

Case Studies in Multi-unit Longitudinal Models with Random Coefficients and Patterned Correlation Structure ¹

Johannes Ledolter

Department of Statistics

Vienna University of Economics and Business Administration

Abstract: Modelling issues in multi-unit longitudinal models with random coefficients and patterned correlation structure are illustrated in the context of three data sets. The first data set deals with short time series data on annual death rates and alcohol consumption of twenty-five European countries. The second data set deals with glaciologic time series data on snow temperature at 14 different locations within a small glacier in the Austrian Alps. The third data set consists of annual economic time series on factor productivity, and domestic and foreign research/development (R&D) capital stocks. A practical model building approach—consisting of model specification, estimation, and diagnostic checking—is outlined in the context of these three data sets.

Zusammenfassung: Diese Arbeit beschäftigt sich mit dem statistischen Modellieren von einer Liste von meist kurzen Zeitreihen. Die Analyse von Querschnitts- und Zeitreihendaten wird dadurch in einem Modell verbunden. Die Modelle bestehen aus linearen Regressionskomponenten, deren Koeffizienten sowohl nicht-stochastisch („fixed“) also auch stochastisch („random“ über die Liste der Reihen) sein können, und stochastischen Modellen zur Beschreibung der Autokorrelation der Zeitreihen. Drei Datensätze werden analysiert: Jahreszeitreihen von Sterberaten und Alkoholkonsum von europäischen Ländern; jährliche Schneetemperaturen erhoben an mehreren Vermessungsstationen im Gletschergebiet Sonnblick; und jährliche Daten über Faktorproduktivität und dem Kapitalstock von OECD Ländern.

Keywords: Bayesian Information Criterion; Mixed Linear Model; Multi-Unit Longitudinal Model; Random Coefficients Model; (Restricted) Maximum Likelihood; Shrinkage Estimation

1 The Model

The model considered in this paper is given by

$$Y_i = Z_i\alpha + X_i\beta_i + \epsilon_i \quad (1)$$

where for each unit or subject i ($i = 1, 2, \dots, N$) the response vector Y_i is a $(T_i \times 1)$ vector. The design matrix Z_i models the effects of the *fixed* regression coefficients α . The design matrix X_i models the effects of the *random* regression coefficients β_i , which are allowed to vary independently across the N units. The regression coefficients β_i are assumed to

¹Data analyzed in this paper are available from the author.

be drawn randomly from a normal distribution with mean vector 0 and covariance matrix D ; non-zero mean vectors of the random regression coefficients are incorporated into the fixed effects. The error vector for unit i , ϵ_i , is assumed normal with mean vector zero and $(T_i \times T_i)$ covariance matrix $\Sigma_i(\theta)$, parameterized by the vector θ of unknown coefficients.

The model is very general. Without the random-effects component (that is, without the term $X_i\beta_i$), the model reduces to the multivariate regression model $Y_i = Z_i\alpha + \epsilon_i$. For $Z_i = X_i$, the model in Equation (1) simplifies to the general linear regression model with random coefficients:

$$Y_i = X_i(\alpha + \beta_i) + \epsilon_i = X_i\alpha + v_i$$

where the errors $v_i = X_i\beta_i + \epsilon_i$ are $N(0, X_iDX_i' + \Sigma_i)$. For known covariance matrices D and Σ_i the maximum likelihood (generalized least squares) estimate of α is a weighted average of the unit-specific generalized least squares estimates $[X_i'(\Sigma_i)^{-1}X_i]^{-1}X_i'(\Sigma_i)^{-1}Y_i$; the weights depend on the covariance matrices D and Σ_i . If the design and covariance matrices are the same across the N units ($X_i = X$ and $\Sigma_i = \Sigma$), the maximum likelihood (generalized least squares) estimate of α is the *average* of the unit-specific generalized least squares estimates.

The covariance structure for the error components ϵ_i , $\Sigma_i(\theta)$, can be parameterized in many different ways, ranging from the independence assumption where $\Sigma = \sigma^2I$ to a completely unstructured covariance matrix Σ ; the unit index i is omitted in order to simplify the notation. Structured parameterizations of the covariance matrix include compound symmetry with $\Sigma = (\sigma)^2I + (\sigma_1)^211'$, first-order autoregressive models with $\Sigma = \sigma_{i,j} = \sigma^2\rho^{|i-j|}$, ante-dependence models, covariance matrices with Toeplitz structure, as well as their heterogeneous generalizations, and spatial covariance structures if locations are involved.

The model in Equation (1) can be written as

$$Y_i = Z_i\alpha + v_i \quad \text{where } v_i = X_i\beta_i + \epsilon_i \quad (2)$$

The combined random component v_i has covariance matrix $\Omega_i = X_iDX_i' + \Sigma_i$. For example, the covariance matrix that is implied by a regression model with random intercepts and independent errors ϵ_i , is given by $\Omega = \sigma^2I + (\sigma_\beta)^211'$. It amounts to assuming a model without random coefficients, but with a compound symmetric error structure. This discussion shows that one can parameterize a compound symmetric structure in two ways: as a regression model without random coefficients and compound symmetric errors, or as a regression model with random intercepts and independent errors.

Certain parametrizations must be avoided, as the resulting overparametrized models cannot be estimated. For example, it is not possible to consider a model with a random intercept and a general unstructured covariance model, as in this case the $T(T+1)/2$ parameters in the covariance matrix of v_i are modeled through $[1+T(T+1)/2]$ parameters (namely, the variance σ_β^2 and the full covariance matrix of the ϵ_i).

The models in (1) and (2) are linear models. They allow for fixed regression effects, random coefficients, and general correlation structures that can accommodate heteroscedasticity and serial correlation in the error vector ϵ_i . The model can handle unequal numbers of observations for each unit, as well as unequal time-spacing of the observations.

2 Model Estimation

The model in Equations (1) or (2) involves two types of parameters: fixed effects α and parameters in the covariance matrices D and $\Sigma_i(\theta)$. PROC MIXED within the SAS Software can be used for their estimation. PROC MIXED calculates either maximum likelihood (ML) or restricted maximum likelihood (REML) estimates of the covariance parameters in the random components of the model; that is, the parameters in the covariance matrix D , possibly structured, as well as the parameters θ in $\Sigma_i(\theta)$. Mixed models theory (for a general treatment, Searle et al. (1992); for an application to multi-unit time series, Tsimikas and Ledolter (1997), Tsimikas and Ledolter (1998)) can be used to obtain the best linear unbiased estimator (BLUE) of α and the best linear unbiased predictors (BLUP) of the random coefficients β_i . One finds that they are given by

$$\begin{aligned} a &= [\Sigma_{i=1,N} Z_i'(\Omega_i)^{-1} Z_i]^{-1} [\Sigma_{i=1,N} Z_i'(\Omega_i)^{-1} Y_i] \\ b_i &= DX_i'(\Omega_i)^{-1} (Y_i - Z_i a) \end{aligned} \quad (3)$$

where unknown parameters in D and $\Sigma_i(\theta)$ are replaced by their respective ML (or restricted ML) estimates. The estimates in (3) can be used to calculate $Z_i a + X_i b_i$, the BLUP of the signal $Z_i \alpha + X_i \beta_i$, as well as the residuals $Y_i - (Z_i a + X_i b_i)$. These residuals are referred to as the Empirical Bayes residuals, as compared to the cross-sectional residuals $Y_i - Z_i a$ which remove only the fixed, but not the random effects.

