced-form error terms. The likelihood ratio criterion for testing the null hypothesis $\mathbf{B} = \bar{\mathbf{B}}$ is given by:

$$\overline{LRC} = \frac{L(\bar{\mathbf{B}})}{L(\mathbf{B})},$$

where $L(\bar{\mathbf{B}})$ is the maximum of the likelihood function of the multivariate model under the null hypothesis, and $L(\mathbf{B})$ is the maximum of the likelihood function under the alternative hypothesis. Similarly, the likelihood ratio criterion for testing the null hypothesis $\beta = \beta_0$ in the LIML framework is given by:

$$LRC = \frac{L(\beta_0)}{L(\beta)},$$

where $L(\beta_0)$ is the maximum of the likelihood function of the LIML model under the null hypothesis, and $L(\beta)$ is the maximum of the likelihood function under the alternative hypothesis. As Dufour (1994) points out correctly,

$$L(\bar{\mathbf{B}}) \le L(\beta_0) \le L(\beta) \le L(\mathbf{B}),$$

hence, it follows that:[7]

$$LRC = \frac{L(\beta_0)}{L(\beta)} \ge \frac{L(\bar{\mathbf{B}})}{L(\mathbf{B})} = \overline{LRC}.$$

$\overline{LRC}^{\frac{2}{T}}$ is the Wilks statistic, whose distribution is the distribution of the product $\prod_{i=1}^{p} b_i$, where p is the number of endogenous variables in the multivariate model, and b_i's are independent with Beta distributions (see Rao (1973), or chapter 8 in Anderson (1984)):

$$b_i \sim Beta(\frac{T-k-p+i}{2}, \frac{k}{2}),\ i = 1, \ldots, p.$$

[7] Dufour (1994) mistakenly claims that $L(\mathbf{B})/L(\bar{\mathbf{B}})$ is the likelihood ratio criterion used to construct the Wilks statistic.

The distribution of the Wilks statistic can be easily determined by simulation. Since $LRC = \hat{k}/k(\beta_0)$, where k is defined in equation (6), a conservative confidence interval for β_0 can be obtained by solving the following inequality:

$$\Pr\{\frac{\hat{k}}{k(\beta_0)} \geq W_\alpha(T-k-2,k)\} \geq \Pr\{\overline{LRC} \geq W_\alpha(T-k-2,k)\} = 1-\alpha,$$

where $W_\alpha(T-k-2,k)$ is the α quantile of the distribution of the Wilks statistic. For example, a conservative 95% confidence interval for β_0 is given by all values of β_0 that satisfy

$$k(\beta_0) \leq \frac{\hat{k}}{W_{0.05}(T-k-2,k)} = C_W.$$

In comparison, conservative 95% LR_{LIML} confidence sets for β_0 based on the $\chi^2(k_2)$ bounding distribution consist of all values of β_0 satisfying

$$k(\beta_0) \leq \hat{k}\exp\{\frac{\chi^2_{95\%}(k_2)}{T}\} = C_{LR}$$

By simulating $W_{0.05}(T-k-2,k)$ for various values of T and k, we find that $1/W_{0.05}(T-k-2,k)$ is always larger than $\exp\{\chi^2_{95\%}(k_2)/T\}$ and so confidence sets formed by inverting LR_{LIML} are always smaller than confidence sets formed by inverting LRC.

More generally, a transformation of the Wilks statistic follows a $\chi^2(2k)$ distribution asymptotically,[8] i.e.,

$$-2\ln\overline{LRC} \Rightarrow \chi^2(2k).$$

This suggests that asymptotically valid confidence sets can be constructed based on LR_{LIML}, using $\chi^2(2k)$ as a bounding distribution. However, this obviously yields "wider" confidence sets than those based on the $\chi^2(k_2)$ bounding distribution.

6 Monte Carlo Results

Monte Carlo experiments are conducted to evaluate the asymptotic approximations to the distributions of LR_{LIML}, $Wald_{LIML}$, $\mathcal{GMM}_0$ and $\mathcal{GMM}_1$,

[8] See section 8.5 in Anderson (1984).

and the coverage probabilities of various confidence sets based on them.[9] The Monte Carlo experiments are set up according to SS's design I. That is, the data are generated from (1) and (2) with $\beta = 0$, $\boldsymbol{\gamma} = 0$, $\sigma_{uu} = \sigma_{vv} = 1$, $\mathbf{Z}_1$ a vector of ones, $\mathbf{Z}_2 \sim iid\ N(0, \mathbf{I}_{k_2})$, $(u_i, v_i) \sim iid\ N(0, \bar{\boldsymbol{\Omega}})$. The asymptotic distributions of LR_{LIML}, $Wald_{LIML}$, $\mathcal{GMM}_0$ and $\mathcal{GMM}_1$ are computed using $20,000$ draws of the random variables appearing in the limiting representations in Theorems 2 and 4. The finite sample distributions are computed from $20,000$ replications. Results are reported for $k_2 = 1, 4$; $\rho = 0.5, 0.99$; $\boldsymbol{\lambda}'\boldsymbol{\lambda}/k_2 = 0, 0.25, 1, 10$, and summarized in tables 1 – 5.

In tables 1 – 5, the finite sample CDFs of the five statistics are evaluated at the 90%, 95% and 99% quantiles of various asymptotic distributions. The entries in the tables can also be interpreted as the actual (finite sample) coverage probabilities of 90%, 95% and 99% confidence sets based on nominal levels from the corresponding asymptotic critical values.

For the just-identified models, the asymptotic distributions of LR_{LIML}, $\mathcal{GMM}_0$ and AR are $\chi^2(1)$ (although not the same $\chi^2(1)$ random variable), whereas the distributions of $Wald_{LIML}$ and $\mathcal{GMM}_1$ depend on the nuisance parameters ρ, $\boldsymbol{\lambda}'\boldsymbol{\lambda}/k_2$ and k_2. For $T/k_2 = 5$, the actual sizes based on the true asymptotic distributions of LR_{LIML}, AR, $\mathcal{GMM}_1$ and $Wald_{LIML}$ are too large, but settle down to their nominal levels at $T/k_2 = 20$. The $\mathcal{GMM}_0$ statistic, however, has the correct size for all values of T. The standard $\chi^2(1)$ asymptotic distribution for $\mathcal{GMM}_1$ and $Wald_{LIML}$, in contrast, does poorly in finite samples, especially when $\rho = 0.99$ and $\boldsymbol{\lambda}'\boldsymbol{\lambda}/k_2 = 0$. The failure of $\chi^2(1)$ asymptotics for $\mathcal{GMM}_1$ is partially explained by the inconsistency of the estimator $\hat{\sigma}_{uu}$, whose asymptotic distribution depends on nuisance parameters. as shown in Theorem 1(b). As a result, the GMM estimator for β is not efficient. We also note that our results for $\mathcal{GMM}_1$ and $Wald_{LIML}$ are slightly different from SS's results for $T/k_2 = 5$.

