Maximum likelihood estimation of the linear model with equicorrelated errors

Giuseppe De Luca
University of Palermo, Palermo, Italy

Jan R. Magnus
Department of Econometrics and Data Science,
Vrije Universiteit Amsterdam and Tinbergen Institute, The Netherlands

Andrey L. Vasnev
University of Sydney, New South Wales, Australia

February 9, 2024

Abstract: We provide a simple proof that the maximum likelihood estimator in the linear model with equicorrelated errors does not exist, and explain why not.

JEL Classification: C20, C21.

Mathematics Subject Classification: 62F10, 62J05, 62J99.

Keywords: equicorrelation, compound symmetry, nonexistence, maximum likelihood, variance components.

Corresponding author: Andrey L. Vasnev, University of Sydney Business School, Abercrombie Building (H70), Sydney, NSW 2006, Australia.

Email addresses: giuseppe.deluca@unipa.it (De Luca),
jan@janmagnus.nl (Magnus),
andrey.vasnev@sydney.edu.au (Vasnev).

Consider a vector $y = (y_1, \dots, y_n)'$ of n observations, normally distributed with mean $X\beta$ and positive definite variance matrix V. We are interested in the case where $\text{corr}(y_i, y_j) = \rho$ for all $i \neq j$, so that the correlation matrix takes the form

$$P = \begin{pmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \vdots & \vdots & \vdots & & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{pmatrix}. \tag{1}$$

We first consider the case where $\text{var}(y_i) = 1$ for all i, in which case $V = P$; later we shall allow $\text{var}(y_i) = \sigma_i^2$.

The matrix P is known as the 'equicorrelation matrix' or the 'compound symmetric matrix,' and the distribution of y is called the 'symmetric normal distribution' (Rao, 1973, Section 3c). Our purpose is to estimate β and ρ by maximum likelihood (ML). This is not possible, and the current note attempts to explain why not.

The possible nonexistence of a ML estimator is a well-known problem in statistics. In the context of variance components, the phenomenon has been studied by Infante (1995), Demidenko and Massam (1999) (with a correction by Grzadziel and Michalski, 2014), Birkes and Wulff (2003), Eck and Geyer (2018), Shi and Xu (2020), and others. All these papers refer to underlying (and sometimes rather complex) theoretical results from other sources. In contrast, our note provides a simple and elementary proof without reference to any exterior source, and therefore allows us to precisely identify and understand the reason why a ML estimator does not exist in this case. We also provide a nontrivial extension which, as far as we know, is new.

Let $\imath$ denote the $n \times 1$ vector of ones. Then, defining the idempotent $n \times n$ matrix $J = \imath\imath'/n$, we can write P as

$$P = \alpha_1 J + \alpha_2 (I_n - J), \tag{2}$$

where $\alpha_1 = (n-1)\rho + 1$ and $\alpha_2 = 1 - \rho$ are the eigenvalues of P with multiplicities 1 and $n-1$, respectively. The matrix P is positive definite if and only if $\alpha_1 > 0$ and $\alpha_2 > 0$, that is, if and only if

$$\frac{-1}{n-1} < \rho < 1. \tag{3}$$

If P is positive definite, then

$$P^{-1} = \frac{1}{\alpha_1} J + \frac{1}{\alpha_2}(I_n - J), \qquad |P| = \alpha_1 \alpha_2^{n-1}. \tag{4}$$

Letting $M = I_n - X(X'X)^{-1}X'$, we see that

$$X'P^{-1}M = -\frac{\rho}{(1-\rho)(n\rho + 1 - \rho)} (X'\imath)(M\imath)', \tag{5}$$

since $MX = 0$. If the regression contains a constant term (or, more generally, if $\imath$ lies in the column space of X), then $M\imath = 0$ and hence

$$X'P^{-1} - X'P^{-1}X(X'X)^{-1}X' = X'P^{-1}M = 0, \tag{6}$$

implying that

$$(X'P^{-1}X)^{-1}X'P^{-1} = (X'X)^{-1}X', \tag{7}$$

so that GLS = OLS in this case; see Amemiya (1985, Theorem 6.1.1) or Lu and Schmidt (2012) for a more general discussion and historical references.

