The Impact of a Hausman Pretest on the Asymptotic Size of a Hypothesis Test

Patrik Guggenberger*
Department of Economics
U.C.L.A.

First Version: November 2006; Revised: September 2008

Abstract

This paper investigates the asymptotic size properties of a two-stage test in the linear instrumental variables model when in the first stage a Hausman (1978) specification test is used as a pretest of exogeneity of a regressor. In the second stage, a simple hypothesis about a component of the structural parameter vector is tested, using a t-statistic that is based on either the ordinary least squares (OLS) or the two-stage least squares estimator (2SLS) depending on the outcome of the Hausman pretest. The asymptotic size of the two-stage test is derived in a model where weak instruments are ruled out by imposing a positive lower bound on the strength of the instruments. The asymptotic size equals 1 for empirically relevant choices of the parameter space. The size distortion is caused by a discontinuity of the asymptotic distribution of the test statistic in the correlation parameter between the structural and reduced form error terms. The Hausman pretest does not have sufficient power against correlations that are local to zero while the OLS based t-statistic takes on large values for such nonzero correlations. Instead of using the two-stage procedure, the recommendation then is to use a t-statistic based on the 2SLS estimator or, if weak instruments are a concern, the conditional likelihood ratio test by Moreira (2003).

JEL Classification: C12.

Keywords: Asymptotic size, exogeneity, Hausman test, pretest, size distortion.

* Patrik Guggenberger, Department of Economics, University of California Los Angeles, Bunche Hall 8385, Box 951477, Los Angeles, CA 90095. Tel.: (310) 825-0849. Email: guggenbe@econ.ucla.edu. I would like to thank the NSF for support under grant number SES-0748922, the Co-Editor Richard Smith, seminar participants, Don Andrews, Badi Baltagi, Ivan Canay, Gary Chamberlain, Victor Chernozhukov, Phoebus Dhrymes, Ivan Fernandez-Val, Raffaella Giacomini, William Greene, Jinyong Hahn, Bruce Hansen, Jerry Hausman, Guido Imbens, Dale Jorgenson, Anna Mikusheva, Marcelo Moreira, Ulrich Müller, Whitney Newey, Pierre Perron, Jack Porter, Zhongjun Qu, John Rust, Jim Stock, and Michael Wolf for comments, and the Economics Department at Harvard and the Cowles Foundation at Yale for their hospitality.
1 Introduction

This paper is concerned with the asymptotic size properties of a two-stage test, where in the first stage a Hausman (1978) specification test is used as a pretest for exogeneity of a regressor in the linear instrumental variables (IV) model. In the second stage, a hypothesis about a component of the structural parameter vector is tested using a $t$-statistic based on either the ordinary least squares (OLS) or the two-stage least squares (2SLS) estimator, depending on the outcome of the pretest. An explicit formula for the asymptotic size of the two-stage test is derived under strong instrument asymptotics in a model, where weak instruments are ruled out by imposing a positive lower bound $\kappa$ on the strength of the instruments. The asymptotic size is a function of the nominal size of the pretest, the nominal size of the second stage test, and $\kappa$. It increases as $\kappa$ or the nominal size of the pretest decreases and equals 1 for empirically relevant choices of the parameter space. It is also shown that, asymptotically, the conditional size of the two-stage test, conditional on the Hausman pretest not rejecting the null hypothesis of exogeneity, equals 1 or is close to 1 in empirically relevant scenarios. Therefore, whenever one attempts to improve the power properties of the test by using the OLS based test statistic in the second stage, one faces the danger of severe size distortion. These asymptotic size results are well reflected in finite samples, see the Supplementary Appendix.

We characterize sequences of nuisance parameters that lead to the highest null rejection probabilities of the two-stage test asymptotically. For sequences of correlations between the structural and reduced form errors that are local to zero of order $n^{-1/2}$, where $n$ denotes the sample size, the Hausman pretest statistic converges to a noncentral chi-squared distribution. The noncentrality parameter is small when the strength of the instruments is small. In this situation, the Hausman pretest has low power against local deviations of the pretest null hypothesis and consequently, with high probability, OLS based inference is done in the second stage. However, the second stage OLS based $t$-statistic often takes on very large values under such local deviations. The latter causes size distortion in the two-stage test.

Note that in the “strong instrument scenario” considered here, a 2SLS based $t$-test has correct asymptotic size while the two-stage procedure is severely size distorted in empirically relevant scenarios. If inference on the structural parameter is the objective and the researcher is concerned about the null rejection probability of the inference procedure, then, based on the above findings, the use of a Hausman test as a pretest can not be recommended. On the other hand, simply using a 2SLS based $t$-statistic is theoretically justified.

Early references on the impact of pretests include Judge and Bock (1978) and Pötscher (1991). For more references, see the survey by Leeb and Pötscher (2008). It is known that pretesting may impact the size properties of two-stage procedures. E.g. Kabaila (1995), Andrews and Guggenberger (2005e, AG henceforth), and Leeb and Pötscher (2005) show that a confidence interval based on a consistent model-selection estimator has asymptotic confidence size equal to 0. AG (2005b) and Leeb and Pötscher (2005) consider tests concerning a parameter in a linear regression model after a “conservative” model selection procedure has been applied and find extreme
size distortion for the two-stage test. The specification tests proposed in Hausman (1978) are routinely used as pretests in applied work. However, to the best of my knowledge, no results are stated anywhere in the literature regarding the negative impact of the Hausman pretest on the size properties of a two-stage test.

This paper is related to the papers by Hahn and Hausman (2002) and Hausman, Stock, and Yogo (2005). The former paper suggests a Hausman-type (pre-)test of the null hypothesis of instrument validity. The latter paper shows that a second stage Wald test is equally size distorted unconditionally and conditional on the Hahn and Hausman (2002) pretest not rejecting the null hypothesis of strong instruments. Another paper concerned with the size effects of pretests is Hall, Rudebusch, and Wilcox (1996). Dhrymes (2008) provides alternatives to Hausman pretests based on estimators of the correlation between the structural and reduced form errors. We consider some of these correlation based pretests in the Appendix.

Next, other common applications of Hausman specification tests as pretests are discussed. In a panel data context, Hausman tests are used to test whether the key assumption needed to justify the use of inference based on the random effects estimator, is satisfied. See Guggenberger (2007) for a discussion of this application. Hausman pretests have also been suggested to test for exogeneity of potential instruments. Staiger and Stock (1997) shows size distortion of the standard Hausman pretest under weakness of instruments and Hahn, Ham, and Moon (2007) introduces a modified version of the Hausman pretest that is robust to weak instruments. But Guggenberger (2008) shows that asymptotically the conditional size of the two-stage test, conditional on their pretest not rejecting, is 1.

The remainder of the paper is organized as follows. Subsections 2.1 and 2.2 describe the model and test statistic. The remaining subsections of Section 2 derive the asymptotic size results of the two-stage test when the Hausman pretest is used to test for exogeneity of a regressor. Some technical details and a discussion of pretests based on correlation estimators are given in the Appendix.

2 The Asymptotic Size of a Test After a Hausman Pretest

This section deals with the asymptotic size of a two-stage test in the linear IV model when in the first stage a Hausman pretest tests for exogeneity of a regressor.

