

# Modelling Unequally Spaced Bivariate Growth Curve with Kalman Filter Approach

Qianyu Dang<sup>†</sup>, Stewart Anderson<sup>†</sup>, Lingshi Tan<sup>‡</sup> and Sati Mazumdar<sup>†</sup>

August 2, 2004

## Abstract

In many clinical studies, patients are followed over time with their responses measured longitudinally. Using mixed model theory, one can characterize these data using a wide array of across subject models. A state-space representation of the mixed effects model and use of the Kalman filter allows one great flexibility in choosing the within error correlation structure even in the presence of missing or unequally spaced observations. Furthermore, using the state-space approach, one can avoid inverting large matrices resulting in efficient computation. The approach also allows one to make detailed inference about the error correlation structure. We consider a bivariate situation where the longitudinal responses are unequally spaced and assume that the within subject errors follows a continuous first order autoregressive (CAR(1)) structure. Since a large number of nonlinear parameters need to be estimated, the modelling strategy and numerical techniques are critical in the process. We developed both a Visual Fortran<sup>®</sup> and a SAS<sup>®</sup> program for modelling such data. A simulation study was conducted to investigate the robustness of the model assumptions. We also use data from a psychiatric study data to demonstrate our model fitting procedure.

*Key Words:* Kalman filter, State space approach, Longitudinal data, Growth curve, Mixed effects model.

<sup>†</sup> Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, U.S.A.

<sup>‡</sup> Pfizer, Inc, New York, New York 10017, U.S.A.

The authors wish to thank Dr. C. F. Reynolds and P. Houck for providing study data and valuable comments on the analysis results.

# 1 Introduction

Longitudinal data are common in randomized clinical trials while patients are followed over time. In many cases, these data are not equally spaced and always measured with correlated within subject errors. In the case of equally spaced model, the first order autoregressive model, AR(1), has been developed for the within subject error structure (Potthoff and Roy, 1964; Chi and Reinsel, 1989). But if there is no general sample interval, the error structures need to be modelled as continuous-time first order autoregressive (CAR(1)). One approach is fitting a random effects model commonly used in longitudinal data (Jones and Ackerson, 1990), and combining with state space approach with Kalman filter to compute the exact likelihood for univariate or multivariate responses (Jones, 1993). In this paper, we applied this to fit bivariate unequally spaced longitudinal measurements with CAR(1) error. Corresponding Visual Fortran<sup>®</sup> and SAS<sup>®</sup> IML programs were developed. A simulation study was conducted to validate the analysis and a psychiatric study data was applied to demonstrate the model fitting strategy.

## 2 The Mixed Effects Model and Kalman Filter

### 2.1 The Multivariate Mixed Effects Model

The univariate linear mixed model is usually defined as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i . \quad (2.1.1)$$

Here  $\mathbf{y}_i$  is a  $n_i \times 1$  column vector of the observations measured on  $i^{th}$  individual,  $\boldsymbol{\beta}$  is a  $p \times 1$  vector of regression coefficients and  $\boldsymbol{\gamma}_i$  is a  $q \times 1$  random effect vector assumed to be independently distributed as  $N(\mathbf{0}, \mathbf{B})$ .  $\mathbf{X}_i$  and  $\mathbf{Z}_i$  are  $n_i \times p$  and  $n_i \times q$  design matrices for fixed and random effects.

It is straight forward to extend the model to handle multivariate outcomes. If we assume for a given individual  $m$  distinct measurements are taken at each time and they are unequally spaced. The observations for different individuals are independent, and every subject can have different number of time points and observations. The mixed model for subject  $i$  at time  $j$  becomes

$$\mathbf{y}_{ij} = \boldsymbol{\phi}\mathbf{x}_{ij} + \boldsymbol{\psi}_i\mathbf{z}_{ij} + \mathbf{u}_{ij} \quad (2.1.2)$$

where  $\mathbf{y}_{ij}$  is a vector of  $m$  observations for subject  $i$  at time  $t_{ij}$ . Let  $\mathbf{y}_i = (y_{i1}, \dots, y_{in(i)})$ , we write the model for subject  $i$  as

$$\mathbf{y}_i = \boldsymbol{\Phi}\mathbf{X}'_i + \boldsymbol{\Psi}_i\mathbf{Z}'_i + \mathbf{U}_i \quad (2.1.3)$$

where  $i = 1, \dots, N$ , and  $\mathbf{X}'_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{in(i)})$ ,  $\mathbf{Z}'_i = (\mathbf{z}_{i1}, \dots, \mathbf{z}_{in(i)})$  are the design matrices for fixed and random effects. Normally for growth curve models,  $\mathbf{x}'_{ij}$  and  $\mathbf{z}'_{ij}$  can have the forms of  $(1, t_{ij}, t_{ij}^2, \dots)$ . Using the “*vec*” operator to let  $\mathbf{y}_i = \text{vec}(\mathbf{y}_i) = (y_{i1}, \dots, y_{in(i)})'$ ,  $\boldsymbol{\beta} = \text{vec}(\boldsymbol{\Phi})$ ,  $\boldsymbol{\gamma}_i = \text{vec}(\boldsymbol{\Psi}_i)$  and  $\boldsymbol{\epsilon}_i = \text{vec}(\mathbf{U}_i)$ , expression (2.1.3) can be rewritten as

$$\mathbf{y}_i = (\mathbf{X}_i \otimes \mathbf{I}_m)\boldsymbol{\beta} + (\mathbf{Z}_i \otimes \mathbf{I}_m)\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i . \quad (2.1.4)$$

