

Schäper, Julius; Winkelmann, Rainer

Working Paper

Random effects panel data models with known heteroskedasticity

Working Paper, No. 445

Provided in Cooperation with:

Department of Economics, University of Zurich

Suggested Citation: Schäper, Julius; Winkelmann, Rainer (2024) : Random effects panel data models with known heteroskedasticity, Working Paper, No. 445, University of Zurich, Department of Economics, Zurich,
<https://doi.org/10.5167/uzh-259907>

This Version is available at:

<https://hdl.handle.net/10419/298812>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



**University of
Zurich**^{UZH}

University of Zurich
Department of Economics

Working Paper Series
ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 445

Random Effects Panel Data Models with Known Heteroskedasticity

Julius Schäper and Rainer Winkelmann

May 2024

Random effects panel data models with known heteroskedasticity*

JULIUS SCHÄPER
University of Zurich

RAINER WINKELMANN
University of Zurich

May 2024

Abstract

The paper introduces two estimators for the linear random effects panel data model with known heteroskedasticity. Examples where heteroskedasticity can be treated as given include panel regressions with averaged data, meta regressions and the linear probability model. While one estimator builds on the additive random effects assumption, the other, which is simpler to implement in standard software, assumes that the random effect is multiplied by the heteroskedastic standard deviation. Simulation results show that substantial efficiency gains can be realized with either of the two estimators, that they are robust against deviations from the assumed specification, and that the confidence interval coverage equals the nominal level if clustered standard errors are used. Efficiency gains are also evident in an illustrative meta-regression application estimating the effect of study design features on loss aversion coefficients.

Keywords: Generalized least squares, linear probability model, meta regression

JEL classification: C23

* Department of Economics, University of Zurich, E-mail of corresponding author: rainer.winkelmann@econ.uzh.ch.

1 Introduction

There is a substantial literature dealing with heteroskedasticity in the linear random effects model for panel data. Part of it has been concerned with heteroskedastic time-invariant error components (e.g. Mazodier and Trognon (1978) and Baltagi and Griffin (1988)), another part with time-invariant heteroskedasticity of the general error term (e.g. Rao et al. (1981), Baltagi (1988) and Wansbeek (1989)) and yet another part with modelling heteroskedasticity as a low-dimensional parametric function of covariates (e.g. Baltagi et al. (2006) and Lejeune (2005)).

Here, we consider a different situation, where the variances of the general errors are left completely unrestricted, but treated as given. Knowledge of the idiosyncratic error variances (or consistent estimators thereof) leads to a considerable simplification of the estimation problem, as the feasible generalised least squares (GLS) estimator of the random effects panel model requires estimation of one additional parameter only, the variance of the time-invariant individual error component. In this way, over-parametrization is avoided and the estimation problem is reduced to finding an efficient, and hopefully also conveniently implementable, GLS estimator for the regression parameters. Also, there is no reason to consider tests for heteroskedasticity, as provided by Holly and Gardiol (2000), Baltagi et al. (2006) or Lejeune (2005), among others, since it is known to be present in our set-up.

It is not uncommon in applied work that the source of heteroskedasticity can be treated as known. One application concerns state and year panel data for proportions or means, where heteroskedasticity is generated by variation in sample sizes, both over time and across states. Another growing field of application is in meta regressions where effect sizes are predicted by study features. In this case, heteroskedasticity is due to variation in standard errors of the estimated effect sizes. A panel dimension can arise if the meta regression includes several estimates from the same study, as is frequently the case in practice.¹ Yet another example is the linear probability model for a binary dependent variable, where the heteroskedastic

¹See Stanley and Doucouliagos (2008) for an introduction to meta analysis for economists.

error variances can be consistently estimated because they are a known function of the linear predictor.²

We discuss two versions of the random effects model with known heteroskedasticity. In the standard version, the variance of the random effect is constant, and heteroskedasticity in the general error term leads to variation in the pairwise within-unit correlations. In a modified model, the random effect is multiplied by the standard deviation of the error term, and thus effectively becomes heteroskedastic as well, since its variance is proportional to that of the general error. In this “scaled random effects model”, the pairwise within-unit correlations remain constant, and we consider this a useful benchmark, as it mirrors the correlation structure implied by the homoskedastic random effects model.

In principle, there are two different ways each for implementing the GLS estimator for the two heteroskedastic random effects models, either by using the inverse of the block-diagonal covariance matrix of the combined errors, or by applying ordinary least squares to suitably transformed data. For the scaled random effects model, there is a particularly simple approach: divide the data by the standard deviation of the general error, and use a standard random effects estimator, available in any econometric software, on the transformed data. No additional programming effort is necessary.

While both models account for dependent observations, and therefore make the estimators potentially more efficient, there is no need to assume “correct specification”, and we recommend using cluster-robust standard errors in addition to weighting (see e.g. Romano and Wolf (2017) for a general discussion of robust weighted least squares estimation, and Pustejovsky and Tipton (2022) for robust estimation in the context of multi-study meta regressions).

In numerical simulation experiments, we document that substantial efficiency gains can be achieved, and correctly sized tests be obtained, by using GLS estimators that account for full heteroskedasticity as well as within-unit correlation, even if the weighting matrix does

²Similarly, heteroskedasticity of known form arises when estimating a linear model for the conditional expectation of a Poisson distributed dependent variable.

not correspond to the true data generating process. The gains are similar for both working models, and their related estimators, supporting the case for the easier-to-use scaled random effects approach that assumes that the variance of the time-invariant error component is proportional to the variance of the general error.

We also conduct an illustrative application to meta-regression. Brown et al. (2024) estimate the effect of study design features on loss aversion coefficients, using 607 estimates from 150 articles. Loss aversion estimates are heteroskedastic, since their standard errors differ depending on sample size and other features of the sampling process. Moreover, an article contains several estimates which can be expected to be correlated. Accounting for correlation and heteroskedasticity, we obtain substantially more precisely estimated meta-regression coefficients than without those adjustments. Consequently, the evidence for the presence of design effects is much stronger than would be inferred from a simple OLS or random effects estimator.

