Problematic Choices: Testing for Correlated Unit Specific Effects in Panel Data

Vera E. Troeger

University of Essex, Government Department, Wivenhoe Park, Colchester CO4 3SQ, UK,
vtroe@essex.ac.uk

Paper prepared for the Political Methodology Meeting in Ann Arbor 2008

Almost a first, very preliminary draft; do not circulate or cite without permission, comments welcome

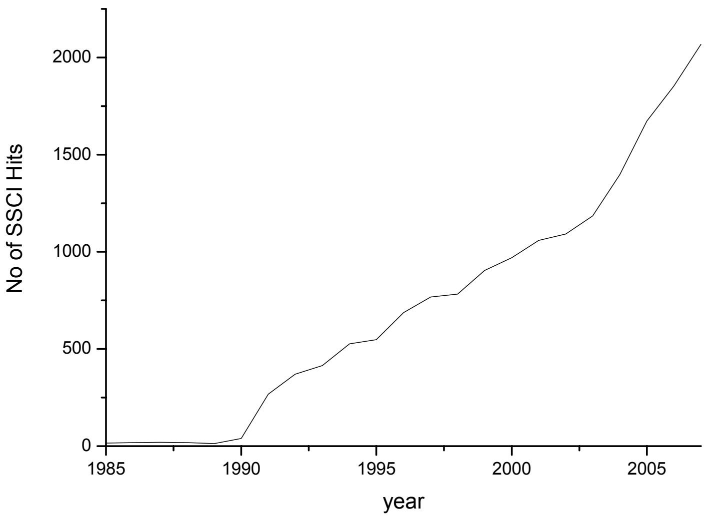
Abstract:
The (generalized) Hausman specification test (Hausman 1978) is the gold-standard for political scientists using time-series cross-section data to check whether unit specific effects are correlated with right-hand-side variables. More than 500 articles (published in SSCI journals) over the last 20 years in Economics and Political Science used the Hausman test to justify the model choice, e.g. whether to employ a fixed effects or random effects/ pooled OLS specification. The asymptotic properties of the Hausman test and its variants are well known and formal power analyses have shown that the Hausman test performs reasonably well. Yet, the differences in the estimates of fixed effects and random effects models in finite samples can originate from two different sources: On the one hand, the Hausman test might rightly pick up differences that are caused by the inconsistency of the random effects estimator if unit specific effects are correlated with any of the explanatory variables and the random effects model therefore produces biased coefficients. On the other hand, differences might also stem from the inefficiency of the fixed effects estimator if explanatory variables are rarely changing and therefore only have a very small within variation. This inefficiency does not only lead to large standard errors but also to very unreliable point estimates that might be far away from the true relationship. While the Hausman test (and especially more recent variants and augmentations of the specification test) acknowledge the inefficiency of the fixed effects model and control for the differences in the asymptotic variances of the two estimators, this inefficiency in combination with correlated unit effects might still lead to unreliable test results. In International Relations and International and Comparative Political Economy where many of our explanatory variables measure institutions which do not change much over time this result might be especially harmful since the fixed effects model in this case produces very unreliable point estimates.
This paper analyses the finite sample properties and power of the Hausman specification test by using Monte Carlo experiments. It shows under what conditions, e.g. the size of the correlation between unit specific effects and explanatory variables, and the between-within variance ratio of right-hand-side variables, the Hausman test generates misleading results.

1. Introduction: Panel Data Analysis and Fixed Effects

During the second half of the 1980s less than 20 Articles were published in social science journals which conducted panel data or pooled cross-section time series analysis. In comparison, 2068 articles analyzing panel data were published in 2007. The use of panel data analysis has increased dramatically over the last two decades. Figure 1 shows that we can observe a steep, almost linear, increase of "panel data" hits over the last 20 years and since the publication of Neal Beck's and Jonathan Katz' work on "Panel Corrected Standard Errors" (Beck and Katz 1995, 1996) there is a further significant acceleration.

There is no evidence that this gain in significance can be stopped in the foreseeable future. Arguably panel data analysis has become the most important tool of data analysis and testing implications of theoretical models in social science.

Figure1: SSCI search for ‚Panel Data'



Analysing pooled data has many advantages over pure time series or cross sectional analyses: first, most social science theories generate predictions across units and over time which cannot be tested by just looking at one of these dimensions. Second, pure cross-sectional analyses do not allow controlling for dynamic processes, which can render estimates spurious. And third, only the analysis of panel data allows controlling for unit heterogeneity which cannot be controlled for by including additional variables. Unit specific effects are particularly problematic if they are correlated with any of the explanatory variables in the model. The fixed effects model which de-means all variables in the model and thereby eliminates the correlated unit specific effects is widely seen as the standard model to control for unit specific heterogeneity in the data. It's popularity as universal solution to correlated unit specific effects increased steadily over time. While during the 1980s only about 12 percent of the analysis using panel data controlled for correlated unit heterogeneity by employing a fixed effects specification, this number tripled for articles published from the late 1990s onwards. Yet, as Plümper and Troeger (2007) have shown using a specification that relies solely on the within variation of all variables can induce serious estimation problems. This is particularly the case when variables do not change at all or only rarely over time or if the theory makes predictions in levels or mainly explains cross-unit variation. Using the fixed effects specification as default in panel data analysis is therefore not without problems. Hence, prior testing for correlation of unit specific effects seems indispensable.

Several tests have been designed in order to test for the necessity to control for correlated unit specific effects by using a fixed effects model or a specification in first differences. The (generalized) Hausman specification test (Hausman 1978) thereby seems to be the gold-standard for social scientists using time-series cross-section data to check whether unit specific effects are correlated with right-hand-side (RHS) variables. About 45 percent of the authors who employ a fixed effects specification use the Hausman-test to justify their model choice (Hausman and Taylor 1981, Cornwell and Rupert 1988, Baltagi and Khanti-Akom 1990, Mann et al. 2004, Egger 2000, Bole and Rebec 2004). The practical relevance of other tests for correlated unit heterogeneity (e.g. augmented variants of the original Hausman-statistic) is negligible[1].

The asymptotic properties of the Hausman test and its variants are well known and formal power analyses have shown that the Hausman test performs reasonably well (for a short overview see Baltagi 2001, 65-70). In addition, several econometricians have conducted Monte Carlo analyses in order to compare the power of the original Hausman statistic and augmented versions of the specification test (Ahn and Low 1996, Bole and Rebed 2004 among others). Theses simulations have shown that both the Hausman-test and its augmented versions show good power in detecting the endogeneity of right-hand-side variables also under conditions of serial correlation, non-stationarity and different forms of heteroskedasticity.

Yet, the differences in the estimates of fixed effects and random effects models in finite samples can originate from two different sources: On the one hand, the Hausman test might rightly pick up differences that are caused by the inconsistency of the random effects estimator if unit specific effects covary with any of the explanatory variables and the random effects model therefore produces biased coefficients. On the other hand, differences might also stem from the inefficiency of the fixed effects estimator if explanatory variables are rarely changing and therefore only have a very small within variation. This inefficiency does not only lead to large standard errors but also to very unreliable point estimates (which are often significantly different from zero) which might be far away from the true relationship. While the Hausman test (and especially more recent variants and augmentations of the specification test) acknowledge the inefficiency of the fixed effects model and control for the differences in the asymptotic variances of the two estimators this problem has not been

---

1 This might be essentially due to the fact that most statistical packages only implemented the basic Hausman-test.

analysed formally – at least not to my knowledge. Indeed, the inefficiency of the fixed effects estimator in case right-hand-side variables are predominantly cross-sectional only poses a problem in finite samples. Therefore, examining the asymptotic properties of the Hausman-test (and its variants) does not help to draw any conclusions about the question how large sampling variation of the fixed effects estimator impacts the power of the specification test. Since in the probability limit the fixed effects estimator is consistent and the variance collapses around the true value, the impact of large sampling variation cannot be examined in a formal asymptotic analysis. Still, no formal finite sample or Monte Carlo analyses do exist that analyze the trade-off between inefficient fixed effects estimation (due to small within variance of right-hand-side variables) and biased random effects estimation (due to correlated unit specific effects).

In International Relations and International and Comparative Political Economy where many of our explanatory variables measure institutions which do not change much over time but might be influenced by country specific characteristics this trade-off should lead to especially unreliable results of the Hausman-test and researchers are left without formal guidance whether to run a fixed effects, random effects or pooled OLS model.

