Ann. Inst. Statist. Math.
Vol. 50, No. 1, 147–164 (1998)

# ANALYSIS OF MULTI-UNIT VARIANCE COMPONENTS MODELS WITH STATE SPACE PROFILES

JOHN V. TSIMIKAS[1]* AND JOHANNES LEDOLTER[2]

[1] *Department of Health Care Policy, Harvard Medical School, 25 Shattuck Str, Boston MA 02115, U.S.A.*
[2] *Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242, U.S.A.*

(Received April 11, 1996; revised March 27, 1997)

**Abstract.** We apply the Kalman Filter to the analysis of multi-unit variance components models where each unit's response profile follows a state space model. We use mixed model results to obtain estimates of unit-specific random effects, state disturbance terms and residual noise terms. We use the signal extraction approach to smooth individual profiles. We show how to utilize the Kalman Filter to efficiently compute the restricted loglikelihood of the model. For the important special case where each unit's response profile follows a continuous structural time series model with known transition matrix we derive an EM algorithm for the restricted maximum likelihood (REML) estimation of the variance components. We present details for the case where individual profiles are modeled as local polynomial trends or polynomial smoothing splines.

*Key words and phrases*: Continuous-time stochastic models, EM algorithm, Kalman Filter, mixed model prediction, restricted maximum likelihood, smoothing splines, unequally spaced observations, variance components.

## 1. Introduction

Time series data on a panel of units are modeled according to

$$y_{it} = x_{it}'\beta + z_{it}'\gamma_i + r_{it} + \nu_{it}, \tag{1.1}$$

where $y_{it}$ denotes the response for the $i$-th unit at time $t$; $\beta$ is a vector of unknown fixed population parameters measuring the impact of covariates; $\gamma_i$ is a vector of unit-specific random effects, usually offsets from the population parameters; $r_{it}$ is a stochastic component, expressed in state space form, that describes the time-evolution of the measurements on unit $i$; and $\nu_{it}$ is an independent measurement error. Furthermore, all random components are independent across subjects.

* Now at Department of Mathematics and Statistics, University of Massachusetts at Amherst, P. O. Box 34515, Amherst, MA 01003, U.S.A.

The model in (1.1) provides a very flexible representation as a wide variety of models can be written in this form. The model in (1.1) without the state space component was introduced by Hartley and Rao (1967), who showed how to compute maximum likelihood estimates of the parameters; for further discussion, see Harville (1977). If $r_{it}$ follows a stationary process, then we can view the model in (1.1) as a Laird-Ware model (Laird and Ware (1982), Laird *et al.* (1987)) with stationary errors. Special cases include independent within-unit errors, and within-unit errors that follow autoregressive moving average (ARMA) models; see Jones (1993). If $r_{it}$ follows a nonstationary process, then the model in (1.1) specifies a semiparametric regression function for each unit. Models of this type can be used to smooth each unit's response profile, in addition to the estimation of covariate effects $\beta$ and the prediction of subject-specific effects $\gamma_i$. A natural way of smoothing the $i$-th unit's profile is to extract the signal vector consisting of elements $s_{it} = x_{it}'\beta + z_{it}'\gamma_i + r_{it}$ for $t = 1, \ldots, n_i$. This is the approach adopted in this paper, and we show how to obtain the Best Linear Unbiased Predictor (BLUP) of the signal. Since this predictor depends on unknown variance components, we follow a two-stage Empirical Bayes approach. We first estimate the parameters by maximizing the restricted loglikelihood of the model and then substitute the restricted maximum likelihood (REML) estimates into the expressions for the BLUPs.

The Kalman Filter is an efficient algorithm for the calculation of the likelihood, and for obtaining predictions of the states, signals and future observations in state space models. It can accomodate unequally spaced observations by operating on the state space representation of the discrete realization of a continuous stochastic process. In the context of a single time series, a number of papers show how to adjust the Kalman Filter for diffuse initial states and fixed effects. Ansley and Kohn (1985, 1990), Kohn and Ansley (1986), De Jong (1991) show how to achieve this by modifying the Kalman Filter recursions. In Section 2 of this paper we follow our earlier approach (Tsimikas and Ledolter (1994, 1997)) which relates the inference in state space models to mixed model prediction and estimation within the REML hierarchy. Our methodology provides a unifying algorithmic framework that ties together the various algorithms that have been proposed for the inference in single-unit state space models.

In the context of the multi-unit Laird-Ware model, Jones (1993) shows how to utilize the Kalman Filter for calculating the likelihood and obtaining predictions of the random effects. In Section 3 of this paper we expand on the work of Jones. Assigning flat prior distributions on fixed effects, we obtain REML estimates of the variance components and best linear unbiased predictors (BLUPs) of the random effects. The REML hierarchy has the advantage that the mean squared errors of the predictions are corrected for the estimation of fixed effects; REML estimates of variance components are also less biased than their maximum likelihood (ML) counterparts. We show how to implement an EM algorithm for REML estimation.

In Section 4 we propose a general stochastic trend model for $r_{it}$ that contains smoothing polynomial splines (Anderson and Jones (1995)) and random-disturbed-highest-derivative polynomials (Wilson (1995)) as special cases. Our model is further enhanced by introducing stochastic seasonal components (Harvey (1989)).

Since we assume an underlying continuous process, our results accomodate quite easily unequally spaced observations

## 2. Smoothing in the single-unit state space model

In this section we review results on single-unit state space model estimation. This review helps us understand the various within-unit state space structures that arise with repeated measures. Detailed accounts of inference in state space models can be found in Harvey (1989), De Jong (1988, 1989, 1991), Koopman and Shephard (1992), Koopman (1993), and Tsimikas and Ledolter (1997).

### 2.1 *The single-unit state space model*

Consider the single-unit state space model with fixed and random time-invariant regressors,

$$\begin{aligned} y_t &= x_t'\beta + z_t'\gamma + h_t'\alpha_t + \nu_t \\ \alpha_t &= \Phi_t\alpha_{t-1} + R_t\xi_t, \end{aligned} \tag{2.1}$$

where $t = 1, 2, \ldots, T$. The vector $\beta$ is a $p \times 1$ vector of fixed effects, $\gamma$ is a $g \times 1$ vector of time-invariant random effects; $x_t'$ is the $1 \times p$ design vector for $\beta$ and $z_t'$ is the $1 \times g$ design vector for $\gamma$; $h_t'$ is a $1 \times q$ vector and $\alpha_t$ is the $q \times 1$ state vector at time $t$, $\nu_t$ is the observation noise, $\Phi_t$ is a $q \times q$ transition matrix, $R_t$ is a $q \times m$ matrix, and $\xi_t$ is the $m \times 1$ disturbance term in the state transition equation. The $\nu_t$'s are uncorrelated and distributed with mean 0 and variance $\sigma^2$, the $\xi_t$'s are uncorrelated and distributed with mean 0 and covariance matrix $\Xi_t$; disturbance terms and observation noise terms are uncorrelated. The vector $\gamma$ has mean 0 and covariance matrix $B_{11}$ and is uncorrelated with the disturbance and noise terms. Time-invariant regressors that appear in the state transition equation are transferred to the observation equation (see Harvey (1989), p. 104).

