The impact of a Hausman pretest on the coverage probability and expected length of confidence intervals

Paul Kabaila *, Rheanna Mainzer and Davide Farchione
Department of Mathematics and Statistics, La Trobe University

June 15, 2015

Abstract

In the analysis of clustered and longitudinal data, which includes a covariate that varies both between and within clusters (e.g. time-varying covariate in longitudinal data), a Hausman pretest is commonly used to decide whether subsequent inference is made using the linear random intercept model or the fixed effects model. We assess the effect of this pretest on the coverage probability and expected length of a confidence interval for the slope parameter. Our results show that for the small levels of significance of the Hausman pretest commonly used in applications, the minimum coverage probability of this confidence interval can be far below nominal. Furthermore, the expected length of this confidence interval is, on average, larger than the expected length of a confidence interval for the slope parameter based on the fixed effects model with the same minimum coverage.

Keywords: Clustered data, Fixed effects model, Hausman specification test, Longitudinal data, Random intercept model.

*Corresponding author. Department of Mathematics and Statistics, La Trobe University, Victoria 3086, Australia. Tel.: +61 3 9479 2594; fax +61 3 9479 2466. E-mail address: P.Kabaila@latrobe.edu.au.
1 Introduction

The linear random intercept model is commonly used in the analysis of clustered and longitudinal data. In clustered data the response variable is measured once on a unit where each unit is nested within a particular cluster of units. For example, analyzing the reading test score of children which are nested within classrooms (clusters). Longitudinal data, which can also be viewed as clustered data (see Rabe-Hesketh and Skrondal (2012, p. 227)), is where the response variable is measured at different time points for each subject and where the measurements across time are nested within each subject (cluster). For example, analyzing the weights of individuals over time where the weight measurements across time are nested within individuals (clusters). When including a covariate in the linear random intercept model that varies both between and within clusters (e.g. time-varying covariate in longitudinal data) a preliminary Hausman (1978) test is commonly used to test the assumption of no correlation between the random intercept and covariate. If the Hausman pretest rejects the null hypothesis of no correlation between the random intercept and covariate then the fixed effects model is chosen for subsequent inference, otherwise the linear random intercept model is chosen. This preliminary model selection procedure has been widely used in econometrics (see e.g. Wooldridge, 2002 and Baltagi, 2005) and has also been adopted in other areas such as medical statistics, see e.g. Gardiner et al. (2009) and Mann et al. (2004). The Hausman pretest has also been implemented in popular statistical computer programmes including SAS, Stata, eViews and R, see Ajmani (2009, Chapter 7.5.3), Rabe-Hesketh and Skrondal (2012, Chapter 3.7.6), Griffiths et al. (2012, Chapter 10.4) and Croissant and Millo (2008), respectively.

The two-stage procedure widely used in the analysis of clustered and longitudinal data is as follows. In the first stage, the Hausman pretest is used to decide whether subsequent inference is made using the random intercept model or the fixed effects model (see e.g. Ebbes et al., 2004 and Jackowicz et al., 2013). The second stage is that the inference of interest is carried out assuming that the model chosen in the first stage had been given to us a priori, as the true model. Guggenberger (2010) considers this two-stage procedure when the inference of interest is a hypothesis test about the slope parameter. He provides both a local asymptotic analysis of the size of this test and a finite sample analysis (via
simulations) of the probability of Type I error.

We consider the case that the inference of interest is a confidence interval for the slope parameter. Kabaila et al., 2015 state 3 new finite sample results (Theorems 1, 2 and 7 of the present paper) on the effect of a Hausman pretest on the coverage probability of the confidence interval for this parameter. They also provide outline proofs of these results and a brief initial analysis of the coverage properties of this confidence interval. These coverage properties are estimated by simulation. In the present paper, we describe how the third of these results can be used to substantially improve the efficiency of these simulations through the use of variance reduction by control variates.

We compare the expected length of the confidence interval resulting from the two-stage procedure with the expected length of the confidence interval based on the fixed effects model, where the latter interval is adjusted to have the same minimum coverage as the former interval. The quantity used for this comparison is the scaled expected length, defined as the expected length of the interval resulting from the two-stage procedure divided by the expected length of the adjusted interval from the fixed effects model. The scaled expected length provides an insight into how useful the Hausman pretest is in this context. In the present paper, we provide 4 new finite sample theorems (Theorems 3, 4, 5 and 6) on the scaled expected length. The scaled expected length is estimated by simulation. We describe how to substantially improve the efficiency of these simulations through the use of variance reduction by control variates.

The results that we present make it easy to assess, for a wide variety of circumstances, the effect of the Hausman pretest on the coverage properties and scaled expected length of the confidence interval for the slope parameter. Our results show that when the usual small nominal level of significance for the Hausman pretest is used, the two-stage procedure can result in a confidence interval with minimum coverage probability far below nominal. However, if the nominal level of significance is increased to 50% then the minimum coverage probability is much closer to the nominal coverage (see Figures 1, 2 and 3). We also show that, in terms of expected length, the confidence interval resulting from the two-stage procedure is consistently outperformed by the confidence interval based on the fixed effects model, regardless of the nominal level of significance chosen for the Hausman pretest (see
Table 1). The results presented in this paper were computed using programs written in the R programming language, which will be made available in a convenient R package.

In Section 2, we consider the practical situation that the random error and random effect variances are estimated from the data. We consider three estimators of these variances: the usual unbiased estimators, the maximum likelihood estimators of Hsiao (1986) and the estimators of Wooldridge (2002). The coverage probability and the scaled expected length of the confidence interval resulting from the two-stage procedure are determined by 4 known quantities and 5 unknown parameters. The known quantities are the number of individuals, the number of time points (for the longitudinal data case), the nominal significance level of the Hausman pretest and the nominal coverage probability of this confidence interval. The unknown parameters are the random error variance, the random effect variance, the variance of the covariate (or time-varying covariate in the case of longitudinal data), a scalar parameter that determines the correlation matrix of the covariates and a non-exogeneity parameter.

If, for given values of the 4 known quantities, we wish to assess the dependence of the coverage probability and the scaled expected length of the confidence interval resulting from the two-stage procedure on the 5 unknown parameters then we might consider, say, five values for each of these unknown parameters, leading to 3125 parameter combinations. Apart from the daunting task of summarizing so many results, it is possible that one might miss important values of the unknown parameters, such as values for which the coverage probability is particularly low or the scaled expected length is particularly large.

Theorems 1 and 5 state that, apart from the known quantities, the coverage probability and scaled expected length are actually determined by only 3 unknown parameters, including the non-exogeneity parameter. If we compute the minimum coverage probability with respect to the non-exogeneity parameter then we have only 2 unknown parameters and our assessment of the coverage properties of the confidence interval resulting from the two-stage procedure is greatly simplified. Theorems 2 and 6 state that the coverage probability and scaled expected length are even functions of the non-exogeneity parameter, so that computation time is halved. We also propose a scaling of the non-exogeneity parameter that takes account of the sample size. In effect, this scaling reduces the number of known
quantities that determine this coverage probability from 4 to 3.

In Section 3, we consider the coverage probability of the confidence interval resulting from the two-stage procedure when the random error and random effect variances are assumed to be known. Theorem 7 shows that the conditional coverage probability of the confidence interval resulting from the two-stage procedure can be found exactly by the evaluation of the bivariate normal cumulative distribution function. This theorem is important because it is used to reduce the variance of the simulation based estimators of the coverage probability and scaled expected length of the confidence interval resulting from the two-stage procedure (when random error and random effect variances are estimated). As we show in Section 4 this variance reduction is achieved by using control variates.

