

# Mixed-Effects State Space Models for Analysis of Longitudinal Dynamic Systems

Dacheng Liu <sup>\*</sup>      Xu-Feng Niu <sup>†</sup>      Hulin Wu <sup>‡</sup>

## Abstract

The rapid development of new biotechnologies allows us to deeply understand the biomedical dynamic systems in more details and at a cellular level. Many of the subject-specific biomedical systems can be described by a set of differential or difference equations which are similar to engineering dynamic systems. In this paper, motivated by HIV dynamic studies, we propose a class of mixed-effects state space models based on the longitudinal feature of the dynamic systems. State space models with mixed-effects components are very flexible in modelling the serial correlation of within-subject observations and between-subject variations. The Bayesian approach and the maximum likelihood methods for standard mixed-effects models and state space models are modified and investigated for estimating unknown parameters in the proposed mixed-effects state space models. In the Bayesian approach the full conditional distributions are derived and the Gibbs sampler is constructed to explore the posterior distributions. For the maximum likelihood method we develop a Monte Carlo EM algorithm with a Gibbs sampler step to approximate the conditional expectations in the E-step. Simulation results indicate that the two methods perform well. We apply the mixed-effects state space model to a data set from an AIDS clinical trial to illustrate the proposed methodologies. The proposed models and methods may also have potential applications in other biomedical system analyses such as tumor dynamics in cancer research and genetic regulatory network modeling.

*Key words and phrases:* State space models, mixed-effects models, parameter estimation, EM algorithm, Gibbs sampler, Kalman filter.

---

<sup>\*</sup>Dacheng Liu is postdoctoral fellow, Department of Biostatistics and Computational Biology, School of Medicine and Dentistry, University of Rochester, 601 Elmwood Ave., Box 630, Rochester, New York 14642.

<sup>†</sup>Xu-Feng Niu is Professor, Department of Statistics, Florida State University, Tallahassee, FL 32306-4330.

<sup>‡</sup>Hulin Wu is Professor, Department of Biostatistics and Computational Biology, School of Medicine and Dentistry, University of Rochester, 601 Elmwood Ave., Box 630, Rochester, New York 14642.

**Short title:** Mixed-Effects State-Space Models.

# 1 Introduction

State space modelling has been used mainly in time series data analysis. It has found its applications in many areas, such as economics, engineering, biology etc. A state space model consists of a state equation and an observation equation. The state equation models the process of the states while the observation equation links the observations to these underlying states. A standard linear state space model can be written as

$$\mathbf{X}_t = \mathbf{F}_t \mathbf{X}_{t-1} + \mathbf{V}_t, \quad (1)$$

$$\mathbf{Y}_t = \mathbf{G}_t \mathbf{X}_t + \mathbf{W}_t, \quad (2)$$

where  $\mathbf{X}_t$  is the state vector,  $\mathbf{Y}_t$  is the observation vector,  $\mathbf{F}_t$  is the state transition matrix, and  $\mathbf{G}_t$  is the observation matrix. It is assumed that vectors  $\mathbf{V}_t$  and  $\mathbf{W}_t$  are independent and identically distributed (i.i.d.), with  $\mathbf{V}_t \sim N(0, \mathbf{Q})$  and  $\mathbf{W}_t \sim N(0, \mathbf{R})$ . The system matrices ( $\mathbf{F}_t, \mathbf{G}_t$ ), and the covariance matrices ( $\mathbf{Q}, \mathbf{R}$ ) may contain unknown parameters  $\boldsymbol{\theta}$  and  $\boldsymbol{\omega}$  respectively.

In this paper, motivated by HIV dynamic studies, we propose a class of mixed-effects state space models (MESSM) based on the longitudinal feature of the dynamic systems. In the models, the dimension and structure of the system matrices  $\mathbf{F}_t$  and  $\mathbf{G}_t$  are specified a priori and may be based on a system of differential/difference equations, which is often used in engineering applications. Two estimation methods for standard mixed-effects models and state space models, the Bayesian approach and the maximum likelihood method, are modified and investigated for estimating unknown parameters in the proposed MESSM. Simulation studies are carried out to assess the finite-sample performance of the two estimation methods, and the results indicate that both methods provide quite accurate estimates for the parameters. The mixed-effects state space models are applied to a data set from an AIDS clinical trial to illustrate the proposed methodologies.

The primary interest of state space modelling is to estimate the state variables based upon the observations. If the parameters are known, the well-known Kalman filter (Kalman, 1960) yields, under the condition of linearity and normality, the optimal estimates of the states. For unknown parameters the maximum likelihood method provides efficient estimates under certain conditions (Schweppe, 1965; Pagan, 1979; Jensen and Petersen, 1999).

The state space model provides rich covariance structures for the observations  $\mathbf{Y}_t$ . It indeed covers the covariance structure of the ARMA (Autoregressive Moving Average) model as the latter can be written into the state space form. Another interesting aspect of the state space model is its flexibility in modeling the underline mechanism (state equation) and the observations (observation equation) separately.

A longitudinal study usually involves a large number of individuals with observations for each individual taken over a period of time. Variation among the observations consists of the between-subject variation and the within-subject variation. Random-effects models have been widely used for longitudinal data analysis. For example, the linear random-effects models (Laird and Ware, 1982) can be written as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i,$$

$$\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D}), \quad \mathbf{e}_i \sim N(\mathbf{0}, \boldsymbol{\Omega}_i)$$

In this model, the random effects  $\mathbf{b}_i$  and regression errors  $\mathbf{e}_i = [e_{i1} \cdots e_{in_i}]'$  account for the between-subject and within-subject variation, respectively. If  $\mathbf{e}_i$  are independent, only the random effects contribute to the serial correlation among observations from a subject. More complicated covariance structures for  $\mathbf{e}_i$  have been considered by several authors (Jones, 1993; Diggle, Liang, and Zeger, 1994). In Jones (1993)  $\mathbf{e}_i$  was assumed to follow the ARMA model, and the state space approach and the Kalman filter were used as convenient computational tools to compute the likelihood.

Stochastic models have been used to model longitudinal data. For example, under the context of studying growth curves, Rahiala (1999) proposed a random-effects autoregressive (AR) model under the context of growth curve modeling. In Rahiala's model the autoregressive coefficients were treated as random effects. The state space approach was briefly discussed for the computation of the likelihood when missing data presented.

Instead of using the state space approach as a computational tool, we consider the state space modeling approach to directly model longitudinal dynamic systems in this paper. We introduce the random effects to the system matrices  $\mathbf{F}_t$  and  $\mathbf{G}_t$  in (1)-(2) to account for between-individual variation in longitudinal studies. It is a stochastic modeling approach, but is more flexible than the random-effects AR model and can easily handle irregularly spaced data and missing data. Such modeling approach inherits the modeling power of state

space models for dynamic systems as well as the idea of random effects for longitudinal studies. Thus it is especially useful in dealing with mechanism-based longitudinal dynamic systems. However, it can be applied to many other applications which we will discuss at the end of the paper.

This paper is organized as follows. In Section 2, we introduce the MESSM. The estimation methods for both the individual and the population state variables are presented in Section 3. The Bayesian approach and the maximum likelihood (ML) method are proposed to estimate unknown parameters in Section 4. In Section 5 we show results from simulation studies, and apply the MESSM to an AIDS clinical trial data. In Section 6 we summarize the results, and discuss possible extensions to the MESSM and related issues.

## 2 Model Specification

A linear MESSM can be written as

$$\mathbf{X}_{i,t+1} = \mathbf{F}(\boldsymbol{\theta}_i)\mathbf{X}_{it} + \mathbf{V}_{it}, \quad \mathbf{V}_{it} \sim \text{N}(0, \mathbf{Q}), \quad (3)$$

$$\mathbf{Y}_{it} = \mathbf{G}(\boldsymbol{\theta}_i)\mathbf{X}_{it} + \mathbf{W}_{it}, \quad \mathbf{W}_{it} \sim \text{N}(0, \mathbf{R}), \quad (4)$$

where  $\mathbf{Y}_{it}$  is the  $q \times 1$  observation vector for the  $i$ th ( $i = 1, \dots, m$ ) subject at time  $t$  ( $t = 1, \dots, n_i$ ),  $\mathbf{X}_{it}$  is the  $p \times 1$  state vector,  $\mathbf{V}_{it}$  is the  $p \times 1$  dynamic disturbance, and  $\mathbf{W}_{it}$  is the  $q \times 1$  observation error.  $\{\mathbf{V}_{it}\}$  and  $\{\mathbf{W}_{it}\}$  are mutually independent within themselves, and are independent to each other. We assume the initial values  $\mathbf{X}_{i0}$ 's are independent and  $\mathbf{X}_{i0} \sim \text{N}(\boldsymbol{\tau}, \mathbf{A})$ . This assumption is critical in deriving the population state in Section 3.2., but not important for the formulation of the parameter estimation methods in Section 4. The  $p \times p$  transition matrix  $\mathbf{F}(\boldsymbol{\theta}_i)$  and the  $q \times p$  observation matrix  $\mathbf{G}(\boldsymbol{\theta}_i)$  are parameterized with the  $r \times 1$  parameter vector  $\boldsymbol{\theta}_i$ , and the  $p \times p$  matrix  $\mathbf{Q}$  and the  $q \times q$  matrix  $\mathbf{R}$  with the  $d \times 1$  vector of parameters  $\boldsymbol{\omega}$ . Recognized as a vector of individual parameters,  $\boldsymbol{\theta}_i$  can be modelled as

$$\boldsymbol{\theta}_i = \boldsymbol{\theta} + \mathbf{b}_i, \quad \mathbf{b}_i \sim \text{N}(0, \mathbf{D}), \quad (5)$$

where the  $r \times 1$  vector  $\boldsymbol{\theta}$  is the fixed effect, or the population parameter, and the  $r \times 1$  vector  $\mathbf{b}_i$  is the random effect.  $\mathbf{b}_i$ 's are assumed to be i.i.d. Model (5) for the mixed effects can

be generalized without much difficulty to more complicated linear or nonlinear models with covariates (Davidian and Giltinan, 1995; Vonesh and Chinchilli, 1996).