2.1 Model and Definitions

Consider the linear IV model

\[ \begin{align*}
    y_1 &= y_2 \theta + X \zeta + u, \\
    y_2 &= Z \pi + X \phi + v,
\end{align*} \]  

(2.1)

where \( y_1, y_2 \in \mathbb{R}^n \), \( X \in \mathbb{R}^{n \times k_1} \) for \( k_1 \geq 0 \) is a matrix of exogenous variables, \( Z \in \mathbb{R}^{n \times k_2} \) for \( k_2 \geq 1 \) is a matrix of IVs, and \( (\theta, \zeta', \phi', \pi')' \in \mathbb{R}^{k_1 \times 1} \) are unknown parameters. Let \( \bar{Z} = [X; Z] \) and \( k = k_1 + k_2 \). For \( j = 1, 2 \), denote by \( y_{j,i}, u_i, v_i, X_i, Z_i, \) and \( Z_i \) the \( i \)-th rows of \( y_j, u, v, X, Z, \) and \( \bar{Z} \), respectively, written as column
vectors (or scalars). The observed data are \( y_1, y_2, X, \) and \( Z \). The data \((u_i, v_i, Z_i)\), \( i = 1, \ldots, n \), are i.i.d.

The paper investigates the asymptotic size of a two-stage test of the null hypothesis
\[
H_0 : \theta = \theta_0,
\]
where in the first stage a Hausman (1978) test is undertaken as a pretest. One- and two-sided alternatives are considered.

The Hausman pretest tests exogeneity of the variable \( y_{2,i} \). If the pretest rejects the exogeneity hypothesis, then, in the second stage, \( H_0 : \theta = \theta_0 \) is tested by using a \( t \)-test based on the 2SLS estimator. If the pretest does not reject the exogeneity hypothesis, a \( t \)-test based on the OLS estimator is used in the second stage. The rationale for the pretest is that if \( y_{2,i} \) is exogenous then a \( t \)-test based on the OLS estimator is more powerful than a \( t \)-test based on the 2SLS estimator. The power advantage is documented in Monte Carlo simulations in Guggenberger (2008) and verified theoretically in fn. 5 below where it is shown that the OLS based \( t \)-test has higher local power against Pitman drifts.

The goal is to calculate the asymptotic size of the two-stage test. By definition, the asymptotic size of a test of the null hypothesis \( H_0 : \theta = \theta_0 \) in the presence of nuisance parameters \( \gamma \in \Gamma \) equals
\[
\text{AsySz}(\theta_0) = \lim_{n \to \infty} \sup_{\gamma \in \Gamma} \sup_{P_{\theta_0,\gamma}} P_{\theta_0,\gamma}(T_n(\theta_0) > c_{1-\alpha}),
\]
where \( \alpha \) is the nominal size, \( T_n(\theta_0) \) is the test statistic, \( c_{1-\alpha} \) the critical value of the test, and \( P_{\theta,\gamma}(\cdot) \) denotes probability when the true parameters are \((\theta, \gamma)\). The test statistics \( T_n(\theta_0) \), critical values \( c_{1-\alpha} \), and parameter space \( \Gamma \) for the present application are defined in the next subsections. By definition, the asymptotic size is simply the limit as \( n \to \infty \) of the exact size \( \sup_{\gamma \in \Gamma} P_{\theta_0,\gamma}(T_n(\theta_0) > c_{1-\alpha}) \). See AG (2005a) and Section 2 in AG (2005d) for a detailed discussion of uniformity and the important distinction between pointwise null rejection probability and size. Uniformity over \( \gamma \in \Gamma \) which is built into the definition of \( \text{AsySz}(\theta_0) \) is crucial for the asymptotic size to give a good approximation for the finite sample size.

### 2.2 Test Statistics and Critical Values

In this subsection, the two-stage test statistic \( T_n(\theta_0) \) for the hypothesis test \( H_0 : \theta = \theta_0 \) is defined. Denote by \( I_n \) the \( n \)-dimensional identity matrix. For a matrix \( W \) with \( n \) rows, define \( P_W = W(W'W)^{-1}W' \), \( M_W = I_n - P_W \), \( W^\perp = M_XW \), and, if no \( X \) appears in (2.1), set \( W^\perp = W \).

The Hausman pretest statistic is defined as
\[
H_n = \frac{n(\hat{\theta}_{2SLS} - \hat{\theta}_{OLS})^2}{V_{2SLS} - V_{OLS}},
\]
(2.4)
where

\[
\hat{\theta}_{2SLS} = y_2 P_{Z^2} y_1 / (y_2' P_{Z^2} y_2),
\]
\[
\hat{\theta}_{OLS} = y_2' y_1 / (y_2' y_2),
\]
\[
\hat{V}_{2SLS} = (y_2 P_{Z^2} y_2 / n)^{-1} \sigma^2_u(\hat{\theta}_{2SLS}),
\]
\[
\hat{V}_{OLS} = (y_2' y_2 / n)^{-1} \sigma^2_u(\hat{\theta}_{OLS}), \quad \text{and}
\]
\[
\hat{\sigma}^2_u(\hat{\theta}_i) = n^{-1} (y_i^1 - y_2^1 \hat{\theta}_i)' (y_i^1 - y_2^1 \hat{\theta}_i)
\]

for \( l = OLS \) and \( 2SLS \). Other definitions of \( H_n \) are possible, that replace \( \hat{\sigma}^2_u(\hat{\theta}_{OLS}) \) by \( \hat{\sigma}^2_u(\hat{\theta}_{2SLS}) \) or vice versa. The results on the asymptotic size do not depend on which definition is used, see (2.17) below. If \( y_2 \) is exogenous and the instruments are strong then \( H_n \to_d \chi_1^2 \) as \( n \to \infty \) under assumptions given in Hausman (1978).

Define the \( t \)-test statistic

\[
T^*_l(\theta) = n^{1/2} (\hat{\theta}_l - \theta) / \hat{V}_l^{1/2}
\]

for \( l = OLS \) and \( 2SLS \). The standard definition of the two-stage test statistic is

\[
T^*_n(\theta_0) = T^*_{OLS}(\theta_0) I(H_n \leq \chi_{1,1-\beta}^2) + T^*_{2SLS}(\theta_0) I(H_n > \chi_{1,1-\beta}^2),
\]

where again \( \beta \) is the nominal size of the pretest, \( I \) is the indicator function, and \( \chi_{1,1-\beta}^2 \) the \( 1-\beta \) quantile of a chi-square random variable with one degree of freedom. Define the two-stage test statistic \( T_n(\theta_0) \) as \( \pm T^*_n(\theta_0) \) or \( |T^*_n(\theta_0)| \) depending on whether the test is a lower/upper one-sided or a symmetric two-sided test, respectively.

The nominal size \( \alpha \) standard fixed critical value (FCV) test rejects \( H_0 \) if

\[
T_n(\theta_0) > c_\infty(1 - \alpha),
\]

where \( c_\infty(1 - \alpha) = z_{1-\alpha}, z_{1-\alpha}, \) and \( z_{1-\alpha}/2 \) for the upper one-sided, lower one-sided, and symmetric two-sided test, respectively and \( z_{1-\alpha} \) is the \( 1-\alpha \) quantile of a standard normal distribution.

### 2.3 Parameter Space

In this subsection, the parameter space \( \Gamma \) of the nuisance parameter vector \( \gamma \) is defined. Following AG (2005a), the parameter \( \gamma \) has three components: \( \gamma = (\gamma_1, \gamma_2, \gamma_3) \). The points of discontinuity of the asymptotic distribution of the test statistic of interest are determined by the first component, \( \gamma_1 \). The parameter space of \( \gamma_1 \) is \( \Gamma_1 \).