For the over-identified models ($k_2 = 4$), the asymptotic approximations to the finite sample distributions of LR_{LIML} and $\mathcal{GMM}_0$ are very good although the approximation for LR_{LIML} is slightly off for $\rho = 0.99$. The standard $\chi^2(1)$ approximation works well for LR_{LIML} and $\mathcal{GMM}_0$ if $\boldsymbol{\lambda}'\boldsymbol{\lambda}/k_2 = 10$. The worst case for all statistics occurs when $T/k_2 = 20$, $\boldsymbol{\lambda}'\boldsymbol{\lambda}/k_2 = 0$ and $\rho = 0.99$, when the actual coverage probabilities of 95% confidence sets for LR_{LIML},

[9] Since LM_{LIML} cannot easily be written as a quadratic equation in β_0 while $\mathcal{GMM}_0$ can and LM_{LIML} is asymptotically equivalent to $\mathcal{GMM}_0$, we use the formula for $\mathcal{GMM}_0$ to compute LM_{LIML}.

$\mathcal{GMM}_0$, $Wald_{LIML}$ and $\mathcal{GMM}_1$ are 72%, 51%, 25% and 1%, respectively. In all cases, the $\chi^2(k_2)$ bounding distribution for LR_{LIML} and $\mathcal{GMM}_0$ is verified. The finite sample coverage probabilities of 95% confidence sets found by inverting the LRC statistic and using the Wilks bounding distribution are, for all cases, greater than the coverage probabilities for LR_{LIML} based on the $\chi^2(k_2)$ bounding distribution and the $\chi^2(2k)$ bounding distribution. Notice that the $\chi^2(k_2)$ distribution does not bound the distributions of $\mathcal{GMM}_1$ and $Wald_{LIML}$.

The LR_{LIML} and $\mathcal{GMM}_0$ confidence sets based on $\chi^2(k_2)$ critical values may be large if k_2 is large, so the power of the LR_{LIML} and $\mathcal{GMM}_0$ tests may be poor if $\boldsymbol{\lambda}'\boldsymbol{\lambda}/k_2$ is reasonably large. To avoid such a possibility, Nelson, Startz, and Zivot (1996) suggest constructing LR_{LIML} and $\mathcal{GMM}_0$ confidence sets based on the significance of the first stage regression. In particular, if $F_{\boldsymbol{\pi}=0} < F(k_2, T-k)$ (or $LM_{\boldsymbol{\pi}=0} < \chi^2(k_2)$), use the $\chi^2(k_2)$ critical values to construct the confidence sets and if $F_{\boldsymbol{\pi}=0} > F(k_2, T-k)$ (or $LM_{\boldsymbol{\pi}=0} > \chi^2(k_2)$), the critical values from the $\chi^2(1)$ are used. The results for this approximation are given in tables 1 and 2 under the column labeled "switching". The coverage probabilities for LR_{LIML} are almost exact and the coverage probabilities for $\mathcal{GMM}_0$ are only slightly too low when $\boldsymbol{\lambda}'\boldsymbol{\lambda}/k_2 = 1$.

7 Conclusion

For nearly non-identified models, Dufour (1994) argues that confidence sets based on the familiar "estimate $\pm\ 2 \cdot$ asymptotic standard error" may be highly misleading. The paper by SS demonstrates this phenomenon in instrumental variables regression with weak instruments. The asymptotic results in SS for t-statistics based on TSLS and LIML, while highly informative, are not straightforward to apply in practice since they involve unknown values of nuisance parameters. In this paper, we have shown how valid confidence intervals for a single structural parameter can be easily constructed by inverting LR_{LIML}, LM_{LIML} or $\mathcal{GMM}_0$ statistics. In contrast to SS, our methods do not involve estimating nuisance parameters, and the confidence intervals are valid in the sense of Dufour (1994), e.g., $(1-\alpha)\cdot 100\%$ confidence sets are unbounded with probability $(1-\alpha)$ asymptotically. Our $\chi^2(k_2)$ bounding distribution is tighter than Dufour's suggested bounding distribution. Hopefully our results will motivate others to consider constructing confidence sets by inverting LR_{LIML}, LM_{LIML} or $\mathcal{GMM}_0$ statistics in other contexts.

8 Appendix

8.1 Proof of Proposition 1

First let's consider Wald statistic:

$$
\begin{aligned}
Wald_{GMM} = & (\hat{\beta}_{GMM} - \beta_0)'(\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp)(\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp)^{-1} \\
& (\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp)(\hat{\beta}_{GMM} - \beta_0)/\hat{\sigma}_{uu} \\
= & (\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp\hat{\beta}_{GMM} - \mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp\beta_0)'(\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp)^{-1} \\
& (\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp\hat{\beta}_{GMM} - \mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp\beta_0)/\hat{\sigma}_{uu} \\
= & [\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)]'(\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp)^{-1}[\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)]/\hat{\sigma}_{uu} \\
= & (\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)'\mathbf{P}_{\hat{x}^\perp}(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)/\hat{\sigma}_{uu}.
\end{aligned}
$$

This proves that $Wald_{GMM}$ statistic is equivalent to LM_{LIML} statistic. For the equivalence of D_{GMM} and $Wald_{GMM}$, note that the second term in D_{GMM} can be written as

$$
\begin{aligned}
& [\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0 - \mathbf{x}^\perp(\hat{\beta}_{GMM} - \beta_0)]'\mathbf{P}_{z_2^\perp}[\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0 - \mathbf{x}^\perp(\hat{\beta}_{GMM} - \beta_0)] \\
& = (\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)'\mathbf{P}_{z_2^\perp}(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0) + (\hat{\beta}_{GMM} - \beta_0)'\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp(\hat{\beta}_{GMM} - \beta_0) \\
& \qquad - 2(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)'\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp(\beta_{GMM} - \beta_0),
\end{aligned}
$$

so D_{GMM} statistic can be rewritten as

$$
\begin{aligned}
D_{GMM} = & [2(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)'\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp(\hat{\beta}_{GMM} - \beta_0) \\
& - (\hat{\beta}_{GMM} - \beta_0)'\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp(\hat{\beta}_{GMM} - \beta_0)]/\hat{\sigma}_{uu} \\
= & [2(\hat{\beta}_{GMM} - \beta_0)'\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp(\hat{\beta}_{GMM} - \beta_0) \\
& - (\beta_{GMM} - \beta_0)'\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp(\hat{\beta}_{GMM} - \beta_0)]/\hat{\sigma}_{uu} \\
= & (\hat{\beta}_{GMM} - \beta_0)'\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp(\hat{\beta}_{GMM} - \beta_0)/\hat{\sigma}_{uu},
\end{aligned}
$$

which is just the $Wald_{GMM}$ statistic.