We partition the initial state vector $\alpha_0 = (\alpha_{10}', \alpha_{20}')'$ into a $q_1 \times 1$ vector $\alpha_{10}$ with a diffuse prior and a $q_2 \times 1$ vector $\alpha_{20}$ with a proper prior distribution with mean 0 and $q_2 \times q_2$ covariance matrix $B_{22}$. Moreover, let $B_{12} = \mathrm{cov}(\gamma, \alpha_{20}')$. We define $\alpha_t^\dagger = \alpha_t - \prod_{i=0}^{t-1}\Phi_{t-i}\alpha_0$ and move $\alpha_0$ into the observation equation; the modified initial state $\alpha_0^\dagger$ is fixed at the value 0. We partition $b_t' = h_t'\prod_{i=0}^{t-1}\Phi_{t-i}$ according to the partition of the initial state $\alpha_0 = (\alpha_{10}', \alpha_{20}')'$; that is, we write $b_t' = (b_{1t}' : b_{2t}')$, where $b_{1t}'$ is a $q_1 \times 1$ vector and $b_{2t}'$ is a $q_2 \times 1$ vector. Furthermore, we form the $p + q_1$ row vector $x_t^{*'} = (x_t' : b_{1t}')$, the design vector at time $t$ for the $(p + q_1) \times 1$ vector of "fixed" effects $\beta^* = (\beta', \alpha_{10}')'$, and the $g + q_2$ row vector $z_t^{*'} = (z_t' : b_{2t}')$, the design vector at time $t$ for the $(g + q_2) \times 1$ vector of random effects $\gamma^* = (\gamma', \alpha_{20}')'$. Then (2.1) can be expressed as:

$$\begin{aligned} y_t &= x_t^{*'}\beta^* + z_t^{*'}\gamma^* + h_t'\alpha_t^\dagger + \nu_t \\ \alpha_t^\dagger &= \Phi_t\alpha_{t-1}^\dagger + R_t\xi_t. \end{aligned} \tag{2.2}$$

2.2 *The state space model as a linear mixed model*

In mixed model form, equation (2.2) can be written as

$$y = X\beta^* + Z\gamma^* + H\alpha^\dagger + \nu, \tag{2.3}$$

where $y = (y_1, \ldots, y_T)'$, $X = (x_1^*, \ldots, x_T^*)'$, $Z = (z_1^*, \ldots, z_T^*)'$, $\nu = (\nu_1, \ldots, \nu_T)'$, $\alpha^\dagger = (\alpha_1^{\dagger'}, \ldots, \alpha_T^{\dagger'})'$ and $H$ is a $T \times mT$ matrix whose $t$ row is the vector $(0, \ldots, h_t', \ldots, 0)$. Model (2.2) can also be expressed as a mixed model in terms of the disturbances $\xi = (\xi_1', \ldots, \xi_T')'$:

$$y = X\beta^* + Z\gamma^* + K\xi + \nu, \tag{2.4}$$

$$K = \begin{bmatrix} k_{11}' & 0 & \cdots & 0 \\ k_{21}' & k_{22}' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ k_{T1}' & k_{T2}' & \cdots & k_{TT}' \end{bmatrix}, \tag{2.5}$$

where $k_{tt}' = h_t' R_t$ and $k_{st}' = h_s' (\prod_{i=0}^{s-t-1} \Phi_{s-i}) R_t$, for $s > t$. The matrix $K$ is of dimension $T \times mT$ and has a triangular structure. The $mT \times 1$ vector of disturbances $\xi$ has mean 0 and covariance matrix $\tilde{\Xi} = \mathrm{diag}(\Xi_1, \ldots, \Xi_T)$. The vector of the time-invariant genuinely random effects $\gamma^*$ has a prior distribution with mean 0 and covariance matrix $B$, which consists of $B_{11}$ and $B_{22}$ on its diagonal and covariance component $B_{12}$. A flat prior is assumed for the fixed effect vector.

2.3 *Estimation and prediction in the state space model*

Let $\Lambda = K\tilde{\Xi}K' + \sigma^2 I$, where $K$ is given in (2.5). The covariance matrix of the observations is given by $V = \mathrm{var}(y) = ZBZ' + \Lambda$, and

$$\ln|V| = \ln|\Lambda| + \ln|B| + \ln|B^{-1} + Z'\Lambda^{-1}Z|. \tag{2.6}$$

Furthermore, under normality, the restricted loglikelihood ($l_R$) and the concentrated loglikelihood ($l_C$) are given by:

$$\begin{aligned} l_R &= -\frac{T - p - q_1}{2} - \frac{1}{2}\ln|X'V^{-1}X| - \frac{1}{2}\ln|V| - \frac{1}{2}y'Wy \\ l_C &= -\frac{T}{2} - \frac{1}{2}\ln|V| - \frac{1}{2}y'Wy, \end{aligned} \tag{2.7}$$

where $W = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$. See Tsimikas and Ledolter (1994, 1997), for example.

The Kalman Filter (KF), applied to the model with observation equation $y_t = h_t'\alpha_t^\dagger + \nu_t$ and with the initial state condition fixed at 0, performs the Cholesky decomposition of $\Lambda^{-1}$. That is, the inverse of $\Lambda$ can be written as $\Lambda^{-1} = L'F^{-1}L$, where the KF operator $L$ is a lower triangular matrix with ones in the diagonal. The diagonal matrix $F = \mathrm{diag}(f_1, \ldots, f_T)$ is also output of the KF recursions. The Cholesky decomposition implies that $|\Lambda| = \prod_{t=1}^{T} f_t$.

For any $T \times s$ matrix $C$ one can obtain the matrix product $LC$ by simultaneously applying the Kalman Filter that corresponds to the model $y_t = h_t'\alpha_t^{\dagger} + \nu_t$ to each column of $C$. That is, let $y = c_{(j)}$, where $c_{(j)}$ is the $j$-th column of $C$. Predictions and updates of the "state" vector are collected in augmented "state" prediction and "state" update matrices,

$$A_{t/r}(C) = [\alpha_{t/r}^{c_{(1)}}, \ldots, \alpha_{t/r}^{c_{(s)}}]; \qquad r = t-1, t, \tag{2.8}$$

where the KF recursions are given by:

$$\begin{aligned}
&\text{(a)} \quad A_{t/t-1}(C) - \Phi_t A_{t-1/t-1}(C),\\
&\text{(b)} \quad P_{t/t-1} = \Phi_t P_{t-1/t-1}\Phi_t' + R_t \Xi_t R_t',\\
&\text{(c)} \quad A_{t/t}(C) = A_{t/t-1}(C) + \frac{1}{f_t} P_{t/t-1} h_t E_t'(C),\\
&\text{(d)} \quad P_{t/t} = P_{t/t-1} - \frac{1}{f_t} P_{t/t-1} h_t h_t' P_{t/t-1},\\
&\text{(e)} \quad f_t = h_t' P_{t/t-1} h_t + \sigma^2.
\end{aligned} \tag{2.9}$$

The $1 \times s$ row vector $E_t'(C) = [c_{t1}, c_{t2}, \ldots, c_{ts}] - h_t' A_{t/t-1}(C)$ collects the "innovations" at time $t$ for the rows of $C$. The initial conditions are $A_{0/0}(C) = 0$ and $P_{0/0} = 0$, matrices of zeroes of respective dimensions. Note that only equations (a) and (c) in (2.9) are augmented. As a by-product of these recursions we obtain $\ln|\Lambda| = \det_T$, where

$$\det_t = \det_{t-1} + \ln f_t, \qquad (\det_0 = 0). \tag{2.10}$$

Computational savings in the recursions occur when $C$ is lower triangular. In that case $c_{(j)}$ has zeroes as its first $j-1$ elements and

$$\alpha_{t/t-1}^{c_{(j)}} - \alpha_{t/t}^{c_{(j)}} - 0 \quad \text{for} \quad t < j \tag{2.11}$$

The matrix $C'\Lambda^{-1}C + G$, where $G$ is a $s \times s$ matrix, is obtained through the recursion

$$Q_t(C;G) = Q_{t-1}(C;G) + \frac{1}{f_t} E_t(C) E_t'(C), \tag{2.12}$$

with starting value $Q_0(C;G) = G$. It follows that $C'\Lambda^{-1}C + G = Q_T(C;G)$. If $C$ is triangular, simplifications arise as (see (2.11)) several elements of $E_t(C)$ are known to be zero.

Evaluation of the loglikelihoods in (2.7), disturbance smoothing, and signal extraction is accomplished through the augmented KF recursions defined in (2.9); see Tsimikas and Ledolter (1997). This is achieved by selecting the appropriate $C$ and $G$ matrices, and by sweeping the end result of the augmented KF recursions on the rows that correspond to the time-invariant effects. The appropriate $C$ matrix

depends on the particular application, but is generally of the form $[X : Z : y : A]$. The appropriate $G$ matrix in this paper is of the form $G = \text{diag}(0, B^{-1}, 0, 0)$.

