The Impact of a Hausman Pretest on the Size of a Hypothesis Test: the Panel Data Case

Patrik Guggenberger*

Department of Economics

UCSD

September 6, 2009

Abstract: The size properties of a two–stage test in a panel data model are investigated where in the first stage a Hausman (1978) specification test is used as a pretest of the random effects specification and in the second stage, a simple hypothesis about a component of the parameter vector is tested, using a t–statistic that is based on either the random effects or the fixed effects estimator depending on the outcome of the Hausman pretest. It is shown that the asymptotic size of the two–stage test depends on the degree of time variation in the regressors and on the variance of the error term relative to the variance of the individual specific effect and equals 1 for empirically relevant specifications of the parameter space. Monte carlo simulations document that the size distortion is well reflected in finite samples. The size distortion is caused mainly by the poor power properties of the pretest that lead to frequent unjustified inference based on the random effects estimator in the second stage. However, it is also shown that the conditional size of the test, conditional on the Hausman pretest rejecting the pretest null hypothesis, exceeds the nominal level of the test. Given the results in the paper, the recommendation then is to use a t–statistic based on the fixed effects estimator instead of using the two–stage procedure.

Keywords: Fixed effects estimation, Hausman specification test, panel data, pretest, random effects estimation, size distortion.

JEL classification: C12, C23, C52.

*Patrik Guggenberger, Department of Economics, University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093. Email: pguggenberger@ucsd.edu. The first version of the paper was written in December 2007. I would like to thank the Sloan foundation and the NSF for support under grant number SES-0748922.

1 Introduction

When deciding between inference based on the random effects or the fixed effects estimator in a panel data model, it is quite standard in applied work to first implement a Hausman (1978) pretest. If the Hausman pretest rejects the pretest null hypothesis that the random effects specification is correct, inference based on the fixed effects estimator is used in the second stage, otherwise inference based on the random effects estimator is used which has favorable power properties. For example, Blonigen (1997) justifies the use of random effects inference based on a Hausman pretest while Hastings (2004) uses fixed effects inference as a result of the Hausman pretest rejecting the random effects specification. The Hausman pretest is a common tool, used in hundreds of applied papers and discussed in most textbooks in Econometrics, see e.g. Wooldridge (2002, chapter 10.7.3).

It is shown in this paper that the asymptotic size of the resulting two–stage test equals 1 for empirically relevant specifications of the parameter space. An explicit formula for the asymptotic size of the two–stage test is derived. It shows that the asymptotic size depends on the degree of time variation in the regressors and also on the relative magnitude of the error variance to the variance of the individual specific effect. Our results explain how these two quantities impact the size of the two–stage test. The result that the two–stage test is size distorted is related to the findings in Guggenberger (2007). In that paper it is shown that the corresponding two–stage test in the linear instrumental variables (IV) model has size 1, where the Hausman pretest is used as a test of exogeneity of a regressor. As outlined in more detail below, the analysis of the panel data example is more complicated than the analysis of the IV example, because in the former case the asymptotic size depends on a higher dimensional nuisance parameter vector than in the latter case.

Based on the general theory developed in Andrews and Guggenberger (2005a, AG(2005a) from now on), we characterize sequences of nuisance parameters that lead to the highest null rejection probabilities of the two–stage test asymptotically. It is shown that under certain local deviations from the random effects specification, the Hausman pretest statistic converges to a noncentral chi–square distribution. The noncentrality parameter is small when the error variance is large relative to the variance of the individual specific effect or when the regressors are positively correlated over time. In this situation, the Hausman pretest has low power against local deviations of the pretest null hypothesis and consequently, with high probability, inference based on the random effects estimator is performed in the second stage which leads to size distortion. However, it is also shown that the conditional size of the two–stage test, conditional on the Hausman pretest rejecting the pretest null hypothesis, exceeds the nominal level of the test.

Given the results in the paper, if controlling the size of a testing procedure is an objective, the use of the two–stage procedure cannot be recommended. Its asymptotic size is severely distorted and the size distortion is well reflected in finite sample simulations. On the other hand, use of a t–statistic based on the fixed effects estimator has correct asymptotic size and performs well in finite samples. If the random effects specification is correct, inference based on the random effects estimator has correct size and has favorable power properties, but of course leads to size distortion otherwise. Given the results in the paper, the random effects specification should not be tested using a Hausman pretest.

It has been long known that pretests have an impact on the risk properties of estimators and the size properties of tests, see Judge and Bock (1978) for an early reference and Guggenberger (2007) for additional references. As documented further below, the specification tests proposed in Hausman (1978) are routinely used as pretests in applied work. However, besides Guggenberger (2007) where the case of the linear IV model is studied, 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.

The remainder of the paper is organized as follows. Section 2 describes the model, the objective, and defines the test statistics. In Section 3, finite sample evidence is provided. Subsection 4.1 is based on AG(2005a). It provides theoretical background on how to calculate the asymptotic size of a test in situations where the test statistic has a limiting distribution that is discontinuous in nuisance parameters. In subsection 4.2 this theory is then applied to the situation of a two–stage test where in the first stage a Hausman pretest is implemented. All technical details are given in the Appendix.

The following notation is used in the remainder of the paper. We denote by 1_T and I_T a T–vector of ones and the T–dimensional identity matrix, respectively. For a matrix A with T rows, let $M_A = I_T - P_A$, where $P_A = A(A'A)^{-1}A'$ is the projection onto the column space of A. By $\chi^2_{1,\beta}$ and z_β we denote the β–quantile of a chi–square distribution with one degree of freedom and of a standard normal distribution, respectively. By $I(\cdot)$ we denote the indicator function that equals 1 if the argument is true and 0 otherwise. Denote by $||\cdot||$ the Euclidean norm. Finally, let $R_+ = \{x \in R : x \geq 0\}$, $R_\infty = R \cup \{\pm\infty\}$, and $R_{+,\infty} = R_+ \cup \{+\infty\}$.

2 Model and Objective

Consider the simple panel data model

$$y_{it} = \lambda + x_{it}\theta + c_i + u_{it}, \tag{1}$$

for $i = 1, ..., N$, $t = 1, ..., T$ with scalar parameters λ and θ and individual specific effect c_i.[1] Denote the regressor vector by $w'_{it} = (1, x_{it})$. By y_i, x_i, w_i, and u_i we denote the matrices (or vectors) with T rows given by y_{it}, x_{it}, w'_{it}, and u_{it}, respectively. The observed data are $(y_i, x_i) \in R^{T\times 2}$, $i = 1, ..., N$. The data (x_i, c_i, u_i), $i = 1, ..., N$ are assumed to be i.i.d. with distribution F and u_{it}, $t = 1, ..., T$ are i.i.d. Assume $E_F x_{it} = E_F c_i = E_F u_{it} = 0$, $E_F c_i u_{it} = 0$, and define $\sigma_u^2 = E_F u_{it}^2$ and $\sigma_c^2 = E_F c_i^2$.[2] Our asymptotic framework has $N \to \infty$, but T fixed.

The object of interest is to test the null hypothesis

$$H_0 : \theta = \theta_0 \tag{2}$$

against a one– or two–sided alternative. One possibility to test (2) is to use a t–statistic $T_{RE}(\theta_0)$ based on the random effects estimator $\widehat{\theta}_{RE}$ of θ. To define these quantities, let $\widehat{\Omega} \in R^{T\times T}$ be an estimator of the variance–covariance matrix

$$\Omega = E_F(u_i + 1_T c_i)(u_i + 1_T c_i)' = \sigma_u^2 I_T + \sigma_c^2 1_T 1'_T \in R^{T\times T} \tag{3}$$

that replaces σ_u^2 and σ_c^2 in Ω by estimated counterparts $\widehat{\sigma}_u^2$ and $\widehat{\sigma}_c^2$. Possible choices for $\widehat{\sigma}_u^2$, $\widehat{\sigma}_c^2$, and for the estimator $\widetilde{\sigma}_u^2$ introduced below are discussed in subsection 5.1 of the Appendix. Then[3]

$$\begin{aligned}
(\widehat{\lambda}_{RE}, \widehat{\theta}_{RE})' &= \left(\sum_{i=1}^N w'_i(\widehat{\Omega})^{-1} w_i\right)^{-1} \sum_{i=1}^N w'_i(\widehat{\Omega})^{-1} y_i, \\
\widehat{V}_{RE} &= (N^{-1}\sum_{i=1}^N x'_i(\widehat{\Omega})^{-1} x_i)^{-1}, \text{ and} \\
T_{RE}(\theta_0) &= \frac{N^{1/2}(\widehat{\theta}_{RE} - \theta_0)}{(\widehat{V}_{RE})^{1/2}}.
\end{aligned} \tag{4}$$

[1] Additional regressors w_{it} could be included into the model at the expense of more complicated notation. The asymptotic results of the paper are identical if the intercept λ is not included into the model.

[2] To simplify notation, we do not index σ_u^2 and σ_c^2 by F.

