THE UNIVERSITY of ADELAIDE

School of Economics

Working Papers

ISSN 2203-6024

Uniform Inference after Pretesting for Exogeneity

Firmin Doko Tchatoka
School of Economics
University of Adelaide

Wenjie Wang
Division of Economics
School of Social Sciences
Nanyang Technological University

Working Paper No. 2020-5
March 2020

Uniform Inference after Pretesting for Exogeneity

Firmin Doko Tchatoka*
The University of Adelaide

Wenjie Wang †
Nanyang Technological University

March 18, 2020

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 residual 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 residual-based bootstrap along with an adjusted Bonferroni size-correction method. We establish uniform validity of this hybrid bootstrap 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 (2019) 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, 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 (2019)'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 (2019)'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 invalid when IVs are weak; see e.g., Andrews and Guggenberger (2010b) and Wang and Doko Tchatoka (2018). Here, by deriving the null limiting distributions of the bootstrap test statistics and their associated asymptotic sizes, we show that the (residual-based) 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).

To address the bootstrap failure, we propose a novel hybrid bootstrap procedure, which makes use of the standard 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 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 is closely related to recent 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 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

$$
\begin{aligned}
y_1 &= y_2\theta + X\zeta + u, && (2.1)\\
y_2 &= Z\pi + X\phi + v, && (2.2)
\end{aligned}
$$

where $y_1, y_2 \in R^n$, $X \in R^{n\times k_1}$ is a matrix of exogenous variables, $Z \in R^{n\times k_2}$ is a matrix of instruments ($k_2 \geq 1$), $(\theta, \zeta', \phi', \pi')' \in R^{1\times k_1\times k_1\times k_2}$ are unknown parameters, and n is the sample size. We assume that the matrix $\bar{Z} = [X : Z] \in R^{n\times k}$ ($k = k_1 + k_2$) has full-column rank with probability one.

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– such as the one proposed by Durbin (1954), Wu (1973, 1974), and Hausman (1978)– 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. As discussed in the introduction, Guggenberger (2010a) shows this two-step methodology introduces substantial size distortions in the second stage t-test when standard asymptotic CVs are used (i.e., chi-squared CV for the pretest and standard normal CV for the second-stage t test), but we believe that this problem could be circumvented by using, for example, an appropriate size-correction technique.

To introduce the test statistics, define $W^{\perp} = M_X W$ for any matrix W with n rows. Let $\hat{\theta}_{2SLS} = y_2^{\perp\prime} P_{Z^{\perp}} y_1^{\perp} / (y_2^{\perp\prime} P_{Z^{\perp}} y_2^{\perp})$, and $\hat{\theta}_{OLS} = y_2^{\perp\prime} y_1^{\perp} / (y_2^{\perp\prime} y_2^{\perp})$ be the OLS and 2SLS estimators of θ respectively in (2.1). Also, define $\hat{V}_{2SLS} = (y_2^{\perp\prime} P_{Z^{\perp}} y_2^{\perp}/n)^{-1} \hat{\sigma}_u^2(\hat{\theta}_{2SLS})$ and $\hat{V}_{OLS} = (y_2^{\perp\prime} y_2^{\perp}/n)^{-1} \hat{\sigma}_u^2(\hat{\theta}_{OLS})$, where $\hat{\sigma}_u^2(\hat{\theta}_l) = n^{-1}(y_1^{\perp} - y_2^{\perp}\hat{\theta}_l)'(y_1^{\perp} - y_2^{\perp}\hat{\theta}_l), l \in \{OLS, 2SLS\}$ are the usual OLS-based 2SLS-based estimators of the variance of the structural error (both without correction for degrees of freedom). We consider the following Durbin-Wu-Hausman pretest statistic[3] for the exogeneity of y_2 in (2.1):

$$H_n = \frac{n(\hat{\theta}_{2SLS} - \hat{\theta}_{OLS})^2}{\hat{V}_{2SLS} - \hat{V}_{OLS}}. \tag{2.3}$$

The pretest reject the null hypothesis that y_2 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 y_2 is exogenous, H_n follows a χ^2_1 distribution asymptotically under assumptions given in Hausman (1978).

The two-stage test statistic associated with a pretest using H_n in the first-stage is given by:

$$\bar{T}_n(\theta_0) = T_{OLS}(\theta_0)\mathbb{1}(H_n \leq \chi^2_{1,1-\beta}) + T_{2SLS}(\theta_0)\mathbb{1}(H_n > \chi^2_{1,1-\beta}), \tag{2.4}$$

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

$$T_l(\theta) = n^{1/2}(\hat{\theta}_l - \theta)/\hat{V}_l^{1/2},\ l \in \{OLS, 2SLS\}. \tag{2.5}$$

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

$$T_n(\theta_0) > c_\infty(1-\alpha), \tag{2.6}$$

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). Importantly, as pointed out in these papers, 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

[3] Alternative formulations of this statistic are given in Hahn, Ham and Moon (2010); Doko Tchatoka and Dufour (2018, 2020) but we only keep that in (2.3) to shorten the presentation.

parameter space of γ_1 is $\varGamma_1$. (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 $\varGamma_2$. (iii) The third component γ_3 does not affect the limiting distribution (by virtue of the CLT). The parameter space of γ_3 is $\varGamma_3$, which in general may depend on γ_1 and γ_2, i.e., $\varGamma_3 \equiv \varGamma_3(\gamma_1, \gamma_2)$. 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 the application examples in Andrews and Guggenberger (2009, 2010a, 2010b)].

Assume that $(u_i, v_i, \bar{Z}_i), i = 1, ..., n$, are i.i.d. with distribution F. Define the vector of nuisance parameters $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ by

$$\gamma_1 = \rho,\ \gamma_2 = \| \varOmega^{1/2}\pi/\sigma_v \|,\ \gamma_3 = (F, \pi, \zeta, \phi), \tag{2.7}$$

where $\sigma_u^2 = E_F u_i^2$, $\sigma_v^2 = E_F v_i^2$, $\rho = Corr_F(u_i, v_i)$, $\varOmega = Q_{ZZ} - Q_{ZX} Q_{XX}^{-1} Q_{XZ}$, $Q = \begin{bmatrix} Q_{XX} & Q_{XZ} \\ Q_{ZX} & Q_{ZZ} \end{bmatrix} = E_F \bar{Z}_i \bar{Z}_i'$. Here, γ_1 measures the degree of endogeneity of y_2, and γ_2 measures the overall strength of the IVs.[5] Let

$$\varGamma_1 = [-1, 1],\ \varGamma_2 = [\underline{\kappa}, \overline{\kappa}] \tag{2.8}$$

for some $0 < \underline{\kappa} < \overline{\kappa} < \infty$. As the lower bound $\underline{\kappa}$ of $\varGamma_2$ is strictly positive, weak IV framework of Staiger and Stock (1997) is ruled out of the scope of the paper.[6] $\varGamma_3(\gamma_1, \gamma_2)$ is defined as follows:

$$\begin{aligned}
&\varGamma_3(\gamma_1, \gamma_2) = \big\{(F, \pi, \zeta, \phi):\ E_F u_i = E_F v_i = 0,\ E_F u_i^2 = \sigma_u^2,\ E_F v_i^2 = \sigma_v^2, \\
&E_F \bar{Z}_i \bar{Z}_i' = Q \text{ for some } \sigma_u^2, \sigma_v^2 > 0,\ \text{pd } Q \in R^{k \times k},\ \text{and } \pi \in R^{k_2} \\
&\text{that satisfy } Corr_F(u_i, v_i) = \gamma_1, ||\varOmega^{1/2}\pi/\sigma_v|| = \gamma_2,\ \zeta, \psi \in R^{k_1}; \\
&E_F u_i \bar{Z}_i = E_F v_i \bar{Z}_i = 0;\ E_F(u_i^2, v_i^2, u_i v_i)\bar{Z}_i \bar{Z}_i' = (\sigma_u^2, \sigma_v^2, \sigma_u \sigma_v \rho) Q; \\
&E_F(u_i^2 v_i \bar{Z}_i) = E_F(u_i v_i^2 \bar{Z}_i) = 0, var(u_i v_i)/(\sigma_u^2 \sigma_v^2) = 1 + \gamma_1^2; \\
&\lambda_{min}(E_F \bar{Z}_i \bar{Z}_i') \geq M^{-1};\ \left\| E_F \left(|u_i/\sigma_u|^{2+\delta}, |v_i/\sigma_v|^{2+\delta}, |u_i v_i/(\sigma_u \sigma_v)|^{2+\delta} \right)' \right\| \leq M, \\
&\left\| E_F \left(||\bar{Z}_i u_i/\sigma_u||^{2+\delta}, ||\bar{Z}_i v_i/\sigma_v||^{2+\delta}, ||\bar{Z}_i||^{2+\delta} \right)' \right\| \leq M \,\big\}
\end{aligned} \tag{2.9}$$

for some constant $\delta > 0$ and $M < \infty$, where pd denotes positive definite.

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

[5]Note that $\gamma_2 = (\mu^2/n)^{1/2}$, where μ^2 denotes the well-known concentration parameter in the IV literature.

[6]However, the Monte Carlo experiments (see Section 4) show that our proposed tests perform very well even when IVs are weak.

Finally, define the parameter space Γ of γ as

$$\Gamma = \{\gamma = (\gamma_1, \gamma_2, \gamma_3) : \gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2, \gamma_2 \in \Gamma_3(\gamma_1, \gamma_2)\}. \tag{2.10}$$

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

$$\underset{n\to\infty}{limsup} P_{\theta_0,\gamma}[T_n(\theta_0) > c_n], \tag{2.11}$$

and the asymptotic size of the test is given by

$$AsySz[c_n] = \underset{n\to\infty}{limsup} \underset{\gamma\in\Gamma}{sup} P_{\theta_0,\gamma}[T_n(\theta_0) > c_n]. \tag{2.12}$$

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 (2010a) shows 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

$$\begin{aligned}
\mathscr{H} &= \left\{h = (h_1, h_2) \in R_\infty^2 : \exists\{\gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3}) \in \Gamma : n \geq 1\} : n^{1/2}\gamma_{n,1} \to h_1,\ \ \gamma_{n,2} \to h_2\right\} \\
&\equiv \mathscr{H}_1 \times \mathscr{H}_2 \equiv R_\infty \times [\underline{\kappa}, \overline{\kappa}]
\end{aligned} \tag{2.13}$$

with $R_\infty = R \cup \{\pm\infty\}$. The relevant drifting sequences $\{\gamma_{n,h}\}$ are defined by Guggenberger (2010a) as follows: $\gamma_{n,h} \equiv (\gamma_{n,h,1}, \gamma_{n,h,2}, \gamma_{n,h,3})$ for $h = (h_1, h_2) \in \mathscr{H}$, where $\gamma_{n,h,1} = Corr_{F_n}(u_i, v_i)$, $\gamma_{n,h,2} = ||\Omega_n^{1/2}\pi_n/(E_{F_n}v_i^2)^{1/2}||$ with $\Omega_n = E_{F_n}Z_iZ_i' - E_{F_n}Z_iX_i'\left(E_{F_n}X_iX_i'\right)^{-1}E_{F_n}Z_iX_i'$ satisfy:

$$n^{1/2}\gamma_{n,h,1} \to h_1,\ \gamma_{n,h,2} \to h_2, \text{ and } \gamma_{n,h,3} = (F_n, \pi_n, \zeta_n, \phi_n) \in \Gamma_3(\gamma_{n,h,1}, \gamma_{n,h,2}). \tag{2.14}$$

