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

Firmin Doko Tchakota       Wenjie Wang

Working Paper No. 2015-01
January, 2015

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

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

January 5, 2015
ABSTRACT

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

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

**JEL classification:** C12; C13; C36.
1. Introduction

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

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

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

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

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

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

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

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

Throughout the paper, $I_q$ stands for the identity matrix of order $q$. For any full-column rank $n \times m$ matrix $A$, $P_A = A(A' A)^{-1} A'$ is the projection matrix on the space of $A$, and $M_A = I_n - P_A$. The notation $\text{vec}(A)$ is the $nm \times 1$ dimensional column vectorization of $A$. $B > 0$ for a squared matrix $B$ means that $B$ is positive definite. Convergence almost surely is symbolized by “a.s.”, “$P$” stands for convergence in probability, while “$d$” 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 $B$, $\partial B$ is the boundary of $B$ and $(\partial B)^\varepsilon$ is the $\varepsilon$-neighborhood of $B$. Finally, $\sup_{\omega \in \Omega} |f(\omega)|$ is the supremum norm on the space of bounded continuous real functions, with topological space $\Omega$.

2. Framework

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

2.1. Model and assumptions

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

\[
y = X\beta + W\gamma + \varepsilon, \quad (2.1)
\]

\[
(X, W) = Z(\Pi_x : \Pi_w) + (V_x, V_w), \quad (2.2)
\]

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

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

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

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

\(^3\)This assumption may also be replaced by another “location assumption,” such as zero medians.
Assumption 2.3 When the sample size \( n \) converges to infinity, the following convergence results hold jointly:

\[ H_0 : \beta = \beta_0 \]  

without any identifying assumptions of model (2.1)-(2.2), where \( \beta_0 \) is a constant vector.

Let \( \tilde{Y} = [y, X, W] = [Y_1, \ldots, Y_n]' \) denote the matrix of endogenous variables, where we define \( \tilde{Y}_i \in \mathbb{R}^3 \) as the \( i \)th row of \( \tilde{Y} \), written as a column vector, and similarly for other model random matrices. Let also \( \mathcal{D}_n = \{ (Y_1', Z_1'), \ldots, (Y_n', Z_n') \} \) and \( R_n = \text{vech} \left[ (\tilde{Y}_n', Z_n') (\tilde{Y}_n', Z_n')' \right] \) \( = (f_1(Y_n', Z_n'), f_2(Y_n', Z_n'), \ldots, f_K(Y_n', Z_n'))' \), where \( f_p(\cdot), p = 1, \ldots, K = \frac{1}{2}(L+2)(L+3) \), are elements of \( (Y_n', Z_n')(Y_n', Z_n')' \) and \( R_n \) has a distribution \( F \).

We make the following generic assumptions on the model variables.

**Assumption 2.1** (i) \( \mathbb{E}(\| R_n \|^2 + r) < \infty \) for some \( r > 0 \), and (ii) \( \limsup_{n \to \infty} \| \mathbb{E}(\exp(it'R_n)) \| < 1 \), where \( i = \sqrt{-1} \).

**Assumption 2.2** \( \mathbb{E}(\langle \varepsilon : V_s : V_w \rangle | Z) = 0. \)

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

\[(i) \quad n^{-1}(\varepsilon : V_s : V_w)'(\varepsilon : V_s : V_w) \overset{D}{\to} \Omega, \quad n^{-1}Z'Z \overset{P}{\to} Q_Z, \quad \text{where} \quad \Omega = \begin{pmatrix} \sigma_{\varepsilon\varepsilon} & \sigma_{\varepsilon\psi} \\ \sigma_{\psi\varepsilon} & \sigma_{\psi\psi} \end{pmatrix}, \quad \Sigma_V = \begin{pmatrix} \sigma_{V\varepsilon} & \sigma_{V\psi} \\ \sigma_{\psi V\varepsilon} & \sigma_{\psi V\psi} \end{pmatrix}, \quad \sigma_{jk} \in \mathbb{R} \text{ for all } j, k \in \{ \varepsilon, V_s, V_w \} ; \]

\[(ii) \quad n^{-1/2}Z'(\varepsilon : V_s : V_w) \overset{D}{\to} (\psi_{Z\varepsilon} : \psi_{ZV_s} : \psi_{ZV_w}), \quad \text{vec} (\psi_{Z\varepsilon} : \psi_{ZV_s} : \psi_{ZV_w}) \sim N(0, \Sigma \otimes Q_Z), \quad \psi_{Z\varepsilon}, \psi_{ZV_s}, \psi_{ZV_w} \text{ are } L \times 1 \text{ random vectors.} \]

Assumption 2.1 is similar to Moreira et al. (2009, Assumptions 2-3) with \( r = s - 2 \) and \( s \geq 3 \). Assumption 2.1-(i) requires that \( R_n \) has second moments or greater while Assumption 2.1-(ii) requires that its characteristic function be bounded above by 1. In particular, the second moments of \( R_n \) exist if \( \mathbb{E}(\| (\tilde{Y}_n', Z_n') \|^2) < \infty \) for some \( r > 0 \). The bound on the characteristic function is the commonly used Cramér’s condition [see Bhattacharya and Ghosh (1978)].
Assumption 2.2 is the usual conditional zero mean assumption of the model errors. The two convergences in Assumption 2.3-(i) are the weak law of large numbers (WLLN) property of \([\varepsilon, V_x, V_w]\) and \(Z\), respectively, while the convergence in Assumption 2.3-(ii) is the central limit theorem (CLT) property.

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

\[
\tilde{Y}(\beta_0) = Z\Pi_w(\gamma: 1) + (v_1, V_w),
\]

(2.4)

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

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

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

\[
n^{-1}(v_1, V_w)'(v_1, V_w) \overset{p}{\rightarrow} \Omega_W, \quad n^{-1}Z'(v_1, V_w) \overset{p}{\rightarrow} 0,
\]

(2.5)

\[
\text{vec} \left((Z'Z/n)^{-1/2}n^{-1/2}Z'(v_1, V_w)\right) \overset{d}{\rightarrow} \text{vec} \left(\psi_{v_1}, \psi_{ZV_w}\right) \sim N(0, \Omega_W \otimes I_L).
\]

(2.6)

We will now present the subset AR statistics studied.

### 2.2. Plug-in subset AR statistics

The plug-in principle usually consists of two steps. The first step is to take an identification-robust statistic [usually the Anderson and Rubin (1949, AR), Kleibergen (2002, K) and Moreira (2003, CLR) type statistics] which results from the test of the joint hypothesis \(H(\beta_0, \gamma_0) : \beta = \beta_0, \gamma = \gamma_0\). The second step consists of replacing the nuisance parameters \(\gamma\) which are not specified by the subset null hypothesis of interest (2.3) by an estimator in the expression of the above statistics. In this paper, we consider the plug-in AR-type statistics that use the restricted LMIL and 2SLS estimators.
of γ under H₀, namely γ_{LIML} and γ_{2SLS}, respectively. For convenience, we adopt the notation γⱼ, j ∈ \{LIML, 2SLS\} in the remainder of the paper. The subset AR statistic for H₀ corresponding to γⱼ can be expressed⁴ as

\[
AR(\beta_0; γ_j) = \frac{1}{L} \|S(\beta_0; γ_j)\|^2,
\]

where \(S(\beta_0; γ_j) = (Z'Z)^{-1/2}Z'Y(\beta_0)\tilde{r}_j(\tilde{r}'_jΩ_W\tilde{r}_j)^{-1/2}Y(\beta_0) = [\gamma(\beta_0): W], \gamma(\beta_0) = Y - X\beta_0, Ω_W = \frac{1}{n-2}\tilde{Y}(β_0)'M_2\tilde{Y}(β_0), \tilde{r}_j = (1, -\tilde{γ}_j)' \). It is well known that γⱼ is given by

\[
\tilde{γ}_j = \left[ W'(P_Z - \tilde{κ}_j M_Z)W \right]^{-1} W'(P_Z - \tilde{κ}_j M_Z)(Y - X\beta_0),
\]

where \(\tilde{κ}_{2SLS} = 0\) and \(\tilde{κ}_{LIML} = (n - L)^{-1} \tilde{κ}_{LIML}, \) \(\tilde{κ}_{LIML}\) is the smallest root of the characteristic polynomial |κΩ_W - \tilde{Y}(β_0)'P_Z\tilde{Y}(β_0)| = 0. Furthermore, we have AR(\β_0; γ_{LIML}) = minγ∈R AR(\β_0; γ) [for example, see Guggenberger et al. (2012)], meaning that AR(\β_0; γ_{2SLS}) ≥ AR(\β_0; γ_{LIML}). This means that the test that uses γ_{2SLS} rejects more often the subset null hypothesis than those that uses γ_{LIML} if the same critical value is applied.

Guggenberger et al. (2012) and Guggenberger and Chen (2011) show that a test of H₀ based on AR(\β_0; γ_{LIML}) has a correct asymptotic size when the errors are homoskedastic even when γ is weakly identified. However, this test is overly conservative even in moderate samples when the identification of γ is weak and that the usual asymptotic \(\chi^2\)-critical values are applied in the inference; see Doko Tchatoka (2014). Moreover, little is known about the size property of the test based on AR(\β_0; γ_{2SLS}), especially when γ is not identified. In this paper, we characterize the asymptotic behavior of both tests, including when γ is not identified, and we investigate whether bootstrapping can improve their size.

In the remainder of the paper, let \(G_{L-1}(\cdot)\) and \(g_{L-1}(\cdot)\) denote the cumulative density function (cdf) and the probability density function (pdf), respectively, of a \(\chi^2\)-distributed random variable with \(L - 1\) degrees of freedom. We now characterize the limiting distribution under H₀ of AR(\β_0; γⱼ) for j ∈ \{LIML, 2SLS\}.