2 The model and the practical two-stage procedure (random error and random effect variances are estimated)

We focus on the case of longitudinal data for which \( i \) denotes the individual (\( i = 1, \ldots, N \)) and \( t \) denotes the time period (\( t = 1, \ldots, T \)). By interpreting \( i \) as the cluster index and \( t \) as the unit of analysis, our results also apply to the analysis of clustered data. Part of the following description of the model and two-stage procedure is taken from Kabaila et al., 2015. Let \( y_{it} \) and \( x_{it} \) denote the response variable and the time-varying covariate, respectively, for the \( i \)'th individual at time \( t \). Suppose that

\[
y_{it} = a + \beta x_{it} + \mu_i + \varepsilon_{it},
\]

where the \( \varepsilon_{it} \)'s and the \( (\mu_i, x_{i1}, \ldots, x_{iT}) \)'s are independent, the \( \varepsilon_{it} \)'s are i.i.d. \( N(0, \sigma^2_{\varepsilon}) \) and the \( \mu_i \)'s are i.i.d. \( N(0, \sigma^2_{\mu}) \). We call \( \beta \) the slope parameter, \( \sigma^2_{\varepsilon} \) the error variance and \( \sigma^2_{\mu} \) the random effect variance. The \( \varepsilon_{it} \)'s and the \( \mu_i \)'s are unobserved. Suppose that the parameter of interest is the slope parameter \( \beta \) and that the inference of interest is a confidence interval for \( \beta \).

Also suppose that the \( (\mu_i, x_{i1}, \ldots, x_{iT}) \)'s are i.i.d. multivariate normally distributed
with zero mean and covariance matrix

\[
\begin{bmatrix}
\sigma_{\mu}^2 & \tau \sigma_{\mu} \sigma_x e' \\
\tau \sigma_{\mu} \sigma_x e & \sigma_x^2 G
\end{bmatrix},
\]

where \( e \) is a \( T \) vector of 1’s, \( G \) is a \( T \times T \) matrix with 1’s on the diagonal and \( \rho \) on the off-diagonal (compound symmetry), and \( \tilde{\tau} \) is a parameter that measures the dependence between \( \mu_i \) and \((x_{i1}, \ldots, x_{iT})\). We define the “non-exogeneity parameter” \( \tau = \tilde{\tau}(T/(1 + (T-1)\rho))^{1/2} \) and note that it is a correlation, so \( \tau \in (-1, 1) \). If \( \tau = 0 \) then the \( x_{it} \)'s are exogenous variables. Let \( x = (x_{11}, \ldots, x_{1T}, x_{21}, \ldots, x_{2T}, \ldots, x_{N1}, \ldots, x_{NT}) \). In simulations, we generate \( x \) as follows. Suppose that \( u_{i1}, \ldots, u_{iT} \) and \( \varepsilon_{i1}, \ldots, \varepsilon_{iT} \) are i.i.d. \( N(0, 1) \). Also suppose that these random variables and the \( \varepsilon_{it} \)'s are independent. Let \( x_{it}/\sigma_x = (1 - \rho)^{1/2} u_{it} + \rho^{1/2} v_i \) for \( i = 1, \ldots, N \) and \( t = 1, \ldots, T \). The distribution of \( \mu_i \) conditional on \((x_{i1}, \ldots, x_{iT})\), found from (2), is used to generate \((\mu_1, \ldots, \mu_N)\). Let \( u = (u_{11}, \ldots, u_{1T}, \ldots, u_{N1}, \ldots, u_{NT}, v_1, \ldots, v_N) \).

Assume, for the moment, that \( \sigma_{\varepsilon} \) and \( \sigma_{\mu} \) are known. When \( \tau = 0 \), a confidence interval for \( \beta \) may be found as follows. Let \( \psi = \sigma_{\mu}/\sigma_{\varepsilon} \). Condition on \( x \) and use the GLS estimator \( \hat{\beta}(\psi) \) of \( \beta \). Let \( z_c = \Phi^{-1}(c) \), where \( \Phi \) denotes the \( N(0, 1) \) cdf. Define the following confidence interval for \( \beta \)

\[
I(\psi) = \left[ \hat{\beta}(\psi) - z_{1-\alpha/2} \left( \text{Var}_0(\hat{\beta}(\psi) | x) \right)^{1/2}, \hat{\beta}(\psi) + z_{1-\alpha/2} \left( \text{Var}_0(\hat{\beta}(\psi) | x) \right)^{1/2} \right],
\]

where \( \text{Var}_0(\hat{\beta}(\psi) | x) \) denotes the variance of \( \hat{\beta}(\psi) \), conditional on \( x \) when \( \tau = 0 \). The confidence interval \( I(\psi) \) has coverage probability \( 1 - \alpha \) when \( \tau = 0 \). Averaging (1) over \( t = 1, \ldots, T \) for each \( i = 1, \ldots, N \) we obtain

\[
\bar{y}_i = a + \beta \bar{x}_i + \mu_i + \bar{\varepsilon}_i, \tag{3}
\]

where

\[
\bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} y_{it}, \quad \bar{x}_i = \frac{1}{T} \sum_{t=1}^{T} x_{it} \quad \text{and} \quad \bar{\varepsilon}_i = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it}.
\]

This model is called the between effects model. When \( \tau = 0 \), an alternative estimator of \( \beta \) is \( \hat{\beta}_B \), the OLS estimator based on the model (3), when we condition on \( x \). This estimator does not require a knowledge of \( \psi \). Subtracting (3) from (1), we obtain

\[
y_{it} - \bar{y}_i = \beta (x_{it} - \bar{x}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i). \tag{4}
\]
This model is called the **fixed effects model**. We estimate \( \beta \) by \( \tilde{\beta}_W \), the OLS estimator based on this model. Define the following confidence interval for \( \beta \)

\[
J(\sigma_\varepsilon) = \left[ \tilde{\beta}_W - z_{1-\alpha/2} \left( \text{Var}(\tilde{\beta}_W | x) \right)^{1/2}, \tilde{\beta}_W + z_{1-\alpha/2} \left( \text{Var}(\tilde{\beta}_W | x) \right)^{1/2} \right],
\]

where \( \text{Var}(\tilde{\beta}_W | x) \) denotes the variance of \( \tilde{\beta}_W \), conditional on \( x \). Irrespective of the value of \( \tau \), the confidence interval \( J(\sigma_\varepsilon) \) has coverage probability \( 1 - \alpha \). In practice, we do not know whether or not \( \tau = 0 \). The usual procedure is to use a Hausman pretest to test \( H_0 : \tau = 0 \) against \( H_a : \tau \neq 0 \). We consider this pretest, based on the test statistic

\[
H(\sigma_\varepsilon, \sigma_\mu) = \frac{(\tilde{\beta}_W - \tilde{\beta}_B)^2}{\text{Var}(\tilde{\beta}_W | x) + \text{Var}_0(\tilde{\beta}_B | x)},
\]

where \( \text{Var}_0(\tilde{\beta}_B | x) \) denotes the variance of \( \tilde{\beta}_B \) conditional on \( x \) and assuming that \( \tau = 0 \). This test statistic has a \( \chi^2_1 \) distribution under \( H_0 \). Suppose that we accept \( H_0 \) if \( H(\sigma_\varepsilon, \sigma_\mu) \leq z_{1-\tilde{\alpha}/2}^2 \); otherwise we reject \( H_0 \). Note that \( \tilde{\alpha} \) is the level of significance of this test. We now describe the two-stage procedure. If \( H_0 \) is accepted then we use the confidence interval \( I(\psi) \); otherwise we use the confidence interval \( J(\sigma_\varepsilon) \). Let \( K(\sigma_\varepsilon, \sigma_\mu) \) denote the confidence interval, with nominal coverage \( 1 - \alpha \), that results from this two-stage procedure.

Of course, in practice, \( \sigma_\varepsilon \) and \( \sigma_\mu \) are not known and need to be estimated. So, in practice, the Hausman pretest is based on the test statistic \( H(\tilde{\sigma}_\varepsilon, \tilde{\sigma}_\mu) \) and the two-stage procedure results in the confidence interval \( K(\tilde{\sigma}_\varepsilon, \tilde{\sigma}_\mu) \) where \( \tilde{\sigma}_\varepsilon \) and \( \tilde{\sigma}_\mu \) denote estimators of \( \sigma_\varepsilon \) and \( \sigma_\mu \), respectively.

### 2.1 The coverage probability of the confidence interval resulting from the two-stage procedure

The coverage probability of the confidence interval constructed from this two-stage procedure is \( P(\beta \in K(\tilde{\sigma}_\varepsilon, \tilde{\sigma}_\mu)) \). The following two theorems (stated by Kabaila et al. 2015) give properties of this coverage probability.