We assume  $\boldsymbol{\gamma}_i$  is independently distributed across subjects as  $N(\mathbf{0}, \mathbf{B})$ ,  $\mathbf{B}$  is a general covariance matrix with off-diagonal elements.  $\boldsymbol{\epsilon}_i$  is a vector of within individual errors, we also assume  $\boldsymbol{\epsilon}_i$ 's are independent cross individuals and distributed as  $N(\mathbf{0}, \mathbf{W}_i)$ . It is well known that for model (2.1.4), given covariance matrices  $\mathbf{B}$  and  $\mathbf{W}_i$ , the  $-2 \log(\text{likelihood})$  is

$$l = \sum_i \{mn_i \log(2\pi) + \log | (\mathbf{Z}_i \otimes \mathbf{I}_m)\mathbf{B}(\mathbf{Z}_i \otimes \mathbf{I}_m)' + \mathbf{W}_i | + [\mathbf{y}_i - (\mathbf{X}_i \otimes \mathbf{I}_m)\boldsymbol{\beta}]' [(\mathbf{Z}_i \otimes \mathbf{I}_m)\mathbf{B}(\mathbf{Z}_i \otimes \mathbf{I}_m)' + \mathbf{W}_i]^{-1} [\mathbf{y}_i - (\mathbf{X}_i \otimes \mathbf{I}_m)\boldsymbol{\beta}] \} \quad (2.1.5)$$

The estimate for the fixed effects coefficient  $\boldsymbol{\beta}$  is

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left\{ \sum_i (\mathbf{X}_i \otimes \mathbf{I}_m)' [(\mathbf{Z}_i \otimes \mathbf{I}_m) \mathbf{B} (\mathbf{Z}_i \otimes \mathbf{I}_m)' + \mathbf{W}_i]^{-1} (\mathbf{X}_i \otimes \mathbf{I}_m) \right\}^{-1} \\ &\quad \times \left\{ \sum_i (\mathbf{X}_i \otimes \mathbf{I}_m)' [(\mathbf{Z}_i \otimes \mathbf{I}_m) \mathbf{B} (\mathbf{Z}_i \otimes \mathbf{I}_m)' + \mathbf{W}_i]^{-1} \mathbf{y}_i \right\} \end{aligned} \quad (2.1.6)$$

and the univariate model is just a special case with  $m = 1$ . The maximum likelihood estimates can be obtained through a nonlinear optimization program with respect to elements in  $\mathbf{B}$  and  $\mathbf{W}_i$ . Since  $\mathbf{B}$  has to be nonnegative definite, so we perform a Cholesky decomposition  $\mathbf{B} = \mathbf{U}'\mathbf{U}$  and let the optimization program working the upper triangular matrix  $\mathbf{U}$ . For most longitudinal data, the error within subjects are considered to be correlated, to solve this problem, Potthoff and Roy(1964) followed the idea of Harville(1976) and specified a first order autoregressive within subject error structure (AR(1)) for equally spaced data. Assuming the errors have mean zero, the error of subject  $i$  at time  $j$  can be represented as

$$\varepsilon_{ij} = \phi \varepsilon_{ij-1} + \eta_{ij} \quad (2.1.7)$$

where  $\eta_{ij}$ 's are a serial of independent identically distributed random variables with mean zero.  $\boldsymbol{\eta}$  is usually assumed to be Gaussian.  $\phi$  is the autoregressive coefficient between two consecutive errors, in order to make the procedure to be stationary, it is necessary to confine that  $-1 < \phi < 1$ . This shows the error is only directly correlated with the error of one previous observation. A more general form is the autoregressive moving average model (ARMA(p,q)), where the within subject error with zero mean is defined as

$$\varepsilon_{ij} = \sum_{k=1}^p \phi_k \varepsilon_{i(j-k)} + \sum_{k=1}^q \theta_k \eta_{j-k} . \quad (2.1.8)$$

Here the error is directly correlated with errors of  $p$  previous observations and is a function of  $q$  previous errors.

## 2.2 The Kalman filter approach for within subject errors

There are many ways to handle AR or ARMA errors for longitudinal data, but most of them include reversing a  $n_i \times n_i$  matrix during the direct approach for obtaining maximum (ML) or restricted maximum likelihood (REML) estimation. The state space representation and Kalman filter was first introduced to avoid that. Kalman (1960) and Kalman and Bucy (1961) developed a recursive optimal estimation procedure for discrete and continuous time state space models. The form of ARMA(p,q) models for univariate time series observations  $x_t$  is as follows:

$$x_t = \sum_{k=1}^p \alpha_k x_{t-k} + \sum_{k=1}^q \delta_k \eta_{t-k} + \eta_t \quad (2.2.1)$$

where  $\eta_t$  is uncorrelated "white noise" with mean 0 and variance  $\sigma_\eta^2$  per unit time. Jones (1986) extended this model with a recursive state equation for an arbitrary continuous time interval, the result is the continuous time ARMA(p,q) model, or the CARMA model. If we let the state of the process vector be  $\mathbf{s}(t) = (x(t|t), x(t|t+1), \dots, x(t+m-1|t))'$  where  $m = \max(p, q+1)$ , then the state equation becomes

$$\mathbf{s}(t) = \mathbf{F}\mathbf{s}(t-1) + \mathbf{G}\boldsymbol{\eta}_t \quad (2.2.2)$$

where  $\mathbf{F} = \begin{bmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & 1 \\ \alpha_m & \alpha_{m-1} & \dots & \alpha_1 \end{bmatrix}$  and  $\mathbf{G} = [1 \quad g_2 \quad \dots \quad g_{m-1} \quad g_m]$ . Here

