Inconsistency of the bootstrap for the subset Anderson-Rubin test and Bonferroni-based size correction

Firmin Doko Tchatoka*
The University of Adelaide

Wenjie Wang †
Hiroshima University

April 22, 2016

ABSTRACT

We study the validity of the bootstrap for the "plug-in" Anderson and Rubin (1949) (AR) test of subvector hypotheses in linear IV regressions where structural parameters may not be identified. Our analysis mainly focuses on two plug-in subset AR statistics–the first uses the restricted limited information maximum likelihood (LIML) estimator and the second utilizes the restricted two-stage least squares (2SLS) estimator. We provide a characterization of the asymptotic distributions of both statistics without and with weak instruments. Our results show that the asymptotic distributions of these statistics are non-standard when the nuisance parameters that are not specified by the subset null hypothesis are not identified, so correction to usual asymptotic critical values are needed. For this, we first provide a bootstrap procedure similar to that of Moreira et al. (2009). We show that this bootstrap provides a high-order refinement of the null distributions of the statistics when the nuisance parameters are identified, but it is inconsistent if these parameters are not identified. We thus proposed a Bonferroni-based size adjustment that yields tests with correct asymptotic size, even when the nuisance parameters are not identified. We present a Monte Carlo experiment that confirms our theoretical findings.

Key words: Subset AR-test; Plug-in estimator; weak instruments; bootstrap inconsistency; Edgeworth expansion; Bonferroni-type size correction.

JEL classification: C12; C13; C36.

* 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

† Department of Economics, Hiroshima University. e-mail: wenjiew@hiroshima-u.ac.jp

1. Introduction

There is now a growing interest on inference procedures for testing subset hypotheses in IV regressions where structural parameters may not be identified.[1] This literature falls generally into two categories: (1) the projection method, and (2) the "plug-in" principle.

The projection method consists of inverting an identification-robust statistic[2] to build confidence regions for the full set of the structural parameters, and then uses the projection technique to obtain confidence sets for the subset of parameters of interest. In addition to being robust to weak instruments, the projection technique based on the Anderson and Rubin (1949) statistic also enjoys robustness to instrument omission in the first-stage regression. However, it can yield a test with low power, especially when too many instruments are used. The plug-in principle consists of replacing the nuisance parameters that are not specified by the hypothesis of interest by estimators.[3] It is now well understood that the plug-in based method outperforms their projection counterpart, and in addition, never over-rejects the true parameter values if the nuisance parameters not specified by the null hypothesis of interest are identified. However, the plug-in based method does not perform well when the corresponding plug-in estimator is inconsistent. This particularly the case when the nuisance parameters are not identified. Recently, Guggenberger et al. (2012) and Guggenberger and Chen (2011) show that the plug-in subset AR test is asymptotically robust to identifying assumptions, while the plug-in Kleibergen (2002) (K) test is sensitive to such assumptions. However, even though the plug-in subset AR test is asymptotically robust to weak instruments (in the sense of level control), it can be overly conservative thus yielding a test with low power when the nuisance parameters are not identified; see Doko Tchatoka (2014).

In this paper, we focus on linear structural models and provide a characterization of the asymptotic distributions of the plug-in AR statistics based on the restricted LIML and 2SLS estimators, without and with weak instruments. Our analysis provides some new insights and extensions of earlier studies. In particular, we show that both AR subset statistics are asymptotically pivotal when the nuisance parameters are identified, but they have nonstandard asymptotic distributions when these parameters are not identified[4]–so correction to usual asymptotic critical values are needed.

For this, we first investigate the validity of the bootstrap similar to that of Moreira, Porter and Suarez (2009). We show that this bootstrap provides a high-order refinement of the null distributions of the statistics when the nuisance parameters are identified, but it is inconsistent if these parameters are not identified. This contrasts with Moreira et al. (2009) who show that bootstrap is valid for the AR statistic of the null hypothesis specified on the full vector of structural parameters, whether

[1] For example, see Stock and Wright (2000), Dufour and Jasiak (2001), Kleibergen (2004, 2008), Dufour and Taamouti (2005, 2007), Startz, Nelson and Zivot (2006), Guggenberger and Chen (2011), Guggenberger, Kleibergen, Mavroeidis and Chen (2012), and Kleibergen (2015).

[2] See Dufour (1997), Dufour and Jasiak (2001), Dufour and Taamouti (2005, 2007).

[3] See Stock and Wright (2000), Kleibergen (2004, 2008), and Startz et al. (2006).

[4] Similar to Guggenberger et al. (2012) and Doko Tchatoka (2014).

identification is strong or weak. The inconsistency of bootstrap for subset AR statistics studied is mainly due to its inability to mimic the *concentration factor* that characterizes the strength of the identification of the nuisance parameters. We thus proposed a Bonferroni-based size adjustment that yields tests with correct asymptotic size, even when the nuisance parameters are not identified. We present a Monte Carlo experiment that confirms our theoretical findings.

This paper is organized as follows. Section 2 presents the setting, the model assumptions, and the subset AR statistics studied. Section 3 characterizes the limiting behavior of these statistics. Section 4 presents the proposed bootstrap method and studies its asymptotic validity. A Monte Carlo experiment on the finite-sample performance of both the standard and bootstrap subset AR tests is presented in Section 5, while Section 6 deals with the Bonferroni-based size adjustment. Conclusions are drawn in Section 7. 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 $m \times m$ squared matrix B means that B is positive definite, and $vech(B)$ is the $\frac{1}{2}m(m+1)$ dimensional half-column vectorization of B. 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. Setting

Let (y_i, X_i, W_i, Z_i'), $i = 1, \ldots, n$ be a sample of n observations, where y_i is observation i on an outcome variable, X_i, W_i are observations i on (possibly) endogenous regressors, and Z_i is a vector of observations i on instrumental variables. The usual linear IV regression, written in matrix form, consists of the following structural and reduced-form equations:

$$y = X\beta + W\gamma + \varepsilon, \tag{2.1}$$

$$(X, W) = Z(\Pi_x : \Pi_w) + (V_x, V_w), \tag{2.2}$$

where $y \in \mathbb{R}^n$, $X \in \mathbb{R}^n$, $W \in \mathbb{R}^n$, and $Z \in \mathbb{R}^{n\times L}$, $(\varepsilon : V_x : V_w) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$ are unobserved errors, $\beta, \gamma \in \mathbb{R}$, and $\Pi_x, \Pi_w \in \mathbb{R}^L$ are unknown parameters. We assume that $L \geq 2$ is fixed, and denote:

$$\bar{Y} = [y, X, W] = [\bar{Y}_1, \ldots, \bar{Y}_n]', \quad \overline{\mathscr{X}}_n = \left\{\mathscr{X}_1', \mathscr{X}_2', \ldots, \mathscr{X}_n'\right\}'$$

$$R_n = vech\left(\mathscr{X}_n'\mathscr{X}_n\right) = \left(f_1(\mathscr{X}_n), f_2(\mathscr{X}_n), \ldots, f_K(\mathscr{X}_n)\right)', \tag{2.3}$$

where $\mathscr{X}_i = (y_i, X_i, W_i, Z_i')$ and $f_p(\cdot)$ $[p = 1, \ldots, K = \frac{1}{2}(L+2)(L+3)]$ are elements of the matrix $\mathscr{X}_n'\mathscr{X}_n$.

We are interested in testing the subset hypothesis

$$H_0 : \beta = \beta_0, \tag{2.4}$$

where $\beta_0 \in \mathbb{R}$ is fixed and γ is a nuisance parameter. In the literature, two procedures have often been used to assess H_0 and build confidence regions for β_0: the projection-based technique,[5] and the conventional plug-in based principle.[6] It is now well known that the plug-in principle outperforms their projection technique counterpart [for example, see Guggenberger et al. (2012)], so we only focus on the plug-in principle in this study.

To be more specific, consider the problem of testing joint hypothesis $H(\beta_0, \gamma_0) :\ \beta = \beta_0,\ \gamma = \gamma_0$ in (2.1)–(2.2). The well known Anderson and Rubin (1949) (AR) test of this joint hypothesis is given by

$$AR_n(\beta_0, \gamma_0) = \frac{1}{L}\left\|\bar{S}_n(\beta_0, \gamma_0)\right\|^2, \tag{2.5}$$

where $\bar{S}_n(\beta_0, \gamma_0) = (Z'Z)^{-1/2}Z'\bar{Y}\bar{r}(\bar{r}'\hat{\Omega}_Y\bar{r})^{-1/2}$, $\hat{\Omega}_Y = \frac{1}{n-L}\bar{Y}'M_Z\bar{Y}$, and $\bar{r} = (1, -\beta_0, -\gamma_0)'$. The plug-in subset AR statistic for H_0 in (2.4) is then defined as

$$AR_n(\beta_0, \tilde{\gamma}) = \min_{\gamma \in \mathbb{R}} AR_n(\beta_0, \gamma), \tag{2.6}$$

where $\tilde{\gamma} =: \arg\min_{\gamma\in\mathbb{R}} AR_n(\beta_0, \gamma)$. It is known from the literature on simultaneous equation that $\tilde{\gamma} = \tilde{\gamma}_{LIML}$ in (2.6), where $\tilde{\gamma}_{LIML}$ is the restricted LIML estimator of γ under H_0, i.e.

$$\tilde{\gamma}_{LIML} = \left[W'(P_Z - \bar{\kappa}_{LIML}M_Z)W\right]^{-1}W'(P_Z - \bar{\kappa}_{LIML}M_Z)(y - X\beta_0), \tag{2.7}$$

$\bar{\kappa}_{LIML} = \tilde{\kappa}_{LIML}/(n-L)$ and $\tilde{\kappa}_{LIML}$ is the smallest root of the characteristic polynomial $|\kappa\hat{\Omega}_W - \tilde{Y}(\beta_0)'P_Z\tilde{Y}(\beta_0)| = 0$. So, the statistic $AR_n(\beta_0, \tilde{\gamma})$ in (2.6) can also be expressed as

$$AR_n(\beta_0, \tilde{\gamma}_{LIML}) = \frac{1}{L}\left\|\tilde{S}_n(\beta_0, \tilde{\gamma}_{LIML})\right\|^2, \tag{2.8}$$

where $\tilde{S}_n(\beta_0, \tilde{\gamma}_{LIML}) = (Z'Z)^{-1/2}Z'\tilde{Y}(\beta_0)\tilde{r}_{LIML}(\tilde{r}_{LIML}'\hat{\Omega}_W\tilde{r}_{LIML})^{-1/2}$, $\tilde{Y}(\beta_0) = [\tilde{y}(\beta_0) : W]$, $\tilde{y}(\beta_0) = y - X\beta_0$, $\hat{\Omega}_W = \frac{1}{n-L}\tilde{Y}(\beta_0)'M_Z\tilde{Y}(\beta_0)$, and $\tilde{r}_{LIML} = (1, -\tilde{\gamma}_{LIML})'$. It is often the case that alternative restricted k-class estimators of γ are used in (2.6); for example, see Startz et al. (2006). When γ is identified, these k-class estimators yield statistics that are asymptotically equivalent to the one in

[5] See Dufour and Jasiak (2001) and Dufour and Taamouti (2005, 2007)

[6] See Stock and Wright (2000), Kleibergen (2004, 2008), Startz et al. (2006), Mikusheva (2010), Guggenberger et al. (2012), Doko Tchatoka (2014), and Kleibergen (2015).

(2.8). However, if γ is not identified, the behavior of these statistics can substantially differ from that of the statistic in (2.8); for example, see Doko Tchatoka (2014) for the case of the restricted 2SLS estimator. Therefore, it is interesting to also study the properties of these statistics, especially when the identification of γ is weak. In this paper, in addition to the subset AR statistic with the restricted LIML estimator, we also consider the one with the restricted 2SLS estimator. Both statistics can be expressed in a unified way as

$$AR_n(\beta_0, \tilde{\gamma}_j) = \frac{1}{L} \left\| \tilde{S}_n(\beta_0, \tilde{\gamma}_j) \right\|^2, \; j \in \{LIML, \, 2SLS\}, \tag{2.9}$$

where $\tilde{\gamma}_{2SLS}$ is obtained by setting $\overline{\kappa}_{2SLS} \equiv \overline{\kappa}_{LIML} = 0$ in (2.7).

In order to characterize the asymptotic null distributions of the statistics in (2.9), it will be useful to consider the following assumptions on the model variables, where $\mathbb{E}[\cdot]$ denotes the expectation with respect to the relevant probability measure and

$$\begin{aligned} \Sigma_U &= \begin{bmatrix} \sigma_{\varepsilon\varepsilon} & \sigma'_{V\varepsilon} \\ \sigma_{V\varepsilon} & \Sigma_V \end{bmatrix}, \; \Sigma = \begin{bmatrix} \sigma_{\varepsilon\varepsilon} & \sigma_{V_w\varepsilon} \\ \sigma_{V_w\varepsilon} & \sigma_{V_wV_w} \end{bmatrix}, \\ \Sigma_V &= \begin{bmatrix} \sigma_{V_xV_x} & \sigma_{V_xV_w} \\ \sigma_{V_xV_w} & \sigma_{V_wV_w} \end{bmatrix}, \; \sigma_{V\varepsilon} = (\sigma_{V_x\varepsilon}, \sigma_{V_w\varepsilon})'. \end{aligned} \tag{2.10}$$

Assumption 2.1 $(u_i, V_{xi}, V_{wi}, Z_i')'$, $i = 1, \ldots, n$, *are i.i.d. with common distribution* F.

Assumption 2.2 *The vectors* $U_i = (u_i, V_{xi}, V_{wi})'$, $i = 1, \ldots, n$, *have zero means and the same (finite) nonsingular covariance matrix:*

$$\mathbb{E}[U_i U_i'] = \Sigma_U > 0 \, , i = 1, \ldots, n,$$

where Σ_U *is defined in* (2.10).

