Size-corrected Bootstrap Test after Pretesting for Exogeneity with Heteroskedastic or Clustered Data

Doko Tchatoka, Firmin and Wang, Wenjie

29 November 2021

Size-corrected Bootstrap Test after Pretesting for Exogeneity with Heteroskedastic or Clustered Data

Firmin Doko Tchatoka*
The University of Adelaide

Wenjie Wang †
Nanyang Technological University

November 29, 2021

ABSTRACT

Pretesting for exogeneity has become a routine in many empirical applications involving instrumental variables to decide whether the ordinary least squares or the two-stage least squares (2SLS) method is appropriate. Guggenberger (2010) shows that the second-stage t-test – based on the outcome of a Durbin-Wu-Hausman type pretest for exogeneity in the first stage – has extreme size distortion with asymptotic size equal to 1 when the standard asymptotic critical values are used. In this paper, we first show that both conditional and unconditional on the data, the standard wild bootstrap procedures are invalid for the two-stage testing and a closely related shrinkage method, and therefore are not viable solutions to such size-distortion problem. Then, we propose a novel size-corrected wild bootstrap approach, which combines certain wild bootstrap critical values along with an appropriate size-correction method. We establish uniform validity of this procedure under either conditional heteroskedasticity or clustering in the sense that the resulting tests achieve correct asymptotic size. Monte Carlo simulations confirm our theoretical findings. In particular, our proposed method has remarkable power gains over the standard 2SLS-based t-test in many settings, especially when the identification is not strong.

Key words: DWH Pretest; Shrinkage; Instrumental Variable; Asymptotic Size; Wild Bootstrap; Bonferroni-based Size-correction; Clustering.

JEL classification: C12; C13; C26.

---

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

† Division of Economics, School of Social Sciences, Nanyang Technological University. HSS-04-65, 14 Nanyang Drive, 637332, SINGAPORE. e-mail: wang.wj@ntu.edu.sg

1. Introduction

Inference after data-driven model selection is widely studied in both statistical and econometric literature. For instance, see Hansen (2005), Leeb and Pötscher (2005), who provide an overview of the importance and difficulty of conducting valid inference after model selection. In particular, it is now well known that widely used model-selection practices such as pretesting may have large impact on the size properties of two-stage procedures and thus invalidate inference on parameter of interest in the second stage. For the classical linear regression model with exogenous covariates, Kabaila (1995) and Leeb and Pötscher (2005) show that confidence intervals (CIs) based on consistent model selection have serious problem of under-coverage, while Andrews and Guggenberger (2009b) show that such CIs have asymptotic confidence size equal to 0. Furthermore, Andrews and Guggenberger (2009a) find extreme size distortion for the two-stage test after "conservative" model selection and propose various least favourable critical values (CVs).

In comparison, the literature on models that contain *endogenous covariates*, such as widely used instrumental variable (IV) regression models, remains relatively sparse. The uniform validity of post-selection inference for structural parameters in linear IV models with homoskedastic errors was studied by Guggenberger (2010a), who advised not to use Hausman-type pretesting to select between ordinary least squares (OLS) and two-stage least squares (2SLS)-based t-tests because such two-stage procedure can be extremely over-sized with asymptotic CVs.[1] Instead, Guggenberger (2010a) recommended to use the standard 2SLS-based t-test. However, it is well known that the 2SLS-based t-test itself may have undesirable finite-sample size properties when IVs are not strong enough. As such, in the quest for statistical power, many empirical practitioners still use pretesting in IV applications despite the important concern raised by Guggenberger (2010a).[2]

Recently, Young (2020) analyzes a sample of 1359 empirical applications involving IV regressions in 31 papers published in the American Economic Association (AEA): 16 in AER, 6 in AEJ: A.Econ., 4 in AEJ: E.Policy, and 5 in AEJ: Macro. He highlights that the IVs often do not appear to be strong in these papers, so that inference methods based-on standard normal CVs can be unreliable, especially in the case with heteroskedastic or clustered errors, and he advocates for the usage of bootstrap methods to improve the quality of inference. Furthermore, he argues that in these papers IV confidence intervals almost always include OLS point estimates and there is little statistical evidence of endogeneity and evidence that OLS is seriously biased, based on the low rejection rates of Hausman-type tests in his data. In his simulations based upon the published regressions (Table XIV), the rejection frequencies can be as low as 0.237 and 0.386 for 1% and 5% significance levels, respectively, for asymptotic Hausman tests, and even as low as 0.105 and

[1] Similar concerns were also raised by Guggenberger and Kumar (2012) about pretesting the instrument exogeneity using a test of overidentifying restrictions, and by Guggenberger (2010b) about pretesting for the presence of random effects before inference on the parameters of interest in panel data models.

[2] Their motivation of implementing the pretesting procedure also lies in the fact that valid IVs (i.e., exogenous IVs) found in practice are often rather uninformative, while strong IVs are typically more or less invalid and such deviation from IV exogeneity may also lead to serious size distortion in the 2SLS-based t-test; e.g., see Conley, Hansen and Rossi (2012), Guggenberger (2012), Andrews, Gentzkow and Shapiro (2017).

0.208, respectively, for bootstrap Hausman tests.[3]

However, Young (2020)'s finding from the AEA data that OLS estimates seem to be not very different from 2SLS estimates may be attributed to the fact that the used IVs are not strong so that 2SLS may be biased towards OLS, and Hausman-type tests also have low power in this case [e.g., see Doko Tchatoka and Dufour (2018, 2020)]. In particular, as shown by Guggenberger (2010a), the Hausman test is not able to reject the null hypothesis of exogeneity in situations where there is only a small degree of endogeneity, i.e., local endogeneity. Then, OLS-based inference is selected in the second stage with high probability. However, the OLS-based t-statistic often takes on very large values even under such local endogeneity, causing extreme size distortions in the two-stage test. Indeed, Guggenberger (2010a) shows that the asymptotic size of this procedure equals 1 even with homoskedastic errors. Such issue with pretesting for exogeneity is highly relevant to empirical practice as endogeneity is mild in many IV applications. For example, Hansen, Hausman and Newey (2008) report that the median, 75th quantile, and 90th quantile of estimated endogeneity parameters are only 0.279, 0.466, and 0.555, respectively, in their investigated AER, JPE, and QJE papers. Angrist and Kolesár (2021) investigate three influential just-identified IV applicaitions: Angrist and Krueger (1991), Angrist and Evans (1998), Angrist and Lavy (1999), and find that the estimated endogeneity is no more than 0.175, 0.075, and 0.460 for different specifications and samples in these papers, respectively [see Section 3.1 and Table 1 in Angrist and Kolesár (2021)].

Motivated by these issues, we study in this paper the possibility of proposing uniformly valid method for the above two-stage testing procedure and a closely related Stein-type shrinkage procedure proposed by Hansen (2017), and we consider an asymptotic framework under conditional heteroskedasticity or clustering, as allowing for non-homoskedastic errors is paramount for the methodology to be useful in practice. Given Young (2020)'s recommendation of using bootstrap, we first study the validity of bootstrapping the two procedures by obtaining the null limiting distributions of the bootstrap statistics and their associated asymptotic null rejection probabilities. Such (unconditional) asymptotic null rejection results are useful for the study of bootstrap validity because even if the bootstrap cannot consistently estimate the distribution of interest conditional on the data (i.e., bootstrap invalidity in the usual sense), it may still be possible that the bootstrap test controls the asymptotic size and thus is valid unconditionally; e.g., see Cavaliere and Georgiev (2020) and the references therein. Here, we find that the standard wild bootstrap procedures are invalid both conditionally and unconditionally for the two-stage and shrinkage procedures even under strong IVs.[4] In particular, the usual intuition for bootstrapping Durbin-Wu-Hausman (DWH) tests is that one should restrict the bootstrap data generating process (DGP) under exogeneity of the regressors. However, we find that such bootstrap DGP can result in extreme size distortion with asymptotic null rejection probabilities close to 1 in some settings, while the bootstrap DGP

[3] Similarly, Keane and Neal (2021, Section 8) argue that a rather strong IV is necessary to give high confidence that 2SLS will outperform OLS (e.g., with a first-stage F higher than 50, which is well above the industry standard of 10).

[4] For the case with weak IVs in the sense of Staiger and Stock (1997), it is well documented in the literature that resampling methods such as bootstrap and subsampling can be inconsistent (i.e., invalid conditional on the data); see e.g., Andrews and Guggenberger (2010b), Wang and Doko Tchatoka (2018) and Wang (2020).

without such restriction typically has much smaller asymptotic size distortions.[5]

To address such bootstrap failure, we then propose a novel size-corrected wild bootstrap procedure, which combines certain standard wild bootstrap CVs with an appropriate Bonferroni-based size-correction method, following the lead of McCloskey (2017). We first show that the resulting CVs are uniformly valid with heteroskedastic errors in the sense that they yield two-stage and shrinkage tests with correct asymptotic size. In particular, since standard wild bootstrap procedures cannot mimic well the key localized endogeneity parameter, particular attention is taken on this parameter when designing bootstrap DGP, and a Bonferroni-based size-correction technique is implemented to deal with the presence of this localization parameter in the limiting distributions of interest. Different from the conventional Bonferroni bound, which may lead to conservative test with asymptotic size strictly less than the nominal level, the size-correction procedure always leads to desirable asymptotic size. Then, we extend the uniform validity result to clustered samples, in which case the rate of convergence of the estimators depends on the regressor, the instruments, the relative cluster size and the intra-cluster correlation structure in a complicated way.

In terms of practical usage of our method, following the aforementioned studies by Hansen et al. (2008), Young (2020), and Angrist and Kolesár (2021), we are particularly interested in the IV applications where the values of endogeneity parameters are relatively small. These are the cases where the pretest would not reject exogeneity and the naive two-stage procedure would lead to extreme size distortion. On the other hand, as the problem of size distortion is circumvented by our method, we may take advantage of the power superiority of the OLS-based t-test over its 2SLS counterpart. In addition, Hansen (2017) shows that his shrinkage estimator has substantially reduced median squared error relative to 2SLS, and Doko Tchatoka and Dufour (2020) show that their pretest estimators based on DWH tests can outperform both OLS and 2SLS estimators in terms of mean squared error if identification is not very strong, even with moderate endogeneity. As such, our proposed method is also attractive from the viewpoint of providing valid inference method for the shrinkage or pretesting estimator. Monte Carlo experiments confirm that our size-corrected bootstrap procedure is able to achieve reliable size correction and remarkable power gains over the 2SLS-based t-test, especially when the identification is not strong. We also note that the size-corrected Hansen-type shrinkage procedure has superior finite-sample power performance than its Hausman-type counterpart in many cases.

Our size-correction procedure follows closely the seminal study by McCloskey (2017), who proposed Bonferroni-based size-correction procedures for general nonstandard testing problems, and McCloskey (2020) applied this method to inference in linear regression model after consistent model selection. Additionally, Han and McCloskey (2019) applied it to inference in moment condition models where the estimating function may exhibit mixed identification strength and a nearly singular Jacobian, and Wang and Doko Tchatoka (2018) applied it to weak-identification-robust

[5] These results are in contrast to the case of bootstrapping the DWH tests only (without the second-stage t-test), which achieves higher-order refinement under strong IVs and remains first-order valid even under weak IVs; e.g., see Doko Tchatoka (2015).

subvector inference in linear IV models. Different from our size-corrected bootstrap procedures, these procedures are based on simulations from null limiting distributions. The motivation of using bootstrap in the current testing problem originates from a growing literature illustrating that when applied to IV regressions, well designed bootstrap procedures typically have superior finite-sample performance than asymptotic approximations; see, e.g., Davidson and MacKinnon (2008, 2010), Wang and Kaffo (2016), Finlay and Magnusson (2019), Young (2020), and Wang and Zhang (2021). Furthermore, we are motivated by the growing literature showing the excellent performance of wild bootstrap methods with heteroskedastic or clustered errors, among them Davidson and Flachaire (2008), Cameron, Gelbach and Miller (2008), MacKinnon and Webb (2017), and Djogbenou, MacKinnon and Nielsen (2019).

The remainder of this paper is organized as follows. Section 2 presents the setting, test statistics, and parameter space of interest. Section 3 presents the main results of both standard and size-corrected wild bootstrap methods. Section 4 investigates the finite sample power performance of our methods using simulations. Conclusions are drawn in Section 5. The proofs and further simulation results are provided in the Appendix and Supplementary Material.

Throughout the paper, for any positive integers n and m, I_n and $0_{n\times m}$ stand for the $n\times n$ identity matrix and $n\times m$ zero matrix, respectively. For any full-column rank $n\times m$ matrix A, $P_A = A(A'A)^{-1}A'$ is the projection matrix on the space spanned by the columns of A, and $M_A = I_n - P_A$. The notation $vec(A)$ is the $nm\times 1$ dimensional column vectorization of A. $\lambda_{min}(A)$ denote the minimum eigenvalue of a square matrix A. $\|U\|$ denotes the usual Euclidean or Frobenius norm for a matrix U. The usual orders of magnitude are denoted by $O_P(.)$ and $o_P(.)$, $\rightarrow^P$ stands for convergence in probability, while $\rightarrow^d$ stands for convergence in distribution. We write P^* to denote the probability measure induced by a bootstrap procedure conditional on the data, and E^* and Var^* to denote the expected value and variance with respect to P^*. For any bootstrap statistic T^* we write $T^* \rightarrow^{P^*} 0$ in probability P if for any $\delta > 0$, $\varepsilon > 0$, $lim_{n\rightarrow\infty} P[P^*(|T^*| > \delta) > \varepsilon] = 0$, i.e., $P^*(|T^*| > \delta) = o_P(1)$; e.g., see Gonçalves and White (2004). Also, we write $T^* = O_{P^*}(n^{\varphi})$ in probability P if and only if for any $\delta > 0$ there exists a $M_\delta < \infty$ such that $lim_{n\rightarrow\infty} P[P^*(|n^{-\varphi}T^*| > M_\delta) > \delta] = 0$, i.e., for any $\delta > 0$ there exists a $M_\delta < \infty$ such that $P^*(|n^{-\varphi}T^*| > M_\delta) = o_P(1)$. Finally, we write $T^* \rightarrow^{d^*} T$ in probability P if, conditional on the data, T^* weakly converges to T under P^*, for all samples contained in a set with probability approaching one.

2. Framework

2.1. Model and test statistics

We consider the following linear IV model

$$y = X\theta + u, \quad X = Z\pi + v, \tag{2.1}$$

where $y \in \mathbb{R}^n$ and $X \in \mathbb{R}^n$ are vectors of dependent and endogenous variables, respectively, $Z \in \mathbb{R}^{n\times k}$ is a matrix of instruments ($k \geq 1$), $(\theta, \pi')' \in \mathbb{R}^{k+1}$ are unknown parameters, and n is the sample size. Denote by u_i, v_i, y_i, X_i, and Z_i the i-th rows of u, v, y, X, and Z respectively, written as column vectors or scalars. Assume that $\{(u_i, v_i, Z_i) : i \leq n\}$ are i.i.d. with distribution F. For notational simplicity, also assume that the other exogenous variables have been partialled out.

The object of inferential interest is the structural parameter θ and we consider the problem of testing the null hypothesis $H_0 : \theta = \theta_0$. We study the two-stage testing procedure for assessing H_0, where an exogeneity test is undertaken in the first stage to decide whether a t-test based on the OLS or 2SLS estimator is appropriate for testing H_0 in the second stage. Assume that the instruments Z are exogenous, i.e., $E_F[u_i Z_i] = 0$, where E_F denotes expectation under the distribution F. Under this orthogonality condition of the instruments, X is endogenous in (2.1) if and only if v and u are correlated. Consider the following linear projection of u on v:

$$u = va + e,\ a = (E_F[v_i^2])^{-1} E_F[v_i u_i], \tag{2.2}$$

where e is uncorrelated with v. Notice that the exogeneity of X in (2.1) can be assessed by testing the null hypothesis $H_a :\ a = 0$ in (2.2). Substituting (2.2) into (2.1), we obtain

$$y = X\theta + va + e, \tag{2.3}$$

where X and v are uncorrelated with e. Therefore, the null hypothesis of exogeneity $H_a :\ a = 0$ can be assessed using a standard Wald statistic in the extended regression (2.3) [e.g., see Doko Tchatoka and Dufour (2014)]. To account for possible conditional heteroskedasticity, we consider the following control function-based Wald statistic:[6]

$$H_n = \hat{a}^2 / \hat{V}_a, \tag{2.4}$$

where $\hat{a} = (\tilde{v}'\tilde{v})^{-1}\tilde{v}'y$, $\hat{V}_a = (n^{-1}\tilde{v}'\tilde{v})^{-1}\left(n^{-2}\sum_{i=1}^n \tilde{v}_i^2 \hat{e}_i^2\right)(n^{-1}\tilde{v}'\tilde{v})^{-1}$ is the Eicker-White heteroskedasticity-robust estimator of the variance of $\hat{a}$, $\tilde{v} = M_X \hat{v}$, $\hat{v} = M_Z X$, and $\hat{e} = M_{[X:\hat{v}]} y$. Note that $\hat{e}$ is the residual vector from the OLS regression of y on X and $\hat{v}$. If θ is strongly identified (Z being strong instruments) and X is exogenous, H_n follows a χ_1^2 distribution asymptotically. The pretest rejects the null hypothesis that X is exogenous in (2.1) if $H_n > \chi^2_{1,1-\beta}$, where $\chi^2_{1,1-\beta}$ is the $(1-\beta)$-th quantile of χ_1^2-distributed random variable for some $\beta \in (0,1)$.

In addition, let $\hat{\theta}_{2sls} = (X'P_Z X)^{-1} X'P_Z y$, and $\hat{\theta}_{ols} = (X'X)^{-1}X'y$ be the 2SLS and OLS estimators of θ in (2.1), respectively. Also, define their corresponding variance estimators as

$$\hat{V}_{2sls} = \left(n^{-1}X'P_Z X\right)^{-1} \hat{\pi}' \left(n^{-2}\sum_{i=1}^n Z_i Z_i' \hat{u}_i^2(\hat{\theta}_{2sls})\right) \hat{\pi} \left(n^{-1}X'P_Z X\right)^{-1},$$

[6] Alternative formulations of this exogeneity statistic are given in Hahn, Ham and Moon (2010), Doko Tchatoka and Dufour (2018, 2020) but the Wald version considered in (2.4) easily accommodates conditional heteroskedasticity or clustering, so we shall use this formulation.

$$\hat{V}_{ols} = \left(n^{-1}X'X\right)^{-1}\left(n^{-2}\sum_{i=1}^{n}X_iX_i'\hat{u}_i^2(\hat{\theta}_{ols})\right)\left(n^{-1}X'X\right)^{-1}, \tag{2.5}$$

where $\hat{u}_i(\hat{\theta}_{2sls}) = y_i - X_i\hat{\theta}_{2sls}$, $\hat{u}_i(\hat{\theta}_{ols}) = y_i - X_i\hat{\theta}_{ols}$, and $\hat{\pi} = (Z'Z)^{-1}Z'X$. Then, the two-stage test statistic associated with the H_n-based pretest of exogeneity in the first stage is given by

$$\bar{T}_{1,n}(\theta_0) = T_{ols}(\theta_0)\mathbb{1}(H_n \leq \chi^2_{1,1-\beta}) + T_{2sls}(\theta_0)\mathbb{1}(H_n > \chi^2_{1,1-\beta}), \tag{2.6}$$

where $T_{ols}(\theta)$ and $T_{2sls}(\theta)$ are the t-statistics with 2SLS and OLS estimates, respectively, i.e.,

$$T_{2sls}(\theta) = (\hat{\theta}_{2sls} - \theta)/\hat{V}_{2sls}^{1/2}, \quad \text{and} \quad T_{ols}(\theta) = (\hat{\theta}_{ols} - \theta)/\hat{V}_{ols}^{1/2}. \tag{2.7}$$

Related to the two-stage procedure, Hansen (2017) proposed a Stein-like shrinkage approach in the context of IV regressions. His estimator follows Maasoumi (1978) in taking a weighted average of the 2SLS and OLS estimators, with the weight depending inversely on the test statistic for exogeneity, and the proposed shrinkage estimator is found to have substantially reduced finite-sample median squared error relative to the 2SLS estimator. Following Hansen (2017)'s approach, we define the Stein-like shrinkage test statistic as follows:

$$\bar{T}_{2,n}(\theta_0) = T_{ols}(\theta_0)w(H_n) + T_{2sls}(\theta_0)(1 - w(H_n)), \tag{2.8}$$

where the weight function takes the form $w(H_n) = \begin{cases} \tau/H_n & \text{if } H_n \geq \tau \\ 1 & \text{if } H_n < \tau \end{cases}$, and τ is a shrinkage parameter chosen by the researcher. The shrinkage statistic has a relatively smooth transition between the OLS and 2SLS test statistics. In Section 4, we evaluate the performance of the shrinkage procedure with different choices of τ. Additionally, denote $T_{l,n}(\theta_0)$ as $\pm\bar{T}_{l,n}(\theta_0)$ or $|\bar{T}_{l,n}(\theta_0)|$ for $l \in \{1,2\}$, depending on whether the test is a lower/upper one-sided or a symmetric two-sided test. The nominal size α test with a standard normal CV rejects $H_0 : \theta = \theta_0$ if $T_{l,n}(\theta_0) > c_\infty(1-\alpha)$, where $c_\infty(1-\alpha) = z_{1-\alpha}$ for the one-sided test and $z_{1-\alpha/2}$ for the symmetric two-sided test, respectively, and $z_{1-\alpha}$ is the $(1-\alpha)$-th quantile of a standard normal distribution. For the conciseness of the paper, in the following sections we will focus on the case with symmetric two-sided test but our results can be extended directly to the one-sided case.

2.2. Parameter space and asymptotic size

To characterize the asymptotic size of the two-stage and shrinkage tests, we define the parameter space Γ of the nuisance parameter vector γ following the seminal studies by Andrews and Guggenberger (2009, 2010a, 2010b), Guggenberger (2012), and Guggenberger and Kumar (2012). For the current testing problem, define the vector of nuisance parameters $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ by

$$\gamma_1 = a, \;\; \gamma_2 = (\gamma_{21}, \gamma_{22}, \gamma_{23}, \gamma_{24}, \gamma_{25}), \;\; \gamma_3 = F, \tag{2.9}$$

where a is defined in (2.2), $\gamma_{21} = \pi$, $\gamma_{22} = E_F e_i^2 Z_i Z_i'$, $\gamma_{23} = E_F e_i^2 v_i^2$, $\gamma_{24} = E_F Z_i Z_i'$, and $\gamma_{25} = E_F v_i^2$. Here, γ_1 measures the degree of endogeneity of X and is the key parameter in the current testing problem as it determines the point of discontinuity of the null limiting distributions of the two-stage and shrinkage test statistics. For the parameter space, let

$$\begin{aligned}\varGamma_1 = \mathbb{R}, \;\; \varGamma_2 = \Big\{&(\gamma_{21}, \gamma_{22}, \gamma_{23}, \gamma_{24}, \gamma_{25}) : \gamma_{21} = \pi \in \mathbb{R}^k, \gamma_{22} = E_F e_i^2 Z_i Z_i' \in \mathbb{R}^{k\times k},\\ &\gamma_{23} = E_F e_i^2 v_i^2 \in \mathbb{R}, \gamma_{24} = E_F Z_i Z_i' \in \mathbb{R}^{k\times k}, \gamma_{25} = E_F v_i^2 \in \mathbb{R},\\ &s.t. \;\; \|\gamma_{21}\| \geq \underline{\kappa}, \lambda_{min}(\gamma_{22}) \geq \underline{\kappa}, \gamma_{23} > 0, \lambda_{min}(\gamma_{24}) \geq \underline{\kappa}, \text{ and } \gamma_{25} > 0\Big\},\end{aligned} \tag{2.10}$$

for some $\underline{\kappa} > 0$ that does not depend on n. As $\|\gamma_{21}\| \geq \underline{\kappa} > 0$, Staiger and Stock (1997)'s weak IV asymptotics is ruled out of the scope of this paper.[7] In addition, $\varGamma_3(\gamma_1, \gamma_2)$ is defined as follows:

$$\begin{aligned}\varGamma_3(\gamma_1, \gamma_2) = \Big\{&F: \; E_F e_i v_i = E_F e_i Z_i = E_F v_i Z_i = 0, \; E_F e_i^2 v_i Z_i = E_F e_i v_i^2 Z_i = E_F e_i v_i Z_i Z_i' = 0,\\ &E_F v_i^2 Z_i Z_i' \in \mathbb{R}^{k\times k} \;\text{ with }\; \lambda_{min}(E_F v_i^2 Z_i Z_i') \geq M^{-1},\\ &\left\| E_F \left(||Z_i e_i||^{2+\xi}, ||Z_i v_i||^{2+\xi}, |v_i e_i|^{2+\xi}, ||Z_i Z_i'||^{2+\xi}, |X_i|^{2(2+\xi)} \right)' \right\| \leq M \Big\},\end{aligned} \tag{2.11}$$

for some constant $\xi > 0$ and $M < \infty$. We then define the whole nuisance parameter space $\varGamma$ of γ as

$$\varGamma = \{\gamma = (\gamma_1, \gamma_2, \gamma_3) : \gamma_1 \in \varGamma_1, \gamma_2 \in \varGamma_2, \gamma_3 \in \varGamma_3(\gamma_1, \gamma_2)\}, \tag{2.12}$$

where $\varGamma_j, \; j = 1, 2, 3$ are given in (2.10) and (2.11). This nuisance parameter space extends the one defined in Guggenberger (2010a) to allows for conditional heteroskedasticity and is similar to those defined in Guggenberger (2012) and Guggenberger and Kumar (2012), which also allow for heteroskedastic errors. In Section 3.3, we further extend analysis to the case with clustered data and show that our size-corrected wild bootstrap is also uniformly valid in that case. The condition that $E_F e_i^2 v_i Z_i = E_F e_i v_i^2 Z_i = E_F e_i v_i Z_i Z_i' = 0$ in (2.11) is similar to that imposed for $\varGamma_3(\gamma_1, \gamma_2)$ in Guggenberger (2010a) [see (A.2) in the Appendix of his paper for related discussions]. This condition simplifies the limiting distributions and its sufficient condition is, for example, independence between (v_i, e_i) and Z_i.

Now we define the asymptotic size. Let c_n denote a (possibly data-dependent) CV being used

[7] As pointed out by a referee, empirical researchers often use first-stage F statistics to detect instrument weakness. While pretesting instrument weakness is intuitively reasonable, it can also result in substantial size distortions [e.g., see Section 4.1 of Andrews, Stock and Sun (2019)]. Therefore, proposing valid testing and inference methods after such pretesting is an important issue. We leave this direction of investigation for further research in the future. However, the simulations in Section 4 and Appendix SA.3 suggest that our proposed tests perform well even with weak IV.

for the two-stage testing or shrinkage procedure. The finite sample null rejection probability (NRP) of the test statistic of interest evaluated at $\gamma \in \Gamma$ is given by $P_{\theta_0,\gamma}\left[T_{l,n}(\theta_0) > c_n\right]$ for $l \in \{1,2\}$, where $P_{\theta_0,\gamma}[E_n]$ denotes the probability of event E_n given γ. Then, the asymptotic NRP of the test evaluated at $\gamma \in \Gamma$ is given by $\underset{n\to\infty}{limsup} P_{\theta_0,\gamma}\left[T_{l,n}(\theta_0) > c_n\right]$, while the asymptotic size is given by

$$AsySz[c_n] = \underset{n\to\infty}{limsup} \underset{\gamma\in\Gamma}{sup} P_{\theta_0,\gamma}\left[T_{l,n}(\theta_0) > c_n\right]. \tag{2.13}$$

In general, asymptotic NRP evaluated at a given $\gamma \in \Gamma$ is not equal to the asymptotic size of the test. To control the asymptotic size, one needs to control the null limiting behaviour of $T_{l,n}(\theta_0)$ under drifting parameter sequences $\{\gamma_n : n \geq 1\}$ indexed by the sample size; e.g., see Andrews and Guggenberger (2009, 2010a, 2010b), Guggenberger (2012), and Guggenberger and Kumar (2012).

Following the arguments used in these papers, to derive $AsySz[c_n]$ we can study the asymptotic NRP along certain parameter sequences of the type $\{\gamma_{n,h}\}$ (defined below) for some $h \in \mathscr{H}$, as the highest asymptotic NRP is materialized under such sequence, where

$$\begin{aligned}\mathscr{H} \;=\; & \Big\{ h = (h_1, h_{21}', vec(h_{22})', h_{23}, vec(h_{24})', h_{25})' \in \mathbb{R}_\infty^{2k^2+k+3} : \exists\{\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3}) \in \Gamma : n \geq 1\} \\ & s.t.\ n^{1/2}\gamma_{n,1} \to h_1 \in \mathbb{R}_\infty,\ \ \gamma_{n,2} \to h_2 = (h_{21}, h_{22}, h_{23}, h_{24}, h_{25}),\ \|h_{21}\| \geq \underline{\kappa},\ \ \lambda_{min}(A) \geq \underline{\kappa} \\ & \text{for } A \in \{h_{22}, h_{24}\},\ h_{23} > 0,\ h_{25} > 0 \Big\} \equiv \mathscr{H}_1 \times \mathscr{H}_{21} \times \mathscr{H}_{22} \times \mathscr{H}_{23} \times \mathscr{H}_{24} \times \mathscr{H}_{25},\end{aligned} \tag{2.14}$$

for some $\underline{\kappa} > 0$ and $\mathbb{R}_\infty = \mathbb{R} \cup \{\pm\infty\}$. Then, for $h \in \mathscr{H}$, the relevant sequence of parameters $\{\gamma_{n,h}\} \subset \Gamma$ is defined following Guggenberger (2010a) as $\gamma_{n,h} = (\gamma_{n,h,1}, \gamma_{n,h,2}, \gamma_{n,h,3})$ where

$$\gamma_{n,h,1} = (E_{F_n}[v_i^2])^{-1} E_{F_n}[v_i u_i],\ \ \gamma_{n,h,2} = (\gamma_{n,h,21}, \gamma_{n,h,22}, \gamma_{n,h,23}, \gamma_{n,h,24}, \gamma_{n,h,25}), \tag{2.15}$$

with $\gamma_{n,h,21} = \pi_n$, $\gamma_{n,h,22} = E_{F_n} e_i^2 Z_i Z_i'$, $\gamma_{n,h,23} = E_{F_n} e_i^2 v_i^2$, $\gamma_{n,h,24} = E_{F_n} Z_i Z_i'$, $\gamma_{n,h,25} = E_{F_n} v_i^2$, s.t.

$$n^{1/2}\gamma_{n,h,1} \to h_1,\ \gamma_{n,h,2} \to h_2, \text{ and } \gamma_{n,h,3} = F_n \in \Gamma_3(\gamma_{n,h,1}, \gamma_{n,h,2}). \tag{2.16}$$

More specifically, under $\{\gamma_{n,h}\}$ satisfying (2.16) with $|h_1| = \infty$ (i.e., strong endogeneity), $H_n \to^P \infty$, and the two-stage and shrinkage test statistics are asymptotically equivalent to the 2SLS-based t-statistic. On the other hand, under $\{\gamma_{n,h}\}$ satisfying (2.16) with $|h_1| < \infty$ (i.e., local endogeneity), the following joint convergence results hold for $T_{2SLS}(\theta_0)$, $T_{OLS}(\theta_0)$, H_n, and the two-stage and shrinkage statistics $T_{l,n}(\theta_0)$ for $l \in \{1,2\}$:

$$\begin{pmatrix} T_{2sls}(\theta_0) \\ T_{ols}(\theta_0) \\ H_n \end{pmatrix} \to^d \begin{pmatrix} \eta_{1,h} \\ \eta_{2,h} \\ \eta_{3,h} \end{pmatrix} = \begin{pmatrix} (h_{21}' h_{22} h_{21})^{-1/2} h_{21}' \psi_{Ze} \\ (h_{21}' h_{22} h_{21} + h_{23})^{-1/2} (h_{21}' \psi_{Ze} + \psi_{ve} + h_{25} h_1) \\ \left(\frac{h_{21}' h_{22} h_{21}}{(h_{21}' h_{24} h_{21})^2} + h_{23} h_{25}^{-2}\right)^{-1} \left(-(h_{21}' h_{24} h_{21})^{-1} h_{21}' \psi_{Ze} + h_{25}^{-1} \psi_{ve} + h_1\right)^2 \end{pmatrix},$$

$$
\begin{aligned}
T_{1,n}(\theta_0) &\to^d \quad \tilde{T}_{1,h} = \left|\eta_{2,h}\mathbb{1}(\eta_{3,h} \leq \chi^2_{1,1-\beta}) + \eta_{1,h}\mathbb{1}(\eta_{3,h} > \chi^2_{1,1-\beta})\right|, \\
T_{2,n}(\theta_0) &\to^d \quad \tilde{T}_{2,h} = \left|\eta_{2,h}w(\eta_{3,h}) + \eta_{1,h}(1-w(\eta_{3,h}))\right|,
\end{aligned} \tag{2.17}
$$

where $\eta_{1,h} \sim N(0,1)$, $\eta_{2,h} \sim N\Big((h'_{21}h_{22}h_{21} + h_{23})^{-1/2}h_{25}h_1, 1\Big)$, $\eta_{3,h} \sim \chi^2_1\Big(\big(\frac{h'_{21}h_{22}h_{21}}{(h'_{21}h_{24}h_{21})^2} + h_{23}h_{25}^{-2}\big)^{-1}h_1^2\Big)$, $w(\eta_{3,h}) = \tau/\eta_{3,h}$ if $\eta_{3,h} \geq \tau$, and $w(\eta_{3,h}) = 1$ if $\eta_{3,h} < \tau$.

