Uniform Inference after Pretesting for Exogeneity with Heteroskedastic Data

Doko Tchatoka, Firmin and Wang, Wenjie

4 March 2021

Uniform Inference after Pretesting for Exogeneity with Heteroskedastic Data

Firmin Doko Tchatoka*
The University of Adelaide

Wenjie Wang †
Nanyang Technological University

March 5, 2021

ABSTRACT

Pretesting for exogeneity has become a routine in many empirical applications involving instrumental variables (IVs) to decide whether the ordinary least squares (OLS) 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 the standard wild bootstrap procedures (with either independent or dependent draws of disturbances) are not viable solutions to such extreme size-distortion problem. Then, we propose a novel hybrid bootstrap approach, which combines the wild bootstrap along with an adjusted Bonferroni size-correction method. We establish uniform validity of this hybrid bootstrap under conditional heteroskedasticity in the sense that it yields a two-stage test with correct asymptotic size. Monte Carlo simulations confirm our theoretical findings. In particular, our proposed hybrid method achieves remarkable power gains over the 2SLS-based t-test, especially when IVs are not very strong.

Key words: DWH Pretest; Instrumental Variable; Asymptotic Size; Bootstrap; Bonferroni-based Size-correction; Uniform Inference.

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 the statistical and econometric literature. For instance, see Hansen (2005), Leeb and Pötscher (2005a, 2009), 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 (2005b) 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, Kabaila and Leeb (2006) derive an upper bound for the large-sample limit minimal coverage probability of CIs after "conservative" model selection such as Akaike Information Criterion (AIC) and various pretesting procedures. 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 was studied by Guggenberger (2010a), who advised not to use Hausman-type pretesting for exogeneity 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 standard asymptotic CVs, even when IVs are strong.[1] Instead, Guggenberger (2010a) recommended to use a t-statistic based on the 2SLS estimator or, if weak IVs are a concern, an identification-robust method[2] to perform inference directly on the structural parameters. However, it is well known that the 2SLS-based t-statistic itself may have undesirable size properties when IVs are not strong (especially if the number of IVs is large), and compared with the t-statistic, identification-robust methods often yield relatively large confidence intervals in such cases. As such, in the quest for statistical power, many empirical practitioners still use Hausman-type pretesting in IV applications despite the important concern raised by Guggenberger (2010a). In particular, their motivation of implementing the two-stage procedure also lies in the fact that valid IVs (i.e., exogenous IVs) found in practice may be 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 standard t-test and identification-robust tests (e.g., see Berkowitz, Caner and Fang (2008, 2012), Doko Tchatoka and Dufour (2008), Conley, Hansen and Rossi (2012), Guggenberger (2012), Andrews, Gentzkow and Shapiro (2017)).

Recently, Young (2020) analyzes a sample of 1359 empirical application involving IV regres-

[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) and Kabaila, Mainzer and Farchione (2015) about pretesting for the presence of random effects before inference on the parameters of interest in panel data models.

[2] Such as Anderson and Rubin (1949, AR), Kleibergen (2002, KLM), and Moreira (2003, CLR) among others.

sions 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 rather 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 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 0.208, respectively, for bootstrap Hausman tests.

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 (2SLS is biased towards OLS under weak IVs), and Hausman-type tests also have low power in this case (e.g., see Doko Tchatoka and Dufour (2018, 2020)). It is therefore unclear whether OLS is not seriously biased in these data. In particular, as shown by Guggenberger (2010a) and Doko Tchatoka and Dufour (2018, 2020), the Hausman test is not able to reject the null hypothesis of exogeneity in situations where there is only a small degree of endogeneity: for sequences of correlations between the structural and reduced form errors that are local to zero of order $n^{-\frac{1}{2}}$ (i.e., local endogeneity), where n is the sample size, the Hausman pretest statistic has a noncentral chi-squared limiting distribution, and its noncentrality parameter is small when IVs strength is not high. Therefore, the pretest has low power and as a result, 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 (2010) shows that the asymptotic size of the naive two-stage test equals 1 for empirically relevant choices of parameter space.

In this paper, we study the possibility of proposing uniformly valid inference method for the two-stage test procedure by using alternative data-dependent CVs. Following Young (2020)'s recommendation of using bootstrap methods for IV models, we first study the validity of bootstrapping the two-stage procedure. It is well documented in the literature that resampling methods such as bootstrap and subsampling can be inconsistent when IVs are weak; see e.g., Andrews and Guggenberger (2010b), Wang and Doko Tchatoka (2018) and Wang (2020). Here, by deriving the null limiting distributions of the bootstrap test statistics and their associated asymptotic sizes, we show that the (wild) bootstrap method is invalid for the two-stage procedure even under strong IVs. In particular, the usual intuition for bootstrapping the Hausman test is that one should restrict the bootstrap data generating process (DGP) under exogeneity of the regressors, which corresponds to the pretest null hypothesis. Interestingly, we find that such bootstrap DGP can still result in extreme size distortion for the two-stage test with asymptotic size close to 1 in some settings, while the bootstrap DGP without the null restriction typically has much smaller size distortions. As such, in general bootstrap is not the solution to guarantee uniform inference for the two-stage test procedure. This is in contrast to the case of bootstrapping the Durbin-Wu-Hausman tests (without the second-stage t-test), which achieves higher-order refinement under strong IVs and remains first-order valid under weak IVs; e.g., see Doko Tchatoka (2015).

Here, to address the bootstrap failure for the two-stage test, we propose a novel hybrid bootstrap procedure, which makes use of the standard wild bootstrap CVs and an appropriate size-correction method. This procedure consists of developing a set of size-corrected bootstrap CVs for the two-stage test statistic, and we show that these CVs are uniformly valid in the sense that they yield tests with correct asymptotic size. In particular, since the standard wild bootstrap CVs cannot mimic well the key localized endogeneity parameter, more attention is required on this parameter when designing the bootstrap DGP. Furthermore, a Bonferroni-based size-correction technique is also implemented to deal with the presence of this localization parameter in the limiting distribution of interest. Different from conventional Bonferroni bound, which may lead to conservative test with asymptotic size strictly less than the nominal level, our technique always leads to correct asymptotic size.

In terms of practical usage of our method, we are particularly interested in the IV applications where the values of endogeneity parameters are relatively small; e.g., Hansen, Hausman and Newey (2008) report that the median of the estimated endogeneity parameters is only 0.279 in their investigated AER, JPE, and QJE papers. These are cases where the Hausman-type 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, Doko Tchatoka and Dufour (2020) show that pretest estimators based on Durbin-Wu-Hausman exogeneity tests can outperform both the 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 inference for this type of models. Monte Carlo experiments confirm that our hybrid bootstrap procedure is able to achieve remarkable power gains over the 2SLS-based t-test and the AR test, especially when the IVs are not very strong.

Our size-correction procedure follows the seminar studies by McCloskey (2017), who proposes Bonferroni-based size-correction procedures for general nonstandard testing problems, and McCloskey (2019) applied this method to post-selection inference in linear regression model. Wang and Doko Tchatoka (2018) proposed size-correction method for subvector inference in linear IV models in which the structural nuisance parameter may be weakly identified, while Han and McCloskey (2019) used it for inference in moment condition models where the estimating function may exhibit mixed identification strength and a nearly singular Jacobian. Different from our hybrid bootstrap procedures, these procedures are based on simulations from limiting distributions. The motivation of using bootstrap in the current two-stage testing problem originates from a growing literature illustrating that when applied to IV regressions (with possibly heteroskedastic or clustered errors), well designed bootstrap procedures typically have superior performance than conventional asymptotic approximations in terms of size control and coverage probability for hypothesis testing and confidence intervals, respectively; see, e.g., Davidson and MacKinnon (2008, 2010, 2014), Wang and Kaffo (2016), Finlay and Magnusson (2019), Young (2020), and Wang (2021).

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 results of both standard and hybrid bootstrap methods for the two-stage testing. Section 4 investigates the finite sample power performance of our methods using Monte Carlo simulations. Conclusions are drawn in Section 5 and the proofs are provided in the Appendix.

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. $B>0$ for a $m\times m$ squared matrix B means that B is positive definite. $\lambda_{min}(A)$, $\lambda_{max}(A)$, and $rank(A)$ denote the minimum and maximum eigenvalues and the rank of matrix A, respectively. $\|U\|$ denotes the usual Euclidean or Frobenius norm for a matrix U. The usual orders of magnitude are denoted by $O_P(.)$ and $o_P(.)$, "$\xrightarrow{P}$" stands for convergence in probability, while "$\xrightarrow{d}$" 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 $T^* \to^{P^*} 0$ in probability if for any $\delta>0$, $\varepsilon>0$, $lim_{n\to\infty}P[P^*(|T^*|>\delta)>\varepsilon]=0$, i.e., $P^*(|T^*|>\delta)=o_P(1)$; see e.g. Gonçalves and White (2004) and Dovonon and Gonçalves (2017). Also, we write $T^* = O_{P^*}(n^{\varphi})$ in probability if and only if for any $\delta>0$ there exists a $M_\delta<\infty$ such that $lim_{n\to\infty}P[P^*(|n^{-\varphi}T^*|>M_\delta)>\delta]=0$, 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 $T^*\to^{d^*} T$ in probability 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

$$y = X\theta + u, \tag{2.1}$$
$$X = Z\pi + v, \tag{2.2}$$

where $y, X \in R^n$, $Z\in\mathbb{R}^{n\times k}$ is a matrix of instruments ($k\geq 1$), $(\theta,\pi')'\in\mathbb{R}^{1\times k}$ are unknown parameters, and n is the sample size. We assume that the matrix Z has full-column rank with probability one. Denote by u_i, v_i, and Z_i the ith rows of u, v, and Z respectively, written as column vectors (or scalars) and similarly for the other random variables. Assume that $\{(u_i, v_i, Z_i) : i \leq n\}$ are i.i.d. with distribution F.

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., $E_F[u_i Z_i] = 0$ for all i, 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:

$$u = va + e, \tag{2.3}$$

where e is uncorrelated with v (i.e., $E_F[e_i v_i] = 0$ for all i) and $a = (E_F[v_i^2])^{-1} E_F[v_i u_i]$. The exogeneity of X in (2.1) can be assessed by testing the null hypothesis $H_a : \ a = 0$ in (2.3); see e.g. Doko Tchatoka and Dufour (2018). Substituting (2.3) into (2.1), we obtain the extended regression

$$y = X\theta + v\gamma + e, \tag{2.4}$$

where X and v are uncorrelated with e, i.e., $E_F[X_i e_i] = 0$ and $E_F[v_i e_i] = 0$ for all $i \leq n$. Therefore, the null hypothesis of exogeneity $H_a : \ a = 0$ can be assessed using a Wald statistic for $a = 0$ in the extended regression (2.4). For any full-column rank matrix A with appropriate sizes, let $M_A W$ denote the residuals from the regression of W on A, $M_A = I - P_A$ and $P_A = A(A'A)^{-1}A'$. To account for possible conditional heteroskedasticity in the model, we consider the following control function-based Wald statistic[3] for $H_a : \ a = 0$ in (2.4):

$$H_n = n\hat{a}^2 / \hat{V}_a, \tag{2.5}$$

where $\hat{a} = (\tilde{v}'\tilde{v})^{-1}\tilde{v}'y$, $\hat{V}_a = (n^{-1}\tilde{v}'\tilde{v})^{-1}\left(n^{-1}\sum_{i=1}^n \tilde{v}_i^2 \hat{e}_i^2\right)(n^{-1}\tilde{v}'\tilde{v})^{-1}$ is the White heteroskedasticity-robust estimator of the covariance matrix 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 residuals vector from the OLS regression of y on X and $\hat{v}$. The pretest reject the null hypothesis that X is exogenous in (2.1) (equivalently, the null hypothesis that OLS is unbiased) if $H_n > \chi^2_{1,1-\beta}$, where $\chi^2_{1,1-\beta}$ is the $(1-\beta)^{th}$ quantile of χ^2_1-distributed random variable for some $\beta \in (0,1)$. If θ is strongly identified (Z being strong instruments) and X is exogenous, H_n follows a χ^2_1 distribution asymptotically.

Let $\hat{\theta}_{2SLS} = X' P_Z y / (X' P_Z X)$, and $\hat{\theta}_{OLS} = X' y / (X' X)$ be the OLS and 2SLS estimators of θ

[3] 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 (2.5) easily accommodates conditional heteroskedasticity, so we shall use this formulation.

respectively in (2.1). Also, define

$$\begin{aligned}
\hat{V}_{2SLS} &= (n^{-1}X'P_ZX)^{-1}(X'Z)(Z'Z)^{-1}\left(n^{-1}\sum_{i=1}^{n} Z_iZ_i'\hat{u}_i^2(\hat{\theta}_{2SLS})\right)(Z'Z)^{-1}(Z'X)(n^{-1}X'P_ZX)^{-1}, \\
\hat{V}_{OLS} &= (n^{-1}X'X)^{-1}\left(n^{-1}\sum_{i=1}^{n} X_iX_i'\hat{u}_i^2(\hat{\theta}_{OLS})\right)(n^{-1}X'X)^{-1}, \qquad (2.6)
\end{aligned}$$

where $\hat{u}_i(l) = y_i - X_i\hat{\theta}_l$, $l \in \{OLS, 2SLS\}$. The two-stage test statistic associated with a pretest using H_n in the first-stage is given by:

$$\bar{T}_n(\theta_0) = T_{OLS}(\theta_0)\mathbb{1}(H_n \leq \chi^2_{1,1-\beta}) + T_{2SLS}(\theta_0)\mathbb{1}(H_n > \chi^2_{1,1-\beta}), \qquad (2.7)$$

where $T_l(\theta), l \in \{OLS, 2SLS\}$ is the usual t-statistic with OLS or 2SLS estimates, i.e.

$$T_l(\theta) = n^{1/2}(\hat{\theta}_l - \theta)/\hat{V}_l^{1/2},\ l \in \{OLS, 2SLS\}. \qquad (2.8)$$

Define $T_n(\theta_0)$ as $\pm\bar{T}_n(\theta_0)$ or $|\bar{T}_n(\theta_0)|$, depending on whether the test is a lower/upper one-sided or a symmetric two-sided test, respectively. The nominal size α test with a standard normal CV rejects $H_0 : \theta = \theta_0$ if

$$T_n(\theta_0) > c_\infty(1-\alpha), \qquad (2.9)$$

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.

2.2. Parameter space and asymptotic size

We define the parameter space Γ of the nuisance parameter vector γ following Andrews and Guggenberger (2009, 2010a, 2010b), Guggenberger (2012), and Guggenberger and Kumar (2012). Importantly, as pointed out in Andrews and Guggenberger (2009, 2010a, 2010b), one may index the model by nuisance parameters that have *three components*: $\gamma = (\gamma_1, \gamma_2, \gamma_3)$. (i) The first component γ_1 determines the point of discontinuity of the limiting distribution of interest. The parameter space of γ_1 is Γ_1. (ii) The second component γ_2 also affects the limiting distribution of interest, but does not affect the distance of the first component to the point of discontinuity. The parameter space of γ_2 is Γ_2. (iii) The third component γ_3 does not affect the limiting distribution. The parameter space of γ_3 is Γ_3, which in general may depend on γ_1 and γ_2, i.e., $\Gamma_3 \equiv \Gamma_3(\gamma_1, \gamma_2)$. To obtain the asymptotic size results, the first and second components need to be finite dimensional, while the third component is allowed to be infinite dimensional [e.g., the error distribution;[4] see

[4] As pointed out by Andrews and Guggenberger (2010b, p.434), the limiting distribution of interest often does not depend on the specific error distribution by virtue of the CLT, and only depends on whether it has certain moments finite and uniformly bounded.

the application examples in Andrews and Guggenberger (2009, 2010a, 2010b)].

Define the vector of nuisance parameters $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ by

$$\gamma_1 = a, \;\; \gamma_2 = (\gamma_{21}, \gamma_{22}, \gamma_{23}, \gamma_{24}, \gamma_{25}), \;\; \gamma_3 = F, \tag{2.10}$$

where $a = (E_F[v_i^2])^{-1}E_F[v_i u_i]$, $\gamma_{21} = \pi$, $\gamma_{22} = E_F e_i^2 Z_i Z_i'$, $\gamma_{23} = E_F e_i^2 v_i^2$, $\gamma_{24} = E_F Z_i Z_i'$, and $\gamma_{25} = E_F v_i^2$. Here, γ_1 measures the degree of endogeneity of X through the *regression endogeneity parameter a* [see Doko Tchatoka and Dufour (2014), and $\gamma_{25}^{-\frac{1}{2}}\|\gamma_{24}^{\frac{1}{2}}\gamma_{21}\|$ measures the overall strength of the IVs.[5] Let

$$\begin{aligned} \Gamma_1 = \mathbb{R}, \;\; \Gamma_2 = \Big\{ & (\gamma_{21}, \gamma_{22}, \gamma_{23}, \gamma_{24}, \gamma_{25}) : \gamma_{21} = \pi \in \mathbb{R}^k, \gamma_{22} = E_F e_i^2 Z_i Z_i' = \Omega_{Ze} \in \mathbb{R}^{k\times k}, \\ & \gamma_{23} = E_F e_i^2 v_i^2 = \Omega_{ve} \in \mathbb{R}, \gamma_{24} = E_F Z_i Z_i' = \Omega_{ZZ} \in \mathbb{R}^{k\times k}, \gamma_{25} = E_F v_i^2 = \sigma_v^2 \in \mathbb{R} \\ & s.t. \;\; \|\gamma_{21}\| \geq \underline{\kappa}, \lambda_{min}(\gamma_{22}) \geq \underline{\kappa}, \gamma_{23} > 0, \lambda_{min}(\gamma_{24}) \geq \underline{\kappa}, \text{ and } \gamma_{25} > 0 \Big\}, \end{aligned} \tag{2.11}$$

for some $\underline{\kappa} > 0$ that does not depend on n, where $\lambda_{min}(A)$ denotes the minimum eigenvalue of the square matrix A. As $\|\gamma_{21}\| \geq \underline{\kappa} > 0$, $\gamma_{25} \geq \underline{\kappa} > 0$, $\lambda_{min}(\gamma_{24}) \geq \underline{\kappa} > 0$, then the measure of instrument strength $\gamma_{25}^{-\frac{1}{2}}\|\gamma_{24}^{\frac{1}{2}}\gamma_{21}\|$ is bounded away from zero.[6] Following Guggenberger (2012) and Guggenberger and Kumar (2012), $\Gamma_3(\gamma_1, \gamma_2)$ is defined to allow for possible conditional heterogeneity of the errors as follows:

$$\begin{aligned} \Gamma_3(\gamma_1, \gamma_2) = \Big\{ & F : \; E_F e_i v_i = E_F e_i Z_i = E_F v_i Z_i = 0, \; E_F e_i^2 v_i Z_i = E_F e_i v_i^2 Z_i = E_F e_i v_i Z_i Z_i' = 0, \\ & E_F e_i v_i^3 = 0, \; E_F v_i^2 Z_i Z_i' = \Omega_{Zv} \in \mathbb{R}^{k\times k} \text{ with } \lambda_{min}(\Omega_{Zv}) \geq \underline{\kappa}, \\ & \left\| E_F\left(|u_i|^{2+\delta}, |v_i|^{2+\delta}, |u_i v_i|^{2+\delta}\right)' \right\| \leq M, \; \left\| E_F\left(||Z_i u_i||^{2+\delta}, ||Z_i v_i||^{2+\delta}, ||Z_i||^{2+\delta}\right)' \right\| \leq M \Big\}, \end{aligned} \tag{2.12}$$

for some constant $\underline{\kappa} > 0$, $\delta > 0$ and $M < \infty$. The definition of $\Gamma_3(\gamma_1, \gamma_2)$ allows for conditional heteroskedasticity by not imposing $E_F e_i^2 E_F Z_i Z_i' = \Omega_{Ze}$, $E_F v_i^2 E_F Z_i Z_i' = \Omega_{Zv}$, or $E_F e_i^2 E_F v_i^2 = \Omega_{ve}$. As the goal is to provide a practical means of controlling size of the two-stage t-test after pretesting for exogeneity, allowing for conditional heteroskedasticity is paramount for the proposed methodology to be useful in applied work.

We then define the whole nuisance parameter space Γ of γ as

$$\Gamma = \{\gamma = (\gamma_1, \gamma_2, \gamma_3) : \gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2, \gamma_3 \in \Gamma_3(\gamma_1, \gamma_2)\}, \tag{2.13}$$

where Γ_j, $j = 1, 2, 3$ are given in (2.11) and (2.12). To characterize the asymptotic behavior of the

[5] Note that $\gamma_{25}^{-\frac{1}{2}}\|\gamma_{24}^{\frac{1}{2}}\gamma_{21}\| = (\mu^2/n)^{1/2}$, where μ^2 denotes the well-known concentration parameter in the IV literature.

[6] Therefore, Staiger and Stock (1997)'s weak IVs asymptotics is ruled out of the scope of this paper. However, the Monte Carlo experiments (see Section 4) show that our proposed tests perform very well even when IVs are weak.

estimators and test statistics, it is useful to consider the following convergence in distribution:

$$\frac{1}{\sqrt{n}}\begin{pmatrix} Z'e \\ Z'v \\ v'e \end{pmatrix} \to^d \begin{pmatrix} \psi_{Ze} \\ \psi_{Zv} \\ \psi_{ve} \end{pmatrix} \sim N\left(0, \begin{pmatrix} \Omega_{Ze} & 0 & 0 \\ 0 & \Omega_{Zv} & 0 \\ 0 & 0 & \Omega_{ve} \end{pmatrix}\right). \tag{2.14}$$

Furthermore, let $\bar{\gamma}_1$ be defined as $\gamma_1 \to \bar{\gamma}_1 \in \mathbb{R}$ as $n \to \infty$. It is easy to see from (2.3) and (2.14) that

$$\frac{1}{\sqrt{n}}\begin{pmatrix} Z'u \\ Z'v \\ (v'u - E_F[v'u]) \end{pmatrix} \to^d \begin{pmatrix} \psi_{Zu}(\bar{\gamma}_1) \\ \psi_{Zv} \\ \psi_{vu}(\bar{\gamma}_1) \end{pmatrix} \sim N\left(0, \begin{pmatrix} \Omega_{Zu}(\bar{\gamma}_1) & 0 & 0 \\ 0 & \Omega_{Zv} & 0 \\ 0 & 0 & \Omega_{vu}(\bar{\gamma}_1) \end{pmatrix}\right), \tag{2.15}$$

where $\psi_{Zu}(\bar{\gamma}_1) = \psi_{Zv}\bar{\gamma}_1 + \psi_{Ze}$, $\psi_{vu}(\bar{\gamma}_1) = \psi_v\bar{\gamma}_1 + \psi_{ve}$ with $n^{-1/2}(v'v - E_F[v'v]) \to^d \psi_v$, $\Omega_{Zu}(\bar{\gamma}_1) = \bar{\gamma}_1^2\Omega_{Zv} + \Omega_{Ze}$, and $\Omega_{vu}(\bar{\gamma}_1) = \bar{\gamma}_1^2\Omega_{vv} + \Omega_{ve}$ with $\Omega_{vv} = E_F v_i^4 - \sigma_v^4$. Clearly, all the variables and covariance matrices in the limiting distribution in (2.15) are continuous functions in $\bar{\gamma}_1 \in \mathbb{R}$. It is also straightforward to see that $\psi_{Zu}(0) \equiv \psi_{Ze}$, $\psi_{vu}(0) \equiv \psi_{ve}$, $\Omega_{Zu}(0) \equiv \Omega_{Ze}$, and $\Omega_{vu}(0) \equiv \Omega_{ve}$.

Let c_n denote a (possibly data-dependent) CV being used for the two-stage testing. The finite sample null rejection probability (NRP) of the two-stage test evaluated at $\gamma \in \Gamma$ is given by $P_{\theta_0,\gamma}[T_n(\theta_0) > c_n]$, 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

$$\underset{n\to\infty}{limsup} P_{\theta_0,\gamma}[T_n(\theta_0) > c_n], \tag{2.16}$$

and the asymptotic size of the test is given by

$$AsySz[c_n] = \underset{n\to\infty}{limsup}\,\underset{\gamma\in\Gamma}{sup} P_{\theta_0,\gamma}[T_n(\theta_0) > c_n]. \tag{2.17}$$

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 the test statistic $T_n(\theta_0)$ 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 these papers, we can show that the asymptotic size of the two-stage test with the standard fixed normal CV (i.e., $AsySz[c_\infty(1-\alpha)]$) is realized under relevant choices of the parameter space. In particular, to derive $AsySz[c_\infty(1-\alpha)]$, it is enough to study the asymptotic NRP along some sequence of the type $\{\gamma_{n,h}\}$ for some $h \in \mathscr{H}$, as the highest asymptotic NRP is materialized among such sequence, where

$$\mathscr{H} = \Big\{ h = (h_1, h_{21}', vec(h_{22})', h_{23}, vec(h_{24})', h_{25})' \in \mathbb{R}_\infty^{2k^2+k+3} : \exists\{\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3}) \in \Gamma : n \geq 1\} :$$

$$
\begin{aligned}
& n^{1/2}\gamma_{n,1} \to h_1 \in R_\infty,\ \gamma_{n,2} \to h_2 = (h_{21},h_{22},h_{23},h_{24},h_{25}):\ (\|h_{21}\|,h_{23},h_{25}) \geq \underline{\kappa} \text{ for some} \\
& \left.\underline{\kappa} > 0 \text{ and } \lambda_{min}(A) \geq \underline{\kappa} \text{ for any } A \in \{\gamma_{22},\gamma_{24}\}\right\} \\
\equiv\ & \mathscr{H}_1 \times \mathscr{H}_{21} \times \mathscr{H}_{22} \times \mathscr{H}_{23} \times \mathscr{H}_{24} \times \mathscr{H}_{25} \qquad (2.18)
\end{aligned}
$$

with $\mathbb{R}_\infty = \mathbb{R} \cup \{\pm\infty\}$. The relevant drifting sequences $\{\gamma_{n,h}\}$ are defined following Guggenberger (2010a) as $\gamma_{n,h} \equiv (\gamma_{n,h,1},\gamma_{n,h,2},\gamma_{n,h,3})$ for $h = (h_1,h_{21}',vec(h_{22})',h_{23},vec(h_{24})',h_{25})' \in \mathscr{H}$, where $\gamma_{n,h,1} = (E_{F_n}[v_i^2])^{-1}E_{F_n}[v_iu_i]$, $\gamma_{n,h,2} = (\gamma_{n,h,21},\gamma_{n,h,22},\gamma_{n,h,23},\gamma_{n,h,24},\gamma_{n,h,25})$ with $\gamma_{n,h,21} = \pi_n$, $\gamma_{n,h,22} = E_{F_n}e_i^2Z_iZ_i'$, $\gamma_{n,h,23} = E_{F_n}e_i^2v_i^2$, $\gamma_{n,h,24} = E_{F_n}Z_iZ_i'$, and $\gamma_{n,h,25} = E_{F_n}v_i^2$ satisfy:

$$
n^{1/2}\gamma_{n,h,1} \to h_1,\ \gamma_{n,h,2} \to h_2 = (h_{21},h_{22},h_{23},h_{24},h_{25}), \text{ and } \gamma_{n,h,3} = F_n \in \Gamma_3(\gamma_{n,h,1},\gamma_{n,h,2}). \qquad (2.19)
$$

More specifically, under H_0 and the drifting sequences of parameters $\{\gamma_{n,h}:\ h \in \mathscr{H}\}$ satisfying (2.19) with $|h_1| < \infty$ (i.e., local endogeneity), we show in Lemma **A.1** that the asymptotic variance $n^{1/2}(\hat{a} - a_{n,h})$ is the same as that of $n^{1/2}\hat{a}$ under exogeneity ($H_a: a = 0$) for any $h_1 \in \mathbb{R}$. Furthermore, the limiting distributions of the estimators $\hat{\theta}_{OLS}$, $\hat{\theta}_{2SLS}$, and $\hat{a}$ are derived in Lemma **A.2**-(a). In particular, by noting that $\psi_{Zu}(0) = \psi_{Ze}$ and $\psi_{vu}(0) = \psi_{ve}$, we have

$$
\begin{pmatrix} n^{1/2}\hat{a} \\ n^{1/2}(\hat{\theta}_{OLS} - \theta) \\ n^{1/2}(\hat{\theta}_{2SLS} - \theta) \end{pmatrix} \to^d \begin{pmatrix} \psi_a \\ \psi_{OLS} \\ \psi_{2SLS}, \end{pmatrix} \equiv \begin{pmatrix} (h_{21}'h_{24}h_{21})^{-1}\left(h_{21}'\psi_{Zu}(0) - h_{25}^{-1}\psi_{vu}(0) + h_1\right. \\ (h_{25} + h_{21}'h_{24}h_{21})^{-1}(h_{21}'\psi_{Zu}(0) + \psi_{vu}(0) + h_{25}h_1) \\ (h_{21}'h_{24}h_{21})^{-1}h_{21}'\psi_{Zu}(0), \end{pmatrix} \qquad (2.20)
$$

from Lemma **A.2**-(a), where $\psi_a \sim N\Big(h_1, \left(h_{21}'h_{24}h_{21}\right)^{-2}h_{21}'h_{22}h_{21} + h_{25}^{-2}h_{23}\Big)$, $\psi_{OLS} \sim N\Big(h_{25}h_1/\left(h_{25} + h_{21}'h_{24}h_{21}\right), (h_{21}'h_{22}h_{21} + h_{23})/\left(h_{25} + h_{21}'h_{24}h_{21}\right)^2\Big)$, and $\psi_{2SLS} \sim N\Big(0, \left(h_{21}'h_{24}h_{21}\right)^{-2}h_{21}'h_{22}h_{21}\Big)$. Interestingly, since we have $\gamma_{n,h,1} \to \bar{\gamma}_1 = 0$ under the drifting parameter sequences $\{\gamma_{n,h}:\ h \in \mathscr{H}\}$ satisfying (2.19) with $|h_1| < \infty$, the limiting distributions under H_0 of all estimators in (2.20) do not depend on the asymptotic behavior of both $n^{-1/2}Z'v$ and $n^{-1/2}\left(v'v - E_F[v'v]\right)$.

Similarly, Lemma **A.2**-(b) and (2.20) imply the following convergence results for $T_{2SLS}(\theta_0)$, $T_{OLS}(\theta_0)$, H_n, and the symmetric two-sided t test statistic $T_n(\theta_0)$:

$$
\begin{aligned}
\begin{pmatrix} T_{2SLS}(\theta_0) \\ T_{OLS}(\theta_0) \\ H_n \end{pmatrix} \xrightarrow{d} \eta_h \quad &= \quad \begin{pmatrix} \left(h_{21}'h_{22}h_{21}\right)^{-1/2}h_{21}'\psi_{Zu}(0) \\ \left(h_{21}'h_{22}h_{21} + h_{23}\right)^{-1/2}(h_{21}'\psi_{Zu}(0) + \psi_{vu}(0) + h_{25}h_1) \\ \left(\frac{h_{21}'h_{22}h_{21}}{(h_{21}'h_{24}h_{21})^2} + h_{23}h_{25}^{-2}\right)^{-1}\left((h_{21}'h_{24}h_{21})^{-1}h_{21}'\psi_{Zu}(0) - h_{25}^{-1}\psi_{vu}(0) + h_1\right)^2 \end{pmatrix} \\
&\equiv \quad \begin{pmatrix} \eta_{1,h} \\ \eta_{2,h} \\ \eta_{3,h} \end{pmatrix} \qquad (2.21)
\end{aligned}
$$

$$
\text{and } T_n(\theta_0) \xrightarrow{d} \tilde{T}_h = \left|\eta_{2,h}\mathbb{1}(\eta_{3,h} \leq \chi^2_{1,1-\beta}) + \eta_{1,h}\mathbb{1}(\eta_{3,h} > \chi^2_{1,1-\beta})\right|, \qquad (2.22)
$$

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 noncentrality parameter $\left(\frac{h_{21}'h_{22}h_{21}}{(h_{21}'h_{24}h_{21})^2}+h_{23}h_{25}^{-2}\right)^{-1}h_1^2$ of the χ^2 limiting distribution of H_n clearly depends on the endogeneity parameter h_1, and is nonzero if and only if $h_1 \neq 0$. Therefore, h_1 determines the power of the first-stage pretest for exogeneity. As in Guggenberger (2010a), we can show from (2.22) that the asymptotic size of $T_n(\theta_0)$ (i.e., $AsySz[c_\infty(1-\alpha)]$) equals 1. That is, the maximal rejection of the two-stage t-test is realized under certain drifting sequence $\{\gamma_{n,h} : h \in \mathscr{H}\}$ with local endogeneity. Such extreme size distortion occurs because when exogeneity is present but is not rejected in the first-stage (which is the case when $h_1 \neq 0$ and $\left(\frac{h_{21}'h_{22}h_{21}}{(h_{21}'h_{24}h_{21})^2}+h_{23}h_{25}^{-2}\right)^{-1}h_1^2 \approx 0$), OLS-based t-test is used in the second stage, but the maximal asymptotic rejection probability for $H_0 : \theta = \theta_0$ with OLS-based t-test equals 1; see the discussion at p.376 in Guggenberger (2010a). This result extends to the one-sided t-test, but we focus on the two-sided t-test to simplify the presentation.

3. Main Results

3.1. Wild bootstrap

In this section, we study the asymptotic behaviour of standard wild bootstrap procedures for the two-stage test, and we show that this bootstrap cannot consistently estimate the distribution of the statistic of interest. To simplify the exposition, we focus on the case of symmetric two-sided test, but our results remain valid for one-sided test.

Wild Bootstrap Algorithm:

1. Given $H_0 : \theta = \theta_0$, compute the residuals from the first-stage and structural equations:

$$\hat{v} = X - Z\hat{\pi}, \tag{3.1}$$
$$\hat{u}(\theta_0) = y - X\theta_0, \tag{3.2}$$

   where $\hat{\pi} = (Z'Z)^{-1}Z'X$ denotes the least squares estimator of π.

2. Generate the bootstrap pseudo-data following

$$X^* = Z\hat{\pi} + v^*, \tag{3.3}$$
$$y^* = X^*\theta_0 + u^*, \tag{3.4}$$

   where there are two options to generate the bootstrap disturbances:

   (a) v^* and u^* are generated independently from each other. Specifically, in the case with heteroskedastic data, we set for each observation

$$\begin{pmatrix} v_i^* \\ u_i^* \end{pmatrix} = \begin{pmatrix} \hat{v}_i e_{1,i}^* \\ \hat{u}_i(\theta_0) e_{2,i}^* \end{pmatrix}, \tag{3.5}$$

where e_{1i}^* and e_{2i}^* are two random variables that has mean 0 and variance 1 and are independent from each other.

(b) (v^*, u^*) are drawn dependently from each other. For heteroskedastic data, we set

$$\begin{pmatrix} v_i^* \\ u_i^* \end{pmatrix} = \begin{pmatrix} \hat{v}_i e_{1,i}^* \\ \hat{u}_i(\theta_0) e_{1,i}^* \end{pmatrix}. \tag{3.6}$$

Following Young (2020), we refer to (a) as *independent transformation* of disturbances and (b) as *dependent transformation* of disturbances. For the purpose of better size control, it is often recommended that to bootstrap exogeneity tests, (u^*, v^*) 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 tests.

3. Compute the bootstrap analogue of the two-stage statistic (for a symmetric test):

$$T_n^*(\theta_0) = \left| T_{OLS}^*(\theta_0) \mathbb{1}(H_n^* \leq \chi_{1,1-\beta}^2) + T_{2SLS}^*(\theta_0) \mathbb{1}(H_n^* > \chi_{1,1-\beta}^2) \right|, \tag{3.7}$$

where $T_{OLS}^*(\theta_0)$, $T_{2SLS}^*(\theta_0)$ 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. Repeat Steps 2-3 B times and obtain $T_n^*(\theta_0)$, $b = 1, \ldots, B$. The bootstrap test rejects H_0 if the bootstrap p-value $\frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\left[T_n^{*(b)}(\theta_0) > T_n(\theta_0) \right]$ is less than α.

To check whether the bootstrap consistently estimates the distribution of the two-stage test statistic, one needs to check whether we have

$$\sup_{x \in R} \left| P^* \left(T_n^*(\theta_0) \leq x \right) - P \left(T_n(\theta_0) \leq x \right) \right| \to^P 0 \tag{3.8}$$

under H_0 and such drifting parameter sequences.

First, we note from Lemma **A.4** in the appendix that the following convergence results hold for the bootstrap statistics (conditional on the sample):

$$n^{-1/2} \begin{pmatrix} Z' u^* \\ Z' v^* \\ \left(u^{*\prime} v^* - E^*[u^{*\prime} v^*] \right) \end{pmatrix} \to^{d^*} \begin{pmatrix} \psi_{Zu}^*(0) \\ \psi_{Zv}^* \\ \psi_{vu}^*(0) \end{pmatrix} \sim N \left(0, \begin{pmatrix} \Omega_{Zu}(0) & 0 & 0 \\ 0 & \Omega_{Zv} & 0 \\ 0 & 0 & \Omega_{vu}(0) \end{pmatrix} \right), \tag{3.9}$$

in probability, i.e., the bootstrap (with dependent or independent transformation) does replicate well the randomness in the original sample. Theorem **3.1** gives the null limiting distributions of the bootstrap two-stage test statistics under the drifting parameter sequences $\{\gamma_{n,h} : h \in \mathscr{H}\}$ satisfying (2.19) with $|h_1| < \infty$.

Theorem 3.1 *Conditional on the sample, the following convergence holds under H_0 and $\{\gamma_{n,h} : h \in \mathscr{H}\}$ satisfying (2.19) with $|h_1| < \infty$:*

$$\begin{pmatrix} T^*_{2SLS}(\theta_0) \\ T^*_{OLS}(\theta_0) \\ H^*_n \end{pmatrix} \to^{d^*} \eta^*_h = \begin{pmatrix} (h'_{21}h_{22}h_{21})^{-1/2} h'_{21}\psi^*_{Zu}(0) \\ (h'_{21}h_{22}h_{21}+h_{23})^{-1/2}\left(h'_{21}\psi^*_{Zu}(0)+\psi^*_{vu}(0)+h_{25}h^b_1\right) \\ \left(\frac{h'_{21}h_{22}h_{21}}{(h'_{21}h_{24}h_{21})^2}+h_{23}h_{25}^{-2}\right)^{-1}\left((h'_{21}h_{24}h_{21})^{-1}h'_{21}\psi^*_{Zu}(0)-\psi^*_{vu}(0)+h^b_1\right)^2 \end{pmatrix},$$

$$\equiv \begin{pmatrix} \eta^*_{1,h} \\ \eta^*_{2,h} \\ \eta^*_{3,h} \end{pmatrix}$$

$$T^*_n(\theta_0) \to^{d^*} \tilde{T}^*_h = \left|\eta^*_{2,h}\mathbb{1}(\eta^*_{3,h} \leq \chi^2_{1,1-\beta}) + \eta^*_{1,h}\mathbb{1}(\eta^*_{3,h} > \chi^2_{1,1-\beta})\right|,$$

in probability, where $h^b_1 = 0$ for the bootstrap based on independent transformation of disturbances, and $h^b_1 = h_1 + \psi_{vu}(0)$ with $\psi_{vu}(0) \sim N(0, \Omega_{vu}(0))$ for the bootstrap based on dependent transformation of disturbances.

According to Theorem **3.1**, 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. In particular, we note that h^b_1 corresponds to the localization parameter of endogeneity in the bootstrap world, and the bootstrap with independent transformation (henceforth dubbed as independent bootstrap) removes all the endogeneity when generating the bootstrap samples. On the other hand, while the bootstrap with dependent transformation (henceforth dubbed as dependent bootstrap) is able to mimic the situation of local endogeneity in the original sample (note that h^b_1 is finite with probability approaching one when h_1 is finite), the approximation is imprecise and results in an extra error term $\psi_{vu}(0) \sim N(0, \Omega_{vu}(0))$, whose value depends on the actual realization of the sample. In particular, these results suggest that, in the bootstrap world, the conditional limiting distribution of H^*_n, given ψ_{uv}, is a central chi-squared distribution under the independent bootstrap, while it is distributed as a noncentral chi-square with one degree of freedom and (random) noncentrality parameter $\left(\frac{h'_{21}h_{22}h_{21}}{(h'_{21}h_{24}h_{21})^2} + h_{23}h_{25}^{-2}\right)^{-1}(h_1 + \psi_{vu}(0))^2$ under the dependent bootstrap. Therefore, the power of the bootstrap pretest statistic H^*_n under either procedure will be different from that of H_n.

From Theorem **3.1**, it is clear that the (conditional) null limiting distribution of the bootstrap two-stage test statistic is different from the null limiting distribution of the original two-stage test statistic in (2.22). Therefore, the bootstrap consistency in (3.8) cannot hold in the current context. However, even if the bootstrap is inconsistent, it might still be able to provide a valid test if its asymptotic NRP does not exceed the nominal size under any sequence in (2.19). To further shed light on the behaviour of the bootstrap statistics, we apply (2.22) and Theorem **3.1** to the case with conditional homoskedasticity as in Guggenberger (2010), and plot the 95% quantiles of $\tilde{T}_h$ and $\tilde{T}_h^*$ in Figure 1 as a function of h_1 with $h_2 \in \{.2, .4, .6, .8, 1, 2\}$, where $h_2 = ||\Omega_{ZZ}^{1/2}\pi/\sigma_v||$, and $\beta = .05$. Notice that the limiting distributions of interest are considerably simplified in this case, only depending on h_1 and h_2. We highlight some interesting findings below.

First, we observe that the quantiles of $\tilde{T}_h^*$ for the independent bootstrap can be much lower than those of $\tilde{T}_h$ when the values of h_1 and/or h_2 are small, suggesting that this bootstrap procedure can seriously overreject the null hypothesis in such cases. Indeed, its quantiles always correspond the case that the endogeneity parameter exactly equals zero as its data generating process totally removes the degree of endogeneity in the bootstrap world. By contrast, the quantiles of $\tilde{T}_h^*$ for the dependent bootstrap turn out to be rather close to those of $\tilde{T}_h$ across various values of h_1 and h_2. However, the figure suggests that this bootstrap procedure may also have some slight over-rejection when the quantiles of $\tilde{T}_h$ are relatively high (e.g., when $h_2 = .4$ and $h_1 = 5$). In addition, we note that the quantiles of $\tilde{T}_h^*$ for the dependent bootstrap converge in each sub-figure to the standard normal CV when the value of h_1 increases: when $|h_1|$ is large, the Hausman pretest rejects with high probability so that the two-stage test becomes the 2SLS-based t-test, and the dependent bootstrap does mimic well such behaviour. On the other hand, we note that the quantiles for the independent bootstrap becomes close to the standard normal CV only when h_2 is fairly large (e.g., when h_2=2). Intuitively, when h_2 becomes large, the term with ψ_{Zu}^* becomes dominant in the limit of $T_{OLS}^*(\theta_0)$ (i.e., $\eta_{2,h}^*$) while the term with ψ_{uv}^* becomes dominant in the limit of H_n^* (i.e., $\eta_{3,h}^*$, which equals $\left(\frac{h_{21}'h_{22}h_{21}}{(h_{21}'h_{24}h_{21})^2} + h_{23}h_{25}^{-2}\right)^{-1}\left((h_{21}'h_{24}h_{21})^{-1}h_{21}'\psi_{Zu}^*(0) - \psi_{vu}^*(0)\right)^2$ for the independent bootstrap), so that conditional on the sample, $\eta_{3,h}^*$ becomes independent from both $\eta_{1,h}^*$ and $\eta_{2,h}^*$ in this case, as ψ_{Zu}^* and ψ_{uv}^* are independent from each other (e.g., see (3.9)).

Furthermore, we can obtain the asymptotic sizes of the two bootstrap tests by applying the results in (2.22) and Theorem **3.1**. Specifically, the asymptotic size of the bootstrap two-stage test can be defined as:

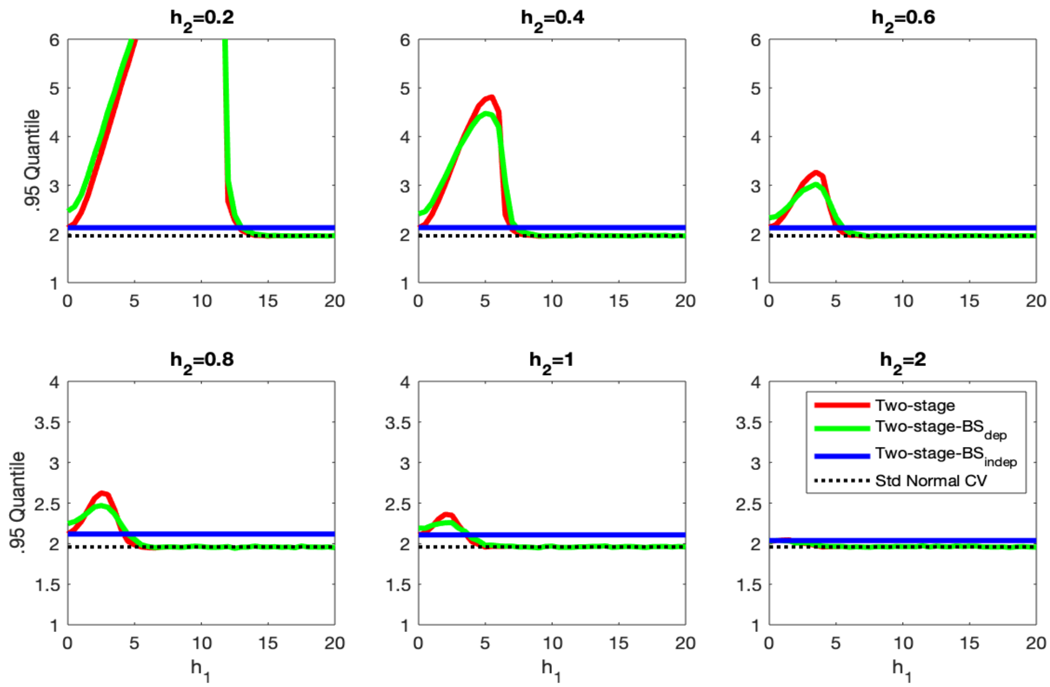
$$AsySz\left[\hat{c}_n^*(1-\alpha)\right] := \limsup_{n\to\infty}\sup_{\gamma\in\Gamma} P_{\theta_0,\gamma}\left[T_n(\theta_0) > \hat{c}_n^*(1-\alpha)\right], \tag{3.10}$$

where $\hat{c}_n^*(1-\alpha)$ denotes the $(1-\alpha)$-th quantile of the distribution of $T_n^*(\theta_0)$, based on the dependent or independent transformation. The next theorem gives an explicit formula of the asymptotic size. Note that the asymptotic size depends on α, β, and $\underline{\kappa}$, but it does not depend on k (number of instruments used).

Theorem 3.2 *$AsySz[\hat{c}_n^*(1-\alpha)]$ equals $\sup_{h\in\mathscr{H}} P[\tilde{T}_h > c_h^*(1-\alpha)]$, where $\tilde{T}_h$ is defined in (2.22) and $c_h^*(1-\alpha)$ is the $(1-\alpha)$-th quantile of $\tilde{T}_h^*$ defined in Theorem* **3.1**.

Following Guggenberger (2010a, Table 1), we report the asymptotic sizes of the (symmetric) two-stage tests based on the standard normal CV, the independent bootstrap CV, and dependent bootstrap CV in Table 1 when $\alpha = .05$ for $\underline{\kappa} \in \{.001, .1, .5, 1, 2, 10\}$ and $\beta \in \{.05, .1, .2, .5\}$. First,

Figure 1. 95% quantiles of $\tilde{T}_h$ and $\tilde{T}_h^*$



Note: The results are based on 100,000 simulation replications.

we note that both the standard normal CVs and the independent bootstrap CVs have asymptotic size much larger than .05; e.g., when $\underline{\kappa} = .001$, the two methods have asymptotic sizes equal to $100\%, 95.1\%, 85.3\%, 55.4\%$ and $97.7\%, 92.8\%, 83.0\%, 53.7\%$, respectively. In addition, it turns out that the independent bootstrap CVs always have smaller size distortion than the standard normal CVs, and this is in line with the results in Figure 1, in which the quantiles of the independent bootstrap limit $\tilde{T}_h^*$ are always higher than the standard normal CVs. On the other hand, we note that although in general also unable to achieve uniform size control, the dependent bootstrap CVs have asymptotic sizes quite close to the nominal level.

Remarks

1. How do these asymptotic size results correspond to real-world data? For instance, as remarked by Guggenberger (2010a), Angrist and Krueger (1991)'s influential study on return to schooling has estimated concentration parameters equal to 95.6 and 257 for the cases with 3 IVs and 180 IVs, respectively. And they correspond to the values of γ_2 equal to .017 and .028, respectively, for the sample size $n = 329,509$ in their study. Therefore, Table 1 suggests that a Hausman-pretest-based two-stage procedure with either the asymptotic CV or the independent bootstrap CV would lead to extreme distortion of null rejection probability for the Angrist and Krueger (1991) data, while the one based on the dependent bootstrap CV would not suffer from serious size distortion.

2. As seen in Table 1, the asymptotic size of the *dependent bootstrap* test can be either higher or lower than the nominal level (thus asymptotically conservative or over-sized), depending on the value of the lower bound of IV strength $\underline{\kappa}$. Still, it has asymptotic sizes quite close to the nominal level across various settings, and is therefore much more desirable than the independent bootstrap in terms of size control for the two-stage test. Note that the extreme size distortion of the *independent bootstrap* is not a surprise, as this scheme assumes exogeneity while the endogeneity is local-to-zero in the true DGP. However, as we will see in Section 4, the dependent bootstrap has relatively low finite-sample power compared with alternative methods considered in the simulations (including our novel hybrid bootstrap procedures that are based on *independent draws* of the structural and reduced-form residuals). In the next section, we will show that the hybrid bootstrap procedures achieve both correct asymptotic size and better finite-sample power properties. In particular, the use of the *independent bootstrap* is paramount for the validity of these procedures since it helps to first remove all the endogeneity in the bootstrap world before applying an appropriate size-correction method to account for the localized endogeneity parameter h_1, which cannot be well estimated by the .

Table 1. Asymptotic size (in %) of two-stage tests for $\alpha = .05$.

	Std Normal CV				BS-independent				BS-dependent			
$\underline{\kappa} \setminus \beta$	.05	.1	.2	.5	.05	.1	.2	.5	.05	.1	.2	.5
.001	100	95.1	85.3	55.4	97.7	92.8	83.0	53.7	1.2	1.2	1.3	2.0
.1	95.5	90.4	80.2	50.7	93.9	88.4	77.6	48.8	1.3	1.5	1.9	3.0
.5	60.4	50.5	39.2	22.2	55.9	45.3	34.5	19.3	6.6	6.6	6.5	6.5
1	27.7	21.7	16.2	9.7	24.7	18.5	12.9	7.8	6.8	6.6	6.5	6.1
2	10.8	9.3	7.7	6.2	10.1	8.3	6.6	5.2	6.1	6.0	5.7	5.3
10	5.3	5.3	5.2	5.2	5.3	5.3	5.3	5.2	5.3	5.3	5.3	5.2

Note: The results are based on 100,000 simulation replications.

3.2. Hybrid bootstrap

In this section, we introduce hybrid bootstrap procedures that are able to achieve correct asymptotic size for the two-stage test. First, we show how to construct a hybrid bootstrap CV in the current context by using Bonferroni bounds. Note that in the case of local endogeneity with $|h_1| < \infty$, the localization parameter h_1 cannot be consistently estimated. However, we may still construct an asymptotically valid confidence set for h_1 by using some appropriate choice of estimator $\hat{h}_{n,1}$. For example, we can define an estimator $\hat{h}_{n,1} = n^{1/2}\hat{a}$, where

$$\hat{a} = \left(\hat{v}' M_X \hat{v}\right)^{-1} \hat{v}' M_X y, \tag{3.11}$$

and $\hat{v} = M_Z X$ is the vector of residuals from the first-stage OLS regression (i.e., the regression of X on instruments Z). Then a confidence set of h_1 can be constructed by using the fact that under the drifting parameter sequences,

$$\hat{h}_{n,1} \to^{d} \tilde{h}_1 \sim N\Big(h_1, \big(h_{21}' h_{24} h_{21}\big)^{-2} h_{21}' h_{22} h_{21} + h_{25}^{-2} h_{23}\Big) \tag{3.12}$$

from (2.20). With the exception of h_1, note that the other parameters appearing in the normal distribution in (3.12) can be consistently estimated.

Alternatively, one may consider using the null-imposed estimator $\hat{h}_{n,1}(\theta_0) = n^{1/2}\hat{a}_n(\theta_0) = \big(\hat{v}'\hat{v}\big)^{-1}\hat{v}'(y - X\theta_0)$, whose null limiting distribution follows $N\Big(h_1, h_{25}^{-2} h_{23}\Big)$. Then, uniformly valid hybrid bootstrap CVs for testing $H_0 : \theta = \theta_0$ under the two-stage procedure can be constructed by using Bonferroni bounds: we first construct a $1-(\alpha-\delta)$ level first-stage confidence set for h_1, then take the maximal $(1-\delta)$-th quantile of appropriately generated bootstrap statistics over the first-stage confidence set. Specifically, let $\hat{h}_{n,2} \equiv (\hat{h}_{n,21}, \hat{h}_{n,22}, \hat{h}_{n,23}, \hat{h}_{n,24}, \hat{h}_{n,25})$ be a consistent estimator of $h_2 \equiv (h_{21}, h_{22}, h_{23}, h_{24}, h_{25})$ and $CI_{\alpha-\delta}(\hat{h}_{n,1})$ denote the $1-(\alpha-\delta)$ level confidence set for h_1 for some $0 < \delta \le \alpha < 1$. The bootstrap-based simple Bonferroni critical value (SBCV) is defined as

$$c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2}) \quad = \sup_{h_1 \in CI_{\alpha-\delta}(\hat{h}_{n,1})} c^*_{(h_1, \hat{h}_{n,2})}(1-\delta), \tag{3.13}$$

where $c^*_{(h_1,\hat{h}_2)}(1-\delta)$ is the $(1-\delta)$-th quantile of the distribution of $T^*_{n,(h_1,\hat{h}_{n,2})}(\theta_0)$, which is the bootstrap analogue of $T_n(\theta_0)$ 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. Therefore, attention has to be taken when considering the bootstrap DGP. In particular, we propose to generate $T^*_{n,(h_1,\hat{h}_{n,2})}(\theta_0)$ as follows:

$$\begin{aligned} & T^*_{n,(h_1,\hat{h}_{n,2})}(\theta_0) \\ = & \left| T^*_{OLS,(h_1,\hat{h}_{n,2})}(\theta_0) \mathbb{1}\left(H^*_{n,(h_1,\hat{h}_{n,2})} \le \chi^2_{1,1-\beta}\right) + T^*_{2SLS}(\theta_0) \mathbb{1}\left(H^*_{n,(h_1,\hat{h}_{n,2})} > \chi^2_{1,1-\beta}\right) \right|, \end{aligned} \tag{3.14}$$

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. To obtain these bootstrap analogues, we first generate the bootstrap counterparts of the OLS and regression endogeneity parameter estimators under h_1:

$$\begin{aligned} \hat{\theta}^*_{OLS,(h_1,\hat{h}_{n,2})} &= \hat{\theta}^*_{OLS} + \big(1 + \hat{h}_{n,21}\hat{h}_{n,22}\hat{h}_{n,21}'\big)^{-1} n^{-1/2} h_1, \\ \hat{a}^*_{(h_1,\hat{h}_{n,2})} &= \hat{a}^* + n^{-1/2} h_1, \end{aligned} \tag{3.15}$$

where $\hat{\theta}^*_{OLS}$ and $\hat{a}^*$ are generated by the standard 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, $\sqrt{n}(\hat{\theta}^*_{OLS,(h_1,\hat{h}_{n,2})} - \theta_0)$ and $\sqrt{n}\hat{a}^*_{(h_1,\hat{h}_{n,2})}$ have appropriate (conditional) null limiting distribution. Then, we obtain $T^*_{OLS,(h_1,\hat{h}_{n,2})}(\theta_0)$ and $H^*_{n,(h_1,\hat{h}_{n,2})}$ as follows:

$$T^*_{OLS,(h_1,\hat{h}_{n,2})}(\theta_0) = \frac{\sqrt{n}(\hat{\theta}^*_{OLS,(h_1,\hat{h}_{n,2})} - \theta_0)}{\hat{V}^{*1/2}_{OLS}}, \tag{3.16}$$

$$H^*_{n,(h_1,\hat{h}_{n,2})} = \frac{n\hat{a}^{*2}_{(h_1,\hat{h}_{n,2})}}{\hat{V}^*_a}, \tag{3.17}$$

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

$$\begin{pmatrix} T^*_{OLS,(h_1,\hat{h}_{n,2})}(\theta_0) \\ H^*_{n,(h_1,\hat{h}_{n,2})} \end{pmatrix} \rightarrow^{d^*} \begin{pmatrix} (h'_{21}h_{22}h_{21}+h_{23})^{-1/2}(h'_{21}\psi^*_{Zu}(0)+\psi^*_{vu}(0)+h_{25}h_1) \\ \left(\frac{h'_{21}h_{22}h_{21}}{(h'_{21}h_{24}h_{21})^2}+h_{23}h_{25}^{-2}\right)^{-1}\left((h'_{21}h_{24}h_{21})^{-1}h'_{21}\psi^*_{Zu}(0)-\psi^*_{vu}(0)+h_1\right)^2 \end{pmatrix}, \tag{3.18}$$

in probability P. This implies that $T^*_{n,(h_1,\hat{h}_{n,2})}(\theta_0)$, the resulting bootstrap counterpart of the two-stage test statistic, has the desired (conditional) limiting distribution evaluated at the value of localization parameter equal to h_1 (different from the limiting distributions in Theorem **3.1**).

As seen from (3.13), the bootstrap SBCV equals the maximal CV $c^*_{(h_1,\hat{h}_{n,2})}(1-\delta)$ 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 theorem for $c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})$.

Theorem 3.3 *Suppose that H_0 holds, and then for any* $0<\delta\leq\alpha<1$, *we have:*

$$AsySz\left[c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})\right] := \limsup_{n\to\infty}\sup_{\gamma\in\Gamma} P_{\theta_0,\gamma}\left[T_n(\theta_0) > c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})\right] \leq \alpha.$$

Theorem **3.3** states that tests based on $c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})$ control the asymptotic size. In practice, $c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})$ can be obtained by using the following algorithm.

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

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

2. Choose α and δ, and compute $CI_{\alpha-\delta}(\hat{h}_{n,1})$.

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

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

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

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

Note that as shown in Theorem **3.3**, although controlling the size, the bootstrap SBCV may yield a conservative test whose asymptotic size does not reach its nominal level. For further refinement on the Bonferroni bound, we propose a size-correction method to adjust the bootstrap SBCV so that the resulting test is not conservative with asymptotic size exactly equal to α. Specifically, the size-correction factor for the bootstrap SBCV is defined as:

$$\hat{\eta}_n \quad = \quad \inf\left\{\eta : \sup_{h_1 \in \mathscr{H}_1} P^* \left[T^*_{n,(h_1,\hat{h}_{n,2})}(\theta_0) > c^{B-S}(\alpha, \alpha-\delta, \hat{h}^*_{n,1}(h_1), \hat{h}_{n,2}) + \eta\right] \leq \alpha\right\}, \tag{3.19}$$

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

$$\hat{h}^*_{n,1}(h_1) = \hat{h}^*_{n,1} + h_1, \tag{3.20}$$

where $\hat{h}^*_{n,1} = n^{1/2}\hat{a}^* = \left(\hat{v}^{*\prime} M_{X^*} \hat{v}^*\right)^{-1} \hat{v}^{*\prime} M_{X^*} y^*$, $\hat{v}^* = M_{Z^*} X^*$, is generated by the standard bootstrap procedure with *independent transformation* (so that the localization parameter equals zero in the bootstrap world). Note that $\hat{h}^*_{n,1}$ converges in distribution to $N\left(0, \left(h'_{21}h_{24}h_{21}\right)^{-2} h'_{21}h_{22}h_{21} + h_{25}^{-2}h_{23}\right)$ in probability P, while $\hat{h}^*_{n,1}(h_1)$ converges in distribution to $N\left(h_1, \left(h'_{21}h_{24}h_{21}\right)^{-2} h'_{21}h_{22}h_{21} + h_{25}^{-2}h_{23}\right)$ in probability P, i.e., the same limiting distribution of $\hat{h}_{n,1}$ in (3.12).

We emphasize that $\hat{h}^*_{n,1}(h_1)$ needs to be generated simultaneously with $T^*_{n,(h_1,\hat{h}_{n,2})}(\theta_0)$ using the same bootstrap samples, so that the dependence structure between the statistics $T_n(\theta_0)$ and $\hat{h}_{n,1}$ is well mimicked by the bootstrap statistics. This is important for the procedure described in (3.19) to correct the conservativeness of the Bonferroni bound. Similarly, for the implementation of the size-correction method, one cannot replace $c^{B-S}(\alpha, \alpha-\delta, \hat{h}^*_{n,1}(h_1), \hat{h}_{n,2})$ in (3.19) with $c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})$, as it also breaks down the dependence structure.

The goal of the size-correction method is to decrease the bootstrap SBCV as much as possible by using the factor η while not violating the inequality in (3.19), so that the asymptotic size of the resulting tests can be controlled. Then, the size-corrected bootstrap CV can be defined as

$$c^{B-C}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2}) \quad = \quad c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2}) + \hat{\eta}_n, \tag{3.21}$$

and one can expect that relatively small $\hat{\eta}_n$ results in relatively less conservative (and more powerful) test. In particular, under a proper algorithm for the size-correction method, and given some fixed $\alpha \in (0,1)$ and $\delta \in (0,\alpha]$, the size-correction factor $\hat{\eta}_n(\cdot)$ is continuous as a function of $\hat{h}_{n,1}$. We can now state the following theorem on the uniform size control of the bootstrap CVs based on the size-correction method.

Theorem 3.4 *Suppose that H_0 holds, and then for any $0 < \delta \leq \alpha < 1$, we have:*

$$AsySz\left[c^{B-C}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})\right] \quad := \quad \limsup_{n\to\infty} \sup_{\gamma \in \Gamma} P_{\theta_0,\gamma}\left[T_n(\theta_0) > c^{B-C}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})\right] = \alpha.$$

Theorem **3.4** shows that $c^{B-C}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})$, the size-corrected bootstrap CVs, yield tests with the correct asymptotic size. To implement such tests in practice, we must compute $c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})$

and $\hat{\eta}_n$. These values can be computed sequentially starting with $c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})$. Then the size-correction factor $\hat{\eta}_n$ can be computed by evaluating (3.19) over a fine grid of $\mathscr{H}_1$ as follows.

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

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 bootstrap procedure with independent transformation of disturbances.

2. Let $c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})$ be the obtained bootstrap SBCV.

3. Create a fine grid of the set $\mathscr{H}_1$ in (3.19) and call it $\mathscr{H}_1^{grid}$.

4. For each $h_1\in\mathscr{H}_1^{grid}$, obtain $T^{*(b)}_{n,(h_1,\hat{h}_{n,2})}(\theta_0)$ and $c^{B-S}(\alpha,\alpha-\delta,\hat{h}^{*(b)}_{n,1}(h_1),\hat{h}_{n,2})$, $b=1,...,B$, 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 repeatedly for each h_1.

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

6. Find all $\eta\in\mathbb{S}_n^{grid}$ such that

$$\sup_{h_1\in\mathscr{H}_1}\frac{1}{B}\sum_{b=1}^{B}\mathbb{1}\left[T^{*(b)}_{n,(h_1,\hat{h}_{n,2})}(\theta_0)>c^{B-S}(\alpha,\alpha-\delta,\hat{h}^{*(b)}_{n,1}(h_1),\hat{h}_{n,2})+\eta\right]\leq\alpha$$

   and set $\hat{\eta}_n$ equal to the smallest η.

7. The size-corrected bootstrap CV is given by

$$c^{B-C}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})=c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})+\hat{\eta}_n.$$

Remarks

1. We note that the computational cost of the proposed hybrid bootstrap procedures is not very high. In particular, the same bootstrap samples can be used in the Algorithms for $c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})$ and $c^{B-C}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})$: there is no need to generate a new set of bootstrap samples to implement the size-correction method in (3.19). Moreover, the same set of bootstrap statistics can be used repeatedly for each value of localization parameter h_1 when constructing the localized quantiles $c^{*}_{(h_1,\hat{h}_{n,2})}(1-\delta)$ in Step 4 of the Algorithm for $c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})$. Similarly, the bootstrap statistics can be used repeatedly for each h_1 when evaluating the size-correction factor in Step 4 of the Algorithm for $c^{B-C}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})$.

2. Recently, Hansen (2017) proposes a Stein-like shrinkage approach in the context of IV regression. His estimator takes a weighted average of the 2SLS and OLS estimators, with the weight depending inversely on the test statistic for exogeneity. Under our current setting, this estimator can be written as

$$\tilde{\theta}=w\hat{\theta}_{OLS}+(1-w)\hat{\theta}_{2SLS},\tag{3.22}$$

where

$$w=\begin{cases}\tau/H_n & \text{if}\quad H_n\geq\tau,\\ 1 & \text{if}\quad H_n<\tau,\end{cases}\tag{3.23}$$

where τ is a shrinkage parameter. We can show that our hybrid bootstrap procedure can be applied to Hansen (2017)'s shrinkage approach as well.

4. Finite sample power performance

In this section, we study the finite-sample power performance of four tests: the 2SLS-based t-test (without Hausman pretest), the two-stage test based on the hybrid bootstrap CVs, and a test that is based on Hansen (2017)'s shrinkage estimator and its corresponding hybrid bootstrap CVs. We do not include the two-stage tests based on the standard normal CVs and the independent bootstrap CVs, as they have extreme size distortion (e.g., see Table 1).

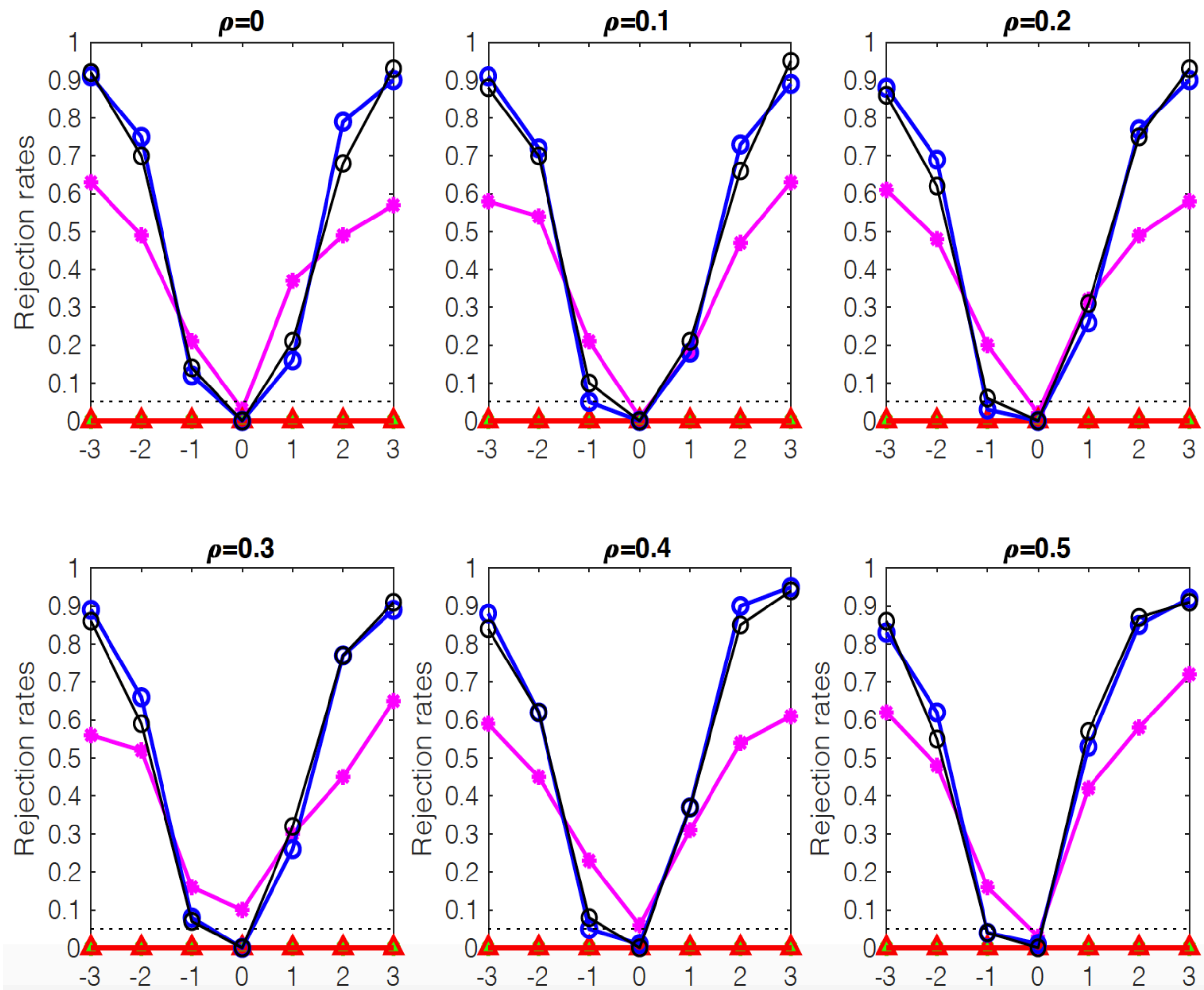
We conduct Monte Carlo simulations by using the linear IV model in (2.1). The sample size is set at $n = 100$, the number of Monte Carlo replications is set at $2,000$, and the number of bootstrap replications is set at $B = 199$. We set $\alpha = .05$ for the nominal levels the 2SLS-based t test, and the two bootstrap-based tests, and set $\beta = .05$ for the nominal level of the Hausman pretest. The size-correction algorithms described in Section 3.2 are executed with $\delta = .025$. The shrinkage parameter τ is set to equal 1. The number of instruments is set at $k = 1$. The errors have unit variance, so the endogeneity parameter, a, equals ρ (the correlation).

Figures 2 - 3 show the finite-sample power curves of the tests. The true values of the endogeneity parameter are set at $\rho \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$. The values of the concentration parameter, which characterizes the overall IV strength, are set at $\mu^2 \in \{2, 10\}$ for Figures 2 - 3, respectively. It is clear that when the IV is not strong, the two hybrid bootstrap-based tests have remarkable power gain over the 2SLS-based t-test. Such power gain originates from the inclusion of the *OLS*-based t-test in the two-stage test. The Monte Carlo simulations suggest that our method could be particularly attractive in the cases where the available instruments 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 distortion.

5. Conclusions

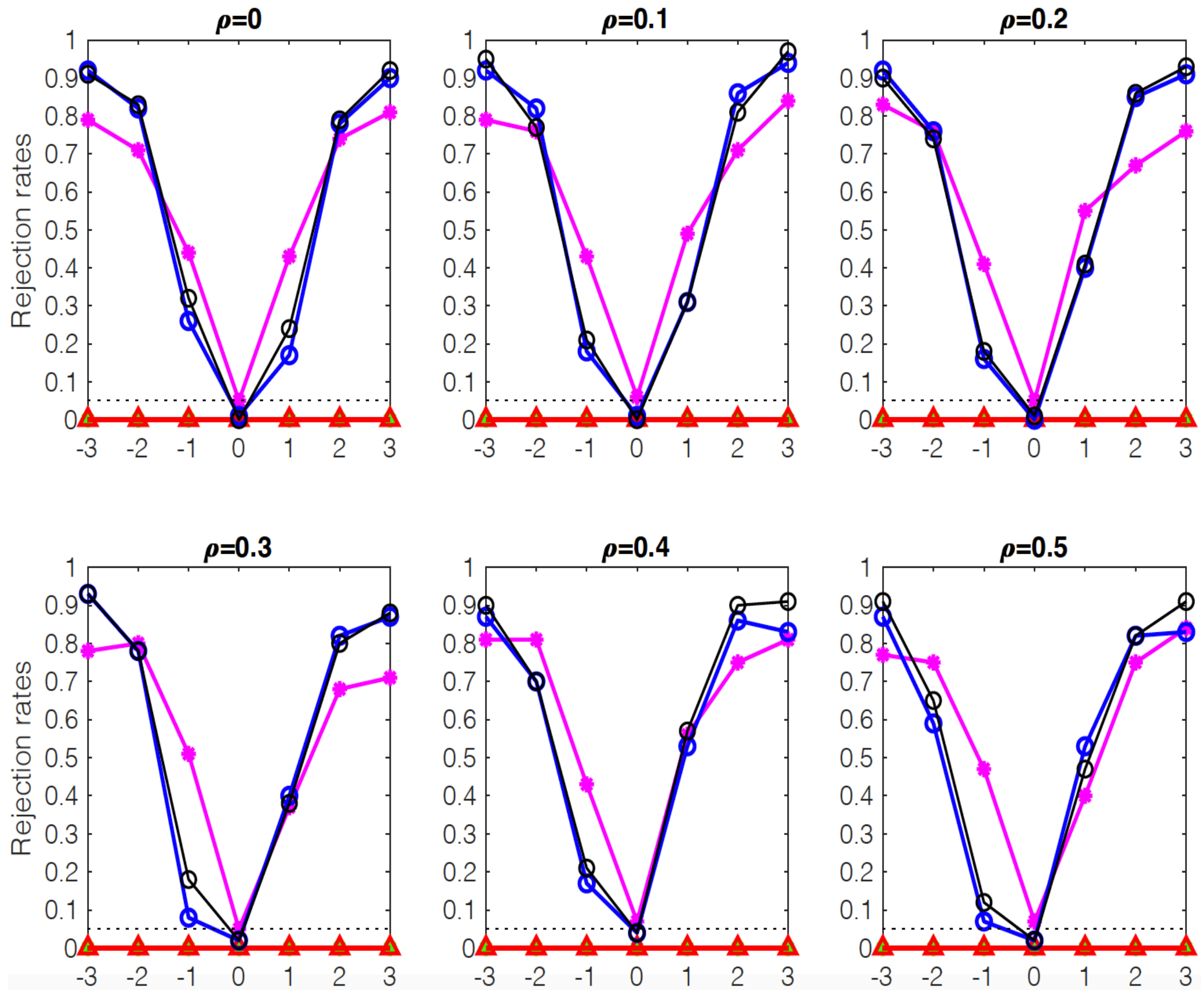
In this paper, we study how to conduct uniformly valid inference for the two-stage procedure by using data-dependent critical values with possibly heteroskedastic data. We first show that standard bootstrap procedures with dependent or independent transformation of disturbances cannot consistently estimate the null distribution of the two-stage test statistics under local endogeneity. In particular, these bootstrap methods cannot mimic well the key localization parameter in the model. We also study the asymptotic sizes of the two bootstrap procedures, and find that the bootstrap two-stage test with independent transformation has extreme size distortion while the one with dependent transformation is much less distorted. Then, we propose a hybrid bootstrap approach, which makes use of the standard bootstrap procedure with independent transformation and a Bonferroni-based size-correction method, which allows us to handle the localization parameter properly. We show that the hybrid bootstrap method is uniformly valid in the sense that it yields correct asymptotic size. Monte Carlo simulations confirm that our proposed method is able to achieve remarkable power gains over the 2SLS-based t-test, especially when the instruments are not very strong.

Figure 2. Power of 2SLS-t and hybrid bootstrap tests: $\mu^2 = 2$



Notes: The power curves for the 2SLS-t test, the two-stage test with hybrid-bootstrap CVs, and the shrinkage test with hybrid-bootstrap CVs are illustrated by the curves in pink, blue, and black, respectively.

Figure 3. Power of 2SLS-*t* and hybrid bootstrap tests: $\mu^2 = 10$



Notes: The power curves for the 2SLS-*t* test, the two-stage test with hybrid-bootstrap CVs, and the shrinkage test with hybrid-bootstrap CVs are illustrated by the curves in pink, blue, and black, respectively.

References

Anderson, T. W., Rubin, H., 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. Annals of Mathematical Statistics 20(1), 46–63.

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.

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

Berkowitz, D., Caner, M., Fang, Y., 2008. Are nearly exogenous instruments reliable?. Economics Letters 101, 20–23.

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

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

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.

Davidson, R., MacKinnon, J. G., 2014. Bootstrap confidence sets with weak instruments. Econometric Reviews 33(5-6), 651–675.

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.

Dovonon, P., Gonçalves, S., 2017. Bootstrapping the GMM overidentification test under first-order underidentification. Journal of Econometrics 201(1), 43–71.

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, 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.

Kabaila, P., Leeb, H., 2006. On the large-sample minimal coverage probability of confidence intervals after model selection. Journal of the American Statistical Association 101(474), 619–629.

Kabaila, P., Mainzer, R., Farchione, D., 2015. The impact of a hausman pretest, applied to panel data, on the coverage probability of confidence intervals. Economics Letters 131, 12–15.

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

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

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

Leeb, H., Pötscher, B. M., 2009. Model selection. In: Handbook of Financial Time Series. Springer,, pp. 889–925.

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

McCloskey, A., 2019. Asymptotically uniform tests after consistent model selection in the linear regression model. Journal of Business & Economic Statistics Forthcoming(just-accepted), 1–33.

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

Ruud, P. A., 2000. An Introduction to Classical Econometric Theory. Oxford University Press, Inc., New York.

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., 2021. Wild bootstrap for instrumental variables regression with weak instruments and few clusters. Technical Report No.106227, University Library of Munich, Germany.

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.

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

A. Appendix

Section A.1 contains the proofs of the theoretical results in the paper.

A.1. Mathematical Proofs

Lemma A.1 *Under the drift sequences of parameters* $\{\gamma_{n,h}\}$ *in* (2.19) *with* $|h_1| < \infty$, *we have*:

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

i.e., the limiting distribution of $n^{1/2}(\hat{a} - a_{n,h})$ *is the same as that of* $\left(n^{-1}\hat{v}'M_X\hat{v}\right)^{-1}\left(n^{-1/2}\hat{v}'M_Xe\right)$. *So, the asymptotic variance of* $n^{1/2}(\hat{a} - a_{n,h})$ *under localized sequences of drifting endogeneity parameter* $n^{1/2}a_{n,h} \to h_1 \in R$ *is the same that as under exogeneity* ($a = 0$).

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

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

It is sufficient to show that the first term in (A.1) is $o_P(1)$. First,

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

which follows from the fact that

$$\begin{aligned} n^{-1}\hat{v}'Z &= n^{-1}(v + (\hat{v} - v))'Z = n^{-1}v'Z + n^{-1}(\hat{v} - v)'Z \qquad \text{(A.3)} \\ &= n^{-1}v'Z + (\pi_{n,h} - \hat{\pi}_{n,h})'(n^{-1}Z'Z) = O_P(n^{-1/2}) + O_P(n^{-1/2})O_P(1) = O_P(n^{-1/2}) \\ n^{-1}\hat{v}'P_XZ &= n^{-1}v'P_XZ + n^{-1}(\hat{v} - v)'P_XZ = (n^{-1}v'Z\pi_{n,h} + n^{-1}v'v)(n^{-1}X'X)^{-1}(n^{-1}X'Z) + \\ &\quad n^{-1}(\hat{v} - v)'P_XZ = \sigma_v^2\left(\sigma_x^2\right)^{-1}\sigma_{xz} + O_P(n^{-1/2}), \qquad \text{(A.4)} \end{aligned}$$

where σ_v^2, σ_x^2 and σ_{xz} denote the probability limits of $n^{-1}v'v$, $n^{-1}X'X$ and $n^{-1}X'Z$, respectively. The $O_P(n^{-1/2})$ term in (A.4) is justified by the fact that

$$n^{-1}(\hat{v} - v)'P_XZ = (\pi_{n,h} - \hat{\pi}_{n,h})'(n^{-1}Z'X)(n^{-1}X'X)^{-1}(n^{-1}X'Z) = O_P(n^{-1/2}). \tag{A.5}$$

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

$$\left(n^{-1}\hat{v}'M_X\hat{v}\right)^{-1}\left(n^{-1/2}\hat{v}'M_X(v - \hat{v})\right)a_{n,h} = o_P(1), \tag{A.6}$$

so that

$$n^{1/2}(\hat{a} - a_{n,h}) = \left(n^{-1}\hat{v}'M_X\hat{v}\right)^{-1}\left(n^{-1/2}\hat{v}'M_Xe\right) + o_P(1), \tag{A.7}$$

as stated.

$\square$

Lemma A.2 *Under the drift sequences of parameters* $\{\gamma_{n,h}\}$ *in (2.19) with* $|h_1| < \infty$, *the following results hold*:

(*a*) *Asymptotic distributions of the estimators*

$$\begin{pmatrix} n^{1/2}\hat{a} \\ n^{1/2}(\hat{\theta}_{OLS}-\theta) \\ n^{1/2}(\hat{\theta}_{2SLS}-\theta) \end{pmatrix} \to^d \begin{pmatrix} \psi_a \\ \psi_{OLS} \\ \psi_{2SLS}, \end{pmatrix} \equiv \begin{pmatrix} -\left(h_{21}'h_{24}h_{21}\right)^{-1}h_{21}'\psi_{Zu}(0)+h_{25}^{-1}\psi_{vu}(0)+h_1 \\ (h_{21}'h_{24}h_{21}+h_{25})^{-1}\left(h_{21}'\psi_{Zu}(0)+\psi_{vu}(0)+h_{25}h_1\right) \\ \left(h_{21}'h_{24}h_{21}\right)^{-1}h_{21}'\psi_{Zu}(0) \end{pmatrix}$$

where

$$\begin{aligned} \psi_a &\sim N\Big(h_1, \left(h_{21}'h_{24}h_{21}\right)^{-2}h_{21}'h_{22}h_{21}+h_{25}^{-2}h_{23}\Big) \\ \psi_{OLS} &\sim N\Big(h_{25}h_1/\left(h_{21}'h_{24}h_{21}+h_{25}\right), (h_{21}'h_{22}h_{21}+h_{23})/\left(h_{21}'h_{24}h_{21}+h_{25}\right)^2\Big) \\ \psi_{2SLS} &\sim N\Big(0, \left(h_{21}'h_{24}h_{21}\right)^{-2}h_{21}'h_{22}h_{21}\Big). \end{aligned}$$

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

$$\begin{aligned} \begin{pmatrix} T_{2SLS}(\theta_0) \\ T_{OLS}(\theta_0) \\ H_n \end{pmatrix} &\overset{d}{\to} \eta_h = \begin{pmatrix} \eta_{1,h} \\ \eta_{2,h} \\ \eta_{3,h} \end{pmatrix} \\ &\equiv \begin{pmatrix} \left(h_{21}'h_{22}h_{21}\right)^{-1/2}h_{21}'\psi_{Zu}(0) \\ (h_{21}'h_{22}h_{21}+h_{23})^{-1/2}\left(h_{21}'\psi_{Zu}(0)+\psi_{vu}(0)+h_{25}h_1\right) \\ \left(\frac{h_{21}'h_{22}h_{21}}{(h_{21}'h_{24}h_{21})^2}+h_{23}h_{25}^{-2}\right)^{-1}\left(-(h_{21}'h_{24}h_{21})^{-1}h_{21}'\psi_{Zu}(0)+h_{25}^{-1}\psi_{uv}(0)+h_1\right)^2 \end{pmatrix} \\ T_n(\theta_0) &\overset{d}{\to} \tilde{T}_h = \Big|\eta_{2,h}\mathbb{1}(\eta_{3,h} \le \chi^2_{1,1-\beta})+\eta_{1,h}\mathbb{1}(\eta_{3,h} > \chi^2_{1,1-\beta})\Big|, \end{aligned}$$

where

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

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 $n^{1/2}(\hat{a}-a_{n,h})$ 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

$$n^{-1}\hat{v}'M_X\hat{v} = n^{-1}\hat{X}'M_X\hat{X} = n^{-1}\hat{X}'\hat{X} - n^{-1}\hat{X}'P_Z\hat{X}, \tag{A.8}$$

where

$$\begin{aligned} n^{-1}\hat{X}'\hat{X} &= n^{-1}X'P_ZX \to^P \pi'\Omega_{ZZ}\pi \equiv \sigma_\pi^2 & \text{(A.9)} \\ n^{-1}\hat{X}'P_X\hat{X} &= (n^{-1}\hat{X}'X)(n^{-1}X'X)^{-1}(n^{-1}X'\hat{X}) \to^P \sigma_\pi^4(\sigma_\pi^2+\sigma_v^2)^{-1}. & \text{(A.10)} \end{aligned}$$

Therefore, we have for the denominator

$$n^{-1}\hat{v}'M_X\hat{v} \quad \to^P \quad \sigma_\pi^2 - \sigma_\pi^4(\sigma_\pi^2+\sigma_v^2)^{-1} = \sigma_\pi^2\sigma_v^2\left(\sigma_\pi^2+\sigma_v^2\right)^{-1} \tag{A.11}$$

For the numerator, note that

$$n^{-1/2}\hat{v}'M_Xe = -n^{-1/2}\hat{X}'M_Xe = -n^{-1/2}\hat{X}'e + n^{-1/2}\hat{X}'P_Xe. \tag{A.12}$$

The first term is such that

$$\begin{aligned} n^{-1/2}\hat{X}'e &= -\left(n^{-1}X'Z\right)\left(n^{-1}Z'Z\right)^{-1}\left(n^{-1/2}Z'e\right) \\ &\to^d -\pi'\Omega_{ZZ}\Omega_{ZZ}^{-1}\psi_{Ze} = -\pi'\psi_{Ze}, \end{aligned} \tag{A.13}$$

and the second term is such that

$$n^{-1/2}\hat{X}'P_Xe \quad = \quad \left(n^{-1}X'P_ZX\right)\left(n^{-1}X'X\right)^{-1}\left(n^{-1/2}X'e\right) \to^d \sigma_\pi^2(\sigma_\pi^2+\sigma_v^2)^{-1}\left(\pi'\psi_{Ze}+\psi_{ve}\right), \tag{A.14}$$

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

$$\begin{aligned} n^{-1/2}\hat{X}'M_Xe &\to^d -\pi'\psi_{Ze} + \sigma_\pi^2(\sigma_\pi^2+\sigma_v^2)^{-1}\left(\pi'\psi_{Ze}+\psi_{ve}\right) \\ &= -\sigma_v^2(\sigma_\pi^2+\sigma_v^2)^{-1}\pi'\psi_{Ze} + \sigma_\pi^2(\sigma_\pi^2+\sigma_v^2)^{-1}\psi_{ve}. \end{aligned} \tag{A.15}$$

By combining (A.11) and (A.15), we obtain

$$\begin{aligned} n^{1/2}(\hat{a}-a_{n,h}) &\to^d -\sigma_\pi^{-2}\pi'\psi_{Ze} + \sigma_v^{-2}\psi_{ve} \\ &\equiv -\left(h_{21}'h_{24}h_{21}\right)^{-1}h_{21}'\psi_{Ze} + h_{25}^{-1}\psi_{ve} \\ &\sim N\left(0, (h_{21}'h_{24}h_{21})^{-2}h_{21}'h_{22}h_{21} + h_{25}^{-2}h_{23}\right). \end{aligned} \tag{A.16}$$

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

$$\begin{aligned} n^{1/2}\hat{a} &\to^d \psi_a = -\left(h_{21}'h_{24}h_{21}\right)^{-1}h_{21}'\psi_{Zu}(0) + h_{25}^{-1}\psi_{vu}(0) + h_1 \\ &\sim N\Big(h_1, (h_{21}'h_{24}h_{21})^{-2}h_{21}'h_{22}h_{21} + h_{25}^{-2}h_{23}\Big), \end{aligned} \tag{A.17}$$

with $\psi_{Zu}(0)=\psi_{Ze}$ and $\psi_{vu}(0)=\psi_{ve}$.

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

$$n^{1/2}(\hat{\theta}_{OLS}-\theta)=\left(n^{-1}X'X\right)^{-1}\left(n^{-1/2}X'u\right), \tag{A.18}$$

where $n^{-1}X'X\to^{P}\sigma_{\pi}^{2}+\sigma_{v}^{2}=h_{21}'h_{24}h_{21}+h_{25}$, and

$$\begin{aligned} n^{-1/2}X'u &= n^{-1/2}(\pi_{n,h}'Z'+v')(va_{n,h}+e) \\ &= \pi_{n,h}'\left(n^{-1/2}Z'e\right)+\pi_{n,h}'\left(n^{-1/2}Z'v\right)a_{n,h}+\left(n^{-1}v'v\right)n^{1/2}a_{n,h}+n^{-1/2}v'e \\ &\to^{d} h_{21}'\psi_{Ze}+\psi_{ve}+h_{25}h_{1}, \end{aligned} \tag{A.19}$$

as $n^{-1/2}Z'e\to^{d}\psi_{Ze}$, $n^{-1/2}v'e\to^{d}\psi_{ve}$, $\pi_{n,h}'(n^{-1/2}Z'v)a_{n,h}=o_{P}(1)$, $n^{-1}(v'v)=h_{25}+o_{P}(1)$, and $n^{1/2}a_{n,h}\to h_{1}$ as $n\to\infty$.

Therefore, we obtain

$$\begin{aligned} n^{1/2}(\hat{\theta}_{OLS}-\theta) &\to^{d} \psi_{OLS}=(h_{21}'h_{24}h_{21}+h_{25})^{-1}(h_{21}'\psi_{Zu}(0)+\psi_{vu}(0)+h_{25}h_{1}) \\ &\sim N\left(\frac{h_{25}h_{1}}{h_{21}'h_{24}h_{21}+h_{25}},\frac{h_{21}'h_{22}h_{21}+h_{23}}{(h_{21}'h_{24}h_{21}+h_{25})^{2}}\right) \end{aligned} \tag{A.20}$$

with $\psi_{Zu}(0)=\psi_{Ze}$ and $\psi_{vu}(0)=\psi_{ve}$.

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

$$n^{1/2}(\hat{\theta}_{2SLS}-\theta)\to^{d}\psi_{2SLS}=(h_{21}'h_{24}h_{21})^{-1}h_{21}'\psi_{Ze}\sim N\left(0,(h_{21}'h_{24}h_{21})^{-2}h_{21}'h_{22}h_{21}\right). \tag{A.21}$$

(b) It also suffices to characterize the asymptotic distributions of each statistic separately.

First, note that $T_{l}(\theta)=n^{1/2}(\hat{\theta}_{l}-\theta)/\hat{V}^{1/2}(\hat{\theta}_{l})$, $l\in\{OLS,2SLS\}$. Since

$$\hat{V}(\hat{\theta}_{OLS})\to^{P}\frac{h_{21}'h_{22}h_{21}+h_{23}}{(h_{21}'h_{24}h_{21}+h_{25})^{2}},\text{ and }\hat{V}(\hat{\theta}_{2SLS})\to^{P}\frac{h_{21}'h_{22}h_{21}}{(h_{21}'h_{24}h_{21})^{2}}, \tag{A.22}$$

the results of $T_{OLS}(\theta)$ and $T_{2SLS}(\theta)$ follow immediately from the proof of part (a) along with the fact that $\psi_{Zu}(0)=\psi_{Ze}$ and $\psi_{vu}(0)=\psi_{ve}$.

Furthermore, we notice that H_{n} is defined as $H_{n}=n\hat{a}^{2}/\hat{V}(\hat{a})$, where the variance estimator $\hat{V}(\hat{a})=\left(n^{-1}\hat{v}'M_{X}\hat{v}\right)^{-1}\left(n^{-1}\sum_{i=1}^{n}\hat{v}_{i}^{2}\hat{e}_{i}^{2}\right)\left(n^{-1}\hat{v}'M_{X}\hat{v}\right)^{-1}$. We can also write H_{n} as:

$$H_{n}=\left(n^{1/2}(\hat{a}-a_{n,h})+n^{1/2}a_{n,h}\right)^{2}/\hat{V}(\hat{a}), \tag{A.23}$$

where $\left(n^{1/2}(\hat{a}-a_{n,h})+n^{1/2}a_{n,h}\right)^{2}\to^{d}\psi_{a}^{2}$ from Lemma **A.2**-(a). Similarly, we can show that

$$\hat{V}(\hat{a})\to^{P}(h_{21}'h_{24}h_{21})^{-2}h_{21}'h_{22}h_{21}+h_{25}^{-2}h_{23}, \tag{A.24}$$

so that

$$H_n \to^d \left(\frac{h_{21}'h_{22}h_{21}}{(h_{21}'h_{24}h_{21})^2}+h_{23}h_{25}^{-2}\right)^{-1}\left(-(h_{21}'h_{24}h_{21})^{-1}h_{21}'\psi_{Zu}(0)+h_{25}^{-1}\psi_{vu}(0)+h_1\right)^2, \tag{A.25}$$

where $\psi_{Zu}(0)=\psi_{Ze}$ and $\psi_{vu}(0)=\psi_{ve}$.

□

Lemma A.3 *If for some* $\delta>0$, $E^*\left[|e_{1i}^*|^{2(2+\delta)},|e_{2i}^*|^{2+\delta}\right]$ *is bounded in probability and* $E_F[w_i^{2+\delta}]<\infty$ *for all* $w_i\in\left\{||Z_iu_i(\theta_0)||,||Z_iv_i||,||Z_iZ_i'||,|u_i(\theta_0)|,|v_i|,|u_i(\theta_0)v_i|\right\}$, *then* $n^{-1}\sum_{i=1}^n E^*\left[||Z_iu_i^*||^{2+\delta}\right]$, $n^{-1}\sum_{i=1}^n E^*\left[||Z_iv_i^*||^{2+\delta}\right]$ *and* $n^{-1}\sum_{i=1}^n E^*\left[||u_i^*v_i^*||^{2+\delta}\right]$ *are bounded in probability.*

PROOF OF LEMMA **A.3**

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

$$\begin{aligned} n^{-1}\sum_{i=1}^n E^*\left[||Z_iu_i^*||^{2+\delta}\right] &= n^{-1}\sum_{i=1}^n E^*\left[||Z_iu_i(\theta_0)e_{1i}^*||^{2+\delta}\right]=n^{-1}\sum_{i=1}^n E^*\left[||Z_iu_i(\theta_0)||^{2+\delta}|e_{1i}^*|^{2+\delta}\right]\\ &= n^{-1}\sum_{i=1}^n ||Z_iu_i(\theta_0)||^{2+\delta}E^*\left[|e_{1i}^*|^{2+\delta}\right]\le C_1 n^{-1}\sum_{i=1}^n ||Z_iu_i(\theta_0)||^{2+\delta}\\ &\to^P C_1E_F\left[||Z_iu_i(\theta_0)||^{2+\delta}\right]<\infty \end{aligned} \tag{A.26}$$

for some large enough constant $C_1<\infty$. Now, consider the bound on $n^{-1}\sum_{i=1}^n E^*\left[||Z_iv_i^*||^{2+\delta}\right]$. As in (A.26) we have

$$\begin{aligned} n^{-1}\sum_{i=1}^n E^*\left[||Z_iv_i^*||^{2+\delta}\right] &= n^{-1}\sum_{i=1}^n E^*\left[||Z_i\hat v_i||^{2+\delta}|e_{ji}^*|^{2+\delta}\right]=n^{-1}\sum_{i=1}^n ||Z_i\hat v_i||^{2+\delta}E^*\left[|e_{ji}^*|^{2+\delta}\right]\\ &\le C_jn^{-1}\sum_{i=1}^n ||Z_i\hat v_i||^{2+\delta} \end{aligned} \tag{A.27}$$

in probability for some large enough constant $C_j<\infty$, where $j=1$ for the dependent bootstrap scheme and $j=2$ for the independent bootstrap scheme. By using Minkowski and Cauchy-Schwartz inequalities, along with the fact that $\hat v_i=v_i-Z_i'(\hat\pi-\pi)$, we obtain

$$\begin{aligned} n^{-1}\sum_{i=1}^n ||Z_i\hat v_i||^{2+\delta} &= n^{-1}\sum_{i=1}^n ||Z_iv_i-Z_iZ_i'(\hat\pi-\pi)||^{2+\delta}\\ &\le C\left\{n^{-1}\sum_{i=1}^n ||Z_iv_i||^{2+\delta}+||\hat\pi-\pi||^{2+\delta}n^{-1}\sum_{i=1}^n ||Z_iZ_i'||^{2+\delta}\right\},\\ &\to^P CE_F\left[||Z_iv_i||^{2+\delta}\right]<\infty, \end{aligned} \tag{A.28}$$

where $C<\infty$ denotes a large enough constant, and (A.28) holds because $n^{-1}\sum_{i=1}^n ||Z_iv_i||^{2+\delta}\to^P E_F\left[||Z_iv_i||^{2+\delta}\right]<\infty$, $n^{-1}\sum_{i=1}^n ||Z_iZ_i'||^{2+\delta}\to^P E_F\left[||Z_iZ_i'||^{2+\delta}\right]<\infty$, and $\hat\pi-\pi\to^P 0$. Therefore, $n^{-1}\sum_{i=1}^n E^*\left[||Z_iv_i^*||^{2+\delta}\right]$ is bounded in probability from (A.27)-(A.28).

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

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n} E^*\left[|u_i^* v_i^*|^{2+\delta}\right] &= n^{-1}\sum_{i=1}^{n} E^*\left[|u_i(\theta_0)\hat{v}_i|^{2+\delta}|e_{1i}^* e_{ji}^*|^{2+\delta}\right] \\
&= n^{-1}\sum_{i=1}^{n} |u_i(\theta_0)\hat{v}_i|^{2+\delta} E^*\left[|e_{1i}^* e_{ji}^*|^{2+\delta}\right] \qquad \text{(A.29)}
\end{aligned}
$$

for $j = 1,2$. For the independent bootstrap scheme, $j = 2$ so that $E^*\left[|e_{1i}^* e_{ji}^*|^{2+\delta}\right] = E^*\left[|e_{1i}^* e_{2i}^*|^{2+\delta}\right] \leq E^*\left[|e_{1i}^*|^{2+\delta}\right] E^*\left[|e_{2i}^*|^{2+\delta}\right] \leq C < \infty$ in probability for some large enough constant C. For the dependent bootstrap scheme, $j = 1$ and we have $E^*\left[|e_{1i}^* e_{ji}^*|^{2+\delta}\right] = E^*\left[|e_{1i}^{*^2}|^{2+\delta}\right] = E^*\left[|e_{1i}^*|^{2(2+\delta)}\right] \leq \bar{C} < \infty$ in probability for some large enough constant $\bar{C}$. Combining both cases into (A.29) along with the fact that $u_i(\theta_0)\hat{v}_i = u_i(\theta_0)v_i - u_i(\theta_0)Z_i'(\hat{\pi} - \pi)$, and using Minkowski and Cauchy-Schwartz inequalities, we get:

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n} E^*\left[|u_i^* v_i^*|^{2+\delta}\right] &\leq Bn^{-1}\sum_{i=1}^{n} |u_i(\theta_0)v_i - u_i(\theta_0)Z_i'(\hat{\pi} - \pi)|^{2+\delta} \\
&\leq D\left\{n^{-1}\sum_{i=1}^{n} |u_i(\theta_0)v_i|^{2+\delta} + ||\hat{\pi} - \pi||^{2+\delta} n^{-1}\sum_{i=1}^{n} ||Z_i u_i(\theta_0)||^{2+\delta}\right\} \\
&\to^P DE_F\left[|u_i(\theta_0)v_i|^{2+\delta}\right] < \infty \qquad \text{(A.30)}
\end{aligned}
$$

for some large enough constants $B < \infty$ and $D < \infty$. $\square$

Lemma A.4 *Suppose that the H_0 holds and the conditions of Lemma* **A.3** *are satisfied. Then, under the sequence $\{\gamma_{n,h}\}$ defined in* (2.19) *with $|h_1| < \infty$ we have*:

$$
\begin{pmatrix} n^{-1/2} Z' u^* \\ n^{-1/2} Z' v^* \\ n^{-1/2}\left(u^{*\prime} v^* - E^*\left[u^{*\prime} v^*\right]\right) \end{pmatrix} \to^{d^*} N\left(0, \begin{pmatrix} \Omega_{Zu}(0) & 0 & 0 \\ 0 & \Omega_{Zu} & 0 \\ 0 & 0 & \Omega_{vu}(0) \end{pmatrix}\right), \qquad \text{(A.31)}
$$

in probability.

PROOF OF LEMMA **A.4**

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

$$
\begin{aligned}
X_{n,i}^* &= \left\{c_1' u_i^* Z_i + c_2' v_i^* Z_i + d\left(u_i^* v_i^* - E^*[u_i^* v_i^*]\right)\right\}/\sqrt{n} \\
&= \left\{c_1' e_{1i}^* \hat{u}_i(\theta_0) Z_i + c_2' e_{ji}^* \hat{v}_i Z_i + d\left(\hat{u}_i(\theta_0)\hat{v}_i e_{1i}^* e_{ji}^* - E^*[\hat{u}_i(\theta_0)\hat{v}_i e_{1i}^* e_{ji}^*]\right)\right\}/\sqrt{n}, \qquad \text{(A.32)}
\end{aligned}
$$

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 Central Limit Theorem hold for $X_{n,i}^*$.

To simplify, we shall give the proof for the case with independent transformation, i.e., $j = 2$ in (A.32). Note that the proof for the case with dependent transformation ($j = 1$) follow similar steps.

(a) $E^*[X_{n,i}^*] = 0$ as $E^*[e_{1i}^* \hat{u}_i(\theta_0) Z_i] = \hat{u}_i(\theta_0) Z_i E^*[e_{1i}^*] = 0$, $E^*[e_{2i}^* \hat{v}_i Z_i] = \hat{u}_i(\theta_0) Z_i E^*[e_{2i}^*] = 0$, and $E^*[\hat{u}_i(\theta_0)\hat{v}_i e_{1i}^* e_{2i}^* - E^*[\hat{u}_i(\theta_0)\hat{v}_i e_{1i}^* e_{2i}^*]] = \hat{u}_i(\theta_0)\hat{v}_i E^*[e_{1i}^* e_{2i}^*] - \hat{u}_i(\theta_0)\hat{v}_i E^*[e_{1i}^* e_{2i}^*] = 0$. Note that $E^*[e_{1i}^* e_{2i}^*] = E^*[e_{1i}^*] E^*[e_{2i}^*] = 0$ under the independent transformation.

(b) Note that $E^*[u_i^{*^2}Z_iZ_i'] = E^*[\hat{u}_i^2(\theta_0)e_{1i}^{*^2}Z_iZ_i'] = \hat{u}_i^2(\theta_0)Z_iZ_i'E^*[e_{1i}^{*^2}] = \hat{u}_i^2(\theta_0)Z_iZ_i'$, $E^*[v_i^{*^2}Z_iZ_i'] = E^*[\hat{v}_i^2e_{2i}^{*^2}Z_iZ_i'] = \hat{v}_i^2Z_iZ_i'$, $E^*[u_i^{*^2}v_i^{*^2}] = E^*[\hat{u}_i^2(\theta_0)\hat{v}_i^2e_{1i}^{*^2}e_{2i}^{*^2}] = \hat{u}_i^2(\theta_0)\hat{v}_i^2E^*[e_{1i}^{*^2}e_{2i}^{*^2}] = \hat{u}_i^2(\theta_0)\hat{v}_i^2$, $E^*[u_i^*v_i^*Z_iZ_i'] = E^*[\hat{u}_i(\theta_0)\hat{v}_ie_{1i}^*e_{2i}^*Z_iZ_i'] = \hat{u}_i(\theta_0)\hat{v}_iZ_iZ_i'E^*[e_{1i}^*e_{2i}^*] = 0$, $E^*[u_i^{*^2}v_i^*Z_i] = E^*[\hat{u}_i^2(\theta_0)\hat{v}_ie_{1i}^{*^2}e_{2i}^*Z_i] = \hat{u}_i^2(\theta_0)\hat{v}_iZ_iE^*[e_{1i}^{*^2}e_{2i}^*] = 0$, and similarly $E^*[v_i^{*^2}u_i^*Z_i] = 0$. So, we have:

$$\begin{aligned}\sum_{i=1}^n E^*[X_{n,i}^{*^2}] &= c_1'\left(n^{-1}\sum_{i=1}^n \hat{u}_i^2(\theta_0)Z_iZ_i'\right)c_1 + c_2'\left(n^{-1}\sum_{i=1}^n \hat{v}_i^2Z_iZ_i'\right)c_2 + d^2\left(n^{-1}\sum_{i=1}^n \hat{u}_i^2(\theta_0)\hat{v}_i^2\right)\\ &= c_1'\Omega_{Zu}(0)c_1 + c_2'\Omega_{Zv}c_2 + d^2\Omega_{vu}(0) + o_P(1) \equiv O_P(1). \qquad \text{(A.33)}\end{aligned}$$

(c) For some $\delta > 0$, we note that

$$\begin{aligned}&\sum_{i=1}^n E^*[\left|X_{n,i}^*\right|^{2+\delta}]\\ \leq\;& Cn^{-\frac{\delta}{2}}n^{-1}\sum_{i=1}^n E^*\left[\left|c_1'u_i^*Z_i + c_2'v_i^*Z_i\right|^{2+\delta} + |du_i^*v_i^*|^{2+\delta}\right]\\ \leq\;& Cn^{-\frac{\delta}{2}}n^{-1}\sum_{i=1}^n E^*\left[\left|c_1'Z_i^*u_i^*\right|^{2+\delta} + \left|c_2'Z_i^*v_i^*\right|^{2+\delta} + |du_i^*v_i^*|^{2+\delta}\right]\\ =\;& Cn^{-\frac{\delta}{2}}n^{-1}\sum_{i=1}^n E^*\left[\left|c_1'Z_i\hat{u}_i(\theta_0)e_{1i}^*\right|^{2+\delta} + \left|c_2'Z_i\hat{v}_ie_{2i}^*\right|^{2+\delta} + |d\hat{v}_i\hat{u}_i(\theta_0)e_{1i}^*e_{2i}^*|^{2+\delta}\right]\\ =\;& Cn^{-\frac{\delta}{2}}\left[n^{-1}\sum_{i=1}^n \left|c_1'Z_i\hat{u}_i(\theta_0)\right|^{2+\delta}E^*[|e_{1i}^*|^{2+\delta}] + n^{-1}\sum_{i=1}^n\left|c_2'Z_i\hat{v}_i\right|^{2+\delta}E^*[|e_{2i}^*|^{2+\delta}]\right.\\ &\left.\quad + n^{-1}\sum_{i=1}^n |d\hat{v}_i\hat{u}_i(\theta_0)|^{2+\delta}E^*[|e_{1i}^*e_{2i}^*|^{2+\delta}]\right]\\ =\;& Cn^{-\frac{\delta}{2}}\left[E^*[|e_{1i}^*|^{2+\delta}]E_F\left|c_1'Z_iu_i\right|^{2+\delta} + E^*[|e_{2i}^*|^{2+\delta}]E_F\left|c_2'Z_iu_i\right|^{2+\delta}\right.\\ &\left.\quad + E^*[|e_{1i}^*e_{2i}^*|^{2+\delta}]E_F\left|dv_iu_i\right|^{2+\delta}\right] + o_P(1) \to^P 0, \qquad \text{(A.34)}\end{aligned}$$

i.e., $\sum_{i=1}^n E^*[\left|X_{n,i}^*\right|^{2+\delta}] \to^P 0$, where the convergence in probability is obtained by using Lemma **A.3**.

From (a)-(c) above, $X_{n,i}^*$ satisfies the Lyapunov CLT conditions [see e.g. Theorem 14 in Ruud (2000)]. Lemma **A.4** follows by applying this CLT. □

PROOF OF THEOREM **3.1**

First, we note that

$$\begin{aligned}n^{-1/2}X^{*'}P_Zu^* &= n^{-1/2}(Z\hat{\pi} + v^*)'P_Zu^*\\ &= n^{-1/2}\hat{\pi}'Z'u^* + n^{-1/2}\left(n^{-1/2}v^{*'}Z\right)\left(n^{-1}Z'Z\right)^{-1}\left(n^{-1/2}Z'u^*\right)\\ &= n^{-1/2}\hat{\pi}'Z'u^* + O_{P^*}\left(n^{-1/2}\right) \to^{d^*} \pi'\psi_{Zu}^*, \qquad \text{(A.35)}\\ & \qquad \text{(A.36)}\end{aligned}$$

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

Second, following the same as above, we have

$$\begin{aligned} n^{-1/2}X^{*'}u^{*} &= n^{-1/2}\hat{\pi}'Z^{'}u^{*}+n^{-1/2}\left(v^{*'}u^{*}-E^{*}[v^{*'}u^{*}]\right)+n^{-1/2}E^{*}[v^{*'}u^{*}] \\ &\to^{d^*} \pi'\psi_{Zu}^{*}+\psi_{uv}^{*}+h_1^{b}, \end{aligned} \tag{A.37}$$

in probability, where $h_1^b=0$ for the independent transformation and $h_1^b=h_1+\psi_{uv}$ for the dependent transformation. This is because $n^{-1/2}E^{*}[v^{*'}u^{*}]=0$ for the independent transformation, and $n^{-1/2}E^{*}[v^{*'}u^{*}]=n^{1/2}\left(n^{-1}\sum_{i=1}^{n}\hat{v}_i\hat{u}_i(\theta_0)\right)$ for the dependent transformation, where

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

Third, we note that

$$\begin{aligned} n^{-1}X^{*'}P_ZX^{*} &= n^{-1}\left(Z\hat{\pi}+v^{*}\right)'P_Z\left(Z\hat{\pi}+v^{*}\right) \\ &= n^{-1}\hat{\pi}'Z^{'}Z\hat{\pi}+n^{-1}\hat{\pi}'Z^{'}v^{*}+n^{-1}v^{*'}Z\hat{\pi}+n^{-1}v^{*'}P_Zv^{*} \\ &= n^{-1}\hat{\pi}'Z^{'}Z\hat{\pi}+O_{P^*}\left(n^{-1/2}\right)+O_{P^*}\left(n^{-1/2}\right)+O_{P^*}\left(n^{-1}\right) \\ &\to^{P^*} \sigma_{\pi}^{2}=\pi'\Omega_{ZZ}\pi=h_{21}^{'}h_{24}h_{21} \text{ in probability.} \end{aligned} \tag{A.38}$$

Using similar arguments, we obtain

$$n^{-1}X^{*'}X^{*} \to^{P^*} h_{25}+h_{21}^{'}h_{24}h_{21}=(1+h_{25}^{-1}h_{21}^{'}h_{24}h_{21})h_{25}, \tag{A.39}$$

in probability. Combining (A.35)-(A.39), along with the expression of the bootstrap OLS and 2SLS estimators, we obtain:

$$\begin{aligned} n^{1/2}(\hat{\theta}_{OLS}^{*}-\theta_0) &\to^{d^*} \psi_{OLS}^{*}=(1+h_{25}^{-1}h_{21}^{'}h_{24}h_{21})^{-1}h_{25}^{-1}(h_{21}^{'}\psi_{Zu}^{*}+\psi_{vu}^{*}+h_{25}h_1^{b}) \\ n^{1/2}(\hat{\theta}_{2SLS}^{*}-\theta_0) &\to^{d^*} \psi_{2SLS}^{*}=(h_{21}^{'}h_{24}h_{21})^{-1}h_{21}^{'}\psi_{Zu}^{*} \end{aligned} \tag{A.40}$$

in probability. Following similar steps as in the derivation of (A.40), we find that

$$\begin{aligned} n^{1/2}\hat{a}^{*} &\to^{d^*} (h_{21}^{'}h_{24}h_{21})^{-1}(h_{21}^{'}\psi_{Zu}^{*}-h_{25}^{-1}h_{21}^{'}h_{24}h_{21}\psi_{vu}^{*})+h_1^{b}, \qquad \text{(A.41)} \\ \hat{V}^{*}(\hat{a}^{*}) &\to^{P^*} (h_{21}^{'}h_{24}h_{21})^{-1}\left((h_{25}^{-1}h_{21}^{'}h_{24}h_{21})^{-2}h_{21}^{'}h_{22}h_{21}+h_{23}\right), \\ \hat{V}^{*}(OLS) &\to^{P^*} (1+h_{25}^{-1}h_{21}^{'}h_{24}h_{21})^{-2}h_{25}^{-2}\left(h_{21}^{'}h_{22}h_{21}+h_{23}\right),\ \hat{V}^{*}(2SLS) \to^{P^*} h_{21}^{'}h_{24}h_{21})^{-2}h_{21}^{'}h_{22}h_{21} \end{aligned}$$

in probability. The desired results are obtained from (A.40)-(A.42), along with the expressions of the different bootstrap statistics. □

PROOF OF THEOREM **3.2**

We follow Andrews and Guggenberger (2010b) [e.g., the proof of Theorem 1; see also Guggenberger (2010a)], and note that there exists a "worst case sequence" $\gamma_n \in \Gamma$ such that:

$$\begin{aligned}
& AsySz[\hat{c}_n^*(1-\alpha)] \\
= \; & \limsup_{n\to\infty} \sup_{\gamma\in\Gamma} P_{\theta_0,\gamma}\left[T_n(\theta_0) > \hat{c}_n^*(1-\alpha)\right] \\
= \; & \limsup_{n\to\infty} P_{\theta_0,\gamma_n}\left[T_n(\theta_0) > \hat{c}_n^*(1-\alpha)\right] \\
= \; & \lim_{n\to\infty} P_{\theta_0,\gamma_{m_n}}\left[T_{m_n}(\theta_0) > \hat{c}_{m_n}^*(1-\alpha)\right] \qquad \text{(A.42)}
\end{aligned}$$

where the first equality holds by the definition of asymptotic size and the second by the choice of the sequence $\{\gamma_n : n \geq 1\}$. And $\{m_n : n \geq 1\}$ is a subsequence of $\{n : n \geq 1\}$; such a subsequence always exists. Furthermore, there exists a subsequence $\{\omega_n : \; n \geq 1\}$ of $\{m_n : \; n \geq 1\}$ such that:

$$\begin{aligned}
& \lim_{n\to\infty} P_{\theta_0,\gamma_{m_n}}\left[T_{m_n}(\theta_0) > \hat{c}_{m_n}^*(1-\alpha)\right] \\
= \; & \lim_{n\to\infty} P_{\theta_0,\gamma_{\omega_n,h}}\left[T_{\omega_n}(\theta_0) > \hat{c}_{\omega_n}^*(1-\alpha)\right] \qquad \text{(A.43)}
\end{aligned}$$

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 $\{\theta_{\omega_n,h} : n \geq 1\}$, we have $\left(T_{\omega_n}(\theta_0), \hat{c}_{\omega_n}^*(1-\alpha)\right) \xrightarrow{d} \left(\tilde{T}_h, c_h^*(1-\alpha)\right)$ jointly. It follows that $AsySz[\hat{c}_n^*(1-\alpha)] = sup_{h\in\mathscr{H}} P[\tilde{T}_h > c_h^*(1-\alpha)]$. □

PROOF OF THEOREM **3.3**

First, note that by following similar arguments as those in the proofs of Theorem **3.1**, we can obtain that the following (conditional) convergence in distribution holds:

$$\begin{pmatrix} T^*_{OLS,(h_1,\hat{h}_2)}(\theta_0) \\ H^*_{n,(h_1,\hat{h}_2)} \end{pmatrix} \rightarrow^{d^*} \begin{pmatrix} (\pi'\Omega_{Ze}\pi + \Omega_{ve})^{-1/2}\left(\pi'\psi^*_{Zu} + \psi^*_{vu} + \sigma_v^2 h_1\right) \\ \left(h_2^{-4}\pi'\Omega_{Ze}\pi + \Omega_{ve}\right)^{-1}\left(h_2^{-2}\pi'\psi^*_{Zu} - \psi^*_{vu} + \sigma_v^2 h_1\right)^2 \end{pmatrix}, \qquad \text{(A.44)}$$

in probability. Then, based on the formula of $T^*_{n,(h_1,\hat{h}_2)}(\theta_0)$, we conclude that the (conditional) null limiting distribution of $T^*_{n,(h_1,\hat{h}_2)}(\theta_0)$ is the same as the null limiting distribution of $T_n(\theta_0)$ with the value of localization parameter equal to h_1, and this implies that $c^*_{(h_1,\hat{h}_2)}(1-\delta) \rightarrow^P c_{(h_1,h_2)}(1-\delta)$, where $c_{(h_1,h_2)}(1-\delta)$ denotes the $(1-\delta)$-th quantile of $\tilde{T}_h$ with $h = (h_1, h_2)$.

Then, the proof is similar to the proof for Theorem **3.2** and those in McCloskey (2017). We note that there exists a "worst case sequence" $\gamma_n \in \Gamma$ such that:

$$\begin{aligned}
& AsySz\left[c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})\right] \\
= \; & \limsup_{n\to\infty} \sup_{\gamma\in\Gamma} P_{\theta_0,\gamma}\left[T_n(\theta_0) > c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})\right] \\
= \; & \limsup_{n\to\infty} P_{\theta_0,\gamma_n}\left[T_n(\theta_0) > c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})\right] \\
= \; & \lim_{n\to\infty} P_{\theta_0,\gamma_{m_n}}\left[T_{m_n}(\theta_0) > c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{m_n,1}, \hat{h}_{m_n,2})\right] \qquad \text{(A.45)}
\end{aligned}$$

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:

$$\begin{aligned} & \lim_{n\to\infty} P_{\theta_0,\gamma_{m_n}} \left[T_{m_n}(\theta_0) > c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{m_n,1}, \hat{h}_{m_n,2})\right] \\ = & \lim_{n\to\infty} P_{\theta_0,\gamma_{\omega_n,h}} \left[T_{\omega_n}(\theta_0) > c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{\omega_n,1}, \hat{h}_{\omega_n,2})\right] \end{aligned} \tag{A.46}$$

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 $\left(T_{\omega_n}(\theta_0), \hat{h}_{\omega_n,1}\right) \xrightarrow{d} \left(\tilde{T}_h, \tilde{h}_1\right)$ jointly. In addition, $c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{\omega_n,1}, \hat{h}_{\omega_n,2})$ 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:

$$\left(T_{\omega_n}(\theta_0), c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{\omega_n,1}, \hat{h}_{\omega_n,2})\right) \quad \xrightarrow{d} \quad \left(\tilde{T}_h, c^{B-S}(\alpha, \alpha-\delta, \tilde{h}_1, h_2)\right) \tag{A.47}$$

where $c^{B-S}(\alpha, \alpha-\delta, \tilde{h}_1, h_2) = \sup_{h_1 \in CI_{\alpha-\delta}(\tilde{h}_1)} c_{(h_1,h_2)}(1-\delta)$. Then, (A.45)-(A.47) imply that

$$\begin{aligned} & AsySz\left[c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})\right] \\ = & \lim_{n\to\infty} P_{\theta_0,\gamma_{\omega_n,h}} \left[T_{\omega_n}(\theta_0) > c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{\omega_n,1}, \hat{h}_{\omega_n,2})\right] \\ = & \sup_{h\in\mathscr{H}} P\left[\tilde{T}_h > c^{B-S}(\alpha, \alpha-\delta, \tilde{h}_1, h_2)\right], \end{aligned} \tag{A.48}$$

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

$$\begin{aligned} & P\left[\tilde{T}_h \geq c^{B-S}(\alpha, \alpha-\delta, \tilde{h}_1, h_2)\right] \\ = & P\left[\tilde{T}_h \geq c^{B-S}(\alpha, \alpha-\delta, \tilde{h}_1, h_2) \geq c_h(1-\delta)\right] \\ + & P\left[\tilde{T}_h \geq c_h(1-\delta) \geq c^{B-S}(\alpha, \alpha-\delta, \tilde{h}_1, h_2)\right] \\ + & P\left[c_h(1-\delta) \geq \tilde{T}_h \geq c^{B-S}(\alpha, \alpha-\delta, \tilde{h}_1, h_2)\right] \\ \leq & P\left[\tilde{T}_h \geq c_h(1-\delta)\right] + P\left[c_h(1-\delta) \geq c^{B-S}(\alpha, \alpha-\delta, \tilde{h}_1, h_2)\right] \\ = & P\left[\tilde{T}_h \geq c_h(1-\delta)\right] + P\left[h_1 \notin CI_{\alpha-\delta}(\tilde{h}_1)\right] \\ = & \delta + (\alpha-\delta) = \alpha, \end{aligned} \tag{A.49}$$

where the inequality and the second equality follow from the form of $c^{B-S}(\alpha, \alpha-\delta, \tilde{h}_1, h_2)$, and the third equality follows from the definition of $CI_{\alpha-\delta}(\tilde{h}_1)$. As (A.49) holds for any $h \in \mathscr{H}$, it is clear from (A.48) that $AsySz[c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})] \leq \alpha$, as stated. □

PROOF OF THEOREM **3.4** As in Theorem **3.3**, we can show that there exists a sequence $\gamma_n \in \Gamma$, a subsequence $\{m_n : n \geq 1\}$ of $\{n : n \geq 1\}$, and a subsubsequnce $\{\omega_n : n \geq 1\}$ of $\{m_n : n \geq 1\}$ such that the following result holds:

$$\begin{aligned} & AsySz\left[c^{B-C}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})\right] \\ = & \limsup_{n\to\infty} \sup_{\gamma\in\Gamma} P_{\theta_0,\gamma}\left[T_n(\theta_0) > c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2}) + \hat{\eta}_n\right] \\ = & \limsup_{n\to\infty} P_{\theta_0,\gamma_n}\left[T_n(\theta_0) > c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2}) + \hat{\eta}_n\right] \end{aligned}$$

$$
\begin{aligned}
&= \lim_{n\to\infty} P_{\theta_0,\gamma_{m_n}}\left[T_{m_n}(\theta_0) > c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{m_n,1},\hat{h}_{m_n,2}) + \hat{\eta}_{m_n}\right] \\
&= \lim_{n\to\infty} P_{\theta_0,\gamma_{\omega_n,h}}\left[T_{\omega_n}(\theta_0) > c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{\omega_n,1},\hat{h}_{\omega_n,2}) + \hat{\eta}_{\omega_n}\right] \qquad \text{(A.50)}
\end{aligned}
$$

for some $h \in \mathscr{H}$. Furthermore, as in the proof of Theorem **3.3**, for any $h \in \mathscr{H}_h$, any subsequence $\{\omega_n : n \geq 1\}$ of $\{n : n \geq 1\}$, and any sequence $\{\gamma_{\omega_n,h} : n \geq 1\}$, we have $\left(T_{\omega_n}(\theta_0), \hat{h}_{\omega_n,1}\right) \xrightarrow{d} \left(\tilde{T}_h, \tilde{h}_1\right)$ jointly. Hence,

$$
\begin{aligned}
& \lim_{n\to\infty} P_{\theta_0,\gamma_{\omega_n,h}}\left[T_{\omega_n}(\theta_0) > c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{\omega_n,1},\hat{h}_{\omega_n,2}) + \hat{\eta}_{\omega_n}\right] \\
&= \sup_{h\in\mathscr{H}} P\left[\tilde{T}_h > c^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2) + \bar{\eta}\right] \qquad \text{(A.51)} \\
&\equiv \sup_{h\in\mathscr{H}} P\left[\tilde{T}_h > c^{B-C}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)\right], \qquad \text{(A.52)}
\end{aligned}
$$

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

$$
NRP[h,\eta] \equiv P[\tilde{T}_h > c^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2) + \eta]. \qquad \text{(A.53)}
$$

It is clear from (A.50)-(A.53) that $AsySz[c^{B-C}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})] = \sup_{h\in\mathscr{H}} NRP[h,\bar{\eta}]$. Hence, it suffices to show that $\sup_{h\in\mathscr{H}} NRP[h,\bar{\eta}] = \alpha$ to establish Theorem **3.4**.

First, from the result of Theorem **3.3** and the definition of the size-correction criterion, it is clear that $\sup_{h\in\mathscr{H}} NRP[h,\bar{\eta}] \leq \alpha$. We proceed to show that $\sup_{h\in\mathscr{H}} NRP[h,\bar{\eta}] < \alpha$ leads to contradiction. Assume that $\sup_{h\in\mathscr{H}} NRP[h,\bar{\eta}] < \alpha$ and define the function $K(\cdot) : \mathbb{R}_- \to [-\alpha,\ 1-\alpha]$ such that

$$
K(x) = \sup_{h\in\mathscr{H}} NRP[h,x] - \alpha. \qquad \text{(A.54)}
$$

As $NRP[h,\cdot]$ is continuous on $\mathbb{R}_-$, the Maximum Theorem entails that $K(\cdot)$ is also continuous on $\mathbb{R}_-$. Moreover, we have

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

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

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

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

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

However, this contradicts the size-correction procedure where

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

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