In this paper we further assume  $\mathbf{F}(\boldsymbol{\theta}_i) = \tilde{\mathbf{F}} + \mathbf{F}^{(m)}(\boldsymbol{\theta}_i)$  where  $\tilde{\mathbf{F}}$  is a matrix containing known entries in  $\mathbf{F}(\boldsymbol{\theta}_i)$ , and the elements in the mixed-effects matrix  $\mathbf{F}^{(m)}(\boldsymbol{\theta}_i)$  are either linear functions of  $\boldsymbol{\theta}_i$  or 0. Under this assumption and model (5), there exists a  $p^2 \times r$  matrix  $\mathbf{U}_1$  such that

$$\text{vec}(\mathbf{F}^{(m)}(\boldsymbol{\theta}_i)) = \mathbf{U}_1 \boldsymbol{\theta}_i = \mathbf{U}_1 \boldsymbol{\theta} + \mathbf{U}_1 \mathbf{b}_i.$$

Here  $\text{vec}(\cdot)$  denotes the stacked vector of a matrix. Let  $\text{mat}_p(\cdot)$  be the inverse operator of  $\text{vec}(\cdot)$ , i.e. for a  $pk \times 1$  stacked vector  $\boldsymbol{\xi} = (\boldsymbol{\xi}'_1 \dots \boldsymbol{\xi}'_k)'$  with  $p \times 1$  sub-vectors  $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_k$ ,  $\text{mat}_p(\boldsymbol{\xi})$  is a  $p \times k$  matrix, and  $\text{mat}_p(\boldsymbol{\xi}) = (\boldsymbol{\xi}_1 \dots \boldsymbol{\xi}_k)$ . Now let  $\mathbf{F}^{(f)}(\boldsymbol{\theta}) = \text{mat}_p(\mathbf{U}_1 \boldsymbol{\theta})$  and  $\mathbf{F}_i^{(r)}(\mathbf{b}_i) = \text{mat}_p(\mathbf{U}_1 \mathbf{b}_i)$ , then  $\mathbf{F}(\boldsymbol{\theta}_i) = \tilde{\mathbf{F}} + \mathbf{F}^{(f)} + \mathbf{F}_i^{(r)}$  with the simplified notations  $\mathbf{F}^{(f)}$  and  $\mathbf{F}_i^{(r)}$  for  $\mathbf{F}^{(f)}(\boldsymbol{\theta})$  and  $\mathbf{F}_i^{(r)}(\mathbf{b}_i)$  respectively. The matrices  $\tilde{\mathbf{F}}$ ,  $\mathbf{F}^{(f)}$  and  $\mathbf{F}_i^{(r)}$  account for the restriction (known effects) on  $\mathbf{F}(\boldsymbol{\theta}_i)$ , the fixed effects, and the random effects respectively. Similarly, we assume there exists a  $q^2 \times r$  matrix  $\mathbf{U}_2$  such that  $\mathbf{G}(\boldsymbol{\theta}_i) = \tilde{\mathbf{G}} + \mathbf{G}^{(m)}(\boldsymbol{\theta}_i)$  and  $\text{vec}(\mathbf{G}^{(m)}(\boldsymbol{\theta}_i)) = \mathbf{U}_2 \boldsymbol{\theta}_i$ . The matrix  $\mathbf{G}(\boldsymbol{\theta}_i)$  can be decomposed as  $\mathbf{G}(\boldsymbol{\theta}_i) = \tilde{\mathbf{G}} + \mathbf{G}^{(f)} + \mathbf{G}_i^{(r)}$  as before. The assumptions on the structures of the matrices  $\mathbf{F}(\boldsymbol{\theta}_i)$  and  $\mathbf{G}(\boldsymbol{\theta}_i)$  facilitate the estimation procedures discussed in Section 4. From the practical point of view, a wide range of application problems are covered under these assumptions. Moreover, the matrices  $\mathbf{F}(\boldsymbol{\theta}_i)$  and  $\mathbf{G}(\boldsymbol{\theta}_i)$  can be generalized to contain elements with more complicated functional forms of  $\boldsymbol{\theta}_i$ . We discuss this point in more detail in Section 6.

We give an illustrating example based on an HIV dynamics model proposed by Perelson *et al.* (1996). The model described the interaction between the HIV viruses and the immune system of patients who were undergoing an antiretroviral drug treatment. The drug was assumed to completely block the reproduction of infectious HIV virions, thus the newly produced virions were all noninfectious. The model was written as a system of differential equations

$$\frac{dT^*}{dt} = kV_I T - \delta T^*, \quad (6)$$

$$\frac{dV_I}{dt} = -cV_I, \quad (7)$$

$$\frac{dV_{NI}}{dt} = N\delta T^* - cV_{NI}, \quad (8)$$

where  $V_I$ ,  $V_{NI}$ ,  $T$  and  $T^*$  represented the plasma concentrations of infectious viruses, noninfectious viruses, uninfected T-cells, and productively infected T-cells, respectively. Wu, Ding, and DeGruttola (1998) and Wu and Ding (1999) introduced mixed-effects model approach for HIV dynamics that was shown to improve the estimation of both the individual and the population parameters. Because in practice HIV dynamic data were collected at discrete times (days), difference equations may be considered in place of the differential equations (6) - (8). If the total plasma HIV concentration ( $V, V = V_{NI} + V_I$ ) and  $T^*$  were observed, the mixed-effects state space model could be formulated as

$$\mathbf{X}_{it} = \mathbf{F}_i \mathbf{X}_{i,t-1} + \mathbf{V}_{it}, \quad (9)$$

$$\mathbf{Y}_{it} = \mathbf{G} \mathbf{X}_{it} + \mathbf{W}_{it}, \quad (10)$$

where  $\mathbf{Y}_{it}$  is the observed ( $T^* V$ ) for the  $i$ th patient at time  $t$ ,

$$\mathbf{X}_{it} = (T_{it}^* \ V_{Iit} \ V_{NIit})', \quad \mathbf{F}_i = \begin{bmatrix} 1 - \delta_i & k_i T_i & 0 \\ 0 & 1 - c_i & 0 \\ N_i \delta_i & 0 & 1 - c_i \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

The mixed effects  $\boldsymbol{\theta}_i = (1 - \delta_i \ k_i T_i \ N_i \delta_i \ 1 - c_i)'$  are assumed to follow

$$\boldsymbol{\theta}_i = \boldsymbol{\theta} + \mathbf{b}_i, \quad \mathbf{b}_i = (b_{i1} \ b_{i2} \ b_{i3} \ b_{i4})' \sim (\mathbf{0}, \mathbf{D}),$$

where the population parameters  $\boldsymbol{\theta} = (1 - \delta \ kT \ N\delta \ 1 - c)' = (\theta_1 \ \theta_2 \ \theta_3 \ \theta_4)'$  are fixed effects.

The MESSM is the hybrid of the mixed-effects model and the state space model. It inherits the strengths of both models. The mixed effects link all subjects via the population structure, which establishes the concept of the population dynamics. The estimation of unknown parameters for individual subjects, especially those with insufficient observations, can be improved by borrowing information from other subjects. State space modelling is especially valuable when the system dynamics can be described by mathematical models. Irregularly spaced data, common in longitudinal studies, could be handled easily by state space models. Also, state space models can flexibly incorporate various correlation structures for serially correlated longitudinal data.

### 3 Estimation of State Variables

For MESSM we consider the estimation of both individual state variables and population state variables. The later is defined in Section 3.2.

#### 3.1 The Kalman filter and the estimation of individual states

We now consider the estimation of an individual state  $\mathbf{X}_{it}$ . The estimation of  $\mathbf{X}_{it}$  can be classified into three types based on the amount of information that is available. The estimation of  $\mathbf{X}_{it}$  is called a one-step-ahead prediction if  $\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{i,t-1}$  are observed; it is called a filtered estimation if  $\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{it}$  are observed; and a smoothed estimation if  $\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{in_i}(n_i > t)$  are observed. If the parameters  $\boldsymbol{\theta}_i$  and  $\boldsymbol{\omega}$  are known in the linear MESSM (3) - (4), then the Kalman filter (Kalman, 1960) provides the basic solutions to the three types of estimation problems. The Kalman filter is a recursive algorithm which produces the linear estimators for the state variables. It is well-known that, under the assumption of normality, these estimators obtained from the Kalman filter are unbiased, and minimize the mean squared error.

Consider the MESSM (3) - (4). If  $\boldsymbol{\theta}_i$  are known, the standard Kalman filter can be applied to estimate  $\mathbf{X}_{it}$ . We review the iterative algorithm of the Kalman filter in the following discussion. Detailed description and properties of the Kalman filter can be found in many books, for examples, Harvey (1989), Brockwell and Davis (1991), and West and Harrison (1997).