Under H_0 and the drifting sequences $\{\gamma_{n,h} :\ h \in \mathscr{H}\}$ satisfying (2.14) with $|h_1| < \infty$ (i.e., local endogeneity), Guggenberger (2010a) shows that (for a symmetric two-sided test):

$$\begin{pmatrix} T_{2SLS}(\theta_0) \\ T_{OLS}(\theta_0) \\ H_n \end{pmatrix} \xrightarrow{d} \eta_h = \begin{pmatrix} s_{k_2}'\psi_u \\ (1+h_2^2)^{-1/2}(h_2 s_{k_2}'\psi_u + \psi_{uv} + h_1) \\ (1+h_2^2)^{-1}\left(s_{k_2}'\psi_u - h_2\psi_{uv} - h_2 h_1\right)^2 \end{pmatrix} \equiv \begin{pmatrix} \eta_{1,h} \\ \eta_{2,h} \\ \eta_{3,h} \end{pmatrix}, \tag{2.15}$$

$$T_n(\theta_0) \quad \xrightarrow{d} \quad \tilde{T}_h = \left|\eta_{2,h}\mathbb{1}(\eta_{3,h} \le \chi^2_{1,1-\beta}) + \eta_{1,h}\mathbb{1}(\eta_{3,h} > \chi^2_{1,1-\beta})\right|, \tag{2.16}$$

with $vec(\psi_u, \psi_{uv}) \sim N\left(0, \begin{pmatrix} I_{k_2} & 0 \\ 0' & 1 \end{pmatrix}\right)$, and $s_{k_2} \in R^{k_2}$ is an arbitrary vector with $||s_{k_2}|| = 1$. Note that $s'_{k_2}\psi_u - h_2\psi_{uv} - h_2 h_1 \sim N(-h_2h_1, 1+h_2^2)$, and the limiting distribution of H_n is therefore a noncentral chi-squared distribution $\chi^2_1(h_1^2h_2^2(h_2^2+1)^{-1})$, where the value of the noncentrality parameter determines the power of the Hausman pretest. As such, it can be shown from (2.16) that the asymptotic size of $T_n(\theta_0)$ (i.e., $AsySz[c_\infty(1-\alpha)]$) equals 1, i.e., the maximal rejection of the two-stage test is realized under certain drifting sequence $\{\gamma_{n,h} : h \in \mathscr{H}\}$ with local endogeneity. Such extreme size distortion occurs because when the Hausman test does not reject the null of exogeneity, OLS-based t-test is used in the second stage, but the maximal asymptotic rejection probability for $H_0 : \theta = \theta_0$ with the OLS-based t-test equals 1 [e.g., see the discussion in p.376 of Guggenberger (2010a)]. Also, similar result is shown for one-sided tests.

3. Main Results

3.1. Standard residual bootstrap

In this section, we study the asymptotic behaviour of standard (residual-based) 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.

Residual Bootstrap Algorithm:

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

$$\hat{v} = y_2^\perp - Z^\perp\hat{\pi}, \tag{3.1}$$
$$\hat{u}(\theta_0) = y_1^\perp - y_2^\perp\theta_0, \tag{3.2}$$

   where $\hat{\pi} = (Z^{\perp\prime}Z^\perp)^{-1}Z^{\perp\prime}y_2^\perp$ denotes the least squares estimator of π. We re-center these residuals by subtracting sample means to yield $(\tilde{u}(\theta_0), \tilde{v})$.

2. Generate the bootstrap pseudo-data following

$$y_2^{\perp *} = Z^{\perp *}\hat{\pi} + v^*, \tag{3.3}$$
$$y_1^{\perp *} = y_2^{\perp *}\theta_0 + u^*, \tag{3.4}$$

   where $Z^{\perp *}$ is drawn from the empirical distribution of $Z^\perp$, and there are two options to generate the bootstrap disturbances:

(a) v^* and u^* are drawn independently from the respective empirical distributions of $\tilde{v}$ and $\tilde{u}(\theta_0)$,

(b) (v^*, u^*) are drawn dependently from the joint empirical distribution of $(\tilde{v}, \tilde{u}(\theta_0))$.

Following Young (2019), we refer to (a) as *independent transformation* of distrubances and (b) as *dependent transformation* of disturbances. Note that for bootstrapping the Hausman test, for better size control it is usually recommended to generate (u^*, v^*) by the independent transformation, so that the bootstrap samples are obtained under the null hypothesis of the Hausman test (i.e., null 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):

$$T_n^*(\theta_0) = \left| T_{OLS}^*(\theta_0)\mathbb{1}(H_n^* \leq \chi^2_{1,1-\beta}) + T_{2SLS}^*(\theta_0)\mathbb{1}(H_n^* > \chi^2_{1,1-\beta}) \right|, \tag{3.5}$$

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 $T_n^*(\theta_0)$, $b = 1, ..., B$. 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.

Guggenberger (2010a) shows that the null limiting distribution of the two-stage test statistic $T_n(\theta_0)$ under the parameter sequences $\{\gamma_{n,h} : h \in \mathscr{H}\}$ satisfying (2.14) with $|h_1| < \infty$ is given by (2.16). To check whether the bootstrap consistently estimates the distribution of the two-stage test statistic, one needs to check whether we have

$$\sup_{x \in R} \left| P^*\left(T_n^*(\theta_0) \leq x\right) - P\left(T_n(\theta_0) \leq x\right) \right| \to^P 0 \tag{3.6}$$

under H_0 and such drifting parameter sequences.

First, we note that the following convergence results hold for the bootstrap statistics, conditional on the sample:

$$\begin{pmatrix} (n^{-1}Z^{\perp *\prime}Z^{\perp *})^{-1/2}\left(n^{-1/2}Z^{\perp *\prime}u^*\right)/E^*(u_i^{*2}) \\ (n^{-1}Z^{\perp *\prime}Z^{\perp *})^{-1/2}\left(n^{-1/2}Z^{\perp *\prime}v^*\right)/E^*(v_i^{*2}) \\ n^{-1/2}\left(u^{*\prime}v^* - E^*\left[u^{*\prime}v^*\right]\right)/\left(E^*(u_i^{*2})E^*(v_i^{*2})\right)^{1/2} \end{pmatrix} \to^{d^*} \begin{pmatrix} \psi_u^* \\ \psi_v^* \\ \psi_{uv}^* \end{pmatrix} \sim N\left(0, \begin{pmatrix} I_{2k_2} & 0 \\ 0' & 1 \end{pmatrix}\right), \tag{3.7}$$

in probability, where $E^*(u_i^{*2}) = n^{-1}\tilde{u}(\theta_0)'\tilde{u}(\theta_0)$ and $E^*(v_i^{*2}) = n^{-1}\tilde{v}'\tilde{v}$. (3.7) shows that both bootstrap procedures (with dependent and independent transformations) do 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 (2.14) 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* (2.14) *with $|h_1| < \infty$:*

$$\begin{pmatrix} T^*_{2SLS}(\theta_0) \\ T^*_{OLS}(\theta_0) \\ H^*_n \end{pmatrix} \to^{d^*} \eta^*_h = \begin{pmatrix} s'_{k_2}\psi^*_u \\ (1+h_2^2)^{-1/2}\left(h_2 s'_{k_2}\psi^*_u + \psi^*_{uv} + h_1^b\right) \\ (1+h_2^2)^{-1}\left(s'_{k_2}\psi^*_u - h_2\psi^*_{uv} - h_2 h_1^b\right)^2 \end{pmatrix},$$

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

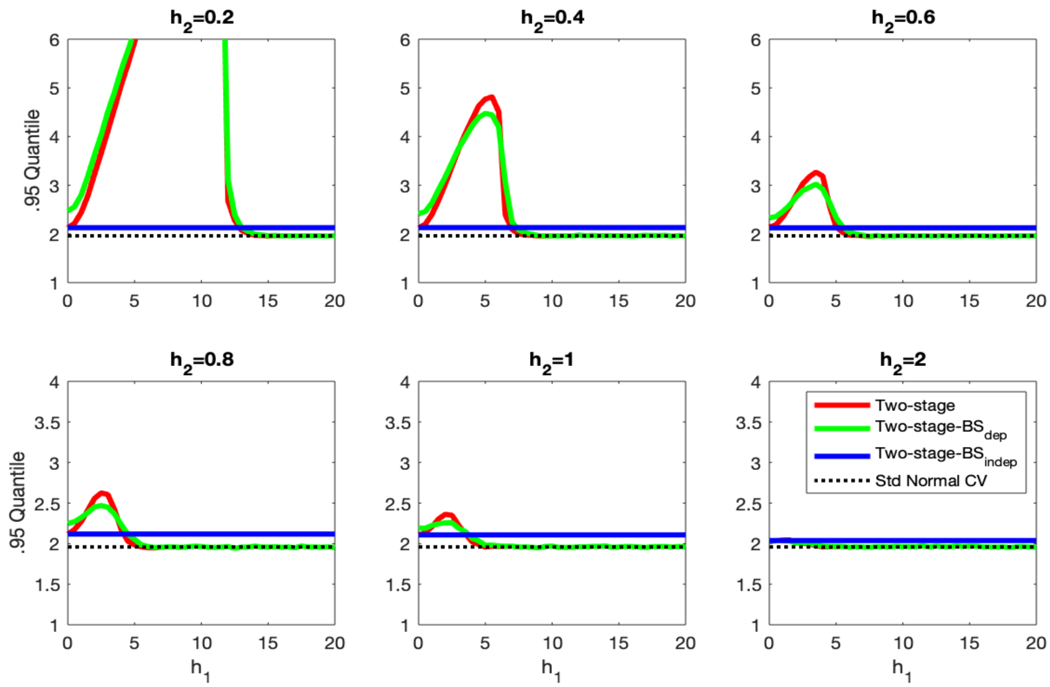
in probability, where $\eta^{*'}_h = (\eta^*_{1,h}, \eta^*_{2,h}, \eta^*_{3,h})$, $h_1^b = 0$ for the bootstrap based on independent transformation of disturbances, and $h_1^b = h_1 + \psi_{uv}$ with $\psi_{uv} \sim N(0,1)$ for the bootstrap based on dependent transformation of disturbances.

