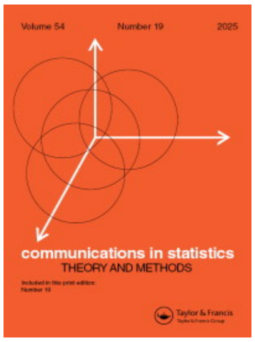

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: www.tandfonline.com/journals/lsta20

Maximum likelihood estimation of the linear model with equicorrelated errors

Giuseppe De Luca, Jan R. Magnus & Andrey L. Vasnev

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2025, VOL. 54, NO. 19, 6295–6302
https://doi.org/10.1080/03610926.2025.2455939

Maximum likelihood estimation of the linear model with equicorrelated errors

Giuseppe De Luca[a], Jan R. Magnus[b], and Andrey L. Vasnev[c]

[a]University of Palermo, Palermo, Italy; [b]Department of Econometrics and Data Science, Vrije Universiteit Amsterdam and Tinbergen Institute, Amsterdam, The Netherlands; [c]University of Sydney, New South Wales, Australia

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.

ARTICLE HISTORY
14 March 2024
13 January 2025

KEYWORDS
Equicorrelation; compound symmetry; non existence; maximum likelihood; variance components

JEL CLASSIFICATION:
C20, C21

MATHEMATICS SUBJECT CLASSIFICATION:
62F10, 62J05, 62J99

1. Introduction

Consider a vector $y = (y_1, \dots, y_n)'$ of n observations, normally distributed with mean $X\beta$ and positive definite variance matrix V, where the elements of V depend on a finite number of unknown parameters $\theta = (\theta_1, \dots, \theta_m)'$. The likelihood is given by

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

and hence maximizing the likelihood is equivalent to minimizing

$$\mathcal{L}(\beta, \theta) = (y - X\beta)'V^{-1}(y - X\beta) + \log|V|. \tag{2}$$

If V is known, then the maximum-likelihood (ML) estimator of β is given by

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y, \tag{3}$$

the generalized least-squares (GLS) estimator. If V is not known, then we may obtain the ML estimators either by minimizing (2) with respect to β and θ or by concentrating out β, in which case we need to minimize

$$\mathcal{L}_c(\theta) = e'V^{-1}e + \log|V| \tag{4}$$

CONTACT Andrey L. Vasnev andrey.vasnev@sydney.edu.au University of Sydney, New South Wales, Australia.

with respect to θ only, where the residual vector

$$e = y - X(X'V^{-1}X)^{-1}X'V^{-1}y \tag{5}$$

now also depends on θ. In the latter case, we first obtain $\hat{\theta}$, then compute $\hat{V} = V(\hat{\theta})$, and finally $\hat{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y$. Both procedures are simple and mathematically transparent, *but they don't always work*. In this note, we shall discuss one important case where the procedure does not work, and explain why it does not work.

The case we are interested in is the case where $\operatorname{corr}(y_i, y_j) = \rho$ for all $i \neq j$, so that the correlation matrix takes the form

$$P = \begin{pmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \vdots & \vdots & \vdots & & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{pmatrix}. \tag{6}$$

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). We first consider the case where $\operatorname{var}(y_i) = 1$ for all i, in which case $V = P$. In Section 4, we shall allow $\operatorname{var}(y_i)$ to depend on i.

The specification $V = P$ is not a weird outlier that never occurs in practice. On the contrary, the idea that all correlations are the same is the simplest form of a correlation structure. As pointed out by Zyskind (1969, p. 1361), the structure arises naturally when we sample without replacement from a population of numerical characteristics $(c_1, \dots, c_N)$. In the population, we define $\mu = \sum_{i=1}^N c_i/N$ and $\sigma^2 = \sum_{i=1}^N (c_i - \mu)^2/N$. Collecting a sample $y = (y_1, \dots, y_n)'$ of size n from this population without replacement gives

