# Specification Tests with Weak and Invalid Instruments

Doko Tchatoka, Firmin Sabro

School of Economics and Finance, University of Tasmania

20 July 2012

# Specification Tests with Weak and Invalid Instruments *

Firmin Doko Tchatoka[†]

School of Economics and Finance

University of Tasmania

June 26, 2012

* The author thanks Mardi Dunguey, Lynda Khalaf and Jean-Marie Dufour for several useful comments and suggestions.

† School of Economics and Finance, University of Tasmania, Private Bag 85, Hobart TAS 7001, Tel: +613 6226 7226, Fax:+61 3 6226 7587; e-mail : Firmin.dokotchatoka@utas.edu.au, homepage: http://www.fdokotchatoka.com/

ABSTRACT

We investigate the size of the Durbin-Wu-Hausman tests for exogeneity when instrumental variables violate the strict exogeneity assumption. We show that these tests are severely size distorted even for a small correlation between the structural error and instruments. We then propose a bootstrap procedure for correcting their size. The proposed bootstrap procedure does not require identification assumptions and is also valid even for moderate correlations between the structural error and instruments, so it can be described as robust to both weak and invalid instruments.

**Key words**: Exogeneity tests; weak instruments; instrument endogeneity; bootstrap technique.

JEL classification: C3; C12; C15; C52.

# 1. Introduction

Instrumental variables (IV) estimation can cure many ills where alternative least squares (OLS) methods yield biased and inconsistent estimators of model coefficients. This is especially the case when explanatory variables are correlated with the error term so that OLS estimators measure only the magnitude of association, rather than the magnitude and direction of causation which is needed for policy analysis. IV estimation of such models provides a way to obtain consistent parameter estimates, once the effect of common driving variables has been eliminated. Even though coefficients estimated in this way may have interesting interpretations from the viewpoint of economic theory, IV procedure usually requires the availability of exogenous instruments, at least as great as the number of coefficients to be estimated, whereas the validity of those initial instruments is not testable. This identifying indisputable exogeneity assumption has been questioned in many applied studies[1]. In most IV applications, the instruments often arrive with a dark cloud of invalidity hanging overhead and researchers usually do not know whether their correlations with the error are exactly zero.

Concerns have intensified in recent years about the effects of instrument endogeneity on standard inference procedures. Bound et al. (1995) provide evidence on how a slight violation of instrument exogeneity can cause severe bias in IV estimates, especially when identification is weak. Hausman and Hahn (2005) show that even in large-sample, the IV estimator can have a substantial bias even when the instruments are almost uncorrelated with the error. Recently, Chaudhuri and Rose (2009) investigate the estimation of the effect of military service on civilian earnings where both veteran status and schooling are suspected endogenous. However, Hausman (1978) test fails to reject joint exogeneity of both variables, though IV estimates seem larger than OLS estimates. Chaudhuri and Rose (2009) then advocate weak instrument issues as a potential cause of the non rejection of exogeneity. Indeed in this implication, Shea (1997)'s partial $R^2$ statistics [see also Hall, Rudebusch and Wilcox (1996)] are low and Stock and Yogo (2005) test suggests that the nominal size of 5 percent Wald test for the joint significance of veteran status and schooling is likely to be more than 25 percent. However, it may be that the non rejection of vet-

[1] See for example, Bound, Jaeger and Baker (1995) and Murray (2006)

eran status and schooling joint exogeneity is due to the fact that some of the *lottery and proximity-to-colleges* instruments used are invalid. Although the C-test and Sargan (1958) test for instrument orthogonality confirm that over-identification restrictions are satisfied, these tests also require the identifying assumption that at least two instruments are indisputably orthogonal to the error. Hence, there is still a reason for concern. Doko Tchatoka and Dufour (2008), and Guggenberger (2011) show that inference on structural parameters based on identification-robust statistics such as Anderson and Rubin (1949, AR-statistic), Kleibergen (2005, KLM-statistic), and Moreira (2003, CLR-statistic), as well as their generalized empirical likelihood versions, is unreliable from the viewpoint of size control when instruments violate "exogeneity" .

Murray (2006) [see also Samuel and Michael (2009)] suggests avoiding invalid instruments in IV procedures. However, since it is difficult to test the validity of all candidate instruments, it might seem that if we want to avoid invalid instruments, there is no hope in trying to use IV methods. Our ultimate gaol is to show how valid inference can still be conducted even if the strict exogeneity of the instruments is fundamentally wrong. Several authors have recently adopted this position and this paper attempts to make a progress in this direction. Imbens (2003) shows that bounds on average treatment effect in program evaluation can be recovered via a sensitivity analysis of the correlations between treatment and unobserved components of the outcomes. Ashley (2009) shows how the discrepancy between OLS and IV estimates can be used to estimate the degree of bias under any given assumption about the degree to which the exclusion restrictions are violated. Kiviet and Niemczyk (2006) show that the realizations of IV estimators based on strong but invalid instruments seem much closer to the true parameter values than those obtained from valid but weak instruments. Imbens et al. (2011) show that consistent point estimators can still be obtained in linear structural models with invalid instruments if the direct effects of these instruments are uncorrelated with the effects of the instruments on the endogenous regressors. The main findings of these studies suggest that valid inference can be conducted even when the instruments are invalid.

In this paper, we focus on linear structural models and we wish to develop valid tests for assessing exogeneity hypotheses allowing for a possibility that some instruments violate "strict exogeneity". To achieve this goal, we find useful to investigate first the sensitivity to instrument endogeneity of the standard exogeneity tests of the type proposed by Durbin (1954), Wu (1973), and Hausman (1978), henceforth DWH tests. Although DWH tests are widely used as pretests in applied work to decide whether OLS or IV method is appropriate, there is little evidence on how they behave when instruments violate "strict exogeneity".

Our theoretical analysis in this paper focuses on *local-to-zero invalid instruments*, i.e., situations where the parameter which controls instrument endogeneity approaches zero at rate $[n^{-\frac{1}{2}}]$ as the sample size $n$ increases. A similar specification has been considered by Hausman and Hahn (2005), Doko Tchatoka and Dufour (2008), Guggenberger (2011), and Berkowitz et al. (2008). However, our simulations extend to fixed endogeneity setup, i.e., cases where instrument endogeneity does not depends on the sample size.

First, we provide a characterization of the limiting distribution of DWH statistics under the null hypothesis of exogeneity and *instrument local-to-zero invalidity*. When model identification is strong, we find (non-degenerate) asymptotic noncentral chi-square distributions for all statistics, though both OLS and IV estimators are still asymptotically consistent. When identification is weak, we show that the statistics converge to a mixture of non-degenerate noncentral chi-square distributions. Therefore, the corresponding tests are seriously size distorted whether identification is strong or weak. However, our results indicate that they are more sensitive to instrument endogeneity when identification is strong than when it is weak.

Second, we propose size correction of the tests by resorting to bootstrap techniques. We present a Monte Carlo experiment indicating that the proposed bootstrap tests have an overall good performance even when instrument endogeneity is moderate, so they can be described as robust to instrument endogeneity. We apply our theoretical framework to the trade and growth model of Frankel and Romer (1999). Our results suggest that the instrument constructed on the basis of countries geographic characteristics in this model is not strictly exogenous. Consequently, standard DWH tests lead to conflicting results. However, all proposed bootstrap tests attribute the discrepancy between the OLS and 2SLS estimates to instrument invalidity.

The remainder of this paper is organized as follows. Section 2 formulates the model and present the statistics studied. Section 3 characterizes the asymptotic distributions of the statistics with locally invalid instruments and also explore their sensitivity to instrument endogeneity through a Monte Carlo experiment. The bootstrap test procedure is presented in Section 4, while Section 5 deals with the empirical application. Finally, Conclusions are drawn in Section 6 and proofs are presented in the Appendix.

## 2. Framework

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

$$y = X\beta + u, \tag{2.1}$$

$$X = Z\mathbf{\Pi} + V \tag{2.2}$$

where $y \in \mathbb{R}^n$ is a vector of observations on a dependent variable, $X \in \mathbb{R}^{n\times m}$ is a matrix of (possibly) endogenous explanatory variables, $Z \in \mathbb{R}^{n\times k}$ is a matrix of instruments ($k \geq m$), $u = (u_1, \ldots, u_n)' \in \mathbb{R}^n$ is a vector of structural disturbances, $V = [V_1, \ldots, V_n]' \in \mathbb{R}^{n\times m}$ is a matrix of reduced form disturbances, $\beta \in \mathbb{R}^m$ is an unknown structural parameter vector, while $\mathbf{\Pi} \in \mathbb{R}^{k\times m}$ is the unknown reduced-form coefficient matrix. Model (2.1)-(2.2) can be modified to include exogenous variables $Z_1$. If so, our analysis will not qualitatively alter by replacing the variables that are currently in (2.1)-(2.2) by the residuals that result from their projection onto the space spanned by the columns of $Z_1$.

Let $u_i$, $V_i$ and $Z_i$ denote the $i$-th row of $u$, $V$, and $Z$ respectively, written as column vectors (or scalars) and similarly for other random variables. We shall assume that $\{(u_i, V_i, Z_i) : 1 \leq i \leq n\}$ is drawn i.i.d across $i$ and the instrument matrix $Z$ has full column rank $k$ with probability one. Our main objective is to develop inference procedures for assessing the exogeneity of $X$ in (2.1)-(2.2), i.e. the hypothesis

$$\mathrm{H}_0 : \mathbf{cov}(X_i, u_i) = \sigma_{Xu} = 0, \tag{2.3}$$

taking into account that the instruments $Z$ may be invalid, at least locally. To achieve this goal, our approach is based essentially on DWH procedures. From that perspective, we find useful to study the sensitivity to instrument endogeneity of these tests first.

### 2.1. DWH statistics for exogeneity

We consider the following unified formulation of DWH statistics [Doko Tchatoka and Dufour (2011b, 2011a)]:

$$
\begin{aligned}
\mathcal{T}_l &= \kappa_l(\tilde{\beta}-\hat{\beta})'\tilde{\Sigma}_l^{-1}(\tilde{\beta}-\hat{\beta})\,, \quad l = 2,\, 3,\, 4, & (2.4)\\
\mathcal{H}_j &= n(\tilde{\beta}-\hat{\beta})'\hat{\Sigma}_j^{-1}(\tilde{\beta}-\hat{\beta})\,, \quad j = 1,\, 2,\, 3 & (2.5)
\end{aligned}
$$

where $\mathcal{T}_l,\, l = 2,\, 3,\, 4,$ are the statistics proposed by Wu (1973, 1974), $\mathcal{H}_j,\, j = 1,\, 2,\, 3,$ are three alternative Hausman (1978) type statistics, $\hat{\beta} = (X'X)^{-1}X'y$ and $\tilde{\beta} = (X'P_ZX)^{-1}X'P_Zy$ are the OLS and IV estimators respectively,

$$
\begin{aligned}
\tilde{\Sigma}_2 &= \tilde{\sigma}_2^2\hat{\Delta},\, \tilde{\Sigma}_3 = \tilde{\sigma}^2\hat{\Delta},\, \tilde{\Sigma}_4 = \hat{\sigma}^2\hat{\Delta},\\
\hat{\Sigma}_1 &= \tilde{\sigma}^2\hat{\Omega}_{IV}^{-1} - \hat{\sigma}^2\hat{\Omega}_{LS}^{-1},\, \hat{\Sigma}_2 = \tilde{\sigma}^2\hat{\Delta},\, \hat{\Sigma}_3 = \hat{\sigma}^2\hat{\Delta},\\
\hat{\Delta} &= \hat{\Omega}_{IV}^{-1} - \hat{\Omega}_{LS}^{-1},\, \hat{\Omega}_{IV} = X'P_ZX/n,\, \hat{\Omega}_{LS} = X'X/n,\\
\tilde{\sigma}^2 &= (y - X\tilde{\beta})'(y - X\tilde{\beta})/n,\, \hat{\sigma}^2 = (y - X\hat{\beta})'(y - X\hat{\beta})/n,\\
\tilde{\sigma}_2^2 &= \hat{\sigma}^2 - (\tilde{\beta}-\hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta}-\hat{\beta}),
\end{aligned}
$$

$\kappa_2 = (n-2m)/m,$ and $\kappa_3 = \kappa_4 = n - m.$ When the instruments are strictly exogenous, all statistics in (2.4)-(2.5) are pivotal under $\mathrm{H}_0$ even in finite-sample, whether identification is strong or weak [see Doko Tchatoka and Dufour (2011b)]. Further, their pivotality in finite-sample does not require the usual Gaussian assumption on model disturbances, i.e., the corresponding tests remain valid in finite-sample even for non-Gaussian errors. A crucial and relevant question is how they behave when the instruments do not satisfy the strictly exogeneity assumption.