2 Random effects models with heteroskedasticity

2.1 Basic setup

To introduce notation and as a point of reference for our extensions, we first restate the textbook results on the random effects model without heteroskedasticity (e.g., Baltagi, 2008). Let

$$y_i = X_i\beta + u_i\iota_{T_i} + \varepsilon_i, \quad i = 1, \dots, N \quad (1)$$

where y_i denotes the $(T_i \times 1)$ vector of outcomes and X_i the $(T_i \times p)$ design matrix that includes a constant. u_i is a scalar random effect, ι_{T_i} a $(T_i \times 1)$ vector of ones, and ε_i a $(T_i \times 1)$ vector of general error terms. There are N independent units. We do not impose the assumption of a balanced panel, so the total number of observations equals $\sum_{i=1}^N T_i$. Both error components are assumed to be homoskedastic, $\text{Var}(u_i) = \tau^2$ and $\text{Var}(\varepsilon_{it}) = \sigma^2$. The

covariance matrix Ω_i of the combined error term $u_i + \varepsilon_i$ can then be written as

$$\Omega_i = (T_i\tau^2 + \sigma^2)\bar{J}_{T_i} + \sigma^2(I_{T_i} - \bar{J}_{T_i}) \quad (2)$$

where $T_i\bar{J}_{T_i}$ is a $(T_i \times T_i)$ matrix of ones, and I_{T_i} is the identity matrix (see Baltagi, 2008). Since \bar{J}_{T_i} and $(I_{T_i} - \bar{J}_{T_i})$ are idempotent and orthogonal, the inverse is given by $\Omega_i^{-1} = (T_i\tau^2 + \sigma^2)^{-1}\bar{J}_{T_i} + \sigma^{-2}(I_{T_i} - \bar{J}_{T_i})$ and the square root of its inverse by $\Omega_i^{-1/2} = (T_i\tau^2 + \sigma^2)^{-1/2}\bar{J}_{T_i} + \sigma^{-1}(I_{T_i} - \bar{J}_{T_i})$. The inverse of the full covariance matrix of $Y = (y_1, \dots, y_N)'$ is block-diagonal with typical element Ω_i^{-1} , and it follows that the GLS estimator β_{GLS} is given by

$$\begin{aligned} \hat{\beta}_{GLS} &= (X'WX)^{-1}X'WY \\ &= \left(\sum_{i=1}^N X_i'W_iX_i \right)^{-1} \left(\sum_{i=1}^N X_i'W_iy_i \right) \end{aligned} \quad (3)$$

where the weighting matrix $W_i = \Omega_i^{-1}$ depends on two parameters only, τ^2 and σ^2 .

Alternatively, the GLS estimator can be conveniently obtained by first transforming the data, observation-by-observation. Premultiplying Y and X by $\sigma\Omega^{-1/2}$, we find that Y^* has typical element $y_{it}^* = y_{it} - \theta_i\bar{y}_i$ where $\theta_i = 1 - \sigma/\sqrt{T_i\tau^2 + \sigma^2}$ and \bar{y}_i is the unit specific mean. The GLS estimator equals the least squares estimator expressed in terms of the transformed data Y^* and X^* . For feasible GLS, τ^2 and σ^2 need to be replaced by consistent estimators $\hat{\tau}^2$ and $\hat{\sigma}^2$.

The GLS formula (3) is completely general, of course, and the heteroskedastic random effects estimators discussed below will only differ in the choice of weighting function W_i . Importantly, under correct specification, when $W_i = \Omega_i^{-1}$, the conditional variance covariance matrix is given by

$$\text{Var}_{GLS}(\hat{\beta}|X) = \left(\sum_{i=1}^N X_i'\Omega_i^{-1}X_i \right)^{-1},$$

and the unconditional variance covariance matrix can be consistently estimated by

$$\text{Var}_{CR}(\hat{\beta}) = \left(\sum_{i=1}^N X_i'W_iX_i \right)^{-1} \left(\sum_{i=1}^N X_i'W_i\hat{e}_i\hat{e}_i'W_iX_i \right) \left(\sum_{i=1}^N X_i'W_iX_i \right)^{-1},$$

where $\hat{e}_i = y_i - X_i \hat{\beta}$ is the vector of unit specific residuals, and the subscript CR stands for “cluster-robust”.

2.2 Introducing heteroskedasticity

The simplifying factorization (2) goes through if u_i is heteroskedastic with variance τ_i , or if the variance of ε_{it} differs across units but is constant within units, such that we can write $\text{Var}(\varepsilon_{it}) = \sigma_i^2$. In this case, the Ω_i submatrices differ not only in the T_i -dimension, but also with respect to the parameters τ_i^2 and σ_i^2 . However, with a sufficiently large T_i , these parameters can be estimated (see Baltagi, 2008, for further detail).

In this paper, we deal with a different case, namely known time- and unit-dependent variances of the general error term, referred to in the following as “full heteroskedasticity”. Consequently, for unit i , the variance-covariance matrix of the general errors is of the form $\Sigma_i = \text{diag}(\sigma_{i1}, \dots, \sigma_{iT_i})$, a $(T_i \times T_i)$ diagonal matrix. For $\tau^2 = 0$ (or in a cross-section application), we would have $\Omega_i = \Sigma_i^2$, and the GLS estimator is obtained by dividing each variable by the standard deviation of the general error that is assumed to be known. However, this transformation ignores the within-unit correlations that may arise from a common unit-specific random effect, or from other forms of correlated errors. As a consequence, simply weighting by the inverse standard deviation cannot be efficient.

We consider two approaches to account for potential within-unit error correlation. A natural first step would be to augment the model by a simple constant-variance random effect. It turns out that such a random effect, in combination with a heteroskedastic general error, implies that high-variance observations have low correlation. Moreover, the model cannot be estimated using standard panel data software. Both issues are addressed by an alternative, scaled random effects specification, that enforces a common within-unit correlation and can be estimated using standard panel data software.