For *disturbance smoothing* we select $C = [X : Z : y : I : K]$. At the end of the recursions we get

$$(2.13)\quad Q_T([X : Z : y : I : K]; G) = \left[\begin{array}{ccc|cc} X'\Lambda^{-1}X & X'\Lambda^{-1}Z & X'\Lambda^{-1}y & X'\Lambda^{-1} & X'\Lambda^{-1}K \\ Z'\Lambda^{-1}X & Z'\Lambda^{-1}Z + B^{-1} & Z'\Lambda^{-1}y & Z'\Lambda^{-1} & Z'\Lambda^{-1}K \\ y'\Lambda^{-1}X & y'\Lambda^{-1}Z & y'\Lambda^{-1}y & y'\Lambda^{-1} & y'\Lambda^{-1}K \\ \hline \Lambda^{-1}X & \Lambda^{-1}Z & \Lambda^{-1}y & \Lambda^{-1} & \Lambda^{-1}K \\ K'\Lambda^{-1}X & K'\Lambda^{-1}Z & K'\Lambda^{-1}y & K'\Lambda^{-1} & K'\Lambda^{-1}K \end{array}\right].$$

If we sweep on the $g+q_2$ rows that correspond to the time-invariant random effects (that is, the second "row" of matrices in (2.13)), we obtain

$$(2.14)\quad Q_T^R([X : Z : y : I : K]; G) = \left[\begin{array}{ccc|cc} X'V^{-1}X & X'V^{-1}ZB & X'V^{-1}y & X'V^{-1} & X'V^{-1}K \\ BZ'V^{-1}X & B - BZ'V^{-1}ZB & BZ'V^{-1}y & BZ'V^{-1} & BZ'V^{-1}K \\ y'V^{-1}X & -y'V^{-1}ZB & y'V^{-1}y & y'V^{-1} & y'V^{-1}K \\ \hline V^{-1}X & -V^{-1}ZB & V^{-1}y & V^{-1} & V^{-1}K \\ K'V^{-1}X & -K'V^{-1}ZB & K'V^{-1}y & K'V^{-1} & K'V^{-1}K \end{array}\right].$$

Further sweeping the matrix in (2.14) on its first $p+q_1$ rows that correspond to the "fixed" effects leads to

$$(2.15)\quad Q_T^{\text{REML}}([X : Z : y : I : K]; G) = \left[\begin{array}{ccc|cc} (X'V^{-1}X)^{-1} & -UZB & \hat{\beta}^* & U & UK \\ -BZ'U' & B - BZ'WZB & \hat{\gamma}^* & BZ'W & BZ'WK \\ -\hat{\beta}^{*'} & \hat{\gamma}^{*'} & y'Wy & y'W & y'WK \\ \hline -U' & -WZB & Wy & W & WK \\ -K'U' & -K'WZB & K'Wy & K'W & K'WK \end{array}\right].$$

where $W = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$ and $U = (X'V^{-1}X)^{-1}X'V^{-1}$.

$|X'V^{-1}X|$ and $|Z'\Lambda^{-1}Z + B^{-1}|$ are obtained during the sweep operations. If the first sweep is on the rows that correspond to the random time-invariant effects and the second one is on the rows that correspond to the "fixed" effects, and if we denote the successive pivots by $\text{pv}_i$, then

$$(2.16)\qquad |Z'\Lambda^{-1}Z + B^{-1}| = \prod_{i=p+q_1+1}^{p+q_1+g+q_2} \text{pv}_i \quad \text{and} \quad |X'V^{-1}X| = \prod_{i=1}^{p+q_1} \text{pv}_i.$$

If, during the second sweep on the rows that correspond to the "fixed" effects, we avoid sweeping the lower right matrix in the partition in (2.14), we obtain $Q_T^{\text{ML}}([X : Z : y : I : K]; G)$, which is identical to $Q_T^{\text{REML}}([X : Z : y : I : K]; G)$, except that $W$ is replaced by $V^{-1}$ in the lower right matrix in the partition shown in (2.15).

Selected components of the matrix $Q_T^{\mathrm{REML}}$ in (2.15) and its maximum likelihood counterpart $Q_T^{\mathrm{ML}}$ are used in the inference of the model in (2.1). Standard theory on linear mixed models (see, for example, Gallas and Harville (1981), Searle *et al.* (1992)) implies the following results:

1) *Estimation of effects*: The Best Linear Unbiased Estimator (BLUE) of $\beta^*$, the Best Linear Unbiased Predictor (BLUP) of $\gamma^*$ and their mean square errors (MSEs) are:

$$\begin{aligned} &\hat{\beta}^* = (X'V^{-1}X)^{-1}X'V^{-1}y; \qquad \mathrm{MSE}(\hat{\beta}^*) = (X'V^{-1}X)^{-1} \\ &\hat{\gamma}^* = BZ'Wy; \qquad \mathrm{MSE}(\hat{\gamma}^*) = B - BZ'WZB. \end{aligned} \tag{2.17}$$

These expressions are obtained by taking $C = [X : Z : y]$ in the augmented KF. $Q_T^{\mathrm{REML}}$ and $Q_T^{\mathrm{ML}}$ are of dimension $p + g + q + 1$.

2) *Computation of restricted and concentrated loglikelihoods*: Using the results in (2.6), (2.10) and (2.16), we obtain the restricted and the concentrated loglikelihoods in (2.7) as

$$\begin{aligned} l_R &= -\frac{T-p-q_1}{2} - \frac{1}{2}\det_T - \frac{1}{2}\sum_{i=1}^{p+g+q} \ln \mathrm{pv}_i - \frac{1}{2}\ln|B| - \frac{1}{2}y'Wy \\ l_C &= -\frac{T}{2} - \frac{1}{2}\det_T - \frac{1}{2}\sum_{i=p+q_1+1}^{p+g+q} \ln \mathrm{pv}_i - \frac{1}{2}\ln|B| - \frac{1}{2}y'Wy \end{aligned} \tag{2.18}$$

They are computed by taking $C = [X : Z : y]$, and are maximized with respect to the relevant parameters.

3) *Signal extraction*: BLUPs of the vector of signals and of the noise components, as well as their respective MSEs, are given by

$$\begin{aligned} &\hat{s} = y - \sigma^2 Wy; \qquad \mathrm{MSE}(\hat{s}) = \sigma^2 I - \sigma^4 W \\ &\hat{\nu} = \sigma^2 Wy; \qquad \mathrm{MSE}(\hat{\nu}) = \sigma^2 I - \sigma^4 W. \end{aligned} \tag{2.19}$$

This is carried out by taking $C = [X : Z : y : I]$. The triangular structure of $I$ is exploited throughout the recursions (see (2.11)). $Q_T^{\mathrm{REML}}$ is of dimension $p+g+q+T+1$. While the number of pivots is small, the number of elements that are changed during each sweep is of order $T^2$. However, if one's only interest is in obtaining $\mathrm{MSE}(\hat{s}_t)$ but not their covariances, then it is not necessary to sweep the off-diagonal elements of $\Lambda^{-1}$. This reduces the number of computations to the order of $T$.

4) *Disturbance smoothing*: The BLUP of the vector of disturbances and its associated MSE are derived from the mixed model representation in (2.4):

$$\begin{aligned} &\hat{\xi} = \tilde{\Xi}K'Wy; \qquad \mathrm{MSE}(\hat{\xi}) = \tilde{\Xi} - \tilde{\Xi}K'WK\tilde{\Xi}; \\ &\hat{\xi}_t = \Xi_t K_t Wy; \qquad \mathrm{MSE}(\hat{\xi}_t) = \Xi_t - \Xi_t K_t W K_t' \Xi_t, \end{aligned} \tag{2.20}$$

where the $m \times T$ matrix $K_t$ is (see (2.5))

$$K_t = [0, \ldots, 0, k_{tt}, k_{t+1,t}, \ldots, k_{Tt}]. \tag{2.21}$$

The quantities needed for disturbance smoothing are stored in the $p+g+q+1+T+Tm$ square matrix $Q_T^{\mathrm{REML}}([X : Z : y : I : K]; G)$. The matrices $K_t W K_t'$ in (2.20) are the diagonal blocks of $K'WK$. $K$ has a triangular structure and simplifications analogous to (2.11) occur in the augmented Kalman filter. Furthermore, if the covariances between the smoothed disturbances are not of interest, one needs to sweep only the $Tm^2$ elements of the diagonal blocks of $K'\Lambda^{-1}K$ in $Q_T$.