A Bayesian treatment of the model in Equation (1) is also possible. The calculation of the posterior distributions of the variance components requires high-dimensional integration, and Monte Carlo methods have been used for the calculations. A procedure that iterates between sampling the random effects for fixed covariances and sampling the covariances for given random effects and observations, has proved useful.

Models with the same fixed effects, but different parametrizations of the random components, can be compared via the Akaike Information Criterion (AIC) and/or the Schwarz Bayesian Information Criterion (BIC); models with large values for these criteria are preferable. The Akaike Information Criterion is given by $AIC = \text{Log-likelihood}(\theta, D) - q$, where q is the number of covariance parameters in D and $\Sigma(\theta)$. The Bayesian Information Criterion $BIC = \text{Log-likelihood}(\theta, D) - (q/2)\log(N^*)$, where N^* is the total number of observations, penalizes models with more covariance parameters even more. One can also use the restricted log-likelihood in place of the ordinary log-likelihood; then N^* in BIC is replaced by $N^* - p$, where p is the number of fixed regression coefficients. These are exactly the calculations that are carried out in Version 6.12 of the SAS PROC MIXED routine.

A Revised Bayesian Information Criterion

The likelihood function in a multi-unit longitudinal model with random coefficients and patterned correlation structure can be written as

$$L(y_1, y_2, \dots, y_N) = L(y^N) = [L(y^N | \xi) \pi(\xi)] / \pi(\xi | y^N) \quad (4)$$

where y^N is the vector of all available observations, $\xi = (\theta, D)$ is the vector of covariance parameters, $\pi(\xi)$ is the prior and $\pi(\xi | y^N)$ is the posterior distribution of ξ . The Schwarz

Bayesian Information Criterion is obtained by approximating the posterior distribution of ξ through a normal distribution, $N(\hat{\xi}, G)$. Then

$$\log L(y^N) \approx \log L(y^N | \hat{\xi}) + (1/2) \log |G| + \log \pi(\hat{\xi}) + (q/2) \log(2\pi) \quad (5)$$

where $q = q^* + q^{**}$ is the total number of parameters; q^* is the number of parameters in the patterned covariance matrix $\Sigma_i(\theta)$, and q^{**} is the number of parameters in the covariance matrix D . The first two terms in the likelihood change with the number of observations. In many situations the asymptotic covariance matrix of the estimate $\hat{\xi}$ is given by $n^{-1}G_0$, where n is the number of observations “relevant” to the estimation of the parameters ξ ; that is, $G \approx n^{-1}G_0$ or $\log |G| \approx \log |G_0| - q \log(n)$. Hence

$$\log L(y^N) \propto \log L(y^N | \hat{\xi}) - (q/2) \log(n). \quad (6)$$

The expression on the right hand side of this equation is known as the Schwarz Bayesian Information Criterion.

The determination of the “relevant number” of observations n in the Schwarz Bayesian Information Criterion is complicated by the fact that multi-unit longitudinal models with random coefficients and patterned correlation structure contain two types of parameters: (1) The parameters θ in the patterned correlation structure of the errors within each of the N units; the precision of their estimates grows with the total number of observations, $N^* = T_1 + \dots + T_N$. (2) The parameters that describe the variation of the model coefficients across the N units (that is, the parameters in the covariance matrix D); the precision of those estimates grows with the number of units N , but not with the number of observations within each unit. This discussion shows that the precisions of the parameter estimates have different growth rates towards their asymptotic covariance matrices. Taking these differences into account, one is led to the following, revised BIC criterion:

$$BIC = \log L(y^N | \hat{\xi}) - (q^*/2) \log(N^*) - (q^{**}/2) \log(N), \quad (7)$$

where q^* is the number of parameters in the patterned covariance matrix $\Sigma_i(\theta)$, q^{**} is the number of parameters in the covariance matrix D , N is the number of units, and N^* is the total number of observations. This definition of BIC appears preferable to the one used in Version 6.12 of SAS PROC MIXED. Note that the latest version of SAS has changed the calculation of the BIC criterion. Version 7 uses the number of units N in place of the total number of observations N^* . Our revised BIC criterion in Equation (7) is a compromise between the definitions in SAS Versions 6 and 7.

Comments: An alternative approach compares models by their posterior probabilities, or by their model likelihoods if equal priors can be assumed. Frühwirth-Schnatter (1995) pursues this approach in the context of linear Gaussian state space models; Frühwirth-Schnatter (1999) addresses the mixed effects models. The computational problem is one difficulty, as this Bayesian approach requires several integrations that can be achieved only through Monte Carlo methods.

3 Model Building Approach

The estimation of multi-unit longitudinal models with random coefficients and patterned covariance matrices is relatively straightforward, and likelihood-based inference proce-

dures are readily available within the SAS software package. However, the first step in any data modeling exercise involves the specification of tentative models. In particular, one must select the fixed regression structure, as well as the covariance structures for the random effects (matrix D) and the noise ($\Sigma_i(\theta)$). In rare situations, models can be based on theory. However, in most instances data must be used for the specification of tentative models. A tentative specification stage with several useful graphical specification tools is desirable, as such a modeling strategy will keep the investigator from fitting too many inappropriate models. A solid model specification step can also avoid the numerical convergence problems that are associated with fitting misspecified or overparametrized models.

Let us first consider the situation where the regression components are functions of time. This is the case in the typical *growth curve model*. A common modeling strategy calculates the average and the standard deviation from all observations at each fixed time period and standardizes the N observations at time t by subtracting the average and dividing by the standard deviation. The standardization removes the mean effects and adjusts for possible heteroscedasticity. It is akin to working with a very unstructured regression component where observations at each time period are allowed to have their own separate level. Parametric models for the mean function can be considered at a later stage, after one has determined the appropriate noise structure. The standardized deviations can be used to form hypotheses about the correlation structure of the noise $v_i = X_i\beta_i + \epsilon_i$. Draftsman displays and parallel axis plots help with the specification.