We notice that under conditional homoskedasticity, the formula of the limiting distribution of $T_{1,n}(\theta_0)$ in (2.17) will be simplified and equal to that derived in Guggenberger (2010a) for the two-stage test with Hausman pretest, whose asymptotic size is equal to 1 with the standard normal CV $c_\infty(1-\alpha)$. Therefore, in the general heteroskedastic case, the asymptotic size of the $T_{1,n}(\theta_0)$-based two-stage test with $c_\infty(1-\alpha)$ is also equal to 1. Similar results can be shown for the $T_{2,n}(\theta_0)$-based shrinkage test. In the next section, we will study the asymptotic behaviours of the two-stage and shrinkage tests under alternative data-dependent CVs.

3. Main Results

3.1. Standard wild bootstrap

In this section, we study the asymptotic behaviour of the standard wild bootstrap for the two-stage testing and shrinkage procedures. As mentioned in Section 2.1, we focus on the symmetric two-sided tests to simplify exposition, but our results can be extended to one-sided tests.

Wild Bootstrap Algorithm:

1. Compute the (null-restricted) residuals from the first-stage and structural equations: $\hat{v} = X - Z\hat{\pi}$, $\hat{u}(\theta_0) = y - X\theta_0$, where $\hat{\pi} = (Z'Z)^{-1}Z'X$ denotes the least squares estimator of π.

2. Generate the bootstrap pseudo-data following $X^* = Z\hat{\pi} + v^*$, $y^* = X^*\theta_0 + u^*$, where there are two options to generate the bootstrap disturbances:

   (a) v^* and u^* are generated independently from each other. Specifically, in the current case with heteroskedastic data, we set for each observation i: $v_i^* = \hat{v}_i\omega^*_{1i}$, and $u_i^* = \hat{u}_i(\theta_0)\omega^*_{2i}$, where ω^*_{1i} and ω^*_{2i} are two random variables with mean 0 and variance 1, i.e., $E^*\left[\omega^*_{1i}\right] = E^*\left[\omega^*_{2i}\right] = 0$ and $Var^*\left[\omega^*_{1i}\right] = Var^*\left[\omega^*_{2i}\right] = 1$, and they are independent from the data and independent from each other.

   (b) v^* and u^* are drawn dependently from each other. We set for each observation i: $v_i^* = \hat{v}_i\omega^*_{1i}$, and $u_i^* = \hat{u}_i(\theta_0)\omega^*_{1i}$.

   Following Young (2020), we refer to (a) as *independent transformation* of disturbances and (b) as *dependent transformation* of disturbances.[8]

[8] For the purpose of better size control, it is often recommended that for bootstrap exogeneity tests, (u^*, v^*) should

3. Compute the bootstrap analogues of the two-stage and shrinkage test statistics:

$$\begin{aligned} T^*_{1,n}(\theta_0) &= \left| T^*_{ols}(\theta_0)\mathbb{1}(H^*_n \le \chi^2_{1,1-\beta}) + T^*_{2sls}(\theta_0)\mathbb{1}(H^*_n > \chi^2_{1,1-\beta}) \right|, \\ T^*_{2,n}(\theta_0) &= \left| T^*_{ols}(\theta_0) w(H^*_n) + T^*_{2sls}(\theta_0)(1 - w(H^*_n)) \right|, \end{aligned} \tag{3.1}$$

where $w(H^*_n) = \begin{cases} \tau / H^*_n & \text{if } H^*_n \ge \tau \\ 1 & \text{if } H^*_n < \tau \end{cases}$, $T^*_{ols}(\theta_0)$, $T^*_{2sls}(\theta_0)$ and H^*_n are the bootstrap analogues of $T_{ols}(\theta_0)$, $T_{2sls}(\theta_0)$ and H_n, respectively, which are obtained from the bootstrap samples generated in Step 2.

4. For $l \in \{1,2\}$, repeat Steps 2-3 B times and obtain $\{T^{*(b)}_{l,n}(\theta_0), b = 1, ..., B\}$. The bootstrap test with the test statistic $T_l(\theta_0)$ rejects H_0 if the corresponding bootstrap p-value $\frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\left[T^{*(b)}_{l,n}(\theta_0) > T_{l,n}(\theta_0)\right]$ is less than the nominal level α.

Following the standard arguments for bootstrap validity, to check whether (conditional on the data) the bootstrap is able to consistently estimate the distribution of the two-stage or shrinkage test statistic, one needs to check whether under H_0 and both cases of strong endogeneity ($|h_1| = \infty$) and local endogeneity ($|h_1| < \infty$), $\sup_{x \in R} \left| P^*\left(T^*_{l,n}(\theta_0) \le x\right) - P\left(T_{l,n}(\theta_0) \le x\right) \right| \to^P 0$, for $l \in \{1,2\}$. However, we notice below that neither bootstrap procedure is able to consistently estimate the distribution of interest under local endogeneity.

More specifically, it holds for the bootstrap statistics with dependent or independent transformation (for the dependent transformation, we further require $E^*\left[\omega^{*3}_{1i}\right] = 0$ and $E^*\left[\omega^{*4}_{1i}\right] = 1$; see Lemma **A.4** in the Appendix for details) that

$$n^{-1/2} \begin{pmatrix} Z' u^* \\ \left(u^{*'} v^* - E^*[u^{*'} v^*]\right) \end{pmatrix} \to^{d^*} \begin{pmatrix} \psi^*_{Ze} \\ \psi^*_{ve} \end{pmatrix}, \tag{3.2}$$

in probability P, where the bootstrap (conditional) weak limit $(\psi^{*'}_{Ze}, \psi^*_{ve})'$ is the same as $(\psi'_{Ze}, \psi_{ve})'$, i.e., the weak limit of $n^{-1/2}((Z'u)', (u'v - E_F[u'v]))'$. Therefore, the bootstrap procedures do replicate well the randomness in the original sample.

On the other hand, under local endogeneity the standard wild bootstraps are not able to mimic well the key localization parameter h_1, thus resulting in the discrepancy between the original and bootstrap samples (see Theorem **A.5** in the Appendix for details). In particular, let h^b_1 denote the localization parameter of endogeneity in the bootstrap world, then $h^b_1 = 0$ for the bootstrap with independent transformation, while $h^b_1 = h_1 + h^{-1}_{25}\psi_{ve}$ for the one with dependent transformation, where $\psi_{ve} \sim N(0, h_{23})$. That is, while the bootstrap with dependent transformation is able to mimic the situation of local endogeneity in the original sample (h^b_1 is finite with probability approaching

be generated using the independent transformation scheme, so that the bootstrap samples are obtained under the null hypothesis of exogeneity. However, as we will see below, this is not necessarily the case for the bootstrap two-stage or shrinkage test statistic.

one when h_1 is finite), the approximation is imprecise and results in an extra error term $h_{25}^{-1}\psi_{ve}$, whose value depends on the actual realization of the sample.

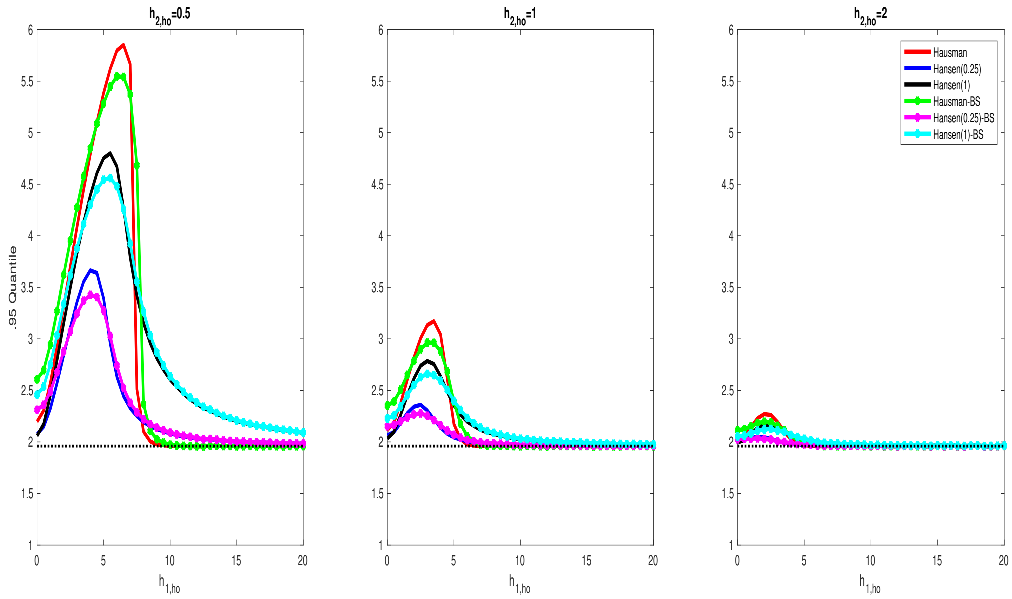
However, even if the bootstrap is inconsistent conditional on the data, it may still be valid in the unconditional sense; e.g., see Cavaliere and Georgiev (2020) and the references therein. More precisely, the bootstrap might still be able to provide a valid test in the current context if its asymptotic NRP does not exceed the nominal level α under any parameter sequence $\{\gamma_{n,h}\}$ in (2.16). To further shed light on the behaviour of the bootstrap statistics with dependent transformation, we apply the results in (2.17) and Theorem **A.5** to plot the quantiles of the null limiting distributions of the original and bootstrap test statistics for the case of conditional homoskedasticity studied in Guggenberger (2010a). The limiting distributions of both two-stage and shrinkage test statistics are substantially simplified in this case and only depend on two scalar parameters, say, $h_{1,ho}$ and $h_{2,ho}$. $h_{1,ho}$ captures the degree of local endogeneity and $h_{2,ho}$ captures the IV strength, respectively.[9] Figure 1 reports the 95% quantiles of $\tilde{T}_{l,h}$ and its bootstrap counterpart $\tilde{T}^*_{l,h}$ for $l \in \{1,2\}$, as a function of $h_{1,ho}$ with $h_{2,ho} \in \{.5,1,2\}$, $\beta = .05$ for the two-stage test statistic, and $\tau \in \{0.25,1\}$ for the shrinkage test statistic. The results are based on 1,000,000 simulation replications.

We highlight some findings below. First, we observe that the quantiles of $\tilde{T}^*_{l,h}$ for the dependent bootstrap turn out to be rather close to those of $\tilde{T}_{l,h}$ across various values of $h_{1,ho}$ and $h_{2,ho}$. However, the figure suggests that this bootstrap procedure may have over-rejection when the quantiles of $\tilde{T}_{l,h}$ are relatively high (e.g., when $h_{2,ho} = .5$ and $h_{1,ho}$ is between 5 and 6). In addition, we note that the quantiles of $\tilde{T}^*_{l,h}$ for the dependent bootstrap converge in each sub-figure to the standard normal CV when the value of $h_{1,ho}$ increases: when $|h_{1,ho}|$ is large, the Hausman pretest rejects with high probability and the weight $w(H_n)$ shrinks toward zero, so that both two-stage and shrinkage tests become the 2SLS-based t-test, and the dependent bootstrap does mimic well such behaviour. Furthermore, the quantiles corresponding to the shrinkage test statistics and their bootstrap analogues (i.e., $\tilde{T}_{2,h}$ and $\tilde{T}^*_{2,h}$) have smoother shapes than their two-stage counterparts (i.e., $\tilde{T}_{1,h}$ and $\tilde{T}^*_{1,h}$), and this may be due to the fact that the two-stage test statistic uses an abrupt transition between the OLS and 2SLS-based t-statistics.

Also based on (2.17) and Theorem **A.5**, we report in Table 1 below the asymptotic null rejection probabilities of the (symmetric) two-stage and shrinkage tests under the standard normal CV, the independent bootstrap CV, and the dependent bootstrap CV for the homoskedastic case, with $\alpha = .05$ for $\underline{\kappa}_{ho} \in \{.001,.1,.5,1,2,10\}$, where $\underline{\kappa}_{ho}$ denotes the lower bound of $h_{2,ho}$, $\beta \in \{.05,.1,.2,.5\}$ for the two-stage test, and $\tau \in \{.1,.25,.5,1\}$ for the shrinkage test. The values of α, β, and $\underline{\kappa}_{ho}$ are set to be the same as those in Table 1 of Guggenberger (2010a). First, we note that both standard normal CVs and independent bootstrap CVs have asymptotic rejection probabilities much larger than .05; e.g., when $\underline{\kappa}_{ho} = .001$, the probabilities are $100\%, 95.0\%, 85.1\%, 55.1\%$ and $97.6\%, 92.7\%, 82.9\%, 53.5\%$, respectively. Second, although the dependent bootstrap CVs result in asymptotic size distortions much smaller than the standard normal CVs and independent

[9] See (9) in Section 2.3 and (12) in Section 2.4 of Guggenberger (2010a) for detailed definition; we note that in Guggenberger (2010a), the parameters $h_{1,ho}$ and $h_{2,ho}$ are denoted as h_1 and h_2, respectively.

Figure 1. 95% quantiles of $\tilde{T}_{l,h}$ and $\tilde{T}^*_{l,h}$ under homoskedasticity



Note: “Hausman", “Hansen(0.25)", and “Hansen(1)" denote the Hausman-type two-stage test statistic, and Hansen (2017)'s shrinkage test statistic with $\tau = 0.25$ and $\tau = 1$, respectively. ‘Hausman-BS", “Hansen(0.25)-BS", and “Hansen(1)-BS" denote their wild bootstrap analogues with the dependent transformation. The results are based on 1,000,000 simulation replications.

bootstrap CVs, in general they are also unable to achieve uniform size control. As the heteroskedastic case nests the homoskedastic case, these results could be regarded as the lower bounds of the asymptotic size distortions under heteroskedasticity. Also notice that the size distortions decrease with the value of β while increase with the value of τ. Furthermore, as pointed out by a referee, the independent and dependent bootstrap CVs are asymptotically equivalent to the asymptotic CVs drawn from the $(1-\alpha)$-th quantile of (2.17) after setting $h_1 = 0$ and $h_1 = \hat{h}_{n,1}$, respectively. This implies that there also exist size distortions when these asymptotic CVs are used.

3.2. Size-corrected wild bootstrap

As the standard wild bootstrap procedures are not able to provide uniform size control, in this section we propose Bonferroni-based size-correction methods for the two-stage testing and shrinkage procedures, following the seminal study by McCloskey (2017). As explained in McCloskey (2017), the idea behind such size-correction is to construct CVs that use the data to determine how far the key nuisance parameter (i.e., the endogeneity parameter in the current testing problem) is from the point that causes the discontinuity in the limiting distributions of the test statistics. Although the key nuisance parameter cannot be consistently estimated under the drifting sequences in (2.16), it is still possible to construct an asymptotically valid confidence set for it and then construct adaptive CVs that control the asymptotic size.

Table 1. Asymptotic rejection probabilities (in %) for $\alpha = .05$ under homoskedasticity.

T_1	Std Normal CV				BS-independent				BS-dependent			
$\underline{\kappa}_{ho} \setminus \beta$	.05	.1	.2	.5	.05	.1	.2	.5	.05	.1	.2	.5
.001	100	95.0	85.1	55.1	97.6	92.7	82.9	53.5	6.6	6.6	6.6	6.5
.1	95.5	90.4	80.2	50.8	93.9	88.3	77.5	48.9	6.6	6.6	6.6	6.5
.5	60.1	50.3	39.0	22.3	55.8	45.4	34.3	19.4	6.6	6.6	6.6	6.5
1	27.6	21.8	16.1	9.7	24.7	18.4	12.9	7.7	6.6	6.6	6.5	5.9
2	10.8	9.1	7.7	6.1	10.0	8.2	6.6	5.1	6.0	5.9	5.6	5.2
10	5.3	5.2	5.1	5.1	5.3	5.1	5.1	5.1	5.1	5.1	5.1	5.1
T_2	Std Normal CV				BS-independent				BS-dependent			
$\underline{\kappa}_{ho} \setminus \tau$	1	.5	.25	.1	1	.5	.25	.1	1	.5	.25	.1
.001	96.2	90.7	83.8	63.3	96.1	90.4	83.4	60.7	6.3	6.3	6.3	6.2
.1	83.1	71.5	61.2	48.3	82.6	70.8	60.2	47.1	6.3	6.3	6.3	6.2
.5	38.3	28.5	22.2	16.8	36.3	26.1	20.0	15.2	6.3	6.3	6.3	6.2
1	15.9	12.0	9.8	8.1	14.5	10.3	8.2	6.9	6.1	6.0	5.8	5.6
2	7.7	6.6	6.1	5.7	7.1	6.0	5.5	5.2	5.5	5.4	5.2	5.2
10	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.1

Note: "T_1" and "T_2" denote the Hausman-type two-stage test statistic and Hansen (2017)'s shrinkage test statistic, respectively. The results are based on 1,000,000 simulation replications.

First, we will construct a size-corrected wild bootstrap CV by using the wild bootstrap CVs with the independent transformation and Bonferroni bounds. Note that although the localization parameter h_1 cannot be consistently estimated, we may still construct an asymptotically valid confidence set for h_1 by defining $\hat{h}_{n,1} = n^{1/2}\hat{a}$, where $\hat{a} = (\tilde{v}'\tilde{v})^{-1}\tilde{v}'y$. A confidence set of h_1 can be constructed by using the fact that under the drifting parameter sequences,

$$\hat{h}_{n,1} \to^d \tilde{h}_1 \sim N\Big(h_1, \left(h_{21}'h_{24}h_{21}\right)^{-2}h_{21}'h_{22}h_{21} + h_{25}^{-2}h_{23}\Big). \tag{3.3}$$

Then, uniformly valid size-corrected bootstrap CVs for testing $H_0 : \theta = \theta_0$ under the two-stage or shrinkage procedure can be constructed by using Bonferroni bounds: we may construct a $1-(\alpha-\delta)$ level first-stage confidence set for h_1, and then take the maximal $(1-\delta)$-th quantile of appropriately generated bootstrap statistics over the first-stage confidence set. Specifically, let $\hat{h}_{n,2} = \Big(\hat{h}_{n,21}', vec(\hat{h}_{n,22})', \hat{h}_{n,23}, vec(\hat{h}_{n,24})', \hat{h}_{n,25}\Big)'$ be a consistent estimator of $h_2 = (h_{21}', vec(h_{22})', h_{23}, vec(h_{24})', h_{25})'$, and define the $1-(\alpha-\delta)$ level confidence set of h_1 for some $0 < \delta \leq \alpha < 1$ as $CI_{\alpha-\delta}(\hat{h}_{n,1}) = \Big[\hat{h}_{n,1} - z_{1-(\alpha-\delta)/2}\cdot\left(n\hat{V}_a\right)^{1/2},\ \hat{h}_{n,1} + z_{1-(\alpha-\delta)/2}\cdot\left(n\hat{V}_a\right)^{1/2}\Big]$, where $\hat{V}_a$ is defined in (2.4). The bootstrap simple Bonferroni critical value (SBCV) is defined as

$$c_l^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2}) \quad = \sup_{h_1 \in CI_{\alpha-\delta}(\hat{h}_{n,1})} c^*_{l,(h_1,\hat{h}_{n,2})}(1-\delta), \tag{3.4}$$

for $l \in \{1,2\}$, where $c^*_{l,(h_1,\hat{h}_2)}(1-\delta)$ is the $(1-\delta)$-th quantile of the distribution of $T^*_{l,n,(h_1,\hat{h}_{n,2})}(\theta_0)$, i.e., the distribution of the bootstrap analogue of $T_{l,n}(\theta_0)$ generated under the value of localization parameter equal to h_1.

As we have seen in the previous section, the standard wild bootstrap procedures cannot mimic well the localization parameter h_1, no matter with independent or dependent transformation. Therefore, attention has to be taken when considering the bootstrap DGP. In particular, we propose to generate the bootstrap statistics under the localization parameter h_1 as follows:

$$\begin{aligned}
T^*_{1,n,(h_1,\hat{h}_{n,2})}(\theta_0) &= \left| T^*_{ols,(h_1,\hat{h}_{n,2})}(\theta_0)\mathbb{1}\left(H^*_{n,(h_1,\hat{h}_{n,2})} \leq \chi^2_{1,1-\beta}\right) + T^*_{2sls}(\theta_0)\mathbb{1}\left(H^*_{n,(h_1,\hat{h}_{n,2})} > \chi^2_{1,1-\beta}\right)\right|, \\
T^*_{2,n,(h_1,\hat{h}_{n,2})}(\theta_0) &= \left| T^*_{ols,(h_1,\hat{h}_{n,2})}(\theta_0)w\left(H^*_{n,(h_1,\hat{h}_{n,2})}\right) + T^*_{2sls}(\theta_0)\left(1-w\left(H^*_{n,(h_1,\hat{h}_{n,2})}\right)\right)\right|,
\end{aligned} \tag{3.5}$$

where $T^*_{ols,(h_1,\hat{h}_{n,2})}(\theta_0)$ and $H^*_{n,(h_1,\hat{h}_{n,2})}$ are the bootstrap analogues of $T_{ols}(\theta_0)$ and H_n, respectively, evaluated at the value of localization parameter equal to h_1. More precisely, to obtain these bootstrap analogues, we first generate the bootstrap counterparts of the OLS and regression endogeneity parameter estimators under h_1:

$$\hat{\theta}^*_{ols,(h_1,\hat{h}_{n,2})} = \hat{\theta}^*_{ols} + \left(\hat{h}'_{n,21}\hat{h}_{n,24}\hat{h}_{n,21} + \hat{h}_{n,25}\right)^{-1}\hat{h}_{n,25}\left(n^{-1/2}h_1\right), \quad \hat{a}^*_{(h_1,\hat{h}_{n,2})} = \hat{a}^* + n^{-1/2}h_1, \tag{3.6}$$

where $\hat{\theta}^*_{ols}$ and $\hat{a}^*$ are generated by the standard wild bootstrap procedure in Section 3.1 with *independent transformation* of disturbances, so that $\hat{\theta}^*_{ols}$ and $\hat{a}^*$ have localization parameter equal to zero in the bootstrap world. By doing so, $\sqrt{n}\left(\hat{\theta}^*_{ols,(h_1,\hat{h}_{n,2})} - \theta_0\right)$ and $\sqrt{n}\hat{a}^*_{(h_1,\hat{h}_{n,2})}$ have appropriate null limiting distribution conditional on the data. Then, we obtain

$$T^*_{ols,(h_1,\hat{h}_{n,2})}(\theta_0) = (\hat{\theta}^*_{ols,(h_1,\hat{h}_{n,2})} - \theta_0)/\hat{V}^{*1/2}_{ols}, \quad H^*_{n,(h_1,\hat{h}_{n,2})} = \hat{a}^{*2}_{(h_1,\hat{h}_{n,2})}/\hat{V}^*_a, \tag{3.7}$$

and we can show that the following (conditional) convergence in distribution holds:

$$\begin{pmatrix} T^*_{ols,(h_1,\hat{h}_{n,2})}(\theta_0) \\ H^*_{n,(h_1,\hat{h}_{n,2})} \end{pmatrix} \to^{d^*} \begin{pmatrix} (h'_{21}h_{22}h_{21} + h_{23})^{-1/2}(h'_{21}\psi^*_{Ze} + \psi^*_{ve} + h_{25}h_1) \\ \left(\frac{h'_{21}h_{22}h_{21}}{(h'_{21}h_{24}h_{21})^2} + h_{23}h_{25}^{-2}\right)^{-1}\left(-(h'_{21}h_{24}h_{21})^{-1}h'_{21}\psi^*_{Ze} + h_{25}^{-1}\psi^*_{ve} + h_1\right)^2 \end{pmatrix},$$

in probability P, where ψ^*_{Ze} and ψ^*_{ve} are the bootstrap analogues of ψ_{Ze} and ψ_{ve}, respectively. This implies that $T^*_{1,n,(h_1,\hat{h}_{n,2})}(\theta_0)$ and $T^*_{2,n,(h_1,\hat{h}_{n,2})}(\theta_0)$, the resulting bootstrap counterparts of the two-stage and shrinkage test statistics, have the desired null limiting distributions evaluated at the value of localization parameter equal to h_1 (different from the results obtained in Theorem **A.5**).

As seen from (3.4), the bootstrap SBCV equals the maximal quantile $c^*_{l,(h_1,\hat{h}_{n,2})}(1-\delta)$ over the values of the localization parameter h_1 in the set $CI_{\alpha-\delta}(\hat{h}_{n,1})$. We can now state the following asymptotic size result for $c_l^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})$, where $l \in \{1,2\}$.

Theorem 3.1 *Suppose that H_0 holds, then we have for any $0 < \delta \leq \alpha < 1$ and for $l \in \{1,2\}$, $AsySz\left[c_l^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})\right] \leq \alpha$.*

Theorem **3.1** states that tests based on $c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})$ control the asymptotic size. In practice, $c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})$ can be obtained by using the following algorithm.

Wild Bootstrap Algorithm for $\boldsymbol{c_l^{\text{B-S}}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})}$:

1. Generate the bootstrap statistics $\left\{\hat{\theta}_{ols}^{*(b)},\hat{\theta}_{2sls}^{*(b)},\hat{a}^{*(b)},\hat{V}_{ols}^{*(b)},\hat{V}_{2sls}^{*(b)},\hat{V}_{a}^{*(b)}\right\},b=1,...,B$, using the standard wild bootstrap procedure with independent transformation of disturbances.

2. Choose α, δ, and compute $CI_{\alpha-\delta}(\hat{h}_{n,1})$. Create a fine grid for $CI_{\alpha-\delta}(\hat{h}_{n,1})$ and call it $\mathscr{C}_{\alpha-\delta}^{grid}$.

3. For $l\in\{1,2\}$ and for $h_1\in\mathscr{C}_{\alpha-\delta}^{grid}$, generate $T_{l,n,(h_1,\hat{h}_{n,2})}^{*(b)}(\theta_0)$, $b=1,...,B$, using the bootstrap statistics generated in Step 1. The same set of $\left\{\hat{\theta}_{ols}^{*(b)},\hat{\theta}_{2sls}^{*(b)},\hat{a}^{*(b)},\hat{V}_{ols}^{*(b)},\hat{V}_{2sls}^{*(b)},\hat{V}_{a}^{*(b)}\right\},b=1,...,B$, can be used repeatedly for each h_1.

4. Compute $c_{l,(h_1,\hat{h}_{n,2})}^{*}(1-\delta)$, the $(1-\delta)$-th quantile of the distribution of $T_{l,n,(h_1,\hat{h}_2)}^{*}(\theta_0)$ from these B draws of bootstrap samples.

5. Find $c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})=\sup_{h_1\in\mathscr{C}_{\alpha-\delta}^{grid}}c_{l,(h_1,\hat{h}_{n,2})}^{*}(1-\delta)$.

Note that as shown in Theorem **3.1**, although controlling the asymptotic size, the bootstrap SBCV may yield a conservative test whose asymptotic size does not reach its nominal level. For further refinement on the Bonferroni bounds, we propose a size-adjustment method to adjust the bootstrap SBCV so that the resulting test is not conservative with asymptotic size exactly equal to α. Specifically, for $l\in\{1,2\}$, the size-adjustment factor for the bootstrap SBCV is defined as:

$$\hat{\eta}_{l,n}=\inf\left\{\eta:\sup_{h_1\in\mathscr{H}_1}P^*\left[T_{l,n,(h_1,\hat{h}_{n,2})}^{*}(\theta_0)>c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1}^{*}(h_1),\hat{h}_{n,2})+\eta\right]\leq\alpha\right\},\qquad(3.8)$$

where $\hat{h}_{n,1}^{*}(h_1)$ denotes the bootstrap analogue of $\hat{h}_{n,1}$ with localization parameter equal to h_1 and is generated by the same bootstrap samples as those for $T_{n,(h_1,\hat{h}_{n,2})}^{*}(\theta_0)$. More precisely, we define

$$\hat{h}_{n,1}^{*}(h_1)=\hat{h}_{n,1}^{*}+h_1,\qquad(3.9)$$

where $\hat{h}_{n,1}^{*}=n^{1/2}\hat{a}^{*}=\left(\hat{v}^{*\prime}M_{X^*}\hat{v}^{*}\right)^{-1}\hat{v}^{*\prime}M_{X^*}y^{*}$, $\hat{v}^{*}=M_ZX^{*}$, is generated by the standard wild bootstrap procedure with independent transformation so that the localization parameter equals zero in the bootstrap world. Notice that we have the following convergence in distribution (jointly with the other bootstrap statistics), $\hat{h}_{n,1}^{*}(h_1)\rightarrow^{d^*}N\Big(h_1,\left(h_{21}'h_{24}h_{21}\right)^{-2}h_{21}'h_{22}h_{21}+h_{25}^{-2}h_{23}\Big)$, in probability P, i.e., the same limiting distribution as that of $\hat{h}_{n,1}$ in (3.3).

The goal of the size-adjustment method is to decrease the bootstrap SBCV as much as possible by using the factor η while not violating the inequality in (3.8), so that the asymptotic size of the resulting tests can be controlled. Then, the bootstrap size-adjusted CV (BACV) can be defined as

$$c_l^{B-A}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2}) = c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})+\hat{\eta}_{l,n} \text{ for } l\in\{1,2\}, \tag{3.10}$$

and one can expect that relatively small $\hat{\eta}_{l,n}$ results in relatively less conservative (and more powerful) test. Under a proper algorithm for the size-adjustment method, and given some fixed $\alpha\in(0,1)$ and $\delta\in(0,\alpha]$, the size-adjustment factor $\hat{\eta}_{l,n}(\cdot)$ is continuous as a function of $\hat{h}^*_{n,1}(h_1)$. Furthermore, we notice that the bootstrap-based size-adjustment method in (3.10) is in the same spirit as the adjusted Bonferroni CV proposed in McCloskey (2017, Section 3.2), which is based on adjusting the quantile level of the underlying localized quantile in the simple Bonferroni CV.

Below we state the theorem on the uniform size control of the wild bootstrap CVs based on the size-adjustment method, and we assume a continuity condition on the NRP function, following similar continuity assumptions in Andrews and Cheng [2012, p.2195, Assumption Rob2(i)] and Han and McCloskey [2019, p.1052, Assumption DF2(ii)]. Define $c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)=\sup_{h_1\in CI_{\alpha-\delta}(\tilde{h}_1)}c_{l,h}(1-\delta)$, where $c_{l,h}(1-\delta)$ is the $(1-\delta)$-th quantile of $\tilde{T}_{l,h}$ and $\tilde{T}_{l,h}$ is the weak limit of $T_{l,n}(\theta_0)$ under the sequence $\{\gamma_{n,h}\}\subset\Gamma$ satisfying (2.16) for $l\in\{1,2\}$.

Assumption 3.2 *$P\left[\tilde{T}_{l,h}=c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)+\eta\right]=0$, $\forall h_1\in H_1$ and $\eta\in[-c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2),0]$, where $l\in\{1,2\}$.*

Theorem 3.3 *Suppose that H_0 and Assumption* **3.2** *hold, then we have for any $0<\delta\leq\alpha<1$ and for $l\in\{1,2\}$: $AsySz\left[c_l^{B-A}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})\right]=\alpha$.*

Furthermore, let $CS_{l,n}(1-\alpha)$ denote the nominal level $1-\alpha$ confidence set for θ constructed by collecting all the values of θ that cannot be rejected by the corresponding size-adjusted two-stage or shrinkage test at nominal level α.