Let  $\mathbf{X}_{it|k}$  denote the estimation of the individual state  $\mathbf{X}_{it}$  given the observations  $\{\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{ik}\}$ . Let  $\mathbf{P}_{it|k}$  be the covariance matrix of the estimation error  $\mathbf{X}_{it|k} - \mathbf{X}_{it}$ . Use the simplified notations  $\mathbf{F}_i$  and  $\mathbf{G}_i$  to represent  $\mathbf{F}(\boldsymbol{\theta}_i)$  and  $\mathbf{G}_i(\boldsymbol{\theta}_i)$  respectively. The Kalman recursion proceeds as follows,

$$\begin{aligned}\mathbf{X}_{it|t-1} &= \mathbf{F}_i \mathbf{X}_{i,t-1|t-1}, \\ \mathbf{P}_{it|t-1} &= \mathbf{F}_i \mathbf{P}_{i,t-1|t-1} \mathbf{F}_i' + \mathbf{Q}, \\ \mathbf{I}_{it} &= \mathbf{Y}_{it} - \mathbf{G}_i \mathbf{X}_{it|t-1}, \\ \mathbf{O}_{it} &= \mathbf{G}_i \mathbf{P}_{it|t-1} \mathbf{G}_i' + \mathbf{R},\end{aligned}$$

$$\begin{aligned}\mathbf{X}_{it|t} &= \mathbf{X}_{it|t-1} + \mathbf{K}_{it}\mathbf{I}_{it}, \\ \mathbf{P}_{it|t} &= \mathbf{P}_{it|t-1} - \mathbf{K}_{it}\mathbf{G}_{it}'\mathbf{P}_{it|t-1},\end{aligned}$$

where

$$\mathbf{K}_{it} = \mathbf{P}_{it|t-1}\mathbf{G}_{it}'\mathbf{O}_{it}^{-1}.$$

Here  $\mathbf{I}_{it}$  is the innovation and  $\mathbf{O}_{it}$  is the covariance matrix of  $\mathbf{I}_{it}$ .  $\mathbf{X}_{it|t-1}$  is the one-step-ahead prediction of  $\mathbf{X}_{it}$ , and  $\mathbf{X}_{it|t}$  is the filtered estimation of  $\mathbf{X}_{it}$ .

Let  $\mathbf{P}_{it,n} = E[(\mathbf{X}_{it} - \mathbf{X}_{it|t-1})(\mathbf{X}_{in} - \mathbf{X}'_{in|n-1})]$ ,  $\mathbf{\Theta}_{in} = \mathbf{F}_i\mathbf{P}_{in|n-1}\mathbf{G}'_i$ , and  $\mathbf{\Delta}_{in} = \mathbf{G}_i\mathbf{P}_{in|n-1}\mathbf{G}'_i + \mathbf{R}$ . The fixed-point smoothing algorithm proceeds as follows:

$$\mathbf{X}_{it|n} = \mathbf{X}_{it|n-1} + \mathbf{P}_{it,n}\mathbf{G}'_i\mathbf{\Delta}_{in}^{-1}\mathbf{I}_{in}, \quad (11)$$

$$\mathbf{P}_{it,n+1} = \mathbf{P}_{it,n}[\mathbf{F}_i - \mathbf{\Theta}_{in}\mathbf{\Delta}_{in}^{-1}\mathbf{G}'_i]',$$

$$\mathbf{P}_{it|n} = \mathbf{P}_{it|n-1} - \mathbf{P}_{it,n}\mathbf{G}'_i\mathbf{\Delta}_{in}^{-1}\mathbf{G}'_i\mathbf{P}'_{it,n}, \quad (12)$$

with the initial condition,  $\mathbf{P}_{it,t} = \mathbf{P}_{it|t-1}$ .

Under the normality assumption, the estimator  $\mathbf{X}_{it|k}$  produced by the Kalman recursion is the mean of  $\mathbf{X}_{it}$  conditional on the data  $(\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{ik})$ . If the normality assumption does not hold, the Kalman filter may not produce the conditional means of the states. The Bayesian approach provides an alternative method for the estimation of the state vectors. We discuss this approach in Section 4.1.

### 3.2 Population state variable and the EM algorithm enhanced smoothing

From (3), we can write  $\mathbf{X}_{it}$  in terms of  $\mathbf{X}_{i0}$  and  $\{\mathbf{V}_{is}, s \leq t\}$ :

$$\mathbf{X}_{it} = \mathbf{F}_i^t \mathbf{X}_{i0} + \sum_{k=0}^{t-1} \mathbf{F}_i^k \mathbf{V}_{i,t-k}.$$

Therefore, under the assumptions that the initial values  $\mathbf{X}_{i0}$ 's are i.i.d. and  $\boldsymbol{\theta}_i$ 's are i.i.d.,  $\mathbf{X}_{it}$ 's are i.i.d. at a fixed time point  $t$ . Thus the individual state variables  $\mathbf{X}_{it}$  ( $i = 1, \dots, m$ ) follow a common mean process  $\mathbf{X}_t$  plus a mean zero process, and can be modelled as

$$\mathbf{X}_{it} = \mathbf{X}_t + \mathbf{Z}_{it}, \quad \mathbf{Z}_{it} \sim (0, \mathbf{B}_t). \quad (13)$$

We call  $\mathbf{X}_t$  the population state variable at time  $t$ . Vector  $\mathbf{Z}_{it}$  represents the deviation of the individual state  $\mathbf{X}_{it}$  from the population variable  $\mathbf{X}_t$  with  $\mathbf{B}_t$  being the covariance matrix. The population state variable  $\mathbf{X}_t$  represents the “average” response of all the individuals at time  $t$ , and the series of population state variables reflect the dynamics of a population.

If  $\mathbf{X}_{it}$  were observable, the maximum pseudo-likelihood estimation of  $\mathbf{X}_t$  would then be the sample mean  $\sum_{i=1}^m \mathbf{X}_{it}/m$ . However, because in the real world  $\mathbf{X}_{it}$  are usually not directly observable, we instead use an estimate of  $\mathbf{X}_{it}$ , denoted by  $\tilde{\mathbf{X}}_{it}$ , to replace  $\mathbf{X}_{it}$  in (13). Suppose the observations are available up to time  $n_i$  ( $n_i > t$ ). Because the Kalman smoothing estimate  $\mathbf{X}_{it|n_i}$  is an unbiased estimator of  $\mathbf{X}_{it}$ , we could use the  $\mathbf{X}_{it|n_i}$  as the estimate  $\tilde{\mathbf{X}}_{it}$ . Let  $\boldsymbol{\varsigma}_{it} = \mathbf{X}_{it|n_i} - \mathbf{X}_{it}$ . one can establish a variance-components model for  $\tilde{\mathbf{X}}_{it}$

$$\tilde{\mathbf{X}}_{it} = \mathbf{X}_t + \mathbf{Z}_{it} + \boldsymbol{\varsigma}_{it}, \quad (14)$$

where  $\boldsymbol{\varsigma}_{it}$  is the estimation error of  $\tilde{\mathbf{X}}_{it}$ , and  $\boldsymbol{\varsigma}_{it} \sim \text{N}(0, \boldsymbol{\Sigma}_{it})$ . Note that  $\boldsymbol{\Sigma}_{it}$  is the covariance matrix  $\mathbf{P}_{it|n_i}$  in the smoothing algorithm (12).

Treating  $\tilde{\mathbf{X}}_{it}$  as “data”, we maximize the pseudo-likelihood function to estimate the population state  $\mathbf{X}_t$  and the covariance matrix  $\mathbf{B}_t$ , that is, to maximize

$$l(\boldsymbol{\Sigma}_t, \mathbf{B}_t | \tilde{\mathbf{X}}_{it}) = - \sum_{i=1}^m \left( \log |\boldsymbol{\Sigma}_i + \mathbf{B}_t| + [\tilde{\mathbf{X}}_{it} - \mathbf{X}_t]' [\boldsymbol{\Sigma}_i + \mathbf{B}_t]^{-1} [\tilde{\mathbf{X}}_{it} - \mathbf{X}_t] \right).$$

The EM algorithm developed by Steimer, Mallet, Golmard, and Boisvieux (1984) can be adapted to find the estimates of the population states, which maximize the above pseudo-likelihood function, as follows.

1. Initialization:

$$\begin{aligned} \hat{\mathbf{X}}_t^{(0)} &= m^{-1} \sum_{i=1}^m \tilde{\mathbf{X}}_{it}, \\ \hat{\mathbf{B}}_t^{(0)} &= m^{-1} \sum_{i=1}^m [\tilde{\mathbf{X}}_{it} - \hat{\mathbf{X}}_t^{(0)}][\tilde{\mathbf{X}}_{it} - \hat{\mathbf{X}}_t^{(0)}]' \end{aligned}$$

2. E-step: Update the individual estimates of state variables:

$$\hat{\mathbf{X}}_{it}^{(k+1)} = [(\hat{\mathbf{B}}_t^{(k)})^{-1} + \boldsymbol{\Sigma}_{it}^{-1}]^{-1} [(\hat{\mathbf{B}}_t^{(k)})^{-1} \hat{\mathbf{X}}_t^{(k)} + \boldsymbol{\Sigma}_{it}^{-1} \tilde{\mathbf{X}}_{it}].$$

where  $\hat{\mathbf{X}}_{it}^{(k+1)}$  is the conditional posterior mean and  $[(\hat{\mathbf{B}}_t^{(k)})^{-1} + \boldsymbol{\Sigma}_{it}^{-1}]^{-1}$  is the conditional posterior variance of  $\mathbf{X}_{it} | \{\tilde{\mathbf{X}}_{it}, \hat{\mathbf{X}}_t^{(k)}, (\hat{\mathbf{B}}_t^{(k)})^{-1}\}$ .