According to Theorem **3.1**, the standard 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_1^b 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_1^b is finite with probability approaching one when h_1 is finite), the approximation is imprecise and results in the extra error term $\psi_{uv} \sim N(0,1)$, whose value depends on the actual realization of the sample. In particular, these results suggest that the (conditional) limiting distribution of H^*_n under the independent bootstrap is a central chi-squared distribution, while that under the dependent bootstrap is a noncentral chi-squared distribution $\chi^2_1\left((h_1+\psi_{uv})^2 h_2^2(h_2^2+1)^{-1}\right)$. Therefore, the power properties of the bootstrap pretest statistic H^*_n under either procedure will be different from those 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 (2.16).Therefore, the bootstrap consistency in (3.6) 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 (2.14). To further shed light on the behaviour of the bootstrap statistics, we apply (2.16) and Theorem **3.1**, and plot the 95% quantiles of $\tilde{T}_h$ and $\tilde{T}^*_h$ in Figure 1 as a function of h_1 with $h_2 \in \{.2, .4, .6, .8, 1, 2\}$ and $\beta = .05$. 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 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,

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



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

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_u^* 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 $(1+h_2^2)^{-1}(s_{k_2}'\psi_u^* - h_2\psi_{uv}^*)^2$ 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_u^* and ψ_{uv}^* are independent from each other (e.g., see (3.7)).

Furthermore, we can obtain the asymptotic sizes of the two bootstrap tests by applying the results in (2.16) and Theorem **3.1**. Specifically, the asymptotic size of the bootstrap two-stage test can be defined as:

$$AsySz\left[\hat{c}_n^*(1-\alpha)\right] := \limsup_{n\to\infty} \sup_{\gamma\in\Gamma} P_{\theta_0,\gamma}\left[T_n(\theta_0) > \hat{c}_n^*(1-\alpha)\right], \tag{3.8}$$

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_1 and k_2.

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 (2.16) 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, 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 standard bootstrap.

3. Besides the residual-based i.i.d. bootstrap procedures described in this section, we may also consider alternative procedures such as the wild bootstrap. Specifically, we may generate the bootstrap disturbances u^* and v^* in (3.3) and (3.4) with dependent transformation by $u_i^* = \hat{u}_i(\theta_0)w_i$ and $v_i^* = \hat{v}_i w_i$ for some randomly generated i.i.d. sequence of weights $\{w_i\}_{i=1}^n$ that is independent from the sample and satisfies $E[w_i] = 0$ and $Var[w_i] = 1$ (e.g., the standard normal weight or the Rademacher weight, which puts probability one half on the values one and negative one). For the independent transformation, we may generate $u_i^* = \hat{u}_i(\theta_0)w_{1i}$ and $v_i^* = \hat{v}_i w_{2i}$ by two independent i.i.d. sequences of weights $\{w_{1i}\}_{i=1}^n$ and $\{w_{2i}\}_{i=1}^n$. Then by using an appropriate bootstrap CLT, we can show that the results in Theorem **3.1** also hold for the wild bootstrap procedures. Furthermore, as having the same null limiting distributions as their i.i.d. bootstrap counterparts, the dependent and independent wild bootstraps will have the same asymptotic size results as those reported in Table 1.

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 a 2SLS-based estimator $\hat{h}_{n,1}(\hat{\theta}_{2SLS}) = n^{1/2}\hat{\rho}(\hat{\theta}_{2SLS})$, where

$$\hat{\rho}_n(\hat{\theta}_{2SLS}) = n^{-1}(y_1^{\perp} - y_2^{\perp}\hat{\theta}_{2SLS})'(y_2^{\perp} - Z^{\perp}\hat{\pi})/(\hat{\sigma}_u(\hat{\theta}_{2SLS})\hat{\sigma}_v). \tag{3.9}$$

Then a confidence set of h_1 can be constructed by using the fact that under the drifting parameter sequences,

$$\hat{h}_{n,1}(\hat{\theta}_{2SLS}) \to^d \tilde{h}_1 \sim N\left(h_1, 1 + h_2^{-2}\right). \tag{3.10}$$

Alternatively, one may consider using the null-imposed estimator $\hat{h}_{n,1}(\theta_0) = n^{1/2}\hat{\rho}_n(\theta_0)$, whose null limiting distribution follows $N(h_1, 1)$.

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 $CI_{\alpha-\delta}(\hat{h}_{n,1})$ denote the $1-(\alpha-\delta)$ level confidence set for h_1 for some $0 < \delta \leq \alpha < 1$. The bootstrap-based simple Bonferroni critical value (SBCV) is defined as

$$c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2}) = \sup_{h_1 \in CI_{\alpha-\delta}(\hat{h}_{n,1})} c^*_{(h_1, \hat{h}_{n,2})}(1-\delta), \tag{3.11}$$

where $\hat{h}_{n,2} = \| \hat{\Omega}^{1/2}\hat{\pi}/\hat{\sigma}_v \|$, $\hat{\Omega} = \hat{Q}_{ZZ} - \hat{Q}_{ZX}\hat{Q}_{XX}^{-1}\hat{Q}_{XZ}$ with $\hat{Q}_{AB} = n^{-1}A'B$, and $c^*_{(h_1,\hat{h}_2)}(1-\delta)$ is the $(1-\delta)$-th quantile of the distribution of $T^*_{n,(h_1,\hat{h}_2)}(\theta_0)$, 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 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 $T^*_{n,(h_1,\hat{h}_2)}(\theta_0)$ as follows:

$$\begin{aligned} & T^*_{n,(h_1,\hat{h}_2)}(\theta_0) \\ = & \left| T^*_{OLS,(h_1,\hat{h}_2)}(\theta_0) \mathbb{1}\left(H^*_{n,(h_1,\hat{h}_2)} \leq \chi^2_{1,1-\beta}\right) + T^*_{2SLS}(\theta_0) \mathbb{1}\left(H^*_{n,(h_1,\hat{h}_2)} > \chi^2_{1,1-\beta}\right) \right|, \end{aligned} \tag{3.12}$$

where $T^*_{OLS,(h_1,\hat{h}_2)}(\theta_0)$ and $H^*_{n,(h_1,\hat{h}_2)}$ 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 counterpart of the OLS estimator under h_1:

$$\hat{\theta}^*_{OLS,(h_1,\hat{h}_2)} = \hat{\theta}^*_{OLS} + (\hat{h}_2^2 + 1)^{-1}\hat{\sigma}_u\hat{\sigma}_v^{-1}h_1, \tag{3.13}$$

where $\hat{\theta}^*_{OLS}$ is generated by the standard bootstrap procedure in Section 3.1 with *independent transformation* of disturbances, so that $\hat{\theta}^*_{OLS}$ has localization parameter equal to zero in the bootstrap world. By doing so, $\sqrt{n}(\hat{\theta}^*_{OLS,(h_1,\hat{h}_2)} - \theta_0)$ has appropriate (conditional) null limiting distribution. Then, we obtain $T^*_{OLS,(h_1,\hat{h}_2)}(\theta_0)$ and $H^*_{n,(h_1,\hat{h}_2)}$ as follows:

$$T^*_{OLS,(h_1,\hat{h}_2)}(\theta_0) = \frac{\sqrt{n}(\hat{\theta}^*_{OLS,(h_1,\hat{h}_2)} - \theta_0)}{\hat{V}^{*1/2}_{OLS}}, \tag{3.14}$$

$$H^*_{n,(h_1,\hat{h}_2)} \quad = \quad \frac{n(\hat{\theta}^*_{2SLS} - \hat{\theta}^*_{OLS,(h_1,\hat{h}_2)})^2}{\hat{V}^*_{2SLS} - \hat{V}^*_{OLS}}, \tag{3.15}$$

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

$$\begin{pmatrix} T^*_{OLS,(h_1,\hat{h}_2)}(\theta_0) \\ H^*_{n,(h_1,\hat{h}_2)} \end{pmatrix} \to^{d^*} \begin{pmatrix} (1+h_2^2)^{-1/2}(h_2 s'_{k_2}\psi^*_u + \psi^*_{uv} + h_1) \\ (1+h_2^2)^{-1}\left(s'_{k_2}\psi^*_u - h_2\psi^*_{uv} - h_2 h_1\right)^2 \end{pmatrix}, \tag{3.16}$$

in probability P. This implies that $T^*_{n,(h_1,\hat{h}_2)}(\theta_0)$, 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 (3.11), the bootstrap SBCV equals the maximal CV $c^*_{(h_1,\hat{h}_2)}(1-\delta)$ 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 $\boldsymbol{c^{\text{B-S}}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})}$:

1. Generate the bootstrap statistics $\left\{\hat{\theta}^{*(b)}_{OLS},\hat{\theta}^{*(b)}_{2SLS},\hat{V}^{*(b)}_{OLS},\hat{V}^{*(b)}_{2SLS}\right\}, b=1,...,B$, 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 $T^{*(b)}_{n,(h_1,\hat{h}_2)}(\theta_0)$, $b=1,...,B$, using the bootstrap statistics generated in Step 1. The same set of $\left\{\hat{\theta}^{*(b)}_{OLS},\hat{\theta}^{*(b)}_{2SLS},\hat{V}^{*(b)}_{OLS},\hat{V}^{*(b)}_{2SLS}\right\}, b=1,...,B$, 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:

$$\hat{\eta}_n \quad = \quad \inf\left\{\eta : \sup_{h_1\in\mathscr{H}_1} P^*\left[T^*_{n,(h_1,\hat{h}_{n,2})}(\theta_0) > c^{B-S}(\alpha,\alpha-\delta,\hat{h}^*_{n,1}(h_1),\hat{h}_{n,2})+\eta\right] \leq \alpha\right\}, \quad (3.17)$$

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

$$\hat{h}^*_{n,1}(h_1) = \hat{h}^*_{n,1} + h_1, \quad (3.18)$$

where $\hat{h}^*_{n,1} = n^{1/2}\hat{\rho}^*_n(\hat{\theta}^*_{2SLS}) = n^{-1/2}(y_1^{\perp *} - y_2^{\perp *}\hat{\theta}^*_{2SLS})'(y_2^{\perp *} - Z^{\perp *}\hat{\pi}^*)/(\hat{\sigma}^*_u(\hat{\theta}^*_{2SLS})\hat{\sigma}^*_v)$ 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 $N\left(0,(1+h_2^{-2})\right)$ in probability, while $\hat{h}^*_{n,1}(h_1)$ converges in distribution to $N\left(h_1,(1+h_2^{-2})\right)$ in probability, i.e., same as the limiting distribution of $\hat{h}_{n,1}$ in (3.10).