**Theorem 1** For \((\tilde{\sigma}_\varepsilon, \tilde{\sigma}_\mu)\) any of the pairs of estimators listed in Appendix B, the unconditional coverage probability \( P(\beta \in K(\tilde{\sigma}_\varepsilon, \tilde{\sigma}_\mu)) \) is determined by \( N \) (the number of individuals), \( T \) (the number of time points), \( \tilde{\alpha} \) (the nominal significance level of the Hausman
pretest), \(1 - \alpha\) (the nominal coverage probability), \(\psi\) (the ratio \(\sigma_\mu/\sigma_\varepsilon\)), \(\rho\) (the parameter that determines \(G\)) and \(\tau\) (the non-exogeneity parameter). Given these quantities, the coverage probability does not depend on either \(\sigma_\varepsilon^2\) (the variance of the random error) or \(\sigma_\mu^2\) (the variance of the random effect) or \(\sigma_x^2\) (the variance of the time-varying covariate \(x_{it}\)).

**Theorem 2** Suppose that \(N, T, \tilde{\alpha}, 1 - \alpha, \psi\) and \(\rho\) are fixed. When \(\sigma_\varepsilon\) and \(\sigma_\mu\) are replaced by any of the pairs of estimators listed in Appendix B, the unconditional coverage probability \(P(\beta \in K(\hat{\sigma}_\varepsilon, \hat{\sigma}_\mu))\) is an even function of \(\tau \in (-1, 1)\).

Outline proofs of these results are provided by Kabaila et al., 2015. Detailed proofs are provided in the supplementary material for the present paper. A remarkable feature of the proofs of these theorems is that they are carried out without relying on a simple expression for the coverage probability \(P(\beta \in K(\hat{\sigma}_\varepsilon, \hat{\sigma}_\mu))\). We use simulations to compute \(P(\beta \in K(\hat{\sigma}_\varepsilon, \hat{\sigma}_\mu))\), employing variance reduction by control variates, as described in Section 4. Using Theorem 2, we only need to consider \(\tau\) in the interval \([0, 1)\), which means that we have reduced the number of simulations needed to estimate the coverage probability function (or its minimum) by half.

We now examine the influence that the nominal level of significance \(\tilde{\alpha}\) of the Hausman pretest has on the coverage probability function \(P(\beta \in K(\hat{\sigma}_\varepsilon, \hat{\sigma}_\mu))\). Suppose that \(\hat{\sigma}_\varepsilon\) and \(\hat{\sigma}_\mu\) are the usual unbiased estimators of \(\sigma_\varepsilon\) and \(\sigma_\mu\), respectively (described in Appendix B). Consider \(\rho = 0.3\), \(N = 100\), \(T = 3\), \(\psi = \sigma_\mu/\sigma_\varepsilon = 1/3\) and the nominal coverage probability \(1 - \alpha = 0.95\). In practice, it is common to use a small value of \(\tilde{\alpha}\), such as 0.05 or 0.01. As noted by Guggenberger (2010), examples of practical applications that have used a small \(\tilde{\alpha}\) for the Hausman pretest are provided by Gaynor et al. (2005, p.245) and Bedard and Deschenes (2006, p.189). Figure 1 presents graphs of the coverage probability \(P(\beta \in K(\hat{\sigma}_\varepsilon, \hat{\sigma}_\mu))\), considered as a function of \(\lambda = N^{1/2}\tau\). Each graph is computed using the variance reduction method and the common random numbers (to produce smoother graphs) described in Section 4. The number of simulation runs used to compute each graph is \(M = 20000\). The bottom (with circle points) graph is for nominal significance level \(\tilde{\alpha} = 0.05\) of the Hausman pretest. This graph falls well below the nominal coverage for a wide interval of values of \(\lambda\), with the minimum of the coverage probability approximately equal to 0.75. Suppose that we choose the significance level of the Hausman pretest to
be quite large, say \( \tilde{\alpha} = 0.50 \). Now the Hausman pretest is more likely to reject the null hypothesis that \( \tau = 0 \) and therefore more likely to choose the fixed effects model for the construction of the confidence interval. The top (with triangle points) graph is for nominal significance level \( \tilde{\alpha} = 0.50 \) of the Hausman pretest. Although this graph is still below the nominal coverage, there has been a large improvement.

![Graph](image)

Figure 1: Graphs of the coverage probability functions of the confidence interval resulting from the two-stage procedure, for nominal coverage probability 0.95. Here \( \lambda = N^{1/2}\tau \), where \( \tau \) is the non-exogeneity parameter, \( \rho = 0.3 \), \( N = 100 \), \( T = 3 \) and \( \psi = 1/3 \). Two nominal levels of significance, \( \tilde{\alpha} = 0.05 \) and \( \tilde{\alpha} = 0.5 \), of the Hausman pretest are considered.

As noted in the Introduction and in Section 2, if we compute the minimum over \( \tau \) of the coverage probability \( P(\beta \in K(\hat{\sigma}_\varepsilon, \hat{\sigma}_\mu)) \) then we are left with only two unknown parameters, \( \psi \) and \( \rho \). If we fix \( \psi \) then the minimum coverage depends only on \( \rho \), where \( \rho \in (-1, 1) \), as it is a correlation. Suppose that \( \hat{\sigma}_\varepsilon \) and \( \hat{\sigma}_\mu \) are the usual unbiased estimators of \( \sigma_\varepsilon \) and \( \sigma_\mu \), respectively. Suppose that \( N = 100 \), \( T = 3 \), \( \psi = \sigma_\mu/\sigma_\varepsilon = 1/3 \) and the nominal coverage probability \( 1 - \alpha = 0.95 \). Figure 2 presents graphs of the coverage probability
\( P(\beta \in K(\hat{\sigma}_\varepsilon, \hat{\sigma}_\mu)) \), minimized over \( \tau \), considered as a function of \( \rho \) (\( \rho \geq 0 \)). It can be seen that the coverage probability, minimized over \( \tau \), is a decreasing function of \( \rho \). This is because as \( \rho \) increases, \( x_{it} - \pi_i \) becomes closer to 0 for \( t = 1, \ldots, T \) (\( i = 1, \ldots, N \)), causing \( \tilde{\beta}_W \) to become a very inaccurate estimator of \( \beta \). This then causes the test statistic \( H(\hat{\sigma}_\varepsilon, \hat{\sigma}_\mu) \) to have very little power to detect non-zero values of \( \tau \). Each estimate of the minimum coverage is found using the common random numbers and the variance reduction method described in Section 4. Similarly to Figure 1, we see a vast improvement in the minimum coverage by letting \( \tilde{\alpha} = 0.50 \) rather than choosing \( \tilde{\alpha} \) to be the commonly used, smaller value 0.05.

Figure 2: Graphs of the coverage probability functions, minimized over the non-exogeneity parameter \( \tau \), of the confidence interval resulting from the two-stage procedure for nominal coverage probability 0.95. This minimum coverage is considered as a function of \( \rho \), the parameter that determines \( G \). Two nominal levels of significance, \( \tilde{\alpha} = 0.05 \) and \( \tilde{\alpha} = 0.5 \), of the Hausman pretest are considered. Here \( N = 100 \), \( T = 3 \) and \( \psi = 1/3 \).
In practice, ψ is not known and must be estimated from the data. However, it is commonly observed in practice that ρ takes small to moderate values, and not values close to 1. This suggests that we fix ρ and plot the graph of the coverage probability \( P(\beta \in K(\hat{\sigma}_\varepsilon, \hat{\sigma}_\mu)) \), minimized over \( \tau \), as a function of ψ. Suppose that \( \hat{\sigma}_\varepsilon \) and \( \hat{\sigma}_\mu \) are the usual unbiased estimators of \( \sigma_\varepsilon \) and \( \sigma_\mu \), respectively. Consider \( N = 100, T = 3 \), the nominal coverage probability \( 1 - \alpha = 0.95 \) and \( \rho = 0.4 \). Figure 3 presents graphs of the coverage probability \( P(\beta \in K(\hat{\sigma}_\varepsilon, \hat{\sigma}_\mu)) \), minimized over \( \tau \), as a function of ψ. For nominal significance level \( \tilde{\alpha} = 0.05 \) of the Hausman pretest, this minimized coverage probability is far below the nominal coverage for \( \psi \) approximately equal to 0.2. However, for nominal significance level \( \tilde{\alpha} = 0.5 \) of the Hausman pretest, we see (once more) a dramatic improvement in the minimum coverage probability.

