On the problem of optimal inference for time heterogeneous data with error components regression structure

Robert Jonsson

University of Goteborg, Sweden

Summary Time heterogeneity, or the fact that subjects are measured at different times, occurs frequently in non-experimental situations. For time heterogeneous data having error components regression structure it is demonstrated that under customary normality assumptions there is no estimation method based on Maximum Likelihood, Least Squares, Within-subject or Between-subject comparisons that is generally superior when estimating the slope of the regression line. However, in some situations it is possible to give guidelines for the choice of an optimal procedure. These are expressed in terms of the variability of the times for the measurements and also of the inter-subject correlation. The results are demonstrated on data from a longitudinal medical study.

Keywords: Error components regression; Time heterogeneity; Optimal estimators; Efficiency; Test power

JEL: C10, C13, C23, C40, C 80

1. Introduction

Consider a longitudinal study where T repeated measurements are made on each of n subjects. Let y_{ij} be the i:th measurement on the j:th subject (i=1…T, j=1…n) and assume that the vector $\mathbf{y}_j = (y_{1j} \ldots y_{Tj})'$ for each j can be written

$$\mathbf{y}_j = \left[\mathbf{1} \mid \mathbf{X}_j\right]\begin{pmatrix} a_j \\ \hline \mathbf{b}_j \end{pmatrix} + \mathbf{u}_j \tag{1}$$

Here, $\mathbf{1} = (1 \ldots 1)'$, $\mathbf{X}_j = \left(\mathbf{x}_{j1} \mid \ldots \mid \mathbf{x}_{jr} \mid \ldots \mid \mathbf{x}_{jp}\right)$ is a design matrix with vectors $\mathbf{x}_{jr} = (x_{1jr} \ldots x_{Tjr})'$, a_j is an intercept, $\mathbf{b}_j = (b_1 \ldots b_p)'$ is a vector of slopes and $\mathbf{u}_j$ is a vector of errors which is assumed to be normally distributed with mean $\mathbf{0}$ and dispersion matrix $\sigma^2\mathbf{I}$. In longitudinal studies the elements of $\mathbf{X}_j$ are times or functions of times, but may also be covariates. A common situation is when measurements are obtained from untreated patients (baseline measurements), then a treatment is given to the patients and new measurements are obtained from the same patients. In this case one may set the times of the baseline measurements, say x_{1j1}, equal to zero, and the times of the subsequent observations, say x_{ij1} for $i = 2 \ldots T$, equal to the times that has elapsed since the baseline measurements were taken. A simple example of this situation is given in Section 5 of this paper.

Different assumptions about a_j and $\mathbf{b}_j$ in (1) give rise to various models. When a_j and $\mathbf{b}_j$ are equal to α and $\boldsymbol{\beta}$, respectively, with probability one, then the classical Gauss-Markov model is obtained. A generalisation of the latter is obtained by allowing the intercepts to vary between the subjects in a wider population. Such models have been called Error Components (EC) models (Swamy, 1971) or random intercept models (Diggle *et. al.*, 1994). If the a_j's are independent of the $\mathbf{u}_j$'s and vary according to a normal distribution with mean α and variance σ_a^2, it follows that the unconditional distribution of $\mathbf{y}_j$ in (1) is normal with mean $\mathbf{1}\alpha + \mathbf{X}_j\boldsymbol{\beta}$ and dispersion matrix $\sigma_a^2\mathbf{1}\mathbf{1}' + \sigma^2\mathbf{I}$. A further generalisation is to allow also the slopes to vary, but such Random Coefficient models will not be considered here.

An example of an EC model with two variables being functions of the times of the measurements can be constructed from Wood's function (Lennox *et al.* (1992)). Here the response at time *t* is $f(t) = A \cdot t^B \exp(-Ct)$, where *A* determines the level of the peak value (PV), while B and C are constants that determine the time to peak (TP). When the variation between the individual PV's is large and the corresponding variation between the TP's can be ignored, then the linearised responses should agree with the following special case of the EC model in (1), $y_{ij} = a_j + \beta_1 x_{ij1} + \beta_2 x_{ij2} + u_{ij}$, where $a_j = \ln(A_j)$, $\beta_1 = -C$, $\beta_2 = B$, $x_{ij1} = t_{ij}$, $x_{ij2} = \ln(t_{ij})$.

Optimal estimators of the parameters in EC and Random Coefficient models under normality assumptions are well known for the case when the design matrices $\mathbf{X}_j$ are the same for all subjects (Rao, 1965). This is the case when the design matrices consist of functions of the times of the measurements and all subjects are measured at the same times. Such data has been termed *balanced* by some (Ware, 1985), while others further require that time intervals between pairs of corresponding observations are the same, for the data to be called balanced (Forcina, 1992). In many non-experimental situations, the design matrices vary between the subjects. For instance, drugs are administrated to patients at a clinic and, for various practical reasons, the effects of the drug are judged after different treatment times. In fact, the situation when the time elements of the design matrix are determined by current needs and resources, rather than by purely statistical considerations, seems to be frequent. Here such data will be termed *time-heterogeneous*, rather than "unbalanced" to avoid confusion. In the latter situation the estimation procedure is more cumbersome and often requires the use of iterative methods (Laird and Ware, 1982). By considering the case when all subjects are measured at T times, as in (1), the expressions for the estimators are simplified (cf. Section 2).

In this paper, some estimators of the slope parameter $\boldsymbol{\beta}$ of the EC model are compared. These estimators are presented in Section 2. Although EC models have been used extensively, very few comparative studies of the merits of different estimators have been published. The variance components σ_a^2 and σ^2 are often of less interest in themselves, but estimates of the latter are crucial for the estimation of $\boldsymbol{\beta}$. It has been shown that more efficient estimators of the variance components need not result in more efficient estimators of $\boldsymbol{\beta}$ (Taylor, 1980). In a frequently cited simulation study, Maddala and Mount (1973) compared bias and mean squared error of 11 estimators of the single slope parameter β when α was set to zero. It was concluded that 'there is nothing much to chose among these estimators', a statement which will be strongly contradicted by the results in Section 3 of this paper where some expressions for the asymptotic efficiencies of some estimators are derived. Section 4 deals with tests and confidence intervals for components of $\boldsymbol{\beta}$. In Section 5 the results are applied to a longitudinal medical study, while Section 6 contains some concluding remarks.

2. Some $\boldsymbol{\beta}$-estimators

It will be convenient to introduce the following sample moments for i=1…T, j=1…n, and r,s= 1…p: Means,

$$\bar{x}_{jr} = \sum_i x_{ijr}/T,\ \bar{x}_r = \sum_j \bar{x}_{jr}/n,\ \bar{y}_j = \sum_i y_{ij}/T,\ \bar{y} = \sum_j \bar{y}_j/n\ .$$

Sums of square (SS) within subjects,

$$w_{rs} = \sum_i \sum_j (x_{ijr} - \bar{x}_{jr})(x_{ijs} - \bar{x}_{js}),\ w_{ry} = \sum_i \sum_j (x_{ijr} - \bar{x}_{jr})(y_{ij} - \bar{y}_j),\ w_{yy} = \sum_i \sum_j (y_{ij} - \bar{y}_j)^2\ .$$

SS between subjects,

$$b_{rs} = T\sum_j (\bar{x}_{jr} - \bar{x}_r)(\bar{x}_{js} - \bar{x}_s),\ b_{ry} = T\sum_j (\bar{x}_{jr} - \bar{x}_r)(\bar{y}_j - \bar{y}),\ b_{yy} = T\sum_j (\bar{y}_j - \bar{y})^2\ .$$

