THE HAUSMAN TEST STATISTIC CAN BE NEGATIVE EVEN ASYMPTOTICALLY

BY SVEN SCHREIBER

GOETHE UNIVERSITY FRANKFURT

ABSTRACT. We show that under the alternative hypothesis the Hausman chi-square test statistic can be negative not only in small samples but even asymptotically. Therefore in large samples such a result is only compatible with the alternative and should be interpreted accordingly. Applying a known insight from finite samples, this can only occur if the different estimation precisions (often the residual variance estimates) under the null and the alternative both enter the test statistic. In finite samples, using the absolute value of the test statistic is a remedy that does not alter the null distribution and is thus admissible.

[add the following paragraph for long summary:]

Even for positive test statistics the relevant covariance matrix difference should be routinely checked for positive semi-definiteness, because we also show that otherwise test results may be misleading. Of course the preferable solution still is to impose the same nuisance parameter (i.e., residual variance) estimate under the null and alternative hypotheses, if the model context permits that with relative ease. We complement the likelihood-based exposition by a formal proof in an omitted-variable context, we present simulation evidence for the test of panel random effects, and we illustrate the problems with a panel homogeneity test.

Keywords: Hausman test, negative chi-square statistic, nuisance parameter

JEL code: C12 (hypothesis testing)

1. INTRODUCTION

The specification test principle proposed by Hausman (1978) is extremely popular because it is versatile and often easy to apply. However, it is well known that its application

Date: August 2008.

I would like to thank (without implicating) Uwe Hassler for helpful comments on an early version of this paper, and Christian Offermanns for providing the data for the illustration.

Author's address: Goethe University Frankfurt, Money & Macro department, Mertonstr. 17 (PF82), D-60054 Frankfurt, Germany; fax +49/69/798-28933; e-mail svetosch@gmx.net.

quite often leads to negative test statistics, caused by estimated parameter variance differences that are not positive semi-definite (not PSD). This experience is clearly worrying given that the test statistic is distributed as χ^2 under the null hypothesis. The consensus (textbook) reaction is to attribute such annoying results to peculiar random data constellations in given finite samples. Since the Hausman test principle relies on asymptotic arguments, negative test statistics are only viewed as troublesome in practical work, not as a more fundamental problem. As a representative example for various textbooks, the authors of the popular Stata software agree that a negative statistic "is not an unusual outcome for the Hausman test" (StataCorp, 2001, vol. 2, p. 13, in the context of multinomial logit models), but as usual they attribute their specific result to their relatively small sample. This paper shows that this view is not entirely correct: We investigate whether negative test statistics may occur even asymptotically and indeed find that this can happen in some model classes if H_1 is true.

We use a likelihood framework for our argument, because most interesting applications employ likelihood methods or can be reformulated as such, although the Hausman test procedure is applicable more generally. We deal with certain fixed alternative hypotheses and ask what can happen if those are true. Our question is actually simpler than the existing local-power studies, but apparently has never been systematically addressed. The result of this gap in the literature is the erroneous belief that the occurrence of non-PSD variance differences is always a small-sample problem. Note that our problem is unrelated to that of Hausman and Taylor (1981), because they analyze the different issue of singular variance differences.

Fortunately, it is clear that the well-known finite-sample solution to use the same estimate of the estimation precision (residual variance) parameter under H_0 as well as under H_1 also works in the limit.[1] However, in some model contexts this remedy is either not available or quite difficult to apply, and in those cases our findings can be useful.[2] We

[1]It has long been known that the problem discussed in this paper is directly linked to the estimation of the residual variance in many contexts, and thus it is a manifestation of the general problem of nuisance parameter estimation, which is a fundamental problem in classical statistics. Therefore this note should in no way be interpreted as a criticism of the Hausman test principle.

[2]For example, in the popular pooled mean-group estimation of dynamic panel models (see Pesaran et al., 1999), a non-parametric technique is normally used to estimate the parameter variance in the unrestricted

also point out that non-PSD variance differences can lead to spuriously small positive test statistics that would imply false acceptances of H_0.

The rest of this paper is divided into three sections. After briefly recapitulating the Hausman test principle and introducing some necessary notation, the following section 2 develops the argument, contains an analytical example, and discusses possible solutions of the problem. In section 3 we present simulation evidence for the widespread panel test of random vs. fixed effects as well as a real-world empirical illustration of the issues involved in a test of homogeneous panel coefficients. Finally, section 4 offers conclusions and recommendations for applied work.

2. THEORY

2.1. **The setup and notation.** This section draws heavily from Holly (1982), but the notation is not identical. Consider a parametric statistical model which for a sample of size T gives a log-likelihood that depends on two parameter vectors: $\ell(\theta,\gamma)$. The respective dimensions of the parameter (column) vectors are k_θ and k_γ. The true values of the parameters are denoted by θ_c and γ_c ("c" for correct), and it is important to note that we keep the true value θ_c fixed and in particular independent of the sample size T. We wish to test the hypothesis $\mathsf{H}_0 : \theta = \theta_0$ with a Hausman test, and we assume $k_\gamma \geq k_\theta$ to ensure that the test works in all directions. The Hausman test analyzes the difference between two estimators of γ, the first of which is simply the unrestricted maximum-likelihood estimator of the entire model, denoted by $\hat{\theta}_u$ and $\hat{\gamma}_u$ ("u" for unrestricted). Under standard regularity conditions it is clear that $\operatorname{plim} \hat{\theta}_u = \theta_c$ and $\operatorname{plim} \hat{\gamma}_u = \gamma_c$, so this estimator is consistent in particular for γ. The second estimator imposes the null hypothesis $\theta = \theta_0$ and therefore does not estimate θ. Denote the resulting H_0-estimate for γ by $\hat{\gamma}_0$. The Hausman test statistic is given by a quadratic form of the scaled vector of contrasts:

$$\hat{m} = (\hat{\gamma}_0 - \hat{\gamma}_u)' \left(\hat{\mathsf{V}}(\hat{\gamma}_u) - \hat{\mathsf{V}}(\hat{\gamma}_0)\right)^{-1} (\hat{\gamma}_0 - \hat{\gamma}_u) \tag{2.1}$$

model, and thus replacing the residual variance estimate is simply not possible. Another important case is the use of robust estimators of parameter (co-) variances where the residual variance estimate is not easily separable from the estimated parameter variance.

where the $\hat{\mathsf{V}}(.)$ are the estimated covariance matrices of the parameter estimates. If H_0 is true then $\hat{\gamma}_0$ is asymptotically more efficient than $\hat{\gamma}_u$, such that $T^{2r_c}\left(\hat{\mathsf{V}}(\hat{\gamma}_u) - \hat{\mathsf{V}}(\hat{\gamma}_0)\right)$ tends to a PSD matrix, where r_c denotes the convergence rate of the specific model context (which will often be 0.5). From Hausman (1978) we know that $\hat{m}$ is asymptotically distributed as $\chi^2(k_\gamma)$ if H_0 is true.

For a (not necessarily consistent) estimator $\hat{\beta}$ we denote with $\mathsf{Avar}(\hat{\beta})$ the asymptotic covariance matrix in the standard sense that it is minus the inverse of the limit of the expected average Hessian: $\mathsf{Avar}(\hat{\beta}) = -\left(\lim_{T\to\infty} \mathsf{E}_\beta(T^{-1}H)\right)^{-1}$, where the Hessian H has typical element $\partial^2\ell/(\partial\beta_i\partial\beta_j)$, and evaluation is at $\mathsf{plim}\,\hat{\beta}$. We can also write this in terms of the asymptotic information matrix: $\mathsf{Avar}(\hat{\beta}) = \mathscr{I}^{-1}$. Therefore, we write the parameter variance of the H_0-model as $\mathsf{Avar}(\hat{\gamma}_0) = \left(\mathscr{I}^0_{\gamma\gamma}\right)^{-1}$, and in the unrestricted maximum-likelihood estimate we can partition the inverse of the parameter variance matrix according to our partitioned parameter vector:

$$\mathsf{Avar}\begin{pmatrix} \hat{\gamma}_u \\ \hat{\theta}_u \end{pmatrix} = \begin{pmatrix} \mathscr{I}^u_{\gamma\gamma} & \mathscr{I}^u_{\gamma\theta} \\ \mathscr{I}^u_{\theta\gamma} & \mathscr{I}^u_{\theta\theta} \end{pmatrix}^{-1} \tag{2.2}$$

We derive $\mathsf{Avar}(\hat{\gamma}_u)$ by applying the formula for partitioned inverses to equation (2.2), which yields:

$$\mathsf{Avar}(\hat{\gamma}_u) = \left(\mathscr{I}^u_{\gamma\gamma} - \mathscr{I}^u_{\gamma\theta}\left(\mathscr{I}^u_{\theta\theta}\right)^{-1}\mathscr{I}^u_{\theta\gamma}\right)^{-1} \tag{2.3}$$

Since Holly (1982) showed that the underlying null hypothesis of the Hausman test is actually $(\mathscr{I}^u_{\gamma\gamma})^{-1}\mathscr{I}^u_{\gamma\theta}(\theta-\theta_0) = 0$, we also assume $\mathscr{I}^u_{\gamma\theta} \neq 0$. Under this assumption it is possible to attribute the test to θ alone, as it is intended.

2.2. **General analytical considerations.** Under H_0 it is obvious that $\mathsf{Avar}(\hat{\gamma}_u) - \mathsf{Avar}(\hat{\gamma}_0)$ is a PSD matrix by construction of the Hausman test principle, and it is well known that the corresponding estimated variance difference may only be non-PSD in finite samples, caused by peculiar random data constellations. Therefore it is only interesting to consider the situation where H_1 is true, and thus the essential question is whether

$\mathsf{Avar}(\hat{\gamma}_u) - \mathsf{Avar}(\hat{\gamma}_0) = \left(\mathscr{I}^u_{\gamma\gamma} - \mathscr{I}^u_{\gamma\theta}\left(\mathscr{I}^u_{\theta\theta}\right)^{-1}\mathscr{I}^u_{\theta\gamma}\right)^{-1} - (\mathscr{I}^0_{\gamma\gamma})^{-1}$ is necessarily a PSD matrix even if $\theta_0 \neq \theta_c$. An equivalent question is whether

$$\mathscr{I}^u_{\gamma\gamma} - \mathscr{I}^u_{\gamma\theta}\left(\mathscr{I}^u_{\theta\theta}\right)^{-1}\mathscr{I}^u_{\theta\gamma} - \mathscr{I}^0_{\gamma\gamma} = \left(\mathscr{I}^u_{\gamma\gamma} - \mathscr{I}^0_{\gamma\gamma}\right) - \left(\mathscr{I}^u_{\gamma\theta}\left(\mathscr{I}^u_{\theta\theta}\right)^{-1}\mathscr{I}^u_{\theta\gamma}\right) \tag{2.4}$$