⁴See Stock and Wright (2000); Zivot, Startz and Nelson (2006); Guggenberger et al. (2012); and Doko Tchatoka (2014)
3. Asymptotic behavior of $AR(\beta_0, \tilde{\gamma}_j)$

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

3.1. Strong identification

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

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

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

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

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

---

5For example, see Stock and Wright (2000); Kleibergen (2004); Startz et al. (2006); Mikusheva (2010); Guggenberger et al. (2012); Guggenberger and Chen (2011); and Doko Tchatoka (2014).
We will now study the null limiting behavior of $AR(\beta_0, \gamma_j)$ under Staiger and Stock’s (1997) local to zero weak instruments setup.

### 3.2. Local to zero weak instruments

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

$$
\Sigma_p = \begin{pmatrix}
1 & \rho_{\psi e} \\
\rho_{\psi e} & 1
\end{pmatrix}, 
\Psi = \sigma_{\psi e}^{-1/2} \Omega_2^{1/2} C_{0w} + \psi_{\psi e}, \quad \psi_{\psi e} = \sigma_{\psi e}^{-1/2} \Omega_2^{-1/2} \psi_{\psi e}, 
\psi_{\psi e} = \sigma_{\psi e}^{-1/2} \Omega_2^{-1/2} \psi_{\psi e}.
$$

(3.1)

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

(a) $\tilde{\gamma}_j - \gamma_0 \xrightarrow{d} \sigma_{\psi e}^{-1/2} \Omega_{2\psi e}^{1/2} \Delta_j$, where $\Delta_j = (\Psi' \Psi - \kappa_j)^{-1} (\Psi' \psi_{\psi e} - \kappa_j \rho_{\psi e})$, $\kappa_{2SLS} = 0$ and $\kappa_{LIML}$ is the smallest root of $| (\psi_{\psi e} : \Psi') (\psi_{\psi e} : \Psi) - \kappa \Sigma_p | = 0$;

(b) $\tilde{r}_j \tilde{\Delta}_w \tilde{r}_j \xrightarrow{d} \sigma_{\psi e} \left( 1 - 2 \rho_{\psi e} \Delta_j + \Delta_j^2 \right)$;

(c) $\tilde{S}(\beta_0, \gamma_j) \xrightarrow{d} \left( 1 - 2 \rho_{\psi e} \Delta_j + \Delta_j^2 \right)^{-1/2} S_j$, where $S_j = \psi_{\psi e} - \Psi \Delta_j$.

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

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

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

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

$$AR(\beta_0, \tilde{\gamma}_j) \xrightarrow{d} \xi_{AR_j} = \frac{1}{L} \left(1 - 2\rho_{\psi_{V_w}}\Delta_j + \Delta_j^2\right)^{-1/2} S_j^2$$

for $j \in \{LIML, 2SLS\}$, where $S_j$, $\rho_{\psi_{V_w}}$, and $\Delta_j$ are defined in Lemma 3.2.

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

(ii) Guggenberger et al. (2012, eq.14) and Guggenberger and Chen (2011) provide an upper bound on the limiting distribution of $AR(\beta_0, \tilde{\gamma}_{LIML})$ and show that a test based on this statistic has correct asymptotic size even with weak instruments. However, it performs poorly even with relatively moderate sample sizes [see Doko Tchatoka (2014)]. Meanwhile, it is not clear from Lemma
3.2 and Theorem 3.3 whether this result on the upper of AR($\beta_0, \gamma_{LIML}$) extends to AR($\beta_0, \gamma_{2SLS}$). The Monte Carlo results show that a test based on AR($\beta_0, \gamma_{2SLS}$) sometimes over-rejects the subset null hypothesis when identification is weak and endogeneity is large, while those based on AR($\beta_0, \gamma_{LIML}$) remains overly conservative with weak instruments; see Table 1.

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

4. Bootstrapping subset AR-test

Moreira et al. (2009) show that a residual-based bootstrap yields a test with correct size for the AR statistic of the joint hypothesis $H(\beta_0, \gamma_0): \beta = \beta_0, \gamma = \gamma_0$ in model (2.1)-(2.2), without identifying assumptions on model parameters. However, unlike the AR statistic of $H(\beta_0, \gamma_0)$, the subset AR statistics of $H_0$ in (2.7) are not pivotal (even asymptotically) when $\gamma$ is weakly identified (see Theorem 3.3). So, the validity of the bootstrap for the subset AR statistics of $H_0$ is not obvious, at least when identification is weak, and further investigation is required. In this section, we examine the validity of the bootstrap similar to that of Moreira et al. (2009) for the subset AR statistics in (2.7).

Before proceeding, let $\hat{\Pi}_x = (Z'Z)^{-1}Z'X$ and $\hat{\Pi}_w = (Z'Z)^{-1}Z'W$ denote the ordinary least squares estimators of $\Pi_x$ and $\Pi_w$ in (2.2). Let also $\tilde{\gamma}_j, j \in \{LIML, 2SLS\}$, denote the restricted LIML and 2SLS estimators of $\gamma$ under $H_0$ given in (2.8). We adapt the bootstrap procedure by Moreira et al. (2009) to the subset AR statistics as follows.

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

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

$$X^* = Z^* \hat{\Pi}_x + V^*_x,$$

(4.1)
\[ W^* = Z^* \hat{\Pi}_w + V^*_w, \]  
\[ y^* = X^* \beta_0 + Z^* \hat{\Pi}_w \tilde{\gamma}_j + v_1^*, \]  
(4.2)  
(4.3)

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

\[ AR^{(b)}(\beta_0, \tilde{\gamma}_j) = \frac{1}{L} \| \hat{S}^{(b)}(\beta_0, \tilde{\gamma}_j) \|^2, \]  
\[ \hat{S}^{(b)}(\beta_0, \tilde{\gamma}_j) = (Z^r Z^r)^{-1/2} Z^r \hat{Y}^r(\beta_0) \tilde{r}_j^r (\hat{\Omega}_1^r \hat{r}_j^r)^{-1/2}, \]  
(4.4)  
(4.5)

where \(\hat{\Omega}^r(\beta_0) = (\hat{s}^r (\beta_0) : W^*)\) and \(\tilde{r}_j^r = (1, -\tilde{\gamma}_j)^t\).

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

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

### 4.1. Bootstrap consistency under strong identification

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

**Lemma 4.1** Suppose that Assumptions 2.1 - 2.3 are satisfied with \(r \geq 1\). Suppose further that \(H_0\)
holds and \( \Pi_w \neq 0 \) is fixed. Then we have:

\[
\sup_{\tau \in \mathbb{R}} |G_{\text{AR}}^j(\tau) - G_{L-1}(\tau) - \sum_{h=1}^{r} n^{-h} p_{\text{AR}}^h(\tau; F_n, \beta_0, \hat{\gamma}_j, \hat{N}_x, \hat{N}_w)g_{L-1}(\tau) | = o(n^{-r}) \text{ a.s.}
\]

for \( j \in \{\text{LIML}, 2\text{SLS}\} \), \( p_{\text{AR}}^h \) is a polynomial in \( \tau \) with coefficients depending on \( \beta_0, \gamma_j, \hat{N}_x, \hat{N}_w \) and the moments of \( F_n \), and \( G_{\text{AR}}^j(\tau) = \mathbb{P}^n[\text{AR}^*(\beta_0, \hat{\gamma}_j) \leq \tau] \) is the empirical cdf of \( \text{AR}^*(\beta_0, \hat{\gamma}_j) \) evaluated at \( \tau \in \mathbb{R} \).

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

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

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

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

where \( c_{\text{AR}}^*(\alpha) \) is the \( 1 - \alpha \) quantile of the empirical distribution of \( \text{AR}^*(\beta_0, \hat{\gamma}_j) \), i.e., the value that minimizes \( |\mathbb{P}^n[\text{AR}^*(\beta_0, \hat{\gamma}_j) \leq \tau] - (1 - \alpha)| \) over \( \tau \in \mathbb{R} \).

Remarks. (i) Theorem 4.2 gives the conditions under which the bootstrap critical values for \( \text{AR}^*(\beta_0, \hat{\gamma}_j) \) yield level for the \( \text{AR}(\beta_0, \hat{\gamma}_j) \) test that is correct through \( O(n^{-1}) \) under \( H_0 \). So, the bootstrap makes an error of size \( O(n^{-1}) \) under \( H_0 \) when \( \gamma \) is identified (\( \Pi_w \neq 0 \) is fixed) for all values of \( \Pi_x \). So, the identification of the parameters specified by the subset null hypothesis (here \( \beta \)) does not affect the results of Theorem 4.2.
(ii) Although Theorem 4.2 focuses on the validity of the bootstrap under $H_0$, there is no impediment to expanding the above results under the alternative hypothesis ($\beta \neq \beta_0$). In this case, we can show that the tests are consistent when the bootstrap critical values are used in the inference and the instruments are strong for both $X$ and $W$ (i.e., if $\Pi_x \neq 0$ and $\Pi_w \neq 0$ are fixed). This is because the asymptotic distribution of $AR(\beta_0, \gamma_j)$ diverges when $\beta \neq \beta_0$, and $\Pi_x \neq 0, \Pi_w \neq 0$ are fixed. Note that test consistency will still hold in this case if the empirical size-corrected critical values were used instead of the bootstrap critical values, as suggested by Horowitz (1994). However, the tests will have low power for weak values of $\Pi_x$ even if $\gamma$ is identified ($\Pi_w \neq 0$ is fixed). These proofs are omitted to shorten the exposition of our results.

We shall now focus on Staiger and Stock’s (1997) _local to zero_ weak instruments framework.