Our theoretical analysis in this paper considers two main setups regarding the strength of the instruments: (**A**) $\mathbf{\Pi}$ is fixed with rank($\mathbf{\Pi}$) $= m$; and (**B**) $\mathbf{\Pi} = \mathbf{\Pi}_0/\sqrt{n}$, where $\mathbf{\Pi}_0 \in \mathbb{R}^{k\times m}$ is a constant matrix (possibly zero). The full rank condition of $\mathbf{\Pi}$ in (**A**) can however be weakened to allow for partial identification[2]. The setup for (**B**) is Staiger and Stock (1997) *local-to-zero weak instrument asymptotic.* In this setup, the parameter which

[2]Further details can be fond in Choi and Phillips (1992), Doko Tchatoka and Dufour (2011a), and Doko Tchatoka (2011).

controls instrument strength approaches zero at rate $[n^{-\frac{1}{2}}]$ as the sample size $n$ increases.

## 2.2. Notations and model assumptions

Throughout this paper, $I_q$ stands for the identity matrix of order $q$. For any full rank $n \times m$ matrix $A$, $P_A = A(A'A)^{-1}A$ is the projection matrix on the space spanned by $A$, $M_A = I_n - P_A$. The notation $\mathbf{vec}(A)$ is the $nm \times 1$ dimensional column vectorization of $A$ and $B > 0$ for a squared matrix $B$ means that $B$ is positive definite (p.d.). The convergence in probability is symbolized by "$\xrightarrow{p}$", "$\xrightarrow{d}$" stands for convergence in distribution while $O_p(.)$ and $o_p(.)$ denote the usual (stochastic) orders of magnitude. Finally, $\|\mathbf{U}\|$ denotes the Euclidian norm of a vector or matrix $\mathbf{U}$, *i.e.*, $\|\mathbf{U}\| = [tr(\mathbf{U}'\mathbf{U})]^{\frac{1}{2}}$.

Along with the basic model (2.1)-(2.2), we shall assume the following generic assumptions on model variables and parameters.

**Assumption 2.1** *The errors* $\{(u_i, V_i) : 1 \leq i \leq n\}$ *have zero mean and the same nonsingular covariance matrix* $\mathbf{\Sigma}$ *given by*

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_u^2 & \sigma'_{Vu} \\ \sigma_{Vu} & \Sigma_V \end{pmatrix} : (m+1) \times (m+1).$$

*Furthermore, we have* $\mathbf{E}(Z_i Z_i') = Q_Z > 0$, $\mathbf{E}[Z_i(u_i, V_i')] = (\sigma_{Zu}, 0)$ *for all* $i$, *where*

$$\sigma_{Zu} = \mathbf{d}/\sqrt{n} \tag{2.6}$$

*for some constant* $k \times 1$ *constant vector* $\mathbf{d}$ *in* $\mathbb{R}^k$.

First, Assumption **2.1** requires the errors $u$ and $V$ to be homoskedastic. It it can however be adapted to account for serial correlated errors. Second, while we impose $Z$ not be correlated with the reduced-form errors $V$, (2.6) implies that $Z$ and $u$ are correlated when $\mathbf{d} \neq 0$. However, their correlation approaches zero at rate $[n^{-\frac{1}{2}}]$ as the sample size increases. Nevertheless, moderate correlations are still allowed, depending on how large is $\mathbf{d}$ with respect to $\sqrt{n}$. A similar local-to-zero instrument invalidity has been used by Berkowitz et al. (2008) [see also Doko Tchatoka and Dufour (2008) and Guggenberger (2011)]. The analysis in Berkowitz et al. (2008) however assume that the support of $\mathbf{d}$ is compact with the same lower and upper bounds for all instruments. Assumption **2.1** does not require these restrictions.

**Assumption 2.2** *When the sample size* $n$ *converges to infinity, the following convergence results hold jointly*:

**(i)** $\frac{1}{n}\sum_{i=1}^{n}(u_i, V_i')'(u_i, V_i') \xrightarrow{p} \mathbf{\Sigma}$, $\frac{1}{n}\sum_{i=1}^{n} Z_i(u_i, V_i') \xrightarrow{p} 0$, $\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i' \xrightarrow{p} Q_Z$;

**(ii)** $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(Z_i u_i - \sigma_{Zu}, Z_i V_i', V_i u_i - \sigma_{Vu}\right) \xrightarrow{d} \psi = (\psi_z, \psi_{Vu})$, *where* $\psi_z = (\psi_{zu}, \psi_{zv})$, $\mathbf{vec}(\psi) \sim \mathbf{N}\left(0, \mathbf{\Omega}\right)$, $\mathbf{vec}(\psi_z) \sim \mathbf{N}\left(0, \mathbf{\Sigma} \otimes Q_Z\right)$ *and* $\psi_{Vu} \sim \mathbf{N}\left(0, \mathbf{\Sigma}_{Vu}\right)$ *with* $\mathbf{\Sigma}_{Vu} \equiv \sigma_u^2 \Sigma_V$ *when* $\sigma_{Vu} = 0$.

The Gaussian assumption on the limiting distributions in Assumption **2.2**-(**ii**) is implied by Assumptions **2.1** and the central limit theorem (CLT). Assumption **2.1**-(**ii**) along with Assumption **2.2**-(**ii**) then entail that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i u_i \xrightarrow{d} \psi_{zu} + \mathbf{d} \sim \mathbf{N}\left(\mathbf{d}\,,\, \sigma_u^2 Q_Z\right). \tag{2.7}$$

So, $Z'u/\sqrt{n}$ converges in distribution to a Gaussian with nonzero mean if $\mathbf{d} \neq 0$, though $Z$ is asymptotically uncorrelated with $u$ $\left[p\,lim_{n\to\infty}\left(\frac{Z'u}{n}\right) = 0\right]$.

We can now examine the asymptotic behavior of the statistics in (2.4)-(2.5).

# 3. Asymptotic behavior with locally invalid instruments

We now wish to analyze the sensitivity to instrument endogeneity of DWH statistics. Section 3.1 characterize their null limiting distribution when identification is strong and weak. Section 3.2 explores the size of the corresponding tests through a Monte Carlo experiment.

## 3.1. Asymptotic null distributions of the statistics

To characterize the null limiting distribution of the statistics, we find useful to examine first the sensitivity to instrument endogeneity of the vector of contrasts $\hat{\beta} - \tilde{\beta}$ that is used in their expressions. Lemmas **A.1-A.2** in the Appendix present the results for both strong and weak identification setups. The results indicate that when model identification is strong, the inconsistency of the vector of contrasts $\hat{\beta} - \tilde{\beta}$ is of order $n^{-1/2}$ under $\mathrm{H}_0$. Therefore,

$\sqrt{n}(\hat{\beta} - \tilde{\beta})$ converges to a Gaussian process with nonzero mean under $\mathrm{H}_0$ when $\mathbf{d} \neq 0$, despite the fact that $plim_{n\to\infty}(\hat{\beta} - \tilde{\beta}) = 0$. Moreover, when identification is weak, $\hat{\beta} - \tilde{\beta}$ converges under $\mathrm{H}_0$, to a mixture of Gaussian processes which have nonzero mean with probability 1 as long as $\mathbf{d} \neq 0$.

We can now prove the following two results on the asymptotic behavior of the statistics.

**Theorem 3.1** *Suppose* (2.1) - (2.2) *and Assumptions* **2.1**-**2.2** *are satisfied and let* $\sigma_{vu} = 0$. *If further rank*$(\mathbf{\Pi}) = m$, *then*

$$\begin{aligned} \mathcal{T}_2 &\overset{d}{\to} \frac{1}{m}\chi^2(m; \|\bar{\tau}_{\mathbf{d}}\|^2),\ \mathcal{T}_l \overset{d}{\to} \chi^2(m; \|\bar{\tau}_{\mathbf{d}}\|^2),\ l = 3,\, 4 \\ \mathcal{H}_j &\overset{L}{\to} \chi^2(m; \|\bar{\tau}_{\mathbf{d}}\|^2),\ j = 1,\, 2,\, 3 \end{aligned}$$

*where* $\bar{\tau}_{\mathbf{d}} = \frac{1}{\sigma_u}[(\mathbf{\Pi}' Q_Z \mathbf{\Pi})^{-1} - (\mathbf{\Pi}' Q_Z \mathbf{\Pi} + \Sigma_V)^{-1}]^{1/2}\mathbf{\Pi}'\mathbf{d}$ *is a* $k \times 1$ *constant vector.*

As seen, if $\mathbf{d} \neq 0$, all DWH statistics converge asymptotically to non-degenerate non-central $\chi^2$ distributions under $\mathrm{H}_0$. Hence, the corresponding tests are invalid (level is not controlled) even when instrument exogeneity $\mathbf{d}$ is small, though both OLS and 2SLS estimators are consistent in such situations when $\sigma_{vu} = 0$. The magnitude of the size distortions of the tests depends on how large the non centrality parameter $\|\bar{\tau}_{\mathbf{d}}\|^2$ is. So, $\mathbf{d}$ may be relatively small but the tests still exhibit large size distortions. Furthermore, as instrument endogeneity $\mathbf{d}$ is unknown and difficult to estimate consistently, it is practically infeasible to use the critical values of $\chi^2(m; \|\bar{\tau}_{\mathbf{d}}\|^2)$ for size correction.

Moreover, since $\mathbf{\Pi}[(\mathbf{\Pi}' Q_Z \mathbf{\Pi})^{-1} - (\mathbf{\Pi}' Q_Z \mathbf{\Pi} + \Sigma_V)^{-1}]\mathbf{\Pi}' > 0$, $\|\bar{\tau}_{\mathbf{d}}\|^2$ is an increasing function of $\|\mathbf{d}\|$, i.e., the size distortions of the tests increases with instrument endogeneity. In addition, it is easy to show that $\|\bar{\tau}_{\mathbf{d}}\|^2 \in [\lambda_p\|\mathbf{d}\|^2,\, \lambda_1\|\mathbf{d}\|^2]$, where $\lambda_p$ and $\lambda_1$ are the smallest and largest eigenvalues of $\sigma_u^{-2}\mathbf{\Pi}[(\mathbf{\Pi}' Q_Z \mathbf{\Pi})^{-1} - (\mathbf{\Pi}' Q_Z \mathbf{\Pi} + \Sigma_V)^{-1}]\mathbf{\Pi}'$, respectively. So, for a given value of instrument endogeneity $\mathbf{d}$, one can establish bounds on test size distortions through the estimates of $\lambda_p$ and $\lambda_1$.

We now derive the asymptotic null distributions of the statistics when instruments are weak.