[3] The definition of $T_{RE}(\theta_0)$ in (4) could also be altered by replacing $N^{-1}\sum_{i=1}^N x'_i(\widehat{\Omega})^{-1}x_i$ by

$$N^{-1}[\sum_{i=1}^N x'_i(\widehat{\Omega})^{-1}x_i - \sum_{i=1}^N x'_i(\widehat{\Omega})^{-1}1_T(\sum_{i=1}^N 1'_T(\widehat{\Omega})^{-1}1_T)^{-1}\sum_{i=1}^N 1'_T(\widehat{\Omega})^{-1}x_i].$$

As verified in (42) in the Appendix, this modification makes no difference asymptotically for the results in the paper.

Alternatively, the test of (2) can be based on the fixed effects estimator $\widehat{\theta}_{FE}$ of θ

$$\widehat{\theta}_{FE} = (\sum_{i=1}^{N} x_i' M_{1_T} x_i)^{-1} \sum_{i=1}^{N} x_i' M_{1_T} y_i \tag{5}$$

by using the t–statistic

$$T_{FE}(\theta_0) = \frac{N^{1/2}(\widehat{\theta}_{FE} - \theta_0)}{(\widehat{V}_{FE})^{1/2}}, \text{ where}$$

$$\widehat{V}_{FE} = (N^{-1} \sum_{i=1}^{N} x_i' M_{1_T} x_i / \widetilde{\sigma}_u^2)^{-1} \tag{6}$$

for an estimator $\widetilde{\sigma}_u^2$ of σ_u^2 that may differ from $\widehat{\sigma}_u^2$.

Let $\overline{x}_i = T^{-1} \sum_{t=1}^{T} x_{it}$ be the time average of the regressor. Inference based on the t–statistic T_{RE} is justified if $\overline{x}_i$ and c_i are uncorrelated but size distorted otherwise. On the other hand if $\overline{x}_i$ and c_i are uncorrelated, inference based on T_{RE} provides power advantages over inference based on T_{FE}. Because of this trade–off between robustness and power, oftentimes in applied work, before testing (2), a Hausman (1978) pretest is undertaken. The pretest tests whether the pretest null hypothesis

$$H_{P,0} : Corr_F(c_i, \overline{x}_i) = 0 \tag{7}$$

is true. If the pretest rejects the pretest hypothesis, then, in the second stage, $H_0 : \theta = \theta_0$ is tested based on $T_{FE}(\theta_0)$, the robust testing procedure when $H_{P,0}$ is false. If the pretest does not reject (7), then in the second stage (2) is tested based on $T_{RE}(\theta_0)$, the more powerful testing procedure when (7) is true. Thus, denoting by

$$H_N = \frac{N(\widehat{\theta}_{FE} - \widehat{\theta}_{RE})^2}{\widehat{V}_{FE} - \widehat{V}_{RE}} \tag{8}$$

the Hausman statistic and by β the nominal size of the pretest, the resulting two–stage test statistic $T_N(\theta_0)$ is given by

$$T_N^*(\theta_0) = T_{RE}(\theta_0) I(H_N \leq \chi^2_{1,1-\beta}) + T_{FE}(\theta_0) I(H_N > \chi^2_{1,1-\beta}) \tag{9}$$

for an upper one–sided test and by $-T_N^*(\theta_0)$ or $|T_N^*(\theta_0)|$ for a lower one–sided or a symmetric two–sided test, respectively. The nominal size α test rejects H_0 if

$$T_N(\theta_0) > c_\infty(1 - \alpha), \tag{10}$$

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.

The goal of this paper is to illustrate the impact of the pretest on the size properties of the two–stage test. We show that the asymptotic size $AsySz(\theta_0)$ of the test differs substantially from its nominal size α and determine the parameters that impact $AsySz(\theta_0)$. Note that if $Corr_F(c_i, \overline{x}_i)$ is nonzero and kept fixed as N goes off to infinity, then the two–stage procedure has asymptotic null rejection probability equal to the nominal size of the test because in this case the Hausman pretest statistic diverges to infinity, and in the second stage T_{FE} is used with probability approaching 1. However, this is only a pointwise justification of the two–stage procedure and it does not hold uniformly.

3 Finite Sample Evidence

For illustration of the overrejection problem of the two–stage test defined in (10), finite sample simulations are conducted in this section. The theoretical results below show that the only parameters that impact the null rejection probability asymptotically are $\gamma_1 = Corr_F(c_i, \overline{x}_i)$, $\gamma_{21} = (T\sigma^2_{\overline{x}_i}/E_F||x_i||^2)^{1/2}$, and $\gamma_{22} = (T\sigma^2_c/\sigma^2_u)^{1/2}$, where

$$(11) \qquad \sigma^2_{\overline{x}_i} = E_F\overline{x}^2_i.$$

The results below also prove that asymptotically, it does not matter whether or not the intercept λ is included in the model (1) and, for simplicity, we therefore consider a model without the intercept. We choose sample size $N = 100$ and $T = 2$ and generate $R = 30,000$ i.i.d. draws from

$$(12) \qquad \begin{pmatrix} x_{i1} \\ x_{i2} \\ c_i \end{pmatrix} \sim N(0, \begin{pmatrix} 1 & p & q \\ p & 1 & q \\ q & q & 1 \end{pmatrix}) \text{ and } \begin{pmatrix} u_{i1} \\ u_{i2} \end{pmatrix} \sim N(0, \begin{pmatrix} \sigma^2_u & 0 \\ 0 & \sigma^2_u \end{pmatrix}).$$

Note that different values for (p, q, σ^2_u) translate into values for γ_1, γ_{21}, and γ_{22} through the relation $\gamma_1 = q/(.5+.5p)^{1/2}$, $\gamma_{21} = (.5+.5p)^{1/2}$, and $\gamma_{22} = (2/\sigma^2_u)^{1/2}$. Finally, we choose nominal sizes $\alpha = \beta = .05$.[4]

[4]Small choices for the pretest nominal size β such as 5% are common in applied work when applying Hausman pretests, both in panel data and linear IV applications. E.g. Gaynor, Seider, and Vogt (2005, p.245) state "A Hausman–Wu test does not reject the null hypothesis of exogeneity (p value = 0.06). ... Given these results ... we treat volume as exogenous hereafter. " and Bedard and Deschênes (2006, p.189) state "... the Jerry R. Hausman (1978) test, testing the null hypothesis that the difference between TSLS and OLS coefficients is due only to sampling error, is rejected at the 5–percent level." However, many times the pretest nominal size β is not even reported in applied papers. E.g. Blonigen (1997, p.453) states "A Hausman test indicated that the

In Table 1a) we list null rejection probabilities of the symmetric two–stage test and in Table 1b) rejection probabilities of the Hausman pretest. Results for the following 30 parameter combinations are reported:

$$(p, q, \sigma_u^2) \in \{.3, .6, .9\} \times \{0, .3, .4, .5, .6\} \times \{1, 5\}. \tag{13}$$

The possible choices for p and q translate into a grid of values for γ_1 in the interval [0,.74] and the two choices for σ_u^2 translate into the values .63 and 1.41 for γ_{22}.

Insert Table 1 here

Table 1a) reveals that the two–stage test overrejects severely. At the nominal size of 5%, the null rejection rates reach values higher than 80% for the parameter combinations considered here. When $q = 0$, i.e. when c_i and the regressors x_{i1} and x_{i2} are independent, the null rejection probabilities are relatively close to the nominal size of the test and fall into the interval [6.2%,8.6%] over the parameter combinations reported here. In this case, inference based on both T_{RE} and T_{FE} is justified. However, when q is nonzero and thus c_i and the regressors x_{i1} and x_{i2} are correlated, only inference based on T_{FE} is justified but inference based on T_{RE} is size–distorted. The two–stage test overrejects because the Hausman pretest does not reject frequently enough (as documented in Table 1b) in these cases but the resulting inference based on T_{RE} in the second stage leads to very frequent rejections. For example, when $p = q = .6$ and $\sigma_u^2 = 5$ the Hausman pretest only rejects in 45.7% of the cases. The resulting frequent use of T_{RE} in the second stage leads to the 51.6% rejection rate of the two–stage test.

The simulations reveal that overrejection increases in p and, most of the times, also in σ_u^2. Picking p close to 1 and/or large enough values for σ_u^2, the null rejection probability of the two–stage test can be made arbitrarily close to 100%. This is consistent with the theoretical results reported in the next section. Such large values of p are probably not empirically relevant, but the important message from the simulations reported here is that severe size distortion of the two–stage test also occurs for empirically relevant choices of the parameters.

---

random effects model estimates are consistent for these data, and thus I report only the more efficient random effects model estimates." and Banerjee and Iyer (2005, p.1205) state "A Hausman test does not reject the null hypothesis that the OLS and IV coefficients are equal." As further indication, that small values of β are common in applied work, consider Bradford (2003, p.1757) that states that the Hausman statistic "which is distributed as a chi–square with two degrees of freedom under the null is calculated at 1.46. This fails to reject the null at any reasonable level of significance. Consequently, these two variables are treated as exogenous regressors hereafter." Note that the p–value in this case is .48. Thus, choices of β of that magnitude are considered unreasonable.