Corollary 3.4 *Suppose that Assumption* **3.2** *holds, then we have for any $0<\delta\leq\alpha<1$ and for $l\in\{1,2\}$: $\liminf_{n\to\infty}\inf_{\gamma\in\Gamma}P_{\theta,\gamma}\left[\theta\in CS_{l,n}(1-\alpha)\right]=1-\alpha$.*

Theorem **3.3** shows that $c_l^{B-A}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})$ yield two-stage and shrinkage tests with the correct asymptotic size, and Corollary **3.4** states that the confidence sets constructed from inverting these tests have correct asymptotic coverage probability.[10] To implement such size-adjusted tests in practice, we must compute $c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})$ and $\hat{\eta}_{l,n}$. These values can be computed sequentially starting with $c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})$. Then the size-adjustment factor $\hat{\eta}_{l,n}$ can be computed by evaluating (3.8) over a fine grid of $\mathscr{H}_1$ as follows.

Wild Bootstrap Algorithm for $c_l^{\text{B-A}}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})$:

1. Generate the bootstrap statistics $\left\{\hat{\theta}^{*(b)}_{ols},\hat{\theta}^{*(b)}_{2sls},\hat{a}^{*(b)},\hat{V}^{*(b)}_{ols},\hat{V}^{*(b)}_{2sls},\hat{V}^{*(b)}_{a},\hat{h}^{*(b)}_{n,1}\right\}, b=1,...,B$, using the standard wild bootstrap procedure with independent transformation.

[10] Also see, e.g., Section 6 in Davidson and MacKinnon (2010) and Section 3.5 in Roodman, Nielsen, MacKinnon and Webb (2019) for detailed guidance on constructing confidence set from inverting a wild bootstrap test.

2. For $l \in \{1,2\}$, let $c_l^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})$ be the obtained SBCV.

3. Create a fine grid of the set $\mathscr{H}_1$ in (3.8) and call it $\mathscr{H}_1^{grid}$. For $l \in \{1,2\}$ and for each $h_1 \in \mathscr{H}_1^{grid}$, obtain $T^{*(b)}_{l,n,(h_1,\hat{h}_{n,2})}(\theta_0)$ and $c_l^{B-S}(\alpha, \alpha-\delta, \hat{h}^{*(b)}_{n,1}(h_1), \hat{h}_{n,2})$, $b = 1,...,B$, using the bootstrap statistics generated in Step 1. Note that the same set of $\left\{\hat{\theta}^{*(b)}_{ols}, \hat{\theta}^{*(b)}_{2sls}, \hat{a}^{*(b)}, \hat{V}^{*(b)}_{ols}, \hat{V}^{*(b)}_{2sls}, \hat{V}^{*(b)}_{a}, \hat{h}^{*(b)}_{n,1}\right\}, b = 1,...,B$, can be used for each h_1.

4. Create a fine grid of $[-c_l^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2}), 0]$ and call it $\mathbb{S}^{grid}$.

5. Find all $\eta \in \mathbb{S}^{grid}$ s.t. $\sup\limits_{h_1 \in \mathscr{H}_1^{grid}} \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\left[T^{*(b)}_{l,n,(h_1,\hat{h}_{n,2})}(\theta_0) > c_l^{B-S}(\alpha, \alpha-\delta, \hat{h}^{*(b)}_{n,1}(h_1), \hat{h}_{n,2}) + \eta\right] \leq \alpha$, and set $\hat{\eta}_{l,n}$ equal to the smallest η.

6. The BACV is given by $c_l^{B-A}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2}) = c_l^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2}) + \hat{\eta}_{l,n}$.

Remarks

1. We emphasize that $\hat{h}^*_{n,1}(h_1)$ needs to be generated simultaneously with $T^*_{l,n,(h_1,\hat{h}_{n,2})}(\theta_0)$ using the same bootstrap samples, so that the dependence structure between the statistics $T_{l,n}(\theta_0)$ and $\hat{h}_{n,1}$ is well mimicked by the bootstrap statistics. This is important for the size-adjustment procedure to correct the conservativeness of the Bonferroni bound. Similarly, for the implementation of the size-adjustment, one cannot replace $c_l^{B-S}(\alpha, \alpha-\delta, \hat{h}^*_{n,1}(h_1), \hat{h}_{n,2})$ in (3.8) with $c_l^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})$, as it also breaks down the dependence structure.

2. We note that the computational cost of the proposed size-corrected wild bootstrap procedures is not very high. In particular, the same bootstrap samples can be used in the algorithms for constructing $c_l^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})$ and $c_l^{B-A}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})$: there is no need to generate a new set of bootstrap samples to implement the size-correction method in (3.8). Moreover, the same set of bootstrap statistics $\left\{\hat{\theta}^{*(b)}_{ols}, \hat{\theta}^{*(b)}_{2sls}, \hat{a}^{*(b)}, \hat{V}^{*(b)}_{ols}, \hat{V}^{*(b)}_{2sls}, \hat{V}^{*(b)}_{a}\right\}, b = 1,...,B$, can be used repeatedly for each value of localization parameter $h_1 \in \mathscr{C}^{grid}_{\alpha-\delta}$ when constructing the localized quantiles $c^*_{l,(h_1,\hat{h}_{n,2})}(1-\delta)$ in Step 3 of the algorithm for $c_l^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})$. Similarly, the same set of bootstrap statistics can be used repeatedly for each $h_1 \in \mathscr{H}_1^{grid}$ when evaluating the size-adjustment factor in Step 3 of the algorithm for $c_l^{B-A}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})$.

3.3. Extension to Clustered Data

Many applications in economics involve error terms that are correlated within clusters (e.g., see Cameron and Miller (2015) and the references therein), and various studies in the literature on cluster-robust inference recommend to use wild cluster bootstrap as a way to obtain more accurate inference, including Cameron et al. (2008), MacKinnon and Webb (2017), Djogbenou et al. (2019), and MacKinnon, Nielsen and Webb (2020), among others. However, by using similar arguments as those for the IV model with heteroskedastic errors, we can show that the standard wild cluster bootstrap is invalid in the current context for the two-stage testing and shrinkage procedures. In this section, we extend the size-corrected wild bootstrap procedure proposed in Section 3.2 to the case with clustered samples.

To proceed, consider the following linear IV model with clustered data:

$$y_g = X_g\theta + u_g, \quad X_g = Z_g\pi + v_g, \tag{3.11}$$

where $y_g = (y_{g1},...,y_{gn_g})'$, $X_g = (X_{g1},...,X_{gn_g})'$, and $Z_g = (Z_{g1},...,Z_{gn_g})'$ denote an $n_g \times 1$ vector of dependent variables, an $n_g \times 1$ vector of endogenous regressors, and an $n_g \times k$ matrix of instruments for the g-th cluster. Let G denote the number of clusters and n denote the total number of observations. Similar to the case with heteroskedastic data, we can define the extended regression $y_g = X_g\theta + v_g a^c + e_g$, where $a^c = \left(n^{-1}\sum_{g=1}^G E_F[v_g'v_g]\right)^{-1}\left(n^{-1}\sum_{g=1}^G E_F[v_g'u_g]\right)$, and the corresponding cluster-robust test statistic for the null of exogeneity $H_a^c : a^c = 0$ takes the form

$$H_n^c \;\; = \;\; (\hat{a}^c)^2/\hat{V}_a^c, \tag{3.12}$$

where $\hat{V}_a^c = (n^{-1}\tilde{v}'\tilde{v})^{-1}\left(n^{-2}\sum_{g=1}^G \tilde{v}_g'\hat{e}_g\hat{e}_g'\tilde{v}_g\right)(n^{-1}\tilde{v}'\tilde{v})$, $\hat{a}^c = (\tilde{v}'\tilde{v})^{-1}\tilde{v}'y$, $\tilde{v}'\tilde{v} = \sum_{g=1}^G \tilde{v}_g'\tilde{v}_g$, and $\tilde{v}'y = \sum_{g=1}^G \tilde{v}_g'y_g$. In addition, we define $\hat{V}_{2sls}^c$ and $\hat{V}_{ols}^c$ following those in the heteroskedastic case.[11] Then, the cluster-robust t-test statistics $T_{ols}^c(\theta_0)$, $T_{2sls}^c(\theta_0)$, the two-stage test statistic $T_{1,n}^c(\theta_0)$ and the shrinkage test statistic $T_{2,n}^c(\theta_0)$ can be defined according to the definitions in Section 2.1.

For the case with clustered data, define the vector of nuisance parameters $\gamma^c = (\gamma_1^c, \gamma_2^c, \gamma_3^c)$ by

$$\gamma_1^c = a^c\left(\frac{\gamma_{21}^{c\prime}\gamma_{22}^c\gamma_{21}^c}{(\gamma_{21}^{c\prime}\gamma_{24}^c\gamma_{21}^c)^2} + \gamma_{23}^c\gamma_{25}^{c-2}\right)^{-1/2}, \;\; \gamma_2^c = (\gamma_{21}^c, \gamma_{22}^c, \gamma_{23}^c, \gamma_{24}^c, \gamma_{25}^c), \;\; \gamma_3^c = F, \tag{3.13}$$

and the corresponding parameter space by $\Gamma^c = \{\gamma^c = (\gamma_1^c, \gamma_2^c, \gamma_3^c) : \gamma_1^c \in \Gamma_1^c, \gamma_2^c \in \Gamma_2^c, \gamma_3^c \in \Gamma_3^c(\gamma_1^c, \gamma_2^c)\}$, where

$$\begin{aligned}
&\Gamma_1^c = \mathbb{R}, \;\; \Gamma_2^c = \Bigg\{(\gamma_{21}^c, \gamma_{22}^c, \gamma_{23}^c, \gamma_{24}^c, \gamma_{25}^c) : \gamma_{21}^c = \pi \in \mathbb{R}^k, \gamma_{22}^c = \mu_n\left(n^{-2}\sum_{g=1}^G E_F Z_g'e_ge_g'Z_g\right) \in \mathbb{R}^{k\times k}, \\
&\gamma_{23}^c = \mu_n\left(n^{-2}\sum_{g=1}^G E_F v_g'e_ge_g'v_g\right) \in \mathbb{R}, \gamma_{24}^c = n^{-1}\sum_{g=1}^G E_F Z_g'Z_g \in \mathbb{R}^{k\times k}, \gamma_{25}^c = n^{-1}\sum_{g=1}^G E_F v_g'v_g \in \mathbb{R}, \\
&s.t. \;\; \|\gamma_{21}^c\| \geq \underline{\kappa}, \lambda_{min}(\gamma_{22}^c) \geq \underline{\kappa}, \gamma_{23}^c > 0, \lambda_{min}(\gamma_{24}^c) \geq \underline{\kappa}, \text{ and } \gamma_{25}^c > 0\Bigg\},
\end{aligned} \tag{3.14}$$

for some $\underline{\kappa} > 0$ that does not depend on n, and $\{\mu_n\}$ is a non-random sequence, which plays the similar role as that used in Djogbenou et al. (2019) and is needed because different from the

[11] Specifically, we define $\hat{V}_{ols}^c = (n^{-1}X'X)^{-1}\left(n^{-2}\sum_{g=1}^G X_g'\hat{u}_g(\hat{\theta}_{ols})\hat{u}_g'(\hat{\theta}_{ols})X_g\right)(n^{-1}X'X)^{-1}$, $\hat{u}_g\left(\hat{\theta}_{ols}^c\right) = y_g - X_g\hat{\theta}_{ols}^c$, $\hat{V}_{2sls}^c = (n^{-1}X'P_ZX)^{-1}\hat{\pi}'\left(n^{-2}\sum_{g=1}^G Z_g'\hat{u}_g(\hat{\theta}_{2sls})\hat{u}_g'(\hat{\theta}_{2sls})Z_g\right)\hat{\pi}(n^{-1}X'P_ZX)^{-1}$, $\hat{u}_g\left(\hat{\theta}_{2sls}^c\right) = y_g - X_g\hat{\theta}_{2sls}^c$, where $X'X = \sum_{g=1}^G X_g'X_g$, $Z'Z = \sum_{g=1}^G Z_g'Z_g$, $Z'X = \sum_{g=1}^G Z_g'X_g$ and $\hat{\pi} = (Z'Z)^{-1}Z'X$, with $\hat{\theta}_{2sls}^c$ and $\hat{\theta}_{ols}^c$ denoting the 2SLS and OLS estimators under the clustered sample, respectively.

model with heteroskedastic errors, the rate of convergence of the estimators $\hat{\theta}^c_{ols}$, $\hat{\theta}^c_{2sls}$, and $\hat{a}^c$ under clustering depends on various factors such as the regressor X, the instruments Z, the relative cluster size, and the intra-cluster correlation [also see Hansen and Lee (2019, Section 4) for related discussions]. As pointed out by Djogbenou et al. (2019), the sequence $\{\mu_n\}$ can be interpreted as the rate at which information accumulates, and because of the studentization of the test statistics, $\{\mu_n\}$ needs not to be known in practice, but only needs to exist. Similarly, we need to standardize the nuisance parameter γ_1^c in (3.13) so that the size-corrected wild bootstrap procedure can be implemented under clustering without knowing $\{\mu_n\}$.

In addition, $\Gamma_3^c(\gamma_1^c,\gamma_2^c)$ is defined as follows:

$$\Gamma_3^c(\gamma_1^c,\gamma_2^c) = \Big\{F:\ E_F e_g' v_g = E_F Z_g' e_g = E_F Z_g' v_g = 0,\ E_F Z_g' e_g e_g' v_g = E_F Z_g' v_g e_g' v_g = E_F Z_g' e_g v_g' Z_g = 0,$$
$$\mu_n \left(n^{-2} \sum_{g=1}^{G} E_F Z_g' v_g v_g' Z_g \right) \in \mathbb{R}^{k\times k} \text{ with } \lambda_{min}\left(\mu_n n^{-2} \sum_{g=1}^{G} E_F Z_g' v_g v_g' Z_g \right) \geq \underline{\kappa},$$
$$\left\| \sup_{g,i} E_F \left(||Z_{gi}e_{gi}||^{2+\xi}, ||Z_{gi}v_{gi}||^{2+\xi}, |v_{gi}e_{gi}|^{2+\xi}, ||Z_{gi}Z_{gi}'||^{2+\xi}, |X_{gi}|^{2(2+\xi)} \right)' \right\| \leq M \Big\}, \tag{3.15}$$

for some constant $\underline{\kappa} > 0$, $\xi > 0$, $M < \infty$, and $\{\mu_n\}$ is the non-random sequence defined above.

We then define the whole nuisance parameter space Γ^c of γ^c as $\Gamma^c = \{\gamma^c = (\gamma_1^c,\gamma_2^c,\gamma_3^c) : \gamma_1^c \in \Gamma_1^c, \gamma_2^c \in \Gamma_2^c, \gamma_3^c \in \Gamma_3^c(\gamma_1^c,\gamma_2^c)\}$. Similar to the heteroskedastic case, to derive the asymptotic size, it suffices to study the asymptotic NRP along certain sequence $\{\gamma_{n,h}^c\}$ for some $h^c \in \mathscr{H}^c$, $\gamma_{n,h}^c = (\gamma_{n,h,1}^c, \gamma_{n,h,2}^c, \gamma_{n,h,3}^c)$ satisfies:

$$\mu_n^{1/2}\gamma_{n,h,1}^c \to h_1^c,\ \gamma_{n,h,2}^c \to h_2^c, \text{ and } \gamma_{n,h,3}^c = F_n \in \Gamma_3^c(\gamma_{n,h,1}^c,\gamma_{n,h,2}^c). \tag{3.16}$$

Also following Djogbenou et al. (2019, Assumption 3), we impose a condition on the number of clusters and the extent of heterogeneity of cluster size n_g (see p.396 of their paper for detailed discussions on this condition).

Assumption 3.5 *For $\{\mu_n\}$ defined in (3.14) and ξ defined in (3.15), $G \to \infty$ and $\mu_n^{\frac{2+\xi}{2+2\xi}} \sup_g \frac{n_g}{n} \to 0$.*

Now we present the algorithm of the wild cluster bootstrap procedure with the independent transformation that will be used to construct the uniformly valid bootstrap CVs under clustering.

Wild Cluster Bootstrap Algorithm:

1. Given $H_0: \theta = \theta_0$, compute the residuals from the first-stage and structural equations: $\hat{v}_g = X_g - Z_g\hat{\pi}$, $\hat{u}_g(\theta_0) = y_g - X_g\theta_0$, where $\hat{\pi} = (Z'Z)^{-1}Z'X = \left(\sum_{g=1}^G Z_g'Z_g\right)^{-1} \sum_{g=1}^G Z_g'X_g$.
2. Generate the cluster-level bootstrap pseudo-data following $X_g^* = Z_g\hat{\pi} + v_g^*$, $y_g^* = X_g^*\theta_0 + u_g^*$, where $v_g^* = \hat{v}_g\omega_{1g}^*$, and $u_g^* = \hat{u}_g(\theta_0)\omega_{2g}^*$, for each $g = 1,\ldots,G$, where ω_{1g}^* and ω_{2g}^* are

two random variables that has mean 0 and variance 1, are independent from the data and independent from each other.

3. Compute $\left\{\hat{\theta}^{c*}_{ols}, \hat{\theta}^{c*}_{2sls}, \hat{a}^{c*}, \hat{V}^{c*}_{ols}, \hat{V}^{c*}_{2sls}, \hat{V}^{c*}_{a}\right\}$ by the bootstrap samples generated in Step 2.

Then, we may generate the bootstrap test statistics under h^c_1 as follows:

$$\begin{aligned} T^{c*}_{ols,(h^c_1,\hat{h}^c_{n,2})}(\theta_0) &= \left(\hat{\theta}^{c*}_{ols} - \theta_0\right)/\hat{V}^{c*1/2}_{ols} + \left(\hat{h}^{c\prime}_{21}\hat{h}^c_{24}\hat{h}^c_{21} + \hat{h}^c_{25}\right)^{-1}\hat{h}^c_{25}\left(\hat{V}^*_a/\hat{V}^*_{ols}\right)^{1/2} h^c_1, \\ H^{c*}_{n,(h^c_1,\hat{h}^c_{n,2})} &= \left(\hat{h}^{c*}_{n,1} + h^c_1\right)^2, \end{aligned} \tag{3.17}$$

where $\hat{h}^c_{n,21} = \left(\sum_{g=1}^G Z'_g Z_g\right)^{-1}\sum_{g=1}^G Z'_g X_g$, $\hat{h}^c_{n,24} = n^{-1}\sum_{g=1}^G Z'_g Z_g$, $\hat{h}^c_{n,25} = n^{-1}\sum_{g=1}^G \hat{v}'_g \hat{v}_g$, and $\hat{h}^{c*}_{n,1} = \hat{a}^{c*}/\hat{V}^{c*1/2}_a$. $T^{c*}_{1,n,(h^c_1,\hat{h}^c_{n,2})}(\theta_0)$ and $T^{c*}_{2,n,(h^c_1,\hat{h}^c_{n,2})}(\theta_0)$, the bootstrap analogues of the two-stage and shrinkage test statistics evaluated at h^c_1, can be obtained subsequently. Notice that because of the studentization, $\{\mu_n\}$ is also not needed in the procedure described by (3.17).

Now, let $CI_{\alpha-\delta}(\hat{h}^c_{n,1})$ denote the $1-(\alpha-\delta)$ level confidence set for h^c_1 for some $0 < \delta \leq \alpha < 1$. The SBCV for clustered data is defined as

$$c^{B-S}_l(\alpha, \alpha-\delta, \hat{h}^c_{n,1}, \hat{h}^c_{n,2}) = \sup_{h^c_1 \in CI_{\alpha-\delta}(\hat{h}^c_{n,1})} c^*_{l,(h^c_1,\hat{h}^c_{n,2})}(1-\delta) \text{ for } l \in \{1,2\}, \tag{3.18}$$

where $c^*_{l,(h^c_1,\hat{h}^c_{n,2})}(1-\delta)$ is the $(1-\delta)$-th quantile of the distribution of $T^*_{l,n,(h^c_1,\hat{h}^c_{n,2})}(\theta_0)$, which is the bootstrap analogue of $T^c_{l,n}(\theta_0)$ generated under the value of localization parameter equal to h^c_1. The specific size-corrected bootstrap algorithm for the SBCV follows closely that for heteroskedastic data in Section 3.2 and is thus omitted for conciseness. The result for the SBCV is stated below.

Theorem 3.6 *Suppose that H_0 and Assumption* **3.5** *hold, then we have for any $0 < \delta \leq \alpha < 1$ and for $l \in \{1,2\}$, $AsySz\left[c^{B-S}_l(\alpha, \alpha-\delta, \hat{h}^c_{n,1}, \hat{h}^c_{n,2})\right] \leq \alpha$.*

For further refinement on the Bonferroni bound, we define the size-adjustment factor $\hat{\eta}^c_{l,n}$ following (3.8). Then, the BACV for the case with clustered data can be defined as

$$c^{B-A}_l(\alpha, \alpha-\delta, \hat{h}^c_{n,1}, \hat{h}^c_{n,2}) = c^{B-S}_l(\alpha, \alpha-\delta, \hat{h}^c_{n,1}, \hat{h}^c_{n,2}) + \hat{\eta}^c_{l,n}. \tag{3.19}$$

Similarly, its algorithm follows closely that described in Section 3.2.

Let $\tilde{T}^c_{l,h}$ denote the weak limit of $T^c_{l,n}(\theta_0)$ under the sequence $\{\gamma^c_{n,h}\} \subset \Gamma^c$ satisfying (3.16) and define $c^{B-S}_l(\alpha, \alpha-\delta, \tilde{h}^c_1, h_2) = \sup_{h^c_1 \in CI_{\alpha-\delta}(\tilde{h}^c_1)} c_{l,h^c}(1-\delta)$, where $c_{l,h^c}(1-\delta)$ is the $(1-\delta)$-th quantile of $\tilde{T}^c_{l,h}$ for $l \in \{1,2\}$. We assume the following continuity condition, similar to that assumed in the heteroskedastic case, and Theorem **3.8** shows that the size-adjusted bootstrap CV achieves correct asymptotic size with clustered samples.

Assumption 3.7 $P\left[\tilde{T}^c_{l,h} = c^{B-S}_l(\alpha, \alpha-\delta, \tilde{h}^c_1, h_2) + \eta\right] = 0$, $\forall h^c_1 \in H^c_1$ *and* $\eta \in [-c^{B-S}_l(\alpha, \alpha - \delta, \tilde{h}^c_1, h_2), 0]$.

Theorem 3.8 *Suppose that H_0, Assumptions* **3.5** *and* **3.7** *hold, then we have for any* $0 < \delta \leq \alpha < 1$ *and for $l \in \{1,2\}$, $AsySz\left[c^{B-A}_l(\alpha, \alpha-\delta, \hat{h}^c_{n,1}, \hat{h}^c_{n,2})\right] = \alpha$.*

Furthermore, let $CS^c_{l,n}(1-\alpha)$ denote the $1-\alpha$ confidence set constructed by collecting all the value of θ that cannot be rejected by the corresponding test at nominal level α under clustering.

Corollary 3.9 *Suppose that Assumptions* **3.5** *and* **3.7** *hold, then we have for any* $0 < \delta \leq \alpha < 1$ *and for $l \in \{1,2\}$:* $\liminf_{n\to\infty} \inf_{\gamma^c \in \Gamma^c} P_{\theta,\gamma^c}\left[\theta \in CS^c_{l,n}(1-\alpha)\right] = 1-\alpha$.

4. Finite sample power performance

In this section, we study the finite-sample power performance of the size-corrected wild bootstrap procedure by conducting simulations for the linear IV model under conditional heteroskedasticity or clustering. For all simulations, the number of Monte Carlo replications is set at $5,000$, and the number of bootstrap replications is set at $B = 599$. We compare the performance of the 2SLS-based wild bootstrap t-test (without pretest or shrinkage), the two-stage test based on the size-adjusted wild bootstrap CVs, and the test that is based on Hansen (2017)'s shrinkage apporach and its corresponding size-adjusted wild bootstrap CVs. We set $\alpha = .05$ for the CVs of the three tests. In addition, we set $\beta = .05$ for the nominal level of the pretest. The algorithms for the size-adjusted wild bootstrap CVs are executed with $\delta = \alpha - \alpha/10 = .045$, following the recommendation in McCloskey (2017). As explained by McCloskey (2017, Section 3.5), this choice of δ tends to have good power performance in both regions of the parameter space in which the key nuisance parameter (i.e., γ_1 or γ^c_1 in the current context) is far from zero and those in which it is close to zero. In Section SA.3 of the Supplementary Material, we provide further simulation results with other choices of β and δ, which show similar patterns as the results reported here. The shrinkage parameter τ in Hansen (2017)'s procedure is set to equal 1, 0.5, or 0.25. The random weights for the wild bootstrap are generated from the standard normal distribution throughout the simulations.

We first study the case with heteroskedastic errors. The simulation model follows the IV model in (2.1), and the DGP is specified as

$$\begin{aligned}&(\tilde{u}_i, \tilde{\varepsilon}_i)' \sim i.i.d.\, N(0, I_2),\ Z_i \sim i.i.d.\, N(0,1) \text{ and is independent from } (\tilde{u}_i, \tilde{\varepsilon}_i)',\\ &\tilde{v}_i = \rho \tilde{u}_i + (1-\rho^2)^{1/2}\tilde{\varepsilon}_i,\ u_i = f(Z_i)\tilde{u}_i, \text{ and } v_i = f(Z_i)\tilde{v}_i,\end{aligned} \tag{4.20}$$

where $i = 1, ..., n$ and $f(x) = x^2$. The sample size is set at $n = 200$ for the heteroskedastic case. The true value of θ is equal to zero, the first-stage coefficient is set at $\pi = \phi^{1/2} \cdot n^{-1/2}$, where we let $\phi \in \{2, 4, 16, 64\}$ to characterize different situations of identification strength for θ, and the true values of the endogeneity parameter are set at $\rho \in \{0.01, 0.05, 0.1, 0.2, 0.4, 0.6\}$.

Figures 2 - 3 show the finite-sample power curves of the tests under heteroskedasticity. The results with $\phi \in \{2,4\}$ are reported in Figures 2 and those with $\phi \in \{16,64\}$ are reported in Figure 3, respectively. We highlight some findings below. First, it is clear that the size-adjusted bootstrap tests have remarkable power gain over the 2SLS-based bootstrap t-test when the IV is relatively weak (e.g., $\phi \in \{2,4\}$) and/or when the degree of endogeneity is low (e.g., $\rho \in \{0.01, 0.05, 0.1, 0.2, 0.4\}$). Such power gain originates from the inclusion of the OLS-based t-test in the two-stage and shrinkage test statistics. Second, we notice that the shrinkage bootstrap tests have power advantage over the two-stage bootstrap test for distant alternative hypotheses. Third, the shrinkage bootstrap test with $\tau = 1$ typically has the best power performance among the size-adjusted bootstrap tests. Fourth, when the IV is very strong ($\phi = 64$), the 2SLS-based bootstrap t-test begins to have power advantage over the size-adjusted bootstrap tests. Furthermore, the 2SLS bootstrap t-test has some modest size distortions when the degree of endogeneity is high ($\rho = 0.6$) and the identification is not strong, but the distortion also disappears when $\phi = 64$. This is in line with the results obtained in Angrist and Kolesár (2021, Section 3) for the 2SLS t-test with one IV, which is the leading case in empirical applications [e.g., 101 out of 230 specifications in Andrews et al. (2019)'s sample and 1,087 out of 1,359 in Young (2020)'s sample only have one IV]. Angrist and Kolesár (2021) argue that with one IV, the asymptotic size distortions of the 2SLS t-test is typically modest unless endogeneity is very high. For example, the distortion is no more than 5% if $|\rho| < 0.76$ even under weak IV. However, although its size is distorted little, its power may also be rather low due to the weak IV, leading to relatively uninformative empirical results. On the other hand, in such cases our proposed tests have the advantage of providing power improvement while controlling the size.[12]

Then, we study the finite-sample power performance of the three tests under clustering. The model for the clustering case follows (3.11), and the disturbances $(u_{gi}, v_{gi})'$ consist of idiosyncratic errors $(\tilde{u}_{gi}, \tilde{v}_{gi})'$ and cluster effects $(\tilde{d}_{u,g}, \tilde{d}_{v,g})'$, which are specified as

$$\begin{aligned}
&(\tilde{u}_{gi}, \tilde{\varepsilon}_{gi})' \sim i.i.d.\, N(0, I_2),\ (\tilde{d}_{u,g}, \tilde{d}_{\varepsilon,g})' \sim i.i.d.\, N(0, I_2),\ Z_{gi} \sim i.i.d.\, N(0,1),\\
&(\tilde{u}_{gi}, \tilde{\varepsilon}_{gi})', (\tilde{d}_{u,g}, \tilde{d}_{\varepsilon,g})', \text{ and } Z_{gi} \text{ are independent from each other},\\
&\tilde{v}_{gi} = \rho \tilde{u}_{gi} + (1-\rho^2)^{1/2} \tilde{\varepsilon}_{gi},\ \tilde{d}_{v,g} = \rho \tilde{d}_{u,g} + (1-\rho^2)^{1/2} \tilde{d}_{\varepsilon,g},\\
&u_{gi} = f(Z_{gi})(\tilde{u}_{gi} + \tilde{d}_{u,g}), \text{ and } v_{gi} = f(Z_{gi})(\tilde{v}_{gi} + \tilde{d}_{v,g}),
\end{aligned} \tag{4.21}$$

where $i = 1, \ldots, n_g$, $g = 1, \ldots, G$, and $f(x) = x^2$. The settings for θ, π, ϕ, and ρ are the same as those for the case with heteroskedastic errors. Additionally, we consider two designs with heterogenous cluster sizes. In design (1), we let $n_1 = 20$ with $G_1 = 20$ (i.e., 20 clusters with cluster-level sample size equal to 20), $n_2 = 15$ with $G_2 = 20$, $n_3 = 10$ with $G_3 = 20$, and $n_4 = 5$ with $G_4 = 20$, so that the total number of clusters is $G = 80$ and the total number of observations is $n = 1{,}000$. In Section SA.3 of the Supplementary Material, we further report the simulation results of design (2).

[12] In the over-identified case with multiple IVs, the size distortions of the 2SLS-based bootstrap t-test become more substantial, while our tests typically have much smaller distortions (the simulation results are available upon request).

Figures 4 - 5 show the finite-sample power curves of the tests under clustering with design (1). We find that when the identification is not strong and/or the degree of endogeneity is low, the size-adjusted wild bootstrap tests also exhibit remarkable power gain over the 2SLS-based bootstrap t-test. Overall, the simulation results show very similar patterns and suggest that our method could be particularly attractive in the cases where the identification may not be strong so that IV-based inference methods could suffer from low power but naively using two-stage procedure to select between the OLS and 2SLS-based t-tests may result in extreme size distortions.

5. Conclusions

We study how to conduct uniformly valid tests for the two-stage and shrinkage procedures in the IV model with heteroskedastic or clustered data. We first show that standard wild bootstrap procedures are invalid both conditionally and unconditionally under local endogeneity, although the one with dependent transformation has much smaller asymptotic size distortions than the one with independent transformation. Then, we propose a size-corrected wild bootstrap approach, which makes use of the standard wild bootstrap with independent transformation and a Bonferroni-based size-correction method. The size-adjustment provides refinement over the Bonferroni bounds so that the resulting tests achieve correct asymptotic size. We show that the size-corrected wild bootstrap is uniformly valid under both heteroskedasticity and clustering. Monte Carlo simulations confirm that our method is able to achieve remarkable power gains over the 2SLS-based bootstrap t-test, especially when the identification is not strong and/or when the degree of endogeneity is low. In addition, the size-corrected wild bootstrap test based on Hansen (2017)'s shrinkage approach has particularly good power performance.

References

Andrews, D. W. , Guggenberger, P. , 2009a. Hybrid and size-corrected subsampling methods. Econometrica 77(3), 721–762.

Andrews, D. W., Guggenberger, P., 2009b. Incorrect asymptotic size of subsampling procedures based on post-consistent model selection estimators. Journal of Econometrics 152(1), 19–27.

Andrews, D. W. , Guggenberger, P. , 2010a. Applications of subsampling, hybrid, and size-correction methods. Journal of Econometrics 158(2), 285–305.

Andrews, D. W., Guggenberger, P., 2010b. Asymptotic size and a problem with subsampling and with the m out of n bootstrap. Econometric Theory 26(2), 426–468.

Andrews, I., Gentzkow, M., Shapiro, J. M., 2017. Measuring the sensitivity of parameter estimates to estimation moments. The Quarterly Journal of Economics 132(4), 1553–1592.

