Linear Algebra and its Applications 388 (2004) 327–340

# On the Kalman filter with possibly degenerate and correlated errors

Debasis Sengupta

*Applied Statistics Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Kolkata 700108, India*

**Abstract**

The Kalman filter is a very popular tool for estimation and prediction in the context of a state-space model. Sometimes it is necessary to formulate the state-space model in such a way that the model errors are correlated. The error dispersion matrix may even be singular. In this paper we establish a connection between prediction in the state-space model in this general set-up, and estimation in the general linear model. Subsequently we use the update equations in the general linear model to derive a generalization of the Kalman filter.

*Keywords:* State space model; Linear model updates; Singular dispersion matrix; Best linear unbiased prediction; Linear zero function

## 1. Introduction

The *state-space model* is a versatile model for a sequence of vector observations. This model has many applications in such diverse areas as control theory, time series analysis and sample surveys (see [3,6,22]). The model is given by the recursive relation

$$x_t = B_t x_{t-1} + u_t, \tag{1.1}$$

$$z_t = H_t x_t + v_t, \tag{1.2}$$

for $t = 1, 2, \ldots$ In the above, the *state vector* $x_t$ is unobservable, but the *measurement vector* $z_t$ is observable. The error vectors $u_t$ and $v_t$ have zero mean with

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{u}_s, \boldsymbol{u}_t) &= \boldsymbol{Q}_u(s,t), \quad s,t = 1,2,\ldots, \\
\mathrm{Cov}(\boldsymbol{v}_s, \boldsymbol{v}_t) &= \boldsymbol{Q}_v(s,t), \quad s,t = 1,2,\ldots, \\
\mathrm{Cov}(\boldsymbol{u}_s, \boldsymbol{v}_t) &= \boldsymbol{Q}_{uv}(s,t), \quad s,t = 1,2,\ldots
\end{aligned}$$

Typically the objective is to predict the state vector $\boldsymbol{x}_t$ or the measurement vector $z_{t+1}$ by a linear function of the observations $z_1, z_2, \ldots, z_t$ and the initial state $\boldsymbol{x}_0$. The linear predictor should have the smallest possible mean squared prediction error. The matrices $\boldsymbol{Q}_u(s,t)$, $\boldsymbol{Q}_v(s,t)$ and $\boldsymbol{Q}_{uv}(s,t)$, $s,t = 1,2,\ldots$ are assumed to be known. The *state transition matrix* $\boldsymbol{B}_t$ and the *measurement matrix* $\boldsymbol{H}_t$, $t = 1,2,\ldots$ are also assumed to be known. The vector $\boldsymbol{x}_0$ may itself be an estimate, where the corresponding estimation error is absorbed in $\boldsymbol{u}_1$.

Note that the combined dispersion matrix of any subset of the errors $\boldsymbol{u}_t$ and $\boldsymbol{v}_t$, $t = 1,2,\ldots$ may be singular. The singularity may arise not only in the case of accurate measurement, but also due to the very nature of the application. See [7] for an illustration where estimation of the parameters of an autoregressive moving average model is done by using a state-space representation of the model, with singular $\boldsymbol{Q}_v(s,t)$.

The *Kalman filter* [11,12] provides a recursive solution to the prediction problem in the special case

$$\boldsymbol{Q}_{uv}(s,t) = 0 \ \forall s,t, \quad \boldsymbol{Q}_u(s,t) = \begin{cases} \boldsymbol{I} & \text{if } s = t, \\ \boldsymbol{0} & \text{otherwise,} \end{cases} \quad \boldsymbol{Q}_v(s,t) = \begin{cases} \boldsymbol{I} & \text{if } s = t, \\ \boldsymbol{0} & \text{otherwise.} \end{cases}$$

Nieto and Guerrero [16] derived a filter for the general case. Haslett [9] used updates in the linear model to derive the filter in the case of non-singular error dispersion matrix. In this article we strengthen a result due to Duncan and Horn [5] which links prediction in the state-space model with estimation in the fixed effects linear model. Subsequently we use this result, together with update formulae of the linear model given by Jammalamadaka and Sengupta [10], to provide an intuitive derivation of the best recursive linear predictor in the state-space model in the most general case.

The paper is organized as follows. Section 2 provides a summary of requisite results on updates in the general linear model, which are already available in the literature. Section 3 provides the explicit link between prediction in the state-space model and estimation in the fixed effects linear model. Section 4 provides the optimal prediction formulae for the general state-space model by building on the results of Sections 2 and 3.

We use the following notations throughout the paper: given a matrix $\boldsymbol{A}$, its transpose, rank, column space and generalized inverse are given by $\boldsymbol{A}'$, $\rho(\boldsymbol{A})$, $\mathscr{C}(\boldsymbol{A})$ and $\boldsymbol{A}^-$, respectively. The orthogonal projection matrix onto $\mathscr{C}(\boldsymbol{A})$ is denoted by $\boldsymbol{P}_A$, while the notation $\boldsymbol{P}_{A^\perp}$ is used in place of $\boldsymbol{I} - \boldsymbol{P}_A$, the orthogonal projection matrix onto $\mathscr{C}(\boldsymbol{A})^\perp$. We will use the number of observations explicitly as a subscript when needed, but drop the subscript whenever there is no scope for ambiguity.

## 2. Updates in the linear model

Consider the general linear model $(y, X\beta, \sigma^2 V)$ where the model matrix $X$ and the dispersion matrix $\sigma^2 V$ are possibly rank deficient. The linear zero functions (LZFs) of this model are functions of the form $l'y$ having zero expectation. The characterization of the best linear unbiased estimators (BLUEs) in this model as those homogeneous linear functions of $y$ which are uncorrelated with *all* LZFs is well-known (see [19]). Bhimasankaram and Sengupta [1] had used this fact to obtain a relatively uncommon but explicit form of the best linear unbiased estimator (BLUE) of $X\beta$. The expression follows from an adjustment for the covariance of $y$, a linear unbiased estimator (LUE) of $X\beta$, with $P_{X^\perp}\beta$, which is in some sense a generating set for all LZFs. The precise nature of the 'covariance adjustment' [19] is given by the following lemma.