The second component, \( \gamma_2 \), of \( \gamma \) also affects the limit distribution of the test statistic, but does not affect the distance of the parameter \( \gamma \) to the point of discontinuity. The parameter space of \( \gamma_2 \) is \( \Gamma_2 \). The third component, \( \gamma_3 \), of \( \gamma \) does not affect the limit distribution of the test statistic. The parameter space for \( \gamma_3 \) is \( \Gamma_3(\gamma_1, \gamma_2) \), which generally may depend on \( \gamma_1 \) and \( \gamma_2 \).
Assume that \( \{ (u_i, v_i, X_i, Z_i) : i \leq n \} \) are i.i.d. with distribution \( F \). Define the vector of nuisance parameters \( \gamma = (\gamma_1, \gamma_2, \gamma_3) \), by

\[
\begin{align*}
\gamma_1 &= \rho, \quad \gamma_2 = \| \Omega^{1/2} \pi / \sigma_v \|, \quad \text{and} \quad \gamma_3 = (F, \pi, \zeta, \phi), \\
\sigma_u^2 &= E_F u_i^2, \quad \sigma_v^2 = E_F v_i^2, \quad \rho = \text{Corr}_F(u_i, v_i),
\end{align*}
\]

\[
\Omega = QZZ - QXZ Q^{-1}_X QXZ, \quad \text{and} \quad Q = \begin{bmatrix} Q_{XX} & Q_{XZ} \\ Q_{ZX} & Q_{ZZ} \end{bmatrix} = E_F Z_i Z_i', \quad (2.9)
\]

and \( \| \cdot \| \) denotes Euclidean norm. The parameter \( \gamma_1 \) measures the degree of endogeneity of \( y_2 \). The parameter \( \gamma_2 \) measures the strength of the instruments. It is related to the well-known concentration parameter \( \mu^2 \) by \( \gamma_2 = (\mu^2 / n)^{1/2} \). Let

\[
\Gamma_1 = [-1, 1], \quad \Gamma_2 = [\kappa, \bar{\kappa}]
\]

(2.10)

for some \( 0 < \kappa < \bar{\kappa} < \infty \). The positive lower bound \( \kappa \) rules out weak instruments. The technical details of the definition of \( \Gamma_3 = \Gamma_3(\gamma_1, \gamma_2) \) are given in the Appendix, see (3.1). Finally, define the parameter space \( \Gamma \) of \( \gamma \) as

\[
\Gamma = \{ \gamma = (\gamma_1, \gamma_2, \gamma_3) : \gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2, \gamma_3 \in \Gamma_3(\gamma_1, \gamma_2) \}.
\]

(2.11)

### 2.4 Asymptotic Distributions and Size

In this subsection, the asymptotic distribution of the test statistic is derived under certain parameter sequences \( \{ \gamma_{n,h} \} \) defined below. Then the asymptotic size of the test is determined.

Let \( R_\infty = R \cup \{ \pm \infty \} \). Define

\[
H = \{ h = (h_1, h_2) \in R_\infty^2 : \exists \{ \gamma_n = (\gamma_{n,1}, \gamma_{n,2}, \gamma_{n,3}) \in \Gamma : n \geq 1 \} \text{ such that } n^{1/2} \gamma_{n,1} \to h_1 \text{ and } \gamma_{n,2} \to h_2 \}.
\]

(2.12)

It follows that

\[
H = H_1 \times H_2 = R_\infty \times [\kappa, \bar{\kappa}].
\]

(2.13)

Two cases are dealt with separately. Case I has \( |h_1| < \infty \) while Case II has \( |h_1| = \infty \). In Case I, \( \rho \to 0 \) and thus \( \text{var}(u_i, v_i) / (\sigma_u^2 \sigma_v^2) \to 1 \), see (3.2). In Case I, \( y_2 \) is only “weakly endogenous” while in Case II it is “strongly endogenous”.

**Definition of \( \{ \gamma_{n,h} \} \)**: For \( h = (h_1, h_2) \in H \), let \( \{ \gamma_{n,h} \} \subset \Gamma \) denote a sequence of parameters with components \( \gamma_{n,1,h}, \gamma_{n,2,h}, \text{ and } \gamma_{n,3,h}, \gamma_{n,h} = (\gamma_{n,1,h}, \gamma_{n,2,h}, \gamma_{n,3,h})' \), where

\[
\begin{align*}
\gamma_{n,1,h} &= \text{Corr}_{F_n}(u_i, v_i), \quad \gamma_{n,2,h} = \| \Omega_n^{1/2} \pi_n / (E_{F_n} v_i^2) \|^{1/2} \|, \quad \text{for} \\
\Omega_n &= E_{F_n} Z_i Z_i' - E_{F_n} Z_i X_i' (E_{F_n} X_i X_i')^{-1} E_{F_n} X_i Z_i', \quad \text{s.t.} \\
n^{1/2} \gamma_{n,1,h} \to h_1, \quad \gamma_{n,2,h} \to h_2, \quad \text{and} \gamma_{n,3,h} = (F_n, \pi_n, \zeta_n, \phi_n) \in \Gamma_3(\gamma_{n,1,h}, \gamma_{n,2,h}).
\end{align*}
\]

(2.14)

As Theorem 2.1 below shows, the highest asymptotic null rejection probability of the test is realized along some sequence of the type \( \{ \gamma_{n,h} \} \). It is therefore enough to
study the asymptotic null rejection rates along sequences \( \{ \gamma_{n,h} \} \). Under any sequence \( \{ \gamma_{n,h} \} \) for which \( \text{Cov}_{F_n}(u_i, v_i) \to \rho \), the following convergence result holds

\[
\begin{pmatrix}
(n^{-1}Z^\top Z)^{-1/2}n^{-1/2}Z^\top u/\sigma_u \\
(n^{-1}Z^\top Z)^{-1/2}n^{-1/2}Z^\top v/\sigma_v \\
n^{-1/2}(u'v - E_{F_n}u'v)/(\sigma_u \sigma_v)
\end{pmatrix} \to_d \begin{pmatrix}
\psi_{u,\rho} \\
\psi_{v,\rho} \\
\psi_{uv,\rho}
\end{pmatrix} \sim N(0, \begin{pmatrix} V_\rho \otimes I_{k_2} & 0 \\ 0 & 1 + \rho^2 \end{pmatrix})
\]

for \( V_\rho = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \). (2.15)

where \( \psi_{u,\rho}, \psi_{v,\rho} \in R^{k_2}, \psi_{uv,\rho} \in R \). See AG (2005c, eq. (2.15)) for similar statements.

Next the limit distribution of the test statistic \( T_n^*(\theta_0) \) is derived under sequences \( \gamma_{n,h} \). To do so, (2.15) and derivations from AG (2005c, Sections 2.3 and 4.1.2) are used. For Case I and \( \xi_h = (\xi_{1,h}, \ldots, \xi_{4,h})' \), \( h = (h_1, h_2)' \)

\[
\begin{pmatrix}
n^{-1/2}y_{2}^\top P_{Z\perp} u/(\sigma_u \sigma_v) \\
n^{-1/2}y_{2}^\top u/(\sigma_u \sigma_v) \\
n^{-1/2}y_{2}^\top y_{2}/\sigma_v^2 \\
n^{-1/2}y_{2}^\top y_{2}/\sigma_v^2
\end{pmatrix} \to_d \xi_h = \begin{pmatrix} h_2 s_{k_2}^j \psi_{u,0} \\
1 + h_2^2 \xi_h \end{pmatrix}
\]