When calculating the estimators $\widehat{\sigma}_u^2$, $\widehat{\sigma}_c^2$, and $\widetilde{\sigma}_u^2$, we use the degrees of freedom adjustment $K = 1$, see Subsection 5.1 below.

In contrast to the size–distorted two–stage procedure, the simple one–stage test that always uses the test statistic T_{FE} has very reliable null rejection probabilities. Over all the parameter combinations in (13), the null rejection probabilities of the one–stage test fall into the interval [5.0%,5.3%]. Note that the corresponding interval of the two–stage test is [5.3%,80.8%]. Therefore, if controlling the size is an objective, use of the two–stage procedure can not be recommended.

4 Asymptotic Size of a Test

In subsection 4.1, the theoretical background is discussed of how to determine the asymptotic size of a fixed critical value test in a situation where the test statistic that the testing procedure is based on, has an asymptotic distribution that may be discontinuous in certain nuisance parameters. This theory is taken from AG(2005a) and illustrated in a motivational example, namely inference in a simple version of the linear instrumental variables model. Then, in subsection 4.2, this theory is applied to the two–stage test with a Hausman pretest in the first stage.

4.1 General Theory and Motivation

Consider a general testing problem of nominal size α with test statistic $T_N(\theta_0)$ and nonrandom critical value $c_\infty(1-\alpha)$. Assume the model contains a possibly infinite–dimensional nuisance parameter vector $\gamma \in \Gamma$. Then, by definition, the asymptotic size of the test of $H_0 : \theta = \theta_0$ equals

$$AsySz(\theta_0) = \limsup_{N\to\infty} \sup_{\gamma\in\Gamma} P_{\theta_0,\gamma}(T_N(\theta_0) > c_\infty(1-\alpha)), \tag{14}$$

where $P_{\theta,\gamma}(\cdot)$ denotes probability when the true parameters are (θ, γ). Uniformity over $\gamma \in \Gamma$ which is built into the definition of $AsySz(\theta_0)$ is crucial for the asymptotic size to give a good approximation for the finite sample size. For illustration, consider the following example.

Example:[5] Consider the simple model given by a structural equation and a reduced–form equation $y_1 = y_2\theta + u$, $y_2 = z\pi + v$, where $y_1, y_2, z \in R^N$ and $\theta, \pi \in R$ are unknown parameters. Assume $\{(u_i, v_i, z_i) : i \leq N\}$ are i.i.d. with distribution F, where a subscript i denotes the i–th component of a vector. To test $H_0 : \theta = \theta_0$ against a two–sided alternative say, the t–statistic $T_N(\theta_0) = |N^{1/2}(\widehat{\theta}_N - \theta_0)/\widehat{\sigma}_N|$ and critical value $c_\infty(1-\alpha) = z_{1-\alpha/2}$ is used, where

[5] See AG(2005c) for the general treatment of this example

$\widehat{\theta}_N = (y_2' P_z y_2)^{-1} y_2' P_z y_1$, $\widehat{\sigma}_N = \widehat{\sigma}_u (N^{-1} y_2' P_z y_2)^{-1/2}$, and $\widehat{\sigma}_u^2 = (N-1)^{-1}(y_1 - y_2\widehat{\theta}_N)'(y_1 - y_2\widehat{\theta}_N)$. The nuisance parameter vector γ equals (F, π), where certain restrictions are imposed on F, such as conditional homoskedasticity, exogeneity of the instrument, and existence of second moments.

Following AG (2005a), the parameter γ is decomposed into 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, γ_1. The parameter space of γ_1 is Γ_1. The second component, γ_2, of γ also affects the limit distribution of the test statistic, but does not affect the distance of the parameter γ to the point of discontinuity. The parameter space of γ_2 is Γ_2. The third component, γ_3, of γ does not affect the limit distribution of the test statistic. The parameter space for γ_3 is $\Gamma_3(\gamma_1, \gamma_2)$, which generally may depend on γ_1 and γ_2. The parameter space Γ for γ satisfies

Assumption A. (i)

$$\Gamma = \{(\gamma_1, \gamma_2, \gamma_3) : \gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2 \subset R^q, \gamma_3 \in \Gamma_3(\gamma_1, \gamma_2)\} \tag{15}$$

and (ii) $\Gamma_1 = \lfloor \gamma_1^\ell, \gamma_1^u \rfloor$ for some $-\infty \leq \gamma_1^\ell < \gamma_1^u \leq \infty$ that satisfy $\gamma_1^\ell \leq 0 \leq \gamma_1^u$, where $\lfloor$ denotes the left endpoint of an interval that may be open or closed at the left end and $\rfloor$ is defined analogously for the right endpoint.

Example (continued): Decompose the nuisance parameter into $\gamma = (\gamma_1, \gamma_2, \gamma_3)$, where $\gamma_1 = |(E_F z_i^2)^{1/2}\pi/\sigma_v|$, $\gamma_2 = \rho$, and $\gamma_3 = (F, \pi)$, where $\sigma_v^2 = E_F v_i^2$, $\sigma_u^2 = E_F u_i^2$, and $\rho = Corr_F(u_i, v_i)$. The parameter spaces for γ_1 and γ_2 are $\Gamma_1 = R_+$ and $\Gamma_2 = [-1, 1]$. The details for the restrictions on the parameter space $\Gamma_3 = \Gamma_3(\gamma_1, \gamma_2)$ for γ_3 are given in AG(2005c) and are such that the following CLT holds under sequences $\gamma = \gamma_N$ for which $\gamma_2 = \gamma_{2,N} \to h_2$:

$$\begin{pmatrix} (N^{-1}z'z)^{-1/2} N^{-1/2} z'u/\sigma_u \\ (N^{-1}z'z)^{-1/2} N^{-1/2} z'v/\sigma_v \end{pmatrix} \to_d \begin{pmatrix} \psi_{u,h_2} \\ \psi_{v,h_2} \end{pmatrix} \sim N(0, \begin{pmatrix} 1 & h_2 \\ h_2 & 1 \end{pmatrix}). \tag{16}$$

In this example, the asymptotic distribution of the statistic $T_N(\theta_0)$ has a discontinuity at $\gamma_1 = 0$. Under different sequences $\gamma_1 = \gamma_{1,N}$ such that $\gamma_{1,N} \to 0$, the limit distribution of $T_N(\theta_0)$ may be different. More precisely, denote by $\gamma_{N,h}$ a sequence of nuisance parameters $\gamma = \gamma_N$ such that $N^{1/2}\gamma_1 \to h_1$ and $\gamma_2 \to h_2$ and $h = (h_1, h_2)$. It is shown below that under $\gamma_{N,h}$, the limit distribution of $T_N(\theta_0)$ depends on h_1 and h_2 and only on h_1 and h_2. As long as h_1 is finite, the sequence γ_1 converges to zero, yet the limit distribution of $T_N(\theta_0)$ does not only depend on the limit point 0 of γ_1, but depends on how precisely γ_1 converges to zero, indexed by the convergence speed $N^{1/2}$ and the localization parameter h_1. In contrast, the limit distribution of $T_N(\theta_0)$ only depends on the limit point h_2 of γ_2 but not on how γ_2 converges to h_2. In that sense, the limit distribution is discontinuous in γ_1 at 0, but continuous on Γ_2 in γ_2. The parameter γ_3 does not influence the limit distribution of $T_N(\theta_0)$ by virtue of the CLT in (16).

If $h_1 < \infty$, it is shown in AG(2005c) that under $\gamma_{N,h}$

$$(17) \qquad \begin{pmatrix} y_2' P_z u/(\sigma_u \sigma_v) \\ y_2' P_z y_2/\sigma_v^2 \\ \widehat{\sigma}_u^2/\sigma_u^2 \end{pmatrix} \to_d \begin{pmatrix} \xi_{1,h} \\ \xi_{2,h} \\ \eta_{u,h}^2 \end{pmatrix} = \begin{pmatrix} (\psi_{v,h_2} + h_1)\psi_{u,h_2} \\ (\psi_{v,h_2} + h_1)^2 \\ (1 - h_2\xi_{1,h}/\xi_{2,h})^2 + (1 - h_2^2)\xi_{1,h}^2/\xi_{2,h}^2 \end{pmatrix}$$

and thus $T_N(\theta_0) \to_d J_h$, where J_h is the distribution of $|\xi_{1,h}/(\xi_{2,h}\eta_{u,h}^2)^{1/2}|$. If $h_1 = \infty$, $T_N(\theta_0) \to_d J_h$, where in this case J_h is the distribution of the absolute value of a standard normal random variable independent of h_2.