Figure 2(a): Power of wild bootstrap tests under heteroskedasticity with $\phi = 2$

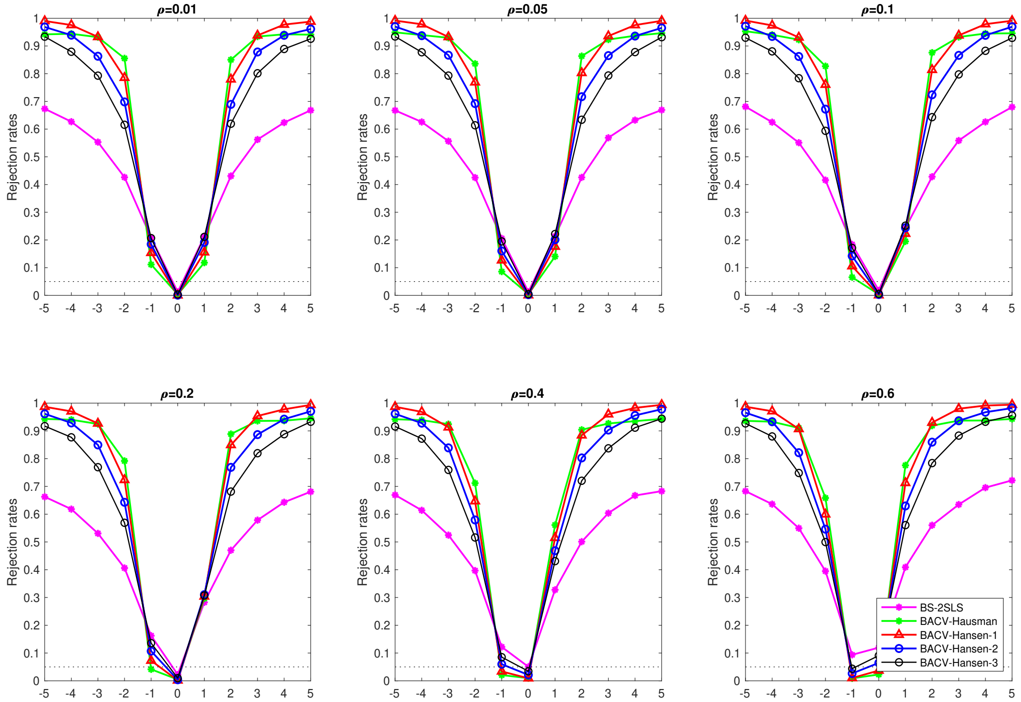
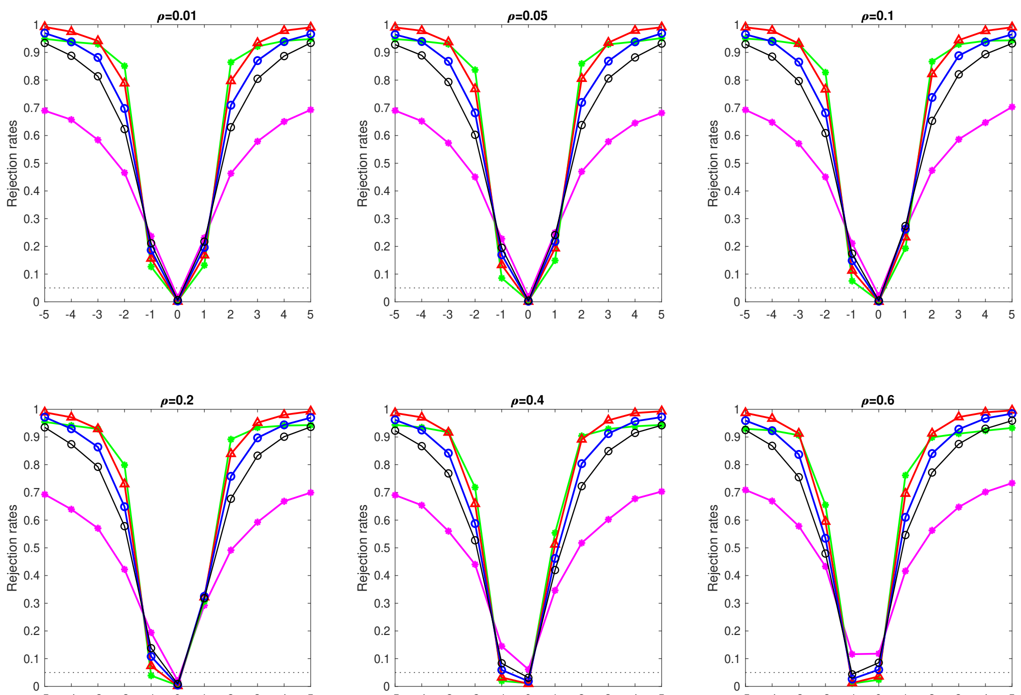


Figure 2(b): Power of wild bootstrap tests under heteroskedasticity with $\phi = 4$



Notes: The power curves for the bootstrap 2SLS-t test, the two-stage test with hybrid-BACVs, and the shrinkage test with BACVs with $\tau = 1, 0.5, 0.25$ are illustrated by the curves in pink, green, red, blue, and black, respectively.

Figure 3(a): Power of wild bootstrap tests under heteroskedasticity with $\phi = 16$

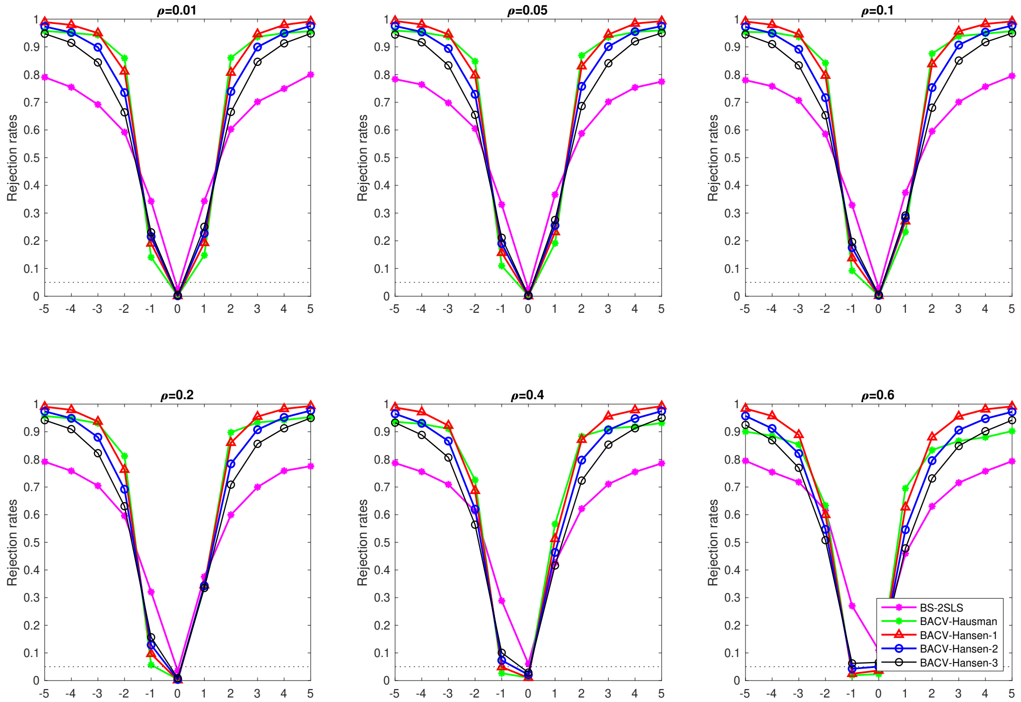
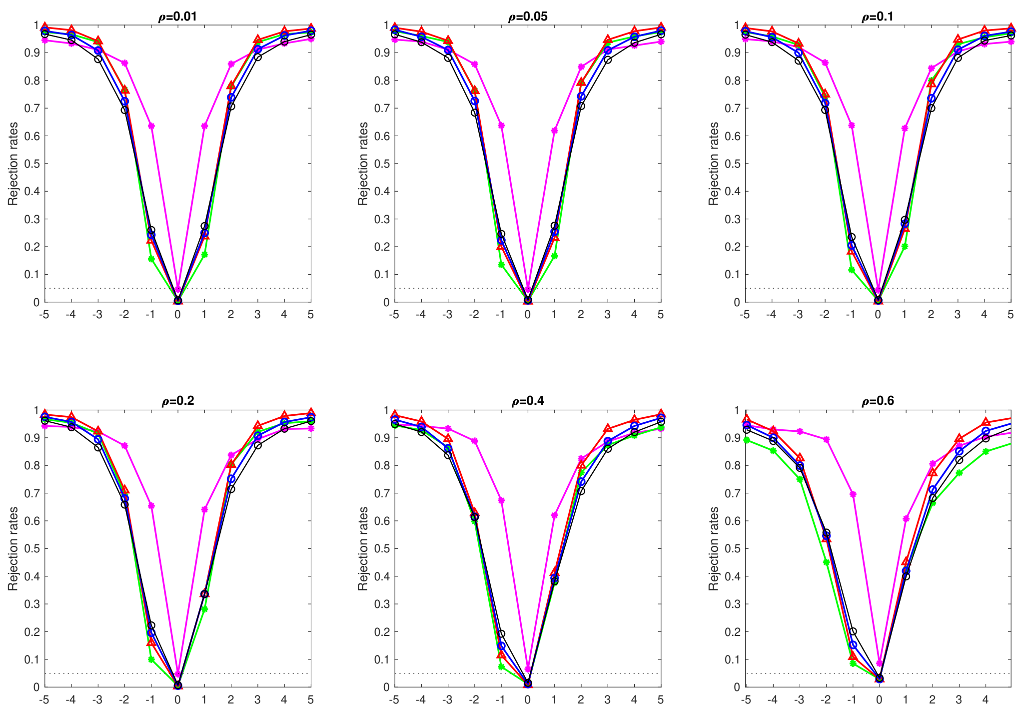


Figure 3(b): Power of wild bootstrap tests under heteroskedasticity with $\phi = 64$



Notes: The power curves for the bootstrap 2SLS-t test, the two-stage test with hybrid-BACVs, and the shrinkage test with BACVs with $\tau = 1, 0.5, 0.25$ are illustrated by the curves in pink, green, red, blue, and black, respectively.

Figure 4(a): Power of wild bootstrap tests under clustering (design 1) with $\phi = 2$

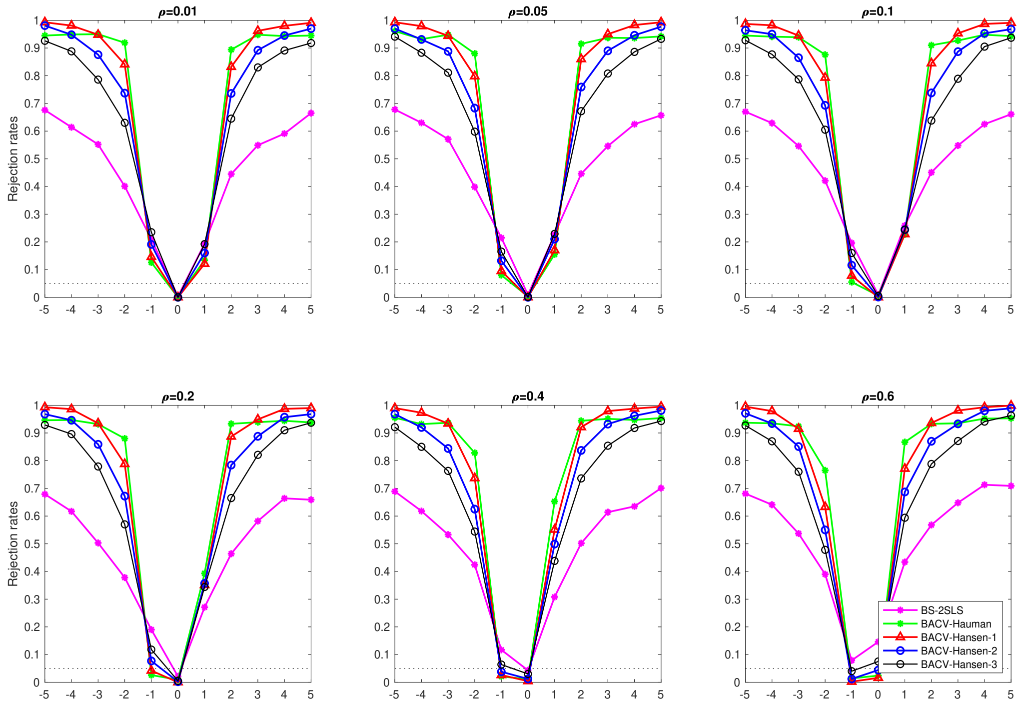
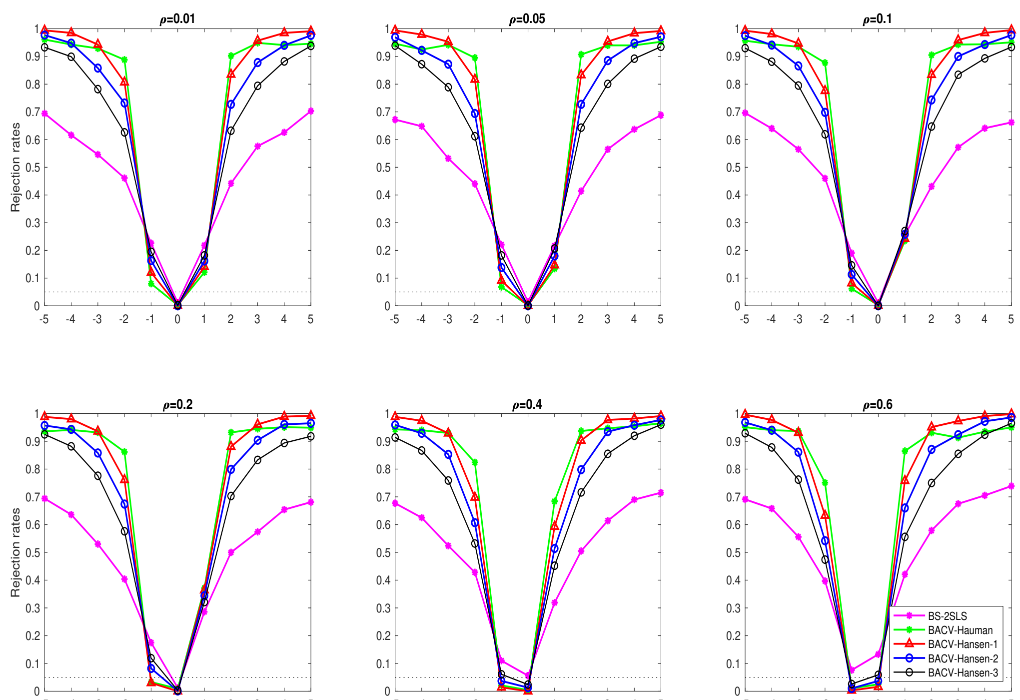


Figure 4(b): Power of wild bootstrap tests under clustering (design 1) with $\phi = 4$



Notes: The power curves for the bootstrap 2SLS-t test, the two-stage test with hybrid-BACVs, and the shrinkage test with BACVs with $\tau = 1, 0.5, 0.25$ are illustrated by the curves in pink, green, red, blue, and black, respectively.

Figure 5(a): Power of wild bootstrap tests under clustering (design 1) with $\phi = 16$

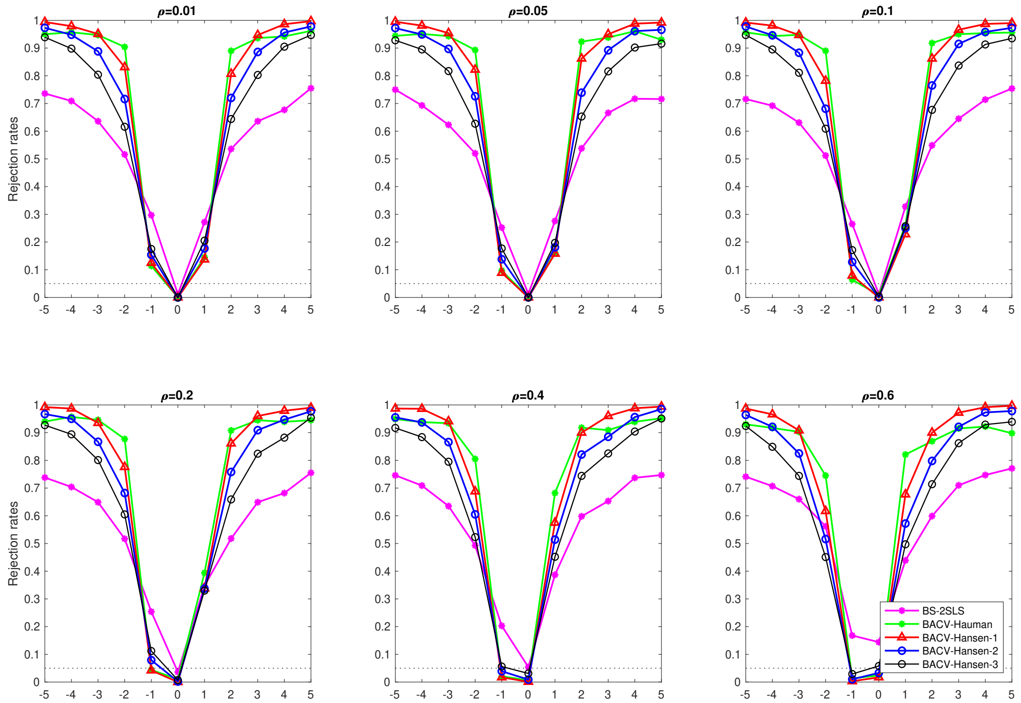
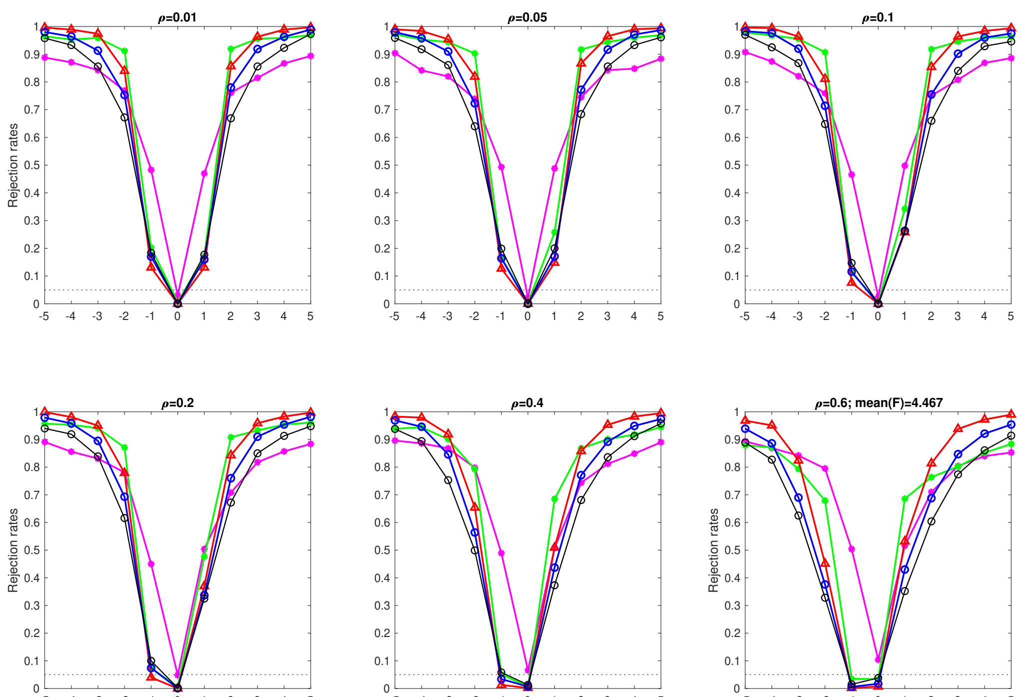


Figure 5(b): Power of wild bootstrap tests under clustering (design 1) with $\phi = 64$



Notes: The power curves for the bootstrap 2SLS-t test, the two-stage test with hybrid-BACVs, and the shrinkage test with BACVs with $\tau = 1, 0.5, 0.25$ are illustrated by the curves in pink, green, red, blue, and black, respectively.

Andrews, I., Stock, J. H., Sun, L., 2019. Weak instruments in instrumental variables regression: Theory and practice. Annual Review of Economics 11, 727–753.

Angrist, J. D., Evans, W. N., 1998. Children and their parents' labor supply: evidence from exogenous variation in family size. The American Economic Review 88(3), 450–477.

Angrist, J. D., Krueger, A. B., 1991. Does compulsory school attendance affect schooling and earning?. Quarterly Journal of Economics 106(4), 979–1014.

Angrist, J. D., Lavy, V., 1999. Using maimonides' rule to estimate the effect of class size on scholastic achievement. The Quarterly journal of economics 114(2), 533–575.

Angrist, J., Kolesár, M., 2021. One instrument to rule them all: The bias and coverage of just-id iv. arXiv preprint arXiv:2110.10556 .

Cameron, A. C., Gelbach, J. B., Miller, D. L., 2008. Bootstrap-based improvements for inference with clustered errors. The Review of Economics and Statistics 90(3), 414–427.

Cameron, A. C., Miller, D. L., 2015. A practitioner's guide to cluster-robust inference. Journal of human resources 50(2), 317–372.

Cavaliere, G., Georgiev, I., 2020. Inference under random limit bootstrap measures. Econometrica 88(6), 2547–2574.

Conley, T. G., Hansen, C. B., Rossi, P. E., 2012. Plausibly exogenous. Review of Economics and Statistics 94(1), 260–272.

Davidson, R., Flachaire, E., 2008. The wild bootstrap, tamed at last. Journal of Econometrics 146(1), 162–169.

Davidson, R., MacKinnon, J. G., 2008. Bootstrap inference in a linear equation estimated by instrumental variables. The Econometrics Journal 11(3), 443–477.

Davidson, R., MacKinnon, J. G., 2010. Wild bootstrap tests for iv regression. Journal of Business & Economic Statistics 28(1), 128–144.

Djogbenou, A. A., MacKinnon, J. G., Nielsen, M. Ø., 2019. Asymptotic theory and wild bootstrap inference with clustered errors. Journal of Econometrics 212(2), 393–412.

Doko Tchatoka, F., 2015. On bootstrap validity for specification tests with weak instruments. The Econometrics Journal 18(1), 137–146.

Doko Tchatoka, F., Dufour, J.-M., 2008. Instrument endogeneity and identification-robust tests: some analytical results. Journal of Statistical Planning and Inference 138(9), 2649–2661.

Doko Tchatoka, F., Dufour, J.-M., 2014. Identification-robust inference for endogeneity parameters in linear structural models. The Econometrics Journal 17(1), 165–187.

Doko Tchatoka, F., Dufour, J.-M., 2018. Instrument endogeneity and identification-robust tests: some analytical results. Journal of Econometrics Forthcoming.

Doko Tchatoka, F., Dufour, J.-M., 2020. Exogeneity tests and weak identification in IV regressions: asymptotic theory and point estimation. Technical report, Department of Economics, McGill University Montréal, Canada.

Finlay, K., Magnusson, L. M., 2019. Two applications of wild bootstrap methods to improve inference in cluster-iv models. Journal of Applied Econometrics 34(6), 911–933.

Gonçalves, S., White, H., 2004. Maximum likelihood and the bootstrap for nonlinear dynamic models. Journal of Econometrics 119(1), 199–219.

Guggenberger, P., 2010a. The impact of a hausman pretest on the asymptotic size of a hypothesis test. Econometric Theory 26(2), 369–382.

Guggenberger, P., 2010b. The impact of a hausman pretest on the size of a hypothesis test: The panel data case. Journal of Econometrics 156(2), 337–343.

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

Guggenberger, P., Kumar, G., 2012. On the size distortion of tests after an overidentifying restrictions pretest. Journal of Applied Econometrics 27(7), 1138–1160.

Hahn, J., Ham, J., Moon, H. R., 2010. The Hausman test and weak instruments. Journal of Econometrics 160, 289–299.

Han, S., McCloskey, A., 2019. Estimation and inference with a (nearly) singular Jacobian. Quantitative Economics 10(3), 1019–1068.

Hansen, B. E., 2005. Challenges for econometric model selection. Econometric Theory 21(1), 60–68.

Hansen, B. E., 2017. Stein-like 2sls estimator. Econometric Reviews 36(6-9), 840–852.

Hansen, B. E., Lee, S., 2019. Asymptotic theory for clustered samples. Journal of econometrics 210(2), 268–290.

Hansen, C., Hausman, J., Newey, W., 2008. Estimation with many instrumental variables. Journal of Business and Economic Statistics 26(4), 398–422.

Kabaila, P., 1995. The effect of model selection on confidence regions and prediction regions. Econometric Theory 11(3), 537–549.

Keane, M., Neal, T. et al., 2021. A practical guide to weak instruments. Technical report, University of New South Wales.

Leeb, H., Pötscher, B. M., 2005. Model selection and inference: Facts and fiction. Econometric Theory 21(1), 21–59.

Maasoumi, E., 1978. A modified stein-like estimator for the reduced form coefficients of simultaneous equations. Econometrica, pp. 695–703.

MacKinnon, J. G., Nielsen, M., Webb, M. D., 2020. Cluster-robust inference: A guide to empirical practice. Technical report, Qed working paper, Queen's University.

MacKinnon, J. G., Webb, M. D., 2017. Wild bootstrap inference for wildly different cluster sizes. Journal of Applied Econometrics 32(2), 233–254.

McCloskey, A., 2017. Bonferroni-based size-correction for nonstandard testing problems. Journal of Econometrics 200(1), 17–35.

McCloskey, A., 2020. Asymptotically uniform tests after consistent model selection in the linear regression model. Journal of Business & Economic Statistics 38(4), 810–825.

Roodman, D., Nielsen, M. Ø., MacKinnon, J. G., Webb, M. D., 2019. Fast and wild: Bootstrap inference in stata using boottest. The Stata Journal 19(1), 4–60.

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

Wang, W., 2020. On the inconsistency of nonparametric bootstraps for the subvector anderson–rubin test. Economics Letters 191, 109157.

Wang, W., Doko Tchatoka, F., 2018. On bootstrap inconsistency and bonferroni-based size-correction for the subset anderson–rubin test under conditional homoskedasticity. Journal of Econometrics 207(1), 188–211.

Wang, W., Kaffo, M., 2016. Bootstrap inference for instrumental variable models with many weak instruments. Journal of Econometrics 192(1), 231–268.

Wang, W., Zhang, Y., 2021. Wild bootstrap for instrumental variables regressions with weak and few clusters. arXiv preprint arXiv:2108.13707 .

Young, A., 2020. Consistency without inference: Instrumental variables in practical application. Technical report, London School of Economics.

A. Appendix

This Appendix contains the proofs of the theoretical results for the heteroskedastic case in the paper, and the proofs for the clustering case are given in Section SA.2 of the Supplementary Material.

The following lemma shows that the limiting distribution of $n^{1/2}(\hat{a}-\gamma_{n,h,1})$ is the same as that of $\left(n^{-1}\hat{v}'M_X\hat{v}\right)^{-1}\left(n^{-1/2}\hat{v}'M_Xe\right)$ under the parameter sequence $n^{1/2}\gamma_{n,h,1}\to h_1\in R$, which implies that the asymptotic variance of $n^{1/2}(\hat{a}-\gamma_{n,h,1})$ under local endogeneity is the same as that under exogeneity ($a=0$).

Lemma A.1 *Under the drift sequences of parameters* $\{\gamma_{n,h}\}$ *in* (2.16) *with* $|h_1|<\infty$, *we have*:

$$n^{1/2}(\hat{a}-\gamma_{n,h,1})=\left(n^{-1}\hat{v}'M_X\hat{v}\right)^{-1}\left(n^{-1/2}\hat{v}'M_Xe\right)+o_P(1).$$

The following lemma gives the limiting distributions of the estimators and test statistics under the sequences of drifting endogeneity parameter $n^{1/2}\gamma_{n,h,1}\to h_1\in R$.

Lemma A.2 *Under the drift sequences of parameters* $\{\gamma_{n,h}\}$ *in* (2.16) *with* $|h_1|<\infty$, *the following results hold*:

(a) *Asymptotic distributions of the estimators:*

$$\begin{pmatrix} n^{1/2}\hat{a} \\ n^{1/2}(\hat{\theta}_{ols}-\theta) \\ n^{1/2}(\hat{\theta}_{2sls}-\theta)\end{pmatrix} \to^d \begin{pmatrix}\psi_a \\ \psi_{ols} \\ \psi_{2sls}\end{pmatrix} = \begin{pmatrix} -\left(h_{21}'h_{24}h_{21}\right)^{-1}h_{21}'\psi_{Ze}+h_{25}^{-1}\psi_{ve}+h_1 \\ (h_{21}'h_{24}h_{21}+h_{25})^{-1}\left(h_{21}'\psi_{Ze}+\psi_{ve}+h_{25}h_1\right) \\ \left(h_{21}'h_{24}h_{21}\right)^{-1}h_{21}'\psi_{Ze}\end{pmatrix}$$

where $\psi_a \sim N\Big(h_1,\left(h_{21}'h_{24}h_{21}\right)^{-2}h_{21}'h_{22}h_{21}+h_{25}^{-2}h_{23}\Big)$, $\psi_{ols}\sim N\Big(h_{25}h_1/\left(h_{21}'h_{24}h_{21}+h_{25}\right),\left(h_{21}'h_{22}h_{21}+h_{23}\right)/\left(h_{21}'h_{24}h_{21}+h_{25}\right)^2\Big)$, *and* $\psi_{2sls}\sim N\Big(0,\left(h_{21}'h_{24}h_{21}\right)^{-2}h_{21}'h_{22}h_{21}\Big)$.

(b) *Asymptotic distributions of the test statistics:*

$$\begin{aligned}\begin{pmatrix}T_{2sls}(\theta_0)\\ T_{ols}(\theta_0)\\ H_n\end{pmatrix} \quad &\to^d \quad \eta_h=\begin{pmatrix}\eta_{1,h}\\ \eta_{2,h}\\ \eta_{3,h}\end{pmatrix}\\ &= \quad \begin{pmatrix}\left(h_{21}'h_{22}h_{21}\right)^{-1/2}h_{21}'\psi_{Ze}\\ (h_{21}'h_{22}h_{21}+h_{23})^{-1/2}\left(h_{21}'\psi_{Ze}+\psi_{ve}+h_{25}h_1\right)\\ \left(\frac{h_{21}'h_{22}h_{21}}{(h_{21}'h_{24}h_{21})^2}+h_{23}h_{25}^{-2}\right)^{-1}\left(-(h_{21}'h_{24}h_{21})^{-1}h_{21}'\psi_{Ze}+h_{25}^{-1}\psi_{ve}+h_1\right)^2\end{pmatrix}\\ T_{1,n}(\theta_0) \quad &\to^d \quad \tilde{T}_{1,h}=\left|\eta_{2,h}\mathbb{1}(\eta_{3,h}\le\chi^2_{1,1-\beta})+\eta_{1,h}\mathbb{1}(\eta_{3,h}>\chi^2_{1,1-\beta})\right|,\\ T_{2,n}(\theta_0) \quad &\to^d \quad \tilde{T}_{2,h}=\left|\eta_{2,h}w(\eta_{3,h})+\eta_{1,h}(1-w(\eta_{3,h}))\right|,\end{aligned}$$

where $\eta_{1,h} \sim N(0,1)$, $\eta_{2,h} \sim N\Big((h_{21}'h_{22}h_{21} + h_{23})^{-1/2}h_{25}h_1, 1\Big)$, *and* $\eta_{3,h} \sim \chi_1^2\Big(\big(\frac{h_{21}'h_{22}h_{21}}{(h_{21}'h_{24}h_{21})^2} + h_{23}h_{25}^{-2}\big)^{-1}h_1^2\Big)$.

The proofs for Lemmas **A.1-A.2** are given in Section SA.1 of the Supplementary Material.

Lemmas **A.3-A.4** are needed for the arguments with regard to the limiting distributions of the bootstrap analogues of the estimators and test statistics.