2.2 Comparison of the two-stage interval with the adjusted interval based on the fixed effects model

The following notation is introduced to describe the expected length of the two-stage confidence interval. Let \( \bar{x} = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} \), \( \text{SSB} = \sum_{i=1}^{N} (\bar{x}_i - \bar{x})^2 \) (“sum of squares between”) and \( \text{SSW} = \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)^2 \) (“sum of squares within”). Let \( r(x) = \text{SSB}/\text{SSW} \) and \( q(\psi, T) = \psi^2 + (1/T) \). Now define \( w = q(\psi, T)/(q(\psi, T) + r(x)) \) and \( \hat{w} = q(\hat{\psi}, T)/(q(\hat{\psi}, T) + r(x)) \), where \( \hat{\psi} = \hat{\sigma}_\mu/\hat{\sigma}_\varepsilon \).

We use the notation

\[
\mathcal{I}(A) = \begin{cases} 
1 & \text{if } A \text{ is true} \\
0 & \text{if } A \text{ is false}
\end{cases}
\]

where \( A \) is an arbitrary statement. Let \( B \) be the statement \( H(\hat{\sigma}_\varepsilon, \hat{\sigma}_\mu) \leq z_{1-\tilde{\alpha}/2}^2 \). The expected length of \( K(\hat{\sigma}_\varepsilon, \hat{\sigma}_\mu) \) is

\[
2z_{1-\alpha/2} E \left( \hat{\sigma}_\varepsilon (\text{SSW})^{-1/2} \left( \hat{w}^{1/2} \mathcal{I}(B) + \mathcal{I}(B^c) \right) \right).
\]

Values of ρ close to 1 are unlikely to be encountered in practice. We therefore restrict attention to \( \rho \in [0, \tilde{\rho}] \) for some given \( \tilde{\rho} \in (0, 1) \). Let \( c_{\text{min}} \) denote the coverage probability of \( K(\hat{\sigma}_\varepsilon, \hat{\sigma}_\mu) \), minimized over \( \tau \in (-1, 1) \), \( \psi \in (0, \infty) \) and \( \rho \in [0, \tilde{\rho}] \).
Figure 3: Graphs of the coverage probability functions, minimized over the non-exogeneity parameter $\tau$, of the confidence interval resulting from the two-stage procedure for nominal coverage probability 0.95. This minimum coverage is considered as a function of $\psi = \frac{\text{random intercept standard deviation}}{\text{random error standard deviation}}$, where $N = 100$, $T = 3$ and $\rho = 0.4$. Two nominal levels of significance, $\tilde{\alpha} = 0.05$ and $\tilde{\alpha} = 0.5$, of the Hausman pretest are considered.

Let

$$J_c(\sigma_\varepsilon) = \left[\tilde{\beta}_W - \Phi^{-1}((c + 1)/2)(\text{Var}(\tilde{\beta}_W | x))^{1/2}, \tilde{\beta}_W + \Phi^{-1}((c + 1)/2)(\text{Var}(\tilde{\beta}_W | x))^{1/2}\right]$$

$$= \left[\tilde{\beta}_W - \Phi^{-1}((c + 1)/2)\frac{\sigma_\varepsilon}{(SSW)^{1/2}}, \tilde{\beta}_W + \Phi^{-1}((c + 1)/2)\frac{\sigma_\varepsilon}{(SSW)^{1/2}}\right].$$

(6)

The following new theorem allows us to easily compute $P(\beta \in J_c(\tilde{\sigma}_\varepsilon))$ for any given $c \in (0, 1)$.

**Theorem 3** $P(\beta \in J_c(\tilde{\sigma}_\varepsilon))$ does not depend on any unknown parameters.

The proof of Theorem 3 is provided in the supplementary material. Define $c^*$ to be the
value of \( c \) such that \( P(\beta \in J_c(\tilde{\sigma}_\varepsilon)) = c_{\min} \). Let \( J_{\text{adj}}(\tilde{\sigma}_\varepsilon) \) be \( J_c(\sigma_\varepsilon) \) for \( c = c^* \). In other words, \( J_{\text{adj}}(\tilde{\sigma}_\varepsilon) \) is the confidence interval based on the fixed effects model, when \( \sigma_\varepsilon \) is estimated from the data, that is adjusted to have the same minimum coverage as \( K(\tilde{\sigma}_\varepsilon, \tilde{\sigma}_\mu) \). The expected length of \( J_{\text{adj}}(\tilde{\sigma}_\varepsilon) \) is

\[
2\Phi^{-1}\left((c^* + 1)/2\right) E\left(\frac{\tilde{\sigma}_\varepsilon (\text{SSW})^{-1/2}}{E}\right).
\]

The scaled expected length of \( K(\tilde{\sigma}_\varepsilon, \tilde{\sigma}_\mu) \) is defined to be the expected length of \( K(\tilde{\sigma}_\varepsilon, \tilde{\sigma}_\mu) \) divided by the expected length of \( J_{\text{adj}}(\tilde{\sigma}_\varepsilon) \). Let \( \tilde{u}_i = T^{-1} \sum_{t=1}^T u_{it} \). A useful expression for this scaled expected length is given by the following result.

**Theorem 4** The scaled expected length of \( K(\tilde{\sigma}_\varepsilon, \tilde{\sigma}_\mu) \) is equal to

\[
\frac{\Phi^{-1}\left((c^* + 1)/2\right)}{E}\left(\frac{\tilde{\sigma}_\varepsilon (\sum_{i=1}^N \sum_{t=1}^T (u_{it} - \tilde{u}_i)^2)^{-1/2}}{E}\left(\frac{\tilde{w}^{1/2}(B + I(B^c))}{\sum_{i=1}^N \sum_{t=1}^T (u_{it} - \tilde{u}_i)^2}^{-1/2}\right)\right).
\]

A proof of this result is provided in the supplementary material. If (7) is less than 1 then \( K(\tilde{\sigma}_\varepsilon, \tilde{\sigma}_\mu) \) is a shorter interval on average than \( J_{\text{adj}}(\tilde{\sigma}_\varepsilon) \). The following two new theorems describe important properties of the scaled expected length. Outline proofs of these theorems are given in Appendix C. Detailed proofs are provided in the supplementary material.

**Theorem 5** For \((\tilde{\sigma}_\varepsilon, \tilde{\sigma}_\mu)\) any of the pairs of estimators listed in Appendix B, the unconditional scaled expected length given in (7) is determined by \( N \) (the number of individuals), \( T \) (the number of time points), \( \tilde{\alpha} \) (the nominal significance level of the Hausman pretest), \( 1 - \alpha \) (the nominal coverage probability), \( \psi \) (the ratio \( \sigma_\mu/\sigma_\varepsilon \)), \( \rho \) (the parameter that determines \( G \)) and \( \tau \) (the non-exogeneity parameter). Given these quantities, the scaled expected length does not depend on either \( \sigma_\varepsilon^2 \) (the variance of the random error) or \( \sigma_\mu^2 \) (the variance of the random effect) or \( \sigma_x^2 \) (the variance of the time-varying covariate \( x_{it} \)).

**Theorem 6** Suppose that \( N, T, \tilde{\alpha}, 1 - \alpha, \psi \) and \( \rho \) are fixed. When \( \sigma_\varepsilon \) and \( \sigma_\mu \) are replaced by any of the pairs of estimators listed in Appendix B, the unconditional scaled expected length given in (7) is an even function of \( \tau \in (-1,1) \).
Suppose that \( \hat{\sigma}_\varepsilon \) and \( \hat{\sigma}_\mu \) are the usual unbiased estimators of \( \sigma_\varepsilon \) and \( \sigma_\mu \), respectively. Consider the case that \( N = 100, T = 3, 1 - \alpha = 0.95 \) and \( \tilde{\alpha} \in \{0.05, 0.50\} \). We find \( c_{\min} \) by minimizing the coverage probability over \( \tau \in [0, 1), \psi \in (0, \infty) \) and \( \rho \in [0, 0.8] \), since it is reasonable to expect that \( \rho \leq 0.8 \) in practice. For given \( \tilde{\alpha} \) and \( \rho \), we define \( \text{min SEL} \) and \( \text{max SEL} \) to be the scaled expected length minimized and maximized, respectively, with respect to \( \tau \in [0, 1) \) and \( \psi \in (0, \infty) \). For each \( \tilde{\alpha} \in \{0.05, 0.50\} \) and \( \rho \in \{0, 0.2, 0.4, 0.6, 0.8\} \), we have computed \( \text{min SEL} \) and \( \text{max SEL} \). This information is shown in Table 1. From Table 1 it is clear that, for this example, \( J_{\text{adj}}(\hat{\sigma}_\varepsilon) \) is the shorter interval on average. The minimum scaled expected length is a decreasing function of \( \rho \) because as \( \rho \) increases, \( \hat{\omega} \) decreases.

