THE UNIVERSITY of ADELAIDE

School of Economics

Working Papers

ISSN 2203-6024

On Bootstrap Validity for Subset Anderson-Rubin Test in IV Regressions

Firmin Doko Tchatoka   Wenjie Wang

Working Paper No. 2015-01
January, 2015

On bootstrap validity for subset Anderson-Rubin test in IV regressions

Firmin Doko Tchatoka[*] Wenjie Wang [†]
The University of Adelaide Kyoto University

January 5, 2015

[*] School of Economics, The University of Adelaide, 10 Pulteney Street, Adelaide SA 5005, Tel:+618 8313 1174, Fax:+618 8223 1460, e-mail: firmin.dokotchatoka@adelaide.edu.au

[†] Graduate School of Economics, Kyoto University. e-mail: wang.jie.33n@st.kyoto-u.ac.jp

ABSTRACT

This paper sheds new light on subset hypothesis testing in linear structural models in which instrumental variables (IVs) can be arbitrarily weak. For the first time, we investigate the validity of the bootstrap for Anderson-Rubin (AR) type tests of hypotheses specified on a subset of structural parameters, with or without identification. Our investigation focuses on two subset AR type statistics based on the plug-in principle. The first one uses the restricted limited information maximum likelihood (LIML) as the plug-in method, and the second exploits the restricted two-stage least squares (2SLS). We provide an analysis of the limiting distributions of both the standard and proposed bootstrap AR statistics under the subset null hypothesis of interest. Our results provide some new insights and extensions of earlier studies. In all cases, we show that when identification is strong and the number of instruments is fixed, the bootstrap provides a high-order approximation of the null limiting distributions of both plug-in subset statistics. However, the bootstrap is inconsistent when instruments are weak. This contrasts with the bootstrap of the AR statistic of the null hypothesis specified on the full vector of structural parameters, which remains valid even when identification is weak; see Moreira et al. (2009). We present a Monte Carlo experiment that confirms our theoretical findings.

Key words: Subset AR-test; bootstrap validity; bootstrap inconsistency; weak instruments; Edgeworth expansion; subsampling.

JEL classification: C12; C13; C36.

1. Introduction

The literature on weak instruments is widespread,[1] and most studies have often focused on testing hypotheses specified on the full set of "structural parameters". There is now a growing interest on inference procedures for subset hypotheses; for examples, see Stock and Wright (2000), Dufour and Jasiak (2001), Kleibergen (2004, 2008), Dufour and Taamouti (2005, 2007), and Startz, Nelson and Zivot (2006). This literature concerned with subset hypotheses testing falls generally into two categories.

The first category is the projection method based on identification-robust statistics.[2] This method consists of inverting robust statistics to build a confidence set for the full set of parameters, and then uses projection techniques to obtain a confidence set for the subset of parameters. In addition to being robust to weak identification, the projection method based on the Anderson and Rubin (1949, AR) statistic also enjoys robustness to instrument exclusion. However, this method has often been criticized for being overly conservative and having low power when too many instruments are used.

The second category includes the robust subset procedures originally suggested by Stock and Wright (2000) and recently developed by Kleibergen (2004, 2008), and Startz et al. (2006). These procedures, known as plug-in based tests, consist of replacing the nuisance parameters that are not specified by the hypothesis of interest by estimators. It is well known that plug-in based tests never over-reject the true parameter values when the nuisance parameters are identified. Guggenberger, Kleibergen, Mavroeidis and Chen (2012), and Guggenberger and Chen (2011) show that the plug-in subset AR test has a *correct asymptotic size* even when identification is weak, i.e., this statistic is robust to identifying assumptions, while the subset test based on Kleibergen (2002) (K) statistic is sensitive to such assumptions. Doko Tchatoka (2014) shows that even for moderate sample size, all conventional plug-in subset tests (including the subset AR test) are overly conservative when the nuisance parameters which are not specified by the hypothesis of interest are weakly identified.

[1] See the reviews by Stock, Wright and Yogo (2002), Dufour (2003), Andrews and Stock (2007), Poskitt and Skeels (2012), and Mikusheva (2013), among others.

[2] See Dufour (1997), Dufour and Jasiak (2001), Dufour and Taamouti (2005, 2007), Dufour, Khalaf and Beaulieu (2010), and Doko Tchatoka and Dufour (2014).

This paper contributes to the literature on weak instruments by investigating whether a distribution-free method such as bootstrap can improve the size property of the subset AR tests, especially when identification is not very strong. To be more specific, we suggest a bootstrap method similar to those of Moreira, Porter and Suarez (2009) for the score test of the null hypothesis in the structural parameters. Testing subset hypotheses is substantially more complex than testing joint hypotheses on the full set of parameters, especially when identification is weak. Thus, the validity of Moreira et al.'s (2009) bootstrap for the subset AR test is not obvious, and further investigation is required to elucidate this issue.

In this paper, our analysis focuses on two plug-in subset AR statistics. The first one uses the restricted LIML as the plug-in estimator of the nuisance parameters, and the second uses the restricted 2SLS as the plug-in method. We observe that most studies of subset hypothesis testing in linear instrumental variables (IV) regressions [except Startz et al. (2006)] usually consider the restricted LIML as the plug-in method. So, not much is known about the size property of the plug-in AR test based on 2SLS estimator when identification is weak (weak instruments). This paper also aims to fill this gap by extending the analysis to the plug-in method based on the 2SLS estimator.

After formulating a general asymptotic framework which allows us to address this issue in a convenient way, we provide an analysis of the limiting distributions of both the standard and bootstrap AR statistics under the subset null hypothesis of interest. Our results provide some new insights and extensions of earlier studies. In all cases, we show that when identification is strong, the bootstrap provides a high-order approximation of the null limiting distributions of both plug-in subset statistics. However, the bootstrap is inconsistent when instruments are weak. This contrasts with the bootstrap of the AR statistic of the null hypothesis specified on the full vector of structural parameters, which remains valid even when identification is weak; see Moreira et al. (2009). The inconsistency of bootstrap under weak instruments is mainly due to the fact that bootstrapping fails to mimic the concentration matrix (or parameter) that characterizes the strength of the identification of the nuisance parameters (which are not specified by the null hypothesis) in such contexts. Moreover, we also show that alternative resampling techniques, such as subsampling, which usually offer a good approximation of the size of the tests in many cases where bootstrap typically fails [see Andrews and Guggenberger (2009)], are invalid under Staiger and Stock's (1997) *local to zero* weak instruments asymptotic; see Section A.1 of the appendix. Finally, we present a Monte Carlo experiment that confirms our theoretical findings.

This paper is organized as follows. In Section 2, the model and assumptions are formulated, and the studied statistics are presented. In Section 3, the limiting behaviors of the standard subset AR tests are provided, while in Section 4, the proposed bootstrap method as well as its asymptotic validity are discussed, with and without weak instruments. In section 5, the Monte Carlo experiment is presented. Conclusions are drawn in Section 6. The subsampling results, the auxiliary lemmata, and proofs are provided in the appendix.

Throughout the paper, I_q stands for the identity matrix of order q. For any full-column rank $n \times m$ matrix A, $P_A = A(A'A)^{-1}A'$ is the projection matrix on the space of A, and $M_A = I_n - P_A$. The notation $vec(A)$ is the $nm \times 1$ dimensional column vectorization of A. $B > 0$ for a squared matrix B means that B is positive definite. Convergence almost surely is symbolized by "$a.s.$", "$\xrightarrow{p}$" stands for convergence in probability, while "$\xrightarrow{d}$" means convergence in distribution. The usual orders of magnitude are denoted by $O_p(.)$, $o_p(.)$, $O(1)$, and $o(1)$. $\|U\|$ denotes the usual Euclidian or Frobenius norm for a matrix U. For any set $\mathscr{B}$, $\partial\mathscr{B}$ is the boundary of $\mathscr{B}$ and $(\partial\mathscr{B})^{\varepsilon}$ is the ε-neighborhood of $\mathscr{B}$. Finally, $\sup_{\omega\in\Omega}|f(\omega)|$ is the supremum norm on the space of bounded continuous real functions, with topological space Ω.

2. Framework

In this section, we present the model and the statistics studied. Sections 3 and 4 will give the concrete results.

2.1. Model and assumptions

We consider a standard linear IV regression with two (possibly) endogenous right hand side (rhs) variables and L instrumental variables (IVs). The sample size is n. The model consists of a structural equation and a reduced-form equation:

$$\begin{aligned} y &= X\beta + W\gamma + \varepsilon, & (2.1) \\ (X, W) &= Z(\Pi_x : \Pi_w) + (V_x, V_w), & (2.2) \end{aligned}$$

where $y \in \mathbb{R}^n$ is a vector of observations on a dependent variable, $X \in \mathbb{R}^n$ and $W \in \mathbb{R}^n$ are vectors of observations on (possibly) endogenous explanatory variables, $Z \in R^{n\times L}$ contains L exogenous variables excluded from (2.1) (instruments), $\varepsilon \in \mathbb{R}^n$ is a vector of structural disturbances, $V_x \in \mathbb{R}^n, V_w \in \mathbb{R}^n$ are vectors of reduced-form disturbances, $\beta \in \mathbb{R}$ and $\gamma \in \mathbb{R}$ are (typically) unknown fixed coefficients (structural parameters), $\Pi_x \in \mathbb{R}^L$ and $\Pi_w \in \mathbb{R}^L$ represent vectors of unknown (fixed) coefficients.

We suppose that Z has full-column rank L with probability one, and that $L \geq 2$. The full-column rank condition of Z ensures the existence of unique least squares estimates in (2.2) when X and W are regressed on each column of Z. As long as Z has full-column rank with probability one and the conditional distributions of X and W, given Z, are absolutely continuous (with respect to the Lebesgue measure), the matrix $[X, W]$ also has full-column rank with probability one. So, the least squares estimates of β and γ in (2.1) are also unique. In practice, it may be relevant to include exogenous regressors in (2.1)-(2.2), whose coefficients remain unrestricted in the inference. If so, the results of this paper do not change qualitatively by replacing the variables that appear currently in (2.1)-(2.2) by the residuals that result from their regression on these exogenous variables.

If the errors ε, V_x and V_w have zero means[3] and Z has full-column rank with probability one, then the usual necessary and sufficient condition for the identification of model (2.1)-(2.2) is $\Pi_{xw} = [\Pi_x : \Pi_w]$ has full-column rank. If $\Pi_{xw} = 0$, Z is irrelevant so that β and γ are completely unidentified. If Π_{xw} is close to zero, β and γ are ill-determined by the data, a situation often called "weak identification" in the literature; see Staiger and Stock (1997), Stock et al. (2002), Dufour (2003), Andrews and Stock (2007), Poskitt and Skeels (2012), and Mikusheva (2013).

We are concerned with the validity of the bootstrap for the plug-in Anderson and Rubin (1949)

[3]This assumption may also be replaced by another "location assumption," such as zero medians.

subset AR-test of the subset null hypothesis

$$H_0 : \beta = \beta_0 \tag{2.3}$$

without any identifying assumptions of model (2.1)-(2.2), where β_0 is a constant vector.

Let $\tilde{Y} = [y, X, W] = [\tilde{Y}_1, \ldots, \tilde{Y}_n]'$ denote the matrix of endogenous variables, where we define $\tilde{Y}_i \in \mathbb{R}^3$ as the ith row of $\tilde{Y}$, written as a column vector, and similarly for other model random matrices. Let also $\mathscr{X}_n = \left\{(\tilde{Y}_1', Z_1'), \ldots, (\tilde{Y}_n', Z_n')\right\}$ and $R_n = vech\left[(\tilde{Y}_n', Z_n')'(\tilde{Y}_n', Z_n')\right] = \left(f_1(\tilde{Y}_n', Z_n'), f_2(\tilde{Y}_n', Z_n'), \ldots, f_K(\tilde{Y}_n', Z_n')\right)'$, where $f_p(\cdot)$, $p = 1, \ldots, K = \frac{1}{2}(L+2)(L+3)$, are elements of $(\tilde{Y}_n', Z_n')'(\tilde{Y}_n', Z_n')$ and R_n has a distribution F.

We make the following generic assumptions on the model variables.

Assumption 2.1 (i) $\mathbb{E}(\|R_n\|^{2+r}) < \infty$ *for some* $r > 0$, *and* (ii) $\limsup_{\|t\| \to \infty} \mid \mathbb{E}\left(exp(it'R_n)\right) \mid < 1$, *where* $i = \sqrt{-1}$.

Assumption 2.2 $\mathbb{E}[(\varepsilon : V_x : V_w) \mid Z] = 0$.

Assumption 2.3 *When the sample size n converges to infinity, the following convergence results hold jointly*:

(i) $n^{-1}(\varepsilon : V_x : V_w)'(\varepsilon : V_x : V_w) \xrightarrow{p} \Omega$, $n^{-1}Z'Z \xrightarrow{p} Q_Z$, *where* $\Omega = \begin{pmatrix} \sigma_{\varepsilon\varepsilon} & \sigma_{V\varepsilon}' \\ \sigma_{V\varepsilon} & \Sigma_V \end{pmatrix}$, $\Sigma_V = \begin{pmatrix} \sigma_{V_xV_x} & \sigma_{V_wV_x} \\ \sigma_{V_wV_x} & \sigma_{V_wV_w} \end{pmatrix}$, $\sigma_{V\varepsilon} = (\sigma_{V_x\varepsilon}, \sigma_{V_w\varepsilon})'$, $\sigma_{jk} \in \mathbb{R}$ *for all* $j, k \in \{\varepsilon, V_x, V_w\}$;

(ii) $n^{-1/2}Z'(\varepsilon : V_x : V_w) \xrightarrow{d} \left(\psi_{Z\varepsilon} : \psi_{ZV_x} : \psi_{ZV_w}\right)$, $vec\left(\psi_{Z\varepsilon} : \psi_{ZV_x} : \psi_{ZV_w}\right) \sim N(0, \Sigma \otimes Q_Z)$, $\psi_{Z\varepsilon}, \psi_{ZV_x}, \psi_{ZV_w}$ *are* $L \times 1$ *random vectors.*

Assumption **2.1** is similar to Moreira et al. (2009, Assumptions 2-3) with $r = s - 2$ and $s \geq 3$. Assumption **2.1**-(i) requires that R_n has second moments or greater while Assumption **2.1**-(ii) requires that its characteristic function be bounded above by 1. In particular, the second moments of R_n exist if $\mathbb{E}(\|(\tilde{Y}_n', Z_n')\|^{2(r+2)}) < \infty$ for some $r > 0$. The bound on the characteristic function is the commonly used Cramér's condition [see Bhattacharya and Ghosh (1978)].

Assumption **2.2** is the usual conditional zero mean assumption of the model errors. The two convergences in Assumption **2.3**-(i) are the weak law of large numbers (WLLN) property of $[\varepsilon, V_x, V_w]$ and Z, respectively, while the convergence in Assumption **2.3**-(ii) is the central limit theorem (CLT) property.

