Journal of Econometrics 192 (2016) 231–268

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Bootstrap inference for instrumental variable models with many weak instruments

Wenjie Wang[a,1], Maximilien Kaffo[b,*]

[a] Graduate School of Social Sciences, Hiroshima University, 1-2-1 Kagamiyama, Higashihiroshima, 7398525, Japan

[b] International Monetary Fund, 700 19th St NW, Washington, DC 20431, USA

ARTICLE INFO

Article history:
Received 9 June 2014
Received in revised form
3 April 2015
Accepted 25 December 2015
Available online 6 January 2016

JEL classification:
C15
C26

Keywords:
Bootstrap
Many instruments
Weak instruments
LIML
FULL
Corrected standard error

ABSTRACT

This study's main contribution is to theoretically analyze the application of bootstrap methods to instrumental variable models when the available instruments may be weak and the number of instruments goes to infinity with the sample size. We demonstrate that a standard residual-based bootstrap procedure cannot consistently estimate the distribution of the limited information maximum likelihood estimator or Fuller (1977) estimator under many/many weak instrument sequence. The primary reason is that the standard procedure fails to capture the instrument strength in the sample adequately. In addition, we consider the restricted efficient (RE) bootstrap of Davidson and MacKinnon (2008, 2010, 2014) that generates bootstrap data under the null (restricted) and uses an efficient estimator of the coefficient of the reduced-form equation (efficient). We find that the RE bootstrap is also invalid; however, it effectively mimics more key features in the limiting distributions of interest, and thus, is less distorted in finite samples than the standard bootstrap procedure. Finally, we propose modified bootstrap procedures that provide a valid distributional approximation for the two estimators with many/many weak instruments. A Monte Carlo experiment shows that hypothesis testing based on the asymptotic normal approximation can have severe size distortions in finite samples. Instead, our modified bootstrap procedures greatly reduce these distortions.

1. Introduction

Empirical applications of instrumental variable (IV) estimation often produce imprecise results. It is now well understood that standard first-order asymptotic theory breaks down when the instruments are weakly correlated with the endogenous regressors. In this case, commonly used IV estimators such as two-stage least square (TSLS) and limited information maximum likelihood (LIML) estimators can lose consistency; cf., Dufour (1997) and Staiger and Stock (1997), among others. However, as demonstrated by Chao and Swanson (2005), having many instruments in such a weakly identified situation can help to improve estimation accuracy. Indeed, using a large number of instruments can enhance the growth of the so-called concentration parameter even if each individual instrument is only weakly correlated with the endogenous explanatory variables. Chao and Swanson (2005) show that for well-centered IV estimators such as LIML, consistency can be established even when instrument weakness is such that the rate of growth of the concentration parameter is much slower than the sample size n. In addition, Hansen et al. (2008) reveal in an application from Angrist and Krueger (1991) that using 180 instruments, rather than 3, substantially improves estimator accuracy.

Moreover, for implementing inferences in the context of many/many weak instruments, Hansen et al. (2008) provide corrected standard errors (CSE). The CSE are an extension of those of Bekker (1994) to the case of non-Gaussian disturbances and are correct under a variety of asymptotic frameworks, including the many weak instrument sequence of Chao and Swanson (2005) and Stock and Yogo (2005), as well as the many instrument sequence of Kunitomo (1980), Morimune (1983) and Bekker (1994). Recently, the CSE are extended further by Chao et al. (2012) and Hausman et al. (2012) to the heteroscedastic case and by Newey and Windmeijer (2009) to continuously updating generalized method of moments (CUE) and other generalized empirical likelihood (GEL) estimators.

* Corresponding author.
E-mail addresses: wenjiew@hiroshima-u.ac.jp (W. Wang), mkaffomelou@imf.org (M. Kaffo).
[1] Tel.: +81 82 424 7262; fax: +81 82 424 7212.

http://dx.doi.org/10.1016/j.jeconom.2015.12.016
0304-4076/

However, our simulation evidence shows that hypothesis testing or confidence intervals (CIs) based on the CSE can be distorted severely in finite samples, especially in the case of strong endogeneity. This provides motivation for the use of the bootstrap instead of the asymptotic normal approximation to improve the quality of inference. Furthermore, the CSE have a rather tedious form, and thus, can be difficult to implement in practice; this also motivates the use of bootstrap methods. In particular, the bootstrap would help to avoid computing the tedious form of the CSE if bootstrap variance estimators or percentile-type bootstrap methods are valid under many/many weak instruments.

The existing literature on bootstrapping IV models turns out to be rather limited. Moreira et al. (2009) provide theoretical proof that guarantees the bootstrap validity of Kleibergen (2002)'s score statistic even under Staiger and Stock (1997)'s weak instrument asymptotics, in which the coefficients of the instruments are specified to be in an $n^{-1/2}$ shrinking neighborhood of zero and the number of instruments is kept fixed. Davidson and MacKinnon (2008, 2010, 2014) study various bootstrap methods (pairs bootstrap and residual-based bootstrap) of hypothesis testing and constructing confidence sets for the IV model. Their extensive simulation results show that the bootstrap approaches typically perform very well relative to the normal approximation, including the case in which instruments are quite weak. However, all these studies focus on the case in which the number of instruments is kept small relative to the sample size.

In this paper, we analyze bootstrap-based inference methods under many/many weak instruments. Based on the excellent results for the cases with a small number of instruments, one may expect the bootstrap also to perform well when the number of instruments becomes large. Surprisingly, we find that the bootstrap typically fails to mimic the limiting distributions of IV estimators in this context. We first consider a standard residual-based bootstrap method, in which the residual of the structural-form equation is obtained by using the LIML or Fuller (1977, FULL) estimator and the residual of the reduced-form equation is obtained by using the least squares estimator. We show analytically that this procedure cannot estimate the limiting distribution of LIML or FULL consistently. In particular, when the number of instruments is of the same order of magnitude as the rate of growth of the concentration parameter, the bootstrap analogue correctly replicates the convergence rate of the estimator, but the bootstrap limiting distribution has an asymptotic variance different from the original one. Furthermore, when the number of instruments grows faster than the concentration parameter, the convergence rate of the bootstrap analogue becomes even faster than that of LIML or FULL.

The primary reason of this bootstrap failure is that the standard procedure generates in the bootstrap sample "pseudo" instrument strength, which has at least the same order of magnitude as the original instrument strength. In addition, in the case with a large number of instruments, the bootstrap d.g.p. cannot mimic well important features of the disturbances in the IV model. Because of these inconsistencies, commonly used bootstrap-based inference approaches such as bootstrap variance estimator or percentile type bootstrap methods will be invalid in the case of many/many weak instruments. Similar results can be shown for other IV estimators such as the TSLS estimator, the bias-corrected TSLS estimator (Nagar, 1959; Rothenberg, 1984), and various jackknife IV estimators (Phillips and Hale, 1977; Angrist et al., 1999; Chao et al., 2012; Hausman et al., 2012; Bekker and Crudu, 2015).

We then consider the restricted efficient (RE) bootstrap procedure of Davidson and MacKinnon (2008, 2010, 2014), which generates bootstrap data under the null hypothesis (restricted) and uses efficient estimates of the reduced-form equation (efficient). These studies demonstrate that the RE bootstrap performs very well relative to the standard procedure. Here, we show that in the current context, the RE bootstrap also cannot estimate the limiting distribution of LIML or FULL consistently. However, we find that it is typically more robust to the instrument weakness than the standard bootstrap, and hence, exhibits relatively less distortion in finite samples.

Finally, we propose modifications to the RE bootstrap and justify that our modified bootstrap procedures provide a valid distributional approximation for LIML or FULL under many/many weak instrument sequences. In particular, we modify the RE bootstrap procedure by accurately rescaling the residuals and by introducing alternative reduced-form estimators, which allows the bootstrap to mimic well the instrument strength in the sample. Furthermore, we show analytically that all the bootstrap procedures analyzed in this study are asymptotically valid under percentile-t type methods. A Monte Carlo experiment demonstrates that the CSE-based normal approximation can have severe size distortions when the concentration parameter is small and/or when the degree of endogeneity is high. Our modified procedures can largely remove these distortions. In particular, one of our modified bootstrap performs best among all the procedures in terms of size control, while our second procedure is relatively balanced between size and power.

To the best of our knowledge, this study is the first to theoretically analyze the bootstrap validity under many/many weak instruments, and we obtain interesting implications of the properties of bootstrap methods that can be overlooked under conventional asymptotics. Indeed, the asymptotic approach taken here forces the distributional approximations to be more sensitive to the number and strength of available instruments and our findings highlight a fragility of bootstrap-based approximations with respect to these key features. In particular, conditions much more restrictive than those for the CSE-based normal approximation are necessary for existing bootstrap methods to estimate the limiting distributions of IV estimators consistently under many/many weak instruments. Furthermore, our results include modified, valid bootstrap procedures for the IV models, which effectively mimics the important features in the limiting distribution of interest.

The remainder of the paper is organized as follows. Section 2 introduces the model and provides a summary of the asymptotic theory for the estimators of interest and the CSE. Section 3 analyzes various residual-based bootstrap procedures and documents the inconsistency of the standard and RE bootstraps under many/many weak instrument sequences. Furthermore, we show that our modified bootstrap procedures provide a valid distributional approximation for LIML or FULL in this context. Section 4 contains the Monte Carlo results, and Section 5 concludes. All proofs are relegated to the Appendix.

2. The model, assumptions and asymptotic theory

We consider a standard linear IV regression given by

$$y = X\beta + \epsilon, \tag{1}$$

$$X = Z\Pi + V, \tag{2}$$

where y and X are an $n \times 1$ vector and an $n \times k$ matrix of observations on the endogenous variables, respectively, and Z is an $n \times l$ matrix of observation on the instruments, which we treat as deterministic. ϵ and V are an $n \times 1$ vector and an $n \times k$ matrix of random disturbances, respectively. We denote $P_Z = Z(Z'Z)^{-1}Z'$ and $M_Z = I_n - P_Z$, where I_n is an identity matrix with dimension n. Throughout this study, we consider the case in which k, the dimension of β, is small relative to n, but we let $l \to \infty$ as $n \to \infty$ in order to model the effect of having many/many weak instruments. In addition, we assume that the included exogenous variables, including a constant term, have already been filtered out from Eqs. (1) and (2).

The model and data are assumed to satisfy the following conditions.

Assumption 1. (i) The errors $\xi_i = \left(\epsilon_i, V_i'\right)'$ are i.i.d. for $i = 1, \ldots, n$ with a mean zero and positive definite variance matrix $\Sigma = \begin{pmatrix} \sigma_{\epsilon\epsilon} & \sigma_{V\epsilon}' \\ \sigma_{V\epsilon} & \Sigma_{VV} \end{pmatrix}$; $E\left(\epsilon_i^8\right)$ and $E\left(\|V_i\|^8\right)$ are bounded.
(ii) $rank(Z) = l$, $\sum_{i=1}^n (1-P_{ii})^2/n \geq C > 0$, where P_{ii} denotes the diagonal elements of the matrix P_Z.

Assumption 1(i) includes moment existence and homoscedasticity assumptions. The consistency of LIML, FULL and the CSE with many instruments depends on the homoscedasticity assumption. The condition $\sum_{i=1}^n (1-P_{ii})^2/n \geq C$ in Assumption 1(ii) implies that $l/n \leq 1 - C$, because $P_{ii} \leq 1$ implies $\sum_{i=1}^n (1-P_{ii})^2/n \leq \sum_{i=1}^n (1-P_{ii})/n = 1 - l/n$.

Assumption 2. (i) As $n \to \infty$, $\lambda_n = l/n \to \lambda$ for some constant λ satisfying $0 \leq \lambda < 1$. There exists a non-decreasing sequence of positive real numbers r_n such that, as $n \to \infty$, $r_n/n \to \kappa$ for some constant κ, with $0 \leq \kappa < \infty$, and such that $\Pi'Z'Z\Pi/r_n \to \Psi$, where Ψ is a positive definite matrix.
(ii) $\sqrt{l}/r_n \to 0$ as $n \to \infty$.
(iii) $\sum_{i=1}^n \|\Pi'Z_i\|^4/r_n^2 \to 0$ as $n \to \infty$.

Assumption 2(i) adopts the many/many weak instruments asymptotic framework. Note that r_n can be interpreted as the rate at which the concentration parameter, $\Sigma_{VV}^{-1/2}\Pi'Z'Z\Pi\Sigma_{VV}^{-1/2}$, grows as n increases. Given that the concentration parameter is a natural measure of instrument strength, the quality of instruments can be characterized by the order of magnitude of r_n, so that the slower the divergence of r_n is, the weaker are the instruments. When $r_n = n$, the concentration parameter grows as fast as the sample size. This case corresponds to the many (strong) instrument sequence considered by Kunitomo (1980), Morimune (1983), Bekker (1994), Donald and Newey (2001), Anderson et al. (2010), Hasselt (2010) and Kuersteiner and Okui (2010), among others. Allowing r_n to grow more slowly than n corresponds to the many weak instrument sequence considered by Chao and Swanson (2005), Stock and Yogo (2005), Hansen et al. (2008), Chao et al. (2012), Hausman et al. (2012), Chao et al. (2014), and Bekker and Crudu (2015), among others. Assumption 2(ii) is required for the consistency of LIML or FULL with many weak instruments, as has been shown by Chao and Swanson (2005) and Hansen et al. (2008). In addition, following Hansen et al. (2008), Chao et al. (2012) and Hausman et al. (2012), we impose Assumption 2(iii); this condition is needed for establishing asymptotic normality under many/many weak instrument sequences.

We emphasize that the many weak instrument asymptotics considered here is very different from the weak instrument asymptotics, in which l is assumed to be fixed and in which the instruments are weak in the Staiger and Stock (1997) sense (i.e., Π is specified to be in an $n^{-1/2}$ shrinking neighborhood of zero). It is well known that under this weak instrument asymptotics, the k-class IV estimators, including LIML and FULL, are inconsistent and that Wald type inferences based on these estimators can have serious size distortions. By contrast, as demonstrated by Chao and Swanson (2005), even if each component of Π is small, the combined effect of using a large number of instruments nevertheless may allow the concentration parameter to grow sufficiently fast, so that consistent estimation can be achieved as $l, n \to \infty$. For example, consider a special case with one endogenous regressor, orthonormal instruments $\left(Z'Z = nI_l\right)$, and the local-to-zero parameterization $\left(\Pi = n^{-\zeta}\iota_l, \iota_l = (1, \ldots, 1)\right)$. In this setting, even when the instruments are weak in the Staiger and Stock (1997) sense ($\zeta = 1/2$), the consistency of LIML or FULL requires only that $l \to \infty$.[2] This illustrates the potential benefit of using many weak instruments.

Now, we describe our estimators of interest and the CSE. The k-class formulation of the LIML estimator is

$$\hat{\beta} = \left(X'P_ZX - \hat{\lambda}X'X\right)^{-1}\left(X'P_Zy - \hat{\lambda}X'y\right),$$

with $\hat{\lambda} = \min_{\|\alpha\|=1} \frac{\alpha'Y'P_ZY\alpha}{\alpha'Y'Y\alpha}$ and $Y = [y, X]$. FULL is proposed by Fuller (1977) as a modification to LIML, and it replaces $\hat{\lambda}$ with $\tilde{\lambda} = \left[\hat{\lambda} - (1-\hat{\lambda})C/n\right] / \left[1 - (1-\hat{\lambda})C/n\right]$ for some constant C. In contrast to LIML, FULL has moments of all orders. In addition, FULL is approximately mean unbiased for $C = 1$, and is second-order admissible for $C \geq 4$ under standard large sample asymptotics.

We focus on these two estimators because they are more robust to the number and the quality of the instruments than is TSLS, the other commonly used IV estimator. It is well established in the literature that TSLS is seriously biased when the number of instruments is large. Indeed, in the context of many weak instruments, LIML and FULL are consistent as long as r_n grows faster than $\sqrt{l}$ while TSLS is consistent only when r_n grows faster than l. Therefore, the choice of estimators becomes more critical in situations with many weak instruments than in situations where standard asymptotics apply.

Following Hansen et al. (2008), Chao et al. (2012), and Hausman et al. (2012), we distinguish between two cases depending on the speed of growth of l relative to r_n, as follows

$$\text{Case (I)} : l/r_n \to \gamma, \quad 0 \leq \gamma < \infty;$$
$$\text{Case (II)} : l/r_n \to \infty.$$

This is necessary because the convergence rates and the limiting distributions of the estimators differ in these two cases. We expect that Case (I) would be more likely in empirical applications. However, it could be difficult to have a clear distinction between the two cases as r_n, the growth rate of the concentration parameter, is typically unobservable in practice. Indeed, our asymptotic results in Sections 3.1 and 3.2 predict that in both Cases (I) and (II), existing residual-based bootstraps will typically lead to overrejection when applied to several commonly used approaches of hypothesis testing (and undercoverage when applied to the corresponding bootstrap CIs), i.e., the bootstrap distortions are in the same direction for both cases. Moreover, the CSE-based normal approximation and our modified bootstrap procedures in Section 3.3 are valid for both cases. Therefore, in practice, it is not necessary for researchers to determine whether the data correspond to Case (I) or Case (II).

Hansen et al. (2008)[3] show the following results and give the formula of the CSE, which are an extension of Bekker (1994)'s standard errors to allow for non-Gaussian disturbances.

Theorem 2.1. *Suppose Assumptions 1–2 hold. Then, in case* (I)

$$\sqrt{r_n}(\hat{\beta} - \beta) \to^d N(0, \Lambda_I), \tag{3}$$

where $\Lambda_I = H^{-1}\Upsilon_I H^{-1}$, $H = (1-\lambda)\Psi$, $\Upsilon_I = (1-\lambda)\sigma_{\epsilon\epsilon}\left\{H + \gamma\Sigma_{\tilde{V}\tilde{V}}\right\} + (1-\lambda)\sqrt{\gamma}\left\{A + A'\right\} + \gamma B$,

$$A = E\left(\epsilon_i^2\tilde{V}_i\right) \times \lim_{n\to\infty}\sum_{i=1}^n \Pi'Z_i(P_{ii} - \lambda_n)/\sqrt{lr_n},$$
$$B = (\phi - \lambda)E\left(\left(\epsilon_i^2 - \sigma_{\epsilon\epsilon}\right)\tilde{V}_i\tilde{V}_i'\right),$$

2 In the current example, $\Pi'Z'Z\Pi = n^{1-2\zeta}l = r_n$, and the condition $\sqrt{l}/r_n \to 0$ is equivalent to $n^{1-2\zeta}\sqrt{l} \to \infty$. When $\zeta = 1/2$, it is satisfied as long as $l \to \infty$.

3 See (5.1)–(5.2) in p. 408.

$\Sigma_{\tilde{V}\tilde{V}} = E\left(\tilde{V}_i\tilde{V}_i'\right)$, $\tilde{V} = V - \epsilon q'$, $q = \sigma_{V\epsilon}/\sigma_{\epsilon\epsilon}$ *and* $\phi = \lim_{n\to\infty}\phi_n$ *where* $\phi_n = \sum_{i=1}^n P_{ii}^2/l$; *in case* (II),

$$\frac{r_n}{\sqrt{l}}(\hat{\beta} - \beta) \to^d N(0, \Lambda_{II}), \tag{4}$$

where $\Lambda_{II} = H^{-1}\Upsilon_{II}H^{-1}$, $\Upsilon_{II} = (1-\lambda)\sigma_{\epsilon\epsilon}\Sigma_{\tilde{V}\tilde{V}} + B$.

Notice that in the formula of the asymptotic variance, H corresponds to the variance term that appears in conventional asymptotics, and the term with $\Sigma_{\tilde{V}\tilde{V}}$ corresponds to the additional term owing to the effect of having many/many weak instruments. Thus, Case (I) with $\gamma > 0$ can be considered to be a knife-edge case in which the additional variance term is of the same order as the usual variance term. If r_n grows faster than l ($\gamma = 0$), the usual variance term will dominate. On the other hand, the additional variance term will dominate in Case (II). Moreover, Hansen et al. (2008)'s result extends Bekker (1994)'s formula of asymptotic variance to the case of non-Gaussian disturbances. In particular, the terms A and B in Theorem 2.1 account for the nonnormality adjustment terms. Hansen et al. (2008, p. 399) also point out that these terms will tend to be quite small when l is small relative to n, and they expect that the CSE and the Bekker (1994) standard error would often give similar results in applied work. In addition, Anderson et al. (2010, 2011) investigate the finite sample properties of LIML in a systematic way and find that the effects of nonnormality of disturbances on the distribution of LIML is often small (e.g. Anderson et al., 2010, p. 196).

To describe the CSE, we need to introduce some additional notation. Let $\epsilon(\beta) = y - X\beta$, $\hat{\sigma}_{\epsilon\epsilon}(\beta) = \epsilon(\beta)'\epsilon(\beta)/(n-k)$, $\hat{\lambda}(\beta) = \epsilon(\beta)'P_Z\epsilon(\beta)/\epsilon(\beta)'\epsilon(\beta)$, $\hat{X} = P_ZX$, $\hat{X}(\beta) = X - \epsilon(\beta)\left(\epsilon(\beta)'X\right)/\epsilon(\beta)'\epsilon(\beta)$, $\hat{V}(\beta) = M_Z\hat{X}(\beta)$,

$$\widehat{H}(\beta) = X'P_ZX - \hat{\lambda}(\beta)X'X,$$

$$\widehat{\Upsilon}_{bkk}(\beta) = \hat{\sigma}_{\epsilon\epsilon}(\beta)\Big\{(1-\hat{\lambda}(\beta))^2\hat{X}(\beta)'P_Z\hat{X}(\beta) + \hat{\lambda}(\beta)^2\hat{X}(\beta)'M_Z\hat{X}(\beta)\Big\},$$

$$\widehat{\Upsilon}(\beta) = \widehat{\Upsilon}_{bkk}(\beta) + \widehat{A}(\beta) + \widehat{A}(\beta)' + \widehat{B}(\beta),$$

$$\widehat{A}(\beta) = \sum_{i=1}^n (P_{ii} - \lambda_n)\hat{X}_i\left(\sum_{j=1}^n \epsilon_j(\beta)^2\hat{V}_j(\beta)/n\right)',$$

$$\widehat{B}(\beta) = \frac{l(\phi_n - \lambda_n)}{n(1 - 2\lambda_n + \lambda_n\phi_n)}\sum_{i=1}^n\left(\epsilon_i(\beta)^2 - \hat{\sigma}_{\epsilon\epsilon}(\beta)\right)\hat{V}_i(\beta)\hat{V}_i(\beta)'.$$

Hansen et al. (2008)'s asymptotic variance estimator is given by $\widehat{\Lambda} = \widehat{H}^{-1}\widehat{\Upsilon}\widehat{H}^{-1}$ with $\widehat{H} = \widehat{H}(\hat{\beta})$ and $\widehat{\Upsilon} = \widehat{\Upsilon}(\hat{\beta})$. Then, the asymptotic normality of the CSE-based t-test can be established for both Cases (I) and (II)

$$t_{cse} = \frac{c'(\hat{\beta} - \beta)}{\sqrt{c'\widehat{\Lambda}c}} \to^d N(0, 1) \tag{5}$$

where c are the linear combination coefficients. Notice that $\widehat{H}^{-1}\widehat{\Upsilon}_{bkk}(\hat{\beta})\widehat{H}^{-1}$ is identical to the Bekker (1994) variance estimator. Furthermore, under many/many weak instruments, $\widehat{\Lambda}$ can be quite different from the usual variance estimator $\hat{\sigma}_{\epsilon\epsilon}\widehat{H}^{-1}$ because $\widehat{\Upsilon}$ can become much larger than $\widehat{H}$ in these cases. Recently, the CSE have been further extended by Chao et al. (2012) and Hausman et al. (2012) to allow for heteroscedasticity, and by Newey and Windmeijer (2009) to general nonlinear models.

However, an important problem with the CSE-based normal approximation is that it can be inaccurate in samples of moderate size, owing to the slower than $n^{-1/2}$ convergence rate of $\hat{\beta}$ in the context of many weak instruments. In fact, as can be seen from our simulation evidence, Wald-type inferences based on the asymptotic normal approximation can be distorted severely, especially in the case of high endogeneity (i.e., when the correlation between ϵ_i and V_i is high). This provides a motivation for the use of the bootstrap as an alternative method of inference. In particular, we can improve the quality of inference by relying on the bootstrap instead of a normal approximation when computing critical values for test statistics. Moreover, the CSE have a rather tedious form that empirical researchers might find difficult to implement. This is another motivation for the use of bootstrap methods. In cases where the analytic standard errors have tedious forms or are believed to be difficult to estimate, the bootstrap often provides a useful empirical alternative. For instance, one may consider using bootstrap-based hypothesis testing without actually computing the analytic standard errors. A widely employed approach is to directly estimate the variance–covariance matrix of $\hat{\beta}$ using the bootstrap, as follows

$$\widehat{\Lambda}^*_{boot} = \frac{1}{B}\sum_{b=1}^B\left(\hat{\beta}^*_b - \frac{1}{B}\sum_{b=1}^B\hat{\beta}^*_b\right)\left(\hat{\beta}^*_b - \frac{1}{B}\sum_{b=1}^B\hat{\beta}^*_b\right)'$$

where B is chosen to be large enough so that $\widehat{\Lambda}^*_{boot}$ approximates well the variance–covariance matrix of interest. Then, a test of the null $H_0 : c'\beta = a, c \in R^k$ can be implemented using $\widehat{\Lambda}^*_{boot}$ instead of $\widehat{\Lambda}$ in Eq. (5). Alternatively, we can compute a percentile-type (symmetric) bootstrap P-value directly, as follows

$$\hat{p}^* = P^*\left(\left|c'\left(\hat{\beta}^* - \hat{\beta}\right)\right| > \left|c'\hat{\beta} - a\right|\right); \tag{6}$$

where P^* denotes the probability measure induced by the bootstrap procedure. To implement a bootstrap test at the α significance level, we reject the null hypothesis whenever $\hat{p}^* < \alpha$. Similarly, for the purpose of constructing CIs, the bootstrap variance estimator-based $100(1-\alpha)$% CIs take the form

$$CI = \left[c'\hat{\beta} - z_{1-\alpha}\sqrt{c'\widehat{\Lambda}^*_{boot}c},\ c'\hat{\beta} + z_{1-\alpha}\sqrt{c'\widehat{\Lambda}^*_{boot}c}\right],$$

where $z_{1-\alpha}$ is such that $P(|Z| \le z_{1-\alpha}) = 1 - \alpha$ with $Z \sim N(0, 1)$, and the percentile-type method leads to the following $100(1-\alpha)$% CIs:

$$CI = \left[c'\hat{\beta} - q^*_{1-\alpha},\ c'\hat{\beta} + q^*_{1-\alpha}\right],$$

where $q^*_{1-\alpha}$ is such that $P^*\left(\left|c'\left(\hat{\beta}^* - \hat{\beta}\right)\right| \le q^*_{1-\alpha}\right) = 1 - \alpha$.

However, as we show in Section 3, all these approaches will typically become invalid under many/many weak instruments if bootstrap data are generated by existing procedures.

3. Main results

In this section, we study the asymptotic validity of the bootstrap methods when applied to LIML or FULL. Various residual-based bootstrap methods adapted to the IV model are considered. We begin with what we call the standard bootstrap procedure, which amounts to re-sampling the residuals obtained by using LIML or FULL for Eq. (1) and using the least squares estimate for the reduced-form Eq. (2) to generate bootstrap data. Then, we consider the RE bootstrap procedure of Davidson and MacKinnon (2008, 2010, 2014), which generates bootstrap data under the null hypothesis and uses an efficient estimator of the coefficient of the reduced-form equation. We demonstrate that these two bootstrap procedures fail to provide valid distributional approximation to the estimators under many/many weak instruments. Furthermore, we propose modifications to the RE bootstrap procedure, and we prove the bootstrap consistency of these procedures.

The following notations are used for the bootstrap asymptotics (see Chang and Park (2003) for similar notation and for several useful bootstrap asymptotic properties): for any bootstrap statistic T^* we write $T^* \to^{P^*} 0$ in probability if for any $\delta > 0$, $\epsilon > 0$, $\lim_{n\to\infty} P\left[P^*\left(|T^*| > \delta\right) > \epsilon\right] = 0$, i.e., $P^*\left(|T^*| > \delta\right) = o_P(1)$. Also, we write $T^* = O_{P^*}(n^{\varphi})$ in probability if and only if for any $\delta > 0$ there exists a $M_\delta < \infty$ such that $\lim_{n\to\infty} P\left[P^*\left(|n^{-\varphi}T^*| > M_\delta\right) > \delta\right] = 0$, i.e., for any $\delta > 0$ there exists a $M_\delta < \infty$ such that $P^*\left(|n^{-\varphi}T^*| > M_\delta\right) = o_P(1)$. Finally, we write $T^* \to^{d^*} T$ in probability if, conditional on the sample, T^* weakly converges to T under P^*, for all samples contained in a set with probability converging to one. Specifically, we write $T^* \to^{d^*} T$ in probability if and only if $E^*\left(f(T^*)\right) \to E\left(f(T)\right)$ in probability for any bounded and uniformly continuous function f. To be concise, we sometimes use the short version $T^* \to^{P^*} 0$ to say that $T^* \to^{P^*} 0$ in probability, and use $T^* = O_{P^*}(n^{\varphi})$ for $T^* = O_{P^*}(n^{\varphi})$ in probability.

3.1. Standard bootstrap procedure

We begin with the standard residual-based bootstrap procedure for LIML. Given the LIML estimate of β and the least square (first-stage) estimate $\hat{\Pi} = (Z'Z)^{-1}Z'X$, the residuals are obtained as

$$\hat{\epsilon} = y - X\hat{\beta} \tag{7}$$

$$\hat{V} = X - Z\hat{\Pi}. \tag{8}$$

Then, (ϵ^*, V^*) are drawn from the empirical distribution function of $\left(\hat{\epsilon}, \hat{V}\right)$. Notice that we do not re-center the residuals here since they already have mean zero under our setting that the constant term is filtered out. Next, we set

$$y^* = X^*\hat{\beta} + \epsilon^* \tag{9}$$

$$X^* = Z\hat{\Pi} + V^*. \tag{10}$$

Finally, we compute the bootstrap analogue of LIML using the pseudo-sample $\{X^*, y^*\}$:

$$\hat{\beta}^*_{std} = \left(X^{*'}P_Z X^* - \hat{\lambda}^* X^{*'}X^*\right)^{-1}\left(X^{*'}P_Z y^* - \hat{\lambda}^* X^{*'}y^*\right),$$

where $\hat{\lambda}^* = \min_{\|\alpha\|=1} \frac{\alpha' Y^{*'} P_Z Y^* \alpha}{\alpha' Y^{*'} Y^* \alpha}$ and $Y^* = [y^*, X^*]$. The standard residual-based bootstrap for FULL can be implemented in a similar way. The procedure in Eqs. (7)–(10) is called an unrestricted inefficient (UI) procedure in Davidson and MacKinnon (2008) since the bootstrap d.g.p. is generated under $\hat{\beta}$ rather than β_0 (thus, unrestricted) and under the least squares estimator $\hat{\Pi}$, instead of a more efficient reduced-form estimator (thus, inefficient). Moreover, Freedman (1984) studies a similar bootstrap procedure for TSLS with resampled instruments, and shows asymptotic validity of this procedure in the case of a fixed number of instruments.

Below we show that the standard residual-based bootstrap fails to estimate the limiting distribution of LIML or FULL consistently in both Cases (I) and (II).

Theorem 3.1. *Suppose Assumptions 1–2 hold. Then, in case* (I) *with* $0 < \gamma < \infty$,

$$\sqrt{r_n}(\hat{\beta}^*_{std} - \hat{\beta}) \to^{d^*} N(0, \bar{\Lambda}_I) \text{ in probability},$$

where

$$\bar{\Lambda}_I = \bar{H}_I^{-1}\bar{\Upsilon}_I\bar{H}_I^{-1}$$
$$\bar{\Upsilon}_I = (1-\lambda)\sigma_{\epsilon\epsilon}\left\{\bar{H}_I + \gamma\bar{\Sigma}_{\tilde{V}\tilde{V}}\right\} + (1-\lambda)\sqrt{\gamma}\left\{\bar{A} + \bar{A}'\right\} + \gamma\bar{B}$$
$$\bar{H}_I = H + (1-\lambda)\gamma\Sigma_{VV}$$
$$\bar{\Sigma}_{\tilde{V}\tilde{V}} = (1-\lambda)\Sigma_{\tilde{V}\tilde{V}} + (\lambda - \lambda^2)\sigma_{V\epsilon}\sigma_{V\epsilon}'/\sigma_{\epsilon\epsilon}$$
$$\bar{A} = (1-\lambda)A$$
$$\bar{B} = (1-2\lambda+\lambda\phi)B + \lambda(\phi-\lambda)^2\left\{qE\left(\epsilon_i^3\tilde{V}_i'\right) + E\left(\epsilon_i^3\tilde{V}_i\right)q' + q\left(E\left(\epsilon_i^4\right) - (\sigma_{\epsilon\epsilon})^2\right)q'\right\};$$

in case (II),

$$\sqrt{l}(\hat{\beta}^*_{std} - \hat{\beta}) \to^{d^*} N(0, \bar{\Lambda}_{II}) \text{ in probability},$$

where

$$\bar{\Lambda}_{II} = \bar{H}_{II}^{-1}\bar{\Upsilon}_{II}\bar{H}_{II}^{-1}$$
$$\bar{\Upsilon}_{II} = (1-\lambda)\sigma_{\epsilon\epsilon}\left\{\bar{H}_{II} + \bar{\Sigma}_{\tilde{V}\tilde{V}}\right\} + \bar{B}$$
$$\bar{H}_{II} = (1-\lambda)\Sigma_{VV}.$$

Theorem 3.1 states the distributional results for the standard residual-based bootstrap procedure under many/many weak instrument sequences. As the theorem makes clear, in Case (I), the bootstrap analogue $\hat{\beta}^*_{std}$ correctly replicates the convergence rate of $\hat{\beta}$ and the bootstrap distribution is asymptotically normal, but its asymptotic variance–covariance matrix is different from the one derived in Theorem 2.1. To see why the bootstrap fails, let us first consider the term H, which characterizes the instrument strength, or the signal, in the IV model. Note that the LIML objective function with bootstrap pseudo-data is

$$\widehat{Q}^*(\beta) = \frac{(y^* - X^*\beta)' P_Z (y^* - X^*\beta)}{(y^* - X^*\beta)'(y^* - X^*\beta)}$$

and the usual Taylor expansion of the first-order condition $\partial\widehat{Q}^*(\hat{\beta}^*_{std})/\partial\beta = 0$ yields

$$\hat{\beta}^*_{std} - \hat{\beta} = \left(\partial^2\widehat{Q}^*(\bar{\beta}^*)/\partial\beta\partial\beta'\right)^{-1}\partial\widehat{Q}^*(\hat{\beta})/\partial\beta$$

where $\bar{\beta}^*$ is an intermediate value on the line joining $\hat{\beta}^*_{std}$ and $\hat{\beta}$. We can show that in Case (I), the bootstrap Hessian term

$$\left(\partial^2\widehat{Q}^*(\bar{\beta}^*)/\partial\beta\partial\beta'\right)/r_n \to^{P^*} \bar{H}_I = (1-\lambda)\left(\Psi + \gamma\Sigma_{VV}\right) = H + (1-\lambda)\gamma\Sigma_{VV}$$

in probability, where P^* denotes the probability measure induced by the standard bootstrap. Thus, the bootstrap fails to mimic well the original Hessian term and results in an approximation error term $(1-\lambda)\gamma\Sigma_{VV}$. Similarly, by applying a bootstrap central limit theorem, we find that although the bootstrap Jacobian term $\sqrt{r_n}\left(\partial\widehat{Q}^*(\hat{\beta})/\partial\beta\right)$ converges in probability to normal distribution, the term H in the asymptotic variance of $\sqrt{r_n}\left(\partial\widehat{Q}(\beta_0)/\partial\beta\right)$ is also replaced by $\bar{H}_I$.

The reason for these bootstrap failures is that under many/many weak instruments, the standard bootstrap cannot capture well the signal in the original sample. Indeed, the bootstrap d.g.p. leads to a "pseudo" increase of the concentration parameter, as long as the rate of growth of l is not slower than that of r_n, and this results in the extra term $(1-\lambda)\gamma\Sigma_{VV}$. For expository purposes, let us assume that $l/n \to 0$, so that the limiting distribution of LIML is largely simplified and we can focus on the bootstrap distortion related to the concentration parameter. Then, in Case (I), the asymptotic variance of $\hat{\beta}$ is $\sigma_{\epsilon\epsilon}H^{-1} + \gamma H^{-1}\Sigma_{\tilde{V}\tilde{V}}H^{-1}$ with $H = \Psi$ in the current context. The asymptotic variance of its bootstrap analogue has a similar formula: $\sigma_{\epsilon\epsilon}\bar{H}_I^{-1} + \gamma\bar{H}_I^{-1}\Sigma_{\tilde{V}\tilde{V}}\bar{H}_I^{-1}$, where $\bar{H}_I = H + \gamma\Sigma_{VV}$; however, it will be smaller than the original one as long as $\gamma > 0$. Thus, it turns out that because of the pseudo increase of the concentration parameter, the bootstrap analogue of LIML will be more efficient than the original LIML in this case.

The bootstrap failure becomes even more severe in Case (II). In this case, the convergence rate of $\hat{\beta}^*_{std}$ turns out to be different from that of $\hat{\beta}$: $\hat{\beta}^*_{std} - \hat{\beta} = O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$ in probability, while $\hat{\beta} - \beta = O_P\left(\frac{\sqrt{l}}{r_n}\right)$. Notice that $\frac{1}{\sqrt{l}}/\frac{\sqrt{l}}{r_n} \to 0$ in Case (II); that is, $\hat{\beta}^*_{std}$ converges under P^* to $\hat{\beta}$, the true value in the bootstrap world, at a higher speed than $\hat{\beta}$ converges to β. Intuitively, the rate of convergence of $\hat{\beta}$ depends on the strength of the available instruments, as reflected in the relative order of magnitude of r_n vis-à-vis l. On the other hand, the rate of $\hat{\beta}^*_{std}$ depends not only on the original instrument strength but also the bootstrap-generated "pseudo" instrument strength. In Case (I), the pseudo instrument strength is of the same order of magnitude as the original instrument strength, so that the convergence speed of $\hat{\beta}^*_{std}$ remains the same as that of $\hat{\beta}$. In Case (II), however, this pseudo strength would be so large as to dominate the original instrument strength, change the identification strength in the bootstrap world,[4] and thus, accelerate the convergence speed of $\hat{\beta}^*_{std}$. For a similar reason, the formula of the bootstrap asymptotic variance $\bar{\Lambda}_{II}$ differs greatly from Λ_{II} in Theorem 2.1. Indeed, the conventional variance term H does not appear in the formula of Υ_{II} because it is dominated by the many/many weak IV adjustment term $(1-\lambda)\sigma_{\epsilon\epsilon}\Sigma_{\tilde{V}\tilde{V}}$ and the nonnormality adjustment term B. By contrast, $\bar{H}_{II}$, the bootstrap analogue of H in Case (II), does appear in the formula of $\bar{\Upsilon}_{II}$ as the pseudo instrument strength guarantees that it is of the same order as the other terms. Moreover, same bootstrap failures can be found for FULL in both Cases (I) and (II), as it can be shown to be asymptotically equivalent to LIML in the bootstrap world.