**Theorem 3.2** *Suppose* (2.1) - (2.2) *and Assumptions* **2.1**-**2.2** *are satisfied and let* $\sigma_{vu} = 0$.

*If further* $\mathbf{\Pi} = \mathbf{\Pi}_0/\sqrt{n}$, *where* $\mathbf{\Pi}_0 \in \mathbb{R}^{k\times m}$ *is a constant matrix (possibly zero), then we have:*

$$
\begin{aligned}
\mathcal{T}_2 &\overset{d}{\to} \frac{1}{m}\int_{\mathbb{R}^{k\times m}} \chi^2(m; \|\bar{\nu}_{\mathbf{d}}\|^2) pdf(x_2)dx_2 \\
\mathcal{T}_3 &\overset{d}{\to} \int_{\mathbb{R}^{k\times m}} \frac{\chi^2(m;\, \|\bar{\nu}_{\mathbf{d}}\|^2)}{1+\chi^2(m;\, \|\bar{\nu}_{\mathbf{d}}\|^2)} pdf(x_2)dx_2 \leq \int_{\mathbb{R}^{k\times m}} \chi^2(m; \|\bar{\nu}_{\mathbf{d}}\|^2) pdf(x_2)dx_2 \\
\mathcal{T}_4 &\overset{d}{\to} \int_{\mathbb{R}^{k\times m}} \chi^2(m; \|\bar{\nu}_{\mathbf{d}}\|^2) pdf(x_2)dx_2 \\
\mathcal{H}_3 &\overset{L}{\to} \int_{\mathbb{R}^{k\times m}} \chi^2(m; \|\bar{\nu}_{\mathbf{d}}\|^2) pdf(x_2)dx_2 \\
\mathcal{H}_j &\overset{d}{\to} \int_{\mathbb{R}^{k\times m}} \frac{\chi^2(m;\, \|\bar{\nu}_{\mathbf{d}}\|^2)}{1+\chi^2(m;\, \|\bar{\nu}_{\mathbf{d}}\|^2)} pdf(x_2)dx_2 \leq \int_{\mathbb{R}^{k\times m}} \chi^2(m; \|\bar{\nu}_{\mathbf{d}}\|^2) pdf(x_2)dx_2 \; \textit{for } j = 1,\, 2
\end{aligned}
$$

*where* $pdf(.)$ *is the probability density function of* $\psi_{zv}$ *and* $\bar{\nu}_{\mathbf{d}} \equiv \bar{\nu}_{\mathbf{d}}(x_2) = \sigma_u^{-1}[(\mathbf{\Pi}_0 + Q_Z^{-1}x_2)'Q_Z(\mathbf{\Pi}_0 + Q_Z^{-1}x_2)]^{-1/2}(\mathbf{\Pi}_0 + Q_Z^{-1}x_2)'\mathbf{d}$.

As indicated the above results, the statistics $\mathcal{T}_2$, $\mathcal{T}_4$ and $\mathcal{H}_3$ converge now to a mixture of noncentral $\chi^2$ distributions under $\mathrm{H}_0$ while the asymptotic distributions of $\mathcal{T}_3$, $\mathcal{H}_1$, and $\mathcal{H}_2$ is bounded away by a mixture of noncentral $\chi^2$ distributions. Hence, $\mathcal{T}_3$, $\mathcal{H}_1$ and $\mathcal{H}_2$ are less size distorted than $\mathcal{T}_2$, $\mathcal{T}_4$ and $\mathcal{H}_3$. Again, we can show here that the size distortions of all tests increase with instrument endogeneity $\mathbf{d}$. An interesting question however is how big or small the distortions are when identification is weak compare to when it is strong. Although we do not provide a formal mathematical proof about any dominance between both setups, it is interesting to observe that test sensitivity to instrument endogeneity is likely more important when identification is strong than when it is weak.

To see why, consider the following two extreme situations: (**1**) the instruments completely irrelevant (i.e., $\mathbf{\Pi}_0 = 0$ in Theorem **3.2**), and (**2**) they are very strong (i.e, $\mathbf{\Pi}$ is large in Theorem **3.1**). For large or relatively moderate values of instrument endogeneity $\mathbf{d}$, case (**2**) implies large values of the non centrality parameter $\|\bar{\tau}_{\mathbf{d}}\|^2$ [see Theorem **3.1**] while in case (**1**), the non centrality parameter $\|\bar{\nu}_{\mathbf{d}}^0\|^2 = \sigma_u^{-2}\mathbf{d}'Q_Z^{-1}\psi_{zv}(\psi_{zv}'Q_Z^{-1}\psi_{zv})^{-1}\psi_{zv}'Q_Z^{-1}\mathbf{d}$ [see Theorem **3.2**] is bounded with probability one, independently of $\mathbf{\Pi}$. If the mean of $\|\bar{\nu}_{\mathbf{d}}^0\|^2$ is less than $\|\bar{\tau}_{\mathbf{d}}\|^2$ with probability one (which seems highly likely to be the case when $\mathbf{\Pi}$ is large), the tests will exhibit less size distortions in Theorem **3.2** than in Theorem **3.1**. By the continuity of $\|\bar{\nu}_{\mathbf{d}}\|^2$ with respect to $\mathbf{\Pi}_0$, the above argument extends to small values of $\mathbf{\Pi}_0$ (weakly identified model). Clearly, the fact that the tests $\mathcal{T}_3$, $\mathcal{H}_1$ and

$\mathcal{H}_2$ are conservative when identification is weak may be viewed in some extends as an asset for statistical inference when instruments are not strictly exogenous.

Section 3.2 illustrates our theoretical findings through a Monte Carlo experiment.

### 3.2. Simulation experiment

We consider a single endogenous variable model described by the following data generating process (DGP):

$$y = X\beta + u, \quad X = Z\mathbf{\Pi} + v \tag{3.1}$$

where $y$ and $X$ are $n \times 1$ vectors, the errors $(u,\, v)$ are such that:

$$u = (1+\rho_v^2)^{-1/2}(\varepsilon_1 + \rho_v v) \tag{3.2}$$

where $(\varepsilon_{1i},\, v_i)' \overset{i.i.d}{\sim} \mathbf{N}\,(0,\, I_2)$ for all $i$. $Z$ contains $k \in \{2,\, 5,\, 15\}$ instruments each generated i.i.d $\mathbf{N}(0,\, 1)$ and its correlation with the error $u$ is $\mathbf{r}_{zu} \in \{0, 0.05, 0.1, 0.2, 0.3\}$, where $\mathbf{r}_{zu} = 0$ characterizes exogenous instruments while $\mathbf{r}_{zu} \neq 0$ are for invalid instruments. In particular, small values of $\mathbf{r}_{zu}$ (for example $\mathbf{r}_{zu} = 0.05$) characterize locally invalid instruments. From (3.1)-(3.2), the exogeneity hypothesis of $X$ is then expressed as $\mathrm{H}_0$ : $\rho_v = 0$. We set the true values of $\beta$ at 2 and $\mathbf{\Pi} = \sqrt{\frac{\eta^2}{n\|Z\mathbf{\Pi}_0\|}}\mathbf{\Pi}_0$, where $\mathbf{\Pi}_0$ is a vector of ones, $\eta^2$ is the concentration parameter. Through this experiment, we vary $\eta^2$ in $\{0,\, 13,\, 1000\}$, where $\eta^2 \leq 613$ correspond to weak instruments and $\eta^2 > 613$ is for strong instruments [see Hansen et al. (2008) ]. The simulations are run with sample sizes $n = 100$; 300, $N = 10,000$ replications and nominal level of 5%.

Tables 1 presents the results. When $\mathbf{r}_{zu} = 0$, i.e., when the instruments are exogenous, all DWH tests have correct level whether identification is weak (columns $\eta = 0$; 13) or strong (columns $\eta = 1000$). More precisely, $\mathcal{T}_2$, $\mathcal{T}_4$ and $\mathcal{H}_3$ have correct size in all cases, but $\mathcal{T}_3$, $\mathcal{H}_1$ and $\mathcal{H}_2$ are overly conservative when identification is weak [see also Staiger and Stock (1997), Guggenberger (2010) and Doko Tchatoka and Dufour (2011b, 2011a)]. Now, when $\mathbf{r}_{zu} \neq 0$, the empirical rejections of all tests exceed the nominal size with a huge margin. As suggested our theoretical findings, the distortions increase with instrument endogeneity

$\mathbf{r}_{zu}$ and the maximal rejection can be as great as 100%, suggesting that the asymptotic size of the tests converges to 1 especially when identification is strong and endogeneity fixed ($\mathbf{r}_{zu} \neq 0$ does not dependent on the sample size). In addition, the effect of instrument endogeneity on the tests increases with model identification strength, i.e., the tests are in general more size distorted when identification is strong ($\eta^2 = 1000$) than when it is weak ($\eta = 0; 13$), thus confirming our prior intuition. As expected, $\mathcal{T}_3$, $\mathcal{H}_1$ and $\mathcal{H}_2$ are less sensitive to instrument endogeneity than $\mathcal{T}_2$, $\mathcal{T}_4$ and $\mathcal{H}_3$, especially when identification is weak (see columns $\eta = 0; 13$ in the table).

Overall, our main recommendation is to use DWH tests only when we have the certitude the instruments are strictly exogeneity. Since it is practically impossible to test the validity of all candidate instruments, it is importance to develop tests that accounts for the violation of instrument strict exogeneity.

Table 1. Size (in %) of DWH-tests at nominal level 5%; $n = 100$; 300

| $n = 100$ | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\mathbf{r}_{zu} = 0$ | | | $\mathbf{r}_{zu} = .05$ | | | $\mathbf{r}_{zu} = .1$ | | | $\mathbf{r}_{zu} = .2$ | | | $\mathbf{r}_{zu} = .3$ | | |
| Statistics | $k_2 \downarrow \eta^2 \rightarrow$ | 0 | 13 | 1000 | 0 | 13 | 1000 | 0 | 13 | 1000 | 0 | 13 | 1000 | 0 | 13 | 1000 |
| $\mathcal{T}_2$ | 2 | 5.2 | 4.8 | 4.9 | 8.3 | 9.4 | 7.6 | 17.3 | 21.4 | 16.1 | 47.7 | 59.8 | 47.1 | 67.4 | 81.0 | 80.9 |
| $\mathcal{T}_3$ | 2 | 0.0 | 0.2 | 4.5 | 0.1 | 0.5 | 7.1 | 0.2 | 1.8 | 15.4 | 1.6 | 8.8 | 45.7 | 4.0 | 19.8 | 80.0 |
| $\mathcal{T}_4$ | 2 | 5.2 | 4.8 | 4.9 | 8.2 | 9.2 | 7.6 | 17.1 | 21.3 | 16.1 | 47.5 | 59.7 | 46.9 | 67.3 | 80.9 | 80.7 |
| $\mathcal{H}_1$ | 2 | 0.0 | 0.1 | 4.2 | 0.1 | 0.4 | 6.7 | 0.2 | 1.5 | 14.7 | 1.5 | 8.1 | 44.3 | 3.6 | 18.6 | 79.2 |
| $\mathcal{H}_2$ | 2 | 0.0 | 0.2 | 4.7 | 0.1 | 0.5 | 7.3 | 0.2 | 1.9 | 15.6 | 1.7 | 9.2 | 46.1 | 4.2 | 20.3 | 80.2 |
| $\mathcal{H}_3$ | 2 | 5.2 | 4.9 | 5.0 | 8.4 | 9.4 | 7.7 | 17.4 | 21.5 | 16.2 | 47.9 | 59.9 | 47.3 | 67.5 | 81.1 | 81.0 |
| $\mathcal{T}_2$ | 5 | 5.0 | 4.9 | 4.7 | 8.9 | 12.8 | 11.2 | 18.9 | 33.3 | 29.6 | 41.7 | 70.2 | 80.4 | 54.2 | 80.9 | 98.7 |
| $\mathcal{T}_3$ | 5 | 0.4 | 0.8 | 4.6 | 0.8 | 3.0 | 10.7 | 2.1 | 11.4 | 28.7 | 9.4 | 39.6 | 79.9 | 15.8 | 52.7 | 98.6 |
| $\mathcal{T}_4$ | 5 | 4.9 | 4.8 | 4.7 | 8.8 | 12.7 | 11.1 | 18.8 | 33.2 | 29.4 | 41.6 | 70.1 | 80.3 | 54.1 | 80.9 | 98.7 |
| $\mathcal{H}_1$ | 5 | 0.3 | 0.7 | 4.2 | 0.7 | 2.6 | 10.2 | 1.8 | 10.6 | 27.4 | 8.8 | 37.9 | 79.1 | 14.6 | 51.1 | 98.5 |
| $\mathcal{H}_2$ | 5 | 0.4 | 0.8 | 4.6 | 0.8 | 3.1 | 10.9 | 2.2 | 11.6 | 29.0 | 9.7 | 40.0 | 80.1 | 16.2 | 53.3 | 98.6 |
| $\mathcal{H}_3$ | 5 | 5.1 | 4.9 | 4.8 | 8.9 | 13.0 | 11.3 | 19.1 | 33.5 | 29.8 | 41.8 | 70.3 | 80.6 | 54.3 | 81.0 | 98.7 |
| $\mathcal{T}_2$ | 15 | 5.2 | 5.1 | 5.2 | 7.8 | 15.3 | 12.7 | 15.1 | 40.0 | 36.0 | 27.7 | 70.3 | 86.0 | 35.0 | 80.0 | 99.1 |
| $\mathcal{T}_3$ | 15 | 2.5 | 3.0 | 5.1 | 4.1 | 10.8 | 12.5 | 9.0 | 32.8 | 35.4 | 19.9 | 65.0 | 85.6 | 27.5 | 75.6 | 99.1 |
| $\mathcal{T}_2$ | 15 | 5.1 | 5.1 | 5.2 | 7.8 | 15.3 | 12.6 | 15.0 | 39.9 | 35.8 | 27.5 | 70.2 | 85.9 | 34.8 | 79.9 | 99.1 |
| $\mathcal{H}_1$ | 15 | 2.2 | 2.6 | 4.8 | 3.7 | 9.9 | 11.9 | 8.3 | 31.6 | 34.3 | 19.0 | 63.9 | 85.0 | 26.4 | 74.8 | 99.0 |
| $\mathcal{H}_2$ | 15 | 2.5 | 3.1 | 5.2 | 4.3 | 11.0 | 12.7 | 9.2 | 33.1 | 35.9 | 20.2 | 65.4 | 86.0 | 27.9 | 75.9 | 99.1 |
| $\mathcal{H}_3$ | 15 | 5.3 | 5.2 | 5.3 | 8.0 | 15.4 | 12.9 | 15.2 | 40.2 | 36.2 | 27.8 | 70.4 | 86.2 | 35.1 | 80.1 | 99.1 |

