DISCUSSION OF PAPER PUBLISHED IN VOLUME LXXXIII

LOSS PREDICTION BY GENERALIZED LEAST SQUARES

LEIGH J. HALLIWELL

DISCUSSION BY KLAUS D. SCHMIDT

Abstract

In a recent paper on loss reserving, Halliwell suggests predicting 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* (Table 1).

The non-observable incremental claims are to be predicted from the observable ones. Whether or not certain predictors are

TABLE 1

Occurrence Year	Development Year 0	1	$\ldots$	k	$\ldots$	$n-i$	$\ldots$	$n-1$	n
0	$Z_{0,0}$	$Z_{0,1}$	$\ldots$	$Z_{0,k}$	$\ldots$	$Z_{0,n-i}$	$\ldots$	$Z_{0,n-1}$	$Z_{0,n}$
1	$Z_{1,0}$	$Z_{1,1}$	$\ldots$	$Z_{1,k}$	$\ldots$	$Z_{1,n-i}$	$\ldots$	$Z_{1,n-1}$	
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$			
i	$Z_{i,0}$	$Z_{i,1}$	$\ldots$	$Z_{i,k}$	$\ldots$	$Z_{i,n-i}$			
$\vdots$	$\vdots$	$\vdots$		$\vdots$					
$n-k$	$Z_{n-k,0}$	$Z_{n-k,1}$	$\ldots$	$Z_{n-k,k}$					
$\vdots$	$\vdots$	$\vdots$							
$n-1$	$Z_{n-1,0}$	$Z_{n-1,1}$							
n	$Z_{n,0}$								

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

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

with real parameters $\mu, \alpha_0, \alpha_1, \ldots, \alpha_n, \gamma_0, \gamma_1, \ldots, \gamma_n$ such that $\sum_{i=0}^{n} \alpha_i = 0 = \sum_{k=0}^{n} \gamma_k$. 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 [2]:

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

$$\mathbf{W} := \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}.$$

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 $\hat{\boldsymbol{\beta}}$ with values in $\mathbf{R}^k$ is

– a *linear estimator* (of $\boldsymbol{\beta}$) if it satisfies $\hat{\boldsymbol{\beta}} = \mathbf{B}\mathbf{Y}_1$ for some matrix $\mathbf{B}$,

– an *unbiased estimator* (of $\boldsymbol{\beta}$) if it satisfies $E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$, and

– an *admissible estimator* (of $\boldsymbol{\beta}$) if it is linear and unbiased.

A linear estimator $\hat{\boldsymbol{\beta}} = \mathbf{B}\mathbf{Y}_1$ 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 due to the *Gauss–Markov Theorem*:

THEOREM 3.1 *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 $\hat{\boldsymbol{\beta}}$, the matrix

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

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

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

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 $\hat{\mathbf{Y}}_2$ with values in $\mathbf{R}^q$ is

– a *linear predictor* (of $\mathbf{Y}_2$) if it satisfies $\hat{\mathbf{Y}}_2 = \mathbf{Q}\mathbf{Y}_1$ for some matrix $\mathbf{Q}$,

– an *unbiased predictor* (of $\mathbf{Y}_2$) if it satisfies $E[\hat{\mathbf{Y}}_2] = E[\mathbf{Y}_2]$, and

– an *admissible predictor* (of $\mathbf{Y}_2$) if it is linear and unbiased.

A linear predictor $\hat{\mathbf{Y}}_2 = \mathbf{Q}\mathbf{Y}_1$ of $\mathbf{Y}_2$ is unbiased if and only if $\mathbf{Q}\mathbf{X}_1 = \mathbf{X}_2$.

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

$$\mathbf{Y}_2(\hat{\boldsymbol{\beta}}) := \mathbf{X}_2\hat{\boldsymbol{\beta}} - \mathbf{D}^{-1}\mathbf{C}(\mathbf{Y}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}})$$

and

$$\mathbf{h}(\hat{\boldsymbol{\beta}}) := -(\mathbf{C}\mathbf{X}_1 + \mathbf{D}\mathbf{X}_2)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + (\mathbf{C}\mathbf{e}_1 + \mathbf{D}\mathbf{e}_2),$$

where $\mathbf{e}_1 := \mathbf{Y}_1 - \mathbf{X}_1\boldsymbol{\beta}$ and $\mathbf{e}_2 := \mathbf{Y}_2 - \mathbf{X}_2\boldsymbol{\beta}$. Then $\mathbf{Y}_2(\hat{\boldsymbol{\beta}})$ is an admissible predictor of $\mathbf{Y}_2$.