*Comments*: The above results on the estimation of effects and the evaluation of the likelihood are given in De Jong (1991); running the augmented KF on $[X : Z : y]$ is equivalent to De Jong's diffuse Kalman Filter.

A disturbance smoother for state space models was first developed by Koopman (1993); his algorithm requires both forward and backward recursions. The contribution of our work consists of combining the Kalman Filter with results from mixed model theory and presenting a general unifying framework that ties together various special algorithms that have been proposed in the statistical literature for state space model inference. Our approach of smoothing the disturbance terms and estimating the signals does not require backward recursions; it is an alternative to the disturbance smoothing algorithm of Koopman.

Our algorithm provides the complete MSE matrices of all estimates; the covariance between any two disturbance or signal estimates is obtained directly, without the use of special recursions (as in De Jong and McKinnon (1988)). Our approach is very general; state smoothing, adjustments for missing observations, and forecasting are easily carried out within our framework by selecting the appropriate $C$ matrix; see Tsimikas and Ledolter (1997).

A drawback of our approach is the large storage requirement that depends on both the number of time points $T$ and the dimension of the state vector $q$. For example, with 100 observations in a model with a state vector of dimension $q = 2$, $Q_T^{\mathrm{REML}}([X : Z : y : I : K]; G)$ is a $(301 + p + g + q) \times (301 + p + g + q)$ matrix. However, considerable savings can be obtained during the matrix recursions by utilizing the simplifications that arise from the special structure of $I$ and $K$. The discussion under points 3) and 4) of this section shows that while the dimension of the matrix may be large, computational savings can be realized by processing only the relevant components; for further discussion, see Tsimikas and Ledolter (1997).

### 2.4 *An EM algorithm for the estimation of variance components*

Let us assume that $\Xi_t$, the variance of $\xi_t$ in model (2.1), does not depend on $t$. The EM algorithm for REML estimation of the variance components, $\sigma^2$ and $\Xi$, in the state space model with known transition matrix $\Phi_t$ can be derived from the theory for normal linear mixed models.

The $(r+1)$ iterates of the *REML estimates* of $\sigma^2$ and $\Xi$ are obtained from the $(r)$ iterates as:

$$
\begin{aligned}
\sigma^2(r+1) &= \frac{\sum[\hat{\nu}_t^2(r) + \mathrm{MSE}(\hat{\nu}_t(r))]}{T-p-q_1} \\
&= \sigma^2(r) + \sigma^4(r)\frac{[(W(r)y)'(W(r)y) - \mathrm{tr}(W(r))]}{T-p-q_1} \\
\Xi(r+1) &= \frac{\sum[\hat{\xi}_t(r)\hat{\xi}_t'(r) + \mathrm{MSE}(\hat{\xi}_t(r))]}{T} \\
&= \Xi(r) + \frac{\sum \hat{\xi}_t(r)\hat{\xi}_t'(r) - \Xi(r)(\sum K_t W(r) K_t')\Xi(r)}{T},
\end{aligned} \tag{2.22}
$$

where $\mathrm{tr}(W)$ denotes the trace of the matrix $W$. We follow Koopman (1993) who has shown how to use the disturbance smoother for EM estimation of variance components in state space models. The disturbance smoother from the previous section is used for the calculations, and all needed quantities are stored in $Q_T^{\mathrm{REML}}([X : Z : y : I : K]; G)$. The indexed notation $(r)$ denotes that the quantity is evaluated with the $(r)$ iterate of the parameters.

The EM iterations for obtaining *ML estimates* of the variance terms are similar, and they are given by:

$$
\begin{aligned}
\sigma^2(r+1) &= \sigma^2(r) + \sigma^4(r)\frac{[(W(r)y)'(W(r)y) - \mathrm{tr}(V^{-1}(r))]}{T} \\
\Xi(r+1) &= \Xi(r) + \frac{\sum \hat{\xi}_t(r)\hat{\xi}_t'(r) - \Xi(r)(\sum K_t V^{-1}(r) K_t')\Xi(r)}{T}.
\end{aligned} \tag{2.23}
$$

The matrix $Q_T^{\mathrm{ML}}([X : Z : y : I : K]; G)$ contains all needed quantities. $Q_T^{\mathrm{REML}}$ in (2.15) involves more sweeping calculations than $Q_T^{\mathrm{ML}}$ and, consequently, REML estimation requires more computer time. This is because the REML hierarchy adjusts the MSEs of the noise and disturbance terms for the estimation of fixed effects.

The EM iterations in (2.22) and (2.23) assume an arbitrary covariance matrix $\Xi$, but in many applications, in particular the structural time series models considered in Section 4, this matrix is diagonal. Assume that $\Xi$ has the following structure

$$
\Xi = \mathrm{diag}(\sigma_1^2 I_{m_1}, \sigma_2^2 I_{m_2}, \ldots, \sigma_h^2 I_{m_h}), \tag{2.24}
$$

where $m_1 + \cdots + m_h = m$. In other words, the disturbance vector at $t$, $\zeta_t = (\xi_{1t}', \xi_{2t}', \ldots, \xi_{ht}')'$, consists of $h$ mutually independent component vectors; the $m_j$-dimensional disturbance component $\xi_{jt}$ has a $N(0, \sigma_j^2 I_{m_j})$ distribution. Denote the $l$-th scalar disturbance in $\xi_{jt}$ by $\xi_{jt}^{(l)}$ and partition $K_t$ in (2.21) according to the partition of the disturbance vector; that is, $K_t = [K_{1t}' : \ldots : K_{ht}']'$. Then the EM iterations for the REML estimate of $\sigma_j^2$ become

$$
\begin{aligned}
\sigma_j^2(r+1) &= \frac{\sum_{t=1}^T \sum_{l=1}^{m_j}[(\hat{\xi}_{jt}^{(l)}(r))^2 + \mathrm{MSE}(\hat{\xi}_{jt}^{(l)}(r))]}{T m_j} \\
&= \sigma_j^2(r) + \sigma_j^4(r) \\
&\quad \cdot \frac{\sum_{t=1}^T [(K_{jt} W(r) y)'(K_{jt} W(r) y) - \mathrm{tr}(K_{jt} W(r) K_{jt}')]}{T m_j}.
\end{aligned} \tag{2.25}
$$

The iteration for the ML estimates of the variance components is similar; the only difference is that $\mathrm{tr}(K_{jt}W(r)K_{jt}')$ in (2.28) is replaced by $\mathrm{tr}(K_{jt}V^{-1}(r)K_{jt}')$.

The disturbance smoother also provides the necessary quantities for the implementation of the Scoring and Newton-Raphson algorithms; for details see Tsimikas and Ledolter (1997). An earlier paper by Koopman and Shephard (1992) shows how to utilize the disturbance smoother for the method of scoring.

## 3. State space models for repeated measures

Consider a multi-unit model for repeated measures in which each unit's response profile follows the state space model in (2.1). More specifically, the observation for the $i$-th unit $(i = 1, \ldots, k)$ at time $t$ $(t = 1, \ldots, n_i)$ is modeled as

$$
\begin{aligned}
&y_{it} = x_{it}'\beta + z_{it}'\gamma_i + h_t'\alpha_{it} + \nu_{it}; \qquad \nu_{it} \sim N(0, \sigma^2) \\
&\alpha_{it} = \Phi_{it}\alpha_{i,t-1} + R_{it}\xi_{it}; \qquad \xi_{it} \sim N(0, \Xi);
\end{aligned} \tag{3.1}
$$

all random components are assumed independent across units. Furthermore, we assume that the $i$-th unit's initial condition $\alpha_{i0}$ is drawn from a $N(\alpha_0, B_{22})$ distribution, and we allow it to be correlated with $\gamma_i$; let $\mathrm{cov}(\gamma_i, \alpha_{i0}') = B_{12}$.