### 4.2. Bootstrap inconsistency under weak instruments

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

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

(a) $\bar{\gamma}_j - \gamma_j \mid \mathcal{F}_n \overset{d}{\to} \sigma_{\gamma_j}^{1/2} \sigma_{\gamma_j}^{-1/2} \Delta_j^B$ a.s.;

(b) $\bar{\gamma}_j, \bar{\gamma}_j, \bar{\Omega}_w^* p_j \mid \mathcal{F}_n \overset{d}{\to} \sigma_{\gamma_j} \left(1 - 2\rho_{\gamma_j} \Delta_j^B + (\Delta_j^B)^2\right)^{1/2}$ a.s.;

(c) $\bar{S}^*(\beta_0, \gamma_j) \mid \mathcal{F}_n \overset{d}{\to} \left(1 - 2\rho_{\gamma_j} \Delta_j^B + (\Delta_j^B)^2\right)^{-1/2}$ $S_j^B$ a.s.

for $j \in \{LIML, 2SLS\}$ where $\Delta_j^B = \left(\psi_j \psi_j - \kappa_j^B\right)^{-1} \left(\psi_j \psi_j - \kappa_j^B \rho_{\gamma_j}\right)$, when $\kappa_j^B = 0$ and $\sigma_{\gamma_j}^2 \neq 0$. $\psi_j$ is the smallest root of the determinantal equation

$\left| \left(\psi_j, \psi_j, \psi_j\right)^\prime \left(\psi_j, \psi_j, \psi_j\right) - \kappa \Sigma_{\gamma_j}\right| = 0,$

$\psi_j = \psi + \psi_{\gamma}, \quad \psi_{\gamma} = \psi V - \psi V \left(\gamma_j + \sigma_{\gamma_j} \sigma_{\gamma_j}^{-1/2} \Delta_j V\right) = \psi V - \sigma_{\gamma_j} \sigma_{\gamma_j}^{-1/2} \Delta_j V,$

$\Sigma_{\gamma_j} = \left(\begin{array}{c} 1 \\ \rho_{\gamma_{\gamma_j}} \end{array}\right), \quad \rho_{\gamma_{\gamma_j}} = \left(\sigma_{\gamma_j} \sigma_{\gamma_j}^{-1/2} \sigma_{\gamma_j}^{-1/2}\right), \quad \sigma_{\gamma_j} = \sigma_{\gamma_j} \left(1 - 2\rho_{\gamma_{\gamma_j}} \Delta_j + (\Delta_j)^2\right),$
\( \sigma_{\nu^{-1}} - \sigma_{\nu^{-1/2}} \Delta_j, \quad S_j = \psi_{\nu^{-1/2}} \Delta_j, \) and \( \Delta \) is given in Lemma 3.2.

**Remarks.** (i) First, we note that the convergence results in Lemma 4.3 (a)-(c) differ from those in Lemma 3.2 (a)-(c). This means that the bootstrap fails to mimic the asymptotic distributions of all statistics \( \tilde{\gamma}_j \) and \( \tilde{\gamma}_j \) under \( H_0 \) when \( \gamma_j \) is used as the pseudo true value of \( \gamma \). Indeed, while \( \gamma_j - \gamma_0 \overset{d}{\to} \sigma_{\nu^{-1/2}} \Delta_j \) in Lemma 3.2 (a), we have \( \gamma_j - \gamma_0 \overset{d}{\to} \sigma_{\nu^{-1/2}} \Delta_j \) but \( \gamma_j \) in Lemma 4.3. Hence, the bootstrap fails to mimic the limiting behavior of \( \gamma_j \) with probability 1.

(ii) The bootstrap failure is mainly due to the fact that replacing \( \pi \) and \( \gamma_j \) by \( \tilde{\pi} \) and \( \tilde{\gamma}_j \), respectively, in (2.4) generates an extra noise term that is added to the original residuals when re-sampled. This can lead to substantial differences between the limits in the original sample and those in the bootstrap sample. To be more explicit, observe that in the local to zero weak instruments setup, we have \( \tilde{\gamma}_j - \tilde{\gamma}_0 \overset{d}{\to} \sigma_{\nu^{-1/2}} \Delta_j \) a.s. by Lemma 4.3 (a). Both limits can differ significantly since \( \Delta_j = (\psi^\gamma - \kappa_j)^{-1} (\psi^\gamma \psi_\nu - \kappa_j \rho_\nu) \neq \sigma_{\nu^{-1/2}} \Delta_j \) and \( \sigma_{\nu^{-1/2}} \neq \sigma_{\nu^{-1/2}} \) with probability 1.

We can prove the following theorem on the limiting distribution of \( AR^*(\tilde{\beta}_0, \tilde{\gamma}_j) \) under \( H_0 \) and weak instruments.

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

\[
AR^*(\tilde{\beta}_0, \tilde{\gamma}_j) \mid \gamma_j \overset{d}{\to} \xi_{AR^*} = \left(1 - 2\rho_\nu \Delta_{\nu^{-1/2}} + (\Delta_{\nu^{-1/2}})^2\right)^{-1/2} S_j^{\nu^{-1/2}} \quad \text{a.s.; } j \in \{LIML, 2SLS\}.
\]

**Remark.** Theorem 4.4 provides a characterization of the limiting distribution of the bootstrap subset AR statistics under \( H_0 \) similar to that in Theorem 3.3. As can be seen, a straightforward comparison confirms our previous analysis in Lemma 4.3, i.e., \( AR^*(\tilde{\beta}_0, \tilde{\gamma}_j) \) and \( AR(\tilde{\beta}_0, \tilde{\gamma}_j) \) have
different asymptotic behavior under $H_0$. So, we can state the following theorem on the inconsistency of the bootstrap for subset AR tests under weak instruments.

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

$$\left| \mathbb{P}[\text{AR}(\beta_0, \gamma_j) > c^*_\text{AR}_j] - \alpha \right| = O_p(1)$$

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

where $c^*_\text{AR}_j = c^*_\text{AR}(\alpha)$ is the $1 - \alpha$ quantile of the empirical distribution of $\text{AR}^*(\beta_0, \gamma_j)$.

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

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

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

**5. Monte Carlo experiment**

We use simulation to examine the performance of both the standard and bootstrap subset AR tests described above. The data generating process is described by (2.1) and (2.2) where $\gamma, X$ and $W$ are $n \times 1$ vectors. The $n$ rows of $[\epsilon : V_z : V_w]$ are i.i.d. normal with zero mean, and covariance matrix

$$
\Sigma = \begin{pmatrix}
1 & \rho & \rho \\
\rho & 1 & \rho \\
\rho & \rho & 1
\end{pmatrix},
$$

where the endogeneity $\rho$ varies in $\{0, 0.1, 0.5, 0.9\}$. The $L$ columns of the instrument matrix $Z$, $L \in \{3, 5, 10, 20\}$, are $N(0, I_L)$, independently from $[\epsilon : V_z : V_w]$. Note that $Z$ is fixed over the simulation experiment but the results do not change qualitatively otherwise. The
true values of $\beta$ and $\gamma$ are set at 2 and $-1$, respectively. The reduced-form coefficient matrix $\Pi_{xw} = [\Pi_x : \Pi_w]$ is chosen as $\Pi_{xw} = (\frac{\mu^2}{m|ZC_0|})^{1/2} \Pi_0$, where $\Pi_0$ is obtained by taking the first 2 columns of an identity matrix of order $L$, and $\mu^2$ is the concentration parameter which describes the strength of $Z$. In this experiment, we vary $\mu^2$ in $\{0, 13, 1000\}$, where $\mu^2 = 0$ is a complete non-identification or irrelevant IVs setup, $\mu^2 = 13$ represents weak instruments (both $\beta$ and $\gamma$ are weakly identified), and $\mu^2 = 1000$ implies strong instruments; see Guggenberger (2010) and Doko Tchatoka (2014) for a similar parametrization. The rejection frequencies are computed using 1,000 replications for the standard subset AR tests, while those of the bootstrap subset AR tests are obtained with $N = 1,000$ replications and $B = 299$ bootstrap pseudo-samples of size $n = 100$. The nominal level of all tests is set at $5\%$.

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

First, we note that when the restricted LIML estimator is used as a plug-in estimator and the usual asymptotic $\chi^2$ critical values are applied, the resulting subset AR test is overly conservative with weak instruments (see columns $\mu^2 \in \{0, 13\}$ in the table). These results are similar to those in Doko Tchatoka (2014). However, the test has rejections close to the nominal $5\%$ level when identification is strong (see columns $\mu^2 = 1000$ in the table), thus confirming our theoretical findings in Section 3.1. Note however that the rejection frequencies of this test are slightly greater than the nominal $5\%$ level as both endogeneity ($\rho$) and the number of instruments ($L$) increase (for example, with $\rho = 0.9$ and $L = 20$, the rejection frequency is around $7\%$). Meanwhile, the subset AR test that uses the restricted 2SLS estimator as the plug-in based estimator is less conservative than those with LIML when instruments are weak and the asymptotic $\chi^2$ critical values are applied (see columns $\mu^2 \in \{0, 13\}$ in the second block of Table 1). This is not surprising because we always have $AR(\beta_0, \hat{Y}_{2SLS}) \geq AR(\beta_0, \hat{Y}_{LIML})$. In particular, the rejection frequencies of this test are close to the nominal $5\%$ level with weak instruments and small endogeneity (see columns $\rho \in$...
\{0,0.1\} and \(\mu^2 \in \{0,13\} \) in the table), but they are greater than 5% for large endogeneity (see column \(\rho = 0.9\) and \(\mu^2 \in \{0,13\}\)). We also observe that this test over-rejects sometimes when identification is strong. For example, the rejection frequencies when \(\mu^2 = 1000\) (strong instruments) are 9.3%, 16.5%, and 40.7% for \(L = 5, 10, 20\) instruments, respectively. So, while the subset AR test that uses the restricted 2SLS estimator seems to outperform the one that uses LIML under weak instruments and small endogeneity, its finite-sample size property is worse than the test with LIML when identification is strong and endogeneity is large.