This paper deals with the econometric issue of testing for the necessity to eliminate unit specific effects in order to obtain unbiased estimation results. I analyse the finite sample properties and power of the Hausman specification test and its augmentations and conditions of inefficient fixed effects estimation and correlated unit specific effects by using Monte Carlo experiments. It shows under what conditions, e.g. the size of the correlation between unit specific effects and explanatory variables, and the between-within variance ratio of right-hand-side variables, the Hausman test generates misleading results and researchers should be careful when using the Hausman-test or its variants to determine whether to employ a fixed effects or random effects specification.

Since many applied researchers also use the F-test for joint significance of unit specific effects in order to determine the existence of unit specific effects, I examine the performance of this test with respect to between and within variance ratio of right-hand-side variables and correlated unit effects. To do so it is important to first establish what unit effects are. Moreover, I will show why fixed effects estimation can be very inefficient and how this inefficiency influences the performance of the different tests.

2. Unit Specific Effects: DGP vs. Estimation

Theoretically fixed effects (FE) are unit specific characteristics which do not change over time and which are hard to measure or operationalize, such as culture, history, geography etc in the case of countries and immeasurable personal characteristics that determine human behaviour in the case of individuals. Yet, the unit specific intercepts estimated by a fixed effects model capture much more than the FE of the data generating process (DGP) they account for the unit mean of all explanatory variables in the model.

Equation 1 depicts the data-generating process of a fixed effects model with time invariant variables:

$$y_{it} = \alpha + \sum_{k=1}^{K} \beta_k x_{kit} + \sum_{m=1}^{M} \gamma_m z_{mi} + u_i + \varepsilon_{it} \ . \qquad (1)$$

where the x-variables are time-varying and the z-variables are time-invariant. u_i denotes the N-1 unit specific effects (fixed effects) of the data generating process and ε_{it} is the iid error term, α is the intercept of the base unit, and β and γ are the parameters to be estimated.

The fixed effects transformation can be obtained by first averaging equation (1) over T:

$$\overline{y}_i = \sum_{k=1}^{K} \beta_k \overline{x}_{ki} + \sum_{m=1}^{M} \gamma_m z_{mi} + \overline{e}_i + u_i \qquad (2)$$

where

$$\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}\ ,\ \overline{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}\ ,\ \overline{e}_i = \frac{1}{T}\sum_{t=1}^{T} e_{it}$$

and e stands for the residual of the estimated model.

Then equation 2 is subtracted from equation 1. As is well known, this transformation removes the individual effects u_i and the time-invariant variables z. We get

$$\begin{aligned} y_{it} - \overline{y}_i &= \beta_k \sum_{k=1}^{K}\left(x_{kit} - \overline{x}_{ki}\right) + \gamma_m \sum_{m=1}^{M}\left(z_{mi} - z_{mi}\right) + \left(e_{it} - \overline{e}_i\right) + \left(u_i - u_i\right) \\ &\equiv \ddot{y}_{it} = \beta_k \sum_{k=1}^{K} \ddot{x}_{kit} + \ddot{e}_{it} \end{aligned} \tag{3}$$

with $\ddot{y}_{it} = y_{it} - \overline{y}_i$, $\ddot{x}_{kit} = x_{kit} - \overline{x}_{ki}$, and $\ddot{e}_{it} = e_{it} - \overline{e}_i$ denoting the demeaned variables of the within transformation. The estimated unit effects $\hat{u}_i$ do not equal the unit effects u_i in the data generating process.[2] Rather, this 'estimated unit effects' include all time invariant variables the overall constant term and the mean effects of the time-varying variables x – or more formally,

$$\hat{u}_i = \overline{y}_i - \sum_{k=1}^{K} \beta_k^{FE} \overline{x}_{ki} - \overline{e}_i \quad , \tag{4}$$

where β_k^{FE} is the pooled OLS estimate of the demeaned model in equation 3.

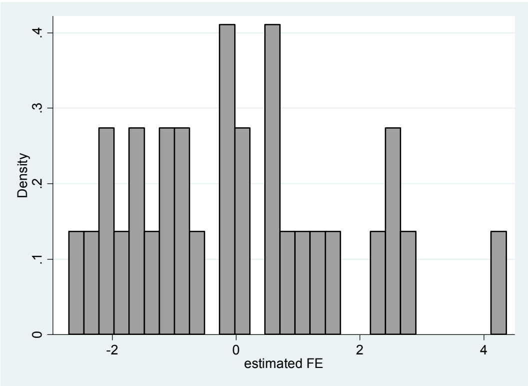
This $\hat{u}_i$ include the unobserved unit specific effects as well as the observed unit specific effects z, the unit means of the residuals $\overline{e}_i$ and the time-varying variables $\overline{x}_{ki}$, whereas u_i only account for unobservable unit specific effects.

In order to clarify this formal representation, figures 2 – 5 show the difference between estimated and "true" fixed effects for different sizes of the unit effects in the DGP, different number of RHS variables and various between to within standard deviation ratios of these RHS variables.

---

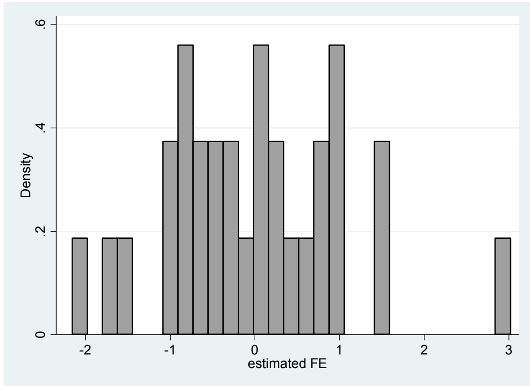
2 This notation follows standard practice. However, from equation 4 it follows that the fixed effects estimate of the unit effects propels much more to the estimated unit effects.

Figure 2: Estimated vs. "true" FE I



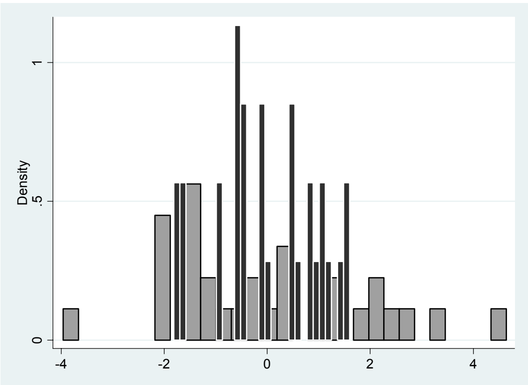
Settings: no FE in DGP, 1 RHS variable, SD(within)=SD(between)=1

Figure 3: Estimated vs. "true" FE II



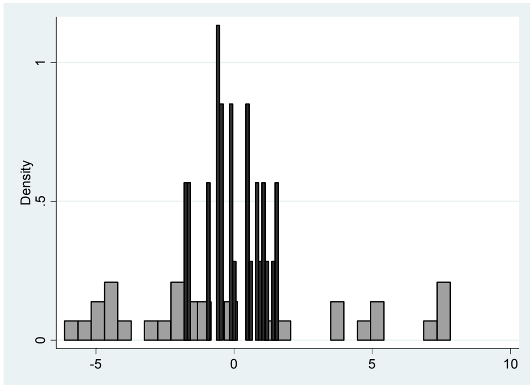
Settings: no FE in DGP, 3 RHS variables, SD(within)=SD(between)=1

Figure 4: Estimated vs. "true" FE III



black: true FE, grey: estimated FE
Settings: FE in DGP ~ N(0,1), 1 RHS variable, SD(within)=SD(between)=1

Figure 5: Estimated vs. "true" FE IV



black: true FE, grey: estimated FE
Settings: FE in DGP ~N(0,1), 1 RHS variable SD(within)=1 SD(between)=3

Figures 2 and 3 show that even if the unit effects in the data generating process equal zero, the estimated fixed effects are larger than zero as long as the RHS variables have some between variation. Figures 4 and 5 suggest that estimated fixed effects have a much larger variance than "true" fixed effects depending on the overall between variance of the RHS variables in the DGP. The correlation between the "true" and estimated fixed effects in figure 4 amounts to 0.53 and in figure 5 the correlation is only 0.26 since the between standard deviation of the RHS variable is larger than in the DGP underlying figure 4.

