

Fixed and Random Effects in Classical and Bayesian Regression

Silvio Rendon

The University of Western Ontario, Canada

July 2003

Summary. This paper proposes a common and tractable framework for analyzing different definitions of fixed and random effects in a constant-slope variable-intercept model. It is shown that, regardless of whether effects (i) are treated as parameters or as an error term, (ii) are estimated in different stages of a hierarchical model, or whether (iii) correlation between effects and regressors is allowed, when the same information on effects is introduced into all estimation methods, the resulting slope estimator is also the same across methods. If different methods produce different results, it is ultimately because different information is being used for each method.

Keywords: Bayes, Panel data, Nuisance Parameters, Fixed Effects, Random Effects.

1 Introduction

Two econometricians, one Classical and one Bayesian, are presented with the same panel data and are asked to estimate the following model:

$$Y_{it} = \alpha_i + \sum_{k=1}^K X_{itk}\beta_k + u_{it}, \quad i = 1, \dots, N \quad \text{and} \quad t = 1, \dots, T, \quad (1)$$

where Y_{it} is a response variable for the i th individual at the t th time period, α_i is a fixed constant varying across individuals, X_{itk} is the k th covariate for each individual at each period, and u_{it} is an error term with zero mean and known variance. Additionally, both econometricians know there is available knowledge on the parameters of the model from previous research, which is contained in a joint distribution of the α_i s and β . When requested to estimate the slope parameter β with these data, both colleagues proceed to estimate fixed and random effects. However, because they have a different understanding of what fixed and random effects are, they perform different estimations.

In Classical econometrics, ‘fixed’ effects are treated as parameters and estimated by a least-square dummy variable (LSDV) or covariance estimation. ‘Random’ effects are assimilated to the error term and estimated by a generalized least squares (GLS) or Aitken estimation. In both instances, effects are basically treated as errors: “the use of dummy variables is an attempt to specify a model with an error term that indeed has zero mean” Wallace and Hussain (1969) (See also Balestra and Nerlove (1966), Maddala (1971), Henderson, Jr. (1971), Mundlak (1978), Chamberlain (1984), Hsiao (1986),

Address for correspondence: Silvio Rendon, Department of Economic, Faculty of Social Science, The University of Western Ontario, London, Ontario, Canada N6A 5C2. Tel: 1 519 661 2111, Ext. 85461. Fax: 1 519 661 3666.

Baltagi (1995)). That is, effects are ‘nuisance’ or ‘incidental’ parameters, which may distort a consistent estimation of the slope (See Lancaster (2000) for a survey of the incidental parameter problem in econometrics.).

In the Bayesian framework both ‘fixed’ and ‘random’ effects are treated as random parameters defined within a three-stage hierarchical model: the dependent variable is distributed around a mean value that depends, together with regressors, on certain parameters; these parameters are, in turn, distributed around a mean value determined by other parameters called ‘hyperparameters,’ which are also random. While a fixed effects estimation updates the distribution of the parameters, a random effects estimation updates the distribution of the hyperparameters. Consequently, “for the Bayesian model the distinction between fixed, random and mixed models reduces to the distinction between different prior assignments in the second and third stages of the hierarchy” Smith (1973) (See also Lindley and Smith (1972), Laird and Ware (1982), Sweeting (1982), Waternaux et al. (1989)).

This paper shows that when in a constant-slope variable-intercept linear regression such as the one in Model (1) all available prior information on effects is taken into account, ‘fixed’ and ‘random’ effects estimations yield the same slope estimator, regardless of the framework used. Within the Classical framework this means that the choice of whether to treat effects as parameters or as part of the error term is just reflecting different uses of the same available information, without any consequence for the slope estimator. In the first case, prior information is being updated; in the second case, it is not. Similarly, in a hierarchical model, where effects are characterized by their ‘layer of randomness’, the choice of the stage in which to update knowledge does not affect the slope estimator either. It is also shown that if the assumption of independence between effects and regressors is relaxed, as long as the same information is used the slope estimator does not change; differences only arise as result of a partial use of prior information. These conclusions stem from the analysis and comparison of different particular cases in a general Bayesian framework based on Lindley and Smith (1972) and Smith (1973) as applied to the constant-slope variable-intercept model.

The model presented in this paper is valid for general forms of effects classifications, error covariance structures, and prior distributions. It is illustrated with the one-way classification model with a spherical error covariance and a diagonal variance in the prior distribution of effects. The rest of the paper is organized as follows: next section describes the general model; Section 3 compares slope estimators when effects are treated as parameters and when they are considered part of the error term; Section 4 compares slope estimators for different stages of a hierarchical model; Section 5 studies a particular application of the general case: the one-way classification model; Section 6 anal-

yses the case when there is correlation between effects and regressors; and Section 7 presents the conclusions.

2 The General Model

We can rewrite Model (1) in matrix form as

$$Y = Z\alpha + X\beta + u, \quad u \sim N(0, \Sigma) \quad (2)$$

where Y is $NT \times 1$, Z is a $NT \times J$ matrix, α is a $J \times 1$ vector, X is a $NT \times K$ matrix of full rank, β is a $K \times 1$ vector, u is $NT \times 1$, and variance Σ is a $NT \times NT$ matrix.

The following analysis is applicable to all structures where observations can be grouped in classes (such as individual and time effects, or “treatments” and “blocks”). That is, Z is left unspecified for the moment and the particular case of Model (1) of a one-way classification where $J = N$ with $Z = I_N \otimes i_T$ will be discussed in Section 5. Thus, α is a vector of ‘effects’ and β is a vector of ‘slope’ parameters common to all observations.