From (2.1)-(2.2), we can write the reduced-form for $\tilde{Y}(\beta_0) = [\tilde{y}(\beta_0), W]$ under $H_0 : \beta = \beta_0$ as:

$$\tilde{Y}(\beta_0) \quad = \quad Z\Pi_w(\gamma : 1) + (v_1, V_w), \tag{2.4}$$

where $v_1 = V_w\gamma + \varepsilon$. As long as Z has full-column rank with probability one and $[\varepsilon, V_w]$ has zero mean, the coefficients of the reduced-form equation (2.4) are identifiable. As a result, γ is identifiable under H_0 whenever $\Pi_w \neq 0$ is fixed. Let

$$\psi_{v_1} \quad = \quad Q_Z^{-1/2}(\psi_{ZV_w}\gamma + \psi_{Z\varepsilon}), \;\; \Omega_W = \begin{pmatrix} \sigma_{11} & \sigma_{V_w1} \\ \sigma_{V_w1} & \sigma_{V_wV_w} \end{pmatrix},$$

where $\sigma_{11} = \sigma_{\varepsilon\varepsilon} + 2\sigma_{V_w\varepsilon}\gamma + \gamma^2\sigma_{V_wV_w}$, and $\sigma_{V_w1} = \sigma_{V_wV_w}\gamma + \sigma_{V_w\varepsilon}$. Under Assumption **2.3**, it is a simple exercise to see that

$$n^{-1}(v_1, V_w)'(v_1, V_w) \xrightarrow{p} \Omega_W, \, n^{-1}Z'(v_1, V_w) \xrightarrow{p} 0, \tag{2.5}$$

$$vec\left((Z'Z/n)^{-1/2}n^{-1/2}Z'(v_1, V_w)\right) \xrightarrow{d} vec\left(\psi_{v_1}, \psi_{ZV_w}\right) \sim N\left(0, \Omega_W \otimes I_L\right). \tag{2.6}$$

We will now present the subset AR statistics studied.

2.2. Plug-in subset AR statistics

The plug-in principle usually consists of two steps. The first step is to take an identification-robust statistic [usually the Anderson and Rubin (1949, AR), Kleibergen (2002, K) and Moreira (2003, CLR) type statistics] which results from the test of the joint hypothesis $H(\beta_0, \gamma_0) : \beta = \beta_0, \, \gamma = \gamma_0$. The second step consists of replacing the nuisance parameters γ which are not specified by the subset null hypothesis of interest (2.3) by an estimator in the expression of the above statistics. In this paper, we consider the plug-in AR-type statistics that use the restricted LMIL and 2SLS estimators of γ under H_0, namely $\tilde{\gamma}_{LIML}$ and $\tilde{\gamma}_{2SLS}$, respectively. For convenience, we adopt the notation $\tilde{\gamma}_j,\ j \in \{LIML,\ 2SLS\}$ in the remainder of the paper. The subset AR statistic for H_0 corresponding to $\tilde{\gamma}_j$ can be expressed[4] as

$$AR(\beta_0, \tilde{\gamma}_j) = \frac{1}{L} \left\| \tilde{S}(\beta_0, \tilde{\gamma}_j) \right\|^2, \tag{2.7}$$

where $\tilde{S}(\beta_0, \tilde{\gamma}_j) = (Z'Z)^{-1/2} Z' \tilde{Y}(\beta_0) \tilde{r}_j (\tilde{r}_j' \hat{\Omega}_W \tilde{r}_j)^{-1/2}$, $\tilde{Y}(\beta_0) = [\tilde{y}(\beta_0) : W]$, $\tilde{y}(\beta_0) = y - X\beta_0$, $\hat{\Omega}_W = \frac{1}{n-L} \tilde{Y}(\beta_0)' M_Z \tilde{Y}(\beta_0)$, $\tilde{r}_j = (1, -\tilde{\gamma}_j)'$. It is well known that $\tilde{\gamma}_j$ is given by

$$\tilde{\gamma}_j = \left[W'(P_Z - \overline{\kappa}_j M_Z) W \right]^{-1} W' \left(P_Z - \overline{\kappa}_j M_Z \right) (y - X\beta_0), \tag{2.8}$$

where $\overline{\kappa}_{2SLS} = 0$ and $\overline{\kappa}_{LIML} = (n-L)^{-1} \tilde{\kappa}_{LIML}$, $\tilde{\kappa}_{LIML}$ is the smallest root of the characteristic polynomial $|\kappa \hat{\Omega}_W - \tilde{Y}(\beta_0)' P_Z \tilde{Y}(\beta_0)| = 0$. Furthermore, we have $AR(\beta_0, \tilde{\gamma}_{LIML}) = \min_{\gamma \in \mathbb{R}} AR(\beta_0, \gamma)$ [for example, see Guggenberger et al. (2012)], meaning that $AR(\beta_0, \tilde{\gamma}_{2SLS}) \geq AR(\beta_0, \tilde{\gamma}_{LIML})$. This means that the test that uses $\tilde{\gamma}_{2SLS}$ rejects more often the subset null hypothesis than those that uses $\tilde{\gamma}_{LIML}$ if the same critical value is applied.

Guggenberger et al. (2012) and Guggenberger and Chen (2011) show that a test of H_0 based on $AR(\beta_0, \tilde{\gamma}_{LIML})$ has a correct *asymptotic size* when the errors are homoskedastic even when γ is weakly identified. However, this test is overly conservative even in moderate samples when the identification of γ is weak and that the usual asymptotic χ^2-critical values are applied in the inference; see Doko Tchatoka (2014). Moreover, little is known about the size property of the test based on $AR(\beta_0, \tilde{\gamma}_{2SLS})$, especially when γ is not identified. In this paper, we characterize the asymptotic behavior of both tests, including when γ is not identified, and we investigate whether bootstrapping can improve their size.

In the remainder of the paper, let $G_{L-1}(\cdot)$ and $g_{L-1}(\cdot)$ denote the cumulative density function (cdf) and the probability density function (pdf), respectively, of a χ^2-distributed random variable with $L-1$ degrees of freedom. We now characterize the limiting distribution under H_0 of $AR(\beta_0, \tilde{\gamma}_j)$ for $j \in \{LIML,\ 2SLS\}$.

[4] See Stock and Wright (2000); Zivot, Startz and Nelson (2006); Guggenberger et al. (2012); and Doko Tchatoka (2014)

3. Asymptotic behavior of $AR(\beta_0, \tilde{\gamma}_j)$

In this section, we study the limiting behavior of $AR(\beta_0, \tilde{\gamma}_j),\ j \in \{LIML,\ 2SLS\}$, under the subset null hypothesis H_0 in both the strong and weak identification setups. Since strong identification is relatively easy to tackle, we will focus on that case first. Section 3.1 presents the results.

3.1. Strong identification

We focus first on the case in which γ is identified under H_0, i.e., $\Pi_w \neq 0$ is fixed. Although this setup is widely considered in many studies of weak instruments,[5] only a first-order approximation of the limiting distribution of the subset AR-statistic is provided. Here, we provide a high-order approximation of the limiting distributions of the statistics under H_0. Theorem **3.1** presents the results.

Theorem 3.1 *Suppose that Assumptions* **2.1 - 2.3** *are satisfied with* $r \geq 1$. *If further* H_0 *holds and* $\Pi_w \neq 0$ *is fixed, then we have*:

$$\sup_{\tau \in \mathbb{R}} \left| G_{AR_j}(\tau) - G_{L-1}(\tau) - \sum_{h=1}^{r} n^{-h} p^h_{AR_j}(\tau; F, \beta_0, \gamma, \Pi_x, \Pi_w) g_{L-1}(\tau) \right| = o(n^{-r})$$

for $j \in \{LIML, 2SLS\}$, *where* $p^h_{AR_j}$ *is a polynomial in* τ *with coefficients depending on* β_0, γ, Π_x, Π_w, *and the moments of the distribution* F *of* R_n, *and* $G_{AR_j}(\tau) = \mathbb{P}[AR(\beta_0, \tilde{\gamma}_j) \leq \tau]$ *is the cdf of* $AR(\beta_0, \tilde{\gamma}_j)$ *evaluated at* $\tau \in \mathbb{R}$.

Remark. Let $c_{AR_j}(\alpha) = \inf\left\{\tau \in \mathbb{R} : G_{AR_j}(\tau) \geq 1 - \alpha\right\}$ define the $1-\alpha$ quantile of the distribution of $AR(\beta_0, \tilde{\gamma}_j),\ j \in \{LIML, 2SLS\}$. Theorem **3.1** gives the conditions under which $c_{AR_j}(\alpha) \approx \chi^2_{L-1}(\alpha) + \sum_{h=1}^{r} n^{-h} q^h_{AR_j}(\chi^2_{L-1}(\alpha))$ uniformly in $\zeta < \alpha < 1 - \zeta$ for any $0 < \zeta < 1/2$ [see Hall (1992)], where the $q^h_{AR_j}$ are polynomials derivable from $p^h_{AR_j}$ and $\chi^2_{L-1}(\alpha)$ is the $1-\alpha$ quantile of a χ^2-distributed random variable with $L-1$ degrees of freedom, i.e., the solution of $G_{L-1}[\chi^2_{L-1}(\alpha)] = 1 - \alpha$. Thus, Theorem **3.1** provides a more greater accurate approximation of the distribution of $AR(\beta_0, \tilde{\gamma}_j)$ than the usual first-order asymptotic χ^2-distribution.

[5] For example, see Stock and Wright (2000); Kleibergen (2004); Startz et al. (2006); Mikusheva (2010); Guggenberger et al. (2012); Guggenberger and Chen (2011); and Doko Tchatoka (2014).

We will now study the null limiting behavior of $AR(\beta_0, \tilde{\gamma}_j)$ under Staiger and Stock's (1997) *local to zero* weak instruments setup.

3.2. Local to zero weak instruments

We now suppose that $\Pi_w = C_{0w}/\sqrt{n}$, where $C_{0w} \in \mathbb{R}^L$ is a fixed vector (possibly zero). Due to identification failure, high-order refinement of the distribution of $AR(\beta_0, \tilde{\gamma}_j)$, such as in Theorem **3.1**, is no longer possible. Because of this, we only derive the asymptotic distributions of the statistics under H_0. Since $AR(\beta_0, \tilde{\gamma}_j)$ depends on $\tilde{\gamma}_j$, $\tilde{S}(\beta_0, \tilde{\gamma}_j)$ and $\tilde{r}_j' \hat{\Omega}_W \tilde{r}_j$ for $j \in \{LIML, 2SLS\}$, it will be illuminating to study the asymptotic behavior of the latter statistics first. Lemma **3.2** presents the results, where

$$
\begin{aligned}
\Sigma_\rho &= \begin{pmatrix} 1 & \rho_{_{V_w\varepsilon}} \\ \rho_{_{V_w\varepsilon}} & 1 \end{pmatrix}, \quad \Psi = \sigma_{_{V_wV_w}}^{-1/2} Q_Z^{1/2} C_{0w} + \psi_{V_w}, \quad \psi_{V_w} = \sigma_{_{V_wV_w}}^{-1/2} Q_Z^{-1/2} \psi_{ZV_w}, \\
\psi_\varepsilon &= \sigma_{\varepsilon\varepsilon}^{-1} Q_Z^{-1/2} \psi_{Z\varepsilon}, \rho_{_{V_w\varepsilon}} = \left(\sigma_{\varepsilon\varepsilon}\sigma_{_{V_wV_w}}\right)^{-1/2} \sigma_{V_w\varepsilon}.
\end{aligned} \tag{3.1}
$$

Lemma 3.2 *Suppose Assumptions* **2.2**-**2.3** *and* H_0 *are satisfied. If* $\Pi_w = C_{0w}/\sqrt{n}$, *where* $C_{0w} \in \mathbb{R}^L$ *is fixed, then the following convergence holds jointly for* $j \in \{LIML, 2SLS\}$:

(*a*) $\tilde{\gamma}_j - \gamma_0 \xrightarrow{d} \sigma_{\varepsilon\varepsilon}^{1/2} \sigma_{_{V_wV_w}}^{-1/2} \Delta_j$, *where* $\Delta_j = (\Psi'\Psi - \kappa_j)^{-1}(\Psi'\psi_\varepsilon - \kappa_j \rho_{_{V_w\varepsilon}})$, $\kappa_{_{2SLS}} = 0$ *and* $\kappa_{_{LIML}}$ *is the smallest root of* $\left|(\psi_\varepsilon : \Psi)'(\psi_\varepsilon : \Psi) - \kappa\Sigma_\rho\right| = 0$;

(*b*) $\tilde{r}_j' \hat{\Omega}_W \tilde{r}_j \xrightarrow{d} \sigma_{\varepsilon\varepsilon,j} = \sigma_{\varepsilon\varepsilon}\left(1 - 2\rho_{_{V_w\varepsilon}}\Delta_j + \Delta_j^2\right)$;

(*c*) $\tilde{S}(\beta_0, \tilde{\gamma}_j) \xrightarrow{d} \left(1 - 2\rho_{_{V_w\varepsilon}}\Delta_j + \Delta_j^2\right)^{-1/2} S_j$, *where* $S_j = \psi_\varepsilon - \Psi\Delta_j$.

Remarks. **(i)** Lemma **3.2** - (a) shows that both the restricted LIML and 2SLS estimators under H_0 are inconsistent under Staiger and Stock's (1997) *local to zero* weak instruments setup [similar to Staiger and Stock (1997) and Doko Tchatoka (2014)]. Under Assumptions **2.2**-**2.3**, it is straightforward to see that $(\psi_\varepsilon', \psi_{V_w}')'$ is Gaussian with zero mean and covariance matrix $\Sigma_\rho \otimes I_L$. So, given ψ_{V_w}, ψ_ε follows a normal distribution with mean $\rho_{_{V_W\varepsilon}}\psi_{V_w}$ and covariance matrix $(1 - \rho_{_{V_W\varepsilon}}^2)I_L$. Therefore, the conditional distribution of $\tilde{\gamma}_{_{2SLS}} - \gamma_0$, given ψ_{V_w}, is Gaussian with mean $\rho_{_{V_W\varepsilon}}(\Psi'\Psi)^{-1}\Psi'\psi_{V_w}$ and covariance matrix $(1-\rho^2_{V_W\varepsilon})(\Psi'\Psi)^{-1}$. Since the mean and covariance matrix of the conditional distribution of $\tilde{\gamma}_{2SLS}-\gamma_0$, given ψ_{V_w}, depend only on ψ_{V_w}, its unconditional distribution is a mixture of Gaussian processes with nonzero means. This is not, however, the case for LIML. Indeed, since $\kappa_{LIML}\neq 0$ with probability one, the conditional distribution of $\tilde{\gamma}_{LIML}-\gamma_0$, given ψ_{V_w}, is not necessarily Gaussian. As a result, $\tilde{\gamma}_{LIML}-\gamma_0$ does not necessarily converge to a mixture of Gaussian processes; see Doko Tchatoka (2014).

(ii) Lemma **3.2** - (b) shows that $\tilde{r}'_j\hat{\Omega}_W\tilde{r}_j$ converges to a random process, rather than a constant scalar for all $j\in\{LIML,2SLS\}$ as in the case of strong identification. Meanwhile, Lemma **3.2** - (c) implies that $S(\beta_0,\tilde{\gamma}_j)$ does not have a standard normal distribution even for $j=2SLS$. Indeed, while $S(\beta_0,\tilde{\gamma}_{2SLS})$ converges to a mixture of Gaussian processes with nonzero mean under H_0, the null limiting distribution of $\tilde{S}(\beta_0,\tilde{\gamma}_j)$ is nonstandard and more complex to characterize because of the presence of κ_{LIML} in it.

We can now state the following theorem on the asymptotic behavior of $AR(\beta_0,\tilde{\gamma}_j)$ under H_0 for all $j\in\{LIML,2SLS\}$, when instruments are weak.

Theorem 3.3 *Suppose that Assumptions* **2.2**-**2.3** *are satisfied and H_0 holds. Let $\Pi_w = C_{0w}/\sqrt{n}$, where $C_{0w}\in\mathbb{R}^L$ is fixed (possibly zero). Then we have*:

$$AR(\beta_0,\tilde{\gamma}_j) \quad \xrightarrow{d} \quad \xi^{\infty}_{AR_j} = \frac{1}{L}\left\|\left(1-2\rho_{V_w\varepsilon}\Delta_j+\Delta_j^2\right)^{-1/2}S_j\right\|^2 \text{ for } j\in\{LIML,2SLS\},$$

where S_j, $\rho_{V_w\varepsilon}$ and Δ_j are defined in Lemma **3.2**.