3. M-step: Update the population estimates of state variable and parameters:

$$\begin{aligned}\hat{\mathbf{X}}_t^{(k+1)} &= m^{-1} \sum_{i=1}^m \hat{\mathbf{X}}_{it}^{(k+1)}, \\ \hat{\mathbf{B}}_t^{(k+1)} &= m^{-1} \left\{ \sum_{i=1}^m [\hat{\mathbf{X}}_{it}^{(k+1)} - \hat{\mathbf{X}}_t^{(k+1)}][\hat{\mathbf{X}}_{it}^{(k+1)} - \hat{\mathbf{X}}_t^{(k+1)}]' + \sum_{i=1}^m [(\hat{\mathbf{B}}_t^{(k)})^{-1} + \boldsymbol{\Sigma}_{it}^{-1}]^{-1} \right\}.\end{aligned}$$

4. Repeat step 2 and 3 until convergence is reached (the difference between the successive estimates of  $\mathbf{X}_t$  and  $\mathbf{B}_t$  is sufficiently small). Estimates of both individual states and population states are collected at the final step.

We have proposed the estimation of the population state variable based on the smoothing estimates of the individual state variables. In a similar fashion, the filtered or predicted individual state variables can be used to estimate the population state variable if the interested time point  $t$  is beyond the last time point of observations.

## 4 Parameter Estimation

In this section we discuss parameter estimation in MESSM from both the Bayesian and the likelihood perspectives. The Bayesian approach and the ML method in mixed-effects models and their counterparts in the state space models have been well established (Davidian and Giltinan, 1995; Vonesh and Chinchilli, 1996; Harvey, 1989; West and Harrison, 1997). The proposed estimation methods for MESSM are combinations of these estimation methods.

### 4.1 Bayesian Approach

The Bayesian approach for state space models has been studied by many authors in recent years. Linear state space models have been presented in West and Harrison (1997). Non-linear non-Gaussian state space models have been studied in, for example, Kitagawa (1987, 1996), Carlin, Polson, and Stoffer (1992), De Jong and Shephard (1994), and Chen and Liu (2000). Most of these studies focus on the estimation of state variables, and the estimation

of unknown parameters has not been paid enough attention. For MESSM we adopt the method proposed by Carlin, Polson and Stoffer (1992), in which the Gibbs sampler is used to draw samples from the full conditional distributions of state variables  $\mathbf{X}_t$  and parameters conditional on the observations  $\mathbf{Y}_t$ . This approach is straightforward and easy to implement, and it has produced good results in our simulation studies. Our method differs from Carlin *et al.* (1992) in that (1) independent individual subjects are involved in the MESSM; (2) we include one more hierarchical structure on the individual parameters; and (3) we emphasize both parameter estimation and state variable estimation. For the Bayesian approach, prior distributions are assigned to the parameters, and the inference is based upon the posterior distributions. Note that, if only one subject  $i$  is considered and the prior distribution  $N(\boldsymbol{\theta}, \mathbf{D})$  is assigned to  $\theta_i$ , then conceptually  $\boldsymbol{\theta}$  and  $\mathbf{D}$  are simply hyperprior parameters for a standard state space model. However, these hyperprior parameters represent the population parameters when all the subjects are considered, and thus have their own physical meanings for the population.

When the prior distributions are carefully chosen according to their conjugacy, the full conditional distributions have closed forms, and are easy to sample from. The Gibbs sampler can then be used to explore the posterior distributions of the unknown parameters.

Let  $W(\boldsymbol{\Upsilon}, \nu)$  denote a Wishart distribution with the scale matrix  $\boldsymbol{\Upsilon}$  and the degrees of freedom parameter  $\nu$ . We use the following prior distributions:

$$\begin{aligned} \boldsymbol{\theta} &\sim N(\boldsymbol{\eta}, \boldsymbol{\Delta}), & \mathbf{D}^{-1} &\sim W((\mathbf{D}^*)^{-1}, \rho), \\ \mathbf{Q}^{-1} &\sim W(\mathbf{Q}^{*-1}, \rho_1), & \mathbf{R}^{-1} &\sim W(\mathbf{R}^{*-1}, \rho_2), & \mathbf{X}_{i0} &\sim N(\boldsymbol{\tau}, \mathbf{A}). \end{aligned} \quad (15)$$

The full conditional distributions can be derived as follows.

1. The conditional distributions of the state vectors are

$$p(\mathbf{X}_{it}, | \mathbf{Y}, \mathbf{X}_{i,k \neq t}, \boldsymbol{\theta}, \mathbf{D}, \boldsymbol{\Theta}, \mathbf{Q}, \mathbf{R}) \sim N(\mathbf{H}_{it}^{-1} \boldsymbol{\lambda}_{it}, \mathbf{H}_{it}^{-1}), \quad (16)$$

where for  $t \neq 0$  or  $n_i$

$$\mathbf{H}_{it} = \mathbf{Q}^{-1} + \mathbf{F}_i' \mathbf{Q}^{-1} \mathbf{F}_i + \mathbf{G}_i' \mathbf{R}^{-1} \mathbf{G}_i, \quad (17)$$

$$\boldsymbol{\lambda}_{it} = \mathbf{Q}^{-1} \mathbf{F}_i \mathbf{X}_{i,t-1} + \mathbf{F}_i' \mathbf{Q}^{-1} \mathbf{X}_{i,t+1} + \mathbf{G}_i' \mathbf{R}^{-1} \mathbf{Y}_{it}; \quad (18)$$

for  $t = 0$

$$\begin{aligned} \mathbf{H}_{i0} &= \mathbf{F}_i' \mathbf{Q}^{-1} \mathbf{F}_i + \mathbf{A}^{-1}, \\ \boldsymbol{\lambda}_{i0} &= \mathbf{F}_i' \mathbf{Q}^{-1} \mathbf{X}_{i1} + \mathbf{A}^{-1} \boldsymbol{\tau}; \end{aligned}$$

and for  $t = n_i$

$$\begin{aligned} \mathbf{H}_{in_i} &= \mathbf{Q}^{-1} + \mathbf{G}_i' \mathbf{R}^{-1} \mathbf{G}_i, \\ \boldsymbol{\lambda}_{in_i} &= \mathbf{Q}^{-1} \mathbf{F}_i \mathbf{X}_{i,n_i-1} + \mathbf{G}_i' \mathbf{R}^{-1} \mathbf{Y}_{in_i}. \end{aligned}$$

2. The conditional distributions of  $\boldsymbol{\theta}_i$ ,  $\boldsymbol{\theta}$ , and  $\mathbf{D}$  are

$$p(\boldsymbol{\theta}_i | \mathbf{X}, \mathbf{Y}, \boldsymbol{\theta}, \mathbf{D}, \boldsymbol{\Theta}_{-i}, \mathbf{Q}, \mathbf{R}) \sim \text{N}(\mathbf{J}_i^{-1} \boldsymbol{\zeta}_i, \mathbf{J}_i^{-1}), \quad (19)$$

$$p(\boldsymbol{\theta} | \mathbf{X}, \mathbf{Y}, \mathbf{D}, \boldsymbol{\Theta}, \mathbf{Q}, \mathbf{R}) \sim \text{N} \left( [m\mathbf{D}^{-1} + \boldsymbol{\Delta}^{-1}]^{-1} \left[ \sum_i^m \mathbf{D}^{-1} \boldsymbol{\theta}_i + \boldsymbol{\Delta}^{-1} \boldsymbol{\eta} \right], [m\mathbf{D}^{-1} + \boldsymbol{\Delta}^{-1}]^{-1} \right),$$

$$p(\mathbf{D}^{-1} | \mathbf{X}, \mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{Q}, \mathbf{R}) \sim \text{W} \left( \left[ \sum_i^m (\boldsymbol{\theta}_i - \boldsymbol{\theta})(\boldsymbol{\theta}_i - \boldsymbol{\theta})' + \mathbf{D}^* \right]^{-1}, m + \rho \right), \quad (20)$$

where

$$\begin{aligned} \mathbf{J}_i &= \sum_{t=1}^{n_i} (\mathbf{T}'_{i,t-1} \mathbf{Q}^{-1} \mathbf{T}_{i,t-1} + \mathbf{L}'_{it} \mathbf{R}^{-1} \mathbf{L}_{it}) + \mathbf{D}^{-1}, \\ \boldsymbol{\zeta}_i &= \sum_{t=1}^{n_i} (\mathbf{T}'_{i,t-1} \mathbf{Q}^{-1} (\mathbf{X}_{it} - \tilde{\mathbf{F}} \mathbf{X}_{i,t-1}) + \mathbf{L}'_{it} \mathbf{R}^{-1} (\mathbf{Y}_{it} - \tilde{\mathbf{G}} \mathbf{X}_{it}) + \mathbf{D}^{-1} \boldsymbol{\theta}), \\ \mathbf{T}_{i,t-1} &= (\mathbf{X}'_{i,t-1} \otimes \mathbf{I}_p) \mathbf{U}_1, \\ \mathbf{L}_{it} &= (\mathbf{X}'_{it} \otimes \mathbf{I}_q) \mathbf{U}_2. \end{aligned}$$

3. The conditional distributions of the covariance matrices  $\mathbf{Q}$  and  $\mathbf{R}$  are

$$p(\mathbf{Q}^{-1} | \mathbf{X}, \mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{Q}, \mathbf{D}) \sim \text{W} \left( [\mathbf{Q}^* + \frac{1}{2} \sum_{i=1}^m \sum_{t=1}^{n_i} (\mathbf{X}_{it} - \mathbf{F}_i \mathbf{X}_{i,t-1})' (\mathbf{X}_{it} - \mathbf{F}_i \mathbf{X}_{i,t-1})]^{-1}, s + \rho_1 \right), \quad (21)$$

$$p(\mathbf{R}^{-1} | \mathbf{X}, \mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{R}, \mathbf{D}) \sim \text{W} \left( [\mathbf{R}^* + \frac{1}{2} \sum_{i=1}^m \sum_{t=1}^{n_i} (\mathbf{Y}_{it} - \mathbf{G}_i \mathbf{X}_{it})' (\mathbf{Y}_{it} - \mathbf{G}_i \mathbf{X}_{it})]^{-1}, s + \rho_2 \right), \quad (22)$$

where  $s = \sum_{i=1}^m n_i$ .

