UNIVERSITY OF GÖTEBORG

Department of Statistics

RESEARCH REPORT 1991:2

ISSN 0349-8034

ON THE PROBLEM OF OPTIMAL INFERENCE IN THE SIMPLE ERROR COMPONENT MODEL FOR PANEL DATA

by

Robert Jonsson

Statistiska institutionen

Göteborgs Universitet
Viktoriagatan 13
S 411 25 Göteborg
Sweden

SUMMARY

For data consisting of cross sections of units observed over time, the Error Component Regression (ECR) model, with random intercept and constant slope, may sometimes be adequate. While most interest has been focused on point estimation of the slope parameter β, little attention has been paid to the problem of making confidence statements and tests about β.

In this paper, the performance of some estimators of β and the corresponding test statistics are investigated. In consideration of bias, efficiency and power of tests, it is shown that the Maximum Likelihood estimator with the corresponding test statistic is outstanding in large samples. But, in the small sample case there are hardly any reasons for the Maximum Likelihood approach. In the latter case, the use of estimators and test statistics based on within- or between group comparisons is suggested.

The results, together with tools for a proper application of the ECR model, are demonstrated on data from a medical follow-up study.

CONTENTS

1. INTRODUCTION 1

2. SOME SAMPLE MOMENTS AND A TRANSFORMATION ..3

3. DERIVATION OF ESTIMATORS 4

3.1 LS APPROACH 4

3.2 ML APPROACH 6

4. COMPARISON BETWEEN THE ESTIMATORS 9

5. TESTS AND CONFIDENCE INTERVALS FOR β 11

6. AN EXAMPLE 14

7. DISCUSSION 16

8. REFERENCES 18

1. INTRODUCTION

Consider a sample of n units from which data y_{ij} are obtained at the times x_{ij}, i=1...t and j=1...n. A simple linear regression model for this pooling of cross section and time series data may be written $\underset{\sim}{y}_j=(\underset{\sim}{1},\underset{\sim}{x}_j)\underset{\sim}{\beta}_j+\underset{\sim}{u}_j$. Here $\underset{\sim}{1}=(1...1)'$ and $\underset{\sim}{x}_j=(x_{1j}...x_{tj})'$ are non-random vectors, $\underset{\sim}{y}_j=(y_{1j}...y_{tj})'$ is a random vector of observations, $\underset{\sim}{\beta}_j=(\beta_{0j},\beta)'$ is a vector of random intercepts β_{0j} and fixed slopes β while $\underset{\sim}{u}_j$ is a random vector of errors. It is assumed that $\underset{\sim}{\beta}_j$ and $\underset{\sim}{u}_j$ are uncorrelated, the $\underset{\sim}{y}_j$'s are uncorrelated and $\beta_{0j}\sim N_1(\alpha,\sigma_\alpha^2)$ while $\underset{\sim}{u}_j\sim N_t(\underset{\sim}{0},\sigma_u^2\underset{\sim}{I})$, where $\sim N_t$ means 'has a t-dimensional normal law' with mean vector and dispersion matrix in the parenthesis.

The model above is an <u>Error Component Regression</u> (ECR) model which is a special case of the <u>Random Coefficient Regression</u> (RCR) model in which both intercept and slope are random. It follows that

$$\underset{\sim}{y}_j\sim N_t\left((\underset{\sim}{1},\underset{\sim}{x}_j)\binom{\alpha}{\beta},\ \begin{pmatrix} ab\ldots b \\ ba\ldots b \\ \vdots\ \ \vdots\ \ \vdots \\ b\ldots ab \\ b\ldots ba \end{pmatrix}\right), \tag{1}$$

where $a=\sigma_\alpha^2+\sigma_u^2$ and $b=\sigma_u^2$.

The ECR model has been used in many fields, especially in econometrics but rather few comparisons between different estimators of the parameters have been reported. In a frequently cited simulation study Maddala and Mount (1973) compared bias and mean squared error between 11 alternative estimators of β, including those given by the Maximum Likelihood (ML) method and several two-step Generalized Least Squares (GLS) methods. It was concluded that 'there is nothing much to chose among these estimators'.

Some effects of substituting estimators for the variance components σ_α^2 and σ_u^2 in the expression for the β-GLS estimator have been studied by Taylor (1980). More efficient estimators of the variance components need not lead to more efficient estimators of β.

While some interest has been focused on the improvement of point estimators of β, little attention has been paid to the problem of making confidence statements and tests about β.
Here, some estimators of β will be compared when they are used as interval estimators and test statistics. The following estimators will be considered: The between-group-, the within-group-, the ordinary Least Squares (LS)- and the ML estimators. Two-step GLS estimators are not considered since their small-sample distributions are complicated and in large samples they are not more efficient than the ML estimator.
The main purpose is to find recommendations for the choice between alternative β-estimators. Estimators of the variance components will only be briefly discussed in connection with the estimation of β.
The results are applied to data consisting of haemoglobin (HbA_{1c}) measurements from diabetic patients.