Here, the two types of SS summarize the total variation within and between subjects. Put $\bar{\mathbf{x}}_j = (\bar{x}_{j1} \ldots \bar{x}_{jp})'$ and $\bar{\mathbf{x}} = \sum_j \bar{\mathbf{x}}_j$. Then the SS's above can be expressed as

$$\mathbf{W}_{xx} = \left(w_{rs}\right) = \sum_j \mathbf{X}_j{}'\mathbf{X}_j - T\sum_j \bar{\mathbf{x}}_j\bar{\mathbf{x}}_j{}',\ \mathbf{B}_{xx} = \left(b_{rs}\right) = T(\sum_j \bar{\mathbf{x}}_j\bar{\mathbf{x}}_j{}' - n\bar{\mathbf{x}}\bar{\mathbf{x}}'),\ \text{and}$$

$$\mathbf{T}_{xx} = \mathbf{W}_{xx} + \mathbf{B}_{xx}\text{, the total SS matrix.}$$

$$\mathbf{w}_{xy} = \left(w_{ry}\right) = \sum_j \mathbf{X}_j{}'\mathbf{y}_j - T\sum_j \bar{\mathbf{x}}_j\bar{y}_j,\ \mathbf{b}_{xy} = \left(b_{ry}\right) = T(\sum_j \bar{\mathbf{x}}_j\bar{y}_j - n\bar{\mathbf{x}}\,\bar{y}),\ \text{and}$$

$$\mathbf{t}_{xy} = \mathbf{w}_{xy} + \mathbf{b}_{xy}\text{. Also, put } t_{yy} = w_{yy} + b_{yy}\text{.}$$

By making an orthogonal transformation of $\mathbf{y}_j$, the normal density is decomposed into two independent parts, one containing within-subject observations and one containing between-subject observations. Put $\mathbf{z}_j = \mathbf{M}\mathbf{y}_j$, where (cf. Rao, 1973, p.197)

$$\mathbf{M} = \left[\begin{array}{c} 1/\sqrt{T} \quad \ldots \quad 1/\sqrt{T} \\ \hdashline \mathbf{L} \end{array}\right]$$

is orthogonal, and since $\mathbf{M}'\mathbf{M} = \mathbf{I}$ it follows that

$$\mathbf{L}'\mathbf{L} = \mathbf{I} - \mathbf{1}\mathbf{1}'T^{-1} \qquad (2)$$

The property in (2) will be used below without having to specify the form of the sub-matrix $\mathbf{L}$.

It is easily verified that $\mathbf{z}_j = (\bar{y}_j\sqrt{T} \mid \mathbf{y}_j'\mathbf{L}')'$ is normally distributed with

$$E(\mathbf{z}_j) = \begin{pmatrix} \sqrt{T}(\alpha + \boldsymbol{\beta}'\bar{\mathbf{x}}_j) \\ \hline \mathbf{L}\mathbf{X}_j\boldsymbol{\beta} \end{pmatrix} \text{ and } V(\mathbf{z}_j) = \sigma_a^2 \left[\begin{array}{c|c} T & \mathbf{0}' \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] + \sigma^2\mathbf{I},$$

so the density of $\mathbf{z}_j$ is, apart from constants

$$(\sigma^2 + \sigma_a^2 T)^{-\frac{1}{2}} \exp\left\{-\frac{T(\bar{y}_j - \alpha - \boldsymbol{\beta}'\bar{\mathbf{x}}_j)^2}{2(\sigma^2 + \sigma_a^2 T)}\right\} \cdot (\sigma^2)^{-\frac{(T-1)}{2}} \exp\left\{-\frac{(\mathbf{L}\mathbf{y}_j - \mathbf{L}\mathbf{X}_j\boldsymbol{\beta})'(\mathbf{L}\mathbf{y}_j - \mathbf{L}\mathbf{X}_j\boldsymbol{\beta})}{2\sigma^2}\right\}$$

(3)

In (3),

$$\sum_j (\mathbf{L}\mathbf{y}_j - \mathbf{L}\mathbf{X}_j\boldsymbol{\beta})'(\mathbf{L}\mathbf{y}_j - \mathbf{L}\mathbf{X}_j\boldsymbol{\beta}) = \sum_j \mathbf{y}_j'\mathbf{L}'\mathbf{L}\mathbf{y}_j - 2\mathbf{y}_j'\mathbf{L}'\mathbf{L}\mathbf{X}_j\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}_j'\mathbf{L}'\mathbf{L}\mathbf{X}_j\boldsymbol{\beta} =$$

$= w_{yy} - 2\mathbf{w}_{xy}'\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{W}_{xx}\boldsymbol{\beta}$, which follows by using (2). Notice that the last expression is obtained without having to specify the form of the sub-matrix $\mathbf{L}$.

Estimators of $\boldsymbol{\beta}$ can be constructed by using either between-subject or within-subject information, or by combining the two approaches. Here, four $\boldsymbol{\beta}$-estimators will be considered: (1) $\hat{\boldsymbol{\beta}}_B$ based on between-subject information only, (2) $\hat{\boldsymbol{\beta}}_W$ based on within-subject information only, (3) $\hat{\boldsymbol{\beta}}_{LS}$, the ordinary least squares estimator, and (4) $\hat{\boldsymbol{\beta}}_{ML}$, the

Maximum Likelihood estimator. As will be seen, the latter two estimators make use of both between- and within-subject information.

Between-subject approach

Data now consists of n independent observations $\sqrt{T}\,\overline{y}_j$, each having a density which is proportional to the first factor of (3). The minimum variance unbiased (MVU) estimators are easily seen to be

$$\begin{pmatrix} \hat{\alpha}_B \\ \hline \hat{\boldsymbol{\beta}}_B \end{pmatrix} = \begin{pmatrix} \overline{y} - \hat{\boldsymbol{\beta}}_B{}'\overline{\mathbf{x}} \\ \hline \mathbf{B}_{xx}^{-1}\mathbf{b}_{xy} \end{pmatrix}, \text{ with } V\begin{pmatrix} \hat{\alpha}_B \\ \hline \hat{\boldsymbol{\beta}}_B \end{pmatrix} = (\sigma^2 + \sigma_a^2 T)\left[\begin{array}{c|c} \frac{1}{n} + \overline{\mathbf{x}}'\mathbf{B}_{xx}^{-1}\overline{\mathbf{x}} & -\overline{\mathbf{x}}'\mathbf{B}_{xx}^{-1} \\ \hline -\mathbf{B}_{xx}^{-1}\overline{\mathbf{x}} & \mathbf{B}_{xx}^{-1} \end{array}\right] \tag{4}$$

The residual SS is $SSE_B = T\sum_j (\overline{y}_j - \hat{\alpha}_B - \hat{\boldsymbol{\beta}}_B{}'\overline{\mathbf{x}}_j)^2 = (b_{yy} - \mathbf{b}_{xy}{}'\hat{\boldsymbol{\beta}}_B)$, which is distributed as $(\sigma^2 + \sigma_a^2 T)\cdot\chi^2(n-p-1)$. Thus, $SSE_B \cdot \mathbf{B}_{xx}^{-1}/(n-p-1)$ is unbiased for $V(\hat{\boldsymbol{\beta}}_B)$.