Note that irregularly spaced data can be treated as data with missing observations, and the full conditional distributions only need slight adjustments at the time points of missing

data. For example, suppose  $\mathbf{Y}_{it}$  is missing, the conditional distribution  $p(\mathbf{X}_{it}|\mathbf{Y}, \mathbf{X}_{i,k \neq t}, \dots)$  in (16) would depend only on  $\mathbf{X}_{i,t-1}$ ,  $\mathbf{X}_{i,t+1}$  and the parameters, but not on  $\mathbf{Y}_{it}$ . Thus, when deriving the conditional distribution, one needs to drop the terms  $\mathbf{G}_i' \mathbf{R}^{-1} \mathbf{G}_i$  in (17) and  $\mathbf{G}_i' \mathbf{R}^{-1} \mathbf{Y}_{it}$  in (18), which are related to the observation.

The construction of the Gibbs sampler is based on the full conditional distributions. Starting from an arbitrary set of initial values, samples are drawn from the conditional distributions sequentially. Let  $\Phi$  be the collection of all the components involved in the full conditional distributions, i.e.,  $\Phi = (\mathbf{X}_{10}, \dots, \mathbf{X}_{1n_1}, \dots, \mathbf{X}_{m0}, \dots, \mathbf{X}_{mn_m}, \theta_1, \dots, \theta_m, \theta, \mathbf{D}, \mathbf{Q}, \mathbf{R})$ . Let  $\Phi - \{\mathbf{u}\}$  be the set of all the components except  $\{\mathbf{u}\}$ . Here  $\{\mathbf{u}\}$  can be any component or a set of components in  $\Phi$ . Starting from the initial value  $\Phi^{(0)}$  the Gibbs sampler draws a sample  $\mathbf{X}_{10}^{(1)}$  from  $p(\mathbf{X}_{10}|\{\Phi - \{\mathbf{X}_{10}\}\}^{(0)}, \mathbf{Y})$ , then draws  $\mathbf{X}_{11}^{(1)}$  from  $p(\mathbf{X}_{11}|\mathbf{X}_{10}^{(1)}, \{\Phi - \{\mathbf{X}_{10}, \mathbf{X}_{10}\}\}^{(0)}, \mathbf{Y})$ , and so on up to  $\mathbf{R}^{(1)}$  from  $p(\mathbf{R}|\{\Phi - \{\mathbf{R}\}\}^{(1)}, \mathbf{Y})$ . Suppose the  $J$ th iteration produces a sample  $\Phi^{(J)}$ . When  $J$  is large,  $\Phi^{(J)}$  is a sample from the posterior distribution  $p(\Phi|\mathbf{Y})$  (Geman and Geman, 1984). The posterior distribution of the  $k$ th component of  $\Phi$  can be approximated by (Carlin, Polson and Stoffer, 1992)

$$p(\Phi_k|\mathbf{Y}) \approx \frac{1}{J} \sum_{j=1}^J p(\Phi_k|\{\Phi - \Phi_k\}^{(j)}, \mathbf{Y}). \quad (23)$$

The posterior mean can be taken as the point estimate of a parameter if the Bayesian loss function is chosen to be the squared-error loss.

Assuming the observations are available up to time  $n_i$ , the Gibbs sampler constructed above can produce samples from the posterior distribution  $p(\mathbf{X}_{it}|\mathbf{Y})$  for  $t \leq n_i$ . For  $t < n_i$  this is indeed a solution to the smoothing problem defined in Section 3.1. For  $t = n_i$  the posterior distribution  $p(\mathbf{X}_{in_i}|\mathbf{Y})$  provides a solution for the filtering problem at time  $n_i$ . To solve the prediction problem, i.e. for  $t > n_i$ , one can treat  $\mathbf{Y}_{i,n_i+1}, \mathbf{Y}_{i,n_i+2}, \dots, \mathbf{Y}_{it}$  as missing and derive the conditional distributions for the state vectors  $\mathbf{X}_{i,n_i+1}, \mathbf{X}_{i,n_i+2}, \dots, \mathbf{X}_{it}$ , then include these state vectors in the Gibbs sampler run. Compared with the Kalman filter, the Gibbs sampler approach is more flexible and can be extended to nonlinear and non-Gaussian cases (Carlin, Polson and Stoffer, 1992); it also provides a unified frame work for the state estimation and parameter estimation. However, the Gibbs sampler approach is computationally more expensive, and does not provide online solutions to the estimation

problems.

## 4.2 Estimation by the EM algorithm

For the standard Gaussian linear state space models, the ML method for the estimation of parameters has been well studied (Harvey, 1989). The idea of the ML method is to calculate the likelihood function via the prediction error decomposition (Schweppe, 1965) in which the Kalman filter is applied to obtain the innovations and their covariances. Due to the complexity of the likelihood function of the standard state space model, numerical methods, specifically the EM algorithm (Watson and Engle, 1983) and the Newton-Raphson algorithm (Engle and Watson, 1981), are used to find the MLE. In this section, we extend the EM algorithm proposed by Watson and Engle (1983) to find the MLE of unknown parameters for the MESSM (3)-(5).

If the state variables  $\mathbf{X}_{it}$  and the random effects  $\mathbf{b}_i$  and  $\mathbf{r}_i$  were observed, the MESSM (3)-(4) would be recognized as a system of linear regressions, and the MLE for the fixed effects  $\boldsymbol{\theta}$ , and the covariance matrices  $\mathbf{D}$ ,  $\mathbf{Q}$  and  $\mathbf{R}$  would be easily derived. These estimators are constructed as follows.

We first rewrite the MESSM (3)-(5) as

$$\mathbf{X}_{it} - (\tilde{\mathbf{F}} + \mathbf{F}_i^{(r)})\mathbf{X}_{i,t-1} = (\mathbf{X}'_{i,t-1} \otimes \mathbf{I}_p)\mathbf{U}_1\boldsymbol{\theta} + \mathbf{V}_{it}, \quad (24)$$

$$\mathbf{Y}_{it} - (\tilde{\mathbf{G}} + \mathbf{G}_i^{(r)})\mathbf{X}_{it} = (\mathbf{X}'_{it} \otimes \mathbf{I}_q)\mathbf{U}_2\boldsymbol{\theta} + \mathbf{W}_{it}, \quad (25)$$

where

$$\mathbf{U}_1\boldsymbol{\theta} = \text{vec}(\mathbf{F}^{(f)}(\boldsymbol{\theta})), \text{ and } \mathbf{U}_2\boldsymbol{\theta} = \text{vec}(\mathbf{G}^{(f)}(\boldsymbol{\theta})).$$

Let

$$\begin{aligned} \mathbf{X}_{it}^* &= \mathbf{X}_{it} - (\tilde{\mathbf{F}} + \mathbf{F}_i^{(r)})\mathbf{X}_{i,t-1}, \\ \mathbf{Y}_{it}^* &= \mathbf{Y}_{it} - (\tilde{\mathbf{G}} + \mathbf{G}_i^{(r)})\mathbf{X}_{it}, \\ \tilde{\mathbf{Y}}_i &= (\mathbf{X}_{i1}^{*'} , \dots , \mathbf{X}_{in_i}^{*'} , \mathbf{Y}_{i1}^{*'} , \dots , \mathbf{Y}_{in_i}^{*'})', \\ \boldsymbol{\Pi}_i &= \begin{pmatrix} \mathbf{I}_p \otimes \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \otimes \mathbf{R} \end{pmatrix} \end{aligned}$$

and

$$\mathbf{M}_i = \begin{pmatrix} \mathbf{M}_i^{(1)} \\ \mathbf{M}_i^{(2)} \end{pmatrix} = \begin{pmatrix} (\mathbf{X}'_{i0} \otimes \mathbf{I}_{n_i})\mathbf{U}_1 \\ \vdots \\ (\mathbf{X}'_{i,n_i-1} \otimes \mathbf{I}_{n_i})\mathbf{U}_1 \\ (\mathbf{X}'_{i1} \otimes \mathbf{I}_p)\mathbf{U}_2 \\ \vdots \\ (\mathbf{X}'_{i,n_i} \otimes \mathbf{I}_p)\mathbf{U}_2 \end{pmatrix}.$$

If the  $\mathbf{X}_{it}$  and  $\mathbf{b}_i$  were observed, the MLE of  $\boldsymbol{\theta}$ ,  $\mathbf{Q}$ ,  $\mathbf{R}$ , and  $\mathbf{D}$  would satisfy

$$\hat{\boldsymbol{\theta}} = \left[ \sum_{i=1}^m \mathbf{M}_i' \hat{\boldsymbol{\Pi}}_i^{-1} \mathbf{M}_i \right]^{-1} \left[ \sum_{i=1}^m \mathbf{M}_i' \hat{\boldsymbol{\Pi}}_i^{-1} \tilde{\mathbf{Y}}_i \right], \quad (26)$$

$$\hat{\mathbf{Q}} = \frac{1}{s} \sum_{i=1}^m \sum_{t=1}^{n_i} \hat{\mathbf{V}}_{it} \hat{\mathbf{V}}_{it}', \quad (27)$$

$$\hat{\mathbf{R}} = \frac{1}{s} \sum_{i=1}^m \sum_{t=1}^{n_i} \hat{\mathbf{W}}_{it} \hat{\mathbf{W}}_{it}', \quad (28)$$

$$\hat{\mathbf{D}} = \frac{1}{m} \sum_{i=1}^m \mathbf{b}_i \mathbf{b}_i', \quad (29)$$

where

$$\begin{aligned} \hat{\mathbf{V}}_{it} &= \mathbf{X}_{it}^* - (\mathbf{X}_{i,t-1}' \otimes \mathbf{I}_p) \mathbf{U}_1 \hat{\boldsymbol{\theta}}, \\ \hat{\mathbf{W}}_{it} &= \mathbf{Y}_{it}^* - (\mathbf{X}_{it}' \otimes \mathbf{I}_q) \mathbf{U}_2 \hat{\boldsymbol{\theta}}. \end{aligned}$$

By iterating (26)-(28) one could find the estimates of  $\boldsymbol{\theta}$ ,  $\mathbf{Q}$  and  $\mathbf{R}$  that satisfy these equations. Equations (26)- (29) define the M-step in the EM algorithm. To establish the E-step we need to find the sufficient statistics for the complete data. Note that we could separate the E-step and the M-step because of the normality assumption on the MESSM (Dempster, Laird and Rubin, 1977).