| \( \tilde{\alpha} \) | 0.05 | 0.50 |
|---|---|---|
| \( \rho \) | 0 | 0.2 | 0.4 | 0.6 | 0.8 | 0 | 0.2 | 0.4 | 0.6 | 0.8 |
| min SEL | 4.06 | 3.67 | 3.23 | 2.70 | 1.99 | 1.65 | 1.57 | 1.49 | 1.38 | 1.24 |
| max SEL | 4.88 | 4.88 | 4.87 | 4.89 | 4.89 | 1.81 | 1.81 | 1.81 | 1.81 | 1.81 |

Table 1: Scaled expected lengths, both minimized (min SEL) and maximized (max SEL) over \( \tau \) and \( \psi \), for given values of \( \tilde{\alpha} \) and \( \rho \).

Remark: For the analysis of longitudinal data it is usually more appropriate to consider \( G \) to be the matrix where the \( (i, j) \)’th element is \( \rho^{|i-j|} \) (first order autoregression). Theorems 1, 2, 5 and 6 are true for \( G \) defined in this way. Details on how to define \( \tau \) in this case are provided in the supplementary material.

3 The coverage probability when random error and random effect variances are assumed known

When \( \sigma_\varepsilon \) and \( \sigma_\mu \) are known, the confidence interval resulting from the two-stage procedure is denoted by \( K(\sigma_\varepsilon, \sigma_\mu) \). In this section we describe how the coverage probability of \( K(\sigma_\varepsilon, \sigma_\mu) \), conditional on \( x \), can be computed exactly using the bivariate normal distribution. The results of this section are used in Section 4 to find control variates (used for variance
reduction) for the estimation by simulation of $P(\beta \in K(\tilde{\sigma}, \tilde{\mu}))$ and the scaled expected length (given by (7)).

Let $P(\beta \in K(\sigma, \mu) \mid x)$ denote the coverage probability of $K(\sigma, \mu)$, conditional on $x$. Observe that $P(\beta \in K(\sigma, \mu) \mid x)$ is equal to

$$P(\beta \in I(\alpha, \beta), H(\alpha, \beta) \leq z_{1-\alpha/2} \mid x) + P(\beta \in J(\alpha, \beta), H(\beta, \chi) > z_{1-\alpha/2} \mid x)$$

$$= P(|g_I| \leq z_{1-\alpha/2}, |h| \leq z_{1-\alpha/2} \mid x) + P(|g_J| \leq z_{1-\alpha/2}, |h| > z_{1-\alpha/2} \mid x),$$

where $g_I = (\tilde{\beta}(\psi) - \beta)/(\text{Var}(\tilde{\beta}(\psi)|x))^{1/2}$, $g_J = (\tilde{\beta}_W - \beta)/(\text{Var}(\tilde{\beta}_W|x))^{1/2}$ and $h = (\tilde{\beta}_W - \tilde{\beta}_H)/(\text{Var}(\tilde{\beta}_W|x) + \text{Var}(\tilde{\beta}_H|x))^{1/2}$. By the law of total probability, (8) is equal to the sum of (1 - $\alpha$) and

$$P(|g_I| \leq z_{1-\alpha/2}, |h| \leq z_{1-\alpha/2} \mid x) - P(|g_J| \leq z_{1-\alpha/2}, |h| \leq z_{1-\alpha/2} \mid x).$$

The first and second terms in this expression are determined by the conditional distributions of the random vectors ($g_I$, $h$) and ($g_J$, $h$), respectively. Theorem 7 gives these distributions, which requires the introduction of $p^2(x) = SSB/\text{Var}(\bar{x}_i)$, where $\text{Var}(\bar{x}_i)$ is given in Appendix A. The following theorem is stated by Kabaila et al 2015, together with an outline proof. A detailed proof is proved in the supplementary material for the present paper.

**Theorem 7** Conditional on $x$, ($g_I$, $h$) and ($g_J$, $h$) have bivariate normal distributions, where $E(g_J \mid x) = 0$, $\text{Var}(g_J \mid x) = 1$,

$$E(g_I \mid x) = \frac{\tau \psi p(x)}{(q(\psi, T) + q^2(\psi, T)/r(x))^{1/2}}, \quad \text{Var}(g_I \mid x) = 1 - \frac{\tau^2 \psi^2}{q(\psi, T) + q^2(\psi, T)/r(x)},$$

$$E(h \mid x) = \frac{-\tau \psi p(x)}{(r(x) + q(\psi, T))^{1/2}}, \quad \text{Var}(h \mid x) = 1 - \frac{\tau^2 \psi^2}{r(x) + q(\psi, T)},$$

$$\text{Cov}(g_I, h \mid x) = \frac{\tau^2 \psi^2}{(q(\psi, T)r(x) + q^2(\psi, T))^{1/2}(1 + q(\psi, T)/r(x))^{1/2}}$$

and $$\text{Cov}(g_J, h \mid x) = \frac{1}{(1 + q(\psi, T)/r(x))^{1/2}}.$$
Thus, when $\sigma_\varepsilon$ and $\sigma_\mu$ are known, $P(\beta \in K(\sigma_\varepsilon, \sigma_\mu) \mid x)$ can be found easily by evaluation of the bivariate normal cumulative distribution function in the expression (9). Similarly to Theorem 1 this probability is determined by $N, T, x, \bar{\alpha}, 1 - \alpha, \psi, \rho$ and $\tau$. Note that the dependence on $\rho$ is through $p(x)$. Also, similarly to Theorem 2 $P(\beta \in K(\sigma_\varepsilon, \sigma_\mu) \mid x)$ is an even function of $\tau \in (-1, 1)$. These results may be proved using similar, but much simpler, arguments to those used in the proofs of Theorems 1 and 2 which are given in the supplementary material.

4 Simulation methods, including the use of variance reduction, when the random error and random effect variances are unknown

In Section 3 we described how to find the coverage probability of the confidence interval resulting from the two-stage procedure, conditional on $x$ when $\sigma_\varepsilon$ and $\sigma_\mu$ are known, using the bivariate normal distribution. In the practically important case that $\sigma_\varepsilon$ and $\sigma_\mu$ are replaced by estimators, the results of Section 3 allow us to find control variates (for variance reduction) for the estimation by simulation of both the coverage probability and the scaled expected length. In Section 4.1 we describe the estimation by simulation of the coverage probability and in Section 4.2 we describe the estimation by simulation of the scaled expected length. The simulation methods described in this section apply to any of the pairs of estimators $(\hat{\sigma}_\varepsilon, \hat{\sigma}_\mu)$ listed in Appendix B.

We consider the model (1) and choose the intercept $a = 0$, the parameter of interest $\beta = 0$ and the values for $N, T, \bar{\alpha}, 1 - \alpha, \sigma_\varepsilon^2, \sigma_\mu^2$ and $\sigma_x^2$. Of course, by Theorem 1 and Theorem 5, the coverage probability and the scaled expected length do not depend on either $a, \beta$ or $\sigma_x^2$ and depend on $\sigma_\varepsilon^2$ and $\sigma_\mu^2$ only through $\psi = \sigma_\mu / \sigma_\varepsilon$. Our simulation methods consist of $M$ independent simulation runs. On the $k$‘th simulation run ($k = 1, \ldots, M$), we generate observations of the $\varepsilon_i$‘s and $(\mu_i, x_{i1}, \ldots, x_{iT})$‘s using the assumptions made in Section 2, i.e. the $\varepsilon_i$‘s are i.i.d. $N(0, \sigma_\varepsilon^2)$ and the $(\mu_i, x_{i1}, \ldots, x_{iT})$‘s are i.i.d. with a multivariate normal distribution with mean 0 and covariance matrix (2). As noted in Section 2, in simulations $x$ is determined by $u$, which we write as $x = x(u)$. Let $x^k$ and $u^k$
denote the observed values of \( x \) and \( u \), respectively, for the \( k \)'th simulation run.