2.2.1 The standard random effects model

Using observation-specific notation, the model can be written as

$$y_{it} = x'_{it}\beta + u_i + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T_i \quad (4)$$

where $\text{Var}(u_i) = \tau^2$ and $\text{Var}(\varepsilon_{it}) = \sigma_{it}^2$. It follows that

$$\text{Var}(y_{it}) = \tau^2 + \sigma_{it}^2$$

$$\text{Cov}(y_{it}, y_{is}) = \tau^2 \quad \text{for } t \neq s$$

and

$$\text{Cor}(y_{it}, y_{is}) = \rho_{its} = \frac{\tau^2}{\sqrt{\tau^2 + \sigma_{it}^2} \sqrt{\tau^2 + \sigma_{is}^2}} \quad \text{for } t \neq s$$

which means that correlations vary unless $\sigma_{it}^2 = \sigma_{is}^2$ for all $t \neq s$. In fact, for a given τ^2 , the correlation is smaller the larger the variances σ_{it}^2 and σ_{is}^2 . This may be an unwanted implication in some application. For example, with correlated errors ε_{it} and ε_{is} , the effect of an increase in the error variances on $\text{Cor}(y_{it}, y_{is})$ becomes ambiguous.

In principle, the GLS estimator (3) can be computed by first inverting each block of the covariance matrix, Ω_i . The inverse of Ω can then be substituted as weighting matrix W into the formula for $\hat{\beta}_{GLS}$. Based on (4), the marginal variance-covariance matrices can be written as

$$\Omega_i = \tau^2 J_i + \Sigma_i^2 \quad (5)$$

where Σ_i^2 is an $T_i \times T_i$ diagonal matrix with i^{th} diagonal entry equal to σ_{ij}^2 and J_i is a $T_i \times T_i$ matrix of ones. The inverses of these matrices have the form

$$\Omega_i^{-1} = \Sigma_i^{-2} - \delta_i \Sigma_i^{-2} J_i \Sigma_i^{-2}, \quad \text{where} \quad \delta_i = \left(\frac{1}{\tau^2} + \sum_{j=1}^{T_i} \frac{1}{\sigma_{ij}^2} \right)^{-1}. \quad (6)$$

Feasible least squares requires estimation of a single parameter, τ^2 , and a natural estimator uses the residuals $\hat{\varepsilon}_{it}$ from an unweighted least squares regression of y_{it} on x_{it} and estimates τ^2 from the off-diagonal elements:

$$\hat{\tau}^2 = \frac{1}{\sum_{i=1}^N T_i(T_i - 1)/2 - p} \sum_{i=1}^N \sum_{t=1}^{T_i-1} \sum_{t'=t+1}^{T_i} \hat{\varepsilon}_{it} \hat{\varepsilon}_{it'} \quad (7)$$

A GLS transformation for this case has been derived by Randolph (1998), allowing in principle to obtain the GLS estimator as an ordinary least squares estimator applied to the transformed data. However, the transformation is quite complicated, and there seem to be no empirical applications.

2.2.2 A scaled random effects model

Let

$$y_{it} = x'_{it}\beta + \sigma_{it}u_i + \varepsilon_{it} \quad (8)$$

The difference to model (4) is that the random effect is now multiplied by the standard deviation of ε_{it} . To the best of our knowledge, such a model has not previously been discussed in the panel data literature, with the exception of Stanley and Doucouliagos (2008) in the related context of meta-regressions.³ In one interpretation, u_i can be thought of as a common factor and σ_{it} as the factor loading. It follows that

$$\text{Var}(y_{it}) = \tau^2 \sigma_{it}^2 + \sigma_{it}^2 = \sigma_{it}^2(\tau^2 + 1)$$

$$\text{Cov}(y_{it}, y_{is}) = \tau^2 \sigma_{it} \sigma_{is} \quad \text{for } t \neq s$$

and

$$\rho_{its} = \frac{\tau^2 \sigma_{it} \sigma_{is}}{\sqrt{\sigma_{it}^2 \sigma_{is}^2 (\tau^2 + 1)^2}} = \frac{\tau^2}{\tau^2 + 1} \quad \text{for } t \neq s$$

so the model shares many properties of the standard random effects model: the within-unit correlations ρ_{its} are constant within units, and they are increasing in τ^2 . The standard model is nested when $\sigma_{it}^2 = \sigma^2 = 1$ for all i , t and s .

The scaling of the random effect makes intuitively sense in cases where the general error is heteroskedastic because of a bounded support of the outcome. For example, when y_{it} is

³Typically, in a meta-regression, σ_{it} itself is included as a regressors to account for publication bias. Stanley and Doucouliagos (2008, p. 69) argue that such a scaled random effects specification is reasonable when "study effects operate largely through an unobserved differential propensity to select for statistical significance."

non-negative, the variance of the error must decrease as the linear index $x'_{it}\beta$ approaches zero. Similarly, for binary or fractional responses, the variance must decrease as predictions approach the upper or lower bound of the unit interval. But what holds for ε_{it} must also hold for the additive random effect, and the σ_{it}^2 -scaling is a simple way to ensure that the variance of the product $u_i\sigma_{it}$ decreases, and eventually goes to zero, as the regression function approaches the respective bounds.

Model (8) can be implemented using standard software, after appropriate weighting, as long as σ_{it}^2 is known (at least up to a multiplicative constant). Divide (8) by σ_{it} :

$$\frac{y_{it}}{\sigma_{it}} = \frac{x'_{it}}{\sigma_{it}}\beta + u_i + \frac{\varepsilon_{it}}{\sigma_{it}} \quad (9)$$

So this is a homoskedastic random effects model for the transformed variables $y_{it}^* = y_{it}/\sigma_{it}$ and $x_{it}^* = x_{it}/\sigma_{it}$. Therefore, estimation requires only two simple steps. In a first step, the transformed variables y_{it}^* and x_{it}^* are obtained. In a second step, the usual random effects transformation (Baltagi, 2008) is applied to y_{it}^* and x_{it}^* . In particular, it can be shown that the random effects estimator is obtained by quasi-demeaning the dependent and independent variables, i.e., regressing

$$y_{it}^* - \hat{\theta}\bar{y}_i^* \text{ on } x_{it}^* - \hat{\theta}\bar{x}_i^*$$

where $\hat{\theta} = 1 - 1/\sqrt{1 + T\hat{\tau}^2}$ and $\hat{\tau}$ is obtained as in (7), using the residuals from a regression of y_{it}^* on x_{it}^* .