| $n = 300$ | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\mathbf{r}_{zu} = 0$ | | | $\mathbf{r}_{zu} = .05$ | | | $\mathbf{r}_{zu} = .1$ | | | $\mathbf{r}_{zu} = .2$ | | | $\mathbf{r}_{zu} = .3$ | | |
| Statistics | $k_2 \downarrow \eta^2 \rightarrow$ | 0 | 13 | 1000 | 0 | 13 | 1000 | 0 | 13 | 1000 | 0 | 13 | 1000 | 0 | 13 | 1000 |
| $\mathcal{T}_2$ | 2 | 5.25 | 5.15 | 5.28 | 12.77 | 15.38 | 18.56 | 36.79 | 44.82 | 55.9 | 70.59 | 80.32 | 98.69 | 80.48 | 88.36 | 100 |
| $\mathcal{T}_3$ | 2 | 0.04 | 0.2 | 4.68 | 0.14 | 0.67 | 17.19 | 0.99 | 3.86 | 53.86 | 4.49 | 15.84 | 98.55 | 7.38 | 21.85 | 100 |
| $\mathcal{T}_4$ | 2 | 5.21 | 5.13 | 5.27 | 12.74 | 15.36 | 18.53 | 36.75 | 44.78 | 55.82 | 70.58 | 80.31 | 98.69 | 80.48 | 88.34 | 100 |
| $\mathcal{H}_1$ | 2 | 0.04 | 0.18 | 4.59 | 0.14 | 0.62 | 16.87 | 0.95 | 3.67 | 53.46 | 4.35 | 15.43 | 98.52 | 7.18 | 21.45 | 100 |
| $\mathcal{H}_2$ | 2 | 0.04 | 0.2 | 4.75 | 0.14 | 0.68 | 17.31 | 1.02 | 3.95 | 54.02 | 4.54 | 15.98 | 98.57 | 7.43 | 21.94 | 100 |
| $\mathcal{H}_3$ | 2 | 5.27 | 5.22 | 5.3 | 12.82 | 15.47 | 18.6 | 36.87 | 44.92 | 55.95 | 70.65 | 80.36 | 98.69 | 80.51 | 88.36 | 100 |
| $\mathcal{T}_2$ | 5 | 4.85 | 4.76 | 5.07 | 14.53 | 21.84 | 38.34 | 36.19 | 55.86 | 91.33 | 62.13 | 79.78 | 100 | 71.24 | 85.7 | 100 |
| $\mathcal{T}_3$ | 5 | 0.31 | 0.55 | 4.75 | 1.49 | 5.06 | 36.97 | 6.89 | 22.97 | 90.83 | 20.97 | 47.3 | 100 | 28.1 | 54.41 | 100 |
| $\mathcal{T}_4$ | 5 | 4.8 | 4.73 | 5.07 | 14.52 | 21.81 | 38.28 | 36.16 | 55.84 | 91.3 | 62.12 | 79.76 | 100 | 71.22 | 85.69 | 100 |
| $\mathcal{H}_1$ | 5 | 0.28 | 0.52 | 4.6 | 1.41 | 4.87 | 36.58 | 6.63 | 22.64 | 90.66 | 20.5 | 46.68 | 100 | 27.56 | 54 | 100 |
| $\mathcal{H}_2$ | 5 | 0.32 | 0.57 | 4.77 | 1.49 | 5.1 | 37.12 | 6.97 | 23.22 | 90.88 | 21.09 | 47.44 | 100 | 28.27 | 54.55 | 100 |
| $\mathcal{H}_3$ | 5 | 4.91 | 4.76 | 5.08 | 14.56 | 21.89 | 38.41 | 36.21 | 55.92 | 91.36 | 62.17 | 79.82 | 100 | 71.26 | 85.71 | 100 |
| $\mathcal{T}_2$ | 15 | 5.28 | 5 | 5.14 | 14.53 | 31.22 | 73.06 | 32.43 | 64.89 | 99.9 | 52.38 | 82.7 | 100 | 59.22 | 86.68 | 100 |
| $\mathcal{T}_3$ | 15 | 2.37 | 2.57 | 4.94 | 8.56 | 23.67 | 72.54 | 23.52 | 57.84 | 99.89 | 44.88 | 79.03 | 100 | 52.05 | 83.63 | 100 |
| $\mathcal{T}_4$ | 15 | 5.25 | 5 | 5.14 | 14.52 | 31.17 | 72.99 | 32.4 | 64.85 | 99.9 | 52.36 | 82.69 | 100 | 59.18 | 86.67 | 100 |
| $\mathcal{H}_1$ | 15 | 2.32 | 2.5 | 4.79 | 8.38 | 23.17 | 72.34 | 23.11 | 57.48 | 99.89 | 44.56 | 78.76 | 100 | 51.73 | 83.54 | 100 |
| $\mathcal{H}_2$ | 15 | 2.39 | 2.59 | 5 | 8.59 | 23.77 | 72.6 | 23.61 | 57.99 | 99.89 | 44.94 | 79.15 | 100 | 52.1 | 83.72 | 100 |
| $\mathcal{H}_3$ | 15 | 5.29 | 5.02 | 5.16 | 14.59 | 31.28 | 73.15 | 32.49 | 64.89 | 99.91 | 52.42 | 82.71 | 100 | 59.26 | 86.7 | 100 |

## 4. Bootstrap exogeneity tests with invalid instruments

We now wish to propose a bootstrap procedure to deal with the invalidity of the standard DWH tests when instruments are strictly exogenous. In addition to the basic assumption (2.1)-(2.2), we shall make the following assumption on model errors $(u,\, V)$ and instruments:

**Assumption 4.1** *$u = Zb + e$, where $b$ is a $k \times 1$ vector of unknown fixed coefficients, $(e,\, V)$ are independent of $Z$ and*

$$(e_i,\, V_i')' \,|\, Z \sim \mathbf{J}\bar{U}_i \quad \textit{for all} \quad i = 1,\, \dots,\, n \tag{4.1}$$

*where $\mathbf{J} : (m+1) \times (m+1)$ is a lower triangular positive matrix of unknown coefficient, $\bar{U}_i : (m+1) \times 1$ are i.i.d. across $i$ with zero mean and a completely specified distribution.*

Assumption **4.1** entails that $(e_i,\, V_i')'$ is homoskedastic with a distribution completely specified up to an unknown factor. It is particularly satisfied in the standard Gaussian assumption context where $(e_i,\, V_i')' \sim \mathbf{IDN}[0,\, \boldsymbol{\Sigma}_*]$ with $\boldsymbol{\Sigma}_* = \mathbf{JJ'} = \begin{pmatrix} \sigma_e^2 & \sigma_{Ve}' \\ \sigma_{Ve} & \Sigma_V \end{pmatrix}$, in which case we have

$$\bar{U}_i \sim \mathbf{IDN}[0,\, \mathbf{I_{m+1}}] \quad \text{for all} \quad i = 1,\, \dots,\, n. \tag{4.2}$$

However, non-Gaussian distributions are covered, including heavy tailed distributions which may lack moments (such as Cauchy distributions). This assumption can be weakened in large-sample [see for example, Assumption **2.2**].

Since $\mathbf{E}(Z_i Z_i') = Q_Z > 0$ for all $i = 1,\, \dots, n$, Assumption **4.1** entails that $cov(Z_i, u_i) = \sigma_{Zu} = Q_Z b$. So, $\sigma_{Zu} = 0$ if and only if $b = 0$, i.e., $Z$ is invalid as long as $b \neq 0$. In the above setup, the local-to-zero endogeneity of $Z$ is expressed as $b = \boldsymbol{\rho}/\sqrt{n}$, where $\boldsymbol{\rho}$ is a $k \times 1$ constant vector. It is therefore straightforward to see that $\mathbf{E}(Z'u/\sqrt{n}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{E}(Z_i Z_i') b = Q_Z \boldsymbol{\rho} \equiv \mathbf{d}$, where $\mathbf{d}$ is the constant vector of Assumptions **2.1**-(2.6).

If the errors $u = (u_1,\, u_2,\, \dots,\, u_n)'$ were observed, we could estimate consistently the instrument endogeneity $b$ from Assumption **4.1**. The difficulty however is that $u$ is not observed. To alleviate this, we adopt the following methodology.

Consider the following regression

$$\hat{u} \quad = \quad Z_* b + e_* \tag{4.3}$$

obtained by post-multiplying the first equation in Assumption **4.1** by $M_X = I_n - X(X'X)^{-1}X'$, so that all $\hat{u}$, $Z_*$ and $e_*$ are residuals from projection onto the space spanned by the columns of $X$[3]. Then the dependent variable $\hat{u} = M_X y$ in (4.3) is also the vector of residuals from the OLS regression of the structural equation (2.1). So, if $H_0$ and Assumptions **2.1-4.1** are satisfied, we have $Z_*' e_*/n = Z' M_X e/n = Z'e/n - (Z'X/n)(X'X/n)^{-1}(X'e/n) \xrightarrow{p} 0$, i.e., $Z_*$ and $e_*$ are asymptotically uncorrelated. Let $\hat{b} = (Z_*' Z_*)^{-1} Z_*' \hat{u}$ denotes the OLS estimator of $b$ from (4.3). It is easy to see that $\hat{b} - b = Z_*' Z_*/n)^{-1}(Z_*' e_*/n) \xrightarrow{p} 0$ under the same above conditions, i.e., $\hat{b}$ is a consistent estimator of $b$. Note that this consistency requires the exogeneity of $X$, at least asymptotically.