**Lemma 2.1** [1]. *Let $z = (u' : v')'$ be a random vector having first and second order moments such that $E(v)$, the expected value of $v$, is contained in the column space of $D(v)$, the dispersion matrix of $v$. Then the linear compound $u + Bv$ is uncorrelated with $v$ if and only if $Bv = -\mathrm{Cov}(v, v)[D(v)]^- v$, where $\mathrm{Cov}(u, v) = E(uv') - E(u)E(v')$.*

By choosing $u = y$ and $v = P_{X^\perp}y$, we have the following expression for the BLUE of $X\beta$.

$$\widehat{X\beta} = \left[I - VP_{X^\perp}\{P_{X^\perp}VP_{X^\perp}\}^- P_{X^\perp}\right]y. \tag{2.1}$$

Note that this expression is invariant under the choice of the generalized inverse, as $P_{X^\perp}y$ almost surely belongs to the column space of its dispersion matrix, and the latter coincides with the column space of $P_{X^\perp}V$. Searle [20] arrived at a similar expression for the BLUE from other considerations. Formula (2.1) lends itself to simple derivation of some results such as (2.4) and (2.5). For the present article however, the method of covariance adjustment with LZFs is more important than the expression of $\widehat{X\beta}$ obtained by it.

The dispersion matrix of the BLUE of (2.1) is

$$D(\widehat{X\beta}) = \sigma^2\left[V - VP_{X^\perp}\{P_{X^\perp}VP_{X^\perp}\}^- P_{X^\perp}V\right]. \tag{2.2}$$

If the residual $y - \widehat{X\beta}$ is denoted by $e$, its dispersion matrix is

$$D(e) = \sigma^2 VP_{X^\perp}\{P_{X^\perp}VP_{X^\perp}\}^- P_{X^\perp}V. \tag{2.3}$$

The expressions of the dispersion matrices given in (2.2) and (2.3) do not depend on the choice of the generalized inverse, as the column spaces of $P_{X^\perp}VP_{X^\perp}$ and $P_{X^\perp}V$ are identical. The column spaces of the two dispersion matrices are as follows [1]:

$$\mathscr{C}(D(\widehat{X\beta})) = \mathscr{C}(X) \cap \mathscr{C}(V), \tag{2.4}$$

$$\mathscr{C}(D(e)) = \mathscr{C}(VP_{X^\perp}). \tag{2.5}$$

Jammalamadaka and Sengupta [10] derived updates in the linear model corresponding to data and model changes, using LZFs and the principle of covariance adjustment. Following their notations for data addition, we will explicitly use the number of observations as a subscript. Let us denote the model with $n$ observations by $\mathscr{M}_n = (y_n, X_n\beta, \sigma^2 V_n)$. If $n = m + l$ with $l > 0$, we can partition the matrices and vectors of $\mathscr{M}_n$ as follows:

$$y_n = \begin{pmatrix} y_m \\ y_l \end{pmatrix}, \quad X_n = \begin{pmatrix} X_m \\ X_l \end{pmatrix}, \quad V_n = \begin{pmatrix} V_m & V_{ml} \\ V'_{ml} & V_l \end{pmatrix}.$$

Consider the 'initial model' $\mathscr{M}_m = (y_m, X_m\beta, \sigma^2 V_m)$ which is augmented by the observation $y_l$ to produce the 'augmented model' $\mathscr{M}_n$. Each LZF in the model $\mathscr{M}_m$ is also an LZF in the model $\mathscr{M}_n$. The number of uncorrelated LZFs exclusive to the augmented model, which are all uncorrelated with the common LZFs, is $[\rho(X_n : V_n) - \rho(X_n)] - [\rho(X_m : V_m) - \rho(X_m)]$. The clue to the update relationships lies in the identification of these LZFs. This number can be written as $l_1 - l_2$ where $l_1 = \rho(X_n : V_n) - \rho(X_m : V_m)$ and $l_2 = \rho(X_n) - \rho(X_m)$, the latter being the number of estimable linearly independent linear parametric functions (LPF) which are exclusive to the augmented model such that no linear combination of these LPFs is estimable in the initial model. Note that $0 < l_2 \leqslant l_1 \leqslant l$. The following cases can arise.

A. $0 < l_2 = l_1$, that is, there are some additional estimable LPFs in the augmented model, but no new LZF.
B. $0 = l_2 < l_1$, that is, $\mathscr{C}(X'_l) \subseteq \mathscr{C}(X'_m)$. This case corresponds to some additional LZFs in the augmented model, but no new estimable LPF.
C. $0 = l_2 = l_1$, that is, there is no new LZF or estimable LPF.
D. $0 < l_2 < l_1$, that is, there are some additional LZFs as well as additional estimable LPFs in the augmented model.

In Case B there is a vector $w_l$ of LZFs in $\mathscr{M}_n$ such that (i) it is uncorrelated with $e_m$ and (ii) every LZF of $\mathscr{M}_n$ is a linear function of $e_m$ and $w_l$. The dispersion matrix of this vector must have rank $l_1 - l_2$. A choice of $w_l$ is

$$w_l = y_l - X_l\hat{\beta}_m - V'_{ml}V_m^-(y_m - X_m\hat{\beta}_m). \tag{2.6}$$

The expression on the right hand side is invariant under the choice of the generalized inverse, as $y_m - X_m\hat{\beta}_m$ almost surely belongs to the column space of its dispersion matrix, and the latter contains the column space of $V_{ml}$. The vector $w_l$ can be interpreted as (i) the prediction error of the best linear unbiased predictor (BLUP) of $y_l$ based on the model $\mathscr{M}_m$, and (ii) an unsealed recursive group residual for $y_l$ in the model $\mathscr{M}_n$ (see [4,13]).

The update equations of the BLUE of $X_m\beta$ and its dispersion in the different cases are given in [21]. A summary of the relevant results is given in the following theorem.