2.2.3 The linear probability model

In a panel linear probability model (LPM) with random effects, we can write

$$\Pr(y_{it} = 1|x_{it}, u_i) = x'_{it}\beta + u_i \quad (10)$$

In order to ensure consistent estimation of β , we need to assume that the restriction $0 \leq x'_{it}\beta + u_i \leq 1$ holds (see Theorem 1 in Horrace and Oaxaca, 2006). Re-writing (10) in

regression format, and assuming an additive error specification without scaling of the random effect, we obtain

$$y_{it} = x'_{it}\beta + u_i + \varepsilon_{it}$$

where $\varepsilon_{it} = 1 - (x'_{it}\beta + u_i)$ when $y_{it} = 1$, $\varepsilon_{it} = -(x'_{it}\beta + u_i)$ when $y_{it} = 0$, and therefore $-1 \leq \varepsilon_{it} \leq 1$.

This is not a typical linear random effects panel data model, because the requirement $0 \leq x'_{it}\beta + u_i \leq 1$ imposes restrictions on both the distribution of x_{it} and that of u_i . To give an example, the model is well defined if x_{it} is uniform between 0.3 and 0.7, with zero constant and unit slope, and u_i is uniform between -0.2 and +0.2; predictor distributions with unbounded support are naturally ruled out in this context.

Under these assumptions, we can derive the implied skedastic function of the general error term ε_{it} . From

$$\text{Var}(\varepsilon_{it}|x_{it}, u_i) = (x'_{it}\beta + u_i)(1 - (x'_{it}\beta + u_i)), \quad (11)$$

it follows that

$$\begin{aligned} \text{Var}(\varepsilon_{it}|x_{it}) &= E_u[\text{Var}(\varepsilon_{it}|x_{it}, u_i)] \\ &= E_u[(x'_{it}\beta + u_i)(1 - (x'_{it}\beta + u_i))|x_{it}] \\ &= E_u[x'_{it}\beta(1 - x'_{it}\beta) - 2x'_{it}\beta u_i + u_i - u_i^2|x_{it}] = v_{it}^2 - \tau^2 \end{aligned} \quad (12)$$

where $v_{it}^2 = x'_{it}\beta(1 - x'_{it}\beta)$ and where we have used the facts that $E(\varepsilon_{it}|x_{it}, u_i) = 0$ and $E(u_i|x_{it}) = 0$. Since the expectation in (12) is taken over a concave function of u_i , it must be smaller than the function of the expectation (which would simply be v_{it}^2). Moreover, it is always the case that $v_{it}^2 \geq \tau^2$, due to the specific restriction the LPM puts on the distribution of u_i . The largest possible variance of the random effect u_i arises for a binary distribution, where u_i takes the values $\{-a, a\}$ with equal probability. In order to satisfy $0 \leq x'_{it}\beta + u_i \leq 1$, a is bound from above by $\min(x'_{it}\beta, 1 - x'_{it}\beta)$ for all i and t . Therefore,

$$\tau^2 \leq [\min_{x_{it}}(x'_{it}\beta, 1 - x'_{it}\beta)]^2$$

which is necessarily smaller than $x'_{it}\beta(1 - x'_{it}\beta)$.

From (12), we get that $\text{Var}(y_{it}|x_{it}) = \text{Var}(u_i + \varepsilon_{it}|x_{it}) = v_{it}^2$, which defines the appropriate weights to correct for heteroskedasticity. The idiosyncratic error variances $\text{Var}(y_{it}|x_{it})$ are largest when $x'_{it}\beta$ equals 0.5. The situation is slightly different, when we combine the LPM with the scaled random effects assumption of the previous Section 2.2.2:

$$\Pr(y_{it} = 1|x_{it}, u_i) = x'_{it}\beta + u_i v_{it} \quad (13)$$

In this case, the variance function (12) can be re-written as $\text{Var}(\varepsilon_{it}|x_{it}) = v_{it}^2 - v_{it}^2 \tau^2 = v_{it}^2(1 - \tau^2)$. So premultiplying the data by $v_{it}^{-1} = 1/\sqrt{x'_{it}\beta(1 - x'_{it}\beta)}$, before performing random effects estimation, yields the correct GLS weights, because the factor of proportionality can be ignored.

2.3 Discussion

The two estimators for the heteroskedastic random effects model rest on a number of assumptions that may not hold in applied work. The two key assumptions are the uncorrelatedness between u_i and x_i , as well as a specific type of serial dependence generated by the error components structure. Typically, the first concern is addressed by using a fixed effects estimator that allows for arbitrary correlation between the random effect and the regressors.⁴ But there are important applications where this is not an option, such as when the main objective of the analysis is predicting new observations (for which a fixed effects estimate is unavailable by definition), when the overall constant is the object of interest (such as in meta analyses) and / or when there are time-invariant predictors.

The assumption of a specific correlation structure implied by (4) and (8) is a working assumption. The idea is that even if the assumption is not entirely correct, it is better (in terms of estimation efficiency) to use it rather than relying on an unweighted estimator. Of course, the standard errors will be affected by any misspecification of the weighting functions,

⁴See Wooldridge, 2002, for a distinction between fixed and random effects estimators.

and hence it is important to complement the analysis with cluster-robust standard errors (see Romano and Wolf, 2017, for a general advocacy of such an approach).