Remarks. **(i)** Theorem **3.3** shows that the null limiting distribution of $AR(\beta_0,\tilde{\gamma}_j)$, $j\in\{LIML,2SLS\}$, is nonstandard and depends on nuisance parameters under Staiger and Stock's (1997) *local to zero* weak instruments framework. Therefore, using the asymptotic χ^2-critical values in the inference is not recommended in such contexts.

(ii) Guggenberger et al. (2012, eq.14) and Guggenberger and Chen (2011) provide an upper bound on the limiting distribution of $AR(\beta_0,\tilde{\gamma}_{LIML})$ and show that a test based on this statistic has *correct asymptotic size* even with weak instruments. However, it performs poorly even with relatively moderate sample sizes [see Doko Tchatoka (2014)]. Meanwhile, it is not clear from Lemma

3.2 and Theorem **3.3** whether this result on the upper of $AR(\beta_0, \tilde{\gamma}_{LIML})$ extends to $AR(\beta_0, \tilde{\gamma}_{2SLS})$. The Monte Carlo results show that a test based on $AR(\beta_0, \tilde{\gamma}_{2SLS})$ sometimes over-rejects the subset null hypothesis when identification is weak and endogeneity is large, while those based on $AR(\beta_0, \tilde{\gamma}_{LIML})$ remains overly conservative with weak instruments; see Table 1.

Based on the above results, it is natural to question whether bootstrapping can improve the size of these subset AR-tests, especially when γ is weakly identified. Section 4 addresses this issue.

4. Bootstrapping subset AR-test

Moreira et al. (2009) show that a residual-based bootstrap yields a test with correct size for the AR statistic of the joint hypothesis $H(\beta_0, \gamma_0) : \beta = \beta_0, \gamma = \gamma_0$ in model (2.1)-(2.2), without identifying assumptions on model parameters. However, unlike the AR statistic of $H(\beta_0, \gamma_0)$, the subset AR statistics of H_0 in (2.7) are not pivotal (even asymptotically) when γ is weakly identified (see Theorem **3.3**). So, the validity of the bootstrap for the subset AR statistics of H_0 is not obvious, at least when identification is weak, and further investigation is required. In this section, we examine the validity of the bootstrap similar to that of Moreira et al. (2009) for the subset AR statistics in (2.7). Before proceeding, let $\hat{\Pi}_x = (Z'Z)^{-1}Z'X$ and $\hat{\Pi}_w = (Z'Z)^{-1}Z'W$ denote the ordinary least squares estimators of Π_x and Π_w in (2.2). Let also $\tilde{\gamma}_j,\ j \in \{LIML, 2SLS\}$, denote the restricted LIML and 2SLS estimators of γ under H_0 given in (2.8). We adapt the bootstrap procedure by Moreira et al. (2009) to the subset AR statistics as follows.

1. For a given β_0 and the observed data, compute $\hat{\Pi}_x$, $\hat{\Pi}_w$ and $\tilde{\gamma}_j,\ j \in \{LIML, 2SLS\}$, along with all other items necessary to obtain the realizations of the statistic $AR(\beta_0, \tilde{\gamma}_j)$ and the residuals from the reduced-form equation (2.4): $\hat{v}_1(\beta_0) \equiv \hat{v}_1 = \tilde{y}(\beta_0) - Z\hat{\Pi}_w\tilde{\gamma}_j$, $\hat{V}_x(\beta_0) \equiv \hat{V}_x = X - Z\hat{\Pi}_x$, and $\hat{V}_w = W - Z\hat{\Pi}_w$. These residuals are then re-centered by subtracting sample means to yield $(\tilde{v}_1, \tilde{V}_x, \tilde{V}_w)$;

2. For each bootstrap sample $b = 1, \ldots, B$, the data are generated following

$$X^* = Z^*\hat{\Pi}_x + V_x^*, \tag{4.1}$$

$$W^* = Z^*\hat{\Pi}_w + V_w^*, \tag{4.2}$$

$$y^* = X^*\beta_0 + Z^*\hat{\Pi}_w\tilde{\gamma}_j + v_1^*, \tag{4.3}$$

where $(Z^*, v_1^*, V_x^*, V_w^*)$ is drawn independently from the joint empirical distribution of $(Z, \tilde{v}_1, \tilde{V}_x, \tilde{V}_w)$. The corresponding bootstrapping subset AR statistics $AR^{*^{(b)}}(\beta_0, \tilde{\gamma}_j^*), b = 1, \ldots, B,$ are computed as

$$AR^{*^{(b)}}(\beta_0, \tilde{\gamma}_j^*) = \frac{1}{L} \parallel \hat{S}^{*^{(b)}}(\beta_0, \tilde{\gamma}_j^*) \parallel^2, \tag{4.4}$$

$$\hat{S}^{*^{(b)}}(\beta_0, \tilde{\gamma}_j^*) = (Z^{*'}Z^*)^{-1/2}Z^{*'}\tilde{Y}^*(\beta_0)\tilde{r}_j^*(\tilde{r}_j^{*\prime}\hat{\Omega}_{1W}^*\hat{r}_j^*)^{-1/2}, \tag{4.5}$$

where $\tilde{Y}^*(\beta_0) = (\tilde{y}^*(\beta_0) : W^*)$ and $\tilde{r}_j^* = (1, -\tilde{\gamma}_j^*)'$;

3. The bootstrap test rejects H_0 if $\frac{1}{B}\sum_{b=1}^{B} \mathbb{1}[AR^{*^{(b)}}(\beta_0, \tilde{\gamma}_j^*) > c_{AR_j}^*(\alpha)]$ is less then α, where $c_{AR_j}^*(\alpha)$ is the $1-\alpha$ quantile of $AR^{*^{(b)}}(\beta_0, \tilde{\gamma}_j^*)$, i.e., the value that minimizes $\mid \mathbb{P}^*[AR^{*^{(b)}}(\beta_0, \tilde{\gamma}_j^*) \leq \tau] - (1-\alpha) \mid$ over $\tau \in \mathbb{R}$; see Andrews (2002).

In the reminder of the paper, F_n denotes the empirical distribution of $R_n^* = vech\left((\tilde{Y}_n^{*'}, Z_n^{*'})'(\tilde{Y}_n^{*'}, Z_n^{*'})\right)$ conditional on $\mathscr{X}_n = \left\{(\tilde{Y}_1', Z_1'), \ldots, (\tilde{Y}_n', Z_n')\right\}$, $\mathbb{P}^*$ is the probability under the empirical distribution function (conditional on $\mathscr{X}_n$), and $\mathbb{E}^*$ its corresponding expectation operator. As in Section 3, we deal separately with the case where identification is strong and the one where it is weak. Section 4.1 presents the results for strong identification.

4.1. Bootstrap consistency under strong identification

We first focus on the case in which γ is identified and we provide an analysis of the behavior of the proposed bootstrap subset AR statistics and the associated tests under H_0. Lemma **4.1** states the validity of high-order expansion for all bootstrap subset AR statistics considered and Theorem **4.2** provides a high-order refinement of the size of the corresponding bootstrap subset AR tests.

Lemma 4.1 *Suppose that Assumptions* **2.1 - 2.3** *are satisfied with* $r \geq 1$. *Suppose further that* H_0

holds and $\Pi_w \neq 0$ is fixed. Then we have:

$$\sup_{\tau\in\mathbb{R}} \left| G_{AR_j^*}(\tau) - G_{L-1}(\tau) - \sum_{h=1}^{r} n^{-h} p^h_{AR_j}(\tau; F_n, \beta_0, \tilde{\gamma}_j, \hat{\Pi}_x, \hat{\Pi}_w) g_{L-1}(\tau) \right| = o(n^{-r}) \ a.s.$$

for $j \in \{LIML, 2SLS\}$, $p^h_{AR_j}$ is a polynomial in τ with coefficients depending on β_0, $\tilde{\gamma}_j$, $\hat{\Pi}_x$, $\hat{\Pi}_w$ and the moments of F_n, and $G_{AR_j^*}(\tau) = \mathbb{P}^*[AR^*(\beta_0, \tilde{\gamma}_j^*) \leq \tau]$ is the empirical cdf of $AR^*(\beta_0, \tilde{\gamma}_j^*)$ evaluated at $\tau \in \mathbb{R}$.

Remark. Lemma **4.1** shows that the bootstrap estimate and the $(r+1)$-term empirical Edgeworth expansion in Theorem **3.1** are asymptotically equivalent up to the $o(n^{-r})$ order under H_0. Furthermore, the bootstrap makes an error of size $O(n^{-1})$ under H_0, which is smaller as $n \to +\infty$ than both $O(n^{-1/2})$ and the error made by the first-order asymptotic approximation. The bootstrap provides a greater accuracy than the $O(n^{-1/2})$ order because each subset AR statistic in (2.7) is a quadratic function of a symmetric pivotal statistic [see Horowitz (2001, Ch. 52, eq. 3.13)] under H_0 if γ is identified ($\Pi_w \neq 0$ is fixed).

We can now prove the following theorem on the size of the tests when bootstrap critical values are used in the inference.

Theorem 4.2 *Suppose that Assumptions* **2.1 - 2.3** *are satisfied with $r \geq 1$. Suppose further that H_0 holds and $\Pi_w \neq 0$ is fixed. Then we have*:

$$\mathbb{P}[AR(\beta_0, \tilde{\gamma}_j) > c^*_{AR_j}] = \alpha + o(n^{-1}) \ \text{ for all } \ j \in \{LIML, 2SLS\},$$

where $c^*_{AR_j} \equiv c^*_{AR_j}(\alpha)$ is the $1-\alpha$ quantile of the empirical distribution of $AR^*(\beta_0, \tilde{\gamma}_j^*)$, i.e., the value that minimizes $|\, \mathbb{P}^*[AR^*(\beta_0, \tilde{\gamma}_j^*) \leq \tau] - (1-\alpha)\,|$ over $\tau \in \mathbb{R}$.

Remarks. **(i)** Theorem **4.2** gives the conditions under which the bootstrap critical values for $AR^*(\beta_0, \tilde{\gamma}_j^*)$ yield level for the $AR(\beta_0, \tilde{\gamma}_j)$ test that is correct through $O(n^{-1})$ under H_0. So, the bootstrap makes an error of size $O(n^{-1})$ under H_0 when γ is identified ($\Pi_w \neq 0$ is fixed) for all values of Π_x. So, the identification of the parameters specified by the subset null hypothesis (here β) does not affect the results of Theorem **4.2**.

(ii) Although Theorem **4.2** focuses on the validity of the bootstrap under H_0, there is no impediment to expanding the above results under the alternative hypothesis ($\beta \neq \beta_0$). In this case, we can show that the tests are consistent when the bootstrap critical values are used in the inference and the instruments are strong for both X and W (i.e., if $\Pi_x \neq 0$ and $\Pi_w \neq 0$ are fixed). This is because the asymptotic distribution of $AR(\beta_0, \tilde{\gamma}_j)$ diverges when $\beta \neq \beta_0$, and $\Pi_x \neq 0$, $\Pi_w \neq 0$ are fixed. Note that test consistency will still hold in this case if the empirical *size-corrected* critical values were used instead of the bootstrap critical values, as suggested by Horowitz (1994). However, the tests will have low power for weak values of Π_x even if γ is identified ($\Pi_w \neq 0$ is fixed). These proofs are omitted to shorten the exposition of our results.

We shall now focus on Staiger and Stock's (1997) *local to zero* weak instruments framework.

4.2. Bootstrap inconsistency under weak instruments

We now assume that $\Pi_w = C_{0w}/\sqrt{n}$, where $C_{0w} \in \mathbb{R}^L$ is fixed (possibly zero), and we study the validity of the bootstrap for the above subset AR tests. As in Section 3.2, we first characterize the asymptotic behavior of the bootstrap statistics $\tilde{\gamma}_j^*$, $\tilde{r}_j^{*\prime}\hat{\Omega}_{1W}^*\tilde{r}_j^*$, and $\hat{S}^*(\beta_0, \tilde{\gamma}_j^*)$ under H_0. Lemma **4.3** presents the results.

Lemma 4.3 *Suppose Assumptions* **2.2-2.3** *are satisfied and* H_0 *holds. Let* $\Pi_w = C_{0w}/\sqrt{n}$, *where* $C_{0w} \in \mathbb{R}^L$ *is fixed. If for some* $\delta > 0$, $\mathbb{E}(\|Z_i\|^{4+\delta}, \|V_i\|^{2+\delta}) < +\infty$, *then we have*:

(*a*) $\tilde{\gamma}_j^* - \tilde{\gamma}_j \mid \mathscr{X}_n \xrightarrow{d} \sigma_{\varepsilon\varepsilon,j}^{1/2}\sigma_{V_wV_w}^{-1/2}\Delta_j^B$ *a.s.*;

(*b*) $\tilde{r}_j^{*\prime}\hat{\Omega}_{1W}^*\tilde{r}_j^* \mid \mathscr{X}_n \xrightarrow{d} \sigma_{\varepsilon\varepsilon,j}\left(1 - 2\rho_{V_w\varepsilon,j}\Delta_j^B + (\Delta_j^B)^2\right)$ *a.s.*;

(*c*) $\hat{S}^*(\beta_0, \tilde{\gamma}_j^*) \mid \mathscr{X}_n \xrightarrow{d} \left(1 - 2\rho_{V_w\varepsilon,j}\Delta_j^B + (\Delta_j^B)^2\right)^{-1/2} S_j^B$ *a.s.*

for $j \in \{LIML, 2SLS\}$ *where* $\Delta_j^B = \left(\Psi^{B\prime}\Psi^B - \kappa_j^B\right)^{-1}\left(\Psi^{B\prime}\psi_{\varepsilon,j} - \kappa_j^B\rho_{V_w\varepsilon,j}\right)$, *when* $\kappa_{2SLS}^B = 0$ *and* κ_{LIML}^B *is the smallest root of the determinantal equation* $\left|\left(\psi_{\varepsilon,j} : \Psi^B\right)'\left(\psi_{\varepsilon,j} : \Psi^B\right) - \kappa\Sigma_{\rho,j}\right| = 0$, $\Psi^B = \Psi + \psi_{V_w}$, $\psi_{\varepsilon,j} = \psi_{V_1} - \psi_{V_w}\left(\gamma_0 + \sigma_{\varepsilon\varepsilon}^{1/2}\sigma_{V_wV_w}^{-1/2}\Delta_j\right) = \psi_\varepsilon - \sigma_{\varepsilon\varepsilon}^{1/2}\sigma_{V_wV_w}^{-1/2}\Delta_j\psi_{V_w}$, $\Sigma_{\rho,j} = \begin{pmatrix} 1 & \rho_{V_w\varepsilon,j} \\ \rho_{V_w\varepsilon,j} & \end{pmatrix}$, $\rho_{V_w\varepsilon,j} = (\sigma_{\varepsilon\varepsilon,j}\sigma_{V_wV_w})^{-1/2}\sigma_{V_w\varepsilon,j}$, $\sigma_{\varepsilon\varepsilon,j} = \sigma_{\varepsilon\varepsilon}\left(1 - 2\rho_{V_w\varepsilon}\Delta_j + (\Delta_j)^2\right)$, $\sigma_{V_w\varepsilon,j} =$

$\sigma_{V_w\varepsilon} - \sigma_{\varepsilon\varepsilon}^{1/2}\sigma_{V_wV_w}^{1/2}\Delta_j,\ S_j^B = \psi_{\varepsilon,j} - \Psi^B\Delta_j^B,$ *and* Δ_j *is given in Lemma* **3.2**.

