On the variance parameter estimator in general linear models

Mathias Lindholm[1] [iD] · **Felix Wahl[1]**

Abstract

In the present note we consider general linear models where the covariates may be both random and non-random, and where the only restrictions on the error terms are that they are independent and have finite fourth moments. For this class of models we analyse the variance parameter estimator. In particular we obtain finite sample size bounds for the variance of the variance parameter estimator which are independent of covariate information regardless of whether the covariates are random or not. For the case with random covariates this immediately yields bounds on the unconditional variance of the variance estimator—a situation which in general is analytically intractable. The situation with random covariates is illustrated in an example where a certain vector autoregressive model which appears naturally within the area of insurance mathematics is analysed. Further, the obtained bounds 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. By using the derived bounds it is simple to show convergence in mean square of the variance parameter estimator for both random and non-random covariates. Moreover, the derivation of the bounds for the above general linear model is based on a lemma which applies in greater generality. This is illustrated by applying the used techniques to a class of mixed effects models.

Keywords General linear models · Non-Gaussian error terms · Moments of variance parameter estimators · Finite sample size bounds · Random covariates · Unconditional bounds

1 The general linear model

The primary model class that will be analysed is defined as follows: Let y be a random $n \times 1$ vector and let X be a random $n \times p_x$ matrix of almost surely full column rank,

✉ Mathias Lindholm
  lindholm@math.su.se

[1] Stockholms Universitet Matematiska Institutionen, Stockholm, Sweden

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

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

where $\boldsymbol{\beta}$ is a $p_x \times 1$ 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 X and $\mathbf{\Sigma}$, mean $\mathbf{0}$ and covariance $\boldsymbol{I}$, together with common central fourth moments $\mu_4 \geq 1$ (since $\mu_4 := \mathbb{E}[e_i^4] \geq \text{Var}(e_i)^2 = 1$). Here $\boldsymbol{I}$ denotes the $n \times n$ identity matrix. The standard generalized least squares estimator of $\boldsymbol{\beta}$, conditional on X and $\mathbf{\Sigma}$, is given by:

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

see for instance (Seber and Lee 2003, Sec. 3.10) for this and more on the general linear model. Moreover, an unbiased estimator of σ^2 (conditional on X and $\mathbf{\Sigma}$), and the estimator we will focus on in the present note, is given by:

$$\hat{\sigma}^2 := \hat{\sigma}_n^2(X, \mathbf{\Sigma}) = \frac{1}{n - p_x}\left(y - X\hat{\boldsymbol{\beta}}\right)' \mathbf{\Sigma}^{-1}\left(y - X\hat{\boldsymbol{\beta}}\right). \tag{2}$$

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_x, for instance simply n. It is important to note that when X is assumed to be random it is assumed to be 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. In Example 3 we comment on the situation when the regression coefficients are allowed to be random instead of the covariates, i.e. we consider mixed effects models.

For the results below it will be useful to define

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

which corresponds to the projection matrix associated with the linear model (1), and to also define the idempotent matrices

$$V := \mathbf{\Sigma}^{-1/2}P\mathbf{\Sigma}^{1/2},$$

and

$$K := I - V.$$

In particular, using the above it is possible to rewrite $\hat{\sigma}^2$ according to

$$\hat{\sigma}^2 = \frac{\sigma^2}{p_k} e' K e, \tag{3}$$

where $p_k = \text{rank}(K) = n - p_x$, see the proof of Proposition 1 for more details.

We will henceforth focus on properties of the variance of the variance parameter estimator $\hat{\sigma}^2$. One can 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. Cramér (1946, Eq. 27.4.2). Other similar results are obtained in the theory of minimum variance component estimation, see e.g. Rao (1970, 1971) and the proof of Seber and Lee (2003, Thm. 3.4). More general results can be found in for instance (Dette et al. 1998) where estimation of the variance parameter in the case of nonparametric regression is treated. Lemma 1 may be seen as a special case of the corresponding expression for the mean squared error (MSE) from Dette et al. (1998, Eq. 6). In Example 3 we discuss extensions to mixed effects models and comment on the results in Li (2012) which extend the analysis in Dette et al. (1998) w.r.t. mixed effects. We will return to these comparisons in more detail below.

The general problem formulation above, of course, relies on the theory of random quadratic forms. For more on this topic, see e.g. Eaton (1983), Mathai et al. (2012) and Seber and Lee (2003) and the references therein.

The results that we obtain for the variance of the variance estimator are based on that $\text{Var}(\hat{\sigma}^2 | X, \Sigma)$ can be calculated explicitly. For the particular situation of interest this variance is obtained using the following result from Plackett (1960, Eq. (2), p. 16) which we state in the following lemma:

Lemma 1 (Plackett 1960) *Let z be an $n \times 1$ dimensional vector of independent random variables with mean 0, and common variance σ^2, and common fourth central moment μ_4, and let W denote an arbitrary $n \times n$ matrix. It then holds that*

$$\text{Var}(z' W z) = \sigma^4 \left(2 \, \text{tr}(W^2) + (\mu_4 - 3) \sum_{i=1}^{n} W_{ii}^2 \right). \tag{4}$$

N.B. The last sum in (4) corresponds to the sum of the squared diagonal elements of W, which should not be confused with $\text{tr}(W^2)$. The proof of Lemma 1 is given in Plackett (1960), and a more general version can be found, without proof, in Atiqullah (1962a), whose proof can be found in Seber and Lee (2003, Thm. 1.6). See also the derivation of the SE expressions given in Dette et al. (1998) and Li (2012, Lemma 3).

Further, the main objective of the current note is to obtain finite sample bounds for the variance of the variance estimator of the general linear model defined by (1) when the covariates may be both random and non-random. In order to prove such bounds we will make use of the following lemma:

Lemma 2 *Given that W from Lemma 1 is idempotent and symmetric it follows that the variance expression (4) may be bounded according to*

$$\begin{cases} \operatorname{Var}(z'Wz) \in [v_n - \kappa_n, v_n] & \text{if } \mu_4 > 3, \\ \operatorname{Var}(z'Wz) \in [v_n, v_n - \kappa_n] & \text{if } 1 \le \mu_4 \le 3, \end{cases}$$

where $v_n := \sigma^4(\mu_4 - 1)(n - p_u)$ *and*

$$\kappa_n := \begin{cases} \sigma^4(\mu_4 - 3)p_u & \text{if } p_u \le n/2, \\ \sigma^4(\mu_4 - 3)(n - p_u) & \text{if } p_u > n/2, \end{cases}$$

and where $U = I - W$ *and* $p_u = \operatorname{rank}(U)$.

Note that Lemma 2 only relies on Lemma 1 in terms of the explicit form of the variance given by (4), a fact which will be exploited further in Example 3 given below. The usefulness of Lemma 2 becomes apparent when the decomposition of W is in terms of a U with p_u being a constant (much) smaller than n. This is what will be exploited in Corollary 1 and 2 , and which is the motivation for why the bounds in Lemma 2 are expressed in terms of p_u instead of $p_w := n - p_u$. Further, note that the split between $p_u \le n/2$ and $p_u > n/2$ ascertains that all bounds are positive.

Proof of Lemma 2 Since $W = I - U$ is idempotent it follows that $W^2 = W$, that $\operatorname{tr}(W) = \operatorname{rank}(W) = p_w = n - p_u$, and that (4) simplifies to

$$\operatorname{Var}(z'Wz) = \sigma^4 \left(2(n - p_u) + (\mu_4 - 3) \sum_{i=1}^{n} (1 - U_{ii})^2 \right).$$

Further, expanding the square and noting that $\sum_{i=1}^{n} U_{ii} = p_u$ yields

$$\operatorname{Var}(z'Wz) = \sigma^4 \left((\mu_4 - 1)(n - p_u) + (\mu_4 - 3) \left(\sum_{i=1}^{n} U_{ii}^2 - p_u \right) \right). \tag{5}$$

Now, since U is idempotent and symmetric it follows that

$$U_{ii} = U_{ii}^2 + \sum_{j \ne i} U_{ij}^2,$$

and in turn that

$$0 \le \sum_{i=1}^{n} U_{ii}^2 \le \sum_{i=1}^{n} U_{ii} = p_u. \tag{6}$$

Inserting the lower and upper bound into (5) finishes the proof of the lemma for $p_u \le n/2$. The corresponding bounds for $p_u > n/2$ are obtained by noting that $\sum W_{ii}^2 \ge 0$. □

We may now state our main result:

Proposition 1 *Consider the general linear model given by* (1) *and let*

(i) $\boldsymbol{\Sigma}$ *be a random symmetric almost surely positive definite* $n \times n$ *matrix and* $\boldsymbol{X}$ *be a random* $n \times p_x$ *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

$$
\begin{cases}
\text{Var}(\hat{\sigma}^2 | \boldsymbol{X}, \boldsymbol{\Sigma}) \in [\nu_n - \kappa_n, \nu_n] & \text{if } \mu_4 > 3, \\
\text{Var}(\hat{\sigma}^2 | \boldsymbol{X}, \boldsymbol{\Sigma}) \in [\nu_n, \nu_n - \kappa_n] & \text{if } 1 \leq \mu_4 \leq 3,
\end{cases}
$$

and

$$
\begin{cases}
\text{Var}(\hat{\sigma}^2) \in [\nu_n - \kappa_n, \nu_n] & \text{if } \mu_4 > 3, \\
\text{Var}(\hat{\sigma}^2) \in [\nu_n, \nu_n - \kappa_n] & \text{if } 1 \leq \mu_4 \leq 3,
\end{cases}
$$

where $\nu_n := \sigma^4 \frac{\mu_4 - 1}{n - p_x}$ *and*

$$
\kappa_n := \begin{cases}
\sigma^4 \frac{\mu_4 - 3}{(n - p_x)^2} p_x & \text{if } p_x \leq n/2, \\
\sigma^4 \frac{\mu_4 - 3}{n - p_x} & \text{if } p_x > n/2.
\end{cases}
$$

The proof of the *conditional* bounds in Proposition 1 is based on that $\hat{\sigma}^2$ may be expressed according to (3) together with an application of Lemmas 1 and 2. The details are given in the appendix. Moreover, note that the conditional variance bounds for $\hat{\sigma}^2$ are deterministic regardless of whether the covariates $\boldsymbol{X}$ are random or not. This fact combined with that $\hat{\sigma}^2$ is unbiased conditional on $\boldsymbol{X}$ and $\boldsymbol{\Sigma}$, together with an application of variance decomposition proves the *unconditional* variance bounds in Proposition 1. The full proof of Proposition 1 is given in the appendix.

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 Proposition 1.

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

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

and

$$
n \, \text{Var}(\hat{\sigma}^2) \to \nu, \quad n \, \text{Var}(\hat{\sigma}^2 | \boldsymbol{X}, \boldsymbol{\Sigma}) \to \nu, \quad \text{uniformly as } n \to \infty \qquad (7)
$$

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

Remark 1 It is of course possible to state the rate of convergence in (7) by noting that

$$
|(n - p_x) \text{Var}(\hat{\sigma}^2 | \boldsymbol{X}, \boldsymbol{\Sigma}) - \nu| \leq \frac{\sigma^4 |\mu_4 - 3|}{n - p_x} p_x,
$$

and

$$|(n - p_x)\operatorname{Var}(\hat{\sigma}^2) - \nu| \leq \frac{\sigma^4 |\mu_4 - 3|}{n - p_x} p_x,$$

hold for all n. Since, Corollary 1 is a convergence result in terms of n where p_x is a fixed quantity, the primary interest is on the situation where $p_x \leq n/2$.

By Corollary 1 it follows that $\operatorname{Var}(\hat{\sigma}^2) \to 0$, uniformly as $n \to \infty$, and therefore we can state the following result.

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

$$\hat{\sigma}^2 := \hat{\sigma}_n^2 \xrightarrow{L^2} \sigma^2,$$

as $n \to \infty$.

Remark 2 Note that $\hat{\sigma}^2$ from (2) is a special case of the variance parameter estimator $\hat{\sigma}_{ON}^2$ defined for the nonparametric mixed effects models analysed in Li (2012), see their Eq. (11). One can also note that the mixed effects models treated in Li (2012) is an extension of the model class treated in Dette et al. (1998). An alternative proof of the consistency of $\hat{\sigma}^2$ stated in Corollary 2 in the case with *fixed covariates*, hence, follows from Li (2012, Thm. 1) by neglecting the random effects part of their model. Further, as already commented upon above, in Dette et al. (1998) and Li (2012) focus is on the *MSE* of the variance parameter estimator. Moreover, their MSE expressions are stated in terms of asymptotic equivalence, i.e. using "$o(1/n)$". Thus, it is not straightforward to use their expressions to obtain bounds similar to those provided by Proposition 1 (or Lemma 2) without carefully inspecting the $o(1/n)$ terms.

2 Examples

Example 1 Consider the following vector autoregressive model:

$$X_{t+1} = X_t \boldsymbol{\beta}_t + \sigma_t \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{e}_{t+1}, \tag{8}$$

where all dimensions are in accordance with the linear model given by (1). Model (8) is closely connected to the distribution free Chain–Ladder model, which is a widely used actuarial reserving model, where t denotes time and X_t denotes the amount of payments made in the time interval $[0, t]$, see e.g. Mack (1993). More specifically, the Chain–Ladder model assumes that X_t is $n \times 1$, $(X_t)_i \geq 0$ and that $\boldsymbol{\Sigma}_t := \operatorname{diag}(X_t)$. In this situation we may analyse the variance of the variance parameter estimator, $\hat{\sigma}_t^2$, both conditional on X_t as well as unconditionally using Proposition 1 and its corollaries. In either situation the above results provide us with finite sample bounds as well as ascertaining that $\hat{\sigma}_t^2$ is a (mean square) consistent estimator of σ_t^2. A practical application of the finite sample bounds is for the Chain–Ladder model w.r.t. the discussion of the appropriateness of using conditional versus unconditional prediction error, see

e.g. Buchwalder et al. (2006) and Lindholm et al. (2019). Moreover, the relevance of using finite sample bounds is also apparent in many real-world insurance applications, when highly aggregated data is used, where the sample size often is around $n = 10$ and $p_x = 1 \ll n$.

Further, the Chain–Ladder model assumes a diagonal structure of $\boldsymbol{\Sigma}_t$ which, even though making $\mathrm{diag}(\boldsymbol{V}) = \mathrm{diag}(\boldsymbol{P})$, does not make the variance bounds any more explicit.

For more on other models than the distribution free Chain–Ladder model that are used in an insurance context, see e.g. Kremer (1984) and Lindholm et al. (2017).

Example 2 One example of a more restricted sub-class of the GLMs from (1) is when we assume that $\boldsymbol{\Sigma} = \boldsymbol{I}$ and explicitly include an intercept, i.e. $\boldsymbol{X}$ contains a column of ones. In this situation it is shown in Seber and Lee (2003, Eq. 10.12) that

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

which makes it is possible to tighten e.g. the lower bound in Lemma 2 when $\mu_4 > 3$ and, hence, tighten the corresponding bound in Proposition 1 to $\nu_n - \kappa_n + \frac{\mu_4 - 3}{(n-p)n}$, and analogously for the upper bound when $\mu_4 \leq 3$.

Example 3 In order to illustrate the usefulness of the techniques of the present paper we will now exploit the essentially model-free aspects of Lemma 1 and, in particular, Lemma 2. As already noted after stating Lemma 2, the results of Lemma 2 are purely algebraic, *given* that you somehow have arrived at a variance expression of the form (4). Purely as an illustration, consider the following special case of a mixed effects model introduced in Atiqullah (1962b):

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\theta} + \boldsymbol{B}\boldsymbol{\tau} + \boldsymbol{\epsilon}, \tag{9}$$

where $\boldsymbol{y}$ is $n \times 1$, and where $\boldsymbol{\theta}$ is $p_a \times 1$, and $\boldsymbol{\epsilon}$ is $n \times 1$, are independent random vectors whose components all have mean 0 and variances σ_θ^2 and σ_ϵ^2, respectively, together with the kurtoses $\mu_4^\theta (= \gamma_{2\theta} + 3)$ and $\mu_4^\epsilon (= \gamma_{2\epsilon} + 3)$, and where $\boldsymbol{\tau}$ is a $p_b \times 1$ vector of fixed effects. Further, $\boldsymbol{A}$ and $\boldsymbol{B}$ are assumed to be of full column rank p_a and p_b, respectively. Note that compared with (1) the model given by (9) is defined in terms of $\boldsymbol{\Sigma} = \boldsymbol{I}$ in accordance with Atiqullah (1962b). Further, the following unbiased estimators of σ_θ^2 and σ_ϵ^2 are given in Atiqullah (1962b):

$$\hat{\sigma}_\theta^2 = \frac{p_h(M_h - M_r)}{\mathrm{tr}(\boldsymbol{U})}, \tag{10}$$

$$\hat{\sigma}_\epsilon^2 = M_r, \tag{11}$$

where

$$M_h = \frac{\boldsymbol{y}'\boldsymbol{H}\boldsymbol{y}}{p_h}, \quad M_r = \frac{\boldsymbol{y}'\boldsymbol{R}\boldsymbol{y}}{p_r},$$

and where

$$G = B(B'B)^{-1}B',$$
$$L = I - G,$$
$$H = LA(A'LA)^{-1}A'L,$$
$$R = I - G - H,$$
$$U = A'HA = A'LA,$$

and where p_g, p_h and p_r denote the rank of G, H and R. Moreover, in Atiqullah (1962b) it is stated that the introduced variance parameter estimators have the following variances:

$$\text{Var}(\hat{\sigma}_\theta^2) = \frac{2\sigma_\epsilon^4((p_r p_h + p_h^2) + p_r\xi(\xi\,\text{tr}(U^2) + 2\,\text{tr}(U)))}{p_r\,\text{tr}(U)^2}$$
$$\times \left(1 + \frac{(\mu_4^\epsilon - 3)(p_r\boldsymbol{h} - p_h\boldsymbol{r})'(p_r\boldsymbol{h} - p_h\boldsymbol{r}) + (\mu_4^\theta - 3)U'U p_r^2\xi^2}{2p_r((p_r p_h + p_h^2) + p_r\xi(\xi\,\text{tr}(U^2) + 2\,\text{tr}(U)))}\right)$$
(12)

$$\text{Var}(\hat{\sigma}_\epsilon^2) = \frac{2\sigma_\epsilon^4}{p_r}\left(1 + \frac{1}{2}(\mu_4^\epsilon - 3)\frac{\boldsymbol{r}'\boldsymbol{r}}{p_r}\right),$$
(13)

where $\xi = \sigma_\theta^2/\sigma_\epsilon^2$ and where $\boldsymbol{h}$ and $\boldsymbol{r}$ corresponds to the vectors of the diagonal elements of H and R. Regarding $\hat{\sigma}_\epsilon^2$ we may start off by noting that (13) may be re-written according to

$$\text{Var}(p_r\hat{\sigma}_\epsilon^2) = \sigma_\epsilon^4\left(2p_r + (\mu_4^\epsilon - 3)\sum_{i=1}^{n} R_{ii}^2\right)$$

which is on the same form as (4) from Lemma 1, since R is idempotent, and Lemma 2 applies. Moreover, by using the same arguments as those used in the proof of Proposition 1, it follows that $p_r = n - p_a - p_b$, where both p_a and p_b are constants, thus ascertaining L^2-consistency and that the corollaries of Proposition 1 hold as well.