is necessarily negative semi-definite (NSD), where only the term $\mathscr{I}^u_{\gamma\gamma} - \mathscr{I}^0_{\gamma\gamma}$ is ambiguous. It is crucial to note that if this latter term is PSD under H_1, this means that the curvature of the likelihood function in the γ-direction is greater at the unrestricted estimate $(\mathsf{plim}\,\hat{\theta}_u, \mathsf{plim}\,\hat{\gamma}_u) = (\theta_c, \gamma_c)$ than at the wrong H_0-value $(\theta_0, \mathsf{plim}\,\hat{\gamma}_0)$. This underestimated curvature at the point of H_0 then means that the stochastic noise components of the model are overestimated when using the H_0-model. If $\mathscr{I}^u_{\gamma\gamma} - \mathscr{I}^0_{\gamma\gamma}$ is PSD "so much" so that it dominates the NSD term $-\mathscr{I}^u_{\gamma\theta}\left(\mathscr{I}^u_{\theta\theta}\right)^{-1}\mathscr{I}^u_{\theta\gamma}$, it may render (2.4) indefinite or even PSD. Therefore, the key problem of the Hausman test is the inconsistent estimation of the noise component in the H_0-model under H_1, where in many models the nuisance parameter called "stochastic noise component" will simply correspond to the residual variance.

To our knowledge it has not been noted so far in the literature that an asymptotic PSD difference $\mathscr{I}^u_{\gamma\gamma} - \mathscr{I}^0_{\gamma\gamma}$ may dominate the last term in (2.4), implying that negative Hausman test statistics may happen systematically even in large samples. Strictly speaking, this finding means that the Hausman test can be regarded as consistent only if additionally the NSD-ness of (2.4) is ensured, because otherwise the test statistic is not guaranteed to diverge to $+\infty$ in the entire parameter region of the alternative hypothesis. To substantiate the claim that such an inconsistency of the test is possible we prove a simple analytical example in section 2.3, and in section 3 we present simulation evidence as well as a real-world illustration. Unfortunately, it is difficult to be more precise without further specification of the model context because the Hausman principle is so widely applicable, but see section 2.5 for a discussion of important exceptions.

2.3. **A proved example in a simple context.** It seems useful to consider an analytical example to prove the asymptotic relevance of our arguments. Let the true regression model with classical properties and zero-mean variables be

$$y_t = \gamma x_{1,t} + \theta x_{2,t} + u_t, \tag{2.5}$$

where the null hypothesis is given by $\mathsf{H}_0 : \theta_0 = 0$, and the estimators of γ are used for a Hausman test of H_0. We use notation where r_{12}^2 is the squared sample correlation between x_1 and x_2, and r_{y2*1}^2 is the squared sample partial correlation between y and x_2 after having accounted for x_1, i.e. $r_{y2*1}^2 = 1 - (SSR_{total}/SSR_{y1})$, where SSR_{total} denotes the sum of squared residuals in the entire unrestricted model after regressing y on x_1 and x_2, SSR_{y1} is the sum of squared residuals from regressing y only on x_1.

Therefore $\mathsf{plim}\, r_{y2*1}^2$ and $\mathsf{plim}\, r_{12}^2$ determine the relative amount of (asymptotically estimated) noise in the two models, influencing the difference of the likelihood curvatures for $\theta_c \neq 0$.[3] Regressing y on x_1 alone gives $\widehat{\gamma}_0$ along with the residual variance estimate $\widehat{\sigma}_0^2$ and $\widehat{\mathsf{V}}(\widehat{\gamma}_0) = \widehat{\sigma}_0^2(\sum x_{1,t}^2)^{-1}$, and then regressing y on both x_1 and x_2 accordingly gives $\hat{\gamma}_u$, $\widehat{\sigma}_u^2$, and $\hat{\mathsf{V}}(\hat{\gamma}_u) = \widehat{\sigma}_u^2\left((1-r_{12}^2)\sum x_{1,t}^2\right)^{-1}$. It is of course essential for the example that we have used different estimates for the residual variance, namely $\widehat{\sigma}_0^2$ and $\widehat{\sigma}_u^2$. We find:

$$\mathsf{plim}\, T\left(\hat{\mathsf{V}}(\hat{\gamma}_u) - \widehat{\mathsf{V}}(\widehat{\gamma}_0)\right) < 0 \Leftrightarrow \mathsf{plim}\, r_{12}^2 < \mathsf{plim}\, r_{y2*1}^2 \tag{2.6}$$

Proof. Using the notation $S_{yy} \equiv \sum y_t^2$, and analogously $S_{11} \equiv \sum x_{1,t}^2$ etc. (S_{12}, S_{22}), and with the textbook-style data matrix X for the unrestricted model we have

$$(X'X)^{-1} = \frac{1}{S_{11}S_{22} - S_{12}^2}\begin{pmatrix} S_{22} & -S_{12} \\ -S_{12} & S_{11} \end{pmatrix}, \tag{2.7}$$

from which we need the upper-left element for $\hat{\mathsf{V}}(\hat{\gamma}_u)$. It follows that

$$\frac{S_{22}}{S_{11}S_{22} - S_{12}^2} = \frac{\frac{S_{22}}{S_{12}^2}}{\frac{1}{r_{12}^2} - 1} = \frac{\frac{S_{22}}{S_{12}^2} r_{12}^2}{1 - r_{12}^2} = \frac{1}{(1-r_{12}^2)S_{11}}. \tag{2.8}$$

[3] Note that the case $\mathsf{plim}\, r_{12}^2 = 0$ is ruled out by the assumption $\mathscr{I}_{\gamma\theta}^u \neq 0$, which was discussed before. Also note that under H_0 it holds that $\mathsf{plim}\, r_{y2*1}^2 = 0$.