Assumption 2.3 $\mathbb{E}[Z_i U_i'] = 0$, $\mathbb{E}[Z_i Z_i'] =: Q_Z > 0$, *and* $\mathbb{E}\left[vec(Z_i U_i')(vec(Z_i U_i'))'\right] = \Sigma_U \otimes Q_Z$ *for all* $i = 1, \ldots, n$.

Assumption 2.4 (i) $\mathbb{E}[\|R_n\|^{2+r}] < \infty$ *for some* $r > 0$, *and* (ii) $\limsup_{\|t\| \to \infty} |\, \mathbb{E}[exp(\boldsymbol{i}t'R_n)] \,| < 1$, *where* $\boldsymbol{i}=\sqrt{-1}$ *and* R_n *is defined in* (2.3).

Assumption 2.5 *As the sample size n converges to infinity, we have*:

$$n^{-1/2} vec\left[Z'(\varepsilon, V_w)\right] \xrightarrow{d} vec\left[\psi_{Z\varepsilon}, \, \psi_{ZV_w}\right] \sim N\left[0, \, \Sigma \otimes Q_Z\right],$$

where Σ *is defined in* (2.10) *and* $\otimes$ *denotes the Kronecker product of two matrices.*

Assumption **2.1** is commonly used in the IV literature; for example, see Guggenberger et al. (2012). The distribution F may depend on n, but for convenience we write F rather than F_n where there is no confusion.

Assumptions **2.2** and **2.3** are also common in the IV literature. While Assumption **2.2** requires that model errors have mean zero and second finite moments, Assumption **2.3** state the (usual) orthogonality condition between the errors and IVs, along with the existence of the same (finite) second moments for the instrument vector $Z_i, i = 1, \ldots, n$.

Assumptions **2.4**-(i) and -(ii) are similar to Assumptions 2-3 in Moreira et al. (2009) with $r = s - 2$ and $s \geq 3$. While (i) requires that R_n has second moments or greater, (ii) imposes that the characteristic function of R_n be bounded above by 1. In particular, the second moments of R_n exist if $\mathbb{E}(\|\mathscr{X}_n\|^{2(r+2)}) < \infty$ for some $r > 0$. The bound on the characteristic function in (ii) is the commonly used Cramér's condition [see Bhattacharya and Ghosh (1978)].

Assumption **2.5** holds by the central limit theorem (CLT) property; for example, see Staiger and Stock (1997), Kleibergen (2002, 2004), Guggenberger et al. (2012).

Now, let $\theta = (\gamma, \Pi_w, F)$ denote the parameters of the model under H_0, where by the notation (γ, Π_w, F) and elsewhere, we allow components of a vector, column vectors, matrices (of different dimensions), and distributions to be tackled. Following Guggenberger et al. (2012), we define the parameter space for θ as

$$\Theta = \Big\{ \theta = (\gamma, \Pi_w, F) \text{ such that Assumptions } \mathbf{2.1}\text{–}\mathbf{2.3} \text{ hold} \Big\}. \tag{2.11}$$

For a given $\theta \in \Theta$, we define the finite sample null rejection probability (NRP) of the subset AR test using $\tilde{\gamma}_j, j \in \{LIML,\ 2SLS\}$, as:

$$\mathrm{NRP}_{AR_j} \quad = \quad \mathbb{P}_\theta \left[AR_n(\beta_0, \tilde{\gamma}_j) > \chi^2_{L-1,1-\alpha} \right], \tag{2.12}$$

where $\chi^2_{L-1}(\alpha)$ is the $1-\alpha$ quantile of a χ^2-distributed random variable with $L-1$ degrees of freedom, and $\mathbb{P}_\theta[A_n]$ denotes the probability of the event A_n. Similarly, the *asymptotic size* of this test is defined as

$$AsySz_{AR_j}\left[\chi^2_{L-1}(\alpha)\right] \quad = \quad \limsup_{n\to\infty} \sup_{\theta \in \Theta} \mathbb{P}_\theta \left[AR_n(\beta_0, \tilde{\gamma}_j) > \chi^2_{L-1}(\alpha) \right], j \in \{LIML, 2SLS\}. \tag{2.13}$$

Under Assumptions **2.2**–**2.3** and **2.5**, Guggenberger et al. (2012) show that $AsySz_{AR_{LIML}}[\chi^2_{L-1}(\alpha)] \leq \alpha$ even when γ is not identified but $AsySz_{AR_{2SLS}}[\chi^2_{L-1}(\alpha)] > \alpha$ under weak instruments. This implies that the subset AR test with restricted LIML has a correct asymptotic size even under weak instruments, while that with restricted 2SLS does not enjoy this property. Even though we have $AsySz_{AR_{LIML}}[\chi^2_{L-1}(\alpha)] \leq \alpha$ even when γ is not identified, $\mathrm{NRP}_{AR_{LIML}}$ in (2.12) can be strictly less than α in small samples when γ is not identified, thus yielding an overly conservative test if the subset statistic $AR_n(\beta_0, \tilde{\gamma}_{LIML})$ is used; see Doko Tchatoka (2014). So, correction to usual asymptotic critical values are needed. To better understand these results, it will be illuminating to summarize the asymptotic properties of the subset AR statistics in (2.9).

3. Preliminary results

In this section, we characterize the asymptotic null distributions of the subset statistics $AR_n(\beta_0, \tilde{\gamma}_j), j \in \{LIML, 2SLS\}$. To do this, we find useful to distinguish the case in which γ is identified to the one where it is not identified. Since the setup in which γ is identified is relatively easy to tackle, we start with that case first.

3.1. High-order approximation when γ is identified

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. Let also F_{R_n} denote the distribution of R_n given in (2.3). Theorem **3.1** provides a high-order refinement of the distributions of the subset AR statistics under H_0.

Theorem 3.1 *Suppose Assumptions* **2.1**–**2.4** *are satisfied. If further H_0 hold and $\Pi_w \neq 0$ is fixed, then for some integer $r \geq 1$, we have*:

$$\sup_{\tau \in \mathbb{R}} \left| \mathbb{P}_\theta[AR_n(\beta_0, \tilde{\gamma}_j) \leq \tau] - G_{L-1}(\tau) - \sum_{h=1}^{r} n^{-h} p^h_{AR_j}(\tau; F_{R_n}, \beta_0, \Pi_w, \tilde{\gamma}_j) g_{L-1}(\tau) \right| = o(n^{-r})$$

for all $j \in \{LIML, 2SLS\}$, where $p^h_{AR_j}$ is a polynomial in τ with coefficients depending on β_0, Π_w, $\tilde{\gamma}_j$, and the moments of F_{R_n}.

It is worth observing that Theorem **3.1** provides a more greater accurate approximation of the distribution of $AR_n(\beta_0, \tilde{\gamma}_j)$ under H_0 than the usual first-order asymptotic χ^2 approximation. In particular, the $1-\alpha$ quantile of the distribution of $AR_n(\beta_0, \tilde{\gamma}_j)$ under H_0 can be approximated uniformly in $\zeta < \alpha < 1-\zeta$ for any $0 < \zeta < 1/2$ by $c_{AR_j}(\alpha) \approx \chi^2_{L-1}(\alpha) + \sum_{h=1}^{r} n^{-h} q^h_{AR_j}(\chi^2_{L-1}(\alpha))$, where $q^h_{AR_j}$ is a polynomial derivable from $p^h_{AR_j}$ and $\chi^2_{L-1}(\alpha)$ is the solution of the equation $G_{L-1}(\tau) = 1-\alpha$; see Hall (1992). Therefore, the corresponding tests have *correct asymptotic size* even if the parameter of interest β is not identified. However, this high-order improvement is achievable only when γ is identified. If γ is not identified, we show in the next section that even the first-order asymptotic χ^2 approximation no longer valid for all subset statistics.

3.2. Asymptotic distribution when γ is not identified

We now study the asymptotic behavior under H_0 of the subset AR statistics when γ is not identified. To proceed, let $\{\theta_n = (\gamma_n, \Pi_{w,n}, F_n) :\ n \geq 1\}$ denote the subsequences of parameters in Θ satisfying:

$$
\begin{aligned}
& n^{1/2} \left(\mathbb{E}_{F_n}[V_{w,i}^2]\right)^{-1/2} \left(\mathbb{E}_{F_n}[Z_i Z_i']\right)^{1/2} \Pi_{w,n} \to h_{ww} \in \bar{\mathbb{R}}^L, \\
& \left(\mathbb{E}_{F_n}[\varepsilon_i^2]\mathbb{E}_{F_n}[V_{w,i}^2]\right)^{-1/2} \mathbb{E}_{F_n}(V_{w,i}\varepsilon_i) \to h_{\varepsilon w} \in [-1,\ 1],
\end{aligned}
\tag{3.1}
$$

where $h_{\varepsilon w} = \left(\sigma_{\varepsilon\varepsilon}\sigma_{V_w V_w}\right)^{-1/2} \sigma_{V_w \varepsilon}$, $\mathbb{E}_{F_n}[\cdot]$ is the expectation with respect to F_n, and $\bar{\mathbb{R}} =: \mathbb{R} \cup \{-\infty, +\infty\}$ is the extended real line. The parameter h_{ww} in (3.1) characterizes the identification strength of γ, and is referred to as the "concentration factor" in the remainder of our analysis. From (3.1), we can partition the space of the concentration factor (h_{ww}) as $\{h_{ww} :\ \|h_{ww}\| < +\infty\} \cup \{h_{ww} :\ \|h_{ww}\| = +\infty\}$. Note that if $\|h_{ww}\| = \infty$ in (3.1), then strong instrument asymptotics apply; for example, see Guggenberger (2012), Guggenberger and Chen (2011), and Guggenberger et al. (2012). However, if $\|h_{ww}\| < +\infty$, it is the easy to see that $h_{ww} = o(n^{1/2})$ and γ is not identified.[7] This case is similar to the *weak IV asymptotic* of Staiger and Stock (1997). Guggenberger et al. (2012) show that the asymptotic behavior of $AR_n(\beta_0, \tilde{\gamma}_{LIML})$ under H_0 is driven only by the subsequences in (3.1), hence we focus on those subsequences in our analysis. In addition, since strong identification is covered in Theorem **3.1**, we deal only with the setup of weak identification, i.e., the case in which $\|h_{ww}\| < +\infty$.

To proceed, let $\lambda_{n,h} =: (h_{n,ww}, h_{n,\varepsilon w}, h_{n,F})$ be sequence of parameters such that

$$
h_{n,ww} = \left(\mathbb{E}_{F_n}[V_{w,i}^2]\right)^{-1/2} \left(\mathbb{E}_{F_n}[Z_i Z_i']\right)^{1/2} \Pi_{w,n}, \tag{3.2}
$$

$$
h_{n,\varepsilon w} = \left(\mathbb{E}_{F_n}[\varepsilon_i^2]\mathbb{E}_{F_n}[V_{w,i}^2]\right)^{-1/2} \mathbb{E}_{F_n}(V_{w,i}\varepsilon_i) \text{ and } h_{n,F} = F_n. \tag{3.3}
$$

If $\|h_{ww}\| < +\infty$, we see from (3.1) that the drifting sequence $(\lambda_{n,h})_{n\geq 1}$ in (3.2)-(3.3) satisfies:

$$
n^{1/2} h_{n,ww} \to h_{ww} \in \mathbb{R}^L,\ h_{n,\varepsilon w} \to h_{\varepsilon w} \in [-1,1], \text{ and } h_{n,F} \to F \text{ as } n \to \infty. \tag{3.4}
$$

Now, define the standardized random variables $\psi_{V_w} = \sigma_{V_w V_w}^{-1/2} Q_Z^{-1/2} \psi_{Z V_w}$ and $\psi_\varepsilon = \sigma_{\varepsilon\varepsilon}^{-1} Q_Z^{-1/2} \psi_{Z\varepsilon}$. Under Assumptions **2.2**–**2.3** and **2.5**, we have $vec[\psi_\varepsilon, \psi_{V_w}] \sim N(0, \Sigma_h \otimes Q_Z)$, where $\Sigma_h = \begin{pmatrix} 1 & h_{\varepsilon w} \\ h_{\varepsilon w} & 1 \end{pmatrix}$. Finally, let $h =: (h_{ww}, h_{\varepsilon w})$. and set $\Psi_h = h_{ww} + \psi_{V_w}$. Also, define

$$
\Delta_{h,j} = (\Psi_h' \Psi_h - \kappa_{h,j})^{-1} (\Psi_h' \psi_\varepsilon - \kappa_{h,j} h_{\varepsilon w}),\ S_{h,j} = \psi_\varepsilon - \Psi_h \Delta_{h,j}, j \in \{LIML, 2SLS\}, \tag{3.5}
$$

where $\kappa_{h,LIML}$ is the smallest root of the determinantal equation $\left|(\psi_\varepsilon : \Psi_h)'(\psi_\varepsilon : \Psi_h) - \kappa_h \Sigma_h\right| = 0$

[7] It worth mentioning that even if the condition $\|h_{ww}\| < \infty$ is viewed as the case of weak IVs, high values of $\|h_{ww}\|$ may indicate that the IVs are not very weak, i.e., γ is in the neighborhood of the identification region.

and $\kappa_{h,2SLS} = 0$.

Theorem **3.2** characterizes the asymptotic distribution under H_0 of the statistics $AR_n(\beta_0, \tilde{\gamma}_j), j \in \{LIML, 2SLS\}$.