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

6. Conclusions

In this paper, we focus on linear IV regressions and study the asymptotic validity of the bootstrap for the plug-in based subset AR type-tests. Specifically, we suggest a bootstrap method similar to those of Moreira et al. (2009) for the score test of the null hypothesis specified on the full set of the structural parameters. We stress the fact that testing subset hypotheses is substantially more complex than testing joint hypotheses on the full set of parameters, so, the validity of Moreira et al.’s (2009) bootstrap for tests of subvectors is not obvious, especially when identification is weak. Our analysis considers two plug-in subset AR statistics. The first one uses the restricted limited information maximum likelihood (LIML) as the plug-in estimator of the nuisance parameters, and
Table 1. Rejection frequencies (in %) at 5% nominal level

<table>
<thead>
<tr>
<th>Statistics</th>
<th>L ↓ μ² →</th>
<th>( \rho = 0 )</th>
<th>( \rho = 0.1 )</th>
<th>( \rho = 0.5 )</th>
<th>( \rho = 0.9 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>13</td>
<td>1000</td>
<td>0</td>
</tr>
<tr>
<td>Asymptotic χ² critical values</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restricted LIML</td>
<td>AR 3</td>
<td>0.4</td>
<td>0.9</td>
<td>5.2</td>
<td>0.3</td>
</tr>
<tr>
<td>- 5</td>
<td>0.3</td>
<td>0.2</td>
<td>5.4</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>- 10</td>
<td>0.3</td>
<td>0.3</td>
<td>6.2</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>- 20</td>
<td>0.4</td>
<td>0.2</td>
<td>4.3</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>Restricted 2SLS</td>
<td>AR 3</td>
<td>3.0</td>
<td>3.5</td>
<td>6.4</td>
<td>2.3</td>
</tr>
<tr>
<td>- 5</td>
<td>3.5</td>
<td>3.4</td>
<td>6.8</td>
<td>3.4</td>
<td>2.8</td>
</tr>
<tr>
<td>- 10</td>
<td>3.8</td>
<td>5.5</td>
<td>5.3</td>
<td>4.9</td>
<td>3.4</td>
</tr>
<tr>
<td>- 20</td>
<td>6.5</td>
<td>6.3</td>
<td>7.3</td>
<td>6.3</td>
<td>6.9</td>
</tr>
<tr>
<td>Bootstrap critical values</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restricted LIML</td>
<td>AR 3</td>
<td>0.7</td>
<td>1.2</td>
<td>5.8</td>
<td>0.4</td>
</tr>
<tr>
<td>- 5</td>
<td>0.2</td>
<td>0.4</td>
<td>5.4</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>- 10</td>
<td>0.3</td>
<td>0.1</td>
<td>5.1</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>- 20</td>
<td>0.2</td>
<td>0.1</td>
<td>3.2</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Restricted 2SLS</td>
<td>AR 3</td>
<td>4.9</td>
<td>5.9</td>
<td>6.8</td>
<td>3.2</td>
</tr>
<tr>
<td>- 5</td>
<td>2.2</td>
<td>3.2</td>
<td>6.7</td>
<td>4.0</td>
<td>2.3</td>
</tr>
<tr>
<td>- 10</td>
<td>1.6</td>
<td>6.7</td>
<td>4.4</td>
<td>1.7</td>
<td>3.7</td>
</tr>
<tr>
<td>- 20</td>
<td>4.5</td>
<td>3.2</td>
<td>5.7</td>
<td>4.4</td>
<td>3.3</td>
</tr>
</tbody>
</table>
the second uses the restricted two-stage least squares (2SLS) as the plug-in method.

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

A. Appendix

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

A.1. Subsampling inconsistency under weak instruments

In this section, we establish the invalidity of subsampling for AR$(\hat{\beta}_0, \tilde{\gamma}_j)$, $j \in \{LIML, 2SLS\}$, under Staiger and Stock’s (1997) local to zero weak instruments asymptotic.

Let $b$ be the subsample size, $G_{AR_j}^b(\tau)$ the empirical distribution function of the subsample statistics $\{AR_j^b(\beta_0, \tilde{\gamma}_{j,b}) : l = 1, ..., q_n\}$, evaluated at $\tau \in \mathbb{R}$, and $c_{AR_j}^b(\alpha) = c_{AR_j}^b(\alpha)$ the $1 - \alpha$ sample quantile of $G_{AR_j}^b(\cdot)$, where $q_n$ is the number of different subsamples of size $b$. With i.i.d. observations, there are $q_n = n!/(n - b)!b!$ different subsamples of size $b$. With time series observations, there are
Lemma A.1 Suppose Assumptions 2.2-2.3 are satisfied and $H_0$ holds. Let $\Pi_n = C_{0n}/\sqrt{n}$, where $C_{0n} \in \mathbb{R}^L$ is fixed. If $b \to \infty$, and $b/n \to 0$ as $n \to \infty$, then the following convergence holds jointly for $j \in \{\text{LIML, 2SLS}\}$:

(a) $\gamma_{j,b} - \gamma_0 \mid \mathcal{F}_n \overset{d}{\to} \sigma_{\varepsilon^2}^{-1/2} \sigma_{\varepsilon^{2n}}^{-1/2} \Delta_j^S a.s.;$

(b) $p'_{j,b} \Omega \hat{\omega}_{\omega,b} \hat{r}_{j,b} \mid \mathcal{F}_n \overset{d}{\to} \sigma_{\varepsilon^{2e}} \left(1 - 2 \rho_{\varepsilon^{2e}} \Delta_j^S + (\Delta_j^S)^2 \right) a.s.;$

(c) $S^b_j(\beta_0, \gamma_{j,b}) \mid \mathcal{F}_n \overset{d}{\to} \left(1 - 2 \rho_{\varepsilon^{2e}} \Delta_j^S + (\Delta_j^S)^2 \right)^{-1/2} S_j^S a.s.,$

$q_n = n - b + 1$ subsamples each consisting of $b$ consecutive observations. We assume that $b > L$.

The subsampled AR statistic, $AR_{b,j}(\beta_0, \gamma_{j,b})$, is defined as

$$AR_{b,j}^b(\beta_0, \gamma_{j,b}) = \frac{1}{L} \| S^b_j(\beta_0, \gamma_{j,b}) \|^2,$$

where

$$S^b_j(\beta_0, \gamma_{j,b}) = \left( Z'_b Z_b \right)^{-1/2} Z'_b \hat{Y}_{\omega,b}(\beta_0) \hat{F}_{j,b} \left( \hat{F}_{j,b} \hat{\omega}_{\omega,b} \hat{F}_{j,b} \right)^{-1/2}, \quad \hat{\omega}_{\omega,b} = \frac{1}{n} \hat{Y}'_b(\beta_0) M_{Z,b} \hat{Y}'_b(\beta_0), \quad \hat{Y}'_b(\beta_0) = [y_b - X_b \beta_0 : W_b], \quad \hat{F}_{j,b} = (1, -\gamma_{j,b})', \quad j \in \{2SLS, LIML\},$$

$$\gamma_{j,b} = [W'_b (P_s - \kappa_{j,b} M_{s}) W_b]^{-1} W'_b (P_s - \kappa_{j,b} M_{s}) (y_b - X_b \beta_0)$$

with $\kappa_{2SLS,b} = 0$ and $\kappa_{LIML,b} = (n - L)^{-1} \kappa_{LIML,b}$ where $\kappa_{LIML,b}$ is the smallest root of the characteristic polynomial $|\kappa \hat{\omega}_{\omega,b} - \hat{Y}_{\omega,b}(\beta_0)' P_{s} \hat{Y}_{\omega,b}(\beta_0)| = 0$. As in the bootstrap case, the subsampling subset AR test based on $AR_{b,j}^b(\beta_0, \gamma_{j,b})$ rejects $H_0$ if

$$AR_{b,j}^b(\beta_0, \gamma_{j,b}) > c_{j,b}(1 - \alpha).$$

In what follows, $\mathbb{P}^b$ denotes the probability under the subsampling empirical distribution function conditional on $\mathcal{F}_n = \{(\hat{Y}'_1, Z'_1), \ldots, (\hat{Y}'_n, Z'_n)\}$ and $\mathbb{E}^b$ its corresponding expectation operator.

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

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

(a) $\gamma_{j,b} - \gamma_0 \mid \mathcal{F}_n \overset{d}{\to} \sigma_{\varepsilon^2}^{-1/2} \sigma_{\varepsilon^{2n}}^{-1/2} \Delta_j^S a.s.;$

(b) $p'_{j,b} \hat{\omega}_{\omega,b} \hat{r}_{j,b} \mid \mathcal{F}_n \overset{d}{\to} \sigma_{\varepsilon^{2e}} \left(1 - 2 \rho_{\varepsilon^{2e}} \Delta_j^S + (\Delta_j^S)^2 \right) a.s.;$

(c) $S^b_j(\beta_0, \gamma_{j,b}) \mid \mathcal{F}_n \overset{d}{\to} \left(1 - 2 \rho_{\varepsilon^{2e}} \Delta_j^S + (\Delta_j^S)^2 \right)^{-1/2} S_j^S a.s.,$

21
where $\Delta^S_j = (\Psi'_{V_s} \Psi_{V_s} - \kappa^S_j)^{-1} (\Psi'_{\tilde{\epsilon}} \Psi_{\tilde{\epsilon}} - \kappa^S_j \rho_{V_{\tilde{\epsilon}}})$, $\Psi^S_j = \Psi_{V_s}^T$, $\kappa^S_{2SLS} = 0$ and $\kappa^S_{LIML}$ is the smallest root of the determinantal equation $\left| (\Psi_{\tilde{\epsilon}} : \Psi_{V_s})' (\Psi_{\tilde{\epsilon}} : \Psi_{V_s}) - \kappa \Sigma \right| = 0$; $S^j = \Psi_{\tilde{\epsilon}} - \Psi_{V_s} \Delta^S_j$.