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 (3.17) 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 (3.17) 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 (3.17), so that the asymptotic size of the resulting tests can be controlled. Then, the size-corrected bootstrap CV can be defined as

$$c^{B-C}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2}) \quad = \quad c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2}) + \hat{\eta}_n, \quad (3.19)$$

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 (3.17) 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 $\left\{\hat{\theta}^{*(b)}_{OLS}, \hat{\theta}^{*(b)}_{2SLS}, \hat{V}^{*(b)}_{OLS}, \hat{V}^{*(b)}_{2SLS}, \hat{h}^{*(b)}_{n,1}\right\}, b = 1, \ldots, B$, 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 (3.17) 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 $c^{B-S}(\alpha, \alpha-\delta, \hat{h}^{*(b)}_{n,1}(h_1), \hat{h}_{n,2})$, $b = 1, \ldots, B$, using the bootstrap statistics generated in Step 1. Note that the same set of $\left\{\hat{\theta}^{*(b)}_{OLS}, \hat{\theta}^{*(b)}_{2SLS}, \hat{V}^{*(b)}_{OLS}, \hat{V}^{*(b)}_{2SLS}, \hat{h}^{*(b)}_{n,1}\right\}, b = 1, \ldots, B$, 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 (3.17). 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. Note that in many empirical studies, the values of the endogeneity parameter are typically quite low. For instance, Hansen et al. (2008, p.407, Table 6) report that the median, the 75% and the 90% quantiles of estimated endogeneity parameters are only 0.279, 0.466, and 0.555, respectively, in the investigated AER, JPE and QJE papers. These are exactly the cases where the Hausman pretest may not reject and naive two-stage procedure would leads to extreme size distortion (this also can be seen from the quantiles of the two-stage test statistics in Figure 1, where the highest quantiles occur at low values of h_1).

Furthermore, we note that it is possible for empirical researchers to reduce the parameter space of the endogeneity parameter ρ, by considering a reasonable range for the values of the structural parameters and by using a data-dependent mapping between the structural parameters and ρ. For example, Staiger and Stock (1997, p.579) argue that the range of the endogeneity parameter is [-0.5, 0.5] in their analysis of the return to education using the dataset of Angrist and Krueger (1991). Their IV model has one RHS endogenous variable ('years of education'), and to restrict the value of the endogeneity parameter, the authors consider the reasonable range of the return to education to be [0, 0.18]; i.e., an additional year of education increases expected weekly earning by at most 18%. (which substantially exceeds economic plausible values).

When applying the size-correction algorithm, researchers can directly incorporate such information on ρ into the construction of CVs, and maximize over a subspace of the endogeneity parameters in $\mathscr{H}_1$. Such reduction of nuisance parameters' space can both reduce computational cost and improve the power of the size-corrected tests.

4. Finite sample power performance

In this section, we study the finite-sample power performance of four tests: the Anderson-Rubin (1949, AR) test, the 2SLS-based t-test (without Hausman pretest), the two-stage test based on the dependent bootstrap CVs, and the two-stage test based on the 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 of the AR test, the 2SLS-based t test, and the two bootstrap-based two-stage 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 number of exogenous regressor is set at $k_1 = 0$, and the number of instruments is set at $k_2 = 1$. Following the discussions in Remark 2 of Section 3.2, we investigate the case where the degree of endogeneity can be restricted by the empirical researcher and the maximization of the CVs can be computed over $\rho \in [-0.5, 0.5]$.

Figures 2 - 5 show the finite-sample power curves of the four tests. The true values of the endogeneity parameter are set at $\rho \in \{0, 0.1n^{-1/2}, 0.5n^{-1/2}, 0.9n^{-1/2}, 0.2, 0.4\}$. The values of the concentration parameter, which characterizes the overall IV strength, are set at $\mu^2 \in \{1, 5, 10, 100\}$ for Figures 2 - 5, respectively. We highlight some findings below. First, it is clear that when the IV is relatively weak (e.g., Figures 2 - 3), the hybrid bootstrap-based two-stage test has remarkable power gain over both AR test and 2SLS-based t-test. Such power gain originates from the inclusion of the *OLS*-based t-test in the two-stage test. In addition, we note that although having much better size control than the independent bootstrap (e.g., Table 1), the dependent bootstrap-based two-stage test has relatively low power in many cases. In particular, it always has lower power than the 2SLS-based t-test and the hybrid bootstrap test. On the other hand, when the IV is very strong (e.g., Figure 5), all the tests have almost the same power properties. In the Appendix, we also report the simulation results for the case with $n = 500$, and the results are very similar to those in Figures 2 - 5. Besides the residual-based i.i.d bootstrap procedure, we also tried simulations with the wild bootstrap procedure (using standard normal weights) discussed in Remark 3 of Section 3.1, and the results were also very similar.

In sum, 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. 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 and AR test, especially when the instruments are not very strong.

Figure 2. Power of AR, 2SLS-t, dependent bootstrap, and hybrid bootstrap tests: $\mu^2 = 1$

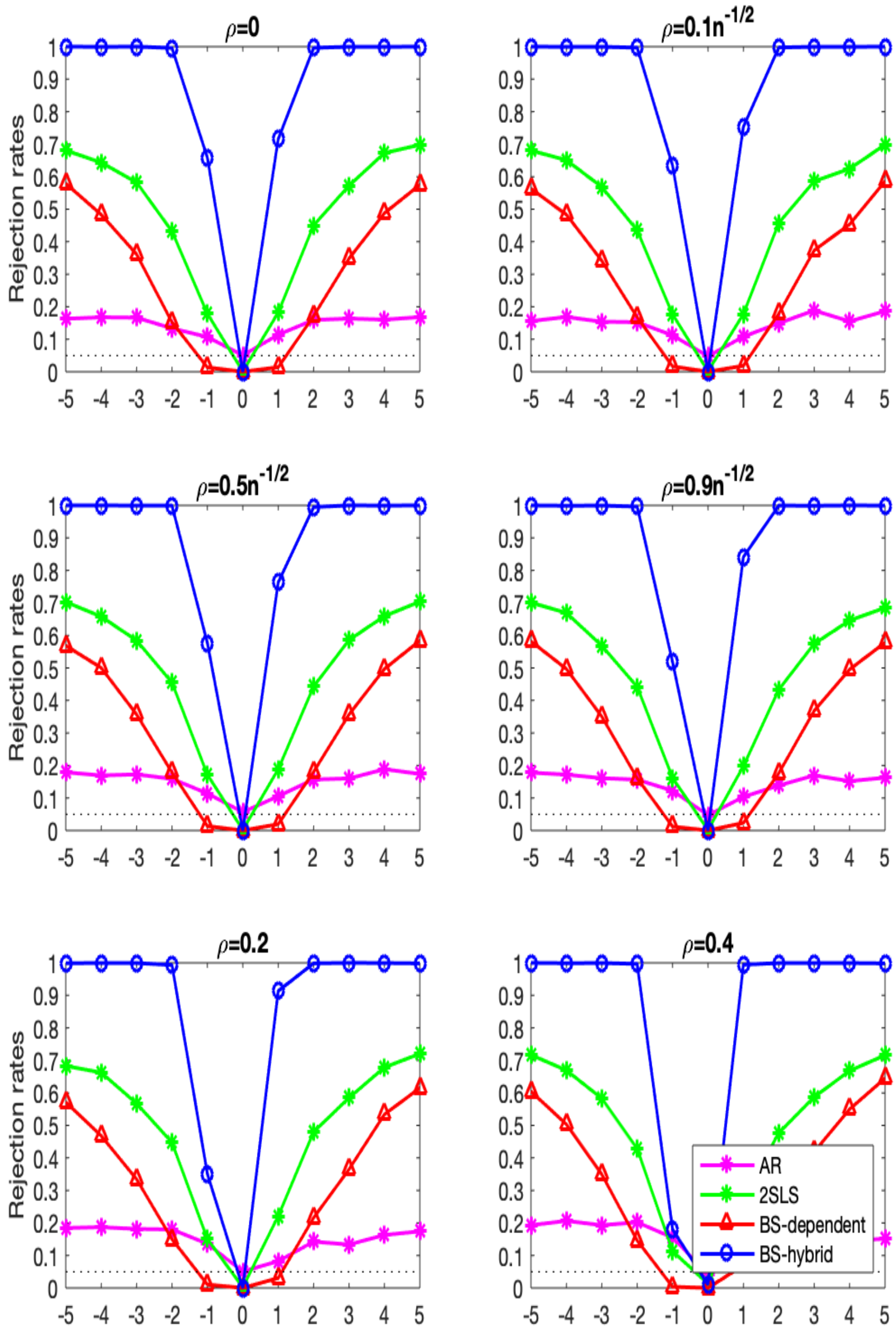


Figure 3. Power of AR, 2SLS-t, dependent bootstrap, and hybrid bootstrap tests: $\mu^2 = 5$

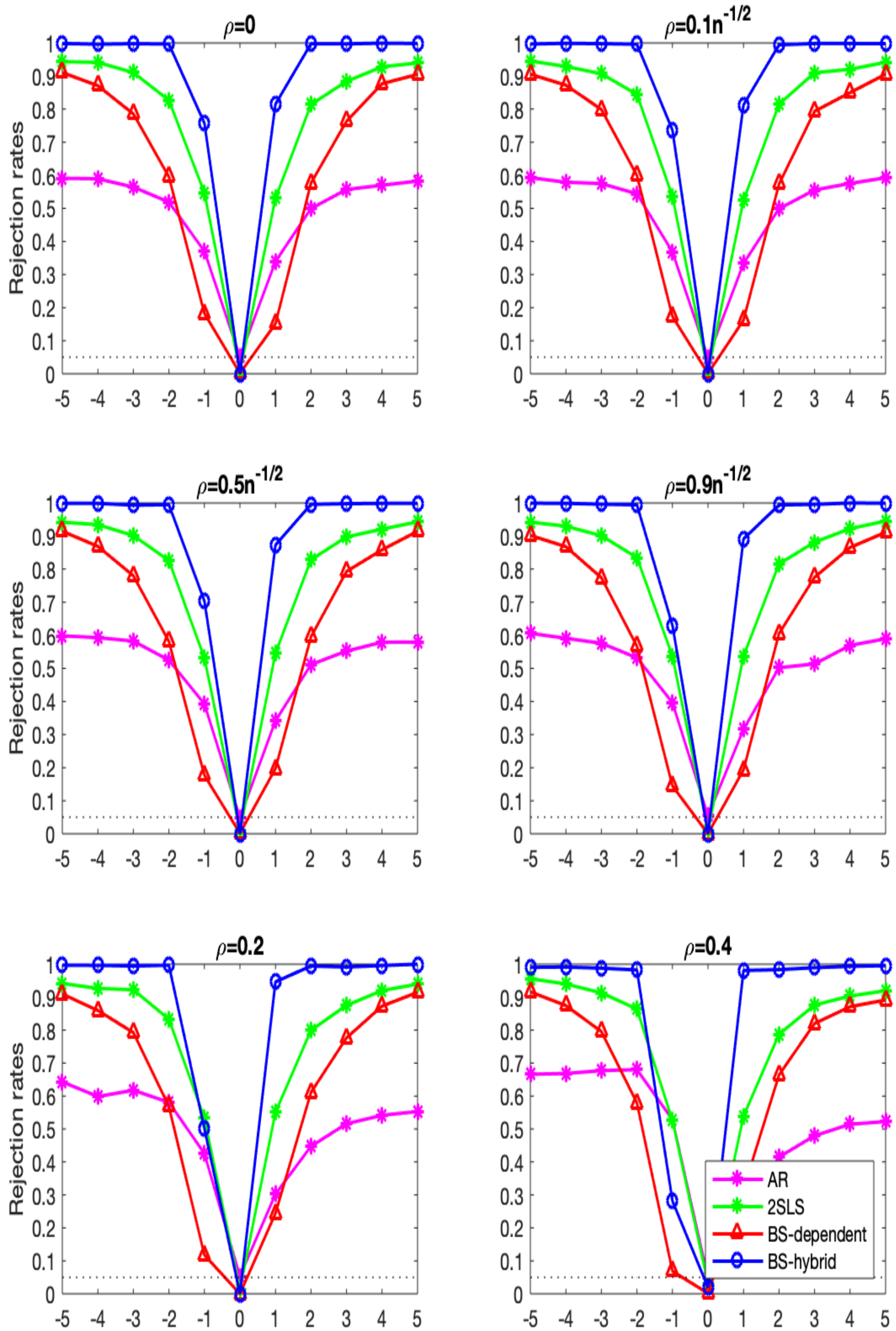


Figure 4. Power of AR, 2SLS-t, dependent bootstrap, and hybrid bootstrap tests:$\mu^2 = 10$