Let $X_Z = \begin{pmatrix} Z & X \end{pmatrix}$ and $\theta' = \begin{pmatrix} \alpha & \beta \end{pmatrix}$, then this model becomes

$$Y = X_Z\theta + u.$$

The econometrician’s prior knowledge on parameters is contained in the joint distribution $\theta \sim N(\bar{\theta}, \Omega)$,

$$\bar{\theta} = \begin{pmatrix} \bar{\alpha} \\ \bar{\beta} \end{pmatrix}, \Omega = \begin{pmatrix} A & P \\ P' & V \end{pmatrix},$$

where $\bar{\theta}$ does not exhibit any further randomness. Thus, the distribution of α conditional on β is $N(\alpha^* + P^*\beta, A^*)$, where

$$A^* \equiv A - P^*VP^{*'}, \quad (3)$$

$$\alpha^* \equiv \bar{\alpha} - P^*\bar{\beta}, \quad (4)$$

$$P^* \equiv PV^{-1}. \quad (5)$$

Assuming a known error covariance structure, as is usual in regression analysis, the posterior distribution, characterized by the precision and the posterior precision times the mean of θ , is denoted

by subscript ‘ p ’ and updated by

$$\Omega_p^{-1} = X'_Z \Sigma^{-1} X_Z + \Omega^{-1}, \quad (6)$$

$$\Omega_p^{-1} \bar{\theta}_p = X'_Z \Sigma^{-1} Y + \Omega^{-1} \bar{\theta}, \quad (7)$$

as proposed by Lindley and Smith (1972). (Parameters of the disturbance term can be added to the joint prior distribution and be also updated or integrated over to isolate the marginal of the location parameters. This extension escapes the purpose of the current paper and as such is left for further research.) From these expressions one obtains the posterior moments for the effects and slope parameters (see Appendix A.1. for derivation of these equations). Thus, the posterior distribution of the slope parameter is given by

$$V_p^{-1} = X^{*'} Q X^* + V^{-1}, \quad (8)$$

$$V_p^{-1} \bar{\beta}_p = X^{*'} Q Y^* + V^{-1} \bar{\beta}, \quad (9)$$

where $X^* = X + ZP^*$, $Y^* = Y - Z\alpha^*$, and

$$Q \equiv Q(Z, \Sigma, A^*) = \Sigma^{-1} - \Sigma^{-1} Z A^* \left\{ A^* + (Z' \Sigma^{-1} Z)^{-1} \right\}^{-1} (Z' \Sigma^{-1} Z)^{-1} Z' \Sigma^{-1}.$$

The posterior moments of the slope parameter have the simple structure of a weighted average of its prior moments and a generalized least-squares estimator. It is assumed that $Z' \Sigma^{-1} Z$ is of full rank. If not, a generalized or Moore-Penrose inverse should stand instead of $(Z' \Sigma^{-1} Z)^{-1}$. The weighting matrix Q can be decomposed in two components: $Q = Q_1 + Q_2$, where

$$Q_1 \equiv Q_1(Z, \Sigma) = \Sigma^{-1} - \Sigma^{-1} Z (Z' \Sigma^{-1} Z)^{-1} Z' \Sigma^{-1},$$

$$Q_2 \equiv Q_2(Z, \Sigma, A^*) = \Sigma^{-1} Z (Z' \Sigma^{-1} Z)^{-1} \left\{ A^* + (Z' \Sigma^{-1} Z)^{-1} \right\}^{-1} (Z' \Sigma^{-1} Z)^{-1} Z' \Sigma^{-1}.$$

The first component Q_1 only depends on data, not on any prior information, and is nonzero for $Z \neq I$; the second component Q_2 does depend on prior information and is nonzero only for an informative prior conditional variance, $A^{*-1} > 0$. Thus, the weighting matrix Q depends on the form of the Z matrix, that is, on the design of the effects and on the availability of an informative prior distribution for the effects.

The posterior conditional moments of the effects have also the form of a weighted average of

prior and sample moments:

$$A_p^{*-1} = Z'\Sigma^{-1}Z + A^{*-1}, \quad (10)$$

$$A_p^{*-1}\alpha_p^* = Z'\Sigma^{-1}Y + A^{*-1}\alpha^*, \quad (11)$$

$$A_p^{*-1}P_p^* = -Z'\Sigma^{-1}X + A^{*-1}P^*. \quad (12)$$

Note that P_p^* is determined by the sample covariance between Z and X and is generally nonzero even if $P^* = 0$. The fact that α_p^* and P_p^* use the same weighting matrix $A_p^*A^{*-1}$ facilitates the computation of the unconditional mean of the effects:

$$\bar{\alpha}_p = A_p^*A^{*-1} \{ \bar{\alpha} + P^* (\bar{\beta}_p - \bar{\beta}) \} + (I - A_p^*A^{*-1}) (Z'\Sigma^{-1}Z)^{-1} Z'\Sigma^{-1} (Y - X\bar{\beta}_p).$$

The other unconditional moments are obtained from definitions (3)-(5).

The present model, with general Z and Σ , is basically a partitioned GLS regression for two parameters, α and β . The next sections present conditions for coincidence of fixed and random effects estimations under different definitions.

3 Classical Approach: Parameter or Error?

The α effects in Eq. (2) can be treated as an intercept or as an error. Fixed and random effects are defined according to this choice, that is,

- **Fixed Effects: α is a parameter (FE)**
- **Random Effects: α is an error (RE).**

The main difference is that an FE updates prior information on effects, whereas an RE estimation treats prior information as final, that is, as if there is no further learning on the distribution of effects. An FE uses dummy variables to account for class effects (be they individual or time related), so it is a LSDV estimation where effects are ‘swept out’ by taking differences. The posterior moments of the slope parameter are obtained from Eq. (8) and Eq. (9):

$$\begin{aligned} V_p^f &= \left\{ (X^{*'}Q_fX^*)^{-1} + V^{-1} \right\}^{-1}, \\ \bar{\beta}_p^f &= (I - V_p^fV^{-1}) (X^{*'}Q_fX^*)^{-1} X^{*'}Q_fY^* + V_p^fV^{-1}\bar{\beta}, \\ \text{where } Q_f &\equiv Q(Z, \Sigma, A^*) \end{aligned}$$

and the subscript f stands for FE.