$$\Pr(y_1 = c_i) = \frac{1}{N}, \qquad \Pr(y_1 = c_i, y_2 = c_j) = \frac{1}{N(N-1)} \ (i \neq j), \tag{7}$$

and hence $\mathrm{E}(y_1) = \mathrm{E}(y_2) = \mu$, $\mathrm{E}(y_1^2) = \mathrm{E}(y_2^2) = \sigma^2 + \mu^2$, and

$$\mathrm{E}(y_1 y_2) = \frac{1}{N(N-1)} \sum_{i \neq j} c_i c_j = -\frac{\sigma^2}{N-1} + \mu^2, \tag{8}$$

so that $\operatorname{var}(y_1) = \operatorname{var}(y_2) = \sigma^2$ and

$$\operatorname{cov}(y_1, y_2) = \mathrm{E}(y_1 y_2) - \mathrm{E}(y_1)\,\mathrm{E}(y_2) = -\frac{\sigma^2}{N-1}. \tag{9}$$

Hence $\operatorname{var}(y) = \sigma^2 P$ with $\rho = -1/(N-1)$, which is positive definite (hence non singular) for every $n < N$.

The structure also arises as a special case ($m = 2$) of the variance-components model (Rao and Kleffe 1988), where

$$V = \alpha_0 I_n + \sum_{i=1}^m \alpha_i A_i, \tag{10}$$

which is widely used in practice. More recently, Engle and Kelly (2011) introduced dynamic equicorrelation in the finance literature, where it has many followers.

The fact that a ML estimator does not exist in such a simple setup is perhaps remarkable, but the case is by no means unique. In the context of variance components, the non existence of a ML estimator 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, we provide a simple and elementary proof without reference to any exterior source, and this allows us to precisely identify and understand the reason why a ML estimator does not exist in this case. In Section 4, we also provide a non trivial extension which, as far as we know, is new.

2. Main result

Let us formally state our main result as follows.

Proposition 1. *Consider the model $y = X\beta + u$, where X is a non random $n \times k$ matrix with full column-rank, u is a normally distributed random error with $\mathrm{E}(u) = 0$ and $\mathrm{var}(u) = P$ as given in (6), and β and ρ are unknown parameters to be estimated. If the regression contains a constant term, then no ML estimator exists.*

To prove the result, 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 the matrix P as

$$P = \alpha_1 J + \alpha_2 (I_n - J), \tag{11}$$

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

$$\frac{-1}{n-1} < \rho < 1. \tag{12}$$

If P is positive definite, then

$$P^{-1} = \frac{1}{\alpha_1} J + \frac{1}{\alpha_2}(I_n - J), \qquad |P| = \alpha_1 \alpha_2^{n-1}. \tag{13}$$

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

$$X'P^{-1}M = -\frac{\rho}{(1-\rho)(n\rho + 1 - \rho)}(X'\imath)(M\imath)', \tag{14}$$

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

$$X'P^{-1} - X'P^{-1}X(X'X)^{-1}X' = X'P^{-1}M = 0, \tag{15}$$

implying that

$$(X'P^{-1}X)^{-1}X'P^{-1} = (X'X)^{-1}X', \tag{16}$$

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.

Letting $u = y - X\beta$, we have

$$u'Ju = n\bar{u}^2, \qquad u'(I_n - J)u = ns_u^2, \tag{17}$$

where

$$\bar{u} = \frac{\iota' u}{n}, \qquad s_u^2 = \frac{(u - \bar{u}\iota)'(u - \bar{u}\iota)}{n}, \tag{18}$$

and hence, using (2),

$$\begin{aligned}\mathcal{L}(\beta, \rho) &= u'P^{-1}u + \log|P| \\ &= \frac{n\bar{u}^2}{\alpha_1} + \frac{ns_u^2}{\alpha_2} + \log\alpha_1 + (n-1)\log\alpha_2.\end{aligned} \tag{19}$$