A *draftsman display* is a matrix of scatterplots of observations (deviations) from the same unit at times 1 and 2, times 1 and 3, . . . , times p and $p-1$. It shows how observations on the same unit, but at different time periods, are related. The matrix of scatterplots supplemented with the correlation matrix, can be used to specify the appropriate noise structure. For example, if the correlations at all lags are of about the same magnitude, then a model with compound symmetry structure is reasonable. If the autocorrelations decay exponentially with the time lag, then an autoregressive model can be considered. Heterogeneous versions of these models should be considered if the standard deviations at the various time points appear different.

Parallel axis plots connect observations from the same unit. For two time periods, a parallel axis plot shows N lines. Lines that cross a lot are an indication of negative correlation. For positive correlation, the lines are more or less parallel and have a tendency not to cross. For zero correlation the situation is somewhere in between; some crossing, but certainly not as much crossing as in the case of negative correlation. This type of plot is easily extended to more than two time periods.

Other tools that help with the specification of the noise structure are the *autocorrelation function* and the *variogram* of the deviations. The theoretical variogram of a stochastic process Y_t , where t is time, is given by $\gamma(u) = (1/2)E[(Y_t - Y_{t-u})^2]$. The sample variogram is calculated from observed half-squared-differences between pairs of observations, $d_{i,u} = (1/2)(Y_{it} - Y_{i,t-u})^2$; it is the average of all observed differences $d_{i,u}$ that correspond to a particular value of u . With irregular sampling schemes (where time is not evenly-spaced, or where irregular space is involved), one can use non-parametric kernel-based smoothing procedures to obtain an estimate of the variogram. The relationships between the noise model and its implied variogram (as well as its autocovariance function

which is functionally related to the variogram) are well known, Diggle et al. (1994). Each noise model implies a certain pattern in the variogram. For example, compound symmetry of the noise implies a horizontal variogram, whereas a first-order autoregressive model leads to a variogram that grows with increasing time lag u towards a new asymptote.

So far our model has not included genuine regressor variables; in growth curve models, the only “explanatory” variable had been time. Going beyond this special situation, we now address the model specification with more general regressor variables. The decision whether or not to include random regression coefficients becomes of interest. One obvious way to answer this question is to fit two models—the model with and the model without random coefficients—and use a likelihood-ratio test or the AIC/BIC criteria to decide whether random coefficients are needed. Another approach is to look for graphical specification tools that can shed light on this issue. Such an approach is pursued in the remainder of this section.

Let us consider the simple linear regression model with random intercept and slope,

$$Y_{it} = (\alpha_0 + \beta_{0i}) + (\alpha_1 + \beta_{1i})x_{it} + \epsilon_{it} \quad (8)$$

where (β_{0i}, β_{1i}) are independent across units, with mean vector zero and (2×2) symmetric covariance matrix D with elements $\{d_{00}, d_{01}, d_{11}\}$. For simplicity, let us assume that all fixed components are known and removed from the model. The cross-sectional residuals (that is, the differences between the observations and the fixed-effects part of the model) are then given by

$$r_{it} = \beta_{0i} + \beta_{1i}x_{it} + \epsilon_{it} \quad (9)$$

Consequences of the random coefficients specification are as follows:

- The variance of the cross-sectional residuals depends on the values of the regressor variable,

$$Var(r_{it}) = d_{00} + d_{11}(x_{it})^2 + 2d_{01}(x_{it}) + (\sigma_\epsilon)^2$$

This variance is smallest when $x = -d_{01}/d_{11}$. Only if the variance is minimal at $x = 0$, would it be advisable to assume that intercept and slope are uncorrelated (that is, $d_{01} = 0$). Uncorrelated intercepts and slopes are appropriate, however, if one adjusts the explanatory variable by shifting its level by the necessary amount (namely, by subtracting d_{01}/d_{11} from the observations x_{it}).

- The cross-sectional residuals r_{it} and $r_{i,t-u}$ on the same unit i are correlated—not just due to the correlation among successive errors ϵ_{it} , but also because of the random nature of the regression coefficients,

$$Cov(r_{it}, r_{i,t-u}) = Cov(\epsilon_{it}, \epsilon_{i,t-u}) + d_{00} + (x_{it}x_{i,t-u})d_{11} + (x_{it} + x_{i,t-u})d_{01}.$$

- Cross-sectional residuals from different units, r_{it} and $r_{j,t-u}$ for $i \neq j$, are always uncorrelated, irrespective whether or not the regression coefficients are assumed random.

An investigation whether cross-sectional residuals exhibit some of these features forms the basis of our graphical specification tool. We recommend a scatterplot of r_{it}^2 against

x_{it} , preferably smoothed, to bring out the features of the plot. Several different smoothing procedures can be used, such as kernel-based smoothing, spline smoothing, and local regression (loess) smoothing approaches. A good summary of smoothing approaches is given in Simonoff (1996) and Bowman and Azzalini (1997); the loess smoother is introduced in Cleveland (1979). In situations where regression coefficients are random, a scatterplot of the squared residuals, r_{it}^2 , against the explanatory variable, x_{it} , will show a (non-linear) relationship. In the absence of random coefficients, no such relationship should be discernable.

One can extend this approach to models with several regressors. If the model includes several regressor variables, then a plot of r_{it}^2 should be carried out against each regressor. Consider the case of two regressors, X and W . If random regression coefficients are present, then one should see patterns in the plots of r_{it}^2 against x_{it} , r_{it}^2 against w_{it} , and r_{it}^2 against the product $x_{it}w_{it}$. However it should be noted at this point that also a heteroscedastic error model with a standard deviation that grows with the explanatory variable leads to similar patterns. Hence it is not always straightforward to distinguish between heteroscedasticity and random regression coefficients.

A comparison of various models with respect to their fits (as indicated by the maximum value of their likelihood functions, or their values on the AIC/BIC criteria) is not always sufficient, as even well-fitting models may not satisfy all assumptions. The validity of the estimated models must be checked before models are put to their use. Residuals are used to check the model assumptions. Commonly-used residual diagnostics include residual plots and the autocorrelation function (or the variogram) of the residuals. These checks will be illustrated in the context of our empirical analysis.

4 Analysis of Three Data Sets

We apply the model building strategy that we have outlined in the previous sections to the analysis of three data sets. The first data set deals with short time series data on annual death rates and annual alcohol consumption of twenty-five European countries during the 1980's. The relationship between death rates and alcohol consumption is at the center of the investigation. The second data set deals with glaciologic time series on snow temperature at 14 different locations within the Wurtenkees glacier, a small glacier in the Sonnblick area of the Austrian Alps. The snowpit locations are at different altitudes, and the role of altitude on snow temperature becomes an interesting question. The third data set consists of annual economic time series on factor productivity, and domestic and foreign research/development (R&D) capital stocks. The relationship between factor productivity and R&D capital stocks is of central importance.