To see how these documented failures would affect bootstrap tests and CIs, let us consider the case of testing $H_0 : \beta = \beta_0$ with one endogenous variable ($k = 1$). In case (I), the bootstrap P-value $\hat{p}^*$ defined in Eq. (6) would tend to overreject the null hypothesis because the limiting distribution of $\hat{\beta}^*_{std}$ has a small asymptotic variance compared to that of $\hat{\beta}$. Similar phenomenon will occur in Case (II), since in this case we have

$$P^*\left(\left|\hat{\beta}^*_{std} - \hat{\beta}\right| > \left|\hat{\beta} - \beta_0\right|\right) = P^*\left(\left|\left(\frac{r_n}{l}\right)\sqrt{l}\left(\hat{\beta}^*_{std} - \hat{\beta}\right)\right| > \left|\frac{r_n}{\sqrt{l}}\left(\hat{\beta} - \beta_0\right)\right|\right) \to^P 0$$

given that $\sqrt{l}\left(\hat{\beta}^*_{std} - \hat{\beta}\right) = O_{P^*}(1)$ in probability, $\frac{r_n}{\sqrt{l}}\left(\hat{\beta} - \beta_0\right) = O_P(1)$, and $r_n/l \to 0$, in this case. Thus, our asymptotic results predict that for both Cases (I) and (II), the bootstrap procedure leads to over-rejection of the null, and this is confirmed by our simulation results. Similar arguments hold for bootstrap variance estimators and CIs.

Furthermore, some algebra shows that by the resampling scheme of the standard bootstrap, the following results hold for the bootstrap disturbances:

$$E^*\left(V_i^*\epsilon_i^*\right) = n^{-1}\sum_{i=1}^{n}\hat{V}_i\hat{\varepsilon}_i \quad \to^P (1-\lambda)\sigma_{V\varepsilon}$$

$$E^*\left(V_i^*V_i^{*'}\right) = n^{-1}\sum_{i=1}^{n}\hat{V}_i\hat{V}_i' \to^P (1-\lambda)\Sigma_{VV}$$

where E^* denotes the expectation under the probability measure induced by the standard bootstrap. Therefore, the standard residual bootstrap cannot consistently estimate the variance–covariance matrix of $(\epsilon_i, V_i')'$ when the number of instruments grows at the same speed as the sample size ($\lambda \neq 0$). Furthermore, this implies that for $\tilde{V}_i^* = V_i^* - \epsilon_i^* q^{b'}$ where $q^b = E^*\left(V_i^*\epsilon_i^*\right)/E^*\left(\epsilon_i^{*2}\right)$, the residuals from the (bootstrap) population regression of V_i^* on ϵ_i^*, we have

$$E^*\left(\tilde{V}_i^*\tilde{V}_i^{*'}\right) = E^*\left[\left(V_i^* - \epsilon_i^* q^{b'}\right)\left(V_i^* - \epsilon_i^* q^{b'}\right)'\right] \to^P (1-\lambda)\Sigma_{\tilde{V}\tilde{V}} + (\lambda - \lambda^2)\frac{\sigma_{V\epsilon}\sigma_{V\epsilon}'}{\sigma_{\epsilon\epsilon}},$$

Thus, as long as the number of instruments grows as fast as the sample size, the standard residual bootstrap fails to estimate consistently $\Sigma_{\tilde{V}\tilde{V}}$, which plays an important role in the formula of the asymptotic variance in Theorem 2.1. In addition, similar results of inconsistency can be shown for other bootstrap moments such as $E^*\left(\epsilon_i^{2*}\tilde{V}_i^*\right)$ and $E^*\left(\epsilon_i^{*2}\tilde{V}_i^*\tilde{V}_i^{*'}\right)$.

Remarks. 1. Since $\hat{\beta}$ attains consistency under our assumptions, folklore may suggest that the bootstrap d.g.p. in Eqs. (7)–(10) will be valid as long as $\hat{\Pi}$ consistently estimates Π. Interestingly, this turns out to be wrong in the current context. Indeed, it is shown in Portnoy (1984) that $\hat{\Pi}$ will be consistent provided that the growth rate of l is not too fast related to the growth rate of n ($l(\log l)/n \to 0$). However, Theorem 3.1 shows that without proper restriction on the relationship between l and r_n, such a condition does not warrant the bootstrap consistency under many weak instruments.

2. Similar results of bootstrap failure can be shown for other IV estimators such as the commonly used TSLS estimator, the bias-adjusted TSLS estimator (e.g. Nagar, 1959; Rothenberg, 1984), and various jackknife IV estimators (JIVEs; e.g. Phillips and Hale, 1977; Angrist et al., 1999; Chao et al., 2012; Hausman et al., 2012; Bekker and Crudu, 2015). We omit these results for the conciseness of the paper.

3. Instead of using residual-based bootstrap, one may consider implementing the nonparametric i.i.d. bootstrap (pairs bootstrap), which amounts to re-sampling the rows of the matrix $[y, X, Z]$. However, the extensive simulation evidence in Davidson and MacKinnon (2008, 2010, 2014) and our trial simulation show that the pairs bootstrap performs substantially worse than residual-based bootstrap methods, even when the number of instruments is small relative to the sample size. Note that under the nonparametric i.i.d. bootstrap, the bootstrap analogue of the slope coefficient in the first-stage regression is characterized by

$$\left[E^*\left(Z_i^*Z_i^{*'}\right)\right]^{-1}E^*\left(Z_i^*X_i^*\right) = \left(n^{-1}Z'Z\right)^{-1}\left(n^{-1}Z'X\right),$$

which is exactly $\hat{\Pi}$ used in the standard residual-based bootstrap.

On the other hand, when l is of lower order of magnitude relative to r_n, the signal part of the limiting distribution of LIML or FULL dominates and

$$\sqrt{r_n}(\hat{\beta} - \beta) \to^d N\left(0, \sigma_{\epsilon\epsilon}\Psi^{-1}\right);$$

the standard residual-based bootstrap does estimate this limiting distribution consistently.

Corollary 3.1. *Suppose that Assumptions 1–2 hold and suppose that $l/r_n \to 0$ ($\gamma = 0$), then*

$$\sup_{x\in R}\left|P^*\left(\sqrt{r_n}(\hat{\beta}^*_{std} - \hat{\beta}) \le x\right) - P\left(\sqrt{r_n}(\hat{\beta} - \beta) \le x\right)\right| \to^P 0$$

where P^* denotes the probability measure induced by the standard bootstrap procedure.

[4] Antoine and Renault (2009, 2012, 2013) discuss in detail the relationship between the identification strength and the rate of convergence of estimators in GMM when the identification is weaker than that under conventional asymptotics. See also Andrews and Cheng (2012), where some cases related to IV regressions are discussed.

Remark. Closely related to our paper is the literature on bootstrapping linear model with increasing dimension. More precisely, consider the following model

$$y_i = X_i'\beta + \epsilon_i, \quad i = 1, \dots, n$$

where X_i's and β are p-dimensional vectors, and ϵ_i's are i.i.d. errors. Asymptotics where p may increase with n have been considered by Bickel and Freedman (1983), Portnoy (1984, 1985, 1988), Mammen (1989, 1993), etc. In particular, Bickel and Freedman (1983) show that residual-based bootstrap consistently estimate the distribution of the least square estimates if $p^2/n \to 0$, and for linear contrasts of these estimates it works if $p/n \to 0$. Mammen (1989) generalizes these results to the case of M estimates. Furthermore, it is shown by these authors that under large p asymptotics, residual-based bootstrap even works in the case that normal approximation typically fails. Apparently, the rate of growth of p with respect to n is crucial for bootstrap consistency under this framework.

In contrast, we show that for the IV model with large l, the bootstrap consistency depends essentially on the relative magnitude of r_n vis-à-vis l, rather than on the relationship between n and l. Additionally, different from bootstrapping under large p asymptotics, conditions more restrictive than those for the normal approximation are necessary for the standard bootstrap to be valid. For the example in Section 2, the bootstrap consistency requires $l/r_n = n^{2\zeta-1} \to 0$ ($\zeta < 1/2$; i.e., the instruments need to be stronger than Staiger and Stock (1997)'s weak instruments), while the CSE-based normal approximation only requires $\sqrt{l}/r_n = l^{-1/2}n^{2\zeta-1} \to 0$.

3.2. Restricted efficient bootstrap procedure

In this section, we study the other residual-based bootstrap procedure recently proposed by Davidson and MacKinnon (2008, 2010, 2014) for the IV model. The RE residual-based bootstrap has two key features. First, the bootstrap pseudo-data is obtained under the null hypothesis $H_0 : \beta = \beta_0$. Second, the RE bootstrap uses a more efficient (reduced-form) estimate instead of $\hat{\Pi}$ in the standard residual bootstrap. Following Davidson and MacKinnon (2008, 2010, 2014), we first obtain the residuals for the RE procedure by

$$\begin{aligned}\epsilon(\beta_0) &= y - X\beta_0 \\ \tilde{V}(\beta_0) &= X - Z\tilde{\Pi}(\beta_0)\end{aligned}$$

where

$$\tilde{\Pi}(\beta_0) = (Z'Z)^{-1}Z'\left(X - \epsilon(\beta_0)\frac{\epsilon'(\beta_0)M_Z X}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}\right).$$

Then, (ϵ^*, V^*) are drawn from the empirical distribution function of $\left(\sqrt{\frac{n}{n-k}}\epsilon(\beta_0), \sqrt{\frac{n}{n-l}}\tilde{V}(\beta_0)\right)$. Next, we set

$$\begin{aligned}y^* &= X^*\beta_0 + \epsilon^* \\ X^* &= Z\tilde{\Pi}(\beta_0) + V^*\end{aligned}$$

and obtain $\hat{\beta}^*_{re}$ using pseudo-data generated by this procedure. Notice that $\tilde{\Pi}(\beta_0)$ is the maximum likelihood estimator of Π when β is constrained to take the null value β_0. It is also used in Kleibergen (2002) and Moreira (2003) to construct their weak identification robust statistics. In particular, they show that using $\tilde{\Pi}(\beta_0)$ rather than $\hat{\Pi}$ leads to their Lagrange Multiplier (LM) test for $H_0 : \beta = \beta_0$ that is asymptotically pivotal even under weak instrument asymptotics of Staiger and Stock (1997).

The RE bootstrap procedure has been applied very successfully in the IV models with relatively small number of instruments. As can be observed from the extensive simulation results in Davidson and MacKinnon (2008, 2010), using this procedure instead of the standard residual bootstrap or the nonparametric i.i.d. bootstrap greatly improves size control for testing the null hypothesis H_0, especially when the available instruments are quite weak (e.g., when $a = 2$ in their setting, which corresponds to the case where the concentration parameter equals 4). The RE bootstrap is also used in Davidson and MacKinnon (2014) to build confidence sets for β in a similar context. Their results demonstrate that in contrast to what is widely believed, even when the instruments are very weak, one can make the Wald-based confidence sets perform well using the RE bootstrap procedure.

However, we find that under many/many weak instrument sequences the RE bootstrap is also invalid in general. The following theorem states the asymptotic distributional results for the RE bootstrap.

Theorem 3.2. *Suppose Assumptions 1–2 hold. Then, in Case (I) with $0 < \gamma < \infty$ and under $H_0 : \beta = \beta_0$,*

$$\sqrt{r_n}(\hat{\beta}^*_{re} - \beta_0) \to^{d^*} N(0, \tilde{\Lambda}_I) \text{ in probability},$$

where

$$\begin{aligned}\tilde{\Lambda}_I &= \tilde{H}_I^{-1}\tilde{\Upsilon}_I\tilde{H}_I^{-1} \\ \tilde{\Upsilon}_I &= (1-\lambda)\sigma_{\epsilon\epsilon}\left\{\tilde{H}_I + \gamma\Sigma_{\tilde{V}\tilde{V}}\right\} + (1-\lambda)\sqrt{\gamma}\left\{\tilde{A} + \tilde{A}'\right\} + \gamma\tilde{B} \\ \tilde{H}_I &= H + (1-\lambda)\gamma\Sigma_{\tilde{V}\tilde{V}} \\ \tilde{A} &= \sqrt{1-\lambda}A \\ \tilde{B} &= \frac{1-2\lambda+\lambda\phi}{1-\lambda}B;\end{aligned}$$

in case (II) and under $H_0 : \beta = \beta_0$,

$$\sqrt{l}(\hat{\beta}^*_{re} - \beta_0) \to^{d^*} N(0, \tilde{\Lambda}_{II}) \text{ in probability},$$

where

$$\begin{aligned}\tilde{\Lambda}_{II} &= \tilde{H}_{II}^{-1}\tilde{\Upsilon}_{II}\tilde{H}_{II}^{-1} \\ \tilde{H}_{II} &= (1-\lambda)\Sigma_{\tilde{V}\tilde{V}} \\ \tilde{\Upsilon}_{II} &= (1-\lambda)\sigma_{\epsilon\epsilon}\left\{\tilde{H}_{II} + \Sigma_{\tilde{V}\tilde{V}}\right\} + \tilde{B}.\end{aligned}$$

Investigating the results in Theorem 3.2, we find that the RE bootstrap is also inconsistent as long as l goes to infinity at a rate not slower than r_n. In particular, using similar arguments as for the standard residual bootstrap, we obtain that in Case (I), the RE-based approximation of the Hessian term

$$\left(\partial^2\widehat{Q}^*(\bar{\beta}^*)/\partial\beta\partial\beta'\right)/r_n \to^{P^*} \tilde{H}_I = H + (1-\lambda)\gamma\Sigma_{\tilde{V}\tilde{V}} \tag{11}$$

in probability under H_0, where P^* denotes the probability measure induced by the RE bootstrap and $\bar{\beta}^*$ denotes an intermediate value on the line joining $\hat{\beta}^*_{re}$ and β_0. Thus, the RE procedure also fails to adequately mimic the instrument strength in the sample, and the pseudo instrument strength in the bootstrap sample results in an approximation error of the same order of magnitude as the key parameter H. A similar problem occurs in $\sqrt{r_n}\left(\partial\widehat{Q}^*(\beta_0)/\partial\beta\right)$, creating a discrepancy between the limiting distribution of the original statistic and its bootstrap analogue. Also, the bootstrap failure becomes more severe in Case (II) as $\hat{\beta}^*_{re}$ converges under the RE bootstrap measure P^* at a rate of $O_{P^*}\left(1/\sqrt{l}\right)$, the same convergence rate as $\hat{\beta}^*_{std}$.

For the moments of the bootstrap disturbances, some algebra shows that

$$E^*\left(V_i^*\epsilon_i^*\right) = \sqrt{\frac{n}{n-k}}\sqrt{\frac{n}{n-l}}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{V}_i(\beta_0)\epsilon_i(\beta_0)\right)$$

$$\to^P \frac{\sigma_{V\epsilon}}{\sqrt{1-\lambda}}$$

$$E^*\left(V_i^* V_i^{*'}\right) = \frac{n}{n-l}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{V}_i(\beta_0)\tilde{V}_i'(\beta_0)\right) \to^P \Sigma_{VV} + \frac{\lambda}{1-\lambda}\frac{\sigma_{V\epsilon}\sigma_{V\epsilon}'}{\sigma_{\epsilon\epsilon}}$$

under H_0, where E^* denotes the expectation under the corresponding bootstrap measure. Similar inconsistency results also can be obtained for the non-normality adjustment terms A and B. Notice that these bootstrap inconsistencies essentially originate from the numerosity of available instruments, and will disappear as long as $l/n \to 0$, different from the inconsistency documented in Eq. (11).

Interestingly, we also find that in contrast to the standard bootstrap, the RE bootstrap consistently estimates $\Sigma_{\tilde{V}\tilde{V}}$, the variance of residuals from the population regression of V_i on ϵ_i:

$$E^*\left(\tilde{V}_i^* \tilde{V}_i^{*'}\right) \to^P \Sigma_{\tilde{V}\tilde{V}}$$

under H_0, including the case that l is of the same order of magnitude as n. Indeed, under the RE bootstrap d.g.p., $E^*\left(V_i^*\epsilon_i^*\right)$ tends to overestimate $\sigma_{V\epsilon}$ while $E^*\left(V_i^* V_i^{*'}\right)$ converges in probability to a sum of Σ_{VV} and an extra term related to $\sigma_{V\epsilon}$. These distortions turn out to cancel each other out asymptotically in $E^*\left(\tilde{V}_i^* \tilde{V}_i^{*'}\right)$. This is remarkable since according to Theorem 2.1, the many/many weak instruments adjustment term crucially depends on $\Sigma_{\tilde{V}\tilde{V}}$. As has been highlighted by Hansen et al. (2008), in practice this adjustment term can be comparable to the usual asymptotic variance term H, while the non-normality adjustment terms will tend to be quite small relative to $\Sigma_{\tilde{V}\tilde{V}}$ and H. Furthermore, because the efficient reduced-form estimator $\tilde{\Pi}(\beta_0)$ is used when generating the bootstrap sample, the pseudo increase of the concentration parameter noted in Eq. (11) will be typically smaller than that in the standard bootstrap, especially in the case of strong endogeneity. Specifically, we have $H \le \tilde{H}_I \le \bar{H}_I$ in Case (I) and $\tilde{H}_{II} \le \bar{H}_{II}$ in Case (II), which follows from $\Sigma_{\tilde{V}\tilde{V}} \le \Sigma_{VV}$. In sum, our asymptotic results in Theorem 3.2 predict that the RE-based distributional approximation for LIML or FULL will be more accurate than the standard bootstrap-based approximation. This is consistent with the superior finite sample performance of the RE bootstrap demonstrated in Davidson and MacKinnon (2008, 2010, 2014), and is also confirmed by our simulation results.

It is also easy to see that the RE bootstrap is consistent when $l/r_n \to 0$.

Corollary 3.2. *Suppose that Assumptions 1–2 hold and that $l/r_n \to 0\ (\gamma = 0)$, then under $H_0 : \beta = \beta_0$,*

$$\sup_{x\in R}\left|P^*\left(\sqrt{r_n}(\hat{\beta}_{re}^* - \beta_0) \le x\right) - P\left(\sqrt{r_n}(\hat{\beta} - \beta_0) \le x\right)\right| \to^P 0$$

where P^* denotes the probability measure induced by the RE bootstrap procedure.

Thus, similar to the standard bootstrap, the RE bootstrap is asymptotically valid only when the available instruments are sufficiently strong so that the concentration parameter grows at a faster rate than the number of instruments.

3.3. Modified RE bootstrap procedure

Now we propose two bootstrap procedures which estimate the distribution of LIML or FULL consistently under much weaker conditions on the growth rate of l vis-à-vis r_n than the standard or RE bootstrap. Our modified RE (MRE) procedures achieve this goal by correctly re-scaling the residuals and by using alternative reduced-form estimators, which are able to remove the pseudo increase of the concentration parameter and thus allow the bootstrap to effectively mimic the instrument strength in the original sample.

More specifically, the residuals for the two MRE bootstrap procedures are obtained as

$$\epsilon(\beta_0) = y - X\beta_0$$
$$\hat{V} = X - Z\hat{\Pi}$$

and (ϵ^*, V^*) is drawn from the empirical distribution function of $\left(\sqrt{\frac{n}{n-l}}M_Z\epsilon(\beta_0), \sqrt{\frac{n}{n-l}}\hat{V}\right)$. Then, different reduced-form estimators are used in the two procedures to generate the bootstrap data.

I. Modified Restricted Efficient 1 (MRE1) Bootstrap Procedure

For the MRE1 bootstrap d.g.p., we use $\tilde{\Pi}_m(\beta_0)$ which is based on modifying the RE reduced-form estimator $\tilde{\Pi}(\beta_0)$ in the following way:

$$\tilde{\Pi}_m(\beta_0) = \tilde{\Pi}(\beta_0)\left(\tilde{\Psi}^{-1/2}(\beta_0)\tilde{\Psi}_m^{1/2}(\beta_0)\right) \tag{12}$$

where

$$\tilde{\Psi}(\beta_0) = \tilde{\Pi}'(\beta_0)Z'Z\tilde{\Pi}(\beta_0)$$
$$\tilde{\Psi}_m(\beta_0) = \left(\tilde{\Psi}(\beta_0) - l\hat{\Sigma}_{\tilde{V}\tilde{V}}(\beta_0), 0\right)^+$$
$$\hat{\Sigma}_{\tilde{V}\tilde{V}}(\beta_0) = \frac{1}{n-l}\tilde{X}'(\beta_0)M_Z\tilde{X}(\beta_0)$$
$$\tilde{X}(\beta_0) = X - \epsilon(\beta_0)\left(\frac{\epsilon'(\beta_0)M_Z X}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}\right)$$

and $(\cdot, 0)^+ = max\,(\cdot, 0)$.[5] Then, we set

$$y^* = X^*\beta_0 + \epsilon^*$$
$$X^* = Z\tilde{\Pi}_m(\beta_0) + V^*$$

and compute $\hat{\beta}_m^*$ using the pseudo-data obtained by this bootstrap procedure. We expect the MRE1 procedure to have excellent performance when $\beta = \beta_0$. However, imposing the null on the reduced-form estimator may entail power losses when the true β is different from the hypothesized value β_0. Therefore, we also consider the following modified procedure.

II. Modified Restricted Efficient 2 (MRE2) Bootstrap Procedure

For the MRE2 bootstrap, we use an unrestricted reduced-form estimator. Specifically, the bootstrap d.g.p. is generated by $\tilde{\Pi}_m(\hat{\beta})$, which replaces β_0 in $\tilde{\Pi}_m(\beta_0)$ with $\hat{\beta}$, i.e.,

$$\tilde{\Pi}_m(\hat{\beta}) = \tilde{\Pi}(\hat{\beta})\left(\tilde{\Psi}^{-1/2}(\hat{\beta})\tilde{\Psi}_m^{1/2}(\hat{\beta})\right). \tag{13}$$

Then, we obtain the bootstrap data by setting

$$y^* = X^*\beta_0 + \epsilon^*$$
$$X^* = Z\tilde{\Pi}_m(\hat{\beta}) + V^*$$

and compute $\hat{\beta}_m^*$.

To motivate the MRE1 procedure, let us first consider the modification introduced in Eqs. (12) and (13). Since the bootstrap sample generated under $\tilde{\Pi}(\beta_0)$ cannot mimic well the original instrument strength and results in an approximation error of order at least as large as the concentration parameter, we introduce $\tilde{\Psi}^{-1/2}(\beta_0)\tilde{\Psi}_m^{1/2}(\beta_0)$ as a correction factor to remove the "extra"

[5] When $k > 1$, we require that the obtained matrix is positive semi-definite so that the (matrix) square root exists.

instrument strength. For example, we can show that under the null $H_0 : \beta = \beta_0$, the key parameter Ψ in the limiting distribution of LIML or FULL is consistently estimated by the MRE1 bootstrap analogue in both Cases (I) and (II):

$$\frac{\tilde{\Pi}_m'(\beta_0)Z'Z\tilde{\Pi}_m(\beta_0)}{r_n} = \frac{\tilde{\Pi}'(\beta_0)Z'Z\tilde{\Pi}(\beta_0)}{r_n} - \left(\frac{l}{r_n}\right)\widehat{\Sigma}_{\tilde{V}\tilde{V}}(\beta_0) \to^P \Psi.$$

Under this modification, the MRE1 procedure ensures that the bootstrap Hessian term correctly converges to H under P^*. Similar result holds for the bootstrap Jacobian term. Moreover, it ensures that the convergence rate of the estimator is correctly replicated by the bootstrap analogue in both Cases (I) and (II). Note that these arguments also hold for MRE2 procedure since $\hat{\beta}$ is consistent under the current many/many weak instrument sequences.

Now consider the bootstrap disturbances generated by the two modified procedures. With our approach of re-scaling the residuals into $\left(\sqrt{\frac{n}{n-l}}M_Z\epsilon(\beta_0), \sqrt{\frac{n}{n-l}}\hat{V}\right)$, the MRE bootstraps are able to mimic well each component of the covariance matrix of $\left(\epsilon_i, V_i'\right)'$ even when $\lambda \neq 0$:

$$E^*\left(\epsilon_i^{*2}\right) = \frac{n}{n-l}\left(\frac{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}{n}\right) \to^P \sigma_{\epsilon\epsilon}$$

$$E^*\left(V_i^*\epsilon_i^*\right) = \frac{n}{n-l}\left(\frac{V'M_Z\epsilon(\beta_0)}{n}\right) \to^P \sigma_{V\epsilon}$$

$$E^*\left(V_i^*V_i^{*'}\right) = \frac{n}{n-l}\left(\frac{V'M_ZV}{n}\right) \to^P \Sigma_{VV}$$

under H_0, where E^* denote the expectation under the MRE bootstrap measure. Here, $\sqrt{\frac{n}{n-l}}$ is employed to both the structural-form and the reduced-form residuals to account for the limiting behavior of the M_Z matrix under a large number of instruments.

Remarks. 1. Our correction factor in the MRE1 procedure is related to the restricted-efficient-corrected (REC) bootstrap in Davidson and MacKinnon (2008, p. 458). The REC bootstrap is motivated by the fact that $\tilde{a}^2 = \tilde{\Pi}'(\beta_0)Z'Z\tilde{\Pi}(\beta_0)/\left(n^{-1}\tilde{V}'(\beta_0)\tilde{V}(\beta_0)\right)$, their RE-based estimator of the concentration parameter (in $k = 1$ case), is inconsistent under Staiger and Stock (1997)'s weak instrument asymptotics, and has a bias of $l \times \left(1 - \sigma_{V\epsilon}^2/\sigma_{\epsilon\epsilon}\right)$. They propose to construct an unbiased estimator of the concentration parameter using $\tilde{a}_{BC}^2 = \left(0, \tilde{a}^2 - l(1-\tilde{\rho}^2)\right)^+$ where $\tilde{\rho}^2 = \epsilon'(\beta_0)\tilde{V}(\beta_0)/\left\{\left(\epsilon'(\beta_0)\epsilon(\beta_0)\right)\left(\tilde{V}'(\beta_0)\tilde{V}(\beta_0)\right)\right\}^{1/2}$. Then, one can generate the bootstrap d.g.p. using $\tilde{\Pi}_{BC}(\beta_0) = \tilde{\Pi}(\beta_0)\left(\frac{\tilde{a}_{BC}}{\tilde{a}}\right)$. Under current many/many weak instrument sequences, it can be shown that in Case (I)

$$\frac{\tilde{\Pi}_{BC}'(\beta_0)Z'Z\tilde{\Pi}_{BC}(\beta_0)}{r_n} \to^P \Psi + \lambda\gamma\sigma_{\tilde{V}\tilde{V}}.$$

Thus, the REC bootstrap will be inconsistent when $\lambda \neq 0$. In Case (II), due to this inconsistency $\tilde{\Pi}_{BC}'(\beta_0)Z'Z\tilde{\Pi}_{BC}(\beta_0)/r_n$ will diverge to infinity, leading the REC-based bootstrap analogue of LIML to also converge too fast.

2. An alternative modified procedure based on $\hat{\beta}$ could also be pursued. More precisely, it amounts to using $\tilde{\Pi}_m(\hat{\beta})$ as in the MRE2 procedure and to using $\hat{\epsilon}$ instead of $\epsilon(\beta_0)$ when generating the bootstrap disturbances. One can show that this procedure is asymptotically equivalent to the MRE2 procedure. However, trial simulation shows that finite-sample size control will be compromised under this procedure.

Before giving the asymptotic results for the MRE bootstraps, we introduce some additional assumptions, which simplify the variance formula of LIML or FULL.

Assumption 3. (i) $\lambda_n \to \lambda \neq 0$ and $n^{-1}\sum_{i=1}^n |P_{ii} - \lambda_n| \to 0$ as $n \to \infty$.
(ii) $\lambda_n \to \lambda = 0$ as $n \to \infty$.

Assumption 3(i) is also used in Anatolyev and Gospodinov (2011) for many instrument sequences and in Anatolyev (2012) for many regressor sequences. As pointed out in their papers, this condition allows that l increases at the same rate as n but requires that (almost) all diagonal elements of the projection matrix P_Z converge to λ (note that under conventional asymptotics they converge to zero). The latter is expected to hold if the instruments $z_i, i = 1, \ldots, n$ are drawn independently from some distribution (i.e., z_i are homogeneous across i; Anatolyev and Gospodinov, 2011, p. 429; Anatolyev, 2012, p. 375). They also point out that Assumption 3(i) will typically fail with heterogeneous instruments, as the diagonal elements of P_Z need not be centered at the same value. On the other hand, Assumption 3(ii) requires that l grows at a slower rate than n. This case is most important in empirical applications, especially in microeconomic studies, where the number of instruments is usually small relative to the sample size. For example, in their study of return to schooling problem, Donald and Newey (2001) and Hansen et al. (2008) used 180 instruments with a sample size of 329, 509. The variance formula in Theorem 2.1 will be simplified under either condition, as the non-normality adjustment terms A and B will disappear.

These conditions are relatively mild because the effects of A and B on the distributions of the estimators will typically be quite small compared with other terms, as emphasized by Hansen et al. (2008) and Anderson et al. (2010, 2011). Under any condition of Assumption 3, the MRE procedures are able to consistently estimate the limiting distribution of interest. We also briefly discuss how Assumption 3 would affect the standard and RE(C) bootstraps. Although A = B = 0 under these conditions, the standard and RE bootstraps will remain inconsistent because of the bootstrap distortions related to the concentration parameter. In addition, the standard bootstrap cannot consistently estimate $\Sigma_{\tilde{V}\tilde{V}}$ under Assumption 3(i). On the other hand, as noted in the Remark, the distortion of the REC bootstrap will vanish if $l/n \to 0$. Thus, although also inconsistent under Assumption 3(i), the REC bootstrap will be consistent under Assumption 3(ii).

The asymptotic distributional results for the MRE1/MRE2-based bootstrap analogues of LIML or FULL are stated in the following theorem.

Theorem 3.3. *Suppose that Assumptions 1–2 hold, also suppose that either Assumption 3(i) or Assumption 3(ii) holds, then under $H_0 : \beta = \beta_0$, in Case (I),*

$$\sup_{x\in R}\left|P^*\left(\sqrt{r_n}(\hat{\beta}_m^* - \beta_0) \le x\right) - P\left(\sqrt{r_n}(\hat{\beta} - \beta_0) \le x\right)\right| \to^P 0$$

and in Case (II),

$$\sup_{x\in R}\left|P^*\left(\frac{r_n}{\sqrt{l}}(\hat{\beta}_m^* - \beta_0) \le x\right) - P\left(\frac{r_n}{\sqrt{l}}(\hat{\beta} - \beta_0) \le x\right)\right| \to^P 0$$

where P^* denotes the probability measure induced by the MRE bootstrap procedures.

Theorem 3.3 shows that contrary to the standard/RE bootstraps, the MRE bootstraps estimate consistently both the limiting distribution of $\sqrt{r_n}(\hat{\beta} - \beta)$ in Case (I) and the limiting distribution of $\frac{r_n}{\sqrt{l}}(\hat{\beta} - \beta)$ in Case (II). In particular, they correctly replicate the convergence rate of the estimators in Case (II) since "pseudo" identification strength has been removed from the bootstrap sample.

3.4. Bootstrapping t-test with corrected standard error

In view of the success of the MRE procedures in providing distributional approximation for LIML or FULL, we can expect that the distribution of t_{cse}, t-test statistic based on the CSE, is also well approximated by our bootstrap procedures. Moreover, since t_{cse} is asymptotically standard normal under many/many weak instrument asymptotics, folklore suggests that the standard and RE bootstraps should also be capable of consistently estimating its distribution even if these bootstrap procedures cannot adequately mimic the limit distribution of LIML or FULL. This conjecture turns out to be correct, because one can show that in Case (I),

$$r_n\widehat{\Lambda}^*_{std} \approx Var^*\left[\sqrt{r_n}\left(\hat{\beta}^*_{std}-\hat{\beta}\right)\right] \quad \text{and}$$
$$r_n\widehat{\Lambda}^*_{re} \approx Var^*\left[\sqrt{r_n}\left(\hat{\beta}^*_{re}-\beta_0\right)\right]$$

with "$A \approx B$" being shorthand for $A^{-1}B \rightarrow^{P^*} I_k$ in probability and Var^* denoting the variance computed under the corresponding bootstrap measure. $\widehat{\Lambda}^*_{std}$ and $\widehat{\Lambda}^*_{re}$ denote the variance estimators computed using the pseudo-data generated by the standard and RE bootstrap d.g.p., respectively. Then, the result of weak convergence in probability for the bootstrap analogues of t_{cse} can be established, i.e.,

$$\frac{c'(\hat{\beta}^*_{std}-\hat{\beta})}{\sqrt{c'\widehat{\Lambda}^*_{std}c}} = \frac{c'\sqrt{r_n}(\hat{\beta}^*_{std}-\hat{\beta})}{\sqrt{c'r_n\widehat{\Lambda}^*_{std}c}} \rightarrow^{d^*} N(0,1)$$

in probability and similar result holds for the RE bootstrap. Analogously, we have for Case (II)

$$l\widehat{\Lambda}^*_{std} \approx Var^*\left[\sqrt{l}\left(\hat{\beta}^*_{std}-\hat{\beta}\right)\right] \quad \text{and}$$
$$l\widehat{\Lambda}^*_{re} \approx Var^*\left[\sqrt{l}\left(\hat{\beta}^*_{re}-\beta_0\right)\right].$$

Thus, these bootstrap-based approximations to t_{cse} converge to standard normal distribution in probability, regardless of the fact that in Case (II) the convergence rate of $\hat{\beta}^*_{std}$ and $\hat{\beta}^*_{re}$ differs from that of $\hat{\beta}$. More precisely, an application of the continuous mapping theorem for weak convergence in probability yields the following result.

Theorem 3.4. *Suppose that Assumptions 1–2 hold, also suppose that $H_0:\beta=\beta_0$ holds for the RE and MRE bootstraps, then*

$$\sup_{x\in R}\left|P^*\left(t^*_{cse,j}\le x\right)-P\left(t_{cse}\le x\right)\right| \rightarrow^{P} 0$$

where $j\in\{s,r,m\}$, $t^*_{cse,s}=c'(\hat{\beta}^*_{std}-\hat{\beta})/\sqrt{c'\widehat{\Lambda}^*_{std}c}$, $t^*_{cse,r}=c'(\hat{\beta}^*_{re}-\beta_0)/\sqrt{c'\widehat{\Lambda}^*_{re}c}$, and $t^*_{cse,m}=c'(\hat{\beta}^*_{m}-\beta_0)/\sqrt{c'\widehat{\Lambda}^*_{m}c}$; P^* denotes the probability measure induced by the corresponding bootstrap method and c are the linear combination coefficients.

Theorem 3.4 gives asymptotic validity for percentile-t type bootstrap methods based on the standard, RE and MRE procedures. Thus, all these bootstrap procedures are expected to have reasonable performances for the purpose of implementing hypothesis testing or constructing CIs. However, the bootstrap is able to provide asymptotic refinement only when the test statistic is asymptotically pivotal and when the bootstrap d.g.p. consistently estimates the original d.g.p. (see, Beran (1988)). Among the standard, RE and MRE bootstrap procedures, one can only expect the MRE bootstraps to provide asymptotic refinement for percentile-t type bootstrap methods since the other two procedures are not able to consistently estimate the original d.g.p., as has been shown previously. We leave a formal study of the MRE bootstraps' higher order property for future work.

4. Simulation results

The goal of this section is to evaluate the finite sample performance of the bootstrap methods studied in the previous sections. Following Davidson and MacKinnon (2008, 2010, 2014), we use the following d.g.p.

$$y = \beta X + \varepsilon$$
$$X = aw + v,$$

where $\beta = 1$, and the $n\times 1$ vector w is normally distributed with mean zero and scaled so that $\|w\|^2 = 1$. In addition, the model is estimated using the $n\times l$ instrument matrix Z, of which one vector is w and the others are standard normal random variables that have no explanatory power. As pointed out in their paper, this setting allows one to measure the strength of the instruments using the parameter a, the square of which equals the concentration parameter. For the disturbances, we set

$$\epsilon = r\epsilon_1 + \rho\epsilon_2$$
$$v = \epsilon_2,$$

with $(\epsilon_1,\epsilon_2)'\sim N(0,\mathbf{I})$, $r^2+\rho^2=1$.[6]

In Figs. 1–8, we present empirical rejection frequencies for asymptotic and bootstrap tests at the 5% significance level. The simulation evidences are based on 1000 replications and $B = 399$ bootstrap samples. The sample size is 100 and we use the LIML or the FULL estimator throughout the simulation. For all bootstrap procedures, we consider both percentile and percentile-t type Wald tests. Notice that we present the properties of percentile type bootstrap tests for two reasons: (i) to see whether the bootstrap inconsistency of the standard/RE procedures predicted by our asymptotic theories can be confirmed by simulation results; (ii) to see whether the MRE bootstrap procedures are able to mimic well the distribution of LIML or FULL in finite samples and thus provide valid inference methods.

Let us start with the standard residual bootstrap procedure. Percentile type bootstrap Wald tests reject $H_0:\beta=\beta_0$ if $|\hat{\beta}-\beta_0|$ is among the $0.05(B+1)$ biggest values in

$$\left\{|\hat{\beta}-\beta_0|,\ |\hat{\beta}^*_{j,1}-\hat{\beta}|,\ \ldots,\ |\hat{\beta}^*_{j,B}-\hat{\beta}|\right\} \tag{14}$$

where $j\in\{std\}$. Notice that normally, we should reject $H_0:\beta=\beta_0$ if $\sqrt{r_n}|\hat{\beta}-\beta_0|$ is among the $0.05(B+1)$ biggest values in

$$\left\{\sqrt{r_n}|\hat{\beta}-\beta_0|,\ \sqrt{r_n}|\hat{\beta}^*_{j,1}-\hat{\beta}|,\ \ldots,\ \sqrt{r_n}|\hat{\beta}^*_{j,B}-\hat{\beta}|\right\}$$

in Case (I) and reject $H_0:\beta=\beta_0$ if $\frac{r_n}{\sqrt{l}}|\hat{\beta}-\beta_0|$ is among the $0.05(B+1)$ biggest values in

$$\left\{\frac{r_n}{\sqrt{l}}|\hat{\beta}-\beta_0|,\ \frac{r_n}{\sqrt{l}}|\hat{\beta}^*_{j,1}-\hat{\beta}|,\ \ldots,\ \frac{r_n}{\sqrt{l}}|\hat{\beta}^*_{j,B}-\hat{\beta}|\right\}$$

in Case (II). However, although we do not know the exact value of r_n in practice, we are still able to use the procedure described by (14) since $\sqrt{r_n}$ and $r_n/\sqrt{l}$ will be canceled out in Cases (I) and (II), respectively. For percentile-t type standard bootstrap Wald tests, we reject $H_0:\beta=\beta_0$ if $\frac{|\hat{\beta}-\beta_0|}{\sqrt{\widehat{\Lambda}}}$ is among the $0.05(B+1)$ biggest values in

$$\left\{\frac{|\hat{\beta}-\beta_0|}{\sqrt{\widehat{\Lambda}}},\ \frac{|\hat{\beta}^*_{j,1}-\hat{\beta}|}{\sqrt{\widehat{\Lambda}^*_{j,1}}},\ \ldots,\ \frac{|\hat{\beta}^*_{j,B}-\hat{\beta}|}{\sqrt{\widehat{\Lambda}^*_{j,B}}}\right\}$$

[6] We also tried other simulation settings such as those in Anatolyev and Gospodinov (2011) and Newey and Windmeijer (2009), the results were very similar.

