Loss Prediction by Generalized Least Squares

Leigh J. Halliwell

Discussion by Klaus D. Schmidt

Abstract

In a recent paper on loss reserving, Halliwell suggests to predict outstanding claims by the method of generalized least squares applied to a linear model. An example is the linear model given by

$$E[Z_{i,k}] = \mu + \alpha_i + \gamma_k$$

where $Z_{i,k}$ is the total claim amount of all claims which occur in year i and are settled in year $i+k$. The predictor proposed by Halliwell is known in econometrics but it is perhaps not well-known to actuaries. The present discussion completes and simplifies the argument used by Halliwell to justify the predictor; in particular, it is shown that there is no need to consider conditional distributions.

1 Loss reserving

For $i, k \in \{0, 1, \ldots, n\}$, let $Z_{i,k}$ denote the total claim amount of all claims which occur in year i and are settled in year $i+k$. We assume that the *incremental claims* $Z_{i,k}$ are observable for $i + k \leq n$ and that they are non-observable for $i + k > n$. The observable incremental claims are represented by the *run-off triangle*:

Occurrence year	Development year 0	1	...	k	...	$n-i$	...	$n-1$	n
0	$Z_{0,0}$	$Z_{0,1}$	...	$Z_{0,k}$	...	$Z_{0,n-i}$	...	$Z_{0,n-1}$	$Z_{0,n}$
1	$Z_{1,0}$	$Z_{1,1}$	...	$Z_{1,k}$	...	$Z_{1,n-i}$	...	$Z_{1,n-1}$	
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$			
i	$Z_{i,0}$	$Z_{i,1}$	...	$Z_{i,k}$	...	$Z_{i,n-i}$			
$\vdots$	$\vdots$	$\vdots$		$\vdots$					
$n-k$	$Z_{n-k,0}$	$Z_{n-k,1}$	...	$Z_{n-k,k}$					
$\vdots$	$\vdots$	$\vdots$							
$n-1$	$Z_{n-1,0}$	$Z_{n-1,1}$							
n	$Z_{n,0}$								

The non-observable incremental claims are to be predicted from the observable ones. Whether or not certain predictors are preferable to others depends on the stochastic mechanism generating the data. It is thus necessary to first formulate a stochastic model and to fix the properties the predictors should have.

For example, we may assume that the incremental claims satisfy the *linear model* given by

$$E[Z_{i,k}] \quad = \quad \mu + \alpha_i + \gamma_k$$

with real parameters $\mu, \alpha_0, \alpha_1, \ldots, \alpha_n, \gamma_0, \gamma_1, \ldots, \gamma_n$ such that $\alpha_0 = 0 = \gamma_0$. This means that the expected incremental claims are determined by an overall mean μ and corrections α_i and γ_k depending on the *occurrence year* i and the *development year* k, respectively.

2 The linear model with missing observations

The model considered in the previous section is a special case of the linear model considered by Halliwell [1996]:

Let $\mathbf{Y}$ be an $(m \times 1)$ random vector satisfying

$$E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$$

and

$$\mathrm{Var}[\mathbf{Y}] = \mathbf{S}$$

for some known $(m \times k)$ design matrix $\mathbf{X}$, some unknown $(k \times 1)$ parameter vector $\boldsymbol{\beta}$, and some known $(m \times m)$ matrix $\mathbf{S}$ which is assumed to be positive definite.

We assume that some but not all coordinates of $\mathbf{Y}$ are observable. Without loss of generality, we may and do assume that the first p coordinates of $\mathbf{Y}$ are observable while the last $q := m - p$ coordinates of $\mathbf{Y}$ are non–observable. We may thus write

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}$$

where $\mathbf{Y}_1$ consists of the observable coordinates of $\mathbf{Y}$ and $\mathbf{Y}_2$ consists of the non–observable coordinates of $\mathbf{Y}$. Accordingly, we partition the design matrix $\mathbf{X}$ into

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$$

We assume that

$$\mathrm{rank}(\mathbf{X}_1) = k \leq p$$

Then $\mathbf{X}$ has full rank and $\mathbf{X}'\mathbf{X}$ is invertible.