Within-subject approach

Data consists of n independent observations $\mathbf{Ly}_j$, each with a density which is proportional to the second factor of (3). In this case the MVU estimators are given by

$$\hat{\boldsymbol{\beta}}_W = \mathbf{W}_{xx}^{-1}\mathbf{w}_{xy} \text{ with } \mathrm{V}(\hat{\boldsymbol{\beta}}_W) = \sigma^2\mathbf{W}_{xx}^{-1} \tag{5}$$

Another way to obtain $\hat{\boldsymbol{\beta}}_W$ in (5) is to first fit separate least squares planes to the data from each subject, giving the estimators $\hat{\boldsymbol{\beta}}_W^{(j)} = \mathbf{W}_{xx}^{(j)-1}\mathbf{w}_{xy}^{(j)}$, and then to form the best linear combination of the latter. The residual SS is $SSE_W = w_{yy} - \mathbf{w}_{xy}'\hat{\boldsymbol{\beta}}_W$ which is distributed as $\sigma^2\chi^2(n(T-1)-p)$. An unbiased estimator of $V(\hat{\boldsymbol{\beta}}_W)$ is $SSE_W \cdot \mathbf{W}_{xx}^{-1}/(n(T-1)-p)$.

From the decomposition in (3) and from fundamental results in LS theory it follows that $\hat{\boldsymbol{\beta}}_B$, $\hat{\boldsymbol{\beta}}_W$, SSE_B and SSE_W are independent.

OLS approach

By fitting a least squares plane to the complete set of observations ($\mathbf{y}_1 \ldots \mathbf{y}_n$) one gets

$$\begin{pmatrix} \hat{\alpha}_{LS} \\ \hat{\boldsymbol{\beta}}_{LS} \end{pmatrix} = \begin{pmatrix} \bar{y} - \hat{\boldsymbol{\beta}}_{LS}'\bar{\mathbf{x}} \\ \mathbf{T}_{xx}^{-1}\mathbf{t}_{xy} \end{pmatrix}, \text{ with } V(\hat{\boldsymbol{\beta}}_{LS}) = (\sigma^2 + \sigma_a^2 T \cdot \mathbf{T}_{xx}^{-1}\mathbf{B}_{xx})\mathbf{T}_{xx}^{-1} \tag{6}$$

$\hat{\boldsymbol{\beta}}_{LS}$ in (6) can also be expressed as a linear combination of between- and within-subject estimators, $\mathbf{T}_{xx}^{-1}(\mathbf{B}_{xx}\hat{\boldsymbol{\beta}}_B + \mathbf{W}_{xx}\hat{\boldsymbol{\beta}}_W)$.

ML approach

Solutions of the ML-equations in order to estimate $\boldsymbol{\beta}$ in the EC model seems first to have been discussed by Balestra and Nerlove (1966). By putting the derivatives of the log-likelihood equal to zero and solving for the unknown parameters, it follows from simple but tedious arguments that the ML estimators can be obtained in the following way: Put $SSB = b_{yy} - 2\boldsymbol{\beta}'\mathbf{b}_{xy} + \boldsymbol{\beta}'\mathbf{B}_{xx}\boldsymbol{\beta}$ and $SSW = w_{yy} - 2\boldsymbol{\beta}'\mathbf{w}_{xy} + \boldsymbol{\beta}'\mathbf{W}_{xx}\boldsymbol{\beta}$. Then the estimators of $\beta_1 \ldots \beta_p$ are obtained by solving the set of equations

$$(b_{ry} - \sum_{s=1}^{p} \beta_s b_{rs})SSW = (T-1)(w_{ry} - \sum_{s=1}^{p} \beta_s w_{rs})SSB,\ r = 1\ldots p \tag{7}$$

The rest of the parameters are then estimated from $\hat{\alpha}_{ML} = \bar{y} - \hat{\boldsymbol{\beta}}_{ML}{}'\bar{\mathbf{x}},\ \hat{\sigma}^2_{ML} = S\hat{S}W / n(T-1)$ and $\hat{\sigma}^2_a = (S\hat{S}B - S\hat{S}E)/nT(T-1)$, where $S\hat{S}B$ and $S\hat{S}E$ are the expressions for SSB and SSE with $\hat{\boldsymbol{\beta}}_{ML}$ inserted for $\boldsymbol{\beta}$.

When p=1, there is just one component to estimate, and (7) reduces to the cubic equation

$$\beta^3 + P\cdot\beta^2 + Q\cdot\beta + R = 0 \tag{8}$$

where $P = -\left[\frac{(2T-1)}{T}\frac{b_{1y}}{b_{11}} + \frac{(T+1)}{T}\frac{w_{1y}}{w_{11}}\right]$, $Q = \left[\frac{w_{yy}}{Tw_{11}} + 2\frac{b_{1y}w_{1y}}{b_{11}w_{11}} + \frac{(T-1)}{T}\frac{b_{yy}}{b_{11}}\right]$ and

$$R = -\left[\frac{(T-1)}{T}\frac{b_{yy}w_{1y}}{b_{11}w_{11}} + \frac{b_{1y}w_{yy}}{Tb_{11}w_{11}}\right].$$

More generally, when n tends to infinity (T remains fixed) the ML estimator is asymptotically normally distributed with mean $\boldsymbol{\beta}$ and dispersion matrix (cf. Hsiao, 1986, p. 40)

$$\left[\frac{\mathbf{B}_{xx}}{\sigma^2 + \sigma_a^2 T} + \frac{\mathbf{W}_{xx}}{\sigma^2}\right]^{-1} \tag{9}$$

Here the dispersion matrix may be estimated by replacing the unknown variance components by the corresponding ML estimates.

Since the results above only hold asymptotically, a simulation study was performed when p=1 for various value of n, T, the inter-subject correlation

$$\rho = Corr(y_{ij}, y_{i'j}) = \frac{\sigma_a^2}{\sigma_a^2 + \sigma^2}, \text{ for } i \neq i', \tag{10}$$

and also for various values of the ratio K_r, defined below in (11), with r=1. In each simulation, data was generated according to the model in (1) by using normally distributed pseudo-random deviates, and a ML estimate was computed by solving (8). This procedure was repeated 10^5 times for each combination of n=(10,50,100), T=(2,10), ρ=(0.1, 0.5, 0.9) and K_1=(0.1, 0.5, 0.9). The means and variances of the ML estimates were then compared with the true value $\beta = 0$ and with the asymptotic variance in (9), respectively.

It was found that the bias when estimating β with n=10 varied between -.0010 and 0.0014. No further reduction in bias was obtained when n was increased to 50 and 100, and no relation could be seen between the bias and the values of n, T and K_1. The variance agreed well with the asymptotic expression in (9), but only for large values of n or ρ and small values of K_1. In Table 1 in Section 3 the two variances ($V(\hat{\beta}_{ML})$ and $asV(\hat{\beta}_{ML})$) are compared when n=10. Here it is seen that the difference can be large. E.g. for $\rho = 0.5$, T=10 and $K_1 = 0.9$ the asymptotic variance given by (9) was .275, compared to the actual value .338. With increasing n the latter was reduced to .286 (n=50) and .281 (n=100). Thus, while the bias of the ML estimator can be neglected with sample sizes as small as 10, the asymptotic expression for the variance in (9) should be used with caution.

3. Efficiency of $\boldsymbol{\beta}$-estimators

When all $\mathbf{X}_j$'s in (1) are equal, then $\mathbf{B}_{xx} = \mathbf{0}$. In this case it is seen from the expressions for the dispersion matrices in (4)-(6) and (9) that the between-subject approach can not be used, while the three other approaches give identical $\boldsymbol{\beta}$-estimators.