When $\rho \to 1$, we have $\alpha_1 \to n$ and $\alpha_2 \to 0^+$, and hence

$$\mathcal{L}(\beta, \rho) \to \bar{u}^2 + \log n + (n-1) \lim_{\alpha_2 \to 0^+} \left(\log\alpha_2 + \frac{ns_u^2/(n-1)}{\alpha_2} \right), \tag{20}$$

and when $\rho \to -1/(n-1)$, we have $\alpha_1 \to 0^+$ and $\alpha_2 \to n/(n-1)$, and hence

$$\mathcal{L}(\beta, \rho) \to (n-1)s_u^2 + (n-1)\log\frac{n}{n-1} + \lim_{\alpha_1 \to 0^+} \left(\log\alpha_1 + \frac{n\bar{u}^2}{\alpha_1} \right). \tag{21}$$

The problem lies in the appearance of $\bar{u}$ in (21). If β is known, so that we only estimate ρ, then there is no problem as long as $\iota'(y - X\beta)$ does not vanish. But if β is unknown, then we need to estimate both β and ρ. In that case, we have to replace the error vector u in (19) by the residual vector

$$e = y - X\hat{\beta} = y - X(X'P^{-1}X)^{-1}X'P^{-1}y = y - X(X'X)^{-1}X'y, \tag{22}$$

in view of (16) and using the fact 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 (20), but it does affect (21). To see why, recall that for any $a \geq 0$,

$$\lim_{x \to 0^+} \left(\log x + \frac{a}{x} \right) = \begin{cases} -\infty & \text{if } a = 0, \\ +\infty & \text{if } a > 0. \end{cases} \tag{23}$$

Given (23), the function (21) approaches $-\infty$ when ρ approaches its lower bound $-1/(n-1)$, and hence achieves no finite minimum. Consequently, a ML estimator does not exist in this case.

3. Discussion

To illustrate the non existence, we consider a simple setup, where we take $X = \iota$ (only constant term), $n = 20$, $\beta = 0$, and $\rho = 0.1$. The function $\mathcal{L}(\beta, \rho)$ is plotted in Figure 1, first for the case where β is known so that only ρ is estimated (left panel), then for the case where β is not known so that both β and ρ are estimated (right panel).

The dashed line indicates the asymptote at $\rho = -1/(n-1) \approx -0.053$. In the left panel (where β is assumed to be known), we do achieve a minimum at $\rho = 0.146$, but in the right panel (where both β and ρ are estimated) the function diverges to $-\infty$ as ρ approaches its lower bound.

The fact that no ML estimator exists raises the question whether this failure is a feature of the method of estimation (maximum likelihood) or of the variance structure (equicorrelation). Let us therefore investigate what happens when we estimate β and ρ by least squares rather than by ML. In the (generalized) least-squares approach, we minimize

$$\mathcal{L}^*(\beta, \theta) = (y - X\beta)'V^{-1}(y - X\beta) \tag{24}$$

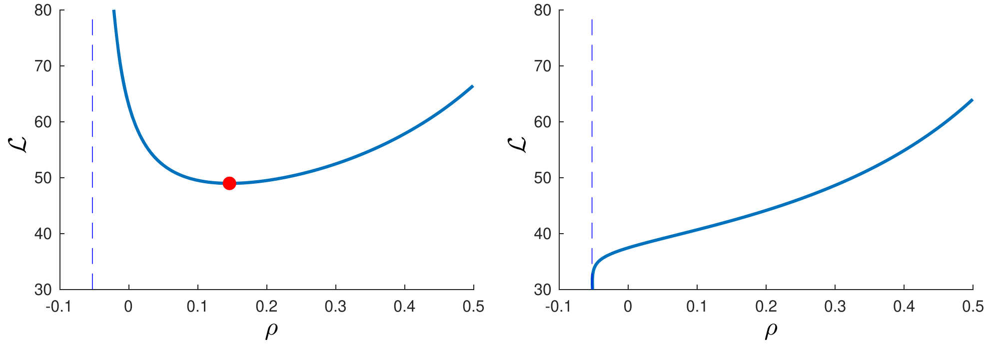

Figure 1. $\mathcal{L}(\beta, \rho)$ as a function of ρ; left panel: β known; right panel: β unknown.