Theorem 3.2 *Suppose that Assumptions* **2.1**–**2.3**, *and* **2.5** *are satisfied. If further H_0 holds and $(\lambda_{n,h})_{n\geq 1}$ satisfies* (3.2)–(3.4), *then we have*:

$$AR_n(\beta_0, \tilde{\gamma}_j) \quad \overset{d}{\to} \quad \xi_j(h) = \frac{1}{L}\left\| \left(1 - 2h_{\varepsilon w}\Delta_{h,j} + \Delta_{h,j}^2\right)^{-1/2} S_{h,j} \right\|^2 \text{ for all } j \in \{LIML, 2SLS\},$$

where $\Delta_{h,j}$ and $S_{h,j}$ are defined in (3.5), *and $h =: (h_{ww}, h_{\varepsilon w})$.*

Theorem **3.2** shows that the limiting distribution under H_0 of $AR_n(\beta_0, \tilde{\gamma}_j)$ is completely characterized by the parameter $h =: (h_{ww}, h_{\varepsilon w})$. More interestingly, the distribution of $\xi_j(h)$ depends on $h =: (h_{ww}, h_{\varepsilon w})$ only through the localized parameters $\|h_{ww}\|$ and $h_{\varepsilon w}$ [see Guggenberger et al. (2012)]. Therefore, the asymptotic size of the test $AR_n(\beta_0, \tilde{\gamma}_j)$ is driven only by $\|h_{ww}\|$ and $h_{\varepsilon w}$. As both $\|h_{ww}\|$ and $h_{\varepsilon w}$ do not depend on the specific value β_0 tested, the asymptotic size of this test does not also depend on β_0. However, it depends on γ through the covariance endogeneity parameter $h_{\varepsilon w}$. More precisely, consider the following reduced-form equations for both $\tilde{y} = y - X\beta_0$ and W under H_0 :

$$\tilde{y} \quad = \quad Z\Pi_w\gamma + v_1,\ W = Z\Pi_w + V_w, \tag{3.6}$$

where $v_1 = V_w\gamma + \varepsilon$. Under Assumptions **2.1**–**2.3**, we have:

$$\begin{aligned}&\mathbb{E}\left[(v_{1i} : V_{w,i})'(v_{1i} : V_{w,i})\right] = \begin{pmatrix} \sigma_{11} & \sigma_{V_w 1} \\ \sigma_{V_w 1} & \sigma_{V_w V_w} \end{pmatrix} : \ \sigma_{V_w 1} = \sigma_{V_w V_w}\gamma + \sigma_{V_w \varepsilon}, \\ &\sigma_{11} = \sigma_{\varepsilon\varepsilon} + 2\sigma_{V_w\varepsilon}\gamma + \gamma^2\sigma_{V_w V_w}.\end{aligned} \tag{3.7}$$

Hence, we can express $h_{\varepsilon w} = \left(\sigma_{\varepsilon\varepsilon}\sigma_{V_w V_w}\right)^{-1/2}\sigma_{V_w \varepsilon}$ as:

$$h_{\varepsilon w} \quad =: \quad h_{\varepsilon w}(\gamma) = (\sigma_{\varepsilon\varepsilon}\sigma_{V_w V_w})^{-1/2}(\sigma_{V_w 1} - \sigma_{V_w V_w}\gamma). \tag{3.8}$$

Moreover, we also have $\sigma_{\varepsilon\varepsilon} =: \sigma_{\varepsilon\varepsilon}(\gamma) = \sigma_{11} - 2\sigma_{V_w 1}\gamma + \gamma^2\sigma_{V_w V_w}$ from (3.7), and $\sigma_{\varepsilon\varepsilon} > 0$ under Assumptions **2.1**–**2.3**. Hence the last term in the right-hand side of (3.8) is a strictly monotonic[8] function of γ.

[8]The partial derivative of $h_{\varepsilon w}(\gamma)$ with respect to γ is $\frac{\partial h_{\varepsilon w}}{\partial \gamma} = \frac{(\sigma_{V_w 1})^2 - \sigma_{11}\sigma_{V_w V_w}}{\sigma_{V_w V_w}^{1/2}(\sigma_{11} - 2\sigma_{V_w 1}\gamma + \gamma^2\sigma_{V_w V_w})^{3/2}}$, which has the sign of $-\begin{vmatrix} \sigma_{11} & \sigma_{V_w 1} \\ \sigma_{V_w 1} & \sigma_{V_w V_w} \end{vmatrix} = (\sigma_{V_w 1})^2 - \sigma_{11}\sigma_{V_w V_w} \neq 0$ (under Assumptions **2.1**–**2.3**) for any value of γ in the parameter space Θ.

As the distribution of $\xi_j(h)$ depends only on $\|h_{ww}\|$ and $h_{\varepsilon w}$, it can be simulated given any localized values of $\|h_{ww}\|$ and $h_{\varepsilon w}$. For the values of $\|h_{ww}\|$ and $h_{\varepsilon w}$ in the neighborhood of their true values, the resulting simulated distribution of $\xi_j(h)$ can produce a good approximation to the finite-sample distribution of $AR_n(\beta_0, \tilde{\gamma}_j)$. The difficulty, however, is that neither $\|h_{ww}\|$ nor $h_{\varepsilon w}$ is known, and both cannot be consistently estimated under H_0 and the subsequences in (3.2)–(3.4). In the next section, we investigate whether bootstrapping can provide a valid approximation of the distribution of $\xi_j(h)$.

4. Bootstrapping subset AR statistics

We adapt the bootstrap of Moreira et al. (2009) to the subset AR statistics. Our bootstrap differs slightly from Moreira et al. (2009) in the sense $H_0 : \beta = \beta_0$ is imposed in the resampling scheme, while theirs uses a " super-consistent" estimator of β_0. We think that there is no need to replace β_0 by an estimator because the asymptotic distributions under H_0 of the subset AR statistics do not depend on the unknown value β_0 tested. Furthermore, in practice, researchers usually rely more on reporting identification-robust confidence regions for β_0 in this type of model, rather than a specific pointwise test outcome. Section 4.1 describes briefly our bootstrap algorithm.

4.1. Bootstrap algorithm

Let $\hat{\Pi}_x = (Z'Z)^{-1}Z'X$ and $\hat{\Pi}_w = (Z'Z)^{-1}Z'W$ denote the ordinary least squares (OLS) estimators of Π_x and Π_w in (2.2). Let also $\tilde{\gamma}_j,\ j \in \{LIML, 2SLS\}$, denote the restricted estimator of γ under H_0. We suggest the following resampling scheme for our bootstrap.

1. For a given β_0 and the observed data, compute $\hat{\Pi}_x$, $\hat{\Pi}_w$ and $\tilde{\gamma}_j$, along with all other items necessary to obtain the realizations of the statistic $AR_n(\beta_0, \tilde{\gamma}_j)$ and the residuals from the reduced-form equation (3.6): $\hat{v}_1 = \tilde{y}(\beta_0) - Z\hat{\Pi}_w\tilde{\gamma}_j$, $\hat{V}_x = X - Z\hat{\Pi}_x$, and $\hat{V}_w = W - Z\hat{\Pi}_w$. Re-centered these residuals by subtracting sample means to yield $(\tilde{v}_1, \tilde{V}_x, \tilde{V}_w)$;

2. For each bootstrap sample $b = 1, \ldots, B,$ generate the data 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 $(Z, \tilde{v}_1, \tilde{V}_x, \tilde{V}_w)$. Compute the corresponding bootstrap subset AR statistics $AR_n^{*(b)}(\beta_0, \tilde{\gamma}_j^*)$, $b =$

$1, \ldots, B,$ as

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

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

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 $\frac{1}{B}\sum_{b=1}^{B} \mathbb{1}[AR_n^{*(b)}(\beta_0, \tilde{\gamma}_j^*) > AR_n(\beta_0, \tilde{\gamma}_j)]$ is less then α.

In the reminder of the paper, $\widetilde{F}_n$ denotes the empirical distribution of $R_n^* = vech\left(\mathscr{X}_n^{*\prime} \mathscr{X}_n^*\right)$ conditional on $\overline{\mathscr{X}}_n$, $\mathbb{P}^*$ is the probability under the empirical distribution function (conditional on $\overline{\mathscr{X}}_n$), and $\mathbb{E}^*$ its corresponding expectation operator. As in Section 3, we deal separately with the case where γ is identified and the one where it is not.

4.2. Bootstrap consistency when γ is identified

In this section, we study the validity of the bootstrap for the subset AR statistics when the nuisance parameter γ is identified. Lemma **4.1** summarizes the results.

Lemma 4.1 *Suppose that Assumptions* **2.1–2.4** *are satisfied. If further H_0 hold and $\Pi_w \neq 0$ is fixed, then for some integer $r \geq 1$, we have*:

$$\sup_{\tau \in \mathbb{R}} \left| \mathbb{P}^*\left[AR_n^*(\beta_0, \tilde{\gamma}_j^*) \leq \tau\right] - G_{L-1}(\tau) - \sum_{h=1}^{r} n^{-h} p_{AR_j}^h(\tau; \widetilde{F}_n, \tilde{\gamma}_j, \hat{\Pi}_w) g_{L-1}(\tau) \right| = o(n^{-r})$$

for all $j \in \{LIML, 2SLS\}$, $p_{AR_j}^h$ is a polynomial in τ with coefficients depending on $\tilde{\gamma}_j$, $\hat{\Pi}_x$, $\hat{\Pi}_w$, and the moments of $\widetilde{F}_n$.

The above lemma shows that the bootstrap estimate and the $(r+1)$-term empirical Edgeworth expansion in Lemma **3.1** are asymptotically equivalent up to the $o(n^{-r})$ order under H_0 when γ is identified. 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 (2.9) is a quadratic function of a symmetric pivotal statistic [see Horowitz (2001, Ch. 52, eq. 3.13)] under H_0 when γ is identified.

Now, let $c_{AR_j}^* =: \min_{\tau \in \mathbb{R}} \left| \mathbb{P}^*[AR_n^*(\beta_0, \tilde{\gamma}_j^*) \leq \tau] - (1-\alpha) \right|$ denotes the $1-\alpha$ quantile of the empirical distribution of $AR_n^*(\beta_0, \tilde{\gamma}_j^*)$. We can state the following theorem on the high-order approximation of the size of the subset AR tests when the bootstrap critical values are used in the inference.

Theorem 4.2 *Suppose that Assumptions* **2.1**–**2.4** *are satisfied. If further* H_0 *hold and* $\Pi_w \neq 0$ *is fixed, then we have*:

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

Theorem **4.2** shows that the bootstrap critical values under H_0 yield correct level for the AR tests through $O(n^{-1})$. Hence, the bootstrap only makes an error of size $O(n^{-1})$ under H_0 if the nuisance parameter γ is identified. It is worth noting that the identification of β plays no role here, so Theorem **4.2** holds even when $\Pi_x = 0$ in (2.2)– complete non-identification of β– or close to zero– weak identification of β.

Moreover, even though Theorem **4.2** focuses on the size properties of the tests, there is no impediment to expanding it to the power analysis. For example, we can show that if $\Pi_w \neq 0$ is fixed, test consistency holds as long as β is identified (i.e., if $\Pi_x \neq 0$ is fixed). However, the bootstrap tests have low power if β is not identified. This proof is omitted in order to shorten the exposition of our results.

We now study the validity of the bootstrap when γ is not identified.

4.3. Bootstrap inconsistency when γ is not identified

As before, we focus on the subsequences of parameters $\theta_n = (\gamma_n, \Pi_{w,n}, F_n)$ satisfying (3.1)–(3.4), and we provide the characterization of the asymptotic distributions under H_0 of the bootstrap subset AR statistics in (4.4).

To ease readability, we set $\Psi^B_h = \Psi_h + \psi_{V_w}$, $\psi_{\varepsilon,j} = \psi_\varepsilon - \varpi^{1/2}_{\varepsilon w}\Delta_{h,j}\psi_{V_w}$, and define

$$\begin{aligned} \Delta^B_{h,j} &= \left(\Psi^{B\prime}_h \Psi^B_h - \kappa^B_{h,j}\right)^{-1}\left(\Psi^{B\prime}\psi_{\varepsilon,j} - \kappa^B_{h,j} h_{\varepsilon w,j}\right),\; S^B_{h,j} = \psi_{\varepsilon,j} - \Psi^B_h \Delta^B_{h,j}\,, \\ h_{\varepsilon w,j} &= (1 - 2h_{\varepsilon w}\Delta_{h,j} + \Delta^2_{h,j})^{-1/2}(h_{\varepsilon w} - \Delta_{h,j}),\; j \in \{LIML, 2SLS\}\,, \end{aligned} \tag{4.6}$$

where $\varpi_{\varepsilon w} = \sigma_{\varepsilon\varepsilon}\sigma^{-1}_{V_w V_w}$, $\kappa^B_{2SLS} = 0$, κ^B_{LIML} is the smallest root of the determinantal equation $\left|\left(\psi_{\varepsilon,j} : \Psi^B_h\right)'\left(\psi_{\varepsilon,j} : \Psi^B_h\right) - \kappa\Sigma_{h,j}\right| = 0$, and $\Sigma_{h,j} = \begin{pmatrix} 1 & h_{\varepsilon w,j} \\ h_{\varepsilon w,j} & 1 \end{pmatrix}$. Lemma **4.3** presents the results.

Lemma 4.3 *Suppose that Assumptions* **2.1**–**2.3**, *and* **2.5** *are satisfied. Suppose also that* H_0 *holds and* $(\lambda_{n,h})_{n\geq 1}$ *satisfies* (3.2)–(3.4). *If further* $\mathbb{E}(\|Z_i\|^{4+\delta}, \|(\varepsilon_i, V_{wi})'\|^{2+\delta}) < +\infty$ *for some* $\delta > 0$, *then we have*:

$$AR^*_n(\beta_0, \tilde{\gamma}^*_j) \mid \overline{\mathscr{X}}_n \;\overset{d}{\to}\; \xi^*_j(h, \varpi_{\varepsilon w}) = \left\|\left(1 - 2h_{\varepsilon w,j}\Delta^B_{h,j} + (\Delta^B_{h,j})^2\right)^{-1/2} S^B_{h,j}\right\|^2 \; a.s.$$

for all $j \in \{LIML, 2SLS\}$.

First, we observe from Lemma **4.3** that the bootstrap statistics $AR^*_n(\beta_0, \tilde{\gamma}^*_j)$, $j \in \{LIML, 2SLS\}$, do not converge to the asymptotic distributions under H_0 of the standard statistics if $\|h_{ww}\| < \infty$

[see Lemma **3.2**]. In particular, we show in Lemma **3.2** that the asymptotic distribution of of $AR_n(\beta_0, \tilde{\gamma}_j)$ is completely characterized by $h =: (h_{ww}, h_{\varepsilon w})$. However, the limit of bootstrap counterpart in Lemma **4.3**, $\xi_j^*(h, \varpi_{\varepsilon w})$, depends not only on h but also on the variance ratio $\varpi_{\varepsilon w}$. This provides evidence of the bootstrap is inconsistent when parameter γ is not identified.