Since the true fixed effects are unobservable and the estimated fixed effects might be very different from the unit specific effects in the data generating process, the possible correlation between RHS variables and unit effects is also not observable. Since these two features are crucial for the right model choice, it is important for the applied researcher to be able to rely on test statistics which inform the decision on estimation procedures.

Sampling Variation of the Fixed Effects Estimator

As can be easily seen from equation 3, the within transformation of the fixed effects estimator eliminates the unit specific effects of the data generating process as well as all time invariant variables. It also propels the unit means of all RHS variables into the estimated fixed effects. Consequently the fixed effects estimator disregards the cross-sectional variance of all variables in the model and only uses variation over time in order to estimate the parameters of the model. As Plümper and Troeger (2007) have shown this might lead to highly inefficient estimation results if RHS variables have small within variance or a between variance that largely exceeds the variation over time. The possible large sampling variation of the FE model results from the fact that it disregards the between variation. Thus, the FE model does not take all the available information into account. In technical terms, the estimation problem stems from the asymptotic variance of the fixed effects estimator that is shown in equation 5:

$$A\,\hat{\mathrm{var}}\left(\hat{\beta}^{FE}\right) = \hat{\sigma}_u^2 \left(\sum_{i=1}^{N} \ddot{X}_i{}'\ddot{X}_i \right)^{-1} . \qquad (5)$$

When the within transformation of the FE model is performed on a variable with little within variance, the variance of the estimates can approach infinity. Thus, if the within variation becomes very small, the point estimates of the fixed effects estimator become unreliable. When the within variance is small, the FE model does not only compute large standard errors. Rather, the sampling variance also gets large and thus the reliability of point predictions is low.

The Random Effects Estimator

As compared to the fixed effects estimator the random effects estimator assumes strict exogeneity of the explanatory variables (similar to a pooled OLS model):

$$E\left(\varepsilon_{it} \middle| x_{it}, u_i\right) = 0$$
$$E\left(u_i \middle| x_i\right) = E\left(u_i\right) = 0 \tag{6}$$

The RE estimator is a feasible GLS estimator that only quasi-demeans all variables in the regression equation:

$$\beta_{RE} = \left(\sum_{i=1}^{N} X_i'\hat{\Omega}^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{N} X_i'\hat{\Omega}^{-1}y_i\right) \tag{7}$$

The Omega matrix in equation 7 (VC matrix of the error term) has a specific RE structure: it depends only on two parameters the variance of the unit specific effects (σ_u^2) and the variance of the iid error term (σ_ε^2) as compared to general GLS analysis where the VC matrix depends on T(T+1)/2 unrestricted variances and covariances. Omega has the so called Random Effects Structure:

$$\Omega = \begin{bmatrix} \sigma_u^2 + \sigma_\varepsilon^2 & \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_\varepsilon^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 & \sigma_u^2 + \sigma_\varepsilon^2 & \cdots & \sigma_u^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 + \sigma_\varepsilon^2 \end{bmatrix} \tag{8}$$

The random effects estimator can be easily compared to the fixed effects estimator by transforming it in the following way:

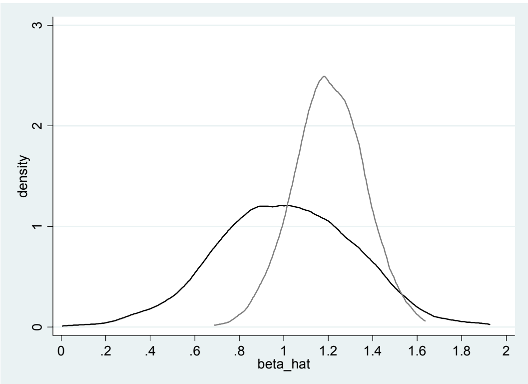
$$\beta_{RE} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \breve{x}_{it}'\breve{x}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} \breve{x}_{it}'\breve{y}_{it}\right)$$
$$\breve{x}_{it} = x_{it} - \hat{\lambda}\overline{x}_i, \qquad \breve{y}_{it} = y_{it} - \hat{\lambda}\overline{y}_i$$
$$\lambda = 1 - \left[\sigma_\varepsilon^2 / \left(\sigma_\varepsilon^2 + T\sigma_u^2\right)\right]^{1/2} \tag{9}$$

If lambda in equation 9 is close to unity, the random effects and fixed effects estimates tend to be close, this is especially the case if T gets large, or the variance of the estimated unit effects gets large as compared to the error variance[3]. Since the random and fixed effects estimator produce similar results as T gets large (Wooldridge 2002, 287) we should expect the Hausman-test to generate more accurate results when T increases.

As the random effects estimator relies on the strict exogeneity assumption it will produce biased estimation results whenever the unit specific effects are correlated with any of the RHS variables. However, in case the unit effects do not covary with the explanatory variables, the random effects estimator generates more efficient results and therefore more reliable point estimates. This trade-off between efficiency and bias should highly influence specification tests that are designed to inform the choice between random and fixed effects models.

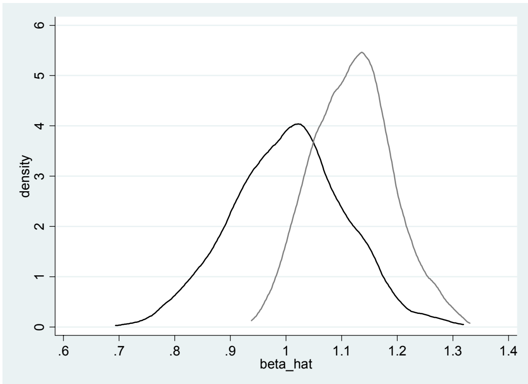
Figures 6 and 7 depict the trade-off between a biased random effects estimator in case of correlated unit effects (grey line) and a fixed effect estimator (black line) with large sampling variation if the RHS variables are dominated by cross-sectional variance.

Figure 6: Trade-off between efficiency and bias



black: FE; grey: RE
Settings : N=T=10, b/w-ratio=3, DGP u~N(0,1), 1000 repetitions, corr (u,x)=0.4, true beta=1

Figure 7: Trade-off between efficiency and bias



black: FE; grey: RE
Settings : N=T=70, b/w-ratio=3, DGP u~N(0,1), 1000 repetitions, corr (u,x)=0.4, true beta=1

[3] I include these characteristics into the MC analyses in order to see whether they impact the power of the specification tests.

Figures 6 and 7 show the distributions of 1000 fixed effects and random effects coefficients estimated from a DGP with one RHS variable that has a 0.4 correlation with the unit effects (which are standard normally distributed) and its between standard deviation is three times larger than its within variation. As expected, the fixed effects estimator produces unbiased results – the mean of the 1000 point estimates equals the true beta of one but the sampling distribution is much wider than the random effects sampling distribution. However, the random effects estimator produces biased estimates due to the correlation between the RHS variable and the unit effects. Further comparing figures 6 and 7 supports the expectation that the sampling variation gets smaller as the number of observations gets larger. Yet, in a finite sample large sampling variation can lead to highly unreliable point estimates. The distribution of 1000 fixed effects coefficients in figure 6 includes values that are 50 to 80 percent smaller or larger than the true value and only 8 percent of the 1000 coefficients (in the extreme tails) turn out to be statistically insignificant. In an analysis with real world data these deviations from the true value can be quite substantial and create misleading inferences. Of course, the estimated standard errors of the fixed effects estimates are about one third larger than the standard errors of the random effects estimates. Still, the fixed effects standard errors remain constant for all point estimates regardless of where in the distribution the point estimate lies. Thus, point estimates that are far away from the true relationship might still be statistically significant and point estimates close to the true value of beta might be insignificant.

I expect this trade-off between biased random effects results and inefficient fixed effects results to largely influence results of specification tests.