Formalizing the additional aspects of the example, we now define the index set for the different asymptotic null distributions of the test statistic $T_N(\theta_0)$ of interest. Let

$$(18) \qquad \begin{aligned} H = \{h = (h_1, h_2) \in R_\infty^{1+q} : \exists \; \{\gamma_N = (\gamma_{N,1}, \gamma_{N,2}, \gamma_{N,3}) \in \Gamma : N \geq 1\} \\ \text{such that } N^r \gamma_{N,1} \to h_1 \text{ and } \gamma_{N,2} \to h_2\}. \end{aligned}$$

Definition of $\{\gamma_{N,h} : N \geq 1\}$**:** Given $r > 0$ and $h = (h_1, h_2) \in H$, let $\{\gamma_{N,h} = (\gamma_{N,h,1}, \gamma_{N,h,2}, \gamma_{N,h,3}) : N \geq 1\}$ denote a sequence of parameters in Γ for which $N^r\gamma_{N,h,1} \to h_1$ and $\gamma_{N,h,2} \to h_2$. In the example, $r = 1/2$ and $H = R_{+,\infty} \times [-1, 1]$. The sequence $\{\gamma_{N,h} : N \geq 1\}$ is defined such that under $\{\gamma_{N,h} : N \geq 1\}$, the asymptotic distribution of $T_N(\theta_0)$ depends on h and only h. This is formalized in the following assumption and has already been illustrated in the above example.

Assumption B. For some $r > 0$, all $h \in H$, all sequences $\{\gamma_{N,h} : N \geq 1\}$, and some distributions J_h, $T_N(\theta_0) \to_d J_h$ under $\{\gamma_{N,h} : N \geq 1\}$.

The next theorem, a special case of Theorem 1(a) in AG(2005a), provides a formula for $AsySz$. In contrast to the formula of $AsySz$ in (14), the formula in the theorem can be used for explicit calculation. It shows that the "worst case" sequence of nuisance parameters, a sequence that yields the highest asymptotic null rejection probability, is of the type $\{\gamma_{N,h} : N \geq 1\}$.

Theorem 1 (*AG(2005a)*) *Suppose Assumptions* A *and* B *hold where* $J_h : R \to [0,1]$ *is a continuous function. Then,* $AsySz(\theta_0) = \sup_{h \in H}[1 - J_h(c_\infty(1 - \alpha))]$.

Example (continued): Here $AsySz(\theta_0) = \sup_{h \in R_{+,\infty} \times [-1,1]}[1 - J_h(z_{1-\alpha/2})]$, where for $h_1 < \infty$, J_h is the distribution of $|\xi_{1,h}/(\xi_{2,h}\eta_{u,h}^2)^{1/2}|$ and for $h_1 = \infty$, J_h is the distribution of the absolute value of a standard normal random variable. The asymptotic size can be easily calculated by simulation of J_h over a fine grid over vectors h in H.

4.2 Asymptotic Size After Hausman Pretest

In this subsection, we return to the panel data application and the two–stage test with a Hausman pretest in the first stage introduced in (10). To apply Theorem 1, we have to determine the decomposition of the nuisance parameter vector γ into "discontinuous" and "continuous" elements, the rate of convergence r, and verify Assumptions A and B. Finally, one needs to derive the limiting distribution J_h of the test statistic $T_N(\theta_0)$ under $\{\gamma_{N,h}\}$.

Let $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ with $\gamma_2 = (\gamma_{21}, \gamma_{22})$,

$$\gamma_1 = Corr_F(c_i, \overline{x}_i),\ \gamma_{21} = (\frac{T\sigma^2_{\overline{x}_i}}{E_F||x_i||^2})^{1/2},\ \gamma_{22} = (T\sigma^2_c/\sigma^2_u)^{1/2}, \tag{19}$$

and $\gamma_3 = (F, \lambda)$. The component γ_{21} measures the expected time variation in the regressor while the component γ_{22} is a function of the ratio of the variances of the individual specific effect and the error term u_{it}. In all the examples studied in AG(2005a–e), γ_2 is one–dimensional and this is the only example where the γ_2–component is two–dimensional. In particular, Guggenberger (2007) studies the asymptotic size of the two–stage test in the linear IV context where the Hausman pretest is used as a test of exogeneity of a regressor. The γ_2–component there is scalar and is a function of the concentration parameter. Here, the situation is more complex and two separate parameters impact the asymptotic size of the two–stage test through γ_2.

The parameter space Γ of γ is defined as in (15) with $q = 2$, $\Gamma_2 = \Gamma_{21} \times \Gamma_{22}$ and

$$\Gamma_1 = [-1, 1],\ \Gamma_{21} = [\kappa_1, \overline{\kappa}_1],\ \text{and}\ \Gamma_{22} = [\kappa_2, \overline{\kappa}_2] \tag{20}$$

for some $0 < \kappa_1 < \overline{\kappa}_1 < 1$ and $0 < \kappa_2 < \overline{\kappa}_2 < \infty$. Let

$$\begin{aligned}
&\Gamma_3(\gamma_1, \gamma_2) = \{(F, \lambda) : \lambda \in R;\ E_F x_{it} = E_F c_i = E_F u_{it} = 0,\\
&\ E_F u^2_{it} = \sigma^2_u,\ E_F c^2_i = \sigma^2_c \text{ for some finite } \sigma^2_u, \sigma^2_c > 0,\\
&\ Corr_F(c_i, \overline{x}_i) = \gamma_1,\ (T\sigma^2_{\overline{x}_i}/E_F||x_i||^2)^{1/2} = \gamma_{21},\ (T\sigma^2_c/\sigma^2_u)^{1/2} = \gamma_{22},\\
&\ E_F x_{it}u_{is} = E_F x_{it}x_{is}u_{iv}c_i = E_F c_i u_{it} = 0,\ E_F(\overline{x}^2_i c^2_i) = \sigma^2_{\overline{x}_i}\sigma^2_c + 2(E_F\overline{x}_i c_i)^2,\\
&\ E_F x_{it}x_{is}u_{iv}u_{iw} = E_F x_{it}x_{is}E_F u_{iv}u_{iw}, \left\|E_F\left(|\overline{x}_i/\sigma_{\overline{x}_i}|^{2+\delta},\ (||x_i||^2/E_F||x_i||^2)^{1+\delta},\right.\right.\\
&\ \left.\left.|x_i'u_i/(\sigma_{\overline{x}_i}\sigma_u)|^{2+\delta},\ |\overline{x}_i\overline{u}_i/(\sigma_{\overline{x}_i}\sigma_u)|^{2+\delta},\ |\overline{x}_i c_i/(\sigma_{\overline{x}_i}\sigma_c)|^{2+\delta},\ (u^2_{it}/\sigma^2_u)^{1+\delta}\right)\right\| \leq M\}
\end{aligned} \tag{21}$$

for some constants $\delta > 0$, $M < \infty$, and subindices $t, s, v, w = 1, ..., T$. The condition $E_F(\overline{x}^2_i c^2_i) = E_F\overline{x}^2_i E_F c^2_i + 2(E_F\overline{x}_i c_i)^2$ holds, for example, if $\overline{x}_i$ and c_i are jointly normal. The remaining conditions are moment restrictions, that imply that Liapounov–type central limit theorems (CLT) or weak law of large numbers (WLLN) for independent $L^{1+\delta}$–bounded random variables hold, and the conditional homoskedasticity–type assumption $E_F x_{it} x_{is} u_{iv} u_{iw} = E_F x_{it} x_{is} E_F u_{iv} u_{iw}$. With the above definitions, Assumption A clearly holds.

For H defined as in (18) it follows that

$$H = H_1 \times H_2 = R_\infty \times [\kappa_1, \overline{\kappa}_1] \times [\kappa_2, \overline{\kappa}_2]. \tag{22}$$

For every $h = (h_1, h_2) \in H$, denote by $\{\gamma_{N,h}\} \subset \Gamma$ a sequence of parameters with components $\gamma_{N,h,1}$, $\gamma_{N,h,2} = (\gamma_{N,h,21}, \gamma_{N,h,22})$, and $\gamma_{N,h,3}$, $\gamma_{N,h} = (\gamma_{N,h,1}, \gamma_{N,h,2}, \gamma_{N,h,3})$, where

$$\begin{aligned} \gamma_{N,h,1} &= Corr_{F_N}(c_i, \overline{x}_i),\ \gamma_{N,h,21} = (\frac{TE_{F_N}\overline{x}_i^2}{E_{F_N}||x_i||^2})^{1/2},\ \gamma_{N,h,22} = (TE_{F_N}c_i^2/E_{F_N}u_{it}^2)^{1/2} \text{ s.t.} \\ N^{1/2}\gamma_{N,h,1} &\to h_1,\ \gamma_{N,h,2} \to h_2,\ \text{and } \gamma_{N,h,3} = (F_N, \lambda_N) \in \Gamma_3(\gamma_{N,h,1}, \gamma_{N,h,2}). \end{aligned} \tag{23}$$

In the Appendix, for every $h \in H$ we derive the limit distribution J_h of the test statistic $T_N(\theta_0)$ under the sequence $\{\gamma_{N,h}\}$, see (51). This verifies Assumption B with $r = 1/2$.