Second, we note that the inconsistency of the bootstrap is mainly due to the fact that under H_0 and the subsequences of drifting parameters $(\theta_n)_{n\geq 1}$ satisfying (3.1)–(3.4), replacing $\Pi_{n,w}$ with $\hat{\Pi}_w$ and γ_n with $\tilde{\gamma}_j$ in the bootstrap DGP adds an extra noise term to the original reduced-form residuals $\hat{v}_1$ and $\hat{V}_w$ during the resampling process. For example, while $(V_w' V_w/n)^{-1/2}(Z'Z/n)^{-1/2}Z'W/\sqrt{n} \xrightarrow{d} \Psi_h = h_{ww} + \psi_{V_w}$, its bootstrap counterpart $(V_w^{*'} V_w^*/n)^{-1/2}(Z^{*'}Z^*/n)^{-1/2}Z^{*'}W^*/\sqrt{n}$ converges almost surely to $\Psi_h^B = \Psi_h + \psi_{V_w}$. Hence, if $\Psi_h < \infty$ with probability one, as is the case when $\|h_{ww}\| < \infty$, the bootstrap fails to mimic the asymptotic behavior of $(V_w' V_w/n)^{-1/2}(Z'Z/n)^{-1/2}Z'W/\sqrt{n}$ under the above subsequences of drifting parameters. A similar result hold for many the arguments of each bootstrap statistic $AR_n^*(\beta_0, \tilde{\gamma}_j^*)$ written as a function of R_n^*.

We now examine the finite-sample performance of both the standard and bootstrap subset AR tests through a Monte Carlo experiment.

5. Performance of the standard and bootstrap subset AR tests

We examine the performance of both the standard and bootstrap subset AR tests in a Monte Carlo experiment. The data generating process is described by (2.1) and (2.2) where y, X and W are $n \times 1$ vectors. The errors ε, V_x, V_w are drawn i.i.d. normal with zero mean and unit variance, and the correlations between them are such that set at $h_{\varepsilon x} = h_{\varepsilon w} = h_{xw}$, where $h_{\varepsilon w} \in \{0, 0.1, 0.5, 0.9\}$. The L columns of the instrument matrix Z, $L \in \{3, 5, 10, 20\}$, are drawn i.i.d. $N(0, I_L)$ independently from $[\varepsilon : V_x : V_w]$. The true values of β and γ are set at 2 and -1 respectively. The reduced-form coefficient matrix Π_x and Π_w is chosen such that the concentration parameters μ_{xx}^2 μ_{ww}^2 which describe the strength of Z satisfy $\mu_{xx}^2 = \mu_{ww}^2 \in \{0, 0.05, 1, 10\}$, where $\mu_{ww}^2 = 0$ is the setup of a complete non-identification (irrelevant IVs), $\mu_{ww}^2 = 0.05$ represents weak instruments, $\mu_{ww}^2 = 1$ designates moderately weak instruments, and $\mu_{ww}^2 = 10$ is for strong instruments. The rejection frequencies are computed using $N = 10,000$ replications for the standard subset AR tests, while those of the bootstrap subset AR tests are obtained with $N = 10,000$ replications and $B = 299$ bootstrap pseudo-samples of size $n = 100$. The nominal level of all tests is set at 5%.

Table 1 shows the empirical rejection frequencies of the tests. The first column of the tables contains the statistics. The second column indicates the number of instruments (L) used in the inference. The other columns show the rejection frequencies for each value of the endogeneity parameter $h_{\varepsilon w}$ and the instrument quality μ_{ww}^2.

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 $\mu^2_{ww} \in \{0, 0.05\}$ 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 $\mu^2_{ww} = 10$ in the table). Note however that the rejection frequencies of this test are slightly greater than the nominal 5% level when both the endogeneity parameter ($h_{\varepsilon w}$) and the number of instruments (L) increase (for example, with $h_{\varepsilon w} = 0.9$ and $L = 20$, the rejection frequency is around 7%). Meanwhile, the subset AR test with the restricted 2SLS is less conservative than those with LIML when IVs are weak and the asymptotic χ^2 critical values are applied (see columns $\mu^2_{ww} \in \{0, 0.05\}$ in the second block of Table 1). This is not surprising because the inequality $AR(\beta_0, \tilde{\gamma}_{2SLS}) \geq AR(\beta_0, \tilde{\gamma}_{LIML})$ is always true. In particular, with weak instruments, the rejection frequencies of this test are close to the nominal 5% level for small endogeneity (see columns $h_{\varepsilon w} \in \{0, 0.1\}$ and $\mu^2_{ww} \in \{0, 0.05\}$ in the table), but they are greater than 5% for large endogeneity (see column $h_{\varepsilon w} = 0.9$ and $\mu^2_{ww} \in \{0, 0.05, 1\}$). We also observe that this test over-rejects sometimes when identification is strong. For example, the rejection frequencies when $h_{\varepsilon w} = 0.9$ and $\mu^2_{ww} = 10$ (strong instruments) are about 8.6%, 16.7%, and 39.6% for $L = 5,\ 10,\ 20$ instruments, respectively. So, while the subset AR test with the restricted 2SLS seems to outperform the one with the restricted LIML under weak instruments and small endogeneity, it over-rejects the null hypothesis when identification is moderate or strong and the endogeneity parameter is large.

Second, we observe that bootstrapping does not improve the size properties of either test when identification is weak, as shown in columns $\mu^2_{ww} \in \{0, .05\}$ of the last block of Table 1 for all values of L and the endogeneity parameters $h_{\varepsilon w}$. This confirms the inconsistency of the bootstrap for subset AR tests when identification is weak (see Section 4.3). 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 ($h_{\varepsilon w} = 0.9$). For example, the rejection frequencies of the test when $\mu^2_{ww} = 10$ and $h_{\varepsilon w} = 0.9$ are about 5.5% and 24.7% for $L = 10, 20$, respectively. This represents a huge drop compared with the usual asymptotic critical values where these rejection frequencies were 16.7%, and 39.6%, respectively.

Table 1. Rejection frequencies (in %) at 5% nominal level, $n_w = 1$ and $B = 299$

Asymptotic χ^2 critical values

Statistics	L ↓ μ^2_{ww} →	$h_{\varepsilon w} = 0$				$h_{\varepsilon w} = 0.1$				$h_{\varepsilon w} = 0.5$				$h_{\varepsilon w} = 0.9$			
		0	0.05	1	10	0	0.05	1	10	0	0.05	1	10	0	0.05	1	10
		Restricted LIML															
AR	3	0.34	0.69	3.90	5.23	0.38	0.58	3.98	5.57	0.29	0.80	4.79	5.17	0.30	1.96	5.18	5.26
-	5	0.23	0.27	2.52	4.99	0.29	0.24	2.39	4.98	0.20	0.50	3.33	5.46	0.22	0.93	5.00	5.47
-	10	0.30	0.26	1.13	4.49	0.39	0.37	1.16	4.99	0.27	0.37	1.44	5.57	0.20	0.66	4.67	5.97
-	20	0.37	0.29	0.95	3.68	0.41	0.42	0.69	3.87	0.30	0.43	1.02	4.55	0.43	0.45	3.90	7.06
		Restricted 2SLS															
AR	3	2.81	2.73	4.48	5.54	2.60	2.74	4.34	5.17	3.00	3.44	6.03	5.48	2.46	11.89	10.98	6.54
-	5	3.42	3.78	4.38	5.26	3.69	3.38	4.82	5.49	3.48	4.06	7.06	6.71	3.32	11.31	19.43	8.63
-	10	4.86	4.38	5.04	6.17	4.33	4.41	5.19	5.84	4.71	4.87	8.11	8.10	4.34	9.84	34.83	16.65
-	20	6.53	6.04	6.47	7.27	6.14	6.78	6.41	7.12	6.22	6.12	9.01	11.65	6.61	8.84	45.00	39.63

Bootstrap critical values

Statistics	L ↓ μ^2_{ww} →	$h_{\varepsilon w} = 0$				$h_{\varepsilon w} = 0.1$				$h_{\varepsilon w} = 0.5$				$h_{\varepsilon w} = 0.9$			
		0	0.05	1	10	0	0.05	1	10	0	0.05	1	10	0	0.05	1	10
		Restricted LIML															
AR	3	0.65	1.28	7.57	4.71	1.46	0.75	2.18	5.75	0.30	1.32	3.79	4.79	1.35	4.31	6.00	3.72
-	5	0.57	0.47	6.01	4.48	1.27	0.62	2.94	3.45	0.92	0.47	3.91	2.52	0.21	0.75	8.56	5.30
-	10	0.35	0.27	1.58	3.31	0.33	0.46	1.82	2.82	0.18	0.23	1.04	3.25	0.33	0.26	5.65	5.24
-	20	0.04	0.14	0.16	2.92	0.39	0.34	0.63	2.31	0.20	0.63	0.68	3.74	0.26	0.20	2.22	4.11
		Restricted 2SLS															
AR	3	2.39	0.56	6.02	6.16	5.36	2.74	4.52	4.64	1.65	2.16	1.93	4.16	2.67	11.79	1.82	5.57
-	5	2.84	1.99	2.25	3.95	2.07	3.87	4.63	7.36	5.78	3.57	3.60	5.68	3.91	13.68	4.35	3.31
-	10	4.57	2.99	4.07	4.34	2.69	4.61	4.60	6.98	1.52	1.54	7.50	4.34	2.58	2.72	30.58	5.52
-	20	3.51	2.91	4.67	3.25	3.20	4.85	6.44	4.05	4.34	3.91	9.33	7.01	2.89	5.87	37.00	24.68

6. Bonferroni-based size correction

In this section, we provide a method to compute critical values for the subset AR statistics that yield tests with correct size uniformly over the nuisance parameters' space Θ. Without loss of generality, we focus on cases in which γ is not identifiable, i.e., the drifting subsequence of parameters $(\theta_n)_{n\geq 1}$ satisfying (3.2)–(3.4) with $\|h_{ww}\| < \infty$.

Let $\mathscr{H}$ be the space of the parameter h that characterize the distribution of $\xi_j(h)$,[9] i.e.

$$\mathscr{H} =: \left\{h \in \mathbb{R}^L : h =: (h_{ww}, h_{\varepsilon w}),\ \|h_{ww}\| < \infty,\ |h_{\varepsilon w}| \leq 1\right\}. \tag{6.1}$$

Let $c_{j,h}(\alpha)$ define the $(1-\alpha)^{th}$ quantile of the distribution of $\xi_j(h)$ for a given $h \in \mathscr{H}$. The least favorable critical value (LFCV) of the subset test $AR_n(\beta_0, \tilde{\gamma}_j)$ is defined [see Andrews and Guggenberger (2009) and McCloskey (2015)] by

$$c_{LF,j}(1-\alpha) \quad = \quad \sup_{h\in\mathscr{H}} c_{j,h}(\alpha),\ j \in \{LIML, 2SLS\}. \tag{6.2}$$

Now, consider the test that rejects H_0 when $AR_n(\beta_0, \tilde{\gamma}_j) > c_{LF,j}(\alpha)$. Then, it easy to see that

$$\begin{aligned}
\limsup_{n\to\infty} \sup_{\theta\in\Theta} \mathbb{P}_\theta\left[AR_n(\beta_0,\tilde{\gamma}_j) > c_{LF,j}(\alpha)\right] &= \limsup_{n\to\infty} \mathbb{P}_{\tilde{\theta}_n}\left[AR_n(\beta_0,\tilde{\gamma}_j) > c_{LF,j}(\alpha)\right] = \\
\lim_{n\to\infty} \mathbb{P}_{\tilde{\theta}_{\omega_{n,h}}}\left[AR_{\omega_{n,h}}(\beta_0,\tilde{\gamma}_j) > c_{LF,j}(\alpha)\right] &\leq \lim_{n\to\infty} \mathbb{P}_{\tilde{\theta}_{\omega_{n,h}}}\left[AR_{\omega_{n,h}}(\beta_0,\tilde{\gamma}_j) > c_{h,j}(\alpha)\right] \\
&= \mathbb{P}_\theta\left[\xi_j(h) > c_{h,j}(\alpha)\right] = \alpha,
\end{aligned} \tag{6.3}$$

where $\left\{\tilde{\theta}_n :\ n \geq 1\right\}$ is a sequence in Θ and $\{\omega_{n,h} :\ n \geq 1\}$ is a subsequence of $\{n :\ n \geq 1\}$ satisfying (3.1)–(3.4).[10] Equation (6.3) clearly shows that using $c_{LF,j}(\alpha)$, $j \in \{LIML, 2SLS\}$, yields tests with *correct asymptotic size* even if γ is not identified. However, there are two drawbacks related to the implementation of such tests. First, $c_{LF,j}(\alpha)$ must be computed over the entire parameter space $\mathscr{H}$, which represents a challenge when $L \geq 2$. Second, the computed $c_{LF,j}(\alpha)$ from (6.2) can be very large, thus yielding overly conservative tests. Therefore, there is a need to both reduce the dimension of $\mathscr{H}$ and adjusting $c_{LF,j}(\alpha)$ computed from the reduced space.

6.1. Reduction of the parameters' space and simple Bonferroni critical values

As seen in Section 3.2, the distribution of $\xi_j(h)$ depends on $h =: (h_{ww}, h_{\varepsilon w})$ only through $\mu_{ww} = \|h_{ww}\|$ and $h_{\varepsilon w}$. This implies that the distribution of $\xi_j(h)$ is invariance to the mapping that transforms $h =: (h_{ww}, h_{\varepsilon w}) \in \mathbb{R}^L \times [-1,1]$ into $\mu_{wh} =: (\mu_{ww}, h_{\varepsilon w}) \in [0,\infty) \times [-1,1]$. So, it suffices to compute

[9] See Lemma **3.2**.

[10] Andrews and Guggenberger (2009) and Guggenberger et al. (2012) show that the sequences $\left\{\tilde{\theta}_n :\ n \geq 1\right\}$ and $\left\{\omega_{n,h} :\ n \geq 1\right\}$ always exist.

$c_{LF,j}(\alpha)$ over the image of $\mathscr{H}$ by this mapping, i.e., over the set

$$\mathscr{H}_\mu =: \left\{\mu_{wh} = (\mu_{ww}, h_{\varepsilon w}) : \mu_{ww} = \|h_{ww}\| \geq 0,\ |h_{\varepsilon w}| \leq 1\right\}. \tag{6.4}$$

Clearly, in cases where the nuisance parameter γ is a scalar, finding $c_{LF,j}(\alpha) = \sup_{\mu_{wh} \in \mathscr{H}_\mu} c_{j,\mu_{wh}}(1-\alpha)$ only involves a maximization over two dimensions, no matter how large is a number of IVs. This obviously is less cumbersome than solving the problem (6.1), especially if $L \geq 3$.

We may even want to further reduce $\mathscr{H}_\mu$ by constructing $c_{LF,j}(\alpha)$ dependently upon the data through the (inconsistent) localization (semi)-estimated drifting sequence of parameters

$$\hat{\mu}_{n,h} =: (\hat{\mu}_{n,wh}, h_{n,F}) = (\hat{\mu}_{n,ww}, h_{n,\varepsilon w}, h_{n,F}), \tag{6.5}$$

where $h_{n,\varepsilon w} = (\sigma_{\varepsilon\varepsilon}\sigma_{V_wV_w})^{-1/2}\mathbb{E}_{F_n}(V_{w,i}\varepsilon_i)$ and $h_{n,F}$ are the sequences in (3.2)–(3.4), $\hat{\mu}_{n,ww} = \|\hat{h}_{n,ww}\| = \|\hat{\sigma}_{V_wV_w}^{-1/2}(Z'Z/n)^{1/2}\hat{\Pi}_{w,n}\|$, $\hat{\Pi}_{w,n}$ and $\hat{\sigma}_{V_wV_w}$ are consistent OLS estimators of $\sigma_{V_wV_w}$ and $\Pi_{w,n}$ in (2.2). Under the drifting subsequence of parameters $(\theta_n)_{n\geq 1}$ in Θ satisfying (3.2)–(3.4), and if further $\|h_{ww}\| < \infty$, both μ_{ww} and $h_{\varepsilon w}$ cannot be consistently estimated because IVs are weak. However, it is easy to provide a simple confidence set for μ_{ww} with a correct asymptotic coverage probability using the (inconsistent) estimator $\hat{\mu}_{hw}$, while such a simple valid confidence set[11] is not available for $h_{\varepsilon w}$. Because of this, we focus on reducing the dimension of $\mathscr{H}_\mu$ in the direction of μ_{ww}.

For this, consider the sequence of parameters $(\lambda_{n,h})_{n\geq 1}$ satisfies (3.2)–(3.4). Under Assumptions **2.1**–**2.3**, and **2.5**, it is straightforward to show that the following convergence holds jointly:

$$(n^{1/2}\hat{h}_{n,ww}, n^{1/2}\hat{\mu}_{n,ww}, h_{n,\varepsilon w}, h_{n,F}) \ \xrightarrow{d}\ (\widetilde{h}_{ww}, \widetilde{\mu}_{ww}, h_{\varepsilon w}, F), \tag{6.6}$$

where $\widetilde{h}_{ww} = h_{ww} + \psi_{V_w}$ and $\widetilde{\mu}_{ww} = \|\widetilde{h}_{ww}\|$. This means that $n^{1/2}\hat{\mu}_{n,ww}$ is not consistent to $\mu_{ww} = \|h_{ww}\|$ when $\|h_{ww}\| < \infty$. Nevertheless, we have $\widetilde{h}_{ww} \sim N(h_{ww}, I_L)$ and the projection-type confidence set

$$CI_\nu(\hat{h}_n) = [\hat{h}_{n,ww} - n^{-1/2}z_{1-\nu/2}\iota_L,\ \hat{h}_{n,ww} + n^{-1/2}z_{1-\nu/2}\iota_L] \tag{6.7}$$

for h_{ww} has an asymptotic coverage probability equal to $(1-\nu)$ for some $\nu \in (0,1)$, i.e., $\mathbb{P}[h_{ww} \in CI_\nu(\hat{h}_n)] \to 1-\nu$ as n increases, where ι_L is a $L \times 1$ column vector of ones and $z_{1-\nu/2}$ is the $(1-\nu)^{\text{th}}$ quantile of the standard normally distributed random variable. We may thus endeavor to adapt the data by maximizing $c_{j,h}(\alpha)$ not over the entire space $\mathscr{H}_\mu$, but only over $CI_\nu(\hat{h}_n) \times [-1,1]$. By the invariance property that leads to (6.4), this also amounts to maximizing $c_{j,\mu_{wh}}(\alpha)$ over $CI_\nu(\hat{\mu}_{n,wh}) \times$

[11] This is mainly because $h_{\varepsilon w}$ depends on the unidentified structural parameters γ and $\sigma_{\varepsilon\varepsilon}$ in a complicated way as given in (3.8), so that the usual Wald-type level confidence sets are not valid. See Doko Tchatoka and Dufour (2014) for further details on these issues.

$[-1,1]$, where

$$CI_{v}(\hat{\mu}_{n,wh}) = [\max(\hat{\mu}_{n,ww} - n^{-1/2} z_{1-v/2}, 0),\ \hat{\mu}_{n,ww} + n^{-1/2} z_{1-v/2}] \tag{6.8}$$

is the $(1-\nu)$ delta-type confidence set for $\mu_{ww} = \|h_{ww}\|$ obtained from (6.7). We then define the simple (as opposed to adjusted) Bonferroni critical value (SBCV) as

$$c_j^{B-S}(\alpha, \alpha-\delta, \hat{\mu}_{n,wh}) \;=\; \sup_{\mu_{wh} \in CI_{\alpha-\delta}(\hat{\mu}_{n,wh}) \times [-1,1]} c_{j,\mu_{wh}}(\delta),\ j \in \{LIML, 2SLS\}, \tag{6.9}$$

where $\nu = \alpha - \delta$ for some $\delta \in [0, \alpha)$. It will be useful to consider the following additional assumptions.

Assumption 6.1 *For some fixed* $\delta \in (0,1)$: *(i)* $c_{j,\mu_{wh}}(\delta)$ *is continuous as a function from* $\mathscr{H}_{\mu}$ *into* $\mathbb{R}$; *and (ii) the distribution of* $\xi_j(\mu_{wh})$ *is continuous at* $c_{j,\mu_{wh}}(\delta)$ *for all* $\mu_{wh} \in \mathscr{H}_{\mu}$.

Assumption 6.2 $\forall \nu \in [0,1]$ *and* $\forall \mu_{wh} \in \mathscr{H}_{\mu}$, $CI_{v} : \mathbb{R}_{\infty}^{+} \times [-1,1] \rightrightarrows \mathbb{R}_{\infty}^{+}$ *is continuous and compact valued with* $\mathbb{P}[\mu_{ww} \in CI_{v}(\tilde{\mu}_{wh})] \geq 1-\nu$, *where* $\tilde{\mu}_{wh} =: (\tilde{\mu}_{ww}, h_{\varepsilon w})$ *and* $\tilde{\mu}_{ww}$ *is defined in* (6.6).

We can now state the following theorem on the asymptotic size of the subset statistics when the simple Bonferroni critical values $c_j^{B-S}(\alpha, \alpha-\delta, \hat{\mu}_{n,wh})$ are used in the inference.

Theorem 6.3 *Suppose that Assumptions* **2.1**–**2.3**, *and* **2.5** *are satisfied. If further* H_0 *and Assumptions* **6.1**–**6.2** *hold with* $\nu = \alpha - \delta \in [0, \alpha)$, *then we have*:

$$AsySz_{AR_j}\left[c_j^{B-S}(\alpha, \alpha-\delta, \hat{\mu}_{n,wh})\right] \leq \alpha \quad \textit{for all } j \in \{LIML, 2SLS\}.$$

Theorem **6.3** shows that the simple Bonferroni critical values yield tests with correct asymptotic size whether γ is identified or not. However, they can be very large CVs so that the corresponding tests are overly conservative, especially small values of μ_{ww} (i.e., when the identification of γ is very weak). In Section (6.2), we seek an adjustment of the simple Bonferroni critical values that yields tests with better size properties in finite-sample, especial when the identification of γ is very weak.

6.2. Size adjusted Bonferroni critical values

Let $\alpha \in (0,\ 1)$ and $\delta \in [0, \alpha)$. Let also $c_{j,\mu_{wh}}(\delta)$ denote the $1-\delta$ quantile of the distribution of $\xi(\mu_{wh}) =: \xi_j(h_{\mu})$ for a given $\mu_{wh} \in CI_{\alpha-\delta}(\hat{\mu}_{n,wh}) \times [-1,1]$. Define the size-corrected factor

$$\widetilde{\eta}_n^j = \inf\Big\{\widetilde{\eta}^j \in \mathbb{R} : \sup_{\mu_{wh} \in CI_{\alpha-\delta}(\widetilde{\mu}_{wh}) \times [-1,1]} \mathbb{P}_{\theta}[\xi_j(\mu_{wh}) > c_j^{B-A}(\alpha, \alpha-\delta, \widetilde{\mu}_{wh}) + \widetilde{\eta}^j] \leq \alpha\Big\}, \tag{6.10}$$

where $\widetilde{\eta}^j =: \widetilde{\eta}^j(\mu_{wh})$ and $c_j^{B-S}(\alpha, \alpha-\delta, \widetilde{\mu}_{wh})$ is given by (6.9). As in Section 6.1, we make the following assumptions on the behavior of $c_{j,\cdot}(\cdot)$ and $\widetilde{\eta}^j(\cdot)$.

Assumption 6.4 *For some pair* $(\delta_*, \delta^*) \in [0, \alpha]$ *with* $0 \leq \delta_* \leq \alpha - \delta^*$, *as a function of* μ_{wh} *and* δ, $c_{j,\mu_{wh}}(\delta)$ *is continuous over* $\mathscr{H}_\mu \times [\delta_*, \alpha - \delta^*]$.

Assumption 6.5 (*i*) $\widetilde{\eta}^j(\cdot) : \mathbb{R}_\infty^+ \times [-1, 1] \rightrightarrows \mathbb{R}$ *is continuous*; (*ii*) $\mathbb{P}_\theta[\xi_j(\mu_{wh}) > c_j^{B-S}(\alpha, \nu, \widetilde{\mu}_{wh}) + \widetilde{\eta}^j(\mu_{wh})] \leq \alpha$ *for all* $\mu_{wh} \in \mathscr{H}_\mu$.

We can now state the following theorem on the uniform validity of the subset AR tests.

Theorem 6.6 *Suppose that Assumptions* **2.1–2.3**, *and* **2.5** *are satisfied. If further* H_0 *and Assumptions* **6.2–6.5** *hold, then*:

$$AsySz_{AR_j}[c_j^{B-A}(\alpha, \alpha-\delta, \widetilde{\mu}_{n,wh}) + \widetilde{\eta}_n^j] \leq \alpha \quad \textit{for all } j \in \{LIML, 2SLS\}\,.$$

Theorem **6.6** shows that the adjusted Bonferroni critical values yield tests with correct asymptotic size no matter how weak the identification of γ is. In practice, we propose the following algorithm for the computation of $c_j^{B-A}(\alpha, \alpha-\delta, \widetilde{\mu}_{n,wh})$ and $\widetilde{\eta}_n^j$ for all $j \in \{LIML, 2SLS\}$.

1. Choose the desired nominal α, ν and δ such that Assumptions **6.2–6.5** hold, and compute $CI_\nu(\hat{\mu}_{n,wh})$ following (6.8).

2. Create a fine gride of the space $CI_\nu(\hat{\mu}_{n,wh}) \times [-1, 1]$ and call it $\mathscr{H}_\nu^{grid}$.

3. For each $\mu_{wh} \in \mathscr{H}_\nu^{grid}$, simulate R draws of the asymptotic distribution $\xi_j(\mu_{wh})$ of the subset statistic $AR_n(\beta_0, \tilde{\gamma}_j)$ and:

   (a) find $c_j^{B-A}(\alpha, \alpha-\delta, \hat{\mu}_{n,wh}) = \sup_{\mu_{wh} \in \mathscr{H}_\nu^{grid}} c_{j,\mu_{wh}}(\delta)$. Then set $S_\eta^j = [-c_j^{B-A}(\alpha, \alpha - \delta, \hat{\mu}_{n,wh}), c_j^{B-A}(\alpha, \alpha-\delta, \hat{\mu}_{n,wh})]$ and create a fine gride of S_η^j : call it S_η^{jgrid};

   (b) compute $\widetilde{\eta}_n^j \in S_\eta^{jgrid}$ such that $\sup \mathbb{P}_\theta[\xi_j(\mu_{wh}) > c_j^{B-A}(\alpha, \alpha-\delta, \hat{\mu}_{n,wh}) + \widetilde{\eta}_n^j]$ over $(\mu_{wh}, \widetilde{\eta}^j) \in CI_{\alpha-\delta}(\hat{\mu}_{n,wh}) \times [-1, 1] \times S_\eta^{jgrid}$ is less or equal to α.

It worth noting that while in theory $\widetilde{\eta}_n^j \in \mathbb{R}$, the simulations show the solution of (6.10) always lies in the interval $S_\eta^j = [-c_j^{B-A}(\alpha, \alpha-\delta, \hat{\mu}_{n,wh}),\, c_j^{B-A}(\alpha, \alpha-\delta, \hat{\mu}_{n,wh})]$. Therefore, the optimization is run over a fine grid of this interval in 3.(b).

6.3. Finite-sample performance with the adjusted critical values

In this section, we examine the performance of the subset AR tests when the above adjusted Bonferroni-type critical values are applied. To do this, we consider again the setup of the Monte

Carlo experiment described in Section 5. The adjusted Bonferroni-type critical values are computed following the algorithm below Theorem **6.6** with $\alpha = 0.05$, $\delta = 0.025$ and $R = 100,000$ draws.

Table 2 presents the results. Despite a relatively small sample size ($n = 100$), it is straightforward to see that the subset tests with adjusted Bonferroni-type critical values outperform the ones with standard and bootstrap critical values in Table 1, especially when the nuisance parameter γ is not identified (i.e., when $\mu^2_{ww} \in \{0, 0.05, 1\}$). More interestingly, the size adjustment works very well even for the subset test with the restricted 2SLS estimator, as the size distortions of this test observed in Table 1 when $\mu^2_{ww} \in \{0, 0.05, 1\}$ have completely disappeared in Table 2.

Table 2. Rejection frequencies (in %) at 5% nominal level with Bonferroni-type adjusted critical values

Restricted LIML

Statistics	L ↓ μ^2_{ww} →	$h_{\varepsilon w}=0$				$h_{\varepsilon w}=0.1$				$h_{\varepsilon w}=0.5$				$h_{\varepsilon w}=0.9$			
		0	0.05	1	10	0	0.05	1	10	0	0.05	1	10	0	0.05	1	10
AR	5	3.98	4.91	4.60	4.99	2.01	3.95	4.04	4.91	2.12	3.01	3.98	3.91	2.99	3.5	4.97	4.80
-	10	4.02	4.95	4.7	5.02	4.00	4.01	4.04	5.03	3.00	3.98	3.99	4.05	3.01	4.01	5.11	4.99
-	20	5.14	5.01	4.85	5.1	4.91	4.71	4.67	4.96	4.50	4.75	4.91	5.01	4.01	4.04	4.97	5.01

Restricted 2SLS

Statistics	L ↓ μ^2_{ww} →	$h_{\varepsilon w}=0$				$h_{\varepsilon w}=0.1$				$h_{\varepsilon w}=0.5$				$h_{\varepsilon w}=0.9$			
		0	0.05	1	10	0	0.05	1	10	0	0.05	1	10	0	0.05	1	10
AR	5	3.00	3.96	4.03	5.03	3.50	4.03	4.00	4.90	4.00	3.98	4.00	4.50	3.00	4.00	4.01	4.51
-	10	4.01	4.02	4.51	5.06	4.00	4.01	3.01	5.01	4.01	4.97	4.10	3.60	2.60	4.21	4.02	4.62
-	20	4.63	4.81	5.02	5.02	4.85	4.98	4.90	5.03	4.05	5.09	4.99	4.94	4.01	4.90	5.04	5.12

7. Conclusions

In this paper, we study the asymptotic validity of the bootstrap for the plug-in subset AR tests based on the restricted limited information maximum likelihood (LIML) and two-stage leas squares (2SLS). We consider linear IV regressions where structural parameters may not be identified, and provide a characterization of the asymptotic distributions of both statistics without and with weak instruments. Our results provide some new insights and extensions of earlier studies. We show that the asymptotic distributions of these statistics are non-standard when the nuisance parameters that are not specified by the subset null hypothesis are not identified, so correction to usual asymptotic critical values are needed. We find that the bootstrap procedures similar to that of Moreira et al. (2009) provide a high-order refinement of the null distributions of the statistics when the nuisance parameters are identified, but is inconsistent if these parameters are not identified. This contrasts with Moreira et al. (2009) who show that bootstrap is valid for the AR statistic of the null hypothesis specified on the full vector of structural parameters, whether identification is strong or weak. The inconsistency of bootstrap for subset AR statistics studied is mainly due to its inability to mimic the *concentration factor* that characterizes the strength of the identification of the nuisance parameters. The inconsistency of bootstrap for subset AR statistics is mainly due to its inability to mimic the *concentration factor* that characterizes the strength of the identification of the nuisance parameters. We thus develop a Bonferroni-based size adjustment that yields tests with correct asymptotic size, whether the nuisance parameters are identified or not. We present a Monte Carlo experiment that confirms our theoretical findings.

A. Appendix

We begin by presenting the supplemental lemmata in Section A.1. Section A.2 contains proofs.

A.1. Supplemental lemmata

Lemma A.1 *Suppose Assumptions* **2.1–2.4** *are satisfied. If further H_0 hold and $\Pi_w \neq 0$ is fixed, then for some integer $r \geq 1$, we have*:

$$\sup_{\tau\in\mathbb{R}}\left|\mathbb{P}_{\theta}[\tilde{S}_n(\beta_0,\tilde{\gamma}_j) \leq \tau] - \Phi(\tau) - \sum_{h=1}^{r} n^{-h/2} p^h_{S_{n,j}}(\tau;F_{R_n},\beta_0,\Pi_w,\tilde{\gamma}_j)\phi(\tau)\right| = o(n^{-r/2})$$

for all $j \in \{LIML, 2SLS\}$, where $p^h_{S_{n,j}}$ is a polynomial in τ with coefficients depending on β_0, Π_w, $\tilde{\gamma}_j$, and the moments of F_{R_n}.

Lemma A.2 *Suppose that Assumptions* **2.1–2.3**, *and* **2.5** *are satisfied. If further H_0 holds and $(\lambda_{n,h})_{n\geq 1}$ satisfies* (3.2)–(3.4), *then we have*:

(a) $\tilde{\gamma}_j - \gamma_0 \xrightarrow{d} \sigma_{\varepsilon\varepsilon}^{1/2}\sigma_{V_wV_w}^{-1/2}\Delta_{h,j}$, *where* $\Delta_{h,j} = (\Psi_h'\Psi_h - \kappa_{h,j})^{-1}(\Psi_h'\psi_\varepsilon - \kappa_{h,j}h_{\varepsilon w})$, $\kappa_{h,2SLS} = 0$ *and* $\kappa_{h,LIML}$ *is the smallest root of* $\left|(\psi_\varepsilon : \Psi_h)'(\psi_\varepsilon : \Psi_h) - \kappa_h\Sigma_h\right| = 0$;

(b) $\tilde{r}_j'\hat{\Omega}_W\tilde{r}_j \xrightarrow{d} \sigma_{\varepsilon\varepsilon,j} = \sigma_{\varepsilon\varepsilon}\left(1 - 2h_{\varepsilon w}\Delta_{h,j} + \Delta_{h,j}^2\right)$;

(c) $\tilde{S}_n(\beta_0, \tilde{\gamma}_j) \xrightarrow{d} \left(1 - 2h_{\varepsilon w}\Delta_{h,j} + \Delta_{h,j}^2\right)^{-1/2} S_{h,j}$, *where* $S_{h,j} = \psi_\varepsilon - \Psi_h\Delta_{h,j}$.

where $h =: (h_{ww}, h_{\varepsilon w})$.

Lemma **A.2** - (a) shows that both the restricted LIML and 2SLS estimators are inconsistent under the sequence $(\lambda_{n,h})_{n\geq 1}$ satisfies (3.2)–(3.4) if $\|h_{ww}\| < \infty$. Under Assumptions **2.1**–**2.3**, and **2.5**, we have $vec(\psi_\varepsilon, \psi_{V_w}) \sim N(0,\ \Sigma_h \otimes I_L)$ so that $\psi_\varepsilon \mid \psi_{V_w} \sim N\big(h_{\varepsilon w}\psi_{V_w},\ (1 - h_{\varepsilon w}^2)I_L\big)$. As a result,

$$\tilde{\gamma}_{2SLS} - \gamma_0 \mid \psi_{V_w} \sim N\Big(h_{\varepsilon w}(\Psi_h'\Psi_h)^{-1}\Psi_h'\psi_{V_w},\ (1 - h_{\varepsilon w}^2)(\Psi_h'\Psi_h)^{-1}\Big). \tag{A.1}$$

We see from (A.1) that the unconditional distribution of $\tilde{\gamma}_{2SLS} - \gamma_0$ is a mixture of Gaussian processes with nonzero means. This is not the case for the LIML estimator. Indeed, since $\kappa_{LIML} \neq 0$ a.s., $\tilde{\gamma}_{LIML} - \gamma_0 \mid \psi_{V_w}$ does not necessarily follow a Gaussian process. Hence, $\tilde{\gamma}_{LIML} - \gamma_0$ does not necessarily converge to a mixture of Gaussian processes [similar to Doko Tchatoka (2014)].

Lemma **A.2** - (b) shows that $\tilde{r}_j'\hat{\Omega}_W\tilde{r}_j$ converges to a random process if γ is weakly identified. This contrast with the case where γ is identified so that $\tilde{r}_j'\hat{\Omega}_W\tilde{r}_j$ converges in probability to a positive scalar. Similarly, $\tilde{S}_n(\beta_0, \tilde{\gamma}_j)$ does not have a standard normal distribution even for $j = 2SLS$, as it is the case with strong instruments [see Lemma **A.2** - (c)]. While $\tilde{S}_n(\beta_0, \tilde{\gamma}_{2SLS})$ follows asymptotically a mixture of Gaussian processes with nonzero mean under H_0, the limiting distribution of $\tilde{S}_n(\beta_0, \tilde{\gamma}_{LIML})$ is nonstandard because of the presence of κ_{LIML} in it.

Lemma A.3 *Suppose Assumptions* **2.1**–**2.4** *are satisfied. If further* H_0 *hold and* $\Pi_w \neq 0$ *is fixed, then for some integer* $r \geq 1$, *we have*:

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

for all $j \in \{LIML, 2SLS\}$, *where* $p_{\tilde{S}_{n,j}}^h$ *is a polynomial in* τ *with coefficients depending on* $\tilde{\gamma}_j$, $\hat{\Pi}_x$, $\hat{\Pi}_w$, *and the moments of* $\widetilde{F}_n$, $\Phi(\cdot)$ *and* $\phi(\cdot)$ *are the cdf and pdf of a standard normally distributed random variable.*

Lemma A.4 *Suppose that Assumptions* **2.1**–**2.3**, *and* **2.5** *are satisfied. If further* H_0 *holds and* $(\lambda_{n,h})_{n\geq 1}$ *satisfies* (3.2)–(3.4), *then the following convergence holds jointly for* $j \in \{LIML, 2SLS\}$:

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

(b) $E^*\left[V^*_{w,i}(v^*_{1,i} - V^{*\prime}_{w,i}\tilde{\gamma}_j)\right] \xrightarrow{d} \sigma_{V_w\varepsilon,j} = \sigma_{V_w\varepsilon} - \sigma^{1/2}_{\varepsilon\varepsilon}\sigma^{1/2}_{V_wV_w}\Delta_{h,j}$ *a.s.*;

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

Lemma A.5 *Suppose that Assumptions* **2.1–2.3**, *and* **2.5** *are satisfied. Suppose also that* H_0 *holds and* $(\lambda_{n,h})_{n\geq 1}$ *satisfies* (3.2)–(3.4). *If further* $\mathbb{E}(\|Z_i\|^{4+\delta}, \|(\varepsilon_i, V_{wi})'\|^{2+\delta}) < +\infty$ *for some* $\delta > 0$, *then we have*:

$$\begin{pmatrix} \left(\frac{Z^{*\prime}Z^*}{n}\right)^{-1/2}\left(\frac{Z^{*\prime}(v_1^*:V_w^*)}{\sqrt{n}}\right)\left(\hat{\Omega}_W\right)^{-1/2} \\ n^{-1/2}\left(n^{-1}M^{*\prime}\mathbb{1}_n - n^{-1}M'\mathbb{1}_n\right)\end{pmatrix} \mid \mathscr{X}_n \to^d \begin{pmatrix}\psi_{V_1} \\ \psi_{V_w} \\ \psi_m\end{pmatrix} \sim N\left(\begin{pmatrix} I_{2L} & 0 \\ 0 & \Omega_{mm}\end{pmatrix}\right) \text{ a.s.},$$

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* $M^* = (m^*_1, \ldots, m^*_n)$, $m^*_i = vech(Z^*_iZ^{*\prime}_i)$ *are the bootstrap counterparts of M and m; and* $\mathbb{1}_n$ *is a* $n\times 1$ *vector of ones.*

Lemma A.6 *Suppose that Assumptions* **2.1–2.3**, *and* **2.5** *are satisfied. Suppose also that* H_0 *holds and* $(\lambda_{n,h})_{n\geq 1}$ *satisfies* (3.2)–(3.4). *If further* $\mathbb{E}(\|Z_i\|^{4+\delta}, \|(\varepsilon_i, V_{wi})'\|^{2+\delta}) < +\infty$ *for some* $\delta > 0$, *then we have*:

(a) $\tilde{\gamma}^*_j - \tilde{\gamma}_j \mid \mathscr{X}_n \xrightarrow{d} \sigma^{1/2}_{\varepsilon\varepsilon,j}\sigma^{-1/2}_{V_wV_w}\Delta^B_{h,j}$ *a.s.*;

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

(c) $\hat{S}^*(\beta_0, \tilde{\gamma}^*_j) \mid \mathscr{X}_n \xrightarrow{d} \left(1 - 2h_{\varepsilon w,j}\Delta^B_{h,j} + (\Delta^B_{h,j})^2\right)^{-1/2} S^B_{h,j}$ *a.s.*

for $j \in \{LIML, 2SLS\}$ *where* $\Delta^B_{h,j} = \left(\Psi^{B\prime}\Psi^B_h - \kappa^B_{h,j}\right)^{-1}\left(\Psi^{B\prime}\psi_{\varepsilon,j} - \kappa^B_{h,j}h_{\varepsilon w,j}\right)$, *when* $\kappa^B_{2SLS} = 0$ *and* κ^B_{LIML} *is the smallest root of the determinantal equation* $\left|\left(\psi_{\varepsilon,j} : \Psi^B_h\right)'\left(\psi_{\varepsilon,j} : \Psi^B_h\right) - \kappa\Sigma_{h,j}\right| = 0$, $\Psi^B_h = \Psi + \psi_{V_w}$, $\psi_{\varepsilon,j} = \psi_{V_1} - \psi_{V_w}\left(\gamma_0 + \sigma^{1/2}_{\varepsilon\varepsilon}\sigma^{-1/2}_{V_wV_w}\Delta_{h,j}\right) = \psi_\varepsilon - \sigma^{1/2}_{\varepsilon\varepsilon}\sigma^{-1/2}_{V_wV_w}\Delta_{h,j}\psi_{V_w}$, $\Sigma_{h,j} = \begin{pmatrix} 1 & h_{\varepsilon w,j} \\ h_{\varepsilon w,j} & \end{pmatrix}$, $h_{\varepsilon w,j} = (\sigma_{\varepsilon\varepsilon,j}\sigma_{V_wV_w})^{-1/2}\sigma_{V_w\varepsilon,j} = (1 - 2h_{\varepsilon w}\Delta_{h,j} + \Delta^2_{h,j})^{-1/2}(h_{\varepsilon w} - \Delta_{h,j})$, $\sigma_{\varepsilon\varepsilon,j} = \sigma_{\varepsilon\varepsilon}(1 - 2h_{\varepsilon w}\Delta_{h,j} + \Delta^2_{h,j})$, $\sigma_{V_w\varepsilon,j} = \sigma_{V_w\varepsilon} - \sigma^{1/2}_{\varepsilon\varepsilon}\sigma^{1/2}_{V_wV_w}\Delta_{h,j}$, $S^B_{h,j} = \psi_{\varepsilon,j} - \Psi^B_h\Delta^B_{h,j}$, *and* $\Delta_{h,j}$ *is given in Lemma* **A.2**.