3 Simulation experiments

In this section, we study the small sample performance of our suggested estimators, with a focus on efficiency and testing of the slope parameter. We consider data generated from a linear model with a continuous dependent variable in subsection 3.1. Afterwards, subsection 3.2 covers the linear probability model for a binary dependent variable. In each case, we estimate the model parameters (intercept and slope) using five different methods: OLS, GLS adjusting for heteroskedasticity only (H), GLS adjusting for the random effect only (RE), and our new estimators that account for within-unit dependence as well as known heteroskedasticity, called HRE1 and HRE2. The difference between HRE1 and HRE2 is that HRE1 is based on the additive random effects model described in Section 2.2.1 (Model 1), whereas HRE2 is based on the scaled random effects model of Section 2.2.2 (Model 2).

The H estimator uses OLS for the transformed data $\{y_{it}^*, x_{it}^*\}$, obtained by dividing y_{it} , x_{it} as well as the constant by σ_{it} . HRE2 uses the same transformed values, but then estimates a standard random effects model, following the procedure outlined in Baltagi (2008).⁵ Lastly, HRE1 reconstructs the covariance matrix Ω using the estimated variance of the random effect based on equation (7) and the known σ_{it}^2 . The estimates are obtained using the standard feasible GLS approach, with $\hat{\Omega}^{-1}$ as weighting matrix and untransformed data.

Naturally, we expect all estimators to be centered at the true parameter values, but to find substantial differences in the variances of the sampling distributions, i.e., in their relative efficiency. HRE2 should dominate HRE1, when the data are generated from Model 2, and vice versa. Estimators using a misspecified variance-covariance matrix, e.g. because they ignore heteroskedasticity and/or the correlation due to the random effect, can be expected to have

⁵Identical results are achieved with random effects implementations in widely available programming languages.

wrong coverage of their implied confidence intervals, unless cluster-robust standard errors are used.

3.1 Linear models

The parameters of the simulations are as follows: there are $N = 100$ units observed for $T = 3$ periods each. The data generating process (DGP) includes a constant and a single time-varying regressor drawn from a χ -squared distribution with 6 degrees of freedom, that is then standardized to have a mean of zero and a variance of 0.25; The true values of the regression parameters are $\beta_0 = 1$ and $\beta_1 = 0.1$. The random effect is normally distributed with mean 0 and variance 4. The general error term is also normally distributed, centered at zero and with a standard deviation that is drawn from a uniform distribution with support over (1,3) to introduce heteroskedasticity. The H, HRE1 and HRE2 estimators utilise the true standard deviation of the general error term in their estimation. This is in the spirit of the proposed methodology, whereby the variances of the general error term are assumed to be known.

The simulation results, based on 5000 replications each, are shown in Tables 1 and 2. Table 1 uses data generated according to DGP 1 where the random effect is assumed to have a constant variance. Table 2 data were generated according to DGP 2, where the variance of the random effect is proportional to the variance of the general error term. As seen in Table 1, all estimators for the intercept β_0 , as well as the slope β_1 , are centered roughly at the true parameter values. However, OLS and heteroskedasticity-only corrected GLS are very inefficient. The variance of these two estimators exceeds that of HRE1 (which in Table 1 is the correct one) by a factor of almost 2. The random-effect-only estimator RE already leads to substantial efficiency gains, but allowing for both heteroskedasticity and random effects is better, in particular, when the “correct” weighting is used, in this case HRE1 rather than HRE2. The efficiency gains are most pronounced for the slope estimator, and less so for the intercept. We also find as expected that cluster robust standard errors lead to better confidence interval coverage. The differences to using unadjusted standard errors are larger

for the intercept than for the slope in these simulations.

Tables 1 and 2 about here

Table 2 shows the results for the second DGP. The parameters of the simulations remain the same as before, with the exception that the random effect, drawn from a Normal(0,4) distribution, is now multiplied with the standard deviation of the general error, drawn from a uniform(1,3) distribution. The results repeat many of the patterns already found in Table 1: All estimators are centered at the true parameter values, and nominal confidence interval coverage for both parameters is achieved if the cluster-robust variance estimator is used.

Regarding efficiency, OLS and H perform very poorly in this context, in particular as far as estimation of the slope is concerned. RE offers a considerable improvement in terms of efficiency, but it is clearly dominated by both HRE1 and HRE2. Their relative ranking is reversed compared to that found in Table 1. This was to be expected, since data are indeed generated from the scaled random effects model, corresponding to the HRE2 estimator. Overall, the efficiency gains are much larger in this DGP relative to the one employed in Table 1: the OLS variance of the slope exceeds the variance of the HRE2 estimator by a factor of more than five.

3.2 Binary dependent variable

Recall that the panel linear probability model with a single predictor and random effect can be written as

$$\Pr(y_{it} = 1|x_{it}, u_i) = p_{it} = \beta_0 + \beta_1 x_{it} + u_i$$

To simulate a binary response, we simply draw y_{it} from a Bernoulli distribution with success probability p_{it} . However, the parameter values and distributional assumptions made in the

previous section are unsuitable in this context, because they would violate the condition that $0 \leq p_{it} \leq 1$.

Hence, we adapt the specification as follows: In a first DGP, based on the additive model with constant-variance random effect, we let $\beta_0 = 0.4$ and $\beta_1 = 0.2$. x_{it} is standard uniform, and u_i has a discrete distribution that takes the values ± 0.35 with probability 0.5 each. This ensures that $0 \leq p_{it} \leq 1$ for all i and t . Because the linear index is bound between 0.4 and 0.6, $v_{it}^2 = x'_{it}\beta(1 - x'_{it}\beta)$ varies only between 0.24 and 0.25. So there is relatively little scope for heteroskedasticity, and substantial within-unit dependence due to the relatively high variance of u_i ($0.35^2 = 0.122$).

In order to implement the simulations for the LPM, we need to keep in mind that while the form of the skedastic function is known, it depends on the parameters β_0 and β_1 which need to be estimated first. Hence, we implement a two-step procedure, whereby an unweighted OLS regression provides estimates $\hat{\beta}_0$ and $\hat{\beta}_1$, which are then used to obtain the weights \hat{v}_{it} . Otherwise, we use the same set-up as before, i.e. there are $N = 100$ cross-sectional units, with $T = 3$ observation years for each unit, and we report the sample means, variances and 95% confidence interval coverage of the estimates obtained for 5000 repeated samples from the DGP.

