THE UNIVERSITY of ADELAIDE

School of Economics

Working Papers

ISSN 2203-6024

On Bootstrap Validity for Specification Tests with Weak Instruments

Firmin Doko Tchatoka

Working Paper No. 2014-06
June, 2014

On bootstrap validity for specification tests with weak instruments

Firmin Doko Tchatoka[*]

The University of Adelaide

June 19, 2014

[*] School of Economics, The University of Adelaide, 10 Pulteney Street, Adelaide SA 5005, Tel:+618 8313 5540, Fax:+618 8223 1460, e-mail: firmin.dokotchatoka@adelaide.edu.au

ABSTRACT

We investigate the asymptotic validity of the bootstrap for Durbin-Wu-Hausman (DWH) tests of exogeneity in linear IV regressions, with or without identification. Our analysis of the properties (size and power) of the proposed bootstrap tests provides some new insights and extensions of earlier studies. More precisely, we show that when identification is strong, the bootstrap: (1) provides an asymptotic refinement of the size of the DWH tests under exogeneity, and (2) is consistent under the alternative hypothesis if the endogeneity parameter is fixed. However, the bootstrap only provides a first-order approximation of the asymptotic distributions of these statistics when instruments are weak. Moreover, we characterize the necessary and sufficient condition under which the proposed bootstrap DWH tests exhibit power under fixed endogeneity and weak instruments. The latter condition may still hold over a wide range of cases, provided that at least one instrument is not irrelevant. But all bootstrap tests have low power when all instruments are irrelevant, a case of little interest in empirical work. We present a Monte Carlo experiment that confirms our theoretical findings.

Key words: Exogeneity testing; Durbin-Wu-Hausman tests; weak instruments; bootstrap validity; asymptotic refinement.

JEL classification: C3; C12; C15; C52.

1. Introduction

Exogeneity tests of the type proposed by Durbin (1954), Wu (1973, 1974), and Hausman (1978), henceforth DWH tests, are widely used in applied work to determine whether ordinary least squares (OLS) or instrumental variables (IVs) method is appropriate. There is now a considerable body of research[1] on this topic, and most studies often impose identifying assumptions on model coefficients, thus leaving out issues associated with *weak instruments*. It is well known that IV estimators can be imprecise and that inference procedures (such as tests and confidence sets) can be highly unreliable in the presence of weak instruments. In recent years, concerns have been raised about the reliability of DWH procedures in the presence of weak instruments because they mainly rely on IV estimators.[2]

Staiger and Stock (1997) show that the limiting distributions of Hausman (1978) type statistics depend on the concentration matrix, which usually determines the strength of the identification. Meanwhile, Wu (1973, 1974) T_2 and T_4 statistics are asymptotically pivotal under exogeneity even when identification is weak. Doko Tchatoka and Dufour (2011b) provide a characterization of the finite-sample distributions of DWH statistics, allowing for the possibility of identification failure and non-Gaussian errors. They show that the statistics T_1, T_2, and T_4 by Wu (1973, 1974) are pivotal under exogeneity without identifying assumptions, even when the errors are non-Gaussian. However, Wu (1973, 1974) T_3 and alternative Hausman (1978) type statistics do not share this property, but they are boundedly pivotal with or without non-Gaussian errors, no matter how weak the instruments are. This suggests that all DWH tests are valid (in the sense that their level is controlled) in the presence of weak instruments if the usual F or asymptotic χ^2 critical values are applied in the inference. However, when applying the asymptotic χ^2 critical values, Wu (1973, 1974)

[1]For examples, see Revankar and Hartley (1973), Farebrother (1976), Revankar (1978), Dufour (1987), Hwang (1980, 1985), Kariya and Hodoshima (1980), Nakamura and Nakamura (1981), Engle (1982), Holly (1982), Reynolds (1982), and Smith (1984, 1985, 1987), among others.

[2]For examples, see Staiger and Stock (1997), Wong (1996, 1997), Guggenberger (2010), Hahn, Ham and Moon (2010), Doko Tchatoka and Dufour (2011a, 2011b, 2014), Kiviet and Niemczyk (2007, 2012), and Kiviet and Pleus (2012), among others.

T_3 and Hausman (1978) statistic can lead to overly conservative procedures when identification is weak. Size correction of the latter statistics is nonetheless achievable by resorting to the exact Monte Carlo tests method such as in Dufour (2006) [see Doko Tchatoka and Dufour (2011b)]. However, the proposed Monte Carlo method requires that the conditional distribution of the structural disturbance, given the instruments, be specified, at least up to an unknown scalar factor. In many empirical applications, researchers usually do not know the distribution of the errors even conditionally on available instruments. So, implementing the exact Monte Carlo tests can be difficult. Therefore, providing a distributional-free method to improve the size and power of the DWH tests, especially when identification is not very strong, can be of great interest in applied work. This paper attempts to make progress in this direction.

To be more specific, we propose a bootstrap method for DWH exogeneity statistics that is robust to identifying assumptions and does not require any distributional assumption on model errors. To do this, we exploit the score interpretation of these statistics [see Engle (1982) and Smith (1983)] to suggest a bootstrap method similar to those of Moreira, Porter and Suarez (2009).[3] We provide an analysis of the limiting distributions of the proposed bootstrap DWH statistics under both the null hypothesis (size) and the alternative hypothesis (power), with or without weak instruments. Our results provide some new insights and extensions of earlier studies. We show that when identification is strong, the proposed bootstrap method offers an asymptotic refinement of the size of the DWH tests under exogeneity. Furthermore, the proposed bootstrap tests are consistent under fixed alternative hypotheses. However, when identification is weak, the bootstrap only provides a first-order approximation of the asymptotic distributions of DWH statistics. Although high-order refinement is no longer possible under weak instruments, the first-order validity of the bootstrap represents an improvement, particulary for the quasi-Wald versions of the

[3] Moreira (2009, footnote 1) shows that testing exogeneity is equivalent to test a null hypothesis of the form $H_\beta : \beta = \beta_0$ in model (2.1)-(2.2), where $\beta_0 = \sigma_{12}/\sigma^2_{v_2}$, $\sigma_{12} = cov(v_{1t}, v_{2t})$, $\sigma^2_{v_2} = var(v_{2t})$ for all $t = 1, \ldots, n$, and v_1 and v_2 are given in (2.3). This suggests that the bootstrap method in Moreira et al. (2009) for the score statistics of $H_\beta : \beta = \beta_0$ can be adapted to exogeneity statistics.

DWH tests which are not asymptotically pivotal (under exogeneity) with weak IVs.

We stress the fact that the validity of the bootstrap for Wu (1973, 1974) T_2 and T_4 statistics and the related Durbin (1954) statistics can be viewed as a generalization of Moreira et al. (2009) to score tests for exogeneity because these statistics are typically score tests; see Engle (1982) and Smith (1983). However, the validity of the bootstrap for the quasi-Wald DWH tests [including the Hausman (1978) test] is not intuitive. Indeed, it is well known that the bootstrap often fails for Wald-type statistics when instruments are weak because their limiting distributions often depend on nuisance parameters; see Dufour (1997, 2003), Andrews and Guggenberger (2010), and Moreira et al. (2009), among others. The bootstrap validity under weak instruments here is mainly justified by the fact that even the quasi-Wald DWH statistics do not directly depend on the unidentified coefficient of the endogenous regressors in the structural equation of interest, whether endogeneity is present or not; see Wu (1973, Section 3), Wu (1974, eqs. 3.11 - 3.16), and Doko Tchatoka and Dufour (2011a, 2011b).

This paper is organized as follows. In section 2, the model and assumptions are formulated, and the studied statistics are presented. In section 3, the proposed bootstrap method is discussed and the properties of the corresponding DWH tests are characterized. In section 4, the Monte Carlo experiment is presented, and the auxiliary lemmas and proofs are provided in the Appendix. Throughout the paper, I_q stands for the identity matrix of order q. For any full-column rank $n \times m$ matrix A, $P_A = A(A'A)^{-1}A$ is the projection matrix on the space spanned by 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 squared matrix B means that B is positive definite. Convergence almost surely is symbolized by "$a.s.$," "$\xrightarrow{p}$" stands for convergence in probability, while "$\xrightarrow{d}$" means convergence in distribution. The usual orders of magnitude are denoted by $O_p(.)$, $o_p(.)$, $O(1)$, and $o(1)$. $\|U\|$ denotes the usual Euclidian or Frobenius norm for a matrix U, while $rank(U)$ is the rank of U. For any set $\mathscr{B}$, $\partial\mathscr{B}$ is the boundary of $\mathscr{B}$ and $(\partial\mathscr{B})^{\varepsilon}$ is the ε-neighborhood of $\mathscr{B}$. Finally, $\sup_{\omega\in\Omega}|f(\omega)|$ is the supremum norm on the space of bounded continuous real functions, with topological space Ω.

2. Framework

We consider a standard linear structural model described by the following equations:

$$y_1 = y_2\beta + Z_1\gamma + u, \tag{2.1}$$

$$y_2 = Z_1\pi_1 + Z_2\pi_2 + v_2, \tag{2.2}$$

where $y_1 \in \mathbb{R}^n$ is a vector of observations on a dependent variable, $y_2 \in \mathbb{R}^n$ is a vector of observations on a (possibly) endogenous explanatory variable, $Z_1 \in \mathbb{R}^{n\times k_1}$ is a matrix of observations on exogenous variables excluded from the structural equation (2.1), $Z_2 \in \mathbb{R}^{n\times k_2}$ is a matrix of excluded exogenous from (2.1) (instruments), $u = (u_1, \ldots, u_n)' \in \mathbb{R}^n$ is a vector of structural disturbances, $v_2 = (v_{21}, \ldots, v_{2n})' \in \mathbb{R}^n$ is a vector of reduced-form disturbances, β and γ are unknown fixed scalar coefficients (structural parameters), and $\pi_1 \in \mathbb{R}^{k_1}$ and $\pi_2 \in \mathbb{R}^{k_2}$ are vectors of reduced-form coefficients.

Let $Z = [Z_1, Z_2] = [Z_{\bullet 1}, \ldots, Z_{\bullet n}]'$, $Y = [y_1, y_2] = [Y_1, \ldots, Y_n]'$, and $\mathscr{R}_n = vech\left[(Y_n', Z_{\bullet n}')'(Y_n', Z_{\bullet n}')\right] = \left(f_1(Y_n', Z_{\bullet n}'), f_2(Y_n', Z_{\bullet n}'), \ldots, f_m(Y_n', Z_{\bullet n}')\right)'$, where $f_p(\cdot)$ $[p = 1, \ldots, m = \frac{1}{2}(k+2)(k+3)$ and $k = k_1 + k_2]$ are elements of $(Y_n', Z_n')'(Y_n', Z_n')$. We suppose that Z has full-column rank k with probability one and $k_2 \geq 1$. The full-column rank condition of Z ensures the existence of unique least squares estimates in (2.2) when y_2 is regressed on each column of Z. As long as Z has full-column rank with probability one and the conditional distribution of y_2 given Z is absolutely continuous (with respect to the Lebesgue measure), $[y_2, Z_1]$ also has full-column rank with probability one. So, the least squares estimates of β and γ in (2.1) are also unique. From (2.1)-(2.2), we can express the reduced-forms for y_1 and y_2 as:

$$(y_1 : y_2) = Z_1(\pi_1\beta + \gamma : \pi_1) + Z_2(\pi_2\beta : \pi_2) + (v_1 : v_2), \tag{2.3}$$

where $v_1 = u + v_2\beta$. If u and v_2 have zero means[4] and Z has full-column rank with probability one, then the usual necessary and sufficient condition for the identification of model (2.1)-(2.2) is $\pi_2 \neq 0$. If $\pi_2 = 0$, Z_2 is irrelevant, and β and γ are completely unidentified. If π_2 is close to zero, β and γ are ill-determined by the data, a situation often called "weak identification" in the literature; see Staiger and Stock (1997), Stock, Wright and Yogo (2002), Dufour (2003), and Andrews and Stock (2007).