Moreover, let $\hat{e}_*$ denotes the vector of residuals from the OLS estimation of (4.3). We can write $\hat{u}$ as: $\hat{u} = Z_* \hat{b} + \hat{e}_*$, so that

$$Z_*' \hat{u} \quad = \quad Z_*' Z_* \hat{b}, \quad \hat{V}' \hat{u} = \hat{V}' Z_* \hat{b} \tag{4.4}$$

where $\hat{V} = M_Z X$ is the matrix of residuals from the first-stage OLS estimation, satisfying $\hat{V}' \hat{e}_* = 0$. This suggests that the structural error-instruments and structural-reduced form errors covariances can be estimated by

$$\hat{\sigma}_{Zu} \quad = \quad \frac{Z_*' \hat{u}}{n-k} = \frac{n}{n-k} \hat{Q}_* \hat{b}, \quad \hat{\sigma}_{Vu} = \frac{n}{n-k} \tilde{Q}_* \hat{b} \tag{4.5}$$

where $\hat{Q}_* = (Z_*' Z_*/n)$ and $\tilde{Q}_* = (\hat{V}' Z_*/n)$.

We are now ready to present the implementation of the bootstrap tests.

[3]It is important to note that if the structural equation (2.1) contains exogenous instruments $Z_1$, the first equation in Assumption **4.1** must be projected on the space orthogonal to $[X, Z_1]$.

### 4.1. Bootstrap tests implementation

To easy the exposition, we shall consider the following linear structural model $\mathbb{M}$ with Gaussian errors:

$$y_i = X_i'\beta + u_i,\; X_i = \mathbf{\Pi}'Z_i + V_i, \tag{4.6}$$

$$u_i = Z_i'b + e_i,\; (e_i,\, V_i')' \sim \mathbf{IDN}[0,\, \mathbf{\Sigma}_*] \quad \text{for all} \quad i = 1,\, \dots,\, n \tag{4.7}$$

where $\mathbf{\Sigma}_* = \begin{pmatrix} \sigma_e^2 & \sigma_{Ve}' \\ \sigma_{Ve} & \Sigma_V \end{pmatrix}$. The null hypothesis of interest is $\mathrm{H}_0 : \sigma_{Ve} = 0$. Thus the model $\mathbb{M}_0$ that represents the null hypothesis is (4.6)-(4.7) where $\sigma_{Ve}$ is replace by 0 in the expression of $\mathbf{\Sigma}_*$, i.e,

$$y_i = X_i'\beta + u_i,\; X_i = \mathbf{\Pi}'Z_i + V_i, \tag{4.8}$$

$$u_i = Z_i'b + e_i,\; (e_i,\, V_i')' \sim \mathbf{IDN}[0,\, \mathbf{diag}(\sigma_e^2,\, \Sigma_V)] \quad \text{for all} \quad i = 1,\, \dots,\, n. \tag{4.9}$$

The first step in constructing the bootstrap tests is to estimate each equation in (4.8)-(4.9) by OLS from observed data, yielding the restricted estimates $\hat{\beta}$, $\hat{\mathbf{\Pi}}$, $\hat{\sigma}_e^2 = (\hat{u} - Z_*\hat{b})'(\hat{u} - Z_*\hat{b})/(n-k)$, $\hat{\Sigma}_V = (X - Z\hat{\mathbf{\Pi}})'(X - Z\hat{\mathbf{\Pi}})/(n-k)$, $\hat{\sigma}^2 = \hat{u}'\hat{u}/(n-m)$ and $\hat{b} = (Z_*'Z_*)^{-1}Z_*'\hat{u}$, where $Z_* = M_X Z$ and $\hat{u} = y - X\hat{\beta}$. Then the bootstrap DGP $\hat{\mu}$ is given by

$$y_i^* = X_i^{*\prime}\hat{\beta} + u_i^*,\; X_i^* = \hat{\mathbf{\Pi}}'Z_i + V_i^*, \tag{4.10}$$

$$u_i^* = Z_i'\hat{b} + e_i^*,\; (e_i^*,\, V_i^{*\prime})' \sim \mathbf{IDN}[0,\, \mathbf{diag}(\hat{\sigma}_e^2,\, \hat{\Sigma}_V)] \quad \text{for all} \quad i = 1,\, \dots,\, n. \tag{4.11}$$

To implement the bootstrap tests, we need to compute the realization of DWH statistics from observed data. For this purpose, we also need to compute the IV estimates $\tilde{\beta}$ and $\tilde{\sigma}^2$ of $\beta$ and $\sigma_u^2$ from observed data. Let $\hat{\mathcal{W}}$ denotes one of the statistics $\mathcal{H}_i$ and $\mathcal{T}_l$ in (2.4)-(2.5), computed from the observed data.

The bootstrap tests can be implemented following the above four steps:

1. Compute the realizations of all data-dependent things needed to setup the bootstrap DGP $\hat{\mu}$ and the test statistic $\hat{\mathcal{W}}$ from the original sample.

2. Generate $B$ bootstrap samples using $\hat{\mu}$, and for each one compute a realization of the

bootstrap statistic, $\mathcal{W}_r^*$, $r = 1, \ldots, B$.

3. Compute the simulated bootstrap $p$-value as the proportion of bootstrap statistics $\mathcal{W}_r^*$ that are more extreme that $\hat{\mathcal{W}}$, i.e.

$$\hat{p}_{bs} \quad = \quad \frac{1}{B}\sum_{r=1}^{B} \mathbb{1}[\mathcal{W}_r^* > \hat{\mathcal{W}}] \tag{4.12}$$

where $\mathbb{1}[C] = 1$ if condition $C$ holds and 0 otherwise.

4. The bootstrap test rejects the null hypothesis of exogeneity at level $\alpha$ if $\hat{p}_{bs} < \alpha$.

As seen, the above algorithm can be generalized easily to non-Gaussian errors satisfying Assumption **4.1**. The algorithm can also be improved in some cases using sophisticated bootstrap techniques.

## 4.2. Performance of the bootstrap tests

We consider again model $\mathbb{M}$ in (4.6)-(4.7) with one endogenous variable ($m = 1$) and $k$ instruments $Z$ generated as $Z_{ji} \overset{\text{i.i.d}}{\sim} \mathbf{N}(0,\, 1)$, independent of the errors $(e_i,\, V_i')'$ for all $i$ and $j$. We generate the errors as $(e_i,\, V_i')' \sim \mathbf{IDN}[0,\, \mathbf{\Sigma}_*]$ with $\mathbf{\Sigma}_* = \begin{pmatrix} 1 & \sigma_{Ve}' \\ \sigma_{Ve} & 1 \end{pmatrix}$ and we test the null hypothesis $\sigma_{Ve} = 0$. The other model parameters true values are set at $\beta = 2$, $\mathbf{\Pi} = \sqrt{\frac{\eta^2}{n\|Z\mathbf{\Pi}_0\|}}\mathbf{\Pi}_0$, where $\mathbf{\Pi}_0$ is a vector of ones and $\eta^2 \in \{0,\, 13,\, 1000\}$. The correlation between $Z_j$ and $u$ is kept constant at $\mathbf{r}_{zu} \in \{0, 0.1, 0.3, 0.4, 0.5\}$ . The simulations are run with sample sizes $n = 50$; 100; 300, and the nominal level is set at 5%. Finally, the number of instruments $k$ varies in $\{5,\, 15\}$ .

We evaluate the empirical rejection frequencies of the bootstrap tests using the *the fast approximation* recently proposed by Davidson and Mackinnon (2007). We use the true DGP $\mu$ to generate $M = 100{,}000$ samples of size $n$. For each replication $r = 1, \ldots, M$, a replication $\mathcal{W}_r$ of the statistic $\mathcal{W}$ is computed from the simulated sample, along with a realization $\hat{\mu}_r$ of the bootstrap DGP. In addition, $\hat{\mu}_r$ is used to draw a *single* bootstrap statistic $\mathcal{W}_r^*$. We estimate the rejection probability (RP) as the proportion of the $\mathcal{W}_r$ that are less than $\hat{Q}^*(\alpha)$, the $1-\alpha$ quantile of the $\mathcal{W}_r^*$. This yields the following estimate of the

RP of the bootstrap test:

$$\hat{RP}_{bs} \;=\; \frac{1}{M}\sum_{r=1}^{M}\mathbb{1}[\mathcal{W}_r < \hat{Q}^*(\alpha)]. \tag{4.13}$$

Table 2 reports the results. As before, $\mathbf{r}_{zu} = 0$ corresponds to exogenous instruments while $\mathbf{r}_{zu} \neq 0$ characterize invalid instruments. Also, $\eta^2 \leq 613$ is for weak identification and $\eta^2 > 613$ characterize strong identification. We see that all tests have an overall correct level whether identification is strong or weak. In addition, level is controlled even for a relatively small sample ($n = 50$) and moderate instrument endogeneity ($\mathbf{r}_{zu} = 0.5$). The maximal size distortion of the bootstrap tests is around 6.7% and is obtained when $n = 50$ and $k = 5$, further with exogenous instruments ($\mathbf{r}_{zu} = 0$). This is a substantial improvement compared to standard DWH tests whose maximal rejection was as great as 100% [see Table 1]. Furthermore, it is interesting to note that when model parameters are completely non identified ($\eta^2 = 0$) or close so, all bootstrap tests have correct level irrespective of the size of instrument endogeneity. However, the tests tend to be conservative when instruments endogeneity is large and identification strong. This is mainly justified by the fact that the effect of instrument endogeneity on the tests increases with identification strength (see Theorems **3.1**-**3.1** and simulations in Table 1). When identification is strong and instrument endogeneity is large, the quantiles of the null distribution of the statistics tend to explode, as discussed in Theorem **3.2**. However, the non centrality parameter is always bounded with probability when identification is weak, even if instrument endogeneity is large (Theorem **3.2**). Therefore, the occurrence of the events $[\mathcal{W}_r < \hat{Q}^*(\alpha)]$ is higher when identification is weak than when it is strong. This is particularly why the tests have correct level in all cases when $\eta^2 = 0$ (irrelevant instruments).

Table 2. Rejection frequencies (in %) of bootstrap DWH-tests at nominal level 5%; $n = 50$; 100; 300