Lemma A.3 *For the independent bootstrap, suppose that* $E^*\left[|\omega_{1i}^*|^{2+\xi}\right] \leq C$ *and* $E^*\left[|\omega_{2i}^*|^{2+\xi}\right] \leq C$*; for the dependent bootstrap, suppose that* $E^*\left[|\omega_{1i}^*|^{2(2+\xi)}\right] \leq C$*, for some* $\xi > 0$ *and some large enough constant* C*. If further* $E_F\left[w_i^{2+\xi}\right] < \infty$ *for all* $w_i \in \left\{||Z_iu_i||, ||Z_iv_i||, ||Z_iZ_i'||, |u_iv_i|\right\}$ *and some* $\xi > 0$*, then under* H_0*,* $n^{-1}\sum_{i=1}^n E^*\left[||Z_iu_i^*||^{2+\xi}\right]$*,* $n^{-1}\sum_{i=1}^n E^*\left[||Z_iv_i^*||^{2+\xi}\right]$ *and* $n^{-1}\sum_{i=1}^n E^*\left[|u_i^*v_i^*|^{2+\xi}\right]$ *are bounded in probability.*

PROOF OF LEMMA **A.3**

The proof is straightforward for $n^{-1}\sum_{i=1}^n E^*\left[||Z_iu_i^*||^{2+\xi}\right]$. Indeed, we have

$$\begin{aligned} & n^{-1}\sum_{i=1}^n E^*\left[||Z_iu_i^*||^{2+\xi}\right] = n^{-1}\sum_{i=1}^n E^*\left[||Z_iu_i(\theta_0)\omega_{1i}^*||^{2+\xi}\right] = n^{-1}\sum_{i=1}^n E^*\left[||Z_iu_i(\theta_0)||^{2+\xi}|\omega_{1i}^*|^{2+\xi}\right] \\ = \; & n^{-1}\sum_{i=1}^n ||Z_iu_i(\theta_0)||^{2+\xi}E^*\left[|\omega_{1i}^*|^{2+\xi}\right] \leq Cn^{-1}\sum_{i=1}^n ||Z_iu_i(\theta_0)||^{2+\xi} = O_P(1), \end{aligned} \tag{A.1}$$

where the last equality follows from $\theta = \theta_0$ under the null hypothesis, $E_F[||Z_iu_i||^{2+\xi}] < \infty$, and $n^{-1}\sum_{i=1}^n ||Z_iu_i||^{2+\xi} - E_F[||Z_iu_i||^{2+\xi}] \to^P 0$ by Law of Large Numbers (LLN). Now, consider $n^{-1}\sum_{i=1}^n E^*\left[||Z_iv_i^*||^{2+\xi}\right]$. As in (A.1) we have for $j = 1$ or 2,

$$n^{-1}\sum_{i=1}^n E^*\left[||Z_iv_i^*||^{2+\xi}\right] = n^{-1}\sum_{i=1}^n ||Z_i\hat{v}_i||^{2+\xi}E^*\left[|\omega_{ji}^*|^{2+\xi}\right] \leq Cn^{-1}\sum_{i=1}^n ||Z_i\hat{v}_i||^{2+\xi}. \tag{A.2}$$

By using Minkowski and Cauchy-Schwartz inequalities, along with $\hat{v}_i = v_i - Z_i'(\hat{\pi} - \pi)$, we obtain

$$\begin{aligned} & n^{-1}\sum_{i=1}^n ||Z_i\hat{v}_i||^{2+\xi} = n^{-1}\sum_{i=1}^n ||Z_iv_i - Z_iZ_i'(\hat{\pi} - \pi)||^{2+\xi} \\ & \leq C_1\left\{n^{-1}\sum_{i=1}^n ||Z_iv_i||^{2+\xi} + ||\hat{\pi} - \pi||^{2+\xi}n^{-1}\sum_{i=1}^n ||Z_iZ_i'||^{2+\xi}\right\} = O_P(1), \end{aligned} \tag{A.3}$$

where C_1 denotes some large enough constant, and (A.3) holds because $\hat{\pi} - \pi \to^P 0$, $E_F\left[||Z_iv_i||^{2+\xi}\right] < \infty$, $E_F\left[||Z_iZ_i'||^{2+\xi}\right] < \infty$, $n^{-1}\sum_{i=1}^n ||Z_iv_i||^{2+\xi} - E_F\left[||Z_iv_i||^{2+\xi}\right] \to^P 0$ and $n^{-1}\sum_{i=1}^n ||Z_iZ_i'||^{2+\xi} - E_F\left[||Z_iZ_i'||^{2+\xi}\right] \to^P 0$ by LLN. Therefore, $n^{-1}\sum_{i=1}^n E^*\left[||Z_iv_i^*||^{2+\xi}\right]$ is bounded in probability from (A.2)-(A.3).

We now show that $n^{-1}\sum_{i=1}^{n} E^*\left[|u_i^* v_i^*|^{2+\xi}\right]$ is bounded in probability. For $j=1$ or 2, we have

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n} E^*\left[|u_i^* v_i^*|^{2+\xi}\right] &= n^{-1}\sum_{i=1}^{n} E^*\left[|u_i(\theta_0)\hat{v}_i|^{2+\xi}|\omega_{1i}^*\omega_{ji}^*|^{2+\xi}\right] \\
&= n^{-1}\sum_{i=1}^{n} |u_i(\theta_0)\hat{v}_i|^{2+\xi} E^*\left[|\omega_{1i}^*\omega_{ji}^*|^{2+\xi}\right]. \qquad \text{(A.4)}
\end{aligned}
$$

Note that $j=2$ for the wild bootstrap scheme with independent transformation, so that $E^*\left[|\omega_{1i}^*\omega_{ji}^*|^{2+\xi}\right] = E^*\left[|\omega_{1i}^*\omega_{2i}^*|^{2+\xi}\right] = E^*\left[|\omega_{1i}^*|^{2+\xi}\right]E^*\left[|\omega_{2i}^*|^{2+\xi}\right] \leq C_2$ for some large enough constant C_2. For the wild bootstrap scheme with dependent transformation, $j=1$, and we have $E^*\left[|\omega_{1i}^*\omega_{ji}^*|^{2+\xi}\right] = E^*\left[|\omega_{1i}^*|^{2(2+\xi)}\right] \leq C$. Combining both cases into (A.4) along with the fact that $u_i(\theta_0)\hat{v}_i = u_i(\theta_0)v_i - u_i(\theta_0)Z_i'(\hat{\pi}-\pi)$, $\theta=\theta_0$ under the null hypothesis, $E_F||Z_iu_i||^{2+\xi}<\infty$, $E_F|u_iv_i|^{2+\xi}<\infty$, and by using the arguments with Minkowski and Cauchy-Schwartz inequalities, we have

$$
n^{-1}\sum_{i=1}^{n} E^*\left[|u_i^* v_i^*|^{2+\xi}\right] \leq C_3\left\{n^{-1}\sum_{i=1}^{n}|u_i(\theta_0)v_i|^{2+\xi} + ||\hat{\pi}-\pi||^{2+\xi} n^{-1}\sum_{i=1}^{n}||Z_iu_i(\theta_0)||^{2+\xi}\right\} = O_P(1),
$$

for some large enough constants C_3. □

Lemma A.4 *Suppose that H_0 holds, the conditions of Lemma* **A.3** *are satisfied, $E^*[\omega_{1i}^*] = E^*[\omega_{2i}^*] = 0$, and $Var^*[\omega_{1i}^*] = Var^*[\omega_{2i}^*] = 1$. For the dependent bootstrap, further suppose that $E^*[\omega_{1i}^{*3}] = 0$ and $E^*[\omega_{1i}^{*4}] = 1$. Then, under the sequence $\{\gamma_{n,h}\}$ defined in* (2.16) *with $|h_1|<\infty$ we have*:

$$
\begin{pmatrix} n^{-1/2}Z'u^* \\ n^{-1/2}\left(u^{*\prime}v^* - E^*\left[u^{*\prime}v^*\right]\right)\end{pmatrix} \to^{d^*} \begin{pmatrix}\psi_{ze}^* \\ \psi_{ve}^*\end{pmatrix} \sim N\left(0, \begin{pmatrix} h_{22} & 0 \\ 0' & h_{23}\end{pmatrix}\right), \qquad \text{(A.5)}
$$

in probability P.

PROOF OF LEMMA **A.4**

Let c_1 denote k-dimensional nonzero vectors, and c_2 denote a nonzero scalar. Define

$$
\begin{aligned}
U_{n,i}^* &= \left\{c_1' u_i^* Z_i + c_2\left(u_i^* v_i^* - E^*[u_i^* v_i^*]\right)\right\}/\sqrt{n} \\
&= \left\{c_1'\omega_{1i}^*\hat{u}_i(\theta_0)Z_i + c_2\left(\hat{u}_i(\theta_0)\hat{v}_i\omega_{1i}^*\omega_{ji}^* - E^*[\hat{u}_i(\theta_0)\hat{v}_i\omega_{1i}^*\omega_{ji}^*]\right)\right\}/\sqrt{n}, \qquad \text{(A.6)}
\end{aligned}
$$

where $j=1$ for the dependent bootstrap scheme and $j=2$ for the independent bootstrap scheme. It suffices to verify that the conditions of the Liapounov CLT hold for $U_{n,i}^*$. For brevity, we shall focus on the proof for the case with independent transformation (i.e., $j=2$). Note that the proof for the case with dependent transformation ($j=1$) follows similar steps.

(a) We have $E^*[U_{n,i}^*]=0$ as $E^*[\omega_{1i}^*\hat{u}_i(\theta_0)Z_i] = \hat{u}_i(\theta_0)Z_iE^*[\omega_{1i}^*] = 0$, and $E^*[\hat{u}_i(\theta_0)\hat{v}_i\omega_{1i}^*\omega_{2i}^* -$

$E^*[\hat{u}_i(\theta_0)\hat{v}_i\omega^*_{1i}\omega^*_{2i}]] = \hat{u}_i(\theta_0)\hat{v}_i E^*[\omega^*_{1i}\omega^*_{2i}] - \hat{u}_i(\theta_0)\hat{v}_i E^*[\omega^*_{1i}\omega^*_{2i}] = 0.$

(b) Note that

$$
\begin{aligned}
E^*[u_i^{*2} Z_i Z_i'] &= E^*[\hat{u}_i^2(\theta_0)\omega_{1i}^{*2} Z_i Z_i'] = \hat{u}_i^2(\theta_0) Z_i Z_i' E^*[\omega_{1i}^{*2}] = \hat{u}_i^2(\theta_0) Z_i Z_i', \\
E^*[u_i^{*2} v_i^{*2}] &= E^*[\hat{u}_i^2(\theta_0)\hat{v}_i^2\omega_{1i}^{*2}\omega_{2i}^{*2}] = \hat{u}_i^2(\theta_0)\hat{v}_i^2 E^*[\omega_{1i}^{*2}\omega_{2i}^{*2}] = \hat{u}_i^2(\theta_0)\hat{v}_i^2 E^*[\omega_{1i}^{*2}]E^*[\omega_{2i}^{*2}] = \hat{u}_i^2(\theta_0)\hat{v}_i^2, \\
E^*[u_i^{*2} v_i^* Z_i] &= E^*[\hat{u}_i^2(\theta_0)\hat{v}_i Z_i \omega_{1i}^{*2}\omega_{2i}^*] = \hat{u}_i^2(\theta_0)\hat{v}_i Z_i E^*[\omega_{1i}^{*2}\omega_{2i}^*] = \hat{u}_i^2(\theta_0)\hat{v}_i Z_i E^*[\omega_{1i}^{*2}]E^*[\omega_{2i}^*] = 0,
\end{aligned}
$$

which implies that under H_0,

$$
\sum_{i=1}^n E^*[U_{n,i}^{*2}] = c_1'\left(n^{-1}\sum_{i=1}^n \hat{u}_i^2(\theta_0) Z_i Z_i'\right)c_1 + c_2^2\left(n^{-1}\sum_{i=1}^n \hat{u}_i^2(\theta_0)\hat{v}_i^2\right) = c_1' h_{22} c_1 + c_2^2 h_{23} + o_P(1) = O_P(1). \tag{A.7}
$$

(c) We note that by Minkowski inequality, for some $\xi > 0$ and some large enough constant C_4,

$$
\sum_{i=1}^n E^*[\left|U_{n,i}^*\right|^{2+\xi}] \le C_4 n^{-\frac{\xi}{2}} n^{-1} \sum_{i=1}^n E^*\left[\left|c_1' Z_i^* u_i^*\right|^{2+\xi} + \left|c_2 u_i^* v_i^*\right|^{2+\xi}\right] \to^P 0, \tag{A.8}
$$

where the convergence in probability is obtained by using Lemma **A.3**.

From (a)-(c) above, $U_{n,i}^*$ satisfies the Lyapunov CLT conditions, and the result of Lemma **A.4** follows for the independent bootstrap. For the dependent bootstrap, notice that for (b),

$$
E^*[u_i^{*2} v_i^{*2}] = \hat{u}_i^2(\theta_0)\hat{v}_i^2 E^*[\omega_{1i}^{*4}] = \hat{u}_i^2(\theta_0)\hat{v}_i^2, \text{ and } E^*[u_i^{*2} v_i^* Z_i] = \hat{u}_i^2(\theta_0)\hat{v}_i Z_i E^*[\omega_{1i}^{*3}] = 0, \tag{A.9}
$$

and the desired result follows.

□

In the following theorem, we give the results of bootstrap inconsistency for the two-stage and shrinkage tests under local endogeneity. For this purpose, we notice that there are two sources of randomness in the bootstrap: the randomness from the original data and the randomness from the bootstrap procedure (i.e., the random weights of the wild bootstrap). Specifically, take the original sample as from the probability space $(\Omega, \mathscr{F}, P)$. In addition, suppose the randomness from the bootstrap is defined on a probability space $(\Lambda, \mathscr{G}, P^*)$, which is independent of $(\Omega, \mathscr{F}, P)$. Then, in the following theorem we view the bootstrap statistics as being defined on the product probability space $(\Omega, \mathscr{F}, P) \times (\Lambda, \mathscr{G}, P^*) = (\Omega \times \Lambda, \mathscr{F} \times \mathscr{G}, \mathbb{P})$, where $\mathbb{P} = P \times P^*$. Theorem **A.5** gives the null limiting distributions of the bootstrap statistics under $\mathbb{P}$. In particular, this framework is needed to characterize the asymptotic behaviour of the bootstrap statistics generated under the dependent transformation of disturbances.

Theorem A.5 *Suppose that H_0 and the conditions of Lemmas* **A.3** *and* **A.4** *hold. Then, under the* *sequence* $\{\gamma_{n,h}\}$ *defined in* (2.16) *with* $|h_1| < \infty$*:*

$$\begin{pmatrix} T^*_{2sls}(\theta_0) \\ T^*_{ols}(\theta_0) \\ H^*_n \end{pmatrix} \rightsquigarrow \eta^*_h \equiv \begin{pmatrix} \eta^*_{1,h} \\ \eta^*_{2,h} \\ \eta^*_{3,h} \end{pmatrix} = \begin{pmatrix} (h'_{21}h_{22}h_{21})^{-1/2} h'_{21}\psi^*_{Ze} \\ (h'_{21}h_{22}h_{21}+h_{23})^{-1/2}\left(h'_{21}\psi^*_{Ze}+\psi^*_{ve}+h_{25}h^b_1\right) \\ \left(\frac{h'_{21}h_{22}h_{21}}{(h'_{21}h_{24}h_{21})^2}+h_{23}h_{25}^{-2}\right)^{-1}\left(-(h'_{21}h_{24}h_{21})^{-1}h'_{21}\psi^*_{Ze}+\psi^*_{ve}+h^b_1\right)^2 \end{pmatrix},$$

$$T^*_{1,n}(\theta_0) \rightsquigarrow \tilde{T}^*_{1,h} = \left|\eta^*_{2,h}\mathbb{1}(\eta^*_{3,h} \le \chi^2_{1,1-\beta}) + \eta^*_{1,h}\mathbb{1}(\eta^*_{3,h} > \chi^2_{1,1-\beta})\right|,$$

$$T^*_{2,n}(\theta_0) \rightsquigarrow \tilde{T}^*_{2,h} = \left|\eta^*_{2,h}w(\eta^*_{3,h}) + \eta^*_{1,h}(1-w(\eta^*_{3,h}))\right|,$$

where $h^b_1 = 0$ *for the bootstrap based on independent transformation of disturbances, and* $h^b_1 = h_1 + h_{25}^{-1}\psi_{ve}$ *with* $\psi_{ve} \sim N(0,h_{23})$*, for the bootstrap based on dependent transformation of disturbances, and* $\rightsquigarrow$ *signifies the weak convergence under* $\mathbb{P}$*.*

PROOF OF THEOREM **A.5**

First, we note that

$$n^{-1}X^{*'}P_ZX^* = n^{-1}(Z\hat{\pi}+v^*)'P_Z(Z\hat{\pi}+v^*) = n^{-1}\hat{\pi}'Z'Z\hat{\pi} + n^{-1}\hat{\pi}'Z'v^* + n^{-1}v^{*'}Z\hat{\pi} + n^{-1}v^{*'}P_Zv^*$$
$$= n^{-1}\hat{\pi}'Z'Z\hat{\pi} + o_{P^*}(1) \to^{P^*} h'_{21}h_{24}h_{21}, \text{ in probability } P, \quad \text{(A.10)}$$

which follows from $\hat{\pi} - h_{21} \to^P 0$, $n^{-1}Z'Z - h_{24} \to^P 0$, and $n^{-1}Z'v^* \to^{P^*} 0$ in probability P. Using similar arguments, we obtain

$$n^{-1}X^{*'}X^* \to^{P^*} h'_{21}h_{24}h_{21} + h_{25}, \quad \text{(A.11)}$$

in probability P. Furthermore, using similar arguments as those for $\hat{V}_a$, $\hat{V}_{ols}$ and $\hat{V}_{2sls}$ in the proof of Lemma **A.2**, we obtain

$$n\hat{V}^*_a \to^{P^*} (h'_{21}h_{24}h_{21})^{-2}h'_{21}h_{22}h_{21} + h_{25}^{-2}h_{23}, \;\; n\hat{V}^*_{ols} \to^{P^*} (h'_{21}h_{24}h_{21}+h_{25})^{-2}\left(h'_{21}h_{22}h_{21}+h_{23}\right),$$
$$n\hat{V}^*_{2sls} \to^{P^*} \left(h'_{21}h_{24}h_{21}\right)^{-2}h'_{21}h_{22}h_{21}, \quad \text{(A.12)}$$

in probability P.

Second, we note that

$$n^{-1/2}X^{*'}P_Zu^* = n^{-1/2}(Z\hat{\pi}+v^*)'P_Zu^* = n^{-1/2}\hat{\pi}'Z'u^* + \left(n^{-1}v^{*'}Z\right)\left(n^{-1}Z'Z\right)^{-1}\left(n^{-1/2}Z'u^*\right)$$
$$= n^{-1/2}\hat{\pi}'Z'u^* + o_{P^*}(1) \to^{d^*} h'_{21}\psi^*_{Ze}, \quad \text{(A.13)}$$

in probability P, where the last equality follows from: (a) by Lemma **A.4**, $n^{-1/2}Z'u^* = O_{P^*}(1)$ in probability P; (b) $n^{-1}Z'v^* \to^{P^*} 0$ in probability P as $E^*[n^{-1}Z'v^*] = 0$; (c) $n^{-1}Z'Z \to^P h_{24}$, which is positive definite, and therefore $\left(n^{-1}Z'Z\right)^{-1} \to^P h_{24}^{-1}$. Then, the (conditional) convergence in distribution in (A.10) follows from Lemma **A.4**, along with the fact that $\hat{\pi} - h_{21} \to^P 0$.

Third, following the same arguments as above, we have $n^{-1/2}X^{*'}u^* = n^{-1/2}\hat{\pi}'Z'u^* +$

$n^{-1/2}\left(v^{*'}u^* - E^*[v^{*'}u^*]\right) + n^{-1/2}E^*[v^{*'}u^*]$, where

$$n^{-1/2}\hat{\pi}'Z'u^* + n^{-1/2}\left(v^{*'}u^* - E^*[v^{*'}u^*]\right) \to^{d^*} h_{21}'\psi_{Ze}^* + \psi_{ve}^*, \tag{A.14}$$

in probability P. Then, for $n^{-1/2}E^*[v^{*'}u^*]$, we notice that it is equal to zero under the independent transformation of disturbances. Under the dependent transformation, $n^{-1/2}E^*[v^{*'}u^*] = n^{1/2}\left(n^{-1}\sum_{i=1}^n \hat{v}_i\hat{u}_i(\theta_0)\right)$, where

$$\begin{aligned} n^{1/2}\left(n^{-1}\sum_{i=1}^n \hat{v}_i\hat{u}_i(\theta_0)\right) &= n^{1/2}\left(n^{-1}\sum_{i=1}^n (v_iu_i(\theta_0) - E_F[v_iu_i(\theta_0)])\right) + n^{1/2}E_F[v_iu_i(\theta_0)] + o_P(1) \\ &\to^d \psi_{ve} + h_{25}h_1. \end{aligned} \tag{A.15}$$

Finally, notice that the results in probability P in (A.10)-(A.14) are invariant to the original data, so they hold under $\mathbb{P}$ as well. Then, by (A.15) and the Continuous Mapping Theorem, we obtain that under H_0,

$$\begin{pmatrix} n^{1/2}\hat{a}^* \\ n^{1/2}(\hat{\theta}_{ols}^* - \theta_0) \\ n^{1/2}(\hat{\theta}_{2sls}^* - \theta_0) \end{pmatrix} \rightsquigarrow \begin{pmatrix} -(h_{21}'h_{24}h_{21})^{-1}h_{21}'\psi_{Ze}^* + h_{25}^{-1}\psi_{ve}^* + h_1^b \\ (h_{21}'h_{24}h_{21} + h_{25})^{-1}(h_{21}'\psi_{Ze}^* + \psi_{ve}^* + h_{25}h_1^b) \\ (h_{21}'h_{24}h_{21})^{-1}h_{21}'\psi_{Ze}^* \end{pmatrix}, \tag{A.16}$$

and the results in the statement of Theorem **A.5** follow. □

PROOF OF THEOREM **3.1**

First, note that by following similar arguments as those in the proofs of Theorem **A.5**, we can obtain that the following (conditional) convergence in distribution holds:

$$\begin{pmatrix} T_{OLS,(h_1,\hat{h}_{n,2})}^*(\theta_0) \\ H_{n,(h_1,\hat{h}_{n,2})}^* \end{pmatrix} \to^{d^*} \begin{pmatrix} (h_{21}'h_{22}h_{21} + h_{23})^{-1/2}(h_{21}'\psi_{Ze}^* + \psi_{ve}^* + h_{25}h_1) \\ \left(\frac{h_{21}'h_{22}h_{21}}{(h_{21}'h_{24}h_{21})^2} + h_{23}h_{25}^{-2}\right)^{-1}\left(-(h_{21}'h_{24}h_{21})^{-1}h_{21}'\psi_{Ze}^* + h_{25}^{-1}\psi_{ve}^* + h_1\right)^2 \end{pmatrix} \tag{A.17}$$

in probability P. Then, based on the formula of $T_{l,n,(h_1,\hat{h}_{n,2})}^*(\theta_0)$ for $l \in \{1,2\}$, we conclude that the (conditional) null limiting distribution of $T_{l,n,(h_1,\hat{h}_{n,2})}^*(\theta_0)$ is the same as the null limiting distribution of $T_{l,n}(\theta_0)$ with the value of localization parameter equal to h_1, and this implies that $c_{l,(h_1,\hat{h}_{n,2})}^*(1-\delta) \to^P c_{l,(h_1,h_2)}(1-\delta)$, where $c_{l,(h_1,h_2)}(1-\delta)$ denotes the $(1-\delta)$-th quantile of $\tilde{T}_{l,h}$ with $h = (h_1, h_2)$.

Then, the arguments for the proof is similar to those in McCloskey (2017). We note that there exists a "worst case sequence" $\gamma_n \in \Gamma$ such that $AsySz\left[c_l^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})\right]$ equals:

$$\limsup_{n\to\infty} \sup_{\gamma\in\Gamma} P_{\theta_0,\gamma}\left[T_{l,n}(\theta_0) > c_l^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})\right]$$

$$
\begin{aligned}
&= \limsup_{n\to\infty} P_{\theta_0,\gamma_n}\left[T_{l,n}(\theta_0) > c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})\right] \\
&= \lim_{n\to\infty} P_{\theta_0,\gamma_{m_n}}\left[T_{l,m_n}(\theta_0) > c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{m_n,1},\hat{h}_{m_n,2})\right] \qquad \text{(A.18)}
\end{aligned}
$$

where $\{m_n : n \geq 1\}$ is a subsequence of $\{n : n \geq 1\}$ and such a subsequence always exists. Furthermore, there exists a subsequence $\{\omega_n :\ n \geq 1\}$ of $\{m_n :\ n \geq 1\}$ such that:

$$
\begin{aligned}
&\lim_{n\to\infty} P_{\theta_0,\gamma_{m_n}}\left[T_{l,m_n}(\theta_0) > c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{m_n,1},\hat{h}_{m_n,2})\right] \\
&= \lim_{n\to\infty} P_{\theta_0,\gamma_{\omega_n,h}}\left[T_{l,\omega_n}(\theta_0) > c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{\omega_n,1},\hat{h}_{\omega_n,2})\right] \qquad \text{(A.19)}
\end{aligned}
$$

for some $h \in \mathscr{H}$. But, for any $h \in \mathscr{H}$, any subsequence $\{\omega_n : n \geq 1\}$ of $\{n : n \geq 1\}$, and any sequence $\{\gamma_{\omega_n,h} : n \geq 1\}$, we have $\left(T_{l,\omega_n}(\theta_0), \hat{h}_{\omega_n,1}\right) \to^d \left(\tilde{T}_{l,h}, \tilde{h}_1\right)$ jointly. In addition, $c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{\omega_n,1},\hat{h}_{\omega_n,2})$ is continuous in $\hat{h}_{\omega_n,1}$ by the definition of the SBCV and Maximum Theorem. Hence, the following convergence holds jointly by the Continuous Mapping Theorem:

$$
\left(T_{l,\omega_n}(\theta_0), c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{\omega_n,1},\hat{h}_{\omega_n,2})\right) \;\to^d\; \left(\tilde{T}_{l,h}, c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)\right) \qquad \text{(A.20)}
$$

where $c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2) = \sup_{h_1 \in CI_{\alpha-\delta}(\tilde{h}_1)} c_{l,(h_1,h_2)}(1-\delta)$. Then, (A.18)-(A.20) imply that

$$
\begin{aligned}
&AsySz\left[c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})\right] \\
&= \lim_{n\to\infty} P_{\theta_0,\gamma_{\omega_{n,h}}}\left[T_{l,\omega_n}(\theta_0) > c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{\omega_{n,1}},\hat{h}_{\omega_{n,2}})\right] \\
&= \sup_{h\in\mathscr{H}} P\left[\tilde{T}_{l,h} > c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)\right], \qquad \text{(A.21)}
\end{aligned}
$$

Now, for any $h \in \mathscr{H}$, we have:

$$
\begin{aligned}
&P\left[\tilde{T}_{l,h} \geq c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)\right] \\
&= P\left[\tilde{T}_{l,h} \geq c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2) \geq c_{l,h}(1-\delta)\right] \\
&+ P\left[\tilde{T}_{l,h} \geq c_{l,h}(1-\delta) \geq c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)\right] \\
&+ P\left[c_{l,h}(1-\delta) \geq \tilde{T}_{l,h} \geq c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)\right] \\
&\leq P\left[\tilde{T}_{l,h} \geq c_{l,h}(1-\delta)\right] + P\left[c_{l,h}(1-\delta) \geq c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)\right] \\
&= P\left[\tilde{T}_{l,h} \geq c_{l,h}(1-\delta)\right] + P\left[h_1 \notin CI_{\alpha-\delta}(\tilde{h}_1)\right] \\
&= \delta + (\alpha-\delta) = \alpha, \qquad \text{(A.22)}
\end{aligned}
$$

where the inequality and the second equality follow from the form of $c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)$, and the third equality follows from the definition of $CI_{\alpha-\delta}(\tilde{h}_1)$. As (A.22) holds for any $h \in \mathscr{H}$, it is clear from (A.21) that $AsySz[c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})] \le \alpha$, as stated. $\square$

PROOF OF THEOREM **3.3**

As in Theorem **3.1**, we can show that there exists a sequence $\gamma_n \in \Gamma$, a subsequence $\{m_n : n \ge 1\}$ of $\{n : n \ge 1\}$, and a subsubsequnce $\{\omega_n : n \ge 1\}$ of $\{m_n : n \ge 1\}$ such that the following result holds for $l \in \{1,2\}$:

$$\begin{aligned}
& AsySz\left[c_l^{B-A}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})\right] \\
= \;& \limsup_{n\to\infty} \sup_{\gamma\in\Gamma} P_{\theta_0,\gamma}\left[T_{l,n}(\theta_0) > c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2}) + \hat{\eta}_{l,n}\right] \\
= \;& \limsup_{n\to\infty} P_{\theta_0,\gamma_n}\left[T_{l,n}(\theta_0) > c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2}) + \hat{\eta}_{l,n}\right] \\
= \;& \lim_{n\to\infty} P_{\theta_0,\gamma_{m_n}}\left[T_{l,m_n}(\theta_0) > c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{m_n,1},\hat{h}_{m_n,2}) + \hat{\eta}_{l,m_n}\right] \\
= \;& \lim_{n\to\infty} P_{\theta_0,\gamma_{\omega_n,h}}\left[T_{l,\omega_n}(\theta_0) > c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{\omega_n,1},\hat{h}_{\omega_n,2}) + \hat{\eta}_{l,\omega_n}\right] \qquad \text{(A.23)}
\end{aligned}$$

for some $h \in \mathscr{H}$. Furthermore, as in the proof of Theorem **3.1**, for any $h \in \mathscr{H}_h$, any subsequence $\{\omega_n : n \ge 1\}$ of $\{n : n \ge 1\}$, and any sequence $\{\gamma_{\omega_n,h} : n \ge 1\}$, we have $\left(T_{l,\omega_n}(\theta_0), \hat{h}_{\omega_n,1}\right) \to^d \left(\tilde{T}_{l,h}, \tilde{h}_1\right)$ jointly. Hence,

$$\begin{aligned}
& \lim_{n\to\infty} P_{\theta_0,\gamma_{\omega_n,h}}\left[T_{l,\omega_n}(\theta_0) > c_l^{B-S}(\alpha,\alpha-\delta,\hat{h}_{\omega_n,1},\hat{h}_{\omega_n,2}) + \hat{\eta}_{l,\omega_n}\right] \\
= \;& \sup_{h\in\mathscr{H}} P\left[\tilde{T}_{l,h} > c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2) + \bar{\eta}_l\right] \qquad \text{(A.24)} \\
\equiv \;& \sup_{h\in\mathscr{H}} P\left[\tilde{T}_{l,h} > c_l^{B-A}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)\right], \qquad \text{(A.25)}
\end{aligned}$$

where $\bar{\eta}_l = \inf\left\{\eta : \sup_{h_1\in\mathscr{H}_1} P\left[\tilde{T}_{l,h} > c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2) + \eta\right] \le \alpha\right\}$. For the simplicity of exposition, define the following asymptotic rejection probability:

$$NRP_l[h,\eta] \;\; \equiv \;\; P[\tilde{T}_{l,h} > c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2) + \eta]. \qquad \text{(A.26)}$$

It is clear from (A.23)-(A.26) that $AsySz[c_l^{B-A}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})] = \sup_{h\in\mathscr{H}} NRP_l[h,\bar{\eta}_l]$. Hence, it suffices to show that $\sup_{h\in\mathscr{H}} NRP_l[h,\bar{\eta}_l] = \alpha$ to establish Theorem **3.3**.

First, from the result of Theorem **3.1** and the definition of the size-correction criterion, it is clear that $\sup_{h\in\mathscr{H}} NRP_l[h,\bar{\eta}_l] \le \alpha$. We proceed to show that $\sup_{h\in\mathscr{H}} NRP_l[h,\bar{\eta}_l] < \alpha$ leads to contradiction. Assume that $\sup_{h\in\mathscr{H}} NRP_l[h,\bar{\eta}_l] < \alpha$ and define the function $K_l(\cdot) : \mathbb{R}_- \to [-\alpha,\ 1-\alpha]$ such that

$$K_l(x) = \sup_{h\in\mathscr{H}} NRP_l[h,x] - \alpha. \qquad \text{(A.27)}$$

Notice that given Assumption **3.2**, $NRP_l[h,\cdot]$ is continuous on $\mathbb{R}_-$. Therefore, the Maximum Theorem entails that $K_l(\cdot)$ is also continuous on $\mathbb{R}_-$. Moreover, we have[13]

$$K_l\left(-c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)\right) = \sup_{h\in\mathscr{H}} NRP_l[h,-c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)]-\alpha = 1-\alpha > 0$$

and $K_l(\bar{\eta}_l) = \sup_{h\in\mathscr{H}} NRP_l[h,\bar{\eta}_l]-\alpha < 0$ (by assumption).