(2.16)

where \( s_{k_2} \in R^{k_2} \) is an arbitrary vector with \( ||s_{k_2}|| = 1 \). Therefore,

\[
\begin{pmatrix}
T_{2SLS}^*(\theta_0) \\
T_{\text{OLS}}^*(\theta_0) \\
H_n \\
\sigma^2_n(\theta_{2SLS})/\sigma_u^2 \\
\sigma^2_n(\theta_{\text{OLS}})/\sigma_u^2
\end{pmatrix} \to_d \begin{pmatrix} s_{k_2}^j \psi_{u,0} \\
1 + h_2^2 \xi_h \\
1 \\
1 \\
1
\end{pmatrix} \sim N(0, \begin{pmatrix} 1 + h_2^2 \xi_h \\
1 + h_2^2 \xi_h \\
1 \\
1 \\
1
\end{pmatrix})
\]

(2.17)

for \( \eta_h = (\eta_{1,h}, \ldots, \eta_{5,h}) \). Because \( \eta_{3,h} = (1 + h_2^2)^{-1}[s_{k_2}^j \psi_{u,0} - h_2 s_{k_2}^j \psi_{uv,0} - h_2 h_1] \)

and

\[
s_{k_2}^j \psi_{u,0} - h_2 \psi_{uv,0} - h_2 h_1 \sim N(-h_2 h_1, 1 + h_2^2)
\]

the limit distribution of \( H_n \) is

\[
\chi^2_1(h_1^2 h_2^2(h_2^2 + 1)^{-1})
\]

(2.18)

where \( \chi^2_1(\cdot) \) denotes a noncentral chi-square distribution with noncentrality parameter given by the expression in brackets. Therefore, \( H_n \to_d \chi^2_1 \) if \( h_1 = 0 \), that is under exogeneity and strong instruments, we obtain Hausman’s (1978) result as a subcase.

If \( h_2 h_1 \neq 0 \), the Hausman test has nonzero local power. The noncentrality parameter \( h_1^2 h_2^2(h_2^2 + 1)^{-1} \) of the limiting distribution in (2.18) is small when \( h_2 \) is small. This leads to the poor power properties of the test. Case II is dealt with in the Appendix.

We have

\[
T_n^*(\theta_0) \to_d J_h^*,
\]

(2.19)

where \( J_h^* \), by definition, is the distribution of

\[
\eta_h^* = \eta_{2,h}^* I(\eta_{3,h} \leq \chi^2_{1,1-\beta}) + \eta_{1,h}^* I(\eta_{3,h} > \chi^2_{1,1-\beta}.
\]

(2.20)

Note that by picking \( s_{k_2} = (1, 0, \ldots, 0)' \) in (2.17), it is evident that \( \eta_h^* \) in (2.20) does not depend on \( k_2 \) and hence the asymptotic size of the two-stage test does not

[6]
depend on $k_2$ either. The distribution $J_{\psi}^*$ depends on $\beta$ but for notational simplicity, this dependence is suppressed. The derivations above imply that Assumption B in AG (2005a) holds with $r = 1/2$.

Next, an explanation is provided for the size distortion of the two-stage test. Simply to gain some intuition, we evaluate the formulas in (2.17) at $h_2 = 0$, i.e. the unidentified case. Strictly speaking, this is not allowed, because for $h_2 = 0$ weak instrument asymptotics could apply. However, by continuity, the same intuition given below applies for small values for $h_2$ rather than $h_2 = 0$. This is confirmed by the theoretical results stated below Theorem 2.1.

The formulas in (2.17) evaluated at $h_2 = 0$ read

$$T_{2SLS}^*(\theta_0) \rightarrow d_s k_2 \psi_{u,0}, \quad T_{OLS}^*(\theta_0) \rightarrow d_s \psi_{uv,0} + h_1,$$
$$H_n \rightarrow d (s_{k_2}^2 \psi_{u,0})^2 \sim \chi_1^2. \quad (2.21)$$

It follows that in this situation, the Hausman pretest rejects with probability equal to $\beta$ asymptotically. When the Hausman test does not reject the pretest hypothesis (which happens with probability $1 - \beta$) and thus the OLS based $t$-statistic is used in the second stage, the maximal asymptotic rejection probability for the null $H_0 : \theta = \theta_0$ equals 1. The latter is seen by picking $h_1$ very large or very negative depending on the type of test and recalling that $\psi_{uv,0}$ and $\psi_{u,0}$ are independent.\footnote{Note that picking a large nominal pretest size $\beta$ does not solve the problem described in the last paragraph. While picking a large $\beta$ reduces the probability at which OLS based inference is performed in the second stage, it does not lower the conditional size of the second stage test, conditional on not rejecting the pretest null hypothesis. In particular, assume the nominal size of the pretest is chosen such that $\beta = \beta_n \rightarrow 1$. Then $\chi_{1,1-\beta_n}^2 \rightarrow 0$ and $P(I(H_n \leq \chi_{1,1-\beta_n}^2) = 1) \rightarrow 0$ under any sequence $\gamma_{n,h}$ and the probability of using OLS based inference in the second stage goes to zero. Thus the two-stage test asymptotically boils down to the much simpler one-stage test that always uses a $t$-statistic based on 2SLS. However, whenever OLS based inference is used in the second stage (which happens with probability going to zero), the conditional size of the test equals 1 by the argument in the previous paragraph based on (2.21). So whenever one tries to gain power by using OLS based inference, the size of the test is completely distorted.}

The next theorem gives an explicit formula for the asymptotic size $AsySz(\theta_0)$ of the two-stage test of $H_0 : \theta = \theta_0$ based on $T_h(\theta_0)$. The results apply to upper, lower one-sided, and symmetric two-sided versions of the test with $\eta_h$ defined as $\eta_h^*, -\eta_h^*$, and $|\eta_h^*|$, respectively.

[7]
Theorem 2.1 For upper, lower, and symmetric FCV tests based on \( T_n(\theta_0) \) of nominal size \( \alpha \), the \( AsySz(\theta_0) \) equals \( \sup_{h \in H} P(\eta_h > c_\infty (1 - \alpha)) \).

The proof follows from Theorem 1(a) in AG (2005a). Note that \( AsySz(\theta_0) \) depends on \( \alpha \), \( \beta \), and \( \kappa \). For notational simplicity, this dependence is suppressed. Note that the results do not depend on \( k_1 \) and \( k_2 \geq 1 \).

Table 1 contains information on the asymptotic size of the two-stage test when \( \alpha = .05 \) for various values of \( \kappa \) and \( \beta \), namely \( \kappa \in \{.001, .1, .5, 1, 2, 10\} \) and \( \beta \in \{.05, .1, .2, .5\} \) using \( \pi = 1000 \) in the simulations. We only report results on upper and symmetric tests because results for lower (and equal-tailed) tests are virtually identical to the upper (and symmetric) ones. Note that a one-stage \( t \)-test based on the 2SLS estimator has asymptotic size equal to 5% whenever \( \kappa > 0 \).

The asymptotic size \( AsySz(\theta_0) \) decreases as \( \kappa \) or \( \beta \) increases. Table 1 shows that \( AsySz(\theta_0) \) by far exceeds the nominal size \( \alpha \) for small numbers of \( \kappa \) and \( \beta \). For example, when \( \kappa = .1 \) and \( \beta = .05 \), \( AsySz(\theta_0) \) equals .93 and .95 for upper and symmetric tests, respectively. On the other hand, when \( \kappa = 10 \) and \( \beta = .05 \), \( AsySz(\theta_0) \) equals .06 and .05 for upper and symmetric tests, respectively, and therefore basically equals the nominal size of the test. For \( \beta = .05 \) the symmetric test has asymptotic size equal to 1 for small lower bounds on the strength of the instrument. These asymptotic size results are well reflected in finite samples, see Guggenberger (2008).