An RE updates the distribution of β taking the distribution of α as known. Given the correlation between these parameters, what is known is the distribution of α conditional on β . Let $\alpha|\beta = \alpha^* + P^*\beta + v$, $v \sim N(0, A^*)$, and $u_r = u + Zv$. Using definition (4), the original model is transformed:

$$\begin{aligned} Y &= Z\alpha + X\beta + u, \\ &= Z\alpha^* + ZP^*\beta + X\beta + u + Zv, \\ Y - Z\alpha^* &= (X + ZP^*)\beta + u_r, \\ Y^* &= X^*\beta + u_r, \end{aligned}$$

where $u_r \sim N(0, \Sigma_r)$, $\Sigma_r = \Sigma + ZA^*Z'$, and where Y^* and X^* are defined as in the previous section. As there is only one parameter β , we can see that $Q_r \equiv Q_1(0, \Sigma_r) = \Sigma_r^{-1} = (\Sigma + ZA^*Z')^{-1}$, where subscript r stands for RE. Hence, the posterior distribution of the common parameter is given by

$$\begin{aligned} V_p^r &= \left\{ (X^{*'}\Sigma_r^{-1}X^*)^{-1} + V^{-1} \right\}^{-1}, \\ \bar{\beta}_p^r &= (I - V_p^r V^{-1}) (X^{*'}\Sigma_r^{-1}X^*)^{-1} X^{*'}\Sigma_r^{-1}Y^* + V_p^r V^{-1}\bar{\beta}. \end{aligned}$$

As before, we have a GLS estimation, only the original error variance is augmented by the variance of the effects.

Theorem 1 *The posterior moments of the slope parameter are invariant to assuming that effects are either parameters or errors, that is: FE and RE yield the same posterior moments: $\bar{\beta}_p^f = \bar{\beta}_p^r$ and $V_p^f = V_p^r$.*

Proof: Note that $Q_f = Q(Z, \Sigma, A^*) = Q_1(Z, \Sigma_r) = Q_r$. ■

This means that the econometrician's choice to update or not to update prior information on effects does not change the common parameter estimator. However, were she interested in learning the most that data can tell, the econometrician would always perform an FE, because it yields the RE slope estimator *and* learning on the effects.

Corollary 2 *A Classical random effects estimation is a special case of a Bayesian fixed effects estimation with ignorance on the common parameter, $V^{-1} = 0$, but **with** an informative prior conditional variance on the effects, $A^{*-1} > 0$.*

By treating effects as an error term in a random effects estimation a Classical econometrician (maybe without realizing it) does introduce prior information into the estimation. (Zellner (1999) indicates several other applications where non-Bayesians do introduce distributions for parameters,

which are usually considered ‘part of the model’ and not prior distributions.) Absence of this prior information in an FE, that is, $A^{*-1} = 0$ implies $Q_f = Q_1(Z, \Sigma) \equiv Q_{1f}$ and produces different estimators:

$$\begin{aligned} FE : V_p^f &= (X'Q_{1f}X)^{-1}, & \bar{\beta}_p^f &= (X'Q_{1f}X)^{-1} X'Q_{1f}Y, \\ RE : V_p^r &= (X'\Sigma_r^{-1}X)^{-1}, & \bar{\beta}_p^r &= (X'\Sigma_r^{-1}X)^{-1} X'\Sigma_r^{-1}Y. \end{aligned}$$

In conclusion, what appears to be a choice of an estimation method is actually a choice about using prior information. Treating effects as parameters or as an error term is not the crucial choice for an econometrician, whether to use or not prior information is.

4 Bayesian approach: fixed or random parameter?

If the dependent variable is distributed around a mean value that depends on parameter θ , which is in its turn distributed around the hyperparameter $\tilde{\theta}$, we have a hierarchical structure in three stages:

$$\begin{aligned} \text{First Stage} & : Y \sim N(X\theta, \Sigma), \\ \text{Second Stage} & : \theta \sim N(\tilde{\theta}, \Lambda), \\ \text{Third Stage} & : \tilde{\theta} \sim N(\hat{\theta}, \Psi). \end{aligned}$$

In this setup, and according to the number of stages used, fixed and random effects are defined as follows:

- **Fixed Effects:** $\tilde{\theta}$ is fixed (FE’),
- **Random Effects:** $\tilde{\theta}$ is random (RE’).

An FE’ is a ‘Model of the First Kind’ or ‘Model I’, which means it does not have a Third Stage (See Nerlove (2002) for a description of the origins of the distinction between ‘Model I’ and ‘Model II’). That is the case when $\tilde{\theta} = \hat{\theta} = \bar{\theta}$ and $\Lambda = \Omega$ as shown in Section 2 and in Section 3, where, regardless of the form of Σ or of Z , neither $\bar{\alpha}$ nor $\bar{\beta}$ are random. Consequently, FE and RE are both FE’. One can also generate an FE’ model by getting rid of the hyperparameter $\tilde{\theta}$ and working directly with the unconditional distribution of θ , $N(\hat{\theta}, \Lambda + \Psi)$. If $\Lambda + \Psi = \Omega$, and $\hat{\theta} = \bar{\theta}$ one obtains the same model of Section 2. An FE’ updates the posterior unconditional moments for the Second

Stage, which are

$$\begin{aligned}\Lambda_p^{-1} &= X'_Z \Sigma^{-1} X_Z + (\Lambda + \Psi)^{-1}, \\ \Lambda_p^{-1} \tilde{\theta}_p &= X'_Z \Sigma^{-1} Y + (\Lambda + \Psi)^{-1} \hat{\theta}.\end{aligned}$$