Following Halliwell, we partition $\mathbf{S}$ into

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix}$$

where

$$\begin{aligned} \mathbf{S}_{11} &:= \mathrm{Cov}[\mathbf{Y}_1, \mathbf{Y}_1] = \mathrm{Var}[\mathbf{Y}_1] \\ \mathbf{S}_{12} &:= \mathrm{Cov}[\mathbf{Y}_1, \mathbf{Y}_2] \\ \mathbf{S}_{21} &:= \mathrm{Cov}[\mathbf{Y}_2, \mathbf{Y}_1] \\ \mathbf{S}_{22} &:= \mathrm{Cov}[\mathbf{Y}_2, \mathbf{Y}_2] = \mathrm{Var}[\mathbf{Y}_2] \end{aligned}$$

Then $\mathbf{S}_{11}$ and $\mathbf{S}_{22}$ are positive definite, and we also have $\mathbf{S}_{21}' = \mathbf{S}_{12}$. Moreover, $\mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$ is positive definite. Then $\mathbf{S}_{11}$ and $\mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$ are invertible and there exist invertible matrices $\mathbf{A}$ and $\mathbf{D}$ satisfying

$$\mathbf{A}'\mathbf{A} = \mathbf{S}_{11}^{-1}$$

and

$$\mathbf{D}'\mathbf{D} = (\mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12})^{-1}$$

Define

$$\mathbf{C} := -\mathbf{D}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}$$

and let

$$\mathbf{W} := \begin{pmatrix} \mathbf{A} & \mathbf{O} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}$$

Then we have

$$\mathbf{W}'\mathbf{W} = \mathbf{S}^{-1}$$

In the following sections, we study the problem of estimating $\boldsymbol{\beta}$ and of predicting $\mathbf{Y}_2$ by estimators or predictors based on $\mathbf{Y}_1$.

3 Estimation

Let us first consider the problem of estimating $\boldsymbol{\beta}$.

A random vector $\widehat{\boldsymbol{\beta}}$ with values in $\mathbf{R}^k$ is a

- *linear estimator* (of $\boldsymbol{\beta}$) if it satisfies $\widehat{\boldsymbol{\beta}} = \mathbf{B}\mathbf{Y}_1$ for some matrix $\mathbf{B}$, it is an
- *unbiased estimator* (of $\boldsymbol{\beta}$) if it satisfies $E[\widehat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$, and it is an
- *admissible estimator* (of $\boldsymbol{\beta}$) if it is linear and unbiased.

A linear estimator $\widehat{\boldsymbol{\beta}} = \mathbf{B}\mathbf{Y}_1$ of $\boldsymbol{\beta}$ is unbiased if and only if $\mathbf{B}\mathbf{X}_1 = \mathbf{I}_k$.

A particular admissible estimator of $\boldsymbol{\beta}$ is the *Gauss–Markov estimator* $\boldsymbol{\beta}^*$ which is defined as

$$\boldsymbol{\beta}^* := (\mathbf{X}_1'\mathbf{S}_{11}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{S}_{11}^{-1}\mathbf{Y}_1$$

Among all admissible estimators of $\boldsymbol{\beta}$, the Gauss–Markov estimator is distinguished by the following properties:

3.1 Theorem (Gauss–Markov Theorem). *The Gauss–Markov estimator $\boldsymbol{\beta}^*$ satisfies*

$$\mathrm{Var}[\boldsymbol{\beta}^*] = (\mathbf{X}_1'\mathbf{S}_{11}^{-1}\mathbf{X}_1)^{-1}$$

Moreover, for each admissible estimator $\widehat{\boldsymbol{\beta}}$, the matrix

$$\mathrm{Var}[\widehat{\boldsymbol{\beta}}] - \mathrm{Var}[\boldsymbol{\beta}^*]$$

is positive semidefinite.

In a sense, the Gauss–Markov Theorem asserts that the Gauss–Markov estimator has minimal variance among all admissible estimators of $\boldsymbol{\beta}$. Since

$$\begin{aligned} E\Big[(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})'(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})\Big] &= E\Big[\mathrm{tr}\Big((\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})'(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})\Big)\Big] \\ &= E\Big[\mathrm{tr}\Big((\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})'\Big)\Big] \\ &= \mathrm{tr}\Big(E\Big[(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})'\Big]\Big) \\ &= \mathrm{tr}\Big(\mathrm{Var}[\widehat{\boldsymbol{\beta}}]\Big) \end{aligned}$$

we see that the Gauss–Markov estimator also minimizes the *expected quadratic estimation error* over all admissible estimators of $\boldsymbol{\beta}$.

4 Prediction

Let us now turn to the problem of predicting $\mathbf{Y}_2$.

A random vector $\widehat{\mathbf{Y}}_2$ with values in $\mathbf{R}^q$ is a

- *linear predictor* (of $\mathbf{Y_2}$) if it satisfies $\widehat{\mathbf{Y}}_2 = \mathbf{QY_1}$ for some matrix $\mathbf{Q}$, it is an
- *unbiased predictor* (of $\mathbf{Y_2}$) if it satisfies $E[\widehat{\mathbf{Y}}_2] = E[\mathbf{Y_2}]$, and it is an
- *admissible predictor* (of $\mathbf{Y_2}$) if it is linear and unbiased.

A linear predictor $\widehat{\mathbf{Y}}_2 = \mathbf{QY_1}$ of $\mathbf{Y_2}$ is unbiased if and only if $\mathbf{QX_1} = \mathbf{X_2}$.

For an admissible estimator $\widehat{\boldsymbol{\beta}}$, define

$$\mathbf{y_2}(\widehat{\boldsymbol{\beta}}) \quad := \quad \mathbf{X_2}\widehat{\boldsymbol{\beta}} - \mathbf{D}^{-1}\mathbf{C}(\mathbf{Y_1}-\mathbf{X_1}\widehat{\boldsymbol{\beta}})$$

and

$$\boldsymbol{h}(\widehat{\boldsymbol{\beta}}) \quad := \quad -(\mathbf{CX_1}+\mathbf{DX_2})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) + (\mathbf{Ce_1}+\mathbf{De_2})$$

Then $\mathbf{y_2}(\widehat{\boldsymbol{\beta}})$ is an admissible predictor of $\mathbf{Y_2}$.

Following Halliwell, we have the following result:

4.1 Lemma. *The identities*

$$\mathbf{Y_2} \quad = \quad \mathbf{y_2}(\widehat{\boldsymbol{\beta}}) + \mathbf{D}^{-1}\boldsymbol{h}(\widehat{\boldsymbol{\beta}})$$

as well as

$$E[\boldsymbol{h}(\widehat{\boldsymbol{\beta}})] \quad = \quad \mathbf{0}$$

and

$$\mathrm{Var}[\boldsymbol{h}(\widehat{\boldsymbol{\beta}})] \quad = \quad (\mathbf{CX_1}+\mathbf{DX_2})\,\mathrm{Var}[\widehat{\boldsymbol{\beta}}]\,(\mathbf{CX_1}+\mathbf{DX_2})' + \mathbf{I_q}$$

hold for each admissible estimator $\widehat{\boldsymbol{\beta}}$; in particular, the matrix

$$\mathrm{Var}[\boldsymbol{h}(\widehat{\boldsymbol{\beta}})] - \mathrm{Var}[\boldsymbol{h}(\boldsymbol{\beta}^*)]$$

is positive semidefinite.

From the last assertion of Lemma 4.1, which is a consequence of the Gauss–Markov theorem, Halliwell concludes that the *Gauss–Markov predictor* $\mathbf{y_2}(\boldsymbol{\beta}^*)$ is the best unbiased linear predictor of $\mathbf{Y_2}$. This conclusion, however, is not justified in his paper. A partial justification is given by the following lemma:

4.2 Lemma. *For each admissible estimator $\widehat{\boldsymbol\beta}$, the matrix*

$$\mathrm{Var}\Big[\mathbf{Y}_2-\mathbf{y}_2(\widehat{\boldsymbol\beta})\Big]-\mathrm{Var}\Big[\mathbf{Y}_2-\mathbf{y}_2(\boldsymbol\beta^*)\Big]$$

is positive semidefinite.