Hansen, Newey, and Hausman (2004, Table 1) reports estimated concentration parameters \( \mu^2 \) for the Angrist and Krueger (1991) data for two different setups with number of instruments equal to 3 and 180, respectively. The estimated concentration parameters are \( \mu^2 = 95.6 \) and 257, respectively. For the sample size \( n = 329,509 \) this implies \( \gamma_2 = .017 \) and .028, respectively. From Table 1, note that when \( \kappa \leq .1 \) and \( \beta \leq .1 \), the asymptotic size of the two-stage test equals about .9 or larger for both one- and two-sided tests. This strongly suggests that the use of a Hausman pretest for the Angrist and Krueger (1991) data set very likely leads to extreme distortion of the null rejection probability of a two-stage procedure.

To gain further insight, the asymptotic probability of the event “pretest does not reject the pretest null hypothesis” and the conditional probability of the event “test rejects the null hypothesis” conditional on the pretest not rejecting the pretest null hypothesis, are investigated. Table 2 contains the results for the case \( h_1 = 5 \) and various values of \( h_2 \) and \( \beta \). For \( h_2 \leq 1 \), this conditional rejection probability is very close to or equal to 1 for both upper and symmetric tests for all nominal sizes \( \beta \) considered. Picking a large \( \beta \) decreases the asymptotic size of the two-stage test because it leads to more frequent use of 2SLS based inference in the second stage, but it does not decrease the size problems of the test if OLS based inference is used in the second stage. The pretest often does not detect the violation of the pretest null hypothesis, however the second stage \( t \)-statistic based on the OLS estimator takes on very large values. The probability of not rejecting the pretest null hypothesis, \( P(H_0 < \chi^2_{1,1-\beta}) \), is decreasing as \( \beta \) or \( h_2 \) increases. For \( \beta = .05 \) and \( h_2 = .1 \), it equals .92. The asymptotic size \( AsySz(\theta_0) \) is large because, the pretest null hypothesis is not rejected with a large probability and conditional on this, the second stage \( t \)-test based on OLS almost certainly rejects the null.

[8]
3 Appendix

Definition of the set $\Gamma_3(\gamma_1, \gamma_2)$: Define

$$\Gamma_3(\gamma_1, \gamma_2) = \{(F, \pi, \zeta, \phi) :$$

$$E_F u_i = E_F v_i = 0, \quad E_F u_i^2 = \sigma_u^2, \quad E_F v_i^2 = \sigma_v^2, \quad E_F Z_i Z_i' = Q = \begin{bmatrix} Q_{XX} & Q_{XZ} \\ Q_{ZX} & Q_{ZZ} \end{bmatrix},$$

for some $\sigma_u^2, \sigma_v^2 > 0$, pd $Q \in R^{k \times k}$, $\pi \in R^{k_2}$ that satisfy

$$Corr_F(u_i, v_i) = \gamma_1, \quad ||Q^{1/2} / \sigma_v|| = \gamma_2$$

for $\Omega = Q_{ZZ} - Q_{XZ}Q_{XX}^{-1}Q_{XZ}$, $\zeta, \phi \in R^{k_1}$; $E_F u_i Z_i = E_F v_i Z_i = 0$; $E_F(u_i^2, v_i^2, u_i v_i) Z_i Z_i' = (\sigma_u^2, \sigma_v^2, \sigma_u \sigma_v) Q$; $E_F(u_i^2 v_i Z_i) = E_F(u_i v_i^2 Z_i) = 0$; $var(u_i v_i)/(\sigma_u^2 \sigma_v^2) = 1 + \gamma_1^2$;

$$\lambda_{\min}(E_F Z_i Z_i') \geq M^{-1}; \quad \left\| E_F \left( |u_i / \sigma_u|^{2+\delta}, \quad |v_i / \sigma_v|^{2+\delta}, \quad |u_i v_i / (\sigma_u \sigma_v)|^{2+\delta} \right) \right\| \leq M, \quad \&$$

$$\left\| E_F \left( ||Z_i u_i / \sigma_u||^{2+\delta}, \quad ||Z_i v_i / \sigma_v||^{2+\delta}, \quad ||Z_i||^{2+\delta} \right) \right\| \leq M \right\} \quad (3.1)$$

for some constants $\delta > 0$ and $M < \infty$, where “pd” denotes “positive definite.” The restrictions in $\Gamma_3(\gamma_1, \gamma_2)$ are similar to those in AG (2005c) and comprise exogeneity restrictions on $\bar{Z}$, moment restrictions that ensure the validity of central limit theorems and conditional homoskedasticity. The additional conditions

$$E_F(u_i^2 v_i \bar{Z}_i) = E_F(u_i v_i^2 \bar{Z}_i) = 0 \quad \text{and} \quad var(u_i v_i)/(\sigma_u^2 \sigma_v^2) = 1 + \rho^2,$$  \quad (3.2)

ensure that under exogeneity and strong instruments $\hat{\theta}_{2SLS} - \hat{\theta}_{OLS}$ is asymptotically uncorrelated with $\hat{\theta}_{OLS}$. Hausman (1978) exploits the latter property when deriving the asymptotic variance of $\hat{\theta}_{2SLS} - \hat{\theta}_{OLS}$ when showing that $H \sim \chi^2_1$ under strong instruments and exogeneity of $y_2$. Sufficient conditions for (3.2) are, for example, independence of $(u_i, v_i)$ and $\bar{Z}_i$, and joint normality of $(u_i, v_i)$ with zero mean.

Limit distribution of test statistic in Case II: Under sequences $\{\gamma_n, h\}$ for which $Corr_{F_n}(u_i, v_i) \to \rho$ and $h = (h_1, h_2)'$ with $|h_1| = \infty$ the following holds jointly

$$\begin{pmatrix} n^{-1/2} y_1' P_{Z_1:u} / (\sigma_u \sigma_v) \\ n^{-1/2} y_2' P_{Z_1:v} / (\sigma_u \sigma_v) \\ n^{-1} y_1' P_{Z_1:u} y_2 / \sigma_v^2 \\ n^{-1} y_1' P_{Z_1:v} y_2 / \sigma_v^2 \end{pmatrix} \to_d \xi_h = \begin{pmatrix} s_{k_2}^{(1)} \psi_{u,\rho} \\ h_2 s_{k_2}^{(2)} \psi_{u,\rho} + \psi_{v,\rho} \\ h_2 s_{k_2}^{(2)} \psi_{v,\rho} + \psi_{u,\rho} \\ h_2 \psi_{v,\rho} + \psi_{u,\rho} \end{pmatrix} \quad (3.3)$$

and

$$\begin{pmatrix} T_{2SLS}^*(\theta_0) \\ T_{OLS}^*(\theta_0) \\ \sigma_u^2 / (\hat{\theta}_{2SLS}) / \sigma_u^2 \\ \sigma_u^2 / (\hat{\theta}_{OLS}) / \sigma_u^2 \end{pmatrix} \to_d \eta_h = \begin{pmatrix} s_{k_2}^{(1)} \psi_{u,\rho} \\ h_1 \\ \infty \\ 1 - \rho^2 / (h_2 + 1) \end{pmatrix} \quad . \quad (3.4)$$
3.1 Asymptotic Equivalence of the Hausman Pretest and Pretests Based on Correlation Estimators

Recently, Dhrymes (2008) introduced alternative exogeneity pretests based on estimators of the correlation between the structural and reduced form error terms. We derive the asymptotic distribution of the two-stage test statistic in (2.7) if instead of the Hausman pretest, a pretest based on a weighted estimated correlation is used. The limit distribution is shown in (3.7) to be identical to the one of the Hausman statistic (see (2.17) above) which implies that the two-stage test has the same asymptotic size properties when based on a Hausman pretest or on one of the alternative pretests discussed here.