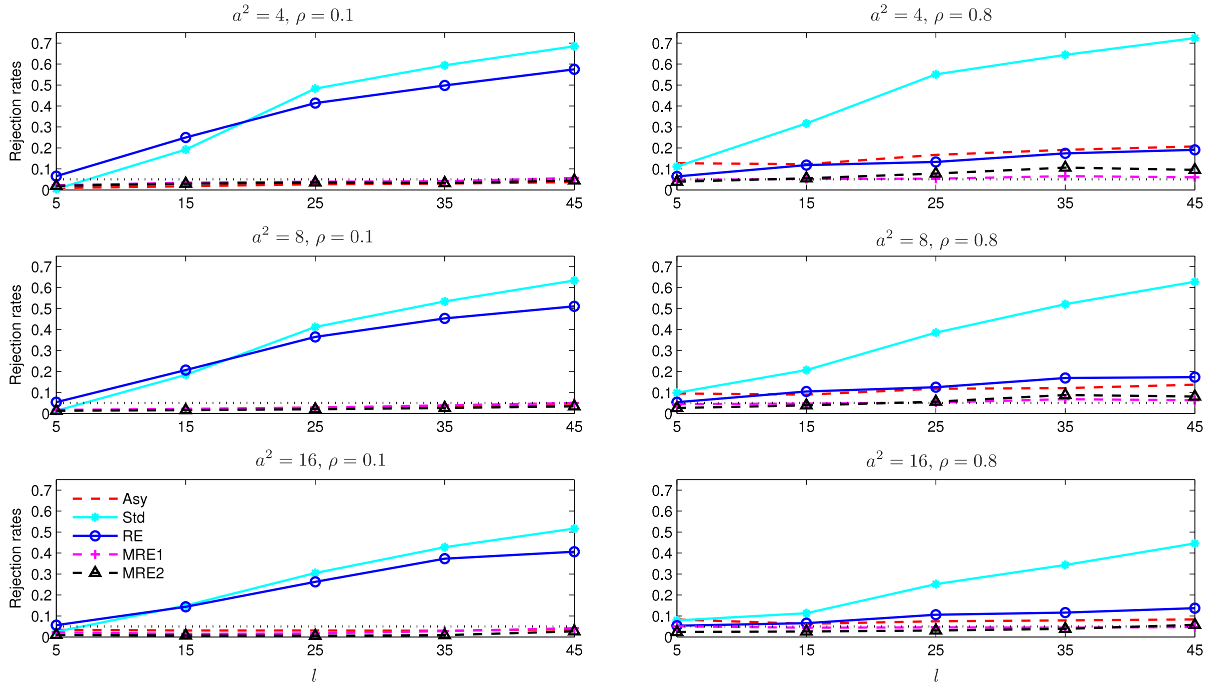

Fig. 1. Rejection rates for percentile type bootstrap Wald tests as a function of l, LIML.

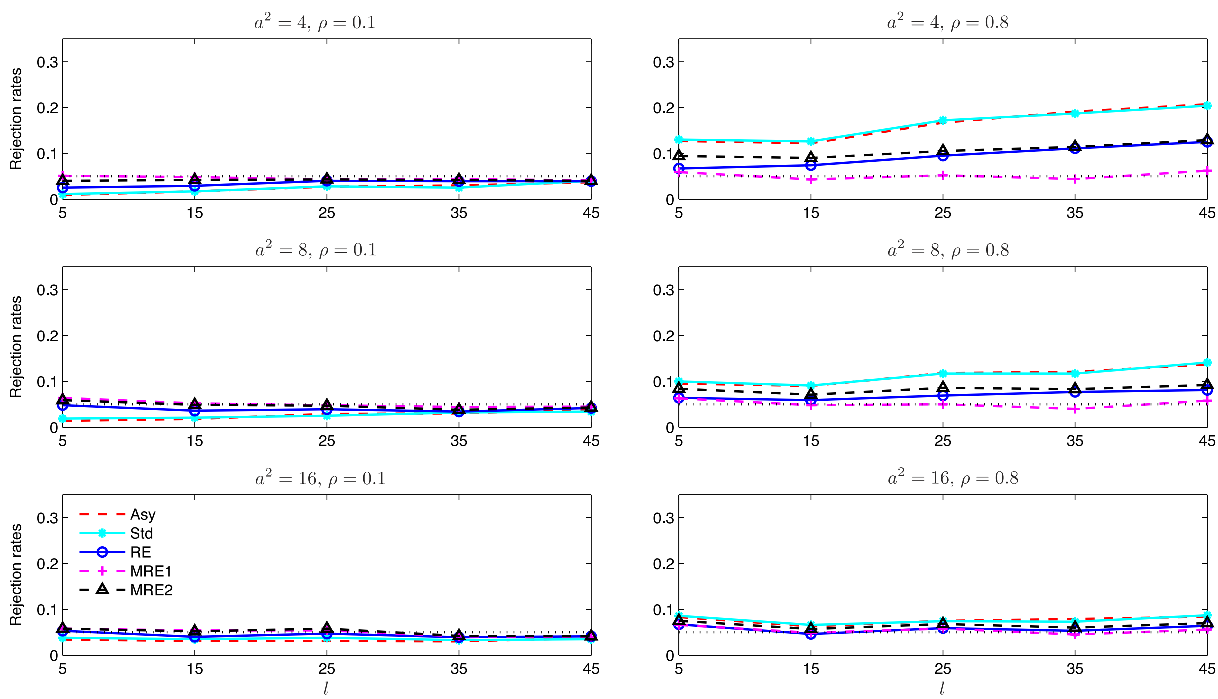

Fig. 2. Rejection rates for percentile-t type bootstrap Wald tests as a function of l, LIML.

where $j \in \{std\}$. In this formula, $\sqrt{\widehat{\Lambda}}$ is the CSE and $\sqrt{\widehat{\Lambda}^*_{std}}$ is its standard bootstrap counterpart.

Also, since bootstrap data of the RE and MRE procedures are generated under the null, percentile type bootstrap Wald tests reject $H_0 : \beta = \beta_0$ if $|\hat{\beta} - \beta_0|$ is among the $0.05(B + 1)$ biggest values in

$$\left\{|\hat{\beta} - \beta_0|, |\hat{\beta}^*_{j,1} - \beta_0|, \ldots, |\hat{\beta}^*_{j,B} - \beta_0|\right\},$$

where $j \in \{re, m\}$. Percentile-t type bootstrap Wald tests reject $H_0 : \beta = \beta_0$ if $\frac{|\hat{\beta}-\beta_0|}{\sqrt{\widehat{\Lambda}}}$ is among the $0.05(B + 1)$ biggest values in

$$\left\{\frac{|\hat{\beta} - \beta_0|}{\sqrt{\widehat{\Lambda}}}, \frac{|\hat{\beta}^*_{j,1} - \beta_0|}{\sqrt{\widehat{\Lambda}^*_{j,1}}}, \ldots, \frac{|\hat{\beta}^*_{j,B} - \beta_0|}{\sqrt{\widehat{\Lambda}^*_{j,B}}}\right\}$$

where $j \in \{re, m\}$.

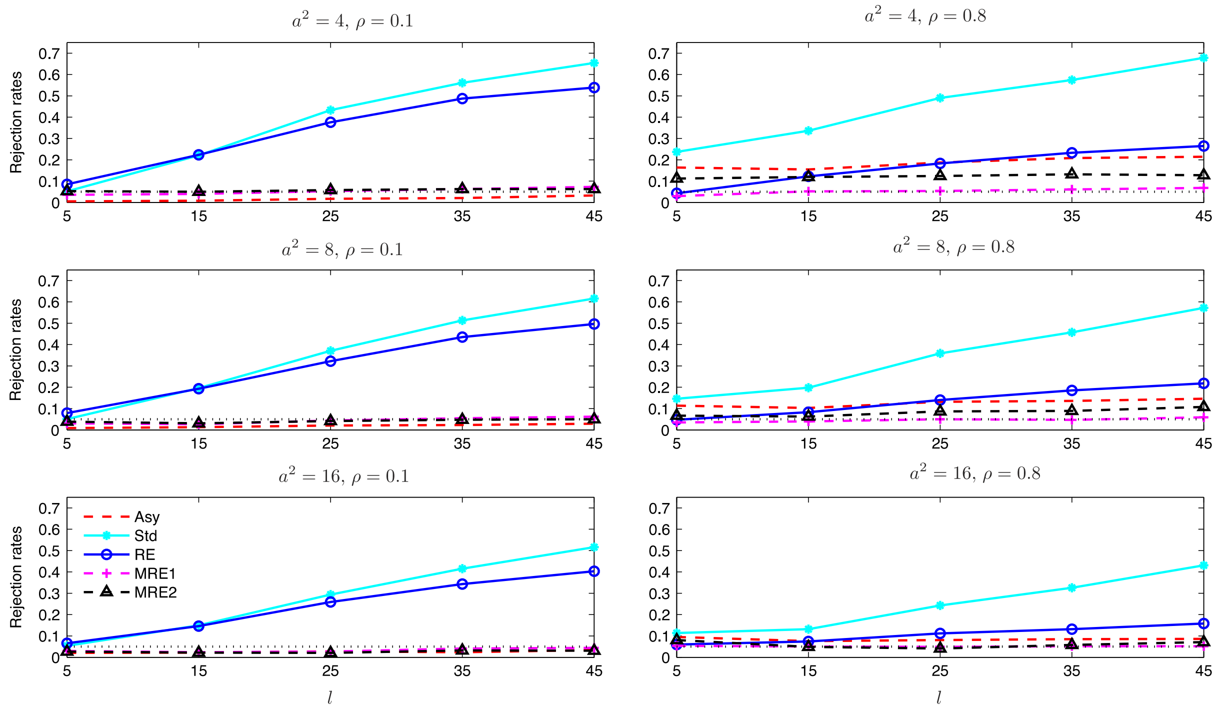

Fig. 3. Rejection rates for percentile type bootstrap Wald tests as a function of *l*, FULL.

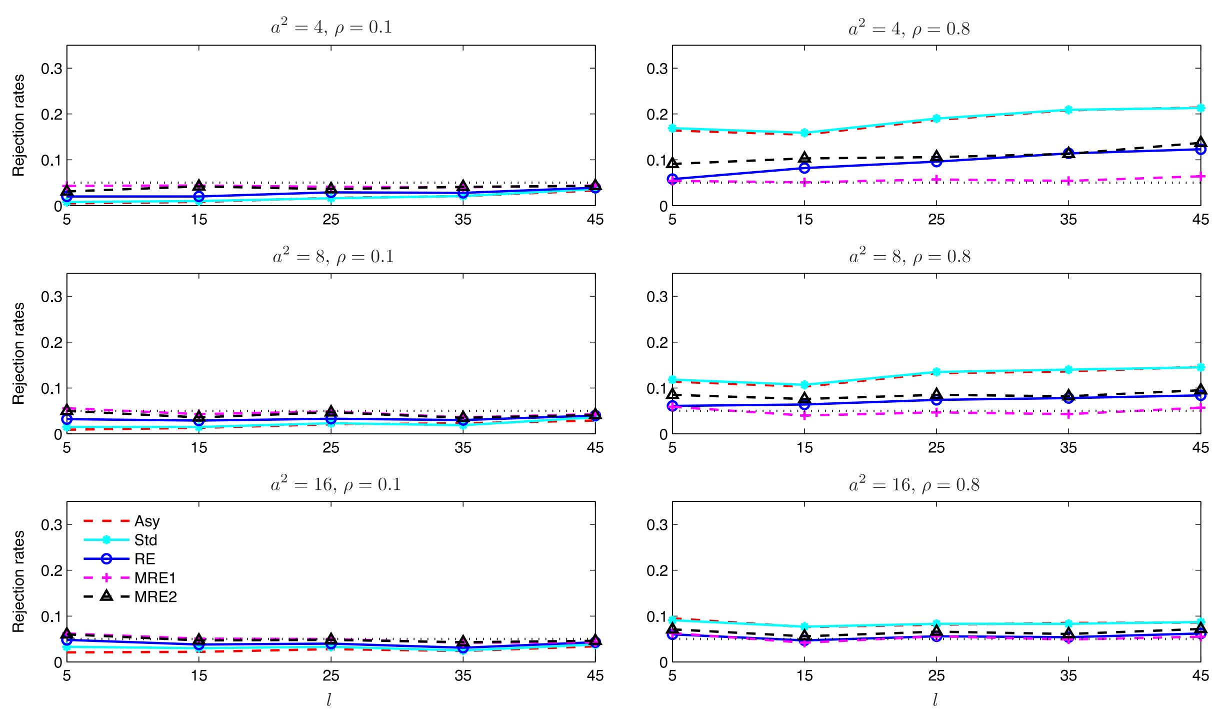

Fig. 4. Rejection rates for percentile-*t* type bootstrap Wald tests as a function of *l*, FULL.

Note that while standard Wald-type CIs could easily be constructed using the procedures that do not impose the null hypothesis (e.g., the standard residual bootstrap), it could be computationally demanding to construct bootstrap CIs using the procedures that impose the null (e.g., the RE, MRE1 and MRE2 bootstraps) because one has to invert the bootstrap test. Davidson and MacKinnon (2014) has a detailed discussion on this issue and presents a method of finding the upper and lower limits of the null imposed bootstrap CIs.

Figs. 1–4 each contains six plots and pertain to percentile and percentile-*t* type bootstrap tests applied to LIML or FULL. They show the effect of varying the number of instruments for three values of a^2 and two values of ρ. Specifically, we vary a^2 across rows ($a^2 \in \{4, 8, 16\}$) and ρ across columns ($\rho \in \{0.1, 0.8\}$). One can interpret $a^2 = 4$ as a very weak instruments case, $a^2 = 8$ as a weak instruments case and $a^2 = 16$ as a moderately strong instruments case. When $\rho = 0.1$, there is not much correlation between the structural and reduced-form disturbances; when $\rho = 0.8$, there is a great deal of correlations. Note that we have different vertical scales for the figures of percentile tests (from 0 to 0.75) and those for percentile-*t* tests (from 0 to 0.35), because otherwise it would be impossible to see many important differences

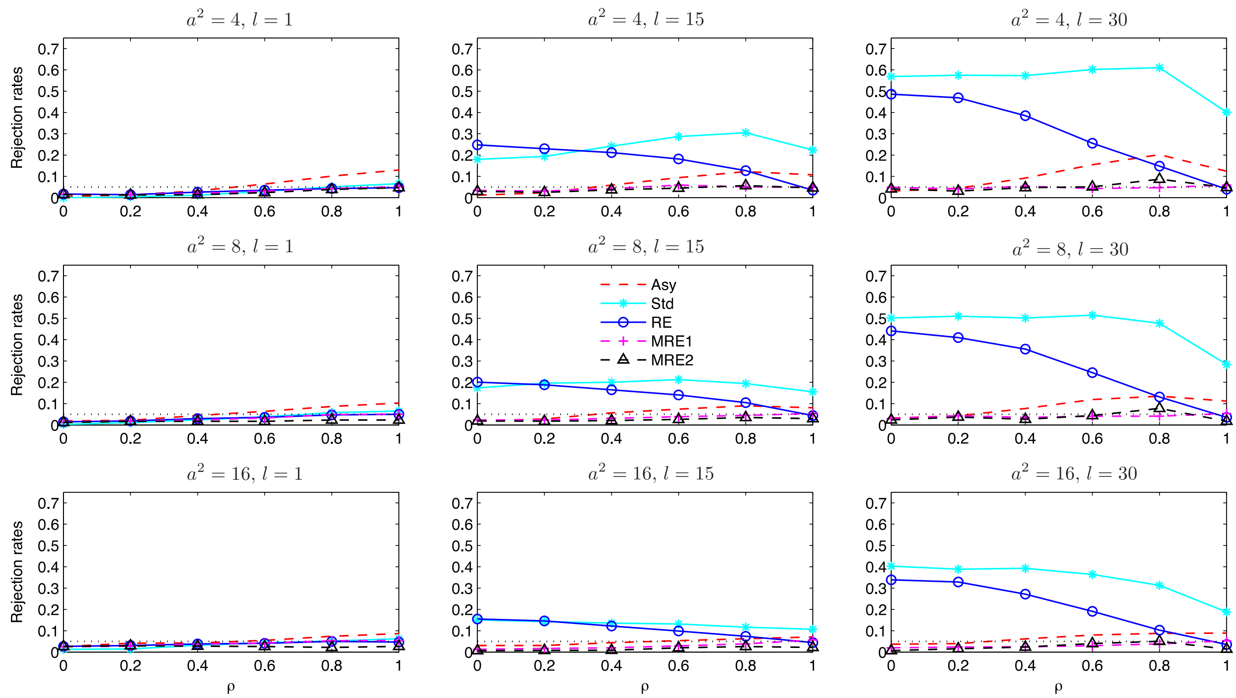

Fig. 5. Rejection rates for percentile type bootstrap Wald tests as a function of ρ, LIML.

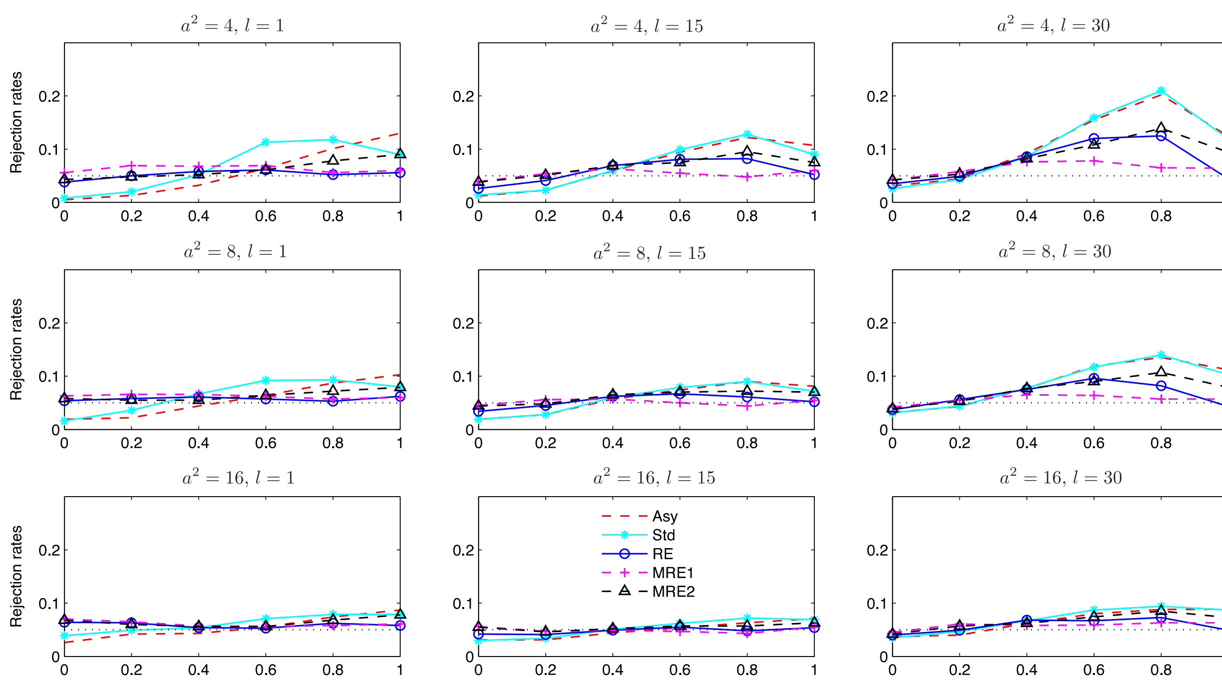

Fig. 6. Rejection rates for percentile-t type bootstrap Wald tests as a function of ρ, LIML.

between alternative methods. We highlight some key findings below.

1. As in the simulation results reported in Davidson and MacKinnon (2008),[7] CSE-based asymptotic Wald (t) test for LIML underrejects when ρ is small and overreject when ρ is large. In particular, asymptotic Wald (t) test has noticeable finite sample distortions for $\rho = 0.8$ and $a^2 = 4$. Indeed, under this setting, the actual rejection rates of the asymptotic Wald (t) test for LIML (Fig. 1) vary between 15% and 20% for values of l between 25 and 45. Similar distortion can be observed for the FULL case (Fig. 3).
2. Figs. 1 and 3 show clearly that percentile bootstrap test based on the standard residual or the RE bootstrap overrejects with many/many weak instruments. Also, it turns out that the distortions of both bootstrap tests increase when the strength of the instruments decreases and/or when the number of instruments increases. Thus, our results in Theorems 3.1 and 3.2 give an excellent approximation to the finite sample behavior

7 See, e.g., Figs. 2 and 4 in Davidson and MacKinnon (2008)

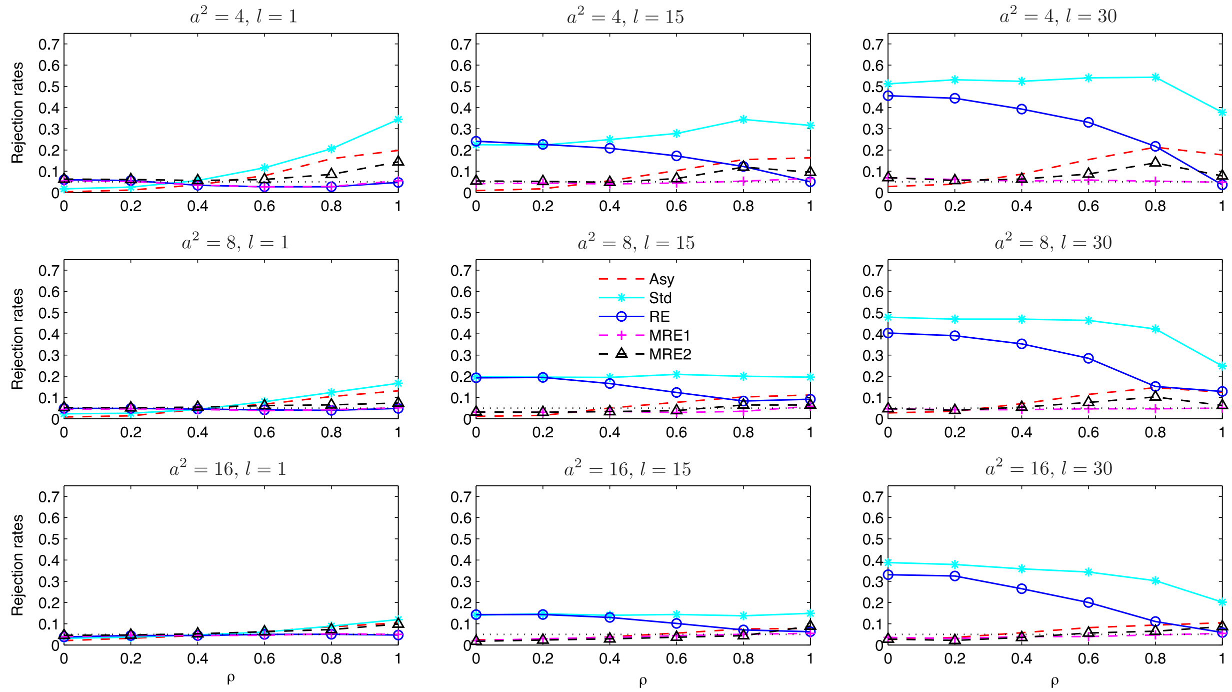

Fig. 7. Rejection rates for percentile type bootstrap Wald tests as a function of ρ, FULL.

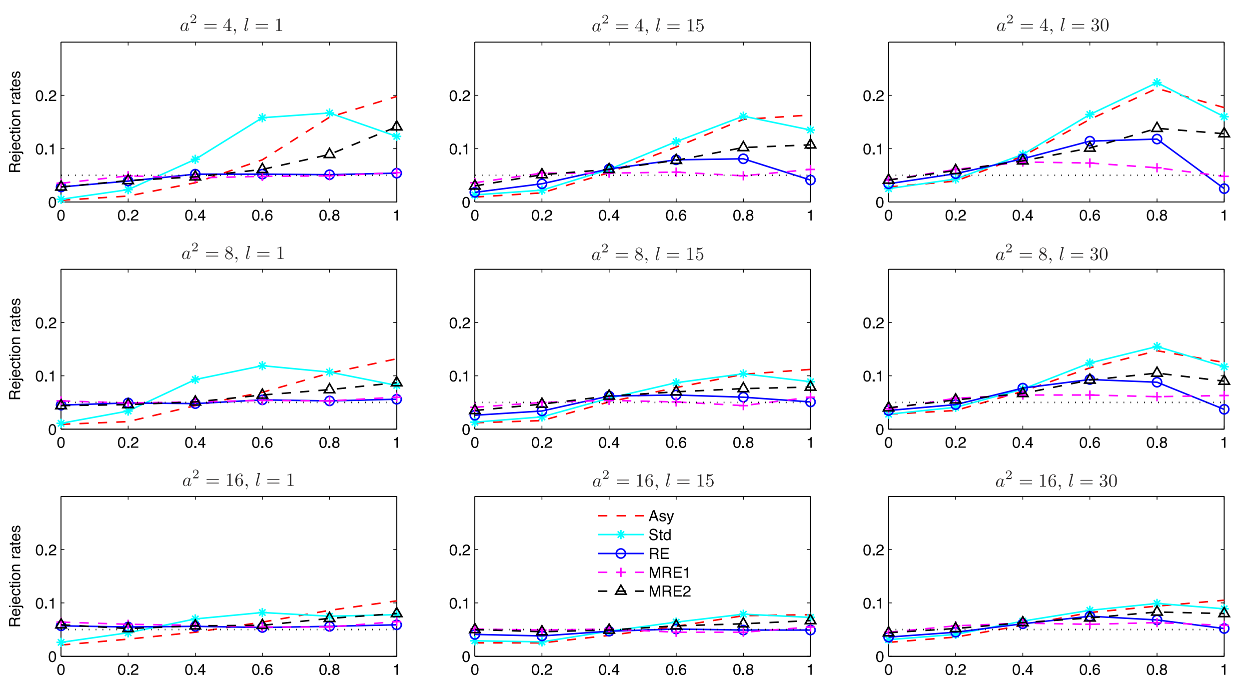

Fig. 8. Rejection rates for percentile-t type bootstrap Wald tests as a function of ρ, FULL.

of these bootstrap procedures. Furthermore, we find that the RE bootstrap tends to dominate the standard bootstrap in terms of size control, confirming our theoretical predictions in Section 3.2 that the RE bootstrap-based approximation of the distribution of LIML or FULL is typically more precise than the standard bootstrap-based approximation. On the other hand, the two MRE percentile bootstrap tests have much better performance for all values of a^2 and ρ. In particular, it is remarkable that in all plots, the MRE1 bootstrap displays very small distortions irrespective of the values of l. In addition, both MRE1 and MRE2 percentile bootstrap tests have improvement over CSE-based asymptotic Wald (t) tests when $\rho = 0.8$.

3. Figs. 2 and 4 show that rejection frequencies of the standard/RE bootstrap percentile-t tests are much better than their corresponding percentile versions. These results are in line with Theorem 3.4 which predicts in particular that percentile-t approximations based on these two bootstrap procedures are asymptotically valid even if their percentile counterparts are not. Also, we find that under the null hypothesis, the standard bootstrap test has almost the same performance as the CSE-based asymptotic Wald (t) test for all configurations of a^2, ρ and

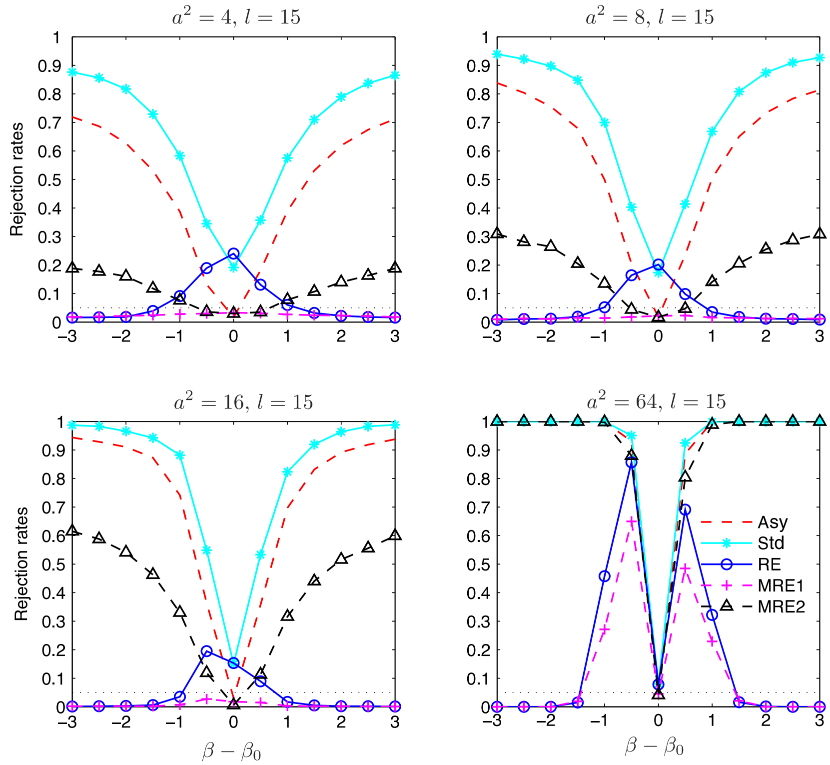

Fig. 9. Power of percentile type bootstrap Wald tests for the LIML estimator, $\rho = 0.1$.

l. The RE and MRE2 bootstraps improve upon the asymptotic theory and the standard bootstrap, especially when $\rho = 0.8$, but is still notably distorted for small values of a^2. The MRE1 bootstrap has the best performance in terms of size control among all the procedures.

Figs. 5–8 each contains nine plots and pertain to percentile and percentile-t type bootstrap tests. They show the effect of varying the value of ρ for three values of a^2 and three values of l. We include the simulations for $l = 1$ to investigate the performance of various procedures when the number of instruments is small. Some comments are in order.

1. Figs. 5 and 7 show that when ρ is small, percentile type bootstrap tests based on the standard and RE bootstraps have poor rejection frequencies in comparison to CSE-based asymptotic Wald (t) tests. For large values of ρ, the distortion of CSE-based asymptotic Wald (t) tests become severe while the rejection frequencies of RE percentile bootstrap tests become better and even improve upon asymptotic Wald (t) tests in some cases. This is natural considering that our theoretical analysis in Section 3.2 (see Eq. (11)) shows that the RE bootstrap approximation error depends crucially on $\Sigma_{\tilde{V}\tilde{V}}$, which equals $1 - \rho^2$ in the current simulation setting. It also turns out that MRE percentile bootstrap tests perform much better than standard/RE percentile bootstrap tests, and improve upon asymptotic Wald (t) tests in many cases.
2. Similar to the findings in Davidson and MacKinnon (2008), our results in Figs. 6 and 8 show that the RE percentile-t test performs very well when $l = 1$. In this case, the MRE1 percentile-t test is somewhat distorted when a^2 and ρ are small, while the MRE2 percentile-t test is relatively distorted when ρ is large. With relatively strong instrument ($l = 1$ and $a^2 = 16$), the behaviors of the RE and MRE1 bootstraps become very similar. On the other hand, when l becomes larger (e.g., $l = 15$), the distortion of the RE bootstrap tends to increase and the MRE1 bootstrap tends to have the best performance. Moreover, all the three bootstrap methods have substantial improvement over the asymptotic theory-based test and the standard bootstrap percentile-t test, which are especially distorted when a^2 is small and/or when ρ is large.

Figs. 9–12 show results of power for bootstrap Wald tests for LIML with $l = 15$, while Figs. 13–16 present similar results for FULL with $l = 15$. To reduce the possible power loss associated with small values of B, we set $B = 599$ in these experiments. Some key results are highlighted below:

1. At first sight, percentile standard bootstrap test outperforms all other percentile bootstrap tests when the null is false but, as already emphasized above, it is unable to control the size and has severe over-rejection when $\beta = \beta_0$, especially when the concentration parameter is small relative to the number of instruments (Figs. 9, 11, 13 and 15).
2. The RE and MRE1 percentile bootstrap tests almost have no power against the null, even when the concentration parameter becomes relatively strong ($a^2 = 16$). In particular, when $\rho = 0.1$ (Figs. 9 and 13), although the RE percentile test tends to over-reject when $\beta = \beta_0$, the rejection frequencies become close to zero when $|\beta - \beta_0|$ becomes large. The MRE1 percentile test has rejection frequencies close to the nominal level under the null, but its behavior under the alternative is very similar to that of the RE percentile test. When $\rho = 0.8$, the power curves of the RE and MRE1 percentile tests are also quite irregular. This power problem is probably due to the fact that both procedures use null-restricted reduced-form estimators when generating bootstrap data.
3. Interestingly, the MRE2 percentile bootstrap test, which uses $\hat{\beta}$ instead of β_0 in the reduced-form estimator, outperforms the RE and MRE1 procedures largely, and its power curves become close to that of the asymptotic theory-based test as the concentration parameter increases. Moreover, MRE2-based percentile bootstrap test for FULL (Figs. 13 and 15) seems to have better power property than that for LIML (Figs. 9 and 11).

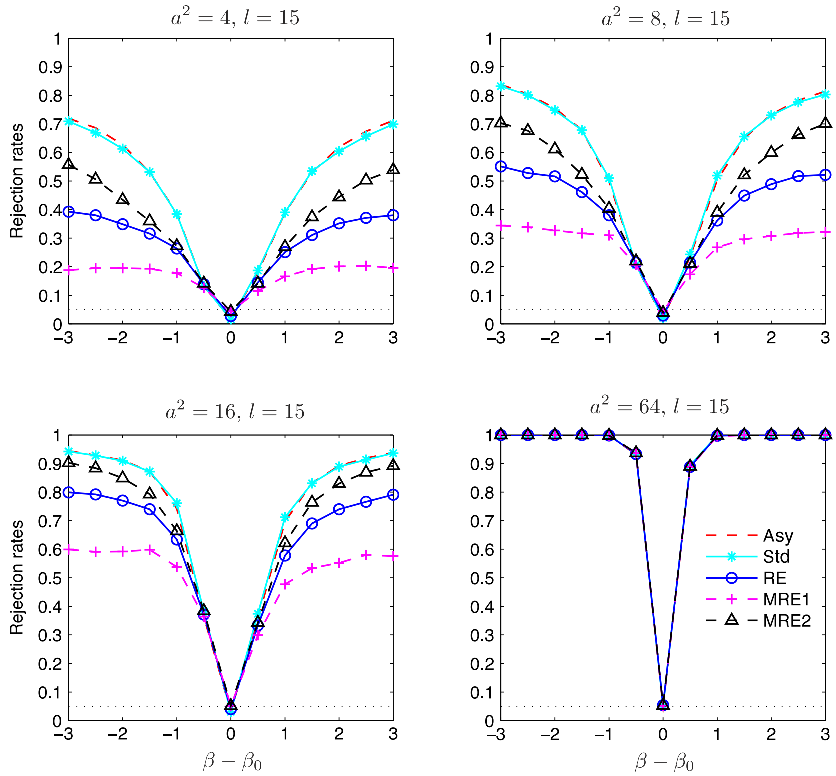

Fig. 10. Power of percentile-t type bootstrap Wald tests for the LIML estimator, $\rho = 0.1$.

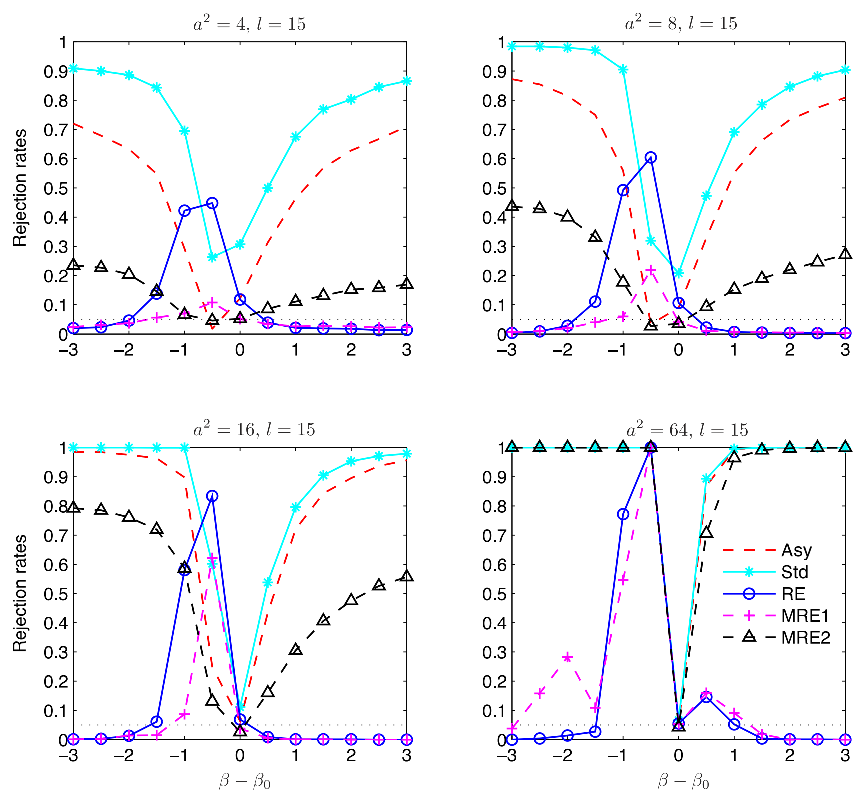

Fig. 11. Power of percentile type bootstrap Wald tests for the LIML estimator, $\rho = 0.8$.

Indeed, when $\rho = 0.8$, the MRE2-based percentile test for FULL even outperforms the asymptotic theory-based test in terms of power in many cases (Fig. 15). We conjecture that this could be related to the fact that LIML does not have moments and has large dispersions in finite samples.