Results for the first DGP are given in the top panel of the Table 3. All five estimators are unbiased for the true parameters, $\beta_0 = 0.4$ and $\beta_1 = 0.2$. Just accounting for heteroskedasticity in H does not improve the variance of the estimator. Because of the minor incidence of heteroskedasticity in this DGP, this was to be expected. However, allowing for the within-unit correlation stemming from the random effect leads to substantial improvements. In this case, it does not matter whether RE, HRE1 or HRE2 is used. For all three estimators, the variance drops by around one third relative to OLS.

The second DGP introduces substantially more heteroskedasticity. To do so, we let $u_i = \pm 0.1$ with equal probability, reducing its variance to a mere 0.01. As a consequence the spread of the linear index can be much larger without violating the bounds of the LPM. Specifically, x_{it} is drawn from a uniform distribution with support between -1.4 and 2.4 , such that

$0.12 \leq x'_{it}\beta \leq 0.88$ and $\text{Var}(\varepsilon_{it}|x_{it})$ has a lower bound of 0.1056 and an upper bound of 0.25.

The middle panel of Table 3 shows the results for this second DGP. Since it is dominated by heteroskedasticity, the two estimators that ignore this aspect, **OLS** and **RE**, don't perform very well in terms of efficiency. The other three estimators reduce the variance by about 10% relative to **OLS**. While they are very similar, **HRE1** based on the correct DGP has the smallest variance. The variance is about 1% below that of its closest competitors.

For our third DGP, based on the scaled random effects model, we use the same β values for the constant and the slope as before. By drawing x_{it} from a uniform distribution bounded between -1 and 2, we maintain a substantial amount of heteroskedasticity. At the same time, we allow for more spread in the random effect, as $u_i = \pm 0.5$. For example, when $\beta_0 + \beta_1 x_{it} = 0.8$ and $u_i = 0.5$, the scaled random effect is $0.5 \times \sqrt{0.8(1-0.2)} = 0.2$, and $\beta_0 + \beta_1 x_{it} + u_i v_{it} = 1$, its largest admissible value. It is easy to verify that the probability bounds are satisfied for all values of x_{it} and u_i

Table 3 about here

The bottom panel of Table 3 gives results for the linear probability model with scaled random effects. None of the five estimators is subject to bias, and the coverage of the 95% confidence interval based on cluster-robust standard errors is quite accurate in all cases. In terms of efficiency, the difference between the least efficient slope estimator, **OLS**, and the most efficient slope estimator, **HRE2**, amounts to 14%. Both heteroskedasticity and within cluster-correlation play a role here, that is: isolated weighting to account only for the random effect reduces the variance, as does isolated weighting to account only for heteroskedasticity, but less so. The full efficiency gain is realized when both departures from i.i.d. errors are accounted for in the weighting. The estimator with the smallest variance is the scaled-random effects estimator, that first transforms the data, dividing each pair (y_{it}, x_{it}) by \hat{v}_{it} , and then applies the standard random effects estimator to the transformed data.

4 Application: Meta-Analysis of Empirical Estimates of Loss Aversion

In a recent paper, Brown et al. (2024) collected and analyzed 607 empirical estimates of loss aversion from 150 articles published between 1992 and 2017.⁶ One goal of the meta-analysis was to determine the extent to which effect size differences can be attributed to observable differences in study design. The number of estimates per study varied between papers, the largest number being 53. It can be expected that study specific unobserved factors generate within-study correlations between these estimates that should be accounted for when running meta-regressions, and the authors used a Bayesian hierarchical model to address this issue.

Alternatively, we can apply our proposed methodology, i.e. perform GLS estimation of linear random effects models with known heteroskedasticity. Table 4 contains results for five different estimators, namely OLS, H, RE, HRE1 and HRE2. We show 13 of the 31 employed predictors in the table. Omitted are 3 indicators of the subject pool, 4 indicators of the utility functional form, 3 indicators of the reference point, 3 indicators of the definition of λ , and 5 geographic indicators plus a constant.

Table 4 about here

Our estimates are not directly comparable to those reported by Brown et al. (2024) because they a) use a Bayesian analysis that requires assumptions on prior distributions, and b) focus on the posterior median rather than posterior mean. Nevertheless, the results are qualitatively similar. In particular the broad conclusion that only few characteristics are substantially correlated with differences in the loss aversion parameter is confirmed. For example, using the individual-level mean as the outcome, rather than the individual-level

⁶The authors made the data available via the OSF repository: <https://osf.io/9un34/>.

median significantly increases the estimated loss-aversion coefficient, which is plausible since loss aversion cannot be negative and its distribution is therefore skewed to the right.

When reading the table, one needs to keep in mind that these are all nominal predictors and the size, sign and statistical significance of the estimate depends on the chosen baseline (which is stated in the table). In terms of point estimates, differences by one unit are not uncommon, which is quite substantial, given that the preferred unadjusted coefficient reported in the study is 1.955; However, estimates tend to be imprecise, so the overall evidence is accordingly weak.

Comparing the five estimation methods, from the left to the right in Table 4, it is striking that OLS and RE produce relatively similar coefficients, but allowing for within study correlation quite uniformly and substantially increases precision. Allowing for heteroskedasticity only, H, has more varied effects. There are sign switches, relative to OLS. Coefficients are in some cases more than double or halve the size, and standard errors tend to go down but not always.

Another key result is that combining heteroskedasticity and random effects, using either HRE1 or HRE2, produces massive further gains in terms of reduced standard errors, compared to what a sole adjustment for heteroscedasticity would do. Among the two, HRE2 often has smaller standard errors, which, combined with larger point estimates (that are often similar to the heteroskedasticity-only estimator), leads to a substantial increase in the proportion of statistically significant coefficients, 8 out of the 13 coefficients displayed in the table. One striking result is that papers published in the field of economics (as opposed to non-econ or unpublished papers) report a loss aversion that surpasses the baseline by around 50%, a statistically significant difference.