After moving the initial states $\alpha_{i0}$ into the observation equations, we obtain

$$
\begin{aligned}
&y_{it} = x_{it}'\beta + z_{it}'\gamma_i + \left( h_t' \prod_{l=0}^{t-1} \Phi_{i,t-l} \right) \alpha_{i0} + h_t'\alpha_{it}^{\dagger} + \nu_{it} \\
&\alpha_{it}^{\dagger} = \Phi_{it}\alpha_{i,t-1}^{\dagger} + R_{it}\xi_{it}; \qquad \alpha_{i0}^{\dagger} = 0.
\end{aligned} \tag{3.2}
$$

We partition the initial condition for the $i$-th unit, $\alpha_{i0} = \alpha_0 + a_{i0}$, into a "fixed" part $\alpha_0$ of dimension $q$ which gets assigned a diffuse prior, and a random offset of dimension $q$ which is modeled as $a_{i0} \sim N(0, B_{22})$ and $\mathrm{cov}(\gamma_i, a_{i0}') = B_{12}$. More general situations where the initial conditions are related to between-unit covariates can be handled similarly. In general, initial conditions are regarded as time-invariant random effects and are handled within the Laird-Ware framework.

Let $\beta^* = (\beta' : \alpha_0')'$, $\gamma_i^* = (\gamma_i' : a_{i0}')'$, $x_{it}^{*'} = (x_{it}' : b_{it}')$, and $z_{it}^{*'} = (z_{it}' : b_{it}')$, where $b_{it}' = h_t' \prod_{l=0}^{t-1} \Phi_{i,t-l}$. Then we can express the $i$-th unit's response vector, $y_i = (y_{i1}, \ldots, y_{in_i})'$, as a mixed model

$$
y_i = X_i\beta^* + Z_i\gamma_i^* + K_i\xi_i + \nu_i, \tag{3.3}
$$

where $X_i = (x_{i1}^*, \ldots, x_{in_i}^*)'$, and $Z_i = (z_{i1}^*, \ldots, z_{in_i}^*)'$, $\xi_i = (\xi_{i1}', \ldots, \xi_{in_i}')'$ and $\nu_i = (\nu_{i1}, \ldots, \nu_{in_i})'$. The $n_i \times n_i m$ matrix $K_i$ is

$$
K_i = \begin{bmatrix} k_{i,11}' & 0 & \cdots & 0 \\ k_{i,21}' & k_{i,22}' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ k_{i,n_i1}' & k_{i,n_i2}' & \cdots & k_{i,n_in_i}' \end{bmatrix}, \tag{3.4}
$$

with $k'_{i,tt} = h'_t R_{it}$ and $k'_{i,st} = h'_s(\prod_{j=0}^{s-t-1} \Phi_{i,s-j}) R_{it}$, for $s > t$. The covariance matrix of $y_i$ is given by

$$V_i = Z_i B Z_i' + \Lambda_i, \tag{3.5}$$

where

$$\Lambda_i = K_i(I_{n_i} \otimes \Xi) K_i' + \sigma^2 I_{n_i}. \tag{3.6}$$

3.1 *REML estimation in the repeated measures model*

Our approach for calculating the restricted loglikelihood of the model in (3.1)–(3.3) processes the units sequentially and does not require the inversion of large matrices. This approach is useful when each unit involves more than just a small number of observations. Our algorithm, described in steps 1 through 4 given below, makes extensive use of the augmented KF in carrying out the Cholesky decomposition of $\Lambda_i^{-1}$. In the ML hierarchy and for the evaluation of the concentrated loglikelihood our algorithm is similar to that of Jones for the Laird-Ware model with ARMA errors. However, we believe that it is preferable to carry out the analysis in the REML hierarchy, since it accounts for the estimation of fixed effects.

*STEP* 1: For each unit $i$ we run the augmented KF recursions that correspond to the $i$-th unit's state space model,

$$\begin{aligned} y_{it} &= h_t' \alpha_{it}^{\dagger} + \nu_{it} \\ \alpha_{it}^{\dagger} &= \Phi_{it} \alpha_{i,t-1}^{\dagger} + R_{it} \xi_{it}; \qquad \alpha_{i0}^{\dagger} = 0 \end{aligned}$$

on the columns of $C = [X_i : Z_i : y_i]$. We let $G = \operatorname{diag}(0, B^{-1}, 0)$ and obtain $Q_{n_i}(i)([X_i : Z_i : y_i]; G)$. As a by-product of the KF recursions we calculate

$$\det_{n_i}(i) = \ln |\Lambda_i| = \sum_{t=1}^{n_i} \ln f_t(i). \tag{3.7}$$

*STEP* 2: We sweep $Q_{n_i}(i)$ on the rows that correspond to the time-invariant random effects (that is, rows $p+q+1$ through $p+q+g+q$). The result of this sweep is stored in $Q_{n_i}^R(i)$. The pivots of this sweep are used to calculate

$$\mathrm{d}(i) = \ln |Z_i' \Lambda_i^{-1} Z_i + B^{-1}| = \sum_{l=p+q+1}^{p+q+g+q} \ln \mathrm{pv}_l(i). \tag{3.8}$$

*STEP* 3: We repeat steps 1 and 2 for $i = 1, \ldots, k$ and accumulate each subject's contribution to the loglikelihood through the recursions

$$\begin{aligned} &\text{(a)}\ S_i = S_{i-1} + Q_{n_i}^R(i), \qquad \text{(b)}\ \mathrm{d}_i = \mathrm{d}_{i-1} + \mathrm{d}(i), \\ &\text{(c)}\ \det_i = \det_{i-1} + \det_{n_i}(i); \end{aligned} \tag{3.9}$$

the above recursions are initialized with zeroes. At the end of the recursions we obtain the matrix $S_k$, whose relevant parts are shown below:

$$S_k = \left[\begin{array}{c|c|c} \sum_{i=1}^{k} X_i' V_i^{-1} X_i & & \sum_{i=1}^{k} X_i' V_i^{-1} y_i \\ \hline & & \\ \hline \sum_{i=1}^{k} y_i' V_i^{-1} X_i & & \sum_{i=1}^{k} y_i' V_i^{-1} y_i \end{array}\right]. \tag{3.10}$$

*STEP* 4: We sweep $S_k$ on its first $p + q$ rows, which correspond to the fixed effects. Elements of the resulting matrix

$$S_k^* = \left[\begin{array}{c|c|c} S_{11}^* & & s_{13}^* \\ \hline & & \\ \hline -s_{13}^{*'} & & s_{33}^* \end{array}\right] \tag{3.11}$$

are used to obtain the BLUE of the fixed effect vector and its variance

$$\begin{aligned} \hat{\beta}^* &= \left(\sum_{i=1}^{k} X_i' V_i^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{k} X_i' V_i^{-1} y_i\right) = s_{13}^* \\ \operatorname{var}(\hat{\beta}^*) &= \left(\sum_{i-1}^{k} X_i' V_i^{-1} X_i\right)^{-1} = S_{11}^*. \end{aligned} \tag{3.12}$$

As a by-product of this sweep we obtain

$$\ln \left|\sum_{i=1}^{k} X_i' V_i^{-1} X_i\right| = \sum_{l=1}^{p+q} \ln \mathrm{pv}_l^*, \tag{3.13}$$

where $\mathrm{pv}_l^*$ is the $l$-th pivot during the sweep.

Let $N = \sum_{i=1}^{k} n_i$. The restricted and concentrated loglikelihoods can be expressed as

$$\begin{aligned} l_R &= -\frac{N-p-q}{2} \ln 2\pi - \frac{1}{2} \sum_{l=1}^{p+q} \ln \mathrm{pv}_l^* - \frac{1}{2} \det_k - \frac{1}{2} \mathrm{d}_k - \frac{k}{2} \ln |B| - \frac{1}{2} s_{33}^* \\ l_C &= -\frac{N}{2} \ln 2\pi - \frac{1}{2} \det_k - \frac{1}{2} \mathrm{d}_k - \frac{k}{2} \ln |B| - \frac{1}{2} s_{33}^*. \end{aligned} \tag{3.14}$$

Maximizing the restricted loglikelihood in (3.14) with respect to the parameters yields REML estimates.

3.2 *BLUPs and the EM algorithm for REML estimation*

Mixed model theory can be used to obtain the BLUPs of the $i$-th unit's time-invariant random effects vector $\gamma_i^*$, the disturbance $\xi_i$, the signal $s_i$ and the noise vector $\nu_i$, as well as their mean square errors. The results of Section 2, and those in De Jong (1991) and Koopman (1993), apply. Define $\hat{e}_i = y_i - X_i\hat{\beta}^*$ where $\hat{\beta}^*$ is given in (3.12); then

$$
\begin{aligned}
&\hat{\gamma}_i^* = BZ_i'V_i^{-1}\hat{e}_i; \qquad \mathrm{MSE}(\hat{\gamma}_i^*) = B - BZ_i'W_{ii}Z_iB \\
&\hat{\xi}_i = (I_{n_i} \otimes \Xi)K_i'V_i^{-1}\hat{e}_i; \\
&\qquad\qquad \mathrm{MSE}(\hat{\xi}_i) = (I_{n_i} \otimes \Xi) - (I_{n_i} \otimes \Xi)K_i'W_{ii}K_i(I_{n_i} \otimes \Xi), \\
&\hat{s}_i = y_i - \sigma^2 V_i^{-1}\hat{e}_i; \qquad \mathrm{MSE}(\hat{s}_i) = \sigma^2 I_{n_i} - \sigma^4 W_{ii} \\
&\hat{\nu}_i = \sigma^2 V_i^{-1}\hat{e}_i; \qquad \mathrm{MSE}(\hat{\nu}_i) = \sigma^2 I_{n_i} - \sigma^4 W_{ii};
\end{aligned}
\tag{3.15}
$$

where $W_{ii} = V_i^{-1} - V_i^{-1}X_i(\sum_{i=1}^k X_i'V_i^{-1}X_i)^{-1}X_i'V_i^{-1}$.

The second equation in (3.15) implies

$$
\mathrm{MSE}(\hat{\xi}_{it}) = \Xi - \Xi K_{it}W_{ii}K_{it}'\Xi \tag{3.16}
$$

where $K_{it} = [0, \ldots, 0, k_{i,tt}, \ldots, k_{i,n_it}]$ is an $m \times n_i$ matrix; its elements are given in (3.4). The BLUPs in (3.15) depend on the variance components and the parameters in the transition matrices $\Phi_{it}$. If these quantities are unknown we estimate them by maximizing the restricted loglikelihood. Substituting these estimates into (3.15) results in *Empirical Bayes estimates. In particular, $\hat{s}_i$ is an Empirical Bayes* smoother for the $i$-th unit's response profile.

For known transition matrices $\Phi_{it}$ the BLUPs and their MSEs in equation (3.15) are sufficient to implement an EM algorithm for REML estimation of the between-unit and the within-unit variance components in model (3.3). Assuming general covariance structures for $B$ and $\Xi$, we obtain the following EM algorithm:

$$
\begin{aligned}
\sigma^2(r+1) &= \sigma^2(r) + \sigma^4(r)\frac{\sum_i ssq(V_i^{-1}(r)\hat{e}_i(r)) - \sum_i \mathrm{tr}(W_{ii}(r))}{(\sum_i n_i - p - q)} \\
B(r+1) &= B(r) + \frac{\sum_i \hat{\gamma}_i^*(r)\hat{\gamma}_i^{*\prime}(r) - B(r)[\sum_i Z_i'W_{ii}(r)Z_i]B(r)}{k} \\
\Xi(r+1) &= \Xi(r) + \frac{\sum_{i,t} \hat{\xi}_{it}(r)\hat{\xi}_{it}'(r) - \Xi(r)[\sum_{i,t} K_{it}W_{ii}(r)K_{it}']\Xi(r)}{\sum_i n_i},
\end{aligned}
\tag{3.17}
$$

where *ssq* denotes the sum of squares of the vector elements. For ML estimation we simply replace $(\sum_i n_i - p - q)$ by $\sum_i n_i$, and $W_{ii}$ by $V_i^{-1}$.

If $\Xi$ has the diagonal structure shown in (2.24), then simplifications similar to the ones shown in (2.25) occur. Partition $K_{it}$ according to the partition of the disturbances, i.e. $K_{it} = [K_{i,1t}', \ldots, K_{i,ht}']'$. The EM iterations for REML estimation of the variance components in $\Xi$ then become

$$
\sigma_j^2(r+1) = \sigma_j^2(r) + \sigma_j^4(r) \cdot \frac{\sum_{i=1}^k \sum_{t=1}^{n_i}[(K_{i,jt}W_{ii}(r)y)'(K_{i,jt}W_{ii}(r)y) - \mathrm{tr}(K_{i,jt}W_{ii}(r)K_{i,jt}')]}{m_j(\sum_{i=1}^k n_i)}. \tag{3.18}
$$

The augmented KF is used extensively in the derivation of the needed quantities. The EM algorithm for REML or ML estimates involves the following steps:

*STEP* 1: Run the KF on $[X_i : Z_i : K_i : I_{n_i} : \hat{e}_i]$ to obtain for the $i$-th unit $Q_{n_i}(i)([X_i : Z_i : K_i : I_{n_i} : \hat{e}_i]; G)$.

*STEP* 2: Sweep $Q_{n_i}(i)([X_i : Z_i : K_i : I_{n_i} : \hat{e}_i]; G)$ on the rows that correspond to the random time-invariant effects and obtain the matrix $Q_{n_i}^{\mathrm{ML}}(i)([X_i : Z_i : K_i : I_{n_i} : \hat{e}_i]; G)$:

(3.19) $Q_{n_i}^{\mathrm{ML}}(i)$

$$= \left[\begin{array}{cccc|c} X_i'V_i^{-1}X_i & -X_i'V_i^{-1}Z_iB & X_i'V_i^{-1}K_i & X_i'V_i^{-1} & X_i'V_i^{-1}\hat{e}_i \\ BZ_i'V_i^{-1}X_i & B - BZ_i'V_i^{-1}Z_iB & BZ_i'V_i^{-1}K_i & BZ_i'V_i^{-1} & \hat{\gamma}_i^* \\ K_i'V_i^{-1}X_i & -K_i'V_i^{-1}Z_iB & K_i'V_i^{-1}K_i & K_i'V_i^{-1} & K_i'V_i^{-1}\hat{e}_i \\ V_i^{-1}X_i & -V_i^{-1}Z_iB & V_i^{-1}K_i & V_i^{-1} & V_i^{-1}\hat{e}_i \\ \hline \hat{e}_i'V_i^{-1}X_i & -\hat{e}_i'V_i^{-1}Z_iB & \hat{e}_i'V_i^{-1}K_i & \hat{e}_i'V_i^{-1} & \hat{e}_i'V_i^{-1}\hat{e}_i \end{array}\right].$$

This matrix contains the contribution of the $i$-th unit to the EM algorithm for ML estimation of the variance components; see equation (3.17). Note that for ML estimation alone it is not necessary to process $X_i$.

*STEP* 3: Replace $X_i'V_i^{-1}X_i$ by $\sum_{i=1}^{k} X_i'V_i^{-1}X_i$, which we obtained during the evaluation of $l_R$, and sweep the left upper part of the matrix in (3.19) (that is, the whole matrix, except for the last row and column) on its first $p+q$ rows (rows that correspond to the fixed effects). This results in $Q_{n_i}^{\mathrm{REML}}(i)([X_i : Z_i : K_i : I_{n_i} : \hat{e}_i]; G)$, with the relevant elements:

(3.20) $Q_{n_i}^{\mathrm{REML}}(i)$

$$= \left[\begin{array}{cccc|c} \left(\sum_{i=1}^{k} X_i'V_i^{-1}X_i\right)^{-1} & & & & \\ & B - BZ_i'W_{ii}Z_iB & & & \hat{\gamma}_i^* \\ & & K_i'W_{ii}K_i & & K_i'V_i^{-1}\hat{e}_i \\ & & & W_{ii} & V_i^{-1}\hat{e}_i \\ \hline & & & & \end{array}\right].$$

It contains the contribution of the $i$-th unit to the EM algorithm for REML estimation of the variance components. Considerable computational savings are achieved by skipping the sweeping operations in *STEP* 2 and *STEP* 3 for the elements that are not on the rows or columns of the pivots and which are not directly needed for the implementation of the EM algorithm.

After convergence of the EM algorithm has been achieved a final pass of the augmented KF over the $i$-th unit performs Empirical Bayes estimation of $\gamma_i^*$ and smoothes the $i$-th unit's profile through Empirical Bayes estimation of $s_i$.

## 4. Local continuous polynomial trends and polynomial smoothing splines

In this section we extend the implementation of the methodology in Sections 2 and 3 to continuous-time structural time series models. Such models are useful as they can accomodate unequally spaced observations. We first consider the *single-unit case*, which amounts to inference on a single unit without any "borrowing" of information from other units. The continuous local polynomial trend of order $m$ satisfies the transition equation

$$\frac{d}{dt}\alpha(t) = A^{(m)}\alpha(t) + \eta(t), \tag{4.1}$$

where

$$A^{(m)} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

is an $m \times m$ matrix and $\eta(t)$ is an $m \times 1$ multivariate disturbance process with mean zero and covariance matrix $Q^{(m)} = \text{diag}(\sigma_1^2, \ldots, \sigma_m^2)$. The observation equation at time $t_\tau$, for $\tau = 1, \ldots, T$, is

$$y_\tau = [1 : 0 : \ldots : 0]\alpha(t_\tau) + \nu_\tau, \tag{4.2}$$

where $\nu_\tau$ is a measurement error term with variance $\sigma^2$. The measurement error is assumed independent of the white noise disturbance process $\eta(t)$.

Let $\delta_\tau = t_\tau - t_{\tau-1}$ be the elapsed time between the $\tau$-th and the $(\tau-1)$-th observation. In discrete time the state space representation of the continuous local polynomial trend model is given by:

$$\begin{aligned} y_\tau &= [1 : 0 : \ldots : 0]\alpha_\tau + \nu_\tau \\ \alpha_\tau &= \Phi_\tau^{(m)}\alpha_{\tau-1} + \eta_\tau; \qquad \eta_\tau \sim N(0, Q_\tau^{(m)}), \end{aligned} \tag{4.3}$$

where (see Harvey (1989), p. 484),

$$\begin{aligned} \Phi_\tau^{(m)} &= \exp(A^{(m)}\delta_\tau) = I + \sum_{k=1}^{\infty} \frac{(A^{(m)}\delta_\tau)^k}{k!} \\ Q_\tau^{(m)} &= \int_0^{\delta_\tau} \exp(A^{(m)}(\delta_\tau - s))Q^{(m)} \exp(A'^{(m)}(\delta_\tau - s))ds. \end{aligned} \tag{4.4}$$

Using the structure of the matrices $A^{(m)}$ and $Q^{(m)}$, and the fact that $(A^{(m)})^k = 0$ for $k > m-1$, we obtain, after some algebra,

$$\Phi_\tau^{(m)} = \begin{bmatrix} 1 & \delta_\tau & \frac{\delta_\tau^2}{2} & \cdots & \frac{\delta_\tau^{m-1}}{(m-1)!} \\ 0 & 1 & \delta_\tau & \cdots & \frac{\delta_\tau^{m-2}}{(m-2)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \delta_\tau \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}; \qquad Q_\tau^{(m)} = \sum_{l=1}^{m} \sigma_l^2 \begin{bmatrix} \Theta_\tau^{(l)} & 0_l \\ 0_l' & 0_{ll} \end{bmatrix}, \tag{4.5}$$

where $\Theta_\tau^{(l)}$ is an $l \times l$ positive definite matrix, with element in the $a$-th row and $b$-th column given by

$$\Theta_\tau^{(l)}(a,b) = \frac{\delta_\tau^{2l-a-b+1}}{(2l-a-b+1)(l-a)!(l-b)!}; \tag{4.6}$$

$0_l$ is an $l \times (m-l)$ matrix of zeroes and $0_{ll}$ is an $(m-l) \times (m-l)$ matrix of zeroes. For $\sigma_l^2 = 0$, for $l = 1, \ldots, m-1$, and $\sigma_m^2 \neq 0$ we obtain the state space representation of a polynomial smoothing spline; see Wecker and Ansley (1983). For an explicit expression for $\sigma_1^2 = 0$ and $\sigma_2^2 \neq 0$, see Anderson and Jones (1995).

The elements in the non-diagonal matrix $Q_\tau^{(m)}$ are functions of $m$ variance components. We can reparameterize this matrix and the corresponding state transition equation in terms of independent random disturbances as follows. Let $L_\tau^{(l)}$ be the upper Cholesky factor of $\Theta_\tau^{(l)}$; that is, $L_\tau^{(l)}$ is an upper triangular $l \times l$ matrix such that $\Theta_\tau^{(l)} = L_\tau^{(l)} L_\tau^{(l)'}$. Then the state space model in (4.3) can be written as

$$\begin{aligned} y_\tau &= [1:0:\ldots:0]\alpha_\tau + \nu_\tau \\ \alpha_\tau &= \Phi_\tau^{(m)}\alpha_{\tau-1} + R_\tau^{(m)}\xi_\tau; \qquad \xi_\tau \sim N(0, \mathrm{diag}(\sigma_1^2, \sigma_2^2 I_2, \ldots, \sigma_m^2 I_m)), \end{aligned} \tag{4.7}$$

where

$$\xi_\tau = \begin{bmatrix} \xi_{1\tau} \\ \vdots \\ \xi_{m\tau} \end{bmatrix}; \qquad R_\tau^{(m)} = [L_\tau^{*(1)} : \ldots : L_\tau^{*(l)} : \ldots : L_\tau^{*(m)}]. \tag{4.8}$$

The disturbance vector $\xi_\tau$ is of dimension $1 + \cdots + m = m(m+1)/2$; each subvector $\xi_{l\tau}$ in the disturbance vector $\xi_\tau$ is of dimension $l$. The matrix $R_\tau^{(m)}$ is an $m \times m(m+1)/2$ matrix; it depends on $\delta_\tau$ and does not contain unknown parameters. The $m \times l$ matrix $L_\tau^{*(l)}$ in the above expression is given by

$$L_\tau^{*(l)} = \begin{bmatrix} L_\tau^{(l)} \\ 0 \end{bmatrix}. \tag{4.9}$$

It is simply the upper Cholesky factor of $\Theta_\tau^{(l)}$ augmented by $(m-l)$ rows of zeroes; furthermore $L_\tau^{*(m)} = L_\tau^{(m)}$.

By moving the initial condition $\alpha_0 = (\alpha_{00}, \alpha_{01}, \ldots, \alpha_{0,m-1}')$ to the observation equation we obtain the following state space representation

$$\begin{aligned} y_\tau &= \alpha_{00} + \sum_{k=1}^{m-1} \alpha_{0k} \frac{t_\tau^k}{k!} + [1:0:\ldots:0]\alpha_\tau^\dagger + \nu_\tau \\ \alpha_\tau^\dagger &= \Phi_\tau^{(m)}\alpha_{\tau-1}^\dagger + R_\tau^{(m)}\xi_\tau; \qquad \xi_\tau \sim N(0, \mathrm{diag}(\sigma_1^2, \sigma_2^2 I_2, \ldots, \sigma_m^2 I_m)), \end{aligned} \tag{4.10}$$

where $\alpha_0^\dagger = 0$.