4.1 Estimating the coverage probability by simulation

For the observed values in the \( k \)'th simulation run, we compute the following three quantities. The confidence interval resulting from the two-stage procedure, when \( \sigma_\epsilon \) and \( \sigma_\mu \) are assumed known, denoted by \( K_k(\sigma_\epsilon, \sigma_\mu) \). The confidence interval resulting from the two-stage procedure, when \( \sigma_\epsilon \) and \( \sigma_\mu \) are estimated by \( \hat{\sigma}_\epsilon \) and \( \hat{\sigma}_\mu \), respectively, denoted by \( K_k(\hat{\sigma}_\epsilon, \hat{\sigma}_\mu) \). The coverage probability of \( K(\sigma_\epsilon, \sigma_\mu) \), conditional on \( x^k \), when \( \sigma_\epsilon \) and \( \sigma_\mu \) are assumed known, is \( P(\beta \in K(\sigma_\epsilon, \sigma_\mu) \mid x^k) \). Note that this conditional coverage probability is computed exactly using the bivariate normal distributions given in Theorem 7.

Let \( CP = P(\beta \in K(\hat{\sigma}_\epsilon, \hat{\sigma}_\mu)) \), the coverage probability of \( K(\hat{\sigma}_\epsilon, \hat{\sigma}_\mu) \). Now define the unbiased estimator

\[
\hat{CP} = \frac{1}{M} \sum_{k=1}^{M} I(\beta \in K_k(\hat{\sigma}_\epsilon, \hat{\sigma}_\mu))
\]

of \( CP \). This is the usual “brute-force” simulation estimator of \( CP \). We estimate the variance of this estimator by noting that it is a binomial proportion. Let \( CPK = P(\beta \in K(\sigma_\epsilon, \sigma_\mu)) \), the coverage probability of \( K(\sigma_\epsilon, \sigma_\mu) \), when \( \sigma_\epsilon \) and \( \sigma_\mu \) are assumed known. Now define the unbiased estimator

\[
\hat{CPK} = \frac{1}{M} \sum_{k=1}^{M} I(\beta \in K_k(\sigma_\epsilon, \sigma_\mu))
\]

of \( CPK \). By the double expectation theorem, \( CPK = E_x(P(\beta \in K(\sigma_\epsilon, \sigma_\mu) \mid x)) \). Thus another unbiased estimator of \( CPK = P(\beta \in K(\sigma_\epsilon, \sigma_\mu)) \) is

\[
\hat{CPK} = \frac{1}{M} \sum_{k=1}^{M} P(\beta \in K(\sigma_\epsilon, \sigma_\mu) \mid x^k),
\]

which is a much more accurate estimator of \( CPK \) than \( \hat{CP} \).

Define the control variate \( \hat{CP} - \hat{CPK} \), which has expected value zero. The simulation-based unbiased estimator of \( CP = P(\beta \in K(\hat{\sigma}_\epsilon, \hat{\sigma}_\mu)) \) that employs variance reduction using this control variate, is

\[
\tilde{CP} = \hat{CP} - \left( \hat{CPK} - \hat{CPK} \right).
\]

We expect that the correlation between \( \tilde{CP} \) and \( \hat{CP} \) will be close to 1. Since \( \hat{CPK} \) is a much more accurate estimator of \( CPK \) than \( \hat{CPK} \), we expect that the correlation between
\(\hat{\text{CP}}\) and the control variate \(\hat{\text{CP}} - \hat{\text{CP}}\) will also be close to 1. Note that

\[
\hat{\text{CP}} = \frac{1}{M} \sum_{k=1}^{M} \left( I(\beta \in K_{k}(\hat{\sigma}_{\epsilon}, \hat{\sigma}_{\mu})) - I(\beta \in K_{k}(\sigma_{\epsilon}, \sigma_{\mu})) + P(\beta \in K(\sigma_{\epsilon}, \sigma_{\mu}) | x^{k}) \right).
\]

We estimate the variance of this estimator by noting that it is an average of i.i.d. random variables.

We evaluate the efficiency gain of using \(\hat{\text{CP}}\) to estimate the coverage probability \(\text{CP}\) over \(\hat{\text{CP}}\), as follows. Let \(\hat{T}\) and \(\bar{T}\) denote the times taken to carry out \(M\) simulation runs when we estimate \(\text{CP}\) by \(\hat{\text{CP}}\) and \(\hat{\text{CP}}\), respectively. The efficiency of the control variate estimator \(\hat{\text{CP}}\) relative to the “brute-force” estimator \(\hat{\text{CP}}\) is

\[
\frac{\hat{T} \text{Var}(\hat{\text{CP}})}{\bar{T} \text{Var}(\hat{\text{CP}})}.
\]

The larger this relative efficiency, the greater the gain in using the control variate estimator \(\hat{\text{CP}}\), by comparison with using the “brute-force” estimator \(\hat{\text{CP}}\). To give an example of the efficiency gained by using \(\hat{\text{CP}}\) compared to \(\hat{\text{CP}}\), when estimating \(\text{CP}\), we set \(\rho = 0\), \(N = 100\), \(T = 3\), \(\tau = 0\), \(\psi = 1/3\), \(\alpha = \bar{\alpha} = 0.05\) and \(M = 10,000\). We obtain \(\hat{T} = 179.37\) seconds, \(\bar{T} = 211.51\) seconds, \(\text{Var}(\hat{\text{CP}}) = 5.613591 \times 10^{-6}\) and \(\text{Var}(\hat{\text{CP}}) = 1.39591 \times 10^{-6}\). The time ratio is \(\hat{T}/\bar{T} = 0.848045\) and the variance ratio is \(\text{Var}(\hat{\text{CP}}) / \text{Var}(\hat{\text{CP}}) = 4.92597\), so the efficiency of \(\hat{\text{CP}}\) relative to \(\hat{\text{CP}}\) is approximately 4.17. In other words, it would take approximately 4.17 times as long to compute the “brute-force” estimator with the same accuracy as the control variate estimator.

We also use \textit{common random numbers} to create smoother plots of the estimated coverage probability, as a function of \(\lambda\). The estimates of the coverage probability are computed for an equally-spaced grid of values of \(\lambda\). On the \(k\)’th simulation run we generate an observation of \((\mu_{i}, x_{i1}, \ldots, x_{iT})\) from observations of the random numbers \(u_{i1}, \ldots, u_{iT}\) and \(v_{i}\). So, on the \(k\)’th simulation run, for each value of \(\lambda\) in the grid, we use the same random numbers that are used to generate the observations of the \(\varepsilon_{i1}\)’s and the \((\mu_{i}, x_{i1}, \ldots, x_{iT})\)’s. These observations are then used to construct our simulation-based estimate of \(\text{CP}\). Therefore on the \(k\)’th simulation run, for each value of \(\lambda\), we have an estimate of the coverage probability using the same random numbers.
4.2 Estimating the scaled expected length by simulation

When estimating (7) by simulation we consider the expected values in the numerator and
denominator separately. We begin with the expected value in the numerator of (7). Let
\( \hat{L}_K^\dagger = \left( \frac{\hat{\sigma}_\varepsilon}{\sigma_\varepsilon} \right) \left( \sum_{i=1}^N \sum_{t=1}^T (u_{it} - \bar{u}_i)^2 \right)^{-1/2} \) and let \( L_K^\dagger_k \) be the value of \( L_K^\dagger \) when \( u = u_k \). An unbiased simulation estimator of the expected value of \( L_K^\dagger \) that
does not use any variance reduction technique is
\[
\hat{L}_K^\dagger = \frac{1}{M} \sum_{k=1}^M L_K^\dagger_k.
\]