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n = 50$ | | | | | | | | | | | | | | | | |
| | | $\mathbf{r}_{zu} = 0$ | | | $\mathbf{r}_{zu} = .1$ | | | $\mathbf{r}_{zu} = .3$ | | | $\mathbf{r}_{zu} = .4$ | | | $\mathbf{r}_{zu} = .5$ | | |
| Statistics | $k_2 \downarrow \eta^2 \rightarrow$ | 0 | 13 | 1000 | 0 | 13 | 1000 | 0 | 13 | 1000 | 0 | 13 | 1000 | 0 | 13 | 1000 |
| $\mathcal{T}_{2bs}$ | 5 | 6.59 | 6.65 | 6.33 | 5.90 | 4.72 | 6.04 | 5.20 | 4.65 | 4.80 | 4.86 | 4.73 | 3.64 | 5.16 | 4.83 | 3.09 |
| $\mathcal{T}_{3bs}$ | 5 | 6.59 | 6.66 | 6.33 | 5.90 | 4.72 | 6.04 | 5.20 | 4.66 | 4.80 | 4.87 | 4.74 | 3.64 | 5.17 | 4.84 | 3.09 |
| $\mathcal{T}_{4bs}$ | 5 | 6.59 | 6.65 | 6.33 | 5.90 | 4.72 | 6.04 | 5.20 | 4.65 | 4.80 | 4.86 | 4.73 | 3.64 | 5.16 | 4.83 | 3.09 |
| $\mathcal{H}_{1bs}$ | 5 | 6.59 | 6.66 | 6.33 | 5.90 | 4.72 | 6.04 | 5.20 | 4.66 | 4.80 | 4.87 | 4.74 | 3.64 | 5.17 | 4.84 | 3.09 |
| $\mathcal{H}_{2bs}$ | 5 | 6.59 | 6.66 | 6.33 | 5.90 | 4.72 | 6.04 | 5.20 | 4.66 | 4.80 | 4.87 | 4.74 | 3.64 | 5.17 | 4.84 | 3.09 |
| $\mathcal{H}_{3bs}$ | 5 | 6.59 | 6.65 | 6.33 | 5.90 | 4.72 | 6.04 | 5.20 | 4.65 | 4.80 | 4.86 | 4.73 | 3.64 | 5.16 | 4.83 | 3.09 |
| $\mathcal{T}_{2bs}$ | 15 | 5.46 | 5.23 | 5.97 | 4.90 | 3.19 | 5.53 | 4.40 | 2.96 | 5.06 | 3.63 | 2.79 | 3.91 | 3.54 | 2.73 | 3.44 |
| $\mathcal{T}_{3bs}$ | 15 | 5.47 | 5.23 | 5.97 | 4.90 | 3.19 | 5.53 | 4.40 | 2.96 | 5.06 | 3.63 | 2.79 | 3.91 | 3.54 | 2.73 | 3.44 |
| $\mathcal{T}_{4bs}$ | 15 | 5.46 | 5.23 | 5.97 | 4.90 | 3.19 | 5.53 | 4.40 | 2.96 | 5.06 | 3.63 | 2.79 | 3.91 | 3.54 | 2.73 | 3.44 |
| $\mathcal{H}_{1bs}$ | 15 | 5.47 | 5.23 | 5.97 | 4.90 | 3.19 | 5.53 | 4.40 | 2.96 | 5.06 | 3.63 | 2.79 | 3.91 | 3.54 | 2.73 | 3.44 |
| $\mathcal{H}_{2bs}$ | 15 | 5.47 | 5.23 | 5.97 | 4.90 | 3.19 | 5.53 | 4.40 | 2.96 | 5.06 | 3.63 | 2.79 | 3.91 | 3.54 | 2.73 | 3.44 |
| $\mathcal{H}_{3bs}$ | 15 | 5.46 | 5.23 | 5.97 | 4.90 | 3.19 | 5.53 | 4.40 | 2.96 | 5.06 | 3.63 | 2.79 | 3.91 | 3.54 | 2.73 | 3.44 |
| $n = 100$ | | | | | | | | | | | | | | | | |
| | | $\mathbf{r}_{zu} = 0$ | | | $\mathbf{r}_{zu} = .1$ | | | $\mathbf{r}_{zu} = .3$ | | | $\mathbf{r}_{zu} = .4$ | | | $\mathbf{r}_{zu} = .5$ | | |
| Statistics | $k_2 \downarrow \eta^2 \rightarrow$ | 0 | 13 | 1000 | 0 | 13 | 1000 | 0 | 13 | 1000 | 0 | 13 | 1000 | 0 | 13 | 1000 |
| $\mathcal{T}_{2bs}$ | 5 | 5.87 | 5.70 | 5.63 | 5.48 | 4.10 | 4.59 | 4.79 | 4.00 | 2.86 | 4.88 | 3.95 | 3.12 | 4.71 | 4.07 | 3.43 |
| $\mathcal{T}_{3bs}$ | 5 | 5.88 | 5.70 | 5.63 | 5.48 | 4.11 | 4.59 | 4.81 | 4.03 | 2.87 | 4.91 | 3.99 | 3.13 | 4.74 | 4.13 | 3.44 |
| $\mathcal{T}_{4bs}$ | 5 | 5.87 | 5.70 | 5.63 | 5.48 | 4.10 | 4.59 | 4.79 | 4.00 | 2.86 | 4.88 | 3.95 | 3.12 | 4.71 | 4.07 | 3.43 |
| $\mathcal{H}_{1bs}$ | 5 | 5.88 | 5.70 | 5.63 | 5.48 | 4.11 | 4.59 | 4.81 | 4.03 | 2.87 | 4.91 | 3.99 | 3.13 | 4.74 | 4.13 | 3.44 |
| $\mathcal{H}_{2bs}$ | 5 | 5.88 | 5.70 | 5.63 | 5.48 | 4.11 | 4.59 | 4.81 | 4.03 | 2.87 | 4.91 | 3.99 | 3.13 | 4.74 | 4.13 | 3.44 |
| $\mathcal{H}_{3bs}$ | 5 | 5.87 | 5.70 | 5.63 | 5.48 | 4.10 | 4.59 | 4.79 | 4.00 | 2.86 | 4.88 | 3.95 | 3.12 | 4.71 | 4.07 | 3.43 |
| $\mathcal{T}_{2bs}$ | 15 | 5.38 | 5.37 | 5.29 | 5.18 | 3.72 | 3.25 | 4.77 | 3.89 | 2.07 | 4.76 | 3.76 | 2.02 | 4.90 | 3.84 | 2.02 |
| $\mathcal{T}_{3bs}$ | 15 | 5.38 | 5.37 | 5.29 | 5.18 | 3.73 | 3.25 | 4.77 | 3.90 | 2.07 | 4.76 | 3.77 | 2.02 | 4.91 | 3.84 | 2.02 |
| $\mathcal{T}_{4bs}$ | 15 | 5.38 | 5.37 | 5.29 | 5.18 | 3.72 | 3.25 | 4.77 | 3.89 | 2.07 | 4.76 | 3.76 | 2.02 | 4.90 | 3.84 | 2.02 |
| $\mathcal{H}_{1bs}$ | 15 | 5.38 | 5.37 | 5.29 | 5.18 | 3.73 | 3.25 | 4.77 | 3.90 | 2.07 | 4.76 | 3.77 | 2.02 | 4.91 | 3.84 | 2.02 |
| $\mathcal{H}_{2bs}$ | 15 | 5.38 | 5.37 | 5.29 | 5.18 | 3.73 | 3.25 | 4.77 | 3.90 | 2.07 | 4.76 | 3.77 | 2.02 | 4.91 | 3.84 | 2.02 |
| $\mathcal{H}_{3bs}$ | 15 | 5.38 | 5.37 | 5.29 | 5.18 | 3.72 | 3.25 | 4.77 | 3.89 | 2.07 | 4.76 | 3.76 | 2.02 | 4.90 | 3.84 | 2.02 |
| $n = 300$ | | | | | | | | | | | | | | | | |
| | | $\mathbf{r}_{zu} = 0$ | | | $\mathbf{r}_{zu} = .1$ | | | $\mathbf{r}_{zu} = .3$ | | | $\mathbf{r}_{zu} = .4$ | | | $\mathbf{r}_{zu} = .5$ | | |
| Statistics | $k_2 \downarrow \eta^2 \rightarrow$ | 0 | 13 | 1000 | 0 | 13 | 1000 | 0 | 13 | 1000 | 0 | 13 | 1000 | 0 | 13 | 1000 |
| $\mathcal{T}_{2bs}$ | 5 | 5.75 | 5.62 | 4.99 | 5.26 | 4.62 | 3.17 | 5.11 | 4.25 | 3.35 | 4.79 | 4.18 | 3.06 | 4.82 | 4.20 | 2.85 |
| $\mathcal{T}_{3bs}$ | 5 | 5.75 | 5.79 | 4.99 | 5.27 | 4.63 | 3.17 | 5.17 | 4.36 | 3.04 | 4.87 | 4.34 | 2.46 | 4.91 | 4.42 | 2.46 |
| $\mathcal{T}_{4bs}$ | 5 | 5.75 | 5.62 | 4.99 | 5.26 | 4.62 | 3.17 | 5.11 | 4.25 | 3.35 | 4.79 | 4.18 | 3.06 | 4.82 | 4.20 | 2.85 |
| $\mathcal{H}_{1bs}$ | 5 | 5.75 | 5.63 | 4.99 | 5.27 | 4.63 | 3.17 | 5.17 | 4.36 | 3.04 | 4.87 | 4.34 | 2.46 | 4.91 | 4.42 | 2.46 |
| $\mathcal{H}_{2bs}$ | 5 | 5.75 | 5.63 | 4.99 | 5.27 | 4.63 | 3.17 | 5.17 | 4.36 | 3.04 | 4.87 | 4.34 | 2.46 | 4.91 | 4.42 | 2.46 |
| $\mathcal{H}_{3bs}$ | 5 | 5.75 | 5.62 | 4.99 | 5.26 | 4.62 | 3.17 | 5.11 | 3.25 | 3.35 | 4.79 | 4.18 | 3.06 | 4.82 | 4.20 | 2.85 |
| $\mathcal{T}_{2bs}$ | 15 | 5.27 | 5.21 | 5.05 | 5.28 | 3.91 | 2.89 | 4.79 | 3.26 | 2.28 | 4.62 | 3.20 | 2.11 | 4.59 | 3.18 | 2.04 |
| $\mathcal{T}_{3bs}$ | 15 | 5.27 | 5.21 | 5.05 | 5.28 | 3.91 | 2.89 | 4.79 | 3.27 | 2.21 | 4.63 | 3.21 | 2.07 | 4.59 | 3.18 | 2.04 |
| $\mathcal{T}_{4bs}$ | 15 | 5.27 | 5.21 | 5.05 | 5.28 | 3.91 | 2.89 | 4.79 | 3.26 | 2.28 | 4.62 | 3.20 | 2.11 | 4.59 | 3.18 | 2.04 |
| $\mathcal{H}_{1bs}$ | 15 | 5.27 | 5.21 | 5.05 | 5.28 | 3.91 | 2.89 | 4.79 | 3.27 | 2.21 | 4.63 | 3.21 | 2.07 | 4.59 | 3.18 | 2.04 |
| $\mathcal{H}_{2bs}$ | 15 | 5.27 | 5.21 | 5.05 | 5.28 | 3.91 | 2.89 | 4.79 | 3.27 | 2.21 | 4.63 | 3.21 | 2.07 | 4.59 | 3.18 | 2.04 |
| $\mathcal{H}_{3bs}$ | 15 | 5.27 | 5.21 | 5.05 | 5.28 | 3.91 | 2.89 | 4.79 | 3.26 | 2.28 | 4.62 | 3.20 | 2.11 | 4.59 | 3.18 | 2.04 |

## 5. Empirical illustration

We illustrate our theoretical results through the trade and growth model [Frankel and Romer (1999), Harrison (1996), Mankiw, Romer and Weil (1992)]. The model studies the relationship between standards of living and openness. Because the trade share (ratio of imports or exports to GDP) commonly used as an indicator of openness is possibly endogenous, Frankel and Romer (1999) suggest instrumental variables method to estimate the income-trade relationship. The basic structural equation studied is given by:

$$ln(w_i) = \beta_0 + \beta \text{Trade}_i + \gamma_1 ln(\text{Pop}_i) + \gamma_2 ln(\text{Area}_i) + u_i,\ i = 1, \ldots, n \tag{5.1}$$

where $w$ is the income per capita, Pop is the population and Area is country area. The instrument $Z$ suggested by Frankel and Romer (1999) is constructed on the basis of geographic characteristics and the first stage specification is

$$\text{Trade}_i = b_0 + b_1 Z_i + c_1 ln(\text{Pop}_i) + c_2 ln(\text{Area}_i) + V_i,\ i = 1, \ldots, n. \tag{5.2}$$

Concern has raised in recent studies that the instrument constructed in this way may be invalid, weak, or both [see the review of Samuel and Michael (2009)].

Here, we want to assess the exogeneity of "Trade" taking into account a possibility that $Z$ may be invalid or weak. The data are from Frankel and Romer (1999) and contain initially 150 countries for year 1985. The first stage $F$-statistic[4] in ( 5.2) is about 13, hence the constructed instrument is not very poor compared to Staiger and Stock (1997)'s rule of thumb of 10. We use both the standard and bootstrap DWH tests to assess the exogeneity of "Trade"Ḟirst, we observe that the OLS and 2SLS estimates of $\beta$ are 0.2809 and 2.0251 respectively, while the magnitude of the regression endogeneity parameter estimate seems relatively small ($\hat{b} = 0.0079$) but is significantly different from zero at level 5% ($t_{\hat{b}} = 2.026 > 1.96$). This corresponds to error-instrument covariance estimate of about $\hat{\sigma}_{Zu} = 12.309$ and the estimate of the covariance between structural and reduced-form errors (here Trade endogeneity parameter) of approximately $\hat{\sigma}_{Zu} = -0.0149$. This suggests that the instrument $Z$ is not strictly exogenous. Hence, it is very likely that the discrepancy

[4]See also Frankel and Romer (1999, Table 2, p.385) and Dufour and Taamouti (2007).

between OLS and 2SLS estimates is more due to instrument endogeneity rather than trade share exogeneity.