Following Halliwell, we have the following

LEMMA 4.1 *The identities*

$$\mathbf{Y}_2 = \mathbf{Y}_2(\hat{\boldsymbol{\beta}}) + \mathbf{D}^{-1}\mathbf{h}(\hat{\boldsymbol{\beta}})$$

as well as

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

and

$$\mathrm{Var}[\mathbf{h}(\hat{\boldsymbol{\beta}})] = (\mathbf{C}\mathbf{X}_1 + \mathbf{D}\mathbf{X}_2)\mathrm{Var}[\hat{\boldsymbol{\beta}}](\mathbf{C}\mathbf{X}_1 + \mathbf{D}\mathbf{X}_2)' + \mathbf{I}_q$$

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

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

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 *For each admissible estimator $\hat{\boldsymbol{\beta}}$, the matrix*

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

is positive semidefinite.

The proof of this lemma is that since $\mathbf{Y}_2(\hat{\boldsymbol{\beta}})$ is an unbiased predictor of $\mathbf{Y}_2$, we have

$$\begin{aligned}\operatorname{Var}[\mathbf{Y}_2 - \mathbf{Y}_2(\hat{\boldsymbol{\beta}})] &= E[(\mathbf{Y}_2 - \mathbf{Y}_2(\hat{\boldsymbol{\beta}}))(\mathbf{Y}_2 - \mathbf{Y}_2(\hat{\boldsymbol{\beta}}))'] \\ &= E[(\mathbf{D}^{-1}\mathbf{h}(\hat{\boldsymbol{\beta}}))(\mathbf{D}^{-1}\mathbf{h}(\hat{\boldsymbol{\beta}}))'] \\ &= \mathbf{D}^{-1}E[\mathbf{h}(\hat{\boldsymbol{\beta}})(\mathbf{h}(\hat{\boldsymbol{\beta}}))'](\mathbf{D}^{-1})' \\ &= \mathbf{D}^{-1}\operatorname{Var}[\mathbf{h}(\hat{\boldsymbol{\beta}})](\mathbf{D}^{-1})'.\end{aligned}$$

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 $\mathbf{Y}_2(\hat{\boldsymbol{\beta}})$ for some admissible estimator $\hat{\boldsymbol{\beta}}$? There may be other unbiased linear predictors $\hat{\mathbf{Y}}_2$ for which

$$\operatorname{Var}[\mathbf{Y}_2 - \mathbf{Y}_2(\boldsymbol{\beta}^*)] - \operatorname{Var}[\mathbf{Y}_2 - \hat{\mathbf{Y}}_2]$$

and hence

$$\operatorname{Var}[\mathbf{Y}_2 - \mathbf{Y}_2(\hat{\boldsymbol{\beta}})] - \operatorname{Var}[\mathbf{Y}_2 - \hat{\mathbf{Y}}_2]$$

is positive semidefinite. The following result improves Lemma 4.2:

THEOREM 4.3 *For each admissible predictor $\hat{\mathbf{Y}}_2$, the matrix*

$$\operatorname{Var}[\mathbf{Y}_2 - \hat{\mathbf{Y}}_2] - \operatorname{Var}[\mathbf{Y}_2 - \mathbf{Y}_2(\boldsymbol{\beta}^*)]$$

is positive semidefinite.

A proof of this theorem can also be presented. Consider a matrix $\mathbf{Q}$ satisfying

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

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

$$\begin{aligned}
&\mathrm{Cov}[\mathbf{Y}_2 - \mathbf{Y}_2(\boldsymbol{\beta}^*), \mathbf{Y}_2(\boldsymbol{\beta}^*) - \hat{\mathbf{Y}}_2] \\
&\quad = \mathrm{Cov}[\mathbf{Y}_2 - \mathbf{Q}^*\mathbf{Y}_1, \mathbf{Q}^*\mathbf{Y}_1 - \mathbf{Q}\mathbf{Y}_1] \\
&\quad = (\mathbf{S}_{21} - \mathbf{Q}^*\mathbf{S}_{11})(\mathbf{Q}^* - \mathbf{Q})' \\
&\quad = -(\mathbf{X}_2 - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{X}_1)(\mathbf{X}_1'\mathbf{S}_{11}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{Q}^* - \mathbf{Q})' \\
&\quad = -(\mathbf{X}_2 - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{X}_1)(\mathbf{X}_1'\mathbf{S}_{11}^{-1}\mathbf{X}_1)^{-1}(\mathbf{Q}^*\mathbf{X}_1 - \mathbf{Q}\mathbf{X}_1)' \\
&\quad = \mathbf{0},
\end{aligned}$$