In the sequel it is assumed that the $\mathbf{X}_j$'s vary between the subjects. The efficiency of the estimators considered in Section 2 will be shown to depend on the correlation in (10) and also on the ratio

$$K_r = \frac{b_{rr}}{b_{rr} + w_{rr}} \tag{11}$$

This measure reflects the dispersion pattern of the r:th independent variable. If the latter is the time at which the measurement is made, then K_r=0 if all subjects are measured at the same times. K_r is large when the times do not overlap. Consider the following simple data sets S_1 and S_2, just for the purpose of illustration:

S_1:			S_2:		
i	j	x_{ij1}	i	j	x_{ij1}
1	1	1	1	1	1
2	1	3	2	1	2
1	2	2	1	2	4
2	2	4	2	2	5
1	3	3	1	3	7
2	3	5	2	3	8

The variation between the x-values in set S_1 is quite moderate, in contrast to the corresponding large variation in set S_2. For the sets S_1 and S_2 one gets $K_1 = 0.40$ and $K_1 = 0.96$, respectively.

As will be shown below, the efficiencies of the $\boldsymbol{\beta}$-estimators depend on K_r and ρ. In practice it may therefore be a good advice to compute K_r and to construct a confidence interval for ρ. The latter is easily obtained from the results in Section 2 by noticing that SSE_B / SSE_W is proportional to the ratio of two independent chi square variables. The 95 percent confidence interval for ρ is thus given by

$$\frac{Q - F_{.975}}{Q + (T-1)F_{.975}} < \rho < \frac{Q - F_{.025}}{Q + (T-1)F_{.025}}, \text{ with } Q = \frac{\big(n(T-1) - p\big) \cdot SSE_B}{\big(n - p - 1\big) \cdot SSE_W} \tag{12}$$

and where F_α denotes the α-percentile of the $F\big(n(T-1) - p, n - p - 1\big)$ distribution.

When there are more than one independent variable in the model, the asymptotic efficiency of the $\boldsymbol{\beta}$-estimators will also depend on the correlations between the x-variables. The expressions for the asymptotic efficiency will in the latter case be quite complicated. Due to these complications only the cases $p=1$ (Section 3.1) and p=2 (Section 3.2) are considered here. Finally, since the result in (9) only holds asymptotically, the relative efficiency in small samples is studied in a Monte Carlo study (Section 3.3).

3.1 Asymptotic efficiency when p=1

The single component of the $\boldsymbol{\beta}$-vector is denoted β and for simplicity the index of K_r in (11) is dropped. From Section 2 the following expressions for the asymptotic efficiencies are obtained:

$$e_B = \frac{V(\hat{\beta}_{ML})}{V(\hat{\beta}_B)} = \left\{1 + \frac{(1-K)}{K}\frac{\left(1+\rho(T-1)\right)}{(1-\rho)}\right\}^{-1}$$

$$e_W = \frac{V(\hat{\beta}_{ML})}{V(\hat{\beta}_W)} = 1 - e_B \qquad (13)$$

$$e_{LS} = \frac{V(\hat{\beta}_{ML})}{V(\hat{\beta}_{LS})} = \left\{1 + \frac{\rho^2 T^2 K(1-K)}{(1-\rho)\left(1+\rho(T-1)\right)}\right\}^{-1}$$

These expressions are plotted in Figure 1 as functions of K for some values of T and ρ.

INSERT FIGURE 1 ABOUT HERE

In Figure 1 it is seen that the efficiency of $\hat{\beta}_B$ decreases with increasing ρ, and to a less extent with increasing T. For $\hat{\beta}_W$ one gets the reversed pattern. Notice that $\hat{\beta}_B$ can be more efficient than $\hat{\beta}_W$ when K is large and ρ is small. This result is perhaps primarily of a theoretical interest. In Section 5 it will be demonstrated that inference based on $\hat{\beta}_B$ can be very risky. The asymptotic efficiency of the OLS estimator has a minimum for K=1/2 and tends to zero as $T \to \infty$ or $\rho \to 1$. The efficiency of $\hat{\beta}_{LS}$ is in fact smaller than that of $\hat{\beta}_W$ when $\rho > \{1 + T(1-K)\}^{-1}$, in which case nothing is gained by also taking between-subject information into consideration. However, the OLS estimator is always better than the between-estimator and it is easily shown that $V(\hat{\beta}_{LS}) \le K \cdot V(\hat{\beta}_B)$.

3.2 Asymptotic efficiency when p=2

Now there are two components β_1 and β_2 of the $\boldsymbol{\beta}$-vector. Consider first the loss of variance when estimating β_1 by including two independent variables x_1 and x_2 in the model instead of only x_1. Let $V(\hat{\beta}_1 \big| p=1)$ and $V(\hat{\beta}_1 \big| p=2)$ be the variances of the β_1-estimator when one and two variables are used in the model, respectively, and introduce the notations

$$r_W = \frac{w_{12}}{\sqrt{w_{11}w_{22}}},\ r_B = \frac{b_{12}}{\sqrt{b_{11}b_{22}}} \tag{14}$$

Then the ratio $V(\hat{\beta}_1 \big| p=1)/V(\hat{\beta}_1 \big| p=2)$ is $1-r_B^2$ when using a between-subject approach, $1-r_W^2$ when using a within-subject approach and

$$1-\frac{\left[(1-\rho)r_B\sqrt{K_1K_2}+\left(1+\rho(T-1)\right)r_W\sqrt{(1-K_1)(1-K_2)}\right]^2}{\left(1-\rho+\rho T(1-K_1)\right)\left(1-\rho+\rho T(1-K_2)\right)}$$

using the ML approach. The last expression tends to $1-r_W^2$ when K_1 and K_2 tend to zero and to $1-r_B^2$ when K_1 and K_2 tend to one.

An estimator of β_1 that is obtained by ignoring x_2 will in general be biased. Consider e.g. the between-subject estimator $\hat{\beta}_{1(B)} = b_{1y}/b_{11}$. This has the expectation

$$b_{11}^{-1}\cdot T\sum_j(\bar{x}_{j1}-\bar{x}_1)(\alpha+\beta_1\bar{x}_1+\beta_2\bar{x}_2) = b_{11}^{-1}\cdot(\alpha\cdot 0+\beta_1 b_{11}+\beta_2 b_{12}) = \beta_1+\beta_2(b_{12}/b_{11})$$

In a similar way it is easily shown that the corresponding within-subject estimator $\hat{\beta}_{1(W)} = w_{1y}/w_{11}$ has the expectation $\beta_1+\beta_2(w_{12}/w_{11})$. To study whether the smaller variance when using only x_1 compensates for the loss of bias, consider the MSE-ratio

$$\frac{MSE\left(\hat{\beta}_1 | p=1\right)}{MSE\left(\hat{\beta}_1 | p=2\right)} = \frac{V\left(\hat{\beta}_1 | p=1\right) + \left(bias\left(\hat{\beta}_1 | p=1\right)\right)^2}{V\left(\hat{\beta}_1 | p=2\right)}$$

For $\hat{\beta}_{1(B)}$ this ratio can be written $(1-r_B^2)[1+r_B^2 \cdot b_{22}\beta_2^2/(\sigma^2+\sigma_a^2 T)]$, and it is easily seen that the MSE-ratio is always smaller than 1, provided that $\beta_2^2 < (\sigma^2+\sigma_a^2 T)/b_{22}$. Notice that the latter is the same as requiring that β_2^2 is smaller that the variance of the β_2-estimator that only uses the second independent variable x_2. The same result holds for $\hat{\beta}_{1(W)}$ by replacing r_B^2 by r_W^2, b_{22} by w_{22} and $\sigma^2+\sigma_a^2 T$ by σ^2.