2. SOME SAMPLE MOMENTS AND A TRANSFORMATION

The following sample moments will be used extensively:

$$\left.\begin{aligned}
&\bar{x}_j=\sum_{i=1}^{t} x_{ij}/t,\quad \bar{y}_j=\sum_{i=1}^{t} y_{ij}/t,\quad \bar{x}=\sum_{j=1}^{n}\bar{x}_j,\quad \bar{y}=\sum_{j=1}^{n}\bar{y}_j/n,\\
&s_{x_jy_j}=\sum_{i=1}^{t}(x_{ij}-\bar{x}_j)(y_{ij}-\bar{y}_j),\ s_{x_jx_j}\text{ and } s_{y_jy_j}.\\
&\text{Within-group sums of squares: } W_{xy}=\sum_{j=1}^{n} s_{x_jy_j},\ W_{xx}\text{ and } W_{yy}.\\
&\text{Between-group sums of squares: } B_{xy}=\sum_{j=1}^{n}(\bar{x}_j-\bar{x})(\bar{y}_j-\bar{y}),\ B_{xx}\text{ and } B_{yy}.\\
&\text{Total sums of squares: } S_{xy}=W_{xy}+B_{xy},\ S_{xx}\text{ and } S_{yy}.
\end{aligned}\right\}\quad (2)$$

The derivations will be much simplyfied by the following transformation:

$$\underset{\sim}{z}_j=\underset{\sim}{M}\underset{\sim}{y}_j,\text{ where } \underset{\sim}{M}=\begin{pmatrix} 1/t^{\frac{1}{2}}\ldots 1/t^{\frac{1}{2}}\\ \hline l_{21}\ldots l_{2t}\\ \vdots \quad\quad \vdots\\ l_{t1}\ldots l_{tt}\end{pmatrix}=\begin{pmatrix}1/t^{\frac{1}{2}}\ldots 1/t^{\frac{1}{2}}\\ \hline \underset{\sim}{L}\end{pmatrix}, \quad (3)$$

where the raws in $\underset{\sim}{M}$ are ortogonal.

Since $\underset{\sim}{M}'\underset{\sim}{M}=\underset{\sim}{I}$ the matrix $\underset{\sim}{L}$ in (3) has the property

$$\underset{\sim}{L}'\underset{\sim}{L}=\underset{\sim}{I}-\underset{\sim}{1}\underset{\sim}{1}'/t. \quad (4)$$

Using (4) the within-group sums of squares can be written

$$W_{xy}=\sum_{j=1}^{n}\underset{\sim}{x}_j'\underset{\sim}{L}'\underset{\sim}{L}\underset{\sim}{y}_j,\quad W_{xx}=\sum_{j=1}^{n}\underset{\sim}{x}_j'\underset{\sim}{L}'\underset{\sim}{L}\underset{\sim}{x}_j\text{ and } W_{yy}=\sum_{j=1}^{n}\underset{\sim}{y}_j'\underset{\sim}{L}'\underset{\sim}{L}\underset{\sim}{y}_j. \quad (5)$$

By the transformation in (3) we obtain the vectors

$$\underset{\sim}{z}_j=\begin{pmatrix}\bar{y}_j t^{\frac{1}{2}}\\ \hline \underset{\sim}{L}\underset{\sim}{y}_j\end{pmatrix}\sim N_t\left(\begin{pmatrix}(\alpha+\beta\bar{x}_j)t^{\frac{1}{2}}\\ \hline \underset{\sim}{L}\underset{\sim}{x}_j\end{pmatrix},\ \left(\begin{array}{c|c} a+b(t-1) & \underset{\sim}{0}'\\ \hline \underset{\sim}{0}' & (a-b)\underset{\sim}{I}\end{array}\right)\right). \quad (6)$$

3. DERIVATION OF ESTIMATORS

GLS- and ML estimation of the slope β has been studied in the more general case when the single slope is replaced by a vector of slopes (Balestra and Nerlove (1966), Maddala (1971)). Simultaneous solution of the resulting equations then becomes complicated (see Hsiao (1986) chap. 3). In the present case it is instructive to show how simple expressions can be derived for the estimators and the standard errors (SE's) of the estimators.

3.1 LS APPROACH

As is seen from (6), the transformation (4) leads to the two uncorrelated components $z_{1j}=\bar{y}_j t^{\frac{1}{2}}$ and $(z_{2j}\ldots z_{tj})'=(\underset{\sim}{L}\underset{\sim}{\bar{y}}_j)'$. The ordinary LS estimators of β and α which are obtained by only using the z_{1j}'s, $j=1\ldots n$, are

$$\tilde{\beta}_b=B_{xy}/B_{xx} \text{ and } \tilde{\alpha}_b=\bar{y}-\tilde{\beta}_b\bar{x}. \tag{7}$$

$\tilde{\beta}_b$ may be called the between-group estimator of β. According to fundamental results in LS theory

$$\begin{aligned}&\tilde{\beta}_b\sim N_1(\beta, \frac{a+b(t-1)}{ntB_{xx}}) \text{ and}\\ &\tilde{\Sigma}_b=\sum_{j=1}^{n}(\bar{y}_j-\tilde{\alpha}_b-\tilde{\beta}_b\bar{x}_j)^2=n(B_{yy}-\tilde{\beta}_b^2B_{xx})\sim(\frac{a+b(t-1)}{t})\chi^2(n-2),\end{aligned} \tag{8}$$

where $\chi^2(n-2)$ denotes a Chi-Square variable with n-2 degrees of freedom (d.f). Furthermore, $\tilde{\beta}_b$ and $\tilde{\Sigma}_b$ are independent.