Then, we note that by the Intermediate Value Theorem, there exists $\dot{\eta}_l$ such that

i) $-c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2) < \dot{\eta}_l < \bar{\eta}_l$ almost surely,

ii) $K_l(\dot{\eta}_l)=0$; *i.e.*, $\sup_{h\in\mathscr{H}} NRP_l[h,\dot{\eta}_l]=\alpha$.

However, this contradicts the size-correction procedure where

$$\bar{\eta}_l = \inf\left\{\eta : \sup_{h_1\in\mathscr{H}_1} P\left[\tilde{T}_{l,h} > c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)+\eta\right] \leq \alpha\right\}.$$

It follows that $\sup_{h\in\mathscr{H}} NRP_l[h,\bar{\eta}_l]=\alpha$; i.e., $AsySz[c_l^{B-A}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})]=\alpha$. $\square$

PROOF OF COROLLARY **3.4** We notice that for $l\in\{1,2\}$,

$$\begin{aligned} & \liminf_{n\to\infty} \inf_{\gamma\in\Gamma} P_{\theta,\gamma}\left[\theta\in CS_{l,n}(1-\alpha)\right] \\ = \; & \liminf_{n\to\infty} \inf_{\gamma\in\Gamma} P_{\theta,\gamma}\left[T_{l,n}(\theta)\leq c_l^{B-A}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})\right], \end{aligned} \tag{A.28}$$

where $c_l^{B-A}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})$ denotes the BACV corresponding to $T_{l,n}(\theta)$. Then, the result follows by Theorem **3.3** and by exploiting the duality between confidence set and inverting the test of each of the individual null hypothesis $H_0:\theta=\theta_0$.

$\square$

[13]We notice that the proof is focused on the symmetric two-sided test and uses the fact that $NRP_l[h,-c_l^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)] = P[\tilde{T}_{l,h}>0]=1$ in this case. This proof can be adapted to the case of a lower/upper one-sided test by noting that for any $\varepsilon>0$ small enough, there exists a large enough positive constant $c\equiv c(\varepsilon)$ such that $NRP_l[h,-c(\varepsilon)]=1-\varepsilon$, for all $h\in\mathscr{H}$. Therefore, $K_l(-c(\varepsilon))=\sup_{h\in\mathscr{H}} NRP_l[h,-c(\varepsilon)]=1-\varepsilon-\alpha$. As this holds for any $\varepsilon>0$ small enough, the result for the case with lower/upper one-sided test follows by choosing ε such that $\varepsilon\to 0$.

Supplementary Material for

"Size-corrected Bootstrap Test after Pretesting for Exogeneity with Heteroskedastic or Clustered Data "

Firmin Doko Tchatoka[a] and Wenjie Wang[b]
[a]School of Economics, The University of Adelaide
[b]Division of Economics, School of Social Sciences, Nanyang Technological University

SA.1. Proofs for Lemmas A.1 and A.2

PROOF OF LEMMA **A.1** Note first that we can write $n^{1/2}(\hat{a}-\gamma_{n,h,1})$ as:

$$\begin{aligned} n^{1/2}(\hat{a}-\gamma_{n,h,1}) &= n^{1/2}\left((\hat{v}'M_X\hat{v})^{-1}\hat{v}'M_X\left((v-\hat{v}+\hat{v})\gamma_{n,h,1}+e\right)-\gamma_{n,h,1}\right) \\ &= \left(n^{-1}\hat{v}'M_X\hat{v}\right)^{-1}\left(n^{-1/2}\hat{v}'M_X(v-\hat{v})\right)\gamma_{n,h,1}+\left(n^{-1}\hat{v}'M_X\hat{v}\right)^{-1}\left(n^{-1/2}\hat{v}'M_Xe\right). \end{aligned} \tag{SA.29}$$

Therefore, to show the result of the lemma, it suffices to show that the first term in (SA.29) is $o_P(1)$. Note that

$$\begin{aligned} n^{-1/2}\hat{v}'M_X(v-\hat{v}) &= n^{-1/2}\hat{v}'M_XZ(Z'Z)^{-1}Z'v = (n^{-1}\hat{v}'M_XZ)(n^{-1}Z'Z)^{-1}(n^{-1/2}Z'v) \\ &= O_P(1)O_P(1)O_P(1) = O_P(1), \end{aligned} \tag{SA.30}$$

which follows from the fact that

$$n^{-1}\hat{v}'M_XZ = n^{-1}\hat{v}'Z - n^{-1}\hat{v}'P_XZ, \tag{SA.31}$$

$$\begin{aligned} n^{-1}\hat{v}'Z &= n^{-1}(v+(\hat{v}-v))'Z = n^{-1}v'Z+n^{-1}(\hat{v}-v)'Z \\ &= n^{-1}v'Z+(\gamma_{n,h,21}-\hat{\pi})'(n^{-1}Z'Z) = O_P(n^{-1/2})+O_P(n^{-1/2})O_P(1) = O_P(n^{-1/2}), \end{aligned} \tag{SA.32}$$

$$\begin{aligned} n^{-1}\hat{v}'P_XZ &= n^{-1}v'P_XZ+n^{-1}(\hat{v}-v)'P_XZ = (n^{-1}v'Z\gamma_{n,h,21}+n^{-1}v'v)(n^{-1}X'X)^{-1}(n^{-1}X'Z)+ \\ & \quad n^{-1}(\hat{v}-v)'P_XZ = \frac{h_{21}'h_{24}h_{25}}{(h_{21}'h_{24}h_{21}+h_{25})}+O_P(n^{-1/2}), \end{aligned} \tag{SA.33}$$

which follows from $n^{-1}Z'v \to^P 0$, $n^{-1}Z'Z \to^P h_{24}$, $n^{-1}v'v \to^P h_{25}$, and $n^{-1}X'X \to^P h_{21}'h_{24}h_{21}+h_{25}$, respectively. The $O_P(n^{-1/2})$ term in the last equality of (SA.33) is justified by the fact that

$$n^{-1}(\hat{v}-v)'P_XZ = (\gamma_{n,h,21}-\hat{\pi})'(n^{-1}Z'X)(n^{-1}X'X)^{-1}(n^{-1}X'Z) = O_P(n^{-1/2}). \tag{SA.34}$$

Therefore, given that $n^{-1/2}\hat{v}'M_X(v-\hat{v}) = O_P(1)$ and $n^{1/2}\gamma_{n,h,1} \to h_1 \in R$, we have

$$\left(n^{-1}\hat{v}'M_X\hat{v}\right)^{-1}\left(n^{-1/2}\hat{v}'M_X(v-\hat{v})\right)\gamma_{n,h,1} = o_P(1), \tag{SA.35}$$

so that $n^{1/2}(\hat{a}-\gamma_{n,h,1}) = \left(n^{-1}\hat{v}'M_X\hat{v}\right)^{-1}\left(n^{-1/2}\hat{v}'M_Xe\right)+o_P(1)$, as stated.

$\square$

PROOF OF LEMMA **A.2** (a) It is sufficient to characterize the asymptotic distributions of estimators separately: (a1) $n^{1/2}\hat{a}$, (a2) $n^{1/2}(\hat{\theta}_{ols}-\theta)$, and (a3) $n^{1/2}(\hat{\theta}_{2sls}-\theta)$.

(a1) Asymptotic distribution of $n^{1/2}\hat{a}$. We know from Lemma **A.1** that $n^{1/2}(\hat{a}-\gamma_{n,h,1})$ is asymptotically equivalent to $\left(n^{-1}\hat{v}'M_X\hat{v}\right)^{-1}\left(n^{-1/2}\hat{v}'M_Xe\right)$, so we focus on characterizing the asymptotic distribution of the latter. First, note that for the denominator,

$$\begin{aligned} n^{-1}\hat{v}'M_X\hat{v} &= n^{-1}\hat{X}'M_X\hat{X} = n^{-1}\hat{X}'\hat{X} - n^{-1}\hat{X}'P_Z\hat{X} \\ &\to^P h_{21}'h_{24}h_{21} - \frac{(h_{21}'h_{24}h_{21})^2}{(h_{21}'h_{24}h_{21}+h_{25})} = \frac{h_{21}'h_{24}h_{21}h_{25}}{(h_{21}'h_{24}h_{21}+h_{25})}, \end{aligned} \tag{SA.36}$$

where $\hat{X}=P_ZX$, the first equality follows from $\hat{v}=X-P_ZX$ and the convergence in probability follows from $n^{-1}\hat{X}'\hat{X}=n^{-1}X'P_ZX\to^P h_{21}'h_{24}h_{21}$, $n^{-1}\hat{X}'P_X\hat{X}=(n^{-1}\hat{X}'X)(n^{-1}X'X)^{-1}(n^{-1}X'\hat{X})\to^P \frac{(h_{21}'h_{24}h_{21})^2}{(h_{21}'h_{24}h_{21}+h_{25})}$. Second, note that for the numerator,

$$n^{-1/2}\hat{v}'M_Xe = -n^{-1/2}\hat{X}'M_Xe = -n^{-1/2}\hat{X}'e + n^{-1/2}\hat{X}'P_Xe. \tag{SA.37}$$

By applying Lyapunov Central Limit Theorem (CLT), we find for the first term in (SA.37),

$$-n^{-1/2}\hat{X}'e = -\left(n^{-1}X'Z\right)\left(n^{-1}Z'Z\right)^{-1}\left(n^{-1/2}Z'e\right)\to^d -h_{21}'\psi_{Ze}, \tag{SA.38}$$

and the second term is such that

$$\begin{aligned} n^{-1/2}\hat{X}'P_Xe &= \left(n^{-1}X'P_ZX\right)\left(n^{-1}X'X\right)^{-1}\left(n^{-1/2}X'e\right) \\ &\to^d (h_{21}'h_{24}h_{21}+h_{25})^{-1}h_{21}'h_{24}h_{21}\left(h_{21}'\psi_{Ze}+\psi_{ve}\right), \end{aligned} \tag{SA.39}$$

where ψ_{Ze} and ψ_{ve} are uncorrelated, $\psi_{Ze}\sim N(0,h_{22})$ and $\psi_{ve}\sim N(0,h_{23})$. Therefore,

$$\begin{aligned} -n^{-1/2}\hat{X}'M_Xe &\to^d -h_{21}'\psi_{Ze}+(h_{21}'h_{24}h_{21}+h_{25})^{-1}h_{21}'h_{24}h_{21}\left(h_{21}'\psi_{Ze}+\psi_{ve}\right) \\ &= -\frac{h_{25}}{(h_{21}'h_{24}h_{21}+h_{25})}h_{21}'\psi_{Ze}+\frac{h_{21}'h_{24}h_{21}}{(h_{21}'h_{24}h_{21}+h_{25})}\psi_{ve}. \end{aligned} \tag{SA.40}$$

By combining (SA.36) and (SA.40), we obtain

$$\begin{aligned} n^{1/2}(\hat{a}-\gamma_{n,h,1}) &\to^d -\left(h_{21}'h_{24}h_{21}\right)^{-1}h_{21}'\psi_{Ze}+h_{25}^{-1}\psi_{ve} \\ &\sim N\left(0,(h_{21}'h_{24}h_{21})^{-2}h_{21}'h_{22}h_{21}+h_{25}^{-2}h_{23}\right). \end{aligned} \tag{SA.41}$$

Since $n^{1/2}\hat{a}=n^{1/2}(\hat{a}-\gamma_{n,h,1})+n^{1/2}\gamma_{n,h,1}$, it follows that

$$n^{1/2}\hat{a} \to^d \psi_a = -\left(h_{21}'h_{24}h_{21}\right)^{-1}h_{21}'\psi_{Ze}+h_{25}^{-1}\psi_{ve}+h_1$$

$$\sim \quad N\Big(h_1, (h_{21}'h_{24}h_{21})^{-2}h_{21}'h_{22}h_{21} + h_{25}^{-2}h_{23}\Big). \quad \text{(SA.42)}$$

(a2) Asymptotic distribution of $n^{1/2}(\hat{\theta}_{OLS} - \theta)$. First, we have

$$n^{1/2}(\hat{\theta}_{OLS} - \theta) = \left(n^{-1}X'X\right)^{-1}\left(n^{-1/2}X'u\right), \quad \text{(SA.43)}$$

where $n^{-1}X'X \to^P h_{21}'h_{24}h_{21} + h_{25}$, and

$$\begin{aligned} n^{-1/2}X'u &= n^{-1/2}(\gamma_{n,h,21}'Z' + v')(v\gamma_{n,h,1} + e) \\ &= \gamma_{n,h,21}'\left(n^{-1/2}Z'e\right) + \gamma_{n,h,21}'\left(n^{-1/2}Z'v\right)\gamma_{n,h,1} + n^{-1/2}v'e + \left(n^{-1}v'v\right)n^{1/2}\gamma_{n,h,1} \\ &\to^d h_{21}'\psi_{Ze} + \psi_{ve} + h_{25}h_1, \end{aligned} \quad \text{(SA.44)}$$

since $\gamma_{n,h,21}'(n^{-1/2}Z'v)\gamma_{n,h,1} = o_P(1)$, $n^{-1}(v'v) = h_{25} + o_P(1)$, and $n^{1/2}\gamma_{n,h,1} \to h_1$ as $n \to \infty$.

Therefore, we obtain

$$\begin{aligned} n^{1/2}(\hat{\theta}_{ols} - \theta) &\to^d \psi_{ols} = (h_{21}'h_{24}h_{21} + h_{25})^{-1}(h_{21}'\psi_{Ze} + \psi_{ve} + h_{25}h_1) \\ &\sim N\Big(\frac{h_{25}h_1}{h_{21}'h_{24}h_{21} + h_{25}}, \frac{h_{21}'h_{22}h_{21} + h_{23}}{(h_{21}'h_{24}h_{21} + h_{25})^2}\Big). \end{aligned} \quad \text{(SA.45)}$$

(a3) Asymptotic distribution of $n^{1/2}(\hat{\theta}_{2sls} - \theta)$. First, note that $n^{1/2}(\hat{\theta}_{2sls} - \theta) = \left(n^{-1}X'P_ZX\right)^{-1}\left(n^{-1/2}X'P_Zu\right)$ and it follows from the proofs above that $n^{-1}X'P_ZX \to^P h_{21}'h_{24}h_{21}$ and $n^{-1/2}X'P_Zu \to^d h_{21}'\psi_{Ze}$. Therefore, we have

$$n^{1/2}(\hat{\theta}_{2sls} - \theta) \to^d \psi_{2sls} = (h_{21}'h_{24}h_{21})^{-1}h_{21}'\psi_{Ze} \sim N\Big(0, (h_{21}'h_{24}h_{21})^{-2}h_{21}'h_{22}h_{21}\Big). \quad \text{(SA.46)}$$

(b) It also suffices to characterize the asymptotic distributions of each statistic separately. Below we first show that $n\hat{V}_{ols} \to^P \frac{h_{21}'h_{22}h_{21}+h_{23}}{\left(h_{21}'h_{24}h_{21}+h_{25}\right)^2}$, and $n\hat{V}_{2sls} \to^P \frac{h_{21}'h_{22}h_{21}}{(h_{21}'h_{24}h_{21})^2}$.

For $\hat{V}_{ols}$ we use the decomposition

$$\frac{\hat{V}_{ols}}{V_{ols}} - 1 = V_{ols}^{-1}\left(\hat{V}_{ols} - V_{ols}\right) = V_{ols}^{-1}\left(A_{ols,1} - 2A_{ols,2} + A_{ols,3}\right) + o_P(1), \quad \text{(SA.47)}$$

where $V_{ols} = n^{-2}Q_{ols}^{-1}\sum_{i=1}^n E_F[X_i^2u_i^2]Q_{ols}^{-1}$, $A_{ols,1} = n^{-2}Q_{ols}^{-1}\sum_{i=1}^n X_i^2u_i^2Q_{ols}^{-1} - n^{-2}Q_{ols}^{-1}\sum_{i=1}^n E_F[X_i^2u_i^2]Q_{ols}^{-1}$, $A_{ols,2} = n^{-2}Q_{ols}^{-1}\sum_{i=1}^n X_i^3u_i(\hat{\theta}_{ols} - \theta)Q_{ols}^{-1}$, $A_{ols,3} = n^{-2}Q_{ols}^{-1}\sum_{i=1}^n X_i^4(\hat{\theta}_{ols} - \theta)^2Q_{ols}^{-1}$, and $Q_{ols} = plim_{n\to\infty}n^{-1}X'X$. Thus, we need to show that $V_{ols}^{-1}A_{ols,m} = o_P(1)$, for $m = 1, 2, 3$.

For $m = 1$, we let $r_i = n^{-1}V_{ols}^{-1/2}Q_{ols}^{-1}X_iu_i$, and we have $E_F\left[\sum_{i=1}^n r_i^2 - 1\right] = E_F\left[V_{ols}^{-1}A_{ols,1}\right] = 0$.

Also define the truncated variable $q_i = r_i \mathbb{1}(|r_i| \leq \varepsilon)$ such that $r_i^2 = q_i^2 + r_i^2 \mathbb{1}(|r_i| > \varepsilon)$. Then,

$$E_F \left| \sum_{i=1}^n r_i^2 - 1 \right| \leq E_F \left| \sum_{i=1}^n \left(q_i^2 - E_F[q_i^2] \right) \right| + E_F \left| \sum_{i=1}^n \left(r_i^2 \mathbb{1}(|r_i| > \varepsilon) - E_F[r_i^2 \mathbb{1}(|r_i| > \varepsilon)] \right) \right|. \quad \text{(SA.48)}$$

by the triangle inequality. The first term is $o(1)$ because

$$Var_F \left[\sum_{i=1}^n q_i^2 \right] = \sum_{i=1}^n Var_F \left[q_i^2 \right] \leq \varepsilon^2 \sum_{i=1}^n Var_F \left[|q_i| \right] \leq \varepsilon^2 \sum_{i=1}^n E_F \left[q_i^2 \right] \leq \varepsilon^2 \sum_{i=1}^n E_F \left[r_i^2 \right] = \varepsilon^2, \quad \text{(SA.49)}$$

where ε is arbitrary. For the second term, we have

$$\begin{aligned} E_F & \left| \sum_{i=1}^n \left(r_i^2 \mathbb{1}(|r_i| > \varepsilon) - E_F(r_i^2 \mathbb{1}(|r_i| > \varepsilon)) \right) \right| \leq 2 \sum_{i=1}^n E_F \left[|r_i|^{2+\xi} |r_i|^{-\xi} \mathbb{1}(|r_i| > \varepsilon) \right] \\ & \leq 2\varepsilon^{-\xi} \sum_{i=1}^n E_F |r_i|^{2+\xi} \to 0, \end{aligned} \quad \text{(SA.50)}$$

where the result of convergence to zero holds by the moment restriction on $E_F[||Z_i e_i||^{2+\xi}]$, $E_F[|v_i e_i|^{2+\xi}]$, $E_F[||Z_i Z_i'||^{2+\xi}]$ and $E_F[|X_i|^{2(2+\xi)}]$, and by $V_{ols} = O(n^{-1})$. For $m = 3$, we have

$$|nA_{ols,3}| = n^{-1} Q_{ols}^{-2} (\hat{\theta}_{ols} - \theta)^2 \sum_{i=1}^n X_i^4 = o_P(1), \quad \text{(SA.51)}$$

where the second equality follows from the moment restriction on $E_F[|X_i|^{2(2+\xi)}]$. Therefore, we obtain that $V_{ols}^{-1} A_{ols,3} = o_P(1)$. For $m = 2$, by the Cauchy-Schwarz inequality,

$$\begin{aligned} \left| V_{ols}^{-1} A_{ols,2} \right| & \leq \left(V_{ols}^{-1} n^{-2} Q_{ols}^{-1} \sum_{i=1}^n X_i^2 u_i^2 Q_{ols}^{-1} \right)^{1/2} \left(V_{ols}^{-1} A_{ols,3} \right)^{1/2} \\ & = \left(1 + V_{ols}^{-1} A_{ols,1} \right)^{1/2} \left(V_{ols}^{-1} A_{ols,3} \right)^{1/2} = o_P(1), \end{aligned} \quad \text{(SA.52)}$$

so that the results follows from those for $m = 1$ and $m = 3$.

Similarly, for $\hat{V}_{2sls}$ we use the decomposition

$$\frac{\hat{V}_{2sls}}{V_{2sls}} - 1 = V_{2sls}^{-1} \left(A_{2sls,1} - 2A_{2sls,2} + A_{2sls,3} \right) + o_P(1), \quad \text{(SA.53)}$$

where $V_{2sls} = n^{-2} Q_{2sls}^{-1} \sum_{i=1}^n E_F \left[Z_i Z_i' u_i^2 \right] Q_{2sls}^{-1}$, $A_{2sls,1} = n^{-2} Q_{2sls}^{-1} \sum_{i=1}^n \left(Z_i Z_i' u_i^2 - E_F \left[Z_i Z_i' u_i^2 \right] \right) Q_{2sls}^{-1}$, $A_{2sls,2} = n^{-2} Q_{2sls}^{-1} \sum_{i=1}^n Z_i Z_i' X_i u_i (\hat{\theta}_{2sls} - \theta) Q_{2sls}^{-1}$, $A_{2sls,3} = n^{-2} Q_{2sls}^{-1} \sum_{i=1}^n Z_i Z_i' X_i^2 (\hat{\theta}_{2sls} - \theta)^2 Q_{2sls}^{-1}$, and $Q_{2sls} = plim_{n\to\infty} \left(n^{-1} X' P_Z X \right)^{-1} (n^{-1} X' Z)(n^{-1} Z' Z)^{-1}$. The result follows by using the same arguments as for $\hat{V}_{ols}$.

Then, it suffices to verify that $nV_{ols} \to^P \frac{h_{21}' h_{22} h_{21} + h_{23}}{\left(h_{21}' h_{24} h_{21} + h_{25} \right)^2}$, and $nV_{2sls} \to^P \frac{h_{21}' h_{22} h_{21}}{(h_{21}' h_{24} h_{21})^2}$, and the results of $T_{ols}(\theta)$ and $T_{2sls}(\theta)$ follow immediately from part (a) of the lemma.

Finally, for $\hat{V}_a$ we use the decomposition

$$\frac{\hat{V}_a}{V_a}-1=V_a^{-1}\left(A_{a,1}-2A_{a,2}+A_{a,3}+A_{a,4}\right)+o_P(1), \tag{SA.54}$$

where $V_a=n^{-2}Q_a^{-1}\sum_{i=1}^n E_F\left[\ell' S_i S_i'\ell\right]Q_a^{-1}$, $A_{a,1}=n^{-2}Q_a^{-1}\sum_{i=1}^n\left(\ell' S_i S_i'\ell-E_F\left[\ell' S_i S_i'\ell\right]\right)Q_a^{-1}$, $A_{a,2}=n^{-2}Q_a^{-1}\sum_{i=1}^n\tilde{v}_i^3 e_i(\hat{a}-a)Q_a^{-1}$, $A_{a,3}=n^{-2}Q_a^{-1}\sum_{i=1}^n\tilde{v}_i^4(\hat{a}-a)^2Q_a^{-1}$, $A_{a,4}=n^{-2}Q_a^{-1}\sum_{i=1}^n(\hat{\ell}'\hat{S}_i\hat{S}_i'\hat{\ell}-\ell' S_i S_i'\ell)Q_a^{-1}$, $\ell=(1,-\pi_v)'$, $\hat{\ell}=(1,-\hat{\pi}_v)'$, $\hat{\pi}_v=(X'X)^{-1}X'\hat{v}$, $S_i=(v_ie_i,X_ie_i)'$, $\hat{S}_i=(\hat{v}_ie_i,X_ie_i)'$, and

$$\begin{aligned}\pi_v &= plim_{n\to\infty}\hat{\pi}_v=h_{25}(h_{21}'h_{24}h_{21}+h_{25})^{-1},\\ Q_a &= plim_{n\to\infty}n^{-1}\tilde{v}'\tilde{v}=h_{25}h_{21}'h_{24}h_{21}/(h_{21}'h_{24}h_{21}+h_{25}).\end{aligned} \tag{SA.55}$$

Then, the arguments for $A_{a,1}, A_{a,2}$, and $A_{a,3}$ follows those for OLS and 2SLS, and we have $V_a^{-1}A_{a,4}=o_P(1)$ by standard arguments. Therefore, $\hat{V}_a/V_a-1=o_P(1)$ and now it suffice to find the probability limit of nV_a to establish the limiting distribution for H_n. Notice that

$$E_F\left[\ell' S_i S_i'\ell\right]=E_F\left[v_i^2e_i^2\right]-2\pi_vE_F\left[X_ie_i^2v_i\right]+\pi_v^2E_F\left[X_i^2e_i^2\right], \tag{SA.56}$$

where $E_F\left[v_i^2e_i^2\right]\to h_{23}$, $E_F\left[X_ie_i^2v_i\right]\to h_{23}$, and $E_F\left[X_i^2e_i^2\right]\to h_{21}'h_{22}h_{21}+h_{23}$. Then, we obtain from the expression of V_a, (SA.55), and (SA.56) that

$$nV_a=Q_a^{-1}n^{-1}\sum_{i=1}^n E_F\left[\ell' S_i S_i'\ell\right]Q_a^{-1}\to(h_{21}'h_{24}h_{21})^{-2}h_{21}'h_{22}h_{21}+h_{25}^{-2}h_{23}, \tag{SA.57}$$

so that $H_n\to^d\left(\frac{h_{21}'h_{22}h_{21}}{(h_{21}'h_{24}h_{21})^2}+h_{25}^{-2}h_{23}\right)^{-1}\left(-(h_{21}'h_{24}h_{21})^{-1}h_{21}'\psi_{Ze}+h_{25}^{-1}\psi_{ve}+h_1\right)^2$.

□

SA.2. Proofs for the Clustering Case

PROOF OF THEOREM **3.6**

The proofs are similar to those for the heteroskedastic case, so we will keep the exposition concise. First, similar to Lemma **A.2**, we have under the drift sequences of parameters $\{\gamma_{n,h}^c\}$ in (3.16) with $|h_1^c|<\infty$, the joint asymptotic distribution of the test statistics are as follows:

$$\begin{pmatrix}T_{2sls}^c(\theta_0)\\ T_{ols}^c(\theta_0)\\ H_n^c\end{pmatrix}\quad\to^d\quad\eta_h^c=\begin{pmatrix}\eta_{1,h}^c\\ \eta_{2,h}^c\\ \eta_{3,h}^c\end{pmatrix}$$

$$
\begin{aligned}
&= \begin{pmatrix} \left(h_{21}^{c'} h_{22}^{c} h_{21}^{c}\right)^{-1/2} h_{21}^{c'} \psi_{Ze}^{c} \\ (h_{21}^{c'} h_{22}^{c} h_{21}^{c} + h_{23}^{c})^{-1/2} \left(h_{21}^{c'} \psi_{Ze}^{c} + \psi_{ve}^{c} + h_{25}^{c} \dot{h}_{1}^{c}\right) \\ \left(\frac{h_{21}^{c'} h_{22}^{c} h_{21}^{c}}{(h_{21}^{c'} h_{24}^{c} h_{21}^{c})^{2}} + h_{23}^{c} h_{25}^{c-2}\right)^{-1} \left(-(h_{21}^{c'} h_{24}^{c} h_{21}^{c})^{-1} h_{21}^{c'} \psi_{Ze}^{c} + h_{25}^{c-1} \psi_{ve}^{c} + \dot{h}_{1}^{c}\right)^{2} \end{pmatrix} \\
T_{1,n}^{c}(\theta_0) \quad &\to^{d} \quad \tilde{T}_{1,h}^{c} = \left|\eta_{2,h}^{c} \mathbb{1}(\eta_{3,h}^{c} \leq \chi_{1,1-\beta}^{2}) + \eta_{1,h}^{c} \mathbb{1}(\eta_{3,h}^{c} > \chi_{1,1-\beta}^{2})\right|, \\
T_{2,n}^{c}(\theta_0) \quad &\to^{d} \quad \tilde{T}_{2,h}^{c} = \left|\eta_{2,h}^{c} w(\eta_{3,h}^{c}) + \eta_{1,h}^{c}(1 - w(\eta_{3,h}^{c}))\right|,
\end{aligned}
$$

where

$$
\dot{h}_{1}^{c} = h_{1}^{c} \left(\frac{h_{21}^{c'} h_{22}^{c} h_{21}^{c}}{(h_{21}^{c'} h_{24}^{c} h_{21}^{c})^{2}} + h_{23}^{c} h_{25}^{c-2}\right)^{1/2}, \text{ and } \begin{pmatrix} \psi_{Ze}^{c} \\ \psi_{ve}^{c} \end{pmatrix} \sim N\left(0, \begin{pmatrix} h_{22}^{c} & 0 \\ 0' & h_{23}^{c} \end{pmatrix}\right). \tag{SA.58}
$$

In particular, we let $U_{n,g} = \mu_n^{1/2} n^{-1} \left\{c_1' Z_g' e_g + c_2 v_g' e_g\right\}$, where c_1 denotes a k-dimensional vector and c_2 denotes a nonzero scalar, and check that the conditions of the Lyapunov CLT hold for $U_{n,g}$:

$$
\begin{aligned}
&(a) \quad E_F\left[U_{n,g}\right] = 0, && \text{(SA.59)} \\
&(b) \quad \sum_{g=1}^{g} E_F\left[U_{n,g}^{2}\right] = c_1' \left(\mu_n n^{-2} \sum_{g=1}^{G} E_F[Z_g' e_g e_g' Z_g]\right) c_1 + c_2^{2} \left(\mu_n n^{-2} \sum_{g=1}^{G} E_F[v_g' e_g e_g' v_g]\right) \\
&\qquad \to c_1' h_{22}^{c} c_1 + c_2^{2} h_{23}^{c}, && \text{(SA.60)} \\
&(c) \quad \text{For some } \xi > 0 \text{ and some large enough constant } C, \\
&\qquad \sum_{g=1}^{G} E_F\left[|U_{n,g}|^{2+\xi}\right] = \left(\mu_n^{1/2} n^{-1}\right)^{2+\xi} \sum_{g=1}^{G} E_F\left[|c_1' Z_g' e_g + c_2 v_g' e_g|^{2+\xi}\right] \\
&\qquad \leq C \left(\mu_n^{1/2} n^{-1}\right)^{2+\xi} \sum_{g=1}^{G} E_F\left[|c_1' Z_g' e_g|^{2+\xi} + [|c_2 v_g' e_g|^{2+\xi}\right] \\
&\qquad = O\left(\mu_n^{1+\xi/2} n^{-1-\xi} \sup_g n_g^{1+\xi}\right), && \text{(SA.61)}
\end{aligned}
$$

where (SA.61) follows from Minkowski Inequality and

$$
\begin{aligned}
\sum_{g=1}^{G} E_F\left[|c_1' Z_g' e_g|^{2+\xi}\right] &= O\left(\sum_{g=1}^{G} n_g^{2+\xi}\right) = O\left(n \sup_g n_g^{1+\xi}\right), \\
\sum_{g=1}^{G} E_F\left[|c_2 v_g' e_g|^{2+\xi}\right] &= O\left(\sum_{g=1}^{G} n_g^{2+\xi}\right) = O\left(n \sup_g n_g^{1+\xi}\right),
\end{aligned} \tag{SA.62}
$$

as we can show that $\sup_g E_F\left[|c_1' Z_g' e_g|^{2+\xi}\right] = O\left(n_g^{2+\xi}\right)$ and $\sup_g E_F\left[|c_2 v_g' e_g|^{2+\xi}\right] = O\left(n_g^{2+\xi}\right)$, by using the arguments similar to those in the proof of Lemma A.2 of Djogbenou et al. (2019) and by using the moment restriction on $\sup_{g,i} E_F\left[||Z_{gi}e_{gi}||^{2+\xi}\right]$ and $\sup_{g,i} E_F\left[|v_{gi}e_{gi}|^{2+\xi}\right]$. Then, by Assumption **3.5**, we obtain that $\sum_{g=1}^{G} E_F\left[|U_{n,g}|^{2+\xi}\right] = o(1)$.