An RE', also called a 'Model of the Second Kind' or 'Model II', updates the distribution of the hyperparameter $\tilde{\theta}$, that is, it estimates the posterior moments for the Third Stage:

$$\begin{aligned}\Psi_p^{-1} &= X'_Z \Sigma_R^{-1} X_Z + \Psi^{-1}, \\ \Psi_p^{-1} \hat{\theta}_p &= X'_Z \Sigma_R^{-1} Y + \Psi^{-1} \hat{\theta}, \\ \text{where } \Sigma_R &\equiv \Sigma + X_Z \Lambda X'_Z.\end{aligned}$$

This is a general result shown by Lindley and Smith (1972) and Smith (1973). To apply it to the constant-slope variable-intercept framework, let

$$\tilde{\theta} = \begin{pmatrix} \tilde{\alpha} \\ \tilde{\beta} \end{pmatrix}, \Lambda = \begin{pmatrix} D & \tilde{P} \\ \tilde{P}' & \tilde{V} \end{pmatrix}, \hat{\theta} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}, \Psi = \begin{pmatrix} C & P \\ P' & V \end{pmatrix},$$

and, to maintain the model with nonrandom slopes, let $\tilde{\beta} = 0, \tilde{V} = 0, \tilde{P} = 0$. The resulting hierarchical structure is thus

$$\begin{aligned}\alpha &\sim N(\tilde{\alpha}, D|\tilde{\alpha}), \\ \begin{pmatrix} \tilde{\alpha} \\ \beta \end{pmatrix} &\sim N\left(\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}, \begin{pmatrix} C & P \\ P' & V \end{pmatrix}\right).\end{aligned}$$

In this setup, $\Lambda + \Psi = \Omega$ implies that $D + C = A$, that is, effects present different 'layers of randomness' but with exactly the same unconditional distribution of Section 2. Let F and R be the subscripts denoting FE' and RE' estimators, respectively. Posterior distributions of the common parameter are obtained using Eq. (8) and Eq. (9). For an FE', the posterior moments of the slope parameters are

$$\begin{aligned}V_p^F &= \left\{ (X^{*'} Q_F X^*)^{-1} + V^{-1} \right\}^{-1}, \\ \tilde{\beta}_p^F &= (I - V_p^F V^{-1}) (X^{*'} Q_F X^*)^{-1} X^{*'} Q_F Y^* + V_p^F V^{-1} \hat{\beta}, \\ \text{where } Q_F &\equiv Q(Z, \Sigma, D + C^*).\end{aligned}$$

For an RE', the posterior moments of the slope parameter are

$$\begin{aligned} V_p^R &= \left\{ (X^{*'} Q_R X^*)^{-1} + V^{-1} \right\}^{-1}, \\ \hat{\beta}_p^R &= (I - V_p^R V^{-1}) (X^{*'} Q_R X^*)^{-1} X^{*'} Q_R Y^* + V_p^R V^{-1} \hat{\beta}, \\ \text{where } Q_R &\equiv Q(Z, \Sigma + Z D Z', C^*). \end{aligned}$$

Here $X \Lambda X' = Z D Z'$. Note that in an RE' the variance of the Second Stage, D , is assimilated to the error term, whereas in an FE' it is in the prior distribution and it is, therefore, updated. In choosing an RE' or an FE' the econometrician is choosing how to 'slice' information, that is, she is choosing what portion of the available knowledge to update.

Theorem 3 *The posterior distribution of the slope parameter is invariant to the stage in which it is estimated, that is, FE' and RE' yield the same posterior moments: $\tilde{\beta}_p^F = \hat{\beta}_p^R$ and $V_p^F = V_p^R$.*

Proof: Note that $A^* = D + C^*$ and $Q_F = Q_R = (\Sigma + Z D Z' + Z C^* Z')^{-1} = Q_f$. ■

The same slope estimator results whether the estimation is done in the Second or in the Third Stage. What matters is that we use all the available information, no matter how we slice it. In other words, what brings about a difference in the resulting estimators is the choice of whether to use or not to use the available information. The later choice of how to 'slice' this information in components that are updated and components that are not, does not entail a difference in the slope estimator obtained.

The posterior distributions for the idiosyncratic parameters in the Second and Third Stages are estimated using Eqs. (10)-(12). Unlike the common parameter, they do not coincide. For example, the posterior conditional precisions of the effects are

$$\begin{aligned} \text{Second Stage} &: D_p^{*-1} = Z' \Sigma^{-1} Z + (D + C^*)^{-1}, \\ \text{Third Stage} &: C_p^{*-1} = \left\{ \left(Z' \Sigma^{-1} Z \right)^{-1} + D \right\}^{-1} + C^{*-1}, \end{aligned}$$

which are only equal in the trivial case, when $D = 0$. So, while FE and FE' coincide also for the effects estimators, RE and RE' do not. In RE there is no learning on effects, while in RE' there is. We can examine a particular case in which the prior distribution is uninformative, that is, where $\Psi^{-1} = 0$ and $D = A$, which implies that $Q_F = Q_1(Z, \Sigma)$ and $Q_R = Q_1(Z, \Sigma + Z A Z') \equiv Q_{1R}$, thus

$$\begin{aligned} FE' : V_p^F &= (X' Q_1 X)^{-1} = V_p^f, & \tilde{\beta}_p^F &= (X' Q_1 X)^{-1} X' Q_1 Y = \bar{\beta}_p^f, \\ RE' : V_p^R &= (X' Q_{1R} X)^{-1} \neq V_p^r, & \hat{\beta}_p^R &= (X' Q_{1R} X)^{-1} X' Q_{1R} Y \neq \bar{\beta}_p^r. \end{aligned}$$

As in the previous section, the difference in the estimators is caused by the use of different information in each of the two estimations; the difference in the estimators does not come at all from having used a hierarchical model with three stages. The next section applies these result to a particular configuration of effects, the one-way classification.