5 Case Study 1: Alcohol Consumption and Mortality Rates

The first case study deals with the analysis of annual death rates (per 100,000 population from all causes, Y) and total absolute alcohol consumption per capita (in liters,

X) of twenty-five European countries ($N = 25$) over the period from 1982 through 1990 ($T = 9$). Austria, Belgium, Bulgaria, Czechoslovakia, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Poland, Portugal, Romania, Spain, Sweden, Switzerland, United Kingdom, the former USSR, and the former Yugoslavia are included in the analysis. The data were provided by Rehm and Her (1997). Our model relates the annual death rate Y_{it} of country (unit) i

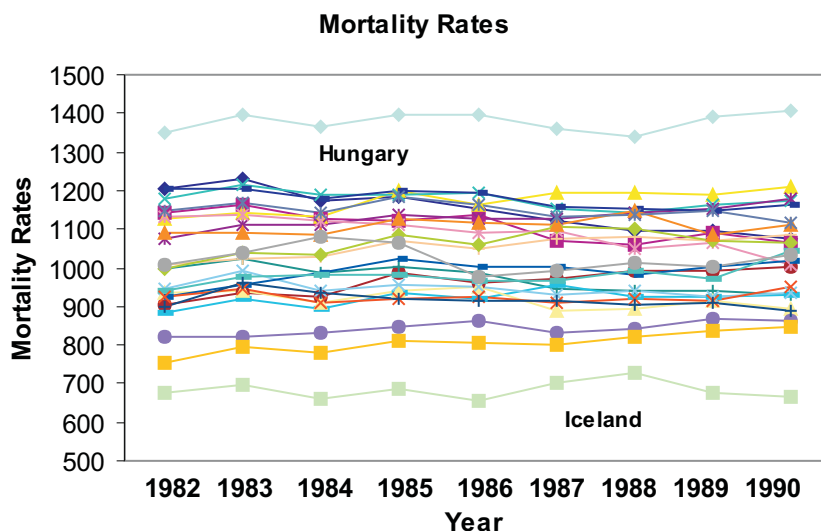


Figure 1: Time-sequence plots of mortality rates.

($i = 1, 2, \dots, 25$) at time t ($t = 1, 2, \dots, 9$) to the annual total absolute alcohol consumption, x_{it} , and a time trend variable, $\text{Time}_t = t$. Time-sequence plots of the mortality rates for the 25 countries in Figure 1 show substantial differences in the levels of mortality. Our initial model

$$Y_{it} = (\alpha_0 + \beta_{0i}) + \alpha_1 \text{Time}_t + \alpha_2 x_{it} + \epsilon_{it} \tag{10}$$

allows for random intercepts. It includes five parameters: two variance components (for the random intercept and for the noise) and three fixed regression coefficients. Empirical Bayes residuals are calculated, and a graph of the squared residuals against time is shown in Figure 2. The pattern in this graph (roughly quadratic, with a minimum value occurring around time 5) suggests random coefficients for the regressor variable “Time”. Model

$$Y_{it} = (\alpha_0 + \beta_{0i}) + (\alpha_1 + \beta_{1i}) \text{Time}_t + \alpha_2 x_{it} + \epsilon_{it} \tag{11}$$

includes seven parameters (three variances, the correlation among random intercept and slope, and three fixed regression coefficients). The maximum of the loglikelihood for this model (-52.63) is a considerable improvement over the one in model (10) which was -108.88. Empirical Bayes residuals from model (11) are calculated and their squares are plotted against alcohol consumption. The graph in Figure 3 shows very little relationship, and it is unclear whether the introduction of a random regression coefficient for alcohol consumption is needed. Nevertheless we proceed by including such a component and consider

$$Y_{it} = (\alpha_0 + \beta_{0i}) + (\alpha_1 + \beta_{1i}) \text{Time}_t + (\alpha_2 + \beta_{2i}) x_{it} + \epsilon_{it}. \tag{12}$$

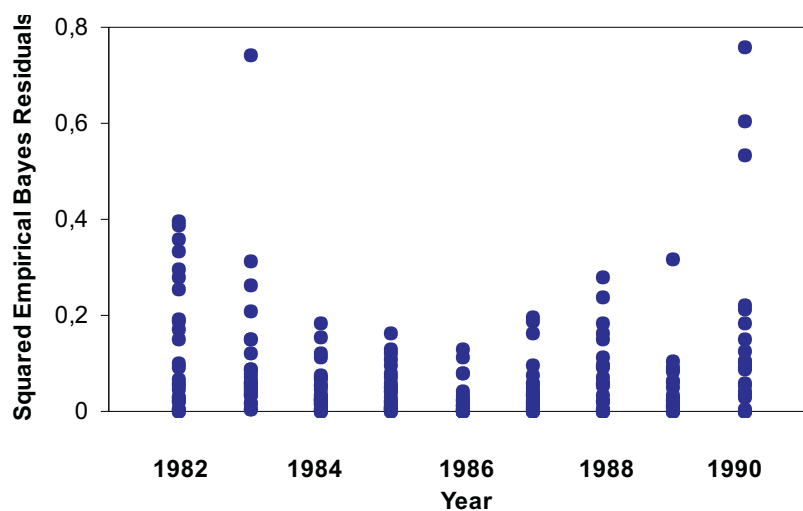


Figure 2: Plot of the squared residuals from model (10) against time.

A model with an unconstrained non-negative definite covariance matrix D is fitted first. The additional three parameters reduce the loglikelihood to -47.11 . The estimated covariance matrix D is very close to being singular, which indicates that a more parsimonious representation of the random structure may be possible. A model where the random regression coefficient for total alcohol is uncorrelated with the coefficients of the random trend component has two fewer parameters, but barely decreases the maximum of the loglikelihood functions (-48.43). This specification is taken as the final model. The estimate of the fixed regression coefficient for alcohol consumption is 13.95 , with an approximate standard error of 4.46 . Alcohol consumption has a significant impact on mortality, even after the adjustment for different country-specific time trends in mortality. The regression coefficient is positive, indicating that with each additional liter of total alcohol the mortality rate increases by 13.95 (per 100,000 population). Periods of higher (lower) than average alcohol consumption coincide with periods of higher (lower) than average mortality rates.