Remarks. **(i)** First, we note that the convergence results in Lemma **4.3** (a)-(c) differ from those in Lemma **3.2** (a)-(c). This means that the bootstrap fails to mimic the asymptotic distributions of all statistics $\tilde{r}_j'\hat{\Omega}_{1W}\tilde{r}_j$ and $\hat{S}(\beta_0,\tilde{\gamma}_j)$ under H_0 when $\tilde{\gamma}_j$ is used as the pseudo true value of γ. Indeed, while $\tilde{\gamma}_j - \gamma_0 \xrightarrow{d} \sigma_{\varepsilon\varepsilon}^{1/2}\sigma_{V_wV_w}^{-1/2}\Delta_j$ in Lemma **3.2**-(a), we have $\tilde{\gamma}_j^* - \tilde{\gamma}_j \mid \mathscr{X}_n \xrightarrow{d} \sigma_{\varepsilon\varepsilon,j}^{1/2}\sigma_{V_wV_w}^{-1/2}\Delta_j^B$ *a.s.* by Lemma **4.3**-(a). Both limits can differ significantly since $\Delta_j = (\Psi'\Psi - \kappa_j)^{-1}\left(\Psi'\psi_\varepsilon - \kappa_j\rho_{V_w\varepsilon}\right) \neq \Delta_j^B = \left(\Psi^{B'}\Psi^B - \kappa_j^B\right)^{-1}\left(\Psi^{B'}\psi_{\varepsilon,j} - \kappa_j^B\rho_{V_w\varepsilon,j}\right)$ and $\sigma_{\varepsilon\varepsilon} \neq \sigma_{\varepsilon\varepsilon,j}$ with probability 1.

(ii) The bootstrap failure is mainly due to the fact that replacing Π_w and γ by $\hat{\Pi}_w$ and $\tilde{\gamma}_j$, respectively, in (2.4) generates an extra noise term that is added to the original residuals when re-sampled. This can lead to substantial differences between the limits in the original sample and those in the bootstrap sample. To be more explicit, observe that in the local to zero weak instruments setup, we have $Z'W/\sqrt{n} = (Z'Z/n)C_{0w} + Z'V_w/\sqrt{n} \xrightarrow{d} Q_ZC_{0w} + \psi_{\mathbf{ZV_w}}$ under Assumptions **2.2-2.3**. Meanwhile, $Z^{*'}W^*/\sqrt{n} = (Z^{*'}Z^*/\sqrt{n})\hat{\Pi}_w + Z^{*'}V_w^*/\sqrt{n} = (Z^{*'}Z^*/n)C_{0w} + (Z^{*'}Z^*/n)(Z'Z/n)^{-1}(Z'V_w/\sqrt{n}) + Z^{*'}V_w^*/\sqrt{n} \mid \mathscr{X}_n \xrightarrow{d} Q_ZC_{0w} + \mathbf{2}\psi_{\mathbf{ZV_w}}$ *a.s.* under the conditions of Lemma **4.3**. Hence, the bootstrap fails to mimic the limiting behavior of $Z'W/\sqrt{n}$ (therefore that of the concentration factor $W'P_ZW$) under Staiger and Stock's (1997) *local to zero* weak instruments asymptotic, and similarly for the other statistics (for example, see the difference between Ψ and Ψ^B in Lemma **4.3**).

We can prove the following theorem on the limiting distribution of $AR^*(\beta_0,\tilde{\gamma}_j^*)$ under H_0 and weak instruments.

Theorem 4.4 *Suppose Assumptions* **2.2-2.3** *are satisfied and* H_0 *holds. Let* $\Pi_w = C_{0w}/\sqrt{n}$, *where* $C_{0w} \in \mathbb{R}^L$ *is fixed. If for some* $\delta > 0$, $\mathbb{E}(\|Z_i\|^{4+\delta}, \|V_i\|^{2+\delta}) < +\infty$, *then we have*:

$$AR^*(\beta_0,\tilde{\gamma}_j^*) \mid \mathscr{X}_n \quad \xrightarrow{d} \quad \xi_{AR_j^*}^\infty = \left\|\left(1 - 2\rho_{V_w\varepsilon}\Delta_j^B + (\Delta_j^B)^2\right)^{-1/2} S_j^B\right\|^2 \ a.s.,\ j \in \{LIML, 2SLS\}\,.$$

Remark. Theorem **4.4** provides a characterization of the limiting distribution of the bootstrap subset AR statistics under H_0 similar to that in Theorem **3.3**. As can be seen, a straightforward comparison confirms our previous analysis in Lemma **4.3**, i.e., $AR^*(\beta_0,\tilde{\gamma}_j^*)$ and $AR(\beta_0,\tilde{\gamma}_j)$ have different asymptotic behavior under H_0. So, we can state the following theorem on the inconsistency of the bootstrap for subset AR tests under weak instruments.

Theorem 4.5 *Suppose Assumptions* **2.2-2.3** *are satisfied and* H_0 *holds. Let* $\Pi_w = C_{0w}/\sqrt{n}$*, where* $C_{0w} \in \mathbb{R}^L$ *is fixed. If for some* $\delta > 0$, $\mathbb{E}(\|Z_i\|^{4+\delta}, \|V_i'\|^{2+\delta}) < +\infty$, *then we have*:

$$\left|\mathbb{P}[AR(\beta_0, \tilde{\gamma}_j) > c^*_{AR_j}] - \alpha\right| = O_p(1) \text{ for all } \; j \in \{LIML, 2SLS\},$$

where $c^*_{AR_j} \equiv c^*_{AR_j}(\alpha)$ *is the* $1-\alpha$ *quantile of the empirical distribution of* $AR^*(\beta_0, \tilde{\gamma}^*_j)$.

Remarks. **(i)** Theorem **4.5** shows that all subset tests do not have a correct size asymptotically when the bootstrap critical values are used in the inference. So, the bootstrap is inconsistent when the nuisance parameter, γ, that is not specified by the subset null hypothesis of interest, is weakly identified.

(ii) In the econometric and statistical literature, subsampling is often considered as a more robust resampling technique than bootstrap, and it is valid in many cases where bootstrap typically fails; for example, see Andrews and Guggenberger (2009). In Section A.1 of the appendix, we show that subsampling is invalid for the $AR(\beta_0, \tilde{\gamma}_j)$ subset tests under Staiger and Stock's (1997) *local to zero* weak instruments asymptotic.

We will now complete our analysis through a Monte Carlo experiment.

5. Monte Carlo experiment

We use simulation to examine the performance of both the standard and bootstrap subset AR tests described above. The data generating process is described by (2.1) and (2.2) where y, X and W are $n \times 1$ vectors. The n rows of $[\varepsilon : V_x : V_w]$ are i.i.d. normal with zero mean, and covariance matrix $\Sigma = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$, where the endogeneity ρ varies in $\{0, 0.1, 0.5, 0.9\}$. The L columns of the instrument matrix Z, $L \in \{3, 5, 10, 20\}$, are $N(0, I_L)$, independently from $[\varepsilon : V_x : V_w]$. Note that Z is fixed over the simulation experiment but the results do not change qualitatively otherwise. The true values of β and γ are set at 2 and -1, respectively. The reduced-form coefficient matrix $\Pi_{xw} = [\Pi_x : \Pi_w]$ is chosen as $\Pi_{xw} = \left(\frac{\mu^2}{n\|ZC_0\|}\right)^{1/2} \Pi_0$, where Π_0 is obtained by taking the first 2 columns of an identity matrix of order L, and μ^2 is the concentration parameter which describes the strength of Z. In this experiment, we vary μ^2 in $\{0, 13, 1000\}$, where $\mu^2 = 0$ is a complete non-identification or irrelevant IVs setup, $\mu^2 = 13$ represents weak instruments (both β and γ are weakly identified), and $\mu^2 = 1000$ implies strong instruments; see Guggenberger (2010) and Doko Tchatoka (2014) for a similar parametrization. The rejection frequencies are computed using $1,000$ replications for the standard subset AR tests, while those of the bootstrap subset AR tests are obtained with $N = 1,000$ replications and $B = 299$ bootstrap pseudo-samples of size $n = 100$. The nominal level of all tests is set at 5%.

Table 1 presents the empirical rejection frequencies of both the standard and bootstrap subset AR tests that use the restricted LIML and 2SLS estimators as plug-in estimators. The first column of the table shows the statistics. The second column indicates the number of instruments (L) used in the plug-in procedure. The other columns show, for each value of endogeneity ρ and instrument quality μ^2, the rejection frequencies.

First, we note that when the restricted LIML estimator is used as a plug-in estimator and the usual asymptotic χ^2 critical values are applied, the resulting subset AR test is overly conservative with weak instruments (see columns $\mu^2 \in \{0, 13\}$ in the table). These results are similar to those in Doko Tchatoka (2014). However, the test has rejections close to the nominal 5% level when identification is strong (see columns $\mu^2 = 1000$ in the table), thus confirming our theoretical findings in Section 3.1. Note however that the rejection frequencies of this test are slightly greater than the nominal 5% level as both endogeneity (ρ) and the number of instruments (L) increase (for example, with $\rho = 0.9$ and $L = 20$, the rejection frequency is around 7%). Meanwhile, the subset AR test that uses the restricted 2SLS estimator as the plug-in based estimator is less conservative than those with LIML when instruments are weak and the asymptotic χ^2 critical values are applied (see columns $\mu^2 \in \{0, 13\}$ in the second block of Table 1). This is not surprising because we always have $AR(\beta_0, \tilde{\gamma}_{_{2SLS}}) \geq AR(\beta_0, \tilde{\gamma}_{_{LIML}})$. In particular, the rejection frequencies of this test are close to the nominal 5% level with weak instruments and small endogeneity (see columns $\rho \in$

$\{0, 0.1\}$ and $\mu^2 \in \{0, 13\}$ in the table), but they are greater than 5% for large endogeneity (see column $\rho = 0.9$ and $\mu^2 \in \{0, 13\}$). We also observe that this test over-rejects sometimes when identification is strong. For example, the rejection frequencies when $\mu^2 = 1000$ (strong instruments) are 9.3%, 16.5%, and 40.7% for $L = 5,\ 10,\ 20$ instruments, respectively. So, while the subset AR test that uses the restricted 2SLS estimator seems to outperform the one that uses LIML under weak instruments and small endogeneity, its finite-sample size property is worse than the test with LIML when identification is strong and endogeneity is large.

Second, we observe that bootstrapping does not improve the size property of either test when identification is weak, as shown in columns $\mu^2 \in \{0, 13\}$ of the last block of Table 1. This confirms the inconsistency of the bootstrap for subset AR tests when identification is weak (see Section 4.2). However, the bootstrap provides a better approximation of the size of the tests than the asymptotic critical values when identification is strong and the number of instruments is moderate, especially in the case of restricted LIML estimator. Note that even in the case of restricted 2SLS estimator, the bootstrap has improved the size of the test for a moderate or large number of instruments ($L = 10, 20$) and large endogeneity ($\rho = 0.9$). For example, the rejection frequencies of the test when $\mu^2 = 1000$ and $\rho = 0.9$ are 6.4% and 14.1% for $L = 10, 20$, respectively. This represents a huge drop compared with the usual asymptotic critical values where these rejection frequencies were 16.5%, and 40.7%, respectively.

6. Conclusions

In this paper, we focus on linear IV regressions and study the asymptotic validity of the bootstrap for the plug-in based subset AR type-tests. Specifically, we suggest a bootstrap method similar to those of Moreira et al. (2009) for the score test of the null hypothesis specified on the full set of the structural parameters. We stress the fact that testing subset hypotheses is substantially more complex than testing joint hypotheses on the full set of parameters, so, the validity of Moreira et al.'s (2009) bootstrap for tests of subvectors is not obvious, especially when identification is weak. Our analysis considers two plug-in subset AR statistics. The first one uses the restricted limited information maximum likelihood (LIML) as the plug-in estimator of the nuisance parameters, and

Table 1. Rejection frequencies (in %) at 5% nominal level

Asymptotic χ^2 critical values													
		$\rho = 0$			$\rho = 0.1$			$\rho = 0.5$			$\rho = 0.9$		
Statistics	L $\downarrow$ $\mu^2 \rightarrow$	0	13	1000	0	13	1000	0	13	1000	0	13	1000
		Restricted LIML											
AR	3	0.4	0.9	5.2	0.3	0.9	5.3	0.1	0.5	5.0	0.3	1.8	5.2
-	5	0.3	0.2	5.4	0.3	0.3	4.1	0.1	0.3	5.6	0.3	0.9	5.6
-	10	0.3	0.3	6.2	0.1	0.0	5.7	0.5	0.2	5.1	0.3	0.6	6.5
-	20	0.4	0.2	4.3	0.3	0.4	4.6	0.4	0.6	6.2	0.2	0.5	7.0
		Restricted 2SLS											
AR	3	3.0	3.5	6.4	2.3	3.4	5.6	2.5	4.5	5.6	3.2	12.9	6.4
-	5	3.5	3.4	6.8	3.4	2.8	7.0	3.1	5.0	5.5	3.5	9.5	9.3
-	10	3.8	5.5	5.3	4.9	3.4	4.8	4.3	4.8	8.4	4.4	8.8	16.5
-	20	6.5	6.3	7.3	6.3	6.9	7.3	5.5	8.2	11.6	7.7	7.7	40.7
Bootstrap critical values													
		$\rho = 0$			$\rho = 0.1$			$\rho = 0.5$			$\rho = 0.9$		
Statistics	L $\downarrow$ $\mu^2 \rightarrow$	0	13	1000	0	13	1000	0	13	1000	0	13	1000
		Restricted LIML											
AR	3	0.7	1.2	5.8	0.4	0.7	4.8	1.3	0.7	4.8	1.6	1.0	4.9
-	5	0.2	0.4	5.4	0.6	0.4	4.3	0.7	0.3	5.6	1.4	1.5	5.4
-	10	0.3	0.1	5.1	0.4	0.0	5.2	0.3	0.4	4.7	0.3	0.3	5.1
-	20	0.2	0.1	3.2	0.1	0.1	3.1	0.0	0.6	3.6	0.2	0.0	3.8
		Restricted 2SLS											
AR	3	4.9	5.9	6.8	3.2	5.4	4.1	3.6	5.5	3.6	3.5	4.3	7.6
-	5	2.2	3.2	6.7	4.0	2.3	4.1	2.2	10.0	2.1	0.6	10.7	7.3
-	10	1.6	6.7	4.4	1.7	3.7	2.8	2.5	4.8	7.4	1.6	4.7	6.4
-	20	4.5	3.2	5.7	4.4	3.3	3.3	2.8	2.7	6.2	4.0	5.6	14.1

the second uses the restricted two-stage least squares (2SLS) as the plug-in method.

We provide a characterization of the asymptotic distributions of both the standard and proposed bootstrap AR statistics under the subset null hypothesis of interest. Our results provide some new insights and extensions of earlier studies. We show that when identification is strong, the bootstrap provides a high-order approximation of the null limiting distributions of both plug-in subset statistics. However, the bootstrap is inconsistent when instruments are weak. This contrasts with the bootstrap of the AR statistic of the null hypothesis specified on the full vector of structural parameters, which remains valid even when identification is weak; see Moreira et al. (2009). The inconsistency of bootstrap is mainly due to its failure to mimic the concentration parameter that characterizes the strength of the identification of the nuisance parameters (which are not specified by the null hypothesis of interest) when instruments are weak. Furthermore, we show that alternative resampling techniques, such as subsampling, which usually offer a good approximation of the size of the tests in many cases where bootstrap typically fails [see Andrews and Guggenberger (2009)], are also invalid under Staiger and Stock's (1997) *local to zero* weak instruments asymptotic; see Section A.1 of the appendix. We present a Monte Carlo experiment that confirms our theoretical findings.