Thus we get the expression for $\hat{\mathsf{V}}(\hat{\gamma}_u)$ in the example. Using estimators for the residual variance without degrees-of-freedom corrections we get

$$\frac{\hat{\mathsf{V}}(\hat{\gamma}_0)}{\hat{\mathsf{V}}(\hat{\gamma}_u)} = \frac{\frac{\widehat{\sigma}_0^2}{S_{11}}}{\frac{\widehat{\sigma}_u^2}{S_{11}(1-r_{12}^2)}} = \frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_u^2}(1-r_{12}^2) = \frac{SSR_{y1}}{SSR_{total}}(1-r_{12}^2) = \frac{1-r_{12}^2}{1-r_{y2*1}^2}, \tag{2.9}$$

and the rest is obvious. □

Hence if the influence of the neglected variable (measured by r_{y2*1}^2) is strong enough relative to any given sensitivity of the Hausman test (measured by r_{12}^2), an asymptotically negative test statistic *must* happen. Since $\operatorname{plim} r_{y2*1}^2$ is a property of the DGP that can be held fixed for all values of $\operatorname{plim} r_{12}^2 \in [0; \operatorname{plim} r_{y2*1}^2)$, we have shown that this result applies to non-trivial regions of the parameter space (in the sense of forming a set of positive measure).

2.4. **Possible solutions.** A good solution for the problem of non-PSD covariance matrix differences is already well known, see for example Hayashi (2000, p. 233): Using only one estimate of the stochastic noise component to calculate the Hausman test statistic guarantees NSD-ness of (2.4). When translated to our asymptotic formulation, this means to impose $\mathscr{I}_{\gamma\gamma}^u - \mathscr{I}_{\gamma\gamma}^0 = 0$, by using $\mathscr{I}_{\gamma\gamma}^{u*} = \mathscr{I}_{\gamma\gamma}^0$ in place of $\mathscr{I}_{\gamma\gamma}^u$, or $\mathscr{I}_{\gamma\gamma}^{0*} = \mathscr{I}_{\gamma\gamma}^u$ in place of $\mathscr{I}_{\gamma\gamma}^0$. This practice is valid because under H_0 it holds that $\mathscr{I}_{\gamma\gamma}^0 =_{|\theta_c=\theta_0} \mathscr{I}_{\gamma\gamma}^u$ anyway, and thus the limiting distribution under H_0 is not affected.

Another general solution is to use a regression-based test approach whenever possible, see for example Davidson and MacKinnon (1990). In that approach the equivalent of the Hausman test is conducted as an F test in an auxiliary regresssion, and that test statistic is obviously well-behaved.[4]

However, in some model contexts it is technically difficult to use only one estimate of the noise component, or to set up the relevant auxiliary regression. In fact, the Hausman test principle is so attractive because the two different specifications under H_0 and H_1 can be estimated separately and the results can be compared afterwards. In those cases the most widespread "solution" of the problem of a negative test statistic is to ignore the

[4]A further abstract solution with very limited applicability is given by Ai (1995).

test result. Often it is then recommended to *not* reject H_0, and sometimes even explicit advice is provided to "round" a negative test statistic to zero, which of course also implies acceptance of H_0. As we have shown, in large samples this strategy will lead to wrong decisions. But even in small or medium-sized samples there seems to be a better solution which allows a rejection of H_0 at least sometimes, and that is to use $|\hat{m}|$ as the relevant test statistic. If H_0 is true and $\hat{m}$ tends to a non-negative number then of course the test will be unaltered by taking the absolute value, which makes it clear that this strategy is admissible (asymptotically, like the Hausman test itself). On the other hand, in a situation such as in subsection 2.3 where the test statistic would be asymptotically negative,[5] using $|\hat{m}|$ would not always lead to rejection of H_0 because in finite samples $\hat{m}$ may be a small negative number. But at least in *some* cases the test would get a chance to actually reject H_0 when H_1 is true. However, taking the absolute value is clearly only a second-best approach, and whenever possible the well-known solutions discussed before should be applied.

2.5. **A model class that is not affected by the problem.** It is important to acknowledge that in some model contexts it always holds that the parameter covariance difference is a PSD matrix irrespective of whether H_0 or H_1 is true, and therefore in those contexts there can be no problems with the Hausman test whatsoever. For example, consider the classic case of the Hausman test for endogeneity bias, where an instrumental-variable (IV) estimator is evaluated against the OLS estimator, and the latter is appropriate only under H_0. Even though OLS is biased under H_1, its asymptotic variance in a matrix sense is always smaller than that of the IV estimator. (The key thing to acknowledge here is that OLS minimizes the estimated residual variance, apart from using the Gauß-Markov theorem; of course, this residual variance estimate is then inconsistent, since we are looking at the case where H_1 is true.) Geometrically, the difference arises because OLS employs orthogonal projections while in IV estimation the projections are non-orthogonal. Similar examples that we can think of are therefore closely related to the geometry of IV estimates, namely Hausman-type tests for endogeneity or over-identification in IV and GMM

[5]Note that in those cases the test statistic would diverge to $-\infty$, because the bias term $\hat{\gamma}_0 - \hat{\gamma}_u$ does not vanish asymptotically, but the estimated parameter variances still converge to zero.

contexts. Therefore, a negative test statistic is always a small-sample problem *in such settings*.