The widely-used polynomial smoothing spline, in which only $\sigma_m^2 \neq 0$, is a special case of this model. Its state space representation is given by

$$
\begin{aligned}
y_\tau &= \alpha_{00} + \sum_{k=1}^{m-1} \alpha_{0k} \frac{t_\tau^k}{k!} + [1 : 0 : \ldots : 0]\alpha_\tau^\dagger + \nu_\tau \\
\alpha_\tau^\dagger &= \Phi_\tau^{(m)} \alpha_{\tau-1}^\dagger + L_\tau^{(m)} \xi_\tau; \qquad \xi_\tau \sim N(0, \sigma_m^2 I_m).
\end{aligned} \tag{4.11}
$$

The model in (4.10) can be extended to handle seasonality with period $s$ by adding a trigonometric stochastic seasonal component. The transition equation for a discrete realization of a continuous time trigonometric seasonal model is

$$
\psi_\tau = \begin{bmatrix} \psi_{1,\tau} \\ \psi_{2,\tau} \end{bmatrix} = \begin{bmatrix} \cos(\lambda_{t_\tau}\delta_\tau) & \sin(\lambda_{t_\tau}\delta_\tau) \\ -\sin(\lambda_{t_\tau}\delta_\tau) & \cos(\lambda_{t_\tau}\delta_\tau) \end{bmatrix} \begin{bmatrix} \psi_{1,\tau-1} \\ \psi_{2,\tau-1} \end{bmatrix} + \delta_\tau^{1/2} \begin{bmatrix} \omega_{1t} \\ \omega_{2t} \end{bmatrix}; \tag{4.12}
$$

$$
(\omega_{1t}, \omega_{2t})' \sim N(0, \sigma_\omega^2 I_2).
$$

$\lambda_{t_\tau}$ in the transition matrix is the seasonal frequency corresponding to $t_\tau$; i.e. $\lambda_{t_\tau} = 2\pi j(t_\tau)/s$, $j(t_\tau)$ indicating the season the $\tau$-th observation was collected; see Harvey ((1989), p. 487). By moving the initial condition to the observation equation and combining the seasonal and trend components one obtains the following structural model

$$
\begin{aligned}
y_\tau &= \psi_{10} \cos\left(\sum_{m=1}^{\tau} \lambda_{t_m}\delta_m\right) + \psi_{20} \sin\left(\sum_{m=1}^{\tau} \lambda_{t_m}\delta_m\right) \\
&\quad + \alpha_{00} + \sum_{k=1}^{m-1} \alpha_{0k} \frac{t_\tau^k}{k!} + [1 : 0 : \ldots : 0 : 1 : 0]\alpha_\tau^\dagger + \nu_\tau \\
\alpha_\tau^\dagger &= \begin{bmatrix} \Phi_\tau^{(m)} & 0 \\ 0 & S_\tau \end{bmatrix} \alpha_{\tau-1}^\dagger + [R_\tau^{(m)} : \delta_\tau^{1/2} I_2]\xi_\tau; \\
\xi_\tau &\sim N(0, \operatorname{diag}(\sigma_1^2, \sigma_2^2 I_2, \ldots, \sigma_m^2 I_m, \sigma_\omega^2 I_2))
\end{aligned} \tag{4.13}
$$

where $S_\tau$ is the transition matrix in (4.13). Further extensions to the continuous structural model are possible, such as including stochastic cycles (Harvey (1989), pp. 487–488).

In the *multi-unit* case we can postulate a Laird-Ware-type model to estimate the unit specific trend and seasonal coefficients, as well as the effects of covariates on the unit profiles. This model specifies

$$
\begin{aligned}
y_{i\tau} &= x_{i\tau}^* \beta^* + z_{i\tau}^* \gamma_i^* + h' \alpha_{i\tau}^\dagger + \nu_{i\tau} \\
\alpha_{i\tau}^\dagger &= \begin{bmatrix} \Phi_{i\tau}^{(m)} & 0 \\ 0 & S_{i\tau} \end{bmatrix} \alpha_{i\tau-1}^\dagger + [R_{i\tau}^{(m)} : \delta_{i\tau}^{1/2} I_2]\xi_{i\tau}
\end{aligned} \tag{4.14}
$$

where $h$ is a vector of zeroes except for ones at the 1st and $(m+1)$-st position. The vector $\beta^*$ includes fixed covariate effects and population coefficients for the trend and seasonal components, whereas $\gamma_i^*$ contains unit-specific random offsets. The

matrices $\Phi_{i\tau}^{(m)}$, $S_{i\tau}$ and $R_{i\tau}^{(m)}$ are functions of $\delta_{i\tau}$. The methodology in Section 3 is directly applicable. We can evaluate the restricted loglikelihood, and we can use the EM recursions in (3.18) to obtain ML or REML estimates of the variance components.

## REFERENCES


Anderson, S. J. and Jones, R. H. (1995). Smoothing splines for longitudinal data, *Statistics in Medicine*, **14**, 1235–1248.

Ansley, C. F. and Kohn, R. (1985). Estimation, filtering and smoothing in state space models with incompletely specified initial conditions, *Ann. Statist.*, **13**, 1286–1316.

Ansley, C. F. and Kohn, R. (1990). Filtering and smoothing in state space models with partially diffuse initial conditions, *J. Time Ser. Anal.*, **11**, 275–293.

De Jong, P. (1988). A cross-validation filter for time series models, *Biometrika*, **75**, 594–600.

De Jong, P. (1989). Smoothing and interpolation with the state space model, *J. Amer. Statist. Assoc.*, **84**, 1085–1088.

De Jong, P. (1991). The diffuse Kalman filter, *Ann. Statist.*, **19**, 1073–1083.

De Jong, P. and McKinnon, M. (1988). Covariances for smoothed estimates in state space models, *Biometrika*, **75**, 601–602.

Hartley, H. O. and Rao, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model, *Biometrika*, **54**, 93–108.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems, *J. Amer. Statist. Assoc.*, **72**, 320–340.

Jones, R. H. (1993). *Longitudinal Data with Serial Correlation: A State Space Approach*, Chapman Hall, London.

Kohn, R. and Ansley, C. F. (1986). Estimation, prediction and interpolation for ARIMA models with missing data, *J. Amer. Statist. Assoc.*, **81**, 751–761.

Koopman, S. J. (1993). Disturbance smoother for state space models, *Biometrika*, **80**, 117–126.

Koopman, S. J. and Shephard, N. (1992). Exact score for time series models in state space form, *Biometrika*, **79**, 823–826.

Laird, N. M. and Ware, J. H. (1982). Random effects models for longitudinal data, *Biometrics*, **38**, 963–974.

Laird, N. M., Lang, N. and Stram, D. (1987). Maximum likelihood computations with repeated measures: application of the EM algorithm, *J. Amer. Statist. Assoc.*, **76**, 860–869.

Sallas, W. M. and Harville, D. A. (1981). Best linear recursive estimation for mixed linear models, *J. Amer. Statist. Assoc.*, **82**, 97–105.

Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance Components*, Wiley, New York.

Tsimikas, J. and Ledolter, J. (1994). REML and best linear unbiased prediction in state space models, *Comm. Statist. Theory Methods*, **23**(8), 2253–2268.

Tsimikas, J. and Ledolter, J. (1997). Mixed model representation of state space models: new smoothing results and their application to REML estimation, *Statistica Sinica* (forthcoming).

Wecker, W. E. and Ansley, C. F. (1983). The signal extraction approach to nonlinear regression and spline smoothing, *J. Amer. Statist. Assoc.*, **78**, 81–89.

Wilson, D. P. (1995). Longitudinal data analysis for linear Gaussian models with random disturbed-highest-derivative-polynomial subject effects, *Statistics in Medicine*, **14**, 1219–1233.