Similarly, the ordinary LS estimator of β obtained from the vectors $(z_{2j}\ldots z_{tj})'$, $j=1\ldots n$, is

$$\tilde{\beta}_w=(\sum_{j=1}^{n}(\underset{\sim}{L}\underset{\sim}{x}_j)'(\underset{\sim}{L}\underset{\sim}{x}_j))^{-1}\sum_{j=1}^{n}(\underset{\sim}{L}\underset{\sim}{x}_j)'(\underset{\sim}{L}\underset{\sim}{y}_j)= W_{xy}/W_{xx}, \tag{9}$$

where the last equality follows from (5). $\tilde{\beta}_w$ may be called the within-group estimator of β. The residual sum of squares $\tilde{\Sigma}_w = \sum_{j=1}^{n} (L\underset{\sim}{y}_j - L\underset{\sim}{x}_j\tilde{\beta}_w)'(L\underset{\sim}{y}_j - L\underset{\sim}{x}_j\tilde{\beta}_w)$ can be written $nt(W_{yy} - \tilde{\beta}_w^2 W_{xx})$ by using (5). It follows that

$$\tilde{\beta}_w \sim N_1\left(\beta, \frac{(a-b)}{ntW_{xx}}\right) \text{ and}$$
$$\tilde{\Sigma}_w = nt(W_{yy} - \tilde{\beta}_w^2 W_{xx}) \sim (a-b)\chi^2(n(t-1)-1). \tag{10}$$

The statistics $\tilde{\beta}_w$ and $\tilde{\Sigma}_w$ are independent and also independent of $\tilde{\beta}_b$ and $\tilde{\Sigma}_b$ in (8).

The ordinary LS estimator of β, obtained from the complete vectors $\underset{\sim}{z}_j$, $j=1\ldots n$, is $\tilde{\beta}_{LS} = S_{xy}/S_{xx}$. This can be written

$$\tilde{\beta}_{LS} = \frac{B_{xx}\tilde{\beta}_b + W_{xx}\tilde{\beta}_w}{B_{xx} + W_{xx}}, \tag{11}$$

which is a weighted average of the between-group estimator and the within-group estimator. This estimator is normally distributed with mean β and variance

$$V(\tilde{\beta}_{LS}) = \{(a+b(t-1))B_{xx} + (a-b)W_{xx}\}/ntS_{xx}^2 . \tag{12}$$

Unbiased estimators of the variances of the estimators can be derived from the residual sums of squares:

$$\tilde{V}(\tilde{\beta}_b) = \frac{\tilde{\Sigma}_b}{n(n-2)B_{xx}}, \quad \tilde{V}(\tilde{\beta}_w) = \frac{\tilde{\Sigma}_w}{nt(n(t-1)-1)W_{xx}}$$
$$\text{and } \tilde{V}(\tilde{\beta}_{LS}) = \{B_{xx}^2\tilde{V}(\tilde{\beta}_b) + W_{xx}^2\tilde{V}(\tilde{\beta}_w)\}/S_{xx}^2. \tag{13}$$

Finally, unbiased estimators of the variance components are

$$\tilde{\sigma}_u^2 = \tilde{b} = \left\{\frac{t\tilde{\Sigma}_b}{n-2} - \frac{\tilde{\Sigma}_w}{n(t-1)-1}\right\}/t \text{ and } \tilde{\sigma}_\alpha^2 = \tilde{a} - \tilde{b} = \frac{\tilde{\Sigma}_w}{n(t-1)-1} . \tag{14}$$

The properties of these estimators can be obtained from (8) and (10) and from the fact that $\tilde{\Sigma}_b$ and $\tilde{\Sigma}_w$ are independent.

3.2 ML APPROACH

The negative log-likelihood of the transformed vectors in (6) is

$$-\log L=\frac{nt}{2}\log 2\pi+\frac{n}{2}\log(a+b(t-1))+\frac{n(t-1)}{2}\log(a-b)$$
$$+\frac{t\sum_{j=1}^{n}(\bar{y}_j-\alpha-\beta\bar{x}_j)^2}{2(a+b(t-1))}+\frac{1}{2(a-b)}\sum_{j=1}^{n}(\underset{\sim}{y}_j-\underset{\sim}{x}_j\beta)'\underset{\sim}{L}'\underset{\sim}{L}(\underset{\sim}{y}_j-\underset{\sim}{x}_j\beta). \tag{15}$$

Differentiating this with respect to α and equating to zero yelds

$$\hat{\alpha}=\bar{y}-\hat{\beta}\bar{x}, \tag{16}$$

where '^' indicates the ML estimator.

Differentiating (15) with respect to a and b leads to the equations

$$\frac{t}{n}\sum_{j=1}^{n}(\bar{y}_j-\hat{\alpha}-\hat{\beta}\bar{x}_j)^2=\hat{a}+\hat{b}(t-1) \text{ and} \tag{17}$$

$$\sum_{j=1}^{n}(\underset{\sim}{y}_j-\underset{\sim}{x}_j\hat{\beta})'\underset{\sim}{L}'\underset{\sim}{L}(\underset{\sim}{y}_j-\underset{\sim}{x}_j\hat{\beta})=n(t-1)(\hat{a}-\hat{b}), \tag{18}$$

while the same procedure for β yelds

$$\frac{\sum_{j=1}^{n}\underset{\sim}{x}_j'\underset{\sim}{L}'\underset{\sim}{L}\underset{\sim}{x}_j\hat{\beta}-\sum_{j=1}^{n}\underset{\sim}{x}_j'\underset{\sim}{L}'\underset{\sim}{L}\underset{\sim}{y}_j}{\sum_{j=1}^{n}\bar{x}_j(\bar{y}_j-\hat{\alpha}-\hat{\beta}\bar{x}_j)}=\frac{(\hat{a}-\hat{b})}{\hat{a}+\hat{b}(t-1)}. \tag{19}$$