5 Application: One-way Classification

The previous sections have shown that for a general form of Z or of Σ the slope estimator is invariant to introducing information on effects as either parameters or errors or at different stages of a hierarchical model. This section examines Model (1), the one-way classification model, where $Z = I_N \otimes i_T$, being i_T a $T \times 1$ vector of ones. Let the error term have a diagonal variance, $\Sigma = \sigma^2 I$, assume that the prior moments are $V^{-1} = 0, P^* = 0, \alpha^* = 0$, and that A^* is diagonal, that is,

$$[A^*]_{ij} = \begin{cases} \sigma_{\alpha i}^{*2}, & i = j, \\ 0, & i \neq j. \end{cases}$$

The posterior of this variance is then:

$$A_p^{*-1} = \Sigma^{-2} Z' Z + A^{*-1},$$

$$\begin{pmatrix} \sigma_{a1,p}^{*-2} & 0 & \cdots & 0 \\ 0 & \sigma_{a2,p}^{*-2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{an,p}^{*-2} \end{pmatrix} = \begin{pmatrix} \sigma^{-2}T & 0 & \cdots & 0 \\ 0 & \sigma^{-2}T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^{-2}T \end{pmatrix} + \begin{pmatrix} \sigma_{a1}^{*-2} & 0 & \cdots & 0 \\ 0 & \sigma_{a2}^{*-2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{an}^{*-2} \end{pmatrix}.$$

Proposition 4 *If $\Sigma = \sigma^2 I$ and the prior conditional variance A^* is diagonal, then the posterior conditional variance A_p^* is also diagonal.*

Proof: *It follows from $Z' \Sigma^{-1} Z = \sigma^{-2} T I_N$ and Eq. (10).■*

Corollary 5 *If $\sigma_{\alpha i}^{*2} = \sigma_{\alpha j}^{*2}$, then $\sigma_{\alpha i,p}^{*2} = \sigma_{\alpha j,p}^{*2}$, $i \neq j$, that is, individuals with the same conditional prior variance have the same conditional posterior variance.*

Corollary 6 *(Conditional independence) Conditional on β , individuals have zero posterior covariance: $\text{cov}_p(\alpha_i, \alpha_j | \beta) = 0$, $i \neq j$.*

As the conditional moments of the effects are used in the computation of the slope estimator, Proposition 4 means that a tractable configuration is preserved. The posterior unconditional variance of effects is, however, not diagonal, because in general the posterior correlation P_p^* is nonzero. This

result also means that individual effects can be estimated independently, because they are correlated with each other only through their individual correlation with β . It is straightforward to extend this result to an unbalanced panel with T_i observations for each individual i , or to a model with an heteroskedastic error term, where the diagonal variance is preserved as well.

To update the slope parameter, we use $Q_1 = \sigma^{-2} (I_{NT} - I_N \otimes \bar{J}_T)$ and $Q_2 = \sigma^{-2} A_p^* A^{*-1} \otimes \bar{J}_T$, where $\bar{J}_T = T^{-1} i_T i_T'$. Then $\sigma_{ai,p}^{*2} = g_i \sigma_{ai}^{*2}$, $\bar{\alpha}_{i,p} = (1 - g_i) (\bar{Y}_i - \bar{X}_i \bar{\beta}_p)$, where $g_i = \sigma^2 (T \sigma_{ai}^{*2} + \sigma^2)^{-1}$ and $\bar{\beta}_p$ is obtained from Eq. (9). This means that in a ‘random effects’ estimation there can also be learning on the idiosyncratic parameters as in a ‘fixed effects’ estimation.

If the prior conditional distribution of effects is not only diagonal but also $\sigma_{\alpha i}^{*2} = \sigma_{\alpha}^{*2}, \forall i$, then $g_i = g \equiv \sigma^2 (T \sigma_a^{*2} + \sigma^2)^{-1}$ and $Q_2 = \sigma^{-2} g I_N \otimes \bar{J}_T$. In this case, Q_1 and Q_2 produce, respectively, the familiar ‘within’ and ‘between’ covariances. The posterior estimator of the slope becomes

$$\begin{aligned} V_p &= (X'Q_1X + X'Q_2X)^{-1} = \sigma^2 (W_{XX} + gB_{XX})^{-1}, \\ \bar{\beta}_p &= (X'Q_1X + X'Q_2X)^{-1} (X'Q_1Y + X'Q_2Y) = (W_{XX} + gB_{XX})^{-1} (W_{XY} + gB_{XY}). \end{aligned}$$

These expressions, obtained by Maddala (1971), are usually seen as different methods of estimation. In the present context, they depend on the availability of prior information on effects. Q_1 , which produces the ‘within’ covariances, does not depend at all on prior information on effects; Q_2 , which produces the ‘between’ covariances, does introduce the prior information on effects, as it contains g .

We can also analyse the one-way classification model applying a hierarchical approach in three stages as in Section 3:

$$\begin{aligned} \text{First Stage} &: Y_{it} \sim N(\alpha_i + X'_{it}\beta, \sigma^2), \\ \text{Second Stage} &: \alpha_i \sim N(\tilde{\alpha}_i, \sigma_{di}^2 | \tilde{\alpha}), \\ \text{Third Stage} &: \tilde{\alpha}_i \sim N(\bar{\alpha}_i, \sigma_{ci}^2), \beta \sim N(\bar{\beta}, V). \end{aligned}$$

In this model C^* and D are diagonal, with σ_{ci}^{*2} and σ_{di}^2 as the prior conditional variances for individual i , and there is independence between idiosyncratic and common parameters. In this context the assumption of Section 4 that $A^* = D + C^*$ means that $\sigma_{ai}^{*2} = \sigma_{di}^2 + \sigma_{ci}^{*2}$, that is, the individual parameter α exhibits ‘layers’ of randomness. For the Second Stage only the sum of these variances matter, but for the Third Stage separate prior information about σ_{di}^2 and σ_{ci}^{*2} is needed. If the prior distribution in the Third Stage is uninformative, then $\sigma_{ai}^{*2} = \sigma_{di}^2$.