The effect of alcohol consumption on mortality is at the very center of this investigation. We conclude this example by discussing the best linear unbiased predictors of $\alpha_2 + \beta_{2i}$, the regression coefficients for total alcohol. The BLUPs are shown on the bottom line in Figure 4. Twenty-three of the 25 estimates are positive, confirming a very strong positive association between mortality and alcohol consumption. The covariance matrix D of the final model determines how the BLUP's shrink the country-specific regression estimates towards a common value. The common estimate of the regression coefficient for total alcohol (the estimate under the assumption that all 25 regression coefficients are the same; that is $\text{Var}(\beta_{2i}) = 0$) is given by 12.81 . The country-specific regression estimates, on the other hand, are obtained by working with a very large value for $\text{Var}(\beta_{2i})$, reestimating the remaining covariance parameters in model (12), and calculating the implied estimates of $\alpha_2 + \beta_{2i}$. The country-specific estimates are shown on the top line of Figure 4. We notice some shrinkage of the regression coefficients, in particular for Ice-

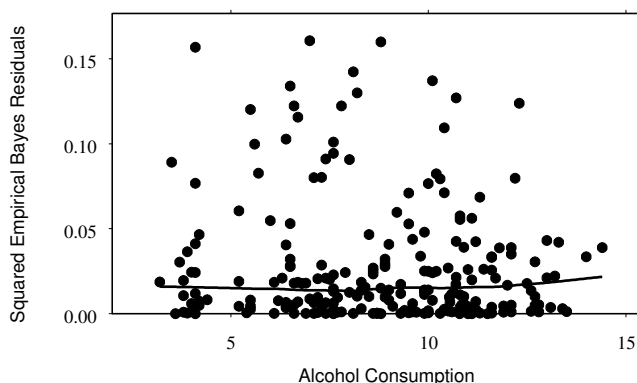


Figure 3: Plot of the squared residuals from model (11) against total alcohol consumption. The line represents the loess smoother.



Figure 4: Estimated regression coefficients for total alcohol: Comparison of (1) country-specific estimates obtained by fitting 25 different regressions, and (2) BLUPs of the random coefficients model in Equation (12).

land. Iceland, with its low alcohol consumption and low mortality rate, is quite different from the other countries.

6 Case Study 2: Temperature of the Winter Snowpack at Wurtenkees Glacier

The glaciological research project Wurtenkees has been carried out since 1982. Wurtenkees, a small glacier in the Sonnblick region of the Austrian Alps, was chosen for intensive investigation because its retreat behavior differs from that of the neighboring glaciers. Wurtenkees retreated during the 1960's and 1970's, when most other Alpine glaciers advanced or stayed the same. Data on many glaciological and meteorological variables were collected under this project, among them the temperature of the winter snowpack. The measurements in Table 1 were taken in the spring of each year (usually early May), over a

period of 11 years (from 1982/83 through 1992/93) . The annual snow temperatures from fourteen sites (snowpits), located at various altitudes, are analyzed in this section. Details on the exact locations of the stations and a discussion of the measurement methods are given in Auer et al. (1995).

Table 1: Mean snow temperatures (degrees Centigrade) for 14 snowpit locations on Wurtenkees glacier, taken over a period of 11 years (around May 1; 1983 through 1993). Snow temperature for station 2 in year 1991 is unusually low (-8 degrees); however, no problems were noted in the original records. Altitude of snowpits is given in meters. Stations are listed in decreasing altitude. The x-y coordinates of the location are given in meters (Gauss-Krüger Koordinaten in Meridianstreifen M41). Reference: Auer et al. (1995).

	82/3	83/4	84/5	85/6	86/7	87/8	88/9	89/90	90/1	91/2	92/3
Station 1	-5.7	-7	0	-4.1	-1	-3	-3.9	-4.1	-3.1	-3	-3
Station 2	-3.6	-4.8	0	-2.1	-1	-4.3	-3.6	-2.3	-8	-2.3	-2.3
Station 3	-3	-5.2	-0.2	-1.5	-0.8	-2.5	-3.1	-1	-4.4	-3.2	-1.9
Station 4	-3.1	-5.5	-0.1	-1	-0.4	-2.2	-3	-1.5	-4.3	-2	-1.6
Station 5	-2.9	-5.6	-0.1	-1.2	-0.9	-2	-2.9	-1.2	-4.6	-2.1	-1.4
Station 6	-3.7	-5	-0.3	-0.2	-0.5	-2	-2.9	-1.2	-4.2	-2.4	-1.1
Station 7	-2.6	-4.8	0	-0.2	-0.3	-1.7	-2.8	-0.9	-3.8	-2.2	-1
Station 8	-2	-4.5	-0.1	-2.3	-0.5	-0.9	-2	0	-2.5	-2	-2.7
Station 9	-2.1	-5	-0.1	-1	-0.6	-0.8	-1.8	-0.7	-2.7	-1.5	-1.8
Station 10	-0.9	-5.2	0	-0.5	-0.9	-0.3	-1.3	0	-0.7	-1.2	0
Station 11	-2	-3.4	0	-1	-1	-0.6	-1.5	-0.3	-1.5	-1.1	-0.4
Station 12	-1.6	-4	0	-0.2	-0.6	-0.1	-1.6	-0.3	-1.2	-0.9	-0.3
Station 13	-1.1	-3.5	0	-1.7	-0.4	-0.5	-0.2	-0.8	-1.3	-1	-0.2
Station 14	-0.5	-3	0	-0.6	-0.2	-0.1	-0.3	0	-0.6	-0.8	0

	Altitude	Location-x	Location-y
Station 1	3061	211199	-23812
Station 2	3023	211001	-23801
Station 3	2946	210824	-23934
Station 4	2898	210751	-24044
Station 5	2840	210653	-24195
Station 6	2813	210486	-24227
Station 7	2743	210521	-24469
Station 8	2683	211073	-24568
Station 9	2677	210817	-24642
Station 10	2655	211175	-24842
Station 11	2647	210921	-24836
Station 12	2592	210848	-25048
Station 13	2593	210954	-25081
Station 14	2533	210778	-25228

Multiple time-sequence plots of snow temperatures are shown in Figure 5; observations from the same station are connected. In 1984/85 most snow temperatures are virtually zero, and there is almost no variability among the measurements. In the spring of 1985 heavy avalanche danger prevented the research team from reaching the snowpits;

measurements were collected later that spring, at a time when the snowpack had started to melt.

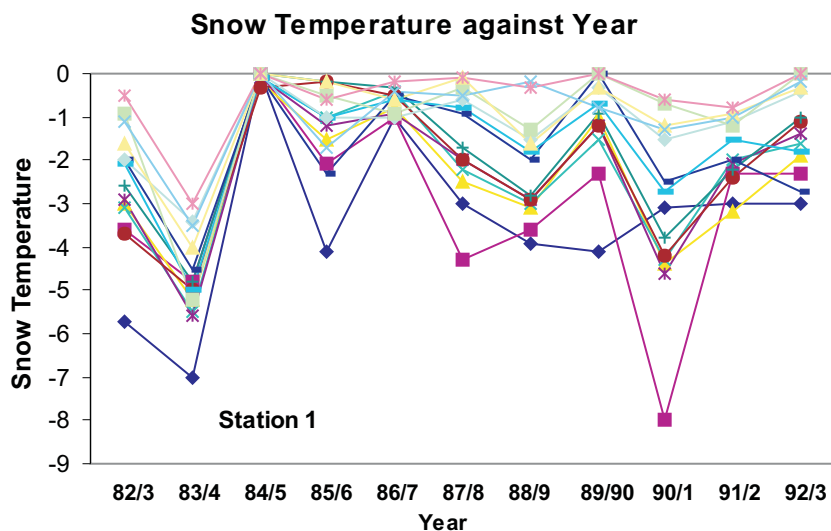


Figure 5: Multiple time-sequence plots of snow temperatures at 14 snowpits on Wurtenkees glacier.