**Theorem 2.1.** *Under the above set-up, let* $X_m\hat{\boldsymbol\beta}_m$ *and* $X_m\hat{\boldsymbol\beta}_n$ *denote the BLUEs of* $X_m\boldsymbol\beta$ *under* $\mathscr{M}_m$ *and* $\mathscr{M}_n$*, respectively. Then*

(a) *In Case A,* $X_m\hat{\boldsymbol\beta}_n = X_m\hat{\boldsymbol\beta}_m$ *and* $D(X_m\hat{\boldsymbol\beta}_n) = D(X_m\hat{\boldsymbol\beta}_m)$*. The BLUE of* $X_l\boldsymbol\beta$ *under* $\mathscr{M}_n$ *and its dispersion are*

$$X_l\hat{\boldsymbol\beta}_n = y_l - V_{lm}V_m^-(y_m - X_m\hat{\boldsymbol\beta}_m),$$
$$D(X_l\hat{\boldsymbol\beta}_n) = \sigma^2 V_l - V_{lm}V_m^- D(y_m - X_m\hat{\boldsymbol\beta}_m)V_m^- V_{ml}.$$

(b) *In Case B,*

$$X_m\hat{\boldsymbol\beta}_n = X_m\hat{\boldsymbol\beta}_m - \mathrm{Cov}(X_m\hat{\boldsymbol\beta}_m, w_l)[D(w_l)]^- w_l,$$
$$D(X_m\hat{\boldsymbol\beta}_n) = D(X_m\hat{\boldsymbol\beta}_m) - \mathrm{Cov}(X_m\hat{\boldsymbol\beta}_m, w_l)[D(w_l)]^- \mathrm{Cov}(X_m\hat{\boldsymbol\beta}_m, w_l)'.$$

(c) *In Case C,* $X_m\hat{\boldsymbol\beta}_n = X_m\hat{\boldsymbol\beta}_m$ *and* $D(X_m\hat{\boldsymbol\beta}_n) = D(X_m\hat{\boldsymbol\beta}_m)$.

In Cases B and C, $\mathscr{C}(X_n') = \mathscr{C}(X_m')$. Therefore, Theorem 2.1 can be used to obtain the update of the BLUE of *any* estimable LPF and its dispersion. In Case D, the elements of $y_l$ can be permuted (together with those of $X_l$, $V_l$, and $V_{ml}$) in such a way that the data augmentation can be split into two stages, corresponding to Cases A and B, respectively. Thus, Theorem 2.1 can be used to handle this case too.

The explicit algebraic expressions for the updates are given by Pordzik [18] and Bhimasankaram et al. [2]. These expressions are somewhat complicated. Simpler expressions can be found in some special cases (see [8,14,15,17]).

## 3. Link between state-space and linear models

Duncan and Horn [5] showed that the minimum mean squared error linear predictor of $x_t$ in the state-space model (1.1) and (1.2) is given by the BLUE of a vector parameter in a fixed effects linear model. We now prove a stronger result with possibly singular dispersion matrices.

**Theorem 3.1.** *Let* $\boldsymbol h$ *be a known non-random vector and* $\boldsymbol x$ *and* $\boldsymbol z$ *be random vectors following the model*

$$\begin{pmatrix} h \\ z \end{pmatrix} = \begin{pmatrix} F \\ G \end{pmatrix} x + \begin{pmatrix} u \\ v \end{pmatrix}, \quad E\begin{pmatrix} u \\ v \end{pmatrix} = \mathbf{0}, \quad D\begin{pmatrix} u \\ v \end{pmatrix} = V, \tag{3.1}$$

*where* $\boldsymbol F$, $\boldsymbol G$ *and* $\boldsymbol V$ *are known matrices which may not have full row or column rank and* $\mathscr{C}(G') \subseteq \mathscr{C}(F')$*. Then for an arbitrary matrix* $\boldsymbol C$ *of appropriate dimension satisfying* $\mathscr{C}(C') \subseteq \mathscr{C}(F')$

(a) *a minimum mean squared error linear predictor of* $Cx$ *having the form* $A_1h + A_2z + a_3$ *must be unbiased in the sense that the expected value of its prediction error is zero for all values of* $E(x)$;
(b) *the BLUE of* $C\beta$ *from the fixed effects model* $(y, X\beta, V)$, *where*

$$y = \begin{pmatrix} h \\ z \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} F \\ G \end{pmatrix},$$

*is a linear predictor of* $Cx$ *based on* $z$ *and* $h$, *having the minimum mean squared error;*
(c) *the mean squared prediction error of the predictor of part* (b) *is the same as the dispersion matrix of the BLUE of* $C\beta$ *from the above fixed effects linear model.*

**Proof.** Let $A_1h + A_2z + a_3$ be a linear predictor of $Cx$. The matrix of mean squared prediction error for this predictor is

$$\begin{aligned} &E[(A_1h + A_2z + a_3 - Cx)(A_1h + A_2z + a_3 - Cx)'] \\ &\quad = E(A_1h + A_2z + a_3 - Cx)E(A_1h + A_2z + a_3 - Cx)' \\ &\qquad + D(A_1h + A_2z + a_3 - Cx) \\ &\quad = [(A_1F + A_2G - C)E(x) + a_3][(A_1F + A_2G - C)E(x) + a_3]' \\ &\qquad + D(A_1h + A_2z - Cx). \end{aligned}$$

Since $h$ is non-random, the dispersion depends only on $A_2$. For a given choice of $A_2$, the bias term can be made equal to zero by choosing $A_1 = (C - A_2G)F^-$ and $a_3 = 0$. Therefore, a linear predictor with minimum mean square prediction error cannot have non-zero bias. This proves part (a).

In order to prove part (b), let $A_1h + A_2z + a_3$ be a linear predictor of $Cx$ and $B = ((C - A_2G)F^- : A_2)$. Let us also write $\epsilon = (u' : v')'$. It follows that

$$\begin{aligned} &E[(A_1h + A_2z + a_3 - Cx)(A_1h + A_2z + a_3 - Cx)'] \\ &\quad \geqslant E[(By - Cx)(By - Cx)'] \\ &\quad = D(By - Cx) = D\left(B\epsilon - ((C - A_2G)F^- : A_2)\begin{pmatrix} F \\ G \end{pmatrix} x - Cx\right) \\ &\quad = D(B\epsilon) = BVB'. \end{aligned}$$

Let $C = LX$ and $B_* = LR$ where

$$R = I - VP_{X^\perp}\{P_{X^\perp}VP_{X^\perp}\}^- P_{X^\perp}. \tag{3.2}$$

According to (2.1), $B_*y$ is the BLUE of $C\beta$ from the model $(y, X\beta, V)$. Moreover, $B_*X = LX = C$. Thus,

$$\begin{aligned} BVB' &= (B - B_* + B_*)V + (B - B_* + B_*)' \\ &= B_*VB_*' + (B - B_*)V(B - B_*)' + B_*V(B - B_*)' + (B - B_*)VB_*'. \end{aligned}$$