rather than (2), and in our case, assuming that the regression contains a constant term, this amounts to minimizing

$$\frac{e'Je}{\alpha_1} + \frac{e'(I_n - J)e}{\alpha_2} = \frac{e'e}{\alpha_2}, \tag{25}$$

where $e = (I_n - X(X'X)^{-1}X')y$ and we used the fact that $Je = 0$. Hence (25) is minimized when $\alpha_2 = 1 - \rho$ is maximized, that is at $\rho = -1/(n-1)$. The least-squares estimators are

$$\hat{\beta} = (X'X)^{-1}X'y, \qquad \hat{\rho} = -\frac{1}{n-1}. \tag{26}$$

Thus, a least-squares solution exists, but the corresponding variance matrix

$$P = \frac{n}{n-1}(I_n - J) \tag{27}$$

is singular. In both cases (ML and least squares), the estimator of ρ equals or approaches the lower bound $-1/(n-1)$, which is not really an estimator, since it does not depend on the sample. Hence, in neither case can we estimate ρ meaningfully, and this is caused by the equicorrelation structure (and the presence of a constant term), not by the method of estimation.

4. Extensions

Let us now 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 (19) is replaced by

$$\mathcal{L}(\beta, \rho) = n + n\log\left(\frac{\bar{u}^2}{\alpha_1} + \frac{s_u^2}{\alpha_2}\right) + \log\alpha_1 + (n-1)\log\alpha_2, \tag{28}$$

after concentrating out σ^2, which, in the special case $\bar{u} = 0$, reduces to

$$\mathcal{L}(\beta, \rho) = n + n\log s_u^2 + \log\alpha_1 - \log\alpha_2. \tag{29}$$

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.

Our second extension merits a more formal statement.

Proposition 2. *Consider the model $y = X\beta + u$, where X is a non random $n \times k$ matrix with full column-rank, u is a normally distributed random error with $\mathrm{E}(u) = 0$ and $\mathrm{var}(u) = V = V_0^{1/2} P V_0^{1/2}$, where*

$$V_0 = \mathrm{diag}(\sigma_1^2, \dots, \sigma_n^2),$$

and P is the correlation matrix defined in (6). We assume that the σ_i^2 are all known, and that β and ρ are the unknown parameters to be estimated. If the regression contains a constant term, then no ML estimator exists.

In contrast to Proposition 1 where $\sigma_i^2 = 1$ for all i and hence $V = P$, the variances σ_i^2 are now not necessarily equal to 1 and not necessarily equal to each other. This case is the 'nontrivial' extension alluded to in the introduction, and it plays a role in meta-analysis (Magnus and Vasnev 2025). 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.

Proof. Write the normal equation $X'V^{-1}X\hat{\beta} = X'V^{-1}y$ as

$$\left(X_*'JX_* + \lambda X_*'(I_n - J)X_*\right)\hat{\beta} = X_*'Jy_* + \lambda X_*'(I_n - J)y_*, \tag{30}$$

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

$$(1 - \lambda)X_*'\imath\imath'(y_* - X_*\hat{\beta}) + n\lambda X_*'(y_* - X_*\hat{\beta}) = 0, \tag{31}$$

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

$$\lim_{\lambda \to 0+} \imath' V_0^{-1/2}(y - X\hat{\beta}) = 0. \tag{32}$$

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, (32) holds and no ML estimator exists. $\square$

In Proposition 2, we have assumed that all σ_i^2 are known. This may in fact happen in practice with grouped data, where $\sigma_i^2 = 1/n_i$ and n_i denotes the (known) number of observations underlying y_i. But it is more common that the σ_i^2 are not known and have to be estimated. If all variances are unknown, then we cannot estimate the parameters, because the number of parameters exceeds the number of observations. So we have to assume that the variances are known functions of a finite set of parameters, in which case an argument similar to extension 1 applies. The simplest such situation occurs when we have two subsamples with two variances, i.e., when $V_0 = \mathrm{diag}(\sigma_1^2\imath_1', \sigma_2^2\imath_2')$, where $\imath_i$ denotes the $n_i \times 1$ vector of ones. Then, (32) still holds when σ_1^2 and σ_2^2 are estimated, so that no ML estimator exists.