Now let \( L_{KK}^\dagger = \left( \frac{w_1}{2} I(C) + I(C^c) \right) \left( \sum_{i=1}^N \sum_{t=1}^T (u_{it} - \bar{u}_i)^2 \right)^{-1/2} \), where \( C \) is the statement
\(-z_{1-\tilde{\alpha}/2} \leq h \leq z_{1-\tilde{\alpha}/2}\). Also let \( L_{KK}^\dagger_k \) be the value of \( L_{KK}^\dagger \) when \( u = u_k \). Then
\( E(L_{KK}^\dagger | u) \) is equal to
\[
E \left( \frac{w_1}{2} I(C) + I(C^c) \right) \left( \sum_{i=1}^N \sum_{t=1}^T (u_{it} - \bar{u}_i)^2 \right)^{-1/2},
\]
where \( x = x(u) \). We compute \( P(C | x) \) using Theorem [7]. An unbiased simulation estimator
of the expected value of \( L_K^\dagger \) that makes use of a control variate for variance reduction is
\[
\hat{L}_K^\dagger = \frac{1}{M} \sum_{k=1}^M \left( L_K^\dagger_k - (L_{KK}^\dagger_k - E(L_{KK}^\dagger | u_k)) \right).
\]

Next we will consider the expected value in the denominator of (7). Let \( \hat{L}_J^\dagger = \left( \frac{\hat{\sigma}_\varepsilon}{\sigma_\varepsilon} \right) \left( \sum_{i=1}^N \sum_{t=1}^T (u_{it} - \bar{u}_i)^2 \right)^{-1/2} \) and let \( L_J^\dagger_k \) be the value of \( L_J^\dagger \) when \( u = u_k \). An
unbiased simulation estimator of the expected value of \( L_J^\dagger \) that does not use any variance
reduction technique is
\[
\hat{L}_J^\dagger = \frac{1}{M} \sum_{k=1}^M L_J^\dagger_k.
\]

Now let \( L_{JK}^\dagger = \left( \sum_{i=1}^N \sum_{t=1}^T (u_{it} - \bar{u}_i)^2 \right)^{-1/2} \) and let \( L_{JK}^\dagger_k \) be the value of \( L_{JK}^\dagger \) when \( u = u_k \). Note that \( \sum_{i=1}^N \sum_{t=1}^T (u_{it} - \bar{u}_i)^2 \sim \chi^2_{N(T-1)} \). It follows from this that
\( E(L_{JK}^\dagger) = 2^{-1/2} (\Gamma ((N(T-1) - 1)/2) / \Gamma ((N(T-1))/2)) \). An unbiased simulation estimator of the
expected value of $LJ^\dagger$ that makes use of a control variate for variance reduction is

$$\tilde{LJ}^\dagger = \frac{1}{M} \sum_{k=1}^{M} \left( LJ_k - \left( LJK_k - 2^{-1/2} \frac{\Gamma \left( (N(T-1) - 1)/2 \right)}{\Gamma \left( (N(T-1))/2 \right)} \right) \right).$$

Therefore, a simulation estimator of the scaled expected length that makes use of control variates for variance reduction is

$$z_{1-\alpha/2} \frac{\tilde{LK}^\dagger}{\Phi^{-1}((c^* + 1)/2)} \frac{\tilde{LK}^\dagger}{\tilde{LJ}^\dagger}.$$

We estimate the variance of $\tilde{LK}^\dagger$, $\tilde{LK}^\dagger$, $\tilde{LJ}^\dagger$ and $\tilde{LJ}^\dagger$ by noting that each estimator is an average of i.i.d. random variables. Using R we can assess the relative efficiency of the estimators that use control variates for variance reduction with their “brute-force” counterparts using a similar method to that discussed in Section 4.1. For example, for known quantities $N = 100, T = 3, \tilde{\alpha} = 0.05$ and unknown quantities $\tau = 0, \rho = 0, \sigma_\epsilon = 1, \sigma_\mu = 1$ and $\sigma_x = 1$, and for $M = 1000$ independent simulation runs, we find that the efficiency of $\tilde{LK}^\dagger$ relative to $\tilde{LK}^\dagger$ is approximately 1 and the efficiency of $\tilde{LJ}^\dagger$ relative to $\tilde{LJ}^\dagger$ is approximately 2. However, if we increase $\rho$ to 0.8 (leaving the other parameters unchanged), the efficiency of $\tilde{LK}^\dagger$ relative to $\tilde{LK}^\dagger$ increases to approximately 3. Hence, estimating the scaled expected length using control variates is well justified.

Common random numbers are also used for the estimation by simulation of the scaled expected length. This is done in a similar way to the method described at the end of Section 4.1.

**Remark:** In this paper we consider the two-stage procedure when the inference of interest is a confidence interval for the slope parameter. As one would expect from the duality between hypothesis tests and confidence intervals, our results have important implications when the subsequent inference is a hypothesis test for the slope parameter. These implications are described in detail in the supplementary material.

## 5 Conclusion

The Hausman pretest is an example of preliminary statistical (i.e. data based) model selection. Other examples include model selection by minimizing a criterion such as the Akaike
Information Criterion or the Bayesian Information Criterion. The effects of preliminary statistical model selection on confidence intervals can range from the benign to the very harmful, depending on the class of models under consideration, the known aspects of the model, the parameter of interest and the model selection procedure employed (Kabaila, 2009 and Kabaila and Leeb, 2006). In other words, each case needs to be considered individually on its merits. Our results show that for the small levels of significance (such as 5% or 1%) of the Hausman pretest commonly used in applications, the minimum coverage probability of the confidence interval for the slope parameter with nominal coverage probability \(1 - \alpha\) can be far below nominal. The methodology that we have described makes it easy to assess, for a wide variety of circumstances, the effect of the Hausman pretest on the minimum coverage probability of this confidence interval. An interesting finding is that if we increase the significance level of the Hausman pretest to, say, 50% then this minimum coverage probability is much closer to the nominal coverage \(1 - \alpha\) for a wide range of parameters. However, regardless of the level of significance of the Hausman pretest, this interval is wider on average than the interval based on the fixed effects model which is adjusted to have the same minimum coverage. Therefore, the Hausman pretest should not be used in practice to choose between the random effects model and fixed effects model for clustered or longitudinal data when the subsequent inference of interest is a confidence interval for the slope parameter.

Appendix A. Definition of the non-exogeneity parameter \(\tau\)

It may be shown that the distribution of \(\mu_i\) conditional on \((x_{i1}, \ldots, x_{iT})\) is normal with mean

\[
\frac{\sigma_\mu \tilde{\tau} T}{(1 + (T - 1)\rho) \sigma_x \bar{x}_i},
\]

where \(\bar{x}_i = T^{-1} \sum_{t=1}^{T} x_{it}\), and variance

\[
\sigma_\mu^2 \left(1 - \frac{\tilde{\tau}^2 T}{1 + (T - 1)\rho}\right).
\]

This suggests that \(\tau = \text{Corr}(\mu_i, \bar{x}_i)\) is a reasonable measure of the dependence between \(\mu_i\) and \((x_{i1}, \ldots, x_{iT})\) i.e. that it is reasonable to designate \(\tau\) as the non-exogeneity parameter.
It may be shown that $\text{Var}(x_i) = \sigma^2 x (1 + (T-1)\rho) / T$ and $\text{Cov}(\mu_i, \bar{x}_i) = \tilde{\tau} \sigma_x \sigma_x$. Thus

$$\tau = \tilde{\tau} \left( \frac{T}{1 + (T-1)\rho} \right)^{1/2}.$$ 

**Appendix B. Description of the estimators of the random error and random effect variances**

It has been suggested in the literature (see e.g. Hsiao, 1986 and Baltagi, 2005) that if a negative estimate of variance is observed then one should do as Maddala and Mount (1973) suggest and replace this negative estimate by 0. We use this kind of approach to ensure that $\hat{\sigma}^2_\varepsilon$ is always positive and $\hat{\sigma}^2_\mu$ is always nonnegative. This ensures that the proofs of Theorems 1, 2, 5 and 6 carry through for each of the three pairs of estimators that we consider in this paper. We consider the following pairs of estimators of $\sigma^2_\varepsilon$ and $\sigma^2_\mu$:

1. The usual unbiased estimators. Define

$$\tilde{\sigma}^2_\varepsilon = \frac{1}{N(T-1) - 1} \sum_{i=1}^{N} \sum_{t=1}^{T} r_{it}^2$$

and $\tilde{\sigma}^2_\mu = \max(0, \tilde{\sigma}^2_\mu)$, where

$$\tilde{\sigma}^2_\mu = \frac{1}{N-2} \sum_{i=1}^{N} \tilde{r}_i^2 - \frac{1}{NT(T-1) - T} \sum_{i=1}^{N} \sum_{t=1}^{T} r_{it}^2.$$ 

The $r_{it}$’s are the OLS residuals from model (4) and the $\tilde{r}_i$’s are the OLS residuals from model (3). Note that $\tilde{\sigma}^2_\mu$ is an unbiased estimator of $\sigma^2_\mu$ only for $\tau = 0$.