To define the alternative pretest statistic, let

$$\rho(\theta) = n^{-1}(y_1^\top - y_2^\top \hat{\theta})(y_2^\top - Z^\top \hat{\pi})/(\hat{\sigma}_u^2 \hat{\sigma}_v),$$

where

$$\hat{\pi} = (Z^\top Z)^{-1}Z^\top y_2^\top,$$

$$\hat{\sigma}_v^2 = n^{-1}(y_2^\top - Z^\top \hat{\pi})(y_2^\top - Z^\top \hat{\pi}),$$

and

$$\hat{\sigma}_u^2 = \hat{\sigma}_v^2(\hat{\theta}_e)$$

for $e = \text{OLS}$ or $2\text{SLS}$. \hfill (3.5)

Here and below we leave out $n$-subscripts to simplify notation. Note that by (2.17), from an asymptotic perspective in Case I, it is irrelevant whether $\hat{\sigma}_u^2$ is defined based on $\hat{\theta}_\text{OLS}$ or $\hat{\theta}_\text{2SLS}$. Define the following exogeneity pretest statistics $\rho_{1n}$ and $\rho_{2n}$ based on weighted estimators of the correlation $\rho$

$$\rho_{1n} = (1 + \hat{\gamma}_2^{-2})(n^{1/2}\rho(\hat{\theta}_{2\text{SLS}}))^2,$$

$$\rho_{2n} = (1 + \hat{\gamma}_2^{-2})(n^{1/2}\rho(\hat{\theta}_{\text{OLS}}))^2,$$

where

$$\hat{\gamma}_2 = ||(\hat{\Omega})^{1/2}\hat{\pi}/\hat{\sigma}_v||$$

and

$$\hat{\Omega} = n^{-1}Z^\top Z.$$ \hfill (3.6)

Note the different weights $(1 + \hat{\gamma}_2^{-2})^{-1}$ and $(1 + \hat{\gamma}_2^{-2})$ in the definition of the two pretest statistics $\rho_{1n}$ and $\rho_{2n}$. The two-stage test to test $H_0 : \theta = \theta_0$ is then defined as in (2.7) with $H_n$ replaced by $\rho_{1n}$ or $\rho_{2n}$.

In Case II, both $\rho_{1n}$ and $\rho_{2n}$ diverge to infinity, as had been shown for $H_n$ in (3.4). The main result of this subsection, proven below, is that in Case I

$$\rho_{1n}, \rho_{2n} \rightarrow d \eta_{3,h}.$$ \hfill (3.7)

Thus, by (3.7), $\rho_{1n}$ and $\rho_{2n}$ have the same limit distribution as $H_n$, in particular, they are asymptotically distributed as $\chi^2_1$ when $h_1 = 0$. Therefore, the two-stage test based on the pretest statistics $\rho_{1n}$ and $\rho_{2n}$ has the same asymptotic size properties as the two-stage test based on $H_n$.

Straightforward calculations show that

$$\eta_{3,h} = (1 + h_2^{-2})(\psi_{uv,0} + h_1 - h_2^{-1}s_{k2}^2(\psi_{u,0}))^2.$$ \hfill (3.8)

Therefore, the following lemma immediately implies (3.7). For the proof see the Supplementary Appendix.
Lemma 3.2 In Case I, under $\gamma_{n,h}$, the following limits hold jointly with those in (2.17).
(a) $\tilde{\gamma}_n \to_P h_2$,
(b) $n^{1/2}\tilde{\rho}(\theta_{2SLS}) \to_d \psi_{uv,0} + h_1 - h_2^{-1}s_k^2\psi_{u,0}$,
(c) $n^{1/2}\tilde{\rho}(\theta_{OLS}) \to_d (1 + h_2^{-2})^{-1}(\psi_{uv,0} + h_1 - h_2^{-1}s_k^2\psi_{u,0})$.

Notes

1See (2.9) and (2.10) below for the precise definition of $\kappa$.

2If instruments are potentially weak, that is, the strength of the instruments is not bounded away from zero, my recommendation is to use one of the robust testing procedures suggested by Anderson and Rubin (1949), Moreira (2001, 2003), Kleibergen (2002), Guggenberger and Smith (2005), or Andrews, Moreira, and Stock (2006).

3Note that in AG (2005a-e), the specification for $\gamma$ has always been chosen such that when $\gamma$ times $n$ diverges to infinity, the “standard fixed critical value” asymptotic distribution is obtained. In this example, when $n^{1/2}\gamma \to \infty$, $y_2$ is not exogenous. Instead, the “standard” result $H_n \to_d \chi_2^2$ is obtained under $n^{1/2}\gamma \to 0$.

4Condition (3.2) in the definition of $\Gamma_3(\gamma_1, \gamma_2)$ ensures that we get the zero entries in the covariance matrix of the asymptotic distribution of $(\psi_{u,p}, \psi_{v,p}, \psi_{uv,p})$ and also that the right lower entry $[\sigma_{u2}^2\sigma_{v2}^2]var(u_i v_i)$ in the covariance matrix equals $1 + p^2$.

5If the true parameter is given by a Pitman drift $\theta = \theta_0 + n^{-1/2}\theta^*\sigma_u/\sigma_v$ for some $\theta^* \in R$, using (2.17) and additional calculations, it can be shown that in Case I, $T_{2SLS}(\theta_0) \to_d \eta_{1,h} + \theta^*h_2$ and $T_{OLS}(\theta_0) \to_d \eta_{2,h} + \theta^*\sqrt{h_2^2 + 1}$. Thus, if $h_1 = 0$, $T_{2SLS}(\theta_0) \to_d N(\theta^*h_2, 1)$ and $T_{OLS}(\theta_0) \to_d N(\theta^*\sqrt{h_2^2 + 1}, 1)$. Thus under exogeneity, the local power of the OLS based $t$-test is higher than the power of the 2SLS based test, especially for small $h_2$, a finding well reflected in simulations in Guggenberger (2008).

6In Case II, the pretest statistic goes off to infinity, $H_n \to_P \infty$, and thus w.p.a.1, $T_{2SLS}(\theta_0)$ is used in the second stage. Because $T_{2SLS}(\theta_0) \to_d N(0, 1)$, there is no size-distortion under the strong endogeneity of Case II.

7Consider, for example, the case of an upper one-sided test. For every $\varepsilon > 0$ there exists a $h_1 = h_1(\varepsilon)$ such that $P(\psi_{uv,0} + h_1 > z_{1-\varepsilon}) > 1 - \varepsilon$. Thus, under $\rho_x = n^{-1/2}h_1$, conditional null rejection probabilities no smaller than $1 - \varepsilon$ are obtained as $n \to \infty$.

8Simulations suggest that if $H_n$, $\rho_{1n}$, and $\rho_{2n}$ are based on the same choice of $\tilde{\sigma}_u(\theta_0)$ then they are even numerically identical, for example, if $\tilde{\sigma}_u^2(\theta_{2SLS})$ is used in $\tilde{V}_{2SLS}$ and $\tilde{V}_{OLS}$ in (2.5) and in (3.5). Also, because $\tilde{\sigma}_u^2(\theta_{2SLS}) \geq \tilde{\sigma}_u^2(\theta_{OLS})$, $H_n$ is smallest if defined as in (2.5) and largest if $\tilde{\sigma}_u^2(\theta_{OLS})$ is used in $\tilde{V}_{2SLS}$ and $\tilde{V}_{OLS}$.
References


——— (2005b) Hybrid and size-corrected subsampling methods. Accepted at Econometrica.