**Remark.** As for the case of bootstrap [see Lemma A.3 (a)-(c)], the convergence results in Lemma A.1 (a)-(c) differ from that of Lemma 3.2 (a)-(c). So, the subsampling fails to mimic the asymptotic distributions of the statistics $\tilde{r}_j' \hat{\Omega}_{1W} \tilde{r}_j$ and $\hat{S}(\beta_0, \tilde{\gamma}_j)$ under $H_0$ and weak instruments.

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

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

$$AR_{b,n} (\beta_0, \tilde{\gamma}_j) \mid \mathcal{N} \xrightarrow{d} \left( 1 - 2 \rho_{\epsilon_{V_s} \epsilon'} + (\Delta^f_j)^2 \right)^{-1/2} S^j \right) \right]^{1/2} \right)$$

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

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

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

1. $\hat{\gamma}_{j,b} - \gamma_0 \mid \mathcal{N} \xrightarrow{d} \sigma_{\epsilon_{V_s} \epsilon}^{-1/2} \sigma_{\epsilon_{V_s} \epsilon} \epsilon' \Delta^f_j \text{ a.s.}$;

2. $\tilde{r}_{j,b} \hat{\Omega}_{1W} \tilde{r}_{j,b} \mid \mathcal{N} \xrightarrow{d} \sigma_{\epsilon_{V_s} \epsilon} \left( 1 - 2 \rho_{\epsilon_{V_s} \epsilon} + (\Delta^f_j)^2 \right) \text{ a.s.}$;
(c) \( S^a_j(\beta_0, \gamma_{j,b}) \mid \mathcal{F}_n \xrightarrow{d} \left( 1 - 2\rho_{w,e} \Delta_j^f + (\Delta_j^f)^2 \right)^{-1/2} S_j^f \text{ a.s.} \),

where \( \Delta_j^f = (\Psi^f \Psi^f - \bar{\kappa}_j^f)^{-1}(\Psi^f \psi_e - \kappa_j^f \rho_{w,e}) \), \( \Psi^f = f \sigma_{V_w}^{-1/2} Q_{Z}^{1/2} C_{0w} + \psi_{V_w} = f \Psi^f + (1 - f) \psi_{V_w} \), \( \kappa_j^{2SLS} = 0 \) and \( \kappa_{jLIML} \) is the smallest root of the determinantal equation \( \left( \psi_e : \Psi^f \right) \left( \psi_e : \Psi^f \right) - \kappa \Sigma_p = 0; S_j^f = \psi_e - \Psi^f \Delta_j^f \).

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

\[
AR_{b,j}(\beta_0, \gamma_{j,b}) \mid \mathcal{F}_n \xrightarrow{d} \left( 1 - 2\rho_{w,e} \Delta_j^f + (\Delta_j^f)^2 \right)^{-1/2} S_j^f \text{ a.s. for all } j \in \{LIML, 2SLS\}.
\]

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

### A.2. Supplemental lemmata

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

\[
\sup_{\tau \in \mathbb{R}} \mathbb{E}[S(\beta_0, \gamma_{j}) \leq \tau] - \Phi(\tau) - \sum_{h=1}^{r} n^{-h/2} p_{j}^{h}(\tau; F, \beta_0, \gamma, \Pi_{c}, \Pi_{w}) \phi(\tau) = o(n^{-r/2})
\]

for all \( j \in \{LIML, 2SLS\} \), where \( \Phi(\cdot) \) and \( \phi(\cdot) \) are the cdf and pdf of a standard normal random variable, \( p_{j}^{h} \) are polynomials in \( \tau \) with coefficients depending on \( \beta_0, \gamma, \Pi_{c}, \Pi_{w} \) and the moments of the distribution \( F \) of \( \mathcal{R}_n \).

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

\[
\sup_{\tau \in \mathbb{R}} \mathbb{E}[S^{*}(\beta_0, \gamma_{j}) \leq \tau] - \Phi(\tau) - \sum_{h=1}^{r} n^{-h/2} p_{j}^{h}(\tau; F_n, \beta_0, \gamma_{j}, \Pi_{c}, \Pi_{w}) \phi(\tau) = o(n^{-r/2})
\]

for all \( j \in \{LIML, 2SLS\} \), where \( \Phi(\cdot) \) and \( \phi(\cdot) \) are the cdf and pdf of a standard normal random variable, \( p_{j}^{h} \) are polynomials in \( \tau \) with coefficients depending on \( \beta_0, \gamma_{j}, \Pi_{c}, \Pi_{w} \) and the moments of the distribution \( F_n \) of \( \mathcal{R}^*_n \).
Lemma A.7 Suppose Assumptions 2.2-2.3 are satisfied and $H_0$ holds. If $\Pi_w = C_{0w}/\sqrt{n}$, where $C_{0w} \in \mathbb{R}^L$ is fixed, then for $j \in \{2SLS, LIML\}$ we have:

(a) $E^s \left[ (v^e_{1,j} - V_{w,i}^e \bar{\nu}_j) \right] \overset{d}{\to} \sigma_{ee,j} = \sigma_{ee} \left( 1 - 2\rho_{v_e} \Delta_j + (\Delta_j)^2 \right)$ a.s.;

(b) $E^s \left[ V_{w,i}^e \left( v^e_{1,j} - V_{w,i}^e \bar{\nu}_j \right) \right] \overset{d}{\to} \sigma_{v_e} \sigma_{v_e} = \sigma_{ee} - \sigma_{ee}^1 \sigma_{v_e} \Delta_j$ a.s.;

(c) $E^s \left[ V_{w,i}^e V_{w,i}^e \right] \overset{ps}{\to} \sigma_{v_i \nu_i}$ a.s.

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

$$
\left( \frac{Z_i^e Z_i^e}{n} \right)^{-1/2} \left( \frac{Z_i^e Z_i^e}{\sqrt{n}} \right) \left( \frac{\hat{\Omega}_w}{\sqrt{n}} \right)^{-1/2} \left( n^{-1/2} \left( n^{-1} M^* sL \Sigma_n - n^{-1} M^* 1_n \right) \right) \overset{d}{\to} N \left( \left( \frac{I_{2L}}{\sqrt{n}} \right) \left( \begin{array}{c} 0 \\ 0 \\ \Omega_{mm} \end{array} \right) \right) \text{a.s.,}
$$

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

A.3. Proofs

Proof of Lemma A.5 To shorten our exposition, we will focus on the proof for $S(\hat{\beta}_0, \bar{\gamma}_{2SLS})$.

The proof for $S(\hat{\beta}_0, \bar{\gamma}_{LIML})$ can be deduced easily by following similar steps to those presented here. First, observe that we can write $S(\hat{\beta}_0, \bar{\gamma}_{2SLS}) = \sqrt{n}/D$ under the subset null hypothesis $H_0$, where

$$
N = \left( \frac{Z_i^e Z_i^e}{n} \right)^{-1/2} \left[ \frac{Z_i^e \bar{\beta}_0}{\sqrt{n}} - \frac{Z_i^e \bar{\beta}_0}{\sqrt{n}} \right] \bar{\gamma}_{2SLS}
$$

$$
= \left( \frac{Z_i^e Z_i^e}{n} \right)^{-1/2} \left[ \frac{Z_i^e Z_i^e}{n} \left( \frac{W_i^e Z_i^e}{n} \left( \frac{Z_i^e Z_i^e}{n} \right)^{-1} \left( \frac{W_i^e Z_i^e}{n} \right)^{-1} \right) \right].
$$

(A.4)

$$
D = \frac{\bar{\gamma} (\bar{\beta}_0) MZ \Sigma (\bar{\beta}_0)}{n-L} - 2\bar{\gamma} (\bar{\beta}_0) MZ + \frac{W_i^e MZ}{n-L} \bar{\gamma}_{2SLS}^2.
$$

(A.5)

$$
\bar{\gamma}_{2SLS} = \left[ \frac{W_i^e Z_i^e}{n} \right] \left( \frac{Z_i^e Z_i^e}{n} \right)^{-1} \left( \frac{W_i^e Z_i^e}{n} \right)^{-1} \left( \frac{Z_i^e Z_i^e}{n} \right)^{-1} \left( \frac{Z_i^e Z_i^e}{n} \right), \text{and} \quad \bar{\gamma} (\bar{\beta}_0) = y - X \beta_0.
$$

So, under $H_0$ and from (A.4)-(A.5), we can write $S(\hat{\beta}_0, \bar{\gamma}_{2SLS})$ as

$$
\bar{S}(\hat{\beta}_0, \bar{\gamma}_{2SLS}) = \sqrt{n} H(\bar{\Sigma}_n) = \sqrt{n} H(\bar{\Sigma}_n - H(\mu))
$$

(A.6)
where \( H(\cdot) \) is a real-valued Borel measurable function on \( \mathbb{R}^K \) with derivatives of order \( s \geq 3 \) and lower, being continuous on the neighborhood of \( \mu = \mathbb{E}(R_n) \) when \( \Pi_w \neq 0 \) is fixed, and \( H(\mu) = 0 \) under \( H_0 \). Note that the derivatives of order \( s \geq 3 \) and lower of \( H(\cdot) \) are not well-defined when \( \Pi_w = 0 \) and does not even exist if \( \Pi_w = C_0 n c_n \) for any sequence \( c_n \downarrow 0 \) [similar to Moreira et al. (2009, footnote 2) and Doko Tchatoka (2013)]. The result in Lemma A.5 follows by applying Bhattacharya and Ghosh (1978, Theorem 2) to (A.6) with \( s = 2 = r \).