4. Figs. 10 and 12 show the results of percentile-t type bootstrap tests for LIML. First, it turns out that the power curves of the asymptotic-theory based test and the standard bootstrap percentile-t test are almost identical. Second, the behaviors of the standard and RE percentile-t tests are more regular than

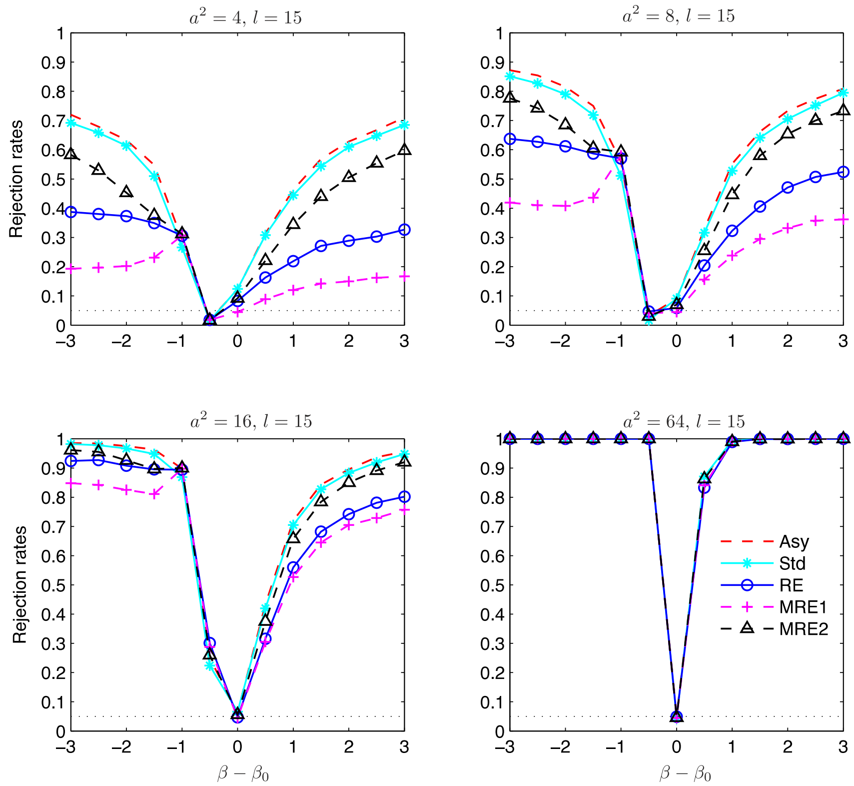

Fig. 12. Power of percentile-t type bootstrap Wald tests for the LIML estimator, $\rho = 0.8$.

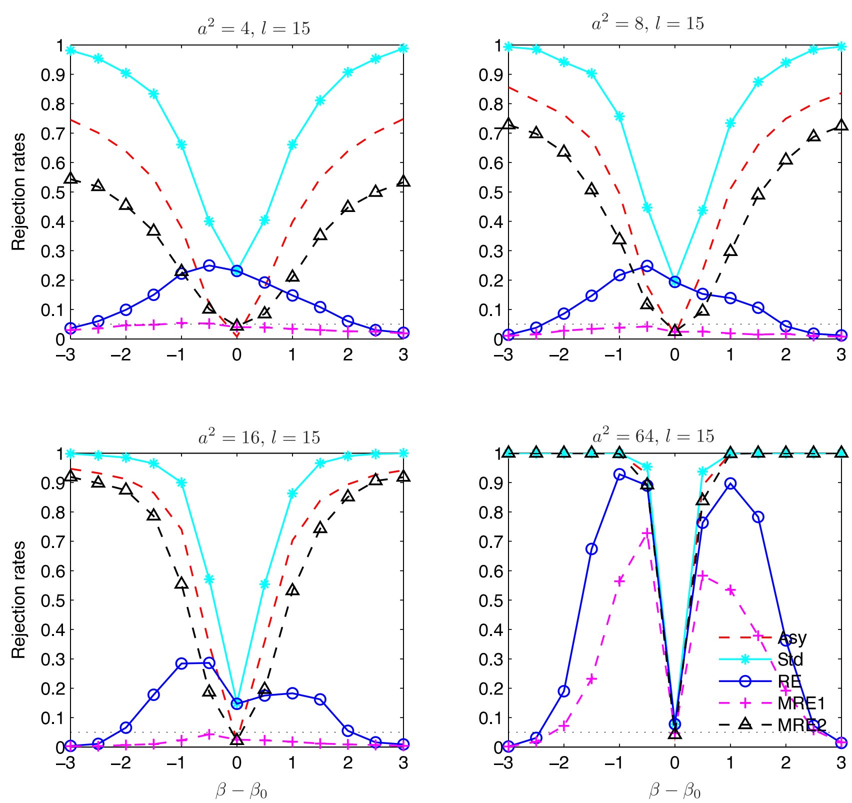

Fig. 13. Power of percentile type bootstrap Wald tests for the FULL estimator, $\rho = 0.1$.

their percentile counterparts. Furthermore, the RE and MRE1 percentile-t tests have nontrivial power against the null, and their power curves become closer to those of the other tests when the concentration parameter increases. Among the four bootstrap procedures, the standard bootstrap has the highest power, followed by the MRE2, RE, and MRE1 bootstraps. Figs. 14 and 16 show the results of percentile-t tests for FULL. We can see that the standard and MRE2 procedures have quite similar

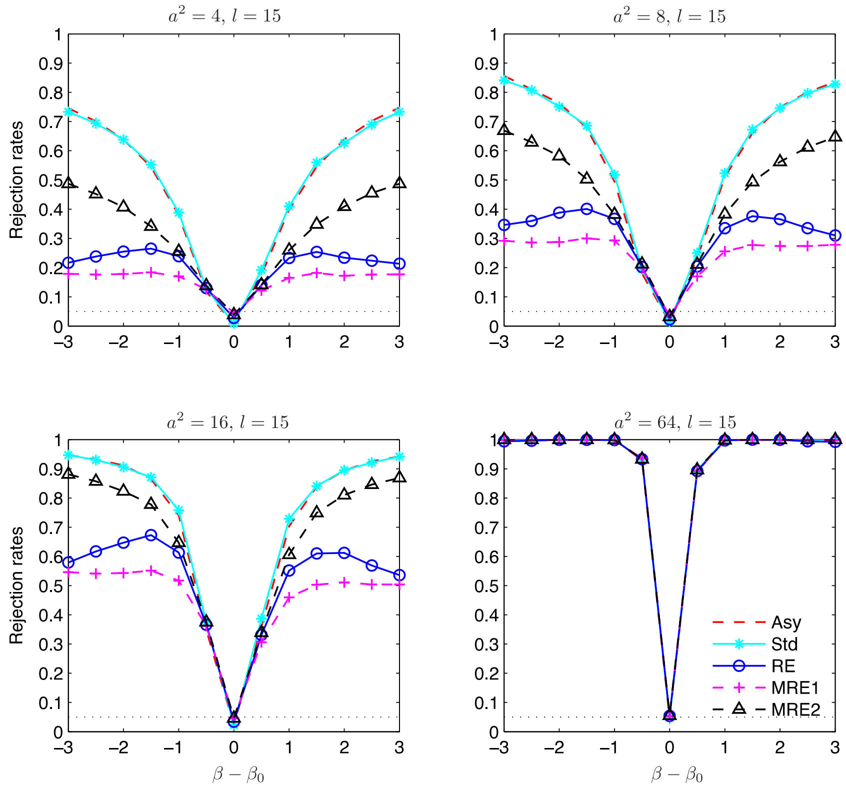

Fig. 14. Power of percentile-t type bootstrap Wald tests for the FULL estimator, $\rho = 0.1$.

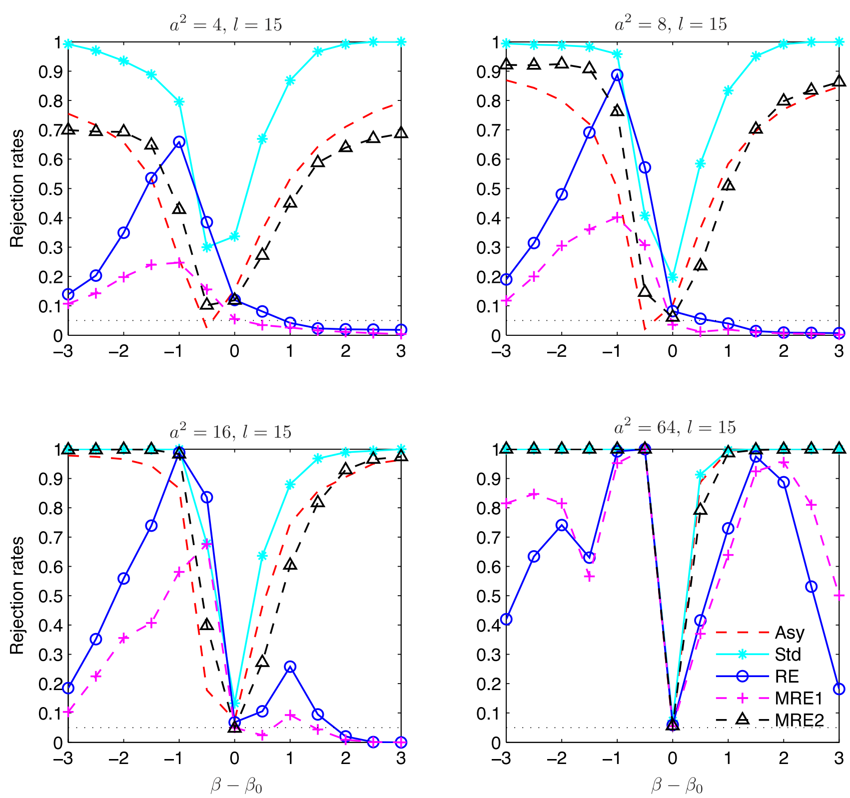

Fig. 15. Power of percentile type bootstrap Wald tests for the FULL estimator, $\rho = 0.8$.

behaviors as in the LIML case. On the other hand, the MRE1 procedure seems to have better power than the RE procedure when $|\beta - \beta_0|$ becomes large.

In summary, from Figs. 9–16, we find that percentile bootstrap tests based on null-restricted reduced-form estimators (i.e., RE and MRE1) typically do not have power. On the other hand, the MRE2

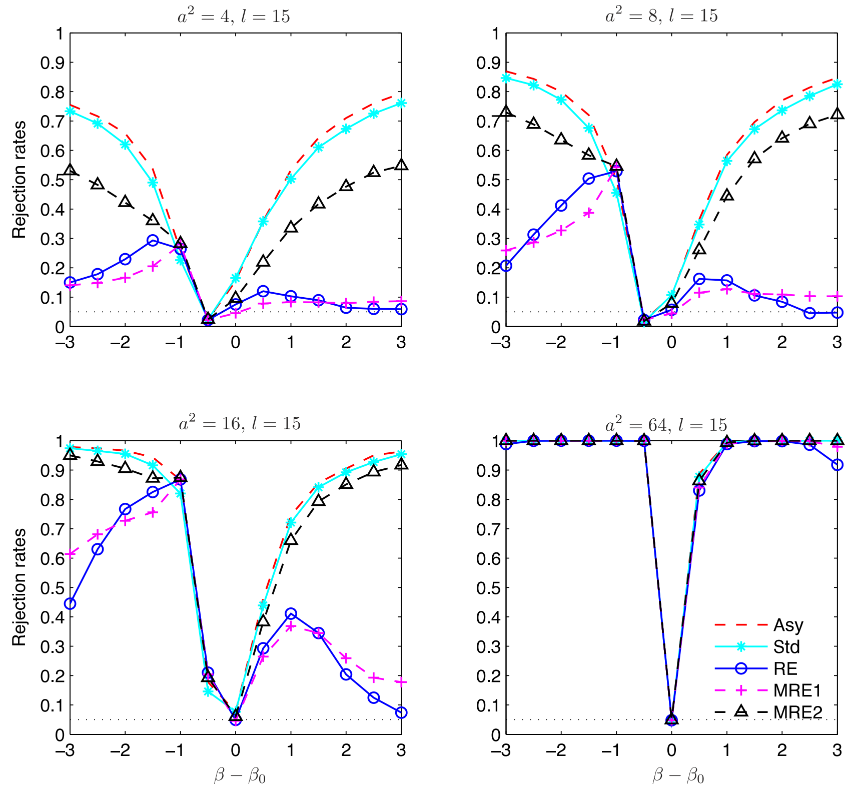

Fig. 16. Power of percentile-t type bootstrap Wald tests for the FULL estimator, $\rho = 0.8$.

procedure improves upon these two procedures in terms of power, and is especially powerful when applied to FULL. For percentile-t tests, the standard bootstrap performs best in terms of power but, as already observed in Figs. 6 and 8, it has severe size distortion when the degree of endogeneity is high. The MRE2 percentile-t test seems to be most balanced between size and power. Moreover, although less powerful than the other methods, the RE and MRE1 bootstraps could provide useful alternatives in the cases that the concentration parameter is quite small so that asymptotic-theory based test may have intolerable size distortions.

Figs. 17–24 show power results for bootstrap Wald tests for LIML and FULL with $l = 30$, and the behaviors of various bootstrap procedures are quite similar to those in the $l = 15$ case.

5. Conclusion

The main contribution of this paper is to study the validity of bootstrap methods for inference in linear IV regression when the available instruments may be weak and the number of instruments may be large. Using the asymptotic framework of many/many weak instruments, we obtain new theoretical results about the bootstrap methods that can be overlooked under the conventional asymptotic framework.

In particular, we show that a standard residual-based bootstrap method is unable to estimate the limiting distribution of LIML or FULL consistently. It cannot mimic well the instrument strength in the original sample and pseudo instrument strength is generated by the bootstrap d.g.p. Also, it fails to adequately mimic certain important properties of the disturbances. Moreover, we show that the RE bootstrap proposed by Davidson and MacKinnon (2008, 2010, 2014) is also invalid in this case. However, the RE bootstrap is found to be more robust to the instrument weakness than the standard bootstrap.

Finally, we propose two modified bootstrap procedures and show that they provide a valid distributional approximation to LIML or FULL under many/many weak instruments. A Monte Carlo experiment demonstrates that our procedures have large improvements over asymptotic theory-based normal approximation in terms of size control. In particular, one of our procedures essentially removes the distortion of the CSE-based approximation in most cases. However, this procedure can be less powerful because its reduced-form estimator imposes the null hypothesis. Our second modified procedure is relatively well balanced in terms of size and power.

A possible extension of this work could include a study in general nonlinear framework on the bootstrap validity (e.g., Hall and Horowitz (1996)'s nonparametric i.i.d. bootstrap, Brown and Newey (2002)'s efficient bootstrap, etc.) for GMM and GEL estimators under the many weak moment sequence proposed by Newey and Windmeijer (2009).

Acknowledgments

The authors are very grateful to the Co-Editor Jianqing Fan, an Associate Editor and three anonymous referees for their insightful comments and suggestions, which helped to substantially improve the paper. They also would like to thank participants at the CEA conference in Montreal (June 2013), the NASM of the Econometric Society at University of Southern California (June 2013), the SETA conference at Sungkyunkwan University (July 2013), and the CESG conference at University of Waterloo (October 2013). Special thanks to Russell Davidson, James MacKinnon, Sílvia Gonçalves, Prosper Dovonon, Yoshihiko Nishiyama, Ryo Okui and Patrick Richard for useful comments and discussions. Wang acknowledges financial support from the JSPS (10J06394). Kaffo acknowledges financial support from the FQRSC (2012-QF-163871).

Appendix A. Proofs of results for the standard bootstrap

All the proofs of the Lemmas are relegated at the end of Appendix A. Let $\hat{\beta}^* = \hat{\beta}^*_{std}$ throughout Appendix A. Let C denote

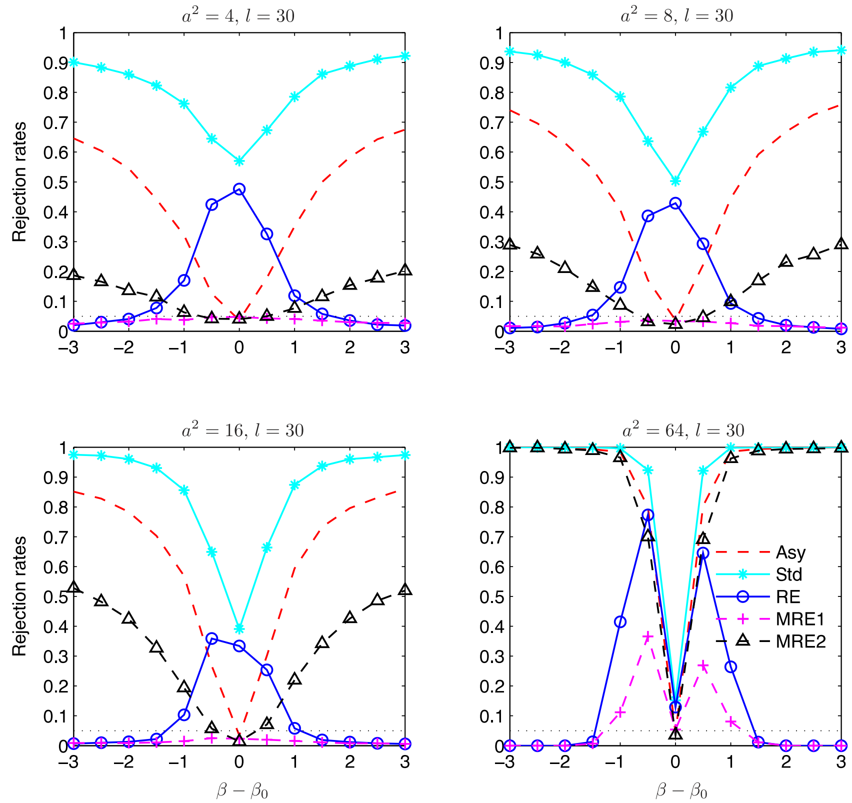

Fig. 17. Power of percentile type bootstrap Wald tests for the LIML estimator, $\rho = 0.1$.

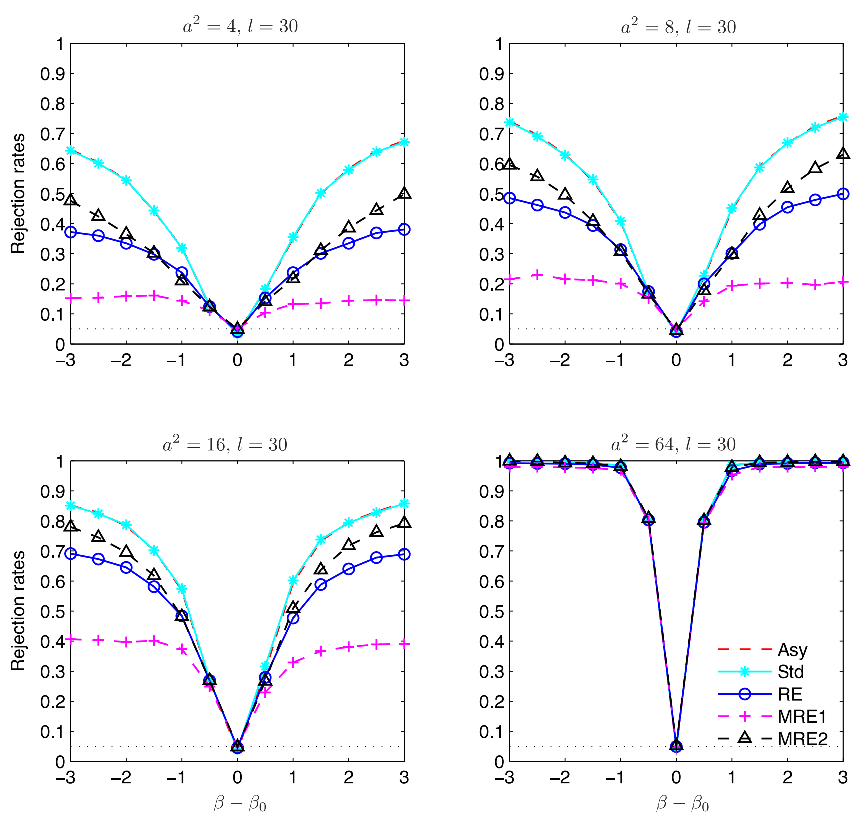

Fig. 18. Power of percentile-t type bootstrap Wald tests for the LIML estimator, $\rho = 0.1$.

a generic positive constant that may be different in different uses. Also, P^* denotes the probability measure induced by the standard residual-based bootstrap procedure and E^* denotes the expectation under P^*.

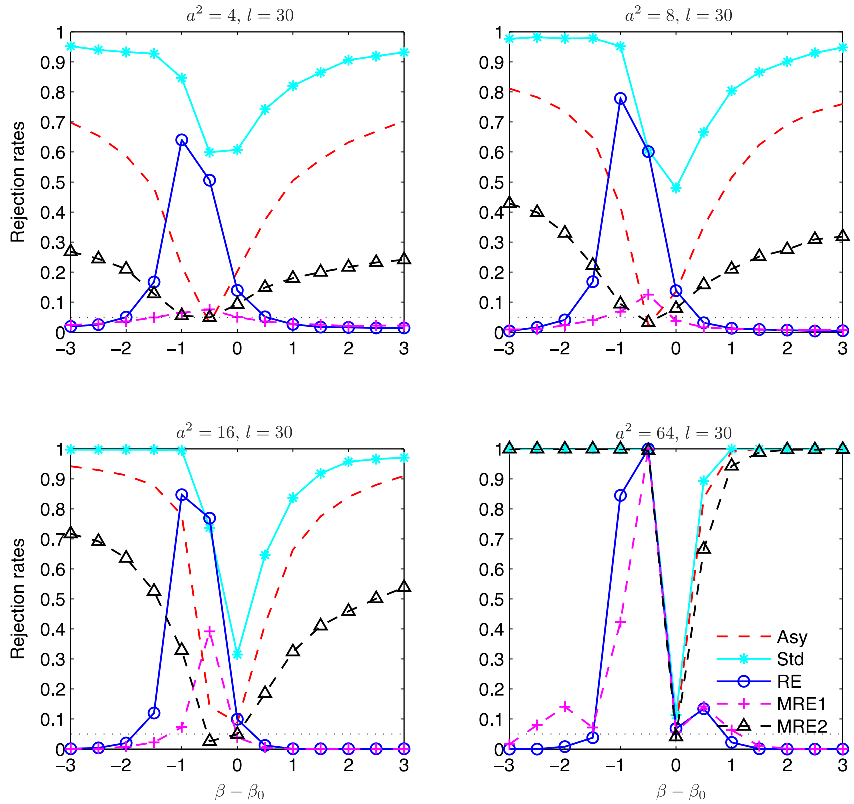

Fig. 19. Power of percentile type bootstrap Wald tests for the LIML estimator, $\rho = 0.8$.

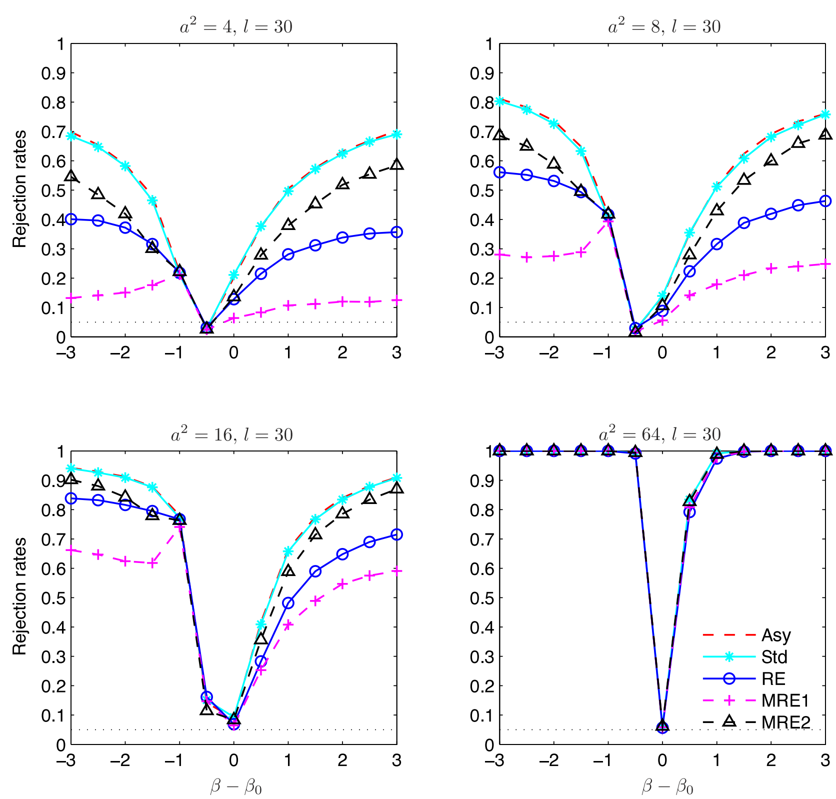

Fig. 20. Power of percentile-t type bootstrap Wald tests for the LIML estimator, $\rho = 0.8$.

Lemma A.1. *Suppose that Assumptions 1–2 hold, then* $E^*\left(\epsilon_i^{*8}\right)$ *and* $E^*\left(\left\|V_i^*\right\|^8\right)$ *are bounded in probability.*

Lemma A.2. *Suppose that Assumptions 1–2 hold, then the following statements are true:*

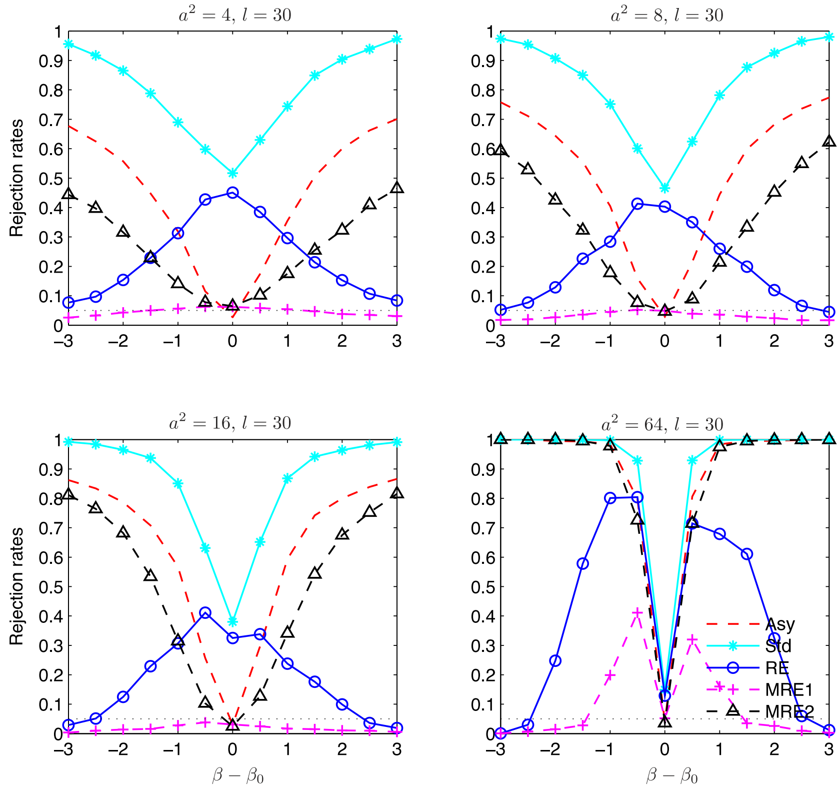

Fig. 21. Power of percentile type bootstrap Wald tests for the FULL estimator, $\rho = 0.1$.

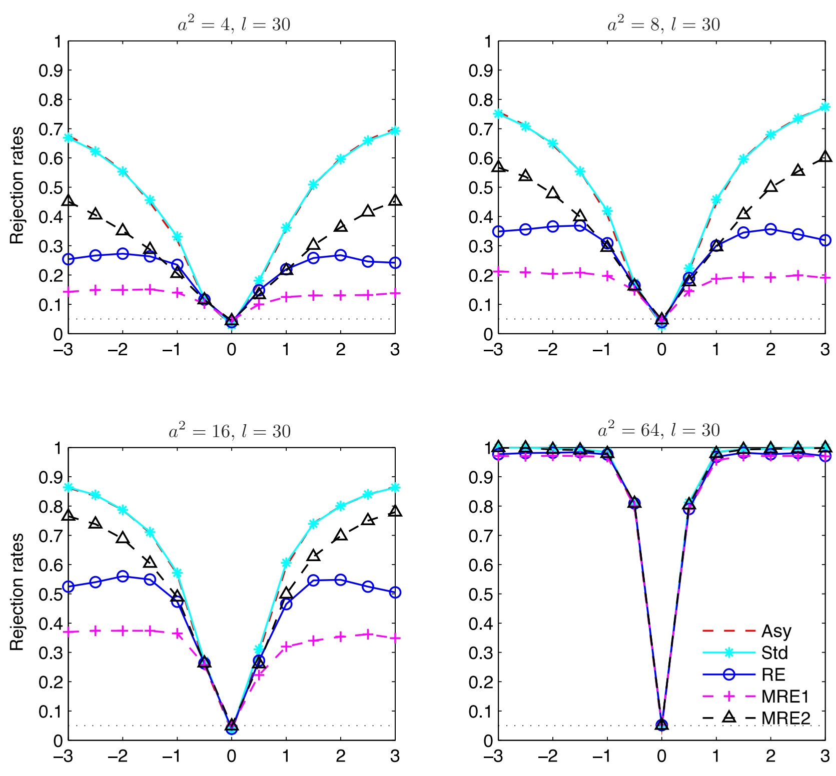

Fig. 22. Power of percentile-t type bootstrap Wald tests for the FULL estimator, $\rho = 0.1$.

(a) $V^{*'}P_Z\epsilon^*/l = \sigma^b_{V\epsilon} + O_{P^*}\left(1/\sqrt{l}\right)$, (b) $V^{*'}P_ZV^*/l = \Sigma^b_{VV} + O_{P^*}\left(1/\sqrt{l}\right)$, (c) $\epsilon^{*'}P_Z\epsilon^*/l = \sigma^b_{\epsilon\epsilon} + O_{P^*}\left(1/\sqrt{l}\right)$, *in probability; in Case* (I), (d) $\hat{\Pi}'Z'V^*/r_n = O_{P^*}\left(1/\sqrt{r_n}\right)$, (e) $\hat{\Pi}'Z'\epsilon^*/r_n = O_{P^*}\left(1/\sqrt{r_n}\right)$, *in probability; in Case* (II), (d′) $\hat{\Pi}'Z'V^*/l = O_{P^*}\left(1/\sqrt{l}\right)$, (e′) $\hat{\Pi}'Z'\epsilon^*/l = O_{P^*}\left(1/\sqrt{l}\right)$, *in probability; where* $\sigma^b_{V\epsilon} \equiv E^*\left(V^*_i\epsilon^*_i\right)$, $\Sigma^b_{VV} \equiv E^*\left(V^*_iV^{*'}_i\right)$ *and* $\sigma^b_{\epsilon\epsilon} \equiv E^*\left(\epsilon^{*2}_i\right)$.

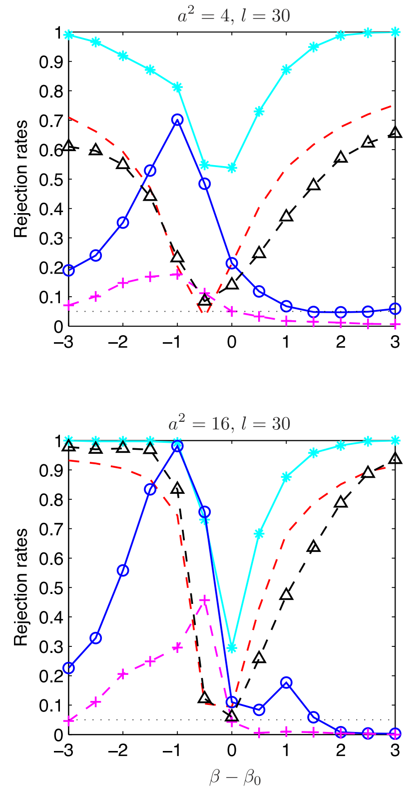

Fig. 23. Power of percentile type bootstrap Wald tests for the FULL estimator, $\rho = 0.8$.

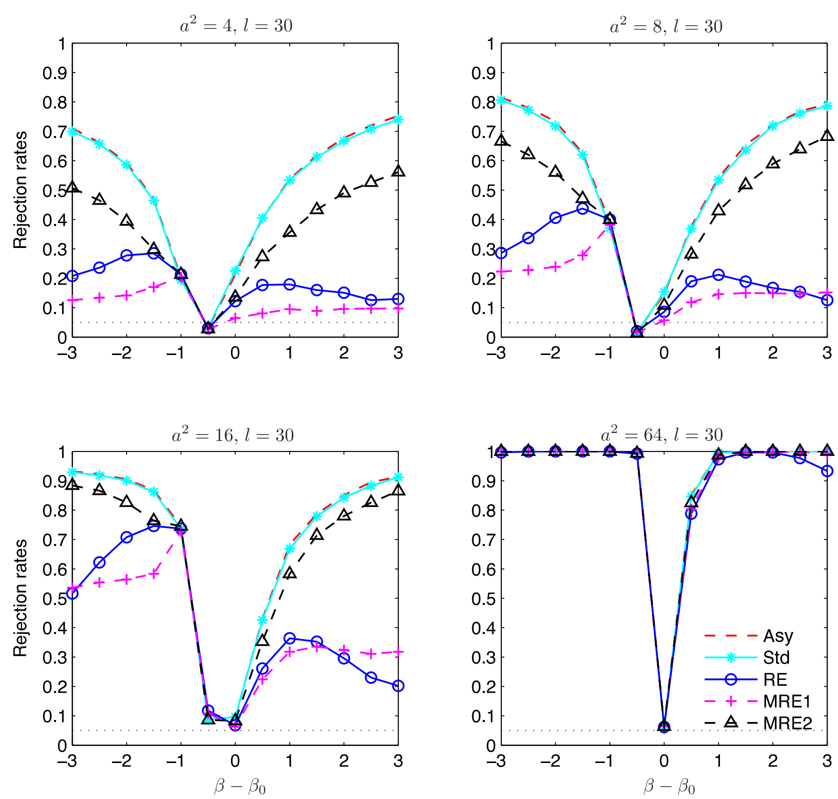

Fig. 24. Power of percentile-t type bootstrap Wald tests for the FULL estimator, $\rho = 0.8$.

To proceed, let $J^* = diag(a_1^*, \ldots, a_n^*)$ where $a_i^* = \epsilon_i^{*2} - \sigma_{\epsilon\epsilon}^b$, $i = 1, \ldots, n$.

Lemma A.3. *Suppose that Assumptions 1–2 hold, then the following statements are true:*

(a) $\tilde{V}^{*'}J^*\tilde{V}^*/n = E^*\left(a_i^*\tilde{V}_i^*\tilde{V}_i^{*'}\right) + O_{P^*}\left(1/\sqrt{n}\right)$; (b) $\tilde{V}^{*'}P_ZJ^*\tilde{V}^*/n = \lambda_n E^*\left(a_i^*\tilde{V}_i^*\tilde{V}_i^{*'}\right) + O_{P^*}\left(1/\sqrt{n}\right)$; (c) $\tilde{V}^{*'}P_ZJ^*P_Z\tilde{V}^*/n = \lambda_n\phi_n E^*\left(a_i^*\tilde{V}_i^*\tilde{V}_i^{*'}\right) + O_{P^*}\left(1/\sqrt{n}\right)$, *in probability.*

Lemma A.4. *Suppose that Assumptions 1–2 hold, then both in Case* (I) *and Case* (II), $\hat{\beta}^* - \hat{\beta} = o_{P^*}(1)$, *in probability.*

Let $\lambda^* = \epsilon^{*'}P_Z\epsilon^*/\epsilon^{*'}\epsilon^*$.

Lemma A.5. *Suppose that Assumptions 1–2 hold, then* $\lambda^* = \lambda_n + O_{P^*}\left(\sqrt{l}/n\right)$, *in probability.*

Lemma A.6. *Suppose that* $\hat{\lambda}^* = \lambda^* + O_{P^*}\left(\delta_n^\lambda\right)$ *and* $\hat{\beta}^* - \hat{\beta} = O_{P^*}\left(\delta_n^\beta\right)$ *in probability, for* $\delta_n^\lambda \to 0$ *and* $\delta_n^\beta \to 0$, *respectively. Then in Case* (I), (a) $(X^{*'}P_ZX^* - \hat{\lambda}^*X^{*'}X^*)/r_n = \bar{H}_{I,n} + O_{P^*}\left(1/\sqrt{r_n} + \delta_n^\lambda n/r_n\right)$; (b) $(X^{*'}P_Z\hat{\epsilon}^*(\hat{\beta}^*) - \hat{\lambda}^*X^{*'}\hat{\epsilon}^*(\hat{\beta}^*))/r_n = O_{P^*}\left(1/\sqrt{r_n} + \delta_n^\beta + \delta_n^\lambda n/r_n\right)$, *in probability, where* $\bar{H}_{I,n} = (1-\lambda_n)\left(\hat{\Pi}'Z'Z\hat{\Pi}/r_n\right)$. *In Case* (II), (a′) $(X^{*'}P_ZX^* - \hat{\lambda}^*X^{*'}X^*)/l = \bar{H}_{II,n} + O_{P^*}\left(1/\sqrt{l} + \delta_n^\lambda n/l\right)$; (b′) $(X^{*'}P_Z\hat{\epsilon}^*(\hat{\beta}^*) - \hat{\lambda}^*X^{*'}\hat{\epsilon}^*(\hat{\beta}^*))/l = O_{P^*}\left(1/\sqrt{l} + \delta_n^\beta + \delta_n^\lambda n/l\right)$, *in probability, where* $\bar{H}_{II,n} = (1-\lambda_n)\left(\hat{\Pi}'Z'Z\hat{\Pi}/l\right)$.

Lemma A.7. *Suppose that Assumptions 1–2 hold. Suppose* $\hat{\beta}^* - \hat{\beta} = O_{P^*}\left(\delta_n^\beta\right)$ *in probability for* $\delta_n \to 0$, *then in Case* (I), $\hat{\lambda}^* = \lambda^* + O_{P^*}\left(\frac{r_n}{n}(\delta_n^\beta)^2\right)$; *in Case* (II), $\hat{\lambda}^* = \lambda^* + O_{P^*}\left(\frac{l}{n}(\delta_n^\beta)^2\right)$, *in probability.*

Let $\hat{D}^*(\beta) = \frac{\partial}{\partial\beta}\left(\frac{(y^*-X^*\beta)'P_Z(y^*-X^*\beta)}{2(y^*-X^*\beta)'(y^*-X^*\beta)}\right) = X^{*'}P_Z\varepsilon^*(\beta) - \frac{\epsilon^*(\beta)'P_Z\epsilon^*(\beta)}{\epsilon^*(\beta)'\epsilon^*(\beta)}X^{*'}\varepsilon^*(\beta)$ where $\epsilon^*(\beta) = y^* - X^*\beta$.

Lemma A.8. *Suppose that Assumptions 1–2 hold. Suppose* $\hat{\beta}^* - \hat{\beta} = O_{P^*}\left(\delta_n^\beta\right)$ *in probability for* $\delta_n^\beta \to 0$, *then in Case* (I), $-\left(\partial\hat{D}^*(\bar{\beta}^*)/\partial\beta\right)/r_n = \bar{H}_{I,n} + O_{P^*}\left(1/\sqrt{r_n} + \delta_n^\beta\right)$; *in Case* (II), $-\left(\partial\hat{D}^*(\bar{\beta}^*)/\partial\beta\right)/l = \bar{H}_{II,n} + O_{P^*}\left(1/\sqrt{l} + \delta_n^\beta\right)$, *in probability, where* $\bar{\beta}^*$ *lies between* $\hat{\beta}$ *and* $\hat{\beta}^*$.