The likelihood takes the form

$$L = (2\pi)^{-n/2}|P|^{-1/2} \exp -\frac{1}{2}(y - X\beta)'P^{-1}(y - X\beta), \tag{8}$$

and, letting $u = y - X\beta$, two times the loglikelihood is

$$\begin{aligned} 2\log L &= -n\log 2\pi - \log|P| - u'P^{-1}u \\ &= -n\log 2\pi - \log\alpha_1 - (n-1)\log\alpha_2 - \frac{n\bar{u}^2}{\alpha_1} - \frac{ns_u^2}{\alpha_2}, \end{aligned} \tag{9}$$

where

$$\bar{u} = \frac{\imath'u}{n}, \qquad s_u^2 = \frac{(u - \bar{u}\imath)'(u - \bar{u}\imath)}{n}. \tag{10}$$

When $\rho \to 1$, we have

$$\begin{aligned} 2\log L \to &-n\log 2\pi - \log n - \bar{u}^2 \\ &- (n-1)\lim_{\alpha_2 \to 0^+}\left(\log\alpha_2 + \frac{ns_u^2/(n-1)}{\alpha_2}\right), \end{aligned} \tag{11}$$

and when $\rho \to -1/(n-1)$, we have

$$\begin{aligned} 2\log L \to &-n\log 2\pi - (n-1)\log(n/(n-1)) - (n-1)s_u^2 \\ &- \lim_{\alpha_1 \to 0^+}\left(\log\alpha_1 + \frac{n\bar{u}^2}{\alpha_1}\right). \end{aligned} \tag{12}$$

The problem lies in the appearance of $\bar{u}$ in (12). If β is known, so that we only estimate ρ, then there is no problem as long as $\imath'(y - X\beta)$ does not vanish. But if β is unknown, then we need to estimate both β and ρ. In that case, the *concentrated* loglikelihood is given by (9) with u replaced by the residual vector

$$e = y - X\hat{\beta} = y - X(X'P^{-1}X)^{-1}X'P^{-1}y = y - X(X'X)^{-1}X'y, \tag{13}$$

in view of (7), assuming that the regression contains a constant term. In general, residuals do not sum to zero in generalized least squares, but here they do, and hence $\bar{e} = 0$. This does not affect (11), but it does affect (12). To see why, recall that for any $a \geq 0$,

$$\lim_{x \to 0^+} \left(\log x + \frac{a}{x} \right) = \begin{cases} -\infty \text{ if } a = 0, \\ +\infty \text{ if } a > 0. \end{cases} \tag{14}$$

Given (14), the loglikelihood (12) approaches $+\infty$ when ρ approaches its lower bound $-1/(n-1)$, and hence a ML estimator does not exist in this case.

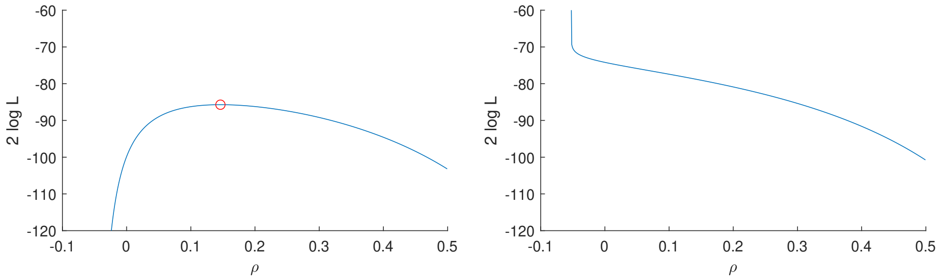

Figure 1: Concentrated loglikelihood as a function of ρ

The behavior of the loglikelihood is illustrated in Figure 1 for the case $X = \imath$ (only constant term), $n = 20$, $\beta = 0$, and $\rho = 0.1$, first for the case where β is known so that only ρ is estimated (left panel, the circle indicates the maximum at $\rho = 0.1461$), then for the case where β is not known so that both β and ρ are estimated (right panel).