**Proof of Theorem 3.1** From (2.7), we have \( L \times AR(\beta_0, \gamma_j) = \|S(\beta_0, \gamma_j)\|^2 \). We want to approximate \( \mathbb{P}[AR(\beta_0, \gamma_j) \leq \tau] \) uniformly in \( \tau \) under \( H_0 \). First, we can write \( \mathbb{P}[AR(\beta_0, \gamma_j) \leq \tau] \) as:

\[
\mathbb{P}[AR(\beta_0, \gamma_j) \leq \tau] = \mathbb{P}[AR(\beta_0, \gamma_j) \in \mathcal{C}_\tau],
\]

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

**Proof of Lemma 3.2** (a) We begin with the result for \( \tilde{\gamma}_{2SLS} \). We have

\[
(\sigma_{v_{w,v_{w}}})^{-1} W'P_2W = (\sigma_{v_{w,v_{w}}})^{-1} (\Pi_w'Z'Z\Pi_w + V_w'Z\Pi_w + \Pi_w'Z'V_w + V_w'P_2V_w)
\]

\[
= (\sigma_{v_{w,v_{w}}})^{-1} (\frac{Z'Z}{n}) (\sigma_{v_{w,v_{w}}})^{-1} (\frac{Z'V_w}{\sqrt{n}}) + (\sigma_{v_{w,v_{w}}})^{-1} (\frac{Z'V_w}{\sqrt{n}}) (\sigma_{v_{w,v_{w}}})^{-1} (\frac{Z'V_w}{\sqrt{n}})
\]

\[
\overset{d}{\rightarrow} C_0^\prime \sigma_{v_{w,v_{w}}}^{-1} Q_Z \sigma_{v_{w,v_{w}}}^{-1} C_0 + \psi_{V_w} \sigma_{v_{w,v_{w}}}^{-1} Q_Z \sigma_{v_{w,v_{w}}}^{-1} C_0 + C_0^\prime \sigma_{v_{w,v_{w}}}^{-1} Q_Z \sigma_{v_{w,v_{w}}}^{-1} \psi_{V_w} + \psi_{V_w} \psi_{V_w}
\]

\[
= (Q_Z^\prime \sigma_{v_{w,v_{w}}}^{-1} C_0 + \psi_{V_w}) (Q_Z^\prime \sigma_{v_{w,v_{w}}}^{-1} C_0 + \psi_{V_w}) = \psi' \psi.
\]
By the same token, we have

\[
(\sigma_{ee}\sigma_{\nu h})^{-1/2}W'PZ\varepsilon = (\sigma_{ee}\sigma_{\nu h})^{-1/2}(\Pi'_WZ'\varepsilon + V'PZ\varepsilon)
\]

\[
= (\sigma_{ee}\sigma_{\nu h})^{-1/2}C_{0w} (Z'Z)^{-1/2} \left(\frac{Z'\varepsilon}{n}\right) + (\sigma_{ee}\sigma_{\nu h})^{-1/2} (Z'Z)^{-1} \left(\frac{\sigma_{ee}}{\sqrt{n}}\right)
\]

\[
\rightarrow C_{0w} (\sigma_{ee}\sigma_{\nu h})^{-1/2}Q_{Z}\varepsilon + \psi_{V\varepsilon} = \Psi'\psi_{\varepsilon}.
\]

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

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

\[
|\kappa_{\hat{\Omega}_W} - (\hat{\gamma}(\beta_0) : W)'PZ(\hat{\gamma}(\beta_0) : W)| = 0.
\]

Observe that

\[
P_Z(\hat{\gamma}(\beta_0) : W) = P_Z \left[ Z\Pi_0(\gamma) + (\varepsilon : V) \begin{pmatrix} 1 & 0 \\ \gamma & 1 \end{pmatrix} \right].
\]

Substituting this into the characteristic polynomial, and pre-multiplying by \(\begin{pmatrix} 1 & 0 \\ -\gamma & 1 \end{pmatrix}\) and post-multiplying by \(\begin{pmatrix} 1 & 0 \\ -\gamma & 1 \end{pmatrix}\)

yields

\[
|\kappa_{\hat{\Sigma}} - (\varepsilon : Z\Pi_W + V_W)'P_Z(\varepsilon : Z\Pi_W + V_W)| = 0,
\]

where \(\hat{\Sigma} = \begin{pmatrix} 1 & 0 \\ -\gamma & 1 \end{pmatrix}\).

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

\[
(\sigma_{ee}\sigma_{\nu h})^{-1/2}(\tilde{\gamma}_{LIML} - \gamma_0) \rightarrow \Delta_{LIML} = \left\{\Psi'\psi - \kappa_{LIML}\right\}^{-1} \left\{\Psi'\psi - \kappa_{LIML}\rho_{V\varepsilon}\right\},
\]

where \(\kappa_{LIML}\) is the smallest root of the characteristic polynomial

\[
\left|(\psi_{V\varepsilon} : \psi)^{\prime}(\psi_{V\varepsilon} : \psi) - \kappa_{\Sigma_{\rho}}\right| = 0, \quad \Sigma_{\rho} = \begin{pmatrix} 1 & \rho_{V\varepsilon} \\ \rho_{V\varepsilon} & 1 \end{pmatrix}.
\]
Thus, the result follows.

(b) Similarly, we have

\[ \hat{r}_j \hat{\Omega}_1 \hat{r}_j = (n - L)^{-1} (\hat{y}(\beta_0) - W \hat{\gamma})' M_Z (\hat{y}(\beta_0) - W \hat{\gamma}) \]

\[ = (n - L)^{-1} \varepsilon' \varepsilon - (n - k)^{-1} (\hat{\gamma}_j - \gamma_0)' W'M_Z \varepsilon - (n - L)^{-1} \varepsilon' M_Z (\hat{\gamma}_j - \gamma_0) \]

\[ + (n - L)^{-1} (\hat{\gamma}_j - \gamma_0)' W'M_Z W (\hat{\gamma}_j - \gamma_0) \]

\[ \Rightarrow \sigma_{\varepsilon \varepsilon} - \left( \sigma_{\varepsilon \varepsilon}^{-1/2} \Delta_j \right)' \sigma_{\varepsilon \varepsilon} - \sigma_{\varepsilon W} \left( \sigma_{\varepsilon \varepsilon}^{-1/2} \sigma_{\varepsilon \varepsilon}^{-1/2} \Delta_j \right) \]

\[ + \left( \sigma_{\varepsilon \varepsilon}^{-1/2} \sigma_{\varepsilon \varepsilon}^{-1/2} \Delta_j \right)' \sigma_{\varepsilon \varepsilon} \left( \sigma_{\varepsilon \varepsilon}^{-1/2} \sigma_{\varepsilon \varepsilon}^{-1/2} \Delta_j \right) \]

\[ = \sigma_{\varepsilon \varepsilon} (1 - 2 \rho_{\varepsilon \varepsilon} \Delta_j + (\Delta_j)^2) \]

(c) First, note that \( \hat{S}(\beta_0, \hat{\gamma}_j) = (Z'Z)^{-1/2} Z' (\hat{y}(\beta_0) : W) \hat{r}_j (\hat{r}_j' \hat{\Omega}_1 \hat{r}_j)^{-1/2} \). However, we have

\[ (Z'Z)^{-1/2} Z' (\hat{y}(\beta_0) : W) \hat{r}_j = \left( \frac{Z'Z}{n} \right)^{-1/2} \frac{Z' \varepsilon}{\sqrt{n}} + \left( \frac{Z'Z}{n} \right)^{-1/2} \frac{Z'W}{\sqrt{n}} (\gamma_0 - \hat{\gamma}_j) \]

\[ = \sigma_{\varepsilon \varepsilon}^{1/2} \left\{ \sigma_{\varepsilon \varepsilon}^{-1/2} \left( \frac{Z'Z}{n} \right)^{-1/2} \frac{Z' \varepsilon}{\sqrt{n}} + \left( \sigma_{\varepsilon \varepsilon}^{-1/2} \sigma_{\varepsilon \varepsilon}^{-1/2} \Delta_j \right) \right\} \]

\[ + \sigma_{\varepsilon \varepsilon}^{-1/2} \left( \frac{Z'Z}{n} \right)^{-1/2} \frac{Z'W}{\sqrt{n}} \sigma_{\varepsilon \varepsilon}^{-1/2} \sigma_{\varepsilon \varepsilon}^{1/2} (\gamma_0 - \hat{\gamma}_j) \}

\[ \Rightarrow \sigma_{\varepsilon \varepsilon}^{1/2} \left\{ \psi - \Psi (\psi' \psi - \kappa_j)^{-1} (\psi' \psi - \kappa_j \rho_{\varepsilon \varepsilon}) \right\} \]

\[ = \sigma_{\varepsilon \varepsilon}^{1/2} (\psi - \Psi \Delta_j) \]

Combing this with (b), the result follows.

\[ \square \]

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

\[ \square \]

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

**Proof of Lemma 4.1** The proof follows the same steps as Theorem 3 of Moreira et al. (2009) and is therefore omitted.

**Proof of Theorem 4.2** The proof is similar to Hall and Horowitz (1996) by exploiting Theorem 3.1 and Lemma 4.1, hence is omitted.