Let $h_i = (T\sigma_{ai}^{*2} + \sigma^2)^{-1} (T\sigma_{di}^2 + \sigma^2)$, then the posterior moments are summarized in Table 1. The first block of this table reports FE' and RE' for both uninformative and informative prior distributions of idiosyncratic parameters. We see that the slope estimator produced by informative prior distributions of effects only depends on their total variance. The partition of this variance in different stages only makes a difference for the posterior distributions of idiosyncratic parameters. It is also clear that for all estimators of idiosyncratic parameters posterior conditional variances are equal if prior variances are equal; however, unconditional variances of effects are not diagonal. Correlation across individuals is built through their individual correlation with the common parameter β .

Table 1: Posterior Moments in the Hierarchical Model

Bayesian	Second Stage: FE' or FE	
Classical	'fixed'	'random'
2nd Stage	$\sigma_{ai}^{*-2} = 0$	$\sigma_{ai}^{*-2} > 0$
3rd Stage	None or $\sigma_{ai}^{*2} = \sigma_{di}^2 + \sigma_{ci}^{*2}$	
$\tilde{\beta}_p$	$(X'HX)^{-1} X'HY$	$(X'QX)^{-1} X'QY$
\tilde{V}_p	$\Sigma^2 (X'HX)^{-1}$	$\sigma^2 (X'QX)^{-1}$
$\sigma_{ai,p}^{*2}$	$T^{-1}\sigma^2$	$g_i\sigma_{ai}^{*2}$
$\bar{\alpha}_{i,p}$	$\bar{Y}_i - \bar{X}_i\tilde{\beta}_p$	$(1 - g_i) (\bar{Y}_i - \bar{X}_i\tilde{\beta}_p)$
$\alpha_{i,p}^*$	\bar{Y}_i	$(1 - g_i) \bar{Y}_i$
$\tilde{P}_{i,p}^*$	$-\bar{X}_i$	$-(1 - g_i) \bar{X}_i$

Bayesian	Third Stage: RE'	
2nd Stage	$\sigma_{ai}^{*2} = \sigma_{di}^2$	$\sigma_{ai}^{*2} = \sigma_{di}^2 + \sigma_{ci}^{*2}$
3rd Stage	$\sigma_{ci}^{*-2} = 0$	$\sigma_{ci}^{*-2} > 0$
$\bar{\beta}_p$	$(X'QX)^{-1} X'QY$	
V_p	$\sigma^2 (X'QX)^{-1}$	
$\sigma_{ci,p}^{*2}$	$T^{-1}\sigma^2 + \sigma_{ai}^{*2}$	$h_i\sigma_{ci}^{*2}$
$\bar{c}_{i,p}$	$\bar{Y}_i - \bar{X}_i\bar{\beta}_p$	$(1 - h_i) (\bar{Y}_i - \bar{X}_i\bar{\beta}_p)$
$c_{i,p}^*$	\bar{Y}_i	$(1 - h_i) \bar{Y}_i$
$P_{i,p}^*$	$-\bar{X}_i$	$-(1 - h_i) \bar{X}_i$

In conclusion, the usual 'fixed' and 'random' effects estimators of the one-way classification model are posterior estimators produced by different prior distributions of individual parameters. Transforming the original model into a hierarchical model by adding 'stages' does not change the slope estimator, as long as the same information on individual effects is used.

6 Independence or Correlation of Effects and Regressors?

If idiosyncratic parameters are correlated not just with slope parameters, but also with regressors, one can postulate the auxiliary regression:

$$\alpha = \Gamma\pi + v,$$

where $\Gamma = \Gamma(X)$ is a known transformation of X of dimension $J \times L$, π is a $L \times 1$ vector, v is a $J \times 1$ vector, $v \sim N(0, \Upsilon)$ and Υ is a known $J \times J$ matrix. The transformation Γ can have several forms. Mundlak (1978) used $\Gamma(X) = (Z'Z)^{-1}Z'X$, so $L = N$. In his setup the idiosyncratic effects depended on the individual mean over time. Chamberlain (1982) used a more general transformation where the effects depended on all the individual X s over time. In any case, one still needs to estimate π , which becomes the main purpose of the estimation. The parameter vector has now the following prior distribution

$$\begin{pmatrix} \pi \\ \beta \end{pmatrix} \sim N \left(\begin{pmatrix} \bar{\pi} \\ \bar{\beta} \end{pmatrix}, \begin{pmatrix} \Pi & R \\ R' & V \end{pmatrix} \right),$$

that is, in terms of the original moments of the effects:

$$\begin{aligned} A^* &= \Gamma\Pi^*\Gamma' + \Upsilon, \\ \alpha^* &= \Gamma\pi^*, \\ P^* &= \Gamma R^*, \end{aligned}$$

where Π^* , π^* , and R^* are defined as in Eqs. (3)-(5). Now, Model (2) becomes

$$Y = Z_m\pi + X\beta + u_m,$$

where $Z_m = Z\Gamma$ and $u_m = u + Zv$, with a known resulting variance $\Sigma_m = \Sigma + Z\Upsilon Z'$. Thereby, effects are divided in two parts: the correlated component, which is maintained as a parameter, and the uncorrelated component, which is assimilated to the error term. We just need to apply the

corresponding expressions derived in Section 2.