$\alpha_m$  is the autocorrelation coefficient between  $x(t+m-1|t)$  and  $x(t-1|t-1)$  while  $\mathbf{G}$  is the matrix multiplying the standard random inputs. The observation equation at time t is

$$\varepsilon_i(t) = [1 \quad \delta_2 \quad \dots \quad \delta_{p-1}] \times \begin{bmatrix} x(t|t) \\ x(t+1|t) \\ \dots \\ x(t+m-1|t) \end{bmatrix} + \nu(t) \quad (2.2.3)$$

where  $\delta_k = 0$  for all  $k > q$  and  $\nu(t)$  is the observational error. The state equation shows how state vectors are correlated with each other while the observation equation

reveals the relationship of state vectors and observations. For multivariate continuous time interval, the state space representation in the derivative form is

$$\frac{d\mathbf{s}(t)}{dt} = \mathbf{F}\mathbf{s}(t) + \mathbf{g}\eta_t, \quad \boldsymbol{\varepsilon}(t) = \mathbf{h}\mathbf{s}(t) + \boldsymbol{\nu}(t). \quad (2.2.4)$$

In this paper, we also consider the random effects in the mixed effects model. Duncan and Horn (1972) suggested incorporating the random effects in the state space vector, so the random effect coefficient  $\boldsymbol{\gamma}_i$  was augmented to the state vector, the complete derivative state representation for subject  $i$  is

$$\begin{aligned} \frac{d}{dt} \begin{bmatrix} \mathbf{s}_i(t) \\ \boldsymbol{\gamma}_i \end{bmatrix} &= \begin{bmatrix} \mathbf{F} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{s}_i(t) \\ \boldsymbol{\gamma}_i \end{bmatrix} + \begin{bmatrix} \mathbf{g} \\ \mathbf{0} \end{bmatrix} \eta_i(t) \\ \boldsymbol{\xi}_i(t_j) &= \begin{bmatrix} \mathbf{h} & \mathbf{Z}_i(t_j) \end{bmatrix} \begin{bmatrix} \mathbf{s}_i(t) \\ \boldsymbol{\gamma}_i \end{bmatrix} + \boldsymbol{\nu}_i(t_j) \end{aligned} \quad (2.2.5)$$

Usually  $\mathbf{s}_i(t)$  represents the vector correlated error terms and  $\boldsymbol{\xi}_i$  is the observation subtracting the fixed effects. The vector of  $\mathbf{Z}_i(t_j)$  is the  $j^{\text{th}}$  row of design matrix  $\mathbf{Z}_i$ . The initial value of state covariance matrix  $\mathbf{B} = \mathbf{U}'\mathbf{U}$  must be given and then optimized by maximum likelihood, so this is basically an empirical Bayes approach.

In order to calculate the continuous time stationary model, Kalman and Bucy (1961) developed a method through integrating the state equation over the intervals between observations, with random effects removed, we have

$$\mathbf{s}(t_j) = e^{(t_j - t_{j-1})\mathbf{F}} \cdot \mathbf{s}(t_{j-1}) \quad (2.2.6)$$

The matrix exponential is defined as

$$e^{\mathbf{M}} = \mathbf{I} + \sum_{k=1}^{\infty} \frac{\mathbf{M}^k}{k!} \quad (2.2.7)$$

In the multivariate situation, we need to evaluate the matrix exponentials numerically, so we write  $\mathbf{F}$  in the diagonal form  $\mathbf{F} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}^{-1}$  where  $\boldsymbol{\Lambda}$  is a diagonal matrix of the eigenvalues.  $\mathbf{C}$  is the matrix of right eigenvectors of  $\mathbf{F}$ . Notice that for univariate AR(1), all these will be scalars. The equation above can be rewritten as

$$\mathbf{s}(t_j) = \mathbf{C}e^{(t_j - t_{j-1})\mathbf{F}}\mathbf{C}^{-1}\mathbf{s}(t_{j-1}) \quad (2.2.8)$$

If we define  $\tilde{\mathbf{s}}(t) = \mathbf{C}^{-1}\mathbf{s}(t)$  as a rotated state vector, then the transformed model becomes

$$\frac{d}{dt}\tilde{\mathbf{s}}(t) = \Lambda\tilde{\mathbf{s}}(t) + \mathbf{C}^{-1}\mathbf{G}\boldsymbol{\eta}_t \quad \text{and} \quad \boldsymbol{\varepsilon}(t) = \mathbf{C}\tilde{\mathbf{s}}(t) + \boldsymbol{\nu}(t) . \quad (2.2.9)$$