Concerning the variance of $\hat{\sigma}_\theta^2$, the expression of (12) can not be approached using Lemma 2. It is, however, possible to make use of trace inequalities to show that when e.g. $\mu_4^\theta, \mu_4^\epsilon > 3$

$$\text{Var}(\hat{\sigma}_\theta^2) \geq \frac{2\sigma_\theta^4}{p_u} > 0$$
(14)

where $p_u := \text{rank}(U)$ is a constant, see the appendix. That is, $\text{Var}(\hat{\sigma}_\theta^2)$ is bounded from below by a positive constant for all n, and $\hat{\sigma}_\theta^2$ is, hence, not L^2-consistent. It is, however, possible to obtain sharper finite sample bounds on $\text{Var}(\hat{\sigma}_\theta^2)$ by using other

trace inequalities—a matter not pursued further in the present note. For other examples and a deeper discussion concerning the performance of estimators of variance components, see e.g. Christensen (2019, Ex. 5.1.1 and Sec. 5.4).

Even though the above mixed effects example is only intended for illustration purposes, without any claim of practical relevance, it is still worth commenting on the relation to the results in Li (2012). In Li (2012) mixed effects nonparametric regression is considered w.r.t. the *MSE* of the *total* variance parameter estimator. That is, the situation with $\sigma_\epsilon = \sigma_\theta$ is considered. Consequently, the decomposition from Atiqullah (1962b) is not covered as a special case. Moreover, by setting $\sigma_\epsilon = \sigma_\theta$ above, the resulting estimator is neither contained in the estimators covered in Li (2012, Thm. 1). For more on differences compared with Li (2012) (and Dette et al. 1998), see Remark 2 above.

Compliance with ethical standards

Appendix A. Proofs

Proof of Proposition 1 We start by reformulating the linear model of (1) as follows:

$$\widetilde{y} = \widetilde{X}\beta + \sigma e,$$

where $\widetilde{y} := \Sigma^{-1/2} y$ and $\widetilde{X} := \Sigma^{-1/2} X$. Now let $\hat{\widetilde{e}} := \widetilde{y} - \widetilde{X}\hat{\beta}$, it then follows that

$$\hat{\sigma}^2 = \frac{1}{n - p_x} \hat{\widetilde{e}}' \hat{\widetilde{e}},$$

which may be expressed in terms of e by noting that

$$
\begin{aligned}
\hat{\widetilde{e}} &= \left(\widetilde{X}\beta + \sigma e\right) - V\widetilde{y} \\
&= \widetilde{X}\beta + \sigma e - V\left(\widetilde{X}\beta + \sigma e\right) \\
&= \sigma\left(I - V\right)e = \sigma K e.
\end{aligned}
$$

Hence, we have

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

Further, since V is idempotent and symmetric it follows that K will inherit these properties as well and we arrive at

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

Finally, since V is idempotent we know that $\text{rank}(V) = p_v = \text{tr}(V)$, where

$$\begin{aligned}
\text{tr}(V) &= \text{tr}(\Sigma^{-1/2} P \Sigma^{1/2}) \\
&= \text{tr}(P \Sigma^{1/2} \Sigma^{-1/2}) \\
&= \text{tr}(P),
\end{aligned}$$

and by repeating this argument if follows that $\text{tr}(P) = \text{tr}(I_{p_x}) = p_x$, that is,

$$\hat{\sigma}^2 = \frac{\sigma^2}{p_k} e' K e.$$

This ascertains that the conditions of Lemmas 1 and 2 are fulfilled, which proves the unconditional part of Proposition 1.

The unconditional variance bounds are obtained by first noting that $\hat{\sigma}^2$ is unbiased and then applying a variance decomposition. This argument of course holds also for biased estimators as long as $\mathbb{E}[\hat{\sigma}^2|X, \Sigma]$ is nonrandom, e.g. when the normalization constant is n and not $n - p_x$. □

Proof of Corollary 1 The first part follows trivially from Proposition 1, but we also get that

$$\begin{cases}
n \, \text{Var}(\hat{\sigma}^2|X, \Sigma) \in [n\nu_n - n\kappa_n, n\nu_n] & \text{if } \mu_4 > 3, \\
n \, \text{Var}(\hat{\sigma}^2|X, \Sigma) \in [n\nu_n, n\nu_n - n\kappa_n] & \text{if } \mu_4 \leq 3,
\end{cases}$$