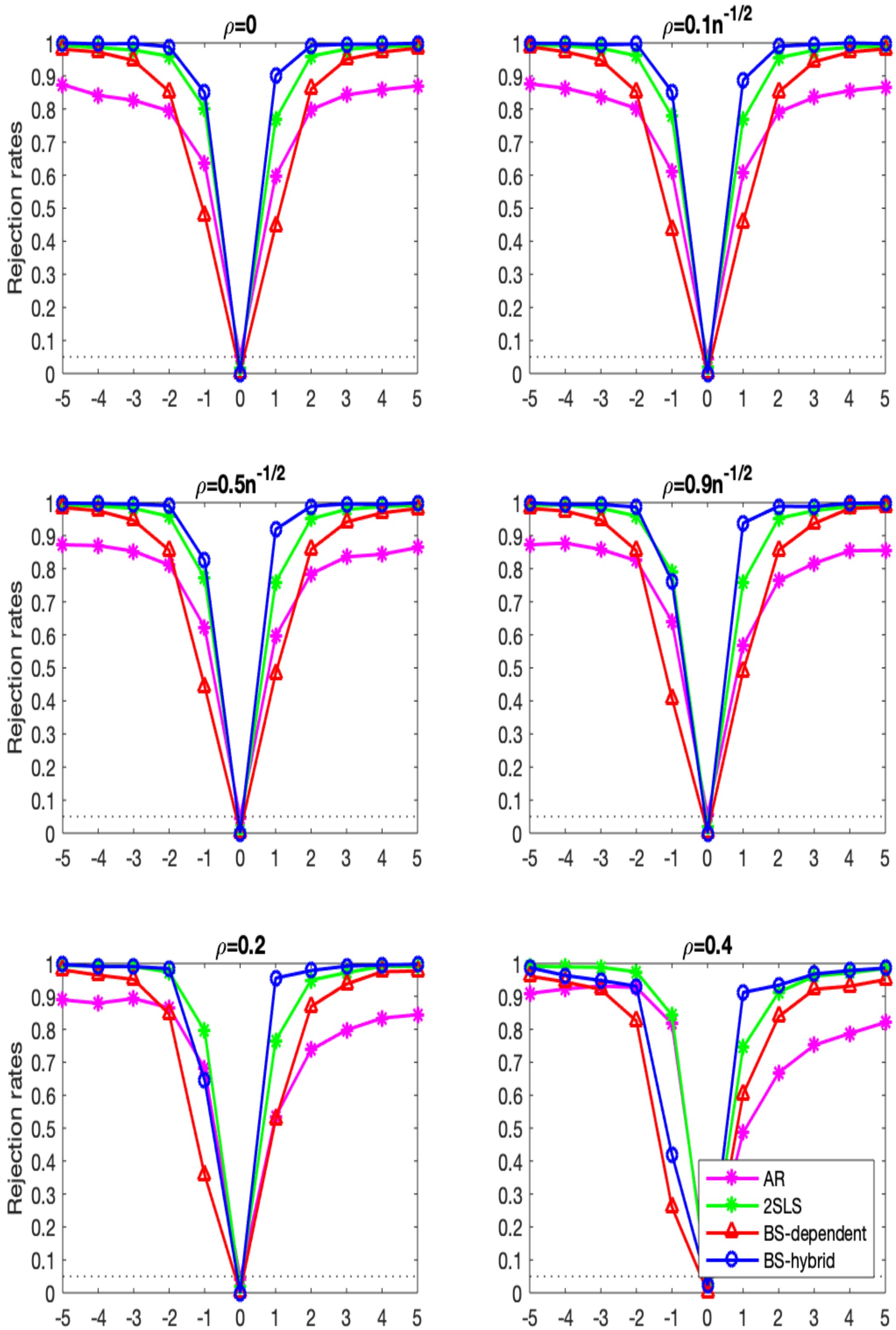
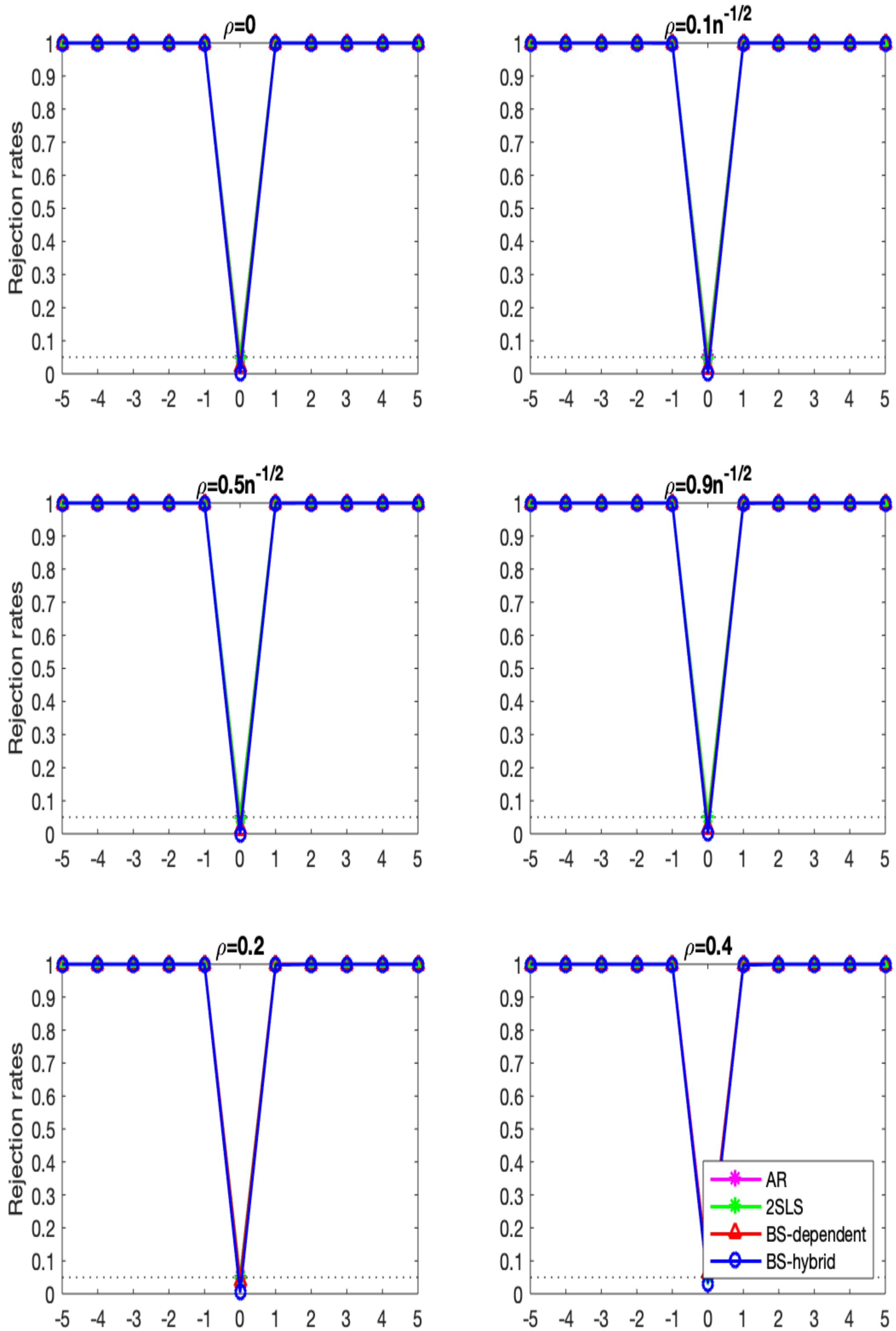


Figure 5. Power of AR, 2SLS-t, dependent bootstrap, and hybrid bootstrap tests: $\mu^2 = 100$



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.

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.

Durbin, J., 1954. Errors in variables. Review of the International Statistical Institute 22, 23–32.

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

Hausman, J., 1978. Specification tests in econometrics. Econometrica 46, 1251–1272.

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.

Moreira, M. J., Porter, J., Suarez, G., 2009. Bootstrap validity for the score test when instruments may be weak. Journal of Econometrics 149(1), 52–64.

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

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.

Wu, D.-M., 1973. Alternative tests of independence between stochastic regressors and disturbances. Econometrica 41, 733–750.

Wu, D.-M., 1974. Alternative tests of independence between stochastic regressors and disturbances: Finite sample results. Econometrica 42, 529–546.

Young, A., 2019. Consistency without inference: Instrumental variables in practical application. Technical report, London School of Economics Montréal, Canada.

A. Appendix

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

A.1. Mathematical Proofs

Lemma A.1 *If for some* $\delta > 0$, $E\left[||Z_i^{\perp}||^{2+\delta}\right] < \infty$, $E\left[|u_i|^{2+\delta}\right] < \infty$ *and* $E\left[|v_i|^{2+\delta}\right] < \infty$, *then for* $j = 1, \ldots, k_2$, $E^*\left[\left|Z_{j,i}^{\perp *} u_i^*\right|^{2+\delta}\right]$ *and* $E^*\left[\left|Z_{j,i}^{\perp *} v_i^*\right|^{2+\delta}\right]$ *are bounded in probability.*

PROOF OF LEMMA **A.1**

The proof follows closely Lemma A.1 of Moreira, Porter and Suarez (2009). We give the proof for $E^*\left[\left|Z_{j,i}^{\perp *} v_i^*\right|^{2+\delta}\right]$. The proof for $E^*\left[\left|Z_{j,i}^{\perp *} u_i^*\right|^{2+\delta}\right]$ is very similar thus omitted.

We note that by independence, $E^*\left[\left|Z_{j,i}^{\perp *} v_i^*\right|^{2+\delta}\right] = E^*\left[\left|Z_{j,i}^{\perp *}\right|^{2+\delta}\right] E^*\left[|v_i^*|^{2+\delta}\right]$. For $j = 1, \ldots, k_2$,

$$E^*\left[\left|Z_{j,i}^{\perp *}\right|^{2+\delta}\right] = n^{-1}\sum_{i=1}^{n}\left|Z_{j,i}^{\perp}\right|^{2+\delta} \to^P E\left[\left|Z_{j,i}^{\perp}\right|^{2+\delta}\right]. \tag{A.1}$$

Let $\bar{v} = n^{-1}\sum_{i=1}^{n} v_i$, and $\bar{Z} = n^{-1}\sum_{i=1}^{n} Z_i^{\perp}$. Using Minkowski and Cauchy-Schwartz inequalities, we obtain

$$\begin{aligned} E^*\left[|v_i^*|^{2+\delta}\right] &= n^{-1}\sum_{i=1}^{n}|\tilde{v}_i|^{2+\delta} \\ &= n^{-1}\sum_{i=1}^{n}\left|v_i - \bar{v} - (Z_i^{\perp} - \bar{Z})'(\hat{\pi} - \pi)\right|^{2+\delta} \\ &\leq C\left\{n^{-1}\sum_{i=1}^{n}|v_i - \bar{v}|^{2+\delta} + ||\hat{\pi} - \pi||^{2+\delta} n^{-1}\sum_{i=1}^{n}||Z_i^{\perp} - \bar{Z}||^{2+\delta}\right\}, \end{aligned} \tag{A.2}$$

where C denotes a large enough constant, which may take different values in different applications. Using the Minkowski inequality again, we get

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}||Z_i^{\perp} - \bar{Z}||^{2+\delta} &\leq C\left\{n^{-1}\sum_{i=1}^{n}||Z_i^{\perp}||^{2+\delta} + ||\bar{Z}||^{2+\delta}\right\} \\ &\to^P C\left\{E\left[||Z_i^{\perp}||^{2+\delta} + ||E[Z_i^{\perp}]||^{2+\delta}\right]\right\}, \end{aligned} \tag{A.3}$$

using $||\bar{Z}|| \to^P ||E[Z_i^{\perp}]|| \leq E[||Z_i^{\perp}||] \leq (E[||Z_i^{\perp}||^{2+\delta}])^{1/(2+\delta)}$ by Jensen's inequality. Similarly, using the Minkowski inequality again, we obtain

$$n^{-1}\sum_{i=1}^{n}|v_i - \bar{v}|^{2+\delta} \leq C\left\{n^{-1}\sum_{i=1}^{n}|v_i|^{2+\delta} + |\bar{v}|^{2+\delta}\right\}$$

$$\to^P \quad C\left\{E[|v_i|^{2+\delta}]\right\}, \tag{A.4}$$

as $\bar{v} \to^P 0$. Since $\hat{\pi} - \pi \to^P 0$, the term $E^*\left[|v_i^*|^{2+\delta}\right]$ is bounded in probability. □

Lemma A.2 *Suppose that the H_0 holds, then under $\{\gamma_{n,h}\}$ defined in (2.14) with $|h_1| < \infty$ we have:*

$$\begin{pmatrix} n^{-1/2} Z^{\perp *\prime} u^* \\ n^{-1/2} Z^{\perp *\prime} v^* \\ n^{-1/2}\left(u^{*\prime} v^* - E^*\left[u^{*\prime} v^*\right]\right) \end{pmatrix} \to^{d^*} N\left(0, \begin{pmatrix} diag(\sigma_u^2, \sigma_v^2) \otimes \Omega & 0 \\ 0' & \sigma_u^2 \sigma_v^2 \end{pmatrix}\right), \tag{A.5}$$

in probability.

PROOF OF LEMMA **A.2**

The proof follows closely Lemma A.2 of Moreira et al. (2009). Let $c = (c_1', c_2')' \in \mathbb{R}^{2k_2}$ be a nonzero vector. Define

$$X_{n,i}^* = \left\{c'\left(U_i^* \otimes Z_i^{\perp *}\right) + d\left(u_i^* v_i^* - E^*[u_i^* v_i^*]\right)\right\} / \sqrt{n}, \tag{A.6}$$

where $U_i^* = (u_i^*, v_i^*)'$ is the i-th bootstrap draw of the (re-centered) residuals, and we use the Cramer-Wald device to verify that the conditions of the Lyapunov Central Limit Theorem hold for $X_{n,i}^*$. We give the proof for the case with independent transformation. The proof for the case with dependent transformation is very similar thus omitted. To proceed, we note that:

(a) $E^*\left[X_{n,i}^*\right] = 0$ follows from the independence of the bootstrap draws between U_i^* and $Z_i^{\perp *}$ and from $E^*[U_i^*] = 0$.

(b) By noting that $E^*\left[U_i^* U_i^{*\prime}\right] = diag\left(n^{-1}\tilde{u}'(\theta_0)\tilde{u}(\theta_0), n^{-1}\tilde{v}'\tilde{v}\right)$, $E^*\left[Z_i^{\perp *} Z_i^{\perp *\prime}\right] = n^{-1} Z^{\perp\prime} Z^{\perp}$, $E^*\left[u_i^{*2} v_i^{*2}\right] = n^{-1}\sum_{i=1}^n \tilde{u}_i^2(\theta_0)\tilde{v}_i^2$, and $E^*[u_i^* v_i^*] = E^*[u_i^*] E^*[v_i^*] = 0$, we have

$$E^*\left[X_{n,i}^{*2}\right] = n^{-1}\left\{c'\left[diag\left(n^{-1}\tilde{u}'(\theta_0)\tilde{u}(\theta_0), n^{-1}\tilde{v}'\tilde{v}\right) \otimes \left(n^{-1} Z^{\perp\prime} Z^{\perp}\right)\right] c + d^2\left(n^{-1}\sum_{i=1}^n \tilde{u}_i^2(\theta_0)\tilde{v}_i^2\right)\right\}.$$

It is clear that $E^*\left[X_{n,i}^{*2}\right]$ is bounded in probability.

(c) Finally, we note that

$$\begin{aligned} & \sum_{i=1}^n E^*\left[\left|X_{n,i}^*\right|^{2+\delta}\right] \\ \leq \quad & C n^{-\frac{\delta}{2}} n^{-1} \sum_{i=1}^n E^*\left[\left|c'\left(U_i^* \otimes Z_i^{\perp *}\right)\right|^{2+\delta} + |d u_i^* v_i^*|^{2+\delta}\right] \\ \leq \quad & C n^{-\frac{\delta}{2}} E^*\left[\left|c_1' Z_i^{\perp *} u_i^*\right|^{2+\delta} + \left|c_2' Z_i^{\perp *} v_i^*\right|^{2+\delta} + |d u_i^* v_i^*|^{2+\delta}\right] \end{aligned}$$

$$\leq \quad Cn^{-\frac{\delta}{2}} \left\{ \sum_{j=1}^{k2} \left[|c_{1j}|^{2+\delta} E^* \left[\left| Z_{j,i}^{\perp *} u_i^* \right|^{2+\delta} + |c_{2j}|^{2+\delta} E^* \left[\left| Z_{j,i}^{\perp *} v_i^* \right|^{2+\delta} \right] \right] + |d|^{2+\delta} E^* \left[|u_i^* v_i^*|^{2+\delta} \right] \right\}$$
$$\rightarrow^P \quad 0,$$

where the convergence in probability is obtained by using the results in Lemma **A.1** and the fact that $E^* \left[|u_i^* v_i^*|^{2+\delta} \right] = E^* \left[|u_i^*|^{2+\delta} \right] E^* \left[|v_i^*|^{2+\delta} \right] = \left(n^{-1} \sum_{i=1}^n [|\tilde{u}_i(\theta_0)|^{2+\delta}] \right) \left(n^{-1} \sum_{i=1}^n [|\tilde{v}_i|^{2+\delta}] \right)$ is bounded in probability.

The required result follows by applying the Lyapunov Central Limit Theorem. □

PROOF OF THEOREM **3.1**

First, we note that

$$\begin{aligned}
& n^{-1/2} y_2^{\perp *'} P_{Z^{\perp *}} u^* / (E^*[u_i^{*2}] E^*[v_i^{*2}])^{1/2} \\
= \quad & \left[n^{-1/2} \left(Z^{\perp *} \hat{\pi} + v^* \right)' P_{Z^{\perp *}} u^* \right] / (E^*[u_i^{*2}] E^*[v_i^{*2}])^{1/2} \\
= \quad & \left[n^{-1/2} \hat{\pi}' Z^{\perp *'} u^* + n^{-1/2} \left(n^{-1/2} v^{*'} Z^{\perp *} \right) \left(n^{-1} Z^{\perp *'} Z^{\perp *} \right)^{-1} \left(n^{-1/2} Z^{\perp *'} u^* \right) \right] / (E^*[u_i^{*2}] E^*[v_i^{*2}])^{1/2} \\
= \quad & n^{-1/2} \hat{\pi}' Z^{\perp *'} u^* / (E^*[u_i^{*2}] E^*[v_i^{*2}])^{1/2} + O_{P^*} \left(n^{-1/2} \right) \\
\rightarrow^{d^*} \quad & h_2 s_{k_2}' \psi_u^*,
\end{aligned} \tag{A.7}$$

in probability, where the last equality follows from: (a) by Lemma **A.2**, $n^{-1/2} v^{*'} Z^{\perp *} = O_{P^*}(1)$ in probability, and $n^{-1/2} Z^{\perp *'} u^* = O_{P^*}(1)$ in probability; (b) by the Markov Law of Large Numbers, $n^{-1} Z^{\perp *'} Z^{\perp *} - n^{-1} Z^{\perp '} Z^{\perp} \rightarrow^{P^*} 0$ in probability; (c) $n^{-1} Z^{\perp '} Z^{\perp} \rightarrow^P \Omega$, which is positive definite, and therefore $\left(n^{-1} Z^{\perp *'} Z^{\perp *} \right)^{-1} \rightarrow^P \Omega^{-1}$ in probability; and (d) $E^*[u_i^{*2}] = n^{-1} \tilde{u}'(\theta_0) \tilde{u}(\theta_0) \rightarrow^P \sigma_u^2$ under H_0, and $E^*[v_i^{*2}] = n^{-1} \tilde{v}' \tilde{v} \rightarrow^P \sigma_v^2$. And the (conditional) convergence in distribution follows from Lemma **A.2**, $\hat{\pi} - \pi \rightarrow^P 0$, and the definition of h_2 and ψ_u^*.