3. The Hausmann Specification Test and its Variants

The key consideration in compare fixed effects and random effects estimators is whether the unit effects are correlated with any of the explanatory variables and therefore the random effects estimates biased. Hausman (1978) proposed a test based on the difference between the

RE and FE estimates. Since fixed effects estimator is consistent when unit effects are correlated with x-variables, but RE is inconsistent, a statistically significant difference is interpreted as evidence against the random effects assumption of strict exogeneity. As seen above random effects estimates are biased whenever the explanatory variables are correlated with the unit specific effects. The Hausman-statistic tests the null-hypothesis that there is no difference between the estimated coefficients of a fixed effects and a random effects estimator. In its basic form it is computed as a chi²-type test for differences in estimates:

$$\chi^2(df) = \left(\hat{\beta}_{FE} - \hat{\beta}_{RE}\right)' \left[A\operatorname{var}\left(\hat{\beta}_{FE}\right) - A\operatorname{var}\left(\hat{\beta}_{RE}\right)\right]^{-1} \left(\hat{\beta}_{FE} - \hat{\beta}_{RE}\right) \tag{10}$$

The hausman statistic is distributed asymptotically as chi². Newey and McFadden (1994) actually provide general sufficient conditions, which are met by the fixed effects and random effects estimator. The fixed effects estimator is consistent under both the null-hypothesis and the alternative hypothesis, the random effects estimator however is inconsistent under the alternative but more efficient under the null. If the null is rejected (the p-value is smaller than 0.05) then at least some of the RHS variables are correlated with the unit effects and random effects produces biased estimation results. Since the fixed effects estimator is consistent when the unit specific effects and right-hand-side variables are correlated, but the random effects estimator is inconsistent, a statistically significant difference is interpreted as evidence against the random effects strict exogeneity assumption.

$\left[A\operatorname{var}\left(\hat{\beta}_{FE}\right) - A\operatorname{var}\left(\hat{\beta}_{RE}\right)\right]$ is positive definite since the random effects estimator only quasi-demeans the right-hand-side variables and therefore the asymptotic variance of the random effects estimate only approaches the asymptotic variance of the fixed effects estimate as T approaches infinity.

The Hausman-test can also be computed as a simple F-statistic which evaluates whether included time demeaned right-hand-side variables into a random effects GLS model still have a significant effect (Wooldridge 2002, 290, Mundlak 1978, Hausman 1978, Hausman and Taylor 1981, Chamberlain 1982).

Wooldridge (2002, 289) points to some caveats of the general Hausman-test. He argues that the Hausman-test assumes that the very strong assumption of constant conditional variance (u_i) and zero conditional covariance (u_i, x) hold under the null-hypothesis. This assumption, though, is just an auxiliary assumption which is not tested by the Hausman-statistic. The Hausman test therefore has no systematic power against the alternative assumption that the weaker assumption of othogonality between the fixed effects and the right hand side variables holds but the stronger assumption does not. The failure of the zero conditional covariance assumption might cause the usual Hausman test to have a non-standard limiting distribution, which means the resulting test could have an asymptotic size larger or smaller than the nominal size (Wooldridge 2002, 289). In this case the above mentioned F-statistic can be replaced by a robust Wald-statistic in the context of pooled OLS. The robust test then accounts for general serial correlation as well as arbitrary heteroskedasticity (Hoechle 2007, Arellano 1993).

The Hausman test statistic can be also derived from an artificial regression of the random effects residuals on the original right-hand side variables and the same variables averaged over time (Baltagi 1997, Baltagi and Liu 2007). Ahn and Low (1996) use this formulation of the Hausman-statistic and apply it to dynamic panel-data GMM estimators. Their generalization is essentially equal to Arellano's (1993) Wald statistic but incorporates a broader set of restrictions reflecting that each of the time-varying regressors is exogenous.

The generalized Hausman-statistic has been also adapted to detect correlated unit specific effects in two-stage least squares instrumental variables fixed and random effects models (e.g. Metcalf 1996).

All the augmentations and generalizations of the original Hausman-statistic produce essentially the same results but are adapted to common specification problems such as serial correlation, heteroskedasticity, simultaneity and non-stationarity. It has been mostly formally shown that the augmentations of the hausman-test produce better results under the conditions (violations of Gauss-Markov assumptions) for which they were developed (Baltagi 2001, 66-69). The general form of the Hausman-test (whether in it's Chi² or F form) performs well with a low frequency of type I errors (Baltagi 2001, 67). Ahn and Low (1996) perform Monte Carlo experiments which show that the original Hausman test and their GMM version have both good power in detecting endogeneity of regressors, thus the correlation between right-hand-side variables and the unit specific effects. They also give evidence that the Ahn and Low test performs better if coefficients of the regressors are nonstationary. Ahn and Low do not, however, discuss or test the impact of rarely changing variables on the performance of their version of the Hausman-test or the original Hausman-test nor do they change the variance of the unit specific effects, they only alter the correlation between right-hand-side variables and unit specific effects. Bole and Rebec (2004) suggest a bootstapped version of the generalized Hausman-statistic and compare the performance of this version to the asymptotic Hausman-statistic by using MC simulations. Their experiments show that the bootstrapped test-statistic generates about 20 % better coverage of the Null-Hypothesis and also performs better in detecting weak fixed effects. Yet, their simulations suffer from the restricted settings of the data generating process, they only look at situations where N equals 25 and T is set to 10. Since the Hausman-test should perform better in larger samples and especially if T gets large, their results only apply to small sample sizes. In addition, they only look at instances where the unit specific effects are relatively small, i.e. have a standard deviation of 0 or 0.5 (as compared to the Standard deviation of the right hand side variable which is set to 1). Again, the main interesting parameter that is varied is the size of the correlation between unit specific effects and right-hand-side variables.

The Hausman test suffers from the same problem as a fixed effects model, it is only consistent in the limit and might be heavily influenced by a large sampling variation if the sample is finite. In the probability limit the fixed effects estimator is consistent if the unit specific effects are correlated with any of the right hand side variables and the random effects estimator is not. Thus, in the limit the possible difference in the coefficients produced by fixed effect and random effect can only result from the bias of the random effects coefficient if the unit specific effects are correlated with right hand side variables. In the finite sample case, however, the difference in point estimates can be caused not only by correlated unit specific effects but also by unreliable point estimates produced by an inefficient fixed effects model. This is especially the case if variables do only change rarely or slowly over time (Beck and Katz 2001; Plümper and Troeger 2007, Wooldridge 2202, 286, 288) or are dominated by their cross-sectional variation. As the figures 6 and 7 show in this case we face a trade-off between the bias of the random effects model (in case of correlated unit effects) and the large sampling variation of the fixed effects estimates. The basic specification of the Hausman–test and all its variants (Baltagi 2001) mirror exactly this trade-off by looking at the ratio between the difference of the two estimates and the difference in the asymptotic variance of the two estimates. In practice this should lead to very weak performance of the Hausman-test in case the model includes variables that are correlated with the unit effects and variables which have small within-variation or – even worse – variables that are rarely changing and at the same time correlated with unobserved unit effects. And I will test exactly this claim. Since the fixed effects model, and consequently the hausman test, are asymptotically consistent it doesn't make much sense to analyze the asymptotic properties of the Hausman-test. In addition many econometricians have done this exercise already and shown that the Hausman test has very favourable asymptotic properties (Baltagi 2005, Baltagi and Lui 2007 among others). I therefore will concentrate on the small sample performance of this test under different conditions.

4. Monte Carlo Simulations of the Performance of Different Tests

The set-up of the Monte Carlo Simulations is straight forward. The data generating process models a cross-section time series model with and without unit specific effects:

$$y_{it} = \alpha + \beta_k \sum_{k=1}^{K} x_{kit} + u_i + \varepsilon_{it} \quad (11)$$

All variables, including the unit specific effects, are drawn from a standard normal distribution, the iid error term is drawn from a normal distribution with mean zero and a standard deviation of 3. All parameters to be estimated (intercept and coefficients) are set equal to one. The following settings are changed during the experiments:

- the number of variables: 1 or 3 variables,
- the size of the unit specific effects: standard deviation = {0,1,3,5},
- the correlation between the variables and the unit specific effects: corr(x,u)={0,0.2,0.4,0.6,0.8},
- efficiency of the fixed effects model: ratio of between to within standard deviation of the right hand side variables: b/w(x)={0.15,1,2,3}, so that the between variation is much smaller, equal or 2-3 times larger than the within variation of the right-hand-side variables,
- number of observations: all permutations of N and T = 10, 30, 70.

In the case of 3 right hand side variables I include experiments where all 3 variables are treated equally, thus where I change the ratio of between to within standard deviation and the correlation with the unit specific effects u for all three variables equally, and where variables are treated differently with respect to between and within standard deviation and correlation. I run theses different experiments for the F-test and the basic version of the Hausman-test[4].