5 Conclusions

The panel data random effects model with its strong exogeneity assumptions is unlikely to be the first choice among researchers interested in estimating causal effects. It nevertheless deserves its place in the applied economist’s toolkit, because it allows for an efficient aggregation of information and estimates all model parameters, including coefficients of time invariant regressors as well as the overall constant. This is important for prediction, and also in situations where the constant is of substantive interest, such as is the case in meta-regressions, for instance.

The paper presented two estimators for the linear random effects panel data model with known heteroskedasticity. Heteroskedasticity can be treated as known, for example, in panel regressions for averaged data, where the variance of the sampling error is inverse proportional to the known number of observations over which the averages are taken, in the linear probability model, and in meta regressions, where the reported standard error of the effect size estimates varies from study to study.

The key difference between the two estimators is whether or not the random effect is allowed to interact with the heteroskedasticity. In some applications, like the linear probability model and meta-regression, such a “scaled random effect” is an appealing assumption, but it is a departure from the standard approach. Simulation results showed that substantial efficiency gains can be realized with either of the two estimators, and correct confidence interval coverage can be obtained if cluster-robust standard errors are used.

Since the relative performance of the two estimators is quite similar, there is good reason to recommend the scaled random effects estimator because it is very simple to implement: it can be computed in a straightforward manner in any standard panel data software without the need for further programming.

While the estimators are robust to misspecification of the feasible GLS weights, they will identify different parameters when the underlying regression model departs from the chosen specification. This points to a limitation of our application to the linear probability model:

If the true DGP is the logit model, then the LPM is misspecified. While the unweighted estimator gives the best linear approximation to the true conditional expectation function, this is no longer the case for the weighted estimators. Moreover, the structural parameter β is not identified, and naturally, in such a situation, efficient estimation is a secondary concern.

References

- Baltagi, B.H. (1993) Useful matrix transformations for panel data analysis: a survey. *Statistical Papers* 34, 281-301.
- Baltagi, B.H. (2008) *Econometric Analysis of Panel Data*, John Wiley & Sons, New York.
- Baltagi, B.H. and J.M. Griffin (1988), A generalized error component model with heteroscedastic disturbances, *International Economic Review*, 29 , 745-753.
- Baltagi, B.H., G. Bresson and A. Pirotte (2006) Joint LM test for homoskedasticity in a one-way error component model, *Journal of Econometrics*, 134(2), 401-417.
- Brown, A.L., T. Imai, F.M. Vieider and C.F. Camerer (2024) Meta-analysis of Empirical Estimates of Loss Aversion, *Journal of Economic Literature* (Forthcoming)
- Holly, A., and L. Gardiol (2000) A score test for individual heteroscedasticity in a one-way error components model, in: Kirshnakumar, J., Ronchetti, E., eds., *Panel Data Econometrics: Future Directions*, Elsevier, Amsterdam. Ch. 10.
- Horrace, W. and R. Oaxaca (2006) Results on the bias and inconsistency of ordinary least squares for the linear probability model, *Economics Letters*, 90, 321-327.
- Lejeune, B. (2005) A full heteroscedastic one-way error components model: Pseudo-maximum likelihood estimation and specification testing, *Contributions to Economic Analysis* 274
- Mazodier, P. and A. Trognon (1978), Heteroscedasticity and stratification in error components models, *Annales de l'INSEE*, Vol. 30-31, 451-482.

- Pustejovsky, J.E. and E. Tipton (2022) Meta-analysis with robust variance estimation: expanding the range of working models. *Prevention Science*, 23, 425-438.
- Randolph, W.C. (1988), A transformation for heteroscedastic error components regression models, *Economics Letters*, 27, 349-354.
- Rao, S.R.S., Kaplan, J. and W.G. Cochran, W.G. (1981), Estimators for the one-way random effects model with unequal error variances, *Journal of the American Statistical Association*, 76, 89-97.
- Romano, J.P. and M. Wolf (2017) Resurrecting weighted least squares, *Journal of Econometrics*, 197, 1-19.
- Stanley, T.D. and H. Doucouliagos (2012) *Meta-Regression Analysis in Economics and Business*, Taylor and Francis, 2012.
- Wansbeek, T. (1989), An alternative heteroscedastic error components model, *Econometric Theory*, 5, 326.
- Wooldridge, J. (2002) *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

Tables

Table 1 (DGP 1)

	OLS	H	RE	HRE1	HRE2
<i>Mean of the estimated coefficients</i>					
Intercept	0.9967	0.9966	0.9967	0.9970	0.9975
Slope	0.0916	0.0964	0.0993	0.1015	0.1021
<i>Variance of the estimated coefficients</i>					
Intercept	0.0544	0.0560	0.0544	0.0512	0.0608
Slope	0.1102	0.1204	0.0727	0.0563	0.0683
<i>95% Coverage (standard variance)</i>					
Intercept	0.8342	0.7908	0.9480	0.9474	0.8792
Slope	0.9494	0.8968	0.9556	0.9542	0.9232
<i>95% Coverage (cluster robust variance)</i>					
Intercept	0.9502	0.9464	0.9496	0.9460	0.9412
Slope	0.9478	0.9438	0.9472	0.9472	0.9384

Notes: DGP 1 refers to the standard random effects model with heteroskedastic general error term. For the true DGP, $\beta_0 = 1$, $\beta_1 = 0.1$, $N = 100$, and $T = 3$. ε_{it} is heteroskedastic, drawn from a uniform distribution with support over (1,3), and the normally distributed additive random effect has mean 0 and variance 4. Columns 2-5 show results for different GLS estimators: H accounts for heteroskedasticity only, RE for the random effect only, and HRE1 and HRE2 for both heteroskedasticity and the within-unit correlation induced by the random effect.