The dispersion of the BLUE $\boldsymbol{Ry}$ given in (2.2) can be written as $\boldsymbol{VR}'$. Eq. (2.4) implies that $\mathscr{C}(\boldsymbol{VR}') \subseteq \mathscr{C}(\boldsymbol{X})$. It follows that $\boldsymbol{VB}_*'$ can be written as $\boldsymbol{XK}$ for some matrix $\boldsymbol{K}$. Hence,

$$(\boldsymbol{B} - \boldsymbol{B}_*)\boldsymbol{VB}_*' = (\boldsymbol{BX} - \boldsymbol{B}_*\boldsymbol{X})\boldsymbol{K} = (\boldsymbol{C} - \boldsymbol{C})\boldsymbol{K} = \boldsymbol{0}.$$

Consequently

$$\begin{aligned} &E[(\boldsymbol{A}_1\boldsymbol{h} + \boldsymbol{A}_2\boldsymbol{z} + \boldsymbol{a}_3 - \boldsymbol{Cx})(\boldsymbol{A}_1\boldsymbol{h} + \boldsymbol{A}_2\boldsymbol{z} + \boldsymbol{a}_3 - \boldsymbol{Cx})'] \\ &\quad \geqslant \boldsymbol{BVB}' = \boldsymbol{B}_*\boldsymbol{VB}_*' + (\boldsymbol{B} - \boldsymbol{B}_*)\boldsymbol{V}(\boldsymbol{B} - \boldsymbol{B}_*)' \\ &\quad \geqslant \boldsymbol{B}_*\boldsymbol{VB}_*' = E[(\boldsymbol{B}_*\boldsymbol{y} - \boldsymbol{Cx})(\boldsymbol{B}_*\boldsymbol{y} - \boldsymbol{Cx})']. \end{aligned}$$

This proves part (b). Part (c) follows from the simplification

$$E[(\boldsymbol{B}_*\boldsymbol{y} - \boldsymbol{Cx})(\boldsymbol{B}_*\boldsymbol{y} - \boldsymbol{Cx})'] = \boldsymbol{B}_*\boldsymbol{VB}_*' = \boldsymbol{LVR}'\boldsymbol{L}',$$

the last expression being the dispersion matrix of the BLUE of $\boldsymbol{C\beta}$ from the model $(\boldsymbol{y}, \boldsymbol{X\beta}, \boldsymbol{V})$. □

Duncan and Horn [5] had proved a version of Theorem 3.1 after assuming that $\boldsymbol{F} = \boldsymbol{I}$, $\boldsymbol{L} = \boldsymbol{I}$ and $\boldsymbol{V}$ is block-diagonal and non-singular. The best linear predictor described in this theorem happens to be the best linear *unbiased* predictor (BLUP). It has the smallest mean squared error among *all* predictors (not necessarily linear or unbiased) when the joint distribution of the errors is normal.

Eqs. (1.1) and (1.2) up to time $t$ can be written as

$$\boldsymbol{y}_t = \boldsymbol{X}_t\boldsymbol{\gamma}_t + \boldsymbol{\epsilon}_t, \tag{3.3}$$

where $\boldsymbol{\gamma}_t = (\boldsymbol{x}_1' : \boldsymbol{x}_2' : \cdots : \boldsymbol{x}_t')'$ and

$$\boldsymbol{y}_t = \begin{pmatrix} \boldsymbol{B}_1\boldsymbol{x}_0 \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \\ z_1 \\ z_2 \\ \vdots \\ z_t \end{pmatrix}, \quad \boldsymbol{X}_t = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ -\boldsymbol{B}_2 & \boldsymbol{I} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & -\boldsymbol{B}_t & \boldsymbol{I} \\ \boldsymbol{H}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}_2 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{H}_t \end{pmatrix}, \quad \boldsymbol{\epsilon}_t = \begin{pmatrix} -\boldsymbol{u}_1 \\ -\boldsymbol{u}_2 \\ \vdots \\ -\boldsymbol{u}_t \\ \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{v}_t \end{pmatrix}.$$

This is a special case of (3.1) with $\boldsymbol{F}$ non-singular. We shall denote $D(\boldsymbol{\epsilon}_t)$ by $\boldsymbol{V}_t$, and use the notation $\mathscr{M}_t$ to describe the model $(\boldsymbol{y}_t, \boldsymbol{X}_t\boldsymbol{\gamma}_t, \boldsymbol{V}_t)$.

The state update and measurement equations up to time $t$ can also be written as

$$\boldsymbol{y}_t = (\boldsymbol{X}_t : \boldsymbol{0})\boldsymbol{\gamma}_{t+1} + \boldsymbol{\epsilon}_t. \tag{3.4}$$

We shall denote by $\mathscr{M}_t^\dagger$ the model $(\boldsymbol{y}_t, (\boldsymbol{X}_t : \boldsymbol{0})\boldsymbol{\gamma}_{t+1}, \boldsymbol{V}_t)$, which also fits into the framework of (3.1). However, the condition $\mathscr{C}(\boldsymbol{C}') \subseteq \mathscr{C}(\boldsymbol{F}')$ of Theorem 3.1 means that the result can be used only to predict linear functions of $\gamma_t$, and not for all functions of $\gamma_{t+1}$.

The state update equations (1.1) up to time $t$ and the measurement equations (1.2) up to time $t-1$ can be combined into the single equation

$$y_{t|t-1} = X_{t|t-1}\gamma_t + \epsilon_{t|t-1}, \tag{3.5}$$

where $\gamma_t$ is as in (3.3) and

$$y_{t|t-1} = \begin{pmatrix} B_1x_0 \\ 0 \\ \vdots \\ 0 \\ z_1 \\ \vdots \\ z_{t-1} \end{pmatrix}, \quad X_{t|t-1} = \begin{pmatrix} I & 0 & \cdots & 0 \\ -B_2 & I & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -B_t & I \\ H_1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & H_{t-1} & 0 \end{pmatrix}, \quad \epsilon_{t|t-1} = \begin{pmatrix} -u_1 \\ -u_2 \\ \vdots \\ -u_t \\ -v_1 \\ \vdots \\ -v_{t-1} \end{pmatrix}$$

We shall denote $D(\epsilon_{t|t-1})$ by $V_{t|t-1}$ and use the notation $\mathscr{M}_{t|t-1}$ for the model $(y_{t|t-1}, X_{t|t-1}\gamma_t, V_{t|t-1})$. This is also a special case of (3.1) with $F$ non-singular.