4 I have conducted the same set of Monte Carlo experiments for the robust Wald-test suggested by Arellano (1993) and the Ahn-Low version of the Wald-statistic (Ahn and Low 1996) but with essentially the same substantial results. This is not surprising since the data generating process used in the presented MC experiments does neither include dynamics nor does it allow for any form of heteroskedasticity or non-stanionarity.

The combination of all settings creates 17280 different experiments. I will only focus on the most important combinations of the above settings and present the most interesting results below. Each experiment is repeated 1000 times.

Monte Carlo Results

Throughout the discussion of the results I will not display the power of the test which is defined as the percentage of results correctly rejecting the null-hypothesis. Since in some settings there exist a trade-off between inefficiency of the fixed effects model and bias of the random effects estimates, I will display the actual percentage of rejecting the null (for significance levels of 5 and 10 percent) – whether the rejection is correct or not – for all experiments. I start by discussing the results for models with strict exogeneity, thus no correlation between the unit specific effects and the right hand side variables and then I will go on to discuss the simulation results for settings where the correlation between explanatory variables and unit specific effects changes.

F-Test

The F test is a simple specification test for the joint significance of the estimated unit specific effects in a fixed effects model. Under the null all fixed effects jointly equal zero. The test compares the R^2 of a pooled OLS and a fixed effects estimation of the same model. The test statistic looks as follows:

$$F = \frac{\left(R_W^2 - R_{OLS}^2\right)/\left(N-1\right)}{\left(1-R_W^2\right)/\left(N\left(T-1\right)-K\right)} \sim F_{N-1,N(T-1)-K} \tag{12}$$

In light of the discussion in section 2 the F-test does not actually test whether the estimated fixed effects are jointly different from zero but whether the "true" unobservable fixed effects add explanatory power to the model. As seen in figures 2 – 5 the estimated fixed effects might be quite different from the unit effects of the DGP. However, since the F-test compares the R^2 of the OLS and the fixed effects specification it does not take into account possible time invariant variables or the unit means of time varying variables in the model because these factors already contribute to the explanatory power of the simple pooled OLS model. From this perspective it should not matter for the performance of the F-test whether the "true" and the estimated fixed effects differ largely. However the F-test should be negatively affected by highly correlated unit specific effects. If RHS variables strongly covary with the unit effects, these might not add largely to the explanatory power of the fixed effects model but might still heavily bias the estimates in a pooled OLS model. Thus, the F-test should produce weaker results in situations where the unit effects are highly correlated. From this perspective, performing an F-test for the existence of unit effects is not sufficient for deciding whether to employ a fixed effects, random effects or pooled OLS specification.

Table 1 depicts the performance of the F-test for a data generating process with one right-hand-side variable, no correlation between the unit specific effects, different sizes of the unit effect and different numbers of units and time points. The F-test performs reasonably well in detecting unit specific effects especially if T gets large and the variance of the unit effects increases. The ratio of between to within standard deviation of the explanatory variable doesn't seem to impact the performance of the F-test systematically as predicted above. This results is due to the fact that the F-statistic compares the difference of the overall R^2 of the pooled OLS and the fixed effects model and thus only cross-sectional variance in addition to the cross-sectional variance of the right-hand-side variables can cause a difference in the R^2 independent of the size of the between variance of the included variables.

If the variance of the unit specific effects equals zero (columns 3 and 4) the F-test correctly accepts the Null of no significant unit effects in most cases. Yet, it seems that if the between variance of the included right-hand-side variable gets larger as compared to the within variance the F-test has a slightly higher probability of wrongly rejecting the Null. Finally, in case the unit specific effects are small, the number of observations is small and/or the number of units largely exceeds the number of time points, the F-test can be highly unreliable by wrongly not rejecting the null of insignificant unit specific effects.

Table 1: Performance of the F-test I

	size	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
N/T	b/w u	0		1		3		5	
10_10	0.15	6	11	19	30	98	99	100	100
	1	8	13	19	30	97	100	100	100
	2	6	12	18	28	99	100	100	100
	3	9	12	21	31	98	100	100	100
10_30	0.15	3	12	55	72	100	100	100	100
	1	3	12	53	66	100	100	100	100
	2	7	15	54	67	100	100	100	100
	3	7	11	62	68	100	100	100	100
10_70	0.15	3	6	98	100	100	100	100	100
	1	6	8	98	99	100	100	100	100
	2	5	6	100	100	100	100	100	100
	3	3	7	97	99	100	100	100	100
30_10	0.15	2	10	38	48	100	100	100	100
	1	5	11	22	37	100	100	100	100
	2	9	13	35	52	100	100	100	100
	3	3	7	37	50	100	100	100	100
30_30	0.15	7	17	96	97	100	100	100	100
	1	2	7	96	100	100	100	100	100
	2	6	10	94	99	100	100	100	100
	3	6	13	94	98	100	100	100	100
30_70	0.15	6	8	100	100	100	100	100	100
	1	6	10	100	100	100	100	100	100
	2	9	14	100	100	100	100	100	100
	3	5	6	100	100	100	100	100	100
70_10	0.15	6	11	61	71	100	100	100	100
	1	5	6	66	78	100	100	100	100
	2	7	14	59	69	100	100	100	100
	3	2	7	60	73	100	100	100	100
70_30	0.15	3	7	100	100	100	100	100	100
	1	5	8	100	100	100	100	100	100
	2	6	9	100	100	100	100	100	100
	3	3	8	100	100	100	100	100	100
70_70	0.15	4	13	100	100	100	100	100	100
	1	1	6	100	100	100	100	100	100
	2	7	11	100	100	100	100	100	100
	3	4	9	100	100	100	100	100	100
Settings: 1 variable, no correlation, SD of unit specific effects {0,1,3,5}; N,T {10,30,70}									

Tables 2 and 3 display the performance of the F-test if we have more than one right-hand-side variable. Again, we first look at situations where the unit specific effects are uncorrelated with the right-hand-side variables. I only depict the results for N=30 and T=30, since the F-test seems to perform well for this number of observations[5]. As for the single variable case the F-test picks up well the existence of unit specific effects and also has good power in not rejecting the null if the variance of the unit effects equals zero.

Table 2: Performance of the F-test II

N/T	size b/w u	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
		0		1		3		5	
30_30	0.15	4	11	96	99	100	100	100	100
	1	6	11	91	96	100	100	100	100
	2	8	8	96	97	100	100	100	100
	3	4	8	94	97	100	100	100	100
Settings: 3 variables, equal treatment, no correlation, SD of unit specific effects {0,1,3,5}; N,T {30}									

Table 3: Performance of the F-test III

N/T	size b/w u	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
		0		1		3		5	
30_30	0.15	7	12	98	98	100	100	100	100
	1	4	7	96	98	100	100	100	100
	2	5	10	98	98	100	100	100	100
	3	5	10	94	97	100	100	100	100
Settings: 3 variables, 2 variables change b/w SD ratio, 1 variable ~ N(0,1), no correlation, SD of unit specific effects {0,1,3,5}; N,T {30}									

Let's now turn to evaluating the performance of the F-test if the unit specific effects are correlated with the right-hand-side variables. Table 4 shows the findings for one right-hand-side variable that is correlated to different degrees with the unit specific effects. The variance of the unit effects is set to three which is quite large in order to assure that the unit heterogeneity is picked up easily by the F-test. As can be seen, as before the performance of the F-test does not appear to be influenced by the level of between and within variance of the explanatory variable. However, the power of the F-test seems to be heavily dependent upon the correlation between the unit effects and the right-hand-side variable and the number of units and time points. If the overall number of observations is low and/or the number of units highly exceeds the number of time points the power of the F-test decreases with the size of the correlation. Note that in this series of experiments the F-test should always reject the Null-

[5] Results for different numbers of observation are available from the author upon request.

hypothesis of no significant unit specific effects since the unit effects have a large variance of 3 which exceeds the variance of the right-hand-side variable. In case the number of observation is very low (100, N=10, T=10), the F-test wrongly accepts the Null in more than 80 percent of the cases for all levels of correlation and even more than 90 percent of the cases if the correlation is very high (0.8). The F-test performs the better, the higher the number of observations and in particular the larger the number of time points.

Table 4: Performance of the F-test with correlated unit specific effects I

	size	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
N/T	b/w rho	0.2		0.4		0.6		0.8	
10_10	0.15	15	28	17	24	12	20	12	19
	1	15	32	12	24	13	29	8	11
	2	16	22	21	36	11	23	5	13
	3	15	20	19	28	10	22	5	12
10_30	0.15	62	70	48	63	34	53	22	32
	1	64	75	57	68	35	51	21	26
	2	53	66	50	68	40	46	26	39
	3	61	72	48	56	37	52	25	39
10_70	0.15	99	100	92	94	81	90	47	62
	1	96	97	93	96	83	88	59	69
	2	95	97	87	92	80	91	50	59
	3	95	98	92	96	90	92	42	57
30_10	0.15	34	40	39	54	19	28	13	22
	1	34	54	28	44	19	31	11	19
	2	40	54	20	31	25	36	12	19
	3	38	50	32	47	27	37	14	18
30_30	0.15	91	98	86	94	79	83	43	50
	1	92	96	89	96	69	82	43	56
	2	91	98	91	94	81	87	32	44
	3	90	92	93	97	80	90	44	53
30_70	0.15	100	100	100	100	99	99	84	90
	1	100	100	100	100	99	100	85	92
	2	100	100	100	100	99	100	82	91
	3	100	100	100	100	100	100	90	94
70_10	0.15	62	70	50	65	38	57	7	18
	1	56	67	56	76	33	44	18	24
	2	65	75	55	67	35	51	14	25
	3	66	72	53	65	40	47	19	34
70_30	0.15	100	100	99	99	93	94	69	84
	1	100	100	100	100	97	98	71	80
	2	100	100	99	100	97	99	68	78
	3	99	100	99	100	96	99	63	72
70_70	0.15	100	100	100	100	100	100	96	99
	1	100	100	100	100	100	100	100	100
	2	100	100	100	100	100	100	99	100
	3	100	100	100	100	100	100	99	100
Settings: 1 variable, corr (u,x) = {0.2, 0.4, 0.6, 0.8}, SD of unit specific effects = 3, N,T = {10,30,70}									

It is quite important to note that the power of the F-test decreases strongly if the correlation between the explanatory variables and the unit specific effects increases. From this perspective, it should be noted that the F-test is by no means a good instrument for detecting unit specific effects if these are correlated with the right-hand side variables. Since the estimates of a random effects or pooled OLS model tend to be heavily biased in case of correlated unit heterogeneity, researcher should never rely solely on the F-test (or a Breusch-Pagan LM test for zero variance of unit effects[6]) to determine the existence of unit specific effects.

It is not surprising that the performance of the F-test deteriorates with the size of the correlation since the F-test relies on the difference between the pooled OLS R^2 and the overall R^2 for the within estimator. A larger correlation indicates that the time-invariant unit means of all right-hand-side variables are closely related to the unobserved effects which reduces the difference in the explanatory power of a fixed effects model and a pooled OLS but it also biases the latter results heavily.

Tables 5 and 6 depict the results for three instead of on explanatory variables. The findings remain basically unchanged. If the correlation between the right-hand-side variables and the unit specific effects increases, the performance of the F-test deteriorates and the probability of wrongly accepting the Null goes up. I also changed the size of the unit effects in this set of experiments. If the unit-effects in the data generating process get really large the F-test has almost perfect power. Unfortunately – as shown above – the unit effects in the DGP and the estimated unit effects differ quite substantially, especially if variables in the model have a large between variance. Since it is impossible to measure or estimate the "true" unit specific

---

[6] I run the exact same MC simulations fort he Breusch-Pagan test for the existence of unit specific effects and the results are virtually the same as for the F-test. Findings can be obtained from the author upon request.

effects and we cannot rely on the estimated unit specific effects as a good predictor of the unit specific effects in the DGP, we do not know whether the F-test performs well.

Table 5: Performance of the F-test with correlated unit specific effects II

u	N/T	size b/w rho	0.05 0.2	0.1	0.05 0.4	0.1	0.05 0.6	0.1	0.05 0.8	0.1
3	30_30	0.15	93	96	79	87	38	58	21	23
		1	92	96	73	83	32	50	16	22
		2	91	95	72	82	41	57	19	31
		3	91	96	74	83	47	54	16	21
5	30_30	0.15	100	100	100	100	100	100	97	98
		1	100	100	100	100	100	100	97	98
		2	100	100	100	100	100	100	100	100
		3	100	100	100	100	100	100	99	100
Settings: 3 variables, equal treatment, corr (u,x) = {0.2, 0.4, 0.6, 0.8}, SD of unit specific effects = {3, 5}, N,T = 30										

In a situation where not all but only one variable is correlated with the unit specific effects (table 6) the F-test performs better than in the case all variables covary with the unit effects. Yet, the tendency remains the same: as the correlation increases, the probability of wrongly accepting the Null-hypothesis increases as well, with less than a 50 percent chance of a correct result when the correlation is very strong (0.8).

Since we do not know ex ante the “true” fixed effects and the correlation between these effects and the right hand side variables, the F-statistic constitutes a highly unreliable test for detecting uncorrelated or correlated unit specific effects.

Table 6: Performance of the F-test with correlated unit specific effects III

u	N/T	size b/w rho	0.05 0.2	0.1	0.05 0.4	0.1	0.05 0.6	0.1	0.05 0.8	0.1
3	30_30	0.15	94	98	86	92	75	87	44	56
		1	89	95	90	95	71	86	45	56
		2	90	93	87	90	70	80	36	51
		3	95	97	88	94	74	80	38	50
5	30_30	0.15	100	100	100	100	100	100	100	100
		1	100	100	100	100	100	100	100	100
		2	100	100	100	100	100	100	100	100
		3	100	100	100	100	100	100	100	100
Settings: 3 variables, equal b/w change, only one variable is correlated, corr (u,x) = {0.2, 0.4, 0.6, 0.8}, SD of unit specific effects = {3, 5}, N,T = 30										

Hausman-Test

I run the same set of experiments to study the performance of the Hausman-test. First, we look at DGPs without correlated unit effects, one explanatory variable with differing ratios of between to within standard deviation, different sizes of unit specific effects and different numbers of observations. Since the correlation is zero, the Hausman-test should never reject the Null-hypothesis of equal random and fixed effects coefficients.

As can be seen from table 7, the Hausman-test performs exceptionally well in such a setting. If there is no correlation between the unit specific effects and the explanatory variable the set-up of the Hausman-test (and its variants) is able to pretty well control for differences in the fixed effects and random effects coefficients that are caused by the inefficiency of the fixed effects estimator in case the between variance of the right-hand-side variable largely exceeds its within variation. The Hausman-test becomes more reliable with the number of observations and the size of the unit specific effects. However, some exceptions are notable: First, if the variance of the unit specific effects equals zero the probability of wrongly rejecting the null gets close to 10 percent (for all settings). Yet, the larger the size of the unit effects the more accurate the Hausman-test. Again, we do not know ex ante, however, the true size of the unit effects in the DGP. Second, if the number of time points largely exceeds the number of units (N=10, T=70), the unit effects are not very large and the within variance highly exceeds the between variance (an optimal case for an efficient fixed effects estimation), the Hausman-test tends to wrongly reject the Null-hypothesis of uncorrelated unit effects. This is very surprising since the fixed effects and random effects estimates should become more similar as T gets large.

Table 7: Performance of the Hausman-test I

	size	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
N/T	b/w u	0		1		3		5	
10_10	0.15	7	9	6	6	7	8	2	2
	1	7	13	3	7	2	2	0	0
	2	7	10	6	10	1	4	0	1

	3	2	3	9	15	3	6	0	2
10_30	0.15	3	4	2	8	0	0	0	0
	1	6	10	1	3	0	0	0	0
	2	4	4	2	5	0	0	0	0
	3	2	7	1	9	2	5	0	0
10_70	0.15	6	8	34	37	0	0	0	0
	1	7	13	0	1	0	0	0	0
	2	4	8	1	2	0	0	0	0
	3	3	7	2	9	0	0	0	0
30_10	0.15	8	9	2	5	1	2	0	0
	1	6	8	3	7	0	0	0	0
	2	10	15	4	8	2	4	0	1
	3	4	9	2	6	2	6	0	1
30_30	0.15	4	6	2	3	0	0	0	0
	1	3	8	1	3	0	0	0	0
	2	7	13	4	7	0	0	0	0
	3	7	11	10	16	2	2	0	0
30_70	0.15	7	12	0	2	0	0	0	0
	1	2	6	1	3	0	0	0	0
	2	3	10	0	0	0	0	0	0
	3	2	8	3	8	0	1	0	0
70_10	0.15	4	11	8	13	2	8	1	7
	1	4	13	1	5	1	1	0	0
	2	7	9	4	5	1	2	0	0
	3	4	10	2	3	1	2	0	3
70_30	0.15	6	10	6	12	0	2	0	0
	1	5	11	1	6	0	0	0	0
	2	2	10	3	7	0	1	0	0
	3	7	13	1	3	0	0	0	0
70_70	0.15	4	8	1	3	0	0	0	0
	1	1	5	0	2	0	0	0	0
	2	2	5	3	6	0	0	0	0
	3	6	8	4	5	0	0	0	0
Settings: 1 variable, corr (u,x) = {0.2, 0.4, 0.6, 0.8}, SD of unit specific effects = 3, N,T = {10,30,70}									

The findings remain essentially unchanged if we increase the number of right-hand-side variables when the correlation between these and the unit specific effects is zero (tables 8 and 9). Under these conditions the Hausman-test performs highly and correctly accepts the null of equal coefficients. Moreover, the larger the size of the unit specific effects and the greater the number of observations – especially the larger T – the higher the accuracy of the test results. This is in accordance with the characteristics of the two estimators. If T gets large and the unit effects exceed the variance of the idiosyncratic error term the random effects estimate approaches the fixed effects estimate.

Table 8: Performance of the Hausman-test II

	size	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
N/T	b/w u	0		1		3		5	
30_30	0.15	6	10	24	33	39	49	2	2
	1	2	4	0	3	0	0	0	0
	2	2	14	1	6	0	0	0	0
	3	2	10	4	7	0	0	0	0
Settings: 3 variables, equal treatment, no correlation, SD of unit specific effects {0,1,3,5}; N,T {30}									

As in the previous set of simulations, in case the within variation largely exceeds the between variation of the right-hand-side variables and the unit specific effects are of a moderate size, Hausman-test produces surprisingly unreliable results and wrongly rejects the null in ca. 30 percent of the cases.

Table 9: Performance of the Hausman-test III

	size	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
N/T	b/w u	0		1		3		5	
30_30	0.15	1	2	32	37	43	46	4	4
	1	3	5	0	0	0	0	0	0
	2	8	12	4	10	0	0	0	0
	3	6	15	0	2	0	0	0	0
Settings: 3 variables, 2 variables change b/w SD ratio, 1 variable ~ N(0,1), no correlation, SD of unit specific effects {0,1,3,5}; N,T {30}									

We now turn to situations that on the one hand really challenge the performance of the Hausman-test and on the other hand come closer to real world data – the combination of differences in the between to within standard deviation ratio of the RHS variables and various levels of correlation between unit specific effects and RHS variables. In the first set of simulations (table 9) we only have one explanatory variable which is correlated with the unit specific effects. Thus, the Hausman-test should always reject the Null of no differences between the fixed effects and random effects coefficient. The unit specific effects have sufficiently large variation (SD=3) so that they are larger than the RHS variable and can be easily detected by both fixed and random effects models.

The number of correct rejections of the Null increases with the size of the correlation and the number of observations as expected. Yet, the percentage of correct rejections decreases when the between variance of the RHS variable increases and exceeds its within variation. And this effect is very strong. Even in case the correlation is very high (0.8) and the number of observations is very large (N=70, T=70) there is a 46 % chance of wrongly not rejecting the Null in case the between standard deviation exceeds the within standard deviation by factor 3. The Hausman-test generates especially weak results when the number of observations is small or N largely exceeds T (even though fixed and random effects should produce greatly different results in this case), when the correlation between unit specific effects and RHS variables is moderate and when the between variation of the RHS variables is much larger than the within variance and fixed effects estimation produces highly inefficient and therefore unreliable point estimates. From table 9 it is quite clear that inefficiency and bias create a trade-off which creates highly variables test results. And indeed, it is not quite clear whether a fixed effects model is the best solution if theoretically interesting RHS variables are both correlated with the unit specific effects and dominantly cross-sectional (Plümper and Troeger 2007). The Hausman-test can certainly not give a definite answer to this question.

Table 9: Performance of the Hausman-test with correlated unit specific effects I

	size	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
N/T	b/w rho	0.2		0.4		0.6		0.8	
10_10	0.15	25	31	26	36	28	40	22	33
	1	6	13	4	11	12	20	8	15
	2	5	7	7	13	7	14	3	7
	3	8	15	7	10	4	6	4	8
10_30	0.15	41	51	61	70	61	78	44	51
	1	4	11	15	25	25	35	11	19
	2	2	11	9	13	14	18	15	25
	3	3	7	4	9	13	18	12	20
10_70	0.15	37	41	100	100	100	100	90	96
	1	5	5	16	27	28	46	49	60
	2	0	3	8	15	17	24	19	32
	3	1	3	7	11	13	19	13	21
30_10	0.15	24	34	59	74	58	68	47	58
	1	8	9	15	23	24	30	22	36
	2	3	6	6	11	12	15	10	19
	3	5	8	3	8	8	13	13	17
30_30	0.15	95	96	100	100	99	99	95	98
	1	6	11	34	46	53	67	58	72
	2	6	15	13	23	30	38	27	36
	3	3	8	14	21	12	19	20	24
30_70	0.15	94	94	100	100	100	100	100	100
	1	3	11	62	78	91	95	90	91

	2	8	13	27	40	47	61	42	62
	3	6	9	19	29	29	35	26	41
70_10	0.15	54	63	94	97	96	99	72	82
	1	6	14	32	48	51	65	47	59
	2	5	11	10	22	17	23	22	34
	3	3	6	7	15	14	21	16	24
70_30	0.15	100	100	100	100	100	100	100	100
	1	16	25	76	86	95	98	84	95
	2	7	13	25	34	56	66	51	60
	3	7	17	15	28	32	39	23	33
70_70	0.15	100	100	100	100	100	100	100	100
	1	23	41	97	99	100	100	100	100
	2	10	24	63	73	83	87	86	93
	3	12	18	36	47	55	70	54	62
Settings: 1 variable, corr (u,x) = {0.2, 0.4, 0.6, 0.8}, SD of unit specific effects = 3, N,T = {10,30,70}									

The problematic trade-off between efficiency and bias also drives the results if more than one RHS variables are included. The choice becomes even more difficult because variables can be at the same time correlated with the unit effects and have a large between variation, or they are either correlated or dominantly cross-sectional. In all cases, the Hausman-test should produce mixed results because there is no perfect solution to the problem since fixed effects produces unreliable point estimates for the variables with large between to within variance ratio and the random effects model generates biased coefficients because of correlated unit effects.

Tables 10 to 14 depict these different combinations and display the performance of the Hausman-test for the different settings. The firs set of simulations examines a situation where the three variables have similar levels of correlation with the unit effects and similar ratios of between to within standard deviation. The unit specific effects have a moderate (SD=3) or large size (SD=5) to provide more favourable conditions for the Hausman-statistic. As in the one RHS variable case the Hausman-test produces very mixed results. Since all three variables are correlated with the unit effects in all experiments the Hausman-test should technically reject the null in all cases. Yet, in situations where the between variation gets large as compared to the within variation and the correlation is small or moderate the rejection rate doesn't reach 20 percent. The Hausman-test performs better if the unit specific effects are very large (SD=5), though. Since we can neither observe the true unit effects nor the correlation between the unit effects and the RHS variables we only can use the observable ratio of between to within standard deviation to partly determine the goodness of the results produced by a Hausman-test if we do real-world data analysis. Thus, if we only have RHS variables with larger within than between variance the test results of the Hausman-statistic are pretty reliable and we can indeed conclude that a rejection of the Null-hypothesis is caused by correlated unit specific effects.