Another way to look at this information is to graph snow temperature against station altitude, and to do this for each year separately. The resulting plot is shown in Figure 6.

Snow temperatures vary from year to year. Altitude of the station has an influence on snow temperature, but the effect of altitude depends on the year. A linear function of altitude, with coefficients that vary from year to year, can be used to approximate the relationship between snow temperature and altitude. One can expect a positive correlation between the random intercept and slope, as in “warm” years the slope coefficient for altitude is less steep (less negative). We consider the model

$$Y_{it} = (\alpha_0 + \beta_{0t}) + (\alpha_1 + \beta_{1t})Alt_i + \epsilon_{it}, \quad (13)$$

where Y_{it} is snow temperature at station i ($i = 1, 2, \dots, 14$) taken at time t ($t = 1, 2, \dots, 11$) and Alt_i is the altitude of station i . In this particular example it is the year that takes the place of the unit, and the regression coefficients are random across years. Altitude is centered by subtracting the midpoint of altitude (2800 meters) from the data. The fitting results for model (13) with independent errors (independence across time periods, as well as stations) are given in Table 2.

A refinement of model (13) replaces the independence assumption across stations, $Cov(\epsilon_{it}, \epsilon_{jt}) = \sigma^2 I$, with a spatial correlation structure that incorporates the Euclidian distances d_{ij} between stations i and j , while retaining the assumption of independence among errors from different years. The coordinates of the snowpits are measured in meters, according to Gauss-Krüger coordinates in meridian strip M41, and they are listed in Table 1. Commonly-used models for spatial correlation are the exponential model with covariance matrix $Cov(\epsilon_{it}, \epsilon_{jt}) = \sigma^2 \exp[-d_{ij}/\rho]$, the Gaussian model with $Cov(\epsilon_{it}, \epsilon_{jt}) = \sigma^2 \exp[-(d_{ij})^2/\rho]$, and the power representation with $Cov(\epsilon_{it}, \epsilon_{jt}) =$

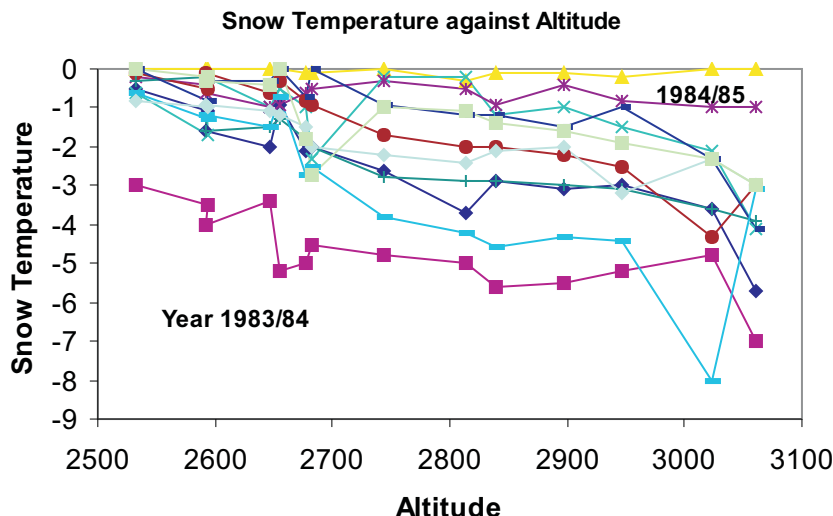


Figure 6: Plots of snow temperature against altitude of the 14 snowpits on Wurtenkees glacier.

$\sigma^2 \exp[\rho^{d_{ij}}]$. A “nugget effect” (that is, an additional variance component when $d = 0$) can also be included. Such models can be fitted easily with SAS PROC MIXED. We found that the addition of a spatial correlation structure, with and without nugget effect, did not lead to an appreciable improvement in the loglikelihood.

Table 2: Estimation Results for Case Study 2. Model $Y_{it} = (\alpha_0 + \beta_{0t}) + (\alpha_1 + \beta_{1t})Alt_i + \epsilon_{it}$.

Intercept	Altitude	Noise	LogLik
Random	NotIncluded	Independent	-251.817
Random	Fixed	Independent	-196.132
Random	Random	Independent	-179.698
(Correlated random coefficients)			

We conclude our discussion of this data set by showing in Figure 7 the BLUPs of the eleven year-specific regression coefficients of the model in Equation (13). One notices the positive association that was anticipated. The coefficients for year 9 (1990/91) are somewhat different from the rest, probably due to one unusual observation at station 2.

7 Case Study 3: Research and Development Spillover

The third example reanalyzes data taken from an article by Coe and Helpman (1995). Coe and Helpman explain total factor productivity (F) in terms of domestic R&D capital stock (CSD) and foreign R&D capital stock weighted by the fraction of imports (CSF). Annual time series data from 1971 through 1990 on 21 OECD countries and Israel are used; logarithms of the data are analyzed. Coe and Helpman base their economic interpretation

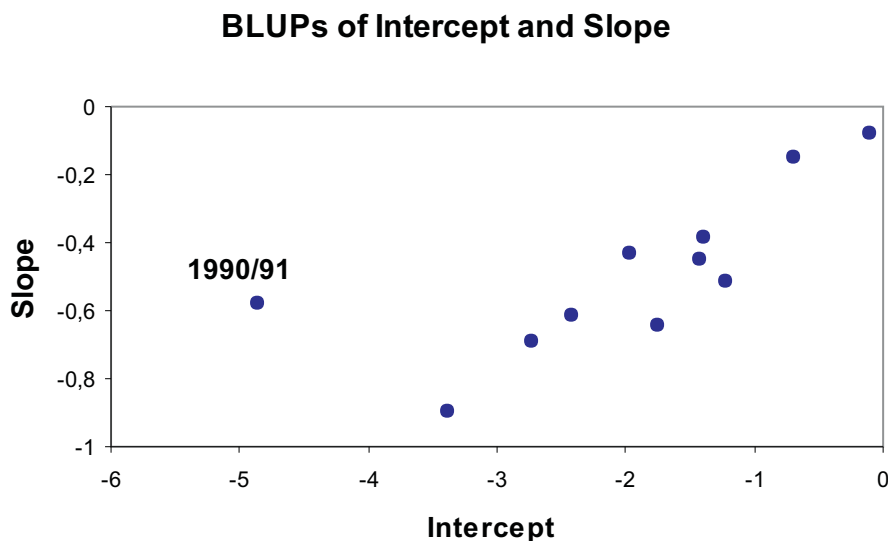


Figure 7: Scatterplot of the BLUPs of the regression coefficients $[(\alpha_0 + \beta_{0t}), (\alpha_1 + \beta_{1t})]$, for $t = 1, 2, \dots, 11$.

on a cross-sectional longitudinal regression model

$$F_{it} = \alpha_{0i} + \alpha_1 CSD_{it} + \alpha_2 CSF_{it} + \epsilon_{it} \quad (14)$$

with country-specific fixed intercepts, constant regression coefficients (constant across countries $i = 1, 2, \dots, 22$) and independent errors ϵ_{it} (independence across time $t = 1, 2, \dots, 20$, as well as countries).