3. Simulation evidence and empirical illustration[6]

3.1. **Simulation of the Hausman test of random effects.** The two most widespread applications of the Hausman test principle are probably the test for endogeneity using an instrumental-variable framework, and the test of whether the random-effects estimator is biased in the context of panel models with unobserved unit-specific effects. In section 2.5 we have already commented that negative test statistics and related problems cannot occur in the context of instrumental-variable estimation. In contrast to that reassuring result, we will now present simulation evidence that the Hausman test of the consistency of the random-effects estimator in panel models may suffer from the problems discussed in this paper. That is, negative test statistics are indeed possible asymptotically in this context, of course depending on correlation and variance parameter constellations.

The standard panel framework is determined by the following data-generating process (DGP) with unobserved unit-specific disturbances (or "effects"):

$$y_{it} = x_{it} + \mu_i + \varepsilon_{it} \tag{3.1}$$

This DGP has the following properties: the overall constant term is set to zero and therefore vanishes from the equation,[7] and the coefficient of the scalar regressor x_{it} is normalized to one. The cross-sectional means of the regressor follow a standard normal distribution (with expectation zero), implying that its variance is normalized to one: $\mathsf{V}_\mathsf{i}(\bar{x}_i) = 1$, where $\bar{x}_i = \mathsf{E}_\mathsf{t} x_{it}$. In constructing x_{it} we add independent normally distributed noise to the $\bar{x}_i$, running three different sets of simulations using different amounts of within-unit variance, namely $\mathsf{V}_\mathsf{t}(x_{it}) \in \{0.2, 1, 5\}$. The unit-specific unobserved disturbances μ_i are constructed as a weighted average of the $\bar{x}_i$ and an independent standard normal distribution, so they also are centered at zero. The weight coefficient is varied

[6]The simulation was programmed in gretl 1.7 (Cottrell and Lucchetti, 2008), and the empirical results were produced with Ox programs (Doornik, 2002). All codes are available from the author.
[7]However, a constant is included and not restricted when estimating the models on the simulated data.

across a certain range in order to simulate different degrees of correlation between the unobserved effects and the regressor, i.e. different "degrees of violation" of the null hypothesis, measured as different amounts of correlation $\mathsf{Corr}(\bar{x}_i, \mu_i)$. Obviously, the null hypothesis holds if the weight on the $\bar{x}_i$ in this linear combination approaches zero, because then the correlation between the regressor and the unobserved unit-specific effects also tends to zero and the well-known random-effects estimator will be consistent (and efficient). For simplicity we use the standard FGLS (feasible generalized least squares) estimator for the random-effects model as a proxy of the more computationally demanding maximum-likelihood estimator, but that does not change the qualitative conclusions of the simulation exercise.[8] Finally, the idiosyncratic disturbances ε_{it} are drawn from another independent standard normal distrubution, again using a range of the variance parameter $\mathsf{V}(\varepsilon_{it})$.

In the simple omitted-variables example of section 2.3 the condition for asymptotically negative Hausman test statistics can be interpreted to say that (a) the null hypothesis must be "sufficiently" violated (r^2_{y2*1} large enough), and that (b) a "helping" assumption must be sufficiently close to being violated (r^2_{12} small enough). In the present context of the Hausman test of random- vs. fixed-effects estimators we have similar interpretations which are however a bit more complicated due to the additional dimension of the panel data: The null hypothesis itself is (a) $\mathsf{Corr}(\bar{x}_i, \mu_i) = 0$, and the helping assumptions are (b) that the idiosyncratic shock variance must be positive, $\mathsf{V}(\varepsilon_{it}) > 0$, and (c) that the within-variation of the regressor relative to the between-variation must not be infinite, $\mathsf{V_t}(x_{it})/\mathsf{V_i}(x_{it}) < \infty$. The reason for (b) is that for $\mathsf{V}(\varepsilon_{it}) = 0$ the random- and fixed-effects estimators are identical, thus the random-effects estimator does not provide additional information, regardless of the status of the null hypothesis. Helping assumption (c) is required for a similar reason; a higher within-variation of the regressor means that more information of the data is available in the within dimension, and thus the fixed-effects estimator precision approaches that of the random-effects estimator.

[8]For the model assuming "fixed" effects and thus applying the standard "within" transformation, the standard OLS estimator is identical to the maximum-likelihood estimator.

To get negative test statistics, we therefore expect that the null hypothesis itself must be sufficiently violated ($|\mathsf{Corr}(\bar{x}_i, \mu_i)|$ large enough), and that the variance of the idiosyncratic disturbance $\mathsf{V}(\varepsilon_{it})$ must be small enough, whereas the within-variation of the regressor $\mathsf{V_t}(x_{it})$ must be large enough. Of course, as in the example from section 2.3 and throughout this paper, both residual variance estimates from the random- and fixed-effect estimators must be used in constructing the Hausman test statistic to make negative test statistics possible.

After simulating the asymptotic properties of the various parameter constellations with large-N panel samples of 20000 observations each ($N = 1000$, $T = 20$), we indeed find that pattern displayed in figure 3.1. We have put the (root of) the variance of the idiosyncratic shocks $\mathsf{V}(\varepsilon_{it})$ on the x-axis and the tested correlation between the regressor and the unobserved effects $\mathsf{Corr}(\bar{x}_i, \mu_i)$ on the y-axis. To capture the third dimension (varying degrees of within-variation of the regressor) we display several graphs, where the within-variation of the regressor is increasing from top to bottom. The outcome of each simulated parameter constellation is symbolized by '$+$' if the Hausman test yields a positive test statistic, or by '$\times$' if the test statistic is negative. We see that negative test statistics consistently occur when the DGP has the "right" parameters. If the variance of the idiosyncratic disturbance is relatively low and the within-variation of the regressor is relatively high compared to the between-variation, relatively small departures from the null hypothesis are sufficient to produce negative test statistics.

3.2. **Empirical illustration with a Hausman test of homogeneous coefficients.** We use the context of a Hausman test against slope heterogeneity in a dynamic panel model for illustration. Accounting for slope heterogeneity is important because ignoring heterogeneity in a dynamic setting leads to asymptotically biased estimates (see Pesaran and Smith, 1995, and Pesaran et al., 1999). But not imposing homogeneity if it exists is inefficient, and therefore a Hausman-type test for homogeneity is a natural approach.

We analyze the exchange rates of European monetary union (EMU) members with respect to Japan (except Portugal for reasons of data availability), using an equation inspired

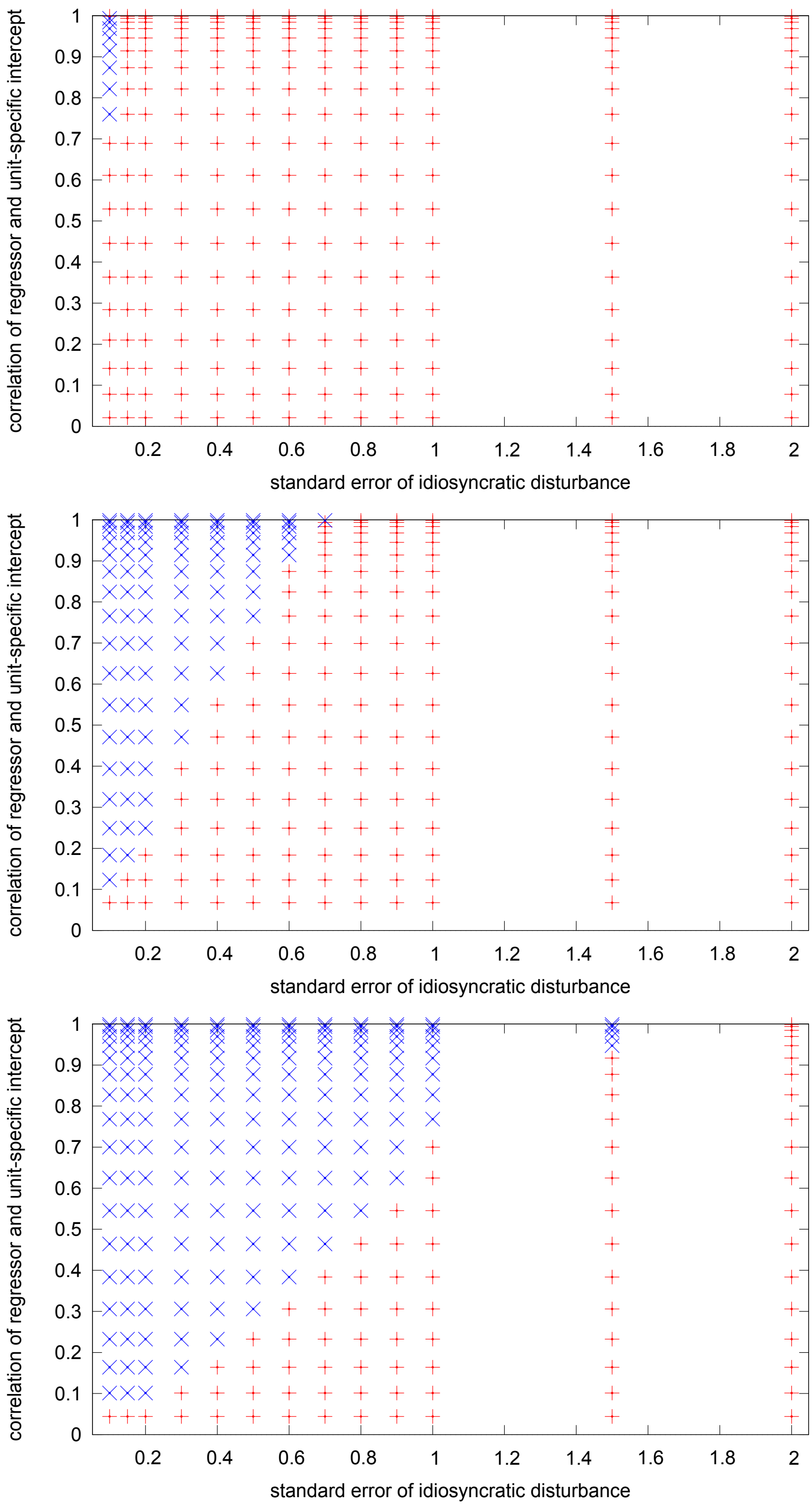

FIGURE 3.1. Simulation of the Hausman test for consistency of random effects. $T = 20$, $N = 1000$, for model specification details see the text. Upper panel: standard error of within-variation of the regressor set to 0.2; middle panel: ... set to 1; lower panel: ... set to 5. Positive Hausman test statistics in a simulation denoted by '$+$', negative Hausman test statistics by '$\times$'.

by purchasing power parity (PPP):

$$s_{it} = \lambda_i s_{i,t-1} + \phi_i' \mathbf{x}_{it} + \sum_{j=1}^{q_1-1} \alpha_{ij} \Delta s_{i,t-j} + \sum_{j=0}^{q_2-1} \delta_{ij}' \Delta \mathbf{x}_{i,t-j} + \mu_i + \varepsilon_{it} \tag{3.2}$$

Here s_{it} is the log nominal exchange rate of country $i = 1...N$ ($N = 11$) against the Japanese Yen in period $t = 1...T_i$, the vector $\mathbf{x}_{it}$ contains the log price level p_{it} of country i and the log price level p_t^* of Japan as the reference country. The implied long-run relation is given by $\bar{s}_i = \zeta_{1,(i)}\overline{p}_i + \zeta_{2,(i)}\overline{p^*}$ (where bars denote steady-state values), the long-run parameters are defined as $\zeta_{(i)} = \phi_{(i)}/(1-\lambda_{(i)})$, and the well-known strong form of PPP would imply $\zeta_i = \zeta = (1,-1)'$. The residual variance is denoted by $\sigma^2 = \mathsf{V}(\varepsilon_{it})$.

The sample consists of monthly observations 1980m1-1998m12 for all countries but Italy, where data for 1980/81 are missing. Since the following formulae are valid for fixed N, $T_i > 200$ seems sufficient to conduct asymptotic inference. For the sake of the illustration we simply assume that the cointegration properties of the data justify estimating (3.2) directly, and that the errors ε_{it} are independent across countries and time. These issues would have to be addressed in a more thorough analysis, and the economic results should therefore be interpreted with caution. Let us define the entire parameter vector as: $\beta_{(i)} = (\lambda_{(i)}, \phi_{(i)}', \alpha_{1,(i)}, ..., \alpha_{q_1-1,(i)}, \delta'_{1,(i)}, ..., \delta'_{q_2-1,(i)})'$.

Under H_0 the model is homogeneous and is estimated with the fixed effects (FE) estimator which imposes $\forall i : \beta_i = \beta$. Under H_1 we allow all individual coefficients to differ and thus employ the mean group (MG) estimator proposed by Pesaran and Smith (1995). The MG method estimates N separate equations and averages over all individual estimates, i.e. $\widehat{\beta}_{MG} = N^{-1}\sum_{i=1}^{N}\widehat{\beta}_i$, where each $\widehat{\beta}_i$ is the OLS estimate of β_i in a separate regression for the i-th country. Both estimators are effectively performed as OLS given the respective assumptions and can thus be interpreted as maximum-likelihood estimators. The following variance formulae hold (cf. Pesaran et al., 1996): Under the null hypothesis we have $\mathsf{V}(\widehat{\beta}_0) = \mathsf{V}(\widehat{\beta}_{FE}) = \sigma^2\left(\sum_{i=1}^{N} W_i' Q_i W_i\right)^{-1}$, where W_i is the $T_i \times (3+q-1+2q)$ data matrix holding the observations of all regressors for country i, and $Q_i = I_{T_i} - T_i^{-1} 1_{T_i} 1_{T_i}'$ (with 1_{T_i} being a T_i-element column vector of ones) is the "within"-transformation matrix taking deviations from country-specific means. In contrast, the unrestricted estimate is $\mathsf{V}(\widehat{\beta}_u) = \mathsf{V}(\widehat{\beta}_{MG}) = \frac{\sigma^2}{N^2}\sum_{i=1}^N (W_i' Q_i W_i)^{-1}$. The variance difference $\mathsf{V}(\widehat{\beta}_{MG}) - \mathsf{V}(\widehat{\beta}_{FE})$ assuming a known σ^2 is basically the difference between the harmonic and arithmetic means of some matrices and is necessarily PSD (again, see Pesaran et al., 1996).

The residual variance σ^2 can be estimated consistently under H_0 by two methods; either using the (restricted) fixed effects residuals $\widehat{\varepsilon}_{FE,it}$:

$$\widehat{\sigma}_0^2 = \frac{1}{\sum_{i=1}^N T_i - N} \sum_{i=1}^N \sum_{t=1}^{T_i} (\widehat{\varepsilon}_{FE,it} - \overline{\widehat{\varepsilon}}_{FE,i})^2 \tag{3.3}$$

where $\overline{\varepsilon}_{FE,i}$ is the individual-specific time-average of the estimated error term; or by averaging the standard variance estimates $\widehat{\sigma}_i^2 = T_i^{-1}\sum_t \widehat{\varepsilon}_{OLS,it}^2$:

$$\widehat{\sigma}_u^2 = N^{-1}\sum_i \widehat{\sigma}_i^2 \tag{3.4}$$

Obviously, $\widehat{\sigma}_u^2 \leq \widehat{\sigma}_0^2$ because less restrictions are imposed, and this difference may become important when those restrictions are not asymptotically justified.