$$\begin{aligned}
 V_p^m &= \left\{ \left(X_m^{*'} Q_m X_m^* \right)^{-1} + V^{-1} \right\}^{-1}, \\
 \bar{\beta}_p^m &= \left(I - V_p^m V^{-1} \right) \left(X_m^{*'} Q_m X_m^* \right)^{-1} \left(X_m^{*'} Q_m Y_m^* \right) + V_p^m V^{-1} \bar{\beta}, \\
 \text{where } Q_m &\equiv Q \left(Z_m, \Sigma_m, \Pi^* \right), \\
 X_m^* &= X + Z_m R^*, \\
 Y_m^* &= Y - Z_m \pi^*.
 \end{aligned}$$

Theorem 7 *The posterior moments of the slope parameter, $\bar{\beta}_p, V_p$, are invariant to allowing for correlation between effects and regressors, regardless of the form of this correlation.*

Proof. *From Eq. (8), Eq. (9), and from*

$$\begin{aligned}
 X_m^* &= X + Z_m R^* = X + Z P^* = X^*, \\
 Y_m^* &= Y - Z_m \pi^* = Y - Z \alpha^* = Y^*, \\
 \text{and } Q_m &\equiv Q \left(Z_m, \Sigma_m, \Pi^* \right), \\
 &= \left(\Sigma + Z \Upsilon Z' + Z \Gamma \Pi^* \Gamma' Z' \right)^{-1} = \left\{ \Sigma + Z \left(\Upsilon + \Gamma \Pi^* \Gamma' \right) Z' \right\}^{-1}, \\
 &= Q \left(Z, \Sigma, A^* \right),
 \end{aligned}$$

it follows that $V_p^m = V_p^f$ and that $\bar{\beta}_p^m = \bar{\beta}_p^f$. ■

Again, it does not matter that the information on effects consists of correlated and uncorrelated components; it only matters that the information is being used. This implies that fixed effects estimators coincide with random effects estimators, if both use the same information, even when nonzero correlation of errors and regressors is allowed. Correlation on itself does not explain possible differences in estimators. Suppose that we have an uninformative prior distribution except for the variance of errors, that is, $\Upsilon^{-1} = 0, P^* = 0, \pi^* = 0$. We can analyse how this result changes for two extreme cases: a degenerate and an uninformative prior distribution for the idiosyncratic effects.

Assuming $\Pi = 0$ is equivalent to excluding π from the regression and assuming with certainty that its value is zero, which implies that $A^* = \Upsilon$ and, thus, $Q_m = \Sigma_m^{-1}$. The uncorrelated component of the effects produces an RE estimator:

$$\begin{aligned}
 V_p^m &= \left(X' \Sigma_m^{-1} X \right)^{-1} = V_p^r, \\
 \bar{\beta}_p^m &= \left(X' \Sigma_m^{-1} X \right)^{-1} X' \Sigma_m^{-1} Y = \bar{\beta}_p^r.
 \end{aligned}$$

On the other hand, ignorance about π , that is, $\Pi^{-1} = 0$, implies $Q_m = Q_{1m}(Z_m, \Sigma_m)$. In this case, the transformation $\Gamma = (Z'Z)^{-1}Z'X$ and a diagonal matrix $\Sigma = \sigma^2 I$ yield the particular result (derived in Appendix A.2.):

$$X'Q_{1m} = XQ_1, \quad (13)$$

where, as in Section 2, $Q_1 = Q_1(Z, \Sigma)$. The estimator is then:

$$\begin{aligned} V_p^m &= (X'Q_1X)^{-1} = V_p^f, \\ \bar{\beta}_p^m &= (X'Q_1X)^{-1} X'Q_1Y = \bar{\beta}_p^f. \end{aligned}$$

This is a special case where prior information on the uncorrelated component is ‘swept out.’ An FE estimator for a completely uninformative prior distribution on effects arises, yet according to the definition of Section 3, it should also be considered an RE estimator.

We conclude that the different use of prior knowledge on the correlated component of the effects π explains the difference in slope estimators. Allowing for correlation between effects and regressors does not, on itself, modify the slope estimator if the same information is used.

7 Conclusion

This paper has developed a common and tractable framework to discuss the Classical and Bayesian approaches to the analysis of the constant-slope variable-intercept model. The main concern has been with the estimation of the slope parameter β that is common to all individuals; the individual parameters α_i have been of interest to the extent that the way they are treated in the estimation affects the slope estimator. It is usually believed that fixed and random effects estimators are the result of different estimation methods. This paper has shown that, rather than the method or approach adopted, it is the difference in information content captured by the prior distribution of individual effects used in the estimations that produces different slope estimators. In other words, by choosing an estimation method, a Classical econometrician is actually choosing a prior distribution of effects, like her Bayesian colleague. The important point here is that the slope estimator is invariant to the introduction of information on effects

- (i) as parameters or as an error term,
- (ii) in different stages of a hierarchical model, or
- (iii) allowing for correlation between effects and regressors.

What matters is to use all available information on idiosyncratic parameters in the estimation, whatever the approach or method. If different methods produce different results, it is ultimately because somehow different prior information is being used.

Acknowledgements

I thank Núria Quella for helpful comments. Financial support of the Spanish Ministry of Science and Technology (Grant SEC 2001-0674) is gratefully acknowledged. All errors and omissions are mine.