The log-likelihood function of the complete data  $(\mathbf{X}, \mathbf{Y}, \mathbf{b})$  can be written as

$$\begin{aligned} l(\boldsymbol{\theta}, \mathbf{D}, \mathbf{Q}, \mathbf{R} | \mathbf{X}, \mathbf{Y}, \mathbf{b}) &\propto -\frac{1}{2} \sum_{i=1}^m \left( |\boldsymbol{\Pi}_i| + (\tilde{\mathbf{Y}}_i - \mathbf{M}_i \boldsymbol{\theta})' \boldsymbol{\Pi}_i^{-1} (\tilde{\mathbf{Y}}_i - \mathbf{M}_i \boldsymbol{\theta}) + |\mathbf{D}| + \mathbf{b}_i' \mathbf{D}^{-1} \mathbf{b}_i \right) \\ &= -\frac{1}{2} \sum_{i=1}^m \left( |\boldsymbol{\Pi}_i| + (\tilde{\mathbf{Y}}_i' \boldsymbol{\Pi}_i^{-1} \tilde{\mathbf{Y}}_i - 2\boldsymbol{\theta}' \mathbf{M}_i' \boldsymbol{\Pi}_i^{-1} \tilde{\mathbf{Y}}_i + \boldsymbol{\theta}' \mathbf{M}_i' \boldsymbol{\Pi}_i^{-1} \mathbf{M}_i \boldsymbol{\theta}) \right) \\ &\quad - \frac{1}{2} \sum_{i=1}^m (|\mathbf{D}| + \text{tr}(\mathbf{b}_i \mathbf{b}_i' \mathbf{D}^{-1})). \end{aligned}$$

The term  $\mathbf{M}_i' \hat{\mathbf{\Pi}}_i^{-1} \mathbf{M}_i$  in the log-likelihood function can be written explicitly as

$$\mathbf{U}'_1(\mathbf{X}_i^{-'} \mathbf{X}_i^- \otimes \hat{\mathbf{Q}}^{-1}) \mathbf{U}_1 + \mathbf{U}'_2(\mathbf{X}_i' \mathbf{X}_i \otimes \hat{\mathbf{R}}^{-1}) \mathbf{U}_2 \quad (30)$$

with

$$\begin{aligned} \mathbf{X}_i' &= (\mathbf{X}_{i1}, \mathbf{X}_{i2}, \dots, \mathbf{X}_{in_i}), \\ \mathbf{X}_i^{-'} &= (\mathbf{X}_{i0}, \mathbf{X}_{i1}, \dots, \mathbf{X}_{i,n_i-1}). \end{aligned}$$

Similarly,  $\mathbf{M}_i' \hat{\mathbf{\Pi}}_i^{-1} \tilde{\mathbf{Y}}_i$  can be written as

$$\mathbf{U}'_1(\mathbf{I}_p \otimes \hat{\mathbf{Q}}^{-1}) \text{vec}(\mathbf{X}_i^{*'} \mathbf{X}_i^-) + \mathbf{U}'_2(\mathbf{I}_q \otimes \hat{\mathbf{R}}^{-1}) \text{vec}(\mathbf{Y}_i^{*'} \mathbf{X}_i) \quad (31)$$

with

$$\mathbf{X}_i^{*'} = (\mathbf{X}_{i1}^*, \mathbf{X}_{i2}^*, \dots, \mathbf{X}_{in_i}^*), \text{ and } \mathbf{Y}_i^{*'} = (\mathbf{Y}_{i1}^*, \mathbf{Y}_{i2}^*, \dots, \mathbf{Y}_{in_i}^*).$$

Plugging (30) and (31) into the log-likelihood function, it can be seen that we need to evaluate the conditional expectations for the following matrices:

$$\begin{aligned} &\mathbf{X}_i', \mathbf{X}_i^{-'}, \mathbf{X}_i' \mathbf{X}_i, \mathbf{X}_i^{-'} \mathbf{X}_i^-, \mathbf{X}_i' \mathbf{X}_i^-, \\ &\mathbf{F}_i^{(r)} \mathbf{X}_i^{-'}, \mathbf{F}_i^{(r)} \mathbf{X}_i^{-'} \mathbf{X}_i^-, \mathbf{F}_i^{(r)} \mathbf{X}_i^{-'} \mathbf{X}_i, \mathbf{F}_i^{(r)} \mathbf{X}_i^{-'} \mathbf{X}_i^- \mathbf{F}_i^{(r)'}, \\ &\mathbf{G}_i^{(r)} \mathbf{X}_i', \mathbf{G}_i^{(r)} \mathbf{X}_i' \mathbf{X}_i, \mathbf{G}_i^{(r)} \mathbf{X}_i' \mathbf{X}_i \mathbf{G}_i^{(r)'}, \text{ and } \mathbf{b}_i \mathbf{b}_i'. \end{aligned} \quad (32)$$

Unfortunately, unlike in the EM algorithm for the standard state space models (Watson and Engle, 1983), the conditional expected values of the above matrices for MESSM can not be computed using an efficient algorithm such as the Kalman filter. However, numerical methods can be used to estimate the conditional expectations of the matrices in (32). This approach is known as the stochastic EM algorithm (McLachlan and Krishnan, 1997). Under mild conditions the stochastic EM algorithm converges to the local maxima of the likelihood function (Delyon, Lavielle and Moulines, 1999). Let  $\mathbf{\Lambda} = (\boldsymbol{\theta}, \mathbf{Q}, \mathbf{R}, \mathbf{D})$  be the collection of the parameters. The Gibbs sampler discussed in Section 4.1. can be modified to draw samples from the distributions of  $\mathbf{X}_{it}$  and  $\mathbf{b}_i$  conditional on the observations  $\mathbf{Y}$  and the current estimate  $\hat{\mathbf{\Lambda}}^{(k)}$ . Specifically, we derive the full conditional distributions  $p(\mathbf{X}_{it} | \mathbf{b}_i, \mathbf{Y}, \hat{\mathbf{\Lambda}}^{(k)})$  and  $p(\mathbf{b}_i | \mathbf{X}_{it}, \mathbf{Y}, \hat{\mathbf{\Lambda}}^{(k)})$ , and draw Gibbs samples  $\mathbf{X}_{it}^{(j)}$  and  $\mathbf{b}_i^{(j)}$  ( $j = 1, \dots, J$ ). Then the

conditional expectations can be evaluated based on these samples. Taking the matrix  $\mathbf{b}_i \mathbf{b}_i'$  in (32) as an example,

$$E(\mathbf{b}_i \mathbf{b}_i' | \mathbf{Y}, \hat{\Lambda}^{(k)}) \approx \frac{1}{J} \sum_{j=1}^J \mathbf{b}_i^{(j)} \mathbf{b}_i'^{(j)}.$$

Note that at the final iteration of the EM algorithm the Gibbs samples for  $\mathbf{b}_i$  generated in the E-step can be used to estimate the random effects  $\mathbf{b}_i$ .

The EM algorithm is now fully defined. In the E-step the samples for  $\mathbf{X}_{it}$  and  $\mathbf{b}_i$  are drawn from the Gibbs sampler, and the conditional expectations of the matrices in (32) are computed. In the M-step the estimates of  $\boldsymbol{\theta}$ ,  $\mathbf{Q}$  and  $\mathbf{R}$  are updated by solving Equations (26)-(28) and the estimate of  $\mathbf{D}$  is directly updated from the E-step outputs, i.e.  $\hat{\mathbf{D}}^{(k+1)} = E(\frac{1}{m} \sum_{i=1}^m \mathbf{b}_i \mathbf{b}_i' | \mathbf{Y}, \hat{\Lambda}^{(k)})$ . As suggested in Watson and Engel (1983) it is not necessary to iterate between (26)-(28) during each M-step. Instead, one can use  $\hat{\mathbf{Q}}^{(k)}$  and  $\hat{\mathbf{R}}^{(k)}$  to construct  $\hat{\boldsymbol{\theta}}^{(k+1)}$  in (26), and then use  $\hat{\boldsymbol{\theta}}^{(k+1)}$  to create  $\hat{\mathbf{Q}}^{(k+1)}$  and  $\hat{\mathbf{R}}^{(k+1)}$ . Upon convergence of the EM algorithm, the final parameter estimates will satisfy (26)-(29) by construction.