Table 3 presents the outcomes for both standard and boostrap tests. The bootstrap tests are run for 999, 9,999 and 99,999 replications. The results indicate that $\mathcal{T}_2$, $\mathcal{T}_4$ and $\mathcal{H}_3$ reject the the exogeneity of the trade share at level 5%, but $\mathcal{T}_3$, $\mathcal{H}_1$ and $\mathcal{H}_2$ fail to reject it even at level 10%, despite the large discrepancy between OLS and 2SLS estimates. The maximal $p$-value is 12.43% and is obtained with $\mathcal{T}_3$. Since model identification is weak, it is more likely that this contradiction among the DWH tests is attributable to instrument invalidity. This is confirmed by the bootstrap tests (that account for instrument endogeneity) which find no evidence against the exogeneity of trade share with p-values above 38%, even for large bootstrap samples. This underscores how the standard DWH tests may be misleading if the instruments violate (even locally) the strict exogeneity assumption.

Table 3. Exogeneity of Trade in growth model ($p$-values in %)

| DWH tests | $\mathcal{T}_2$ | $\mathcal{T}_3$ | $\mathcal{T}_4$ | $\mathcal{H}_1$ | $\mathcal{H}_2$ | $\mathcal{H}_3$ |
|---|---|---|---|---|---|---|
| p-value | 04.95 | 12.43 | 04.99 | 12.22 | 11.93 | 04.69 |
| Boostrap tests with $B = 999$ | $\mathcal{T}_{2bs}$ | $\mathcal{T}_{3bs}$ | $\mathcal{T}_{4bs}$ | $\mathcal{H}_{1bs}$ | $\mathcal{H}_{2bs}$ | $\mathcal{H}_{3bs}$ |
| p-value | 50.15 | 38.14 | 50.15 | 38.14 | 38.14 | 50.15 |
| Boostrap tests with $B = 9,999$ | $\mathcal{T}_{2bs}$ | $\mathcal{T}_{3bs}$ | $\mathcal{T}_{4bs}$ | $\mathcal{H}_{1bs}$ | $\mathcal{H}_{2bs}$ | $\mathcal{H}_{3bs}$ |
| p-value | 49.6 | 38.14 | 49.6 | 38.14 | 38.14 | 49.6 |
| Boostrap tests with $B = 99,999$ | $\mathcal{T}_{2bs}$ | $\mathcal{T}_{3bs}$ | $\mathcal{T}_{4bs}$ | $\mathcal{H}_{1bs}$ | $\mathcal{H}_{2bs}$ | $\mathcal{H}_{3bs}$ |
| p-value | 50.34 | 38.2 | 50.34 | 38.2 | 38.2 | 50.34 |

## 6. Conclusion

This paper focuses on linear structural models and investigate the size of the standard Durbin-Wu-Hausman tests for exogeneity when instruments violate “strict exogeneity”. We characterize the asymptotic distributions of the tests under *local-to-zero endogeneity* which clearly show that these tests are severely size distorted even for a small correlation between instruments and the structural error. A Monte carlo experiment suggests that the maximal size distortion can be as great as 100% when instrument endogeneity is fixed and identification strong.

We then propose a bootstrap procedure to correct the size of the tests. The simulations indicate that the bootstrap tests have an overall good performance even for moderate instrument endogeneity. We apply our theoretical framework to the trade and growth model of Frankel and Romer (1999). We find that the instrument constructed on the basis of countries geographic characteristics is not strictly exogenous. Therefore, the use of the standard DWH tests for assessing the exogeneity of the trade variable lead to conflicting results, though the structural parameter is identified [see Dufour and Taamouti (2007)]. More precisely, the tests $\mathcal{T}_2$, $\mathcal{T}_4$ and $\mathcal{H}_3$ find evidence against the exogeneity of the trade variable with a $p$-values less than 5%, while $\mathcal{T}_3$, $\mathcal{H}_1$ and $\mathcal{H}_2$ cannot reject the exogeneity even at level 10%. However, all proposed bootstrap tests that account for instrument endogeneity conclude the non rejection of exogeneity with $p$-values above 38%. This suggests that the large discrepancy between OLS and 2SLS estimates is more attributable to instrument invalidity rather than trade share endogeneity.

APPENDIX

# A. Proofs

**Lemma A.1** *Suppose Assumptions* **2.1**-**2.2** *are satisfied and rank*$(\mathbf{\Pi}) = m$. *If* $\sigma_{vu} = 0$, *then*:

(**i**) $\hat{\beta} \xrightarrow{p} \beta$, $\tilde{\beta} \xrightarrow{p} \beta$, $\tilde{\beta} - \hat{\beta} \xrightarrow{p} 0$;

(**ii**)
$$\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta) &\xrightarrow{d} \mathbf{N}\left[(\mathbf{\Pi}' Q_Z \mathbf{\Pi} + \Sigma_V)^{-1}\mathbf{\Pi}'\mathbf{d}, \sigma_u^2(\mathbf{\Pi}' Q_Z \mathbf{\Pi} + \Sigma_V)^{-1}\right],\\
\sqrt{n}(\tilde{\beta} - \beta) &\xrightarrow{d} \mathbf{N}\left[(\mathbf{\Pi}' Q_Z \mathbf{\Pi})^{-1}\mathbf{\Pi}'\mathbf{d}, \sigma_u^2(\mathbf{\Pi}' Q_Z \mathbf{\Pi})^{-1}\right],\\
\sqrt{n}(\tilde{\beta} - \hat{\beta}) &\xrightarrow{d} \mathbf{N}\left(\tau_{\mathbf{d}}, \sigma_u^2[(\mathbf{\Pi}' Q_Z \mathbf{\Pi})^{-1} - (\mathbf{\Pi}' Q_Z \mathbf{\Pi} + \Sigma_V)^{-1}]\right)
\end{aligned}$$
*where* $\tau_{\mathbf{d}} = [(\mathbf{\Pi}' Q_Z \mathbf{\Pi})^{-1} - (\mathbf{\Pi}' Q_Z \mathbf{\Pi} + \Sigma_V)^{-1}]\mathbf{\Pi}'\mathbf{d}$.

Proof of Lemma **A.1** First, observe that $\hat{\beta} - \tilde{\beta}$ as:

$$\begin{aligned}
\hat{\beta} - \beta &= \hat{\Omega}_{LS}^{-1} X'u/n, \quad \tilde{\beta} - \beta = \hat{\Omega}_{IV}^{-1} X' P_Z u/n,\\
\tilde{\beta} - \hat{\beta} &= \hat{\Omega}_{IV}^{-1} X' P_Z u/n - \hat{\Omega}_{LS}^{-1} X'u/n. \qquad \text{(A.1)}
\end{aligned}$$

If Assumptions **2.1**-**2.2** hold and if further rank$(\mathbf{\Pi}) = m$ and $\sigma_{vu} = 0$, then $\hat{\Omega}_{LS} \xrightarrow{p} \mathbf{\Pi}' Q_Z \mathbf{\Pi} + \Sigma_V$, $\hat{\Omega}_{IV} \xrightarrow{p} \mathbf{\Pi}' Q_Z \mathbf{\Pi}$, $X'u/n = \mathbf{\Pi}' Z'u/n + V'u/n \xrightarrow{p} 0$, and $X' P_Z u/n = \mathbf{\Pi}' Z'u/n + V' P_Z u/n \xrightarrow{p} 0$. Thus Lemma **A.1**-(**i**) holds.

Now from (A.1), we have: $\sqrt{n}(\hat{\beta} - \beta) = \hat{\Omega}_{LS}^{-1} X'u/\sqrt{n}$, $\sqrt{n}(\tilde{\beta} - \beta) = \hat{\Omega}_{IV}^{-1} X' P_Z u/\sqrt{n}$, and $\sqrt{n}(\tilde{\beta} - \hat{\beta}) = \hat{\Omega}_{IV}^{-1} X' P_Z u/\sqrt{n} - \hat{\Omega}_{LS}^{-1} X'u/\sqrt{n}$. Under the condition of the lemma, we have $X'u/\sqrt{n} = \mathbf{\Pi}' \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i u_i + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (V_i u_i - \sigma_{Vu}) \xrightarrow{d} \mathbf{\Pi}'(\psi_{zu} + \mathbf{d}) + \psi_{Vu}$, and $X' P_Z u/\sqrt{n} = \mathbf{\Pi}' Z'u/\sqrt{n} + V' P_Z u/\sqrt{n} = \mathbf{\Pi}' Z'u/\sqrt{n} + o_p(1) \xrightarrow{d} \mathbf{\Pi}'(\psi_{zu} + \mathbf{d})$. So, we find

$$\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta) &\xrightarrow{d} (\mathbf{\Pi}' Q_Z \mathbf{\Pi} + \Sigma_V)^{-1}[\mathbf{\Pi}'(\psi_{zu} + \mathbf{d}) + \psi_{Vu}], \quad \sqrt{n}(\tilde{\beta} - \beta) \xrightarrow{d} (\mathbf{\Pi}' Q_Z \mathbf{\Pi})^{-1}\mathbf{\Pi}'(\psi_{zu} + \mathbf{d})\\
\sqrt{n}(\tilde{\beta} - \hat{\beta}) &\xrightarrow{d} \Psi_{\mathbf{d}} = (\mathbf{\Pi}' Q_Z \mathbf{\Pi})^{-1}\mathbf{\Pi}'(\psi_{zu} + \mathbf{d}) - (\mathbf{\Pi}' Q_Z \mathbf{\Pi} + \Sigma_V)^{-1}[\mathbf{\Pi}'(\psi_{zu} + \mathbf{d}) + \psi_{Vu}].
\end{aligned}$$

Under $\mathrm{H}_0$ and Assumption **2.2**, we have $(\psi_{zu} + \mathbf{d}, \psi_{Vu}) \sim \mathbf{N}\left[\mathbf{d},\ \sigma_u^2 \begin{pmatrix} Q_Z & 0 \\ 0 & \Sigma_V \end{pmatrix}\right]$ jointly, so that Lemma **A.1**-(**ii**) follows directly.

□

**Lemma A.2** *Suppose Assumptions* **2.1**-**2.2** *are satisfied and let* $\sigma_{vu} = 0$. *If further* $\mathbf{\Pi} = \mathbf{\Pi}_0/\sqrt{n}$, *where* $\mathbf{\Pi}_0 \in \mathbb{R}^{k \times m}$ *is a constant matrix (possibly zero), then*:

**(i)** $\hat{\beta} \xrightarrow{p} \beta, \quad \tilde{\beta} - \beta \xrightarrow{d} \int_{\mathbb{R}^{k\times m}} \mathbf{N}\left(\nu_{\mathbf{d}},\, \sigma_u^2[(\mathbf{\Pi}_0 + Q_Z^{-1}x_2)'Q_Z(\mathbf{\Pi}_0 + Q_Z^{-1}x_2)]^{-1}\right) pdf(x_2)dx_2;$

**(ii)** $\tilde{\beta} - \hat{\beta} \xrightarrow{d} \int_{\mathbb{R}^{k\times m}} \mathbf{N}\left(\nu_{\mathbf{d}},\, \sigma_u^2[(\mathbf{\Pi}_0 + Q_Z^{-1}x_2)'Q_Z(\mathbf{\Pi}_0 + Q_Z^{-1}x_2)]^{-1}\right) pdf(x_2)dx_2$

*where* $\nu_{\mathbf{d}} \equiv \nu_{\mathbf{d}}(x_2) = [(\mathbf{\Pi}_0+Q_Z^{-1}x_2)'Q_Z(\mathbf{\Pi}_0+Q_Z^{-1}x_2)]^{-1}(\mathbf{\Pi}_0+Q_Z^{-1}x_2)'\mathbf{d}$ *and pdf(.) is the probability density function of* $\psi_{zv}$.