Our third and final extension considers the case where the observations constitute a panel, say y_{it}, where equicorrelation applies to one dimension but not the other. In that case the problem disappears and a ML estimator does exist; see Greene (2018, pp. 607–608).

5. Conclusions

In this note, we have attempted to explain why a ML estimator in the linear model with equicorrelated errors does not exist. Since this model is frequently used in practice, this finding should serve as a warning not to use the equicorrelation structure in a simple regression framework. Things are different in a panel-data framework, where equicorrelation does not cause any estimation problems.

The fact that we cannot estimate ρ (at least not meaningfully) is caused by the equicorrelation structure, not by the method of estimation, because problems also occur when we estimate by least squares.

Acknowledgments

We thank the editor and reviewer for their helpful comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Giuseppe De Luca acknowledges financial support from MUIR, Italy, PRIN PRJ-1446 and PRIN PRJ-1547 at the University of Palermo.

ORCID

Giuseppe De Luca https://orcid.org/0000-0002-1411-2543
Jan R. Magnus https://orcid.org/0000-0002-3390-639X
Andrey L. Vasnev https://orcid.org/0000-0001-5213-2855

References

Amemiya, T. 1985. Advanced econometrics. Hoboken, NJ: Basil Blackwell.

Birkes, D., and S. S. Wulff. 2003. Existence of maximum likelihood estimates in normal variance-components models. *Journal of Statistical Planning and Inference* 113 (1):35–47. doi: 10.1016/S0378-3758(01)00302-0.

Demidenko, E., and H. Massam. 1999. On the existence of the maximum likelihood estimate in variance components models. *Sankhyā, Series A* 61:431–43.

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* 15 (1):2105–56. ArXiv:1803.11240v2.

Engle, R., and B. Kelly. 2012. Dynamic equicorrelation. *Journal of Business & Economic Statistics* 30 (2):212–28. doi: 10.1080/07350015.2011.652048.

Greene, W. H. 2018. *Econometric analysis*, 8th ed. London, UK: 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 (1-2):159. doi: 10.7151/dmps.1164.

Infante, A. M. 1995. The difficulties of estimation of dispersion parameters in linear models: An illustration. *Statistical Papers* 36 (1):183–9. doi: 10.1007/BF02926031.

Lu, C., and P. Schmidt. 2012. Conditions for the numerical equality of the OLS, GLS and Amemiya–Cragg estimators. *Economics Letters* 116 (3):538–40. doi: 10.1016/j.econlet.2012.01.015.

Magnus, J. R., and A. L. Vasnev. 2025. *More information, less precision: Meta-analysis through random effects. Submitted for publication*. Available at https://www.janmagnus.nl/wips/meta-analysis.pdf.

Rao, C. R. 1973. *Linear statistical inference and its applications*, 2nd ed. New York: Wiley.

Rao, C. R., and J. Kleffe. 1988. *Estimation of variance components and applications*. Amsterdam: North Holland.

Shi, Y., and P. Xu. 2020. *Nonexistence of maximum likelihood estimation of variance components in some stochastic models*, Vol. 146, 6020001. Reston, VA: ASCE.

Zyskind, G. 1969. parametric augmentations and error structures under which certain simple least squares and analysis of variance procedures are also best. *Journal of the American Statistical Association* 64 (328):1353–68. doi: 10.1080/01621459.1969.10501062.