Then, applying Theorem 1 the asymptotic size of the two–stage test equals

$$AsySz(\theta_0) = \sup_{h \in H}[1 - J_h(c_\infty(1-\alpha))] \tag{24}$$

with H defined in (22) and J_h defined in (51). The formula applies to upper, lower one–sided, and symmetric two–sided versions of the test with $c_\infty(1-\alpha) = z_{1-\alpha}$, $z_{1-\alpha}$, and $z_{1-\alpha/2}$, respectively. Note that $AsySz(\theta_0)$ depends on α, β, and on the boundaries in the definition of Γ_2. For notational simplicity, this dependence is suppressed.

Figure 2 plots the asymptotic maximal rejection probability of the symmetric two–stage test, where the maximum is taken over $h_1 \in H_1$, as a function of h_{21} and h_{22}, i.e. the figure plots the function

$$f(h_{21}, h_{22}) = \sup_{h_1 \in H_1} [1 - J_{(h_1, h_{21}, h_{22})}(c_\infty(1-\alpha))] \tag{25}$$

for $\alpha = \beta = .05$.[6] For small values of h_{21}, $f(h_{21}, h_{22})$ is close to the nominal size 5% of the test. For $h_{21} \leq .4$, $f(h_{21}, h_{22}) < .1$. The size distortion increases as h_{21} increases and the asymptotic maximal null rejection probability gets arbitrarily close to 1 as $h_{21} \to 1$ (as documented in

[6]For each h, the results are based on $R = 30,000$ random draws from the distribution of J_h. We consider h_1 values in [-2000,2000] using a grid with stepsize .01 on [0,.1], stepsize .1 on [.1,1] stepsize 1 on [1,10], stepsize 10 on [10,100], and stepsize 50 on [100,2000] and the analogous grid for negative h_1 values.

additional simulations). For fixed h_{21}, the function $f(h_{21}, h_{22})$ decreases as h_{22} increases and as $h_{22} \to \infty$ it decreases to the nominal size of the test. However, the slope of the function $f(h_{21}, \cdot)$ is rather small and it takes rather large values of h_{22} to make $f(h_{21}, h_{22})$ small when h_{21} is close to 1. For example, $f(.95, h_{22})$ equals 63.7, 40.0, 25.9, 16.9, 13.0, 10.8, and 9.4% for $h_{22} = 1, 2, ..., 7$.

Insert Figure 2 here

What is the reason for the size distortion? It is shown in (47) that $H_N \to_d \chi_1^2(h_1^2 \overline{h}(\overline{h}+1)^{-1})$ for $\overline{h} = h_{22}^2(1 - h_{21}^2)$ under $\gamma_{N,h}$, where $\chi_1^2(\cdot)$ denotes a noncentral chi–square distribution with one degree of freedom with noncentrality parameter given by the expression in brackets. If $h_1 \overline{h} \neq 0$, the Hausman pretest has nonzero local power. However, the noncentrality parameter $h_1^2 \overline{h}(\overline{h}+1)^{-1}$ of the limiting distribution in (47) is small when $\overline{h}$ is small which is the case if h_{21} is close to 1 or if h_{22} is close to 0. In these cases, the pretest has poor power properties and the two–stage test frequently uses inference based on T_{RE} in the second stage. But the test based on T_{RE} tends to reject frequently under moderate failures of the pretest hypothesis (7) which leads to size distortion of the two–stage test. The parameter h_{21} is close to 1when there is little time variation in the regressor, i.e. in the extreme case where $x_{it} = x_{is}$ for all $s, t = 1, ..., T$, $h_{21} = 1$. In the case where x_{it}, $t = 1, ..., T$ is i.i.d., $h_{21} = T^{-1/2}$; for example, $h_{21} = .71$ and $h_{21} = .58$ when $T = 2$ or 3, respectively. Note that if $\overline{\kappa}_1 = .71$, the simulations for Figure 2 show that $AsySz(\theta_0)$ is about 30% if κ_2 is small. So, even in the case where the regressor x_{it} is uncorrelated for different time indices t, the two–stage test is extremely size distorted.[7] The parameter h_{22} is small when the ratio of the variances of the individual specific effect and of the error term is small.

Insert Table 3 here

Table 3 reports conditional rejection probabilities of the symmetric two–stage test, conditional on the Hausman pretest rejecting the pretest null hypothesis, $R - C - R = P(T_N(\theta_0) > c_\infty(1-\alpha)|H_N > \chi^2_{1,1-\beta})$, and conditional on the pretest not rejecting the pretest null hypothesis, $R - C - NR = P(T_N(\theta_0) > c_\infty(1 - \alpha)|H_N < \chi^2_{1,1-\beta})$, when $\alpha = \beta = .05$ and $h_1 = 15$ for a grid of h_{21} and h_{22} values. Table 3 also reports rejection probabilities of the Hausman pretest. Even though $h_1 = 15$ is quite large, these latter rejection probabilities can be quite small, especially when h_{21} is close to one and/or h_{22} is close to 0. This is consistent with the local power result of the Hausman pretest described in the previous paragraph because when h_{21} is close to one and/or h_{22} is close to 0 then the noncentrality parameter $h_1^2 \overline{h}(\overline{h} + 1)^{-1}$ is close to

[7] Values of h_{21} close to zero are possible if x_{it} is negatively correlated, a case which is probably of lesser importance in applied work.

0. For example, when $h_{21} = .75$ and $h_{22} = .1$, then the Hausman pretest rejects in 16.9% of the cases. However, in cases where the Hausman pretest does not reject – despite the fact that h_1 is 15– the rejection probability in the second stage can be very high. This is because then in the second stage inference is based on T_{RE} which takes on relatively large values when h_1 is nonzero. For example, in the case $h_{21} = .75$ and $h_{22} = .2$, conditional on the Hausman pretest not rejecting (which happens in 49.5% of the cases), the test rejects with probability 59.1% in the second stage. Perhaps more surprisingly, size distortion of the two–stage test is also caused by the two–stage test rejecting at high frequency conditional on the Hausman pretest rejecting in the first stage. This is despite the fact, that then in the second stage inference is based on the statistic T_{FE} and the unconditional size of the one–stage test based on T_{FE} is $\alpha = .05$. For example, in the case $h_{21} = .75$ and $h_{22} = .05$, conditional on the Hausman pretest rejecting (which happens in 8.0% of the cases), the test rejects with probability 26.0% in the second stage. The reason for this overrejection is that the Hausman statistic and the t–statistic T_{FE} are correlated and if the former statistic takes on large values (and therefore the pretest hypothesis is rejected and the two–stage test is based on T_{FE} in the second stage) the latter statistic is likely to take on a large value too. This correlation increases as h_{21} approaches one and/or h_{22} approaches zero.[8]

5 Appendix

The Appendix provides possible choices for the variance estimators $\widehat{\sigma}^2_u$, $\widehat{\sigma}^2_c$, and $\widetilde{\sigma}^2_u$ and contains the derivation of the asymptotic distribution of the two–stage test statistic under sequences $\{\gamma_{N,h}\}$.

5.1 Possible choices for $\widehat{\sigma}^2_u$, $\widehat{\sigma}^2_c$, and $\widetilde{\sigma}^2_u$

Following Wooldridge (2002, p.260 and 271), one alternative to define the variance estimators is as follows. Let

$$(\widehat{\lambda}_{OLS}, \widehat{\theta}_{OLS})' = \left(\sum_{i=1}^{N} w_i' w_i\right)^{-1} \sum_{i=1}^{N} w_i' y_i \tag{26}$$

be the pooled OLS estimator of $(\gamma, \theta)'$ and

$$\widehat{v}_{it} = y_{it} - w_{it}'(\widehat{\lambda}_{OLS}, \widehat{\theta}_{OLS})' \tag{27}$$

[8]Note that if we evaluate the limiting distribution $\xi_{H,h}$ of H_n in (45) at $h_{21} = 1$ and $h_{22} = 0$ the result is $\xi^2_{FE,h}$ which is the squared limiting distribution of T_{FE}, derived in (40).

the residuals from pooled OLS regression estimating $v_{it} = c_i + u_{it}$. Then let

$$\begin{aligned}
\widehat{\sigma}_v^2 &= (NT-K)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{v}_{it}^2,\\
\widehat{\sigma}_c^2 &= (NT(T-1)/2-K)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\widehat{v}_{it}\widehat{v}_{is},\\
\widehat{\sigma}_u^2 &= \widehat{\sigma}_v^2-\widehat{\sigma}_c^2 \qquad (28)
\end{aligned}$$

for $K = 0$ or 2 depending on whether a degrees–of–freedom correction is desired. We can also estimate σ_u^2 based on the fixed effects estimator $\widehat{\theta}_{FE}$ in (5). Let $\overline{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$ and let

$$\widetilde{u}_{it} = (y_{it}-\overline{y}_i)-(x_{it}-\overline{x}_i)\widehat{\theta}_{FE} \qquad (29)$$

be the fixed effects residuals estimating u_{it} and define the estimator

$$\widetilde{\sigma}_u^2 = (N(T-1)-K)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{u}_{it}^2. \qquad (30)$$

For the asymptotic results below, the specific choice of variance estimators $\widehat{\sigma}_u^2$ or $\widetilde{\sigma}_u^2$ does not matter as long as they are consistent in the sense that $\widehat{\sigma}_u^2/\sigma_u^2 \to_p 1$ and $\widetilde{\sigma}_u^2/\sigma_u^2 \to_p 1$ under sequences $\{\gamma_{N,h}\}$ with finite h_1.