Lemma A.9. *Suppose that Assumptions 1–2 hold, then the following statements are true: in Case* (I), $\hat{D}^*\left(\hat{\beta}\right)/\sqrt{r_n} = ((1-\lambda_n)Z\hat{\Pi} + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*)'\epsilon^*/\sqrt{r_n} + O_{P^*}\left(1/\sqrt{r_n}\right)$; *in Case* (II), $\hat{D}^*\left(\hat{\beta}\right)/\sqrt{l} = ((1-\lambda_n)Z\hat{\Pi} + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*)'\epsilon^*/\sqrt{l} + O_{P^*}\left(1/\sqrt{l}\right)$, *in probability, where* $\tilde{V}^* = V^* - \epsilon^*q^{b'}$ *and* $q^b = \sigma_{V\epsilon}^b/\sigma_{\epsilon\epsilon}^b$.

Proof of Theorem 3.1. Consider first the case in which $\hat{\beta}^*$ is the bootstrap analogue of LIML. Notice that the first-order conditions for $\hat{\beta}^*$ can be written as $\hat{D}^*(\hat{\beta}^*) = 0$. Expanding around $\hat{\beta}$ gives

$$0 = \hat{D}^*(\hat{\beta}) + \frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}(\hat{\beta}^* - \hat{\beta}) \tag{15}$$

where $\bar{\beta}^*$ lies on the line joining $\hat{\beta}^*$ and $\hat{\beta}$.

We first show the results for Case (I). The proof is decomposed into three steps:

(1) We verify the conditions of a martingale central limit theorem;

(2) We derive the asymptotic variance–covariance matrix for $\sum_{i=1}^n W_{I,i}^*$;

(3) We obtain the limiting distribution of $\hat{\beta}^*$.

(1) Notice that in Case (I), $-\frac{1}{r_n}\left(\frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right) = O_{P^*}(1)$ in probability by Lemmas A.4 and A.8 and $\bar{H}_{I,n} = O_P(1)$. Then solving Eq. (15) gives

$$\sqrt{r_n}(\hat{\beta}^* - \hat{\beta}) = -\left(\frac{1}{r_n}\frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right)^{-1}\frac{\hat{D}^*(\hat{\beta})}{\sqrt{r_n}}.$$

To proceed, define $W_{I,i}^* \equiv \begin{pmatrix}(1-\lambda_n)\hat{\Pi}'Z_i\varepsilon_i^*/\sqrt{r_n}\\(P_{ii}-\lambda_n)\tilde{V}_i^*\varepsilon_i^*/\sqrt{l}\end{pmatrix}$, and $\tilde{V}_i^* \equiv V_i^* - \varepsilon_i^*q^b$. We check the conditions of Lemma A2 in Hansen et al. (2008) hold with $\left(W_{I,1}^*, \tilde{V}_1^*, \epsilon_1^*\right), \ldots, \left(W_{I,n}^*, \tilde{V}_n^*, \epsilon_n^*\right)$ conditionally on the original sample with probability converging to one.[8] Notice that by the current bootstrap d.g.p., $E^*\left(\epsilon_i^*\right) = 0$ and $E^*\left(\tilde{V}_i^*\right) = 0$. $E^*\left(\epsilon_i^{*4}\right)$ and $E^*\left(\tilde{V}_i^{*4}\right)$ are bounded in probability by similar arguments as in the proofs of Lemma A.1. It is also easy to see that $\sum_{i=1}^n E^*\left(W_{I,i}^*W_{I,i}^{*'}\right)$ is bounded in probability. Then, it suffices to show that

$$\begin{cases}\sum_{i=1}^n E^*\left(\left\|\frac{1}{\sqrt{r_n}}(1-\lambda_n)\hat{\Pi}'Z_i\varepsilon_i^*\right\|^4\right) \to^P 0\\ \sum_{i=1}^n E^*\left(\left\|\frac{1}{\sqrt{l}}(P_{ii}-\lambda_n)\tilde{V}_i^*\varepsilon_i^*\right\|^4\right) \to^P 0.\end{cases}$$

For the first term, by the Minkowski inequality

$$\begin{aligned}E^*\left(\sum_{i=1}^n\left\|\frac{\hat{\Pi}'Z_i\epsilon_i^*}{\sqrt{r_n}}\right\|^4\right) &= E^*\left(\sum_{i=1}^n\left\|\frac{\Pi'Z_i\epsilon_i^*}{\sqrt{r_n}} + \frac{V'Z(Z'Z)^{-1}Z_i\epsilon_i^*}{\sqrt{r_n}}\right\|^4\right)\\ &\le C\sum_{i=1}^n\left\|\frac{\Pi'Z_i}{\sqrt{r_n}}\right\|^4 E^*\left(\epsilon_i^{*4}\right) + C\sum_{i=1}^n\left\|\frac{V'Z(Z'Z)^{-1}Z_i}{\sqrt{r_n}}\right\|^4 E^*\left(\epsilon_i^{*4}\right)\\ &\equiv D_1 + D_2\end{aligned}$$

$D_1 = O_P(1) \times C\left(\frac{1}{r_n^2}\sum_{i=1}^n \|\Pi'Z_i\|^4\right) \to^P 0$ by $E^*\left(\epsilon_i^{*4}\right) = O_P(1)$ and by Assumption 2. Similarly, we have $D_2 = O_P(1) \times C\left(\frac{1}{r_n^2}\sum_{i=1}^n\left\|V'Z(Z'Z)^{-1}Z_i\right\|^4\right)$; let $v = (v_1, \ldots, v_n)'$ be an arbitrary column of V, then by Marcinkiewicz–Zygmund inequality,

$$E\left[\left|v'Z(Z'Z)^{-1}Z_i\right|^4\right] = E\left[\left|\sum_{j=1}^n v_jP_{ji}\right|^4\right] \le CE\left[\left|\sum_{j=1}^n v_j^2P_{ji}^2\right|^2\right].$$

Furthermore, by Jensen's inequality and $\sum_{j=1}^n P_{ji}^2 = P_{ii}$ we have

$$E\left[\left|\sum_{j=1}^n v_j^2P_{ji}^2\right|^2\right] \le P_{ii}\left(\sum_{j=1}^n E\left(|v_j|^4\right)P_{ji}^2\Big/\sum_{j=1}^n P_{ji}^2\right) \le CP_{ii}.$$

[8] Alternatively, CLTs such as that in Heyde and Brown (1970) could be used to establish the asymptotic normality.

Combining the last two equations gives $E\left[\left|v'Z(Z'Z)^{-1}Z_i\right|^4\right] \leq CP_{ii}$. Therefore

$$\frac{1}{r_n^2}\sum_{i=1}^n E\left(\left|v'Z(Z'Z)^{-1}Z_i\right|^4\right) \leq C\left(\frac{\sum_{i=1}^n P_{ii}}{r_n^2}\right) = C\left(\frac{l}{r_n^2}\right) \to 0$$

given that $\sum_{i=1}^n P_{ii} = l$ and $\sqrt{l}/r_n \to 0$. Thus, we conclude that $\frac{1}{r_n^2}\sum_{i=1}^n \left|v'Z(Z'Z)^{-1}Z_i\right|^4 \to^P 0$ by the Markov inequality. Showing the result for each column of V, we obtain that $D_2 \to^P 0$, and the desired result for $\sum_{i=1}^n E^*\left(\left\|\frac{1}{\sqrt{r_n}}(1-\lambda_n)\hat{\Pi}'Z_i\varepsilon_i^*\right\|^4\right)$ follows.

For the term $\sum_{i=1}^n E^*\left\|\frac{1}{\sqrt{l}}(P_{ii}-\lambda_n)\tilde{V}_i^*\epsilon_i^*\right\|^4$, we have by Lemma A.1 that $E^*\left(\epsilon_i^{*8}\right)$ and $E^*\left(V_i^{*8}\right)$ are bounded in probability so that

$$\sum_{i=1}^n E^*\left\|\frac{1}{\sqrt{l}}(P_{ii}-\lambda_n)\tilde{V}_i^*\epsilon_i^*\right\|^4 \leq O_P(1)\left(\frac{\sum_{i=1}^n P_{ii}^4 + n\lambda_n^4}{l^2}\right) \leq O_P(1)\left(\frac{1}{l}+\frac{\lambda_n^2}{n}\right) \to^P 0$$

which follows from the fact that $P_{ii}^4 \leq P_{ii}$ and $\sum_{i=1}^n P_{ii} = l$. Thus, conditionally on the sample, the conditions in Lemma A2 of Hansen et al. (2008) are satisfied with probability approaching one.

(2) Now we derive the asymptotic variance–covariance matrix for $\sum_{i=1}^n W_{l,i}^*$. First, notice the equation in Box I, where $\bar{H}_l = (1-\lambda)(\Psi+\gamma\Sigma_{VV})$; $\bar{\Sigma}_{\tilde{V}\tilde{V}} = (1-\lambda)\Sigma_{\tilde{V}\tilde{V}} + (\lambda-\lambda^2)\frac{\sigma_{V\epsilon}\sigma_{V\epsilon}'}{\sigma_{\epsilon\epsilon}}$; $\bar{A} = (1-\lambda)A$; $\bar{B} = (1-2\lambda+\lambda\phi)B + \lambda(\phi-\lambda)^2\left\{2E\left(\epsilon_i^3\tilde{V}_i\right)q' + q\left[E\left(\epsilon_i^4\right)-(\sigma_{\epsilon\epsilon})^2\right]q'\right\}$, $q = \sigma_{V\epsilon}/\sigma_{\epsilon\epsilon}$.

(2.1) These results follow by first showing that the diagonal term

$$\begin{aligned}
&\sum_{i=1}^n \frac{(1-\lambda_n)^2}{r_n}\hat{\Pi}'Z_iZ_i'\hat{\Pi}E^*\left(\varepsilon_i^{*2}\right)\\
&= (1-\lambda_n)^2\left(\frac{\hat{\epsilon}'\hat{\epsilon}}{n}\right)\left(\frac{\Pi'Z'Z\Pi}{r_n}+\frac{\Pi'Z'V}{r_n}+\frac{V'Z\Pi}{r_n}+\frac{V'P_ZV}{r_n}\right)\\
&= (1-\lambda_n)^2\left(\frac{\hat{\epsilon}'\hat{\epsilon}}{n}\right)\left\{\frac{\Pi'Z'Z\Pi}{r_n}+O_P\left(\frac{1}{\sqrt{r_n}}\right)+O_P\left(\frac{1}{\sqrt{r_n}}\right)\right.\\
&\quad\left.+\left(\frac{l}{r_n}\right)\Sigma_{VV}+O_P\left(\frac{\sqrt{l}}{r_n}\right)\right\}\\
&\to^P (1-\lambda)^2\sigma_{\epsilon\epsilon}\left(\Psi+\gamma\Sigma_{VV}\right) \equiv (1-\lambda)\sigma_{\epsilon\epsilon}\bar{H}_l
\end{aligned}$$

by $\hat{\epsilon}'\hat{\epsilon}/n \to^P \sigma_{\epsilon\epsilon}$, $\Pi'Z'Z\Pi/r_n \to \Psi$, $\Pi'Z'V/r_n = O_P\left(1/\sqrt{r_n}\right)$, and $\sqrt{l}/r_n \to 0$, as $n\to\infty$.

(2.2) For the off-diagonal term $\sum_{i=1}^n \frac{1-\lambda_n}{\sqrt{lr_n}}(P_{ii}-\lambda_n)\hat{\Pi}'Z_iE^*\left(\varepsilon_i^{*2}\tilde{V}_i^{*'}\right)$, note that under the current bootstrap d.g.p.,

$$\begin{aligned}
E^*\left(\epsilon_i^{*2}\tilde{V}_i^{*'}\right) &= E^*\left(\epsilon_i^{*2}V_i^{*'}\right) - E^*\left(\epsilon_i^{*3}\right)q^{b'}\\
&= \frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i^2\hat{V}_i' - \frac{1}{n}\sum_{i=1}^n\hat{\epsilon}_i^3\left(\frac{1}{n}\sum_{i=1}^n\hat{\epsilon}_i\hat{V}_i'\right)\left(\frac{1}{n}\sum_{i=1}^n\hat{\epsilon}_i^2\right)^{-1}.
\end{aligned}$$

Let $\hat{a} = \left(\hat{\epsilon}_1^2-\hat{\sigma}_{\epsilon\epsilon},\ldots,\hat{\epsilon}_n^2-\hat{\sigma}_{\epsilon\epsilon}\right)'$, $\hat{\sigma}_{\epsilon\epsilon} = \hat{\epsilon}'\hat{\epsilon}/n$, and $a = \left(\epsilon_1^2-\sigma_{\epsilon\epsilon},\ldots,\epsilon_n^2-\sigma_{\epsilon\epsilon}\right)'$. Note that $\frac{1}{n}\sum_{i=1}^n\hat{V}_i = 0$ by the setting of our model that exogenous regressors, including a constant, have to be filtered out. Thus, we have $\frac{1}{n}\sum_{i=1}^n\hat{\epsilon}_i^2\hat{V}_i' = \frac{\hat{a}'\hat{V}}{n}$. Moreover, $\frac{(\hat{a}-a)'\hat{V}}{n} = O_P\left(\frac{1}{\sqrt{r_n}}\right)$ by $\frac{\|\hat{a}-a\|^2}{n} = O_P\left(\frac{1}{r_n}\right)$, $\frac{\hat{V}'\hat{V}}{n} = \frac{V'M_ZV}{n} = (1-\lambda)\Sigma_{VV}+o_P(1) = O_P(1)$ and Cauchy–Schwarz inequality. In addition, we have $\frac{a'\hat{V}}{n} \to^P (1-\lambda)E\left(\epsilon_i^2V_i'\right)$ by following the same steps as in the proof of Lemma A9 and Lemma A10 in Hansen et al. (2008). By the triangle inequality, we conclude that $E^*\left(\epsilon_i^{*2}V_i^{*'}\right) \to^P (1-\lambda)E\left(\epsilon_i^2V_i'\right)$. For the term $E^*\left(\epsilon_i^{*3}\right)q^{b'}$, we have

$$\frac{1}{n}\sum_{i=1}^n\hat{\epsilon}_i^3\left(\frac{1}{n}\sum_{i=1}^n\hat{\epsilon}_i\hat{V}_i'\right)\left(\frac{1}{n}\sum_{i=1}^n\hat{\epsilon}_i^2\right)^{-1} \to^P (1-\lambda)E\left(\epsilon_i^3\right)q'$$

which follows by using the fact that $\frac{1}{n}\sum_{i=1}^n\hat{\epsilon}_i^3 \to^P E\left(\epsilon_i^3\right)$, $\frac{1}{n}\sum_{i=1}^n\hat{\epsilon}_i\hat{V}_i' = \frac{\hat{\epsilon}'M_ZV}{n} \to^P (1-\lambda)E\left(\epsilon_iV_i'\right)$, and $\frac{1}{n}\sum_{i=1}^n\hat{\epsilon}_i^2 \to^P E\left(\epsilon_i^2\right)$, as $n\to\infty$. Together with the result for $\frac{1}{n}\sum_{i=1}^n\hat{\epsilon}_i^2\hat{V}_i'$, we conclude that

$$\begin{aligned}
E^*\left(\epsilon_i^{*2}\tilde{V}_i^{*'}\right) &\to^P (1-\lambda)E\left(\epsilon_i^2V_i'\right) - (1-\lambda)E\left(\epsilon_i^3\right)q'\\
&= (1-\lambda)E\left(\epsilon_i^2\tilde{V}_i'\right)
\end{aligned}$$

for the standard residual bootstrap.

Now, for the term $\sum_{i=1}^n\frac{(1-\lambda_n)}{\sqrt{lr_n}}(P_{ii}-\lambda_n)\hat{\Pi}'Z_i$, let $d_i = (P_{ii}-\lambda_n)/\sqrt{l}$ and $d = (d_1,\ldots,d_n)'$. Note that $\|d\|^2 \leq 1$, and $E\left(\left\|V'P_Zd\right\|^2\right) \leq Cd'd \leq C$. Thus, $V'P_Zd = O_P(1)$ by standard arguments. Then, we have

$$\begin{aligned}
\sum_{i=1}^n\frac{(1-\lambda_n)}{\sqrt{lr_n}}(P_{ii}-\lambda_n)\hat{\Pi}'Z_i &= \frac{(1-\lambda_n)}{\sqrt{r_n}}X'P_Zd\\
&= \frac{(1-\lambda_n)}{\sqrt{r_n}}\Pi'Z'd + O_P\left(\frac{1}{\sqrt{r_n}}\right)\\
&= \sum_{i=1}^n\frac{(1-\lambda_n)}{\sqrt{lr_n}}(P_{ii}-\lambda_n)\Pi'Z_i + O_P\left(\frac{1}{\sqrt{r_n}}\right)
\end{aligned}$$

and the desired result for the off-diagonal terms follows.

(2.3) For the other diagonal term $\sum_{i=1}^n(P_{ii}-\lambda_n)^2E^*\left(\varepsilon_i^{*2}\tilde{V}_i^*\tilde{V}_i^{*'}\right)/l$, note that

$$E^*\left(\epsilon_i^{*2}\tilde{V}_i^*\tilde{V}_i^{*'}\right) = \sigma_{\epsilon\epsilon}^bE^*\left(\tilde{V}_i^*\tilde{V}_i^{*'}\right) + E^*\left(\left(\epsilon_i^{*2}-\sigma_{\epsilon\epsilon}^b\right)\tilde{V}_i^*\tilde{V}_i^{*'}\right).$$

First, we have by the current bootstrap d.g.p. that

$$\begin{aligned}
E^*\left(\tilde{V}_i^*\tilde{V}_i^{*'}\right) &= E^*\left\{\left(V_i^*-\epsilon_i^*q^{b'}\right)\left(V_i^*-\epsilon_i^*q^{b'}\right)'\right\}\\
&= \frac{\hat{V}'\hat{V}}{n} - \left(\frac{\hat{V}'\hat{\epsilon}}{n}\right)\left(\frac{\hat{\epsilon}'\hat{V}}{n}\right)\left(\frac{\hat{\epsilon}'\hat{\epsilon}}{n}\right)^{-1}\\
&\to^P (1-\lambda)\Sigma_{VV} - (1-\lambda)^2\frac{\sigma_{V\epsilon}\sigma_{V\epsilon}'}{\sigma_{\epsilon\epsilon}}
\end{aligned}$$

since $\hat{V}'\hat{V}/n = V'M_ZV/n \to^P (1-\lambda)\Sigma_{VV}$, $\hat{V}'\hat{\epsilon}/n = V'M_Z\hat{\epsilon}/n \to^P (1-\lambda)\sigma_{V\epsilon}$, and $\hat{\epsilon}'\hat{\epsilon}/n \to^P \sigma_{\epsilon\epsilon}$, as $n\to\infty$.

For the second term $E^*\left(\left(\epsilon_i^{*2}-\sigma_{\epsilon\epsilon}^b\right)\tilde{V}_i^*\tilde{V}_i^{*'}\right)$, we let $\tilde{v}$ and v be arbitrary columns of $\tilde{V}$ and V, respectively. In addition, let $\tilde{v}^*$ and v^* be an arbitrary column of $\tilde{V}^*$ and V^*, respectively. By the standard residual bootstrap scheme,

$$E^*\left(\left(\epsilon_i^{*2}-\sigma_{\epsilon\epsilon}^b\right)\tilde{v}_i^{*2}\right) = \frac{1}{n}\sum_{i=1}^n\left\{\left(\hat{\epsilon}_i^2-\sigma_{\epsilon\epsilon}^b\right)\left[\hat{v}_i-\hat{\epsilon}_i\left(\frac{\sigma_{\epsilon v}^b}{\sigma_{\epsilon\epsilon}^b}\right)\right]^2\right\}$$

$$\sum_{i=1}^{n} E^*\left(W^*_{I,i}W^{*'}_{I,i}\right) = \begin{pmatrix} \sum_{i=1}^{n}\frac{(1-\lambda_n)^2}{r_n}\hat{\Pi}'Z_iZ_i'\hat{\Pi}E^*\left(\varepsilon_i^{*2}\right) & \sum_{i=1}^{n}\frac{1-\lambda_n}{\sqrt{lr_n}}(P_{ii}-\lambda_n)\hat{\Pi}'Z_iE^*\left(\varepsilon_i^{*2}\tilde{V}_i^{*'}\right) \\ E^*\left(\varepsilon_i^{*2}\tilde{V}_i^*\right)\sum_{i=1}^{n}\frac{1-\lambda_n}{\sqrt{lr_n}}(P_{ii}-\lambda_n)Z_i'\hat{\Pi} & \sum_{i=1}^{n}\frac{(P_{ii}-\lambda_n)^2}{l}E^*\left(\varepsilon_i^{*2}\tilde{V}_i^*\tilde{V}_i^{*'}\right)\end{pmatrix}$$
$$\to^P \begin{pmatrix}(1-\lambda)\sigma_{\epsilon\epsilon}\bar{H}_I & (1-\lambda)\bar{A}' \\ (1-\lambda)\bar{A} & (\phi-\lambda)\sigma_{\epsilon\epsilon}\bar{\Sigma}_{\tilde{V}\tilde{V}}+\bar{B}\end{pmatrix}$$

Box I.

where $\hat{v} = M_Z v$, $\sigma^b_{\epsilon\epsilon} = \hat{\epsilon}'\hat{\epsilon}/n$, and $\sigma^b_{\epsilon v} = \hat{\epsilon}'\hat{v}/n$. Also note that $\hat{v} - \hat{\epsilon}\left(\frac{\sigma^b_{\epsilon v}}{\sigma^b_{\epsilon\epsilon}}\right) = M_Z(\tilde{v} + \epsilon q_v) - \hat{\epsilon}\hat{q}_v$, where we denote q_v, $\hat{q}_v$ for $\sigma_{\epsilon v}/\sigma_{\epsilon\epsilon}$ and $\hat{\epsilon}'M_Z v/\hat{\epsilon}'\hat{\epsilon}$, respectively. Let $\hat{J} = diag(\hat{a}_1, \ldots, \hat{a}_n)$, and we obtain that

$$\begin{aligned} &E^*\left(\left(\epsilon_i^{*2} - \sigma^b_{\epsilon\epsilon}\right)\tilde{v}_i^{*2}\right) \\ &= \frac{1}{n}\left\{M_Z\tilde{v} + M_Z\epsilon q_v - \hat{\epsilon}\hat{q}_v\right\}'\hat{J}\left\{M_Z\tilde{v} + M_Z\epsilon q_v - \hat{\epsilon}\hat{q}_v\right\} \\ &\to^P (1-2\lambda+\lambda\phi)E\left(\left(\epsilon_i^2 - \sigma_{\epsilon\epsilon}\right)\tilde{v}_i^2\right) \\ &\quad + \lambda(\phi-\lambda)\left\{2q_vE\left(\epsilon_i^3\tilde{v}_i\right) + q_v^2\left(E\left(\epsilon_i^4\right) - (\sigma_{\epsilon\epsilon})^2\right)\right\}\end{aligned}$$

by showing that

$$\begin{aligned} &\tilde{v}'M_Z\hat{J}M_Z\tilde{v}/n \to^P (1-2\lambda+\lambda\phi)E\left(\left(\epsilon_i^2 - \sigma_{\epsilon\epsilon}\right)\tilde{v}_i^2\right), \\ &\epsilon'M_Z\hat{J}M_Z\tilde{v}/n \to^P (1-2\lambda+\lambda\phi)E\left(\epsilon_i^3\tilde{v}_i\right), \\ &\epsilon'M_Z\hat{J}M_Z\epsilon/n \to^P (1-2\lambda+\lambda\phi)\left(E\left(\epsilon_i^4\right) - (\sigma_{\epsilon\epsilon})^2\right), \\ &\hat{\epsilon}'\hat{J}M_Z\tilde{v}/n \to^P (1-\lambda)E\left(\epsilon_i^3\tilde{v}_i\right), \\ &\hat{\epsilon}'\hat{J}M_Z\epsilon/n \to^P (1-\lambda)\left(E\left(\epsilon_i^4\right) - (\sigma_{\epsilon\epsilon})^2\right),\end{aligned}$$

$\hat{\epsilon}'\hat{J}\hat{\epsilon}/n \to^P E\left(\epsilon_i^4\right) - (\sigma_{\epsilon\epsilon})^2$, and $\hat{q}_v \to^P q_v$, using similar arguments as in the proofs of Lemma A11 in Hansen et al. (2008). We apply this result to each column of $\tilde{V}$ and the result for $E^*\left((\epsilon_i^{*2} - \sigma^b_{\epsilon\epsilon})\tilde{V}_i^*\tilde{V}_i^{*'}\right)$ follows. Note that $B = (\phi-\lambda)E\left((\epsilon_i^2 - \sigma_{\epsilon\epsilon})\tilde{V}_i\tilde{V}_i'\right)$ and $\lim_{n\to\infty}\sum_{i=1}^n(P_{ii}-\lambda_n)^2/l = \phi - \lambda$, then we have

$$\begin{aligned} &\frac{1}{l}\sum_{i=1}^{n}(P_{ii}-\lambda_n)^2E^*\left(\left(\epsilon_i^{*2} - \sigma^b_{\epsilon\epsilon}\right)\tilde{V}_i^*\tilde{V}_i^{*'}\right) \\ &\to^P (1-2\lambda+\lambda\phi)B + \lambda(\phi-\lambda)^2 \\ &\quad\times\left\{qE\left(\epsilon_i^3\tilde{V}_i'\right) + E\left(\epsilon_i^3\tilde{V}_i\right)q' + q\left(E\left(\epsilon_i^4\right) - (\sigma_{\epsilon\epsilon})^2\right)q'\right\}\end{aligned}$$

as required.

(3) Finally, to obtain the asymptotic distributional result for $\hat{\beta}^*$, we define $U^*_{I,n} = \begin{pmatrix}\sum_{i=1}^n W^*_{I,i} \\ \sum_{i\neq j}^n \tilde{V}_i^*P_{ij}\varepsilon_j^*/\sqrt{l}\end{pmatrix}$, and we have by Lemma A.9 and by Lemma A2 of Hansen et al. (2008) that

$$\begin{aligned}\frac{\hat{D}^*(\hat{\beta})}{\sqrt{r_n}} &= \frac{1}{\sqrt{r_n}}\left((1-\lambda_n)Z\hat{\Pi} + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right)'\epsilon^* + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) \\ &= \frac{1}{\sqrt{r_n}}\left(\sum_{i=1}^{n}(1-\lambda_n)\hat{\Pi}'Z_i\epsilon_i^* + \sum_{i=1}^{n}(P_{ii}-\lambda_n)\tilde{V}_i^*\epsilon_i^*\right. \\ &\quad\left. + \sum_{i\neq j}^{n}\tilde{V}_i^*P_{ij}\epsilon_j^*\right) + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) \\ &= F_{I,n}U^*_{I,n} + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) \to^{d^*} N(0, \bar{\Upsilon}_I)\end{aligned}$$

in probability, where $F_{I,n} = \left[I_k, \sqrt{l/r_n}I_k, \sqrt{l/r_n}I_k\right]$ and $\bar{\Upsilon}_I = (1-\lambda)\sigma_{\epsilon\epsilon}\left\{\bar{H}_I + \gamma\bar{\Sigma}_{\tilde{V}\tilde{V}}\right\} + (1-\lambda)\sqrt{\gamma}\left\{\bar{A} + \bar{A}'\right\} + \gamma\bar{B}$; together with the result in Lemma A.8 and the continuous mapping theorem for weak convergence in probability (e.g., Xiong and Li (2008), Theorem 3.1), we obtain that in case (I),

$$\sqrt{r_n}(\hat{\beta}^* - \hat{\beta}) = \left(\frac{1}{r_n}\frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right)^{-1}\frac{\hat{D}^*(\hat{\beta})}{\sqrt{r_n}} \to^{d^*} N\left(0, \bar{\Lambda}_I\right)$$

in probability, where $\bar{\Lambda}_I = \bar{H}_I^{-1}\bar{\Upsilon}_I\bar{H}_I^{-1}$.

Analogously, for Case (II), the proof is decomposed into three steps:

(1) We verify the conditions of a martingale central limit theorem;

(2) We derive the asymptotic variance–covariance matrix for $\sum_{i=1}^n W^*_{II,i}$;

(3) We obtain the limiting distribution of $\hat{\beta}^*$.

(1) It holds by Lemmas A.4 and A.8 and by $\bar{H}_{II,n} = O_P(1)$ that

$$\sqrt{l}(\hat{\beta}^* - \hat{\beta}) = -\left(\frac{1}{l}\frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right)^{-1}\frac{\hat{D}^*(\hat{\beta})}{\sqrt{l}}.$$

Let $W^*_{II,i} = \begin{pmatrix}(1-\lambda_n)\hat{\Pi}'Z_i\varepsilon_i^*/\sqrt{l} \\ (P_{ii}-\lambda_n)\tilde{V}_i^*\varepsilon_i^*/\sqrt{l}\end{pmatrix}$, and we have $\sum_{i=1}^n E^*\left(\left\|W^*_{II,i}\right\|^4\right) \to^P 0$ by applying the same arguments as those used in Case (I). Also, other conditions in Lemma A2 of Hansen et al. (2008) are verified to be satisfied conditionally on the original sample with probability converging to one.

(2) Then, we derive the asymptotic variance–covariance matrix of $\sum_{i=1}^n W^*_{II,i}$ as in Box II, where $\bar{H}_{II} = (1-\lambda)\Sigma_{VV}$. Note that in Case (II),

$$\sum_{i=1}^{n}\left((1-\lambda_n)(P_{ii}-\lambda_n)\hat{\Pi}'Z_i/l\right) \to^P 0$$

by $\frac{\Pi'Z'd}{\sqrt{l}} = O_P\left(\sqrt{\frac{r_n}{l}}\right) = o_P(1)$ and $\frac{V'P_Zd}{\sqrt{l}} = O_P\left(\frac{1}{\sqrt{l}}\right)$, where $d_i = (P_{ii}-\lambda_n)/\sqrt{l}$ and $d = (d_1, \ldots, d_n)'$. Therefore, the off-diagonal terms converge in probability to zero in this case.

(3) Now, we define $U^*_{II,n} = \begin{pmatrix}\sum_{i=1}^n W^*_{II,i} \\ \sum_{i\neq j}^n \tilde{V}_i^*P_{ij}\varepsilon_j^*/\sqrt{l}\end{pmatrix}$, then it follows by using similar arguments as in Case (I) that

$$\begin{aligned}\frac{\hat{D}^*(\hat{\beta})}{\sqrt{l}} &= \frac{1}{\sqrt{l}}\left((1-\lambda_n)Z\hat{\Pi} + P_Z\tilde{V}^* - \lambda\tilde{V}^*\right)'\epsilon^* + O_{P^*}\left(\frac{1}{\sqrt{l}}\right) \\ &= \frac{1}{\sqrt{l}}\left(\sum_{i=1}^{n}(1-\lambda_n)\hat{\Pi}'Z_i\epsilon_i^* + \sum_{i=1}^{n}(P_{ii}-\lambda_n)\tilde{V}_i^*\epsilon_i^*\right. \\ &\quad\left. + \sum_{i\neq j}^{n}\tilde{V}_i^*P_{ij}\epsilon_j^*\right) + O_{P^*}\left(\frac{1}{\sqrt{l}}\right) \\ &= F_{II}U^*_{II,n} + O_{P^*}\left(\frac{1}{\sqrt{l}}\right) \to^{d^*} N\left(0, \bar{\Upsilon}_{II}\right)\end{aligned}$$

$$\sum_{i=1}^{n} E^*\left(W_{II,i}^* W_{II,i}^{*'}\right) = \begin{pmatrix} \sum_{i=1}^{n} \frac{(1-\lambda_n)^2}{l} \hat{\Pi}' Z_i Z_i' \hat{\Pi} E^*\left(\varepsilon_i^{*2}\right) & \sum_{i=1}^{n} \frac{1-\lambda_n}{l} (P_{ii}-\lambda_n) \hat{\Pi}' Z_i E^*\left(\varepsilon_i^{*2} \tilde{V}_i^{*'}\right) \\ E^*\left(\varepsilon_i^{*2} \tilde{V}_i^*\right) \sum_{i=1}^{n} \frac{1-\lambda_n}{l} (P_{ii}-\lambda_n) Z_i' \hat{\Pi} & \sum_{i=1}^{n} \frac{(P_{ii}-\lambda_n)^2}{l} E^*\left(\varepsilon_i^{*2} \tilde{V}_i^* \tilde{V}_i^{*'}\right) \end{pmatrix}$$

$$\to^P \begin{pmatrix} (1-\lambda)\sigma_{\epsilon\epsilon} \bar{H}_{II} & 0 \\ 0 & (\phi-\lambda)\sigma_{\epsilon\epsilon} \bar{\Sigma}_{\tilde{V}\tilde{V}} + \bar{B} \end{pmatrix}$$

Box II.

in probability, where $F_{II} = [I_k, I_k, I_k]$ and $\bar{\Upsilon}_{II} = (1-\lambda)\sigma_{\epsilon\epsilon}\left\{\bar{H}_{II} + \bar{\Sigma}_{\tilde{V}\tilde{V}}\right\} + \bar{B}$. Together with Lemma A.8, this leads to the conclusion that in case (II),

$$\sqrt{l}\left(\hat{\beta}^* - \hat{\beta}\right) = \left(\frac{1}{l}\frac{\partial \hat{D}^*(\bar{\beta}^*)}{\partial \beta}\right)^{-1} \frac{\hat{D}^*(\hat{\beta})}{\sqrt{l}} \to^{d^*} N\left(0, \bar{\Lambda}_{II}\right)$$

in probability, where $\bar{\Lambda}_{II} = \bar{H}_{II}^{-1}\bar{\Upsilon}_{II}\bar{H}_{II}^{-1}$.

The results for the bootstrap analogue of FULL follow immediately by using the fact that $\tilde{\lambda}^* = \hat{\lambda}^* + O_{P^*}(1/n)$ in probability. ■

Proof of Corollary 3.1. By Theorem 3.1, we have when $l/r_n \to 0$, $\sqrt{r_n}(\hat{\beta}^* - \hat{\beta}) \to^{d^*} N(0, \sigma_{\epsilon\epsilon}Q^{-1})$ in probability, which is the same as the limiting distribution of $\sqrt{r_n}(\hat{\beta} - \beta)$. The result follows from Polya's Theorem, given that the normal distribution is everywhere continuous. ■

Now we give the proofs for the Lemmas.

Proof of Lemma A.1. We first derive the results for $E^*\left(\|V_i^*\|^8\right)$. Using the Minkowski inequality,

$$\begin{aligned} E^*\left(\|V_i^*\|^8\right) &= \frac{1}{n}\sum_{i=1}^{n}\left\|V_i - \left(\hat{\Pi} - \Pi\right)' Z_i\right\|^8 \\ &\le C\left\{\frac{1}{n}\sum_{i=1}^{n}\|V_i\|^8 + \frac{1}{n}\sum_{i=1}^{n}\left\|\left(\hat{\Pi} - \Pi\right)' Z_i\right\|^8\right\} \\ &= C\left\{\frac{1}{n}\sum_{i=1}^{n}\|V_i\|^8 + \frac{1}{n}\sum_{i=1}^{n}\left\|V'Z\left(Z'Z\right)^{-1} Z_i\right\|^8\right\}. \end{aligned}$$

Since by Assumption 1, $\frac{1}{n}\sum_{i=1}^{n}\|V_i\|^8 = O_P(1)$, it suffices to show that $\frac{1}{n}\sum_{i=1}^{n}\left\|V'Z\left(Z'Z\right)^{-1}Z_i\right\|^8 = O_P(1)$. We are going to show that this holds for each element of the vector $V'Z\left(Z'Z\right)^{-1}Z_i$. Let v be an arbitrary column of V. Then, by the Marcinkiewicz–Zygmund inequality

$$E\left|v'Z\left(Z'Z\right)^{-1}Z_i\right|^8 = E\left|\sum_{j=1}^{n} v_j P_{ji}\right|^8 \le CE\left|\sum_{j=1}^{n} v_j^2 P_{ji}^2\right|^4.$$

Also notice that $\left(\sum_{j=1}^{n} P_{ji}^2\right)^{-4} \ge 1$, then by Jensen's inequality,

$$\begin{aligned} E\left|\sum_{j=1}^{n} v_j^2 P_{ji}^2\right|^4 &\le E\left|\sum_{j=1}^{n} v_j^2 P_{ji}^2 \Big/ \left(\sum_{j=1}^{n} P_{ji}^2\right)\right|^4 \\ &\le E\left[\sum_{j=1}^{n} v_j^8 P_{ji}^2 \Big/ \left(\sum_{j=1}^{n} P_{ji}^2\right)\right] \\ &= \sum_{j=1}^{n} E\left(v_j^8\right) P_{ji}^2 \Big/ \left(\sum_{j=1}^{n} P_{ji}^2\right) \le C. \end{aligned}$$

Thus, $E\left\|v'Z(Z'Z)^{-1}Z_i\right\|^8 = O(1)$. Applying this result to each column of V, we obtain $E\left\|V'Z(Z'Z)^{-1}Z_i\right\|^8 = O(1)$. Thus, by the Markov inequality, $\frac{1}{n}\sum_{i=1}^{n}\left\|V'Z(Z'Z)^{-1}Z_i\right\|^8 = O_P(1)$, and the desired result follows.

For $E^*\left(\epsilon_i^{*8}\right)$, note that by the standard bootstrap d.g.p., $E^*\left(\epsilon_i^{*8}\right) = n^{-1}\sum_{i=1}^{n}\hat{\epsilon}_i$ with $\hat{\epsilon}_i = y_i - X_i'\hat{\beta}$, then the result follows by standard arguments. ■

Proof of Lemma A.2. To prove part (a), note that it suffices to prove that $v^{*'}P_Z\epsilon^*/l = \sigma_{v\epsilon}^b + O_{P^*}\left(1/\sqrt{l}\right)$ as $n \to \infty$, where v^* denotes an arbitrary column of V^* and $\sigma_{v\epsilon}^b$ denotes the corresponding element of $\sigma_{V\epsilon}^b$. Also, let Σ_{vv}^b denote the corresponding element of Σ_{VV}^b.

From the bootstrap d.g.p., we have

$$\begin{aligned} E^*\left(\frac{v^{*'}P_Z\epsilon^*}{l}\right) &= \left(\frac{1}{l}\right) trace\left(P_Z E^*\left(\epsilon^* v^{*'}\right)\right) \\ &= \left(\frac{\sigma_{v\epsilon}^b}{l}\right) trace(P_Z) = \sigma_{v\epsilon}^b \end{aligned}$$

because $E^*\left(\epsilon_i^* v_j^*\right) = E^*\left(\epsilon_i^*\right)E^*\left(v_j^*\right) = 0$ for $i \neq j$ by the property of i.i.d. bootstrap scheme.

Furthermore, note that

$$\begin{aligned} & E^*\left(\frac{v^{*'}P_Z\epsilon^*}{l} - \sigma_{v\epsilon}^b\right)^2 \\ &= \frac{1}{l^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} P_{ij}P_{kl}E^*\left(v_i^*\epsilon_j^*v_k^*\epsilon_l^*\right) \\ &\quad - \left(\frac{2\sigma_{v\epsilon}^b}{l}\right)\sum_{i=1}^{n}\sum_{j=1}^{n} P_{ij}E^*\left(v_i^*\epsilon_j^*\right) + \left(\sigma_{v\epsilon}^b\right)^2 \\ &= \frac{1}{l^2}E^*\left(v_i^{*2}\epsilon_i^{*2}\right)\left[\sum_{i=1}^{n} P_{ii}^2\right] + \frac{2}{l^2}\left(\Sigma_{vv}^b\sigma_{\epsilon\epsilon}^b\right)\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1} P_{ij}^2\right] \\ &\quad + \left\{\frac{2}{l^2}\left(\sigma_{v\epsilon}^b\right)^2\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1}\left(P_{ii}P_{jj} + P_{ij}^2\right)\right] - \left(\sigma_{v\epsilon}^b\right)^2\right\} \\ &\equiv L_1 + L_2 + L_3. \end{aligned}$$

The second equality follows from noting that $E^*\left(v_i^*\epsilon_j^*V_k^*\epsilon_l^*\right)$ equals zero except in the case where either $(i = j = k = l)$ or $(i = k, j = l)$ or $(i = j, k = l)$ or $(i = l, j = k)$ and from using $\sum_{i=1}^{n} P_{ii} = l$.