By in turn putting (16) into (17) and (18) and then (17) and (18) into (19) a cubic equation in $\hat{\beta}$ is obtained, which can be simplyfied to

$$\hat{\beta}^3+P\hat{\beta}^2+Q\hat{\beta}+R=0, \text{ where} \tag{20}$$

$$P=-\left\{\frac{(2t-1)}{t}\frac{B_{xy}}{B_{xx}}+\frac{(t+1)}{t}\frac{W_{xy}}{W_{xx}}\right\},\quad Q=\frac{W_{yy}}{tW_{xx}}+2\frac{B_{xy}W_{xy}}{B_{xx}W_{xx}}+\frac{(t-1)}{t}\frac{B_{yy}}{B_{xx}}$$

$$\text{and } R=-\left\{\frac{(t-1)}{t}\frac{B_{yy}W_{xy}}{B_{xx}W_{xx}}+\frac{B_{xy}W_{yy}}{tB_{xx}W_{xx}}\right\}.$$

The solution of $\hat{\beta}$ into (16) gives $\hat{\alpha}$. The ML estimators of a and b are then obtained from the expressions

$$\hat{b}=\frac{\hat{\Sigma}_b}{n}-\frac{\hat{\Sigma}_w}{nt(t-1)} \quad \text{and} \quad \hat{a}=\frac{t}{n}\hat{\Sigma}_b-(t-1)\hat{b}, \tag{21}$$

where $\hat{\Sigma}_b = n(B_{yy}+\hat{\beta}^2B_{xx}-2\hat{\beta}B_{xy})$ and $\hat{\Sigma}_w=nt(W_{yy}+\hat{\beta}^2W_{xx}-2\hat{\beta}W_{xy})$.

The ML approach thus rests upon the solution of $\hat{\beta}$ in (20). Put $F=(3Q-P^2)/3$ and $G=(2P^3-9PQ+27R)/27$ and consider $D=F^3/27 + G^2/4$. If $D>0$ it is well known that (20) has one real solution, but for $D\leq 0$ there may be two or more unequal real roots in which case these have to be inserted into the likelihood function in order to find the ML estimator.

From (20) it is seen that $\hat{\beta}$ is a non-linear function of the jointly sufficient LS-statistics $\tilde{\beta}_b$, $\tilde{\beta}_w$, $\tilde{\Sigma}_b$ and $\tilde{\Sigma}_w$.

As $n\to\infty$ the vector of estimators $(\hat{\alpha}, \hat{\beta}, \hat{a}, \hat{b})'$ tends to a normal distribution with mean vector $(\alpha, \beta, a, b)'$ and dispersion matrix with the following non-zero elements, obtained from the 2:nd partial derivatives of $-\log L$ (cf. Kendall and Stuart (1961), chap. 18):

$$\begin{aligned}
V(\hat{\beta}) &= \frac{1}{nt}\left\{\frac{B_{xx}}{a+b(t-1)} + \frac{W_{xx}}{a-b}\right\}^{-1} = -\frac{1}{\bar{x}}\mathrm{Cov}(\hat{\alpha},\hat{\beta}),\\
V(\hat{\alpha}) &= \frac{1}{nt}(a+b(t-1)) + \bar{x}^2V(\hat{\beta}),\\
V(\hat{b}) &= \frac{2}{nt^2}\left(\frac{(a-b)^2}{t-1} + (a+b(t-1))^2\right),\\
V(\hat{a}) &= \frac{2}{nt^2}((t-1)(a-b)^2 + (a+b(t-1))^2),\\
\mathrm{Cov}(\hat{a},\hat{b}) &= \frac{2}{nt}((a-b)^2 - (a+b(t-1))^2).
\end{aligned} \tag{22}$$

From these expressions the variances of the estimators of the variance components σ_α^2 and σ_u^2 can be calculated.

Notice that the asymptotic variance of $\hat{\beta}$ in (22) is the same as the variance of the best linear combination of $\tilde{\beta}_b$ and $\tilde{\beta}_w$ when a and b are known.

To study some properties of $\hat{\beta}$ in finite samples a simulation study was performed in which 400 000 simulations were made for each choice of n=10 and 100, t=2 and 10, b=0.1 and 0.9, a=1 and β=0. The result of the study is presented in Table 1. The bias was very small and the observed st.d. agreed well with the asymptotic theoretical st.d., obtained from (22). When the latter was estimated by inserting the ML estimates for the unknown parameters, somewhat lower values were obtained.

				St.d.		
b	n	t	Bias	Obs	As	EAs
0.1	10	2	0.0001	0.169	0.157	0.138
"	"	10	0.0001	0.081	0.078	0.073
"	100	2	0.0001	0.050	0.050	0.049
"	"	10	0.0001	0.025	0.025	0.025
0.9	10	2	0.0001	0.070	0.069	0.063
"	"	10	0.0000	0.031	0.031	0.031
"	100	2	0.0000	0.022	0.022	0.022
"	"	10	0.0000	0.010	0.010	0.010

Table 1. Bias and precision of the ML estimator $\hat{\beta}$ expressed in standard deviation (st.d.). Obs= Observed from simulations, As= According to asymptotic theory and EAs= Estimated from simulations according to asymptotic theory.

4. COMPARISON BETWEEN THE ESTIMATORS

A comparison between the proposed estimators leads to somewhat different conclusions than were reached in the simulation study by Maddala and Mount (1973) mentioned in the introduction.