——— (2005d) Validity of subsampling and “plug-in asymptotic” inference for parameters defined by moment inequalities. Accepted at Econometric Theory.


Table 1.

AsySz($\theta_0$) (in %) of Two-stage FCV Test for $\alpha = .05$

<table>
<thead>
<tr>
<th>$\kappa \backslash \beta$</th>
<th>.05</th>
<th>.1</th>
<th>.2</th>
<th>.5</th>
<th>.05</th>
<th>.1</th>
<th>.2</th>
<th>.5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
<td>94.9</td>
<td>85.0</td>
<td>55.6</td>
<td>100</td>
<td>94.9</td>
<td>85.0</td>
<td>55.6</td>
</tr>
<tr>
<td>.001</td>
<td>97.4</td>
<td>94.8</td>
<td>84.9</td>
<td>55.4</td>
<td>100</td>
<td>94.9</td>
<td>85.0</td>
<td>55.6</td>
</tr>
<tr>
<td>.1</td>
<td>93.0</td>
<td>88.4</td>
<td>80.3</td>
<td>51.5</td>
<td>95.2</td>
<td>89.9</td>
<td>80.1</td>
<td>51.0</td>
</tr>
<tr>
<td>.5</td>
<td>62.4</td>
<td>52.9</td>
<td>40.6</td>
<td>23.1</td>
<td>58.6</td>
<td>50.0</td>
<td>38.9</td>
<td>21.4</td>
</tr>
<tr>
<td>1</td>
<td>30.0</td>
<td>24.2</td>
<td>18.5</td>
<td>10.5</td>
<td>27.0</td>
<td>20.4</td>
<td>15.8</td>
<td>9.9</td>
</tr>
<tr>
<td>2</td>
<td>13.5</td>
<td>11.1</td>
<td>8.8</td>
<td>6.5</td>
<td>10.7</td>
<td>9.3</td>
<td>7.7</td>
<td>6.2</td>
</tr>
<tr>
<td>10</td>
<td>5.9</td>
<td>5.6</td>
<td>5.4</td>
<td>5.2</td>
<td>5.3</td>
<td>5.3</td>
<td>5.2</td>
<td>5.2</td>
</tr>
</tbody>
</table>

Table 2.

Asymptotic Rejection Probabilities (in %) of Two-stage Test Conditional on Pretest Not Rejecting for $\alpha = .05$ and $h_1 = 5$

<table>
<thead>
<tr>
<th>$h_2 \backslash \beta$</th>
<th>.05</th>
<th>.1</th>
<th>.2</th>
<th>.5</th>
<th>.05</th>
<th>.1</th>
<th>.2</th>
<th>.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99.9</td>
<td>99.9</td>
<td>99.9</td>
<td>99.9</td>
</tr>
<tr>
<td>.001</td>
<td>99.7</td>
<td>99.8</td>
<td>99.8</td>
<td>99.8</td>
<td>99.4</td>
<td>99.4</td>
<td>99.5</td>
<td>99.5</td>
</tr>
<tr>
<td>.1</td>
<td>97.5</td>
<td>97.4</td>
<td>97.3</td>
<td>96.9</td>
<td>94.5</td>
<td>94.8</td>
<td>95.2</td>
<td>94.8</td>
</tr>
<tr>
<td>.5</td>
<td>71.7</td>
<td>71.2</td>
<td>71.4</td>
<td>74.5</td>
<td>60.8</td>
<td>59.8</td>
<td>60.3</td>
<td>61.4</td>
</tr>
<tr>
<td>10</td>
<td>13.0</td>
<td>14.8</td>
<td>14.8</td>
<td>14.3</td>
<td>8.5</td>
<td>10.0</td>
<td>9.0</td>
<td>8.6</td>
</tr>
</tbody>
</table>

Notes: The results in Tables 1 and 2 are based on $R = 50,000$ simulation repetitions. If conditional events occur less than 100 times, the number of repetitions is increased.
4 Supplementary Appendix

Proof of Lemma 3.2. (a) Note that by (2.15) in AG (2005c), we have
\[ \Omega_n^{-1} \hat{\Theta} \rightarrow_p I_{k_2}. \]  
(4.9)
Furthermore,
\[ n^{1/2} \Omega_n^{1/2} (\pi - \hat{\pi})/\sigma_v = n^{1/2} \Omega_n^{1/2} (\pi - (Z^\prime Z)^{-1} Z^\prime y_2^*)/\sigma_v \]
\[ = -n^{1/2} \Omega_n^{1/2} (Z^\prime Z)^{-1} Z^\prime v/\sigma_v \]
\[ = -\Omega_n^{1/2} (n^{-1} Z^\prime Z)^{-1/2} [(n^{-1} Z^\prime Z)^{-1/2} n^{-1/2} Z^\prime v/\sigma_v] \]
\[ \rightarrow d - \psi_{v,0}. \]  
(4.10)
where the last line holds by (2.15) and (4.9). Finally,
\[ \hat{\sigma}_v^2/\sigma_v^2 = n^{-1} (y_2 - Z\hat{\pi})^\prime (y_2 - Z\hat{\pi})^\prime/\sigma_v^2 \]
\[ = n^{-1} (Z(\pi - \hat{\pi}) + v)^\prime (Z(\pi - \hat{\pi}) + v)^\prime/\sigma_v^2 \]
\[ = n^{-1} [v^\prime v + 2(\pi - \hat{\pi})^\prime Z^\prime v + (\pi - \hat{\pi})^\prime Z^\prime (\pi - \hat{\pi})]/\sigma_v^2 \]
\[ \rightarrow p. \]  
(4.11)
The last step holds by the moment assumptions in (3.1) and by straightforward calculations using (4.10) that establish \( n^{-1} (\pi - \hat{\pi})^\prime Z^\prime v/\sigma_v^2 \rightarrow_p 0 \) and \( n^{-1} (\pi - \hat{\pi})^\prime Z^\prime Z (\pi - \hat{\pi})/\sigma_v^2 \rightarrow_p 0 \). Result (a) follows from (4.9), (4.10), and (4.11).

To prove parts (b) and (c), note that for \( e = OLS \) or \( 2SLS \)
\[ n^{1/2} \rho(\hat{\theta}_e) = n^{-1/2} (y_1 - y_2 \hat{\theta}_e)^\prime (y_2 - Z\hat{\pi})^\prime/(\hat{\sigma}_u \hat{\sigma}_v) \]
\[ = n^{-1/2} [(Z\pi + v) (\theta - \hat{\theta}_e) + u]^\prime [(Z(\pi - \hat{\pi}) + v)]^\prime/(\sigma_u \sigma_v) + o_p(1), \]  
(4.12)
where the second equality holds by by lines 4 and 5 of (2.17) and by (4.11). Note that by (4.10) and (2.15)
\[ n^{-1/2} u^\prime (Z(\pi - \hat{\pi}) + v)/\sigma_v \]
\[ = n^{-1/2} u'v/(\sigma_u \sigma_v) + O_p(n^{-1/2}) \]
\[ \rightarrow d\psi_{uv,0} + h_1. \]  
(4.13)
To deal with the additional terms in (4.12), note that it follows from (2.16) that
\[ n^{1/2} (\sigma_u/\sigma_u)(\theta - \tilde{\theta}_{OLS}) \rightarrow d - (1 + h_2^2)^{-1} (\psi_{uv,0} + h_1 + h_2 s_{k_2} \psi_{u,0}), \]
\[ n^{1/2} (\sigma_u/\sigma_u)(\theta - \tilde{\theta}_{2SLS}) \rightarrow d - h_2^{-1} s_{k_2} \psi_{u,0}. \]  
(4.14)
Therefore, by straightforward calculations,
\[ n^{1/2} [(Z\pi + v)(\theta - \tilde{\theta}_e)]^\prime [Z(\pi - \hat{\pi}) + v]^\prime/(\sigma_u \sigma_v) \]
\[ \rightarrow d n^{1/2} (\sigma_u/\sigma_u)(\theta - \hat{\theta}_e) + o_p(1). \]  
(4.15)
Combining (4.13) and (4.15), parts (b) and (c) follow. \[ \square \]
4.1 Finite Sample Evidence