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

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

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

Proof of (14) We will now show that $\hat{\sigma}_\theta^2$ is not in general L^2 consistent. If we restrict our attention to $\mu_4^\epsilon, \mu_4^\theta \geq 3$ it follows that (12) may be bounded according to

$$\text{Var}(\hat{\sigma}_\theta^2) \geq \frac{2\sigma_\epsilon^4((p_r p_h + p_h^2) + p_r \xi(\xi \, \text{tr}(U^2) + 2 \, \text{tr}(U)))}{p_r \, \text{tr}(U)^2},$$

since the second factor consists of quadratic forms and non-negative functions/constants. Further, note that by construction it is assumed that $(A'LA)^{-1}$ exists, i.e. $A'LA$ is positive definite and of full column rank $p_a = \text{rank}(A)$, from which it follows that $0 < p_h = p_u = p_a = O(1)$. Moreover, an application of the inequality

$$\frac{\text{tr}(U)^2}{\text{tr}(U^2)} \leq p_u,$$

see e.g. Wolkowicz and Styan (1980, Eq. 2.35), yields

$$\text{Var}(\hat{\sigma}_\theta^2) \geq 2\sigma_\epsilon^4 \left(\frac{p_r p_h + p_h^2}{p_r \text{tr}(U)^2} + \frac{\xi^2}{p_u} + \frac{2\xi}{\text{tr}(U)} \right).$$

Furthermore, since $p_r \geq 0$ it follows that

$$\text{Var}(\hat{\sigma}_\theta^2) \geq \frac{2\sigma_\epsilon^4 \xi^2}{p_u} = \frac{2\sigma_\theta^4}{p_u} > 0,$$

i.e. $\text{Var}(\hat{\sigma}_\theta^2)$ is bounded from below by a constant value greater than 0. $\square$

References

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

Atiqullah M (1962b) On the effect of non-normality on the estimation of components of variance. J R Stat Soc Ser B (Methodol) 24:140–147

Buchwalder M, Bühlmann H, Merz M, Wüthrich MV (2006) The mean square error of prediction in the chain ladder reserving method (Mack and Murphy revisited). ASTIN Bull 36(2):521–542

Christensen R (2019) Advanced linear modeling: statistical learning and dependent data, 3rd edn. Springer, New York

Cramér H (1946) Mathematical methods of statistics. Princeton University Press, Princeton

Dette H, Munk A, Wagner T (1998) Estimating the variance in nonparametric regression—What is a reasonable choice? J R Stat Soc Ser B (Stat Methodol) 60(4):751–764

Eaton ML (1983) Multivariate statistics: a vector space approach. Wiley, New York

Kremer E (1984) A class of autoregressive models for predicting the final claims amount. Insur Math Econ 3(2):111–119

Li Z (2012) A comparison of error variance estimates in nonparametric mixed models. Commun Stat Theory Methods 41(4):778–790

Lindholm M, Lindskog F, Wahl F (2017) Valuation of non-life liabilities from claims triangles. Risks 5(3):39

Lindholm M, Lindskog F, Wahl F (2019) Estimation of conditional mean squared error of prediction for claims reserving. Ann Actuar Sci. https://doi.org/10.1017/S174849951900006X

Mack T (1993) Distribution-free calculation of the standard error of chain ladder reserve estimates. ASTIN Bull 23(2):213–225

Mathai AM, Provost SB, Hayakawa T (2012) Bilinear forms and zonal polynomials, vol 102. Springer, Berlin

Plackett RL (1960) Principles of regression analysis. Clarendon Press, Oxford

Rao CR (1970) Estimation of heteroscedastic variances in linear models. J Am Stat Assoc 65(329):161–172

Rao CR (1971) Minimum variance quadratic unbiased estimation of variance components. J Multivar Anal 1:445–456

Seber G, Lee A (2003) Linear regression analysis, 2nd edn. Wiley, Hoboken

Wolkowicz H, Styan GPH (1980) Bounds for eigenvalues using traces. Linear Algebra Appl 29:471–506

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.