8.2 Proof of Theorem 2

When the model is over-identified, let's use LM_{GMM} statistic for $\mathcal{GMM}$.

$$
\begin{aligned}
\mathcal{GMM}_0 &= \frac{(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)'\mathbf{P}_{\hat{x}^\perp}(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)}{(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)'(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)/T} \\
&= \frac{\mathbf{u}^{\perp\prime}\hat{\mathbf{x}^\perp}(\hat{\mathbf{x}^\perp}{}'\hat{\mathbf{x}^\perp})^{-1}\hat{\mathbf{x}^\perp}\mathbf{u}^\perp}{\mathbf{u}^{\perp\prime}\mathbf{u}^\perp/T} \\
&\Rightarrow [\mathbf{z}_u{}'(\boldsymbol{\lambda} + \mathbf{z}_v)]^2/\nu_3 \equiv \nu_4{}^2/\nu_3,
\end{aligned}
$$

where the convergence follows from Lemma 1(a), 1(c), 1(h), and

$$
\begin{aligned}
\mathcal{GMM}_1 &= \frac{(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)'\mathbf{P}_{x^\perp}(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)}{(\mathbf{y}^\perp - \mathbf{x}^\perp\hat{\beta}_{GMM})'(\mathbf{y}^\perp - \mathbf{x}^\perp\hat{\beta}_{GMM})/T} \\
&\Rightarrow \mathbf{z}_u{}'\mathbf{z}_u[1 - 2\rho(\frac{\sigma_{vv}}{\sigma_{uu}})^{\frac{1}{2}}\beta^* + (\frac{\sigma_{vv}}{\sigma_{uu}})\beta^{*2}]^{-1},
\end{aligned}
$$

where the convergence follows from Lemma 1(c), 1(h) and Theorem 1(b).

When the model is just-identified, let's use the D_{GMM} statistic for $\mathcal{GMM}$. Because the model is just-identified, the second term in D_{GMM} statistic will always be equal to zero, so

$$
\begin{aligned}
\mathcal{GMM}_0 &= \frac{(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)'\mathbf{P}_{z_2^\perp}(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)}{(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)'(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)/T} \\
&= \frac{\mathbf{u}^{\perp\prime}\mathbf{Z}_2^\perp(\mathbf{Z}_2^{\perp\prime}\mathbf{Z}_2^\perp)^{-1}\mathbf{Z}_2^{\perp\prime}\mathbf{u}^\perp}{\mathbf{u}^{\perp\prime}\mathbf{u}^\perp/T} \Rightarrow \mathbf{z}_u^2,
\end{aligned}
$$

where the convergence follows from Lemma 1(a) and 1(e). Similarly, for $\mathcal{GMM}_1$:

$$
\begin{aligned}
\mathcal{GMM}_1 &= \frac{(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)'\mathbf{P}_{z_2^\perp}(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)}{(\mathbf{y}^\perp - \mathbf{x}^\perp\hat{\beta}_{GMM})'(\mathbf{y}^\perp - \mathbf{x}^\perp\hat{\beta}_{GMM})/T} \\
&= \frac{\mathbf{u}^{\perp\prime}\mathbf{Z}_2^\perp(\mathbf{Z}_2^{\perp\prime}\mathbf{Z}_2^\perp)^{-1}\mathbf{Z}_2^{\perp\prime}\mathbf{u}^\perp}{\hat{\mathbf{u}^\perp}{}'\hat{\mathbf{u}^\perp}/T} \\
&\Rightarrow \mathbf{z}_u^2[1 - 2\rho(\frac{\sigma_{vv}}{\sigma_{uu}})^{\frac{1}{2}}\beta^* + (\frac{\sigma_{vv}}{\sigma_{uu}})\beta^{*2}]^{-1},
\end{aligned}
$$

where the convergence follows from Lemma 1(e) and Theorem 1(b).

8.3 Proof of Theorem 4

(a) For LR_{LIML} statistic, first let's consider the first term:

$$
\begin{aligned}
T\ln k(\beta_0) &= T\ln\frac{(\mathbf{y}^\perp-\mathbf{x}^\perp\beta_0)'(\mathbf{y}^\perp-\mathbf{x}^\perp\beta_0)}{(\mathbf{y}^\perp-\mathbf{x}^\perp\beta_0)'\mathbf{M}_{z_2^\perp}(\mathbf{y}^\perp-\mathbf{x}^\perp\beta_0)}\\
&= T\ln\frac{\mathbf{u}^{\perp\prime}\mathbf{u}^\perp}{\mathbf{u}^{\perp\prime}\mathbf{u}^\perp-\mathbf{u}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{u}^\perp}\\
&= -T\ln\left(1-\frac{\mathbf{u}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{u}^\perp}{\mathbf{u}^{\perp\prime}\mathbf{u}^\perp}\right)\\
&\approx T\frac{\mathbf{u}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{u}^\perp}{\mathbf{u}^{\perp\prime}\mathbf{u}^\perp}\\
&= \frac{\mathbf{u}^{\perp\prime}\mathbf{Z}_2^\perp(\mathbf{Z}_2^{\perp\prime}\mathbf{Z}_2^\perp)^{-\frac{1}{2}}\mathbf{Z}_2^{\perp\prime}\mathbf{u}^\perp}{\mathbf{u}^{\perp\prime}\mathbf{u}^\perp/T}\\
&\Rightarrow \mathbf{z}_u{}'\mathbf{z}_u,
\end{aligned}
$$

where the approximation on the fourth line follows from a Taylor series expansion, and the convergence follows from Lemma 1(a) and 1(e).

If the model is just-identified, the second term will be equal to zero, so $LR_{LIML}\Rightarrow\mathbf{z}_u^2$; if the model is over-identified, then the second term can be rewritten as:

$$
\begin{aligned}
T\log\hat{k} &= T\log(1+\frac{\zeta_T}{T})\\
&\approx T\frac{\zeta_T}{T}\Rightarrow\zeta^*,
\end{aligned}
$$

where the approximation on the second line follows from a Taylor series expansion, and the convergence follows from Theorem 4(a); hence, $LR_{LIML}\Rightarrow \mathbf{z}_u{}'\mathbf{z}_u-\zeta^*$.