In this section, the finite sample size properties of the two-stage test are investigated in a simulation study based on parameter choices for the concentration parameter \( \mu^2 = n\pi'EZ_iZ_i'/\pi/Ev_i^2 \) and the correlation \( \rho = \text{Corr}(u_i, v_i) \) that were estimated from data sets in applied papers published in the last five years in the American Economic Review (AER), Journal of Political Economy (JPE), and the Quarterly Journal of Economics (QJE), see Hansen, Hausman, and Newey (2004). [Note that the concentration parameter \( \mu^2 \) equals \( n\gamma_2 \) when there are no included exogenous variables. In general, the concentration parameter is defined as \( n^2 \) where \( \gamma_2 \) is defined in (2.9).] Their Table 7 is reproduced here; it reports several percentiles Q10, ..., Q90 for the concentration and correlation parameters in these data sets:

| \( \mu^2 \) | 28 | 8.95 | 12.7 | 23.6 | 105 | 588 |
| \( \rho \) | 22 | .022 | .0735 | .279 | .466 | .555 |

In the simulations, the nominal sizes of the pretest and the second stage test are \( \alpha = \beta = .05 \). Furthermore, \( EZ_iZ_i' = I_{k_2} \) and \( Ev_i^2 = 1 \). This implies \( ||\pi|| = \sqrt{\mu^2n^{-1/2}} \). The vector \( \pi \) is chosen to have all components equal, \( \pi = \pi_0(1, ..., 1)' \in R^{k_2} \) for \( \pi_0 \in R \). The vector \( (u_i, v_i, Z_i) \) is chosen as i.i.d. normal with zero mean and unit variances and \( Z_i \) is independent of \( u_i \) and \( v_i \). The asymptotic results do not depend on \( k_1 \), the number of included exogenous variables, and therefore \( k_1 = 0 \) in the simulations.

Two Monte Carlo experiments based on the information in Table 7 of Hansen, Hausman, and Newey (2004) are implemented.

In the first experiment, the values of \( \mu^2 \) and \( \rho \) are fixed at the estimated median values over the data sets, namely \( \mu^2 = 23.6 \) and \( \rho = .279 \). Empirical null rejection probabilities of the two-stage test are reported for various values of the sample size \( n \) and the number of instruments \( k_2 \), namely \( n \in \{100, 1000, 10000\} \) and \( k_2 = \{1, 5, 20\} \). In Table Ia below, columns 4 and 5 with headings “Upper” and “Sym” report these finite sample null rejection probabilities for upper and symmetric two-stage tests. Column 6 with heading “HPre” reports null rejection probabilities of the Hausman pretest. Finally, columns 7 and 8 with headings “CondlUpper” and “CondlSym” report conditional probabilities of rejecting the null hypothesis of the second stage test, conditional on the Hausman pretest not rejecting the pretest null hypothesis.

For all configurations, the two-stage test overrejects severely, with null rejection probabilities in the range \([.62, .85]\). The pretest null hypothesis is only rejected with probabilities ranging roughly between 10% and 20% even though \( \rho = .279 \). However, conditional on not rejecting the pretest null hypothesis and thus using an OLS based \( t \)-statistic in the second stage, the null rejection probabilities equal 100% in most scenarios. The OLS based \( t \)-statistic takes on very large values under the failure of the pretest null hypothesis while the Hausman pretest does not.
In the second experiment, the sample size and the number of instruments are fixed at $n = 1000$, $k_2 = 5$ and various values of the concentration parameter $\mu^2$ and $\rho$ are considered that cover the whole range of values reported in Hansen, Hausman, and Newey (2004), namely $\mu^2 \in \{0, 13, 50, 113, 200, 313, 450, 613\}$ and $\rho \in \{0, .05, .1, .2, .3, .4, .5, .6\}$. Therefore, the results cover all the cases of combinations of $\mu^2$ and $\rho$ that were found in the applied papers in the last five years in AER, JPE, and QJE considered in the table above. For each such combination, Table Ib below reports null rejection probabilities of the symmetric two-stage test and of the symmetric $t$-test based on the 2SLS estimator. The results strongly suggest that in terms of null rejection probabilities, simply using the one-stage $t$-test, is the better of the two methods. In situations, where the two-stage test has good null rejection probabilities (the cases where $\rho = 0$ or ($\rho \geq .3$ and $\mu^2 \geq 200$)), the same is true for the one-stage $t$-test. However, in all other situations the two-stage test overrejects, oftentimes severely, while the one-stage test has relatively good size properties (except when $\rho \geq .5$ and $\mu^2 \leq 13$). For example, for the cases (.1 $\leq \rho \leq .4$ and $\mu^2 \leq 13$) the null rejection probabilities of the two-stage test fall into the interval [.84, 1.00] while the corresponding interval for the one-stage test is [0, .1]. For $\rho = .1$ the null rejection probability of the two-stage test is .87 when $\mu = 13$ and .38 when $\mu = 613$ while for the one-stage test, the corresponding probabilities are .01 and .05.
### TABLE Ia

Finite Sample Null Rejection Probabilities (in %) of Two-stage Test

$k_1 = 0$, $\alpha = \beta = .05$, $||\pi|| = \sqrt{23.6} n^{-1/2}$, $p = .279$; based on 50,000 repetitions

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k_2$</th>
<th>$\pi_0$</th>
<th>Upper</th>
<th>Sym</th>
<th>HPre</th>
<th>CondlUpper</th>
<th>CondlSym</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>1</td>
<td>.49</td>
<td>69.6</td>
<td>62.4</td>
<td>15.4</td>
<td>82.2</td>
<td>73.0</td>
</tr>
<tr>
<td>1000</td>
<td>1</td>
<td>.15</td>
<td>78.9</td>
<td>79.4</td>
<td>21.1</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10000</td>
<td>1</td>
<td>.05</td>
<td>78.6</td>
<td>79.1</td>
<td>21.5</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>.22</td>
<td>70.9</td>
<td>63.2</td>
<td>14.0</td>
<td>82.4</td>
<td>73.0</td>
</tr>
<tr>
<td>1000</td>
<td>5</td>
<td>.07</td>
<td>80.7</td>
<td>81.0</td>
<td>19.3</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10000</td>
<td>5</td>
<td>.02</td>
<td>80.3</td>
<td>80.7</td>
<td>19.7</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>.11</td>
<td>74.3</td>
<td>66.2</td>
<td>10.3</td>
<td>82.5</td>
<td>73.4</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>.03</td>
<td>85.3</td>
<td>85.4</td>
<td>14.8</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10000</td>
<td>20</td>
<td>.01</td>
<td>84.2</td>
<td>84.3</td>
<td>15.9</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

### TABLE Ib

Finite Sample Null Rejection Probabilities (in %) of Symmetric Two-stage Test and 2SLS Based $t$-Test

$k_1 = 0$, $\alpha = \beta = .05$, $n = 1000$, $k_2 = 5$; based on 50,000 repetitions