Focusing on L_1 first, notice that

$$\begin{aligned} L_1 &\le \frac{1}{l^2}\left(E^*\left(v_i^{*4}\right)\right)^{1/2}\left(E^*\left(\epsilon_i^{*4}\right)\right)^{1/2}\left[\sum_{i=1}^{n} P_{ii}^2\right] \\ &\le \frac{1}{l}\left(E^*\left(v_i^{*4}\right)\right)^{1/2}\left(E^*\left(\epsilon_i^{*4}\right)\right)^{1/2} = O_P\left(\frac{1}{l}\right) \end{aligned}$$

where the first inequality follows from the Cauchy–Schwarz inequality, and the second inequality follows from using $\sum_{i=1}^{n}(P_{ii})^2 \le \sum_{i=1}^{n} P_{ii} = l$. The last equality follows from using the same arguments as in Lemma A.1.

Next, for L_2, we have

$$L_2 \le \left(\frac{\Sigma_{vv}^b \sigma_{\epsilon\epsilon}^b}{l^2}\right)\left[\sum_{i=1}^n P_{ii}^2 + 2\sum_{i=2}^n \sum_{j=1}^{i-1} P_{ij}^2\right] = \frac{\Sigma_{vv}^b \sigma_{\epsilon\epsilon}^b}{l} = O_P\left(\frac{1}{l}\right)$$

because $\sum_{i=1}^n (P_{ii})^2 + 2\sum_{i=2}^n \sum_{j=1}^{i-1}(P_{ij})^2 = Tr(P_Z' P_Z) = Tr(P_Z) = l$ given that P_Z is symmetric and idempotent.

Finally, for L_3, we note that

$$|L_3| = \left|\frac{(\sigma_{v\epsilon}^b)^2}{l^2}\left[(Tr(P_Z))^2 + Tr(P_Z'P_Z) - 2\sum_{i=1}^n P_{ii}^2\right] - (\sigma_{v\epsilon}^b)^2\right|$$

$$= \left|\frac{(\sigma_{v\epsilon}^b)^2}{l^2}\left(l - 2\sum_{i=1}^n P_{ii}^2\right)\right| \le \frac{(\sigma_{v\epsilon}^b)^2}{l} + \frac{2(\sigma_{v\epsilon}^b)^2 \sum_{i=1}^n P_{ii}}{l^2}$$

$$= O_P\left(\frac{1}{l}\right).$$

Therefore, we obtain $E^*\left(\frac{v^{*'}P_Z\epsilon^*}{l} - \sigma_{v\epsilon}^b\right)^2 = O_P\left(\frac{1}{l}\right)$.

But, for any T^* such that $Var^*(T^*) = O_P(1/l)$, where Var^* denotes the variance computed under P^*, by the Chebychev's inequality, we have for any $\delta > 0$ and any fixed $M_\delta > 0$,

$$P^*\left(|\sqrt{l}T^*| > M_\delta\right) \le \frac{1}{M_\delta^2} Var^*\left(\sqrt{l}T^*\right) = \left(\frac{1}{M_\delta^2}\right) O_P(1).$$

Also, by the definition of $O_P(1)$, for δ, there exists a $M_\delta' < \infty$ such that

$$\lim_{n\to\infty} P\left(|\ O_P(1)\ | > M_\delta'\right) = 0.$$

If we take $M_\delta = \sqrt{\frac{M_\delta'}{\delta}}$, i.e. $M_\delta^2 = \frac{M_\delta'}{\delta}$, then,

$$\lim_{n\to\infty} P\left(\left|\frac{1}{M_\delta^2} O_P(1)\right| > \delta\right) = \lim_{n\to\infty} P\left(\frac{\delta}{M_\delta'}|O_P(1)| > \delta\right)$$
$$= \lim_{n\to\infty} P\left(|O_P(1)| > M_\delta'\right) = 0.$$

Then, it follows that $P^*\left(|\sqrt{l}T^*| > M_\delta\right) = o_P(1)$, i.e. $T^* = O_{P^*}\left(1/\sqrt{l}\right)$. Therefore it follows that $v^{*'}P_Z\epsilon^*/l - \sigma_{v\epsilon}^b = O_{P^*}\left(1/\sqrt{l}\right)$ in probability, as required. This proves part (a). Parts (b) and (c) follow from proof similar to that of part (a).

The proofs for parts (d) and (e) are similar, so we will only prove (d). To proceed, note that by the properties of Expectation and Trace operator,

$$E^*\left(\left\|\frac{V^{*'}Z\hat{\Pi}}{r_n}\right\|^2\right) = E^*\left(trace\left(\frac{\hat{\Pi}'Z'V^*V^{*'}Z\hat{\Pi}}{r_n^2}\right)\right)$$
$$= \frac{trace\left(\Sigma_{VV}^b\right)}{r_n}\left(trace\left(\frac{\hat{\Pi}'Z'Z\hat{\Pi}}{r_n}\right)\right).$$

When $l/r_n \to \gamma < \infty$, by $\hat{\Pi} = (Z'Z)^{-1}Z'X$ we have

$$\frac{\hat{\Pi}'Z'Z\hat{\Pi}}{r_n} = \frac{\Pi'Z'Z\Pi}{r_n} + \frac{V'Z\Pi}{r_n} + \frac{\Pi'Z'V}{r_n} + \frac{V'P_ZV}{r_n}$$
$$= Q + \left(\frac{l}{r_n}\right)\Sigma_{VV} + O_P\left(\frac{1}{\sqrt{r_n}}\right) = O_P(1).$$

Thus, $E^*\left(\left\|\frac{V^{*'}Z\hat{\Pi}}{r_n}\right\|^2\right) = \left(\frac{1}{r_n}\right)O_P(1)O_P(1) = O_P\left(\frac{1}{r_n}\right)$ because Σ_{VV}^b is bounded in probability by using argument similar to the proofs of Lemma A.1. It follows that $\hat{\Pi}'Z'V^*/r_n = O_{P^*}\left(1/\sqrt{r_n}\right)$.

The proofs for parts (d') and (e') are similar, so we will only prove (d'). Similar to part (d), we have

$$E^*\left(\left\|\frac{V^{*'}Z\hat{\Pi}}{l}\right\|^2\right) = \frac{trace\left(\Sigma_{VV}^b\right)}{l}\left(trace\left(\frac{\hat{\Pi}'Z'Z\hat{\Pi}}{l}\right)\right).$$

When $l/r_n \to \infty$,

$$\frac{\hat{\Pi}'Z'Z\hat{\Pi}}{l} = \frac{\Pi'Z'Z\Pi}{l} + \frac{V'Z\Pi}{l} + \frac{\Pi'Z'V}{l} + \frac{V'P_ZV}{l}$$
$$= \Sigma_{VV} + O_P\left(\frac{r_n}{l}\right) + O_P\left(\frac{\sqrt{r_n}}{l}\right) + O_P\left(\frac{1}{\sqrt{l}}\right) = O_P(1).$$

Thus, we obtain $E^*\left(\left\|\frac{V^{*'}Z\hat{\Pi}}{l}\right\|^2\right) = O_P\left(\frac{1}{l}\right)$, and it follows that $\hat{\Pi}'Z'V^*/l = O_{P^*}\left(1/\sqrt{l}\right)$ in this case. ■

Proof of Lemma A.3. The proof follows closely from Lemma A11 of Hansen et al. (2008) by replacing their a_i with a_i^* and $\tilde{V}$ with $\tilde{V}^*$. ■

Proof of Lemma A.4. For Case (I), write

$$\hat{\beta}^* - \hat{\beta} = \left(\frac{X^{*'}P_ZX^*}{r_n} - \hat{\lambda}^*\frac{X^{*'}X^*}{r_n}\right)^{-1}\left(\frac{X^{*'}P_Z\epsilon^*}{r_n} - \hat{\lambda}^*\frac{X^{*'}\epsilon^*}{r_n}\right).$$

First, we needs to show that $\hat{\lambda}^* = \lambda_n + o_{P^*}(r_n/n)$ in probability.

Let $Y^* = [y^*, X^*]$, $\eta^* = \left[\epsilon^* + V^*\hat{\beta}, V^*\right]$, $\widehat{Z} = Z\hat{\Pi}\left[\hat{\beta}, I_k\right]$, and define $\Omega_{\eta\eta}^b \equiv E^*\left(\eta_i^*\eta_i^{*'}\right)$. Note that using similar arguments as in the proof of Lemma A.2, we can obtain: $\eta^{*'}\eta^*/n = \Omega_{\eta\eta}^b + O_{P^*}\left(1/\sqrt{n}\right)$; $\eta^{*'}P_Z\eta^*/l = \Omega_{\eta\eta}^b + O_{P^*}\left(1/\sqrt{l}\right)$; $\widehat{Z}'\eta^*/r_n = \left[\hat{\beta}, I_k\right]'\left[\hat{\Pi}'Z'\epsilon^*/r_n, \hat{\Pi}'Z'V^*/r_n\right]\begin{pmatrix}1 & 0\\ \hat{\beta} & I_k\end{pmatrix} = O_{P^*}\left(1/\sqrt{r_n}\right)$, in probability.

Now, let $\widehat{\Delta}^* = Y^{*'}P_ZY^*/l$, $\widehat{\Theta}^* = Y^{*'}Y^*/n$; $\Delta^b = \Omega_{\eta\eta}^b + (r_n/l)\widehat{\Phi}_I$, $\Theta^b = \Omega_{\eta\eta}^b + (r_n/n)\widehat{\Phi}_I$, where $\widehat{\Phi}_I = \left[\hat{\beta}, I_k\right]'\{\Psi + (l/r_n)\Sigma_{VV}\}\left[\hat{\beta}, I_k\right]$.

Note that for the bootstrap analogue of LIML, $\hat{\lambda}^* = \lambda_n \times \min_{\|\alpha\|=1}\frac{\alpha'\widehat{\Delta}^*\alpha}{\alpha'\widehat{\Theta}^*\alpha}$. Thus, it suffices to show that $\left|\min_{\|\alpha\|=1}\frac{\alpha'\widehat{\Delta}^*\alpha}{\alpha'\widehat{\Theta}^*\alpha} - 1\right| = o_{P^*}(r_n/l)$, in probability. By the triangle inequality,

$$\left\|\widehat{\Delta}^* - \Delta^b\right\| \le \left\|\frac{\eta^{*'}P_Z\eta^*}{l} - \Omega_{\eta\eta}^b\right\| + 2\left\|\frac{\widehat{Z}'\eta^*}{l}\right\| + \frac{r_n}{l}\left\|\frac{\widehat{Z}'\widehat{Z}}{r_n} - \widehat{\Phi}_I\right\|$$
$$= O_{P^*}\left(\frac{1}{\sqrt{l}}\right) + O_{P^*}\left(\frac{\sqrt{r_n}}{l}\right) + o_P\left(\frac{r_n}{l}\right)$$
$$= O_{P^*}\left(\left(\frac{r_n}{l}\right)\left(\frac{\sqrt{l}}{r_n}\right)\right) + o_{P^*}\left(\frac{r_n}{l}\right) = o_{P^*}\left(\frac{r_n}{l}\right)$$

in probability, given that $\sqrt{l}/r_n \to 0$ as $n \to \infty$. Using similar arguments, we obtain that

$$\left\|\widehat{\Theta}^* - \Theta^b\right\| \le \left\|\frac{\eta^{*'}\eta^*}{n} - \Omega_{\eta\eta}^b\right\| + 2\left\|\frac{\widehat{Z}'\eta^*}{n}\right\| + \frac{r_n}{n}\left\|\frac{\widehat{Z}'\widehat{Z}}{r_n} - \widehat{\Phi}_I\right\|$$
$$= O_{P^*}\left(\frac{1}{\sqrt{n}}\right) + O_{P^*}\left(\frac{\sqrt{r_n}}{n}\right) + o_P\left(\frac{r_n}{n}\right) = o_{P^*}\left(\frac{r_n}{l}\right)$$

in probability. Moreover, using the Cauchy–Schwarz inequality, we obtain that, for all α with $\|\alpha\| = 1$,

$$\left|\alpha'\widehat{\Delta}^*\alpha - \alpha'\Delta^b\alpha\right| = \left|trace\left[(\widehat{\Delta}^* - \Delta^b)\alpha\alpha'\right]\right|$$
$$\le \sqrt{trace\left[(\widehat{\Delta}^* - \Delta^b)(\widehat{\Delta}^* - \Delta^b)'\right]}\sqrt{trace(\alpha\alpha'\alpha\alpha')}$$
$$= \left\|\widehat{\Delta}^* - \Delta^b\right\| = o_{P^*}\left(\frac{r_n}{l}\right)$$

in probability, given that $\sqrt{trace(\alpha\alpha'\alpha\alpha')} = 1$. Similarly, we have $\left|\alpha'\widehat{\Theta}^*\alpha - \alpha'\Theta^b\alpha\right| = o_{P^*}(r_n/l)$, in probability.

It follows that

$$\left|\frac{\alpha'\widehat{\Delta}^*\alpha}{\alpha'\widehat{\Theta}^*\alpha} - \frac{\alpha'\Delta^b\alpha}{\alpha'\Theta^b\alpha}\right| \le \left|\alpha'\widehat{\Theta}^*\alpha\right|^{-1}\left\{\left|\alpha'\widehat{\Delta}^*\alpha - \alpha'\Delta^b\alpha\right| + \left(\frac{\alpha'\Delta^b\alpha}{\alpha'\Theta^b\alpha}\right)\left|\alpha'\widehat{\Theta}^*\alpha - \alpha'\Theta^b\alpha\right|\right\} \to^{P^*} 0 \tag{16}$$

in probability, given that $\alpha'\widehat{\Theta}^*\alpha = O_{P^*}(1)$ in probability, and $\alpha'\Delta^b\alpha/\alpha'\Theta^b\alpha = O_P(1)$. In addition, note that $\Delta^b = \Theta^b + (1 - \lambda_n)(r_n/l)\widehat{\Phi}_I$, so that

$$\alpha'\Delta^b\alpha/\alpha'\Theta^b\alpha = 1 + \left((1-\lambda_n)(r_n/l)\alpha'\widehat{\Phi}_I\alpha\right)/\alpha'\Theta^b\alpha \ge 1,$$

with probability approaching one. Define $\hat{\alpha} = (1, -\hat{\beta}')'/\|(1, -\hat{\beta}')\|$; we have $\widehat{\Phi}_I\hat{\alpha} = 0$ for all n, thus $\min_{\|\alpha\|=1}\frac{\alpha'\Delta^b\alpha}{\alpha'\Theta^b\alpha} = \frac{\hat{\alpha}'\Delta^b\hat{\alpha}}{\hat{\alpha}'\Theta^b\hat{\alpha}} = 1$, with probability approaching one.

Let $\hat{\alpha}^* = argmin_{\|\alpha\|=1}\frac{\alpha'\widehat{\Delta}^*\alpha}{\alpha'\widehat{\Theta}^*\alpha}$, and we obtain that

$$\left|\min_{\|\alpha\|=1}\frac{\alpha'\widehat{\Delta}^*\alpha}{\alpha'\widehat{\Theta}^*\alpha} - 1\right| = \left|\frac{\hat{\alpha}^{*'}\widehat{\Delta}^*\hat{\alpha}^*}{\hat{\alpha}^{*'}\widehat{\Theta}^*\hat{\alpha}^*} - 1\right|$$
$$= \left|\frac{\hat{\alpha}^{*'}\widehat{\Delta}^*\hat{\alpha}^*}{\hat{\alpha}^{*'}\widehat{\Theta}^*\hat{\alpha}^*} - \frac{\hat{\alpha}'\Delta^b\hat{\alpha}}{\hat{\alpha}'\Theta^b\hat{\alpha}}\right| \le \max_{\|\alpha\|=1}\left|\frac{\alpha'\widehat{\Delta}^*\alpha}{\alpha'\widehat{\Theta}^*\alpha} - \frac{\alpha'\Delta^b\alpha}{\alpha'\Theta^b\alpha}\right|$$
$$= o_{P^*}\left(\frac{r_n}{l}\right)$$

in probability. This verifies the condition that $\hat{\lambda}^* = \lambda_n + o_{P^*}(r_n/n)$, in probability.

Then, we obtain that

$$\left(\hat{\lambda}^* - \lambda_n\right)\frac{X^{*'}X^*}{r_n} = o_{P^*}\left(\frac{r_n}{n}\right)O_{P^*}\left(\frac{n}{r_n}\right) \to^{P^*} 0$$

in probability, by showing that $X^{*'}X^* = O_{P^*}(n)$ in probability, using standard arguments. Similarly, $\left(\hat{\lambda}^* - \lambda_n\right)\frac{X^{*'}\epsilon^*}{r_n} \to^{P^*} 0$, in probability. Therefore, for the denominator

$$\frac{X^{*'}P_ZX^*}{r_n} - \hat{\lambda}^*\frac{X^{*'}X^*}{r_n} = \frac{X^{*'}P_ZX^*}{r_n} - \lambda_n\frac{X^{*'}X^*}{r_n} + o_{P^*}(1)$$
$$= (1-\lambda_n)\frac{\hat{\Pi}'Z'Z\hat{\Pi}}{r_n} + (1-\lambda_n)\left(\frac{\hat{\Pi}'Z'V^*}{r_n} + \frac{V^{*'}Z\hat{\Pi}}{r_n}\right) + \frac{V^{*'}P_ZV^*}{r_n} - \left(\frac{l}{r_n}\right)\frac{V^{*'}V^*}{n} + o_{P^*}(1)$$
$$= (1-\lambda_n)\frac{\hat{\Pi}'Z'Z\hat{\Pi}}{r_n} + \left(\frac{l}{r_n}\right)\Sigma^b_{VV} - \left(\frac{l}{r_n}\right)\Sigma^b_{VV} + o_{P^*}(1)$$
$$= O_{P^*}(1)$$

in probability, where the last equality follows from Lemma A.2 and from $\hat{\Pi}'Z'Z\hat{\Pi}/r_n = O_P(1)$ in Case (I). Similarly,

$$\frac{X^{*'}P_Z\epsilon^*}{r_n} - \hat{\lambda}^*\frac{X^{*'}\epsilon^*}{r_n} = \frac{X^{*'}P_Z\epsilon^*}{r_n} - \lambda_n\frac{X^{*'}\epsilon^*}{r_n} + o_{P^*}(1)$$
$$= (1-\lambda_n)\frac{\hat{\Pi}'Z'\epsilon^*}{r_n} + \frac{V^{*'}P_Z\epsilon^*}{r_n} - \lambda_n\frac{V^{*'}\epsilon^*}{r_n} + o_{P^*}(1)$$
$$= o_{P^*}(1) + \left(\frac{l}{r_n}\right)\sigma^b_{V\epsilon} - \left(\frac{l}{r_n}\right)\sigma^b_{V\epsilon} + o_{P^*}(1) \to^{P^*} 0$$

in probability. Putting these results together, we obtain $\hat{\beta}^* - \hat{\beta} = O_{P^*}(1)o_{P^*}(1) \to^{P^*} 0$ in probability, as required.

For Case (II), write

$$\hat{\beta}^* - \hat{\beta} = \left(\frac{X^{*'}P_ZX^*}{l} - \hat{\lambda}^*\frac{X^{*'}X^*}{l}\right)^{-1}\left(\frac{X^{*'}P_Z\epsilon^*}{l} - \hat{\lambda}^*\frac{X^{*'}\epsilon^*}{l}\right),$$

and the desired result follows by using similar arguments as in Case (I).

For the bootstrap analogue of FULL, note that $\tilde{\lambda}^* = \hat{\lambda}^* + O_{P^*}(1/n)$, in probability. Then, the desired result follows immediately. ■

Proof of Lemma A.5. Notice that by the definition of λ^*, we have

$$\lambda^* - \lambda_n = \frac{l}{\epsilon^{*'}\epsilon^*}\left\{\frac{\epsilon^{*'}P_Z\epsilon^*}{l} - \sigma^b_{\epsilon\epsilon} - \left(\frac{\epsilon^{*'}\epsilon^*}{n} - \sigma^b_{\epsilon\epsilon}\right)\right\}$$
$$= O_{P^*}\left(\frac{l}{n}\right)\left\{O_{P^*}\left(\frac{1}{\sqrt{l}}\right) + O_{P^*}\left(\frac{1}{\sqrt{n}}\right)\right\} = O_{P^*}\left(\frac{\sqrt{l}}{n}\right)$$

in probability, where the second equality follows from $\epsilon^{*'}\epsilon^* = O_{P^*}(n)$ in probability and from using similar arguments as for Lemma A.2. ■

Proof of Lemma A.6. Note that by standard arguments, $X^{*'}X^* = O_{P^*}(n)$ and $X^{*'}\epsilon^* = O_{P^*}(n)$. Therefore, in Case (I), $(\hat{\lambda}^* - \lambda^*)\frac{X^{*'}X^*}{r_n} = O_{P^*}\left(\frac{\delta^{\lambda}_n n}{r_n}\right)$ and $(\hat{\lambda}^* - \lambda^*)\frac{X^{*'}\epsilon^*}{r_n} = O_{P^*}\left(\frac{\delta^{\lambda}_n n}{r_n}\right)$. Also, by Lemma A.5 and by $l/r_n \to \gamma < \infty$ in Case (I)

$$(\lambda^* - \lambda_n)\frac{X^{*'}X^*}{r_n} = O_{P^*}\left(\frac{\sqrt{l}}{n}\cdot\frac{n}{r_n}\right) = O_{P^*}\left(\frac{\sqrt{l}}{r_n}\right) = O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right).$$

Similarly, $(\lambda^* - \lambda_n)\frac{X^{*'}\epsilon^*}{r_n} = O_{P^*}\left(\frac{\sqrt{l}}{n}\cdot\frac{n}{r_n}\right) = O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$. Also, we obtain by $X^* = Z\hat{\Pi} + V^*$ and by Lemma A.2 that $\frac{1}{r_n}\left(X^{*'}P_ZX^* - \lambda_nX^{*'}X^*\right) = \bar{H}_{I,n} + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$ and $\frac{1}{r_n}(X^{*'}P_Z\epsilon^* - \lambda_nX^{*'}\epsilon^*) = O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$. Putting these results together, we obtain

$$\frac{1}{r_n}\left(X^{*'}P_ZX^* - \hat{\lambda}^*X^{*'}X^*\right) = \bar{H}_{I,n} + O_{P^*}\left(\frac{1}{\sqrt{r_n}} + \frac{\delta^{\lambda}_n n}{r_n}\right)$$
$$\frac{1}{r_n}\left(X^{*'}P_Z\hat{\epsilon}^*(\hat{\beta}^*) - \hat{\lambda}^*X^{*'}\hat{\epsilon}^*(\hat{\beta}^*)\right) = O_{P^*}\left(\frac{1}{\sqrt{r_n}} + \delta^{\beta}_n + \frac{\delta^{\lambda}_n n}{r_n}\right)$$

by the triangle inequality and by the fact that

$$\frac{1}{r_n}\left(X^{*'}P_Z\hat{\epsilon}^*(\hat{\beta}^*) - \hat{\lambda}^*X^{*'}\hat{\epsilon}^*(\hat{\beta}^*)\right) = \frac{1}{r_n}\left(X^{*'}P_Z\epsilon^* - \hat{\lambda}^*X^{*'}\epsilon^*\right) - \left(\frac{1}{r_n}(X^{*'}P_ZX^* - \hat{\lambda}^*X^{*'}X^*)\right)(\hat{\beta}^* - \hat{\beta}).$$

For Case (II), we have $(\hat{\lambda}^* - \lambda^*)\frac{X^{*'}X^*}{l} = O_{P^*}\left(\frac{\delta^{\lambda}_n n}{l}\right)$, $(\hat{\lambda}^* - \lambda^*)\frac{X^{*'}\epsilon^*}{l} = O_{P^*}\left(\frac{\delta^{\lambda}_n n}{l}\right)$, $(\lambda^* - \lambda_n)\frac{X^{*'}X^*}{l} = O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$, and $(\lambda^* - \lambda_n)\frac{X^{*'}\epsilon^*}{l} = O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$. Also, by Lemma A.2, $\frac{1}{l}\left(X^{*'}P_ZX^* - \lambda_nX^{*'}X^*\right) = \bar{H}_{II,n} + O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$ and $\frac{1}{l}(X^{*'}P_Z\epsilon^* - \lambda_nX^{*'}\epsilon^*) = O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$. Then, the conclusion follows by the triangle inequality. ■

Proof of Lemma A.7. Let $\hat{\sigma}^*_{\epsilon\epsilon}(\hat{\beta}^*) = \hat{\epsilon}^{*'}(\hat{\beta}^*)\hat{\epsilon}^*(\hat{\beta}^*)/n$, then for Case (I),

$$\begin{aligned}
&\frac{\hat{\epsilon}^{*'}(\hat{\beta}^*)P_Z\hat{\epsilon}^*(\hat{\beta}^*)}{\hat{\epsilon}^{*'}(\hat{\beta}^*)\hat{\epsilon}^*(\hat{\beta}^*)} - \frac{\epsilon^{*'}P_Z\epsilon^*}{\epsilon^{*'}\epsilon^*}\\
&= \frac{1}{\hat{\epsilon}^{*'}(\hat{\beta}^*)\hat{\epsilon}^*(\hat{\beta}^*)}\left\{\left(\hat{\epsilon}^{*'}(\hat{\beta}^*)P_Z\hat{\epsilon}^*(\hat{\beta}^*) - \epsilon^{*'}P_Z\epsilon^*\right) - \lambda^*\left(\hat{\epsilon}^{*'}(\hat{\beta}^*)\hat{\epsilon}^*(\hat{\beta}^*) - \epsilon^{*'}\epsilon^*\right)\right\}\\
&= \frac{r_n}{n\hat{\sigma}^*_{\epsilon\epsilon}(\hat{\beta}^*)}\left\{(\hat{\beta}^* - \hat{\beta})'\left(\frac{X^{*'}P_ZX^* - \lambda^*X^{*'}X^*}{r_n}\right)(\hat{\beta}^* - \hat{\beta}) - 2(\hat{\beta}^* - \hat{\beta})'\left(\frac{X^{*'}P_Z\epsilon^* - \lambda^*X^{*'}\epsilon^*}{r_n}\right)\right\} = O_{P^*}\left(\frac{r_n(\delta_n^\beta)^2}{n}\right)
\end{aligned}$$

because $\left(\hat{\sigma}^*_{\epsilon\epsilon}(\hat{\beta}^*)\right)^{-1} = O_{P^*}(1)$, $\frac{1}{r_n}\left(X^{*'}P_ZX^* - \lambda^*X^{*'}X^*\right) = O_{P^*}(1)$, $\frac{1}{r_n}\left(X^{*'}P_Z\epsilon^* - \lambda^*X^{*'}\epsilon^*\right) = O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$ by Lemma A.6 with $\hat{\lambda}^* = \lambda^*$ and $\delta_n^\lambda = \delta_n^\beta = 0$.

In analogy, we obtain that for Case (II), $\frac{\hat{\epsilon}^{*'}(\hat{\beta}^*)P_Z\hat{\epsilon}^*(\hat{\beta}^*)}{\hat{\epsilon}^{*'}(\hat{\beta}^*)\hat{\epsilon}^*(\hat{\beta}^*)} - \frac{\epsilon^{*'}P_Z\epsilon^*}{\epsilon^{*'}\epsilon^*} = O_{P^*}\left(\frac{l(\delta_n^\beta)^2}{n}\right)$. ■

Proof of Lemma A.8. Let $\bar{\epsilon}^* = y^* - X^*\bar{\beta}$ and $\bar{q}^* = X^*\bar{\epsilon}^*/\bar{\epsilon}^{*'}\bar{\epsilon}^*$. Suppose $\bar{\beta}^*$ lies between $\hat{\beta}^*$ and $\hat{\beta}$. Then differentiating gives

$$\begin{aligned}
-\left(\partial\hat{D}^*(\bar{\beta}^*)/\partial\beta\right) &= X^{*'}P_ZX^* - \frac{\bar{\epsilon}^{*'}P_Z\bar{\epsilon}^*}{\bar{\epsilon}^{*'}\bar{\epsilon}^*}X^{*'}X^* - X^{*'}\bar{\epsilon}^*\frac{\bar{\epsilon}^{*'}P_ZX^*}{\bar{\epsilon}^{*'}\bar{\epsilon}^*}\\
&\quad - \frac{X^{*'}P_Z\bar{\epsilon}^*}{\bar{\epsilon}^{*'}\bar{\epsilon}^*}\bar{\epsilon}^{*'}X^* + 2\frac{\bar{\epsilon}^{*'}P_Z\bar{\epsilon}^*}{(\bar{\epsilon}^{*'}\bar{\epsilon}^*)^2}X^{*'}\bar{\epsilon}^*\bar{\epsilon}^{*'}X^*\\
&= X^{*'}P_ZX^* - \bar{\lambda}^*X^{*'}X^* + \bar{q}^*\hat{D}^*(\bar{\beta}^*)' + \hat{D}^*(\bar{\beta}^*)\bar{q}^{*'}
\end{aligned}$$

where $\bar{\lambda}^* = \bar{\epsilon}^{*'}P_Z\bar{\epsilon}^*/\bar{\epsilon}^{*'}\bar{\epsilon}^*$. By Lemma A.7, we have $\bar{\lambda}^* = \lambda^* + O_{P^*}\left(\frac{(\delta_n^\beta)^2r_n}{n}\right)$ for Case (I). Then by Lemma A.6 with $\delta_n^\lambda = \frac{(\delta_n^\beta)^2r_n}{n}$, we obtain that $\frac{1}{r_n}\left(X^{*'}P_ZX^* - \bar{\lambda}^*X^{*'}X^*\right) = \bar{H}_{I,n} + O_{P^*}\left(\frac{1}{\sqrt{r_n}} + (\delta_n^\beta)^2\right)$ and $\frac{\hat{D}^*(\bar{\beta}^*)}{r_n} = O_{P^*}\left(\frac{1}{\sqrt{r_n}} + \delta_n^\beta\right)$. Note that by standard argument $\bar{q}^* = O_{P^*}(1)$, hence $\frac{1}{r_n}\left(\hat{D}^*(\bar{\beta}^*)\bar{q}^*\right) = \frac{1}{r_n}\hat{D}^*(\bar{\beta}^*)O_{P^*}(1) = O_{P^*}\left(\frac{1}{\sqrt{r_n}} + \delta_n^\beta\right)$.
The conclusion then follows by the triangle inequality. For Case (II), note that by similar arguments, one can obtain $\frac{1}{l}\left(X^{*'}P_ZX^* - \bar{\lambda}^*X^{*'}X^*\right) = \bar{H}_{II,n} + O_{P^*}\left(\frac{1}{\sqrt{l}} + (\delta_n^\beta)^2\right)$, $\frac{1}{l}\hat{D}^*(\bar{\beta}^*) = O_{P^*}\left(\frac{1}{\sqrt{l}} + \delta_n^\beta\right)$, and $\frac{1}{l}\left(\hat{D}^*(\bar{\beta}^*)\bar{q}^*\right) = O_{P^*}\left(\frac{1}{\sqrt{l}} + \delta_n^\beta\right)$. ■

Proof of Lemma A.9. Note that by Lemma A.5, $\lambda^* = \lambda_n + O_{P^*}\left(\sqrt{l}/n\right)$. Also, $\tilde{V}^{*'}\epsilon^*/n = O_{P^*}\left(1/\sqrt{n}\right)$ in probability, since $E^*\left(\tilde{V}_i^*\epsilon_i^*\right) = 0$. Moreover, for Case (I) we have $\hat{\Pi}'Z'\epsilon^*/\sqrt{r_n} = O_{P^*}\left(1/\sqrt{r_n}\right)$ in probability. Thus

$$\begin{aligned}
\frac{\hat{D}^*(\hat{\beta})}{\sqrt{r_n}} &= \frac{1}{\sqrt{r_n}}\left(X^{*'}P_Z\varepsilon^* - \lambda^*X^{*'}\varepsilon^*\right)\\
&= \frac{1}{\sqrt{r_n}}\left\{\left(X^* - \epsilon^*q^{b'}\right)'P_Z\epsilon^* - \lambda^*\left(X^* - \epsilon^*q^{b'}\right)'\epsilon^*\right\}\\
&= \frac{1}{\sqrt{r_n}}\left\{\hat{\Pi}'Z'\epsilon^* + \tilde{V}^{*'}P_Z\epsilon^* - \left(Z\hat{\Pi} + \tilde{V}^*\right)'\epsilon^*\left[\lambda_n + O_{P^*}\left(\frac{\sqrt{l}}{n}\right)\right]\right\}\\
&= \frac{1}{\sqrt{r_n}}\left((1-\lambda_n)Z\hat{\Pi} + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right)'\epsilon^* + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)
\end{aligned}$$

in probability, where the last equality follows by noting that

$$\begin{aligned}
\frac{1}{\sqrt{r_n}}\left(Z\hat{\Pi} + \tilde{V}^*\right)'\epsilon^*O_{P^*}\left(\frac{\sqrt{l}}{n}\right) &= O_{P^*}\left(\sqrt{\frac{n}{r_n}}\right)O_{P^*}\left(\frac{\sqrt{l}}{n}\right)\\
&= O_{P^*}\left(\sqrt{\frac{\lambda_n}{r_n}}\right)
\end{aligned}$$

and $\lambda_n \to \lambda \in [0, 1)$ as $n \to \infty$. Using similar arguments, we obtain for Case (II)

$$\frac{\hat{D}^*(\hat{\beta})}{\sqrt{l}} = \frac{1}{\sqrt{l}}\left((1-\lambda_n)Z\hat{\Pi} + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right)'\epsilon^* + O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$$

in probability. ■

Appendix B. Proofs of results for the RE bootstrap

All the proofs of the Lemmas are relegated at the end of Appendix B. Let $\hat{\beta}^* = \hat{\beta}^*_{re}$ throughout Appendix B. let C denote a generic positive constant that may be different in different uses. Also, P^* denotes the probability measure induced by the RE bootstrap procedure and E^* denotes the expectation under P^*.

Lemma B.1. *Suppose that Assumptions 1–2 hold, then under H_0 : $\beta = \beta_0$, $E^*\left(\epsilon_i^{*8}\right)$ and $E^*\left(\left\|V_i^*\right\|^8\right)$ are bounded in probability.*

Lemma B.2. *Suppose that Assumptions 1–2 hold, then the following statements are true under H_0:*

(a) $V^{*'}P_Z\epsilon^*/l = \sigma^b_{V\epsilon} + O_{P^*}\left(1/\sqrt{l}\right)$, (b) $V^{*'}P_ZV^*/l = \Sigma^b_{VV} + O_{P^*}\left(1/\sqrt{l}\right)$, (c) $\epsilon^{*'}P_Z\epsilon^*/l = \sigma^b_{\epsilon\epsilon} + O_{P^*}\left(1/\sqrt{l}\right)$, in probability; in Case (I), (d) $\tilde{\Pi}'(\beta_0)Z'V^*/r_n = O_{P^*}\left(1/\sqrt{r_n}\right)$, (e) $\tilde{\Pi}'(\beta_0)Z'\epsilon^*/r_n = O_{P^*}\left(1/\sqrt{r_n}\right)$, in probability; in Case (II), (d′) $\tilde{\Pi}'(\beta_0)Z'V^*/l = O_{P^*}\left(1/\sqrt{l}\right)$, (e′) $\tilde{\Pi}'(\beta_0)Z'\epsilon^*/l = O_{P^*}\left(1/\sqrt{l}\right)$, in probability; where $\sigma^b_{V\epsilon} \equiv E^*\left(V_i^*\epsilon_i^*\right)$, $\Sigma^b_{VV} \equiv E^*\left(V_i^*V_i^{*'}\right)$ and $\sigma^b_{\epsilon\epsilon} \equiv E^*\left(\epsilon_i^{*2}\right)$.

Let $J^* = diag(a_1^*, \ldots, a_n^*)$ where $a_i^* = \epsilon_i^{*2} - \sigma^b_{\epsilon\epsilon}$, $i = 1, \ldots, n$.

Lemma B.3. *Suppose that Assumptions 1–2 hold, then the following statements are true under H_0:*

(a) $\tilde{V}^{*'}J^*\tilde{V}^*/n = E^*\left(a_i^*\tilde{V}_i^*\tilde{V}_i^{*'}\right) + O_{P^*}\left(1/\sqrt{n}\right)$; (b) $\tilde{V}^{*'}P_ZJ^*\tilde{V}^*/n = \lambda_nE^*\left(a_i^*\tilde{V}_i^*\tilde{V}_i^{*'}\right) + O_{P^*}\left(1/\sqrt{n}\right)$; (c) $\tilde{V}^{*'}P_ZJ^*P_Z\tilde{V}^*/n = \lambda_n\phi_nE^*\left(a_i^*\tilde{V}_i^*\tilde{V}_i^{*'}\right) + O_{P^*}\left(1/\sqrt{n}\right)$, in probability.

Lemma B.4. *Suppose that Assumptions 1–2 hold, then under H_0, $\hat{\beta}^* - \beta_0 = o_{P^*}(1)$, in probability.*

Let $\lambda^* = \epsilon^{*'}P_Z\epsilon^*/\epsilon^{*'}\epsilon^*$.

Lemma B.5. *Suppose that Assumptions 1–2 hold, then under H_0, $\lambda^* = \lambda_n + O_{P^*}\left(\sqrt{l}/n\right)$, in probability.*

Lemma B.6. *Suppose that under H_0, $\hat{\lambda}^* = \lambda^* + O_{P^*}(\delta_n^{\lambda})$ and $\hat{\beta}^* - \beta_0 = O_{P^*}(\delta_n^{\beta})$ in probability for $\delta_n^{\lambda} \to 0$ and for $\delta_n^{\beta} \to 0$, respectively. Then in Case* (I), (a) $\left(X^{*'}P_Z X^* - \hat{\lambda}^* X^{*'}X^*\right)/r_n = \tilde{H}_{I,n} + O_{P^*}\left(1/\sqrt{r_n} + \delta_n^{\lambda} n/r_n\right)$; (b) $\left(X^{*'}P_Z\hat{\epsilon}^*(\hat{\beta}^*) - \hat{\lambda}^* X^{*'}\hat{\epsilon}^*(\hat{\beta}^*)\right)/r_n = O_{P^*}\left(1/\sqrt{r_n} + \delta_n^{\beta} + \delta_n^{\lambda} n/r_n\right)$, *in probability, where* $\tilde{H}_{I,n} = (1-\lambda_n)\left(\tilde{\Pi}'(\beta_0)Z'Z\tilde{\Pi}(\beta_0)/r_n\right)$. *In Case* (II), (a′) $\left(X^{*'}P_Z X^* - \hat{\lambda}^* X^{*'}X^*\right)/l = \tilde{H}_{II,n} + O_{P^*}\left(1/\sqrt{l} + \delta_n^{\lambda} n/l\right)$; (b′) $(X^{*'}P_Z\hat{\epsilon}^*(\hat{\beta}^*) - \hat{\lambda}^* X^{*'}\hat{\epsilon}^*(\hat{\beta}^*))/l = O_{P^*}\left(1/\sqrt{l} + \delta_n^{\beta} + \delta_n^{\lambda} n/l\right)$, *in probability, where* $\tilde{H}_{II,n} = (1-\lambda_n)\left(\tilde{\Pi}'(\beta_0)Z'Z\tilde{\Pi}(\beta_0)/l\right)$.