Thus, the rotated state can be used in the recursion with the prediction in the interval,  $[t_{j-1}, t_j]$ , being

$$\tilde{\mathbf{s}}(t_j|t_{j-1}) = \boldsymbol{\Phi}(t_j - t_{j-1})\tilde{\mathbf{s}}(t_j - t_{j-1}) \quad (2.2.10)$$

where  $\boldsymbol{\Phi}(t_j - t_{j-1}) = e^{(t_j - t_{j-1})\Lambda}$ . The rotated state vector can be used during the recursive process and then rotated back at the end by the inverse equation  $\mathbf{s}(t) = \mathbf{C}\tilde{\mathbf{s}}(t)$ . Besides predicting the state vector over any arbitrary interval, it is also necessary to compute the prediction error introduced by  $\boldsymbol{\eta}_t$ , the complex prediction error over a time step of length  $t_j - t_{j-1}$  is

$$\int_{t_{j-1}}^{t_j} \boldsymbol{\Phi}(t_j - t_{j-1})\mathbf{C}^{-1}\mathbf{G}\boldsymbol{\eta}_t dt \quad (2.2.11)$$

with covariance matrix

$$\mathbf{Q}(t_j - t_{j-1}) = \int_{t_{j-1}}^{t_j} \mathbf{C}^{-1}\mathbf{G}\mathbf{G}'(\mathbf{C}^*)^{-1}\boldsymbol{\Phi}^*(t_j - t_{j-1})dt \quad (2.2.12)$$

where ‘\*’ denotes the complex conjugate. Let  $\mathbf{K} = \mathbf{C}^{-1}\mathbf{G}\mathbf{G}'(\mathbf{C}^*)^{-1}$  so that the  $k^{th}$  row and  $l^{th}$  column element of  $\mathbf{Q}(t_j - t_{j-1})$  is

$$\mathbf{Q}_{kl}(t_j - t_{j-1}) = \mathbf{K}_{kl} \frac{e^{(\lambda_k + \lambda_l^*)(t_j - t_{j-1})} - 1}{(\lambda_k + \lambda_l^*)} \quad \text{if } \lambda_k + \lambda_l^* \neq \mathbf{0} \quad (2.2.13)$$

$$\mathbf{Q}_{kl}(t_j - t_{j-1}) = \mathbf{K}_{kl}(t_j - t_{j-1}) \quad \text{if } \lambda_k + \lambda_l^* = \mathbf{0} .$$

By using the rotated state vectors, the state space representation (2.2.5) can be rewritten as

$$\begin{aligned} \frac{d}{dt} \begin{bmatrix} \tilde{\mathbf{s}}_i(t_j) \\ \boldsymbol{\gamma}_i \end{bmatrix} &= \begin{bmatrix} \boldsymbol{\Phi}(t_j - t_{j-1}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{s}}_i(t_{j-1}) \\ \boldsymbol{\gamma}_i \end{bmatrix} + \begin{bmatrix} \mathbf{G} \\ \mathbf{0} \end{bmatrix} \boldsymbol{\eta}_i(t) \\ \boldsymbol{\xi}_i(t_j) &= [\mathbf{h}\mathbf{C} \quad \mathbf{Z}_i(t_j)] \begin{bmatrix} \tilde{\mathbf{s}}_i(t_j) \\ \boldsymbol{\gamma}_i \end{bmatrix} + \nu_i(t_j) \end{aligned} \quad (2.2.14)$$

where  $\boldsymbol{\xi}(t_j)$  is normal with mean  $\mathbf{0}$  and covariance matrix  $\mathbf{Q}(t_j - t_{j-1})$ . The initial values of the state vector covariance can be defined as  $\mathbf{Q}(\infty)$ , as if the time interval goes to infinity. The elements of the initial rotated state covariance matrix are

$$\mathbf{P}_r(0|0) = \mathbf{Q}_{kl}(\infty) = -\frac{\mathbf{K}_{kl}}{\lambda_k + \lambda_l^*} . \quad (2.2.15)$$

Jones(1993) mentioned that if  $\lambda_k + \lambda_l^*$  has non-negative real part, then the process is non-stationary, and the corresponding element will be a diffuse prior. One solution is to set it to a very large number, like  $10^4$  times of the sample variance. All the formulas above can be applied to both univariate and multivariate models.

To integrate the random effects into the recursive procedure, we denote the state and state transition matrix to be:

$$\mathbf{S}(t) = \begin{bmatrix} \tilde{\mathbf{s}}_i(t_j) \\ \boldsymbol{\gamma}_i \end{bmatrix} \quad \boldsymbol{\Phi}(\delta t) = \begin{bmatrix} e^{(\boldsymbol{\Lambda}\delta t)} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

and the initial state covariance matrix and the covariance matrix of the random input to the state is

$$\mathbf{P}(0|0) = \begin{bmatrix} \mathbf{P}_r(0|0) & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{B} \end{bmatrix} \quad \mathbf{Q}(\delta t) = \begin{bmatrix} \mathbf{Q}_r(\delta t) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

The steps of Kalman recursion for continuous time multivariate first order autoregressive model (CAR(1)) can be written as

1. The one-step prediction to next observation time, the state transition matrix should be diagonal.

$$\mathbf{S}(t_j|t_j - \delta t) \leftarrow \boldsymbol{\Phi}(\delta t) \times \mathbf{S}(t_j - \delta t|t_j - \delta t)$$

2. The covariance matrix for the prediction is

$$\mathbf{P}(t_j|t_j - \delta t) \leftarrow \boldsymbol{\Phi}(\delta t)\mathbf{P}(t_j - \delta t|t_j - \delta t)\boldsymbol{\Phi}^*(\delta t) + \mathbf{Q}(\delta t)$$



3. The innovation matrix  $\mathbf{I}$  at time  $t_j$  for subject  $i$ . Set  $\mathbf{H}(t_j) \leftarrow [\mathbf{C} \quad \mathbf{Z}_i(t_{ij})]$

where  $\mathbf{C}$  is the right eigenvalue of  $\mathbf{F}$  as mentioned above.

$$\mathbf{I}(t_j) \leftarrow [\mathbf{X}_i(t_j) \quad \mathbf{y}_i(t_j)] - \mathbf{H}(t_j)\mathbf{S}(t_j|t_j - \delta t)$$

where  $\mathbf{y}_i(t_j)$  is the row vector of  $j^{th}$  observation, and  $\mathbf{X}_i(t_j)$  is the corresponding row  $j$  of design matrix.

4. The innovation variance is

$$\mathbf{V}(t_j) = \mathbf{H}(t_j)\mathbf{P}(t_j|t_j - \delta t)\mathbf{H}^*(t_j) + \mathbf{R}(t_j)$$

5. Update the matrix needed to calculate  $\boldsymbol{\beta}$  and accumulate the determinant term in the  $-2 \log(\text{likelihood})$ ,

$$\mathbf{M} \leftarrow \mathbf{M} + \mathbf{I}^*(t_j)\mathbf{V}^{-1}(t_j)\mathbf{I}(t_j)$$

$$\text{DET} \leftarrow \text{DET} + \ln|\mathbf{V}(t_j)|$$

where the matrix  $\mathbf{M}$  is  $\begin{bmatrix} \tilde{\mathbf{X}}'\tilde{\mathbf{X}} & \tilde{\mathbf{X}}'\tilde{\mathbf{y}} \\ \tilde{\mathbf{y}}'\tilde{\mathbf{X}} & \tilde{\mathbf{y}}'\tilde{\mathbf{y}} \end{bmatrix}$

6. Calculate the Kalman Gain,

$$\mathbf{KG}(t_j) \leftarrow \mathbf{P}(t_j|t_j - \delta t)\mathbf{H}^*(t_j)\mathbf{V}^{-1}(t_j)$$

7. Update the estimated state vector,

$$\mathbf{S}(t_j|t_j) \leftarrow \mathbf{S}(t_j|t_j - \delta t) + \mathbf{KG}(t_j)\mathbf{I}(t_j)$$

8. Update the covariance matrix of the state vector,

$$\mathbf{P}(t_j|t_j) \leftarrow \mathbf{P}(t_j|t_j - \delta t) - \mathbf{KG}^*(t_j)\mathbf{KG}(t_j)\mathbf{V}(t_j)$$

For a bivariate CAR(1) model,  $\mathbf{V}(t_j)$  is a  $2 \times 2$  matrix while in the univariate model, it is just a scalar. The recursion above should be carried out with double precision to avoid the accumulation of roundoff errors. The state vector and its covariance matrix will be re-initialized for every subject while  $\mathbf{M}$  and  $\text{DET}$  are accumulated over all the subjects to compute the overall value of  $-2 \log(\text{likelihood})$ . At the end of each recursion, one performs a Cholesky factorization of  $\mathbf{M}$ , and then replaces the upper

portion of  $\mathbf{M}$  by  $[\mathbf{T} \quad \mathbf{b}]$  where  $\tilde{\mathbf{X}}'\tilde{\mathbf{X}} = \tilde{\mathbf{T}}'\tilde{\mathbf{T}}$  and  $\mathbf{b} = (\tilde{\mathbf{T}}')^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}}$ . Thus, the unadjusted regression sums of squares can be obtained as  $\mathbf{b}'\mathbf{b} = \tilde{\mathbf{y}}'\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}}$ . Consequently, the residual sum of squares can be obtained as  $\text{RSS} = \tilde{\mathbf{y}}'\tilde{\mathbf{y}} - \mathbf{b}'\mathbf{b}$ . The  $-2 \log(\text{likelihood})$  then becomes

$$l = n_T \log(2\pi) + \text{DET} + \text{RSS} \quad (2.2.16)$$

where  $n_T$  is the total number of observations across all subjects. Notice that in the last step of the recursive procedure, a subtraction is performed. This operation can cause the estimated covariance matrix,  $\mathbf{P}(t_j|t_j)$ , to be negative definite. When that situation occurs, one must reassign the initial values for the  $\mathbf{F}$ ,  $\mathbf{B}$  and  $\mathbf{G}$  matrices.

### 3 A Model Fitting Strategy

For bivariate processes where the outcomes are linear functions of time and which have a CAR(1) error structure, the number of nonlinear parameters that need to be estimated ranges from 10 to 16. Hence, numerical stability may be a problem if we fit all of the parameters at once. Also, in most of the cases we considered, the Hessian matrices for the nonlinear parameters were not positive definite and so significance testing could not be performed.

The best way to build the model is to start from two independent univariate models and then add off-diagonal elements to the  $\mathbf{B}$ ,  $\mathbf{F}$  and  $\mathbf{G}$  matrices in sequence. The best model is ultimately selected based on minimizing AIC values. Since two independent univariate models are initially determined for the response variables, the covariance between the intercepts and slopes of these two variables are added one at a time in the  $\mathbf{B}$  matrix. The matrix,  $\mathbf{F}$ , is usually not symmetric which means that previous values of variable 1 as correlated with present values of variable 2 are not necessarily the same as previous values of variable 2 as correlated with with present values of

variable 1. The last step in the estimation of these three matrices is the addition of off-diagonal elements in the  $\mathbf{G}$  matrix.

Finally, we have found that including an independent observational error matrix,  $\mathbf{R}$ , often significantly decreases the AIC value, but it also caused numerical instabilities in the Kalman recursion. Hence, it is better to fit diagonal elements of  $\mathbf{R}$  at the end of the model selection procedure. Due the independent behavior usually associated with observational errors, it is reasonable to assume there are no off-diagonal elements in the  $\mathbf{R}$  matrix under most circumstances.

## 4 Model Validation by Simulation

We examined characteristics of the bivariate model estimation procedure by use of a simulation where we employed the IML language in SAS<sup>®</sup>8e. Our procedure for generating bivariate data with a CAR(1) error structure relied on generating a bivariate continuous process which was measured at time intervals of unequal length and which was observed with a Gaussian error structure. First, we generated the numbers of observations for the subjects as uniformly distributed values ranging from 5 to 20. Each unequal time interval was simulated from a random exponential distribution with a scale parameter of 10 and then rounded up to the next highest integer. The random effects vector,  $\boldsymbol{\gamma}$ , which consisted of an intercept and slope for each of two outcome variables, were created from a multinormal distribution,  $N(\mathbf{0}, \mathbf{B})$ , where the covariance matrix. To generate the process,  $\mathbf{B}$  was factorized as  $\mathbf{B} = \mathbf{U}'\mathbf{U}$  where  $\mathbf{U}$  is an upper triangular matrix. We chose the  $\mathbf{U}$  to be

$$\mathbf{U} = \begin{bmatrix} 1 & 0.2 & 0.05 & 0 \\ 0 & 1 & 0 & 0.05 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 \end{bmatrix}.$$

Table 1: simulation results for nonlinear parameters

|                   |                   |                   |                   |                   |                   |                   |
|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Parameters        | $\mathbf{F}(1,1)$ | $\mathbf{F}(2,2)$ | $\mathbf{F}(1,2)$ | $\mathbf{G}(1,1)$ | $\mathbf{G}(2,2)$ | $\mathbf{U}(1,1)$ |
| Input value       | -0.6              | -0.4              | 0.1               | 1.5               | 1.5               | 1                 |
| Mean of estimates | -0.620            | -0.397            | 0.102             | 1.518             | 1.493             | 0.985             |
| Standard error    | 0.0845            | 0.0474            | 0.0301            | 0.0956            | 0.0807            | 0.116             |
| Parameters        | $\mathbf{U}(1,2)$ | $\mathbf{U}(1,3)$ | $\mathbf{U}(2,2)$ | $\mathbf{U}(2,4)$ | $\mathbf{U}(3,3)$ | $\mathbf{U}(4,4)$ |
| Input value       | 0.2               | 0.05              | 1                 | 0.05              | 0.5               | 0.5               |
| Mean of estimates | 0.203             | 0.0574            | 0.960             | 0.0576            | 0.0491            | 0.487             |
| Standard error    | 0.185             | 0.0656            | 0.157             | 0.0742            | 0.0359            | 0.0410            |

Table 2: simulation results for linear parameters

|                   |                           |                           |                           |                           |
|-------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Parameters        | $\boldsymbol{\beta}(1,1)$ | $\boldsymbol{\beta}(1,2)$ | $\boldsymbol{\beta}(2,1)$ | $\boldsymbol{\beta}(2,2)$ |
| Input value       | 10                        | 0.8                       | 15                        | 0.6                       |
| Mean of estimates | 10.005                    | 0.8022                    | 14.995                    | 0.5984                    |
| Standard error    | 0.1367                    | 0.0512                    | 0.1503                    | 0.0494                    |

The fixed effect,  $\boldsymbol{\beta}$ , was chosen as  $\begin{bmatrix} 10 & 0.8 \\ 15 & 0.6 \end{bmatrix}$ . The parameters of the time series were chosen as having a CAR(1) autocorrelation matrix,  $\mathbf{F} = \begin{bmatrix} -0.6 & -0.1 \\ 0 & -0.4 \end{bmatrix}$ , with eigenvector  $\mathbf{C} = \begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix}$  and eigenvalues  $\boldsymbol{\Lambda} = [-0.6 \quad -0.4]$ . The  $\mathbf{G}$  matrix was  $\begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix}$ .

The final simulated responses were computed as  $Y_j = \boldsymbol{\beta}\mathbf{T}_j + \boldsymbol{\gamma}\mathbf{T}_j + \mathbf{E}_j$  where  $\mathbf{T}_j = [1 \quad t_j]'$ . Notice that  $\mathbf{Y}_j$  and  $\mathbf{E}_j$  are  $2 \times 1$  vectors and  $\boldsymbol{\beta}, \boldsymbol{\gamma}$  are  $2 \times 2$  matrix. A total of 500 sets of data were simulated, each consisting of 100 subjects. As was mentioned above, each subject had 5 to 20 observations.

The results of our simulations for the nonlinear and linear parameters are summarized in Tables 1 and 2, respectively. Each table consists of the true value and the mean and standard errors of over all the 500 simulated sets of data.

The simulation results show that both the nonlinear and linear estimates are fairly close to their true value. The standard errors of the off-diagonal elements estimates are

larger than on-diagonal elements for both  $\mathbf{F}$  and  $\mathbf{U}$  matrix, so estimating interactions between two variables is less stable compared to other nonlinear and linear parameters. However, there were about 30 sets of the data in which the Kalman recursion did not converge initially. Accordingly, by changing the initial values of the nonlinear parameters, all of them converged successfully.