Appendix

A.1. Derivation of Eqs. (8)-(12) in Section 2

Note that

$$\begin{aligned}\Omega^{-1} &= \begin{pmatrix} A & P \\ P' & V \end{pmatrix}^{-1} = \begin{pmatrix} A^{*-1} & -A^{*-1}P^* \\ -P^{*'}A^{*-1} & V^{-1} + P^{*'}A^{*-1}P^* \end{pmatrix}, \text{ and} \\ X_Z'\Sigma^{-1}X_Z &= \begin{pmatrix} Z'\Sigma^{-1}Z & Z'\Sigma^{-1}X \\ X'\Sigma^{-1}Z & X'\Sigma^{-1}X \end{pmatrix};\end{aligned}$$

Eq. (10) and Eq. (12) follow then from Eq. (6). For Eq. (8):

$$\begin{aligned}V_p^{-1} &= -P_p^{*'}A_p^{*-1}P_p^* + X'\Sigma^{-1}X + P^{*'}A^{*-1}P^* + V^{-1}, \\ &= \left(X^{*'}\Sigma^{-1}Z (Z'\Sigma^{-1}Z)^{-1} A_p^* A^{*-1} - X'\Sigma^{-1}Z (Z'\Sigma^{-1}Z)^{-1} \right) (Z'\Sigma^{-1}X - A^{*-1}P^*) \\ &\quad + X'\Sigma^{-1}X + P^{*'}A^{*-1}P^* + V^{-1}, \\ &= X^{*'}\Sigma^{-1}Z (Z'\Sigma^{-1}Z)^{-1} \{ A_p^* A^{*-1}Z'\Sigma^{-1}X + (I - A_p^* A^{*-1}) A^{*-1}P^* \} + X'Q_1X + V^{-1}, \\ &= X^{*'}Q_2X^* + X'Q_1X + V^{-1} = X^{*'}QX^* + V^{-1}.\end{aligned}$$

Note that $Z'Q_1 = 0$, therefore $X^{*'}Q_1X^* = X'Q_1X$, and $X^{*'}Q_1Y^* = X'Q_1Y$. Eq. (11) follows from Eqs. (3)-(5) and Eq. (7), which also determine Eq. (9):

$$\begin{aligned}V_p\bar{\beta}_p &= \left(P_p^{*'}Z'\Sigma^{-1} + X'\Sigma^{-1} \right) Y + \left(P_p^{*'} - P^{*'} \right) A^{*-1}\bar{\alpha} \\ &\quad + \left(V^{-1} - \left(P_p^{*'} - P^{*'} \right) A^{*-1}P^* \right) \bar{\beta}, \\ &= X'Q_1Y + X^{*'}\Sigma^{-1}Z (Z'\Sigma^{-1}Z)^{-1} A_p^* A^{*-1}Z'\Sigma^{-1} \{ Y + Z(\bar{\alpha} - P^*\bar{\beta}) \} + V^{-1}\bar{\beta}, \\ &= X'Q_1Y + X^{*'}Q_2Y^* + V^{-1}\bar{\beta} = X^{*'}QY^* + V^{-1}\bar{\beta}.\end{aligned}$$

A.2. Derivation of Eq. (13) in Section 6

Note that:

$$\begin{aligned}\Sigma_m^{-1} &= (\Sigma + ZVZ')^{-1} = Q_1 + Q_{2v}, \\ Z'_m \Sigma_m^{-1} &= X'Q_{2v},\end{aligned}$$

where it is used that $Z'Q_1 = 0$. Note that if $\Sigma = \sigma^2 I$, then $Z'_m \Sigma_m^{-1} Z_m = Z'_m \Sigma_m^{-1} X = X'Q_{2v}X$. Then Eq. (13) follows:

$$\begin{aligned}Q_{1m} &= \Sigma_m^{-1} - \Sigma_m^{-1} Z_m \left(Z'_m \Sigma_m^{-1} Z_m \right)^{-1} Z'_m \Sigma_m^{-1}, \\ X'Q_{1m} &= X' \Sigma_m^{-1} - X'Q_{2v} = X'Q_1.\end{aligned}$$

References

- Balestra, P. and M. Nerlove (1966). Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas. *Econometrica* 34, 585–612.
- Baltagi, B. H. (1995). *Econometric Analysis of Panel Data*. Chichester: John Wiley.
- Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics* 18, 5–46.
- Chamberlain, G. (1984). Panel Data. In Z. Griliches and M. Intriligator (Eds.), *Handbook of Econometrics Vol. 2*, pp. 1247–1318. Amsterdam: North Holland.
- Henderson, Jr., C. R. (1971). Comment on "The Use of Error Components Models in Combining Cross Section with Time Series Data". *Econometrica* 39, 397–401.
- Hsiao, C. (1986). *Analysis of Panel Data*. Cambridge: Cambridge University Press.
- Laird, N. M. and J. H. Ware (1982). Random-Effects Models for Longitudinal Data. *Biometrics* 38, 963–974.
- Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics* 95, 391–413.

- Lindley, D. and A. Smith (1972). Bayes Estimates for the Linear Model. *Journal of the Royal Statistical Society. Series B (Methodological)* 34, 1–41.
- Maddala, G. S. (1971). The Use of Variance Components Models in Pooling cross section and Time Series Data. *Econometrica* 39, 341–358.
- Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica* 49, 483–509.
- Nerlove, M. (2002). Essays in panel data econometrics. forthcoming in Cambridge University Press.
- Smith, A. (1973). A General Bayesian Linear Model. *Journal of the Royal Statistical Society. Series B (Methodological)* 35, 67–75.
- Sweeting, T. J. (1982). A Bayesian Analysis of Some Pharmacological Data Using a Random Coefficient Regression Model. *Applied Statistics* 31, 205–213.
- Wallace, T. and A. Hussain (1969). The Use of Error Components Models in Combining Cross Section with Time Series Data. *Econometrica* 37, 55–72.
- Waternaux, C., N. M. Laird, and J. H. Ware (1989). Methods for Analysis of Longitudinal Data: Blood Concentrations and Cognitive Development. *Journal of the American Statistical Association* 84, 33–41.
- Zellner, A. (1999). Bayesian and non-bayesian approaches to scientific modeling and inference in economics and econometrics. mimeo. U. of Chicago.