Lemma B.7. *Suppose that Assumptions 1–2 hold and that under H_0, $\hat{\beta}^* - \beta_0 = O_{P^*}\left(\delta_n^{\beta}\right)$ in probability for $\delta_n^{\beta} \to 0$. Then in Case* (I), $\hat{\lambda}^* = \lambda^* + O_{P^*}\left(\frac{r_n}{n}(\delta_n^{\beta})^2\right)$; *in Case* (II), $\hat{\lambda}^* = \lambda^* + O_{P^*}\left(\frac{l}{n}(\delta_n^{\beta})^2\right)$, *in probability.*

Let $\hat{D}^*(\beta) = X^{*'}P_Z\varepsilon^*(\beta) - \frac{\epsilon^*(\beta)'P_Z\epsilon^*(\beta)}{\epsilon^*(\beta)'\epsilon^*(\beta)}X^{*'}\varepsilon^*(\beta)$.

Lemma B.8. *Suppose that Assumptions 1–2 hold and that under H_0, $\hat{\beta}^* - \beta_0 = O_{P^*}\left(\delta_n^{\beta}\right)$ in probability for $\delta_n^{\beta} \to 0$. Then in Case* (I), $-\left(\partial\hat{D}^*(\bar{\beta}^*)/\partial\beta\right)/r_n = \tilde{H}_{I,n} + O_{P^*}\left(1/\sqrt{r_n} + \delta_n^{\beta}\right)$; *in Case* (II), $-\left(\partial\hat{D}^*(\bar{\beta}^*)/\partial\beta\right)/l = \tilde{H}_{II,n} + O_{P^*}\left(1/\sqrt{l} + \delta_n^{\beta}\right)$, *in probability, where $\bar{\beta}^*$ lies between β_0 and $\hat{\beta}^*$.*

Lemma B.9. *Suppose that Assumptions 1–2 hold, then the following statements are true under H_0: in Case* (I), $\hat{D}^*(\beta_0)/\sqrt{r_n} = \left((1-\lambda_n)Z\tilde{\Pi}(\beta_0) + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right)'\epsilon^*/\sqrt{r_n} + O_{P^*}\left(1/\sqrt{r_n}\right)$; *in Case* (II), $\hat{D}^*(\beta_0)/\sqrt{l} = \left((1-\lambda_n)Z\tilde{\Pi}(\beta_0) + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right)'\epsilon^*/\sqrt{l} + O_{P^*}\left(1/\sqrt{l}\right)$, *in probability, where* $\tilde{V} = V^* - \epsilon^* q^{b'}$.

Proof of Theorem 3.2. The proof is similar to that of Theorem 3.1. Consider first the case in which $\hat{\beta}^*$ is the bootstrap analogue of LIML. We first show the results for Case (I). The proof is decomposed into three steps:

(1) We verify the conditions of a martingale central limit theorem;

(2) We derive the asymptotic variance–covariance matrix for $\sum_{i=1}^n W_{I,i}^*$;

(3) We obtain the limiting distribution of $\hat{\beta}^*$.

(1) In Case (I), by the first order condition of LIML and by Lemma B.8, we have

$$\sqrt{r_n}(\hat{\beta}^* - \beta_0) = -\left(\frac{1}{r_n}\frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right)^{-1}\frac{\hat{D}^*(\beta_0)}{\sqrt{r_n}}$$

under H_0, where $\bar{\beta}^*$ lies between $\hat{\beta}^*$ and β_0. Then, we let $W_{I,i}^* = \begin{pmatrix}(1-\lambda_n)\tilde{\Pi}'(\beta_0)Z_i\epsilon_i^*/\sqrt{r_n} \\ (P_{ii}-\lambda_n)\tilde{V}_i^*\varepsilon_i^*/\sqrt{l}\end{pmatrix}$, and we check the conditions of Lemma A2 in Hansen et al. (2008) hold conditionally on the original sample with probability converging to one. In particular, under the RE bootstrap scheme and under H_0, we can show that $E^*\left(\epsilon_i^*\right) = 0$, $E^*\left(\tilde{V}_i^*\right) = 0$; $E^*\left(\epsilon_i^{*4}\right)$ and $E^*\left(\tilde{V}_i^{*4}\right)$ are bounded in probability; $\sum_{i=1}^n E^*\left(W_{I,i}^* W_{I,i}^{*'}\right)$ is bounded in probability. Furthermore, by applying the same reasoning as that used in the proof of Theorem 3.1, we obtain that under H_0

$$\begin{cases}\sum_{i=1}^n E^*\left(\left\|\frac{1}{\sqrt{r_n}}(1-\lambda_n)\tilde{\Pi}'(\beta_0)Z_i\varepsilon_i^*\right\|^4\right) \to^P 0\\ \sum_{i=1}^n E^*\left(\left\|\frac{1}{\sqrt{l}}(P_{ii}-\lambda_n)\tilde{V}_i^*\varepsilon_i^*\right\|^4\right) \to^P 0.\end{cases}$$

(2) Then, we derive the asymptotic variance–covariance matrix of $\sum_{i=1}^n W_{I,i}^*$.

$$\sum_{i=1}^n E^*\left(W_{I,i}^* W_{I,i}^{*'}\right) \to^P \begin{pmatrix}(1-\lambda)\sigma_{\epsilon\epsilon}\tilde{H}_I & (1-\lambda)\tilde{A}' \\ (1-\lambda)\tilde{A} & (\phi-\lambda)\sigma_{\epsilon\epsilon}\Sigma_{\tilde{V}\tilde{V}} + \tilde{B}\end{pmatrix}$$

where $\tilde{H}_I = (1-\lambda)\left(\Psi + \gamma\Sigma_{\tilde{V}\tilde{V}}\right)$, $\tilde{A} = \sqrt{1-\lambda}A$, and $\tilde{B} = (1-2\lambda+\lambda\phi)B/(1-\lambda)$.

(2.1) These results follow by the fact that under the RE bootstrap d.g.p. and under H_0,

$$E^*\left(\epsilon_i^{*2}\right) = \frac{n}{n-k}\left\{\frac{\epsilon'(\beta_0)\epsilon(\beta_0)}{n}\right\} \to^P \sigma_{\epsilon\epsilon};$$

$$E^*\left(V_i^*\epsilon_i^*\right) = \sqrt{\frac{n}{n-k}}\sqrt{\frac{n}{n-l}}\left\{\frac{\tilde{V}'(\beta_0)\epsilon(\beta_0)}{n}\right\}$$
$$= \sqrt{\frac{n}{n-l}}\left\{\frac{V'M_Z\epsilon(\beta_0)}{n} + \frac{V'M_Z\epsilon(\beta_0)}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}\frac{\epsilon'(\beta_0)P_Z\epsilon(\beta_0)}{n}\right\} + o_P(1) \to^P \frac{\sigma_{V\epsilon}}{\sqrt{1-\lambda}};$$

$$E^*\left(V_i^*V_i^{*'}\right) = \frac{n}{n-l}\left\{\frac{\tilde{V}'(\beta_0)\tilde{V}(\beta_0)}{n}\right\}$$
$$= \frac{n}{n-l}\left\{\frac{V'M_Z V}{n} + \frac{V'M_Z\epsilon(\beta_0)}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}\frac{\epsilon'(\beta_0)P_Z\epsilon(\beta_0)}{n} \times \frac{\epsilon'(\beta_0)M_Z V}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}\right\}$$
$$\to^P \Sigma_{VV} + \frac{\lambda}{1-\lambda}\frac{\sigma_{V\epsilon}\sigma_{V\epsilon}'}{\sigma_{\epsilon\epsilon}},$$

which follows by using the results that under H_0, $V'M_Z\epsilon(\beta_0)/n \to^P (1-\lambda)\sigma_{V\epsilon}$, $\epsilon'(\beta_0)M_Z\epsilon(\beta_0)/n \to^P (1-\lambda)\sigma_{V\epsilon}$, $V'M_Z V/n \to^P (1-\lambda)\Sigma_{VV}$, and $\epsilon'(\beta_0)P_Z\epsilon(\beta_0)/n \to^P \lambda\sigma_{\epsilon\epsilon}$, as $n \to \infty$.

Putting the result of $E^*\left(\epsilon_i^{*2}\right)$, $E^*\left(V_i^*\epsilon_i^*\right)$ and $E^*\left(V_i^*V_i^{*'}\right)$ together, we conclude that under H_0,

$$E^*\left(\tilde{V}_i^*\tilde{V}_i^{*'}\right) \to^P \Sigma_{VV} + \left(\frac{\lambda}{1-\lambda}\frac{\sigma_{V\epsilon}\sigma_{V\epsilon}'}{\sigma_{\epsilon\epsilon}} - \frac{1}{1-\lambda}\frac{\sigma_{V\epsilon}\sigma_{V\epsilon}'}{\sigma_{\epsilon\epsilon}}\right) = \Sigma_{\tilde{V}\tilde{V}}$$

i.e., the RE bootstrap d.g.p. estimates $\Sigma_{\tilde{V}\tilde{V}}$ consistently.

Using similar arguments, we find that under H_0,

$$\sum_{i=1}^n \frac{(1-\lambda_n)^2}{r_n}\tilde{\Pi}'(\beta_0)Z_iZ_i'\tilde{\Pi}(\beta_0)E^*\left(\epsilon_i^{*2}\right)$$
$$= (1-\lambda_n)^2\left(\frac{\epsilon'(\beta_0)\epsilon(\beta_0)}{n}\right)\left(\frac{\tilde{\Pi}'(\beta_0)Z'Z\tilde{\Pi}(\beta_0)}{r_n}\right)$$
$$= (1-\lambda_n)^2\left(\frac{\epsilon'(\beta_0)\epsilon(\beta_0)}{n}\right)\left\{\frac{X'P_Z X}{r_n} - \frac{V'M_Z\epsilon(\beta_0)}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)} \times \frac{\epsilon'(\beta_0)P_Z X}{r_n} - \frac{X'P_Z\epsilon(\beta_0)}{r_n}\frac{\epsilon'(\beta_0)M_Z V}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)} + \frac{V'M_Z\epsilon(\beta_0)}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}\frac{\epsilon'(\beta_0)P_Z\epsilon(\beta_0)}{r_n}\frac{\epsilon'(\beta_0)M_Z V}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}\right\}$$

$$= (1-\lambda_n)^2\left(\frac{\epsilon'(\beta_0)\epsilon(\beta_0)}{n}\right)\left\{\frac{\Pi'Z'Z\Pi}{r_n}+\left(\frac{l}{r_n}\right)\Sigma_{VV} - \left(\frac{l}{r_n}\right)\frac{\sigma_{V\epsilon}\sigma_{V\epsilon}'}{\sigma_{\epsilon\epsilon}}+O_P\left(\frac{1}{\sqrt{r_n}}\right)\right\}$$
$$\to^P (1-\lambda)^2\sigma_{\epsilon\epsilon}\left(\Psi+\gamma\Sigma_{\tilde V\tilde V}\right)\equiv(1-\lambda)\sigma_{\epsilon\epsilon}\tilde H_I.$$

(2.2) For the off-diagonal term $\sum_{i=1}^n \frac{1-\lambda_n}{\sqrt{lr_n}}(P_{ii}-\lambda_n)\tilde\Pi'(\beta_0)Z_iE^*\left(\varepsilon_i^{*2}\tilde V_i^{*'}\right)$, note that under the RE bootstrap d.g.p.,

$$E^*\left(\epsilon_i^{*2}\tilde V_i^{*'}\right) = E^*\left(\epsilon_i^{*2}V_i^{*'}\right) - E^*\left(\epsilon_i^{*3}\right)q^{b'}$$
$$= \left(\frac{n}{n-k}\right)\left(\sqrt{\frac{n}{n-l}}\right)\left\{\frac{1}{n}\sum_{i=1}^n\epsilon_i^2(\beta_0)\tilde V_i'(\beta_0)\right\} - \left(\frac{n}{n-k}\right)\left(\sqrt{\frac{n}{n-l}}\right)$$
$$\times\left\{\frac{1}{n}\sum_{i=1}^n\epsilon_i^3(\beta_0)\left(\frac{1}{n}\sum_{i=1}^n\epsilon_i(\beta_0)\tilde V_i'(\beta_0)\right)\right\}\left\{\frac{1}{n}\sum_{i=1}^n\epsilon_i^2(\beta_0)\right\}^{-1}.$$

For the first term, let $\hat a(\beta_0) = (\epsilon_1^2(\beta_0)-\hat\sigma_{\epsilon\epsilon}(\beta_0),\ldots,\epsilon_n^2(\beta_0)-\hat\sigma_{\epsilon\epsilon}(\beta_0))'$ where $\hat\sigma_{\epsilon\epsilon}(\beta_0)=\epsilon'(\beta_0)\epsilon(\beta_0)/n$.

Note that under H_0,

$$\frac{\hat a'(\beta_0)\tilde V(\beta_0)}{n} = \frac{\hat a'(\beta_0)\hat V}{n}+\frac{\hat a'(\beta_0)P_Z\epsilon(\beta_0)}{n}\left(\frac{\epsilon'(\beta_0)M_ZV}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}\right) \to^P (1-\lambda)E\left(\epsilon_i^2V_i'\right)+\lambda E\left(\epsilon_i^3\right)q'$$

by using similar arguments as those for the standard bootstrap procedure. Moreover, $n^{-1}\sum_{i=1}^n\tilde V_i(\beta_0)=0$ by our setup that exogenous regressors, including a constant, have been filtered out. Therefore, $n^{-1}\sum_{i=1}^n\epsilon_i^2(\beta_0)\tilde V_i'(\beta_0) = n^{-1}\hat a'(\beta_0)\tilde V(\beta_0)$, and it follows that

$$\left(\frac{n}{n-k}\right)\left(\sqrt{\frac{n}{n-l}}\right)\left\{\frac{1}{n}\sum_{i=1}^n\epsilon_i^2(\beta_0)\tilde V_i'(\beta_0)\right\} \to^P \sqrt{1-\lambda}E\left(\epsilon_i^2V_i'\right)+\frac{\lambda}{\sqrt{1-\lambda}}E\left(\epsilon_i^3\right)q'.$$

For the second term,

$$\left(\frac{n}{n-k}\right)\left(\sqrt{\frac{n}{n-l}}\right)\left\{\frac{1}{n}\sum_{i=1}^n\epsilon_i^3(\beta_0)\left(\frac{1}{n}\sum_{i=1}^n\epsilon_i(\beta_0)\tilde V_i'(\beta_0)\right)\right\}\times\left\{\frac{1}{n}\sum_{i=1}^n\epsilon_i^2(\beta_0)\right\}^{-1}\to^P\frac{E\left(\epsilon_i^3\right)q'}{\sqrt{1-\lambda}}.$$

Putting these results together, we obtain that $E^*\left(\epsilon_i^{*2}\tilde V_i^{*'}\right)\to^P\sqrt{1-\lambda}E\left(\epsilon_i^2\tilde V_i'\right)$ under H_0 and the RE bootstrap d.g.p.

Now, for the term $\sum_{i=1}^n\frac{(1-\lambda_n)}{\sqrt{lr_n}}(P_{ii}-\lambda_n)\tilde\Pi'(\beta_0)Z_i$, we define $d_i=(P_{ii}-\lambda_n)/\sqrt{l}$ and $d=(d_1,\ldots,d_n)'$, as in the proofs for the standard bootstrap procedure. By $\|d\|^2\le 1$, we have $E\left(\left\|V'P_Zd\right\|^2\right)\le Cd'd\le C$ and $E\left(\left\|\epsilon'P_Zd\right\|^2\right)\le C$. Thus, $V'P_Zd=O_P(1)$ and $\epsilon'P_Zd=O_P(1)$ by standard arguments, and

$$\sum_{i=1}^n\frac{(1-\lambda_n)}{\sqrt{lr_n}}(P_{ii}-\lambda_n)\tilde\Pi'(\beta_0)Z_i = \frac{(1-\lambda_n)}{\sqrt{r_n}}\left(X-\epsilon(\beta_0)\frac{\epsilon'(\beta_0)M_ZV}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}\right)'P_Zd$$
$$= \frac{(1-\lambda_n)}{\sqrt{r_n}}\Pi'Z'd+O_P\left(\frac{1}{\sqrt{r_n}}\right) = \sum_{i=1}^n\frac{(1-\lambda_n)}{\sqrt{lr_n}}(P_{ii}-\lambda_n)\Pi'Z_i+O_P\left(\frac{1}{\sqrt{r_n}}\right)$$

under H_0. The desired result for the off-diagonal terms follows.

(2.3) Furthermore, for the term $\sum_{i=1}^n(P_{ii}-\lambda_n)^2E^*\left(\varepsilon_i^{*2}\tilde V_i^*\tilde V_i^{*'}\right)/l$, we have $E^*\left(\epsilon_i^{*2}\tilde V_i^*\tilde V_i^{*'}\right)=\sigma_{\epsilon\epsilon}^bE^*\left(\tilde V_i^*\tilde V_i^{*'}\right)+E^*((\epsilon_i^{*2}-\sigma_{\epsilon\epsilon}^b)\tilde V_i^*\tilde V_i^{*'})$, and we have shown that $\sigma_{\epsilon\epsilon}^b = E^*\left(\epsilon_i^{*2}\right)\to^P\sigma_{\epsilon\epsilon}$ and $E^*\left(\tilde V_i^*\tilde V_i^{*'}\right)\to^P\Sigma_{\tilde V\tilde V}$ under H_0.

Proceeding as in the proofs for the standard bootstrap procedure, we note that under the RE bootstrap d.g.p.,

$$E^*\left(\left(\epsilon_i^{*2}-\sigma_{\epsilon\epsilon}^b\right)\tilde V_i^*\tilde V_i^{*'}\right) = \left(\frac{n}{n-k}\right)\left(\frac{n}{n-l}\right)\left\{\frac{1}{n}\sum_{i=1}^n\left(\epsilon_i^2(\beta_0)-\hat\sigma_{\epsilon\epsilon}(\beta_0)\right)\times\left(\tilde V_i(\beta_0)-\epsilon_i(\beta_0)\hat q'(\beta_0)\right)\left(\tilde V_i(\beta_0)-\epsilon_i(\beta_0)\hat q'(\beta_0)\right)'\right\}$$
$$= \left(\frac{n}{n-k}\right)\left(\frac{n}{n-l}\right)\frac{1}{n}\left\{M_Z\tilde V+M_Z\epsilon q'-M_Z\epsilon(\beta_0)\hat q'(\beta_0)\right\}'\times\hat J(\beta_0)\left\{M_Z\tilde V+M_Z\epsilon q'-M_Z\epsilon(\beta_0)\hat q'(\beta_0)\right\}$$
$$\to^P\frac{1-2\lambda+\lambda\phi}{1-\lambda}E\left((\epsilon_i^2-\sigma_{\epsilon\epsilon})\tilde V_i\tilde V_i'\right)$$

where $\hat J(\beta_0) = diag\left(\hat a_1(\beta_0),\ldots,\hat a_n(\beta_0)\right)$ and $\hat q(\beta_0)=\tilde V'(\beta_0)\epsilon(\beta_0)/n$. Then, the desired result for $\sum_{i=1}^n(P_{ii}-\lambda_n)^2E^*\left(\varepsilon_i^{*2}\tilde V_i^*\tilde V_i^{*'}\right)/l$ follows.

(3) Now by Lemma B.9 and by following closely the line of arguments for the standard bootstrap, we obtain that in Case (I) and under H_0

$$\frac{\hat D^*(\beta_0)}{\sqrt{r_n}} = \frac{1}{\sqrt{r_n}}\left\{(1-\lambda_n)Z\tilde\Pi(\beta_0)+P_Z\tilde V^*-\lambda_n\tilde V^*\right\}'\epsilon^* + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$$
$$= \frac{1}{\sqrt{r_n}}\left(\sum_{i=1}^n\tilde\Pi'(\beta_0)Z_i\epsilon_i^*+\sum_{i=1}^n(P_{ii}-\lambda_n)\tilde V_i^*\epsilon_i^* + \sum_{i\neq j}\tilde V_i^*P_{ij}\epsilon_j^*\right)+O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$$
$$\to^{d^*}N\left(0,\tilde\Upsilon_I\right)$$

in probability, where $\tilde\Upsilon_I = (1-\lambda)\sigma_{\epsilon\epsilon}\left\{\tilde H_I+\gamma\Sigma_{\tilde V\tilde V}\right\}+(1-\lambda)\sqrt{\gamma}\left\{\tilde A+\tilde A'\right\}+\gamma\tilde B$. Then, by Lemma B.8 and by continuous mapping theorem for weak convergence in probability,

$$\sqrt{r_n}\left(\hat\beta^*-\beta_0\right) = \left(\frac{1}{r_n}\frac{\partial\hat D^*(\bar\beta^*)}{\partial\beta}\right)^{-1}\frac{\hat D^*(\beta_0)}{\sqrt{r_n}}\to^{d^*}N\left(0,\tilde\Lambda_I\right)$$

in probability, where $\tilde\Lambda_I=\tilde H_I^{-1}\tilde\Upsilon_I\tilde H_I^{-1}$.

For Case (II), we let $W^*_{II,i} = \begin{pmatrix} (1-\lambda_n)\tilde{\Pi}'(\beta_0)Z_i\varepsilon^*_i/\sqrt{l} \\ (P_{ii}-\lambda_n)\tilde{V}^*_i\varepsilon^*_i/\sqrt{l} \end{pmatrix}$, and we obtain that under H_0,

$$\sum_{i=1}^{n} E^*\left(W^*_{II,i}W^{*'}_{II,i}\right) \to_P \begin{pmatrix} (1-\lambda)\sigma_{\epsilon\epsilon}\tilde{H}_{II} & 0 \\ 0 & (\phi-\lambda)\sigma_{\epsilon\epsilon}\Sigma_{\tilde{V}\tilde{V}} + \tilde{B} \end{pmatrix}$$

where $\tilde{H}_{II} = (1-\lambda)\Sigma_{\tilde{V}\tilde{V}}$. Then, by using Lemmas B.8 and B.9 and proceeding as in Case (I), we obtain that

$$\frac{\hat{D}^*(\beta_0)}{\sqrt{l}} = \frac{1}{\sqrt{l}}\left\{(1-\lambda_n)Z\tilde{\Pi}(\beta_0) + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right\}'\epsilon^* + o_{P^*}\left(\frac{1}{\sqrt{l}}\right) \to^{d^*} N\left(0, \tilde{\Upsilon}_{II}\right)$$

in probability, and $\sqrt{l}\left(\hat{\beta}^* - \beta_0\right) = \left(\frac{1}{l}\frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right)^{-1}\frac{\hat{D}^*(\beta_0)}{\sqrt{l}} \to^{d^*} N\left(0, \tilde{\Lambda}_{II}\right)$ in probability, where $\tilde{\Upsilon}_{II} = (1-\lambda)\sigma_{\epsilon\epsilon}\left\{\tilde{H}_{II} + \Sigma_{\tilde{V}\tilde{V}}\right\} + \tilde{B}$ and $\tilde{\Lambda}_{II} = \tilde{H}_{II}^{-1}\tilde{\Upsilon}_{II}\tilde{H}_{II}^{-1}$.

The results for the bootstrap analogue of FULL follow immediately by using the fact that $\tilde{\lambda}^* = \hat{\lambda}^* + O_{P^*}(1/n)$ in probability. ■

Proof of Corollary 3.2. By Theorem 3.2, we have when $l/r_n \to 0$, $\sqrt{r_n}(\hat{\beta}^*_{re} - \beta_0) \to^{d^*} N(0, \sigma_{\epsilon\epsilon}Q^{-1})$, which is the same as the limiting distribution of $\sqrt{r_n}(\hat{\beta} - \beta_0)$. The result then follows by Polya's Theorem. ■

The proofs for Lemmas B.1–B.9 are similar to those in Appendix A thus omitted.

Appendix C. Proofs of results for the modified RE bootstrap

All the proofs of the Lemmas are relegated at the end of Appendix C. Let $\hat{\beta}^* = \hat{\beta}^*_m$ throughout Appendix C. Let C denote a generic positive constant that may be different in different uses. Also, P^* denotes the probability measure induced by the MRE bootstrap procedures and E^* denotes the expectation under P^*. Below, we show the proofs for the MRE1 procedure; the results for the MRE2 procedure follow immediately by noting that the LIML or FULL estimate $\hat{\beta}$ is consistent under our assumptions.

Lemma C.1. *Suppose that Assumptions 1–2 hold, then under H_0 : $\beta = \beta_0$, $E^*\left(\epsilon_i^{*8}\right)$ and $E^*\left(\left\|V^*_i\right\|^8\right)$ are bounded in probability.*

Lemma C.2. *Suppose that Assumptions 1–2 hold, then the following statements are true under H_0: (a) $V^{*'}P_Z\epsilon^*/l = \sigma^b_{V\epsilon} + O_{P^*}\left(1/\sqrt{l}\right)$; (b) $V^{*'}P_ZV^*/l = \Sigma^b_{VV} + O_{P^*}\left(1/\sqrt{l}\right)$; (c) $\epsilon^{*'}P_Z\epsilon^*/l = \sigma^b_{\epsilon\epsilon} + O_{P^*}\left(1/\sqrt{l}\right)$; (d) $\tilde{\Pi}'_m(\beta_0)Z'V^*/r_n = O_{P^*}\left(1/\sqrt{r_n}\right)$; (e) $\tilde{\Pi}'_m(\beta_0)Z'\epsilon^*/r_n = O_{P^*}\left(1/\sqrt{r_n}\right)$; in probability, where $\sigma^b_{V\epsilon} \equiv E^*\left(V^*_i\epsilon^*_i\right)$, $\Sigma^b_{VV} \equiv E^*\left(V^*_iV^{*'}_i\right)$ and $\sigma^b_{\epsilon\epsilon} \equiv E^*\left(\epsilon_i^{*2}\right)$.*

Let $J^* = diag(a^*_1, \ldots, a^*_n)$ where $a^*_i = \epsilon_i^{*2} - \sigma^b_{\epsilon\epsilon}$, $i = 1, \ldots, n$.

Lemma C.3. *Suppose that Assumptions 1–2 hold, then the following statements are true under H_0:*

(a) $\tilde{V}^{*'}J^*\tilde{V}^*/n = E^*\left(a^*_i\tilde{V}^*_i\tilde{V}^{*'}_i\right) + O_{P^*}\left(1/\sqrt{n}\right)$; (b) $\tilde{V}^{*'}P_ZJ^*\tilde{V}^*/n = \lambda_nE^*\left(a^*_i\tilde{V}^*_i\tilde{V}^{*'}_i\right) + O_{P^*}\left(1/\sqrt{n}\right)$; (c) $\tilde{V}^{*'}P_ZJ^*P_Z\tilde{V}^*/n = \lambda_n\phi_nE^*\left(a^*_i\tilde{V}^*_i\tilde{V}^{*'}_i\right) + O_{P^*}\left(1/\sqrt{n}\right)$, in probability.

Lemma C.4. *Suppose that Assumptions 1–2 hold, then under H_0, $\hat{\beta}^* - \beta_0 = o_{P^*}(1)$, in probability.*

Let $\lambda^* = \epsilon^{*'}P_Z\epsilon^*/\epsilon^{*'}\epsilon^*$.

Lemma C.5. *Suppose that Assumptions 1–2 hold, then under H_0, $\lambda^* = \lambda_n + O_{P^*}\left(\sqrt{l}/n\right)$, in probability.*

Lemma C.6. *Suppose that under H_0, $\hat{\lambda}^* = \lambda^* + O_{P^*}\left(\delta^\lambda_n\right)$ and $\hat{\beta}^* - \beta_0 = O_{P^*}\left(\delta^\beta_n\right)$ in probability for $\delta^\lambda_n \to 0$ and $\delta^\beta_n \to 0$, respectively. Then, (a) $\left(X^{*'}P_ZX^* - \hat{\lambda}^*X^{*'}X^*\right)/r_n = \tilde{H}_{m,n} + O_{P^*}\left(\sqrt{l}/r_n + \delta^\lambda_n n/r_n\right)$; (b) $\left(X^{*'}P_Z\hat{\epsilon}^* - \hat{\lambda}^*X^{*'}\hat{\epsilon}^*(\hat{\beta}^*)\right)/r_n = O_{P^*}\left(\sqrt{l}/r_n + \delta^\beta_n + \delta^\lambda_n n/r_n\right)$, in probability, where $\tilde{H}_{m,n} = (1 - \lambda_n)\left(\tilde{\Pi}'_m(\beta_0)Z'Z\tilde{\Pi}_m(\beta_0)/r_n\right)$.*

Lemma C.7. *Suppose that Assumptions 1–2 hold and that under H_0, $\hat{\beta}^* - \beta_0 = O_{P^*}\left(\delta^\beta_n\right)$ in probability for $\delta^\beta_n \to 0$. Then, $\hat{\lambda}^* = \lambda^* + O_{P^*}\left(r_n(\delta^\beta_n)^2/n\right)$, in probability.*

Let $\hat{D}^*(\beta) = X^{*'}P_Z\varepsilon^*(\beta) - \frac{\epsilon^*(\beta)'P_Z\epsilon^*(\beta)}{\epsilon^*(\beta)'\epsilon^*(\beta)}X^{*'}\varepsilon^*(\beta)$.

Lemma C.8. *Suppose that Assumptions 1–2 hold and that under H_0, $\hat{\beta}^* - \beta_0 = O_{P^*}\left(\delta^\beta_n\right)$ in probability. Then, $-\left(\partial\hat{D}^*(\bar{\beta}^*)/\partial\beta\right)/r_n = \tilde{H}_{m,n} + O_{P^*}\left(\sqrt{l}/r_n + \delta^\beta_n\right)$ in probability, where $\bar{\beta}^*$ lies between β_0 and $\hat{\beta}^*$.*

Lemma C.9. *Suppose that Assumptions 1–2 hold, then the following statements are true under H_0: $\hat{D}^*(\beta_0)/\sqrt{r_n} = \{(1-\lambda_n)Z\tilde{\Pi}_m(\beta_0) + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\}'\epsilon^*/\sqrt{r_n} + O_{P^*}\left(1/\sqrt{r_n}\right)$ in probability, where $\tilde{V}^* = V^* - \epsilon^*q^{b'}$.*

Proof of Theorem 3.3. We begin with Case (I), and we show the results for the bootstrap analogue of LIML. The results for the bootstrap analogue of FULL follow immediately by using the fact that $\tilde{\lambda}^* = \hat{\lambda}^* + O_{P^*}(1/n)$ in probability.

Similar to the proofs for standard and RE bootstraps, the proofs for the MRE1 bootstrap are decomposed in three steps:

(1) We verify the conditions of a martingale central limit theorem;

(2) We derive the asymptotic variance–covariance matrix for $\sum_{i=1}^n W^*_{i,m}$;

(3) We obtain the limiting distribution of $\hat{\beta}^*$.

(1) By the first order condition of LIML, by Lemmas C.4 and C.8, and by $\tilde{H}_{m,n} = O_P(1)$, we have

$$\sqrt{r_n}(\hat{\beta}^* - \beta_0) = -\left(\frac{1}{r_n}\frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right)^{-1}\frac{1}{\sqrt{r_n}}\hat{D}^*(\beta_0).$$

Similar to the proofs for the standard and RE bootstraps, we define $W^*_{i,m} = \begin{pmatrix} (1-\lambda_n)\tilde{\Pi}'_m(\beta_0)Z_i\varepsilon^*_i/\sqrt{r_n} \\ (P_{ii}-\lambda_n)\tilde{V}^*_i\varepsilon^*_i/\sqrt{l} \end{pmatrix}$, and we find that the conditions of Lemma A2 in Hansen et al. (2008) hold with $\left(W^*_{1,m}, \tilde{V}^*_1, \epsilon^*_1\right), \ldots, \left(W^*_{n,m}, \tilde{V}^*_n, \epsilon^*_n\right)$, conditionally on the original sample with probability converging to one.

(2) Furthermore, for $\sum_{i=1}^n E^*\left(W^*_{i,m}W^{*'}_{i,m}\right)$, we obtain by the MRE1 bootstrap scheme that

$$\sum_{i=1}^{n}\frac{(1-\lambda_n)^2}{r_n}\tilde{\Pi}'_m(\beta_0)Z_iZ'_i\tilde{\Pi}_m(\beta_0)E^*\left(\varepsilon_i^{*2}\right) \to^P (1-\lambda)\sigma_{\epsilon\epsilon}H$$

under H_0, which follows from the fact that

$$\frac{\tilde{\Pi}_m'(\beta_0)Z'Z\tilde{\Pi}_m(\beta_0)}{r_n} = \frac{\tilde{\Psi}(\beta_0) - l\widehat{\Sigma}_{\tilde{V}\tilde{V}}(\beta_0)}{r_n}$$
$$= \frac{\Pi'Z'Z\Pi}{r_n} + O_P\left(\frac{1}{\sqrt{r_n}}\right) + \frac{l}{r_n}\left\{\Sigma_{\tilde{V}\tilde{V}} + O_P\left(\frac{1}{\sqrt{l}}\right)\right\}$$
$$- \frac{l}{r_n}\left\{\Sigma_{\tilde{V}\tilde{V}} + O_P\left(\frac{1}{\sqrt{l}}\right)\right\}$$
$$= \frac{\Pi'Z'Z\Pi}{r_n} + O_P\left(\frac{1}{\sqrt{r_n}}\right) + O_P\left(\frac{\sqrt{l}}{r_n}\right) \to^P \Psi.$$

The second equality follows by $\Psi(\beta_0)/r_n = \Pi'Z'Z\Pi/r_n + O_P\left(1/\sqrt{r_n}\right) + (l/r_n)\left\{\Sigma_{\tilde{V}\tilde{V}} + O_P\left(1/\sqrt{l}\right)\right\}$ and by

$$\widehat{\Sigma}_{\tilde{V}\tilde{V}}(\beta_0) = \frac{V'M_ZV}{n-l} - \frac{V'M_Z\epsilon(\beta_0)}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}\left(\frac{\epsilon'(\beta_0)M_ZV}{n-l}\right)$$
$$- \left(\frac{V'M_Z\epsilon(\beta_0)}{n-l}\right)\frac{\epsilon'(\beta_0)M_ZV}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}$$
$$+ \frac{V'M_Z\epsilon(\beta_0)}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}\left(\frac{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}{n-l}\right)\frac{\epsilon'(\beta_0)M_ZV}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}$$
$$= \Sigma_{\tilde{V}\tilde{V}} + O_P\left(\frac{1}{\sqrt{l}}\right)$$

under H_0.

On the other hand, $E^*\left(\varepsilon_i^{*2}\tilde{V}_i^*\right)\sum_{i=1}^n \frac{1-\lambda_n}{\sqrt{lr_n}}(P_{ii} - \lambda_n)Z_i'\tilde{\Pi}_m(\beta_0)$ and $\sum_{i=1}^n \frac{1}{l}(P_{ii}-\lambda_n)^2 E^*\left(\varepsilon_i^{*2}\tilde{V}_i^*\tilde{V}_i^{*'}\right)$ converge to zero in probability under Assumption 3(a) or Assumption 3(b).

Also notice that by the MRE bootstrap d.g.p., we have

$$E^*\left(\epsilon_i^{*2}\right) = \frac{n}{n-l}\left(\frac{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}{n}\right) \to^P \sigma_{\epsilon\epsilon};$$
$$E^*\left(V_i^*\epsilon_i^*\right) = \frac{n}{n-l}\left(\frac{\hat{V}'M_Z\epsilon(\beta_0)}{n}\right) \to^P \sigma_{V\epsilon};$$
$$E^*\left(V_i^*V_i^{*'}\right) = \frac{n}{n-l}\left(\frac{\hat{V}'\hat{V}}{n}\right) + o_P(1) \to^P \Sigma_{VV};$$
$$E^*\left(\tilde{V}_i^*\tilde{V}_i^{*'}\right) = \frac{n}{n-l}\left\{\left(\frac{\hat{V}'\hat{V}}{n}\right) - \left(\frac{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}{n}\right)^{-1}\right.$$
$$\left.\left(\frac{\hat{V}'M_Z\epsilon(\beta_0)}{n}\right)\left(\frac{\hat{V}'M_Z\epsilon(\beta_0)}{n}\right)'\right\} \to^P \Sigma_{\tilde{V}\tilde{V}},$$

under H_0.

(3) Now defining $U_{m,n}^* = \begin{pmatrix} \sum_{i=1}^n W_{m,i}^* \\ \sum_{i\neq j}^n \tilde{V}_i^* P_{ij}\epsilon_j^*/\sqrt{l} \end{pmatrix}$ and proceeding as in the proof for Theorems 3.1 and 3.2, we obtain under H_0 and in case (I),

$$\frac{\hat{D}^*(\beta_0)}{\sqrt{r_n}} = \frac{1}{\sqrt{r_n}}\left\{(1-\lambda_n)Z\tilde{\Pi}_m(\beta_0) + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right\}'\epsilon^*$$
$$+ O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$$
$$= \frac{1}{\sqrt{r_n}}\left(\sum_{i=1}^n (1-\lambda_n)\tilde{\Pi}_m'(\beta_0)Z_i\epsilon_i^* + \sum_{i=1}^n (P_{ii}-\lambda_n)\tilde{V}_i^*\epsilon_i^*\right.$$
$$\left.+ \sum_{i\neq j}^n \tilde{V}_i^* P_{ij}\epsilon_j^*\right) + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$$
$$= F_{I,n}U_{m,n}^* + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) \to^{d^*} N\left(0, \tilde{\Upsilon}_{m,I}\right)$$

in probability, where $F_{I,n} = \left[I_k, \sqrt{l/r_n}I_k, \sqrt{l/r_n}I_k\right]$ and $\tilde{\Upsilon}_{m,I} = (1-\lambda)\sigma_{\epsilon\epsilon}\left\{H + \gamma\Sigma_{\tilde{V}\tilde{V}}\right\}$; it follows by using Lemma C.8 and by using continuous mapping theorem for weak convergence in probability that

$$\sqrt{r_n}\left(\hat{\beta}^* - \beta_0\right) = \left(\frac{1}{r_n}\frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right)^{-1}\frac{\hat{D}^*(\beta_0)}{\sqrt{r_n}} \to^{d^*} N\left(0, \tilde{\Lambda}_{m,I}\right)$$

in probability, where $\tilde{\Lambda}_{m,I} = H^{-1}\tilde{\Upsilon}_{m,I}H^{-1}$.