Recursive prediction of the state vector consists of the following cycle of steps.

(I) Given the prediction of $x_{t-1}$ based on $x_0, z_1, \ldots, z_{t-1}$, and the dispersion of the prediction error, predict $x_t$ and the dispersion matrix.
(II) Given the above quantities, update these by taking into account the additional measurement $z_t$.

The foregoing discussion and Theorem 3.1 implies that the best linear predictor of the state vector and at every stage is given by a BLUE in a suitable 'equivalent' linear model. This is where the update equations of Theorem 2.1 have a role to play. Using the 'BLUE' of $x_{t-1}$ and its dispersion under the linear model (3.3) (with $t$ replaced by $t-1$), we can find the 'BLUE' of $x_t$ and its dispersion under the model (3.3) recursively by tracking the following three transitions:

(Ia) from $\mathscr{M}_{t-1}$ to $\mathscr{M}^{\dagger}_{t-1}$,
(Ib) from $\mathscr{M}^{\dagger}_{t-1}$ to $\mathscr{M}_{t|t-1}$, and
(II) from $\mathscr{M}_{t|t-1}$ to $\mathscr{M}_t$.

Haslett [8] considered the updates in steps Ia and Ib together, and consequently needed a more complex update result, after assuming non-singularity of the dispersion matrices. Theorem 2.1 will be adequate for our purpose, even though the set-up is more general (with possible singularity of the dispersion matrices). Because of the correlation of errors, it is necessary to update the prediction of all the state vectors $(x_1, \ldots, x_t)$ when $z_t$ becomes available, even though the interest is mainly in the latest state, $x_t$.

We use the following additional notations:

| Quantity | Notation, when computed from $\mathscr{M}_{t\|t-1}$ | Notation, when computed from $\mathscr{M}_t$ |
|---|---|---|
| BLUE of $\gamma_t$ | $\hat{\gamma}_{t\|t}$ | $\hat{\gamma}_{t\|t-1}$ |
| Dispersion of above | $\boldsymbol{P}_{1\ldots t\|t}$ | $\boldsymbol{P}_{1\ldots t\|t-1}$ |
| BLUE of $\boldsymbol{x}_s\,(s \leqslant t)$ | $\hat{\boldsymbol{x}}_{s\|t}$ | $\hat{\boldsymbol{x}}_{s\|t-1}$ |
| Dispersion of above | $\boldsymbol{P}_{s\|t}$ | $\boldsymbol{P}_{s\|t-1}$ |

The update equations for prediction are given in the next section.

## 4. Optimal recursive prediction in the state-space model

*Step Ia: transition from $\mathscr{M}_{t-1}$ to $\mathscr{M}^{\dagger}_{t-1}$.* Using Theorem 3.1 for (3.3), with $t$ replaced by $t-1$ and $\boldsymbol{C} = \boldsymbol{I}$, $\hat{\gamma}_{t-1|t-1}$ and $\boldsymbol{P}_{1\ldots t-1|t-1}$ may be identified as the minimum mean squared linear predictor of $\gamma_{t-1}$ based on $\boldsymbol{x}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_{t-1}$, and the dispersion matrix of the corresponding prediction error.

The transition from $\mathscr{M}_{t-1}$ to $\mathscr{M}^{\dagger}_{t-1}$ should involve *no change* in the BLUE or the dispersion matrix, since the model $\mathscr{M}^{\dagger}_{t-1}$ is only a reparametrization of $\mathscr{M}_{t-1}$. Applying Theorem 3.1 to (3.4), with $t$ replaced by $t-1$, we observe that the best linear predictor and the dispersion matrix of the prediction error remain the same.

*Step Ib: transition from $\mathscr{M}^{\dagger}_{t-1}$ to $\mathscr{M}_{t|t-1}$.* Applying Theorem 3.1 to (3.5), $\hat{\gamma}_{t|t-1}$ and $\boldsymbol{P}_{1\ldots t|t-1}$ may be identified as the minimum mean squared error linear predictor of $\gamma_t$ based on $\boldsymbol{x}_0, \boldsymbol{z}_1, \cdots, \boldsymbol{z}_{t-1}$, and the dispersion matrix of the corresponding prediction error. Their relationship with the corresponding quantities in the preceding step are given in the following theorem.

**Theorem 4.1.** *Under the set-up described in the foregoing discussion and the notations of Section 3, the minimum mean squared error predictor of $\gamma_t$ on the basis of $\boldsymbol{x}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_{t-1}$ is*

$$\hat{\gamma}_{t|t-1} = \begin{pmatrix} \hat{\gamma}_{t-1|t-1} \\ \hat{\boldsymbol{x}}_{t|t-1} \end{pmatrix}, \tag{4.1}$$

*where*

$$\hat{\boldsymbol{x}}_{t|t-1} = \boldsymbol{B}_t\hat{\boldsymbol{x}}_{t-1|t-1} - \boldsymbol{K}'_{t|t-1}\boldsymbol{V}^{-}_{t-1} \begin{pmatrix} \boldsymbol{B}_1\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_{1|t-1} \\ \boldsymbol{B}_2\hat{\boldsymbol{x}}_{1|t-1} - \hat{\boldsymbol{x}}_{2|t-1} \\ \vdots \\ \boldsymbol{B}_{t-1}\hat{\boldsymbol{x}}_{t-2|t-1} - \hat{\boldsymbol{x}}_{t-1|t-1} \\ \boldsymbol{z}_1 - \boldsymbol{H}_1\hat{\boldsymbol{x}}_{1|t-1} \\ \vdots \\ \boldsymbol{z}_{t-1} - \boldsymbol{H}_{t-1}\hat{\boldsymbol{x}}_{t-1|t-1} \end{pmatrix}, \tag{4.2}$$