Table 10: Performance of the Hausman-test with correlated unit specific effects II

u	N/T	size b/w rho	0.05 0.2	0.1	0.05 0.4	0.1	0.05 0.6	0.1	0.05 0.8	0.1
3	30_30	0.15	97	98	97	99	82	87	37	47
		1	9	17	36	50	35	46	17	29
		2	7	13	19	24	15	28	11	22
		3	3	11	11	20	8	17	7	10
5	30_30	0.15	100	100	100	100	100	100	93	93
		1	42	66	98	99	100	100	99	100
		2	4	14	70	77	86	94	59	71
		3	2	11	45	62	57	71	36	48
Settings: 3 variables, equal treatment, corr (u,x) = {0.2, 0.4, 0.6, 0.8}, SD of unit specific effects = {3, 5}, N,T = 30										

Tables 11 and 12 depict situations where all variables have similar between to within variance ratios but only one (in table 11) or two (in table 12) variables are correlated with the unit specific effects. Comparing the two tables we can see that the Hausman-test performs worse, i.e. gives more mixed results, if variables are at the same time dominantly cross-sectional and correlated with the unit effects. If variables are either correlated or dominantly cross-sectional the Hausman-test has a higher probability of rejecting the Null.

Table 11: Performance of the Hausman-test with correlated unit specific effects III

u	N/T	size b/w rho	0.05 0.2	0.1	0.05 0.4	0.1	0.05 0.6	0.1	0.05 0.8	0.1
3	30_30	0.15	84	89	93	94	92	95	80	86
		1	5	9	18	27	45	56	40	50
		2	8	14	7	17	21	27	11	16
		3	1	2	7	13	10	13	7	15
5	30_30	0.15	54	57	70	75	100	100	100	100
		1	35	42	44	70	100	100	100	100

	2	10	17	8	27	66	83	79	88
	3	4	11	12	19	38	52	52	69
Settings: 3 variables, equal b/w change, only one variable is correlated, corr (u,x) = {0.2, 0.4, 0.6, 0.8}, SD of unit specific effects = {3, 5}, N,T = 30									

Table 12: Performance of the Hausman-test with correlated unit specific effects IV

N/T	size b/w rho	0.05 0.2	0.1	0.05 0.4	0.1	0.05 0.6	0.1	0.05 0.8	0.1
30_30	0.15	91	93	91	92	87	92	56	69
	1	8	23	34	49	38	50	32	43
	2	4	15	9	15	19	24	4	15
	3	6	9	15	25	8	21	9	15
Settings: 3 variables, equal b/w change, two variables are correlated, corr (u,x) = {0.2, 0.4, 0.6, 0.8}, SD of unit specific effects = 3, N,T = 30									

The same observation holds for tables 13 and 14, the Hausman-test produces more consistent results if variables are either correlated or dominantly cross-sectional but not both at the same time. In table 13 two out of the three variables are at the same time correlated and have different levels of between to within variance ratios while the third variable is an uncorrelated standard normal. The probability of not rejecting the null is higher when variables are at the same time correlated with the unit effects and dominantly cross-sectional.

Table 13: Performance of the Hausman-test with correlated unit specific effects V

N/T	size b/w rho	0.05 0.2	0.1	0.05 0.4	0.1	0.05 0.6	0.1	0.05 0.8	0.1
30_30	0.15	89	89	87	89	87	93	58	71
	1	7	13	48	58	37	51	29	36
	2	6	14	15	25	14	25	19	25
	3	6	9	10	13	11	19	10	17
Settings: 3 variables, 2 variables treated (b/w ratio and correlation), 1 variable standard normal, corr (u,x) = {0.2, 0.4, 0.6, 0.8}, SD of unit specific effects = 3, N,T = 30									

In table 14, in comparison two out of the three variables are uncorrelated with the unit specific effects but have different levels of between and within variance while the third variable is standard normally distributed, e.g. the within variation is larger than the between variation but covaries with the unit effects to different degrees. The Hausman-test now produces pretty reliable results since it attributes the possible differences in fixed effects and random effects estimates for the first two variables to the larger sampling variation of the fixed effects estimator and it correctly attributes the possible difference in the fixed effects and random effects coefficients of the third variable to bias caused by unit heterogeneity. However, since with non-artificial data we do not know the covariance between the true unit effects and the RHS variables it is almost impossible to determine the reliability of a one shot Hausman-test.

Table 14: Performance of the Hausman-test with correlated unit specific effects VI

	size	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
N/T	b/w rho	0.2		0.4		0.6		0.8	
30_30	0.15	75	78	96	97	94	95	76	83
	1	63	76	96	96	99	99	78	86
	2	80	85	99	99	93	97	86	92
	3	78	88	100	100	98	98	84	90
Settings: 3 variables, 2 variables treated (b/w ratio), 1 variable standard normal and correlated, corr (u,x) = {0.2, 0.4, 0.6, 0.8}, SD of unit specific effects = 3, N,T = 30									

Overall, the Hausman-test produces very mixed and unreliable results in case the data generating process contains RHS variables which covary with the unit specific effects and are characterized at the same time by a larger between variation as compared to their within variation. These problems, however, become less pronounced if the unit specific effects are big and the number of time points becomes large. In the latter case, however, fixed and random effects estimates grow closer anyway.

4. Conclusion

The Monte Carlo results show that it is often unwise to rely on a simple test statistic in order to decide whether to employ a fixed effects or random effects specification in panel data analysis. Since especially for variables that only change rarely over time the fixed effects estimator produces highly inefficient and therefore unreliable point estimates, fixed effects should not be used as default estimator in pooled data analysis just because it is consistent in the limit. Especially if the number of observations is generally small and/or the number of units highly exceeds the number of time periods the Hausman-statistic produces weak and unreliable test results. This problem becomes more serious if the data generating process contains RHS variables that at the same time covary with the unit specific effects and are dominantly cross-sectional.

In this case it is advisable to examine the single variables and the ratio of their between and within standard deviation. Another possibility is to follow Wooldridge's suggestion (2002, 290) and use a simple t-test for single parameters in order to evaluate for each variable separately whether the probability is high that the random effects estimate suffers from bias. This t-statistic can be easily obtained by dividing the difference of the fixed effects and random effects estimates by the difference of their standard errors.

MORE TO COME…

5. References

[INCOMPLETE]

Ahn and Low. 1996.

Baltagi, Badi H., Georges Bresson and Alain Pirotte. 2003. Fixed Effects, Random Effects or Hausman-Taylor? A Pretest Estimator. Economics Letters 79, 361-369.

Baltagi, BH Panel Econometrics

Baltagi, BH. 2005. A Hausman test based on the difference between fixed effects two-stage least squares and error components two-stage least squares - Solution. ECONOMETRIC THEORY 21 (2):483-484,

Baltagi, BH. 2006a. An alternative derivation of Mundlak's fixed effects results using system estimation. ECONOMETRIC THEORY 22 (6):1191-1194,

Baltagi, BH. 2006b. Estimating an economic model of crime using panel data from North Carolina. JOURNAL OF APPLIED ECONOMETRICS 21 (4):543-547,

Baltagi, BH; Liu, L. 2007. Alternative ways of obtaining Hausman's test using artificial regressions STATISTICS & PROBABILITY LETTERS 77 (13):1413-1417

Beck, Nathaniel and Jonathan Katz. 1995. What to do (and not to do) with Time-Series Cross-Section Data. American Political Science Review 89: 634-647.

Beck, Nathaniel and Jonathan Katz. 2001. Throwing Out the Baby with the Bath Water: A Comment on Green, Kim, and Yoon. International Organization 55: 487-495.

Bole and Rebec. 2004.

Breusch, T S & Pagan, A R, 1980. "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics," Review of Economic Studies, Blackwell Publishing, vol. 47(1), pages 239-53

Egger, P. 2000. A note on the proper econometric specification of the gravity equation. ECONOMICS LETTERS 66 (1):25-31, art.no.-ISSN 0165-1765.

Hausman, J 1978

Hoechle, D. 2007. Robust standard errors for panel regressions with cross-sectional dependence. STATA JOURNAL 7 281-312, art.no.-ISSN 1536-867X.

Hsiao, C; Pesaran, MH; Tahmiscioglu, AK. 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. JOURNAL OF ECONOMETRICS 109 (1):107-150, art.no.-PII S0304-4076 (01)00143-9|ISSN 0304-4076.

Mann, V; De Stavola, BL; Leon, DA. 2004. Separating within and between effects in family studies: an application to the study of blood pressure in children. STATISTICS IN MEDICINE 23 (17):2745-2756.

Metcalf, GE. 1996. Specification testing in panel data with instrumental variables. JOURNAL OF ECONOMETRICS 71 (1-2):291-307,

Plümper and Troeger 2007

Tchetgen, EJ; Coull, BA. 2006. A diagnostic test for the mixing distribution in a generalised linear mixed model. BIOMETRIKA 93 (4):1003-1010

Wooldridge, 2002