2. Hsiao’s (1986) maximum likelihood estimators $\hat{\sigma}^2_\varepsilon$ and $\hat{\sigma}^2_\mu$. We assume, of course, that the maximum likelihood estimator is obtained by maximizing the log-likelihood function subject to the parameter constraints $\sigma^2_\varepsilon \geq 0$ and $\sigma^2_\mu \geq 0$.

3. Wooldridge’s (2002) estimators. Define $\tilde{\sigma}^2_\varepsilon = \max(-\epsilon \tilde{\sigma}^2_\varepsilon, \tilde{\sigma}^2_\varepsilon)$ where $\epsilon$ is a very small positive number and

$$\tilde{\sigma}^2_\varepsilon = \frac{1}{NT - K} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{r}_{it}^2 - \frac{1}{NT(T-1)/2 - K} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \tilde{r}_{it} \tilde{r}_{is}.$$ 

Also define $\tilde{\sigma}^2_\mu = \max(0, \tilde{\sigma}^2_\mu)$ where

$$\tilde{\sigma}^2_\mu = \frac{1}{NT(T-1)/2 - K} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \tilde{r}_{it} \tilde{r}_{is}.$$ 

22
Here, the \( \hat{r}_{it} \)’s are the residuals from pooled OLS estimation for the model (1) and \( K = 0 \) (no d.o.f. correction) or \( K = 2 \) (d.o.f. correction).

**Appendix C. Outline proofs of Theorems 5 and 6**

Outline proofs of Theorems 1, 2 and 7 have been provided by Kabaila et al., 2015. Here we provide outline proofs of Theorems 5 and 6. Detailed proofs of Theorems 1 – 7 can be found in the supplementary material. Theorems 5 and 6 hold for the three test statistics given by Hausman and Taylor (1981) and the three pairs of estimators described in Appendix B. For the sake of brevity, the outline proofs of these results are given only for the Hausman test statistic \( H(\hat{\sigma}_\varepsilon, \hat{\sigma}_\mu) \) and the unbiased estimators of \( \sigma_\varepsilon \) and \( \sigma_\mu \) described in Appendix B.

**Proof of Theorem 5**

Let \( \hat{h} \) denote the statistic \( h \), when \( \sigma_\varepsilon \) and \( \sigma_\mu \) are replaced by the unbiased estimators \( \hat{\sigma}_\varepsilon \) and \( \hat{\sigma}_\mu \), respectively. Then \( \mathcal{B} \) is equivalent to the statement that \( -z_{1-\bar{\alpha}/2} \leq \hat{h} \leq z_{1-\bar{\alpha}/2} \).

Let \( x_{it}^\dagger = x_{it} / \sigma_x \), \( \varepsilon_{it}^\dagger = \varepsilon_{it} / \sigma_\varepsilon \) and \( \mu_i^\dagger = \mu_i / \sigma_\mu \). The joint distribution of the \( \varepsilon_{it}^\dagger \)'s and the \( (\mu_i, x_{i1}, \ldots, x_{iT})' \)'s is determined by \( \rho \) and \( \tau \). Now express \( \hat{h} \) and \( \hat{\sigma}_\varepsilon / \sigma_\varepsilon \) in terms of the \( x_{it}^\dagger \)'s, \( \varepsilon_{it}^\dagger \)'s, \( \mu_i^\dagger \)'s and \( \psi \). Since \( c_{\min} \) and the joint distribution of the \( u_{it} \)'s do not depend on any parameters, it follows that (7) is determined by the known quantities \( N, T, 1 - \alpha \) and \( \bar{\alpha} \), and unknown parameters \( \psi, \rho \) and \( \tau \).

**Proof of Theorem 6**

Assume that \( (\mu_i, x_{i1}, \ldots, x_{iT}) \) has a multivariate normal distribution with mean 0 and covariance matrix (2). Then \( \bar{\tau} = \tau ((1 + (T - 1)\rho)) / T^{1/2} \). By Theorem 5, the scaled expected length is a function of \( \tau \). Let \( x_{it}^* = -x_{it} \) for \( i = 1, \ldots, N \) and \( t = 1, \ldots, T \). For \( \tau = d \), \( (\mu_i, x_{i1}, \ldots, x_{iT}) \) has a multivariate normal distribution with mean 0 and covariance matrix (2), where \( \bar{\tau} = d ((1 + (T - 1)\rho)) / T^{1/2} \). For \( \tau = -d \), \( (\mu_i, x_{i1}^*, \ldots, x_{iT}^*) \) has the same distribution. It can be shown that the scaled expected length is the same function of the \( x_{it}^* \)'s, \( \varepsilon_{it}^* \)'s and \( \mu_i^* \)'s as it is of the \( x_{it} \)'s, \( \varepsilon_{it} \)'s and \( \mu_i \)'s. Hence the scaled expected length is an even function of \( \tau \).
SUPPLEMENTARY MATERIAL

Supplementary Appendix: A remark on the two-stage procedure when the inference of interest is a hypothesis test and a remark on the choice of the scaling $\lambda = N^{1/2}\tau$. This supplementary appendix also includes detailed proofs of Theorems 1 – 7 and the definition of the non-exogeneity parameter $\tau$ in the case of first order autoregression. (PDF file)

References

Ajmani, V.B., 2009. Applied Econometrics Using the SAS system. John Wiley, Hoboken, N.J.

Baltagi, B.H., 2005. Econometric Analysis of Panel Data, 3rd edition. John Wiley & Sons, Ltd.

Bedard, K., Deschenes, O., 2006. The long-term impact of military service on health: Evidence from work war II and Korean war veterans. American Economic Review 96, 176-194.

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

Croissant, Y., Millo, G., 2008. Panel data econometrics in R: The plm package. Journal of Statistical Software, 27(2), 1-43.

Ebbes, P., Bockenholt, U., Wedel, M., 2004. Regressor and random-effects dependencies in multilevel models. Statistica Neerlandica 58, 161–178.

Gardiner, J.C., Luo, Z., Roman, L.A., 2009. Fixed effects, random intercepts and GEE: What are the differences? Statistics in Medicine 28, 221–239.

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

Griffiths, W. E., Hill, R. C., Lim, G., 2012. Using EViews for Principles of Econometrics. Danvers, MA: John Wiley & Sons.

Guggenberger, P., 2010. The impact of a Hausman pretest on the size of a hypothesis test: The panel data case. Journal of Econometrics 156, 337–343.

Hastings, J.S., 2004. Vertical relationships and competition in retail gasoline markets: Empirical evidence from contract changes in Southern California. American Economic Review 94, 317–328.
Hsiao, C., 1986. Analysis of Panel Data. Cambridge University Press, Cambridge.
Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46, 1251–1271.
Hausman, J.A., Taylor, W.E., 1981. Panel data and unobservable individual effects. Econometrica 49, 1377–1398.
Jackowicz, K., Kowalewski, O., Kozlowski, L., 2013. The influence of political factors on commercial banks in Central European countries. Journal of Financial Stability 9, 759–777.
Kabaila, P., 2009. The coverage properties of confidence regions after model selection. International Statistical Review 77, 405–414.
Kabaila, P., Leeb, H., 2006. On the large-sample minimal coverage probability of confidence intervals after model selection. Journal of the American Statistical Association 101, 619–629.
Kabaila, P., Mainzer, R., Farchione, D., 2015. The impact of a Hausman pretest, applied to panel data, on the coverage probability of confidence intervals. Economics Letters 131, 12–15.
Maddala, G.S., Mount, T.D., 1973. A comparative study of alternative estimators for variance component models used in econometric applications. Journal of the American Statistical Association 68, 324–328.
Mann, V., De Stavola, B.L., Leon, D.A., 2004. Separating within and between effects in family studies: an application to the study of blood pressure in children. Statistics in Medicine 23, 2745–2756.
Rabe-Hesketh, S., Skrondal, A., 2012. Multilevel and Longitudinal Modeling Using Stata, 3rd edition. Stata Press, Texas.
Wooldridge, J. M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge.