Mathematical Statistics
Stockholm University

Finite Sample Size Bounds on the Variance Estimator in Non-Gaussian General Linear Models

Mathias Lindholm and Felix Wahl

Research Report 2018:2

ISSN 1650-0377

Postal address:
Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:
http://www.math.su.se

Mathematical Statistics
Stockholm University
Research Report **2018:2**,
http://www.math.su.se

Finite Sample Size Bounds on the Variance Estimator in Non-Gaussian General Linear Models

Mathias Lindholm and Felix Wahl

January 2018

Abstract

We consider bounds on the variance of the standard unbiased variance estimator in a general non-Gaussian linear model for finite sample sizes. In particular we obtain bounds that are sharp in the sense that both the lower and upper bound will converge to the same asymptotic limit when scaled with the sample size. Further, these bounds are independent of covariate information. Due to this we may also obtain unconditional variance bounds for the situation with random covariates. The above results rely on a result in Atiqullah (1962) which is stated without proof. We provide a proof of this result, both for easy reference, but also since the derivation of the variance bounds rely on an observation in the construction of this proof.

Keywords: General linear models, non-Gaussian error terms, moments of variance estimators, finite sample size bounds, random covariates, unconditional bounds

1 Introduction

In the present note we will consider the variance of the standard unbiased variance estimator in a general linear model (GLM) with non-Gaussian error terms. For this class of models the covariance of the generalized least squares estimator $\hat{\boldsymbol{\beta}}$ is well-known, but an explicit expression for the variance of $\hat{\sigma}^2$ is not as widely known. In [2] an expression for this variance is provided without proof for a mixed GLM with non-Gaussian homoscedastic error terms. This result allows us to obtain finite sample size bounds on the variance of $\hat{\sigma}^2$ which are independent of the covariates in the non-Gaussian GLM. Moreover, the bounds which we obtain are sharp in the sense that

both the lower and upper bound, when scaled by the sample size n, converges to the asymptotic variance of $\sqrt{n}\hat{\sigma}^2$ as $n \to \infty$. Further, we note that these results remain valid if the covariates are random, but conditioned upon. Consequently, since the variance bounds are independent of covariate information it is possible to obtain *unconditional* variance bounds for the situation when the covariates themselves are random, given suitable regularity conditions. This is interesting since the unconditional variance of the standard unbiased variance estimator is in general not explicitly computable due to the randomness of the covariates.

The problem formulation above of course relies on the theory of random quadratic forms. For more on this topic, see e.g. [4, 5, 8] and the references therein. One can also note that the special case of a sample variance in the non-Gaussian setting, i.e. an intercept only GLM, was treated already in e.g. [3, Eq. (27.4.2)]. Other similar results are obtained in the theory of minimum variance component estimation, see e.g. [6, 7] and the proof of [8, Thm. 3.4].

One direct application of the bounds derived in the present note is that they provide a simple way of assessing consistency of $\hat{\sigma}^2$.

Another motivating example is the situation when we want to calculate conditional moments of a non-linear function of the parameter estimators. This is typically computationally infeasible, but the results in the present note provide exact expressions for second order Taylor approximations of conditional moments. Analogous approximations of unconditional moments will not result in exact expressions, but we may obtain sharp finite sample size bounds.

2 The General Linear Model

Let $\boldsymbol{Y}$ be a random $n \times 1$ vector and let $\boldsymbol{X}$ be a random $n \times p$ matrix of almost surely full column rank, $\operatorname{rank}(\boldsymbol{X}) = p$, with $n > p$. Further, let $\boldsymbol{\Sigma}$ be a symmetric almost surely strictly positive definite $n \times n$ matrix. This ensures that we may define $\boldsymbol{\Sigma}^{1/2}$ in the standard way using orthogonalization. The class of GLMs which will be studied in the present note are of the form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \sigma\boldsymbol{\Sigma}^{1/2}\boldsymbol{e}, \tag{1}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector, $\sigma > 0$ is a scalar and $\boldsymbol{e}$ is some random $n \times 1$ vector whose elements are independent. Moreover, we assume that $\boldsymbol{e}$ has, *conditional* on $\boldsymbol{X}$ and $\boldsymbol{\Sigma}$, mean $\boldsymbol{0}$ and covariance $\boldsymbol{I}_n$, together with common central fourth moments μ_4. Here $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix. The

standard generalized least squares estimator of $\boldsymbol{\beta}$, conditional on $\boldsymbol{X}$ and $\boldsymbol{\Sigma}$, is given by:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y},$$

see for instance [8, Sec. 3.10] for this and more on the general linear model. Moreover, an unbiased estimator of σ^2 (conditional on $\boldsymbol{X}$ and $\boldsymbol{\Sigma}$), and the estimator we will focus on in the present note, is given by:

$$\hat{\sigma}^2 = \frac{1}{n-p}\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right). \tag{2}$$

The results below will of course remain valid, with the obvious changes, if we consider an estimator normalized with some other, non-degenerate, function of n and p, for instance simply n. It is important to note that $\boldsymbol{X}$ is random, but we assume that it is possible to observe perfectly. That is, we are not dealing with an errors-in-variables model, which would lead to problems such as biased estimators.

As stated in the introduction, the result of [2] concerns general linear mixed models with non-Gaussian homoscedastic errors. In order to see how this result may be used in the above setting, let $\widetilde{\boldsymbol{Y}} := \boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y}$ and $\widetilde{\boldsymbol{X}} := \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}$, which gives us that we can restate (1) as the linear model

$$\widetilde{\boldsymbol{Y}} = \widetilde{\boldsymbol{X}}\boldsymbol{\beta} + \sigma\boldsymbol{e}, \tag{3}$$

and may hence rephrase the estimator $\hat{\sigma}^2$ in terms of (3). Further, define

$$\boldsymbol{U} := \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1/2},$$

which corresponds to the projection matrix associated with the linear model (3), and we note for future reference that $\mathrm{tr}(\boldsymbol{U}) = \mathrm{rank}(\boldsymbol{X}) = p$. If, in addition, we introduce

$$\boldsymbol{K} := \boldsymbol{I}_n - \boldsymbol{U}$$

we obtain the following result, which is a special case of a result stated without proof in [2]:

Proposition 1. *(Atiqullah (1962) for the special case* (1)*, see [2]) Let*

(i) $\boldsymbol{\Sigma}$ *be a random symmetric almost surely positive definite* $n \times n$ *matrix and* $\boldsymbol{X}$ *be a random* $n \times p$ *matrix of almost surely full column rank,*

(ii) the error term components defining the random $n \times 1$ vector $\boldsymbol{e}$ be independent with, conditional on $\boldsymbol{X}$ and $\boldsymbol{\Sigma}$, mean 0, variance 1 and common fourth central moments μ_4.

It then holds that

$$\mathrm{Var}(\hat{\sigma}^2|\boldsymbol{X}, \boldsymbol{\Sigma}) = \frac{\sigma^4}{n-p}\left(2 + \frac{\mu_4 - 3}{n-p}\sum_{i=1}^{n} \boldsymbol{K}_{ii}^2\right). \tag{4}$$

The proof of Proposition 1 relies on that (2) may be rewritten according to

$$\hat{\sigma}^2 = \frac{\sigma^2}{n-p}\boldsymbol{e}'\boldsymbol{K}\boldsymbol{e}, \tag{5}$$

which together with an application of [8, Thm. 1.6] yields the desired result. Further, the derivation of the variance bounds stated below in Corollary 1 and 2, rely on that $\boldsymbol{U}$ and $\boldsymbol{K}$ are idempotent, facts used in the derivation of (5).

Proof. (Proof of Proposition 1) In order to see how this result is obtained, one can note that if we let $\hat{\tilde{\boldsymbol{e}}} := \widetilde{\boldsymbol{Y}} - \widetilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}$, it follows that

$$\hat{\sigma}^2 = \frac{1}{n-p}\hat{\tilde{\boldsymbol{e}}}'\hat{\tilde{\boldsymbol{e}}},$$

which may be expressed in terms of $\boldsymbol{e}$ by noting that

$$\begin{aligned}
\hat{\tilde{\boldsymbol{e}}} &= \left(\widetilde{\boldsymbol{X}}\boldsymbol{\beta} + \sigma\boldsymbol{e}\right) - \boldsymbol{U}\widetilde{\boldsymbol{Y}} \\
&= \widetilde{\boldsymbol{X}}\boldsymbol{\beta} + \sigma\boldsymbol{e} - \boldsymbol{U}\left(\widetilde{\boldsymbol{X}}\boldsymbol{\beta} + \sigma\boldsymbol{e}\right) \\
&= \sigma\left(\boldsymbol{I}_n - \boldsymbol{U}\right)\boldsymbol{e} = \sigma\boldsymbol{K}\boldsymbol{e}.
\end{aligned}$$

Hence, we have

$$\hat{\sigma}^2 = \frac{\sigma^2}{n-p}\boldsymbol{e}'\boldsymbol{K}'\boldsymbol{K}\boldsymbol{e}.$$

Further, since $\boldsymbol{U}$ is a projection matrix, it follows that $\boldsymbol{U}$ is idempotent, and moreover, $\boldsymbol{U}$ is, by definition, symmetric. Thus, $\boldsymbol{K}$ will inherit these properties as well, since $(\boldsymbol{I}_n - \boldsymbol{U})(\boldsymbol{I}_n - \boldsymbol{U}) = \boldsymbol{I}_n - 2\boldsymbol{U} + \boldsymbol{U}\boldsymbol{U}$ and, trivially, $\boldsymbol{I}_n - \boldsymbol{U}$ is symmetric if $\boldsymbol{U}$ is symmetric. Due to this we arrive at

$$\hat{\sigma}^2 = \frac{\sigma^2}{n-p}\boldsymbol{e}'\boldsymbol{K}\boldsymbol{e},$$

which is exactly (5). Continuing, recall from above that $\mathrm{tr}(\boldsymbol{U}) = p$, which immediately yields $\mathrm{tr}(\boldsymbol{K}) = \mathrm{tr}(\boldsymbol{I}_n - \boldsymbol{U}) = \mathrm{tr}(\boldsymbol{I}_n) - \mathrm{tr}(\boldsymbol{U}) = n - p$. Finally, an application of [8, Thm. 1.6], which is also stated without proof in [1], gives us

$$\begin{aligned}\mathrm{Var}(\hat{\sigma}^2|\boldsymbol{X}, \boldsymbol{\Sigma}) &= \frac{\sigma^4}{(n-p)^2}\,\mathrm{Var}(\boldsymbol{e}'\boldsymbol{K}\boldsymbol{e}|\boldsymbol{X}, \boldsymbol{\Sigma}) \\ &= \frac{\sigma^4}{n-p}\left(2 + \frac{\mu_4 - 3}{n-p}\sum_{i=1}^{n}\boldsymbol{K}_{ii}^2\right).\end{aligned}$$

which corresponds to the result of Proposition 1. □

Remark 1. *Note that for the application of [8, Thm. 1.6] in the proof of Proposition 1 we need not assume that the error terms have common (conditional) third moments. This can be seen by inspecting the proof of [8, Thm. 1.6] and noting that any parts involving the third moments disappear since the first moment is 0.*

Further, note that it is possible to generalize Proposition 1 in the obvious way when the error terms have finite, but not necessarily common, fourth moments. This however is not very illuminating and is hence omitted.

3 Results: bounds on the variance of $\hat{\sigma}^2$

Based on Proposition 1 it is natural to approach $\boldsymbol{K}$ in order to obtain bounds on the variance of $\hat{\sigma}^2$. If we exploit the properties of $\boldsymbol{K}$ directly we arrive at the following naive bounds:

Corollary 1. *If the assumptions of Proposition 1 hold: Then*

$$\mathrm{Var}(\hat{\sigma}^2|\boldsymbol{X}, \boldsymbol{\Sigma}), \mathrm{Var}(\hat{\sigma}^2) \in \left[\frac{2\sigma^4}{n-p}, \frac{\sigma_4(\mu_4 - 1)}{n-p}\right].$$

Proof. (Proof of Corollary 1) Since $\boldsymbol{K}$ is idempotent and symmetric it follows that

$$\boldsymbol{K}_{ii} = \boldsymbol{K}_{ii}^2 + \sum_{j \neq i}\boldsymbol{K}_{ij}^2,$$

and in turn that

$$0 \leq \sum_{i=1}^{n}\boldsymbol{K}_{ii}^2 \leq \sum_{i=1}^{n}\boldsymbol{K}_{ii} = n - p.$$

This together with (4) directly yields

$$\frac{2\sigma^4}{n-p} \leq \mathrm{Var}(\hat{\sigma}^2|\boldsymbol{X}, \boldsymbol{\Sigma}) \leq \frac{\sigma^4(\mu_4 - 1)}{n-p}.$$

It now trivially follows that the inequality holds for $\mathrm{Var}(\hat{\sigma}^2)$ as well. □

Note that the bounds in Corollary 1 are not sufficient in order to show that the variance of $\sqrt{n}\hat{\sigma}^2$ will converge to a limit point. However, by using that $\boldsymbol{K} = \boldsymbol{I}_n - \boldsymbol{U}$ and exploiting the properties of $\boldsymbol{U}$ the lower variance bound may be tightened:

Corollary 2. *If the assumptions of Proposition 1 hold: Then*

$$\mathrm{Var}(\hat{\sigma}^2|\boldsymbol{X}, \boldsymbol{\Sigma}), \mathrm{Var}(\hat{\sigma}^2) \in [\nu_n - \kappa_n, \nu_n]$$

with $\nu_n := \sigma^4 \frac{\mu_4 - 1}{n-p}$ *and* $\kappa_n := \sigma^4 \frac{(\mu_4 - 3)p}{(n-p)^2}$.

Proof. (Proof of Corollary 2) Since $\boldsymbol{K} = \boldsymbol{I}_n - \boldsymbol{U}$ we can rewrite (4) as follows

$$\mathrm{Var}(\hat{\sigma}^2|\boldsymbol{X}, \boldsymbol{\Sigma}) = \frac{\sigma^4}{n-p}\left(2 + \frac{\mu_4 - 3}{n-p}\sum_{i=1}^{n}(1 - \boldsymbol{U}_{ii})^2\right).$$

Expanding the square and noting that $\sum_{i=1}^{n} \boldsymbol{U}_{ii} = p$ yields

$$\mathrm{Var}(\hat{\sigma}^2|\boldsymbol{X}, \boldsymbol{\Sigma}) = \frac{\sigma^4}{n-p}\left(\mu_4 - 1 - \frac{(\mu_4 - 3)p}{n-p} + \frac{\mu_4 - 3}{n-p}\sum_{i=1}^{n}\boldsymbol{U}_{ii}^2\right).$$

Now, as before for $\boldsymbol{K}$, since $\boldsymbol{U}$ is idempotent and symmetric we know that

$$0 \leq \sum_{i=1}^{n} \boldsymbol{U}_{ii}^2 \leq \sum_{i=1}^{n} \boldsymbol{U}_{ii} = p, \tag{6}$$

and therefore

$$\frac{\sigma^4(\mu_4 - 1)}{n-p} - \frac{\sigma^4(\mu_4 - 3)p}{(n-p)^2} \leq \mathrm{Var}(\hat{\sigma}^2|\boldsymbol{X}, \boldsymbol{\Sigma}) \leq \frac{\sigma^4(\mu_4 - 1)}{n-p}.$$

As before the inequality trivially follows for $\mathrm{Var}(\hat{\sigma}^2)$. □

We can also state a finite sample upper bound on the difference between the conditional and unconditional variances together with convergence of these using the bounds in Corollary 2.