3.2.1. *A negative test statistic that means rejection.* Arbitrarily choosing lags up to one year ($q_1 = q_2 = 12$) produces some negative entries on the diagonal of the parameter variance difference, which is thus clearly indefinite; of its 38 eigenvalues only 21 are positive. A researcher who applies the Hausman test using different σ^2-estimates for the H_0- and the H_1-model would find that the Hausman test statistic is actually negative, see table 1. Using $\widehat{\sigma}_u^2$ everywhere instead produces an extremely large positive test statistic which leads us to reject joint homogeneity of all parameters. The bottom line here is that discarding the Hausman test as "not applicable" or not rejecting H_0 because of the initially negative test statistic would have been a clear mistake.

3.2.2. *The problem of hidden indefinite variance differences.* Finally, let us point out a problem even when the Hausman test statistic is positive. We can see this by testing β for homogeneity using the lag structure $q_1 = 2$, $q_2 = 1$. This extreme choice is of course not meant to be a serious PPP model specification, but the test results in table 2 are illuminating. The test in the "natural" setup with different residual variance estimates produces a positive but insignificant test statistic. However, the parameter variance difference is

TABLE 1. A systematically negative test statistic

	using different estimates for σ^2	with $\widehat{\sigma}_u^2$ everywhere
# of neg. eigenvalues of variance difference	17 (of 38)	0
Hausman statistic $\widehat{m}$	-13.7	78.6
p-value	n.a.	0.00012

Note: Results refer to (3.2) with the lag structure described in section 3.2.1. Estimated residual variances as defined in the text are $\widehat{\sigma}_u^2 = 0.000484$, $\widehat{\sigma}_0^2 = 0.000543$.

TABLE 2. A pitfall even with a positive test statistic

	using different estimates for σ^2	using $\widehat{\sigma}_u^2$ everywhere
# of neg. eigenvalues of variance difference	1 (of 6)	0
Hausman statistic $\widehat{m}$	5.59	35.3
p-value	0.47	0.0000038

Note: Results refer to (3.2) with the lag structure described in section 3.2.2. Estimated residual variances as defined in the text are $\widehat{\sigma}_u^2 = 0.000613$, $\widehat{\sigma}_0^2 = 0.000633$.

indefinite again, and imposing the same residual variance estimate leads to an extremely high test statistic implying unequivocal rejection. This example highlights the fact that there is a potential problem whenever the parameter variance difference is not PSD, no matter whether the test statistic itself is negative or not. Therefore, Hausman test applications should routinely check the definiteness of the parameter variance difference, otherwise false test decisions may occur.

4. CONCLUSIONS

In this paper we have pointed out that the problem of non-PSD estimated variance differences in Hausman tests can be systematic and is not always only a finite-sample issue. A well-known symptom of this problem is the occurrence of negative test statistics that defy the χ^2 null distribution, but a non-PSD variance difference can also produce misleading *positive* test statistics. Quite a bit of applied work might be affected by this latter problem even though the respective researchers are not aware of it, because it has not been acknowledged in the literature that a non-PSD variance difference may persist in infinite samples.

Therefore, whenever it is technically infeasible to apply the general solutions –namely to impose the PSD-ness of the variance difference a priori or to use the equivalent auxiliary regression approach– we recommend to routinely check the estimated variance difference numerically for PSD-ness. If this check fails, it would be a sign of a considerably worse fit of the H_0-model, and it may then be worthwhile to apply a variance encompassing test (see e.g. Mizon and Richard, 1986) or similar diagnostic tools. As a first remedy in case of a negative test statistic, the researcher can use the absolute value of the statistic instead, which leaves the test statistic asymptotically unchanged under H_0. However, this approach clearly cannot solve the problem of misleading *positive* test statistics.

In any case, finding a non-PSD parameter variance difference (and especially a negative test statistic) should not *per se* be interpreted as evidence in favor of H_0. This insight applies especially to large samples because we have shown that an asymptotically negative test statistic can happen, but only if H_1 is true.

REFERENCES

Ai, C. (1995). A positive semi-definite covariance matrix for Hausman specification tests of conditional and marginal densities. *Oxford Bulletin of Economics and Statistics*, 57(2):277–281.

Cottrell, A. and Lucchetti, R. (2008). *Gretl User's Guide (1.7.x)*. Available at http://gretl.sourceforge.net/#man.

Davidson, R. and MacKinnon, J. G. (1990). Specification tests based on artificial regressions. *Journal of the American Statistical Association*, 85(409):220–227.

Doornik, J. (2002). *Object-Oriented Matrix Programming Using Ox*. Timberlake Consultants Press, London, 3rd edition.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6):1251–1272.

Hausman, J. A. and Taylor, W. E. (1981). A Generalized Specification Test. *Economics Letters*, 8(3):239–245.

Hayashi, F. (2000). *Econometrics*. Princeton University Press.

Holly, A. (1982). A remark on Hausman's specification test. *Econometrica*, 50(3):749–760.

Mizon, G. E. and Richard, J. F. (1986). The encompassing principle and its application to testing non-nested hypotheses. *Econometrica*, 54:657–678.

Pesaran, M. H., Shin, Y., and Smith, R. (1999). Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association*, 94(446):621–634.

Pesaran, M. H. and Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68:79–113.

Pesaran, M. H., Smith, R., and Im, K. S. (1996). Dynamic linear models for heterogeneous panels. In Matyas, L. and Sevestre, P., editors, *The Econometrics of Panel Data: Handbook of the Theory with Applications*, chapter 8. Kluwer.

StataCorp (2001). *Stata Statistical Software: Release 7.0.* College Station (TX), USA.