The relative efficiency of two estimators $\tilde{\beta}_1$ and $\tilde{\beta}_2$ can be expressed by the ratio between their variances, $V(\tilde{\beta}_1)/V(\tilde{\beta}_2)$. Put $K=B_{xx}/S_{xx}$ and $\rho=b/a$, the latter being the correlation between two observations y_{ij} and $y_{i'j}$. Then the following asymptotic results are obtained from sect. 3:

$$\begin{aligned} V(\hat{\beta})/V(\tilde{\beta}_b) &= \{1+\frac{(1-K)}{K}\frac{(1+\rho(t-1))}{(1-\rho)}\}^{-1}, \\ V(\hat{\beta})/V(\tilde{\beta}_w) &= 1-V(\hat{\beta})/V(\tilde{\beta}_b), \\ V(\hat{\beta})/V(\tilde{\beta}_{LS}) &= \{1+\frac{\rho^2 t^2 K(1-K)}{(1-\rho)(1+\rho(t-1))}\}^{-1}. \end{aligned} \tag{23}$$

These three asymptotic efficiencies are plotted in Figure 1 as functions of K for some values of t and ρ.
The within-group estimator $\tilde{\beta}_w$ is more efficient than the between-group estimator $\tilde{\beta}_b$ if $\rho>(2K-1)\{K+(1-K)(t-1)\}^{-1}$. The asymptotic relative efficiency of the LS estimator $\tilde{\beta}_{LS}$ has a minimum for K=1/2 and decreases to zero as $\rho\to 1$ or $t\to\infty$. When $\rho>\{1+t(1-K)\}^{-1}$ the efficiency of $\tilde{\beta}_{LS}$ is in fact smaller than that of $\tilde{\beta}_w$, in which case nothing is gained by also considering between-group variability.

The results suggest that in large samples, say n>100 or n>10 and t>10 (cf. Table 1), much can be gained in precision by using the ML estimator.

In small samples the variance of the ML estimator can be larger than the asymptotic expression in (22) (cf. Table 1) and even larger than the variance of some of the LS estimators. Consider e.g. the case $a=1$, $b=0.1$, $n=10$, $t=2$ and $B_{xx}=1=W_{xx}$. Since $K=B_{xx}/S_{xx}$ the asymptotic relative efficiency of the ordinary LS estimator has a minimum. Yet, in this small-sample case, the LS estimator has smaller st.d. than the ML estimator, 0.158 compared to 0.169.

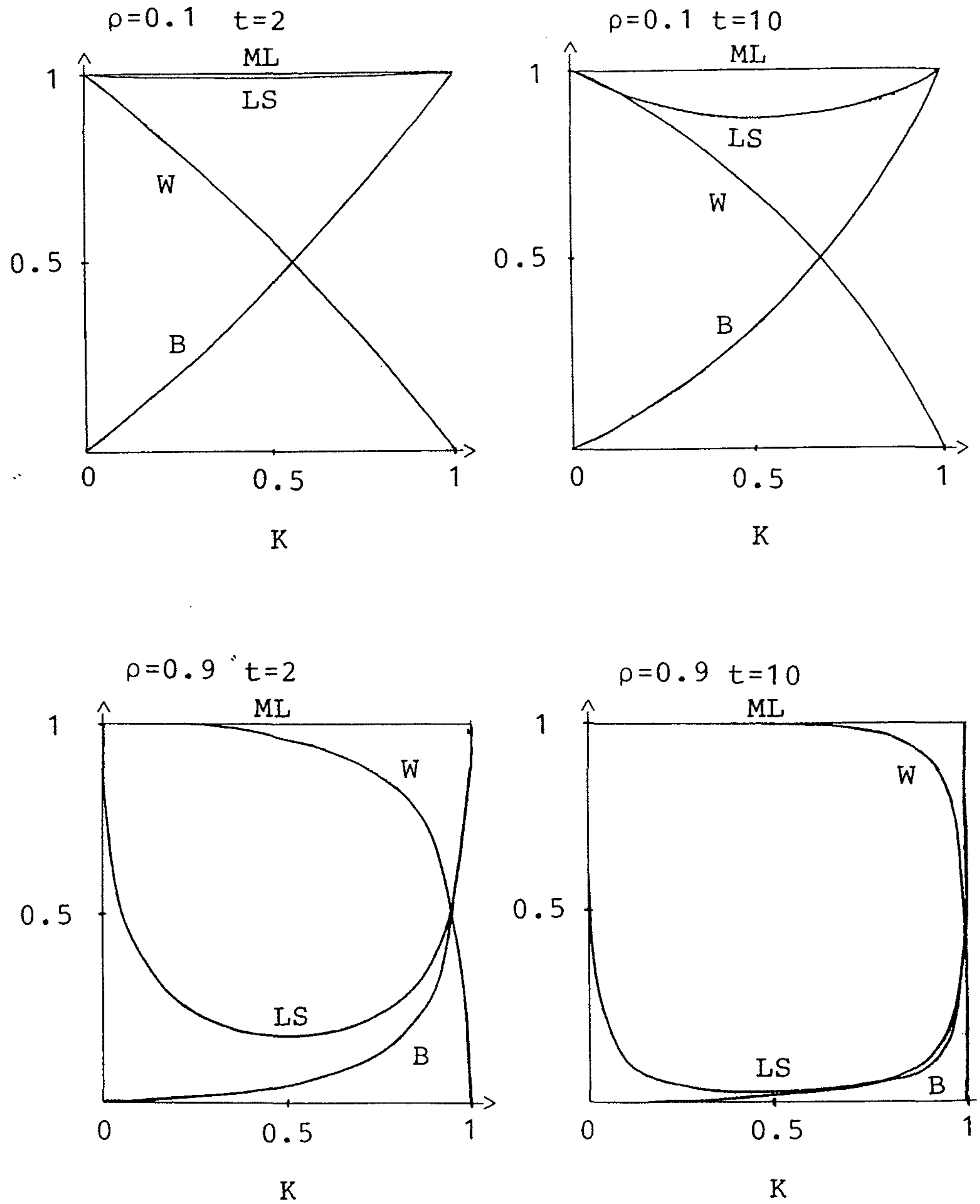

Figure 1. Relative efficiencies of some β-estimators plotted versus $K=B_{xx}/S_{xx}$ for two values of $\rho=b/a$ and t. ML: ML estimator, LS: Ordinary LS estimator, B: Between-group estimator and W: Within-group estimator. Notice how the efficiency of the LS estimator decreases with increasing ρ and t. The between-group estimator is never more efficient than the LS estimator, but it may be more efficient than the within-group estimator when K is large.