The expressions for the asymptotic efficiency when p=2 are quite complicated, but they can be simplified by using the results in Section *3.1*. Let $e^{(1)}$ and $e^{(2)}$ denote the asymptotic efficiency of an estimator of β_1 and β_2, respectively, which is obtained by using a single independent variable in the model. The latter are given in (13). Then the following expressions are obtained for the asymptotic efficiency of the β_1-estimator with two independent variables:

$$e_B = \frac{V(\hat{\beta}_{1,ML})}{V(\hat{\beta}_{1,B})} = \frac{e_B^{(1)}(1-r_B^2)}{1-\left[r_B\sqrt{e_B^{(1)}e_B^{(2)}} + r_W\sqrt{e_W^{(1)}e_W^{(2)}}\right]}$$

$$e_W = \frac{e_W^{(1)}(1-r_W^2)}{e_B^{(1)}(1-r_B^2)} \cdot e_B \qquad (15)$$

It is not possible to express the asymptotic efficiency of the OLS estimator in this neat way. In (15) one may notice that $e_W^{(1)} > e_B^{(1)}$ does not guarantee that $e_W > e_B$.

If the linearised version of Wood's function given in the introduction, with $x_{ij1} = t_{ij}$ and $x_{ij2} = \ln(t_{ij})$, are used for the data sets S_1 and S_2 presented in the beginning of this section, then $r_W = 0.9524$ and $r_B = 0.9921$ for S_1 while $r_W = 0.9140$ and $r_B = 0.9765$ for S_2. In these cases e_W / e_B will be roughly 6-7 times larger than $e_W^{(1)} / e_B^{(1)}$.

3.3 Efficiency in small samples

Table 1 shows the actual variances of the four estimators when *n*=10 together with the asymptotic variance of the ML estimator given in (9). Here one may notice that the estimation equation (8)

TABLE 1 INSERTED ABOUT HERE

sometimes failed to produce ML estimates and that the failure rate was very low when *T*=10 and ρ=0.9. An interesting pattern is that the variance of the OLS estimator is constantly smaller than that of the ML estimator when ρ is small. In the latter case the actual variance of the ML estimator can be considerably larger than the variance given by the asymptotic expression in (9). The precision of the within-subject and ML estimators are improved as ρ increases and *K* decreases. As a curious fact one may notice that even the between-subject estimator can have smaller variance than the ML estimator in small samples.

4. Tests for β

The four estimators in the preceding sections can be used for constructing tests, as well as confidence intervals, for β. Here the performance of the test statistics

T_B, T_W, T_{LS} and T_{ML} for testing $H_0 : \beta = \beta_0$ will be compared, where each statistic has the form $T = (\hat{\beta} - \beta_0) / SE(\hat{\beta})$ and where $SE(\hat{\beta})$ denotes the square root of the estimated variance of $\hat{\beta}$.

From Section 2 one easily finds that

$$T_B = \frac{(\hat{\beta}_B - \beta_0)}{\left[SSE_B / b_{11}(n-2)\right]^{1/2}} \text{ and } T_W = \frac{(\hat{\beta}_W - \beta_0)}{\left[SSE_W / w_{11}(n(T-1)-1)\right]^{1/2}} \tag{16}$$

have Student's T distributions with $n-2$ and $n(T-1)-1$ degrees of freedom, respectively. The non-centrality parameters needed for power calculations are both of the form $(\beta - \beta_0)/\sqrt{V(\hat{\beta})}$. The distribution of T_{LS} is more complicated since $SE(\hat{\beta}_{LS})$ in the numerator is a linear combination of chi square variables (c.f. Ch.18.8 in Johnson *et al.*, 1994). From Section 2,

$$(SE(\hat{\beta}_{LS}))^2 = \frac{b_{11}}{t_{11}} \frac{SSE_B}{(n-2)} + \frac{w_{11}}{t_{11}} \frac{SSE_W}{(n(T-1)-1)}$$

Following Welch (1947) one may try to approximate the distribution of T_{LS} by the Student's T distribution with degrees of freedom equal to

$$\left[\frac{b_{11}}{t_{11}} \frac{SSE_B}{n-2} + \frac{w_{11}}{t_{11}} \frac{SSE_W}{n(T-1)-1}\right]^2 \left[\left(\frac{b_{11}}{t_{11}} \frac{SSE_B}{n-2}\right)^2 \frac{1}{n} + \left(\frac{w_{11}}{t_{11}} \frac{SSE_W}{n(T-1)-1}\right)^2 \frac{1}{(n(T-1)+1)}\right]^{-1} - 2$$

The distribution of T_{ML}, where the ML estimator is the solution of (8), is far more complicated. However, the ML estimator has an asymptotic normal distribution as $n \to \infty$, while the distribution of the OLS estimator is exactly normal. It therefore follows that both T_{ML} and T_{LS} has standard normal distributions in large samples, since the SE's of both estimators are consistent. It is also to be expected that the rate of convergence is faster for the OLS estimator.

To study the distributions of T_{LS} and T_{ML} a Monte Carlo study was performed. The approach to normality was found to be unaffected by the magnitude of b_{11} and w_{11}, but especially for the OLS estimator the approach to normality was found to be heavily dependent on K_1 in (11). As expected, the slowest rate of convergence to normality for the OLS estimator was obtained for $K_1 = 1/2$ (cf. Figure 1), and for this value some percentiles of the distribution of T_{LS} and T_{ML} are presented in Table 2, together with the corresponding percentiles obtained from the distribution of Welch *T*. As can be seen from the table, the statistic T_{ML} converges slower to

TABLE 2 INSERTED ABOUT HERE

normality than T_{LS} when *T* (the number of times the measurements are made) is small. But, T_{ML} converges faster to normality when *T* and ρ is large. A further conclusion is that there seems to be little to gain by using Welch's adjusted degrees of freedom, unless *n* is at least 100.

Under quite general conditions there is a close connection between the efficiency of estimators and the efficiency of the corresponding test statistics (Stuart and Ord, Ch. 25, 1991). From Figure 2, where some power curves are compared, it is evident that similar conclusions can be drawn here. In large samples the power of the ML statistic always dominates the power of the other statistics. However, the statistic based on within-subject information may be a good alternative when ρ is large.

FIGURE 2 INSERTED ABOUT HERE

5. Application to a longitudinal study

A screening program for diabetic patients has been running since 1982 at Sahlgren's hospital in Gothenburg (Kalm, 1993). To study whether an attempt to decrease the patients level of HbA_{1c} (glycosulated haemoglobin) had been successful, a sample of n=461 patients with exactly T=2 visits at the hospital was selected. The measurements at the first and second visits were obtained from patients before and after, respectively, they had participated in a training program aiming to improve the metabolic control. The training program started immediately after the measurements at the first visit. Due to the large intra-subject variability of the measurements, a mean of 6 HbA_{1c}-values was calculated for each patient at each visit. In terms of the notations in Section 1 y_{ij} represents the mean HbA_{1c}-level of patient j obtained at the times $x_{1j} = 0$ (first visit) and x_{2j}=Time after first visit (in years).