Proof. Since $\mathbf{y}_2(\widehat{\boldsymbol\beta})$ is an unbiased predictor of $\mathbf{Y}_2$, we have

$$\begin{aligned}\mathrm{Var}\Big[\mathbf{Y}_2-\mathbf{y}_2(\widehat{\boldsymbol\beta})\Big] &= E\Big[\Big(\mathbf{Y}_2-\mathbf{y}_2(\widehat{\boldsymbol\beta})\Big)\Big(\mathbf{Y}_2-\mathbf{y}_2(\widehat{\boldsymbol\beta})\Big)'\Big]\\ &= E\Big[(\mathbf{D}^{-1}\boldsymbol h(\widehat{\boldsymbol\beta}))\,(\mathbf{D}^{-1}\boldsymbol h(\widehat{\boldsymbol\beta}))'\Big]\\ &= \mathbf{D}^{-1}E\Big[\boldsymbol h(\widehat{\boldsymbol\beta})\,(\boldsymbol h(\widehat{\boldsymbol\beta}))'\Big](\mathbf{D}^{-1})'\\ &= \mathbf{D}^{-1}\mathrm{Var}[\boldsymbol h(\widehat{\boldsymbol\beta})](\mathbf{D}^{-1})'\end{aligned}$$

Now the assertion follows from Lemma 4.1. □

We may even push the discussion a bit further: Why should we confine ourselves to predictors which can be written as $\mathbf{y}_2(\widehat{\boldsymbol\beta})$ for some admissible estimator $\widehat{\boldsymbol\beta}$? There may be other unbiased linear predictors $\widehat{\mathbf{Y}}_2$ for which

$$\mathrm{Var}\Big[\mathbf{Y}_2-\mathbf{y}_2(\boldsymbol\beta^*)\Big]-\mathrm{Var}\Big[\mathbf{Y}_2-\widehat{\mathbf{Y}}_2\Big]$$

and hence

$$\mathrm{Var}\Big[\mathbf{Y}_2-\mathbf{y}_2(\widehat{\boldsymbol\beta})\Big]-\mathrm{Var}\Big[\mathbf{Y}_2-\widehat{\mathbf{Y}}_2\Big]$$

is positive semidefinite. The following result improves Lemma 4.2:

4.3 Theorem. *For each admissible predictor $\widehat{\mathbf{Y}}_2$, the matrix*

$$\mathrm{Var}\Big[\mathbf{Y}_2-\widehat{\mathbf{Y}}_2\Big]-\mathrm{Var}\Big[\mathbf{Y}_2-\mathbf{y}_2(\boldsymbol\beta^*)\Big]$$

is positive semidefinite.

Proof. Consider a matrix $\mathbf{Q}$ satisfying

$$\widehat{\mathbf{Y}}_2 = \mathbf{Q}\mathbf{Y}_1$$

and hence $\mathbf{Q}\mathbf{X}_1=\mathbf{X}_2$. Letting

$$\mathbf{Q}^* := \mathbf{S}_{21}\mathbf{S}_{11}^{-1}+(\mathbf{X}_2-\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{X}_1)(\mathbf{X}_1'\mathbf{S}_{11}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{S}_{11}^{-1}$$

we obtain

$$\begin{aligned}
\mathbf{y}_2(\boldsymbol{\beta}^*) &= \mathbf{X}_2\boldsymbol{\beta}^* - \mathbf{D}^{-1}\mathbf{C}(\mathbf{Y}_1 - \mathbf{X}_1\boldsymbol{\beta}^*) \\
&= \mathbf{X}_2\boldsymbol{\beta}^* + \mathbf{S}_{21}\mathbf{S}_{11}^{-1}(\mathbf{Y}_1 - \mathbf{X}_1\boldsymbol{\beta}^*) \\
&= \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{Y}_1 + (\mathbf{X}_2 - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{X}_1)\boldsymbol{\beta}^* \\
&= \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{Y}_1 + (\mathbf{X}_2 - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{X}_1)(\mathbf{X}_1'\mathbf{S}_{11}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{S}_{11}^{-1}\mathbf{Y}_1 \\
&= \mathbf{Q}^*\mathbf{Y}_1
\end{aligned}$$