Furthermore, we show the consistency of the cluster-robust variance estimators as follows. For $\hat{V}^c_{ols}$, we use the decomposition

$$\frac{\hat{V}^c_{ols}}{V^c_{ols}} - 1 = V^{c-1}_{ols}\left(\hat{V}^c_{ols} - V^c_{ols}\right) = V^{c-1}_{ols}\left(A^c_{ols,1} - A^c_{ols,2} - A^{c'}_{ols,2} + A^c_{ols,3}\right) + o_P(1), \qquad \text{(SA.63)}$$

where $V^c_{ols} = n^{-2}Q^{-1}_{ols}\sum_{g=1}^{G} E_F[X_g'u_gu_g'X_g]Q^{-1}_{ols}$,

$$\begin{aligned} A^c_{ols,1} &= n^{-2}Q^{-1}_{ols}\sum_{g=1}^{G} X_g'u_gu_g'X_gQ^{-1}_{ols} - n^{-2}Q^{-1}_{ols}\sum_{g=1}^{G} E_F[X_g'u_gu_g'X_g]Q^{-1}_{ols}, \\ A^c_{ols,2} &= n^{-2}Q^{-1}_{ols}\sum_{g=1}^{G} X_g'u_g(\hat{\theta}^c_{ols} - \theta)X_g'X_gQ^{-1}_{ols}, \\ A^c_{ols,3} &= n^{-2}Q^{-1}_{ols}\sum_{g=1}^{G} X_g'X_g(\hat{\theta}^c_{ols} - \theta)^2X_g'X_gQ^{-1}_{ols}, \qquad \text{(SA.64)} \end{aligned}$$

and $Q_{ols} = plim_{n\to\infty}n^{-1}X'X$. Thus, we need to show that $V^{c-1}_{ols}A^c_{ols,m} = o_P(1)$, for $m = 1,2,3$.

For $m = 1$, we let $r_g = n^{-1}V^{c-1/2}_{ols}Q^{-1}_{ols}X_g'u_g$, and we have $E_F\left[\sum_{g=1}^{G} r_g^2 - 1\right] = E_F\left[V^{c-1}_{ols}A^c_{ols,1}\right] = 0$. Also define the truncated variable $q_g = r_g\mathbb{1}(|r_g| \leq \varepsilon)$ such that $r_g^2 = q_g^2 + r_g^2\mathbb{1}(|r_g| > \varepsilon)$. Then, by the triangle inequality,

$$E_F\left|\sum_{g=1}^{G} r_g^2 - 1\right| \leq E_F\left|\sum_{g=1}^{G}\left(q_g^2 - E_F[q_g^2]\right)\right| + E_F\left|\sum_{g=1}^{G}\left(r_g^2\mathbb{1}(|r_g| > \varepsilon) - E_F(r_g^2\mathbb{1}(|r_g| > \varepsilon))\right)\right| \qquad \text{(SA.65)}$$

The first term is $o_P(1)$ because

$$\begin{aligned} &Var_F\left(\sum_{g=1}^{G} q_g^2\right) = \sum_{g=1}^{G} Var_F\left(q_g^2\right) \leq \varepsilon^2\sum_{g=1}^{G} Var_F\left(|q_g|\right) \leq \varepsilon^2\sum_{g=1}^{G} E_F\left(q_g^2\right) \\ &\leq \varepsilon^2\sum_{g=1}^{G} E_F\left(r_g^2\right) = \varepsilon^2, \qquad \text{(SA.66)} \end{aligned}$$

where ε is arbitrary. For the second term, we have

$$\begin{aligned} &E_F\left|\sum_{g=1}^{G}\left(r_g^2\mathbb{1}(|r_g| > \varepsilon) - E_F[r_g^2\mathbb{1}(|r_g| > \varepsilon)]\right)\right| \leq 2\sum_{g=1}^{G} E_F\left[|r_g|^{2+\xi}|r_g|^{-\xi}\mathbb{1}(|r_g| > \varepsilon)\right] \\ &\leq 2\varepsilon^{-\xi}\sum_{g=1}^{G} E_F|r_g|^{2+\xi} \leq C\mu_n^{1+\xi/2}n^{-1-\xi}\sup_g n_g^{1+\xi} \to 0, \qquad \text{(SA.67)} \end{aligned}$$

where C is some large enough constant, the convergence to zero follows from Assumption **3.5**, and the last inequality follows from the fact that $V_{ols}^c = O\left(\mu_n^{-1}\right)$ and

$$\sum_{g=1}^{G} E|X_g' u_g|^{2+\xi} = O\left(n \sup_g n_g^{1+\xi}\right), \tag{SA.68}$$

since $\sup_g E_F\left[|X_g' u_g|^{2+\xi}\right] = O\left(n_g^{2+\xi}\right)$, by similar arguments as those in the proof of Lemma A.2 of Djogbenou et al. (2019) and the moment restriction on $\sup_{g,i} E_F\left[||Z_{gi} e_{gi}||^{2+\xi}\right]$, $\sup_{g,i} E_F\left[|v_{gi} e_{gi}|^{2+\xi}\right]$, $\sup_{g,i} E_F\left[||Z_{gi} Z_{gi}'||^{2+\xi}\right]$, and $\sup_{g,i} E_F\left[|X_{gi}|^{2(2+\xi)}\right]$.

For $m = 3$, we have

$$|\mu_n A_{ols,3}^c| = \mu_n n^{-2} Q_{ols}^{-2} (\hat{\theta}_{ols}^c - \theta)^2 \sum_{g=1}^{G} (X_g' X_g)^2 = O_P\left(\mu_n n^{-2} \sup_g n_g^2\right) = o_P(1), \tag{SA.69}$$

where the second equality follows from the moment restriction on $\sup_{g,i} E_F\left[|X_{g,i}|^{2(2+\xi)}\right]$, which implies that $\sup_g E_F\left[|X_g' X_g|^2\right] = O\left(n_g^2\right)$, and thus

$$\sum_{g=1}^{G} (X_g' X_g)^2 = O_P\left(\sum_{g=1}^{G} n_g^2\right) = O_P\left(n \sup_g n_g\right), \tag{SA.70}$$

and by using $|\hat{\theta}_{ols}^c - \theta| = O_P\left(V_{ols}^{c1/2}\right) = O_P\left(n^{-1/2} \sup_g n_g^{1/2}\right)$. For $m = 2$, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
\left|V_{ols}^{c-1} A_{ols,2}^c\right| &\leq \left(V_{ols}^{c-1} n^{-2} Q_{ols}^{-1} \sum_{g=1}^{G} X_g' u_g u_g' X_g Q_{ols}^{-1}\right)^{1/2} \left(V_{ols}^{c-1} A_{ols,3}^c\right)^{1/2} \\
&= \left(1 + V_{ols}^{c-1} A_{ols,1}^c\right)^{1/2} \left(V_{ols}^{c-1} A_{ols,3}^c\right)^{1/2},
\end{aligned} \tag{SA.71}$$

so that the results follows from those for $m = 1$ and $m = 3$.

For $\hat{V}_{2sls}^c$, we use the decomposition

$$\frac{\hat{V}_{2sls}^c}{V_{2sls}^c} - 1 = V_{2sls}^{c-1}\left(A_{2sls,1}^c - A_{2sls,2}^c - A_{2sls,2}^{c'} + A_{2sls,3}^c\right) + o_P(1), \tag{SA.72}$$

where $V_{2sls}^c = n^{-2} Q_{2sls}^{-1} \sum_{g=1}^{G} E_F\left[Z_g' u_g u_g' Z_g\right] Q_{2sls}^{-1}$,

$$\begin{aligned}
A_{2sls,1}^c &= n^{-2} Q_{2sls}^{-1} \sum_{g=1}^{G} Z_g' u_g u_g' Z_g Q_{2sls}^{-1} - n^{-2} Q_{2sls}^{-1} \sum_{g=1}^{G} E_F\left[Z_g' u_g u_g' Z_g\right] Q_{2sls}^{-1}, \\
A_{2sls,2}^c &= n^{-2} Q_{2sls}^{-1} \sum_{g=1}^{G} Z_g' u_g (\hat{\theta}_{2sls} - \theta) X_g' Z_g Q_{2sls}^{-1},
\end{aligned}$$

$$A^c_{2sls,3} \;\; = \;\; n^{-2}Q^{-1}_{2sls}\sum_{g=1}^{G} Z'_g X_g(\hat{\theta}_{2sls}-\theta)^2 X'_g Z_g Q^{-1}_{2sls}, \qquad \text{(SA.73)}$$

and $Q_{2sls} = plim_{n\to\infty}\left(n^{-1}X'P_Z X\right)^{-1}(n^{-1}X'Z)(n^{-1}Z'Z)^{-1}$. The result follows by using similar arguments as those for $\hat{V}^c_{ols}$. In particular, we have $\sum_{g=1}^{G} E_F\left[||Z'_g u_g||^{2+\xi}\right] = O\left(n\sup_g n_g^{1+\xi}\right)$ by the moment restriction on $\sup_{g,i} E_F\left[||Z_{gi}v_{gi}||^{2+\xi}\right]$ and $\sup_{g,i} E_F\left[||Z_{gi}e_{gi}||^{2+\xi}\right]$, and $\sum_{g=1}^{G}||Z'_g X_g||^2 = O_P\left(n\sup_g n_g\right)$ by the moment restriction on $\sup_{g,i} E_F\left[||Z_{gi}v_{gi}||^{2+\xi}\right]$ and $\sup_{g,i} E_F\left[||Z_{gi}Z'_{gi}||^{2+\xi}\right]$.

Additionally, for $\hat{V}^c_a$, we use the decomposition

$$\frac{\hat{V}^c_a}{V^c_a} - 1 = V^{c-1}_a\left(A^c_{a,1} - A^c_{a,2} - A^{c'}_{a,2} + A^c_{a,3} + A^c_{a,4}\right) + o_P(1), \qquad \text{(SA.74)}$$

where $V^c_a = n^{-2}Q^{-1}_a\sum_{g=1}^{G} E\left[\ell' S_g S'_g \ell\right]Q^{-1}_a$,

$$\begin{aligned}
A^c_{a,1} &= n^{-2}Q^{-1}_a\sum_{g=1}^{G}\ell' S_g S'_g \ell Q^{-1}_a - n^{-2}Q^{-1}_a\sum_{g=1}^{G}E_F\left[\ell' S_g S'_g \ell\right]Q^{-1}_a,\\
A^c_{a,2} &= n^{-2}Q^{-1}_a\sum_{g=1}^{G}\tilde{v}'_g e_g(\hat{a}^c - a)\tilde{v}'_g\tilde{v}_g Q^{-1}_a,\\
A^c_{a,3} &= n^{-2}Q^{-1}_a\sum_{g=1}^{G}\tilde{v}'_g\tilde{v}_g(\hat{a}^c - a)^2\tilde{v}'_g\tilde{v}_g Q^{-1}_a,\\
A^c_{a,4} &= n^{-2}Q^{-1}_a\sum_{g=1}^{G}\hat{\ell}'\hat{S}_g\hat{S}'_g\hat{\ell} Q^{-1}_a - n^{-2}Q^{-1}_a\sum_{g=1}^{G}\ell' S_g S'_g \ell Q^{-1}_a, \qquad \text{(SA.75)}
\end{aligned}$$

$\ell = (1, -\pi_v)'$, $\hat{\ell} = (1, -\hat{\pi}_v)'$, $\hat{\pi}_v = (X'X)^{-1}X'\hat{v}$, $\pi_v = plim_{n\to\infty}\hat{\pi}_v$, $S_g = (v'_g e_g, X'_g e_g)'$, $\hat{S}_g = (\hat{v}'_g e_g, X'_g e_g)'$, and $Q_a = plim_{n\to\infty} n^{-1}\tilde{v}'\tilde{v}$. Then, the arguments for $A^c_{a,1}, A^c_{a,2}$, and $A^c_{a,3}$ follows those for OLS and 2SLS, and we have $V^{c-1}_a A^c_{a,4} = o_P(1)$ by standard arguments.

Now, to show the results for the bootstrap analogues of the test statistics, we first show that under H_0 and the drift sequences of parameters $\{\gamma^c_{n,h}\}$ in (3.16) with $|h^c_1| < \infty$,

$$\begin{pmatrix} \mu_n^{1/2} n^{-1} Z' u^* \\ \mu_n^{1/2} n^{-1}\left(u^{*\prime} v^* - E^*\left[u^{*\prime} v^*\right]\right)\end{pmatrix} \to^{d^*} \begin{pmatrix}\psi^{c*}_{Ze} \\ \psi^{c*}_{ve}\end{pmatrix} \sim N\left(0, \begin{pmatrix} h^c_{22} & 0 \\ 0' & h^c_{23}\end{pmatrix}\right), \qquad \text{(SA.76)}$$

in probability P.

Let c_1 denote a k-dimensional nonzero vector and c_2 a nonzero scalar. Define

$$\begin{aligned}
U^*_{n,g} &= \mu_n^{1/2} n^{-1}\left\{c'_1 Z'_g u^*_g + c_2\left(u^{*\prime}_g v^*_g - E^*\left[u^{*\prime}_g v^*_g\right]\right)\right\}\\
&= \mu_n^{1/2} n^{-1}\left\{c'_1 Z'_g \hat{u}_g(\theta_0)\omega^*_{1g} + c_2\left(\hat{u}_g(\theta_0)\hat{v}_g\omega^*_{1g}\omega^*_{2g} - E^*\left[\hat{u}_g(\theta_0)\hat{v}_g\omega^*_{1g}\omega^*_{2g}\right]\right)\right\},
\end{aligned} \qquad \text{(SA.77)}$$

and it suffices to verify that the conditions of the Lyapunov CLT hold for $U^*_{n,g}$:

(a) $E^*\left[U^*_{n,g}\right]=0$ as $E^*\left[Z_g'\hat{u}_g(\theta_0)\omega^*_{1g}\right]=0$ and $E^*\left[\hat{u}_g'(\theta_0)\hat{v}_g\omega^*_{1g}\omega^*_{2g}-E^*[\hat{u}_g'(\theta_0)\hat{v}_g\omega^*_{1g}\omega^*_{2g}]\right]=0$.

(b) Note that

$$
\begin{aligned}
E^*\left[Z_g'u^*_g u^{*'}_g Z_g'\right] &= Z_g'\hat{u}_g(\theta_0)\hat{u}_g'(\theta_0)Z_g E^*\left[\omega^{*2}_{1g}\right]=Z_g'\hat{u}_g(\theta_0)\hat{u}_g'(\theta_0)Z_g,\\
E^*\left[u^{*'}_g v^*_g v^{*'}_g u^*_g\right] &= \hat{u}_g'(\theta_0)\hat{v}_g\hat{v}_g'\hat{u}_g(\theta_0)E^*\left[\omega^{*2}_{1g}\right]E^*\left[\omega^{*2}_{2g}\right]=\hat{u}_g'(\theta_0)\hat{v}_g\hat{v}_g'\hat{u}_g(\theta_0),\\
E^*\left[Z_g'u^*_g u^{*'}_g v^*_g\right] &= Z_g'\hat{u}_g(\theta_0)\hat{u}_g'(\theta_0)\hat{v}_g E^*\left[\omega^{*2}_{1g}\right]E^*\left[\omega^*_{2g}\right]=0.
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
\sum_{g=1}^{G}E^*\left[U^{*2}_{n,g}\right] &= c_1'\left(\mu_n n^{-2}\sum_{g=1}^{G}Z_g'\hat{u}_g(\theta_0)\hat{u}_g'(\theta_0)Z_g\right)c_1+c_2^2\left(\mu_n n^{-2}\sum_{g=1}^{G}\hat{u}_g'(\theta_0)\hat{v}_g\hat{v}_g'\hat{u}_g(\theta_0)\right)\\
&= c_1'\gamma^c_{n,h,22}c_1+c_2^2\gamma^c_{n,h,23}=c_1'h^c_{22}c_1+c_2^2h^c_{23}+o_P(1)=O_P(1). \qquad \text{(SA.78)}
\end{aligned}
$$

(c) For some $\xi>0$ and some large enough constant C_1, we note that by Minkowski Inequality,

$$
\sum_{g=1}^{G}E^*\left[\left|U^*_{n,g}\right|^{2+\xi}\right] \le C_1\left(\mu_n^{1/2}n^{-1}\right)^{2+\xi}\sum_{g=1}^{G}E^*\left[\left|c_1'Z_g'u^*_g\right|^{2+\xi}+\left|c_2u^{*'}_g v^*_g\right|^{2+\xi}\right]. \qquad \text{(SA.79)}
$$

Furthermore, notice that for some large enough constant C_2,

$$
\begin{aligned}
&\sum_{g=1}^{G}E^*\left[\left|c_1'Z_g'u^*_g\right|^{2+\xi}\right]=\sum_{g=1}^{G}E^*\left[\left|c_1'Z_g'\hat{u}_g(\theta_0)\omega^*_{1g}\right|^{2+\xi}\right]\\
&\le C_2\sum_{g=1}^{G}|c_1'Z_g'\hat{u}_g(\theta_0)|^{2+\xi}=O_P\left(n\sup_g n_g^{1+\xi}\right), \qquad \text{(SA.80)}
\end{aligned}
$$

where the inequality follows from the moment restriction on $E^*\left[|\omega^*_{1g}|^{2+\xi}\right]$. By similar argument,

$$
\sum_{g=1}^{G}E^*\left[|c_2u^{*'}_g v^*_g|^{2+\xi}\right]=O_P\left(n\sup_g n_g^{1+\xi}\right). \qquad \text{(SA.81)}
$$

Therefore, we have

$$
\sum_{g=1}^{G}E^*\left[\left|U^*_{n,g}\right|^{2+\xi}\right]=O_P\left(\mu_n^{1+\xi/2}n^{-1-\xi}\sup_g n_g^{1+\xi}\right)=o_P(1), \qquad \text{(SA.82)}
$$

where the first equality follows from (SA.79)-(SA.81) and the second equality follows from Assumption **3.5**.

Then, following similar steps as in the derivation for the bootstrap test statistics in the heteroskedastic case, we find that

$$\begin{aligned}
\mu_n^{1/2} n^{-1} X^{*'} P_Z u^* &= \hat{\pi}' \left(\mu_n^{1/2} n^{-1} Z' u^*\right) + \left(n^{-1} v^{*'} Z\right) \left(n^{-1} Z' Z\right)^{-1} \left(\mu_n^{1/2} n^{-1} Z' u^*\right) \\
&= \hat{\pi}' \left(\mu_n^{1/2} n^{-1} Z' u^*\right) + o_{P^*}(1) \to^{d^*} h_{21}^{c'} \psi_{Ze}^{c*}, \\
\mu_n^{1/2} n^{-1} X^{*'} u^* &= \hat{\pi}' \left(\mu_n^{1/2} n^{-1} Z' u^*\right) + \mu_n^{1/2} n^{-1} \left(u^{*'} v^* - E^*[u^{*'} v^*]\right) + \mu_n^{1/2} n^{-1} E^*[u^{*'} v^*] \\
&\to^{d^*} h_{21}^{c'} \psi_{Ze}^{c*} + \psi_{ve}^{c*}, \qquad \text{(SA.83)}
\end{aligned}$$

in probability P, where the last (conditional) convergence in distribution follows from $\mu_n^{1/2} n^{-1} E^*[u^{*'} v^*] = \mu_n^{1/2} n^{-1} \sum_{g=1}^{G} E^*[u_g^{*'} v_g^*] = \mu_n^{1/2} n^{-1} \sum_{g=1}^{G} E^*[\omega_{1g}^*] E^*[\omega_{2g}^*] \hat{u}_g'(\theta_0) \hat{v}_g = 0$ by the independent bootstrap scheme. Additionally, we find that

$$n^{-1} X^{*'} P_Z X^* \to^{P^*} h_{21}^{c'} h_{24}^{c} h_{21}^{c}, \quad n^{-1} X^{*'} X^* \to^{P^*} h_{21}^{c'} h_{24}^{c} h_{21}^{c} + h_{25}^{c}, \qquad \text{(SA.84)}$$

in probability P.

Furthermore, by using similar arguments as those for the consistency of the cluster-robust variance estimators, we can show the consistency of their bootstrap counterparts, i.e.,

$$\frac{\hat{V}_{2sls}^{c*}}{V_{2sls}^{c*}} - 1 \to^{P^*} 0, \quad \frac{\hat{V}_{ols}^{c*}}{V_{ols}^{c*}} - 1 \to^{P^*} 0, \quad \frac{\hat{V}_{a}^{c*}}{V_{a}^{c*}} - 1 \to^{P^*} 0, \qquad \text{(SA.85)}$$

in probability P, where $V_{2sls}^{c*} = n^{-2} Q_{2sls}^{-1} \sum_{g=1}^{G} E^*[Z_g' u_g^* u_g^{*'} Z_g] Q_{2sls}^{-1}$, $V_{ols}^{c*} = n^{-2} Q_{ols}^{-1} \sum_{g=1}^{G} E^*[X_g^{*'} u_g^* u_g^{*'} X_g^*] Q_{ols}^{-1}$, and $V_a^{c*} = n^{-2} Q_a^{-1} \sum_{g=1}^{G} E^*[\ell' S_g^* S_g^{*'} \ell] Q_a^{-1}$ with $S_g^* = \left(v_g^{*'} e_g^*, X_g^{*'} e_g^*\right)'$.

Combining these arguments together, we obtain for $T_{2sls}^{c*}(\theta_0)$ that

$$\begin{aligned}
&\frac{\hat{\theta}_{2sls}^{c*} - \theta_0}{\hat{V}_{2sls}^{c*1/2}} = \frac{\mu_n^{1/2} (\hat{\theta}_{2sls}^{c*} - \theta_0)}{(\mu_n V_{2sls}^{c*})^{1/2}} \left(\frac{V_{2sls}^{c*}}{\hat{V}_{2sls}^{c*}}\right)^{1/2} = \frac{(n^{-1} X^{*'} P_Z X^*)^{-1} \mu_n^{1/2} n^{-1} X^{*'} P_Z u^*}{(\mu_n V_{2sls}^{c*})^{1/2}} \left(\frac{V_{2sls}^{c*}}{\hat{V}_{2sls}^{c*}}\right)^{1/2} \\
&\to^{d^*} (h_{21}^{c'} h_{22}^{c} h_{21}^{c})^{-1/2} h_{21}^{c'} \psi_{Ze}^{c}, \qquad \text{(SA.86)}
\end{aligned}$$

in probability P, where the (conditional) convergence in distribution follows from (SA.83)-(SA.85) and $\mu_n V_{2sls}^{c*} \to^{P} (h_{21}^{c'} h_{24}^{c} h_{21}^{c})^{-2} h_{21}^{c'} h_{22}^{c} h_{21}^{c}$. The (conditional) convergence in distribution of $T_{ols}^{c*}(\theta_0)$ and H_n^{c*} follows similar arguments. The result of the theorem then follows by using the same arguments as those in the proof of Theorem **3.1**.

$\square$

PROOF OF THEOREM **3.8** The proof of the thoerem follows the same arguments as those in the proof of Thoerem **3.3**, and is thus omitted.

$\square$

PROOF OF COROLLARY **3.9** The proof of the corollary follows the same arguments as those in the proof of Corollary **3.4**, and is thus omitted.

□

SA.3. Further Simulation Results

In this section, we report further simulation results for the finite-sample power performance. Figures SA.1-SA.2 report the results for the case with heteroskedastic errors with $\beta = 0.025$ and $\delta = 0.045$. Figures SA.3 - SA.4 report the results for the case with heteroskedastic errors with $\beta = 0.1$ and $\delta = 0.045$. Figures SA.5 - SA.6 report the results for the case with heteroskedastic errors with $\beta = 0.05$ and $\delta = 0.025$. Figures SA.7 - SA.8 report the results for the case with design (2) of clustering, where $n_1 = 20$ with $q_1 = 5$, $n_2 = 15$ with $q_2 = 10$, $n_3 = 10$ with $q_3 = 20$, and $n_4 = 5$ with $q_4 = 30$.

Figure SA.1(a): Power of wild bootstrap tests under heteroskedasticity ($\beta = 0.025, \delta = 0.045$) with $\phi = 2$

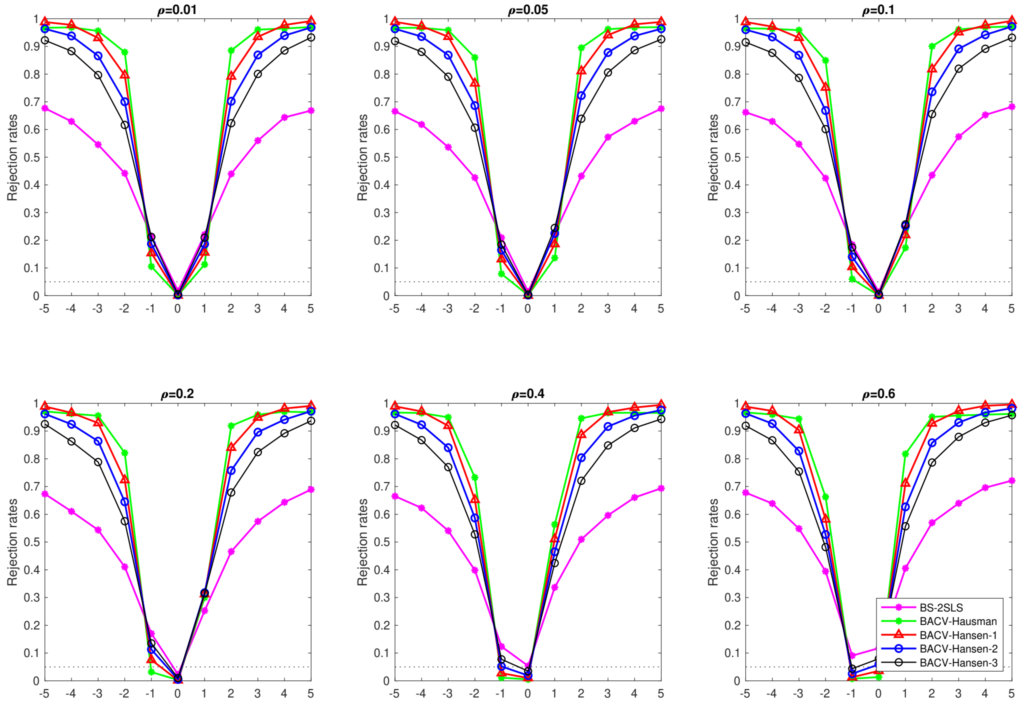
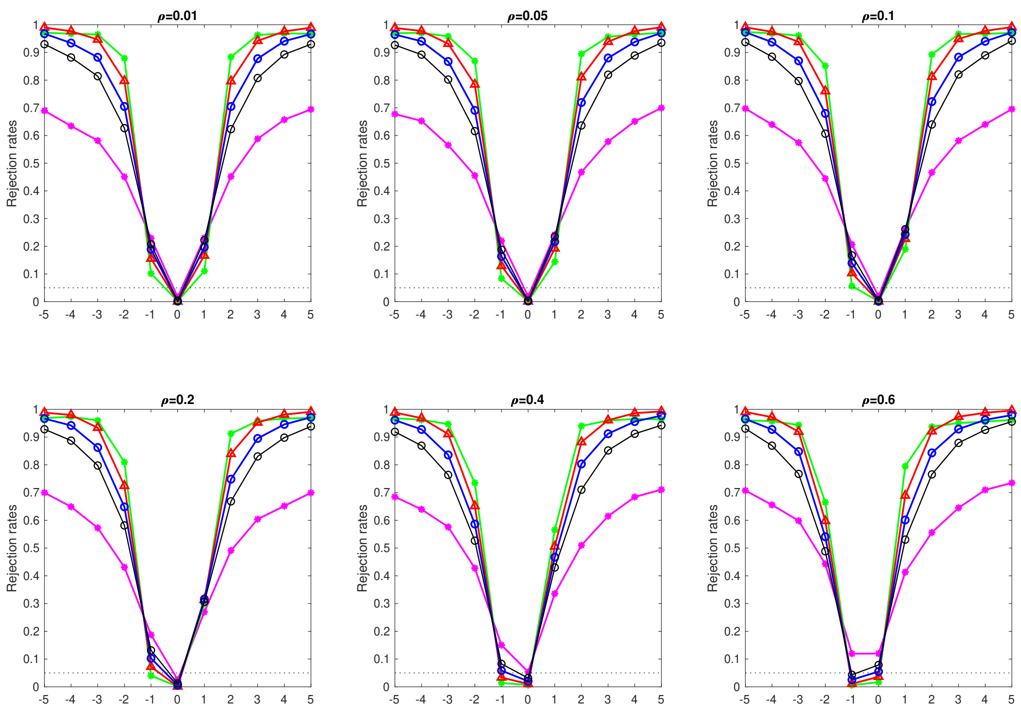


Figure SA.1(b): Power of wild bootstrap tests under heteroskedasticity ($\beta = 0.025, \delta = 0.045$) with $\phi = 4$



Notes: The power curves for the bootstrap 2SLS-t test, the two-stage test with hybrid-BACVs, and the shrinkage test with BACVs with $\tau = 1, 0.5, 0.25$ are illustrated by the curves in pink, green, red, blue, and black, respectively.

Figure SA.2(a): Power of wild bootstrap tests under heteroskedasticity ($\beta = 0.025, \delta = 0.045$) with $\phi = 16$

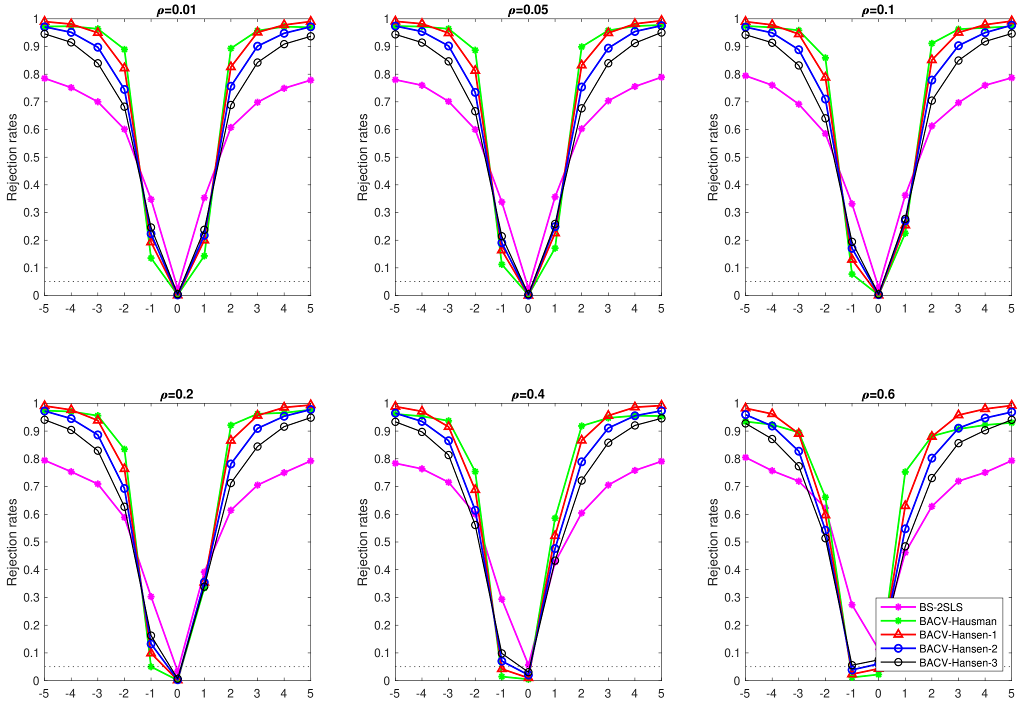
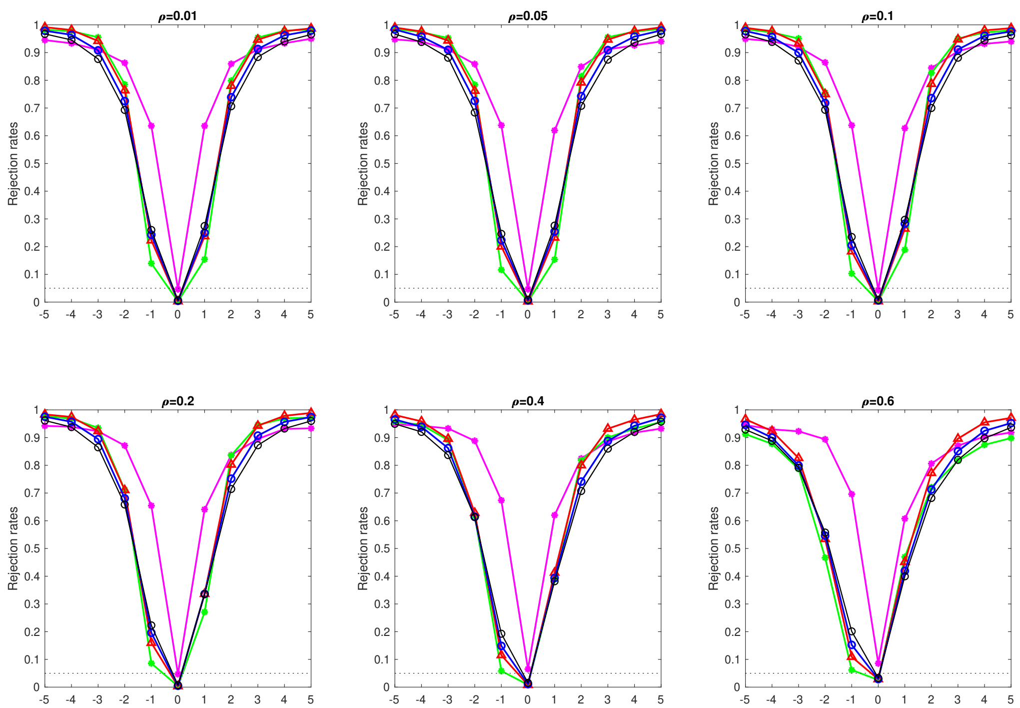


Figure SA.2(b): Power of wild bootstrap tests under heteroskedasticity ($\beta = 0.025, \delta = 0.045$) with $\phi = 64$



Notes: The power curves for the bootstrap 2SLS-t test, the two-stage test with hybrid-BACVs, and the shrinkage test with BACVs with $\tau = 1, 0.5, 0.25$ are illustrated by the curves in pink, green, red, blue, and black, respectively.