Proof of Lemma **A.2** Let $\mathbf{\Pi} = \mathbf{\Pi}_0/\sqrt{n}$ where $\mathbf{\Pi}_0$ is a $k \times m$ constant matrix. When $\sigma_{vu} = 0$, we have $\hat{\Omega}_{LS} \xrightarrow{p} \Sigma_V > 0$ and $X'u/n \xrightarrow{p} 0$ so that $\hat{\beta} - \beta = o_p(1)$ and $\tilde{\beta} - \hat{\beta} = (n\hat{\Omega}_{IV})^{-1}X'P_Zu - \hat{\Omega}_{LS}^{-1}X'u/n = (n\hat{\Omega}_{IV})^{-1}X'P_Zu + 0_p(1) = \tilde{\beta} - \beta + o_p(1)$. Since $n\hat{\Omega}_{IV} = X'P_ZX = (X'Z/\sqrt{n})(Z'Z/n)^{-1}(Z'X/\sqrt{n})$ and $Z'Z/n \xrightarrow{p} Q_Z$, $Z'X/\sqrt{n} \xrightarrow{d} Q_Z\mathbf{\Pi}_0 + \psi_{zv}$, it is clear that $n\hat{\Omega}_{IV} \xrightarrow{d} (\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})'Q_Z(\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})$. By the same way, we have $X'P_Zu \xrightarrow{d} (\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})'(\psi_{zu} + \mathbf{d})$. Therefore $\hat{\beta} - \tilde{\beta} = \tilde{\beta} - \beta + o_p(1) \xrightarrow{d} \Psi_{Zv,\mathbf{d}}$ where

$$\Psi_{Zv,\mathbf{d}} \;=\; [(\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})'Q_Z(\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})]^{-1}(\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})'(\psi_{zu} + \mathbf{d}).$$

Because $\psi_{zv}$ is independent of $\psi_{zu}$ under $H_0$, we have $\Psi_{Zv,\mathbf{d}}|_{\psi_{zv}} \sim \mathbf{N}\left(\nu_{\mathbf{d}},\, \sigma_u^2[(\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})'Q_Z(\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})]^{-1}\right)$, where $\nu_{\mathbf{d}} = [(\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})'Q_Z(\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})]^{-1}(\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})'\mathbf{d}$. By integrating with respect to $\psi_{zv}$, the result follows. □

Proof of Theorem **3.1** Let rank$(\mathbf{\Pi}) = m$ and suppose Assumptions **2.1**-**2.2** hold with $\sigma_{vu} = 0$. From Lemma **A.1**, we have

$$\hat{\sigma}^2 \;=\; u'u/n - (u'X/n)\hat{\Omega}_{LS}^{-1}(X'u/n) \xrightarrow{p} \sigma_u^2, \quad \tilde{\sigma}_2^2 = \hat{\sigma}^2 + 0_p(1) \xrightarrow{p} \sigma_u^2 \tag{A.2}$$

$$\tilde{\sigma}^2 \;=\; u'u/n - 2(u'X/n)\hat{\Omega}_{IV}^{-1}(X'P_Zu/n) + (u'P_ZX/n)\hat{\Omega}_{IV}^{-1}(X'P_Zu/n) \xrightarrow{p} \sigma_u^2\,. \tag{A.3}$$

Lemma **A.1** along with (A.2)-(A.3) then entail that

$$\mathcal{T}_2 \;\xrightarrow{d}\; \frac{1}{m\sigma_u^2}\Psi_{\mathbf{d}}'[(\mathbf{\Pi}'Q_Z\mathbf{\Pi})^{-1} - (\mathbf{\Pi}'Q_Z\mathbf{\Pi} + \Sigma_V)^{-1}]^{-1}\Psi_{\mathbf{d}} \sim \frac{1}{m}\chi^2(m;\mu_{\mathbf{d}}),$$

$$\mathcal{T}_l,\, \mathcal{H}_j \;\xrightarrow{d}\; \frac{1}{m}\Psi_{\mathbf{d}}'[(\mathbf{\Pi}'Q_Z\mathbf{\Pi})^{-1} - (\mathbf{\Pi}'Q_Z\mathbf{\Pi} + \Sigma_V)^{-1}]^{-1}\Psi_{\mathbf{d}} \sim \chi^2(m;\mu_{\mathbf{d}}),\ ,\ l = 3,\,4 \quad \text{and} \quad j = 1,\,2,\,3$$

where $\mu_{\mathbf{d}} = \frac{1}{\sigma_u^2}\tau_{\mathbf{d}}'[(\mathbf{\Pi}'Q_Z\mathbf{\Pi})^{-1} - (\mathbf{\Pi}'Q_Z\mathbf{\Pi} + \Sigma_V)^{-1}]^{-1}\tau_{\mathbf{d}} = \|\bar{\tau}_{\mathbf{d}}\|^2$, $\bar{\tau}_{\mathbf{d}} = \frac{1}{\sigma_u}[(\mathbf{\Pi}'Q_Z\mathbf{\Pi} + \Sigma_V)^{-1} - (\mathbf{\Pi}'Q_Z\mathbf{\Pi})^{-1}]^{-1/2}\tau_{\mathbf{d}} = \frac{1}{\sigma_u}[(\mathbf{\Pi}'Q_Z\mathbf{\Pi})^{-1} - (\mathbf{\Pi}'Q_Z\mathbf{\Pi} + \Sigma_V)^{-1}]^{1/2}\mathbf{\Pi}'\mathbf{d}$.

□

Proof of Theorem **3.2** As in the above proof of Theorem **3.1**, we still have:

$$\hat{\sigma}^2 \;=\; u'u/n - (u'X/n)\hat{\Omega}_{LS}^{-1}(X'u/n) \xrightarrow{p} \sigma_u^2, \quad \tilde{\sigma}_2^2 = \hat{\sigma}^2 + 0_p(1) \xrightarrow{p} \sigma_u^2. \tag{A.4}$$

However, we can see from Lemma **A.2** that we now have

$$\begin{aligned}\tilde{\sigma}^2 &= u'u/n - 2(u'X/n)\hat{\Omega}_{IV}^{-1}(X'P_Z u/n) + (u'P_Z X/\sqrt{n})(n\hat{\Omega}_{IV})^{-1}(X'P_Z u/\sqrt{n}) \\ &\xrightarrow{d} \bar{\sigma}_u^2 = \sigma_u^2 + \Psi_{Zv,\mathbf{d}}'(\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})'Q_Z(\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})\Psi_{Zv,\mathbf{d}} \geq \sigma_u^2 \,. \qquad \text{(A.5)}\end{aligned}$$

The independence between $\psi_{zu}$ and $\psi_{zv}$ under $\mathrm{H}_0$, along with Lemma **A.2** then imply that $\Psi_{Zv,\mathbf{d}}'(\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})'Q_Z(\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})\Psi_{Zv,\mathbf{d}}|_{\psi_{zv}} \sim \sigma_u^2\chi^2(m;\, \|\bar{\nu}_{\mathbf{d}}\|^2)$. Therefore, $\tilde{\sigma}^2|_{\psi_{zv}} \xrightarrow{d} \sigma_u^2[1 + \chi^2(m;\, \|\bar{\nu}_{\mathbf{d}}\|^2)] \geq \sigma_u^2$, where $\bar{\nu}_{\mathbf{d}} = \frac{1}{\sigma_u}[(\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})'Q_Z(\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})]^{-1/2}(\mathbf{\Pi}_0 + Q_Z^{-1}\psi_{zv})'\mathbf{d}$. By the same way, we get $\mathcal{T}_2|_{\psi_{zv}} \xrightarrow{d} \frac{1}{m}\chi^2(m;\, \|\bar{\nu}_{\mathbf{d}}\|^2)$, $\mathcal{T}_4,\, \mathcal{H}_3|_{\psi_{zv}} \xrightarrow{d} \chi^2(m;\, \|\bar{\nu}_{\mathbf{d}}\|^2)$, and $\mathcal{T}_3,\, \mathcal{H}_j|_{\psi_{zv}} \xrightarrow{d} \frac{\chi^2(m;\, \|\bar{\nu}_{\mathbf{d}}\|^2)}{1+\chi^2(m;\, \|\bar{\nu}_{\mathbf{d}}\|^2} \leq \chi^2(m;\, \|\bar{\nu}_{\mathbf{d}}\|^2$ for $j = 1, 2$. By integrating with respect to $\psi_{zv}$, the results follow.

□

## 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, 46–63.

Ashley, R., 2009. Assessing the credibilility of instrumental variables inference with imperfect instruments via sensitivity analysis. Journal of Applied Econometrics 24(2), 325–337.

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

Bound, J., Jaeger, D. A., Baker, R. M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American Statistical Association 90, 443–450.

Chaudhuri, S., Rose, E., 2009. Estimating the veteran effect with endogenous schooling when instruments are potentially weak. Technical report, IZA Discussion Papers No. 4203.

Choi, I., Phillips, P. C. B., 1992. Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations. Journal of Econometrics 51, 113–150.

Davidson, R., Mackinnon, J. G., 2007. Testing for consistency using artificial regressions. Computational Statistics and Data Analysis 50, 3259–3281.

Doko Tchatoka, F., 2011. Subset hypotheses testing and instrument exclusion in the linear IV regression. Technical report, School of Economics and Finance, University of Tasmania Hobart, Australia.

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., 2011a. Exogeneity tests and estimation in IV regressions. Technical report, Department of Economics, McGill University, Canada 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, Canada Montréal, Canada.

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.

Frankel, J. A., Romer, D., 1999. Does trade cause growth?. American Economic Review 89(3), 379–399.

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

Guggenberger, P., 2011. On the asymptotic size distortion of tests when instruments locally violate the exogeneity assumption. Econometric Theory forthcoming.

Hall, A. R., Rudebusch, G. D., Wilcox, D. W., 1996. Judging instrument relevance in instrumental variables estimation. International Economic Review 37, 283–298.

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

Harrison, A., 1996. Oponness and growth: a time-series, cross-country analysis for developing countries. Journal of Development Economics 48, 419–447.

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

Hausman, J., Hahn, J., 2005. Estimation with valid and invalid instruments. Annales d'Économie et de Statistique 79–80, 25–57.

Imbens, G. W., 2003. Sensitivity to exogeneity assumptions in program evaluation. American Economic Review 93(2), 126–132.

Imbens, G. W., Kolesár, M., Chetty, R., Friedman, J., Glaeser, E., 20011. Inference and identification with many invalid instruments. Technical report, Department of Economics, Havard University Boston, MA.

Kiviet, J. F., Niemczyk, J., 2006. On the limiting and empirical distribution of IV estimators when some of the instruments are invalid. Technical report, Department of Quantitative Economics, University of Amsterdam Amsterdam, The Netherlands.

Kleibergen, F., 2005. Testing parameters in GMM without assuming that they are identified. Econometrica 73, 1103–1124.

Mankiw, N. G., Romer, D., Weil, D. N., 1992. A contribution to the empirics of economic growth. The Quarterly Journal of Economics 107(2), 407–437.

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

Murray, P. M., 2006. Avoiding invalid instruments and coping with weak instruments. The Journal of Economic Perspectives 20(4), 111–132.

Samuel, B., Michael, A. C., 2009. Blunt instruments: A cautionary note on establishing the causes of economic growth. Technical report, Center for Global Development N0. 171.

Sargan, J., 1958. The estimation of economic relationships using instrumental variables. Econometrica 26(3), 393–415.

Shea, J., 1997. Instrument relevance in multilinear models: A simple measure. Review of Economics and Statistics 79, 348–352.

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

Stock, J. H. , Yogo, M. , 2005. Testing for weak instruments in linear IV regression. In: D. W. Andrews , J. H. Stock, eds, Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg . Cambridge University Press, Cambridge, U.K. , chapter 6 , pp. 80–108.

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.