Table 2 (DGP 2)

	OLS	H	RE	HRE1	HRE2
<i>Mean of the estimated coefficients</i>					
Intercept	0.9949	0.9946	0.9949	0.9960	0.9991
Slope	0.0826	0.0917	0.0998	0.1030	0.1036
<i>Variance of the estimated coefficients</i>					
Intercept	0.1771	0.1201	0.1772	0.1388	0.0730
Slope	0.2841	0.1859	0.1005	0.0790	0.0511
<i>95% Coverage (standard variance)</i>					
Intercept	0.7874	0.7844	0.9442	0.9674	0.9458
Slope	0.9482	0.9428	0.9612	0.9234	0.9556
<i>95% Coverage (cluster robust variance)</i>					
Intercept	0.9472	0.9486	0.9462	0.9484	0.9418
Slope	0.9496	0.9464	0.9500	0.9462	0.9432

Notes: See Table 1. DGP 2 refers to the scaled random effects model where general error term and random effect are both heteroskedastic.

TABLE 3: Linear Probability Models

	OLS	H	RE	HRE1	HRE2
DGP 1 (little heteroskedasticity)					
<i>Mean of the estimated coefficients</i>					
Intercept	0.3988	0.3988	0.3991	0.3990	0.3990
Slope	0.2017	0.2017	0.2012	0.2014	0.2014
<i>Variance of the estimated coefficients</i>					
Intercept	0.0041	0.0041	0.0032	0.0033	0.0032
Slope	0.0098	0.0098	0.0065	0.0067	0.0065
<i>95% Coverage (cluster robust variance)</i>					
Intercept	0.9424	0.9418	0.9436	0.9416	0.9406
Slope	0.9442	0.9448	0.9476	0.9410	0.9428
DGP 2 (substantial heteroskedasticity)					
<i>Mean of the estimated coefficients</i>					
Intercept	0.3999	0.4005	0.3999	0.4005	0.4005
Slope	0.1999	0.1989	0.1999	0.1989	0.1988
<i>Variance of the estimated coefficients (10^{-3})</i>					
Intercept	0.8393	0.7821	0.8494	0.7818	0.7837
Slope	0.4509	0.4177	0.4499	0.4119	0.4165
<i>95% Coverage (cluster robust variance)</i>					
Intercept	0.9440	0.9420	0.9426	0.9398	0.9372
Slope	0.9448	0.9374	0.9410	0.9320	0.9316
DGP 3 (scaled random effect)					
<i>Mean of the estimated coefficients</i>					
Intercept	0.3991	0.3994	0.3991	0.3994	0.3994
Slope	0.2003	0.1998	0.2004	0.1999	0.1999
<i>Variance of the estimated coefficients ($\times 10^{-2}$)</i>					
Intercept	0.1278	0.1247	0.1262	0.1232	0.1228
Slope	0.0857	0.0833	0.0778	0.0754	0.0752
<i>95% Coverage (cluster robust variance)</i>					
Intercept	0.9456	0.9454	0.9456	0.9416	0.9406
Slope	0.9472	0.9448	0.9446	0.9424	0.9414

Notes: Simulation results based on 5000 replications. In all three panels, $N = 100$, $T = 3$, $\beta_0 = 0.4$ and $\beta_1 = 0.2$. x has a continuous uniform distribution with support $[a, b]$, while u has discrete uniform distribution with masspoints $(-c, c)$ were a , b , and c were chosen such that the probabilities implied by the linear model are always between 0 and 1. See the text for further detail.

TABLE 4. Meta-regression for loss aversion coefficients

	OLS	H	RE	HRE1	HRE2
<i>Type of estimates (Individual-level mean is baseline)</i>					
Individual-level median	-1.0792 (0.4273)*	-0.4530 (0.2249)*	-1.2628 (0.4263)*	-0.3303 (0.1582)*	-0.4363 (0.1411)*
Aggregate-level	-0.8441 (0.4545)	-0.1393 (0.2742)	-0.9148 (0.3625)*	-0.4410 (0.1231)*	-0.1056 (0.1288)
<i>Type of data (Lab experiment is baseline)</i>					
Field experiment	0.5716 (0.5143)	0.4136 (0.5314)	0.5494 (0.3437)	0.1828 (0.2600)	0.3874 (0.2280)
Class experiment	-0.5358 (0.6817)	-0.3644 (0.7273)	-0.5857 (0.4002)	-0.1710 (0.2051)	-0.4256 (0.2418)
Online experiment	-0.3494 (0.6125)	0.6519 (0.6255)	-0.2753 (0.3417)	0.3542 (0.2438)	0.6017 (0.2478)*
Other field data	-0.2550 (0.4978)	-0.4382 (0.4381)	-0.2472 (0.3654)	-0.4953 (0.2439)*	-0.4324 (0.2044)*
<i>Reward (Hypothetical money is baseline)</i>					
Real money	-0.5511 (0.4517)	0.0684 (0.3626)	-0.5936 (0.3107)	-0.2115 (0.1140)	0.0503 (0.1605)
Non-money	-0.1681 (0.4841)	0.0121 (0.3693)	-0.2619 (0.3036)	0.3013 (0.1771)	0.0201 (0.1715)
<i>Method (Binary choice is baseline)</i>					
Survey	0.2998 (0.5738)	0.7367 (0.4008)	0.3818 (0.3661)	0.2830 (0.2936)	0.7620 (0.2089)*
Matching	-0.0243 (0.8150)	0.7618 (0.5580)	-0.0465 (0.6104)	0.5030 (0.2757)	0.7826 (0.2243)*
Other	0.1826 (0.3460)	0.2592 (0.2619)	0.2022 (0.2494)	0.3348 (0.1772)	0.2520 (0.1186)*
<i>Publication status (Published (econ) is baseline)</i>					
Published (non-econ)	0.0443 (1.4795)	-0.9609 (0.8186)	0.0319 (0.9536)	-0.4536 (0.7058)	-0.9881 (0.3001)*
Unpublished	-0.4407 (1.5146)	-1.2459 (0.8044)	-0.4593 (0.9813)	-0.7957 (0.7215)	-1.2641 (0.2902)*

Source: See Table D.1 as well as Figure 8 in Brown et al. (2024), and also <https://osf.io/9un34/>. Robust standard errors in parenthesis. * indicates statistical significance at the 5% level.