Second, note that using similar arguments and Lemma **A.2**, we have

$$\begin{aligned}
& n^{-1/2} y_2^{\perp *'} u^* / (E^*[u_i^{*2}] E^*[v_i^{*2}])^{1/2} \\
= \quad & \left[n^{-1/2} \hat{\pi}' Z^{\perp *'} u^* + n^{-1/2} \left(v^{*'} u^* - E^*[v^{*'} u^*] \right) + n^{-1/2} E^*[v^{*'} u^*] \right] / (E^*[u_i^{*2}] E^*[v_i^{*2}])^{1/2} \\
\rightarrow^{d^*} \quad & h_2 s_{k_2}' \psi_u^* + \psi_{uv}^* + h_1^b,
\end{aligned} \tag{A.8}$$

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

Third, we note that

$$n^{-1} y_2^{\perp *'} P_{Z^{\perp *}} y_2^{\perp *} / E^*[v_i^{*2}]$$

$$
\begin{aligned}
&= n^{-1}\left(Z^{\perp *}\hat{\pi}+v^{*}\right)' P_{Z^{\perp *}}\left(Z^{\perp *}\hat{\pi}+v^{*}\right)/E^{*}[v_i^{*2}] \\
&= \left[n^{-1}\hat{\pi}'Z^{\perp *'}Z^{\perp *}\hat{\pi}+n^{-1}\hat{\pi}'Z^{\perp *'}v^{*}+n^{-1}v^{*'}Z^{\perp *}\hat{\pi}+n^{-1}v^{*'}P_{Z^{\perp *}}v^{*}\right]/E^{*}[v_i^{*2}] \\
&= n^{-1}\hat{\pi}'Z^{\perp *'}Z^{\perp *}\hat{\pi}/E^{*}[v_i^{*2}]+O_{P^{*}}\left(n^{-1/2}\right)+O_{P^{*}}\left(n^{-1/2}\right)+O_{P^{*}}\left(n^{-1}\right)\to^{P^{*}} h_2^2, \quad \text{(A.9)}
\end{aligned}
$$

in probability. Using similar arguments, we obtain

$$
n^{-1}y_2^{\perp *'}y_2^{\perp *}/E^{*}[v_i^{*2}]\to^{P^{*}} h_2^2+1, \tag{A.10}
$$

in probability. Combining (A.7)-(A.10), we get the desired result. □

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:

$$
\begin{aligned}
& AsySz[\hat{c}_n^{*}(1-\alpha)] \\
=\ & \limsup_{n\to\infty}\sup_{\gamma\in\Gamma} P_{\theta_0,\gamma}\left[T_n(\theta_0)>\hat{c}_n^{*}(1-\alpha)\right] \\
=\ & \limsup_{n\to\infty} P_{\theta_0,\gamma_n}\left[T_n(\theta_0)>\hat{c}_n^{*}(1-\alpha)\right] \\
=\ & \lim_{n\to\infty} P_{\theta_0,\gamma_{m_n}}\left[T_{m_n}(\theta_0)>\hat{c}_{m_n}^{*}(1-\alpha)\right] \qquad \text{(A.11)}
\end{aligned}
$$

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:

$$
\begin{aligned}
& \lim_{n\to\infty} P_{\theta_0,\gamma_{m_n}}\left[T_{m_n}(\theta_0)>\hat{c}_{m_n}^{*}(1-\alpha)\right] \\
=\ & \lim_{n\to\infty} P_{\theta_0,\gamma_{\omega_n,h}}\left[T_{\omega_n}(\theta_0)>\hat{c}_{\omega_n}^{*}(1-\alpha)\right] \qquad \text{(A.12)}
\end{aligned}
$$

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:

$$
\begin{pmatrix} T^{*}_{OLS,(h_1,\hat{h}_2)}(\theta_0) \\ H^{*}_{n,(h_1,\hat{h}_2)} \end{pmatrix} \to^{d^{*}} \begin{pmatrix} (1+h_2^2)^{-1/2}(h_2 s_{k_2}'\psi_u^{*}+\psi_{uv}^{*}+h_1) \\ (1+h_2^2)^{-1}\left(s_{k_2}'\psi_u^{*}-h_2\psi_{uv}^{*}-h_2h_1\right)^2 \end{pmatrix}, \tag{A.13}
$$

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 $c^*_{(h_1,\hat{h}_2)}(1-\delta) \to^P c_{(h_1,h_2)}(1-\delta)$, 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:

$$\begin{aligned}
& AsySz\left[c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})\right] \\
= \; & \limsup_{n\to\infty} \sup_{\gamma\in\Gamma} P_{\theta_0,\gamma}\left[T_n(\theta_0) > c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})\right] \\
= \; & \limsup_{n\to\infty} P_{\theta_0,\gamma_n}\left[T_n(\theta_0) > c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})\right] \\
= \; & \lim_{n\to\infty} P_{\theta_0,\gamma_{m_n}}\left[T_{m_n}(\theta_0) > c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{m_n,1},\hat{h}_{m_n,2})\right] \qquad (A.14)
\end{aligned}$$

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:

$$\begin{aligned}
& \lim_{n\to\infty} P_{\theta_0,\gamma_{m_n}}\left[T_{m_n}(\theta_0) > c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{m_n,1},\hat{h}_{m_n,2})\right] \\
= \; & \lim_{n\to\infty} P_{\theta_0,\gamma_{\omega_n,h}}\left[T_{\omega_n}(\theta_0) > c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{\omega_n,1},\hat{h}_{\omega_n,2})\right] \qquad (A.15)
\end{aligned}$$

for some $h\in\mathscr{H}$. But, for any $h\in\mathscr{H}$, any subsequence $\{\omega_n : n\geq 1\}$ of $\{n : n\geq 1\}$, and any sequence $\{\gamma_{\omega_n,h} : n\geq 1\}$, we have $\left(T_{\omega_n}(\theta_0), \hat{h}_{\omega_n,1}\right) \xrightarrow{d} \left(\tilde{T}_h, \tilde{h}_1\right)$ 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:

$$\left(T_{\omega_n}(\theta_0), c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{\omega_n,1},\hat{h}_{\omega_n,2})\right) \xrightarrow{d} \left(\tilde{T}_h, c^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)\right) \qquad (A.16)$$

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.14)-(A.16) imply that

$$\begin{aligned}
& AsySz\left[c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})\right] \\
= \; & \lim_{n\to\infty} P_{\theta_0,\gamma_{\omega_n,h}}\left[T_{\omega_n}(\theta_0) > c^{B-S}(\alpha,\alpha-\delta,\hat{h}_{\omega_n,1},\hat{h}_{\omega_n,2})\right] \\
= \; & \sup_{h\in\mathscr{H}} P\left[\tilde{T}_h > c^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)\right], \qquad (A.17)
\end{aligned}$$

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

$$\begin{aligned}
& P\left[\tilde{T}_h \geq c^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2)\right] \\
= \; & P\left[\tilde{T}_h \geq c^{B-S}(\alpha,\alpha-\delta,\tilde{h}_1,h_2) \geq c_h(1-\delta)\right]
\end{aligned}$$

$$
\begin{aligned}
&+ \quad P\left[\tilde{T}_h \geq c_h(1-\delta) \geq c^{B-S}(\alpha, \alpha-\delta, \tilde{h}_1, h_2)\right] \\
&+ \quad P\left[c_h(1-\delta) \geq \tilde{T}_h \geq c^{B-S}(\alpha, \alpha-\delta, \tilde{h}_1, h_2)\right] \\
&\leq \quad P\left[\tilde{T}_h \geq c_h(1-\delta)\right] + P\left[c_h(1-\delta) \geq c^{B-S}(\alpha, \alpha-\delta, \tilde{h}_1, h_2)\right] \\
&= \quad P\left[\tilde{T}_h \geq c_h(1-\delta)\right] + P\left[h_1 \notin CI_{\alpha-\delta}(\tilde{h}_1)\right] \\
&= \quad \delta + (\alpha - \delta) = \alpha, \qquad \text{(A.18)}
\end{aligned}
$$

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.18) holds for any $h \in \mathscr{H}$, it is clear from (A.17) 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:

$$
\begin{aligned}
& AsySz\left[c^{B-C}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2})\right] \\
= \quad & \limsup_{n\to\infty} \sup_{\gamma\in\Gamma} P_{\theta_0,\gamma}\left[T_n(\theta_0) > c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2}) + \hat{\eta}_n\right] \\
= \quad & \limsup_{n\to\infty} P_{\theta_0,\gamma_n}\left[T_n(\theta_0) > c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{n,1}, \hat{h}_{n,2}) + \hat{\eta}_n\right] \\
= \quad & \lim_{n\to\infty} P_{\theta_0,\gamma_{m_n}}\left[T_{m_n}(\theta_0) > c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{m_n,1}, \hat{h}_{m_n,2}) + \hat{\eta}_{m_n}\right] \\
= \quad & \lim_{n\to\infty} P_{\theta_0,\gamma_{\omega_n,h}}\left[T_{\omega_n}(\theta_0) > c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{\omega_n,1}, \hat{h}_{\omega_n,2}) + \hat{\eta}_{\omega_n}\right] \qquad \text{(A.19)}
\end{aligned}
$$

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,

$$
\begin{aligned}
& \lim_{n\to\infty} P_{\theta_0,\gamma_{\omega_n,h}}\left[T_{\omega_n}(\theta_0) > c^{B-S}(\alpha, \alpha-\delta, \hat{h}_{\omega_n,1}, \hat{h}_{\omega_n,2}) + \hat{\eta}_{\omega_n}\right] \\
= \quad & \sup_{h\in\mathscr{H}} P\left[\tilde{T}_h > c^{B-S}(\alpha, \alpha-\delta, \tilde{h}_1, h_2) + \bar{\eta}\right] \qquad \text{(A.20)} \\
\equiv \quad & \sup_{h\in\mathscr{H}} P\left[\tilde{T}_h > c^{B-C}(\alpha, \alpha-\delta, \tilde{h}_1, h_2)\right], \qquad \text{(A.21)}
\end{aligned}
$$

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:

$$
NRP[h, \eta] \quad \equiv \quad P[\tilde{T}_h > c^{B-S}(\alpha, \alpha-\delta, \tilde{h}_1, h_2) + \eta]. \qquad \text{(A.22)}
$$

It is clear from (A.19)-(A.22) 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 $K(\cdot):\ \mathbb{R}_{-} \rightarrow [-\alpha,\ 1-\alpha]$ such that

$$K(x) = \sup_{h\in\mathscr{H}} NRP[h,x] - \alpha. \tag{A.23}$$

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., $AsySz[c^{B-C}(\alpha,\alpha-\delta,\hat{h}_{n,1},\hat{h}_{n,2})] = \alpha$. $\square$

A.2. Further Simulation Results

In this section, we report further simulation results for the finite-sample power performance of the four tests: the AR test, the 2SLS-based t-test (without Hausman pretest), the two-stage test based on the dependent bootstrap CVs, and the two-stage test based on the hybrid bootstrap CVs. We follow the same set ups as those in Section 4 but with the sample size $n = 500$. Figures A.1 - A.4 show the finite-sample power curves of the four tests, and the results are very similar to those in Figures 2 - 5.