The EM algorithm proposed here is closely related to the Bayesian approach as both contain a Gibbs sampling scheme, and the full conditional distributions for the Gibbs sampler are very similar. The differences are also clear. The EM algorithm is for the likelihood approach, thus the parameters  $(\boldsymbol{\theta}, \mathbf{Q}, \mathbf{R}, \mathbf{D})$  are not treated as random and no prior distributions are needed. Computationally, while in the Bayesian approach the parameters  $(\boldsymbol{\theta}, \mathbf{Q}, \mathbf{R}, \mathbf{D})$  are constantly updated by sampling from their conditional distributions, in the EM algorithm they remain fixed during the sampling step, and are updated only in the M-step. Therefore, for one iteration of the EM algorithm, the sampling step is computationally less expensive than the Bayesian approach, but an extra computational cost occurs in the M-step. From our experience, running one iteration of the EM algorithm is more time consuming than running the Gibbs sampler in the Bayesian approach, assuming equal number of Gibbs iterations. The convergence of the EM algorithm is also slow, usually more than 50 iterations are needed for convergence. Overall the computational cost for the EM algorithm is much higher.

## 5 Numerical Examples

**Example 1:** Univariate Model: Simulation Studies. For illustration and comparison purposes, we carried out the simulation studies using the two estimation methods discussed in Section 4 for the following univariate MESSM

$$X_{it} = \theta_i X_{i,t-1} + V_{it}, \quad V_{it} \sim \text{i.i.d. } N(0, Q), \quad (33)$$

$$Y_{it} = X_{it} + W_{it}, \quad W_{it} \sim \text{i.i.d. } N(0, R), \quad (34)$$

where

$$\theta_i = \theta + b_i, \quad b_i \sim N(0, D), \quad i = 1, \dots, m. \quad (35)$$

Note that model (33)-(35) can be recognized as a mixed-effects AR(1) model with measurement errors. Formulation of the mixed-effects AR(p) models with measurement errors is also straightforward under the framework of MESSM. These models are a natural extension of the measurement error AR(p) models (Staudenmayer and Buonaccorsi, 2004) in the context of longitudinal studies.

We set  $\theta = 0.8057$ ,  $D = 0.0064$ ,  $Q = 1.44$  and  $R = 1$  for the simulation studies. Data  $Y_{it}$  were generated from the model (33)-(35) with  $m = 20, 60$  and  $n_1 = n_2 = \dots = n_m = n$  where  $n = 10, 30$ . We assumed the following priors for the Bayesian approach

$$\begin{aligned} \theta &\sim N(\eta, \Delta), \quad D^{-1} \sim G(\beta_0, \beta_1), \\ Q^{-1} &\sim G(\nu_0, \nu_1), \quad R^{-1} \sim G(\omega_0, \omega_1), \quad x_{i0} \sim N(\tau, A). \end{aligned} \quad (36)$$

Here  $G(., .)$  stands for the Gamma distribution. The full conditional distributions were derived and shown in the Appendix. We specified the priors by setting  $\eta = 0.5$ ,  $\Delta = 4$ ,  $\beta_0 = 0.5$ ,  $\beta_1 = 0.5$ ,  $\nu_0 = 0.5$ ,  $\nu_1 = 0.5$ ,  $\omega_0 = 0.5$ ,  $\omega_1 = 0.5$ ,  $\tau = 20$ , and  $A = 100$ . Under this set-up the prior distributions for  $D$ ,  $Q$  and  $R$  were improper priors.

**(Display FIGURE 1 and FIGURE 2 Here)**

**Table 1.** Parameter estimation for the univariate model using the Bayesian approach and the EM algorithm. 100 replicates were simulated.  $m = 20, 60$  and  $n = 10, 30$ .

m		20				60			
		10		30		10		30	
n		B	EM	B	EM	B	EM	B	EM
$\hat{\theta}$	Mean	0.8012	0.8017	0.7992	0.8061	0.8011	0.8032	0.7982	0.8057
	Bias	-0.0045	-0.004	-0.0065	0.0003	-0.0047	-0.0026	-0.0075	0
	MSE	0.0004	0.0003	0.0003	0.0002	0.0001	0.0001	0.0002	0.0001
	RE	0.025	0.0217	0.0217	0.0177	0.0125	0.0125	0.0177	0.0125
$\hat{D}$	Mean	0.0647	0.0066	0.064	0.006	0.0254	0.0067	0.0252	0.0053
	Bias	0.0583	0.00019	0.0576	-0.0004	0.019	0.0003	0.0188	-0.0011
	MSE	0.0034	$3 \times 10^{-6}$	0.0033	$5 \times 10^{-6}$	0.0004	$2 \times 10^{-6}$	0.0004	$3 \times 10^{-6}$
	RE	9.1109	0.27	8.9759	0.34	3.125	0.21	3.125	0.25
$\hat{Q}$	Mean	1.5313	0.0844	1.4864	1.1257	1.4173	0.0246	1.4583	1.1535
	Bias	0.0913	-1.3556	0.0464	-0.314	-0.0228	-1.4154	0.0183	-0.2865
	MSE	0.2787	1.88281	0.0458	0.1351	0.0874	2.0043	0.0139	0.0989
	RE	0.3666	0.9529	0.1486	0.2552	0.2053	0.9831	0.0819	0.2184
$\hat{R}$	Mean	0.9568	2.3202	0.9728	1.1427	0.9989	2.3543	0.9864	1.1636
	Bias	-0.0432	1.3202	-0.0272	0.1427	-0.0011	1.3543	-0.0136	0.1636
	MSE	0.1211	1.8831	0.0306	0.0475	0.0429	1.8572	0.0074	0.0413
	RE	0.348	1.3723	0.1749	0.2179	0.2071	1.3628	0.086	0.2032

The simulation studies were performed on a Dell Dimension 4600 with 3.0G Hz CPU and 1GB memory. All programs were written in Visual C++, and built with the Release option. We simulated 100 replicates, and for each replicate 20000 Gibbs samples were generated for both the Bayesian approach and the EM algorithm. With 60 subjects and 30 observations the Bayesian approach required about 24 seconds for each replicate. The EM algorithm was much slower: for each replicate, around 60 iterations were required for convergence with each iteration consuming about 38 seconds. Thus the overall computational cost of the EM algorithm was about 95 times as much as the Bayesian approach for this example.

Table 1 shows the results of parameter estimation by the Bayesian approach and the EM algorithm. In Table 1,  $RE = \sqrt{MSE}/True\ Parameter$  represents the relative error. The Bayesian estimates of  $\theta$ ,  $Q$  and  $R$  were very good in all cases. The estimates of  $D$  were not satisfactory with large relative errors. We speculated the reason was that only  $m$  “data” ( $\theta_i$ ’s) contributed to the conditional distribution of  $D^{-1}$  in (20). Thus the prior distribution of  $D^{-1}$  could have strong influence on the sampling of  $D^{-1}$ . However, as a large

number ( $2mn$ ) of data were presented in the conditional distribution for  $Q^{-1}$  or  $R^{-1}$  (21, 22), the prior distributions could be less influential on the estimation of  $Q$  and  $R$ . For a fixed number of subjects, increasing the number of observations from 10 to 30 did not seem to improve the estimation of  $D$  in the Bayesian approach: for  $m = 10$ , the relative errors for  $\hat{D}$  were 9.11 and 8.98 when  $n = 10$  and  $n = 30$ , respectively; for  $m = 60$  the relative errors were the same for different  $n$ 's. However, the estimate of  $D$  improved noticeably as the number of subjects increased from 20 to 60. On the other hand, the EM algorithm gave good estimates for  $\theta$  and  $D$  in all cases, but the estimates for  $Q$  and  $R$  were not as good as their Bayesian counterparts. The estimates for  $Q$  and  $R$  were especially poor for  $n = 10$ , and were significantly improved as  $n$  was increased to 30: for  $m = 60$  the relative errors for  $\hat{R}$  were 1.36 and 0.203 for  $n = 10$  and  $n = 30$  respectively.

As mentioned previously we observed slow convergence of the EM algorithm. The slow convergence was possibly caused by the linear convergence rate of the EM algorithm, and by the large amount of data ( $X_{it}$  and  $b_i$ ) considered as missing (McLachlan and Krishnan, 1997).

**Example 2:** Bivariate Model. Consider a bivariate model with  $p = 2$ ,  $q = 1$ ,  $r = 2$ ,  $\mathbf{F}_i = \begin{pmatrix} 0.7 & 0 \\ \theta_{1i} & \theta_{2i} \end{pmatrix}$  and  $G = (1 \ 1)$ . Set  $\boldsymbol{\theta} = (\theta_1 \ \theta_2) = (0.3 \ 0.6)$ ,  $\mathbf{D} = \sigma_D^2 \mathbf{I}$ ,  $\mathbf{Q} = \sigma_Q^2 \mathbf{I}$ . The values and prior distributions of  $\sigma_D^2$ ,  $\sigma_Q^2$  and  $R$  were chosen as in Example 1. We assumed a normal prior distribution  $N(\boldsymbol{\eta}, \boldsymbol{\Delta})$  for  $\boldsymbol{\theta}$  with  $\boldsymbol{\eta} = (0.5, 0.5)'$  and  $\boldsymbol{\Delta} = \sigma_{\Delta}^2 \mathbf{I} = 2\mathbf{I}$ . For this bivariate MESSM,  $\tilde{\mathbf{F}} = \begin{pmatrix} 0.7 & 0 \\ 0 & 0 \end{pmatrix}$  contained the known parameters in the system matrix  $\mathbf{F}_i$ . A simulation study of 50 replicates was carried out for  $m = 20$  and  $n = 20$ . The Gibbs sampler was run for 20000 iterations for each replicate. The estimation results are shown in Table 2. For both the Bayesian approach and the EM algorithm the estimates for  $\theta_1$  and  $\theta_2$  were close to the true values. The Bayesian estimates of  $\sigma_Q^2$  and  $R$  were better than the EM based estimates. However the EM algorithm again provided better estimation for  $\sigma_D^2$ .