**Proof of Lemma A.7** (a). First, we have

\[
E^* \left( (v_{1,i} - V_{w,i}'\tilde{\gamma}_j) \right)^2
= n^{-1} \sum_{i=1}^n \left( \tilde{\gamma}_{1,i} - \tilde{\gamma}_{w,i} \right)^2 \\
= n^{-1} \tilde{\gamma}_1 - 2 \left( n^{-1} \tilde{\gamma}_w \right) \gamma_j + \gamma_j \left( n^{-1} \tilde{\gamma}_w \right)^2 \\
\xrightarrow{d} \sigma_{\epsilon x} \left( 1 - 2 p_{V_w} A_j + (A_j)^2 \right).
\]

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

\[
E^* \left[ V_{w,i}' (v_{1,i} - V_{w,i}'\tilde{\gamma}_j) \right] = n^{-1} \tilde{\gamma}_1 - \left( n^{-1} \tilde{\gamma}_w \right) \gamma_j \\
= n^{-1} \tilde{\gamma}_1 - \left( n^{-1} \tilde{\gamma}_w \right) \gamma_0 + \left( n^{-1} \tilde{\gamma}_w \right) \left( \gamma_0 - \tilde{\gamma}_j \right) \\
\xrightarrow{d} \sigma_{V_w} - \sigma_{\epsilon x}^{1/2} \sigma_{V_w}^{1/2} A_j.
\]

(c). By the same means, we find \( E^* \left( V_{w,i}' V_{w,i}' \right) = n^{-1} \tilde{\gamma}_w \tilde{\gamma}_w \xrightarrow{P} \sigma_{V_w} a.s. \)

**Proof of Lemma A.8** The proof follows closely Lemma A.2 of Moreira et al. (2009). Let \((c',d')'\) be a nonzero vector with \(c = (c_1', c_w')' \in \mathbb{R}^{2L}\) and \(d \in \mathbb{R}^{L(L+1)/2}\). Define

\[
X_{n,i}^* = \{ c' (V_i^* \otimes Z_i^*) + d' (m_i^* - \bar{m}) \} / \sqrt{n},
\]

where \(V_i^* = \left( v_{i,j}, V_{w,i}^* \right)'\) is the \(i\)th 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 \( X_{n,i}^s. \)

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

(b) By noting that \( \mathbb{E}^*[V_i^s V_i^s] = n^{-1} \tilde{\nu} \tilde{\nu}' \), where \( \tilde{\nu} = (\tilde{\nu}_i : \tilde{\nu}_w) \), and \( \mathbb{E}^*[Z_i^s Z_i^s] = n^{-1} Z_i^s Z_i^s \); and that

\[ n^{-1} \tilde{\nu} \tilde{\nu}' \xrightarrow{p} \Omega_w, \text{ and } n^{-1} Z_i^s Z_i^s \xrightarrow{p} Q_Z, \] it follows immediately that

\[ \mathbb{E}^*[X_{n,i}^s] = n^{-1} \left\{ c' \left( (n^{-1} \tilde{\nu} \tilde{\nu}') \otimes (n^{-1} Z_i^s Z_i^s) \right) c + d' \hat{\Omega}_{mn} \right\} < +\infty \]

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

(c) By using the same argument as in Lemma A.2 of Moreira et al. (2009), we have

\[ \lim_{n \to \infty} \sum_{i=1}^n \mathbb{E}^* \left[ X_{n,i}^s \right]^{2+\delta} = 0 \text{ a.s.} \] Since \( \mathbb{E}^* \left[ n^{-1} Z_i^s Z_i^s \right] = n^{-1} Z_i^s Z_i^s \) and \( n^{-1} Z_i^s Z_i^s - n^{-1} Z_i^s Z_i^s \sim \mathcal{N}_n \rightarrow 0 \) by the Markov law of large numbers (LLN), we also have \( n^{-1} Z_i^s Z_i^s \sim \mathcal{N}_n \xrightarrow{p} Q_Z \text{ a.s.} \) In addition, it is easy to see that \( \hat{\Omega}_W \xrightarrow{p} \Omega_w. \) Lemma A.8 follows by the Liapunov Central Limit Theorem.

**Proof of Lemma 4.3**

(a). As before, we begin with the 2SLS estimator. Let \( C_{0w}^B = C_{0w} + (n^{-1} Z_i^s Z_i^s)^{-1} (n^{-1/2} Z_i^s V_i^s), \) then

\[
\left( \mathbb{E}^* \left[ V_{w,i}^s V_{w,j}^s \right] \right)^{-1} W^s P_Z W^s
= \left( \mathbb{E}^* \left[ V_{w,i}^s V_{w,j}^s \right] \right)^{-1} \left( \hat{\Pi}_w Z_i^s Z_i^s \hat{\Pi}_w + V_{w,i}^s Z_i^s \hat{\Pi}_w + \hat{\Pi}_w Z_i^s V_{w,i}^s + V_{w,i}^s P_Z V_{w,i}^s \right)
= \left( \mathbb{E}^* \left[ V_{w,i}^s V_{w,j}^s \right] \right)^{-1} \left( c_{0w}^B \left( \frac{Z_i^s Z_i^s}{n} \right) C_{0w}^B + \left( \frac{V_{w,i}^s Z_i^s}{\sqrt{n}} \right) C_{0w}^B + C_{0w}^B \left( \frac{Z_i^s V_{w,i}^s}{\sqrt{n}} \right) \right)
+ \left( \frac{V_{w,i}^s Z_i^s}{\sqrt{n}} \right) \left( \frac{Z_i^s Z_i^s}{n} \right)^{-1} \left( \frac{Z_i^s V_{w,i}^s}{\sqrt{n}} \right)
\]

and similarly, we have

\[
\left( \mathbb{E}^* \left[ (V_{i,i}^s - V_{w,i}^s \tilde{y}_j)^2 \right] \right)^{-1/2} W^s P_Z \left( V_{i,i}^s - V_{w,i}^s \tilde{y}_j \right)
= \left( \mathbb{E}^* \left[ (V_{i,i}^s - V_{w,i}^s \tilde{y}_j)^2 \right] \right)^{-1/2} \left( \hat{\Pi}_w Z_i^s (V_{i,i}^s - V_{w,i}^s \tilde{y}_j) + V_{w,i}^s P_Z (V_{i,i}^s - V_{w,i}^s \tilde{y}_j) \right)
\]

29
Similarly, we have
\[ \psi \]

where we have
\[ \hat{\kappa} \]

From Lemmata A.7-A.8, we have
\[ n^{-1} Z' Z^* \rightharpoonup \mathcal{N} \rightarrow 0 \quad \text{and} \quad C_{0w}^B = C_{0w} + (n^{-1/2}Z'Z)^{-1} (n^{-1/2}Z'V_w), \]

it follows that
\[ \left( \mathbb{E}^* \left[ V_{w,i}^* V_{w,i}' \right] \right)^{-1} W' P_Z W \rightharpoonup n^{-1/2} Z' C_{0w}^B + 2 \psi V_w \quad \text{a.s.} \]

Similarly, we have
\[ \left( \mathbb{E}^* \left[ (V_{i,d}' - V_{w,i}^* \bar{Y}_{j})^2 \right] \right)^{-1} W' P_Z (V_{i}^* - V_{w}^* \bar{Y}_{j}) \rightharpoonup \mathcal{N} \rightarrow 0 \quad \text{as.s.,} \]

where
\[ \psi_{eSLS} = \psi_{V_1} - \psi_{V_2} (y_0 + \sigma_{ee}^{1/2} \sigma_{V_1V_2}^{-1/2} \Delta_{SLS}) = \psi_e - \sigma_{ee}^{1/2} \sigma_{V_1V_2}^{-1/2} \Delta_{SLS} \psi_{V_2}. \]

For LIML, observe that \( \hat{\kappa}^*_{SLS} \) is the smallest root of
\[ \kappa \Sigma^*_{SLS} - (V_{i}^* - V_{w}^* \tilde{\gamma}_{SLS} : Z' \hat{\Pi}_w + V_w^*) \rightharpoonup \mathcal{N} \rightarrow 0, \]

where
\[ \Sigma_{SLS}^* = \left( \begin{array}{cc} 1 & 0 \\ 0 & \hat{\Omega}_{w}^* \end{array} \right) \]

\[ \hat{\gamma}_{SLS} = \left( \begin{array}{c} \hat{\gamma}_{LIML} \\ \hat{\gamma}_{LIML} \end{array} \right) \]

So, by following the same steps as in the case of 2SLS and using Lemma 3.2, we get (conditionally on \( \mathcal{N} \))
\[ \left( \mathbb{E}^* \left[ V_{w,i}^* V_{w,i}' \right] \right)^{-1/2} (\hat{\gamma}_{LIML} - \gamma_{LIML}) \rightharpoonup (\psi^B \psi^B - \kappa_{j}^B)^{-1} (\psi^B \psi_{e,LIML} - \kappa_{j}^B \rho_{V_2,e,LIML}) \]
This establishes Lemma 4.3-(a).

(b) Similarly, we have

\[
\begin{align*}
\bar{p}_j^* \hat{\Omega}_W^* \bar{p}_j^* &= (n - L)^{-1} \left( \bar{y}^* (\beta_0) - W^* \bar{y}_j^* \right)' M_Z^* \left( \bar{y}^* (\beta_0) - W^* \bar{y}_j^* \right) \\
&= (n - L)^{-1} \left( V_1^* - V_w^* \bar{y}_j^* \right)' \left( V_1^* - V_w^* \bar{y}_j^* \right) - (n - L)^{-1} \left( \bar{y}_j^* - \bar{y}_j^* \right)' W^* M_Z^* \left( V_1^* - V_w^* \bar{y}_j^* \right) \\
&\quad - (n - L)^{-1} \varepsilon^* M_Z^* W^* \left( \bar{y}_j^* - \bar{y}_j^* \right) + (n - L)^{-1} \left( \bar{y}_j^* - \bar{y}_j^* \right)' W^* M_Z^* W^* \left( \bar{y}_j^* - \bar{y}_j^* \right).
\end{align*}
\]

Since \( n^{-1} Z' Z \mid \mathcal{F}_n \xrightarrow{a.s.} Q_Z \) and \( \hat{\Omega}_W^* \xrightarrow{a.s.} \Omega_W \), hence we have

\[
\bar{p}_j^* \hat{\Omega}_W^* \bar{p}_j^* \mid \mathcal{F}_n \xrightarrow{d} \sigma_{e\epsilon,j} \left( 1 - 2 \rho_{eW,j}^B (\Delta_j^B + (\Delta_j^B)^2) \right) \ x.a.s.
\]

(c) Again, we have

\[
\begin{align*}
\left( Z' Z \right)^{-1/2} Z^* \left( \bar{y}^* (\beta_0) : W^* \right) \bar{p}_j^* \\
= \left( \frac{Z' Z}{n} \right)^{-1/2} Z' \left( V_1^* - V_w^* \bar{y}_j^* \right) + \left( \frac{Z' Z}{n} \right)^{-1/2} Z^* \left( \bar{y}_j^* - \bar{y}_j^* \right) \\
= \sigma_{\epsilon\epsilon,j}^{1/2} \left\{ \sigma_{\epsilon\epsilon,j}^{-1/2} \left( \frac{Z' Z}{n} \right)^{-1/2} Z' \left( V_1^* - V_w^* \bar{y}_j^* \right) + \sigma_{\epsilon\epsilon,j}^{-1/2} \left( \frac{Z' Z}{n} \right)^{-1/2} \varepsilon^* M_Z^* W^* \left( \bar{y}_j^* - \bar{y}_j^* \right) \sigma_{\epsilon\epsilon,j}^{-1/2} \left( \frac{Z' Z}{n} \right)^{-1/2} \sigma_{\epsilon\epsilon,j}^{1/2} \left( \bar{y}_j^* - \bar{y}_j^* \right) \right\} + o_p(1).
\end{align*}
\]

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

\[
\left( Z' Z \right)^{-1/2} Z^* \left( \bar{y}^* (\beta_0) : W^* \right) \bar{p}_j^* \mid \mathcal{F}_n \xrightarrow{d} \sigma_{\epsilon\epsilon,j}^{1/2} \left\{ \psi_{\epsilon,j} - \psi_{\epsilon,j}^B \Delta_j^B \right\} \ x.a.s.,
\]

\[\text{a.s.}, \text{ where } \kappa_j^B \text{ is the smallest root of the characteristic polynomial} \]

\[
\left( \psi_{\epsilon,LIML} : \psi_{\epsilon,LIML}^B \right)' \left( \psi_{\epsilon,LIML} : \psi_{\epsilon,LIML}^B \right) - \kappa \Sigma_{\epsilon,LIML} = 0, \Sigma_{\epsilon,LIML} = \begin{pmatrix} 1 & \rho_{\epsilon\epsilon,LIML} \\ \rho_{\epsilon\epsilon,LIML} & 1 \end{pmatrix}.
\]
where \( \Delta_j^B = \left( \Psi^{B'}\Psi^B - \kappa_j^B \right)^{-1} \left( \Psi^{B'}\kappa_{e,j} - \kappa_j^B\kappa_{i,e,j} \right) \). The final result follows by combining this with (b).

**Proof of Theorem 4.4** The proof follows immediately from equation (4.4) and Lemma 4.3.

**Proof of Theorem 4.5** From Theorems 3.3 and 4.4, it is clear that

\[
|\mathbb{P}[|AR^x(\beta_0, \gamma_j) \leq \tau]\mathbb{P}[AR(\beta_0, \gamma_j) \leq \tau] - |\mathbb{P}[|AR^x(\beta_0, \gamma_j) > \tau]\mathbb{P}[AR(\beta_0, \gamma_j) > \tau]| = O_p(1) \text{ a.s.}
\]

for all \( \tau \in \mathbb{R} \). By choosing \( \tau = c_{AR_j}(\alpha) \equiv c_{AR} \), we have \( \mathbb{P}[|AR^x(\beta_0, \gamma_j) > c_{AR_j}] = \alpha \) so that we get

\[
|\mathbb{P}[AR(\beta_0, \gamma_j) > c_{AR_j}] - \alpha| = O_p(1) \text{ a.s.}
\]

**Proof of Lemma A.1** (a). From the subsampling algorithm, we have that as \( b \to +\infty \) and \( f_n = b/n \to 0 \), as \( n \to +\infty \). Thus, we get \( b^{-1/2}Z^bW_{w,b} \overset{a.s.}{\to} Q_Z \) given the realized sample \( \mathcal{R}_n \). Also, conditional on \( \mathcal{R}_n \), we have

\[
b^{-1/2}Z^bW_{w,b} = b^{1/2} \left( b^{-1} \sum_{i=1}^b Z_{i,b}V_{w_i,b} - \mathbb{E}^{b}(Z_{i,b}V_{w_i,b}) \right) + b^{1/2}\mathbb{E}^{b}(Z_{i,b}V_{w_i,b})
\]

\[
= b^{1/2} \left( b^{-1} \sum_{i=1}^b Z_{i,b}V_{w_i,b} - n^{-1} \sum_{i=1}^n Z_iV_{w_i} \right) + f_n^{1/2} n^{1/2} \sum_{i=1}^n Z_iV_{w_i}
\]

\( \overset{d}{\to} \psi_{ZV_{w}} \text{ a.s.} \)

Similarly, we have \( b^{-1/2}Z^bW_{b} \mid \mathcal{R}_n \overset{d}{\to} \psi_{ZW} \text{ a.s.} \) and

\[
b^{-1/2}Z^bW_{b} = b^{-1/2}Z^b(Z_b\Pi + V_{w,b})
\]

\[
= f_n^{1/2} (b^{-1}Z^b_0)C_{0w} + b^{-1/2}Z^b_0W_{w,b}
\]

32
\[ \mathcal{X}_n \overset{d}{\to} \psi_{ZV_a} a.s. \]

So, we get

\[
\hat{\gamma}_{2SLS,b} - \gamma_0 = (W_b'P_{Z,b}W_b)^{-1}W_b'P_{Z,b} \varepsilon_b \\
= \left\{ \left( b^{-1/2}W_b'Z_b \right) \left( b^{-1/2}Z_b'W_b \right)^{-1} \left( b^{-1/2}W_b'Z_b \right) \left( b^{-1/2}Z_b'W_b \right)^{-1} \right\}^{-1} \left( b^{-1/2}W_b'Z_b \right) \left( b^{-1/2}Z_b'W_b \right)^{-1} \left( b^{-1/2}Z_b' \varepsilon_b \right)
\]

\[ | \mathcal{X}_n | \overset{d}{\to} \left( \psi_{ZV_a}Q_Z^{-1}\psi_{ZV_a} \right)^{-1} \psi_{ZV_a}Q_Z^{-1}\psi_{ZV_a} = \sigma_{ee}^{1/2} \sigma_{V_aV_a}^{-1/2} \Delta_{2SLS}^S a.s. \]

where \( \Delta_{2SLS}^S = (\psi_{V_aV_a}^{-1})^{-1} \psi_{V_aV_a} \psi_e \) The result for LIML is deduced similarly.

(b). From (a), we have

\[
\hat{r}_{j,b} = (b - L)^{-1} \varepsilon_b \varepsilon_b - (b - L)^{-1} (\hat{\gamma}_{j,b} - \gamma_0) W_b'M_{Z,b} \varepsilon_b - (b - L)^{-1} \varepsilon_b'M_{Z,b}W_b \left( \hat{\gamma}_{j,b} - \gamma_0 \right)
\]

\[ | \mathcal{X}_n | \overset{d}{\to} \sigma_{ee} - \left( \sigma_{ee}^{1/2} \sigma_{V_aV_a}^{-1/2} \Delta^S_j \right) \sigma_{V_eV_e} - \sigma_{eV_e} \left( \sigma_{ee}^{1/2} \sigma_{V_aV_a}^{-1/2} \Delta^S_j \right)
\]

\[ + \sigma_{V_aV_a} \left( \sigma_{ee}^{1/2} \sigma_{V_aV_a}^{-1/2} \Delta^S_j \right)^2 a.s. 
\]

\[ = \sigma_{ee} \left( 1 - 2\rho_{V_eV_e} \Delta^S_j + (\Delta^S_j)^2 \right). \]

(c). We have

\[
(Z_b'Z_b)^{-1/2}Z_b \left( \hat{y}_b(\beta_0) : W_b \right) \hat{r}_{j,b} = \left( Z_b'Z_b \right)^{-1/2}Z_b' \varepsilon_b + \left( Z_b'Z_b \right)^{-1/2}Z_b'W_b \left( \gamma_0 - \hat{\gamma}_{j,b} \right)
\]

\[ | \mathcal{X}_n | \overset{d}{\to} \sigma_{ee}^{1/2} \left[ \psi_e - \psi^S \left( \psi^S \psi_e - \kappa_e^S \right)^{-1} \left( \psi^S \psi_e - \kappa_e^S \rho_{V_eV_e} \right) \right] a.s. 
\]

\[ = \sigma_{ee}^{1/2} \left( \psi_e - \psi^S \Delta^S_j \right) \]

The result follows by combining this with (b).

\[
\square
\]

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

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

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

References


Mikusheva, A., 2013. Survey on statistical inferences in weakly-identified instrumental variable

71(4), 1027–1048.

may be weak. Journal of Econometrics 149, 52–64.

FTEcx 6(1), 26–44.

65(3), 557–586.

variables regression. in (Corbae, Durlauf and Hansen 2006), chapter 5.


18, 1438–1452.

regression with weak instruments. in (Corbae et al. 2006), chapter 5, pp. 125–163.