By applying the same reasoning for Case (II), we obtain that under H_0,

$$\left(\sqrt{\frac{r_n}{l}}\right)\frac{\hat{D}^*(\beta_0)}{\sqrt{r_n}} = \left(\sqrt{\frac{r_n}{l}}\right)\left\{\frac{1}{\sqrt{r_n}}\left[(1-\lambda_n)Z\tilde{\Pi}_m(\beta_0)\right.\right.$$
$$\left.\left.+ P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right]'\epsilon^* + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)\right\}$$
$$= \left(\sqrt{\frac{r_n}{l}}\right)\left(F_{I,n}U_{m,n}^* + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)\right)$$
$$\to^{d^*} N\left(0, \tilde{\Upsilon}_{m,II}\right)$$

in probability, where $\tilde{\Upsilon}_{m,II} = (1-\lambda)\sigma_{\epsilon\epsilon}\Sigma_{\tilde{V}\tilde{V}}$;

$$\frac{r_n}{\sqrt{l}}\left(\hat{\beta}^* - \beta_0\right) = \left(\frac{1}{r_n}\frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right)^{-1}\frac{\hat{D}^*(\beta_0)}{\sqrt{l}} \to^{d^*} N\left(0, \tilde{\Lambda}_{m,II}\right)$$

in probability, where $\tilde{\Lambda}_{m,II} = H^{-1}\tilde{\Upsilon}_{m,II}H^{-1}$.

In perfect analogy, we have $\sqrt{r_n}(\hat{\beta} - \beta) \to^d N\left(0, \tilde{\Lambda}_{m,I}\right)$ in Case (I) and $\frac{r_n}{\sqrt{l}}(\hat{\beta} - \beta) \to^d N\left(0, \tilde{\Lambda}_{m,II}\right)$ in Case (II). Therefore, the asymptotic validity of the MRE1 bootstrap follows by applying Polya's Theorem to both cases. ■

Now we give the proofs for the Lemmas.

Proof of Lemma C.1. It is similar to the proof of Lemma A.1.

Proof of Lemma C.2. The proof for (a)–(c) is similar to those in the proof of Lemma A.2. The proof for (d) and (e) follows from noting that for both Cases (I) and (II), $E^*\left\|\frac{\tilde{\Pi}_m'(\beta_0)Z'V^*}{r_n}\right\|^2 = O_P\left(\frac{1}{r_n}\right)$ because $\tilde{\Pi}_m'(\beta_0)Z'Z\tilde{\Pi}_m(\beta_0) = O_P(r_n)$ for both cases and

$$E^*\left(\left\|\frac{\tilde{\Pi}_m'(\beta_0)Z'V^*}{r_n}\right\|^2\right)$$
$$= \frac{trace\left(\Sigma_{VV}^b\right)}{r_n}\left(trace\left(\frac{\tilde{\Pi}_m'(\beta_0)Z'Z\tilde{\Pi}_m(\beta_0)}{r_n}\right)\right).$$

Similarly, we have $E^*\left\|\frac{\tilde{\Pi}_m'(\beta_0)Z'\epsilon^*}{r_n}\right\|^2 = O_P\left(\frac{1}{r_n}\right)$. ■

Proof of Lemma C.3. It is similar to the proof of Lemma A.3.

Proof of Lemma C.4. It can be verified that under the MRE bootstrap scheme, $\hat{\lambda}^* = \lambda_n + o_{P^*}(r_n/n)$ in probability, for both Cases (I) and (II). Then, write $\hat{\beta}^* - \beta_0 = \left(\frac{X^{*'}P_ZX^*}{r_n} - \hat{\lambda}^*\frac{X^{*'}X^*}{r_n}\right)^{-1}\left(\frac{X^{*'}P_Z\epsilon^*}{r_n}\right.$ $\left.- \hat{\lambda}^*\frac{X^{*'}\epsilon^*}{r_n}\right)$, and the desired result follows by using similar arguments as in the proof of Lemma A.4.

Proof of Lemma C.5. It is similar to the proof of Lemma A.5.

Proof of Lemma C.6. Notice that by Lemma C.2, for both Cases (I) and (II) $\left(X^{*'}P_ZX^* - \lambda_nX^{*'}X^*\right)/r_n = \tilde{H}_{m,n} + O_{P^*}\left(\sqrt{l}/r_n\right)$ in probability. Then, the results follow by applying same arguments as in the proof of Lemma A.5. ■

Proof of Lemma C.7. It is similar to the proof of Lemma A.7.

Proof of Lemma C.8. Let $\bar{\epsilon}^* = y^* - X^*\bar{\beta}^*$ and $\bar{\gamma}^* = X^*\bar{\epsilon}^*/\bar{\epsilon}^{*'}\bar{\epsilon}^*$, where $\bar{\beta}^*$ lies between $\hat{\beta}^*$ and β_0. Differentiating gives

$$-\left(\partial\hat{D}^*(\bar{\beta}^*)/\partial\beta\right) = X^{*'}P_Z X^* - \bar{\lambda}^* X^{*'}X^* + \bar{\gamma}^*\hat{D}^*(\bar{\beta}^*)' + \hat{D}^*(\bar{\beta}^*)\bar{\gamma}^{*'},$$

where $\bar{\lambda}^* = \bar{\epsilon}^{*'}P_Z\bar{\epsilon}^*/\bar{\epsilon}^{*'}\bar{\epsilon}^*$. Notice that for both Cases (I) and (II), by Lemmas C.6 and C.7, we have $\left(X^{*'}P_Z X^* - \bar{\lambda}^* X^{*'}X^*\right)/r_n = \tilde{H}_{m,n} + O_{P^*}\left(\sqrt{l}/r_n + (\delta_n^\beta)^2\right)$ and $\hat{D}^*\left(\bar{\beta}^*\right)/r_n = O_{P^*}\left(\sqrt{l}/r_n + \delta_n^\beta\right)$, in probability. Also, by standard argument we have $\bar{\gamma}^* = O_{P^*}(1)$ in probability. The conclusion then follows by the triangle inequality. ■

Proof of Lemma C.9. It is similar to the proof of Lemma A.9.

Appendix D. Proofs of results for Theorem 3.4

We give the proof for the case of standard bootstrap. The proofs for the RE/MRE bootstraps are similar thus omitted. Let $\hat{\beta}^* = \hat{\beta}^*_{std}$ and $\hat{\Lambda}^* = \hat{\Lambda}^*_{std}$. Also, denote $\tilde{X}^*(\hat{\beta}^*) = X^* - \hat{\epsilon}^*(\hat{\beta}^*)\frac{\hat{\epsilon}^{*'}(\hat{\beta}^*)X^*}{\hat{\epsilon}^{*'}(\hat{\beta}^*)\hat{\epsilon}^*(\hat{\beta}^*)}$ and $\tilde{X}^* = X^* - \epsilon^* q^{b'}$ where $q^b = \sigma^b_{V\epsilon}/\sigma^b_{\epsilon\epsilon}$.

The proof is decomposed into three steps:

(1) We derive asymptotic results for the bootstrap analogue of Bekker (1994)'s variance estimator;

(2) We derive asymptotic results for the bootstrap analogues of the nonnormality adjustment terms;

(3) We derive asymptotic results for the bootstrap analogue of the CSE-based t-ratio.

(1) To obtain the asymptotic behavior of $\hat{\Lambda}^*$, the bootstrap analogue of the CSE, we start with the term $\hat{\Upsilon}^*_{bkk}$. For Case (I), notice that $\left\|\hat{\beta}^* - \hat{\beta}\right\| = O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$, then

$$\begin{aligned} n^{-1}\left\|\hat{\epsilon}^*(\hat{\beta}^*) - \epsilon^*\right\|^2 &\le n^{-1}\left\|X^*\right\|^2\left\|\hat{\beta}^* - \hat{\beta}\right\|^2 \\ &\le \left(n^{-1}\left\|X^*\right\|^2\right) O_{P^*}\left(\frac{1}{r_n}\right) \\ &= O_{P^*}\left(\frac{1}{r_n}\right) \end{aligned}$$

where the last equality follows from $X^{*'}X^* = O_{P^*}(n)$. It follows by standard arguments that

$$\left\|\frac{X^{*'}\hat{\epsilon}^*(\hat{\beta}^*)}{\hat{\epsilon}^{*'}(\hat{\beta}^*)\hat{\epsilon}^*(\hat{\beta}^*)} - q^b\right\|^2 = O_{P^*}\left(\frac{1}{r_n}\right)$$

and

$$\begin{aligned} \left\|\tilde{X}^*(\hat{\beta}^*) - \tilde{X}^*\right\| &= \left\|\hat{\epsilon}^*(\hat{\beta}^*)\frac{\hat{\epsilon}^{*'}(\hat{\beta}^*)X^*}{\hat{\epsilon}^{*'}(\hat{\beta}^*)\hat{\epsilon}^*(\hat{\beta}^*)} - \epsilon^* q^{b'}\right\| \\ &= O_{P^*}\left(\sqrt{\frac{n}{r_n}}\right); \end{aligned}$$

then, we obtain by $\left\|\tilde{X}^*\right\| = O_{P^*}\left(\sqrt{n}\right)$ that

$$\left\|\tilde{X}^{*'}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*) - \tilde{X}^{*'}\tilde{X}^*\right\| = O_{P^*}\left(\frac{n}{\sqrt{r_n}}\right).$$

By Lemmas A.5 and A.7, $\hat{\lambda}^* = \lambda_n + O_{P^*}\left(\sqrt{l}/n\right)$, we obtain

$$\begin{aligned} &\left\|\hat{\lambda}^*\left(\frac{\tilde{X}^{*'}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{r_n} - \frac{\tilde{X}^{*'}\tilde{X}^*}{r_n}\right)\right\| \\ &= O_{P^*}\left(\frac{l}{n}\right) O_{P^*}\left(\frac{1}{r_n}\right) O_{P^*}\left(\frac{n}{\sqrt{r_n}}\right) = O_{P^*}\left(\frac{l}{r_n\sqrt{r_n}}\right). \end{aligned}$$

Furthermore, it follows by $\tilde{X}^{*'}\tilde{X}^* = O_{P^*}(n)$ that

$$\begin{aligned} \left\|\left(\hat{\lambda}^* - \lambda_n\right)\frac{\tilde{X}^{*'}\tilde{X}^*}{r_n}\right\| &= O_{P^*}\left(\frac{\sqrt{l}}{n}\right) O_{P^*}\left(\frac{n}{r_n}\right) = O_{P^*}\left(\frac{\sqrt{l}}{r_n}\right) \\ &= O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) \end{aligned}$$

in Case (I). Putting these results together, we obtain

$$\begin{aligned} \hat{\lambda}^*\left(\frac{\tilde{X}^{*'}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{r_n}\right) &= \lambda_n\left(\frac{\tilde{X}^{*'}\tilde{X}^*}{r_n}\right) + \left(\hat{\lambda}^* - \lambda_n\right)\frac{\tilde{X}^{*'}\tilde{X}^*}{r_n} \\ &\quad + \hat{\lambda}^*\left(\frac{\tilde{X}^{*'}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{r_n} - \frac{\tilde{X}^{*'}\tilde{X}^*}{r_n}\right) \\ &= \lambda_n\left(\frac{\tilde{X}^{*'}\tilde{X}^*}{r_n}\right) + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) + O_{P^*}\left(\frac{l}{r_n\sqrt{r_n}}\right) \\ &= \lambda_n\left(\frac{\tilde{X}^{*'}\tilde{X}^*}{r_n}\right) + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) \end{aligned}$$

given that $l/r_n \to \gamma < \infty$ in this case. Similarly, we can find that for Case (II)

$$\begin{aligned} \hat{\lambda}^*\left(\frac{\tilde{X}^{*'}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{l}\right) &= \lambda_n\left(\frac{\tilde{X}^{*'}\tilde{X}^*}{l}\right) + \left(\hat{\lambda}^* - \lambda_n\right)\frac{\tilde{X}^{*'}\tilde{X}^*}{l} \\ &\quad + \hat{\lambda}^*\left(\frac{\tilde{X}^{*'}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{l} - \frac{\tilde{X}^{*'}\tilde{X}^*}{l}\right) \\ &= \lambda_n\left(\frac{\tilde{X}^{*'}\tilde{X}^*}{l}\right) + O_{P^*}\left(\frac{1}{\sqrt{l}}\right) \end{aligned}$$

given that $\left\|(\hat{\lambda}^* - \lambda_n)\frac{\tilde{X}^{*'}\tilde{X}^*}{l}\right\| = O_{P^*}\left(\frac{\sqrt{l}}{n}\cdot\frac{n}{l}\right) = O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$ and $\left\|\hat{\lambda}^*\left(\frac{\tilde{X}^{*'}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{l} - \frac{\tilde{X}^{*'}\tilde{X}^*}{l}\right)\right\| = O_{P^*}\left(\frac{l}{n}\cdot\frac{1}{l}\cdot\frac{n}{\sqrt{l}}\right) = O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$.

Moreover, it follows from arguments similar to Lemma A.2 that for Case (I)

$$\begin{aligned} \lambda_n\left(\frac{\tilde{V}^{*'}\tilde{V}^*}{r_n}\right) &= \left(\frac{l}{r_n}\right)\left(\Sigma^b_{\tilde{V}\tilde{V}} + O_{P^*}\left(\frac{1}{\sqrt{n}}\right)\right) \\ &= \left(\frac{l}{r_n}\right)\Sigma^b_{\tilde{V}\tilde{V}} + O_{P^*}\left(\frac{l}{r_n\sqrt{n}}\right) \\ &= \left(\frac{l}{r_n}\right)\Sigma^b_{\tilde{V}\tilde{V}} + O_{P^*}\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

where $\Sigma^b_{\tilde{V}\tilde{V}} = E^*\left(\tilde{V}_i^*\tilde{V}_i^{*'}\right)$, and $\lambda_n\left(\frac{\hat{\Pi}'Z'\tilde{V}^*}{r_n}\right) = O_{P^*}\left(\frac{l}{n}\right)O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) = O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$. Then, together with previous arguments, we obtain by $\tilde{X}^* = Z\hat{\Pi} + \tilde{V}^*$ that for Case (I)

$$\begin{aligned} \hat{\lambda}^*\left(\frac{\tilde{X}^{*'}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{r_n}\right) &= \lambda_n\left(\frac{\tilde{X}^{*'}\tilde{X}^*}{r_n}\right) + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) \\ &= \lambda_n\left(\frac{\hat{\Pi}'Z'Z\hat{\Pi}}{r_n}\right) + \left(\frac{l}{r_n}\right)\Sigma^b_{\tilde{V}\tilde{V}} + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right); \end{aligned}$$

using similar arguments, we obtain for Case (II) that

$$\hat{\lambda}^*\left(\frac{\tilde{X}^{*'}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{l}\right)=\lambda_n\left(\frac{\hat{\Pi}'Z'Z\hat{\Pi}}{l}\right)+\Sigma^b_{\tilde{V}\tilde{V}}+O_{P^*}\left(\frac{1}{\sqrt{l}}\right).$$

Proceeding similarly for the term $\tilde{X}^{*'}(\hat{\beta}^*)P_Z\tilde{X}^*(\hat{\beta}^*)/r_n$, we can show that for Case (I)

$$\begin{aligned}&\frac{\tilde{X}^{*'}(\hat{\beta}^*)P_Z\tilde{X}^*(\hat{\beta}^*)}{r_n}-\hat{\lambda}^*\left(\frac{\tilde{X}^{*'}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{r_n}\right)\\&=\frac{\tilde{X}^{*'}P_Z\tilde{X}^*}{r_n}-\hat{\lambda}^*\left(\frac{\tilde{X}^{*'}\tilde{X}^*}{r_n}\right)+O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)\\&=\bar{H}_{I,n}+O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)\end{aligned}$$

where $\bar{H}_{I,n}=(1-\lambda_n)\left(\hat{\Pi}'Z'Z\hat{\Pi}/r_n\right)$; for Case (II),

$$\frac{\tilde{X}^{*'}(\hat{\beta}^*)P_Z\tilde{X}^*(\hat{\beta}^*)}{l}-\hat{\lambda}^*\left(\frac{\tilde{X}^{*'}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{l}\right)=\bar{H}_{II,n}+O_{P^*}\left(\frac{1}{\sqrt{l}}\right),$$

where $\bar{H}_{II,n}=(1-\lambda_n)\left(\hat{\Pi}'Z'Z\hat{\Pi}/l\right)$.

In addition, we obtain by standard arguments that $\hat{\sigma}^*_{\epsilon\epsilon}(\hat{\beta}^*)=\hat{\epsilon}^{*'}(\hat{\beta}^*)\hat{\epsilon}^*(\hat{\beta}^*)/n=\sigma^b_{\epsilon\epsilon}+O_{P^*}\left(1/\sqrt{r_n}\right)$ for Case (I) and $\hat{\sigma}^*_{\epsilon\epsilon}(\hat{\beta}^*)=\sigma^b_{\epsilon\epsilon}+O_{P^*}\left(1/\sqrt{l}\right)$ for Case (II).

Putting these results together, we find that for Case (I)

$$\begin{aligned}r_n^{-1}\hat{\Upsilon}^*_{bkk}(\hat{\beta}^*)&=\hat{\sigma}^*_{\epsilon\epsilon}(\hat{\beta}^*)\left\{\left(1-2\hat{\lambda}^*(\hat{\beta}^*)\right)\left(\frac{\tilde{X}^{*'}(\hat{\beta}^*)P_Z\tilde{X}^*(\hat{\beta}^*)}{r_n}\right.\right.\\&\quad\left.-\hat{\lambda}^*(\hat{\beta}^*)\frac{\tilde{X}^{*'}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{r_n}\right)\\&\quad\left.+\hat{\lambda}^*(\hat{\beta}^*)\left(1-\hat{\lambda}^*(\hat{\beta}^*)\right)\frac{\tilde{X}^{*'}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{r_n}\right\}\\&=\sigma^b_{\epsilon\epsilon}\left\{(1-\lambda_n)\left[\bar{H}_{I,n}+\left(\frac{l}{r_n}\right)\Sigma^b_{\tilde{V}\tilde{V}}\right]\right\}\\&\quad+O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right),\end{aligned}$$

and for Case (II)

$$\begin{aligned}l^{-1}\hat{\Upsilon}^*_{bkk}(\hat{\beta}^*)&=\hat{\sigma}^*_{\epsilon\epsilon}(\hat{\beta}^*)\left\{\left(1-2\hat{\lambda}^*(\hat{\beta}^*)\right)\left(\frac{\tilde{X}^{*'}(\hat{\beta}^*)P_Z\tilde{X}^*(\hat{\beta}^*)}{l}\right.\right.\\&\quad\left.-\hat{\lambda}^*(\hat{\beta}^*)\frac{\tilde{X}^{*'}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{l}\right)\\&\quad\left.+\hat{\lambda}^*(\hat{\beta}^*)\left(1-\hat{\lambda}^*(\hat{\beta}^*)\right)\frac{\tilde{X}^{*'}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{l}\right\}\\&=\sigma^b_{\epsilon\epsilon}\left\{(1-\lambda_n)\left[\bar{H}_{II,n}+\Sigma^b_{\tilde{V}\tilde{V}}\right]\right\}+O_{P^*}\left(\frac{1}{\sqrt{l}}\right).\end{aligned}$$

(2) Now we show the results for the non-normality adjustment terms $\hat{A}^*(\hat{\beta}^*)$ and $\hat{B}^*(\hat{\beta}^*)$. Let $\hat{a}^*(\hat{\beta}^*)=(\hat{\epsilon}^{*2}_1(\hat{\beta}^*)-\sigma^b_{\epsilon\epsilon},\ldots,\hat{\epsilon}^{*2}_n(\hat{\beta}^*)-\sigma^b_{\epsilon\epsilon})'$, $a^*=\left(\epsilon^{*2}_1-\sigma^b_{\epsilon\epsilon},\ldots,\epsilon^{*2}_n-\sigma^b_{\epsilon\epsilon}\right)'$, and $\bar{V}^*=M_Z\tilde{V}^*$.

Note that $\hat{V}^*(\hat{\beta}^*)=M_Z\left(X^*-\hat{\epsilon}^*(\hat{\beta}^*)\frac{\hat{\epsilon}^{*'}(\hat{\beta}^*)X^*}{\hat{\epsilon}^{*'}(\hat{\beta}^*)\hat{\epsilon}^*(\hat{\beta}^*)}\right)$ and $\bar{V}^*=M_Z(V^*-\epsilon^*q^{b'})$, so that

$$\hat{V}^*(\hat{\beta}^*)-\bar{V}^*=M_Z\left(\epsilon^*q^{b'}-\hat{\epsilon}^*(\hat{\beta}^*)\frac{\hat{\epsilon}^{*'}(\hat{\beta}^*)X^*}{\hat{\epsilon}^{*'}(\hat{\beta}^*)\hat{\epsilon}^*(\hat{\beta}^*)}\right).$$

Then, in Case (I), it follows by $n^{-1}\left\|\hat{\epsilon}^*(\hat{\beta}^*)-\epsilon^*\right\|^2=O_{P^*}\left(\frac{1}{r_n}\right)$ and $\left\|\frac{\hat{\epsilon}^{*'}(\hat{\beta}^*)X^*}{\hat{\epsilon}^{*'}(\hat{\beta}^*)\hat{\epsilon}^*(\hat{\beta}^*)}-q^{b'}\right\|^2=O_{P^*}\left(\frac{1}{r_n}\right)$ that

$$n^{-1}\left\|\hat{V}^*(\hat{\beta}^*)-\bar{V}^*\right\|^2=O_{P^*}\left(\frac{1}{r_n}\right).$$

By using similar arguments, we obtain $n^{-1}\left\|\hat{a}^*(\hat{\beta}^*)-a^*\right\|^2=O_{P^*}\left(\frac{1}{r_n}\right)$.

Furthermore, note that by using arguments similar to Lemma A.2 and by Markov inequality, we have $n^{-1}a^{*'}a^*=n^{-1}\sum_{i=1}^n(\epsilon^{*2}_i-\sigma^b_{\epsilon\epsilon})^2=O_{P^*}(1)$ and $n^{-1}\bar{V}^{*'}\bar{V}^*=O_{P^*}(1)$. Then, we obtain by Cauchy–Schwarz inequality that in Case (I)

$$\begin{aligned}&n^{-1}\hat{a}^*(\hat{\beta}^*)'\hat{V}^*(\hat{\beta}^*)-n^{-1}a^{*'}\bar{V}^*\\&=n^{-1}\left(\hat{a}^*(\hat{\beta}^*)-a^*\right)'\left(\hat{V}^*(\hat{\beta}^*)-\bar{V}^*\right)+n^{-1}\left(\hat{a}^*(\hat{\beta}^*)-a^*\right)'\bar{V}^*\\&\quad+n^{-1}a^{*'}(\hat{V}^*(\hat{\beta}^*)-\bar{V}^*)=O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right).\end{aligned}$$

In addition, we have $n^{-1}a^{*'}\bar{V}^*=(1-\lambda_n)E^*\left(\epsilon^{*2}_i\tilde{V}^*_i\right)+O_{P^*}\left(\frac{1}{\sqrt{n}}\right)$ by Lemma A.3, and we have $n^{-1}\sum_{i=1}^n\hat{V}^*_i(\hat{\beta}^*)=O_{P^*}\left(1/\sqrt{n}\right)$ by standard arguments. It follows by the triangle inequality that

$$n^{-1}\sum_{i=1}^n\hat{\epsilon}^{*2}_i(\hat{\beta}^*)\hat{V}^{*'}_i(\hat{\beta}^*)=(1-\lambda_n)E^*\left(\epsilon^{*2}_i\tilde{V}^*_i\right)+O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right).\tag{17}$$

Now, let $d_i=(P_{ii}-\lambda_n)/\sqrt{l}$ and $d=(d_1,\ldots,d_n)'$. Notice that $\|d\|^2\le 1$ and $E^*\left\|V^{*'}P_Zd\right\|^2=O_P(1)d'd=O_P(1)$. Thus, $V^{*'}P_Zd=O_{P^*}(1)$ by Markov inequality. Then, $\frac{1}{\sqrt{r_n}}\sum_{i=1}^n\hat{\Gamma}^*_i\left(\frac{P_{ii}-\lambda_n}{\sqrt{l}}\right)=\frac{1}{\sqrt{r_n}}\left(\hat{\Pi}'Z'd+V^{*'}P_Zd\right)=\frac{1}{\sqrt{r_n}}\hat{\Pi}'Z'd+O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$, where $\hat{\Gamma}^*=P_ZX^*$. Then we obtain with Eq. (17) that

$$r_n^{-1}\hat{A}^*(\hat{\beta}^*)=(1-\lambda_n)\left(\sqrt{\frac{l}{r_n}}\right)\sum_{i=1}^n\left(\frac{(P_{ii}-\lambda_n)\hat{\Pi}'Z_i}{\sqrt{lr_n}}\right)E^*\left(\epsilon^{*2}_i\tilde{V}^*_i\right)+O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right).\tag{18}$$

For Case (II), we obtain $n^{-1}\left\|\hat{a}^*(\hat{\beta}^*)-a^*\right\|^2=O_{P^*}\left(\frac{1}{l}\right)$, $n^{-1}\left\|\hat{V}^*(\hat{\beta}^*)-\bar{V}^*\right\|^2=O_{P^*}\left(\frac{1}{l}\right)$, $n^{-1}a^{*'}\bar{V}^*=(1-\lambda_n)E^*\left(\epsilon^{*2}_i\tilde{V}^*_i\right)+O_{P^*}\left(\frac{1}{\sqrt{n}}\right)$, $n^{-1}\sum_{i=1}^n\hat{\epsilon}^{*2}_i(\hat{\beta}^*)\hat{V}^*_i(\hat{\beta}^*)=(1-\lambda_n)E^*\left(\epsilon^{*2}_i\tilde{V}^*_i\right)+O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$, and

$$l^{-1}\hat{A}^*(\hat{\beta}^*)=(1-\lambda_n)\sum_{i=1}^n\left(\frac{(P_{ii}-\lambda_n)\hat{\Pi}'Z_i}{l}\right)E^*\left(\epsilon^{*2}_i\tilde{V}^*_i\right)+O_{P^*}\left(\frac{1}{\sqrt{l}}\right)\to^{P^*}0,$$

in probability, because $\sum_{i=1}^n\left(\frac{(P_{ii}-\lambda_n)\hat{\Pi}'Z_i}{l}\right)\to^P 0$.

For $\hat{B}^*(\hat{\beta}^*)$ term, it follows similarly to previous arguments that in Case (I)

$$\left\| n^{-1} \sum_{i=1}^{n} \left(\hat{\epsilon}_i^{*2}(\hat{\beta}^*) - \hat{\sigma}_{\epsilon\epsilon}^*(\hat{\beta}^*) - (\epsilon_i^{*2} - \sigma_{\epsilon\epsilon}^b) \right) \bar{V}_i^* \bar{V}_i^{*'} \right\| = O_{P^*}\left(\frac{1}{\sqrt{r_n}} \right);$$

$$\left\| n^{-1} \sum_{i=1}^{n} \left(\hat{\epsilon}_i^{*2}(\hat{\beta}^*) - \hat{\sigma}_{\epsilon\epsilon}^*(\hat{\beta}^*) \right) \left(\hat{V}_i^*(\hat{\beta}^*) \hat{V}_i^{*'}(\hat{\beta}^*) - \bar{V}_i^* \bar{V}_i^{*'} \right) \right\| = O_{P^*}\left(\frac{1}{\sqrt{r_n}} \right),$$

in probability, by which we obtain that

$$\begin{aligned} n^{-1} \sum_{i=1}^{n} \left(\hat{\epsilon}_i^{*2}(\hat{\beta}^*) - \hat{\sigma}_{\epsilon\epsilon}^*(\hat{\beta}^*) \right) \hat{V}^*(\hat{\beta}^*) \hat{V}^{*'}(\hat{\beta}^*) &= n^{-1} \sum_{i=1}^{n} (\epsilon_i^{*2} - \sigma_{\epsilon\epsilon}^b) \bar{V}_i^* \bar{V}_i^{*'} + O_{P^*}\left(\frac{1}{\sqrt{r_n}} \right) \\ &= (1 - 2\lambda_n + \lambda_n \phi_n) E^* \left((\epsilon_i^{*2} - \sigma_{\epsilon\epsilon}^b) \tilde{V}_i^* \tilde{V}_i^{*'} \right) + O_{P^*}\left(\frac{1}{\sqrt{r_n}} \right) \end{aligned}$$

in probability, where the second equality follows from Lemma A.3 and from arguments similar to the proofs of Lemma A11 in Hansen et al. (2008). It follows that

$$r_n^{-1} \hat{B}^*(\hat{\beta}^*) = \frac{l(\phi_n - \lambda_n)}{r_n(1 - 2\lambda_n + \lambda_n \phi_n)} \times \left(n^{-1} \sum_{i=1}^{n} \left(\hat{\epsilon}_i^{*2}(\hat{\beta}^*) - \hat{\sigma}_{\epsilon\epsilon}^*(\hat{\beta}^*) \right) \hat{V}^*(\hat{\beta}^*) \hat{V}^{*'}(\hat{\beta}^*) \right) \tag{19}$$

$$= (\phi_n - \lambda_n) \left(\frac{l}{r_n} \right) E^* \left((\epsilon_i^{*2} - \sigma_{\epsilon\epsilon}^b) \tilde{V}_i^* \tilde{V}_i^{*'} \right) + O_{P^*}\left(\frac{1}{\sqrt{r_n}} \right) \tag{20}$$

in probability. Similarly, we have for Case (II)

$$n^{-1} \sum_{i=1}^{n} \left(\hat{\epsilon}_i^{*2}(\hat{\beta}^*) - \hat{\sigma}_{\epsilon\epsilon}^*(\hat{\beta}^*) \right) \hat{V}^*(\hat{\beta}^*) \hat{V}^{*'}(\hat{\beta}^*) = (1 - 2\lambda_n + \lambda_n \phi_n) E^* \left((\epsilon_i^{*2} - \sigma_{\epsilon\epsilon}^b) \tilde{V}_i^* \tilde{V}_i^{*'} \right) + O_{P^*}\left(\frac{1}{\sqrt{l}} \right),$$

and $l^{-1} \hat{B}^*(\hat{\beta}^*) = (\phi_n - \lambda_n) E^* \left((\epsilon_i^{*2} - \sigma_{\epsilon\epsilon}^b) \tilde{V}_i^* \tilde{V}_i^{*'} \right) + O_{P^*}\left(\frac{1}{\sqrt{l}} \right)$ in probability.

(3) Finally, we show the result for t_{cse}^*, the bootstrap analogue of the t-ratio based on the CSE. Notice that in Case (I)

$$\begin{aligned} r_n \times \hat{\Lambda}^*(\hat{\beta}^*) &= \left(r_n^{-1} \hat{H}^*(\hat{\beta}^*) \right)^{-1} \left\{ r_n^{-1} \left(\hat{\Upsilon}_{bkk}^*(\hat{\beta}^*) + \hat{A}^*(\hat{\beta}^*) + \hat{A}^{*'}(\hat{\beta}^*) + \hat{B}^*(\hat{\beta}^*) \right) \right\} \left(r_n^{-1} \hat{H}^*(\hat{\beta}^*) \right)^{-1} \\ &\to^{P^*} \bar{\Lambda}_I \end{aligned}$$

in probability, which follows from $\hat{\beta}^* - \hat{\beta} = O_{P^*}(1/\sqrt{r_n})$, Lemmas A.6 and A.7, and the previous results for $\hat{\Upsilon}_{bkk}^*(\hat{\beta}^*)$, $\hat{A}^*(\hat{\beta}^*)$, $\hat{B}^*(\hat{\beta}^*)$. It then follows that

$$\frac{c' \sqrt{r_n} \left(\hat{\beta}^* - \hat{\beta} \right)}{\sqrt{c' r_n \hat{\Lambda}^* \left(\hat{\beta}^* \right) c}} \to^{d^*} N(0, 1)$$

in probability, by Theorem 3.1 and by continuous mapping theorem for weak convergence in probability (e.g., Xiong and Li (2008), Theorem 3.1).

For Case (II), we have $l \times \hat{\Lambda}^*(\hat{\beta}^*) \to^{P^*} \bar{\Lambda}_{II}$ in probability, and the desired result follows by using similar arguments as in Case (I). ■

References

Anatolyev, S., 2012. Inference in regression models with many regressors. J. Econometrics 170 (2), 368–382.
Anatolyev, S., Gospodinov, N., 2011. Specification testing in models with many instruments. Econometric Theory 27 (2), 427–441.
Anderson, T.W., Kunitomo, N., Matsushita, Y., 2010. On the asymptotic optimality of the LIML estimator with possibly many instruments. J. Econometrics 157 (2), 191–204.
Anderson, T.W., Kunitomo, N., Matsushita, Y., 2011. On finite sample properties of alternative estimators of coefficients in a structural equation with many instruments. J. Econometrics 165 (1), 58–69.
Andrews, D.W., Cheng, X., 2012. Estimation and inference with weak, semi-strong, and strong identification. Econometrica 80 (5), 2153–2211.
Angrist, J., Imbens, G., Krueger, A., 1999. Jackknife instrumental variables estimation. J. Appl. Econometrics 14 (1), 57–67.
Angrist, J.D., Krueger, A.B., 1991. Does compulsory school attendance affect schooling and earnings?. Q. J. Econ. 106 (4), 979–1014.
Antoine, B., Renault, E., 2009. Efficient GMM with nearly-weak instruments. Econom. J. 12 (1), 135–171.
Antoine, B., Renault, E., 2012. Efficient minimum distance estimation with multiple rates of convergence. J. Econometrics 170 (2), 350–367.
Antoine, B., Renault, E., 2013. Testing identification strength, Discussion Paper dp12-17. Department of Economics, Simon Fraser University.
Bekker, P.A., 1994. Alternative approximations to the distributions of instrumental variable estimators. Econometrica 62 (3), 657–681.
Bekker, P.A., Crudu, F., 2015. Jackknife instrumental variable estimation with heteroskedasticity. J. Econometrics 185 (2), 332–342.
Beran, R., 1988. Prepivoting test statistics: A bootstrap view of asymptotic refinements. J. Amer. Statist. Assoc. 83 (403), 687–697.
Bickel, P.J., Freedman, D.A., 1983. Bootstrapping regression models with many parameters. Festschrift for Erich L. Lehmann 28–48.
Brown, B.W., Newey, W.K., 2002. Generalized method of moments, efficient bootstrapping, and improved inference. J. Bus. Econom. Statist. 20 (4), 507–517.
Chang, Y., Park, J.Y., 2003. A sieve bootstrap for the test of a unit root. J. Time Series Anal. 24 (4), 379–400.
Chao, J.C., Hausman, J.A., Newey, W.K., Swanson, N.R., Woutersen, T., 2014. Testing overidentifying restrictions with many instruments and heteroskedasticity. J. Econometrics 178, 15–21. Part 1.
Chao, J.C., Swanson, N.R., 2005. Consistent estimation with a large number of weak instruments. Econometrica 73 (5), 1673–1692.
Chao, J.C., Swanson, N.R., Hausman, J.A., Newey, W.K., Woutersen, T., 2012. Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments. Econometric Theory 28 (1), 42–86.
Davidson, R., MacKinnon, J.G., 2008. Bootstrap inference in a linear equation estimated by instrumental variables. Econom. J. 11 (3), 443–477.
Davidson, R., MacKinnon, J.G., 2010. Wild bootstrap tests for IV regression. J. Bus. Econom. Statist. 28 (1), 128–144.
Davidson, R., MacKinnon, J.G., 2014. Bootstrap confidence sets with weak instruments. Econometric Rev. 33 (5–6), 651–675.
Donald, S.G., Newey, W.K., 2001. Choosing the number of instruments. Econometrica 69 (5), 1161–1191.
Dufour, J.-M., 1997. Some impossibility theorems in econometrics with applications to structural and dynamic models. Econometrica 65 (6), 1365–1387.
Freedman, D., 1984. On bootstrapping two-stage least-squares estimates in stationary linear models. Ann. Statist. 12 (3), 827–842.
Fuller, W.A., 1977. Some properties of a modification of the limited information estimator. Econometrica 45 (4), 939–953.
Hall, P., Horowitz, J.L., 1996. Bootstrap critical values for tests based on generalized-method-of-moments estimators. Econometrica 64 (4), 891–916.
Hansen, C., Hausman, J., Newey, W., 2008. Estimation with many instrumental variables. J. Bus. Econom. Statist. 26 (4), 398–422.
Hasselt, M.v., 2010. Many instruments asymptotic approximations under nonnormal error distributions. Econometric Theory 26 (2), 633–645.
Hausman, J.A., Newey, W.K., Woutersen, T., Chao, J.C., Swanson, N.R., 2012. Instrumental variable estimation with heteroskedasticity and many instruments. Quant. Econ. 3 (2), 211–255.
Heyde, C., Brown, B., 1970. On the departure from normality of a certain class of martingales. Ann. Math. Stat. 41 (6), 2161–2165.
Kleibergen, F., 2002. Pivotal statistics for testing structural parameters in instrumental variables regression. Econometrica 70 (5), 1781–1803.
Kuersteiner, G., Okui, R., 2010. Constructing optimal instruments by first-stage prediction averaging. Econometrica 78 (2), 697–718.
Kunitomo, N., 1980. Asymptotic expansions of the distributions of estimators in a linear functional relationship and simultaneous equations. J. Amer. Statist. Assoc. 75 (371), 693–700.
Mammen, E., 1989. Asymptotics with increasing dimension for robust regression with applications to the bootstrap. Ann. Statist. 17 (1), 382–400.

Mammen, E., 1993. Bootstrap and wild bootstrap for high dimensional linear models. Ann. Statist. 21 (1), 255–285.
Moreira, M.J., 2003. A conditional likelihood ratio test for structural models. Econometrica 71 (4), 1027–1048.
Moreira, M.J., Porter, J.R., Suarez, G.A., 2009. Bootstrap validity for the score test when instruments may be weak. J. Econometrics 149 (1), 52–64.
Morimune, K., 1983. Approximate distributions of k-class estimators when the degree of overidentifiability is large compared with the sample size. Econometrica 51 (3), 821–841.
Nagar, A., 1959. The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. Econometrica 27 (4), 575–595.
Newey, W.K., Windmeijer, F., 2009. Generalized method of moments with many weak moment conditions. Econometrica 77 (3), 687–719.
Phillips, G.D.A., Hale, C., 1977. The bias of instrumental variable estimators of simultaneous equation systems. Internat. Econom. Rev. 18 (1), 219–228.
Portnoy, S., 1984. Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. Ann. Statist. 12 (4), 1298–1309.
Portnoy, S., 1985. Asymptotic behavior of M estimators of p regression parameters when p^2/n is large; II. Normal approximation. Ann. Statist. 13 (4), 1403–1417.
Portnoy, S., 1988. Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. Ann. Statist. 16 (1), 356–366.
Rothenberg, T.J., 1984. Approximating the distributions of econometric estimators and test statistics. In: Handbook of Econometrics, vol. 2. pp. 881–935.
Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65 (3), 557–586.
Stock, J.H., Yogo, M., 2005. Asymptotic distributions of instrumental variables statistics with many instruments. In: Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. pp. 109–120.
Xiong, S., Li, G., 2008. Some results on the convergence of conditional distributions. Statist. Probab. Lett. 78 (18), 3249–3253.