Since the data consists of means it may be reasonable to assume normality for the observations. The next step is to check whether an EC model is appropriate, or if a Random Coefficient model, where also the slopes vary randomly, is more adequate. This may be done formally by performing formal tests (cf. Hsiao, 1986, Ch. 6.2.2.d; Petzold and Jonsson, 2003). Since the latter require that $T \geq 3$, a less formal approach is used here. When the slopes b_j in (1) has a normal distribution with $V(b_j) = \sigma_b^2$ and $Cov(a_j, b_j) = \sigma_{ab}$ it follows that

$$V\begin{pmatrix} y_{1j} \\ y_{2j} \end{pmatrix} = \begin{bmatrix} \sigma^2 + \sigma_a^2 & \sigma_a^2 + \sigma_{ab}x_{2j} \\ \sigma_a^2 + \sigma_{ab}x_{2j} & \sigma^2 + \sigma_a^2 + 2\sigma_{ab}x_{2j} + \sigma_b^2(x_{2j})^2 \end{bmatrix} \text{ and } V(y_{2j} - y_{1j}) = 2\sigma^2 + \sigma_b^2(x_{2j})^2 . \tag{17}$$

If (17) is compared with the corresponding quantities that have been estimated from data in Table 3, it is evident that the EC model, in which $\sigma_b^2 = 0 = \sigma_{ab}$, suffices. Especially the lack

TABLE 3 INSERTED ABOUT HERE

of a monotonous quadratic increase in $V(y_{2j} - y_{1j})$ argues against using the more general RC model.

From the data the following summary statistics were calculated:

$$\begin{aligned} &\bar{y} = 8.29 \\ &b_{11} = 0.1596,\ b_{1y} = 0.0486,\ b_{yy} = 1.9482 \\ &w_{11} = 0.8988,\ w_{1y} = -0.1007,\ w_{yy} = 0.4314 \end{aligned}$$

The ratio K_1 in (11) is 0.15, so the degree of time heterogeneity is quite small. The 95 percent confidence interval for ρ given in (12) is $0.59 < \rho < 0.69$. According to the results in Section 3.1 it is to be expected that in this situation the ML and the within-subject estimators should perform well and that the OLS estimator is less good. The between-subject estimator should be poor. The estimates of β are, with SE in parentheses: $\hat{\beta}_{ML}$ = -.097 (.031), $\hat{\beta}_W$ = -.112 (.032), $\hat{\beta}_{LS}$ = -.049 (.037) and $\hat{\beta}_B$ = +.304 (.162) The estimated variance components are $\hat{\sigma}_a^2 = 1.55$ and $\hat{\sigma}^2 = 0.84$, using the ML approach.

The hypothesis $\beta = 0$ is strongly rejected by two-sided tests using T_{ML} and T_W, whereas the statistics T_{LS} and T_B fail to detect significant departures from the hypothesis at the 5 % level. The conclusion is that the training program has resulted in a weak, but statistically significant decrease of the HbA_{1c}-level. It is worth remarking that the test based on T_B is not far from suggesting a significant increase of the HbA_{1c}-level.

6. Discussion and conclusions

In non-experimental situations time heterogeneity occurs frequently. This heterogeneity, as measured by K_r in (11), can vary between 0 and 1. Values of K_r being as large as 0.90 have been experienced by the author for data consisting of times between examinations of patients with osseointegrated oral implants. The choice of $\boldsymbol{\beta}$-estimators will be important in such cases, although this seems to have been overlooked since the simulation study by Maddala and Mount (1973). It should be noted that the conclusions about the small differences in efficiency between various estimation methods that were made in the latter study were based on only 100 simulations. By repeating these simulation experiments with the same parameter settings (K_r=0.76, T=20, n=25 and ρ=0.002, 0.11, 0.50) it is obvious that at least 10^4-10^5 simulations would have been needed to draw any definite conclusions.

The present paper has shown that there can be large differences in efficiency between various $\boldsymbol{\beta}$-estimators. In samples with large n, inference based on the ML approach is optimal. The estimation equations may exceptionally fail to produce ML estimates due to boundary solutions, but this is perhaps of less practical importance, since in large samples the probability of getting boundary solutions will be very small (Maddala, 1971).

In smaller samples the problem arises whether the sample is large enough for asymptotic results to hold. Although the OLS estimator may be more efficient than the other estimators considered in this paper, the OLS approach is less suitable due to distributional problems if tests and confidence statements are required. In this case the within-subject or sometimes even the between-subject approach may be a good alternative. A guidance for choosing a proper estimator is then to calculate K_r and to construct a confidence interval for correlation coefficient ρ by means of (12).

Acknowledgements

I wish to thank Dr Helle Kalm at the Department of Ophtalmology, Sahlgren's Hospital, Gothenburg for providing me with the data. The author also wants to thank Dr Gunnar Rosenqvist at the Department of Statistics, Swedish School of Economics and Business Administration, Helsinki, as well as an anonymous referee, for valuable comments.

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.

Diggle, P. J., Liang, K-Y. and Zeger, S. L. (1994) *Analysis of Longitudinal Data*. Oxford: Clarendon Press.

Forcina, A. (1992) Modelling balanced longitudinal data: Maximum Likelihood Estimation and analysis of variance. *Biometrics*, **48**, 743-750.

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

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994) *Continuous univariate distributions*, vol 1. Wiley: New York.

Kalm, H. (1993) *Diabetic Retinopathy Screening*. M. D. Thesis, Department of Ophtalmology, Sahlgren's Hospital, Gothenburg University.

Laird, N. and Ware, J. H. (1982) Random effects models for longitudinal data. *Biometrics*, **38**, 963-974.

Lennox, S. D., Goodall, E. A. and Mayne, C. S. (1992) A mathematical model of the lactation curve of a dairy cow to incorporate metabilizable energy intake. *Statistician*, **41**, 285-293.

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. *J. Am. Statist. Ass.*, **68**, 324-328.

Petzold, M. and Jonsson, R. (2003) Maximum Likelihood ratio based small-sample tests for random coefficients in linear regression. *Working Papers in Economics*, **102**, Department of Economics, Gothenburg University.

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.

Rao, C. R. (1973) *Linear Statistical Inference and Its Applications*, 2nd edn.. Wiley: New York.

Stuart, A. and Ord, J. K. (1991) *Kendalls Advanced Theory of Statistics*, vol 2, 5th edn.. London: Hodder and Stoughton.

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. *J. Econometr.*. **13**, 203-223.

Ware, J. H. (1985) Linear models for the analysis of longitudinal studies. *Am. Statistician*, **39**, 95-101.