We make the following generic assumptions on the model variables, where $\Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{v_2u} \\ \sigma_{v_2u} & \sigma_{v_2}^2 \end{pmatrix}$, $Q_Z = \begin{pmatrix} Q_{Z_1} & Q'_{Z_1Z_2} \\ Q_{Z_1Z_2} & Q_{Z_2} \end{pmatrix}$, and $i = \sqrt{-1}$.

Assumption 2.1 $\mathbb{E}(\|\mathscr{R}_n\|^s) < \infty$ *for some* $s \geq 3$ *and* $\limsup_{\|t\|\to\infty} |\, \mathbb{E}(exp(it'\mathscr{R}_n)) \,| < 1$.

Assumption 2.2 *When the sample size n goes to infinity, we have*:

$n^{-1}\sum_{t=1}^{n}(u_t, v_{2t})'(u_t, v_{2t}) \xrightarrow{p} \Sigma > 0$, $n^{-1}\sum_{t=1}^{n} Z_{\bullet t}Z'_{\bullet t} \xrightarrow{p} Q_Z > 0$ *and* $n^{-1/2}Z_{\bullet t}(u_t, v_{2t}) \xrightarrow{d} (\psi_{Zu}, \psi_{Zv_2})$, *where* $\psi_{Zu} = (\psi'_{Z_1u}, \psi'_{Z_2u})' : k \times 1$, $\psi_{Zv_2} = (\psi'_{Z_1v_2}, \psi'_{Z_2v_2})' : k \times 1$, *and* $vec(\psi_{Zu}, \psi_{Zv_2}) \sim N(0, \Sigma \otimes Q_Z)$.

Assumption **2.1** is similar to Moreira et al. (2009, Assumptions 2-3). It requires that $\mathscr{R}_n$ has third moments or greater and that its characteristic function be bounded above by 1. In particular, the third moments of $\mathscr{R}_n$ exist if $\mathbb{E}(\|(Y'_n, Z'_n)\|^{2s}) < \infty$ for some $s \geq 3$. The bound on the characteristic function of $\mathscr{R}_n$ is the commonly used Cramér's condition [see Bhattacharya and Ghosh (1978)]. The first two convergence in Assumption **2.2** are the weak law of large numbers (WLLN) property of $[u, v_2]$ and Z, respectively, while the last one is the central limit theorem (CLT) property.

Under Assumption **2.2**, the exogeneity of y_2 in (2.1) - (2.2) can be expressed as

$$H_0 : \sigma_{v_2u} = 0. \tag{2.4}$$

In this paper, we are concerned with the validity of the bootstrap for the DWH tests of H_0

[4]This assumption may also be replaced by another "location assumption," such as zero medians.

without any identifying assumptions of model (2.1) - (2.2). To investigate this, we consider six alternative versions of the DWH statistics, namely, the three statistics T_l $(l = 2, 3, 4)$ by Wu (1973, 1974) and three alternative Durbin-Hausman statistics $(DW_j,\ j = 1, 2, 3)$. All statistics can be expressed in a unified way [see Staiger and Stock (1997) and Doko Tchatoka and Dufour (2011b)] as:

$$T_l = \kappa_l(\tilde{\beta} - \hat{\beta})^2/\tilde{\omega}_l^2, \quad DW_j = n(\tilde{\beta} - \hat{\beta})^2/\hat{\omega}_j^2, l = 2,3,4,\ j = 1,2,3, \tag{2.5}$$

where $\hat{\beta} = (y_2' M_{Z_1} y_2)^{-1} y_2' M_{Z_1} y_1$ and $\tilde{\beta} = [y_2'(M_{Z_1} - M_Z)y_2]^{-1} y_2'(M_{Z_1} - M_Z)y_1$ are the OLS and IV estimators of β, respectively, $\kappa_2 = n - 2 - k_1$, $\kappa_3 = \kappa_4 = n - 1 - k_1$, and

$$\begin{aligned}
\tilde{\omega}_2^2 &= \tilde{\sigma}_2^2 \hat{\Delta},\ \tilde{\omega}_3^2 = \tilde{\sigma}^2 \hat{\Delta},\ \tilde{\omega}_4^2 = \hat{\sigma}^2 \hat{\Delta}, \quad \hat{\omega}_1^2 = \tilde{\sigma}^2 \hat{\omega}_{IV}^{-1} - \hat{\sigma}^2 \hat{\omega}_{LS}^{-1},\ \hat{\omega}_2^2 = \tilde{\sigma}^2 \hat{\Delta},\ \hat{\omega}_3^2 = \hat{\sigma}^2 \hat{\Delta}, \\
\hat{\Delta} &= \hat{\omega}_{IV}^{-1} - \hat{\omega}_{LS}^{-1},\ \hat{\omega}_{IV} = y_2'(M_{Z_1} - M_Z)y_2/n,\ \hat{\omega}_{LS} = y_2' M_{Z_1} y_2/n,\ \tilde{\sigma}_2^2 = \hat{\sigma}^2 - (\tilde{\beta} - \hat{\beta})^2/\hat{\Delta} \\
\tilde{\sigma}^2 &= (y_1 - y_2\tilde{\beta})' M_{Z_1}(y_1 - y_2\tilde{\beta})/n,\ \hat{\sigma}^2 = (y_1 - y_2\hat{\beta})' M_{Z_1}(y_1 - y_2\hat{\beta})/n.
\end{aligned}$$

Engle (1982) and Smith (1983) show that T_2, T_4, and DW_3 in (2.5) are score (LM) statistics, while T_3, DW_1, and DW_2 are quasi-Wald statistics.[5] The regression interpretation is also provided in Hausman (1978), Dufour (1979, 1987), and Davidson and Mackinnon (1993) for the statistic T_2, and in Doko Tchatoka and Dufour (2011b) for all statistics. Finite-sample distributions are available in Wu (1973) for T_2 when the errors are Gaussian and IVs are strong. Doko Tchatoka and Dufour (2011b) provide a characterization of the finite-sample distributions of all statistics [including the Hausman (1978) statistic], allowing for the possibility of identification failure and non-Gaussian errors. If instruments are sufficiently strong, all statistics in (2.5) have the usual asymptotic χ^2-distribution under H_0 and Assumption **2.2**. The problem, however, is that T_3, DW_1, and DW_2 can be overly conservative when IVs are weak.[6] This paper examines whether the bootstrap can improve the

[5] See Smith (1983) for the score interpretation (eqs. 6 and 9) and for the quasi-Wald interpretation (eqs. 7, 8 and 10).

[6] See Staiger and Stock (1997), Guggenberger (2010), and Doko Tchatoka and Dufour (2011a, 2011b).

size and power of the above DWH tests, especially when identification is not very strong. Wong (1996) suggests that bootstrapping the Hausman (1978) exogeneity test improves its size and power. Li (2006) extends this result to models with serially correlated errors. However, neither Wong (1996) nor Li (2006) provides a formal proof of the validity of their bootstraps even when IVs are strong. Both studies are Monte Carlo study experiments and the designs of the experiments only consider cases where the strength of the instruments is moderate. In this paper, our analysis allows for any arbitrary level of instrument strength.

It is also worth noting that the exogeneity tests in (2.5) have their own shortcomings. Indeed, Moreira (2009, footnote 1) shows that testing $H_0 : \sigma_{v_2 u} = 0$ is equivalent to test $H_\beta : \beta = \beta_0$ in model (2.1)-(2.2), where $\beta_0 = \sigma_{12}/\sigma^2_{v_2}$, $\sigma_{12} = cov(v_{1t}, v_{2t})$, $\sigma^2_{v_2} = var(v_{2t})$ for all $t = 1, \ldots, n,$ and v_1 and v_2 are given in (2.3). This means that doing a pre-test on $\sigma_{v_2 u}$ may imply important size distortions when making inference on β using a t-type test after the pre-test. Guggenberger (2010) shows that the asymptotic size of the two-stage t-test where a DWH pre-test is used in the first-stage equals 1 for empirically relevant choices of the parameter space. Despite this important issue, the DWH tests are still widely used in many empirical work: for example, the Baum, Schaffer and Stillman (2003) versions of these tests is now implemented in Stata and the study has more than three hundred citations in *RePEc*. Wong (1997) shows through a Monte Carlo study experiment that bootstrapping substantially improves inference over the conventional pretesting procedure. Our goal in this paper is to make progress in this direction.

3. Bootstrap validity for the DWH tests

We now wish to describe the proposed bootstrap method for the DWH exogeneity statistics. Let $\hat{\pi} = (Z'Z)^{-1}Z'y_2$ denote the first-stage OLS estimate of $\pi = (\pi_1', \pi_2')'$ in (2.2) and $\hat{\beta}$ and $\hat{\gamma}$ be the OLS estimates of β and γ from the structural equation (2.1). We adapt the bootstrap procedure by Moreira et al. (2009) to DWH exogeneity statistics as follows: **(1)**

from the observed data, compute $\hat{\pi}$ and $\hat{\beta}$ along with all other things necessary to obtain the realizations of the statistics T_l, DW_j, and the residuals from the reduced-form equation (2.3): $\hat{v}_1 = y_1 - Z_1(\hat{\pi}_1\hat{\beta} + \hat{\gamma}) - Z_2\hat{\pi}_2\hat{\beta}$, $\hat{v}_2 = y_2 - Z\hat{\pi}$. These residuals are then re-centered by subtracting sample means to yield $(\tilde{v}_1, \tilde{v}_2)$; **(2)** for each bootstrap sample $r = 1, \ldots, B$, the data are generated following

$$y_1^* \;=\; Z_1^*(\hat{\pi}_1\hat{\beta} + \hat{\gamma}) + Z_2^*\hat{\pi}_2\hat{\beta} + v_1^*, \quad y_2^* = Z^*\hat{\pi} + v_2^*, \tag{3.1}$$

where $Z^* = [Z_1^* : Z_2^*]$ and (v_1^*, v_2^*) are drawn independently from the joint empirical distribution of Z and $(\tilde{v}_1, \tilde{v}_2)$. The corresponding bootstrap statistics $T_l^{*^r}$ and $DW_j^{*^r}$ are then computed for each bootstrap sample $r = 1, \ldots, B$; **(3)** the simulated bootstrap p-value of each statistic is obtained as the proportion of bootstrap statistics that are more extreme than the computed statistic from the observed data; **(4)** the corresponding bootstrap test rejects exogeneity at level α if its p-value is less than α.