Since $\mathbf{Q}^*\mathbf{X}_1 = \mathbf{X}_2 = \mathbf{Q}\mathbf{X}_1$, we have

$$\begin{aligned}
&\mathrm{Cov}\Big[\mathbf{Y}_2 - \mathbf{y}_2(\boldsymbol{\beta}^*)\,,\, \mathbf{y}_2(\boldsymbol{\beta}^*) - \widehat{\mathbf{Y}}_2\Big] \\
&= \mathrm{Cov}\Big[\mathbf{Y}_2 - \mathbf{Q}^*\mathbf{Y}_1\,,\, \mathbf{Q}^*\mathbf{Y}_1 - \mathbf{Q}\mathbf{Y}_1\Big] \\
&= (\mathbf{S}_{21} - \mathbf{Q}^*\mathbf{S}_{11})(\mathbf{Q}^* - \mathbf{Q})' \\
&= -(\mathbf{X}_2 - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{X}_1)(\mathbf{X}_1'\mathbf{S}_{11}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{Q}^* - \mathbf{Q})' \\
&= -(\mathbf{X}_2 - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{X}_1)(\mathbf{X}_1'\mathbf{S}_{11}^{-1}\mathbf{X}_1)^{-1}(\mathbf{Q}^*\mathbf{X}_1' - \mathbf{Q}\mathbf{X}_1)' \\
&= \mathbf{O}
\end{aligned}$$

and hence

$$\begin{aligned}
\mathrm{Var}\Big[\mathbf{Y}_2 - \widehat{\mathbf{Y}}_2\Big] &= \mathrm{Var}\Big[\big(\mathbf{Y}_2 - \mathbf{y}_2(\boldsymbol{\beta}^*)\big) + \big(\mathbf{y}_2(\boldsymbol{\beta}^*) - \widehat{\mathbf{Y}}_2\big)\Big] \\
&= \mathrm{Var}\Big[\mathbf{Y}_2 - \mathbf{y}_2(\boldsymbol{\beta}^*)\Big] + \mathrm{Var}\Big[\mathbf{y}_2(\boldsymbol{\beta}^*) - \widehat{\mathbf{Y}}_2\Big]
\end{aligned}$$

The assertion follows. □

Theorem 4.3 asserts that the Gauss–Markov predictor minimizes the variance of the prediction error over all admissible predictors of $\mathbf{Y}_2$. Since

$$E\Big[\big(\mathbf{Y}_2 - \widehat{\mathbf{Y}}_2\big)'\big(\mathbf{Y}_2 - \widehat{\mathbf{Y}}_2\big)\Big] = \mathrm{tr}\Big(\mathrm{Var}\Big[\mathbf{Y}_2 - \widehat{\mathbf{Y}}_2\Big]\Big)$$

we see that the Gauss–Markov predictor also minimizes the *expected quadratic prediction error* over all admissible predictors of $\mathbf{Y}_2$.

5 A related optimization problem

To complete the discussion of the predictor proposed by Halliwell, we consider the following optimization problem:

Minimize

$$E\Big[\big(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}\big)'\mathbf{S}^{-1}\big(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}\big)\Big]$$

over all admissible estimators $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$.

We thus aim at minimizing an objective function in which there is no discrimination between the observable and the non–observable part of $\mathbf{Y}$; this distinction, however, is present in the definition of an admissible estimator.