(b) For LM_{LIML} statistic, we will use the following approximation results:

$$
\begin{aligned}
\mathcal{G}(\beta_0) &\approx T\frac{\hat{\mathbf{x}^\perp}{}'\mathbf{u}^\perp}{\mathbf{u}^{\perp\prime}\mathbf{u}^\perp},\\
\mathcal{I}(\beta_0) &\approx \frac{\mathbf{u}^{\perp\prime}\mathbf{P}_{\hat{X^\perp}}\mathbf{u}^\perp}{\mathbf{u}^{\perp\prime}\mathbf{u}^\perp},
\end{aligned}
$$

then the LM_{LIML} statistic reduces to:

$$LM_{LIML} = \frac{\mathbf{u}^{\perp\prime}\mathbf{P}_{\hat{X}^{\perp}}\mathbf{u}^{\perp}}{\mathbf{u}^{\perp\prime}\mathbf{u}^{\perp}/T},$$

which is identical with $\mathcal{GMM}_0$. For the rest of the proof, see the proof of Theorem 2.

(c) For the asymptotic distribution of Wald statistic, a similar proof is given in SS.

8.4 Proof of Theorem 5

From Theorem 4, we know that

$$LR_{LIML} \Rightarrow \mathbf{z}_u{}'\mathbf{z}_u - \zeta^*,$$

where ζ^* is positive because, from Theorem 3, it is an eigenvalue of the positive definite matrix $\bar{\mathbf{\Omega}}^{\frac{1}{2}} G^* \bar{\mathbf{\Omega}}^{\frac{1}{2}\prime}$, provided $\rho \neq \pm 1$. Thus it follows trivially that

$$LR_{LIML} \Rightarrow \mathbf{z}_u{}'\mathbf{z}_u - \zeta^* < \mathbf{z}_u{}'\mathbf{z}_u.$$

From Theorems 2 and 4 we know that asymptotically both LM_{LIML} and $\mathcal{GMM}_0$ converge to ν_4^2/ν_3, where

$$\begin{aligned}\frac{\nu_4^2}{\nu_3} &= \frac{[(\boldsymbol{\lambda} + \mathbf{z}_v)'\mathbf{z}_u]^2}{(\boldsymbol{\lambda} + \mathbf{z}_v)'(\boldsymbol{\lambda} + \mathbf{z}_v)} \\ &= \frac{\mathbf{z}_u{}'\mathbf{z}_u}{\mathbf{z}_u{}'\mathbf{z}_u} \frac{[(\boldsymbol{\lambda} + \mathbf{z}_v)'\mathbf{z}_u]^2}{(\boldsymbol{\lambda} + \mathbf{z}_v)'(\boldsymbol{\lambda} + \mathbf{z}_v)} \\ &= \mathbf{z}_u{}'\mathbf{z}_u \cos^2(\eta) \leq \mathbf{z}_u{}'\mathbf{z}_u,\end{aligned}$$

and

$$\cos^2(\eta) = \frac{[(\boldsymbol{\lambda} + \mathbf{z}_v)'\mathbf{z}_u]^2}{(\mathbf{z}_u{}'\mathbf{z}_u)(\boldsymbol{\lambda} + \mathbf{z}_v)'(\boldsymbol{\lambda} + \mathbf{z}_v)},$$

with η being the angle between any realizations of $\boldsymbol{\lambda} + \mathbf{z}_v$ and $\mathbf{z}_u$.

8.5 Proof of Proposition 2

(a) Since the AR statistic follows an F distribution under the assumption of normality, the $(1-\alpha)\cdot 100\%$ confidence interval for β_0 is given by:

$$\frac{(\mathbf{y}^{\perp}-\mathbf{x}^{\perp}\beta_0)'\mathbf{P}_{z_2^{\perp}}(\mathbf{y}^{\perp}-\mathbf{x}^{\perp}\beta_0)}{(\mathbf{y}^{\perp}-\mathbf{x}^{\perp}\beta_0)'\mathbf{M}_{z_2^{\perp}}(\mathbf{y}^{\perp}-\mathbf{x}^{\perp}\beta_0)}\frac{T-k}{k_2} \le F_{1-\alpha}(k_2, T-k),$$

where $F_{1-\alpha}(k_2, T-k)$ is the $(1-\alpha)$ quantile of F distribution with k_2 and $T-k$ degrees of freedom. From the last inequality it follows that

$$\frac{(\mathbf{y}^{\perp}-\mathbf{x}^{\perp}\beta_0)'(\mathbf{y}^{\perp}-\mathbf{x}^{\perp}\beta_0)}{(\mathbf{y}^{\perp}-\mathbf{x}^{\perp}\beta_0)'\mathbf{M}_{z_2^{\perp}}(\mathbf{y}^{\perp}-\mathbf{x}^{\perp}\beta_0)} \le C_{AR},$$

where $C_{AR} = 1 + F_{1-\alpha}(k_2, T-k)\frac{k_2}{T-k}$. This can be rewritten more compactly as:

$$\begin{pmatrix} 1 & -\beta_0 \end{pmatrix} Q_{AR} \begin{pmatrix} 1 \\ -\beta_0 \end{pmatrix} \le 0,$$

where

$$Q_{AR} = \begin{pmatrix} \mathbf{y}^{\perp\prime}\mathbf{y}^{\perp} & \mathbf{y}^{\perp\prime}\mathbf{x}^{\perp} \\ \mathbf{x}^{\perp\prime}\mathbf{y}^{\perp} & \mathbf{x}^{\perp\prime}\mathbf{x}^{\perp} \end{pmatrix} - C_{AR}\begin{pmatrix} \mathbf{y}^{\perp\prime}\mathbf{M}_{z_2^{\perp}}\mathbf{y}^{\perp} & \mathbf{y}^{\perp\prime}\mathbf{M}_{z_2^{\perp}}\mathbf{x}^{\perp} \\ \mathbf{x}^{\perp\prime}\mathbf{M}_{z_2^{\perp}}\mathbf{y}^{\perp} & \mathbf{x}^{\perp\prime}\mathbf{M}_{z_2^{\perp}}\mathbf{x}^{\perp} \end{pmatrix}.$$

Therefore the $(1-\alpha)\cdot 100\%$ confidence interval for β_0 will be unbounded when the $(2,2)$ element of matrix Q_{AR} is negative, i.e., when

$$\frac{\mathbf{x}^{\perp\prime}\mathbf{x}^{\perp}}{\mathbf{x}^{\perp\prime}\mathbf{M}_{z_2^{\perp}}\mathbf{x}^{\perp}} < C_{AR}.$$

The above condition can be rewritten as

$$\frac{\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^{\perp}}\mathbf{x}^{\perp}}{\mathbf{x}^{\perp\prime}\mathbf{M}_{z_2^{\perp}}\mathbf{x}^{\perp}}\frac{T-k}{k_2} < F_{1-\alpha}(k_2, T-k).$$

This condition is equivalent to the statement that the first stage F statistic is insignificant at level α, with the first stage regression given in equation (4).

(b) When $\boldsymbol{\pi} = 0$,

$$\Pr\{\frac{\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^{\perp}}\mathbf{x}^{\perp}}{\mathbf{x}^{\perp\prime}\mathbf{M}_{z_2^{\perp}}\mathbf{x}^{\perp}}\frac{T-k}{k_2} < F_{1-\alpha}(k_2, T-k)\} = 1-\alpha.$$

This implies that the confidence interval for β_0 is unbounded with probability $(1-\alpha)$ at significance level α, using the result in (a).

(c) Using the expression for LR_{LIML} given in section (5) and Theorem 5, we know that the $(1-\alpha)\cdot 100\%$ (conservative) confidence interval for β_0 is the set of values of β_0 that satisfy the following inequality:

$$T \ln \frac{(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)'(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)}{(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)'\mathbf{M}_{z_2^\perp}(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)} - T\ln\hat{k} \leq \chi^2_{1-\alpha}(k_2),$$

where $\chi^2_{1-\alpha}(K_2)$ is the $(1-\alpha)$ quantile from $\chi^2(k_2)$ distribution. The last inequality can be rewritten as

$$\frac{(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)'(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)}{(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)'\mathbf{M}_{z_2^\perp}(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)} \leq C_{LR},$$

where $C_{LR} = \hat{k}\exp\{\chi^2_{1-\alpha}(k_2)/T\}$. We can further rewrite the inequality in a quadratic form:

$$\begin{pmatrix} 1 & -\beta_0 \end{pmatrix} Q_{LR} \begin{pmatrix} 1 \\ -\beta_0 \end{pmatrix} \leq 0,$$

where

$$Q_{LR} = \begin{pmatrix} \mathbf{y}^{\perp\prime}\mathbf{y}^\perp & \mathbf{y}^{\perp\prime}\mathbf{x}^\perp \\ \mathbf{x}^{\perp\prime}\mathbf{y}^\perp & \mathbf{x}^{\perp\prime}\mathbf{x}^\perp \end{pmatrix} - C_{LR}\begin{pmatrix} \mathbf{y}^{\perp\prime}\mathbf{M}_{z_2^\perp}\mathbf{y}^\perp & \mathbf{y}^{\perp\prime}\mathbf{M}_{z_2^\perp}\mathbf{x}^\perp \\ \mathbf{x}^{\perp\prime}\mathbf{M}_{z_2^\perp}\mathbf{y}^\perp & \mathbf{x}^{\perp\prime}\mathbf{M}_{z_2^\perp}\mathbf{x}^\perp \end{pmatrix}.$$

Therefore the confidence interval obtained by solving this inequality will be unbounded if the $(2,2)$ element of Q_{LR} is negative, i.e., if

$$\frac{\mathbf{x}^{\perp\prime}\mathbf{x}^\perp}{\mathbf{x}^{\perp\prime}\mathbf{M}_{z_2^\perp}\mathbf{x}^\perp} < C_{LR}.$$

This inequality can be rewritten as

$$\frac{\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp}{\mathbf{x}^{\perp\prime}\mathbf{M}_{z_2^\perp}\mathbf{x}^\perp}\frac{T-k_1-k_2}{k_2} < (\hat{k}\exp\{\chi^2_{1-\alpha}(k_2)/T\}-1)\frac{T-k_1-k_2}{k_2}.$$

Note that $\hat{k}$ is approximately equal to 1, and $\exp\{\chi^2_{1-\alpha}(k_2)/T\} \approx 1+\chi^2_{1-\alpha}(k_2)/T$, so the right-hand-side of the inequality is approximately equal to $\chi^2_{1-\alpha}(k_2)/k_2$. From part (a) of this proof, the left-hand-side of the inequality is the first stage F statistic. We know that asymptotically the F statistic converges to $\chi^2_{1-\alpha}(k_2)/k_2$, thus, the $(1-\alpha)\cdot 100\%$ confidence interval obtained by inverting LR_{LIML} will be unbounded when the first stage F statistic is insignificant at level α, using the asymptotic distribution.

(d) Since LM_{LIML} is approximately equal to $\mathcal{GMM}_0$, as shown in part 3(b) of this appendix, let's just consider $\mathcal{GMM}_0$. The $(1-\alpha)\cdot 100\%$ (conservative) confidence interval for β_0 can be obtained by inverting $\mathcal{GMM}_0$, i.e., solving the following inequality:

$$\frac{(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)'\mathbf{P}_{\hat{x^\perp}}(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)}{(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)'(\mathbf{y}^\perp - \mathbf{x}^\perp\beta_0)/T} \le \chi^2_{1-\alpha}(k_2),$$

which can be rewritten compactly in the quadratic form:

$$\begin{pmatrix} 1 & -\beta_0 \end{pmatrix} Q_{LM} \begin{pmatrix} 1 \\ -\beta_0 \end{pmatrix} \le 0,$$

where

$$Q_{LM} = \begin{pmatrix} \mathbf{y}^{\perp\prime}\mathbf{P}_{\hat{x^\perp}}\mathbf{y}^\perp & \mathbf{y}^{\perp\prime}\mathbf{P}_{\hat{x^\perp}}\mathbf{x}^\perp \\ \mathbf{x}^{\perp\prime}\mathbf{P}_{\hat{x^\perp}}\mathbf{y}^\perp & \mathbf{x}^{\perp\prime}\mathbf{P}_{\hat{x^\perp}}\mathbf{x}^\perp \end{pmatrix} - C_{LM} \begin{pmatrix} \mathbf{y}^{\perp\prime}\mathbf{y}^\perp & \mathbf{y}^{\perp\prime}\mathbf{x}^\perp \\ \mathbf{x}^{\perp\prime}\mathbf{y}^\perp & \mathbf{x}^{\perp\prime}\mathbf{x}^\perp \end{pmatrix},$$

and

$$C_{LM} = \chi^2_{1-\alpha}(k_2)/T.$$

Therefore the confidence interval will be unbounded when the $(2,2)$ element of Q_{LM} is negative, i.e., when

$$\frac{\mathbf{x}^{\perp\prime}\mathbf{P}_{\hat{x^\perp}}\mathbf{x}^\perp}{\mathbf{x}^{\perp\prime}\mathbf{x}^\perp} < C_{LM} \quad \text{or} \quad T\frac{\mathbf{x}^{\perp\prime}\mathbf{P}_{z_2^\perp}\mathbf{x}^\perp}{\mathbf{x}^{\perp\prime}\mathbf{x}^\perp} < \chi^2_{1-\alpha}(k_2).$$

Note that the left-hand-side of the inequality is T times the uncentered R^2 from the first stage regression, i.e., the first stage LM statistic for testing $H_0 : \boldsymbol{\pi} = 0$, which follows a $\chi^2(k_2)$ distribution asymptotically. Therefore, the $(1-\alpha)\cdot 100\%$ confidence interval obtained by inverting $\mathcal{GMM}_0$ (LM_{LIML}) will be unbounded if the first stage LM statistic is insignificant at level α.

(e) The proof is similar to the proof in (b), if we use the asymptotic (bounding) distribution for LR_{LIML} and LM_{LIML}.

References

ANDERSON, T. W. (1984): *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, Inc.

ANDERSON, T. W., AND H. RUBIN (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *The Annals of Mathematical Statistics*, 20, 46–63.

DAVIDSON, R., AND J. G. MACKINNON (1993): *Estimation and Inference in Econometrics*. Oxford University Press.

DUFOUR, J.-M. (1990): "Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors," *Econometrica*, 58(2), 475–494.

——— (1994): "Some Impossibility Theorems in Econometrics with Applications to Instrumental Variables, Dynamic Models, and Cointegration," Unversité of Montréal, Working Paper.

MADDALA, G. S., AND J. JEONG (1992): "On the Exact Small Sample Distribution of the IV Estimator," *Econometrica*, 60, 181–183.

NELSON, C. R., AND R. STARTZ (1990a): "The Distribution of the Instrumental Variables Estimator and Its t-ratio When the Instrument is a Poor One," *Journal of Business*, 63, S125–S140.

——— (1990b): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator," *Econometrica*, 58, 967–976.

NELSON, C. R., R. STARTZ, AND E. ZIVOT (1996): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," Department of Economics, University of Washington, Working Paper.

NEWEY, W. K., AND K. D. WEST (1987): "Hypothesis Testing with Efficient Method of Moments Estimators," *International Economic Review*, 28, 777–787.

PHILLIPS, P. C. B. (1983): "Exact Small Sample Theory in the Simultaneous Equations Model," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. 1, chap. 8. North-Holland, Amsterdam.

Rao, C. R. (1973): *Linear Statistical Inference and its Applications.* John Wiley & Sons.

Staiger, D., and J. H. Stock (1994): "Asymptotics for Instrumental Variables Regressions with Weakly Correlated Instruments," NBER Technical Working Paper 151.

Table 1: Finite Sample CDF of LR Statistic:
Evaluated at Selected Quantiles of Various Asymptotic Distributions

$T / k_2 = 5$

Parameters			$\chi^2(1)$			Switching			$\chi^2(k_2)$			Asymptotic		
ρ	$\lambda' \lambda / k_2$	k_2	90%	95%	99%	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.50	0.00	1	0.76	0.84	0.93									
0.50	0.25	1	0.77	0.84	0.94									
0.50	1.00	1	0.76	0.84	0.94									
0.50	10.00	1	0.76	0.84	0.93									
0.50	0.00	4	0.61	0.72	0.88	0.87	0.93	0.98	0.92	0.95	0.99	0.85	0.91	0.97
0.50	0.25	4	0.65	0.75	0.90	0.87	0.93	0.98	0.93	0.96	0.99	0.86	0.92	0.98
0.50	1.00	4	0.73	0.82	0.93	0.87	0.93	0.98	0.96	0.98	0.99	0.91	0.95	0.99
0.50	10.00	4	0.86	0.92	0.98	0.86	0.92	0.98	0.99	0.99	1.00	0.97	0.98	1.00
0.99	0.00	1	0.77	0.84	0.94									
0.99	0.25	1	0.76	0.84	0.94									
0.99	1.00	1	0.76	0.84	0.93									
0.99	10.00	1	0.76	0.84	0.94									
0.99	0.00	4	0.61	0.72	0.88	0.85	0.91	0.97	0.92	0.95	0.99	0.84	0.91	0.97
0.99	0.25	4	0.86	0.92	0.98	0.92	0.96	0.99	0.99	0.99	1.00	0.97	0.99	1.00
0.99	1.00	4	0.87	0.93	0.98	0.93	0.96	0.99	0.99	1.00	1.00	0.98	0.99	1.00
0.99	10.00	4	0.87	0.93	0.98	0.88	0.94	0.99	0.99	1.00	1.00	0.97	0.99	1.00

$T / k_2 = 20$

Parameters			$\chi^2(1)$			Switching			$\chi^2(k_2)$			Asymptotic		
ρ	$\lambda' \lambda / k_2$	k_2	90%	95%	99%	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.50	0.00	1	0.88	0.93	0.99									
0.50	0.25	1	0.87	0.93	0.98									
0.50	1.00	1	0.88	0.93	0.98									
0.50	10.00	1	0.88	0.94	0.98									
0.50	0.00	4	0.66	0.77	0.92	0.91	0.96	0.99	0.95	0.97	0.99	0.89	0.94	0.99
0.50	0.25	4	0.70	0.80	0.93	0.91	0.95	0.99	0.96	0.98	1.00	0.90	0.95	0.99
0.50	1.00	4	0.78	0.87	0.96	0.90	0.95	0.99	0.98	0.99	1.00	0.94	0.97	0.99
0.50	10.00	4	0.89	0.94	0.99	0.88	0.94	0.99	0.99	1.00	1.00	0.98	0.99	1.00
0.99	0.00	1	0.87	0.93	0.98									
0.99	0.25	1	0.88	0.93	0.98									
0.99	1.00	1	0.88	0.93	0.98									
0.99	10.00	1	0.88	0.93	0.98									
0.99	0.00	4	0.66	0.77	0.92	0.89	0.94	0.99	0.95	0.98	0.99	0.89	0.94	0.99
0.99	0.25	4	0.88	0.94	0.99	0.93	0.97	0.99	0.99	1.00	1.00	0.98	0.99	1.00
0.99	1.00	4	0.89	0.95	0.99	0.94	0.97	0.99	0.99	1.00	1.00	0.98	0.99	1.00
0.99	10.00	4	0.89	0.95	0.99	0.89	0.95	0.99	0.99	1.00	1.00	0.98	0.99	1.00

Table 2: Finite Sample CDF of GMM_0 Statistic:
Evaluated at Selected Quantiles of Various Asymptotic Distributions

$T / k_2 = 5$

Parameters			$\chi^2(1)$			Switching			$\chi^2(k_2)$			Asymptotic		
ρ	$\lambda' \lambda / k_2$	k_2	90%	95%	99%	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.50	0.00	1	0.84	0.95	1.00									
0.50	0.25	1	0.85	0.95	1.00									
0.50	1.00	1	0.84	0.95	1.00									
0.50	10.00	1	0.84	0.95	1.00									
0.50	0.00	4	0.75	0.85	0.97	0.94	0.98	1.00	0.99	1.00	1.00	0.89	0.96	1.00
0.50	0.25	4	0.78	0.87	0.97	0.94	0.98	1.00	0.99	1.00	1.00	0.89	0.95	1.00
0.50	1.00	4	0.83	0.91	0.98	0.91	0.97	1.00	0.99	1.00	1.00	0.89	0.95	0.99
0.50	10.00	4	0.88	0.94	0.99	0.89	0.94	0.99	1.00	1.00	1.00	0.89	0.95	0.99
0.99	0.00	1	0.85	0.95	1.00									
0.99	0.25	1	0.84	0.95	1.00									
0.99	1.00	1	0.84	0.95	1.00									
0.99	10.00	1	0.85	0.95	1.00									
0.99	0.00	4	0.34	0.51	0.83	0.89	0.96	1.00	0.91	0.97	1.00	0.91	0.96	1.00
0.99	0.25	4	0.45	0.62	0.88	0.84	0.93	1.00	0.94	0.98	1.00	0.90	0.96	1.00
0.99	1.00	4	0.64	0.78	0.95	0.74	0.86	0.98	0.97	0.99	1.00	0.90	0.96	1.00
0.99	10.00	4	0.86	0.93	0.99	0.86	0.93	0.99	1.00	1.00	1.00	0.89	0.94	0.99

$T / k_2 = 20$

Parameters			$\chi^2(1)$			Switching			$\chi^2(k_2)$			Asymptotic		
ρ	$\lambda' \lambda / k_2$	k_2	90%	95%	99%	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.50	0.00	1	0.89	0.95	0.99									
0.50	0.25	1	0.89	0.95	0.99									
0.50	1.00	1	0.89	0.95	0.99									
0.50	10.00	1	0.89	0.95	0.99									
0.50	0.00	4	0.77	0.86	0.96	0.94	0.97	1.00	0.98	0.99	1.00	0.90	0.95	0.99
0.50	0.25	4	0.80	0.88	0.97	0.93	0.97	1.00	0.98	0.99	1.00	0.90	0.95	0.99
0.50	1.00	4	0.85	0.91	0.98	0.91	0.95	0.99	0.99	1.00	1.00	0.90	0.95	0.99
0.50	10.00	4	0.89	0.95	0.99	0.89	0.94	0.99	1.00	1.00	1.00	0.90	0.95	0.99
0.99	0.00	1	0.88	0.94	0.99									
0.99	0.25	1	0.89	0.95	0.99									
0.99	1.00	1	0.89	0.94	0.99									
0.99	10.00	1	0.89	0.95	0.99									
0.99	0.00	4	0.40	0.57	0.85	0.89	0.95	0.99	0.91	0.95	0.99	0.90	0.95	0.99
0.99	0.25	4	0.49	0.66	0.89	0.81	0.90	0.98	0.93	0.96	0.99	0.90	0.95	0.99
0.99	1.00	4	0.68	0.80	0.94	0.73	0.83	0.95	0.97	0.99	1.00	0.90	0.95	0.99
0.99	10.00	4	0.87	0.93	0.99	0.87	0.93	0.99	0.99	1.00	1.00	0.90	0.95	0.99

Table 3: Finite Sample CDF of $Wald_{LIML}$ Statistic:
Evaluated at Selected Quantiles of Various Asymptotic Distributions

$T / k_2 = 5$

Parameters			$\chi^2(1)$			$\chi^2(k_2)$			Asymptotic		
ρ	$\lambda' \lambda / k_2$	k_2	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.50	0.00	1	0.84	0.88	0.93				0.76	0.82	0.90
0.50	0.25	1	0.84	0.88	0.93				0.76	0.83	0.90
0.50	1.00	1	0.83	0.87	0.92				0.76	0.83	0.90
0.50	10.00	1	0.76	0.82	0.89				0.72	0.80	0.89
0.50	0.00	4	0.82	0.88	0.96	0.97	0.99	0.99	0.86	0.91	0.97
0.50	0.25	4	0.83	0.89	0.96	0.97	0.98	0.99	0.87	0.92	0.97
0.50	1.00	4	0.85	0.90	0.96	0.98	0.98	0.99	0.89	0.94	0.98
0.50	10.00	4	0.87	0.92	0.97	0.98	0.99	1.00	0.88	0.93	0.98
0.99	0.00	1	0.28	0.31	0.36				0.75	0.81	0.89
0.99	0.25	1	0.50	0.53	0.59				0.75	0.82	0.91
0.99	1.00	1	0.64	0.67	0.73				0.76	0.83	0.92
0.99	10.00	1	0.75	0.79	0.85				0.77	0.85	0.92
0.99	0.00	4	0.23	0.25	0.29	0.30	0.32	0.35	0.86	0.91	0.97
0.99	0.25	4	0.74	0.77	0.82	0.84	0.86	0.89	0.99	1.00	1.00
0.99	1.00	4	0.84	0.87	0.91	0.93	0.94	0.96	0.98	0.99	1.00
0.99	10.00	4	0.88	0.92	0.96	0.97	0.98	0.99	0.90	0.94	0.99

$T / k_2 = 20$

Parameters			$\chi^2(1)$			$\chi^2(k_2)$			Asymptotic		
ρ	$\lambda' \lambda / k_2$	k_2	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.50	0.00	1	0.95	0.97	0.99				0.88	0.93	0.98
0.50	0.25	1	0.94	0.97	0.99				0.88	0.93	0.98
0.50	1.00	1	0.93	0.96	0.99				0.88	0.93	0.98
0.50	10.00	1	0.90	0.94	0.98				0.87	0.93	0.98
0.50	0.00	4	0.85	0.91	0.98	0.99	0.99	1.00	0.89	0.94	0.99
0.50	0.25	4	0.85	0.91	0.98	0.99	0.99	1.00	0.90	0.95	0.99
0.50	1.00	4	0.87	0.92	0.97	0.99	0.99	1.00	0.91	0.95	0.99
0.50	10.00	4	0.90	0.95	0.98	0.99	1.00	1.00	0.90	0.95	0.99
0.99	0.00	1	0.32	0.36	0.42				0.86	0.92	0.98
0.99	0.25	1	0.61	0.65	0.71				0.88	0.93	0.98
0.99	1.00	1	0.75	0.79	0.84				0.87	0.93	0.98
0.99	10.00	1	0.86	0.90	0.94				0.88	0.93	0.98
0.99	0.00	4	0.24	0.26	0.30	0.31	0.33	0.36	0.89	0.94	0.99
0.99	0.25	4	0.76	0.80	0.85	0.87	0.89	0.91	1.00	1.00	1.00
0.99	1.00	4	0.85	0.88	0.93	0.94	0.95	0.97	0.99	1.00	1.00
0.99	10.00	4	0.91	0.94	0.97	0.98	0.99	0.99	0.92	0.96	0.99

Table 4: Finite Sample CDF of GMM_1 Statistic:
Evaluated at Selected Quantiles of Various Asymptotic Distributions:

$T / k_2 = 5$

Parameters			$\chi^2(1)$			$\chi^2(k_2)$			Asymptotic		
ρ	$\lambda' \lambda / k_2$	k_2	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.50	0.00	1	0.84	0.88	0.93				0.76	0.82	0.90
0.50	0.25	1	0.84	0.88	0.93				0.76	0.83	0.90
0.50	1.00	1	0.83	0.87	0.92				0.76	0.83	0.90
0.50	10.00	1	0.76	0.82	0.89				0.72	0.80	0.90
0.50	0.00	4	0.70	0.81	0.93	0.95	0.97	0.99	0.84	0.90	0.96
0.50	0.25	4	0.73	0.82	0.93	0.95	0.97	0.99	0.84	0.90	0.96
0.50	1.00	4	0.78	0.85	0.94	0.96	0.97	0.99	0.87	0.92	0.97
0.50	10.00	4	0.85	0.91	0.97	0.98	0.99	0.99	0.86	0.92	0.98
0.99	0.00	1	0.28	0.31	0.36				0.75	0.81	0.89
0.99	0.25	1	0.50	0.53	0.59				0.75	0.82	0.91
0.99	1.00	1	0.64	0.67	0.73				0.76	0.83	0.92
0.99	10.00	1	0.75	0.79	0.85				0.77	0.85	0.92
0.99	0.00	4	0.01	0.01	0.02	0.02	0.03	0.04	0.82	0.88	0.96
0.99	0.25	4	0.11	0.13	0.19	0.21	0.24	0.30	0.97	0.99	1.00
0.99	1.00	4	0.36	0.42	0.53	0.57	0.62	0.70	0.95	0.99	1.00
0.99	10.00	4	0.79	0.85	0.92	0.94	0.95	0.98	0.87	0.93	0.98

$T / k_2 = 20$

Parameters			$\chi^2(1)$			$\chi^2(k_2)$			Asymptotic		
ρ	$\lambda' \lambda / k_2$	k_2	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.50	0.00	1	0.95	0.97	0.99				0.88	0.93	0.90
0.50	0.25	1	0.94	0.97	0.99				0.88	0.93	0.90
0.50	1.00	1	0.93	0.96	0.99				0.88	0.93	0.90
0.50	10.00	1	0.90	0.94	0.98				0.87	0.93	0.98
0.50	0.00	4	0.75	0.86	0.96	0.98	0.99	1.00	0.89	0.94	0.98
0.50	0.25	4	0.77	0.86	0.96	0.98	0.99	1.00	0.88	0.94	0.99
0.50	1.00	4	0.80	0.88	0.96	0.98	0.99	1.00	0.89	0.94	0.99
0.50	10.00	4	0.88	0.93	0.98	0.99	0.99	1.00	0.89	0.95	0.99
0.99	0.00	1	0.32	0.36	0.42				0.86	0.92	0.98
0.99	0.25	1	0.61	0.65	0.71				0.88	0.93	0.98
0.99	1.00	1	0.75	0.79	0.84				0.87	0.93	0.98
0.99	10.00	1	0.86	0.90	0.94				0.88	0.93	0.98
0.99	0.00	4	0.01	0.01	0.02	0.03	0.03	0.04	0.88	0.94	0.98
0.99	0.25	4	0.12	0.14	0.20	0.22	0.25	0.30	0.94	0.98	1.00
0.99	1.00	4	0.38	0.44	0.55	0.59	0.62	0.70	0.92	0.97	1.00
0.99	10.00	4	0.82	0.87	0.94	0.95	0.97	0.98	0.90	0.95	0.99

Table 5: Finite Sample CDF of AR Statistic,
Evaluated at Selected Quantiles of the Asymptotic Distribution

Parameters			$T / k_2 = 5$			$T / k_2 = 20$		
ρ	$\lambda' \lambda / k_2$	k_2	90%	95%	99%	90%	95%	99%
0.50	0.00	1	0.80	0.85	0.92	0.89	0.94	0.98
0.50	0.25	1	0.81	0.86	0.92	0.88	0.93	0.98
0.50	1.00	1	0.80	0.85	0.92	0.88	0.93	0.98
0.50	10.00	1	0.80	0.85	0.92	0.89	0.94	0.98
0.50	0.00	4	0.85	0.90	0.96	0.89	0.94	0.99
0.50	0.25	4	0.85	0.90	0.96	0.89	0.94	0.98
0.50	1.00	4	0.85	0.90	0.96	0.90	0.94	0.99
0.50	10.00	4	0.85	0.90	0.96	0.89	0.94	0.99
0.99	0.00	1	0.80	0.86	0.92	0.88	0.93	0.98
0.99	0.25	1	0.80	0.85	0.92	0.88	0.93	0.98
0.99	1.00	1	0.80	0.85	0.92	0.88	0.93	0.98
0.99	10.00	1	0.81	0.86	0.92	0.88	0.93	0.98
0.99	0.00	4	0.85	0.90	0.96	0.89	0.94	0.99
0.99	0.25	4	0.85	0.90	0.96	0.89	0.94	0.98
0.99	1.00	4	0.85	0.90	0.96	0.89	0.94	0.99
0.99	10.00	4	0.85	0.90	0.96	0.89	0.94	0.99

Notes for Tables:

- The blanked rows in the tables correspond to the results in the rows under the "$\chi^2(1)$" heading.
- The results under the column labeled "Switching" are based on the switching confidence intervals as explained in section 6.
- For GMM_1, the results under the heading "$\chi^2(1)$" correspond to the results in SS table 5 on the coverage probabilities for t_{TSLS}. The results under "Asymptotic" correspond to the results in SS Table 2 under the column "95%".
- For $Wald_{LIML}$, the results under the heading "$\chi^2(1)$" correspond to the results in SS table 5 on the coverage probabilities for t_{LIML}. The results under "Asymptotic" correspond to the results in SS Table 4 under the column "95%".