Corollary 3. *If the assumptions of Proposition 1 hold: Then*

$$|\operatorname{Var}(\hat{\sigma}^2|\boldsymbol{X},\boldsymbol{\Sigma}) - \operatorname{Var}(\hat{\sigma}^2)| \leq \kappa_n, \text{ for all } n,$$

and

$$n\operatorname{Var}(\hat{\sigma}^2) \to \nu, \quad n\operatorname{Var}(\hat{\sigma}^2|\boldsymbol{X},\boldsymbol{\Sigma}) \to \nu, \quad \textit{uniformly as } n \to \infty$$

where $\nu = \sigma^4(\mu_4 - 1)$.

Proof. (Proof of Corollary 3) The first part follows trivially from Corollary 2, but we also get that

$$n\operatorname{Var}(\hat{\sigma}^2|\boldsymbol{X},\boldsymbol{\Sigma}) \in [n\nu_n - n\kappa_n, n\nu_n]$$

and, since $\lim_n n\nu_n = \nu$ and $\lim_n n\kappa_n = 0$, it follows that

$$\lim_{n\to\infty} n\operatorname{Var}(\hat{\sigma}^2|\boldsymbol{X},\boldsymbol{\Sigma}) = \nu,$$

due to the assumptions on $\boldsymbol{X}$ and $\boldsymbol{\Sigma}$. That is, $n\operatorname{Var}(\hat{\sigma}^2|\boldsymbol{X},\boldsymbol{\Sigma}) \to \nu$ uniformly. By the same argument it follows that $n\operatorname{Var}(\hat{\sigma}^2) \to \nu$ uniformly in n. □

Remark 2. *One example of a more restricted sub-class of the GLM from* (1) *is when we explicitly include an intercept, i.e.* $\widetilde{\boldsymbol{X}}$ *contains a column of ones. In [8, Eq. (10.12)] it is shown that in this situation it holds that*

$$1/n \leq \boldsymbol{U}_{ii} \leq 1$$

which gives us that the lower bound in (6) *can be tightened according to*

$$\frac{1}{n} \leq \sum_{i=1}^{n} \boldsymbol{U}_{ii}^2.$$

Hence, in this situation we can tighten the lower bound in Corollary 2 to $\nu_n - \kappa_n + \frac{\mu_4 - 3}{(n-p)n}$.

Another natural sub-class of models contained in the GLM from (1) *is to assume that* $\boldsymbol{\Sigma}$ *is diagonal. In this situation it holds that* $\operatorname{diag}(\boldsymbol{U}) = \operatorname{diag}(\boldsymbol{P})$ *where*

$$\boldsymbol{P} := \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}$$

is the projection matrix of the general linear model in (1). *Thus, assuming diagonal* $\boldsymbol{\Sigma}$ *does not make any parts of the variance formula, or its bounds, any more explicit.*

Acknowledgements

The authors acknowledge beneficial discussion with Rolf Sundberg.

References

[1] Atiqullah, M. (1962). The estimation of residual variance in quadratically balanced least-squares problems and the robustness of the F-test. *Biometrika, 49*(1-2), pp. 83–91.

[2] Atiqullah, M. (1962). On the effect of non-normality on the estimation of components of variance. *Journal of the Royal Statistical Society. Series B (Methodological)*, 140–147.

[3] Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton: Princeton university press.

[4] Eaton, M.L. (1983). Multivariate Statistics: A Vector Space Approach. Wiley, New York.

[5] Mathai, A.M., Provost, S.B., & Hayakawa, T. (2012). *Bilinear Forms and Zonal Polynomials* (Vol. 102). Springer Science & Business Media.

[6] Rao, C.R. (1970) Estimation of Heteroscedastic Variances in Linear Models. *Journal of the American Statistical Association, 65*(329), pp. 161–172.

[7] Rao, C.R. (1971) Minimum Variance Quadratic Unbiased Estimation of Variance Components. *Journal of Multivariate Analysis, 1*, pp. 445–456.

[8] Seber, G., & Lee, A. (2003). *Linear Regression Analysis* (2nd ed.). John Wiley & Sons, Inc.