Figure A.1. Power of AR, 2SLS-t, dependent bootstrap, and hybrid bootstrap tests: $\mu^2 = 1$

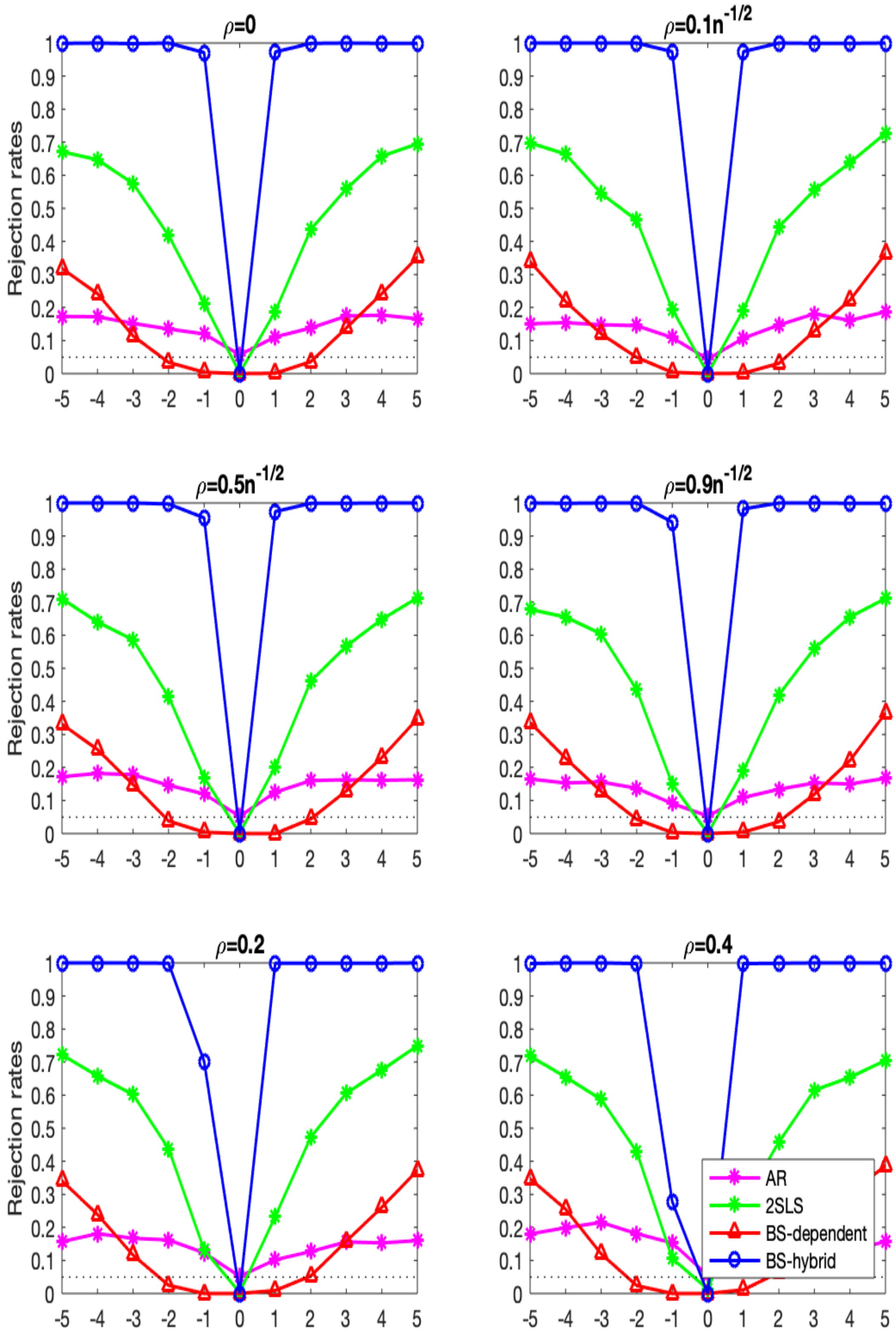


Figure A.2. Power of AR, 2SLS-t, dependent bootstrap, and hybrid bootstrap tests: $\mu^2 = 5$

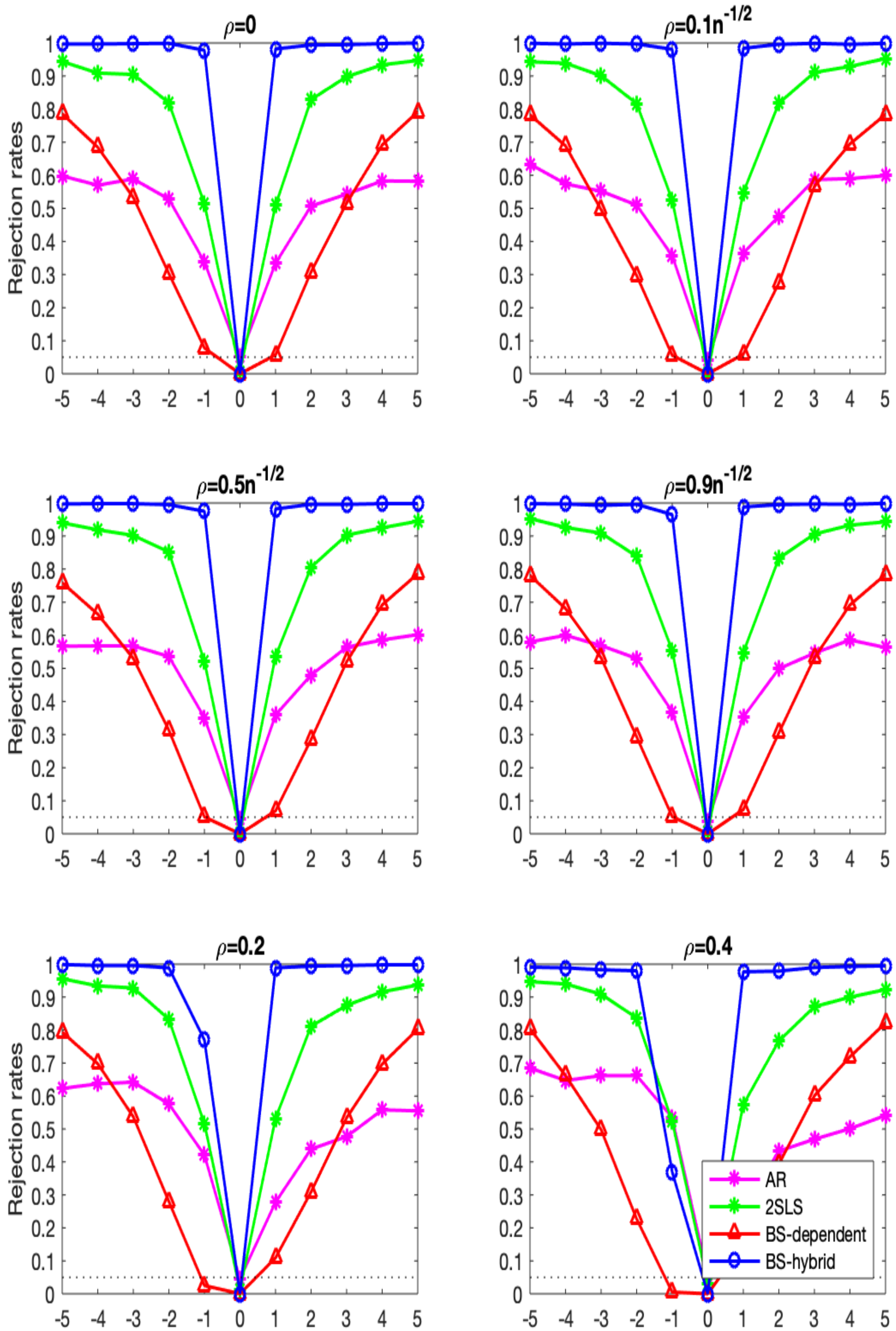


Figure A.3. Power of AR, 2SLS-t, dependent bootstrap, and hybrid bootstrap tests:$\mu^2 = 10$

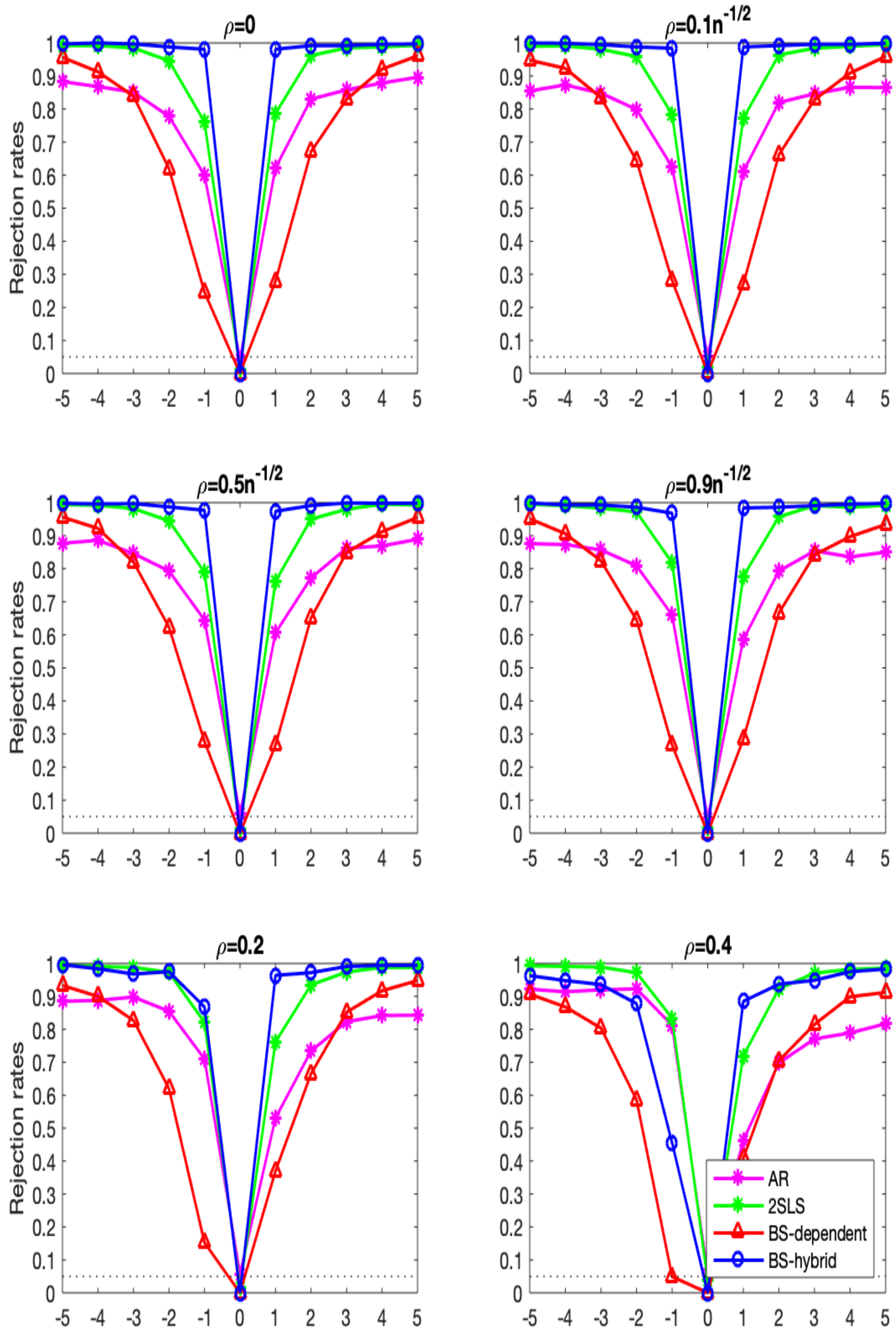


Figure A.4. Power of AR, 2SLS-t, dependent bootstrap, and hybrid bootstrap tests: $\mu^2 = 100$