*and*

$$K_{t|t-1} = \mathrm{Cov}(\epsilon_{t-1}, -u_t) = \big(Q_u(t,1) : \cdots : Q_u(t-1,t) : -Q_{uv}(t,1) : \cdots : -Q_{uv}(t,1) : \cdots : -Q_{uv}(t,t-1)\big)'.$$

*The corresponding prediction error matrix is*

$$P_{1\ldots t|t-1} = \begin{pmatrix} P_{1\ldots t-1|t-1} & P_{1\ldots t-1|t-1}M_t' \\ M_tP_{1\ldots t-1|t-1} & P_{t|t-1} \end{pmatrix}, \tag{4.3}$$

*where*

$$P_{t|t-1} = M_tP_{1\ldots t-1|t-1}M_t' + Q_u(t,t) - K_{t|t-1}'V_{t-1}^{-}K_{t|t-1}, \tag{4.4}$$

*and*

$$M_t = (\mathbf{0} : \mathbf{0} : \cdots : B_t) + K_{t|t-1}'V_{t-1}^{-}X_{t-1}.$$

**Proof.** The model $\mathscr{M}_{t|t-1}$ is obtained from the model $\mathscr{M}_{t-1}^{\dagger}$ by including some additional observations. Note that $\rho(X_{t|t-1}) - \rho(X_{t-1} : \mathbf{0})$ is equal to the size of the vector $x_t$. Therefore, there is no *new* LZF. The BLUE of $\gamma_{t-1}$ and its dispersion remain unchanged. This justifies the form of $\hat{\gamma}_{t|t-1}$ given in (4.1) and the top left block on the right hand side of (4.3). We can use part (a) of Theorem 2.1 in order to obtain $\hat{x}_{t|t-1}$ and $P_{t|t-1}$ in terms of the previously computed quantities. Specifically, we have

$$\begin{aligned} -B_t\hat{x}_{t-1|t-1} + \hat{x}_{t|t-1} &= \mathbf{0} - \mathrm{Cov}(-u_t, \epsilon_{t-1})[D(\epsilon_{t-1})]^{-} \\ &\quad \times (y_{t-1} - X_{t-1}\hat{\gamma}_{t-1|t-1}) \\ &= -K_{t|t-1}'V_{t-1}^{-}(y_{t-1} - X_{t-1}\hat{\gamma}_{t-1|t-1}), \end{aligned}$$

$$\begin{aligned} D(-B_t\hat{x}_{t-1|t-1} + \hat{x}_{t|t-1}) &= Q_u(t,t) - K_{t|t-1}'V_{t-1}^{-}D(y_{t-1} - X_{t-1}\hat{\gamma}_{t-1|t-1}) \\ &\quad \times V_{t-1}^{-}K_{t|t-1} \\ &= Q_u(t,t) - K_{t|t-1}'V_{t|t-1}^{-}K_{t|t-1} + K_{t|t-1}'V_{t-1}^{-}X_{t-1} \\ &\quad \times P_{1\ldots t-1|t-1}X_{t-1}'V_{t-1}^{-}K_{t|t-1}. \end{aligned}$$

The first equation implies (4.2). In order to simplify the second equation, we borrow the notation $y_l$ from Theorem 2.1, which happens to be numerically equal to $\mathbf{0}$ in this case. We have, by virtue of zero correlation between the LZF $(y_{t-1} - X_{t-1}\hat{\gamma}_{t-1|t-1})$ and the BLUE $\hat{x}_{t-1|t-1}$,

$$\begin{aligned} &\mathrm{Cov}(\hat{x}_{t|t-1}, \hat{x}_{t-1|t-1}) \\ &\quad = \mathrm{Cov}(y_l + B_t\hat{x}_{t-1|t-1} - K_{t|t-1}'V_{t-1}^{-}(y_{t-1} - X_{t-1}\hat{\gamma}_{t-1|t-1}), \hat{x}_{t-1|t-1}) \\ &\quad = \mathrm{Cov}(y_l, \hat{x}_{t-1|t-1}) + B_tP_{t-1|t-1} \\ &\quad = \mathrm{Cov}(y_l, y_{t-1})[D(y_{t-1})]^{-}\mathrm{Cov}(y_{t-1}, \hat{x}_{t-1|t-1}) + B_tP_{t-1|t-1}. \end{aligned}$$

The last expression follows from the fact that $\hat{x}_{t-1|t-1}$ is a linear function of $y_{t-1}$. It follows that

$$\begin{aligned}
&\mathrm{Cov}(\hat{x}_{t|t-1}, \hat{x}_{t-1|t-1}) \\
&= K'_{t|t-1} V^{-}_{t-1} \mathrm{Cov}(X_{t-1}\hat{\gamma}_{t-1|t-1}, \hat{x}_{t-1|t-1}) + B_t P_{t-1|t-1} \\
&= K'_{t|t-1} V^{-}_{t-1} X_{t-1} P_{1\ldots t-1|t-1} (\mathbf{0} : \mathbf{0} : \cdots : I)' + B_t P_{t-1|t-1}.
\end{aligned}$$

Consequently

$$\begin{aligned}
&Q_u(t,t) - K'_{t|t-1} V^{-}_{t-1} K_{t|t-1} + K'_{t|t-1} V^{-}_{t-1} X_{t-1} P_{1\ldots t-1|t-1} X'_{t-1} V^{-}_{t-1} K_{t|t-1} \\
&= D(B_t \hat{x}_{t-1|t-1}) + P_{t|t-1} - \mathrm{Cov}(B_t \hat{x}_{t-1|t-1}, \hat{x}_{t|t-1}) \\
&\quad - \mathrm{Cov}(\hat{x}_{t|t-1}, B_t \hat{x}_{t-1|t-1}) \\
&= P_{t|t-1} + B_t P_{t-1|t-1} B'_t - (\mathbf{0} : \mathbf{0} : \cdots : B_t) P_{1\ldots t-1|t-1} X'_{t-1} V^{-}_{t-1} K_{t|t-1} \\
&\quad - K'_{t|t-1} V^{-}_{t-1} X_{t-1} P_{1\ldots t-1|t-1} (\mathbf{0} : \mathbf{0} : \cdots : B_t)' - 2 B_t P_{t-1|t-1} B'_t.
\end{aligned}$$

By rearranging the terms of the first and final expressions, we have (4.4).