Figure SA.3(a): Power of wild bootstrap tests under heteroskedasticity ($\beta = 0.1, \delta = 0.045$) with $\phi = 2$

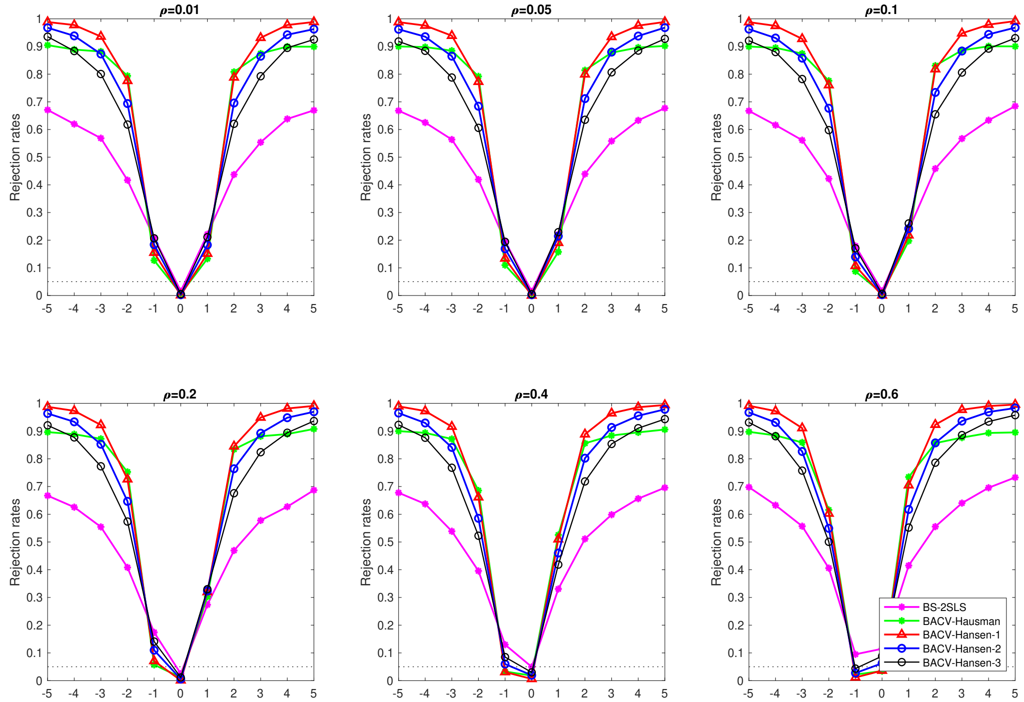
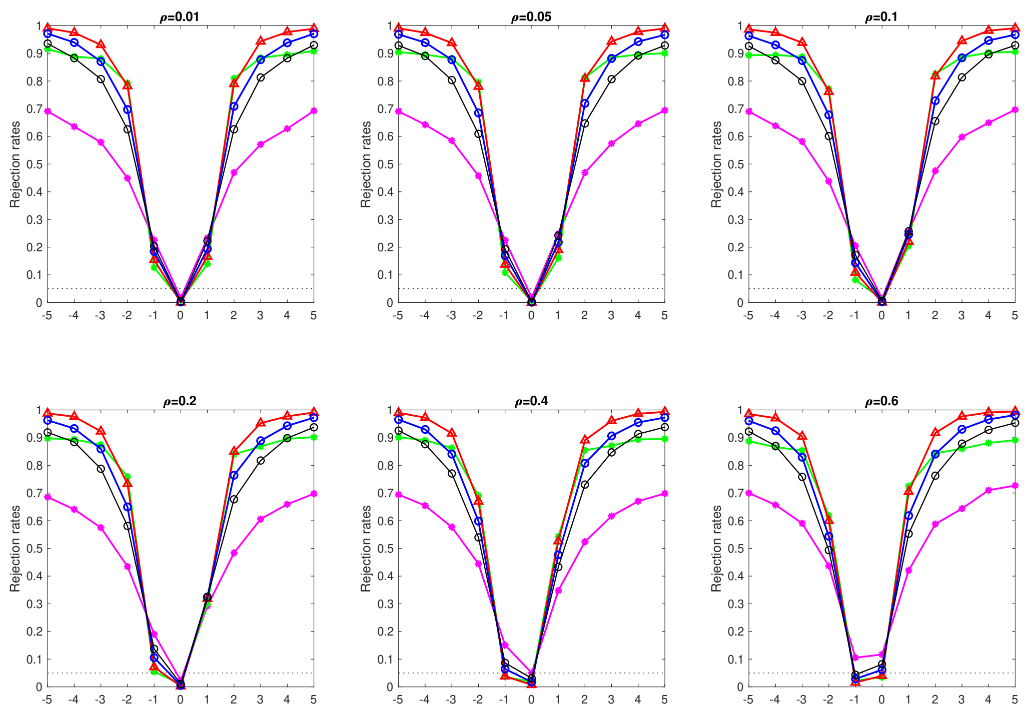


Figure SA.3(b): Power of wild bootstrap tests under heteroskedasticity ($\beta = 0.1, \delta = 0.045$) with $\phi = 4$



Notes: The power curves for the bootstrap 2SLS-t test, the two-stage test with hybrid-BACVs, and the shrinkage test with BACVs with $\tau = 1, 0.5, 0.25$ are illustrated by the curves in pink, green, red, blue, and black, respectively.

Figure SA.4(a): Power of wild bootstrap tests under heteroskedasticity ($\beta = 0.1, \delta = 0.045$) with $\phi = 16$

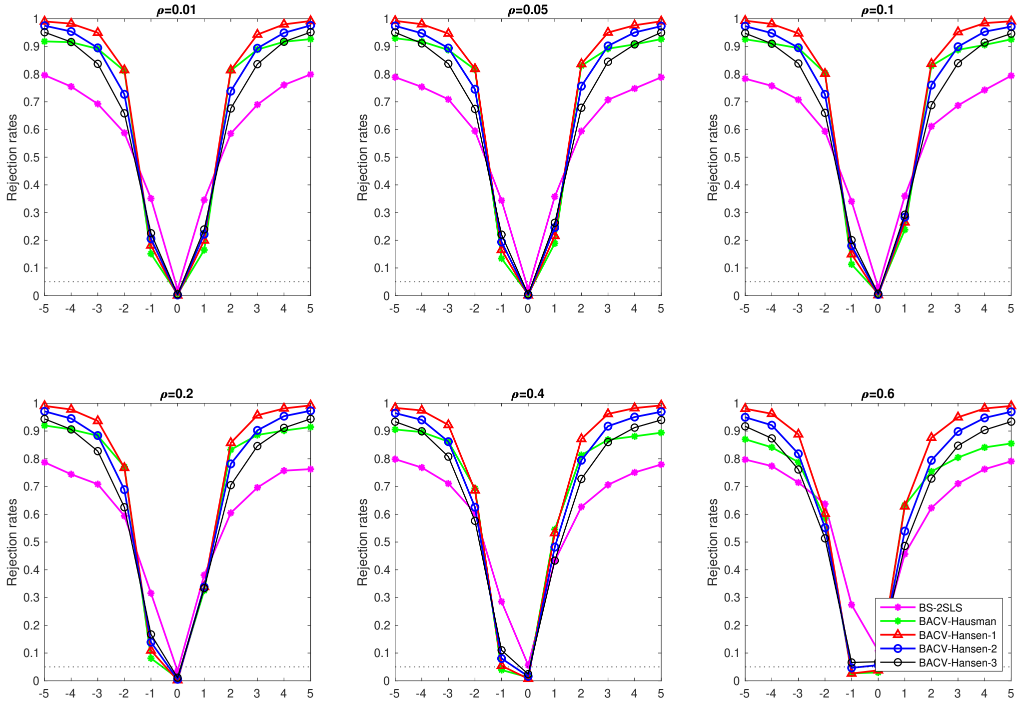
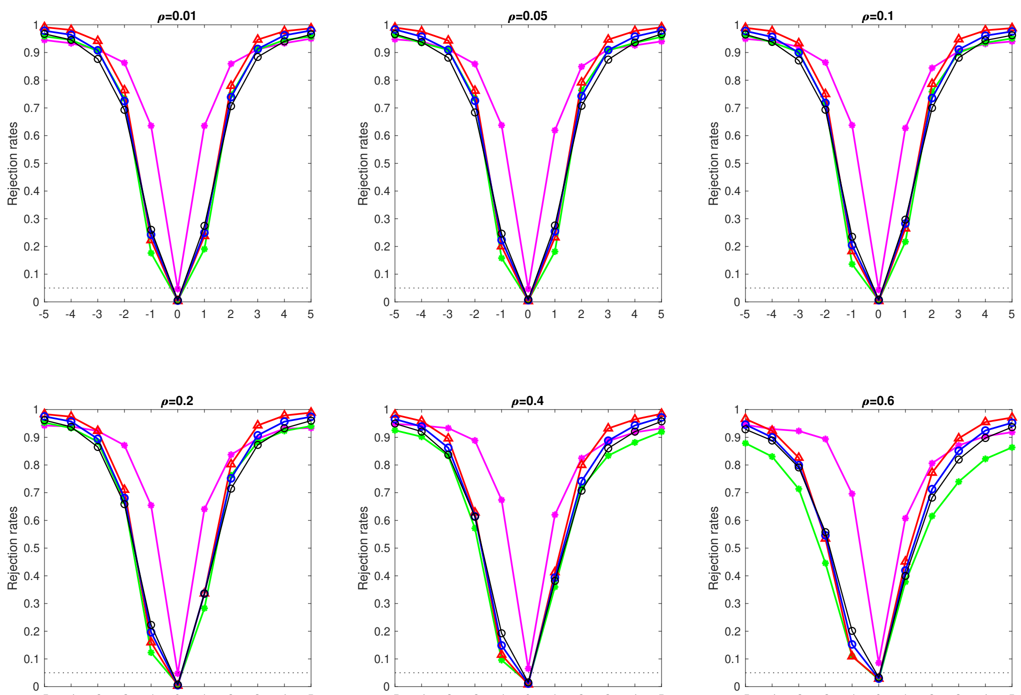


Figure SA.4(b): Power of wild bootstrap tests under heteroskedasticity ($\beta = 0.1, \delta = 0.045$) with $\phi = 64$



Notes: The power curves for the bootstrap 2SLS-t test, the two-stage test with hybrid-BACVs, and the shrinkage test with BACVs with $\tau = 1, 0.5, 0.25$ are illustrated by the curves in pink, green, red, blue, and black, respectively.

Figure SA.5(a): Power of wild bootstrap tests under heteroskedasticity ($\beta = 0.05, \delta = 0.025$) with $\phi = 2$

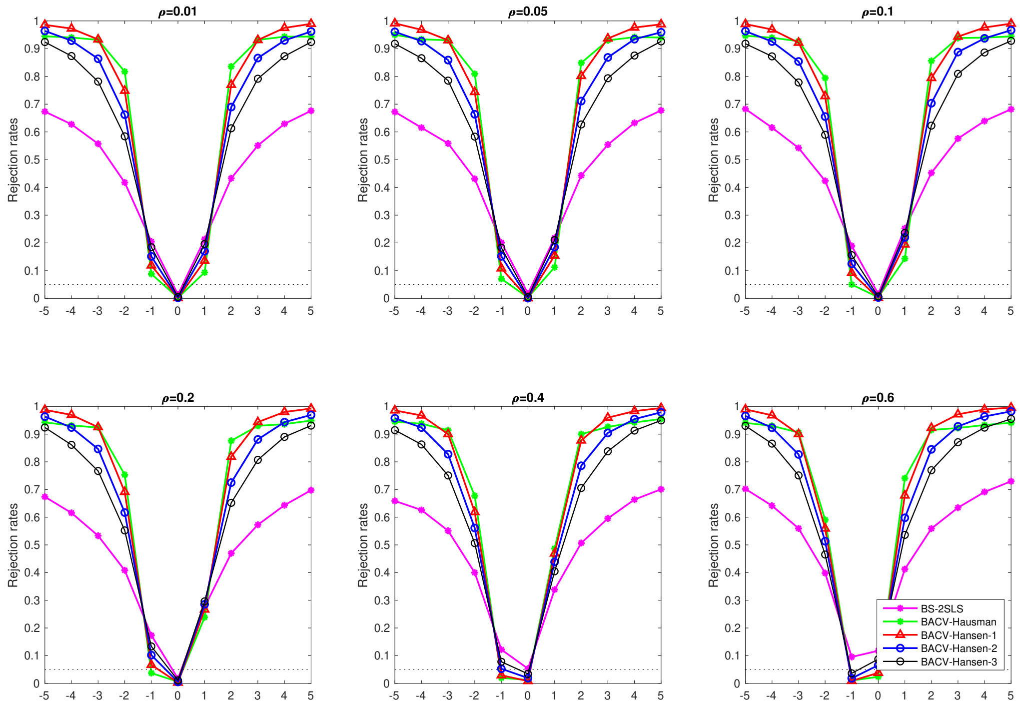
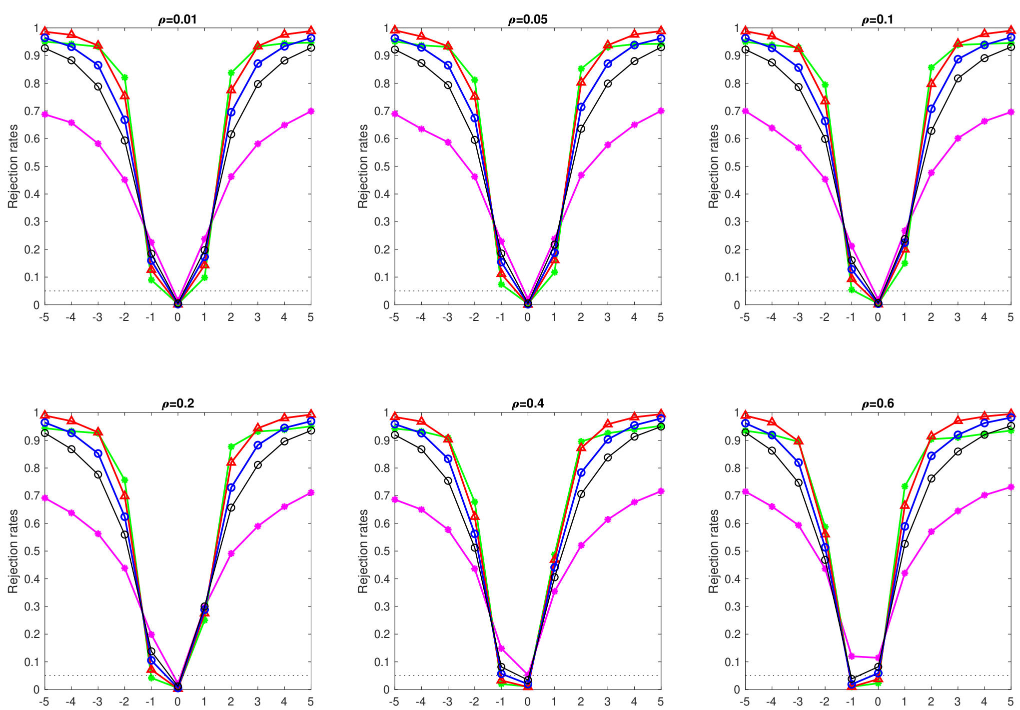


Figure SA.5(b): Power of wild bootstrap tests under heteroskedasticity ($\beta = 0.05, \delta = 0.025$) with $\phi = 4$



Notes: The power curves for the bootstrap 2SLS-t test, the two-stage test with hybrid-BACVs, and the shrinkage test with BACVs with $\tau = 1, 0.5, 0.25$ are illustrated by the curves in pink, green, red, blue, and black, respectively.

Figure SA.6(a): Power of wild bootstrap tests under heteroskedasticity ($\beta = 0.05, \delta = 0.025$) with $\phi = 16$

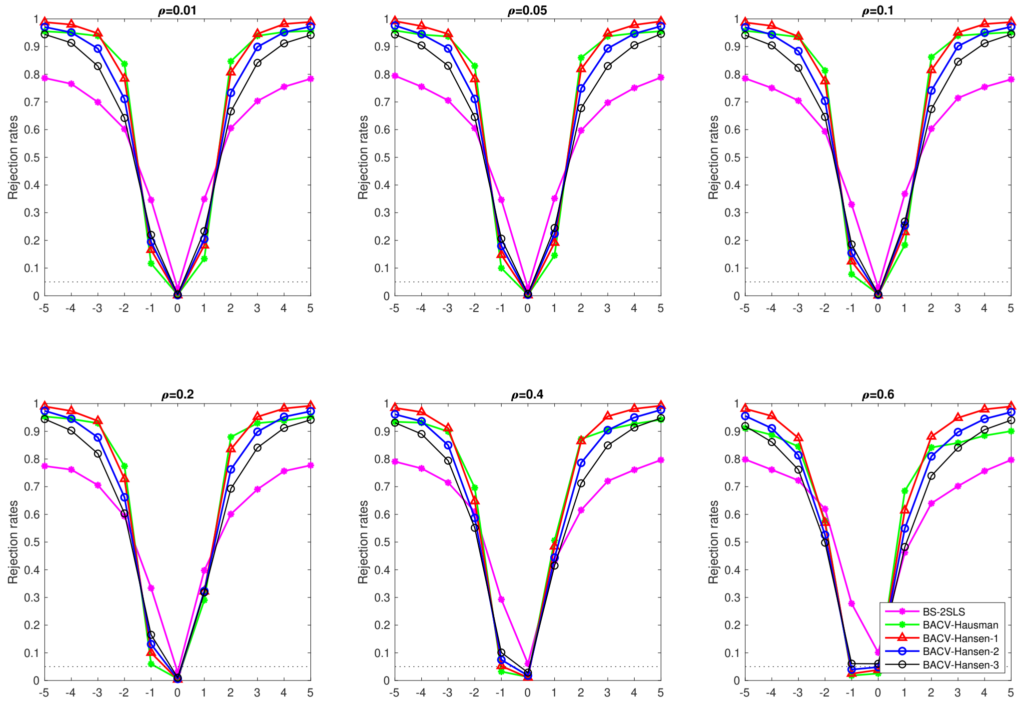
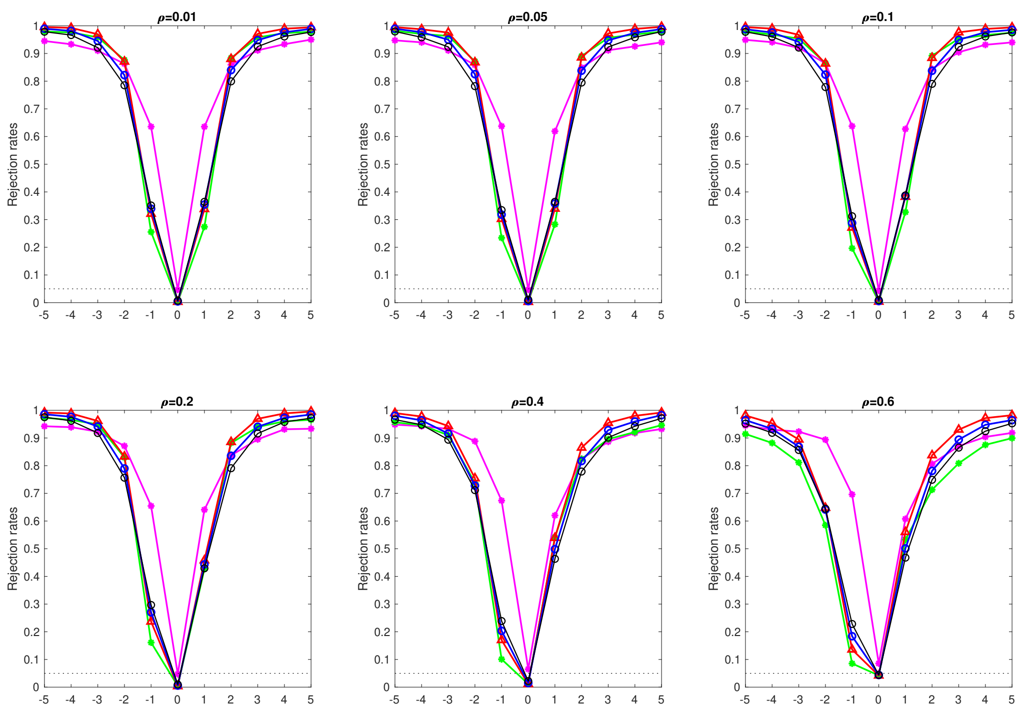


Figure SA.6(b): Power of wild bootstrap tests under heteroskedasticity ($\beta = 0.05, \delta = 0.025$) with $\phi = 64$



Notes: The power curves for the bootstrap 2SLS-t test, the two-stage test with hybrid-BACVs, and the shrinkage test with BACVs with $\tau = 1, 0.5, 0.25$ are illustrated by the curves in pink, green, red, blue, and black, respectively.

Figure SA.7(a): Power of wild bootstrap tests under clustering (design 2) with $\phi = 2$

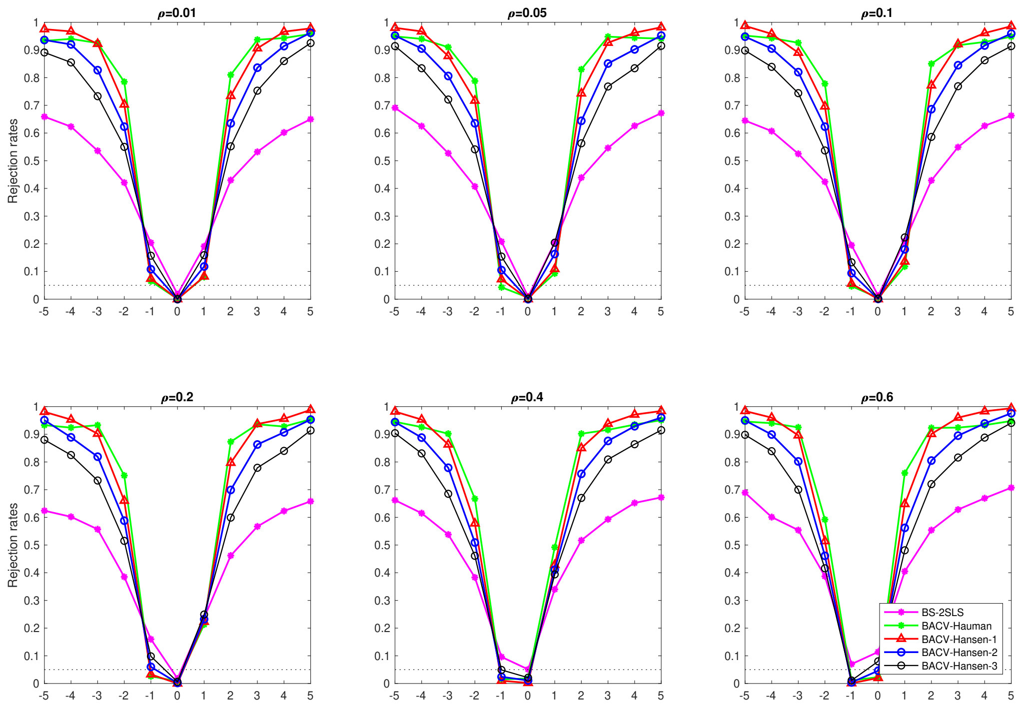
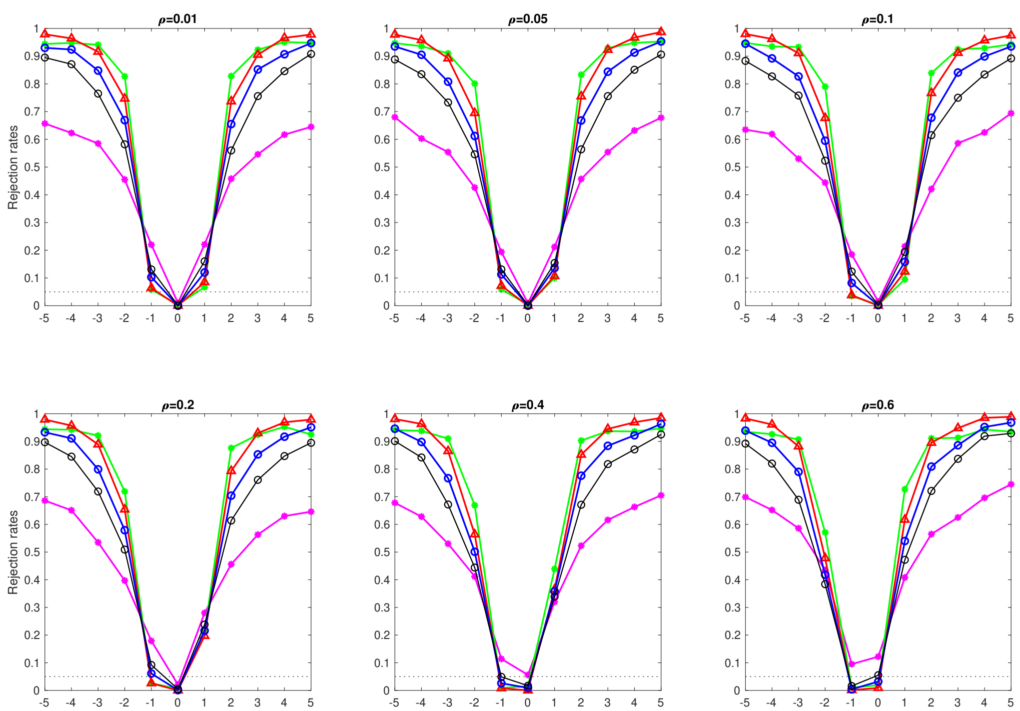


Figure SA.7(b): Power of wild bootstrap tests under clustering (design 2) with $\phi = 4$



Notes: The power curves for the bootstrap 2SLS-t test, the two-stage test with hybrid-BACVs, and the shrinkage test with BACVs with $\tau = 1, 0.5, 0.25$ are illustrated by the curves in pink, green, red, blue, and black, respectively.

Figure SA.8(a): Power of wild bootstrap tests under clustering (design 2) with $\phi = 16$

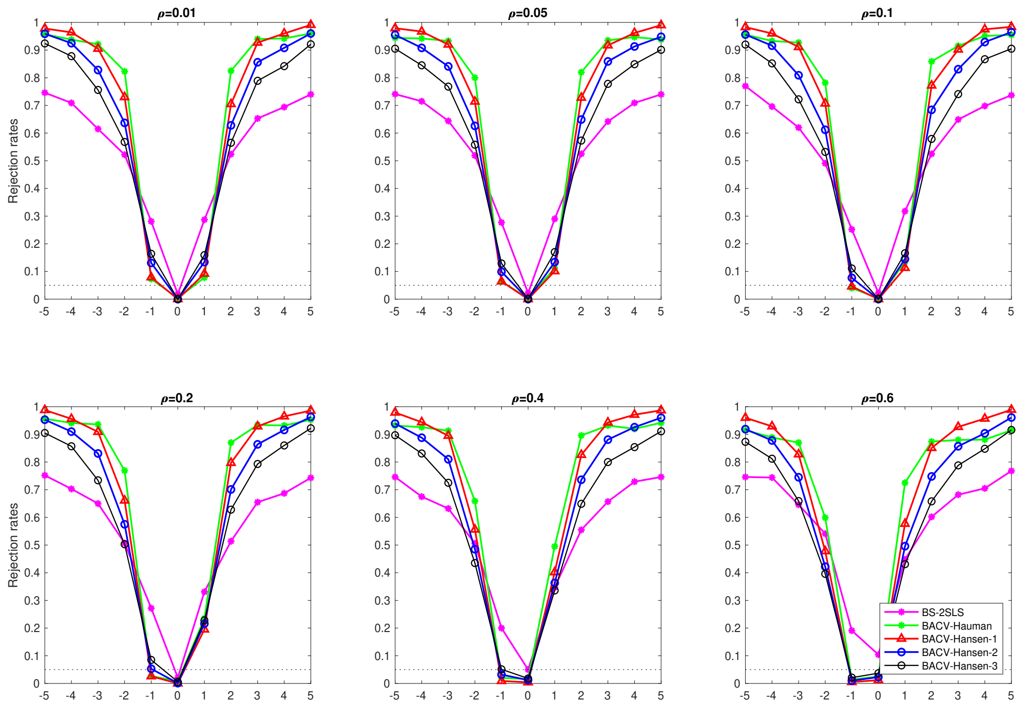
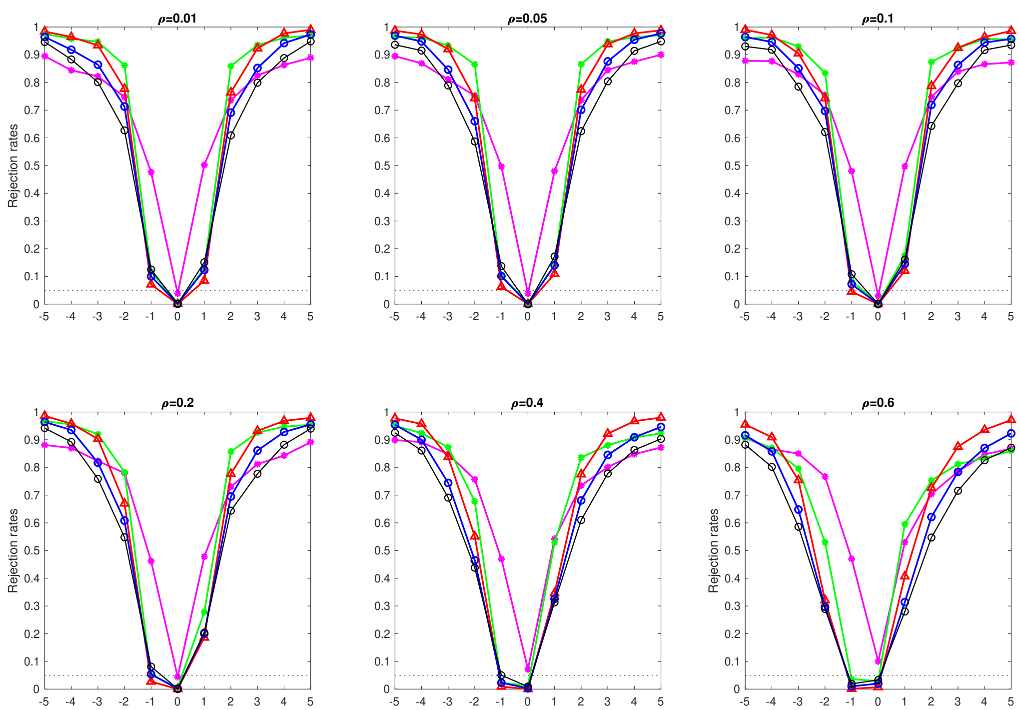


Figure SA.8(b): Power of wild bootstrap tests under clustering (design 2) with $\phi = 64$



Notes: The power curves for the bootstrap 2SLS-t test, the two-stage test with hybrid-BACVs, and the shrinkage test with BACVs with $\tau = 1, 0.5, 0.25$ are illustrated by the curves in pink, green, red, blue, and black, respectively.