5.2 Derivation of J_h

In this subsection, the limit distribution of the test statistic $T_N(\theta_0)$ is derived under sequences $\{\gamma_{N,h}\}$. Two cases are dealt with separately. Case I has $|h_1| < \infty$ while Case II has $|h_1| = \infty$. In Case I, $\gamma_1 \to 0$. Recall that if F_N is the true distribution, then $\sigma_c^2 = E_{F_N}c_i^2$, $\sigma_u^2 = E_{F_N}u_{it}^2$, and $\sigma_{\overline{x}_i}^2 = E_{F_N}\overline{x}_i^2$. Under any sequence $\{\gamma_{N,h}\}$ for which $\gamma_{N,h,1} = Corr_{F_N}(c_i,\overline{x}_i) \to \overline{\gamma}_1$,

(31)

$$\begin{pmatrix} (\sigma_{\overline{x}_i}^2\sigma_c^2)^{-1/2}N^{-1/2}\sum_{i=1}^{N}(\overline{x}_ic_i - E_{F_N}\overline{x}_ic_i)\\ (\sigma_{\overline{x}_i}^2\sigma_u^2/T)^{-1/2}N^{-1/2}\sum_{i=1}^{N}\overline{x}_i\overline{u}_i\\ (E_{F_N}||x_i||^2\sigma_u^2)^{-1/2}N^{-1/2}\sum_{i=1}^{N}x_i'u_i \end{pmatrix} \to_d \begin{pmatrix} \psi_{c,\overline{\gamma}_1}\\ \psi_{\overline{u},h_{21}}\\ \psi_{u,h_{21}} \end{pmatrix} \sim N(0, \begin{pmatrix} 1+\overline{\gamma}_1^2 & 0 & 0\\ 0 & 1 & h_{21}\\ 0 & h_{21} & 1 \end{pmatrix}).$$

The result holds by the Liapounov CLT for independent, mean zero, $L^{2+\delta}$–bounded random variables using the moment restrictions in (21) noting that $E_{F_N}\overline{x}_i^2\overline{u}_i^2 = \sigma_{\overline{x}_i}^2\sigma_u^2/T$ and $E_{F_N}(x_i'u_i)^2 = E_{F_N}||x_i||^2\sigma_u^2$. In particular, the condition $E_{F_N}(\overline{x}_i^2c_i^2) = \sigma_{\overline{x}_i}^2\sigma_c^2 + 2(E_{F_N}\overline{x}_ic_i)^2$ from (21) yields

$$E_{F_N}(\overline{x}_ic_i - E_{F_N}\overline{x}_ic_i)^2 = E_{F_N}(\overline{x}_ic_i)^2 - (E_{F_N}\overline{x}_ic_i)^2 = \sigma_{\overline{x}_i}^2\sigma_c^2 + (E_{F_N}\overline{x}_ic_i)^2 \qquad (32)$$

and thus $(\sigma^2_{\overline{x}_i}\sigma^2_c)^{-1}E_{F_N}(\overline{x}_ic_i - E_{F_N}\overline{x}_ic_i)^2 = 1 + Corr_{F_N}(c_i, \overline{x}_i)^2$. The limiting distribution $\psi_{c,\overline{\gamma}_1}$ is independent of $\psi_{\overline{u},h_{21}}, \psi_{u,h_{21}}$ because of the conditions $E_{F_N}x_{it}u_{is} = E_{F_N}x_{it}x_{is}u_{iv}c_i = 0$ for $t, s, v = 1, ..., T$. Finally, the covariance between $\psi_{\overline{u},h_{21}}$ and $\psi_{u,h_{21}}$ is h_{21}. This holds because $E_{F_N}\overline{x}_i\overline{u}_ix_i'u_i = E_{F_N}u_i^2E_{F_N}\overline{x}_i^2$ which holds because $E_{F_N}x_{it}x_{is}u_{iv}u_{iw} = E_{F_N}x_{it}x_{is}E_{F_N}u_{iv}u_{iw}$ for $t, s, v, w = 1, ..., T$.

Next, the joint limiting distribution of the t–statistics and the Hausman statistic are derived. We first assume that Ω and σ_u^2 are known and replace $\widehat{\Omega}$ and $\widetilde{\sigma}_u^2$ by Ω and σ_u^2 in the test statistics. We then show that this modification does not matter asymptotically. The relevant ingredients in the t–statistics and Hausman statistic are the expressions $\sum_{i=1}^N x_i'\Omega^{-1}v_i$, $\sum_{i=1}^N x_i'M_{1_T}u_i$, $\sum_{i=1}^N x_i'\Omega^{-1}x_i$, and $\sum_{i=1}^N x_i'M_{1_T}x_i$, where

$$v_i = u_i + 1_Tc_i. \tag{33}$$

We first derive the appropriate normalizations of these expressions and their limit expressions.

We first consider Case I in which case $\overline{\gamma}_1^2 = 0$. Note that

$$\Omega^{-1} = \sigma_u^{-2}I_T - \sigma 1_T1_T', \text{ for } \sigma = \sigma_c^2\sigma_u^{-2}(\sigma_u^2 + \sigma_c^2T)^{-1}. \tag{34}$$

Then,

$$\begin{aligned} &\textstyle\sum_{i=1}^N x_i'\Omega^{-1}v_i \\ &= \textstyle\sum_{i=1}^N x_i'\Omega^{-1}u_i + \sum_{i=1}^N(x_i'\Omega^{-1}1_Tc_i - E_{F_N}x_i'\Omega^{-1}1_Tc_i + E_{F_N}x_i'\Omega^{-1}1_Tc_i) \\ &= \textstyle\sigma_u^{-2}\sum_{i=1}^N x_i'u_i - T^2\sigma\sum_{i=1}^N \overline{x}_i\overline{u}_i + (\sigma_u^2 + \sigma_c^2T)^{-1}T\sum_{i=1}^N(\overline{x}_ic_i - E_{F_N}\overline{x}_ic_i + E_{F_N}\overline{x}_ic_i) \end{aligned} \tag{35}$$

and thus by (31)

$$T^{-1}\sigma_{\overline{x}_i}^{-1}\sigma_c^{-1}(\sigma_u^2+\sigma_c^2T)N^{-1/2}\textstyle\sum_{i=1}^N x_i'\Omega^{-1}v_i \to_d (h_{22}+h_{22}^{-1})h_{21}^{-1}\psi_{u,h_{21}} - h_{22}\psi_{\overline{u},h_{21}} + \psi_{c,0} + h_1. \tag{36}$$

Furthermore, because $\sum_{i=1}^N x_i'M_{1_T}u_i = \sum_{i=1}^N(x_i'u_i - T\overline{x}_i\overline{u}_i)$,

$$(E_{F_N}||x_i||^2\sigma_u^2)^{-1/2}N^{-1/2}\textstyle\sum_{i=1}^N x_i'M_{1_T}u_i \to_d \psi_{u,h_{21}} - h_{21}\psi_{\overline{u},h_{21}}. \tag{37}$$

Also, because $\sum_{i=1}^N x_i'\Omega^{-1}x_i = \sum_{i=1}^N(\sigma_u^{-2}||x_i||^2 - \sigma T^2\overline{x}_i^2)$ it follows by a WLLN (using the last two lines in (21)) and straightforward calculations that

$$\sigma^{-1}\sigma_{\overline{x}_i}^{-2}T^{-2}N^{-1}\textstyle\sum_{i=1}^N x_i'\Omega^{-1}x_i \to_p (h_{22}^{-2}+1)h_{21}^{-2} - 1. \tag{38}$$

Finally because $\sum_{i=1}^N x_i'M_{1_T}x_i = \sum_{i=1}^N(||x_i||^2 - T\overline{x}_i^2)$ it follows that

$$(E_{F_N}||x_i||^2)^{-1}N^{-1}\textstyle\sum_{i=1}^N x_i'M_{1_T}x_i \to_p 1 - h_{21}^2. \tag{39}$$

Results (37), (39) and the continuous mapping theorem immediately imply that

$$(40)\qquad T_{FE}(\theta_0) = (\sigma_u^2 N^{-1}\sum_{i=1}^N x_i' M_{1_T} x_i)^{-1/2} N^{-1/2}\sum_{i=1}^N x_i' M_{1_T} u_i \to_d \xi_{FE,h} \sim \frac{\psi_{u,h_{21}} - h_{21}\psi_{\overline{u},h_{21}}}{(1-h_{21}^2)^{1/2}}.$$

Note that the distribution $\xi_{FE,h}$ does not depend on h_1 and we extend the definition of $\xi_{FE,h}$ in (40) to the case $|h_1| = \infty$.