The asymptotic behavior of the bootstrap statistics in Lemma **A.6** (a)-(c) differ substantially from those in Lemma **A.2** (a)-(c) if $\|h_{ww} < \infty\|$. Therefore, the bootstrap fails to mimic the asymptotic distributions of these statistics under H_0 and the subsequence of parameters $(\lambda_{n,h})_{n\geq 1}$ satisfying (3.2)–(3.4) when γ is not identified (i.e., when $\|h_{ww} < \infty\|$). For example, while $\tilde{\gamma}_j - \gamma_0 \xrightarrow{d} \sigma^{1/2}_{\varepsilon\varepsilon}\sigma^{-1/2}_{V_wV_w}\Delta_{h,j}$ in Lemma **A.2**-(a), we have $\tilde{\gamma}^*_j - \tilde{\gamma}_j \mid \mathscr{X}_n \xrightarrow{d} \sigma_{\varepsilon\varepsilon}(1 - 2h_{\varepsilon w}\Delta_{h,j} + \Delta^2_{h,j})^{1/2}\sigma^{1/2}_{\varepsilon\varepsilon}\sigma^{-1/2}_{V_wV_w}\Delta^B_{h,j}$ *a.s.* in Lemma **A.6**-(a), where $\Delta_{h,j} \neq \Delta^B_{h,j}$ with probability one if $\|h_{ww} < \infty\|$.

A.2. Proofs

PROOF OF LEMMA **A.1** We show the proof for the statistic $\tilde{S}_n(\beta_0, \tilde{\gamma}_{2SLS})$. The proof for $\tilde{S}_n(\beta_0, \tilde{\gamma}_{LIML})$ can be deduced similarly. First, observe that we can express $\tilde{S}_n(\beta_0, \tilde{\gamma}_{2SLS})$ under H_0 as $\tilde{S}_n(\beta_0, \tilde{\gamma}_{2SLS}) = \sqrt{n}N/D$, where

$$
\begin{aligned}
N &= \left(\frac{Z'Z}{n}\right)^{-1/2}\left[\frac{Z'\tilde{y}(\beta_0)}{n} - \left(\frac{Z'W}{n}\right)\tilde{\gamma}_{2SLS}\right] \\
&= \left(\frac{Z'Z}{n}\right)^{-1/2}\left[\frac{Z'\varepsilon}{n} - \left(\frac{Z'W}{n}\right)\left[\left(\frac{W'Z}{n}\right)\left(\frac{Z'Z}{n}\right)^{-1}\left(\frac{Z'W}{n}\right)\right]^{-1}\left(\frac{W'Z}{n}\right)\left(\frac{Z'Z}{n}\right)^{-1}\left(\frac{Z'\varepsilon}{n}\right)\right], \quad (A.2) \\
D &= \frac{\tilde{y}'(\beta_0)M_Z\tilde{y}(\beta_0)}{n-L} - 2\frac{\tilde{y}'(\beta_0)M_Z W}{n-L}\tilde{\gamma}_{2SLS} + \frac{W'M_Z W}{n-L}(\tilde{\gamma}_{2SLS})^2, \quad (A.3) \\
\tilde{\gamma}_{2SLS} &= \left[\left(\frac{W'Z}{n}\right)\left(\frac{Z'Z}{n}\right)^{-1}\left(\frac{Z'W}{n}\right)\right]^{-1}\left(\frac{W'Z}{n}\right)\left(\frac{Z'Z}{n}\right)^{-1}\left(\frac{Z'\tilde{y}(\beta_0)}{n}\right), \text{ and } \quad \tilde{y}(\beta_0) = y - X\beta_0.
\end{aligned}
$$

So,from (A.2)-(A.3), we can write $S(\beta_0, \tilde{\gamma}_{2SLS})$ under H_0 as

$$\tilde{S}_n(\beta_0, \tilde{\gamma}_{2SLS}) = \sqrt{n}H(\bar{R}_n) = \sqrt{n}[H(\bar{R}_n) - H(\mu)] \quad (A.4)$$

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 under the sequences $\{\Pi_{w,n} = 0 : n \geq 1\}$ and does not even exist under the sequences $\{\Pi_{w,n} = \Pi_{w0}c_n : c_n \downarrow 0 \forall n \geq 1, \Pi_{w0} \in \mathbb{R}^L \text{ is fixed}\}$; see Moreira et al. (2009, footnote 2) and Doko Tchatoka (2013) for a similar result. Lemma **A.1** follows by applying Bhattacharya and Ghosh (1978, Theorem 2) to (A.4) with $s - 2 = r$. □

PROOF OF THEOREM **3.1** From (2.9), we have $L \times AR_n(\beta_0, \tilde{\gamma}_j) = \left\|\tilde{S}_n(\beta_0, \tilde{\gamma}_j)\right\|^2$. We want to approximate $\mathbb{P}_\theta[AR_n(\beta_0, \tilde{\gamma}_j) \leq \tau]$ uniformly in Θ under H_0. First, we can write $\mathbb{P}_\theta[AR_n(\beta_0, \tilde{\gamma}_j) \leq \tau]$ as:

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

where $\mathscr{C}_\tau = \{x \in \mathbb{R}; x^2 \leq \tau\}$ 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 $W_n = \tilde{S}_n(\beta_0, \tilde{\gamma}_j),\ j \in \{2SLS, LIML\}$. By using the approximation of $\mathbb{P}_\theta[\tilde{S}_n(\beta_0, \tilde{\gamma}_j) \leq \tau]$ in Lemma **A.1** 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 **A.2** (a) We start with the proof for $j = 2SLS$. We have

$$\begin{aligned}
\frac{W'P_Z W}{\sigma_{V_wV_w}} &= \sigma^{-1}_{V_wV_w}\left(\Pi'_{w,n}Z'Z\Pi_{w,n} + V'_wZ\Pi_{w,n} + \Pi'_{w,n}Z'V_w + V'_wP_ZV_w\right) \\
&= \left(n^{1/2}\sigma^{-1/2}_{V_wV_w}Q_Z^{1/2}\Pi_{w,n}\right)' Q_Z^{-1/2}\left(\frac{Z'Z}{n}\right)Q_Z^{-1/2}\left(n^{1/2}\sigma^{-1/2}_{V_wV_w}Q_Z^{1/2}\Pi_{w,n}\right) + \\
&\quad \left(\sigma^{-1/2}_{V_wV_w}\frac{Z'V_w}{\sqrt{n}}\right)' Q_Z^{-1/2}\left(n^{1/2}\sigma^{-1/2}_{V_wV_w}Q_Z^{1/2}\Pi_{w,n}\right) + \left(n^{1/2}\sigma^{-1/2}_{V_wV_w}Q_Z^{1/2}\Pi_{w,n}\right)' Q_Z^{-1/2}\left(\sigma^{-1/2}_{V_wV_w}\frac{Z'V_w}{\sqrt{n}}\right) \\
&\quad \left(\sigma^{-1/2}_{V_wV_w}\frac{Z'V_w}{\sqrt{n}}\right)'\left(\frac{Z'Z}{n}\right)^{-1}\left(\sigma^{-1/2}_{V_wV_w}\frac{Z'V_w}{\sqrt{n}}\right) \\
&\xrightarrow{d} h'_{ww}h_{ww} + \psi'_{V_w}h_{ww} + h'_{ww}\psi_{V_w} + \psi'_{V_w}\psi_{V_w} = (h_{ww}+\psi_{V_w})'(h_{ww}+\psi_{V_w}) = \Psi'_h\Psi_h.
\end{aligned}$$

By the same token, we have

$$\begin{aligned}
\left(\sigma_{\varepsilon\varepsilon}\sigma_{V_wV_w}\right)^{-1/2}W'P_Z\varepsilon &= \left(\sigma_{\varepsilon\varepsilon}\sigma_{V_wV_w}\right)^{-1/2}\left(\Pi'_{w,n}Z'\varepsilon + V'_wP_Z\varepsilon\right) \\
&= \left(n^{1/2}\sigma^{-1/2}_{V_wV_w}Q_Z^{1/2}\Pi_{w,n}\right)'\left(\sigma^{-1/2}_{\varepsilon\varepsilon}Q_Z^{-1/2}\frac{Z'\varepsilon}{\sqrt{n}}\right) \\
&\quad + \left(\sigma^{-1/2}_{V_wV_w}\frac{Z'V_w}{\sqrt{n}}\right)'\left(\frac{Z'Z}{n}\right)^{-1}\left(\sigma^{-1/2}_{\varepsilon\varepsilon}\frac{Z'\varepsilon}{\sqrt{n}}\right) \\
&\xrightarrow{d} h'_{ww}\psi_\varepsilon + \psi'_{V_w}\psi_\varepsilon = \Psi'_h\psi_\varepsilon.
\end{aligned}$$

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 $P_Z(\tilde{y}(\beta_0):W) = P_Z\left[Z\Pi_{w,n}(\gamma:) + (\varepsilon:V_W)\begin{pmatrix}1 & 0\\ \gamma & 1\end{pmatrix}\right]$. 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

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

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

$$\left(\sigma^{-1}_{\varepsilon\varepsilon}\sigma_{V_wV_w}\right)^{1/2}\left(\tilde{\gamma}_{LIML} - \gamma_0\right) \xrightarrow{d} \Delta_{h,LIML} = \left\{\Psi'_h\Psi_h - \kappa_{h,LIML}\right\}^{-1}\left\{\Psi'_h\psi_\varepsilon - \kappa_{h,LIML}h_{\varepsilon w}\right\},$$

where $\kappa_{h,LIML}$ is the smallest root of the characteristic polynomial

$$\left|(\psi_\varepsilon : \Psi_h)'(\psi_\varepsilon : \Psi_h) - \kappa_h \Sigma_h\right| = 0,\ \Sigma_h = \begin{pmatrix} 1 & h_{\varepsilon w} \\ h_{\varepsilon w} & 1 \end{pmatrix}.$$

Thus, the result follows.

(b) Similarly, we have

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

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

$$\begin{aligned}
(Z'Z)^{-1/2} Z' (\tilde{y}(\beta_0) : W) \tilde{r}_j &= \left(\frac{Z'Z}{n}\right)^{-1/2} \frac{Z'\varepsilon}{\sqrt{n}} + \left(\frac{Z'Z}{n}\right)^{-1/2} \frac{Z'W}{\sqrt{n}} \left(\gamma_0 - \tilde{\gamma}_j\right) \\
&= \sigma_{\varepsilon\varepsilon}^{1/2} \left\{ \sigma_{\varepsilon\varepsilon}^{-1/2} \left(\frac{Z'Z}{n}\right)^{-1/2} \frac{Z'\varepsilon}{\sqrt{n}} + \left(n^{1/2} \sigma_{V_w V_w}^{-1/2} \left(\frac{Z'Z}{n}\right)^{1/2} \Pi_{w,n} \right.\right. \\
&\quad \left.\left. + \sigma_{V_w V_w}^{-1/2} \left(\frac{Z'Z}{n}\right)^{-1/2} \frac{Z'V_W}{\sqrt{n}} \right) \sigma_{\varepsilon\varepsilon}^{-1/2} \sigma_{V_w V_w}^{1/2} \left(\gamma_0 - \tilde{\gamma}_j\right) \right\} \\
&\xrightarrow{d} \sigma_{\varepsilon\varepsilon}^{1/2} \left\{ \psi_\varepsilon - \Psi_h \left(\Psi_h' \Psi_h - \kappa_{h,j}\right)^{-1} \left(\Psi_h' \psi_\varepsilon - \kappa_{h,j} h_{\varepsilon w}\right) \right\} \\
&= \sigma_{\varepsilon\varepsilon}^{1/2} (\psi_\varepsilon - \Psi_h \Delta_{h,j}).
\end{aligned}$$

Combing this with (b), the result follows.