Because of $\mathbf{S}^{-1} = \mathbf{W}'\mathbf{W}$ and the structure of $\mathbf{W}$, it is easy to see that the objective function of the optimization problem can be decomposed into an approximation part and a prediction part:

5.1 Lemma. *The identity*

$$\begin{aligned} & E\Big[\big(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}\big)'\mathbf{S}^{-1}\big(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}\big)\Big] \\ &= E\Big[\big(\mathbf{Y}_1-\mathbf{X}_1\widehat{\boldsymbol{\beta}}\big)'\mathbf{S}_{11}^{-1}\big(\mathbf{Y}_1-\mathbf{X}_1\widehat{\boldsymbol{\beta}}\big)\Big] \\ &\quad + E\Big[\big(\mathbf{Y}_2-\mathbf{y}_2(\widehat{\boldsymbol{\beta}})\big)'\mathbf{D}'\mathbf{D}\big(\mathbf{Y}_2-\mathbf{y}_2(\widehat{\boldsymbol{\beta}})\big)\Big] \end{aligned}$$

holds for each admissible estimator $\widehat{\boldsymbol{\beta}}$.

Moreover, using similar arguments as before, the three expectations occurring in Lemma 5.1 can be represented as follows:

5.2 Theorem. *The identities*

$$\begin{aligned} & E\Big[\big(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}\big)'\mathbf{S}^{-1}\big(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}\big)\Big] \\ &= (p+q) - 2k + \operatorname{tr}\Big((\mathbf{W}\mathbf{X})\operatorname{Var}[\widehat{\boldsymbol{\beta}}]\,(\mathbf{W}\mathbf{X})'\Big) \end{aligned}$$

as well as

$$\begin{aligned} & E\Big[\big(\mathbf{Y}_1-\mathbf{X}_1\widehat{\boldsymbol{\beta}}\big)'\mathbf{S}_{11}^{-1}\big(\mathbf{Y}_1-\mathbf{X}_1\widehat{\boldsymbol{\beta}}\big)\Big] \\ &= p - 2k + \operatorname{tr}\Big((\mathbf{A}\mathbf{X}_1)\operatorname{Var}[\widehat{\boldsymbol{\beta}}]\,(\mathbf{A}\mathbf{X}_1)'\Big) \end{aligned}$$

and

$$\begin{aligned} & E\Big[\big(\mathbf{Y_2}-\mathbf{y_2}(\widehat{\boldsymbol\beta})\big)'\mathbf{D}'\mathbf{D}\big(\mathbf{Y_2}-\mathbf{y_2}(\widehat{\boldsymbol\beta})\big)\Big] \\ &= q+\mathrm{tr}\Big((\mathbf{CX_1}+\mathbf{DX_2})\,\mathrm{Var}[\widehat{\boldsymbol\beta}]\,(\mathbf{CX_1}+\mathbf{DX_2})'\Big) \end{aligned}$$

hold for each admissible estimator $\widehat{\boldsymbol\beta}$.

Because of Theorem 5.2, each of the three expectations occurring in Lemma 5.1 is minimized by the Gauss–Markov estimator $\boldsymbol\beta^*$. We have thus again justified the restriction to predictors of $\mathbf{Y_2}$ which can be written as $\mathbf{y_2}(\widehat{\boldsymbol\beta})$ for some admissible estimator $\widehat{\boldsymbol\beta}$.

The technical details concerning the proofs of the results of this section can be found in Schmidt [1998].

6 Conditioning

Following the example of $\mathbf{Y}$ having a multivariate normal distribution, Halliwell uses arguments related to the conditional distribution of $\mathbf{Y_2}$ with respect to $\mathbf{Y_1}$; in particular, he claims that $\mathbf{y_2}(\boldsymbol\beta^*)$ is the conditional expectation $E(\mathbf{Y_2}|\mathbf{Y_1})$ of $\mathbf{Y_2}$ with respect to $\mathbf{Y_1}$. This is not true in general: Without particular assumptions on the distribution of $\mathbf{Y}$, the conditional expectation $E(\mathbf{Y_2}|\mathbf{Y_1})$ may fail to be linear in $\mathbf{Y_1}$, and the unbiased linear predictor of $\mathbf{Y_2}$ based on $\mathbf{Y_1}$ minimizing the expected quadratic loss may fail to be the conditional expectation $E(\mathbf{Y_2}|\mathbf{Y_1})$.