To derive the limit distribution of $T_{RE}(\theta_0)$, first note that by partitioned regression, it follows that

$$(41)\qquad N^{1/2}(\widehat{\theta}_{RE} - \theta) = \frac{N^{-1/2}[\sum_{i=1}^N x_i'\Omega^{-1}v_i - \sum_{i=1}^N x_i'\Omega^{-1}1_T(\sum_{i=1}^N 1_T'\Omega^{-1}1_T)^{-1}\sum_{i=1}^N 1_T'\Omega^{-1}v_i]}{N^{-1}[\sum_{i=1}^N x_i'\Omega^{-1}x_i - \sum_{i=1}^N x_i'\Omega^{-1}1_T(\sum_{i=1}^N 1_T'\Omega^{-1}1_T)^{-1}\sum_{i=1}^N 1_T'\Omega^{-1}x_i].}$$

By (36) and (38) the normalizations for the numerator and denominator in (41) are $T^{-1}\sigma_{\overline{x}_i}^{-1}\sigma_c^{-1}(\sigma_u^2 + \sigma_c^2 T)$ and $\sigma^{-1}\sigma_{\overline{x}_i}^{-2}T^{-2}$, respectively. However, by straightforward calculations

(42)

$$\sigma^{-1}\sigma_{\overline{x}_i}^{-2}T^{-2}N^{-1}[\sum_{i=1}^N x_i'\Omega^{-1}1_T(\sum_{i=1}^N 1_T'\Omega^{-1}1_T)^{-1}\sum_{i=1}^N 1_T'\Omega^{-1}x_i] = \frac{\sigma_u^2}{T\sigma_c^2}(N^{-1}\sum_{i=1}^N \overline{x}_i/\sigma_{\overline{x}_i})^2 = o_p(1),$$

where the last equality holds because $T^{-1}\sigma_c^{-2}\sigma_u^2 \to h_{22}^{-2}$ and $N^{-1}\sum_{i=1}^N \overline{x}_i/\sigma_{\overline{x}_i} = o_p(1)$ by a WLLN for independent $L^{1+\delta}$–bounded random variables and (21). Using an analogous argument for the numerator in (41), it follows that

$$(43)\quad T\sigma_{\overline{x}_i}\sigma_c\sigma_u^{-2}N^{1/2}(\widehat{\theta}_{RE} - \theta) = T\sigma_{\overline{x}_i}\sigma_c\sigma_u^{-2}\left(N^{-1}\sum_{i=1}^N x_i'\Omega^{-1}x_i\right)^{-1} N^{-1/2}\sum_{i=1}^N x_i'\Omega^{-1}v_i + o_p(1).$$

Therefore, using (43), (36), and (38)

$$(44)\qquad \begin{aligned} T_{RE}(\theta_0) &= \left(N^{-1}\sum_{i=1}^N x_i'\Omega^{-1}x_i\right)^{-1/2} N^{-1/2}\sum_{i=1}^N x_i'\Omega^{-1}v_i + o_p(1) \\ &\to_d \xi_{RE,h} \sim \frac{(h_{22}+h_{22}^{-1})h_{21}^{-1}\psi_{u,h_{21}} - h_{22}\psi_{\overline{u},h_{21}} + \psi_{c,0} + h_1}{((h_{22}^{-2}+1)h_{21}^{-2}-1)^{1/2}(1+h_{22}^2)^{1/2}}. \end{aligned}$$

It is easy to verify that $T_{RE}(\theta_0) \to_d N(0,1)$ when $h_1 = 0$. Finally, using (36)–(39) and straightforward calculations, the limiting distribution of the Hausman statistic under $\{\gamma_{N,h}\}$ with

$|h_1| < \infty$ is given by

$$\begin{aligned} H_N &= \frac{N((\sum_{i=1}^N x_i' M_{1_T} x_i)^{-1} \sum_{i=1}^N x_i' M_{1_T} u_i - (\sum_{i=1}^N x_i' \Omega^{-1} x_i)^{-1} \sum_{i=1}^N x_i' \Omega^{-1} v_i)^2}{(N^{-1} \sum_{i=1}^N x_i' M_{1_T} x_i / \sigma_u^2)^{-1} - (N^{-1} \sum_{i=1}^N x_i' \Omega^{-1} x_i)^{-1}} + o_p(1) \\ &= (TN^{-1} \sum_{i=1}^N \overline{x}_i^2)^{-1} \frac{\sigma_u^2 + \sigma_c^2 T}{\sigma_u^2} \left(T_{FE} (\sigma_u^2 N^{-1} \sum_{i=1}^N x_i' \Omega^{-1} x_i)^{1/2} - T_{RE} (N^{-1} \sum_{i=1}^N x_i' M_{1_T} x_i)^{1/2} \right)^2 + o_p(1) \\ &\to_d \xi_{H,h} \sim (1 + h_{22}^2)[\xi_{FE,h} (h_{21}^{-2} - h_{22}^2 (h_{22}^2 + 1)^{-1})^{1/2} - \xi_{RE,h} (h_{21}^{-2} - 1)^{1/2}]^2, \end{aligned} \tag{45}$$

where the last step holds because

$$(TN^{-1} \sum_{i=1}^N \overline{x}_i^2)^{-1} \sigma_u^2 N^{-1} \sum_{i=1}^N x_i' \Omega^{-1} x_i = (TN^{-1} \sum_{i=1}^N \overline{x}_i^2)^{-1} N^{-1} \sum_{i=1}^N x_i' x_i - T\sigma_c^2 (\sigma_u^2 + \sigma_c^2 T)^{-1},$$

$$(TN^{-1} \sum_{i=1}^N \overline{x}_i^2)^{-1} N^{-1} \sum_{i=1}^N x_i' M_{1_T} x_i = (TN^{-1} \sum_{i=1}^N \overline{x}_i^2)^{-1} N^{-1} \sum_{i=1}^N x_i' x_i - 1, \tag{46}$$

and because $(TN^{-1} \sum_{i=1}^N \overline{x}_i^2)^{-1} N^{-1} \sum_{i=1}^N x_i' x_i \to_p h_{21}^{-2}$ and $T\sigma_c^2 (\sigma_u^2 + \sigma_c^2 T)^{-1} \to h_{22}^2 (h_{22}^2 + 1)^{-1}$. It follows from (45) that the limit distribution of H_N is

$$\chi_1^2 (h_1^2 \overline{h} (\overline{h} + 1)^{-1}) \text{ for } \overline{h} = h_{22}^2 (1 - h_{21}^2) \tag{47}$$

and thus $H_N \to_d \chi_1^2$ if $h_1 = 0$.

In Case II, under sequences $\{\gamma_{N,h}\}$ for which $|h_1| = \infty$ the following limits hold jointly

$$\begin{pmatrix} T_{FE}(\theta_0) \\ H_N \end{pmatrix} \to_d \begin{pmatrix} \xi_{FE,h} \\ \infty \end{pmatrix} \tag{48}$$

and thus in this case, with probability approaching 1, fixed effect inference is conducted in stage two and since $\xi_{FE,h} \sim N(0,1)$ it follows that the asymptotic rejection probability of the two–stage test equals α in this case.

In summary,

$$T_N^*(\theta_0) \to_d J_h^*, \tag{49}$$

under $\{\gamma_{N,h}\}$, where J_h^* is the distribution of the random variable

$$\xi_h^* = \xi_{RE,h} I(\xi_{H,h} \leq \chi_{1,1-\beta}^2) + \xi_{FE,h} I(\xi_{H,h} > \chi_{1,1-\beta}^2) \tag{50}$$

and $\xi_{FE,h}$, $\xi_{RE,h}$, and $\xi_{H,h}$ have been defined in (40), (44), and (45), respectively.

Define $-J_h^*$, and $|J_h^*|$ as the distribution of the random variable $-\xi_h^*$ and $|\xi_h^*|$, respectively. For an upper one–sided, lower one–sided, and symmetric two–sided test, define

$$J_h = J_h^*, -J_h^*, \text{ and } |J_h^*|, \tag{51}$$

respectively. The distribution J_h depends on β but for notational simplicity, this dependence is suppressed. The derivations above imply that Assumption B holds with $r = 1/2$.