□

PROOF OF THEOREM **3.2** The proof follows immediately from equation (2.9) and Lemma **A.2**. Therefore, it is omitted. □

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

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

PROOF OF THEOREM **4.2** The proof is similar to that of Hall and Horowitz (1996, Theorem 3) upon exploiting the results of Theorem **3.1** and Lemma **4.1**, hence it is omitted. □

PROOF OF LEMMA **A.4** (a). First, we have

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

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

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

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

□

PROOF OF LEMMA **A.5** 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

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

where $V^*_i=\left(v^*_{1,i},V^{*\prime}_{w,i}\right)'$ is the ith bootstrap draw of the (re-centered) reduced-form residuals, $\bar{m}=n^{-1}\sum_{i=1}^{n}m_i$. 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 $\mathbb{E}^*[V^*_iV^{*\prime}_i]=\tilde{V}'\tilde{V}/n$, where $\tilde{V}=(\tilde{v}_i:\tilde{V}_w)$, and $\mathbb{E}^*[Z^*_iZ^{*\prime}_i]=Z'Z/n$, we have

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

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

(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 $\mathbb{E}^*\left[n^{-1}Z^{*\prime}Z^*\right]=n^{-1}Z'Z$ and $n^{-1}Z^{*\prime}Z^*-n^{-1}Z'Z\mid\mathscr{X}_n\stackrel{a.s.}{\to}$

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 $\hat{\Omega}_W \xrightarrow{p} \Omega_W$. Therefore, Lemma **A.5** follows by applying the Liapunov Central Limit Theorem. □

PROOF OF LEMMA **A.6**

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

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

and similarly, we have

$$\begin{aligned}
& \left(\mathbb{E}^*\left[(v^*_{1,i} - V^*_{w,i}\tilde{\gamma}_j)^2\right]\mathbb{E}^*\left[V^*_{w,i}V^{*'}_{w,i}\right]\right)^{-1/2} W^{*'}P_{Z^*}\left(V^*_1 - V^*_w\tilde{\gamma}_j\right) \\
= & \left(\mathbb{E}^*\left[(V^*_{1,i} - V^*_{w,i}\tilde{\gamma}_j)^2\right]\mathbb{E}^*\left[V^*_{w,i}V^{*'}_{w,i}\right]\right)^{-1/2}\left\{\hat{\Pi}'_w Z^{*'}\left(V^*_1 - V^*_w\tilde{\gamma}_j\right) + V^{*'}_w P_{Z^*}\left(V^*_1 - V^*_w\tilde{\gamma}_j\right)\right\} \\
= & \left(\mathbb{E}^*\left[(V^*_{1,i} - V^*_{w,i}\tilde{\gamma}_j)^2\right]\mathbb{E}^*\left[V^*_{w,i}V^{*'}_{w,i}\right]\right)^{-1/2}\left\{\Pi^{B'}_{w,n}\left(\frac{Z^{*'}\left(V^*_1 - V^*_w\tilde{\gamma}_j\right)}{\sqrt{n}}\right)\right. \\
& \left. + \left(\frac{V^{*'}_w Z^*}{\sqrt{n}}\right)\left(\frac{Z^{*'}Z^*}{n}\right)^{-1}\left(\frac{Z^{*'}\left(V^*_1 - V^*_w\tilde{\gamma}_j\right)}{\sqrt{n}}\right)\right\}.
\end{aligned}$$

Since $n^{-1}Z^{*'}Z^* \mid \mathscr{X}_n \xrightarrow{a.s.} \mathbb{E}(Z_iZ'_i) = Q_Z$ from Lemmata **A.4-A.5**, we have

$$\begin{aligned}
\left(\mathbb{E}^*\left[V^*_{w,i}V^{*'}_{w,i}\right]\right)^{-1} W^{*'}P_{Z^*}W^* \quad \mid \mathscr{X}_n \xrightarrow{d} & \quad (h_{ww} + 2\psi_{V_w})'(h_{ww} + 2\psi_{V_w})\,a.s. \\
= & \quad \left(\Psi_h + \psi_{V_w}\right)'\left(\Psi_h + \psi_{V_w}\right) \equiv \Psi^{B'}_h\Psi^B_h
\end{aligned}$$

Similarly, we have

$$\left(\mathbb{E}^*\left[(V^*_{1,i} - V^*_{w,i}\tilde{\gamma}_j)^2\right]\mathbb{E}^*\left[V^*_{w,i}V^{*'}_{w,i}\right]\right)^{-1/2} W^{*'}P_{Z^*}\left(V^*_1 - V^*_w\tilde{\gamma}_j\right) \mid \mathscr{X}_n \xrightarrow{d} \Psi^{B'}_h\psi_{\varepsilon,2SLS}\ a.s.,$$

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

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 **A.2**, we get (conditionally on $\mathscr{X}_n$)

$$\left(\frac{\mathbb{E}^*\left[V^*_{w,i}V^{*\prime}_{w,i}\right]}{\mathbb{E}^*(V^*_{1,i} - V^*_{w,i}\tilde{\gamma}_{LIML})^2}\right)^{1/2}\left(\tilde{\gamma}^*_{LIML} - \tilde{\gamma}_{LIML}\right) \xrightarrow{d} (\Psi_h^{B\prime}\Psi_h^B - \kappa^B_{h,j})^{-1}(\Psi_h^{B\prime}\psi_{\varepsilon,LIML} - \kappa^B_{h,j}h_{\varepsilon V_w,LIML})$$

$a.s.$, where $\kappa^B_{h,j}$ is the smallest root of the characteristic polynomial

$$\left|\left(\psi_{\varepsilon,LIML} : \Psi_h^B\right)'\left(\psi_{\varepsilon,LIML} : \Psi_h^B\right) - \kappa_h\Sigma_{h,LIML}\right| = 0,\ \Sigma_{h,LIML} = \begin{pmatrix} 1 & h_{V_w\varepsilon,LIML} \\ h_{V_w\varepsilon,LIML} & 1 \end{pmatrix}.$$

This establishes Lemma **A.6**-(a).

(b) Similarly, we have

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

Since $\hat{\Omega}^*_{1W} \mid \mathscr{X}_n \xrightarrow{a.s.} \Omega_W$, it follows that

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

(c) Again, we have

$$\begin{aligned}
&\left(Z^{*\prime}Z^*\right)^{-1/2} Z^{*\prime}\left(\tilde{y}^*(\beta_0) : W^*\right)\tilde{r}_j^* \\
&= \left(\frac{Z^{*\prime}Z^*}{n}\right)^{-1/2}\frac{Z^{*\prime}\left(V_1^* - V_w^*\tilde{\gamma}_j\right)}{\sqrt{n}} + \left(\frac{Z^{*\prime}Z^*}{n}\right)^{-1/2}\frac{Z^{*\prime}W^*\left(\tilde{\gamma}_j - \tilde{\gamma}_j^*\right)}{\sqrt{n}} \\
&= \sigma^{1/2}_{\varepsilon\varepsilon,j}\left\{\sigma^{-1/2}_{\varepsilon\varepsilon,j}\left(\frac{Z^{*\prime}Z^*}{n}\right)^{-1/2}\frac{Z^{*\prime}\left(V_1^* - V_w^*\tilde{\gamma}_j\right)}{\sqrt{n}} + \left(\sigma^{-1/2}_{V_wV_w}\left(\frac{Z^{*\prime}Z^*}{n}\right)^{1/2}\Pi^B_{w,n}\right.\right. \\
&\quad \left.\left. +\sigma^{-1/2}_{V_wV_w}\left(\frac{Z^{*\prime}Z^*}{n}\right)^{-1/2}\frac{Z^{*\prime}V_w^*}{\sqrt{n}}\right)\sigma^{-1/2}_{\varepsilon\varepsilon,j}\sigma^{1/2}_{V_wV_w}\left(\tilde{\gamma}_j - \tilde{\gamma}_j^*\right)\right\} + o_p(1).
\end{aligned}$$

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

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

where $\Delta_{h,j}^B=\left(\Psi_h^{B'}\Psi_h^B-\kappa_{h,j}^B\right)^{-1}\left(\Psi_h^{B'}\psi_{\varepsilon,j}-\kappa_{h,j}^B h_{\varepsilon w,j}\right)$. The final result follows from the above limits and (b). □

PROOF OF LEMMA **4.3** Lemma **4.3** follows from equation (4.4) and the results of Lemma **A.6**. Therefore, the proof is omitted. □

PROOF OF THEOREM **6.3** First, we can express $AsySz_{AR_j}[c_j^{B-S}(\alpha,\alpha-\delta,\hat{\mu}_{n,wh})]$ as:

$$\begin{aligned} AsySz_{AR_j}[c_j^{B-S}(\alpha,\alpha-\delta,\hat{\mu}_{n,wh})] &=: \limsup_{n\to\infty}\sup_{\theta\in\Theta}\mathbb{P}_\theta[AR_n(\beta_0,\tilde{\gamma}_j)>c_j^{B-S}(\alpha,\alpha-\delta,\hat{\mu}_{n,wh})] \\ &= \limsup_{n\to\infty}\mathbb{P}_{\tilde{\theta}_n}[AR_n(\beta_0,\tilde{\gamma}_j)>c_j^{B-S}(\alpha,\alpha-\delta,\hat{\mu}_{n,wh})] \quad \text{(A.5)} \\ &= \lim_{n\to\infty}\mathbb{P}_{\tilde{\theta}_{\omega_{n,h}}}[AR_{\omega_{n,h}}(\beta_0,\tilde{\gamma}_j)>c_j^{B-S}(\alpha,\alpha-\delta,\hat{\mu}_{\omega_{n,h},wh})] \end{aligned}$$

where $\{\tilde{\theta}_n: n\geq 1\}$ is a sequence in Θ and $\{\omega_{n,h}: n\geq 1\}$ is a subsequence of $\{n: n\geq 1\}$ satisfying (3.1)–(3.4). Andrews and Guggenberger (2009) show that such sequence and subsequence always exists. So, it suffices to show that $\limsup_{n\to\infty}\lim_{n\to\infty}\mathbb{P}_{\theta_{\omega_{n,h}}}[AR_{\omega_{n,h}}(\beta_0,\tilde{\gamma}_j) > c_j^{B-S}(\alpha,\alpha-\delta,\hat{\mu}_{\omega_{n,h},wh})]\leq\alpha$ for all $\mu_{wh}\in\mathscr{H}_\mu$, subsequence $\{\omega_{n,h}: n\geq 1\}$ of $\{n: n\geq 1\}$ and sequences $\{\theta_{\omega_{n,h}}: n\geq 1\}$ satisfying (3.1)–(3.4). Let $c_j(\delta,\hat{\mu}_{\omega_{n,h},wh})$ be the value of $c_{j,\mu_{wh}}(\delta)$ evaluated at $\mu_{wh}=:\hat{\mu}_{\omega_{n,h},wh}$. Then, for any $\mu_{wh}\in\mathscr{H}_\mu$, we have

$$\begin{aligned} &\limsup_{n\to\infty}\mathbb{P}_{\theta_{\omega_{n,h}}}[AR_{\omega_{n,h}}(\beta_0,\tilde{\gamma}_j)>c_j^{B-S}(\alpha,\alpha-\delta,\hat{\mu}_{n,wh})] \quad \text{(A.6)} \\ =\ &\lim_{n\to\infty}\mathbb{P}_{\tilde{\theta}_{\omega_{n,h}}}[AR_{\omega_{n,h}}(\beta_0,\tilde{\gamma}_j)>c_j^{B-S}(\alpha,\alpha-\delta,\hat{\mu}_{\omega_{n,h},wh})]=\mathbb{P}_\theta[\xi_j(h)\geq c_j^{B-S}(\alpha,\alpha-\delta,\widetilde{\mu}_{wh})] \end{aligned}$$

under H_0, Assumptions **2.1**–**2.3**, **2.5** and Assumptions **6.1**–**6.2**, where $c_j^B(\alpha,\alpha-\delta,\widetilde{\mu}_{wh})$ is a random variable because $\widetilde{\mu}_{wh}=:(\widetilde{\mu}_{ww},h_{\varepsilon w})$ and $\widetilde{\mu}_{ww}=\|h_{ww}+\psi_{V_w}\|$ is a random variable. Now, we have

$$\begin{aligned} \mathbb{P}_\theta[\xi_j(h)\geq c_j^{B-S}(\alpha,\alpha-\delta,\widetilde{\mu}_{wh})] &= \mathbb{P}_\theta[\xi_j(h)\geq c_j^{B-S}(\alpha,\alpha-\delta,\widetilde{\mu}_{wh})\geq c_{j,\mu_{wh}}(\delta)] \\ &+ \mathbb{P}_\theta[\xi_j(h)\geq c_{j,\mu_{wh}}(\delta)\geq c_j^{B-S}(\alpha,\alpha-\delta,\widetilde{\mu}_{wh})] \quad \text{(A.7)} \\ &+ \mathbb{P}_\theta[c_{j,\mu_{wh}}(\delta)\geq\xi_j(h)\geq c_j^{B-S}(\alpha,\alpha-\delta,\widetilde{\mu}_{wh})] \\ &\leq \mathbb{P}_\theta[\xi_j(h)\geq c_{j,\mu_{wh}}(\delta)]+\mathbb{P}_\theta[c_{j,\mu_{wh}}(\delta)\geq c_j^{B-S}(\alpha,\alpha-\delta,\widetilde{\mu}_{wh})] \\ &\leq \mathbb{P}_\theta[\xi_j(h)\geq c_{j,\mu_{wh}}(\delta)]+\mathbb{P}_\theta[\mu_{wh}\notin CI_{\alpha-\delta}(\widetilde{\mu}_{wh})] \\ &\leq \delta+(\alpha-\delta)=\alpha. \end{aligned}$$

It follows from (A.6)–(A.8) that $AsySz_{AR_j}[c_j^{B-S}(\alpha, \alpha-\delta, \hat{\mu}_{n,wh})] \leq \alpha$ as stated.

□

PROOF OF THEOREM **6.6** The proof is similar to those in **6.3** and it is omitted. □

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. Hybrid and size-corrected subsampling methods. Econometrica 77(3), 721–762.

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.

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., 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., 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., 2012. On the asymptotic size distortion of tests when instruments locally violate the exogeneity assumption. Econometric Theory 28, 387–421.

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., 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.

Kleibergen, F., 2015. Efficient size correct subset inference in linear instrumental variables regression. Technical report, Department of Quantitative Economics, University of Amsterdam.

McCloskey, A., 2015. Instrumental varaible regressions with honestly uncertain exclusion restrictions. Technical report, Brown University Providence, RI.

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

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.

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: D. Corbae, S. N. Durlauf, B. E. Hansen, eds, Econometric Theory and Practice: Frontiers of Analysis and Applied Research. Cambridge University Press, Cambridge, U.K., chapter 5.

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