## 5 A Bivariate Example

Maintenance therapies in late life depression (MTLD) (Reynolds et al, 1999) were treatments for patients aged 60 to 90 years old who had depression. The study consisted of three phases. The first was an acute treatment phase where 187 patients with recurrent depression were enrolled and treated with a full dose of nortriptyline (NT) and weekly interpersonal psychotherapy (IPT) to achieve a remission of depressive symptoms. The maximum length of this phase was 26 weeks. The patients' responses were measured by Hamilton Depression Rating Scale (17-item). Among all the patients initially enrolled, only those who achieved "full remission" (HDRS score = 10 for 3 consecutive weeks) were eligible to enter the next phase. Two response variables were selected to represent the recovery pattern. The first was the General Life Functioning (GLF) score, which measured the patients' quality of life profiles and was an important indicator for treatment response. The second response was the blood test results for NT concentration, which was an indicator of the patients' metabolism and also was a measure of compliance of their medicine taking. These two measurements were taken at the time the participants came for their weekly treatment. Because patients often missed their appointments for various reasons, the observations are unequally spaced. Moreover, these two variables are most likely to be correlated.

Following the model building strategy previously described, we first fit two uni-

Table 3: Summary of Models

| Fixed effect | Nonlinear parameters included  | AIC value |
|--------------|--|-----------|
| linear       | $G_{11}, G_{22}, F_{11}, F_{22}, B_{11}, B_{22}, B_{33}, B_{44}, B_{13}, B_{24}$                                 | 32031.19  |
| quadratic    | $G_{11}, G_{22}, F_{11}, F_{22}, B_{11}, B_{22}, B_{33}, B_{44}, B_{13}, B_{24}$                                 | 31897.51  |
| quadratic    | $G_{11}, G_{22}, F_{11}, F_{22}, B_{11}, B_{22}, B_{33}, B_{44}, B_{13}, B_{24}, B_{12}$                         | 31897.23  |
| quadratic    | $G_{11}, G_{22}, F_{11}, F_{22}, B_{11}, B_{22}, B_{33}, B_{44}, B_{13}, B_{24}, B_{12}, B_{34}$                 | 31898.01  |
| quadratic    | $G_{11}, G_{22}, F_{11}, F_{22}, B_{11}, B_{22}, B_{33}, B_{44}, B_{13}, B_{24}, B_{34}$                         | 31897.49  |
| quadratic    | $G_{11}, G_{22}, F_{11}, F_{22}, F_{12}, B_{11}, B_{22}, B_{33}, B_{44}, B_{13}, B_{24}, B_{12}$                 | 31883.82  |
| quadratic    | $G_{11}, G_{22}, F_{11}, F_{22}, F_{21}, B_{11}, B_{22}, B_{33}, B_{44}, B_{13}, B_{24}, B_{12}$                 | 31891.91  |
| quadratic    | $G_{11}, G_{22}, F_{11}, F_{22}, F_{21}, F_{12}, B_{11}, B_{22}, B_{33}, B_{44}, B_{13}, B_{24}, B_{12}$         | 31885.56  |
| quadratic    | $G_{11}, G_{22}, G_{21}, F_{11}, F_{22}, F_{12}, B_{11}, B_{22}, B_{33}, B_{44}, B_{13}, B_{24}, B_{12}$         | 31885.56  |
| quadratic    | $G_{11}, G_{22}, G_{12}, F_{11}, F_{22}, F_{12}, B_{11}, B_{22}, B_{33}, B_{44}, B_{13}, B_{24}, B_{12}$         | 31885.56  |
| quadratic    | $G_{11}, G_{22}, F_{11}, F_{22}, F_{12}, B_{11}, B_{22}, B_{33}, B_{44}, B_{13}, B_{24}, B_{12}, R_{11}, R_{22}$ | 31762.89  |

variate models separately and found that quadratic models for the fixed effect yielded the best results. The random effects contained intercepts and slopes to accommodate general situations. The value of  $-2 \log(\text{likelihood})$  associated with the independent univariate models was 31865.51. Next, the off-diagonal elements of the  $\mathbf{B}$ ,  $\mathbf{F}$  and  $\mathbf{G}$  matrices were added in the model one by one. We found that by the addition of only  $\mathbf{B}(1, 2)$  and  $\mathbf{F}(1, 2)$ , the AIC value was minimized. Finally, adding the observational error term,  $\mathbf{R}$ , further lowered the AIC. The results of model selection are summarized in table 3:

The last model displayed above had the lowest AIC value and the resulting autoregressive matrix  $\mathbf{F}$  was  $\mathbf{F} = \begin{bmatrix} -0.01587 & 0.0016194 \\ 0 & -0.07935 \end{bmatrix}$  and the eigenvalue vector,  $\mathbf{C} = \begin{bmatrix} 1 & -0.02551 \\ 0 & 1 \end{bmatrix}$ .

Since the scheduled time interval between visits was 7 days, a discrete time autoregressive coefficients matrix between visits of every weeks can be written as

$$\begin{aligned} \Phi(7) &= \begin{bmatrix} 1 & -0.02551 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e^{-0.01587 \times 7} & 0 \\ 0 & e^{-0.07935 \times 7} \end{bmatrix} \begin{bmatrix} 1 & -0.02551 \\ 0 & 1 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 0.895 & 0.0082 \\ 0 & 0.574 \end{bmatrix} \end{aligned} \quad (5.1)$$