and hence

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

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

$$E[(\mathbf{Y}_2 - \hat{\mathbf{Y}}_2)'(\mathbf{Y}_2 - \hat{\mathbf{Y}}_2)] = \mathrm{tr}(\mathrm{Var}[\mathbf{Y}_2 - \hat{\mathbf{Y}}_2]),$$

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:

$$\text{Minimize} \qquad E[(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{S}^{-1}(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})]$$

over all admissible estimators $\hat{\boldsymbol{\beta}}$ 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:

LEMMA 5.1 *The identity*

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

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

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

THEOREM 5.2 *The identities*

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

as well as

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

and

$$E[(\mathbf{Y}_2 - \mathbf{Y}_2(\hat{\boldsymbol{\beta}}))'\mathbf{D}'\mathbf{D}(\mathbf{Y}_2 - \mathbf{Y}_2(\hat{\boldsymbol{\beta}}))]$$
$$= q + \mathrm{tr}((\mathbf{C}\mathbf{X}_1 + \mathbf{D}\mathbf{X}_2)\mathrm{Var}[\hat{\boldsymbol{\beta}}](\mathbf{C}\mathbf{X}_1 + \mathbf{D}\mathbf{X}_2)')$$

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

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 $\mathbf{Y}_2(\hat{\boldsymbol{\beta}})$ for some admissible estimator $\hat{\boldsymbol{\beta}}$.

The technical details concerning the proofs of the results of this section can be found in Schmidt [4].

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 $E(\mathbf{Y}_2 \mid \mathbf{Y}_1)$ 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 $E(\mathbf{Y}_2 \mid \mathbf{Y}_1)$ 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 $E(\mathbf{Y}_2 \mid \mathbf{Y}_1)$.

Moreover, since the identities of Lemma 4.1 hold for each admissible estimator $\hat{\boldsymbol{\beta}}$ (and not only for the Gauss–Markov estimator $\boldsymbol{\beta}^*$), Halliwell's arguments [2, p. 482] suggest that each admissible estimator $\hat{\boldsymbol{\beta}}$ satisfies

$$E(\mathbf{Y}_2 \mid \mathbf{Y}_1) = \mathbf{X}_2\hat{\boldsymbol{\beta}} - \mathbf{D}^{-1}\mathbf{C}(\mathbf{Y}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}})$$

and

$$\mathrm{Var}(\mathbf{Y}_2 \mid \mathbf{Y}_1) = \mathbf{D}^{-1}\mathrm{Var}[\mathbf{h}(\hat{\boldsymbol{\beta}})](\mathbf{D}^{-1})'.$$

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 $\hat{\boldsymbol{\beta}} = \mathbf{B}\mathbf{Y}_1$.

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 $\mathbf{Y}_1$.

REMARKS

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

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

where all coordinates of $\mathbf{Y}$ are observable. It also involves considerations concerning the variance of $\hat{\boldsymbol{\beta}}$, and it 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*

$$E[(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{S}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})],$$

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 var$[\hat{\boldsymbol{\beta}}]$ such that minimization of the expected quadratic loss and minimization of var$[\hat{\boldsymbol{\beta}}]$ 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 [2] by a direct approach which avoids conditioning. This predictor was first proposed by Goldberger [1] (see also Rao and Toutenburg [3; Theorem 6.2]).

ACKNOWLEDGEMENT

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

REFERENCES

[1] Goldberger, Arthur S., "Best Linear Unbiased Prediction in the Generalized Linear Regression Model," *Journal of the American Statistical Association* 57, 1962, pp. 369–375.

[2] Halliwell, Leigh J., "Loss Prediction by Generalized Least Squares," *PCAS* LXXXIII, 1996, pp. 436–489.

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

[4] Schmidt, Klaus D., "Prediction in the Linear Model: A Direct Approach," *Metrika* 48, 1998, pp. 141–147.