In order to obtain the off-diagonal blocks of the right hand side of (4.3), we write the first equation of this proof as

$$\begin{aligned}
-B_t \hat{x}_{t-1|t-1} + \hat{x}_{t|t-1} &= y_t - K'_{t|t-1} V^{-}_{t-1} (y_{t-1} - X_{t-1} \hat{\gamma}_{t-1|t-1}) \\
&= y_t - K'_{t|t-1} V^{-}_{t-1} (y_{t-1} - X_{t-1} \gamma_{t-1}) \\
&\quad + K'_{t|t-1} V^{-}_{t-1} X_{t-1} (\hat{\gamma}_{t-1|t-1} - \gamma_{t-1}).
\end{aligned}$$

It is easy to see that the combination of the first two terms of the last expression is uncorrelated with $y_{t-1}$, and hence, with $\hat{\gamma}_{t-1|t-1}$ Therefore,

$$\begin{aligned}
K'_{t|t-1} V^{-}_{t-1} X_{t-1} P_{1\ldots t-1|t-1} &= \mathrm{Cov}(K'_{t|t-1} V^{-}_{t-1} X_{t-1} (\hat{\gamma}_{t-1|t-1} - \gamma_{t-1}), \hat{\gamma}_{t-1|t-1}) \\
&= \mathrm{Cov}(-B_t \hat{x}_{t-1|t-1} + \hat{x}_{t|t-1}, \hat{\gamma}_{t-1|t-1}) \\
&= -(\mathbf{0} : \mathbf{0} : \cdots : B_t) P_{1\ldots t-1|t-1} \\
&\quad + \mathrm{Cov}(\hat{x}_{t|t-1}, \hat{\gamma}_{t-1|t-1}).
\end{aligned}$$

Therefore,

$$\mathrm{Cov}(\hat{x}_{t|t-1}, \hat{\gamma}_{t-1|t-1}) = [K'_{t|t-1} V^{-}_{t-1} X_{t-1} + (\mathbf{0} : \mathbf{0} : \cdots : B_t)] P_{1\ldots t-1|t-1}.$$

This completes the proof of (4.3). □

*Step II: transition from $\mathcal{M}_{t|t-1}$ to $\mathcal{M}_t$.* Note that $\hat{\gamma}_{t|t}$ is the minimum mean squared error linear predictor of $\gamma_t$ based on $x_0, z_1, \ldots, z_t$, and $P_{1\ldots t|t}$ is the is the dispersion matrix of the corresponding prediction error. In order to obtain these from $\hat{\gamma}_{t|t-1}$ and $P_{1\ldots t|t-1}$ we once again use Theorem 3.1. The updates are obtained as a special case of part (b) of Theorem 2.1. The results are summarized in the following theorem.

**Theorem 4.2.** *Under the set-up described in the foregoing discussion and the notations of Section 3, the minimum mean squared error predictor of $\gamma_t$ on the basis of $x_0, z_1, \ldots, z_t$ is*

$$\hat{\gamma}_{t|t} \quad = \quad \hat{\gamma}_{t|t-1} + P_{1\ldots t|t-1} R_t' [D(w_t)]^- w_t, \tag{4.5}$$

*where*

$$w_t = z_t - \hat{z}_{t|t}, \tag{4.6}$$

$$\hat{z}_{t|t} = H_t \hat{x}_{t|t-1} + K_t' V_{t|t-1}^- \begin{pmatrix} B_1 x_0 - \hat{x}_{1|t-1} \\ B_2 \hat{x}_{1|t-1} - \hat{x}_{2|t-1} \\ \vdots \\ B_t \hat{x}_{t-1|t-1} - \hat{x}_{t|t-1} \\ z_1 - H_1 \hat{x}_{1|t-1} \\ \vdots \\ z_{t-1} - H_{t-1} \hat{x}_{t-1|t-1} \end{pmatrix}, \tag{4.7}$$

$$\begin{aligned} K_t &= \mathrm{Cov}(\epsilon_{t|t-1}, v_t) \\ &= (-Q_{uv}(1,t)' : \cdots : -Q_{uv}(t,t)' : Q_v(t,1) : \cdots : Q_v(t,t-1))', \\ D(w_t) &= Q_v(t,t) - K_t' V_{t|t-1}^- K_t + R_t P_{1\ldots t|t-1} R_t' \\ \text{and } R_t &= (0 : \cdots 0 : H_t) - K_t' V_{t|t-1}^- X_{t|t-1}. \end{aligned} \tag{4.8}$$

*The corresponding prediction error matrix is*

$$P_{1\ldots t|t-1} - P_{1\ldots t|t-1} - P_{1\ldots t|t-1} R_t' [D(w_t)]^- R_t P_{1\ldots t|t-1}. \tag{4.9}$$

**Proof.** The model $\mathscr{M}_t$ is obtained from the model $\mathscr{M}_{t|t-1}$ by including some additional observations. Since $X_{t|t-1}$ has full column rank, there is no newly estimable LPF. In the present case, the recursive group residual of (2.6) is identified as

$$w_t = z_t - H_t \hat{x}_{t|t-1} - \mathrm{Cov}(v_t, \epsilon_{t|t-1})[D(\epsilon_{t|t-1})]^- (y_{t|t-1} - X_{t|t-1} \hat{\gamma}_{t|t-1}),$$

which simplifies to (4.6). By rewriting $w_t$ as

$$\begin{aligned} w_t &= v_t - K_t' V_{t|t-1}^- \epsilon_{t|t-1} - H_t (\hat{x}_{t|t-1} - x_t) + K_t' V_{t|t-1}^- X_{t|t-1} (\hat{\gamma}_{t|t-1} - \gamma_t) \\ &= [v_t - K_t' V_{t|t-1}^- \epsilon_{t|t-1}] - R_t (\hat{\gamma}_{t|t-1} - \gamma_t), \end{aligned}$$

we obtain its dispersion matrix by (4.8). Finally,

$$\begin{aligned} \mathrm{Cov}(\hat{\gamma}_{t|t-1}, w_t) &= \mathrm{Cov}(\hat{\gamma}_{t|t-1}, v_t - K_t' V_{t|t-1}^- \epsilon_{t|t-1} - R_t (\hat{\gamma}_{t|t-1} - \gamma_t)) \\ &= -\mathrm{Cov}(\hat{\gamma}_{t|t-1}, R_t (\hat{\gamma}_{t|t-1} - \gamma_t)) = -D(\hat{\gamma}_{t|t-1}) R_t'. \end{aligned}$$

Substitution of $w_t$, $D(w_t)$ and $\mathrm{Cov}(\hat{\gamma}_{t|t-1}, w_t)$, in part (b) of Theorem 2.1 produces (4.5) and (4.9). □

Under the special case given in Section 1, the update equations of Theorem 4.1 and 4.2 simplify considerably. In this case it is enough to use updates of the prediction of $\boldsymbol{x}_t$ (instead of the entire $\boldsymbol{\gamma}_t$) and the corresponding prediction error matrix. The equations obtained from Theorem 4.1 are

$$\hat{\boldsymbol{x}}_{t|t-1} = \boldsymbol{B}_t\hat{\boldsymbol{x}}_{t-1|t-1}, \tag{4.10}$$

$$\boldsymbol{P}_{t|t-1} = \boldsymbol{B}_t\boldsymbol{P}_{t-1|t-1}\boldsymbol{B}_t' + \boldsymbol{Q}_u(t,t). \tag{4.11}$$

The equations obtained from Theorem 4.2 are

$$\hat{\boldsymbol{x}}_{t|t} = \hat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{H}_t'(\boldsymbol{Q}_v(t,t) + \boldsymbol{H}_t\boldsymbol{P}_{t|t-1}\boldsymbol{H}_t')^-(z_t - \boldsymbol{H}_t\hat{\boldsymbol{x}}_{t|t-1}), \tag{4.12}$$

$$\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{H}_t'(\boldsymbol{Q}_v(t,t) + \boldsymbol{H}_t\boldsymbol{P}_{t|t-1}\boldsymbol{H}_t')^-\boldsymbol{H}_t\boldsymbol{P}_{t|t-1}. \tag{4.13}$$

The recursive relations (4.10)–(4.13) constitute the usual Kalman filter. The generalization of this filter for correlated errors is given by (4.1), (4.3), (4.5) and (4.9). These relations hold for $t \geqslant 2$. The initial iterates are $\hat{\boldsymbol{x}}_{1|0} = \boldsymbol{B}_1\boldsymbol{x}_0$ and $\boldsymbol{P}_{1|0} = \boldsymbol{Q}_u(1,1)$. The minimum mean squared error linear predictor of the measurement vector $z_t$ (in terms of $\boldsymbol{x}_0, z_1, \ldots, z_{t-1}$) and the dispersion matrix of the corresponding prediction error are given by (4.7) and (4.8), respectively.

Derivation of the usual Kalman Filter (in the simplest special case) from the update equations of the general linear model appears in [21], along with many other applications of these equations.

## References


[1] P. Bhimasankaram, D. Sengupta, The linear zero functions approach to linear models, Sankhyā Ser. B 58 (1996) 338–351.

[2] P. Bhimasankaram, D. Sengupta, S. Ramanathan, Recursive inference in a general linear model, Sankhyā Ser. A 57 (1995) 227–255.

[3] H. Bolfarine, S. Zacks, Prediction Theory for Finite Populations, Springer Series in Statistics, Springer-Verlag, New York, 1992.

[4] R.L. Brown, J. Durbin, J.M. Evans, Methods of investigating whether a regression relationship is constant over time (with discussion), J. Roy. Statist. Soc. Ser. B 37 (1975) 149–192.

[5] D.B. Duncan, S.D. Horn, Linear dynamic recursive estimation in the general linear model, J. Amer. Statist. Assoc. 67 (1972) 815–821.

[6] J. Durbin, S.J. Koopman, Time series analysis by state space methods, Oxford Statistical Science Series, Oxford University Press, Oxford, 2001.

[7] A.C. Harvey, D.A. Phillips, Maximum likelihood estimation of regression models with autoregressive-moving average disturbances, Biometrika 66 (1979) 49–58.

[8] S. Haslett, Recursive estimation of the generral linear model with dependent errors and multiple additional observations, Aust. J. Statist. 27 (1985) 183–188.

[9] S. Haslett, Updating linear models with dependent errors to include additional data and/or parameters, Linear Algebra Appl. 237/238 (1996) 329–349.

[10] S.R. Jammalamadaka, D. Sengupta, Changes in the general linear model: a unified approach, Linear Algebra Appl. 289 (1999) 225–242.

[11] R.E. Kalman, A new approach to linear filtering and prediction problem, ASME Trans. J. Basic Engrg. 82-D (1960) 35–45.

[12] R.E. Kalman, R.S. Bucy, New results in linear filtering and prediction theory, ASME Trans. J. Basic Engrg. 83-D (1961) 95–108.
[13] F. Kianifard, W. Swallow, A review of the development and application of recursive residuals in linear models, Amer. Statist. Assoc. 91 (1996) 391–400.
[14] C.A. McGilchrist, R.L. Sandland, Recursive estimation of the general linear model with dependent errors, J. Roy. Statist. Soc. Ser. B 41 (1979) 65–68.
[15] S.K. Mitra, P. Bhimasankaram, Generalized inverses of partitioned matrices and recalculation of least squares estimators for data and model changes, Sankhyā Ser. A 33 (1971) 395–410.
[16] F.H. Nieto, V.M. Guerrero, Kalman filter for singular and conditional state-space models when the system state and the observational error are correlated, Statist. Probab. Lett. 22 (1995) 303–310.
[17] R.L. Placket, Some theorems in least squares, Biometrika 37 (1950) 149–157.
[18] P.R. Pordzik, A lemma on $g$-inverse of the bordered matrix and its application to recursive estimation in the restricted model, Comput. Statist. 7 (1992) 31–37.
[19] C.R. Rao, Representations of best linear unbiased estimators in the Gauss–Markoff model with a singular dispersion matrix, J. Multivariate Anal. 3 (1973) 276–292.
[20] S.R. Searle, Extending some results and proofs for the singular linear model, Linear Algebra Appl. 210 (1994) 139–151.
[21] D. Sengupta, S.R. Jammalamadaka, Linear Models: An Integrated Approach, Series on multivariate analysis, World Scientific, Singapore, 2003.
[22] H.W. Sorenson, Parameter Estimation: Principles and Problems, Series on Control and Systems Theory, Marcel Dekker, New York, 1980.