5. TESTS AND CONFIDENCE INTERVALS FOR β

Tests and confidence intervals for β can be based on the four estimators in sect. 4. Here, the performance of the following statistics will be compared: $T_b=(\tilde{\beta}_b-\beta)/(\tilde{V}(\tilde{\beta}_b))^{\frac{1}{2}}$, $T_w=(\tilde{\beta}_w-\beta)/(\tilde{V}(\tilde{\beta}_w))^{\frac{1}{2}}$, $T_{LS}=(\tilde{\beta}_{LS}-\beta)/(\tilde{V}(\tilde{\beta}_{LS}))^{\frac{1}{2}}$ and $T_{ML}=(\hat{\beta}-\beta)/(\hat{V}(\hat{\beta}))^{\frac{1}{2}}$, where the estimated variances are given in (13) and (22). In the latter case with the estimators inserted for the parameters.

The distributions of T_b and T_w are simple. Tests of $H_0:\beta=\beta_0$ and confidence intervals for β can be based on the fact that T_b and T_w have Student-T distributions with degrees of freedom (df) n-2 and n(t-1)-1, respectively. The non-centrality parameters which are needed for calculating the powers of the tests are $\delta_b=(\beta-\beta_0)\ V(\tilde{\beta}_b)^{\frac{1}{2}}$ for T_b and $\delta_w=(\beta-\beta_0)V(\tilde{\beta}_w)^{\frac{1}{2}}$ for T_w.

The distribution of T_{LS} is more complicated. From (8), (10) and (13) it is seen that the denominator of T_{LS} consists of a linear combination of chi-square variables. T_{LS} may thus be considered to be approximately Student-T distributed with df in the range n-2 to nt-3 (cf. Walsh (1947)). The distribution of T_{ML} is even more cumbersome.

Since T_{LS} and T_{ML} has asymptotic standard normal distributions it is of interest to study the rate at which the convergence to the asymptotic distribution takes place. To this end 400 000 simulated values of T_{LS} and T_{ML} were generated for each choice of the parameters $\beta=0$, a=1, b=0.1 and 0.9, $B_{xx}/S_{xx}=0.1$, 0.3, 0.5, 0.7 and 0.9, n=10, 100 and 200 and finally, t=2 and 10. The comparison with the standard normal distribution was restricted to the 95%- and 99% percentiles.

For symmetry reasons there was no need to consider the 5%- and 1% percentiles.

The agreement was found to be uneffected by the absolute magnitudes of B_{xx} and W_{xx} in the range 1 to 10000, but dependent on the ratio $K=B_{xx}/S_{xx}$. The largest discrepancy was observed for K=1/2. The 95%- and the 99% percentiles in this least favourable case are presented in Table 2. It is seen that the approach to normality with increasing n is somewhat faster for T_{LS} than for T_{ML}. Table 2 suggests that T_{ML} can be treated as a standard normal variable when n>100 and t>10 or n>200 and t>2.

			T_{ML}		T_{LS}	
b	n	t	95%	99%	95%	99%
0.1	10	2	2.11	3.23	1.75	2.58
"	"	10	1.86	2.75	1.73	2.54
"	100	2	1.68	2.38	1.65	2.34
"	"	10	1.66	2.34	1.65	2.34
"	200	2	1.65	2.34	1.65	2.34
"	"	10	1.65	2.33	1.65	2.33
0.9	10	2	1.95	2.98	1.85	2.88
"	"	10	1.68	2.40	1.84	2.82
"	100	2	1.67	2.36	1.66	2.36
"	"	10	1.65	2.34	1.65	2.34
"	200	2	1.65	2.34	1.65	2.33
"	"	10	1.65	2.33	1.65	2.33
Standard normal:			1.645	2.33	1.645	2.33

Table 2. 95- and 99% percentiles of the distributions of the statistics T_{ML} and T_{LS} for some values of b, n and t when $\beta=0$ and a=1. Each percentile is computed from a distribution based on 400 000 simulations. The percentiles are to be compared with those of the standard normal distribution.

The powers of two-sided tests of the hypothesis $\beta=0$ at the 5% significance level based on the four statistics were compared. As expected from the correspondence between efficiency of estimators and tests (cf. Kendall and Stuart (1961), chap. 25), similar conclusions about the powers could be drawn as for the estimators in sect. 4. Figure 2 shows some examples of power curves. In large samples the powers of the ML statistic always dominated those of the other statistics. The within-group statistic may however be a good alternative when the observations are highly correlated within groups.