Although the above bootstrap steps are similar to those in Moreira et al. (2009), it is worth noting that there is a substantial difference. In contrast to Moreira et al. (2009), where the two-stage least squares (2SLS) or the limited information maximum likelihood (LIML) estimators are suggested as the pseudo-true value of β under the bootstrap DGP, our algorithm uses the OLS estimator of β in (2.1). Indeed, Moreira et al. (2009) show that the validity of their bootstrap requires using a strong consistent estimator of β; i.e., an estimator $\hat{\beta}$ that satisfies $\hat{\pi} \xrightarrow{p} \pi$ and $\hat{\beta}\hat{\pi} \xrightarrow{p} \beta\pi$. In a linear classical setting of this paper, both the 2SLS and LIML estimators satisfy the sufficient conditions for strong consistency [see Moreira et al. (2009, Proposition 4 and fn.3, p.55)]. The OLS estimator $\hat{\beta}$ is not qualified for strong consistency when $\sigma_{v_2u} \neq 0$ (endogeneity). However, when $\sigma_{v_2u} = 0$ (exogeneity), $\hat{\beta}$ is consistent and efficient even when IVs are weak. Based on this fact, $\hat{\beta}$ is preferred to an alternative 2SLS or LIML estimator because the choice of the latter should imply a sizable efficiency loss under H_0. Moreover, choosing $\hat{\beta}$ as the pseudo-true value of

β when H_0 is false is suggested in Horowitz (1994, Section 2.3) to approximate the power function of the bootstrap tests.[7]

In the reminder of the paper, let F_n denote the empirical distribution of $\mathscr{R}_n^* = vech\left((Y_n^{*\prime}, Z_n^{*\prime})'(Y_n^{*\prime}, Z_n^{*\prime})\right)$ conditional on $\mathscr{F}_n = \{(Y_1', Z_1'), \ldots, (Y_n', Z_n')\}$, $\mathbb{P}^*$ be the probability under the empirical distribution function (conditional on$\mathscr{F}_n$), and $\mathbb{E}^*$ its corresponding expectation operator. Also, let $G_1(\cdot)$ and $g_1(\cdot)$ be the cumulative density function (cdf) and the probability density function (pdf) of a χ^2-distributed random variable with one degree of freedom. In order to ease the exposition of our results, we shall deal separately with the case where identification is strong and the one where it is weak. Since strong identification is relatively easy to tackle, we will focus on that case first. Section 3.1 presents the results.

3.1. Strong identification

In this section, we focus on the case where identification is strong and provide an analysis of the size and power properties of the proposed bootstrap DWH tests. Let $c^*_{T_l,\alpha}$ and $c^*_{DW_j,\alpha}$ denote the $1-\alpha$ quantile of T_l^* ($l = 2, 3, 4$) and DW_j^* ($j = 1, 2, 3$), respectively. Following Andrews (2002), we define $c^*_{T_l,\alpha}$ to be the value that minimizes $\mid \mathbb{P}^*(T_l^* \leq \tau) - (1-\alpha) \mid$ over $\tau \in \mathbb{R}$, and similarly for $c^*_{DW_j,\alpha}$. Theorem **3.1** establishes: (1) asymptotic refinements of the bootstrap DWH tests under exogeneity ($\sigma_{v_2u} = 0$), and (2) the bootstrap DWH tests consistency under fixed endogeneity ($\sigma_{v_2u} \neq 0$ is fixed).

Theorem 3.1 *Suppose that Assumptions* **2.1 - 2.2** *are satisfied and that* $\pi_2 \neq 0$ *is fixed. Then for all* $l = 2, 3, 4$ *and* $j = 1, 2, 3$, *we have*:

(a) $\mid \mathbb{P}(T_l > c^*_{T_l,\alpha}) - \alpha \mid = o(n^{-1}), \mid \mathbb{P}(DW_j > c^*_{DW_j,\alpha}) - \alpha \mid = o(n^{-1})$ *if* $\sigma_{v_2u} = 0$;

(b) $\mathbb{P}(T_l > c^*_{T_l,\alpha}) \to 1, \mathbb{P}(DW_j > c^*_{DW_j,\alpha}) \to 1$ *as* $n \to +\infty$ *if* $\sigma_{Vu} \neq 0$ *is fixed* .

[7] Also, see Bickel and Freedman (1981) for further developments on the asymptotic validity of the bootstrap.

Theorem **3.1**-(a) gives the conditions under which the bootstrap critical values for T_l^* and DW_j^* yield levels for the T_l and DW_j tests that are correct through $O(n^{-1})$ under H_0. So, the bootstrap makes an error of size $O(n^{-1})$ under exogeneity, which is smaller as $n \to +\infty$ than both $O(n^{-1/2})$ and the error made by the first-order asymptotic approximations. The bootstrap provides a greater accuracy than the $O(n^{-1/2})$ order because all DWH statistics are quadratic functions of symmetric pivotal statistics [see Horowitz (2001, Ch. 52, eq. 3.13)] under exogeneity ($\sigma_{v_2u} = 0$) and strong identification ($\pi_2 \neq 0$). Theorem **3.1**-(b) shows that test consistency holds if the bootstrap critical values are used in the inference. This is because the asymptotic distributions of all DWH statistics diverge under fixed endogeneity ($\sigma_{v_2u} \neq 0$ is fixed) and strong identification ($\pi_2 \neq 0$). Hence, test consistency will still hold if the empirical *size-corrected* critical values are used instead of the bootstrap ones, as suggested by Horowitz (1994). Although Theorem **3.1**-(b) considers the case of fixed endogeneity, there is no impediment to expanding it to local to zero alternatives of the form $\mathrm{H}_{1c} : \sigma_{v_2u} = c/\sqrt{n}$ for some scalar $c \neq 0$. In this case, test consistency does no longer hold because $\sigma_{v_2u} \to 0$ as $n \to +\infty$ so that the limiting distributions of the bootstrap DWH statistics are not far from their limiting distributions under exogeneity ($\sigma_{v_2u} = 0$). However, we can provide high-order refinements of the power functions of the tests similar to Theorem **3.1**-(a), where the Edgeworth expansion is run at $\mu_c = \mathbb{E}(\mathscr{R}_n) \neq 0$; for example, see Taniguchi (1988). This proof is omitted in order to shorten the exposition.

3.2. Local to zero weak instruments

We now analyze the Staiger and Stock (1997) local to zero weak instruments framework. To be more specific, we assume that $\pi_2 = \pi_{02}/\sqrt{n}$, where $\pi_{02} \in \mathbb{R}^{k_2}$ is a fixed vector (possibly zero). As argued previously, high-order refinements of the size of the tests in Theorem **3.1**-(a) are no longer achievable due to identification failure. This is because the functions $H(\cdot)$ and $\tilde{H}(\cdot)$ in eqs.(A.2) - (A.3) of the appendix are no longer smooth around $\mu = \mathbb{E}(\mathscr{R}_n)$. For example, both functions depend on $y_2'(M_{Z_1} - M_Z)y_2/n$ but their derivatives with respect to $y_2'(M_{Z_1} - M_Z)y_2/n$ is not well-defined when $\pi_2 = 0$ or does not even exist if $\pi_2 = \pi_{02}c_n$ for any sequence $c_n \downarrow 0$ [similar to Moreira et al. (2009, footnote 2)]. This is particularly the case in the Staiger and Stock (1997) setup where $c_n = 1/\sqrt{n}$. Nonetheless, we can establish the following theorem on the first-order validity of the bootstrap.

Theorem 3.2 *Suppose that Assumption* **2.2** *is satisfied and that* $\pi_2 = \pi_{02}/\sqrt{n}$, *where* $\pi_{02} \in \mathbb{R}^{k_2}$ *is fixed. If further for some* $\delta > 0$, $\mathbb{E}(\|Z_i\|^{4+\delta}, \|v_i\|^{2+\delta}) < \infty$, *then we have*:

(*a*) $|\,\mathbb{P}(T_l > c^*_{T_l,\alpha}) - \alpha\,| = o(1), |\,\mathbb{P}(DW_j > c^*_{DW_j,\alpha}) - \alpha\,| = o(1)$ *if* $\pi_{02}\sigma_{v_2u} = 0$;

(*b*) $|\,\mathbb{P}(T_l > c^*_{T_l,\alpha}) - (1-\alpha_{T_l})\,| = o(1), |\,\mathbb{P}(DW_j > c^*_{DW_j,\alpha}) - (1-\alpha_{DW_j})\,| = o(1)$ *if* $\pi_{02}\sigma_{Vu} \neq 0$,

where $\alpha_{T_l} = \mathbb{E}[G_1(c^*_{T_l,\alpha}; \|\mu\|^2)]$ *for* $l = 2,4$, $\alpha_{T_3} = \mathbb{E}[(\sigma_u^2/\bar{\sigma}_u^2)G_1(c^*_{T_3,\alpha}; \|\mu\|^2)]$, $\alpha_{DW_j} = \mathbb{E}[(\sigma_u^2/\bar{\sigma}_u^2)G_1(c^*_{DW_j,\alpha}; \|\mu\|^2)]$ *for* $j = 1,2$, $\alpha_{DW_3} = \mathbb{E}[G_1(c^*_{DW_3,\alpha}; \|\mu\|^2)]$, $G_1(.; \|\mu\|^2)$ *is the cdf of a noncentral* χ^2 *distributed random variable with one degree of freedom and non-centrality parameter* $\|\mu\|^2$, $\bar{\sigma}_u^2$ *and* μ *are defined in Lemma* **A.4**.

Theorem **3.2**-(a) holds in particular when $H_0 : \sigma_{v_2u} = 0$ is satisfied. So, the bootstrap critical values for T_l^* and DW_j^* yield levels for the T_l and DW_j tests that are asymptotically correct under exogeneity, despite the lack of identification. Since the statistics T_2, T_4, and DW_3 are score statistics [see Engle (1982) and Smith (1983)], and that testing H_0 is equivalent to test $H_\beta : \beta = \beta_0 = \sigma_{12}/\sigma_{v_2}^2$ [see Moreira (2009, footnote 1)], the bootstrap validity for these statistics under weak instruments is not surprising, as established in Moreira et al. (2009). However, the bootstrap validity for the quasi-Wald tests T_3, DW_1, and DW_2 is not intuitive. In general, the bootstrap often fails for Wald-type statistics when instruments are weak because their limiting distributions often involve nuisance parameters; see Dufour (1997, 2003, 2006), Andrews and Guggenberger (2010), and Moreira et al. (2009), among others. The validity of the bootstrap for these statistics here is mainly justified by the fact that they do not directly depend on the structural coefficient β even when endogeneity is present; see Wu (1973, Section 3), Wu (1974, eqs. 3.11 - 3.16), and Doko Tchatoka and Dufour (2011a, 2011b).

Theorem **3.2**-(b) indicates that the DWH tests may still exhibit power under weak instruments if the bootstrap critical values are used, provided that $\pi_{02}\sigma_{v_2u} \neq 0$. However, the bootstrap critical values yield tests with power no higher than the nominal size for any level of endogeneity if $\pi_{02}\sigma_{v_2u} = 0$, as showed Theorem **3.2**-(a). This is particularly the case when all instruments are irrelevant ($\pi_{02} = 0$) so that β is completely unidentified. This is because when $\pi_{02}\sigma_{v_2u} = 0$, the limiting distributions of all DWH statistics (both the standard and bootstrap statistics) are the same as under exogeneity, despite the fact that $\sigma_{v_2u} \neq 0$; see Lemmas **A.4** and **A.7** in Appendix. This result is intuitive because σ_{v_2u} is not identifiable if β is not identifiable [see Doko Tchatoka and Dufour (2014)], which is the case[8] when $\pi_{02}\sigma_{v_2u} = 0$. In many empirical applications, not all instruments are often irrelevant, so, the bootstrap tests may have power if at least one instrument is not very poor.

4. Monte Carlo experiment

We use simulation to examine the finite-sample performance of the proposed bootstrap DWH tests. The DGP is described by eqs. (2.1) and (2.2) with $k_1 = 0$ [no exogenous instruments in (2.1)], Z_2 contains $k_2 = 5$ instruments, each generated $i.i.d.$ $N(0,1)$ and independent of $(u_t, v_{2t})'$ for all $t = 1, \ldots, n$. The true value of β is set at 2 and the reduced-form coefficient π_2 is chosen as $\pi_2 = \lambda\pi_{02}$, where $\pi_{02} \in \mathbb{R}^{k_2}$ is a vector of ones and $\lambda \in \{0, 0.05, 0.1, 1\}$ characterizes the strength of the instruments.[9] More precisely, $\lambda = 0$ is a design of complete non-identification (irrelevant instruments), $\lambda = 0.05$ designs weak identification, $\lambda = 0.1$ is the setup of moderate identification, and finally $\lambda = 1$ symbolizes strong identification. The errors $(u_t, v_{2t})'$ are generated as $(u_t, v_{2t})' = J\varepsilon_t$, where $\varepsilon_t \sim \mathbf{N}(0, I_2)$, $J = \begin{pmatrix} 1 & \frac{\rho_{v_2u}}{(1+\rho_{v_2u}^2)^{1/2}} \\ 0 & \frac{1}{(1+\rho_{v_2u}^2)^{1/2}} \end{pmatrix}$, and ρ_{v_2u} measures the endogeneity of y_2.

[8] When $\pi_{02}\sigma_{v_2u} = 0$, all values of $(\beta, \sigma_{v_2u}) \in \mathbb{R}^2$ are *observationally equivalent* so that the bootstrap tests have no power to discriminate between $\sigma_{v_2u} \neq 0$ and $\sigma_{v_2u} = 0$.

[9] Dufour and Taamouti (2007) and Doko Tchatoka (2014) consider similar data-generating processes for the first-stage regression (2.2).

The exogeneity of y_2 is then expressed as $H_0 : \rho_{v_2u} = 0$. In this experiment, we vary ρ_{v_2u} in $\{-0.8, 0, 0.2, 0.9\}$, but the results do not change qualitatively for alternative values of ρ_{v_2u}. The rejection frequencies are computed using $1,000$ replications for the standard DWH tests, while those of the bootstrap tests are obtained with $N = 1,000$ replications and $B = 999$ bootstrap pseudo-samples of size $n = 300$.

Table 1 presents the results. In the first part of the table, the rejection frequencies of the DWH tests are reported using both the bootstrap and usual asymptotic critical values. The second part of the table reports the rejection frequencies of Wong's (1997) second-stage bootstrap tests as well as the usual second-stage two-sided t-tests of $H_\beta : \beta = \beta_0$ (after pre-testing exogeneity). The nominal levels are set at 0.05 for both the pre-tests and the second-stage tests. The first column of the table contains the statistics, while the others present, for each value of the endogeneity ρ_{v_2u} and IV quality λ, the empirical rejection frequencies of the tests. The column $\rho_{v_2u} = 0$ represents exogeneity, while those with $\rho_{v_2u} \neq 0$ indicate endogeneity.

First, we observe that the empirical rejection frequencies of all DWH tests when the bootstrap critical values are used are close to the 5% nominal level when $\rho_{v_2u} = 0$ (exogeneity), irrespective of the value of λ (quality of the instruments), thus confirming our theoretical findings in Theorems **3.1**-(a) and Theorem **3.2**-(a). Meanwhile, only the LM versions of these tests, namely, T_2, T_4, and DW_3, have correct size asymptotically for $\lambda \in \{0, 0.05, 0.1\}$ (weak or moderate instruments) when the usual asymptotic critical values are used in the inference. The quasi-Wald versions of the DWH tests (T_3, DW_1, and DW_2) are overly conservative under weak instruments [$\lambda \in \{0, 0.05\}$]. As expected, the rejection frequencies of all tests are similar and close to the 5% nominal level under exogeneity ($\rho_{v_2u} = 0$) and strong identification ($\lambda = 1$), with both the bootstrap and asymptotic χ^2 critical values. Clearly, bootstrapping substantially improves the size of T_3, DW_1, and DW_2, especially when identification is weak.

Second, when $\rho_{v_2u} \neq 0$ (endogeneity), we observe that all tests have empirical rejections approaching or equal to 100% when identification is moderate or strong and endogeneity is large [see columns $\rho_{v_2u} \in \{-0.8, 0.9\}$ and $\lambda \in \{0.1, 1\}$], whether the bootstrap or asymptotic χ^2 critical values are used, thus confirming our theoretical findings in Theorem **3.1** - (b). However, all tests have low power when identification is very weak [columns $\rho_{v_2u} \neq 0$ and $\lambda \in \{0, 0.05\}$] even if the bootstrap critical values are used, as expected from the theoretical results in Theorem **3.2** - (a). Nevertheless, even the quasi-Wald DWH tests exhibit power with weak instruments and large endogeneity [see columns $\lambda = 0.05$ and $\rho_{v_2u} \in \{-0.8, 0.9\}$] if the bootstrap critical values are used, thus confirming the analysis in Theorem **3.2** - (b). So, in contrast to the usual asymptotic χ^2-critical values, the bootstrap improves the power of the quasi-Wald versions of the DWH tests, provided that identification is not completely irrelevant. For examples, the empirical rejection frequencies of T_3, DW_1, and DW_2 with the bootstrap critical values when $\rho_{v_2u} = 0.9$ and $\lambda = 0.05$ are around 24%. These represents nearly the triple of their rejection frequencies when the asymptotic χ^2-critical values are used (about 8%). Furthermore, the tests T_3, DW_1, and DW_2 become competitive with the bootstrap critical values in terms of power compared with T_2, T_4, and DW_3. Indeed, while the usual asymptotic LM tests (T_2, T_4, and DW_3) strictly dominate the quasi-Wald ones (T_3, DW_1, and DW_2), the empirical rejection frequencies using the bootstrap critical values are very close for all tests even when identification is not strong.

Finally, the second part of the table show that the size of the usual two-stage t-tests, where a DWH-type pre-test is used in the first-stage, is close to 1 when $\lambda \in \{0, 0.05, 0.1\}$ (weak instruments) and endogeneity is present ($\rho_{v_2u} \neq 0$).[10] The usual two-stage t-tests have approximately good size asymptotically only when identification is very strong $\lambda = 1$ (see the bottom part of Table 1). In contrast, the bootstrap tests of $H_\beta : \beta = \beta_0$, similar to those of Wong (1997), substantially improves the size over the conventional two-stage t-tests. For example, the maximal rejection frequencies of the proposed bootstrap tests of $H_\beta : \beta = \beta_0$ are around 12%, while those of the conventional two-stage t-tests can

[10]Similar to Guggenberger (2010).

be as high as 99.9%. More interestingly, while the usual asymptotic LM versions of the DWH tests (T_2, T_4, and DW_3) have correct size and are (almost uniformly) more powerful than their bootstrap's counterparts, their use as pre-tests lead to serious size distortions of the second stage *t*-tests. Meanwhile, using the bootstrap versions of all DWH tests (including the quasi-Wald DWH tests) substantially improves the conventional pre-testing based-inference.

Acknowledgments

We are grateful to Professor Richard J. Smith, Managing Editor of *The Econometrics Journal* and two anonymous referees for their constructive comments and suggestions. We would also like to thank Professors Jean-Marie Dufour, Mardi Dungey, and Jan Kiviet for several useful comments. This project is supported by a Faculty of Business and School of Economics and Finance (University of Tasmania) joint *research grant*, and I am grateful for their support.

APPENDIX

A. Auxiliary Lemmata and Proofs

In the remainder of this appendix, we define $\tilde{S}_{l,n} = \sqrt{n}(\tilde{\beta} - \hat{\beta})/\tilde{\omega}_l$ and $\hat{S}_{j,n} = \sqrt{n}(\tilde{\beta} - \hat{\beta})/\hat{\omega}_j$, $l = 2,3,4$; $j = 1,2,3$, where $\tilde{\beta}$, $\hat{\beta}$, $\tilde{\omega}_l$, and $\hat{\omega}_j$ are given in (2.5). It is then easy to see that the statistics T_l ($l = 2,3,4$) and DW_j ($j = 1,2,3$) in (2.5) can be expressed as

$$T_l \quad = \quad n^{-1}\kappa_l \parallel \tilde{S}_{l,n} \parallel^2, \;\; DW_j = \parallel \hat{S}_{j,n} \parallel^2 . \tag{A.1}$$

We shall now state the following auxiliary lemmas that are used in the proofs of the main results.

Table 1. Rejection frequencies (in percentage) at nominal level 5%

Rejection frequencies of T_l and DW_j tests with the bootstrap and asymptotic χ^2 critical values

	$\rho_{v_2u} = -0.8$				$H_0: \rho_{v_2u} = 0$				$\rho_{v_2u} = 0.5$				$\rho_{v_2u} = 0.9$			
Statistics ↓ $\lambda \rightarrow$	0	0.05	0.1	1	0	0.05	0.1	1	0	0.05	0.1	1	0	0.05	0.1	1
	Bootstrap critical values															
T_2	6.8	17.9	73.4	100	5.2	6.6	5.8	4.5	4.6	11.2	35.7	100	6.2	24.1	82.2	100
T_3	5.5	21.0	71.7	100	4.4	4.4	6.6	4.4	4.1	5.9	35.5	100	4.0	24.4	77.1	100
T_4	6.8	17.9	73.4	100	5.2	6.6	5.8	4.5	4.6	11.2	35.7	100	6.2	24.1	82.2	100
DW_1	5.5	21.0	71.7	100	4.4	4.4	6.6	4.4	4.1	5.9	35.5	100	4.0	24.4	77.1	100
DW_2	5.5	21.0	71.7	100	4.4	4.4	6.6	4.4	4.1	5.9	35.5	100	4.0	24.4	77.1	100
DW_3	6.8	17.9	73.4	100	5.2	6.6	5.8	4.5	4.6	11.2	35.7	100	6.2	24.1	82.2	100
	χ^2-critical values															
T_2	5.0	19.3	74.8	100	4.8	6.2	6.0	4.9	3.8	12.0	38.4	100	5.1	25.6	82.8	100
T_3	0.2	6.7	64.3	100	0.2	0.9	3.0	4.9	0.4	2.1	26.9	100	0.4	8.0	74.4	100
T_4	4.9	19.0	74.7	100	4.8	6.2	5.9	4.9	3.8	11.4	38.2	100	4.9	25.3	82.7	100
DW_1	0.2	6.6	64.0	100	0.2	0.8	3.0	4.8	0.4	2.0	26.7	100	0.4	7.7	73.7	100
DW_2	0.2	6.7	64.4	100	0.2	0.9	3.0	4.9	0.4	2.3	27.2	100	0.4	8.1	74.8	100
DW_3	5.0	19.0	74.7	100	4.8	6.2	5.9	4.9	3.8	11.7	38.3	100	4.9	25.3	82.7	100

Two-sided t-tests and bootstrap tests of $H_\beta : \beta = \beta_0$ after pretesting exogeneity

	$\rho_{v_2u} = -0.8$				$\rho_{v_2u} = 0$				$\rho_{v_2u} = 0.5$				$\rho_{v_2u} = 0.9$			
Statistics ↓ $\lambda \rightarrow$	0	0.05	0.1	1	0	0.05	0.1	1	0	0.05	0.1	1	0	0.05	0.1	1
Pre-tests	Bootstrap tests															
T_2^*	11.7	10.7	7.5	5.3	4.8	5.2	5.1	5.2	10.9	10.1	7.8	5.5	9.3	8.7	7.8	4.8
T_3^*	11.9	11.1	7.5	5.3	4.9	5.0	5.4	5.2	11.2	10.5	8.3	5.5	10.4	10.0	7.8	4.8
T_4^*	11.7	10.7	7.5	5.3	4.8	5.2	5.1	5.2	10.9	10.1	7.8	5.5	9.3	8.7	7.8	4.8
DW_1^*	11.9	11.1	7.5	5.3	4.9	5.0	5.4	5.2	11.2	10.5	8.3	5.5	10.4	10.0	7.8	4.8
DW_2^*	11.9	11.1	7.5	5.3	4.9	5.0	5.4	5.2	11.2	10.5	8.3	5.5	10.4	10.0	7.8	4.8
DW_3^*	11.7	10.7	7.5	5.3	4.8	5.2	5.1	5.2	10.9	10.1	7.8	5.5	9.3	8.7	7.8	4.8
Pre-tests	Two-sided t-tests															
T_2	96.9	81.2	27.6	6.1	5.0	6.1	7.7	6.5	97.3	89.0	61.9	5.7	96.6	75.3	22.7	4.9
T_3	99.9	93.3	37.7	6.1	5.0	6.0	6.8	6.5	99.7	97.9	73.2	5.7	99.8	92.1	29.9	4.9
T_4	96.9	81.2	27.6	6.1	5.0	6.1	7.7	6.5	97.3	89.2	61.9	5.7	96.6	75.3	22.7	4.9
DW_1	99.9	93.4	38.0	6.1	5.0	6.0	6.8	6.5	99.7	98.0	73.4	5.7	99.8	92.3	30.6	4.9
DW_1	99.9	93.3	37.6	6.1	5.0	6.0	6.8	6.5	99.7	97.7	72.9	5.7	99.8	92.0	29.6	4.9
DW_1	96.8	81.2	27.6	6.1	5.0	6.1	7.7	6.5	97.3	89.0	61.8	5.7	96.6	75.3	22.7	4.9

A.1. Auxiliary Lemmata

Lemma A.1 *Suppose that Assumptions* **2.1 - 2.2** *are satisfied and that* $\pi_2 \neq 0$ *is fixed. Then we have*:

(a) $\sup_{\tau\in\mathbb{R}} \mid \mathbb{P}(\tilde{S}_{l,n} \leq \tau) - \Phi(\tau) - \sum_{h=1}^{s-2} n^{-h/2} p^h_{\tilde{S}_{l,n}}(\tau; F, \pi_2)\phi(\tau) \mid = o(n^{-\frac{s-2}{2}})$,

$\sup_{\tau\in\mathbb{R}} \mid \mathbb{P}(\hat{S}_{j,n} \leq \tau) - \Phi(\tau) - \sum_{h=1}^{s-2} n^{-h/2} p^h_{\hat{S}_{j,n}}(\tau; F, \pi_2)\phi(\tau) \mid = o(n^{-\frac{s-2}{2}})$ *if* $\sigma_{v_2 u} = 0$;

(b) $\sup_{\tau\in\mathbb{R}} \mathbb{P}(|\tilde{S}_{l,n}| \leq \tau) \to 0$, $\sup_{\tau\in\mathbb{R}} \mathbb{P}(|\hat{S}_{j,n}| \leq \tau) \to 0$ *as* $n \to +\infty$ *if* $\sigma_{Vu} \neq 0$ *is fixed,*

where $\Phi(\cdot)$ *and* $\phi(\cdot)$ *are the cdf and pdf of a standard normal random variable,* $p^h_{\tilde{S}_{l,n}}$ *and* $p^h_{\hat{S}_{j,n}}$ *are polynomials in* τ *with coefficients depending on* $\hat{\beta}$, $\hat{\pi}_2$, *and the moments of the distribution* F *of* $\mathscr{R}_n$.

Lemma A.2 *Suppose that Assumptions* **2.1 - 2.2** *are satisfied and that* $\pi_2 \neq 0$ *is fixed. Then for some* $r \in \mathbb{N}$, *we have*:

(a) $\sup_{\tau\in\mathbb{R}} \mid \mathbb{P}(T_l \leq \tau) - G_1(\tau) - \sum_{h=1}^{r} n^{-h} p^h_{T_{l,n}}(\tau; F, \pi_2) g_1(\tau) \mid = o(n^{-r})$,

$\sup_{\tau\in\mathbb{R}} \mid \mathbb{P}(DW_j \leq \tau) - G_1(\tau) - \sum_{h=1}^{r} n^{-h} p^h_{DW_{j,n}}(\tau; F, \pi_2) g_1(\tau) \mid = o(n^{-r})$ *if* $\sigma_{v_2 u} = 0$;

(b) $\sup_{\tau\in\mathbb{R}} \mathbb{P}(T_l \leq \tau) \to 0$, $\sup_{\tau\in\mathbb{R}} \mathbb{P}(DW_j \leq \tau) \to 0$ *as* $n \to +\infty$ *if* $\sigma_{Vu} \neq 0$ *is fixed,*

where $G_1(\cdot)$ *and* $g_1(\cdot)$ *are the cdf and pdf of a* $\chi^2(1)$*-distributed random variable,* $p^h_{T_{l,n}}$ *and* $p^h_{DW_{j,n}}$ *are polynomials in* τ *with coefficients depending on* $\hat{\beta}$, $\hat{\pi}_2$, *and the moments of the distribution* F *of* $\mathscr{R}_n$.

Lemma A.3 *Suppose that Assumptions* **2.1 - 2.2** *are satisfied and that* $\pi_2 \neq 0$ *is fixed. Then for some* $r \in \mathbb{N}$, *we have*:

(a) $\sup_{\tau\in\mathbb{R}} \mid \mathbb{P}^*(T_l^* \leq \tau) - G_1(\tau) - \sum_{h=1}^{r} n^{-h} p^h_{T_l}(\tau; F_n, \hat{\beta}, \hat{\pi}_2) g_1(\tau) \mid = o(n^{-r})$,

$\sup_{\tau\in\mathbb{R}} \mid \mathbb{P}^*(DW_j^* \leq \tau) - G_1(\tau) - \sum_{h=1}^{r} n^{-h} p^h_{DW_j}(\tau; F_n, \hat{\beta}, \hat{\pi}_2) g_1(\tau) \mid = o(n^{-r})$ *a.s. if* $\sigma_{v_2 u} = 0$;

(b) $\sup_{\tau\in\mathbb{R}} \mathbb{P}^*(T_l^* \leq \tau) \to 0$, $\sup_{\tau\in\mathbb{R}} \mathbb{P}^*(DW_j^* \leq \tau) \to 0$ *a.s. as* $n \to +\infty$ *if* $\sigma_{Vu} \neq 0$ *is fixed,*

$\hat{p}^h_{T_l}$, $\hat{p}^h_{DW_j}$ *are polynomials in* τ *with coefficients depending on* $\hat{\beta}$, $\hat{\pi}_2$ *and the moments of* F_n.

Lemma A.4 *Suppose Assumption* **2.2** *is satisfied and that* $\pi_2 = \pi_{02}/\sqrt{n}$, $\pi_{02} \in \mathbb{R}^{k_2}$ *is fixed. Then we have*:

(a) $T_2, T_4, H_3 \xrightarrow{d} \chi^2(1)$ *and* $T_3, H_1, H_2 \mid \tilde{\psi}_{Z_2 v_2} \xrightarrow{d} \frac{\sigma_u^2}{\bar{\sigma}_u^2}\chi^2(1) \leq \chi^2(1)$ *when* $\pi_{02}\sigma_{v_2 u} = 0$;

(b) $T_2, T_4, H_3 \mid \tilde{\psi}_{Z_2 v_2} \xrightarrow{d} \chi^2(1, \|\mu\|^2)$ *and* $T_3, H_1, H_2 \xrightarrow{d} \frac{\sigma_u^2}{\bar{\sigma}_u^2}\chi^2(1, \|\mu\|^2) \leq \chi^2(1, \|\mu\|^2)$

when $\pi_{02}\sigma_{v_2 u} \neq 0$, *where* $\mu = \sigma_u^{-1}\sigma_{v_2}^{-2}\tilde{\Psi}_{Z_2 v_2}^{-1/2}\bar{Q}_{Z_2}^{-1}\pi_{02}\sigma_{v_2 u}$

with $\tilde{\Psi}_{Z_2v_2} = (\tilde{\psi}_{Z_2v_2} + \bar{Q}_{Z_2}\pi_{02})'\bar{Q}_{Z_2}^{-1}(\tilde{\psi}_{Z_2v_2} + \bar{Q}_{Z_2}\pi_{02}),\ \tilde{\psi}_{Z_2v_2} = \psi_{Z_2v_2} - Q_{Z_2Z_1}Q_{Z_1}^{-1}\psi_{Z_1v_2},\ \tilde{\psi}_{Z_2u} = \psi_{Z_2u} - Q_{Z_2Z_1}Q_{Z_1}^{-1}\psi_{Z_1u},\ \bar{\sigma}_u^2 = \sigma_u^2 - 2\sigma_{v_2u}\tilde{\Psi}_{Z_2v_2}^{-1}(\tilde{\psi}_{Z_2v_2} + \bar{Q}_{Z_2}\pi_{02})'\bar{Q}_{Z_2}^{-1}\tilde{\psi}_{Z_2u} + \sigma_{v_2}\tilde{\psi}'_{Z_2u}\bar{Q}_{Z_2}^{-1}(\tilde{\psi}_{Z_2v_2} + \bar{Q}_{Z_2}\pi_{02})\tilde{\Psi}_{Z_2v_2}^{-2}(\tilde{\psi}_{Z_2v_2} + \bar{Q}_{Z_2}\pi_{02})'\bar{Q}_{Z_2}^{-1}\tilde{\psi}_{Z_2u}$, *and* $\bar{Q}_{Z_2} = Q_{Z_2} - Q_{Z_2Z_1}Q_{Z_1}^{-1}Q'_{Z_2Z_1}$.

Lemma **A.4**-(a) holds in particular when $\sigma_{v_2u} = 0$ (exogeneity). So, T_2, T_4 and H_3 are pivotal under weak instruments and exogeneity, while H_1, H_2 and T_3 are boundly pivotal. Since Lemma **A.4**-(a) may still hold even if $\sigma_{v_2u} \neq 0$, hence all exogeneity tests have no power against endogeneity if $\pi_{02}\sigma_{v_2u} = 0$. This is the case if $\pi_{02} = 0$ (irrelevant IVs) so that (a) holds for any value of σ_{v_2u}: the power of all tests cannot exceed the nominal level asymptotical; see Doko Tchatoka and Dufour (2011b). However, all tests exhibit power against endogeneity even when IVs are weak, provided that $\pi_{02}\sigma_{v_2u} \neq 0$, as showed Lemma **A.4**-(b).

Lemma A.5 *Suppose that Assumption* **2.2** *is satisfied and that* H_0 *holds. If for some* $\delta > 0$, *we have* $\mathbb{E}(\|Z_t\|^{2+\delta}, \|v_t\|^{2+\delta}) < \infty$, *then* $\mathbb{E}^*(|Z_{jt}^* v_{mt}^*|^{2+\delta})$ *and* $\mathbb{E}^*(|Z_{ji}^* u_i^*|^{2+\delta})$ *are bounded a.s. for all* $j = 1, \ldots, k$ *and* $m = 1, 2$; *where* Z^* *and* $v^* = [v_1^* : v_2^*]$ *are the bootstrap draws from the empirical distribution of* Z *and the re-centered residuals* $\tilde{v} = [\tilde{v}_1 : \tilde{v}_2]$, *and* $u^* = v_1^* - v_2^*\beta$.

Lemma A.6 *Suppose that Assumption* **2.2** *is satisfied. If for some* $\delta > 0$, $\mathbb{E}(\|Z_i\|^{4+\delta}, \|v_i\|^{2+\delta}) < \infty$, *then under* H_0, *we have* $\frac{1}{\sqrt{n}}\left(Z^*u^*, Z^*v_2^*, (\frac{W^{*\prime}\mathbb{1}}{n} - \frac{W'\mathbb{1}}{n})\right) \mid \mathscr{F}_n \xrightarrow{d} N\left[0, \begin{pmatrix} diag(\sigma_u^2, \sigma_{v_2}) \otimes Q_Z & 0 \\ 0 & \Sigma_w \end{pmatrix}\right]$ *a.s, where* $W = (w_1, \ldots, w_n)$, $w_i = vech(Z_iZ_i')$, $W^* = (w_1^*, \ldots, w_n^*)$, $w_i^* = vech(Z_i^*Z_i^{*\prime}) \in \mathbb{R}^{k(k+1)/2}$, $\Sigma_w = var(w_i)$, *and* $\mathbb{1}$ *is a* $(n$ *by* $1)$ *constant vector of ones.*

Lemma A.7 *Suppose Assumption* **2.2** *is satisfied and let* $\pi_2 = \pi_{02}/\sqrt{n}$, $\pi_{02} \in \mathbb{R}^{k_2}$ *is fixed. If for some* $\delta > 0$, $\mathbb{E}(\|Z_i\|^{4+\delta}, \|v_i\|^{2+\delta}) < \infty$, *then, conditional on* $\mathscr{F}_n$, *we have*:

(a) $T_2^*, T_4^*, H_3^* \xrightarrow{d} \chi^2(1)$ *and* $T_3^*, H_1^*, H_2^* \mid \tilde{\psi}_{Z_2v_2} \xrightarrow{d} \frac{\sigma_u^2}{\bar{\sigma}_u^2}\chi^2(1) \leq \chi^2(1)$ *a.s. when* $\pi_{02}\sigma_{v_2u} = 0$;

(b) $T_2^*, T_4^*, H_3^* \mid \tilde{\psi}_{Z_2v_2} \xrightarrow{d} \chi^2(1, \|\mu\|^2)$ *and* $T_3^*, H_1^*, H_2^* \mid \tilde{\psi}_{Z_2v_2} \xrightarrow{d} \frac{\sigma_u^2}{\bar{\sigma}_u^2}\chi^2(1, \|\mu\|^2) \leq \chi^2(1, \|\mu\|^2)$ *a.s. when* $\pi_{02}\sigma_{v_2u} \neq 0$.

Lemma A.8 *Suppose Assumption* **2.2** *is satisfied and let* $\pi_2 = \pi_{02}/\sqrt{n}$, *where* $\pi_{02} \in \mathbb{R}^{k_2}$ *is fixed. If for some* $\delta > 0$, $\mathbb{E}(\|Z_i\|^{4+\delta}, \|v_i\|^{2+\delta}) < \infty$, *then we have*: $\sup_{\tau \in \mathbb{R}} | \mathbb{P}^*(T_l^* \leq \tau) - \mathbb{P}(T_l \leq \tau) | = o_p(1)$

and $\sup_{\tau\in\mathbb{R}} \mid \mathbb{P}^*(DW_j^* \leq \tau) - \mathbb{P}(DW_j \leq \tau) \mid = o_p(1)$ *for any value of* $\sigma_{v_2 u}$.

A.2. Proofs

PROOF OF LEMMA **A.1** (a) Suppose first that H_0 is satisfied (i.e., $\sigma_{v_2 u} = 0$). We can express $\tilde{S}_{l,n}$ and $\hat{S}_{j,n}$ as:

$$
\begin{aligned}
\tilde{S}_{l,n} &=_{(1)} \sqrt{n}\frac{(y_2'M_{Z_1}y_2/n)^{-1}(y_2'M_{Z_1}y_1/n) - (y_2'(M_{Z_1}-M_Z)y_2/n)^{-1}(y_2'(M_{Z_1}-M_Z)y_1/n)}{\sqrt{\frac{y_1'M_{y_2}y_1}{n}[(\frac{y_2'(M_{Z_1}-M_Z)y_2}{n})^{-1} - (\frac{y_2'M_{Z_1}y_2}{n})^{-1}] - [(\frac{y_2'M_{Z_1}y_2}{n})^{-1}(\frac{y_2'M_{Z_1}y_1}{n}) - (\frac{y_2'(M_{Z_1}-M_Z)y_2}{n})^{-1}(\frac{y_2'(M_{Z_1}-M_Z)y_1}{n})]^2}} \\
&=_{(2)} \sqrt{n}\frac{(y_2'M_{Z_1}y_2/n)^{-1}(y_2'M_{Z_1}u/n) - (y_2'(M_{Z_1}-M_Z)y_2/n)^{-1}(y_2'(M_{Z_1}-M_Z)u/n)}{\sqrt{\frac{y_1'M_{y_2}y_1}{n}[(\frac{y_2'(M_{Z_1}-M_Z)y_2}{n})^{-1} - (\frac{y_2'M_{Z_1}y_2}{n})^{-1}] - [(\frac{y_2'M_{Z_1}y_2}{n})^{-1}(\frac{y_2'M_{Z_1}y_1}{n}) - (\frac{y_2'(M_{Z_1}-M_Z)y_2}{n})^{-1}(\frac{y_2'(M_{Z_1}-M_Z)y_1}{n})]^2}} \\
&=_{(3)} \sqrt{n}H(\bar{\mathscr{R}}_n) = \sqrt{n}\left[H(\bar{\mathscr{R}}_n) - H(\mu)\right] \qquad \text{(A.2)}
\end{aligned}
$$

$$
\begin{aligned}
\hat{S}_{j,n} &=_{(1)} \sqrt{n}\frac{(y_2'M_{Z_1}y_2/n)^{-1}(y_2'M_{Z_1}y_1/n) - (y_2'(M_{Z_1}-M_Z)y_2/n)^{-1}(y_2'(M_{Z_1}-M_Z)y_1/n)}{\sqrt{\frac{y_1'M_{y_2}y_1}{n}[(\frac{y_2'(M_{Z_1}-M_Z)y_2}{n})^{-1} - (\frac{y_2'M_{Z_1}y_2}{n})^{-1}]}} \\
&=_{(2)} \sqrt{n}\frac{(y_2'M_{Z_1}y_2/n)^{-1}(y_2'M_{Z_1}u/n) - (y_2'(M_{Z_1}-M_Z)y_2/n)^{-1}(y_2'(M_{Z_1}-M_Z)u/n)}{\sqrt{\frac{y_1'M_{y_2}y_1}{n}[(\frac{y_2'(M_{Z_1}-M_Z)y_2}{n})^{-1} - (\frac{y_2'M_{Z_1}y_2}{n})^{-1}]}} \\
&=_{(3)} \sqrt{n}\tilde{H}(\bar{\mathscr{R}}_n) = \sqrt{n}[\tilde{H}(\bar{\mathscr{R}}_n) - \tilde{H}(\mu)] \qquad \text{(A.3)}
\end{aligned}
$$

where $H(\tilde{y}_2\tilde{y}_2, \tilde{y}_2\tilde{y}_1, \ddot{y}_2\ddot{y}_2, \ddot{y}_2\ddot{y}_1, \dot{y}_1\dot{y}_1) = \frac{(\tilde{y}_2'\tilde{y}_2)^{-1}\tilde{y}_2'\tilde{y}_1 - (\ddot{y}_2'\ddot{y}_2)^{-1}\ddot{y}_2'\ddot{y}_1}{\sqrt{\dot{y}_1'\dot{y}_1[(\ddot{y}_2'\ddot{y}_2)^{-1} - (\tilde{y}_2'\tilde{y}_2)^{-1}] - [(\tilde{y}_2'\tilde{y}_2)^{-1}\tilde{y}_2'\tilde{y}_1 - (\ddot{y}_2'\ddot{y}_2)^{-1}\ddot{y}_2'\ddot{y}_1]}}$ and $\tilde{H}(\tilde{y}_2\tilde{y}_2, \tilde{y}_2\tilde{y}_1, \ddot{y}_2\ddot{y}_2, \ddot{y}_2\ddot{y}_1, \dot{y}_1\dot{y}_1) = \frac{(\tilde{y}_2'\tilde{y}_2)^{-1}\tilde{y}_2'\tilde{y}_1 - (\ddot{y}_2'\ddot{y}_2)^{-1}\ddot{y}_2'\ddot{y}_1}{\sqrt{\dot{y}_1'\dot{y}_1[(\ddot{y}_2'\ddot{y}_2)^{-1} - (\tilde{y}_2'\tilde{y}_2)^{-1}]}}$ are real-valued Borel measurable functions on $\mathbb{R}^m$ with derivatives of order $s \geq 3$ and lower, being continuous on the neighborhood of $\mu = \mathbb{E}(\mathscr{R}_n)$ when $\pi_2 \neq 0$ is fixed, $H(\mu) = 0$ and $\tilde{H}(\mu) = 0$ under H_0. Note that the derivatives of order $s \geq 3$ and lower of $H(\cdot)$ and $\tilde{H}(\cdot)$ with respect to $y_2'M_{Z_1}y_2/n$, $y_2'M_{Z_1}y_1/n$, $y_2'(M_{Z_1} - M_Z)y_1/n$, and $y_1'M_{y_2}y_1/n$ are well define for any value of π_2. However, their derivatives with respect to $y_2'(M_{Z_1} - M_Z)y_2/n$ is not well-defined when $\pi_2 = 0$ and does not even exist if $\pi_2 = \pi_{02}c_n$ for any sequence $c_n \downarrow 0$ [similar to Moreira et al. (2009, footnote 2)]. The results in Lemma **A.1** - (a) follow from (A.2)-(A.2) by using Bhattacharya and Ghosh (1978, Theorem 2).

(b) Suppose now that $\sigma_{v_2 u} \neq 0$ is fixed. Since $\pi_2 \neq 0$ is fixed and Assumption **2.2** holds, we can write $\tilde{S}_{l,n} = \sqrt{n}(n^{-1/2}\tilde{S}_{l,n} - \mu_{l,S}) + \sqrt{n}\mu_{l,S}$, where $\mu_{l,S} = p\lim_{n\to+\infty} n^{-1/2}\tilde{S}_{l,n} = -\frac{1}{\sigma_u}[\pi_2'(Q_{Z_2} - Q_{Z_1Z_2}'Q_{Z_1}^{-1}Q_{Z_1Z_2})\pi_2 + \sigma_{v_2}]^{-1}\sigma_{v_2 u} \neq 0$ and $\sqrt{n}(n^{-1/2}\tilde{S}_{l,n} - \mu_{l,S}) \xrightarrow{d} \psi_{l,S} \sim N\left\{0, [\pi_2'(Q_{Z_2} - Q_{Z_1Z_2}'Q_{Z_1}^{-1}Q_{Z_1Z_2})\pi_2]^{-1} - [\pi_2'(Q_{Z_2} - Q_{Z_1Z_2}'Q_{Z_1}^{-1}Q_{Z_1Z_2})\pi_2 + \sigma_{v_2}]^{-1}\right\}$, and similarly for $\hat{S}_{j,n}$. Therefore, for any $\tau \in \mathbb{R}$, we have $\lim_{n\to+\infty}\mathbb{P}(|\tilde{S}_{l,n}| \leq \tau) = \lim_{n\to+\infty}\mathbb{P}(|\sqrt{n}(n^{-1/2}\tilde{S}_{l,n} - \mu_{l,S}) + \sqrt{n}\mu_{l,S}| \leq \tau) \leq \lim_{n\to+\infty}\mathbb{P}(|\sqrt{n}(n^{-1/2}\tilde{S}_{l,n} - \mu_{l,S})| + \sqrt{n}|\mu_{l,S}| \leq \tau) \to \mathbb{P}(|\psi_{l,S}| + \infty \leq \tau) = 0$, and similarly we have $\lim_{n\to+\infty}\mathbb{P}(|\hat{S}_{j,n}| \leq \tau) \to 0$. It immediately follows that $\sup_{\tau\in\mathbb{R}} \mid \mathbb{P}(|\tilde{S}_{l,n}| \leq \tau) \mid\to 0$ and $\sup_{\tau\in\mathbb{R}} \mid \mathbb{P}(|\hat{S}_{j,n}| \leq \tau) \mid\to 0$, as stated. ■

PROOF OF LEMMA **A.2** (a) Suppose first that H_0 is satisfied. Since we have $T_l = n^{-1}\kappa_l \parallel \tilde{S}_{l,n} \parallel^2$ and $DW_j = \parallel \hat{S}_{j,n} \parallel^2$ from (A.1) for all l and j, it suffices to approximate $\mathbb{P}(T_l \leq \tau)$ and $\mathbb{P}(DW_j \leq \tau)$ uniformly in τ to complete the proof. First, we can write both $\mathbb{P}(T_l \leq \tau)$ and $\mathbb{P}(DW_j \leq \tau)$ as:

$$\mathbb{P}(T_l \leq \tau) = \mathbb{P}(\tilde{S}_{l,n} \in \mathscr{C}_\tau), \mathbb{P}(DW_j \leq \tau) = \mathbb{P}(\hat{S}_{j,n} \in \mathscr{C}_\tau), \tag{A.4}$$

where $\mathscr{C}_\tau = \{x \in \mathbb{R}; x^2 \leq \tau\}$ are convex sets. From Bhattacharya and Rao (1976, Corollary 3.2), we have $\sup_{\tau\in\mathbb{R}} \Phi((\partial\mathscr{C}_\tau)^\varepsilon) \leq d.\varepsilon$ for some constant d and $\varepsilon > 0$. So, Bhattacharya and Ghosh (1978, Theorem 1) holds with $B = \mathscr{C}_\tau$ and $W_n \in \{\tilde{S}_{l,n}, \hat{S}_{j,n}\}$. By using the approximation of $\mathbb{P}(\tilde{S}_{l,n} \leq \tau)$ and $\mathbb{P}(\hat{S}_{j,n} \leq \tau)$ in Lemma **A.1** - (a) and the definition of $\mathscr{C}_\tau$ in (A.4), Lemma **A.2** - (a) follows directly from the fact that the odd terms of the quadratic expansion are even [see also Horowitz (2001, Ch.52, eq.3.13) for the high-order approximation of pivotal symmetric statistics]. ■

PROOF OF LEMMA **A.3** The proof of (a) follows the same steps as Theorem 3 of Moreira et al. (2009) and is therefore omitted. The proof of (b) follows similar steps to those of Lemma **A.1** - (b), thus it is also omitted. ■

PROOF OF THEOREM **3.1** The proof of Theorem **3.1**-(a) is similar Hall and Horowitz (1996) by exploiting Lemmas **A.2**-**A.3**, hence is therefore omitted. Note that Lemmas **A.3**-(a) shows that the bootstrap estimates and the $(r+1)$-term empirical Edgeworth expansion in Lemma **A.2**-(a) for all statistics are asymptotically equivalent up to the $o(n^{-r})$ order under H_0. Theorem **3.1**-(b) holds mainly because the asymptotic distributions of all DWH statistics diverge under fixed endogeneity ($\sigma_{v_2u} \neq 0$ is fixed) and strong identification ($\pi \neq 0$); as showed Lemma **A.2**-(b). ■

PROOF OF LEMMA **A.4** Lemma **A.4** is a special case of Doko Tchatoka and Dufour (2011a) for $G = 1$, therefore the proof is omitted. ■

PROOF OF LEMMA **A.5** The proof of Lemma **A.5** is similar to those of Lemma A.1 in Moreira et al. (2009) and is therefore omitted. ■

PROOF OF LEMMA **A.6** The proof of Lemma **A.6** is similar to those of Lemma A.2 in Moreira et al. (2009) and is also omitted. ■

PROOF OF LEMMA **A.7** First, we can write the bootstrap DWH statistics as

$$T_l^* \;=\; n^{-1}\kappa_l \parallel \tilde{S}_{l,n}^* \parallel^2,\;\; DW_j^* = \parallel \hat{S}_{j,n}^* \parallel^2, \tag{A.5}$$

where $\tilde{S}_{l,n}^* = \sqrt{n}(\tilde{\beta}^* - \hat{\beta}^*)/\tilde{\omega}_l^*$ and $\hat{S}_{j,n}^* = \sqrt{n}(\tilde{\beta}^* - \hat{\beta}^*)/\hat{\omega}_j^*$, and $\tilde{\beta}^*, \hat{\beta}^*, \tilde{\omega}_l^*, \hat{\omega}_j^*$ are the bootstrap counterparts of $\tilde{\beta}, \hat{\beta}, \tilde{\omega}_l$, and $\hat{\omega}_j$, respectively. Now, observe that $\mathbb{E}^*(Z^{*'}Z^*/n) = Z'Z/n$, $\mathbb{E}^*(Z^{*'}u^*/n) = Z'\tilde{u}/n$, $\mathbb{E}^*(Z^{*'}v_2^*/n) = Z'\tilde{v}_2/n$, and $\mathbb{E}^*[(u^* : v_2^*)'(u^* : v_2^*)/n] = (\tilde{u} : \tilde{v}_2)'(\tilde{u} : \tilde{v}_2)/n$. So, conditional on $\mathscr{F}_n$, we have by the Markov law of large numbers: $Z^{*'}Z^*/n - Z'Z/n \to 0$, $Z^{*'}u^*/n - Z'\tilde{u}/n \to 0$, $Z^{*'}v_2^*/n - Z'\tilde{v}_2/n \to 0$, and $(u^* : v_2^*)'(u^* : v_2^*)/n - (\tilde{u} : \tilde{v}_2)'(\tilde{u} : \tilde{v}_2)/n \to 0$ $a.s.$ Since $Z'Z/n \xrightarrow{p} Q_Z$, $Z'\tilde{v}_2/n \xrightarrow{p} 0$, $Z'\tilde{u}/n \xrightarrow{p} 0$, and $(\tilde{u} : \tilde{v}_2)'(\tilde{u} : \tilde{v}_2)/n \xrightarrow{p} \Sigma$, hence we have $Z^{*'}Z^*/n \to Q_Z$, $Z^{*'}u^*/n \to 0$, $Z^{*'}v_2^*/n \to 0$, and $(u^* : v_2^*)'(u^* : v_2^*)/n \to \Sigma$ $a.s.$ under Assumption **2.2**. Now, if $\pi_2 = \pi_{02}/\sqrt{n}$ and Lemma **A.6** holds, we can show that (conditional on $\mathscr{F}_n$): $\tilde{S}_{2,n}^*, \tilde{S}_{4,n}^*, \hat{S}_{3,n}^* \xrightarrow{d} \tilde{\Psi}_\beta = \frac{1}{\sigma_u}\tilde{\Psi}_{Z_2v_2}^{-1/2}(\tilde{\psi}_{Z_2v_2} + \bar{Q}_{Z_2}\pi_{02})'\bar{Q}_{Z_2}^{-1}\tilde{\psi}_{Z_2u} - \frac{1}{\sigma_u}\tilde{\Psi}_{Z_2v_2}^{1/2}\sigma_{v_2}^{-2}\sigma_{v_2u}$ and $\tilde{S}_{3,n}^*, \hat{S}_{1,n}^*, \hat{S}_{2,n}^* \xrightarrow{d} (\sigma_u/\bar{\sigma}_u)\tilde{\Psi}_\beta$ $a.s.$, where $\tilde{\Psi}_{Z_2v_2} = (\tilde{\psi}_{Z_2v_2} + \bar{Q}_{Z_2}\pi_{02})'\bar{Q}_{Z_2}^{-1}(\tilde{\psi}_{Z_2v_2} + \bar{Q}_{Z_2}\pi_{02})$, $\tilde{\psi}_{Z_2v_2} = \psi_{Z_2v_2} - Q_{Z_2Z_1}Q_{Z_1}^{-1}\psi_{Z_1v_2} \sim N(0, \sigma_{v_2}^2\bar{Q}_{Z_2})$, $\tilde{\psi}_{Z_2u} = \psi_{Z_2u} - Q_{Z_2Z_1}Q_{Z_1}^{-1}\psi_{Z_1u} \sim N(0, \sigma_u^2\bar{Q}_{Z_2})$, $\bar{Q}_{Z_2} = Q_{Z_2} - Q_{Z_2Z_1}Q_{Z_1}^{-1}Q_{Z_2Z_1}'$, and $\bar{\sigma}_u^2 = \sigma_u^2 - 2\sigma_{v_2u}\tilde{\Psi}_{Z_2v_2}^{-1}(\tilde{\psi}_{Z_2v_2} + \bar{Q}_{Z_2}\pi_{02})'\bar{Q}_{Z_2}^{-1}\tilde{\psi}_{Z_2u} + \sigma_{v_2}\tilde{\psi}_{Z_2u}'\bar{Q}_{Z_2}^{-1}(\tilde{\psi}_{Z_2v_2} + \bar{Q}_{Z_2}\pi_{02})\tilde{\Psi}_{Z_2v_2}^{-2}(\tilde{\psi}_{Z_2v_2} + \bar{Q}_{Z_2}\pi_{02})'\bar{Q}_{Z_2}^{-1}\tilde{\psi}_{Z_2u}$. Moreover, we have $\tilde{\Psi}_\beta \mid \tilde{\psi}_{Z_2v_2} \xrightarrow{d} N(\mu, 1)$, where $\mu = \sigma_u^{-1}\tilde{\Psi}_{Z_2v_2}^{-1/2}\bar{Q}_{Z_2}^{-1}\pi_{02}\sigma_{v_2}^{-2}\sigma_{v_2u} = \sigma_u^{-1}\sigma_{v_2}^{-2}\tilde{\Psi}_{Z_2v_2}^{-1/2}\bar{Q}_{Z_2}^{-1}\pi_{02}\sigma_{v_2u}$. We will now distinguish the following two cases: (a) $\pi_{02}\sigma_{v_2u} = 0$ and (b) $\pi_{02}\sigma_{v_2u} \neq 0$.

(a) Suppose first that $\pi_{02}\sigma_{v_2u} = 0$. Then, we have $\mu = 0$ so that $\tilde{\Psi}_\beta \mid \tilde{\psi}_{Z_2v_2} \xrightarrow{d} N(0,1)$. As a result, we also have $T_2^*, T_4^*, H_3^* \mid \tilde{\psi}_{Z_2v_2} \xrightarrow{d} \chi^2(1)$ and $T_3^*, H_1^*, H_2^* \mid \tilde{\psi}_{Z_2v_2} \xrightarrow{d} \Psi_0 = \frac{\sigma_u^2}{\bar{\sigma}_u^2}\chi^2(1) \leq \chi^2(1)$ from (A.5). Hence, Lemma **A.7**-(a) follows immediately.

(b) Suppose now that $\pi_{02}\sigma_{v_2u} \neq 0$. Then, we have $\mu \neq 0$ so that $T_2^*, T_4^*, H_3^* \mid \tilde{\psi}_{Z_2v_2} \xrightarrow{d} \chi^2(1, \|\mu\|^2)$, $T_3^*, H_1^*, H_2^* \mid \tilde{\psi}_{Z_2v_2} \xrightarrow{d} \frac{\sigma_u^2}{\sigma_u^2}\chi^2(1, \|\mu\|^2) \leq \chi^2(1, \|\mu\|^2)$, and Lemma **A.7**-(b) follows. ■

PROOF OF LEMMA **A.8** Lemma **A.8** follows directly from Lemmas **A.4-A.7**. ■

PROOF OF THEOREM **3.2** From Lemma **A.8**, we have $\mathbb{P}(T_l > \tau) - \mathbb{P}^*(T_l^* > \tau) = o_p(1)$ uniformly over $\tau \in \mathbb{R}$. So, the results follow from Lemma **A.7** and an expected value argument. ■

References

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

Andrews, D. W. K., 2002. High-order improvement of a computationally attractive k-step bootstrap for extremum estimators. Econometrica 70(1), 119–162.

Andrews, D. W. K., Stock, J. H., 2007. Inference with weak instruments. In: R. Blundell, W. Newey , T. Pearson, eds, Advances in Economics and Econometrics, Theory and Applications, 9th Congress of the Econometric Society Vol. 3. Cambridge University Press, Cambridge, U.K., chapter 6.

Baum, C., Schaffer, M., Stillman, S., 2003. Instrumental variables and GMM: Estimation and testing. Stata Journal 3(1), 1–30.

Bhattacharya, R. N., Ghosh, J., 1978. On the validity of the formal Edgeworth expansion. The Annals of Statistics 6, 434–451.

Bhattacharya, R. N., Rao, R., 1976. Normal approximation and asymptotic expansions. In: R. Bhattacharya, R. Rao, eds, Normal Approximation and Asymptotic Expansions. Wiley Series in Probability and Mathematical Analysis, New York.

Bickel, P. J., Freedman, D. A., 1981. Some asymptotic theory for the bootstrap. The Annals of Statistics 9(6), 1196–1217.

Davidson, R., Mackinnon, J., 1993. Econometric Theory and Methods. Oxford University Press, New York, New York.

Doko Tchatoka, F., 2014. Subset hypotheses testing and instrument exclusion in the linear IV regression. Econometric Theory forthcoming.

Doko Tchatoka, F., Dufour, J.-M., 2011a. Exogeneity tests and estimation in IV regressions. Technical report, Department of Economics, McGill University Montréal, Canada.

Doko Tchatoka, F., Dufour, J.-M., 2011b. On the finite-sample theory of exogeneity tests with possibly non-Gaussian errors and weak identification. Technical report, Department of Economics, McGill University Montréal, Canada.

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

Dufour, J.-M., 1979. Methods for Specification Errors Analysis with Macroeconomic Applications PhD thesis University of Chicago. 257 + XIV pages.

Dufour, J.-M., 1987. Linear Wald methods for inference on covariances and weak exogeneity tests in structural equations. In: I. B. MacNeill, G. J. Umphrey, eds, Advances in the Statistical Sciences: Festschrift in Honour of Professor V.M. Joshi's 70th Birthday. Volume III, Time Series and Econometric Modelling. D. Reidel, Dordrecht, The Netherlands, pp. 317–338.

Dufour, J.-M., 1997. Some impossibility theorems in econometrics, with applications to structural and dynamic models. Econometrica 65, 1365–1389.

Dufour, J.-M., 2003. Identification, weak instruments and statistical inference in econometrics. Canadian Journal of Economics 36(4), 767–808.

Dufour, J.-M., 2006. Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics. Journal of Econometrics 138, 2649–2661.

Dufour, J.-M., Taamouti, M., 2007. Further results on projection-based inference in IV regressions with weak, collinear or missing instruments. Journal of Econometrics 139(1), 133–153.

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

Engle, R. F., 1982. A general approach to Lagrange multiplier diagnostics. Journal of Econometrics 20, 83–104.

Farebrother, R. W., 1976. A remark on the Wu test. Econometrica 44, 475–477.

Guggenberger, P., 2010. The impact of a Hausman pretest on the size of the hypothesis tests. Econometric Theory 156, 337–343.

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

Hall, P., Horowitz, J. L., 1996. Bootstrap critical values for tests based on generalized-method-of-moments estimators. Econometrica 64(4), 891–916.

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

Holly, A., 1982. A remark on Hausman's test. Econometrica 50, 749–759.

Horowitz, J. L., 1994. Bootstrap-based critical values for the IM test. Journal of Econometrics 61, 395–411.

Horowitz, J. L., 2001. The bootstrap. In: J. Heckman, E. E. Leamer, eds, Handbook of Econometrics. Elsvier Science, Amsterdam, The Netherlands.

Hwang, H.-S., 1980. Test of independence between a subset of stochastic regressors and disturbances. International Economic Review 21, 749–760.

Hwang, H.-S., 1985. The equivalence of Hausman and Lagrange multiplier tests of independence between disturbance and a subset of stochastic regressors. Economics Letters 17, 83–86.

Kariya, T., Hodoshima, H., 1980. Finite-sample properties of the tests for independence in structural systems and LRT. The Quarterly Journal of Economics 31, 45–56.

Kiviet, J. F., Niemczyk, J., 2007. The asymptotic and finite-sample distributions of OLS and simple IV in simultaneous equations. Computational Statistics and Data Analysis 51, 3296–3318.

Kiviet, J. F., Niemczyk, J., 2012. Comparing the asymptotic and empirical (un)conditional distributions of OLS and IV in a linear static simultaneous equation. Computational Statistics and Data Analysis 56, 3567–3586.

Kiviet, J. F., Pleus, M., 2012. The performance of tests on endogeneity of subsets of explanatory variables scanned by simulation. Technical report, Amsterdam School of Economics Amsterdam, The Netherlands.

Li, J., 2006. The block bootstrap test of Hausman's exogeneity in the presence of serial correlation. Economics Letters 91, 76–82.

Moreira, M. J., 2009. Tests with correct size when instruments can be abitrarily weak. Journal of Econometrics 152, 131–140.

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

Nakamura, A., Nakamura, M., 1981. On the relationships among several specification error tests presented by Durbin, Wu and Hausman. Econometrica 49, 1583–1588.

Revankar, N. S., 1978. Asymptotic relative efficiency analysis of certain tests in structural sysytems. International Economic Review 19, 165–179.

Revankar, N. S., Hartley, M. J., 1973. An independence test and conditional unbiased predictions in the context of simultaneous equation systems. International Economic Review 14, 625–631.

Reynolds, R. A., 1982. Posterior odds for the hypothesis of independence between stochastic regressors and disturbances. International Economic Review 23(2), 479–490.

Smith, R. J., 1983. On the classical nature of the Wu-Hausman statistics for independence of stochastic regressors and disturbance. Economics Letters 11, 357–364.

Smith, R. J., 1984. A note on likelihood ratio tests for the independence between a subset of stochastic regressors and disturbances. International Economic Review 25, 263–269.

Smith, R. J., 1985. Wald tests for the independence of stochastic variables and disturbance of a single linear stochastic simultaneous equation. Economics Letters 17, 87–90.

Smith, R. J., 1987. Testing for exogeneity in limited dependent variable models using a simplified likelihood ratio statistic. Journal of Applied Econometrics 2(3), 237–245.

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

Stock, J. H., Wright, J. H., Yogo, M., 2002. A survey of weak instruments and weak identification in generalized method of moments. Journal of Business and Economic Statistics 20(4), 518–529.

Taniguchi, M., 1988. Asymptotic expansions of the distriutions of some tests statistics for Gausssian ARMA processes. Journal of Multivariate Analysis 27, 494–511.

Wong, K., 1996. Bootstrapping Hausman's exogeneity test. Economics Letters 53, 139–143.

Wong, K., 1997. Effect on inference of pretesting the exogeneity of a regressor. Economics Letters 56, 267–271.

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.