Moreover, since the identities of Lemma 4.1 hold for each admissible estimator $\widehat{\boldsymbol\beta}$ (and not only for the Gauss–Markov estimator $\boldsymbol\beta^*$), Halliwell's arguments (on p. 482 of his paper) suggest that each admissible estimator $\widehat{\boldsymbol\beta}$ satisfies

$$E(\mathbf{Y_2}|\mathbf{Y_1}) = \mathbf{X_2}\widehat{\boldsymbol\beta}-\mathbf{D}^{-1}\mathbf{C}(\mathbf{Y_1}-\mathbf{X_1}\widehat{\boldsymbol\beta})$$

and

$$\mathrm{Var}(\mathbf{Y_2}|\mathbf{Y_1}) = \mathbf{D}^{-1}\mathrm{Var}[\boldsymbol{h}(\widehat{\boldsymbol\beta})](\mathbf{D}^{-1})'$$

Again, this cannot be true since in both cases the left hand side depends only on $\mathbf{Y}_1$ whereas the right hand side also varies with the matrix $\mathbf{B}$ defining the admissible estimator $\widehat{\boldsymbol{\beta}} = \mathbf{B}\mathbf{Y}_1$.

More generally: When only unconditional moments of the distribution of the random vector $\mathbf{Y}$ are specified, it is impossible to obtain any conclusions concerning the conditional distribution of its non–observable part $\mathbf{Y}_2$ with respect to its observable part $\mathbf{Y}_1$.

7 Remarks

Traditional least squares theory aims at minimizing the *quadratic loss*

$$\left(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}\right)'\mathbf{S}^{-1}\left(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}\right)$$

where all coordinates of $\mathbf{Y}$ are observable. It also involves considerations concerning the variance of $\widehat{\boldsymbol{\beta}}$ and usually handles prediction as a separate problem which has to be solved after estimating $\boldsymbol{\beta}$.

In Section 5 of the present paper, we proposed instead to minimize the *expected quadratic loss*

$$E\left[\left(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}\right)'\mathbf{S}^{-1}\left(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}\right)\right]$$

where some but not all of the coordinates of $\mathbf{Y}$ are observable and the admissible estimators of $\boldsymbol{\beta}$ are unbiased and linear in the observable part $\mathbf{Y}_1$ of $\mathbf{Y}$. This approach has several advantages:

- The expected quadratic loss can be expressed in terms of $\mathrm{var}[\widehat{\boldsymbol{\beta}}]$ such that minimization of the expected quadratic loss and minimization of $\mathrm{var}[\widehat{\boldsymbol{\beta}}]$ turns out to be the same problem; see Theorem 5.2.
- The expected quadratic loss can be decomposed in a canonical way into an approximation part and a prediction part such that the expected quadratic loss and its two components are simultaneously minimized by the Gauss–Markov estimator; see Lemma 5.1.

- Inserting the Gauss–Markov estimator in the prediction part of the expected quadratic loss provides an unbiased linear predictor for the non–observable part $\mathbf{Y}_2$ of $\mathbf{Y}$.

We thus obtain the predictor proposed by Halliwell [1996] by a direct approach which avoids conditioning. This predictor was first proposed by Goldberger [1962]; see also Rao and Toutenburg [1995; Theorem 6.2].

Acknowledgement. I would like to thank Michael D. Hamer who provided Theorem 4.3 and its proof.

References

Goldberger, A. S. [1962]: *Best linear unbiased prediction in the generalized linear regression model.* J. Amer. Statist. Assoc. **57**, 369–375.

Halliwell, L. J. [1996]: *Loss prediction by generalized least squares.* Proc. Casualty Actuar. Soc. **83**, 436–489.

Rao, C. R., and Toutenburg, H. [1995]: *Linear Models: Least Squares and Alternatives.* Berlin – Heidelberg – New York: Springer.

Schmidt, K. D. [1998]: *Prediction in the linear model – A direct approach.* Metrika **48**, 141–147.

Klaus D. Schmidt
Lehrstuhl für Versicherungsmathematik
Technische Universität Dresden
D–01062 Dresden
Germany

e–mail: klausd@math.tu-dresden.de

October 13, 1999