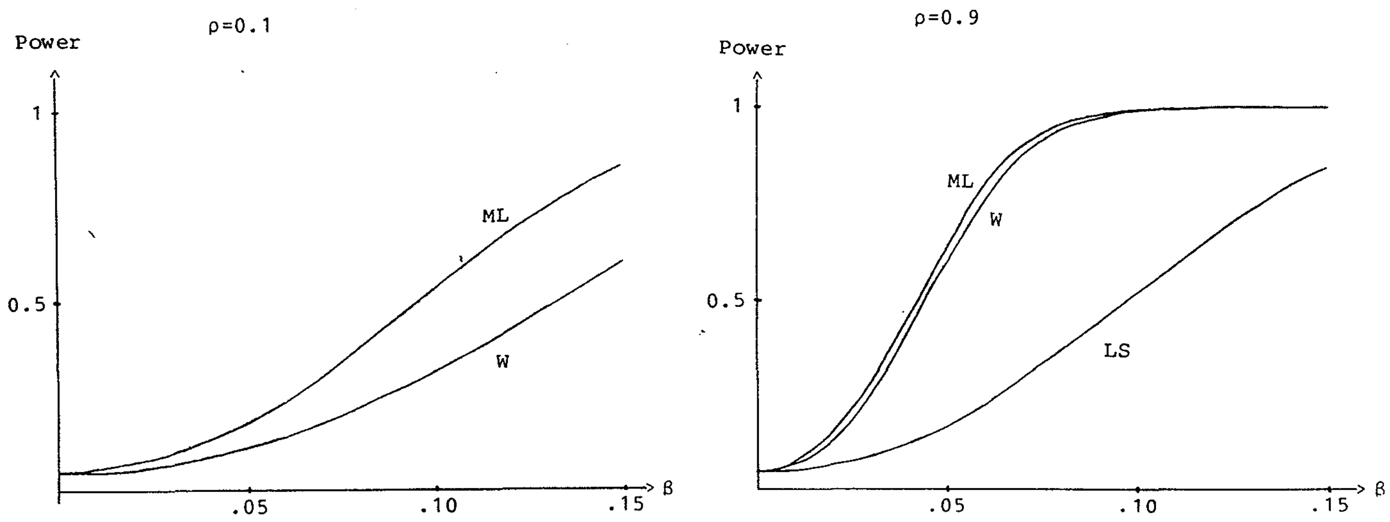

Figure 2. Positive parts of power curves for two-sided tests of the hypothesis $\beta=0$ at the 5% significance level when $\rho=b/a=0.1$ and 0.9 while $K=B_{xx}/S_{xx}=1/2$. When $\rho=0.1$ the powers of the ML statistic (ML) and the within-group statistic (W) are slightly larger than the powers of the LS statistic (LS) and the between-group statistic, respectively and therefore the latter are not shown. When $\rho=0.9$ the power of the between-group statistic is not shown because it is very small and is of less practical interest.

6. AN EXAMPLE

A large number of diabetic patients were screened at Sahlgren's Hospital in Gothenburg during 1982-88. (Details about the patient data set can be communicated by the author or by Dr H. Kalm, Dept. of Ophthalmology, Sahlgren's Hospital, S-413 45 Gothenburg, Sweden). To study whether there was an over-all reduction in HbA_{1c} (glycosulated haemoglobin) among the participants in the screening, a sample of 461 patients with exactly two visits at the hospital was selected. Due to the large intra-patient variability of the HbA_{1c} measurements, a mean of six values was calculated for each patient at each visit. With the present notations y_{ij} represent the mean HbA_{1c} level of the j:th patient, j=1...461, obtained at the times $x_{1j}=0$ and x_{2j}=time after first visit (years).

Since means of HbA_{1c} values were considered it may be reasonable to assume normally distributed y_{ij}'s, as in (1). It remains to check whether the ECR model with constant slopes and random intercepts is valid, or if both slopes and intercepts should be treated as random as in the RCR model. In the latter case the dispersion matrix of $\underset{\sim}{Y}_j$ in (1) is a quadratic form in $(\underset{\sim}{1},\underset{\sim}{X}_j)$ (cf. Swamy (1970), chap. 4.3) and it follows that the variance $V(y_{2j}-y_{1j})$ increases quadratically with x_{2j} in the RCR model while the variance remains constant in the ECR model. The following estimates were obtained:

Time after first visit	Sample size	Estimated dispersion matrix of $(y_{1j},y_{2j})'$	Estimate of $V(y_{2j}-y_{1j})$
$x_{2j}=1$	216	$\begin{pmatrix} 2.6 & 1.5 \\ 1.6 & 2.5 \end{pmatrix}$	1.9
$x_{2j}=2$	172	$\begin{pmatrix} 2.5 & 1.5 \\ 1.5 & 2.1 \end{pmatrix}$	1.2
$x_{2j}=3$	58	$\begin{pmatrix} 2.1 & 1.6 \\ 1.6 & 3.4 \end{pmatrix}$	2.1
$x_{2j}\geq 4$	15	Not computed due to small samples	

The roughly constant elements of the dispersion matrices and the absence of a quadratic increase in $V(y_{2j}-y_{1j})$ with x_{2j} indicate that the ECR model is adequate.

The following summary statistics were obtained from the data:

$n=461$, $t=2$, $\bar{x}=0.86$, $\bar{y}=8.29$,

$W_{xx}=0.8988$, $W_{xy}=-0.1007$, $W_{yy}=0.4314$,

$B_{xx}=0.1596$, $B_{xy}=0.0486$, $B_{yy}=1.9482$.

In this case $B_{xx}/S_{xx}=0.15$ and $\rho=b/a$ is about 0.6 as estimated from the dispersion matrices. According to the results in sect. 4 the within-group estimator $\tilde{\beta}_w$ can be expected to be nearly as efficient as the ML estimator $\hat{\beta}$. The LS estimator $\tilde{\beta}_{LS}$ shall be less efficient while the between-group estimator $\tilde{\beta}_b$ shall be very poor.