Müller and Nettekoven (1999) consider a random coefficients version of this model,

$$F_{it} = (\alpha_0 + \beta_{0i}) + (\alpha_1 + \beta_{1i})CSD_{it} + (\alpha_2 + \beta_{2i})CSF_{it} + \epsilon_{it} \quad (15)$$

where $\beta_i = (\beta_{0i}, \beta_{1i}, \beta_{2i})$, for countries $i = 1, 2, \dots, 22$, are independent drawings from a normal distribution with mean vector zero and unconstrained covariance matrix D . The errors ϵ_{it} are assumed independent across time and countries, with country-specific variances $(\sigma_i)^2$.

Müller and Nettekoven find that the estimated mean elasticity of foreign R&D capital stock (α_2) in the random coefficients regression model in Equation (15) is negative, while the estimated elasticity of CSF in the Coe and Helpman model in Equation (14) has a positive sign.

In this section of the paper we reanalyze this data set. We consider the following models:

- The regression model with country-specific fixed regression coefficients. This is equivalent to assuming a very large covariance matrix D in model (15).
- The random regression coefficients model in Equation (15), considered by Müller and Nettekoven.
- A revised random coefficients model where the autocorrelations of the errors ϵ_{it} (for $t = 1, 2, \dots, 20$) in Equation (15) are modeled by a first-order autoregression with correlation $Corr(\epsilon_{it}, \epsilon_{it-k}) = \phi^k$.

We start our analysis by estimating the regression model in Equation (14), assuming this time constant intercepts as well as constant regression coefficients. Residuals from that model are calculated and in Figure 8 and the squared residuals are plotted against the explanatory variables CSD and CSF . We notice curvature in these graphs, especially in the plot against CSD . This provides some indication for a random coefficients model.

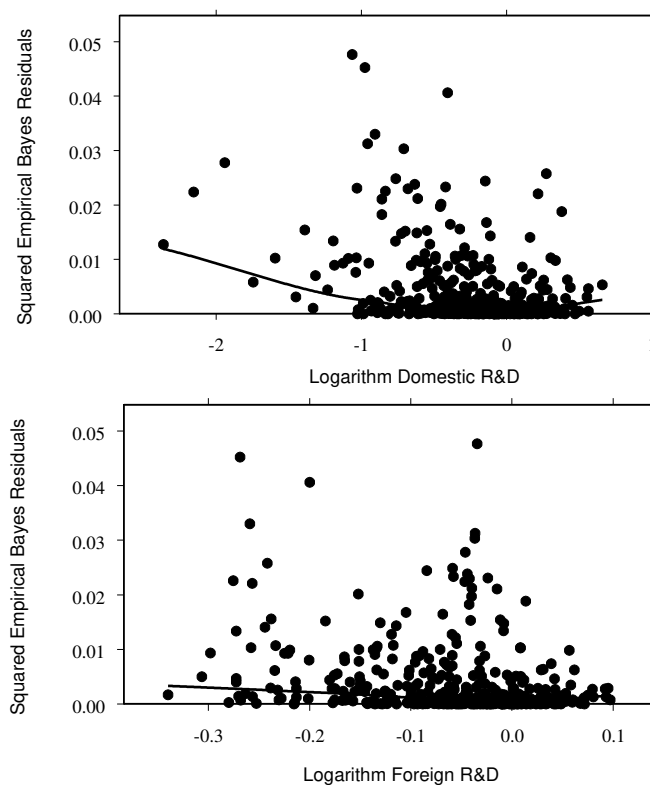


Figure 8: Plot of the squared residuals from the model in Equation (14), assuming constant intercepts as well as constant regression coefficients, against the logarithms of domestic and foreign R&D capital stock. Loess smoother is given by the solid line.

The Müller and Nettekoven model in Equation (15) is reestimated. The maximum likelihood estimates, obtained with the SAS PROC MIXED routine, are very similar to the iterative estimates calculated in Müller and Nettekoven.

Diagnostics of the Müller and Nettekoven model reveal problems as the residuals are autocorrelated. The autocorrelations (lags 1 through 3) of the 22 residual time sequences of length 20 are shown in Figure 9. The patterns in the autocorrelations show that the model in Equation (15) can be improved by adding a first-order autoregressive correlation structure with $Corr(\epsilon_{it}, \epsilon_{it-k}) = \phi^k$. Estimation results for the random coefficients regression model in Equation (15) with and without autocorrelated (first-order autoregressive) errors are compared in Table 3.

The added correlation structure leads to substantial improvements; it increases the maximum of the loglikelihood from 975.05 to 1,036.05. While there are some changes in the estimated mean elasticities and their standard errors, the conclusions regarding their sign and significance are largely unchanged. Adjusting for the effect of domestic R&D

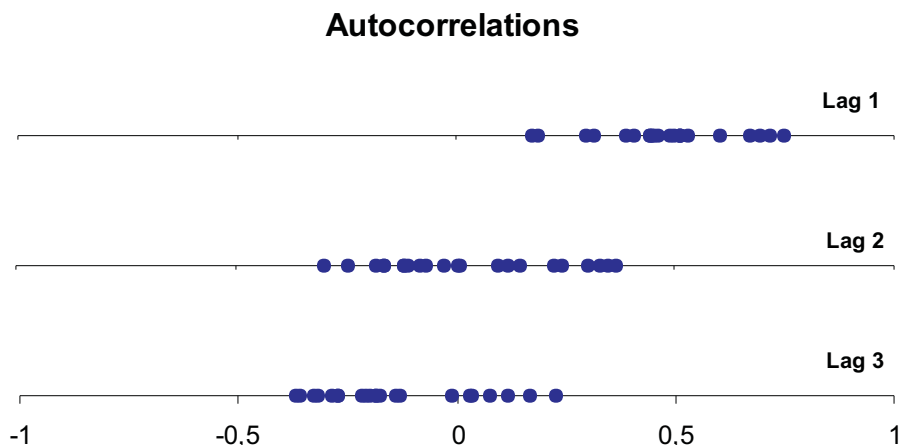


Figure 9: Autocorrelations (lags 1 through 3) of the residuals from the model in Equation (15) with independent errors; 22 countries.