A. Appendix

We begin by establishing the inconsistency of the subsampling method. The subsequent subsections will present the supplemental lemmata and proofs.

A.1. Subsampling inconsistency under weak instruments

In this section, we establish the invalidity of subsampling for $AR(\beta_0, \tilde{\gamma}_j),\ j \in \{LIML, 2SLS\}$, under Staiger and Stock's (1997) *local to zero* weak instruments asymptotic.

Let b be the subsample size, $G^b_{AR_j}(\tau)$ the empirical distribution function of the subsample statistics $\{AR^b_l(\beta_0, \tilde{\gamma}_{j,b}) : l = 1, \ldots, q_n\}$, evaluated at $\tau \in \mathbb{R}$, and $c^b_{AR_j} \equiv c^b_{AR_j}(\alpha)$ the $1-\alpha$ sample quantile of $G^b_{AR_j}(\cdot)$, where q_n is the number of different subsamples of size b. With i.i.d. observations, there are $q_n = n!/(n-b)!b!$ different subsamples of size b. With time series observations, there are

$q_n = n - b + 1$ subsamples each consisting of b consecutive observations. We assume that $b > L$. The subsampled AR statistic, $AR_{b,l}(\beta_0, \tilde{\gamma}_{j,b})$, is defined as

$$AR_l^b(\beta_0, \tilde{\gamma}_{j,b}) = \frac{1}{L} \parallel S_l^b(\beta_0, \tilde{\gamma}_{j,b}) \parallel^2, \tag{A.1}$$

where $S_l^b(\beta_0, \tilde{\gamma}_{j,b}) = \left(Z_b' Z_b\right)^{-1/2} Z_b' \tilde{Y}_b(\beta_0) \tilde{r}_{j,b} \left(\tilde{r}_{j,b}' \hat{\Omega}_{W_b} \tilde{r}_{j,b}\right)^{-1/2}$, $\hat{\Omega}_{W_b} = \frac{1}{n-L} \tilde{Y}_b'(\beta_0) M_{Z,b} \tilde{Y}_b'(\beta_0)$, $\tilde{Y}_b(\beta_0) = [y_b - X_b\beta_0 : W_b]$, $\tilde{r}_{j,b} = (1, -\tilde{\gamma}_{j,b})'$, $j \in \{2SLS, LIML\}$, and

$$\tilde{\gamma}_{j,b} = \left[W_b' \left(P_{Z_b} - \overline{\kappa}_{j,b} M_{Z_b}\right) W_b\right]^{-1} W_b' \left(P_{Z_b} - \overline{\kappa}_{j,b} M_{Z_b}\right) (y_b - X_b\beta_0) \tag{A.2}$$

with $\overline{\kappa}_{2SLS,b} = 0$ and $\overline{\kappa}_{LIML,b} = (n-L)^{-1} \tilde{\kappa}_{LIML,b}$ where $\tilde{\kappa}_{LIML,b}$ is the smallest root of the characteristic polynomial $|\kappa \hat{\Omega}_{W_b} - \tilde{Y}_b(\beta_0)' P_{Z_b} \tilde{Y}_b(\beta_0)| = 0$. As in the bootstrap case, the subsampling subset AR test based on $AR_l^b(\beta_0, \tilde{\gamma}_{j,b})$ rejects H_0 if

$$AR_l^b(\beta_0, \tilde{\gamma}_{j,b}) > c_{j,b}(1-\alpha). \tag{A.3}$$

In what follows, $\mathbb{P}^b$ denotes the probability under the subsampling empirical distribution function conditional on $\mathscr{X}_n = \left\{(\tilde{Y}_1', Z_1'), \ldots, (\tilde{Y}_n', Z_n')\right\}$ and $\mathbb{E}^b$ its corresponding expectation operator.

Lemma **A.1** characterizes the asymptotic null distributions of the subsampled statistics $\tilde{\gamma}_{j,b}$, $\tilde{r}_{j,b}' \hat{\Omega}_{W_b} \tilde{r}_{j,b}$, and $S_{b,l}(\beta_0, \tilde{\gamma}_{j,b})$ under Staiger and Stock's (1997) *local to zero* weak instruments setup, while Theorem **A.2** is concerned with the behavior of the subsampled subset AR statistics $AR_{b,l}(\beta, \tilde{\gamma}_{j,b})$.

Lemma A.1 *Suppose Assumptions* **2.2-2.3** *are satisfied and H_0 holds. Let $\Pi_w = C_{0w}/\sqrt{n}$, where $C_{0w} \in \mathbb{R}^L$ is fixed. If $b \to \infty$, and $b/n \to 0$ as $n \to \infty$, then the following convergence holds jointly for $j \in \{LIML, 2SLS\}$:*

(a) $\tilde{\gamma}_{j,b} - \gamma_0 \mid \mathscr{X}_n \xrightarrow{d} \sigma_{\varepsilon\varepsilon}^{1/2} \sigma_{V_w V_w}^{-1/2} \Delta_j^S$ *a.s.*;

(b) $\tilde{r}_{j,b}' \hat{\Omega}_{W_b} \tilde{r}_{j,b} \mid \mathscr{X}_n \xrightarrow{d} \sigma_{\varepsilon\varepsilon} \left(1 - 2\rho_{V_w\varepsilon} \Delta_j^S + (\Delta_j^S)^2\right)$ *a.s.*;

(c) $S_l^b(\beta_0, \tilde{\gamma}_{j,b}) \mid \mathscr{X}_n \xrightarrow{d} \left(1 - 2\rho_{V_w\varepsilon} \Delta_j^S + (\Delta_j^S)^2\right)^{-1/2} S_j^S$ *a.s.*, *where* $\Delta_j^S = (\psi_{V_w}'\psi_{V_w} - \kappa_j^S)^{-1}(\psi_{V_w}'\psi_\varepsilon - \kappa_j^S\rho_{V_w\varepsilon})$, $\Psi^S = \psi_{V_W}$, $\kappa_{2SLS}^S = 0$ *and* κ_{LIML}^S *is the smallest root of the determinantal equation* $\left|\left(\psi_\varepsilon : \psi_{V_w}\right)'\left(\psi_\varepsilon : \psi_{V_w}\right) - \kappa\Sigma_\rho\right| = 0$; $S_j^S = \psi_\varepsilon - \psi_{V_w}\Delta_j^S$.

Remark. As for the case of bootstrap [see Lemma **4.3** (a)-(c)], the convergence results in Lemma **A.1** (a)-(c) differ from that of Lemma **3.2** (a)-(c). So, the subsampling fails to mimic the asymptotic distributions of the statistics $\tilde{r}_j'\hat{\Omega}_{1W}\tilde{r}_j$ and $\hat{S}(\beta_0, \tilde{\gamma}_j)$ under H_0 and weak instruments.

We can now state the following theorem on the inconsistency of subsampling for the subset AR tests.

Theorem A.2 *Suppose Assumptions* **2.2-2.3** *are satisfied and* H_0 *holds. Let* $\Pi_w = C_{0w}/\sqrt{n}$, *where* $C_{0w} \in \mathbb{R}^L$ *is fixed. If* $b \to +\infty$, *and* $b/n \to 0$ *as* $n \to +\infty$, *then we have*:

$$AR_{b,l}(\beta_0, \tilde{\gamma}_{j,b}) \mid \mathscr{X}_n \overset{d}{\to} \left\| \left(1 - 2\rho_{V_w\varepsilon}\Delta_j^S + (\Delta_j^S)^2\right)^{-1/2} S_j^S \right\|^2 \text{ a.s. for all } j \in \{LIML, 2SLS\}.$$

Remark. Theorem **A.2** confirms our previous analysis in Lemma **A.1**. It is straightforward to see that the limiting distributions of Theorem **A.2** are different from those in Theorem **3.3**, thus indicating the subsampling invalidity with weak instruments.

An alternative subsampling method that is used in the literature is the jackknife resampling technique; see Wu (1990). This method differs from the previous subsampling procedure only by assuming that $b \to +\infty$ and $b/n \to f > 0$, as $n \to +\infty$. Berkowitz, Caner and Fang (2012) show that this method yields a test that controls the level of the AR test for full vector hypothesis without identifying assumptions on model parameters when IVs violate locally the orthogonality conditions. Lemma **A.3** and Theorem **A.4** show that this method is also invalid for the subset AR type-tests under weak instruments.

Lemma A.3 *Suppose Assumptions* **2.2-2.3** *are satisfied and* H_0 *holds. Let* $\Pi_w = C_{0w}/\sqrt{n}$, *where* $C_{0w} \in \mathbb{R}^L$ *is fixed. If* $b \to +\infty$, *and* $b/n \to f > 0$ *as* $n \to +\infty$, *then the following convergence holds jointly for* $j \in \{LIML, 2SLS\}$:

(*a*) $\tilde{\gamma}_{j,b} - \gamma_0 \mid \mathscr{X}_n \overset{d}{\to} \sigma_{\varepsilon\varepsilon}^{1/2}\sigma_{V_wV_w}^{-1/2}\Delta_j^J$ *a.s.*;

(*b*) $\tilde{r}_{j,b}'\hat{\Omega}_{W,b}\tilde{r}_{j,b} \mid \mathscr{X}_n \overset{d}{\to} \sigma_{\varepsilon\varepsilon}\left(1 - 2\rho_{V_w\varepsilon}\Delta_j^J + (\Delta_j^J)^2\right)$ *a.s.*;

(c) $S_l^b(\beta_0, \tilde{\gamma}_{j,b}) \mid \mathscr{X}_n \xrightarrow{d} \left(1 - 2\rho_{V_w\varepsilon}\Delta_j^J + (\Delta_j^J)^2\right)^{-1/2} S_j^J$ *a.s.*,

where $\Delta_j^J = (\Psi^{J\prime}\Psi^J - \tilde{\kappa}_j^J)^{-1}(\Psi^{J\prime}\psi_\varepsilon - \kappa_j^J\rho_{V_w\varepsilon})$, $\Psi^J = f\sigma_{V_wV_w}^{-1/2}Q_Z^{1/2}C_{0W} + \psi_{V_w} = f\Psi + (1-f)\psi_{V_w}$, $\kappa_{2SLS}^J = 0$ *and* κ_{LIML}^J *is the smallest root of the determinantal equation* $\left|(\psi_\varepsilon : \Psi^J)'(\psi_\varepsilon : \Psi^J) - \kappa\Sigma_\rho\right| = 0$; $S_j^J = \psi_\varepsilon - \Psi^J\Delta_j^J$.

Theorem A.4 *Suppose Assumptions* **2.2**-**2.3** *are satisfied and* H_0 *holds. Let* $\Pi_w = C_{0w}/\sqrt{n}$, *where* $C_{0w} \in \mathbb{R}^L$ *is fixed. If* $b \to +\infty$, *and* $b/n \to f > 0$ *as* $n \to +\infty$, *then we have*:

$$AR_{b,l}(\beta_0, \tilde{\gamma}_{j,b}) \mid \mathscr{X}_n \quad \xrightarrow{d} \quad \left\| \left(1 - 2\rho_{V_w\varepsilon}\Delta_j^J + (\Delta_j^J)^2\right)^{-1/2} S_j^J \right\|^2 \text{ a.s. for all } j \in \{LIML, 2SLS\}.$$

The inconsistency of the jackknife method follows straightforwardly by observing that the limiting distribution of $AR_{b,l}(\beta_0, \tilde{\gamma}_{j,b})$ in Theorem **A.4** differs significantly from those of Theorem **3.3** for all $j \in \{LIML, 2SLS\}$.

A.2. Supplemental lemmata

Lemma A.5 *Suppose that Assumptions* **2.2**-**2.3** *a and* H_0 *are satisfied and that* $\Pi_w \neq 0$ *is fixed. Then for some integer* $r \geq 1$, *we have*:

$$\sup_{\tau\in\mathbb{R}} \left| \mathbb{P}[S(\beta_0, \tilde{\gamma}_j) \leq \tau] - \Phi(\tau) - \sum_{h=1}^{r} n^{-h/2} p_{s_j}^h(\tau; F, \beta_0, \gamma, \Pi_x, \Pi_w)\phi(\tau) \right| = o(n^{-r/2})$$

for all $j \in \{LIML, 2SLS\}$, *where* $\Phi(\cdot)$ *and* $\phi(\cdot)$ *are the cdf and pdf of a standard normal random variable,* $p_{s_j}^h$ *are polynomials in* τ *with coefficients depending on* β_0, γ, Π_x, Π_w *and the moments of the distribution* F *of* $\mathscr{R}_n$.

Lemma A.6 *Suppose that Assumptions* **2.2**-**2.3** *a and* H_0 *are satisfied and that* $\Pi_w \neq 0$ *is fixed. Then for some integer* $r \geq 1$, *we have*:

$$\sup_{\tau\in\mathbb{R}} \left| \mathbb{P}^*[S^*(\beta_0, \tilde{\gamma}_j^*) \leq \tau] - \Phi(\tau) - \sum_{h=1}^{r} n^{-h/2} p_{s_j}^h(\tau; F_n, \beta_0, \tilde{\gamma}_j, \hat{\Pi}_x, \hat{\Pi}_w)\phi(\tau) \right| = o(n^{-r/2})$$

for all $j \in \{LIML, 2SLS\}$, *where* $\Phi(\cdot)$ *and* $\phi(\cdot)$ *are the cdf and pdf of a standard normal random variable,* $p_{s_j}^h$ *are polynomials in* τ *with coefficients depending on* β_0, $\tilde{\gamma}_j$, $\hat{\Pi}_x$, $\hat{\Pi}_w$ *and the moments of the distribution* F_n *of* $\mathscr{R}_n^*$.

Lemma A.7 *Suppose Assumptions* **2.2-2.3** *are satisfied and* H_0 *holds. If* $\Pi_w = C_{0w}/\sqrt{n}$, *where* $C_{0w} \in \mathbb{R}^L$ *is fixed, then for* $j \in \{2SLS, LIML\}$ *we have*:

(a) $E^*\left[(v_{1,i}^* - V_{w,i}^{*'}\tilde{\gamma}_j)^2\right] \xrightarrow{d} \sigma_{\varepsilon\varepsilon,j} = \sigma_{\varepsilon\varepsilon}\left(1 - 2\rho_{V_w\varepsilon}\Delta_j + (\Delta_j)^2\right)$ *a.s.*;

(b) $E^*\left[V_{w,i}^*(v_{1,i}^* - V_{w,i}^{*'}\tilde{\gamma}_j)\right] \xrightarrow{d} \sigma_{V_w\varepsilon,j} = \sigma_{V_w\varepsilon} - \sigma_{\varepsilon\varepsilon}^{1/2}\sigma_{V_wV_w}^{1/2}\Delta_j$ *a.s.*;

(c) $E^*\left[V_{w,i}^*V_{w,i}^{*'}\right] \xrightarrow{p} \sigma_{V_wV_w}$ *a.s.*

Lemma A.8 *Suppose Assumptions* **2.2-2.3** *are satisfied and* H_0 *holds. Let* $\Pi_w = C_{0w}/\sqrt{n}$, *where* $C_{0w} \in \mathbb{R}^L$ *is fixed. If, for some* $\delta > 0$, $E(\| Z_i \|^{4+\delta}, \| V_i \|^{2+\delta}) < \infty$, *then we have*:

$$\begin{pmatrix} \left(\frac{Z^{*'}Z^*}{n}\right)^{-1/2}\left(\frac{Z^{*'}V^*}{\sqrt{n}}\right)\left(\hat{\Omega}_W\right)^{-1/2} \\ n^{-1/2}\left(n^{-1}M^{*'}\mathbb{1}_n - n^{-1}M'\mathbb{1}_n\right) \end{pmatrix} \mid \mathscr{X}_n \to^d \begin{pmatrix} \psi_{V_1} \\ \psi_{V_w} \\ \psi_m \end{pmatrix} \sim N\left(\begin{pmatrix} I_{2L} & 0 \\ 0 & \Omega_{mm} \end{pmatrix}\right) \text{ a.s.},$$

where $M = (m_1, \ldots, m_n)$, $m_i = vech(Z_iZ_i^*) \in R^{L(L+1)/2}$, $\Omega_{mm} = Var(m_i)$, *and* $M^* = (m_1^*, \ldots, m_n^*)$, $m_i^* = vech(Z_i^*Z_i^{*'})$ *are the bootstrap counterparts of M and m; and* $\mathbb{1}_n$ *is a* $n \times 1$ *vector of ones.*

A.3. Proofs

PROOF OF LEMMA **A.5** To shorten our exposition, we will focus on the proof for $S(\beta_0, \tilde{\gamma}_{2SLS})$. The proof for $S(\beta_0, \tilde{\gamma}_{LIML})$ can be deduced easily by following similar steps to those presented here. First, observe that we can write $S(\beta_0, \tilde{\gamma}_{2SLS}) = \sqrt{n}N/D$ under the subset null hypothesis H_0, where

$$\begin{aligned} N &= \left(\frac{Z'Z}{n}\right)^{-1/2}\left[\frac{Z'\tilde{y}(\beta_0)}{n} - \left(\frac{Z'W}{n}\right)\tilde{\gamma}_{2SLS}\right] \\ &= \left(\frac{Z'Z}{n}\right)^{-1/2}\left[\frac{Z'\varepsilon}{n} - \left(\frac{Z'W}{n}\right)\left[\left(\frac{W'Z}{n}\right)\left(\frac{Z'Z}{n}\right)^{-1}\left(\frac{Z'W}{n}\right)\right]^{-1}\left(\frac{W'Z}{n}\right)\left(\frac{Z'Z}{n}\right)^{-1}\left(\frac{Z'\varepsilon}{n}\right)\right], \qquad \text{(A.4)} \\ D &= \frac{\tilde{y}'(\beta_0)M_Z\tilde{y}(\beta_0)}{n-L} - 2\frac{\tilde{y}'(\beta_0)M_ZW}{n-L}\tilde{\gamma}_{2SLS} + \frac{W'M_ZW}{n-L}(\tilde{\gamma}_{2SLS})^2, \qquad \text{(A.5)} \\ \tilde{\gamma}_{2SLS} &= \left[\left(\frac{W'Z}{n}\right)\left(\frac{Z'Z}{n}\right)^{-1}\left(\frac{Z'W}{n}\right)\right]^{-1}\left(\frac{W'Z}{n}\right)\left(\frac{Z'Z}{n}\right)^{-1}\left(\frac{Z'\tilde{y}(\beta_0)}{n}\right), \text{ and } \quad \tilde{y}(\beta_0) = y - X\beta_0. \end{aligned}$$

So, under H_0 and from (A.4)-(A.5), we can write $S(\beta_0, \tilde{\gamma}_{2SLS})$ as

$$\tilde{S}(\beta_0, \tilde{\gamma}_{2SLS}) = \sqrt{n}H(\bar{R}_n) = \sqrt{n}[H(\bar{R}_n) - H(\mu)] \qquad \text{(A.6)}$$

where $H(\cdot)$ is a real-valued Borel measurable function on $\mathbb{R}^K$ with derivatives of order $s \geq 3$ and lower, being continuous on the neighborhood of $\mu = \mathbb{E}(R_n)$ when $\Pi_w \neq 0$ is fixed, and $H(\mu) = 0$ under H_0. Note that the derivatives of order $s \geq 3$ and lower of $H(\cdot)$ are not well-defined when $\Pi_w = 0$ and does not even exist if $\Pi_w = C_{0w} c_n$ for any sequence $c_n \downarrow 0$ [similar to Moreira et al. (2009, footnote 2) and Doko Tchatoka (2013)]. The result in Lemma **A.5** follows by applying Bhattacharya and Ghosh (1978, Theorem 2) to (A.6) with $s - 2 = r$. □

PROOF OF THEOREM **3.1** From (2.7), we have $L \times AR(\beta_0, \tilde{\gamma}_j) = \left\| \tilde{S}(\beta_0, \tilde{\gamma}_j) \right\|^2$. We want to approximate $\mathbb{P}[AR(\beta_0, \tilde{\gamma}_j) \leq \tau]$ uniformly in τ under H_0. First, we can write $\mathbb{P}[AR(\beta_0, \tilde{\gamma}_j) \leq \tau]$ as:

$$\mathbb{P}[AR(\beta_0, \tilde{\gamma}_j) \leq \tau] = \mathbb{P}[AR(\beta_0, \tilde{\gamma}_j) \in \mathscr{C}_\tau],$$

where $\mathscr{C}_\tau = \left\{ x \in \mathbb{R}; x^2 \leq \tau \right\}$ are convex sets. From Bhattacharya and Rao (1976, Corollary 3.2), we have $\sup_{\tau \in \mathbb{R}} \Phi((\partial \mathscr{C}_\tau)^\varepsilon) \leq d.\varepsilon$ for some constant d and $\varepsilon > 0$. So, Bhattacharya and Ghosh (1978, Theorem 1) holds with $B = \mathscr{C}_\tau$ and $W_n = \tilde{S}(\beta_0, \tilde{\gamma}_j),\ j \in \{2SLS, LIML\}$. By using the approximation of $\mathbb{P}[\tilde{S}(\beta_0, \tilde{\gamma}_j) \leq \tau]$ in Lemma **A.5** and the definition of $\mathscr{C}_\tau$, Theorem **3.1** follows directly from the fact that the odd terms of the quadratic expansion are even. □

PROOF OF LEMMA **3.2** (a) We begin with the result for $\tilde{\gamma}_{2SLS}$. We have

$$\begin{aligned}
\left(\sigma_{V_w V_w}\right)^{-1} W' P_Z W &= \left(\sigma_{V_w V_w}\right)^{-1} \left(\Pi_w' Z' Z \Pi_w + V_w' Z \Pi_w + \Pi_w' Z' V_w + V_w' P_Z V_w \right) \\
&= \left(\sigma_{V_w V_w}^{-1/2} C_{0w} \right)' \left(\frac{Z'Z}{n} \right) \left(\sigma_{V_w V_w}^{-1/2} C_{0w} \right) + \left(\sigma_{V_w V_w}^{-1/2} \frac{Z' V_w}{\sqrt{n}} \right)' \left(\sigma_{V_w V_w}^{-1/2} C_{0w} \right) \\
&\quad + \left(\sigma_{V_w V_w}^{-1/2} C_{0w} \right)' \left(\sigma_{V_w V_w}^{-1/2} \frac{Z' V_w}{\sqrt{n}} \right) + \left(\sigma_{V_w V_w}^{-1/2} \frac{Z' V_w}{\sqrt{n}} \right)' \left(\frac{Z'Z}{n} \right)^{-1} \left(\sigma_{V_w V_w}^{-1/2} \frac{Z' V_w}{\sqrt{n}} \right) \\
&\xrightarrow{d} C_{0w}' \sigma_{V_w V_w}^{-1/2} Q_Z \sigma_{V_w V_w}^{-1/2} C_{0w} + \psi_{V_w}' \sigma_{V_w V_w}^{-1/2} Q_Z^{1/2} C_{0w} + C_{0w}' \sigma_{V_w V_w}^{-1/2} Q_Z^{1/2} \psi_{V_w} + \psi_{V_w}' \psi_{V_w} \\
&= (Q_Z^{1/2} \sigma_{V_w V_w}^{-1/2} C_{0w} + \psi_{V_w})' (Q_Z^{1/2} \sigma_{V_w V_w}^{-1/2} C_{0w} + \psi_{V_w}) = \boldsymbol{\Psi}' \boldsymbol{\Psi}.
\end{aligned}$$

By the same token, we have

$$\begin{aligned}
\left(\sigma_{\varepsilon\varepsilon}\sigma_{V_wV_w}\right)^{-1/2} W'P_Z\varepsilon &= \left(\sigma_{\varepsilon\varepsilon}\sigma_{V_wV_w}\right)^{-1/2}\left(\Pi_w'Z'\varepsilon + V_w'P_Z\varepsilon\right) \\
&= \left(\sigma_{V_wV_w}^{-1/2}C_{0w}\right)'\left(\frac{Z'Z}{n}\right)^{1/2}\left(\frac{Z'Z}{n}\right)^{-1/2}\left(\sigma_{\varepsilon\varepsilon}^{-1/2}\frac{Z'\varepsilon}{\sqrt{n}}\right) \\
&\quad + \left(\sigma_{V_wV_w}^{-1/2}\frac{Z'V_w}{\sqrt{n}}\right)'\left(\frac{Z'Z}{n}\right)^{-1}\left(\sigma_{\varepsilon\varepsilon}^{-1/2}\frac{Z'\varepsilon}{\sqrt{n}}\right) \\
&\xrightarrow{d} C_{0w}'\sigma_{V_wV_w}^{-1/2}Q_Z^{1/2}\psi_\varepsilon + \psi_{V_w}'\psi_\varepsilon = \Psi'\psi_\varepsilon.
\end{aligned}$$

Because, $\tilde{\gamma}_{2SLS} - \gamma_0 = (W'P_ZW)^{-1}W'P_Z\varepsilon$, the result follows immediately.

For the case of $\tilde{\gamma}_{LIML}$, we note that $\tilde{\kappa}_{LIML}$ is the smallest root of the characteristic polynomial

$$\left|\kappa\hat{\Omega}_W - (\tilde{y}(\beta_0):W)'P_Z(\tilde{y}(\beta_0):W)\right| = 0.$$

Observe that $P_Z(\tilde{y}(\beta_0):W) = P_Z\left[Z\Pi_w(\gamma:) + (\varepsilon:V_W)\begin{pmatrix}1 & 0\\ \gamma & 1\end{pmatrix}\right]$. Substituting this into the characteristic polynomial, and pre-multiplying by $\left|\begin{pmatrix}1 & 0\\ -\gamma & 1\end{pmatrix}'\right|$ and post-multiplying by $\left|\begin{pmatrix}1 & 0\\ -\gamma & 1\end{pmatrix}\right|$ yields

$$\left|\kappa\hat{\Sigma} - (\varepsilon:Z\Pi_W + V_W)'P_Z(\varepsilon:Z\Pi_W + V_W)\right| = 0,$$

where $\hat{\Sigma} = \begin{pmatrix}1 & 0\\ -\gamma & 1\end{pmatrix}'\hat{\Omega}_W\begin{pmatrix}1 & 0\\ -\gamma & 1\end{pmatrix} \equiv \begin{pmatrix}\hat{\sigma}_{\varepsilon\varepsilon} & \hat{\sigma}_{\varepsilon V_w}\\ \hat{\sigma}_{V_w\varepsilon} & \hat{\Sigma}_{V_w}\end{pmatrix}$.

From Theorem 1(a) and Theorem 2 in Staiger and Stock (1997), we get

$$\left(\sigma_{\varepsilon\varepsilon}^{-1}\sigma_{V_wV_w}\right)^{1/2}(\tilde{\gamma}_{LIML} - \gamma_0) \xrightarrow{d} \Delta_{LIML} = \left\{\Psi'\Psi - \kappa_{LIML}\right\}^{-1}\left\{\Psi'\psi_\varepsilon - \kappa_{LIML}\rho_{V_w\varepsilon}\right\},$$

where κ_{LIML} is the smallest root of the characteristic polynomial

$$\left|(\psi_\varepsilon:\Psi)'(\psi_\varepsilon:\Psi) - \kappa\Sigma_\rho\right| = 0,\ \Sigma_\rho = \begin{pmatrix}1 & \rho_{V_w\varepsilon}\\ \rho_{V_w\varepsilon} & 1\end{pmatrix}.$$

Thus, the result follows.

(b) Similarly, we have

$$\begin{aligned}
\tilde{r}_j'\hat{\Omega}_{1W}\tilde{r}_j &= (n-L)^{-1}\left(\tilde{y}(\beta_0)-W\tilde{\gamma}\right)'M_Z\left(\tilde{y}(\beta_0)-W\tilde{\gamma}\right) \\
&= (n-L)^{-1}\varepsilon'\varepsilon-(n-k)^{-1}\left(\tilde{\gamma}_j-\gamma_0\right)'W'M_Z\varepsilon-(n-L)^{-1}\varepsilon'M_ZW\left(\tilde{\gamma}_j-\gamma_0\right) \\
&\quad +(n-L)^{-1}\left(\tilde{\gamma}_j-\gamma_0\right)'W'M_ZW\left(\tilde{\gamma}_j-\gamma_0\right) \\
&\xrightarrow{d} \sigma_{\varepsilon\varepsilon}-\left(\sigma_{\varepsilon\varepsilon}^{1/2}\sigma_{V_wV_w}^{-1/2}\Delta_j\right)'\sigma_{V_w\varepsilon}-\sigma_{\varepsilon V_W}\left(\sigma_{\varepsilon\varepsilon}^{1/2}\sigma_{V_wV_w}^{-1/2}\Delta_j\right) \\
&\quad +\left(\sigma_{\varepsilon\varepsilon}^{1/2}\sigma_{V_wV_w}^{-1/2}\Delta_j\right)'\sigma_{V_wV_w}\left(\sigma_{\varepsilon\varepsilon}^{1/2}\sigma_{V_wV_w}^{-1/2}\Delta_j\right) \\
&= \sigma_{\varepsilon\varepsilon}\left(1-2\rho_{V_w\varepsilon}\Delta_j+(\Delta_j)^2\right)
\end{aligned}$$

(c). First, note that $\tilde{S}(\beta_0,\tilde{\gamma}_j)=(Z'Z)^{-1/2}Z'\left(\tilde{y}(\beta_0):W\right)\tilde{r}_j(\tilde{r}_j'\hat{\Omega}_{1W}\tilde{r}_j)^{-1/2}$. However, we have

$$\begin{aligned}
\left(Z'Z\right)^{-1/2}Z'\left(\tilde{y}(\beta_0):W\right)\tilde{r}_j &= \left(\frac{Z'Z}{n}\right)^{-1/2}\frac{Z'\varepsilon}{\sqrt{n}}+\left(\frac{Z'Z}{n}\right)^{-1/2}\frac{Z'W}{\sqrt{n}}\left(\gamma_0-\tilde{\gamma}_j\right) \\
&= \sigma_{\varepsilon\varepsilon}^{1/2}\left\{\sigma_{\varepsilon\varepsilon}^{-1/2}\left(\frac{Z'Z}{n}\right)^{-1/2}\frac{Z'\varepsilon}{\sqrt{n}}+\left(\sigma_{V_wV_w}^{-1/2}\left(\frac{Z'Z}{n}\right)^{1/2}C_w\right.\right. \\
&\quad \left.\left.+\sigma_{V_wV_w}^{-1/2}\left(\frac{Z'Z}{n}\right)^{-1/2}\frac{Z'V_W}{\sqrt{n}}\right)\sigma_{\varepsilon\varepsilon}^{-1/2}\sigma_{V_wV_w}^{1/2}\left(\gamma_0-\tilde{\gamma}_j\right)\right\} \\
&\xrightarrow{d} \sigma_{\varepsilon\varepsilon}^{1/2}\left\{\psi_\varepsilon-\Psi\left(\Psi'\Psi-\kappa_j\right)^{-1}\left(\Psi'\psi_\varepsilon-\kappa_j\rho_{V_w\varepsilon}\right)\right\} \\
&= \sigma_{\varepsilon\varepsilon}^{1/2}(\psi_\varepsilon-\Psi\Delta_j).
\end{aligned}$$

Combing this with (b), the result follows.

□

PROOF OF THEOREM **3.3** The proof follows immediately from equation (2.7) and Lemma **3.2**.

□

PROOF OF LEMMA **A.6** The proof follows the same steps as Theorem 3 of Moreira et al. (2009)

and is therefore omitted. □

PROOF OF LEMMA **4.1** The proof follows the same steps as Theorem 3 of Moreira et al. (2009) and is therefore omitted. □

PROOF OF THEOREM **4.2** The proof is similar to Hall and Horowitz (1996) by exploiting Theorem **3.1** and Lemma **4.1**, hence is omitted. □

PROOF OF LEMMA **A.7** (a). First, we have

$$\begin{aligned}
\mathbb{E}^*\left[(v^*_{1,i} - V^{*\prime}_{W,i}\tilde{\gamma}_j)^2\right] &= n^{-1}\sum_{i=1}^{n}\left(\tilde{v}_{1,i} - \tilde{V}'_{w,i}\tilde{\gamma}_j\right)^2 \\
&= n^{-1}\tilde{v}'_1\tilde{v}_1 - 2\left(n^{-1}\tilde{v}'_1\tilde{V}_W\right)\tilde{\gamma}_j + \tilde{\gamma}_j\left(n^{-1}\tilde{V}'_w\tilde{V}_w\right)(\tilde{\gamma}_j)^2 \\
&\xrightarrow{d} \sigma_{\varepsilon\varepsilon}\left(1 - 2\rho_{V_w\varepsilon}\Delta_j + (\Delta_j)^2\right).
\end{aligned}$$

(b). Similarly to (a), we have

$$\begin{aligned}
\mathbb{E}^*\left[V^*_{w,i}(v^*_{1,i} - V^{*\prime}_{w,i}\tilde{\gamma}_j)\right] &= n^{-1}\tilde{V}'_w\tilde{v}_1 - \left(n^{-1}\tilde{V}'_w\tilde{V}_w\right)\tilde{\gamma}_j \\
&= n^{-1}\tilde{V}'_w\tilde{v}_1 - \left(n^{-1}\tilde{V}'_w\tilde{V}_w\right)\gamma_0 + \left(n^{-1}\tilde{V}'_w\tilde{V}_w\right)(\gamma_0 - \tilde{\gamma}_j) \\
&\xrightarrow{d} \sigma_{V_w\varepsilon} - \sigma^{1/2}_{\varepsilon\varepsilon}\sigma^{1/2}_{V_wV_w}\Delta_j.
\end{aligned}$$

(c). By the same means, we find $E^*\left(V^*_{w,i}V^{*\prime}_{w,i}\right) = n^{-1}\tilde{V}'_w\tilde{V}_w \xrightarrow{p} \sigma_{V_wV_w}$ $a.s.$

□

PROOF OF LEMMA **A.8** The proof follows closely Lemma A.2 of Moreira et al. (2009). Let $(c', d')'$ be a nonzero vector with $c = (c'_1, c'_w)' \in \mathbb{R}^{2L}$ and $d \in \mathbb{R}^{L(L+1)/2}$. Define

$$X^*_{n,i} = \left\{c'\left(V^*_i \otimes Z^*_i\right) + d'\left(m^*_i - \overline{m}\right)\right\}/\sqrt{n},$$

where $V^*_i = \left(v^*_{1,i}, V^{*\prime}_{w,i}\right)'$ is the ith bootstrap draw of the (re-centered) reduced-form residuals, $\overline{m} =$

$n^{-1}\sum_{i=1}^{n} m_i$. We use the Cramer-Wald device to verify the condition of the Liapunov Central Limit Theorem for $X_{n,i}^*$.

(a) $\mathbb{E}^*[X_{n,i}^*] = 0$ follows from the independence of the bootstrap draws and $E^*[V_i^*] = 0$.

(b) By noting that $\mathbb{E}^*[V_i^* V_i^{*'}] = n^{-1}\tilde{V}'\tilde{V}$, where $\tilde{V} = (\tilde{v}_i : \tilde{V}_w)$, and $\mathbb{E}^*[Z_i^* Z_i^{*'}] = n^{-1}Z'Z$; and that $n^{-1}\tilde{V}'\tilde{V} \xrightarrow{p} \Omega_W$, and $n^{-1}Z'Z \xrightarrow{p} Q_Z$, it follows immediately that

$$\mathbb{E}^*[X_{n,i}^{*^2}] = n^{-1}\left\{c'\left[\left(n^{-1}\tilde{V}'\tilde{V}\right) \otimes \left(n^{-1}Z'Z\right)\right]c + d'\tilde{\Omega}_{mm}\right\} < +\infty$$

where $\tilde{\Omega}_{mm} = n^{-1}\sum_{i=1}^{n}(m_i - \overline{m})(m_i - \overline{m})'$.

(c) By using the same argument as in Lemma A.2 of Moreira et al. (2009), we have $\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}^*\left[\left|X_{n,i}^*\right|^{2+\delta}\right] = 0\ a.s.$ Since $\mathbb{E}^*\left[n^{-1}Z^{*'}Z^*\right] = n^{-1}Z'Z$ and $n^{-1}Z^{*'}Z^* - n^{-1}Z'Z \mid \mathscr{X}_n \xrightarrow{a.s.} 0$ by the Markov law of large numbers (LLN), we also have $n^{-1}Z^{*'}Z^* \mid \mathscr{X}_n \xrightarrow{p} Q_Z\ a.s.$ In addition, it is easy to see that $\hat{\Omega}_W \xrightarrow{p} \Omega_W$. Lemma **A.8** follows by the Liapunov Central Limit Theorem. □

PROOF OF LEMMA **4.3**

(a). As before, we begin with the 2SLS estimator. Let $C_{0w}^B = C_{0w} + \left(n^{-1}Z'Z\right)^{-1}\left(n^{-1/2}Z'V_w\right)$, then

$$\begin{aligned}
&\left(\mathbb{E}^*\left[V_{w,i}^* V_{w,i}^{*'}\right]\right)^{-1} W^{*'} P_{Z^*} W^* \\
= &\left(\mathbb{E}^*\left[V_{w,i}^* V_{w,i}^{*'}\right]\right)^{-1}\left(\hat{\Pi}_w' Z^{*'} Z^* \hat{\Pi}_w + V_w^{*'} Z^* \hat{\Pi}_w + \hat{\Pi}_w' Z^{*'} V_w^* + V_w^{*'} P_{Z^*} V_w^*\right) \\
= &\left(\mathbb{E}^*\left[V_{w,i}^* V_{w,i}^{*'}\right]\right)^{-1}\left\{C_{0w}^{B'}\left(\frac{Z^{*'}Z^*}{n}\right)C_{0w}^B + \left(\frac{V_w^{*'}Z^*}{\sqrt{n}}\right)C_{0w}^B + C_{0w}^{B'}\left(\frac{Z^{*'}V_w^*}{\sqrt{n}}\right)\right. \\
&\left. + \left(\frac{V_w^{*'}Z^*}{\sqrt{n}}\right)\left(\frac{Z^{*'}Z^*}{n}\right)^{-1}\left(\frac{Z^{*'}V_w^*}{\sqrt{n}}\right)\right\}
\end{aligned}$$

and similarly, we have

$$\begin{aligned}
&\left(\mathbb{E}^*\left[(v_{1,i}^* - V_{w,i}^*\tilde{\gamma}_j)^2\right]\mathbb{E}^*\left[V_{w,i}^* V_{w,i}^{*'}\right]\right)^{-1/2} W^{*'} P_{Z^*}\left(V_1^* - V_w^*\tilde{\gamma}_j\right) \\
= &\left(\mathbb{E}^*\left[(V_{1,i}^* - V_{w,i}^*\tilde{\gamma}_j)^2\right]\mathbb{E}^*\left[V_{w,i}^* V_{w,i}^{*'}\right]\right)^{-1/2}\left\{\hat{\Pi}_w' Z^{*'}\left(V_1^* - V_w^*\tilde{\gamma}_j\right) + V_w^{*'} P_{Z^*}\left(V_1^* - V_w^*\tilde{\gamma}_j\right)\right\}
\end{aligned}$$

$$= \left(\mathbb{E}^*\left[(V_{1,i}^* - V_{w,i}^*\tilde{\gamma}_j)^2\right]\mathbb{E}^*\left[V_{w,i}^* V_{w,i}^{*'}\right]\right)^{-1/2}\left\{C_{0w}^{B'}\left(\frac{Z^{*'}\left(V_1^* - V_w^*\tilde{\gamma}_j\right)}{\sqrt{n}}\right)\right.$$

$$\left. + \left(\frac{V_w^{*'}Z^*}{\sqrt{n}}\right)\left(\frac{Z^{*'}Z^*}{n}\right)^{-1}\left(\frac{Z^{*'}\left(V_1^* - V_w^*\tilde{\gamma}_j\right)}{\sqrt{n}}\right)\right\}.$$

From Lemmata **A.7-A.8**, we have $n^{-1}Z^{*'}Z^* \mid \mathscr{X}_n \xrightarrow{a.s.} \mathbb{E}(Z_i Z_i') = Q_Z$, and since $C_{0w}^B = C_{0w} + \left(n^{-1}Z'Z\right)^{-1}\left(n^{-1/2}Z'V_w\right)$, it follows that

$$\begin{aligned}\left(\mathbb{E}^*\left[V_{w,i}^* V_{w,i}^{*'}\right]\right)^{-1} W^{*'}P_{Z^*}W^* \quad \mid \mathscr{X}_n \xrightarrow{d} & \left(\sigma_{V_wV_w}^{-1/2}Q_Z^{1/2}C_{0w} + 2\psi_{V_w}\right)'\left(\sigma_{V_wV_w}^{-1/2}Q_Z^{1/2}C_{0w} + 2\psi_{V_w}\right) a.s. \\ = & \left(\Psi + \psi_{V_w}\right)'\left(\Psi + \psi_{V_w}\right) \equiv \Psi^{B'}\Psi^B\end{aligned}$$

Similarly, we have

$$\left(\mathbb{E}^*\left[(V_{1,i}^* - V_{w,i}^*\tilde{\gamma}_j)^2\right]\mathbb{E}^*\left[V_{w,i}^* V_{w,i}^{*'}\right]\right)^{-1/2} W^{*'}P_{Z^*}\left(V_1^* - V_w^*\tilde{\gamma}_j\right) \mid \mathscr{X}_n \xrightarrow{d} \Psi^{B'}\psi_{\varepsilon,2SLS}\ a.s.,$$

where $\psi_{\varepsilon,2SLS} = \psi_{V_1} - \psi_{V_w}(\gamma_0 + \sigma_{\varepsilon\varepsilon}^{1/2}\sigma_{V_wV_w}^{-1/2}\Delta_{2SLS}) = \psi_\varepsilon - \sigma_{\varepsilon\varepsilon}^{1/2}\sigma_{V_wV_w}^{-1/2}\Delta_{2SLS}\psi_{V_w}$.

For LIML, observe that $\tilde{\kappa}_{LIML}^*$ is the smallest root of

$$\left|\kappa\hat{\Sigma}_{LIML}^* - \left(V_1^* - V_w^*\tilde{\gamma}_{LIML} : Z^*\hat{\Pi}_w + V_w^*\right)' P_{Z^*}\left(V_1^* - V_w^*\tilde{\gamma}_{LIML} : Z^*\hat{\Pi}_w + V_w^*\right)\right| = 0,$$

where we have $\hat{\Sigma}_{LIML}^* = \begin{pmatrix} 1 & 0 \\ -\tilde{\gamma}_{LIML} & 1 \end{pmatrix}' \hat{\Omega}_W^* \begin{pmatrix} 1 & 0 \\ -\tilde{\gamma}_{LIML} & 1 \end{pmatrix}$.

So, by following the same steps as in the case of 2SLS and using Lemma **3.2**, we get (conditionally on $\mathscr{X}_n$)

$$\left(\frac{\mathbb{E}^*\left[V_{w,i}^* V_{w,i}^{*'}\right]}{\mathbb{E}^*(V_{1,i}^* - V_{w,i}^*\tilde{\gamma}_{LIML})^2}\right)^{1/2}\left(\tilde{\gamma}_{LIML}^* - \tilde{\gamma}_{LIML}\right) \xrightarrow{d} (\Psi^{B'}\Psi^B - \kappa_j^B)^{-1}(\Psi^{B'}\psi_{\varepsilon,LIML} - \kappa_j^B\rho_{V_w\varepsilon,LIML})$$

$a.s.$, where κ_j^B is the smallest root of the characteristic polynomial

$$\left|\left(\psi_{\varepsilon,LIML}:\Psi^B\right)'\left(\psi_{\varepsilon,LIML}:\Psi^B\right)-\kappa\Sigma_{\rho,LIML}\right|=0,\ \Sigma_{\rho,LIML}=\begin{pmatrix}1 & \rho_{V_w\varepsilon,LIML}\\ \rho_{V_w\varepsilon,LIML} & 1\end{pmatrix}.$$

This establishes Lemma **4.3**-(a).

(b) Similarly, we have

$$\begin{aligned}\tilde{r}_j^{*'}\hat{\Omega}_{1W}^*\tilde{r}_j^* &= (n-L)^{-1}\left(\tilde{y}^*(\beta_0)-W^*\tilde{\gamma}_j^*\right)'M_{Z^*}\left(\tilde{y}^*(\beta_0)-W^*\tilde{\gamma}_j^*\right)\\ &= (n-L)^{-1}\left(V_1^*-V_w^*\tilde{\gamma}_j\right)'\left(V_1^*-V_w^*\tilde{\gamma}_j\right)-(n-L)^{-1}\left(\tilde{\gamma}_j^*-\tilde{\gamma}_j\right)'W^{*'}M_{Z^*}\left(V_1^*-V_w^*\tilde{\gamma}_j\right)\\ &\quad -(n-L)^{-1}\varepsilon^{*'}M_{Z^*}W^*\left(\tilde{\gamma}_j^*-\tilde{\gamma}_j\right)+(n-L)^{-1}\left(\tilde{\gamma}_j^*-\tilde{\gamma}_j\right)'W^{*'}M_{Z^*}W^*\left(\tilde{\gamma}_j^*-\tilde{\gamma}_j\right).\end{aligned}$$

Since $n^{-1}Z^{*'}Z^*\mid\mathscr{X}_n\overset{a.s.}{\to}Q_Z$ and $\hat{\Omega}_{1W}^*\mid\mathscr{X}_n\overset{a.s.}{\to}\Omega_W$, hence we have

$$\tilde{r}_j^{*'}\hat{\Omega}_{1W}^*\tilde{r}_j^*\mid\mathscr{X}_n\overset{d}{\to}\sigma_{\varepsilon\varepsilon,j}\left(1-2\rho_{\varepsilon V_W,j}\Delta_j^B+(\Delta_j^B)^2\right)\ a.s.$$

(c) Again, we have

$$\begin{aligned}&\left(Z^{*'}Z^*\right)^{-1/2}Z^{*'}\left(\tilde{y}^*(\beta_0):W^*\right)\tilde{r}_j^*\\ =&\ \left(\frac{Z^{*'}Z^*}{n}\right)^{-1/2}\frac{Z^{*'}\left(V_1^*-V_w^*\tilde{\gamma}_j\right)}{\sqrt{n}}+\left(\frac{Z^{*'}Z^*}{n}\right)^{-1/2}\frac{Z^{*'}W^*\left(\tilde{\gamma}_j-\tilde{\gamma}_j^*\right)}{\sqrt{n}}\\ =&\ \sigma_{\varepsilon\varepsilon,j}^{1/2}\left\{\sigma_{\varepsilon\varepsilon,j}^{-1/2}\left(\frac{Z^{*'}Z^*}{n}\right)^{-1/2}\frac{Z^{*'}\left(V_1^*-V_w^*\tilde{\gamma}_j\right)}{\sqrt{n}}+\left(\sigma_{V_wV_w}^{-1/2}\left(\frac{Z^{*'}Z^*}{n}\right)^{1/2}C_{0w}^B\right.\right.\\ &\left.\left.+\sigma_{V_wV_w}^{-1/2}\left(\frac{Z^{*'}Z^*}{n}\right)^{-1/2}\frac{Z^{*'}V_w^*}{\sqrt{n}}\right)\sigma_{\varepsilon\varepsilon,j}^{-1/2}\sigma_{V_wV_w}^{1/2}\left(\tilde{\gamma}_j-\tilde{\gamma}_j^*\right)\right\}+o_p(1).\end{aligned}$$

Thus, by proceeding as in (a) and (b), we get

$$\left(Z^{*'}Z^*\right)^{-1/2}Z^{*'}\left(\tilde{y}^*(\beta_0):W^*\right)\tilde{r}_j^*\mid\mathscr{X}_n\overset{d}{\to}\sigma_{\varepsilon\varepsilon,j}^{1/2}\left\{\psi_{\varepsilon,j}-\Psi^B\Delta_j^B\right\}\ a.s.,$$

where $\Delta_j^B = \left(\Psi^{B'}\Psi^B - \kappa_j^B\right)^{-1}\left(\Psi^{B'}\psi_{\varepsilon,j} - \kappa_j^B \rho_{V_w\varepsilon,j}\right)$. The final result follows by combining this with (b). $\square$

PROOF OF THEOREM **4.4** The proof follows immediately from equation (4.4) and Lemma **4.3**. $\square$

PROOF OF THEOREM **4.5** From Theorems **3.3** and **4.4**, it is clear that

$$|\,\mathbb{P}^*[AR^*(\beta_0,\tilde{\gamma}_j^*) \le \tau] - \mathbb{P}[AR(\beta_0,\tilde{\gamma}_j) \le \tau]\,| = |\,\mathbb{P}[AR(\beta_0,\tilde{\gamma}_j) > \tau] - \mathbb{P}^*[AR^*(\beta_0,\tilde{\gamma}_j^*) > \tau]\,| = O_p(1)\; a.s.$$

for all $\tau \in \mathbb{R}$. By choosing $\tau = c^*_{AR_j^*}(\alpha) \equiv c^*_{AR_j^*}$, we have $\mathbb{P}^*[AR^*(\beta_0,\tilde{\gamma}_j^*) > c^*_{AR_j^*}] = \alpha$ so that we get

$$\left|\mathbb{P}[AR(\beta_0,\tilde{\gamma}_j) > c^*_{AR_j^*}] - \alpha\right| = O_p(1)\; a.s.$$

$\square$

PROOF OF LEMMA **A.1** (a). From the subsampling algorithm, we have that as $b \to +\infty$ and $f_n = b/n \to 0$, as $n \to +\infty$. Thus, we get $b^{-1}Z_b'Z_b \overset{a.s.}{\to} Q_Z$ given the realized sample $\mathscr{X}_n$. Also, conditional on $\mathscr{X}_n$, we have

$$\begin{aligned}
b^{-1/2}Z_b'V_{w,b} &= b^{1/2}\left(b^{-1}\sum_{i=1}^{b} Z_{i,b}V_{wi,b} - \mathbb{E}^b(Z_{i,b}V_{wi,b})\right) + b^{1/2}\mathbb{E}^b(Z_{i,b}V_{wi,b}) \\
&= b^{1/2}\left(b^{-1}\sum_{i=1}^{b} Z_{i,b}V_{wi,b} - n^{-1}\sum_{i=1}^{n} Z_iV_{wi}\right) + f_n^{1/2}n^{1/2}\sum_{i=1}^{n} Z_iV_{wi} \\
&\overset{d}{\to} \psi_{ZV_w}\; a.s.
\end{aligned}$$

Similarly, we have $b^{-1/2}Z_b'\varepsilon_b \mid \mathscr{X}_n \overset{d}{\to} \psi_{Z\varepsilon}\; a.s.$ and

$$\begin{aligned}
b^{-1/2}Z_b'W_b &= b^{-1/2}Z_b'\left(Z_b\Pi_W + V_{w,b}\right) \\
&= f_n^{1/2}\left(b^{-1}Z_b'Z_b\right)C_{0w} + b^{-1/2}Z_b'V_{w,b}
\end{aligned}$$

$$| \mathscr{X}_n \quad \xrightarrow{d} \quad \psi_{ZV_w} \; a.s.$$

So, we get

$$\begin{aligned}
& \tilde{\gamma}_{2SLS,b} - \gamma_0 \\
= \; & \left(W_b' P_{Z,b} W_b\right)^{-1} W_b' P_{Z,b} \varepsilon_b \\
= \; & \left\{ \left(b^{-1/2} W_b' Z_b\right) \left(b^{-1} Z_b' Z_b\right)^{-1} \left(b^{-1/2} Z_b' W_b\right) \right\}^{-1} \left(b^{-1/2} W_b' Z_b\right) \left(b^{-1} Z_b' Z_b\right)^{-1} \left(b^{-1/2} Z_b' \varepsilon_b\right) \\
| \mathscr{X}_n \xrightarrow{d} \; & \left(\psi_{ZV_w}' Q_Z^{-1} \psi_{ZV_w}\right)^{-1} \psi_{ZV_W}' Q_Z^{-1} \psi_{Z\varepsilon} = \sigma_{\varepsilon\varepsilon}^{1/2} \sigma_{V_w V_w}^{-1/2} \Delta_{2SLS}^S \; a.s.
\end{aligned}$$

where $\Delta_{2SLS}^S = \left(\psi_{V_w}' \psi_{V_w}\right)^{-1} \psi_{V_w}' \psi_\varepsilon$. The result for LIML is deduced similarly.

(b). From (a), we have

$$\begin{aligned}
\tilde{r}_{j,b}' \hat{\Omega}_{1W,b} \tilde{r}_{j,b} \; = \; & (b-L)^{-1} \varepsilon_b' \varepsilon_b - (b-L)^{-1} \left(\tilde{\gamma}_{j,b} - \gamma_0\right) W_b' M_{Z,b} \varepsilon_b - (b-L)^{-1} \varepsilon_b' M_{Z_b} W_b \left(\tilde{\gamma}_{j,b} - \gamma_0\right) \\
& + (b-L)^{-1} W_b' M_{Z,b} W_b \left(\tilde{\gamma}_{j,b} - \gamma_0\right)^2 \\
| \mathscr{X}_n \xrightarrow{d} \; & \sigma_{\varepsilon\varepsilon} - \left(\sigma_{\varepsilon\varepsilon}^{1/2} \sigma_{V_w V_w}^{-1/2} \Delta_j^S\right) \sigma_{V_w \varepsilon} - \sigma_{\varepsilon V_w} \left(\sigma_{\varepsilon\varepsilon}^{1/2} \sigma_{V_w V_w}^{-1/2} \Delta_j^S\right) \\
& + \sigma_{V_w V_w} \left(\sigma_{\varepsilon\varepsilon}^{1/2} \sigma_{V_w V_w}^{-1/2} \Delta_j^S\right)^2 \; a.s. \\
= \; & \sigma_{\varepsilon\varepsilon} \left(1 - 2\rho_{V_w \varepsilon} \Delta_j^S + (\Delta_j^S)^2\right).
\end{aligned}$$

(c). We have

$$\begin{aligned}
\left(Z_b' Z_b\right)^{-1/2} Z_b' \left(\tilde{y}_b(\beta_0) : W_b\right) \tilde{r}_{j,b} \; = \; & \left(\frac{Z_b' Z_b}{b}\right)^{-1/2} \frac{Z_b' \varepsilon_b}{\sqrt{b}} + \left(\frac{Z_b' Z_b}{b}\right)^{-1/2} \frac{Z_b' W_b}{\sqrt{b}} \left(\gamma_0 - \tilde{\gamma}_{j,b}\right) \\
| \mathscr{X}_n \xrightarrow{d} \; & \sigma_{\varepsilon\varepsilon}^{1/2} \left[\psi_\varepsilon - \Psi^S (\Psi^{S\prime} \Psi^S - \kappa_j^S)^{-1} (\Psi^{S\prime} \psi_\varepsilon - \kappa_j^S \rho_{V_w \varepsilon})\right] \; a.s. \\
= \; & \sigma_{\varepsilon\varepsilon}^{1/2} (\psi_\varepsilon - \Psi^S \Delta_j^S)
\end{aligned}$$

The result follows by combining this with (b).

□

PROOF OF THEOREM **A.2** The proof follows immediately from equation (A.3) and Lemma **A.1**.

□

Proof of Lemma **A.3**

The proof is similar to those of Lemma **A.1** by noting that $b/n \to f > 0$ as $n \to +\infty$. Therefore, it is omitted. □

Proof of Theorem **A.4** The proof follows immediately from equation (A.3) and Lemma **A.3**. □

References

Anderson, T. W., Rubin, H., 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. Annals of Mathematical Statistics 20, 46–63.

Andrews, D. W., Guggenberger, P., 2009. Validity of subsampling and " plug-in asymptotic" inference for parameters defined by moment inequalities. Econometric Theory 25, 669–709.

Andrews, D. W. K., 2002. High-order improvement of a computationally attractive k-step bootstrap for extremum estimators. Econometrica 70(1), 119–162.

Andrews, D. W. K., Stock, J. H., 2007. Inference with weak instruments. In: R. Blundell, W. Newey , T. Pearson, eds, Advances in Economics and Econometrics, Theory and Applications, 9th Congress of the Econometric Society Vol. 3. Cambridge University Press, Cambridge, U.K., chapter 6.

Berkowitz, D., Caner, M., Fang, Y., 2012. The validity of instruments revisited. Journal of Econometrics 166, 255–266.

Bhattacharya, R. N., Ghosh, J., 1978. On the validity of the formal Edgeworth expansion. The Annals of Statistics 6, 434–451.

Bhattacharya, R. N., Rao, R., 1976. Normal approximation and asymptotic expansions. In: R. Bhattacharya, R. Rao, eds, Normal Approximation and Asymptotic Expansions. Wiley Series in Probability and Mathematical Analysis, New York.

Corbae, D., Durlauf, S. N., Hansen, B. E., eds, 2006. Econometric Theory and Practice: Frontiers of Analysis and Applied Research. Cambridge University Press, Cambridge, U.K.

Doko Tchatoka, F., 2013. On bootstrap validity for specification tests with weak instruments. Technical report, School of Economics and Finance, University of Tasmania Hobart, Australia.

Doko Tchatoka, F., 2014. Subset hypotheses testing and instrument exclusion in the linear IV regression. Econometric Theory forthcoming.

Doko Tchatoka, F., Dufour, J.-M., 2014. Identification-robust inference for endogeneity parameters in linear structural models. The Econometrics Journal 17, 165–187.

Dufour, J.-M., 1997. Some impossibility theorems in econometrics, with applications to structural and dynamic models. Econometrica 65, 1365–1389.

Dufour, J.-M., 2003. Identification, weak instruments and statistical inference in econometrics. Canadian Journal of Economics 36(4), 767–808.

Dufour, J.-M., Jasiak, J., 2001. Finite sample limited information inference methods for structural equations and models with generated regressors. International Economic Review 42, 815–843.

Dufour, J.-M., Khalaf, L., Beaulieu, M.-C., 2010. Multivariate residual-based finite-sample tests for serial dependence and GARCH with applications to asset pricing models. Journal of Applied Econometrics 25, 263–285.

Dufour, J.-M., Taamouti, M., 2005. Projection-based statistical inference in linear structural models with possibly weak instruments. Econometrica 73(4), 1351–1365.

Dufour, J.-M., Taamouti, M., 2007. Further results on projection-based inference in IV regressions with weak, collinear or missing instruments. Journal of Econometrics 139(1), 133–153.

Guggenberger, P., 2010. The impact of a Hausman pretest on the size of the hypothesis tests. Econometric Theory 156, 337–343.

Guggenberger, P., Chen, L., 2011. On the asymptotic size of subvector tests in the linear instrumental variables model. Technical report, Department of Economics, UCSD.

Guggenberger, P., Kleibergen, F., Mavroeidis, S., Chen, L., 2012. On the asymptotic sizes of subset anderson-rubin and lagrange multiplier tests in linear instrumental variables regression. Econometrica 80(6), 2649–2666.

Hall, P., 1992. The bootstrap and Edgeworth expansion. In: P. Hall, ed., The bootstrap and Edgeworth expansion. Springer-Verlag New York, Inc, New York.

Hall, P., Horowitz, J. L., 1996. Bootstrap critical values for tests based on generalized-method-of-moments estimators. Econometrica 64(4), 891–916.

Horowitz, J. L., 1994. Bootstrap-based critical values for the IM test. Journal of Econometrics 61, 395–411.

Horowitz, J. L., 2001. The bootstrap. In: J. Heckman, E. E. Leamer, eds, Handbook of Econometrics. Elsvier Science, Amsterdam, The Netherlands.

Kleibergen, F., 2002. Pivotal statistics for testing structural parameters in instrumental variables regression. Econometrica 70(5), 1781–1803.

Kleibergen, F., 2004. Testing subsets of structural coefficients in the IV regression model. Review of Economics and Statistics 86, 418–423.

Kleibergen, F., 2008. Subset statistics in the linear IV regression model. Technical report, Department of Economics, Brown University, Providence, Rhode Island Providence, Rhode Island.

Mikusheva, A., 2010. Robust confidence sets in the presence of weak instruments. Journal of Econometrics 157, 236–247.

Mikusheva, A., 2013. Survey on statistical inferences in weakly-identified instrumental variable models. Applied Econometrics 29(1), 117–131.

Moreira, M. J., 2003. A conditional likelihood ratio test for structural models. Econometrica 71(4), 1027–1048.

Moreira, M. J., Porter, J., Suarez, G., 2009. Bootstrap validity for the score test when instruments may be weak. Journal of Econometrics 149, 52–64.

Poskitt, D., Skeels, C., 2012. Inference in the presence of weak instruments: A selected survey. FTEcx 6(1), 26–44.

Staiger, D., Stock, J. H., 1997. Instrumental variables regression with weak instruments. Econometrica 65(3), 557–586.

Startz, R., Nelson, C. R. N., Zivot, E., 2006. Improved inference in weakly identified instrumental variables regression. *in* (Corbae, Durlauf and Hansen 2006), chapter 5.

Stock, J. H., Wright, J. H., 2000. GMM with weak identification. Econometrica 68, 1055–1096.

Stock, J. H., Wright, J. H., Yogo, M., 2002. A survey of weak instruments and weak identification in generalized method of moments. Journal of Business and Economic Statistics 20(4), 518–529.

Wu, C., 1990. On the asymptotic properties of the jackknife histogram. The Annals of Statistics 18, 1438–1452.

Zivot, E., Startz, R., Nelson, C. R., 2006. Inference in partially identified instrumental variables regression with weak instruments. *in* (Corbae et al. 2006), chapter 5, pp. 125–163.