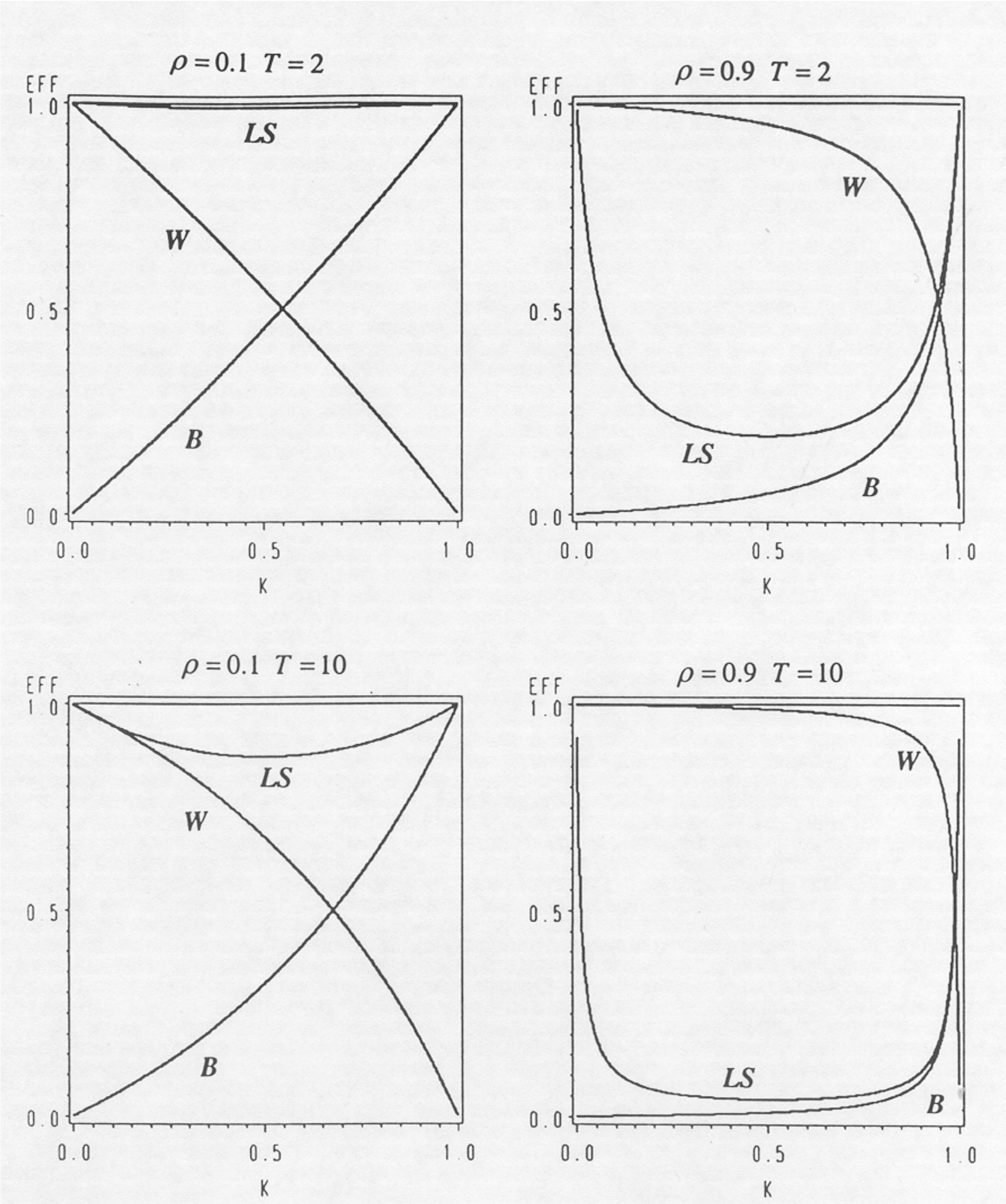
Welch, B. L. (1947) The generalizations of ‘Student’s’ problem when several different population variances are involved. *Biometrika*, **34**, 28-35.

Legends to figures

Figure 1. Asymptotic relative efficiencies of some β-estimators plotted versus $K = b_{11}/(b_{11} + w_{11})$ for two values of T and $\rho = \sigma_a^2/(\sigma_a^2 + \sigma^2)$. ***B***: Between-group estimator, ***W***: Within-group estimator, ***LS***: Ordinary Least Squares estimator.

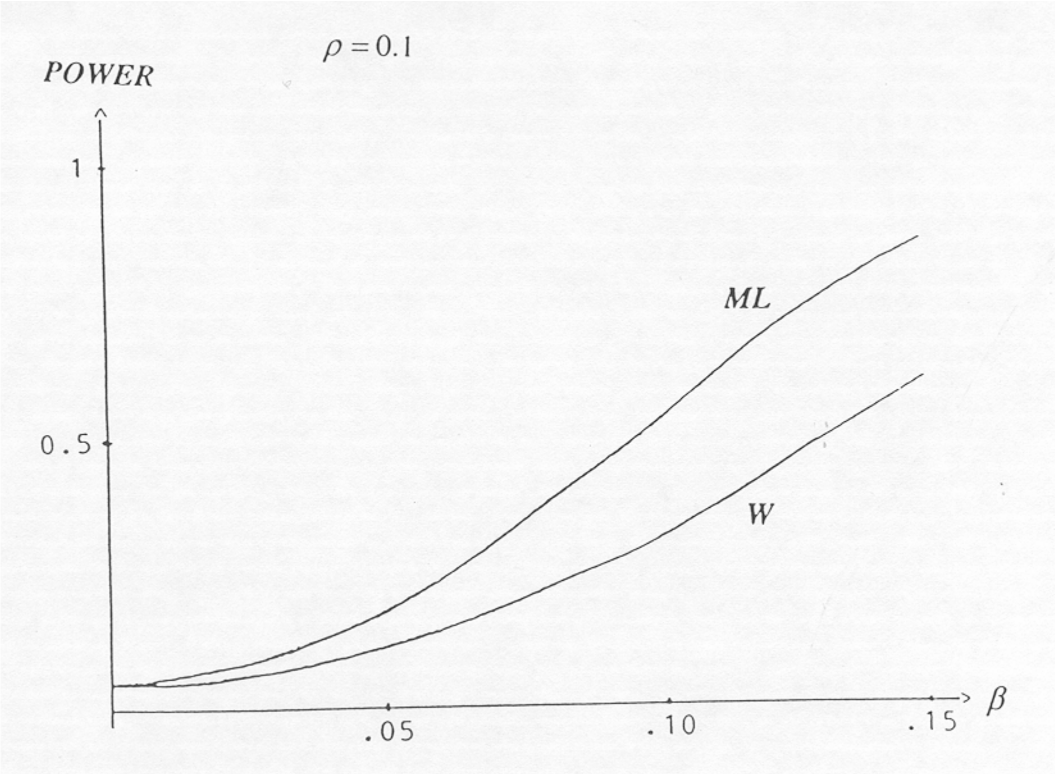
Figure 2. Positive parts of simulated power curves for two-sided tests of $\beta = 0$ at the 5% significance level when $\rho = 0.1$ and 0.9. In both cases K=1/2, T=2 and n=100. For ρ=0.1 the power of the ML statistic (***ML***) is only slightly greater than that of the Least Squares statistic (***LS***) and therefore the latter is not shown. Also the power of the within-subject statistic (***W***) is slightly greater than that of the between-subject statistic, which is not shown. When $\rho = 0.9$ the power of the between-subject statistic has been omitted because it is very low and of less interest. All results are based on the outcomes in 10^5 simulations.