Calculations give the following estimates of β with estimated SE's within parentheses:

$\tilde{\beta}_b=0.304(0.162)$, $\tilde{\beta}_w=-0.112(0.032)$, $\tilde{\beta}_{LS}=-0.049(0.037)$, $\hat{\beta}=-0.097(0.032)$.

Estimates of a and b are: $\tilde{a}=2.36$, $\tilde{b}=1.52$ from LS approach and $\hat{a}=2.39$, $\hat{b}=1.55$ from ML approach.

The hypothesis $\beta=0$ is strongly rejected by two-sided tests based on the statistics T_w and T_{ML} whereas the statistics T_{LS} and T_b fail to detect significant departures from the hypothesis at the 5% level.

To conclude, there has been a significant decrease of over-all mean HbA_{1c} level during the screening period of about 0.1 unit per year.

7. DISCUSSION

LS theory has played a dominant role in estimating the parameters of the RCR model having a vector of random regression coefficients. If all measurements are taken at the same times for all n units things become easy. Under fairly general assumptions the LS approach leads to minimum variance unbiased estimators with simple distributions (C.R. Rao (1965). But, during non-experimental conditions it is rarely feasible to collect data at the same times for all sample units, e.g. firms or patients. Then it becomes difficult to find optimal parameter estimates and the distributional problems become severe (cf. Swamy (1970) chap. 4). In such cases it may be fruitful to check if the RCR model can be reduced to an ECR model, in which only the intercepts vary randomly between the sample units, as was done in the example of sect. 6.

As has been demonstrated, the choice of estimator of the slope parameter in the ECR model is indeed not a question of less account. With large samples there should be just one candidate, the ML estimator. Exceptionally, the estimation equation (20) may fail to produce an ML estimate due to boundary solutions when $\rho=b/a$ is 0 or 1 (cf. Maddala (1971)). This is of less practical importance since the probability of boundary solutions tends to zero as t or n tends to infinity. In the simulations in sect. 3.2 the ML approach failed with a frequency of about 1/1000 when $t=2$, $n=10$ and $\rho=0.1$. In large samples, say $t>10$ and $n>100$ or $t>2$ and $n>200$, the statistic T_{ML} based on the ML estimator behaves like a standard normal variable, at least at the extreme tails. This makes the ML approach easy to use.

On the other hand, in small samples there are hardly any reasons for using the ML approach. If tests or confidence statements about the slope parameter are required the use of the LS estimator leads to distributional problems. One possibility is to use the between- or the within group estimator $\tilde{\beta}_b$ and $\tilde{\beta}_w$, respectively and the corresponding statistics T_b and T_w. As shown in Figure 1 the efficiencies of the latter estimators are critically dependent on $K=B_{xx}/S_{xx}$. As K approaches 1 $\tilde{\beta}_b$ becomes more efficient than $\tilde{\beta}_w$ as far as $\rho=b/a$ is small. But, it should be kept in mind that $\tilde{\beta}_b$ can be very poor, as was demonstrated in the example of sect. 6.

In practice it may be useful to first compute a confidence interval for ρ which, together with the information about K, can be used as a guidline for the choice between $\tilde{\beta}_b$ and $\tilde{\beta}_w$. From the results in sect. 3.1 it follows that a 95% confidence interval for ρ is given by

$$\frac{T-F_{.975}}{T+(t-1)F_{.975}} < \rho < \frac{T-F_{.025}}{T+(t-1)F_{.025}}, \text{ with } T=\frac{t(n(t-1)-1)}{n-2}\frac{(B_{yy}-\tilde{\beta}_b^2 B_{xx})}{(W_{yy}-\tilde{\beta}_w^2 W_{xx})}$$

and where F_p is the p-percentile of the $F(n(t-1)-1,n-2)$-distribution.

REFERENCES

Balestra, P., and Nerlove, M. (1966), "Pooling Cross Section and Time Series Data in Estimation of a Dynamic Model: The Demand for Natural Gas," Econometrica, 34, 585-612.

Hsiao, C. (1986), Analysis of Panel Data, Cambridge: Cambridge University Press.

Kendall, M., Stuart, A., and Ord, J.K. (1983), The Advanced Theory of Statistics (Vol. 2, 4th ed.), London: Charles Griffin & Company Ltd.

Maddala, G.S. (1971), "The Use of Variance Components in Pooling Cross Section and Time Series Data," Econometrica, 39, 341-358.

Maddala, G.S., and Mount, T.D. (1973), "A Comparative Study of Alternative Estimators for Variance Component Models Used in Econometric Applications," Journal of the American Statistical Association, 68, 324-328.

Rao, C.R. (1965), "The Theory of Least Squares When the Parameters Are Stochastic and Its Application to the Analysis of Growth Curves," Biometrika, 52, 447-458.

Swamy, P.A.V.B. (1971), Statistical Inference in Random Coefficient Regression Models, Berlin: Springer-Verlag.

Taylor, W.E. (1980), "Small Sample Considerations in Estimation from Panel Data," Journal of Econometrics, 13 203-223.

Welch, B.L. (1947), "The Generalization of 'Student's' Problem When Several Different Population Variances Are Involved," Biometrika, 34, 28-35.

1990:1	Holm, S.	Abstract bootstrap confidence intervals in linear models.
1990:2	Holm, S. & Dahlbom, U	On tests of equivalence
1991:1	Olofsson, Jonny	On some prediction methods for categorical data
1991:2	Jonsson, Robert	On the problem of optimal inference in the simple error component model for panel data