From the estimates of the  $\mathbf{F}$  and  $\mathbf{\Phi}$  matrices, we can see that the correlations of the errors between consecutive weekly visits are quite high for both variables and that the GLF score has less variability than do the NT levels. In addition, since the upper right hand corner of the  $\mathbf{F}$  matrix is nonzero and the component in the lower left hand corner is  $\mathbf{0}$ , the GLF score is affected by *both* the GLF and NT levels of the previous time, but the NT level is *only* affected by earlier results of NT levels. So the “driving force” of these two measurements is the NT level. In other words, the NT level could cause the GLF score to change, but is not a feedback system where both observations affect each other. It is difficult to obtain such a relationship between two response variables through the use of traditional mixed models. The matrix multiplying the random input into the state equation was

$$\mathbf{G} = \begin{bmatrix} 1.0618 & 0 \\ 0 & 11.6104 \end{bmatrix}$$

The variance-covariance matrix of random effects was

$$\mathbf{B} = \mathbf{U}'\mathbf{U} = \begin{bmatrix} 6.3267 & -14.571 & 0.01992 & 0 \\ -14.571 & 131.285 & -0.04587 & -0.42174 \\ 0.01992 & -0.04587 & 0.000075 & 0 \\ 0 & -0.42174 & 0 & 0.003121 \end{bmatrix}$$

Finally, the observational error was

$$\mathbf{R} = \begin{bmatrix} 9.4305 & 0 \\ 0 & 15.813 \end{bmatrix},$$

and the fixed effects were estimated as

| Parameters | Intercept | Slope   | Quadratic term |
|------------|-----------|---------|----------------|
| GLF score  | 37.203    | 0.09665 | -0.000287      |
| NT levels  | 74.915    | 0.22885 | -0.000963      |

From the estimates of the fixed effects, both variables show an upward trend and get flatter over time, indicating an improving and stabilizing process. But some of the elements of the variance–covariance matrix for the random effects are quite large, so that there was much variation between individuals in addition to the common trend.

Specific predictions can be made for each individual based on the model estimations above.

## 6 Discussion

Using the Kalman filter to compute the likelihood for a bivariate growth curve was introduced by Jones (1993) and its use for computing a bivariate mixed model was also employed by Tan (1993). Until now, however, not many programs and applications have been developed. This may be due to the poor numerical stability of the multivariate Kalman filter. Theoretically, a multivariate ARMA( $p, q$ ) model can be fit by the same algorithm, but numerically, it is very complicated because of the instability of the procedure and difficulty of finding universal minimum with too many unknown parameters.

Part of the numerical instability comes from the last step of the Kalman filter which involves a subtraction of one matrix from another which produces a new variance-covariance matrix for the state. However, one can not guarantee that for each time, a positive-definite matrix will be produced. So with a non-positive-definite variance matrix, the recursion will quickly disintegrate. We found that it happened more often when the initial values for the nonlinear parameters were far away from their true values. Consequently, good guesses for the initial parameters, especially the  $\mathbf{B}$  and  $\mathbf{G}$  matrices, are very important. Otherwise, one could try a “grid search” of different initial values. In addition, introducing an observational error  $\mathbf{R}$  in the early stages of model fitting could increase the instability of the algorithm. Thus, it is better to include this term at the end of the estimation process after all other parameters have been selected. Usually, the more unknown parameters that need to be estimated, the more difficult it is for the procedure to converge. We found that if the number of



nonlinear parameters was more than 15, then the procedure was often very unstable. One ultimate solution for this problem of numerical instability might be found in Bierman (1977) who suggested implementing the Kalman filter in a factorized form while sacrificing efficiency in computing time. But Bierman's algorithm was only implemented for a univariate model. Thus, some theoretical work still needs to be done to extend his procedure to fit multivariate models.

Despite all of these problems, the Kalman filter is still a useful tool for fitting multivariate growth curve models. Its use allows one to avoid inverting matrices with large dimensions and it also gives estimates of both the cross correlation and autocorrelation associated with two response variables. Also, both fixed and random effects can be easily estimated. Future work is to improve the numerical stability, so that this approach can handle more complicated models.

## References

- POTTHOFF , R.F. AND ROY, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve models. *Biometrika*. 51, 313-326.
- CHI, E.M. AND REINSEL, G.C. (1989). Models for longitudinal data with random effects and AR(1) errors. *J. Am. Statist. Assoc.* 84, 452-459.
- JONES, R.H. AND ACKERSON, L.M. (1990). Serial correlation in unequally spaced longitudinal data. *Biometrika*. 77, 721-731.
- JONES, R.H. (1993). *Longitudinal Data with Serial Correlation: A State-space Approach*. First edition, Chapman and Hall, London.
- HARVILLE D.A.(1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. *Annals of Statistics*. 4, 384-95.

- KALMAN R.E. (1960). A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng.* 82D, 35-45.
- KALMAN R.E. AND BUCY R.L. (1961). New results in linear filtering and prediction theory. *Trans. ASME J. Basic Eng.* 83D,95-108.
- JONES, R.H. (1986). Time series regression with unequally spaced data. *Handbook of Statistics, Vol.5: Time Series in the Time Domain* (E.J. Hannan, P.R. Krishnaiah and M.M. Rao, Eds.), North-Holland, 157-177.
- DUNCAN, D.B. AND HORN, S.D.(1972). Linear dynamic recursive estimation from the viewpoint of regression analysis. *Journal of American Statistical Association.* 67, 815-21.
- REYNOLDS C.F., FRANK E., PEREL J.M., IMBER S.D., CORNES C., MILLER M.D., ET AL (1999). Nortriptyline and interpersonal psychotherapy as maintenance therapies for recurrent major depression. *Journal of American Medical Association.* 281, 39-45.
- BIERMAN, G.J.(1977). *Factorization methods for discrete sequential estimation, Series: Mathematics in Science and Engineering*, Vol. 128, Academic Press, New York.
- LINGSHI TAN (1993). A multivariate growth curve model with random effects and CAR(1) errors. Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Doctoral Dissertation.