Let us finally consider three related cases. First, if $V = \sigma^2 P$ and σ^2 is not known, so that it needs to be estimated together with β and ρ, then (9) is replaced by

$$2 \log L = -n \log 2\pi - n - n \log \left(\frac{\bar{u}^2}{\alpha_1} + \frac{s_u^2}{\alpha_2} \right) - \log \alpha_1 - (n-1) \log \alpha_2, \tag{15}$$

which reduces, in the special case $\bar{u} = 0$, to

$$2 \log L = -n \log 2\pi - n - n \log s_u^2 - \log \alpha_1 + \log \alpha_2. \tag{16}$$

Since $\log \alpha_1 \to -\infty$ as $\alpha_1 \to 0^+$, a ML estimator does not exist in this case either, if the regression contains a constant term.

Second, if the variances σ_i^2 are not all equal to one, then the variance matrix is given by

$$V = V_0^{1/2} P V_0^{1/2}, \qquad V_0 = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2), \tag{17}$$

where P is the correlation matrix defined in (1) and we assume that the σ_i^2 are all known. This case is the 'nontrivial' extension alluded to in the introduction, and it plays a role in meta-analysis. It would seem that, upon premultiplying the regression by $V_0^{-1/2}$, and noting that $\imath$ is, in general, not a linear combination of the columns of $V_0^{-1/2}X$, we should conclude that a ML does exist in this case. This conclusion is, however, incorrect — no ML estimator exists. To see why this happens, let us write the normal equation $X'V^{-1}X\hat{\beta} = X'V^{-1}y$ as

$$\left(X_*'JX_* + \epsilon X_*'(I_n - J)X_*\right)\hat{\beta} = X_*'Jy_* + \epsilon X_*'(I_n - J)y_*, \tag{18}$$

where $X_* = V_0^{-1/2}X$, $y_* = V_0^{-1/2}y$, and $\epsilon = \alpha_1/\alpha_2$. Using the fact that $J = \imath\imath'/n$, we then obtain

$$(1-\epsilon)X_*'\imath\imath'(y_* - X_*\hat{\beta}) + n\epsilon X_*'(y_* - X_*\hat{\beta}) = 0, \tag{19}$$

where we note that $X_*'(y_* - X_*\hat{\beta})$ will, in general, not vanish. Hence, if $X'V_0^{-1/2}\imath \neq 0$, it follows that

$$\lim_{\epsilon \to 0^+} \imath' V_0^{-1/2}(y - X\hat{\beta}) = 0. \tag{20}$$

If V_0 is equal or proportional to the identity matrix, then GLS = OLS and the residuals sum to 0. But even if V_0 is not proportional to the identity matrix, (20) holds and no ML estimator exists.

Third, if the observations constitute a panel, say y_{it}, where equicorrelation applies to one dimension but not the other, then the problem disappears and a ML estimator does exist; see Greene (2018, pp. 607–608).

References

Amemiya, T. (1985). *Advanced Econometrics*, Basil Blackwell, Oxford, UK.

Birkes, D. and S. S. Wulff (2003). Existence of maximum likelihood estimates in normal variance-components models, *Journal of Statistical Planning and Inference*, 113, 35–47.

Demidenko, E. and Massam, H. (1999). On the existence of the maximum likelihood estimate in variance components models, *Sankhyā, Series A*, 61, 431–443.

Eck, D. J. and C. J. Geyer (2018). Computationally efficient likelihood inference in exponential families when the maximum likelihood estimator does not exist, *Electronic Journal of Statistics*, ArXiv:1803.11240v2.

Greene, W. H. (2018). *Econometric Analysis*, 8th edition, Pearson.

Grzadziel, M. and A. Michalski (2014). A note on the existence of the maximum likelihood estimate in variance components models, *Discussiones Mathematicae Probability and Statistics*, 34, 159–167.

Infante, A. M. (1995). The difficulties of estimation of dispersion parameters in linear models: An illustration, *Statistical Papers*, 36, 183–189.

Lu, C. and P. Schmidt (2012). Conditions for the numerical equality of the OLS, GLS and Amemiya–Cragg estimators, *Economics Letters*, 116(3), 538–540.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd edition, Wiley, New York.

Shi, Y. and P. Xu (2020). Nonexistence of maximum likelihood estimation of variance components in some stochastic models, *American Society of Civil Engineers (ASCE)*, 146(2), 06020001.