Table 3: Estimates of fixed effects (mean elasticities) in the random coefficients model in Equation (15) with and without autocorrelated errors.

	Müller - Nettekoven Model			Model with AR(1) errors		
	BLUE	StdErr	t-ratio	BLUE	StdErr	t-ratio
Intercept(α_0)	-0.005	0.003	-1.53	-0.005	0.005	-0.94
CSD (α_1)	0.267	0.045	5.89	0.240	0.031	7.83
CSF (α_2)	-0.187	0.062	-3.03	-0.160	0.060	-2.67

capital stock, the estimate of the mean elasticity of foreign R&D capital stock remains negative.

The effects of the random specification on the estimates of the elasticities of domestic and foreign R&D capital stock are illustrated in Figure 10. It shows how the two random coefficients models with and without autocorrelated errors shrink the country-specific regression estimates that are obtained by fitting separate regressions to each of the 22 countries. The addition of the autocorrelation component enhances the shrinkage effect. For elasticities of domestic R&D capital stock the shrinkage is especially noticeable for UK and Switzerland; for foreign capital stock it is largest for Italy and Spain.

8 Concluding Remarks

The three case studies in this paper illustrate a flexible strategy for building regression-type models from multi-unit (panel) data. All considered models can be estimated quite easily within the PROC MIXED routine of the SAS statistical software.

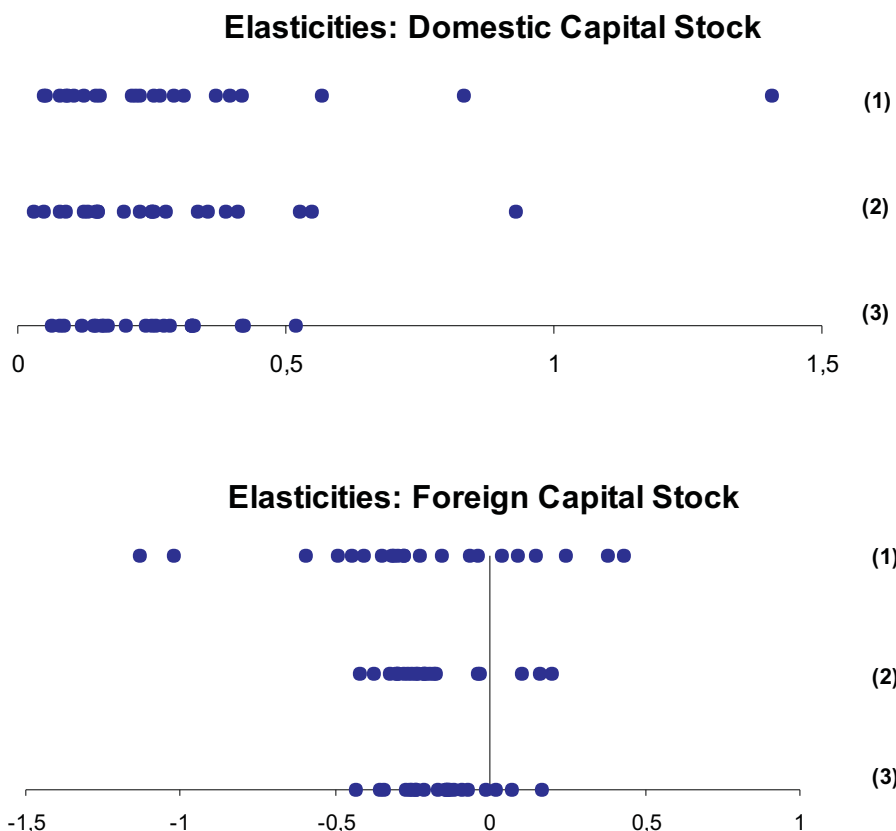


Figure 10: Estimated regression coefficients for domestic and foreign capital stock. Comparison of (1) country-specific estimates obtained by fitting 22 different regressions, (2) BLUPs under the random coefficients model without autocorrelated errors, and (3) BLUPs under the random coefficient model with AR(1) errors.

References

I. Auer, R. Böhm, N. Hammer, W. Schöner, T. Wiesinger, and W. Winiwarter. *Glaciological Studies in the Region of Sonnblick (Austria): Research Programme Wurtenkees*. Zentralanstalt für Meteorologie and Geodynamik, Wien, 1995.

A.W. Bowman and A. Azzalini. *Applied Smoothing Techniques for Data Analysis*. Clarendon Press, Oxford, 1997.

W.S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74:829–836, 1979.

D.T. Coe and E. Helpman. International R&D spillovers. *European Economic Review*, 39:859–887, 1995.

P.J. Diggle, K. Liang, and S.L. Zeger. *Analysis of Longitudinal Data*. Oxford University Press, Oxford, UK, 1994.

- S. Frühwirth-Schnatter. Bayesian model discrimination and Bayes factors for linear gaussian state space models. *Journal of the Royal Statistical Society, Series B*, 57:237–246, 1995.
- S. Frühwirth-Schnatter. *Bayesian Estimation and Model Selection for Fixed Effects Models*. Unpublished preprint, Vienna University of Economics and Business Administration, 1999.
- W. G. Müller and M. Nettekoven. Research&development spillover. *Economics Letters*, 64:37–41, 1999.
- J. Rehm and M. Her. *Alcohol and All-Cause Mortality in Europe 1982-1990: A Pooled Cross-Section Time-Series Analysis*. Unpublished preprint, Addiction Research Foundation, Toronto, Canada, 1997.
- S.R. Searle, G. Casella, and C.E. McCulloch. *Variance Components*. Wiley, New York, 1992.
- J.S. Simonoff. *Smoothing Methods in Statistics*. Springer Verlag, New York, 1996.
- J. Tsimikas and J. Ledolter. Mixed model representation of state space models: New smoothing results and their application to REML estimation. *Statistica Sinica*, 7:973–991, 1997.
- J. Tsimikas and J. Ledolter. Analysis of multi-unit variance components models with state space profiles. *Annals of the Institute of Statistical Mathematics*, 50:147–164, 1998.

Author's address:

Univ.-Prof. Dr. Johannes Ledolter
Department of Statistics
Vienna University of Economics and Business Administration
Augasse 2-6
A-1090 Wien
Tel. +43 1 31336 / 4756
Fax +43 1 31336 / 774
Email: Johannes.Ledolter@wu-wien.ac.at
<http://www.wu-wien.ac.at/wwwu/institute/stat1/tafel1.html>