To conclude the derivation of the asymptotic distribution of $T_N(\theta_0)$, it has to be verified that replacing $\widehat{\Omega}$ and $\widetilde{\sigma}_u^2$ by Ω and σ_u^2 does not matter asymptotically. We only do so in Case I, Case II can be dealt with analogously. For Case I, it is clear that it is sufficient to show that

$$\begin{aligned}(N^{-1}\sum_{i=1}^N x_i'(\widehat{\Omega})^{-1}x_i)^{-1}N^{-1}\sum_{i=1}^N x_i'\Omega^{-1}x_i &= 1 + o_p(1),\\ (N^{-1/2}\sum_{i=1}^N x_i'\Omega^{-1}v_i)^{-1}N^{-1/2}\sum_{i=1}^N x_i'(\widehat{\Omega})^{-1}v_i &= 1 + o_p(1), \text{ and}\\ \widetilde{\sigma}_u^2/\sigma_u^2 &= 1 + o_p(1).\end{aligned} \tag{52}$$

We verify (52) for the estimators $\widehat{\sigma}_u^2$, $\widehat{\sigma}_c^2$, and $\widetilde{\sigma}_u^2$ defined in Subsection 5.1. We only verify the first of the three conditions, the other conditions are verified analogously. To do so, note that

$$(N^{-1}\sum_{i=1}^N x_i'(\widehat{\Omega})^{-1}x_i)^{-1}N^{-1}\sum_{i=1}^N x_i'\Omega^{-1}x_i = \frac{\widehat{\sigma}_u^2}{\sigma_u^2}\frac{1 - \frac{\sigma_c^2 T}{\sigma_u^2+\sigma_c^2 T}TN^{-1}\sum_{i=1}^N \overline{x}_i^2/(N^{-1}\sum_{i=1}^N x_i'x_i)}{1 - \frac{\widehat{\sigma}_c^2 T}{\widehat{\sigma}_u^2+\widehat{\sigma}_c^2 T}TN^{-1}\sum_{i=1}^N \overline{x}_i^2/(N^{-1}\sum_{i=1}^N x_i'x_i)} \tag{53}$$

and thus it is enough to show that

$$\frac{\widehat{\sigma}_u^2}{\sigma_u^2} = 1 + o_p(1) \text{ and that } \frac{\widehat{\sigma}_c^2}{\sigma_c^2} = 1 + o_p(1), \tag{54}$$

because using $h_2 \in [\kappa_2, \overline{\kappa}_2]$ for $0 < \kappa_2 < \overline{\kappa}_2 < \infty$, the second condition in (54) implies $\sigma_u^{-2}\widehat{\sigma}_c^2 T - h_{22}^2 = o_p(1)$ and thus (54) implies $(\widehat{\sigma}_u^2 + \widehat{\sigma}_c^2 T)^{-1}\widehat{\sigma}_c^2 T - (\sigma_u^2 + \sigma_c^2 T)^{-1}\sigma_c^2 T = o_p(1)$. We only show the first condition $\sigma_u^{-2}\widehat{\sigma}_u^2 = 1 + o_p(1)$, the second condition in (54) can be verified analogously. For notational convenience, assume $K = 0$ in (28). By definition

$$\widehat{\sigma}_u^2 = \widehat{\sigma}_v^2 - \widehat{\sigma}_c^2 = (NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \widehat{v}_{it}^2 - (NT(T-1)/2)^{-1}\sum_{i=1}^N\sum_{t=1}^{T-1}\sum_{s=t+1}^T \widehat{v}_{it}\widehat{v}_{is}, \tag{55}$$

where

$$\widehat{v}_{it} = y_{it} - w_{it}'(\widehat{\lambda}_{OLS}, \widehat{\theta}_{OLS})' = w_{it}'((\lambda, \theta)' - (\widehat{\lambda}_{OLS}, \widehat{\theta}_{OLS})') + c_i + u_{it}. \tag{56}$$

Multiplying out in (55), it follows that all the contributions with a c_i–factor cancel out. For the contributions with only u_{it}–factors we have $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}^2/\sigma_u^2 = 1 + o_p(1)$ by a WLLN for independent $L^{1+\delta}$–bounded random variables and (21) and

$$(NT(T-1)/2)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T} u_{it}u_{is}/\sigma_u^2 = o_p(1) \tag{57}$$

also by the WLLN because $E_{F_N} u_{it}u_{is} = 0$. Finally, the terms involving $w_{it}'((\lambda,\theta)' - (\widehat{\lambda}_{OLS}, \widehat{\theta}_{OLS})')$–components are negligible. For example, consider the cross term

$$\begin{pmatrix} \lambda - \widehat{\lambda}_{OLS} \\ \theta - \widehat{\theta}_{OLS} \end{pmatrix}' \left(2(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} w_{it}u_{it} - (NT(T-1)/2)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}(w_{it}u_{is} + w_{is}u_{it})\right)/\sigma_u^2. \tag{58}$$

It is $o_p(1)$ using a WLLN for the mean zero vectors $w_{it}u_{it}$ and $w_{it}u_{is}$ and by consistency of the pooled OLS estimators. This concludes the proof of showing that replacing $\widehat{\Omega}$ and $\widetilde{\sigma}_u^2$ by Ω and σ_u^2 does not affect the limiting distribution J_h.

References

Andrews, D. W. K. & P. Guggenberger (2005a) Asymptotic size and a problem with subsampling and with the m out of n bootstrap. Forthcoming in *Econometric Theory.*

——— (2005b) Hybrid and size–corrected subsampling methods. Forthcoming in *Econometrica.*

——— (2005c) Applications of subsampling, hybrid and size–correction methods. Forthcoming in *Journal of Econometrics.*

——— (2005d) Validity of subsampling and "plug–in asymptotic" inference for parameters defined by moment inequalities. Forthcoming in *Econometric Theory.*

——— (2005e) Incorrect asymptotic size of subsampling procedures based on post–consistent model selection estimators. Forthcoming in *Journal of Econometrics.*

Banerjee, A. & L. Iyer (2005) History, institutions, and economic performance: The legacy of colonial land tenure systems in India. *American Economic Review* **95**, 1190–1213.

Bedard, K. & O. Deschênes (2006) The long–term impact of military service on health: evidence from world war II and Korean war veterans. *American Economic Review* **96**, 176–194.

Blonigen, B.A. (1997) Firm–specific assets and the link between exchange rates and foreign direct investment. *American Economic Review* **87**, 447–465.

Bradford, W.D. (2003) Pregnancy and the demand for cigarettes. *American Economic Review* **93**, 1752–1763.

Gaynor, M., H. Seider, & W.B. Vogt (2005) The volume–outcome effect, scale economies, and learning–by–doing. *American Economic Review* **95**, 243–247.

Guggenberger, P. (2007) The impact of a Hausman pretest on the size of a hypothesis test. Forthcoming in *Econometric Theory.*

Hastings, J.S. (2004) Vertical relationships and competition in retail gasoline markets: Empirical evidence from contract changes in southern California, *American Economic Review* **94**, 317–328.

Hausman J.A. (1978) Specification tests in Econometrics. *Econometrica* **46**, 1251–1271.

Judge, G.G. & M.E. Bock (1978) The Statistical Implications of Pre–Test and Stein Rule Estimators in Econometrics. North–Holland, Amsterdam.

Wooldridge, J.M. (2002) Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge and London.

Tables and Figures

Table 1[9]

Finite Sample (Null) Rejection Probabilities (in %) of a) Symmetric Two–stage Test and b) Hausman Pretest for $N = 100, T = 2, \alpha = \beta = .05$

a)	$\sigma_u^2 = 5$					$\sigma_u^2 = 1$				
$p\backslash q$	0	.3	.4	.5	.6	0	.3	.4	.5	.6
.9	8.6	37.3	56.0	71.6	80.8	8.2	59.2	71.5	66.6	55.2
.6	7.7	32.7	45.9	53.3	51.6	7.0	35.0	28.3	15.2	7.2
.3	6.9	27.7	34.6	33.1	24.2	6.2	18.6	10.2	5.5	5.3
b)										
.9	5.1	7.2	8.9	11.4	14.1	5.2	15.8	23.9	35.1	47.2
.6	5.2	15.3	23.7	33.9	45.7	5.4	44.9	69.6	87.4	96.8
.3	5.5	26.3	42.5	59.7	76.3	5.9	70.3	91.9	99.0	99.9

Table 3[10]

Asymptotic Rejection Probabilities (in %) of Symmetric Two–stage Test Conditional on Pretest (Not) Rejecting $\alpha = \beta = .05,\ h_1 = 15$

	$R - C - R$				$R - C - NR$				$P(H_N > \chi^2_{1,1-\beta})$			
$h_{21}\backslash h_{22}$	.05	.1	.2	.3	.05	.1	.2	.3	.05	.1	.2	.3
.75	26.0	13.7	5.0	3.5	8.8	20.1	59.1	88.7	8.0	16.9	50.5	83.2
.8	31.0	16.5	5.8	3.5	9.3	22.3	64.8	92.5	7.4	14.8	43.3	75.9
.85	37.5	20.7	7.3	3.9	9.8	24.6	70.2	95.2	6.9	12.5	35.1	65.1
.9	46.6	27.7	10.0	5.1	10.4	27.0	75.2	97.1	6.3	10.1	25.8	49.6
.95	60.8	41.7	17.5	8.9	11.0	29.5	79.8	98.3	5.7	7.6	15.6	28.9

[9] The results are based on $R = 30,000$ simulation repetitions.

[10] The results are based on $R = 3,000,000$ simulation repetitions.

Fig. 2. Asy max rej prob over h_1 of sym test as function of h_{21} and h_{22}