Figure 1

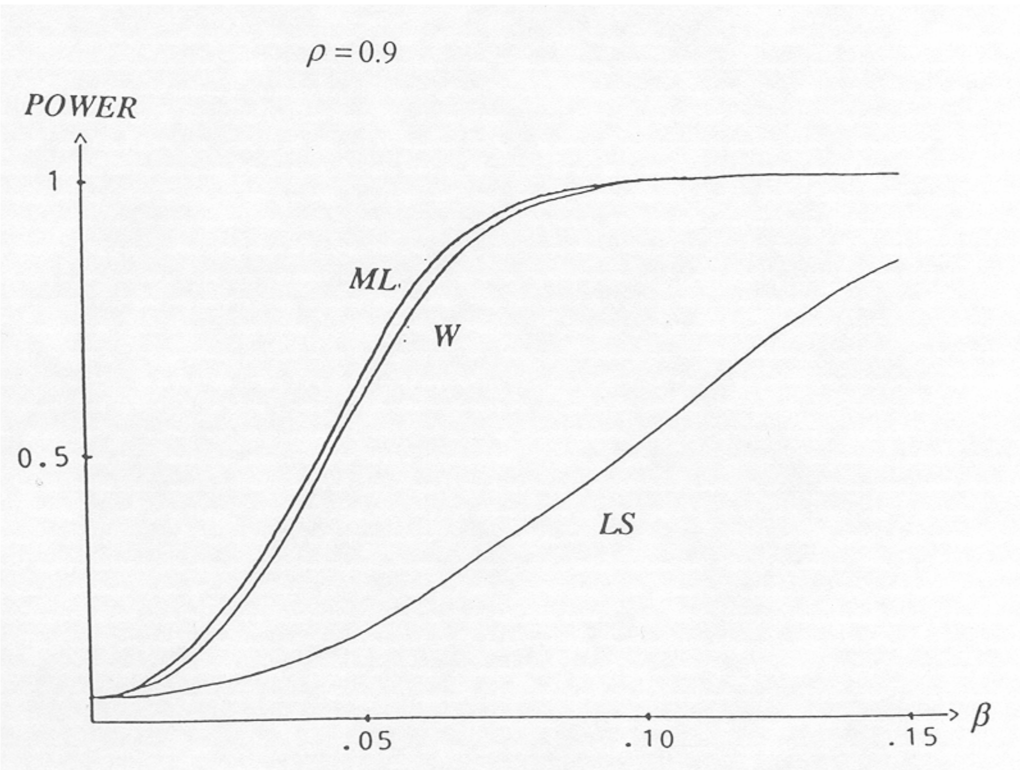


$\rho = 0.1$ $T = 2$
EFF
1.0
0.5
0.0
LS
W
B
0.0
0.5
1.0
K
$\rho = 0.9$ $T = 2$
EFF
1.0
0.5
0.0
W
LS
B
0.0
0.5
1.0
K
$\rho = 0.1$ $T = 10$
EFF
1.0
0.5
0.0
LS
W
B
0.0
0.5
1.0
K
$\rho = 0.9$ $T = 10$
EFF
1.0
0.5
0.0
W
LS
B
0.0
0.5
1.0
K

Figure 2.



$\rho = 0.1$
POWER
1
0.5
ML
W
.05
.10
.15
β



$\rho = 0.9$
POWER
1
0.5
ML
W
LS
.05
.10
.15
β

Table 1. Small sample variances (n=10) of the estimators $\hat{\beta}_B$, $\hat{\beta}_W$, $\hat{\beta}_{LS}$ and $\hat{\beta}_{ML}$, and the asymptotic variance of the ML estimator given in (9). The last column shows the percentage of the cases in which the estimation equation (8) failed to produce real-valued solutions. The small sample variances being obtained from 10^5 simulations.

ρ	T	K	$V(\hat{\beta}_B)$	$V(\hat{\beta}_W)$	$V(\hat{\beta}_{LS})$	$V(\hat{\beta}_{ML})$	$asV(\hat{\beta}_{ML})$	% missing
0.1	2	0.9	.122	.900	.108	.131	.108	2.0
"	"	0.5	.220	.180	.100	.154	.099	1.0
"	"	0.1	1.100	.100	.092	.110	.092	2.8
"	10	0.9	.211	.900	.180	.196	.171	0.0
"	"	0.5	.380	.180	.140	.177	.122	0.5
"	"	0.1	1.900	.100	.100	.101	.095	2.3
0.5	2	0.9	.167	.500	.140	.178	.125	1.3
"	"	0.5	.300	.100	.100	.113	.075	1.7
"	"	0.1	1.500	.056	.060	.057	.054	3.2
"	10	0.9	.611	.500	.500	.388	.275	0.2
"	"	0.5	1.100	.100	.300	.104	.092	2.0
"	"	0.1	5.500	.056	.100	.055	.055	2.7
0.9	2	0.9	.211	.100	.172	.105	.068	1.9
"	"	0.5	.380	.020	.100	.021	.019	3.1
"	"	0.1	1.900	.011	.028	.011	.011	3.7
"	10	0.9	1.011	.100	.820	.105	.091	1.3
"	"	0.5	1.820	.020	.460	.020	.020	2.8
"	"	0.1	9.100	.011	.100	.011	.011	2.9

Table 2. Percentiles of the distributions of the statistics T_{ML} and T_{LS} together with the corresponding percentiles for Student's T statistic where the degrees of freedom has been adjusted in a way suggested by Welch. The figures in each raw are based on 10^5 simulations.

			T_{ML}				T_{LS}				Welch T			
ρ	n	T	.01	.05	.95	.99	.01	.05	.95	.99	.01	.05	.95	.99
0.1	10	2	–3.25	–2.09	2.10	3.25	–2.79	–1.81	1.82	2.80	–2.60	–1.74	1.74	2.61
"	"	10	–3.15	–2.04	2.02	3.14	–2.80	–1.80	1.82	2.80	–2.51	–1.71	1.72	2.51
"	100	2	–2.38	–1.68	1.68	2.37	–2.36	–1.66	1.66	2.35	–2.35	–1.65	1.65	2.35
"	"	10	–2.40	–1.68	1.68	2.40	–2.37	–1.66	1.65	2.38	–2.35	–1.66	1.65	2.33
"	200	2	–2.36	–1.66	1.66	2.36	–2.35	–1.66	1.65	2.35	–2.34	–1.65	1.66	2.35
"	"	10	–2.37	–1.66	1.65	2.35	–2.34	–1.65	1.65	2.33	–2.35	–1.66	1.64	2.33
0.9	10	2	–3.18	–2.07	2.06	3.17	–2.88	–1.84	1.86	2.90	–2.61	–1.76	1.75	2.60
"	"	10	–2.46	–1.73	1.73	2.49	–2.91	–1.85	1.87	2.92	–2.52	–1.72	1.72	2.51
"	100	2	–2.37	–1.67	1.68	2.40	–2.35	–1.65	1.66	2.39	–2.30	–1.64	1.64	2.35
"	"	10	–2.34	–1.66	1.65	2.34	–2.37	–1.66	1.66	2.37	–2.36	–1.65	1.66	2.36
"	200	2	–2.36	–1.66	1.65	2.34	–2.36	–1.65	1.65	2.34	–2.34	–1.64	1.65	2.34
"	"	10	–2.34	–1.65	1.64	2.33	–2.33	–1.65	1.65	2.34	–2.33	–1.65	1.64	2.30

Table 3. Estimated dispersion matrices of the observational vector in (17) and of the variance of the difference between the observations computed at various times after first visit, x_{2j}. Estimates at times large than 3 were not computed due to small sample sizes.

x_{2j}	Sample size	Estimated dispersion matrix	$\hat{V}(y_{2j} - y_{1j})$
1	216	$\begin{bmatrix} 2.6 & 1.6 \\ 1.6 & 2.5 \end{bmatrix}$	1.9
2	172	$\begin{bmatrix} 2.5 & 1.5 \\ 1.5 & 2.1 \end{bmatrix}$	1.2
3	58	$\begin{bmatrix} 2.1 & 1.6 \\ 1.6 & 3.4 \end{bmatrix}$	2.1
4-	15	-	-