Figure 1 shows the histogram of the Gibbs samples for  $\theta_1$  and the estimated marginal posterior density of  $\theta_1$  for one replicate. Notice that  $p(\theta_1 | \{\boldsymbol{\Phi} - \theta_1\}, \mathbf{Y}) = p(\theta_1 | \theta_{11}, \dots, \theta_{1m}, \sigma_D^2)$  with  $\boldsymbol{\Phi}$  being the collection of all the parameters and the state variables. Using the formula

(23) the estimated posterior density can be derived as

$$\hat{p}(\theta_1|\mathbf{Y}) \approx \frac{1}{J} \sum_{j=1}^J \text{N} \left( \theta_1 \mid \frac{\sigma_\Delta^2}{m\sigma_\Delta^2 + \sigma_D^2(j)} \sum_{i=1}^m \theta_{1i}^{(j)} + \frac{\sigma_D^2(j)}{m\sigma_\Delta^2 + \sigma_D^2(j)} \eta_1, \frac{\sigma_\Delta^2 \sigma_D^2(j)}{m\sigma_\Delta^2 + \sigma_D^2(j)} \right)$$

where  $\eta_1$  is the first element of  $\boldsymbol{\eta}$ , and  $\text{N}(\theta_1|a, b)$  stands for the normal density with mean  $a$  and variance  $b$ . The histogram shows the Gibbs samples for  $\theta_1$  are tightly centered around 0.3, and agrees very well with the estimated posterior density. To monitor the convergence of the Gibbs samples the trace plot is also shown in Figure 1. The trace plot does not display any signs of patterns, suggesting the convergence to the stationary distribution. The similar plots for  $\theta_2$  are displayed in Figure 2. It can be seen that the Gibbs samples for  $\theta_2$  are concentrated around 0.6, the true parameter value of  $\theta_2$ .

**Table 2.** Parameter estimation for the bivariate model. 50 replicates,  $m = 20$  and  $n = 20$ .

	$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\sigma}_D^2$		$\hat{\sigma}_Q^2$		$\hat{R}$	
	B	EM	B	EM	B	EM	B	EM	B	EM
Mean	0.306	0.293	0.592	0.606	0.0296	0.006	1.03	1.164	1.343	1.122
Bias	0.006	-0.007	-0.008	0.006	0.02	-0.0004	0.032	0.164	-0.09	-0.318
MSE	0.0016	0.0012	0.0017	0.0014	0.0006	$3 \times 10^{-6}$	0.07	0.215	0.166	0.345
RE	0.133	0.115	0.069	0.062	3.827	0.271	0.264	0.463	0.283	0.401

**Example 3:** HIV Dynamics. We applied the MESSM to a data set from the AIDS Clinical Trial Group (ACTG) Protocol 315 (Wu and Ding, 1999). In this protocol, 53 HIV infected patients were treated with potent antiviral drugs (ritonavir, 3TC, and AZT). Among the 53 treated patients, 5 dropped out during the treatment. Treatments started on day 0 and the plasma concentration of HIV-1 RNA was repeatedly measured on days 0, 2, 7, 10, 14, 21, 28 and weeks 8, 12, 24, 48 after the initiation of the treatment. Note that because only the plasma HIV RNA concentration was observed in this study, the MESSM based on Perelson's 3-compartment model shown in Section 2 could not be identified. Here we considered the MESSM (33)-(34) for the ACTG 315 data. This model was closed related to the simple

HIV dynamic model proposed by Ho *et al.* (1995). Although this is an oversimplified HIV dynamic model, it is good enough for us to illustrate our methodologies. In this example,  $Y_{it}$  is the base 10 logarithm of the measured viral load from patient  $i$  at time  $t$ , and  $X_{it}$  represents the base 10 logarithm of the actual viral load.  $\theta_i$  indicates the viral clearance rate for patient  $i$ .

The same prior distributions as in the Example 1 were utilized for the Bayesian approach. Table 3 gives the estimation values of both the population parameters ( $\theta$   $D$   $Q$   $R$ ) and the individual parameters  $\theta_i$  ( $i = 1, \dots, 48$ ). The population parameters were estimated using the Bayesian approach and the MLE. It could be seen that the two methods gave similar estimation results. The individual parameters were estimated by the Bayesian approach. Similar estimation results for the individual parameters could be obtained using the EM approach.

**[Display FIGURE 3 Here]**

Figure 3 shows the one-step-ahead Kalman forecasts (dashed curve) of the individual states for four selected patients using the estimated parameters listed in Table 3. The estimated posterior means for the individual states using the Gibbs sampler approach are also shown in solid curve. While the estimation based on the Gibbs sampler approach gives smooth curves, the Kalman forecasts contain jumps at the observation time points, indicating on-line adjustments based on the observations. Table 4 shows the one-step-ahead forecasts of the viral loads for these four patients on the original scale. We observed from Figure 3 and Table 4 that the forecasts tended to overestimate the observed viral loads in the early stage and underestimate them in the later stage of the treatment. The main reason was that the HIV dynamics usually contained two or more phases of decay after the initiation of a potent antiviral therapy (Perelson *et al.* , 1997). The first phase featured a rapid drop in viral load due to the exponential death rate of the productively infected cells. A slower second phase was due to the slow clearance rate of the long lived infected T-cells or latently infected T-cells. This implied that the true viral decay rate should be time-varying. Thus when fitting the model with constant decay rate  $\theta_i$ , we underestimated the early decay rate,

but overestimated the later decay rate. We also fit a two-phase MESSM to the ACTG 315 data and the forecasting results were much improved (results not shown).

**Table 3.** Parameter Estimation for the HIV dynamic model

Estimation of population parameters				
$\hat{\theta}$	$\hat{D}$	$\hat{Q}$	$\hat{R}$	
Bayes	0.958	0.041	0.0443	0.653
EM	0.953	0.036	0.0143	0.662

Estimation of the individual parameters  $\theta_i$  for 48 patients

0.9801	0.9894	0.9729	0.9844	0.9821	0.9836	0.972	0.9735
0.9022	0.9719	0.0714	0.9779	0.986	0.9811	0.9885	0.9733
0.9808	0.9687	0.9778	0.9836	0.9837	0.9644	0.9595	0.9829
0.9734	0.9871	0.9831	0.9827	0.9786	0.9814	0.9878	0.9865
0.9614	0.9756	0.9759	0.9717	0.9552	0.9829	0.9695	0.9865
0.9889	0.9868	0.987	0.9782	0.9805	0.9846	0.976	0.9542

**Table 4.** Forecasts of viral loads for four patients

Patient 1									
Day	0	2	7	8	16	22	29	57	91
RNA	23000	23000	3400	1200	950	380	440	130	100
Forecast	NA	15486	6423	4501	966	440	188	24	8
Patient 2									
Day	0	2	7						
RNA	6200	2400	100						
Forecast	NA	1221	73						
Patient 3									
Day	0	7	9	14	21	28	56		
RNA	59000	13000	7000	1700	800	600	270		
Forecast	NA	14682	9853	3858	1051	395	36		
Patient 4									
Day	0	2	7	10	14	22	28	56	
RNA	8500	9000	410	410	420	440	160	100	
Forecast	NA	5290	1897	708	312	111	88	9	

## 6 Discussion

In this paper, we have proposed a class of mixed-effects state space models for analysis of longitudinal dynamic systems that may arise from biomedical longitudinal studies. In many biomedical studies, the underlying biological mechanisms are well studied, and mathematical representations, usually a set of differential or difference equations, are available. In these cases, we introduce the mixed-effects modelling idea into state space models for longitudinal dynamic system analysis. With the intention of giving a comprehensive treatment of this subject, we investigated two methods for estimation of unknown parameters in the proposed models by borrowing ideas from standard mixed-effects models. The concept of population state was introduced, and the EM-enhanced smoothing was proposed. Filtering, smoothing and forecasting algorithms for estimation of individual states were also discussed based on the Kalman filter methods.

We introduced the mixed effects in the state space models under the assumptions of linearity and normality. We also assumed the system matrices  $\mathbf{F}_i$  and  $\mathbf{G}_i$  in (3)-(4) were linear of the individual parameter  $\theta_i$ . If the structures of  $\mathbf{F}_i$  and  $\mathbf{G}_i$  are complicated, as shown in the following examples, several strategies could be applied. In the first example adopted from Stengel (1994), we consider  $\mathbf{F}_i = \begin{pmatrix} 0 & 1 \\ -\theta_{1i}^2 & -2\theta_{2i}\theta_{1i} \end{pmatrix}$  and assume  $\mathbf{G}_i$  are known. Simple reparameterization  $\theta_{1i}^2 = p_{1i}, \theta_{2i} = \frac{p_{2i}}{2\sqrt{p_{1i}}}$  leads to the linear functional form of  $(p_{1i} \ p_{2i})$  in the  $\mathbf{F}_i$ , and the two estimation methods in Section 4 could be directly applied. Note that for the Bayesian approach, once the posterior distribution of  $(p_1 \ p_2)$  is estimated, the posterior distribution of  $(\theta_1 \ \theta_2)$  can be computed by adjusting the Jacobian matrix for the reparameterization. In the second example, we consider  $\mathbf{F}_i = \begin{pmatrix} \theta_{1i}\sin(\theta_{2i}) & \theta_{1i} \\ 0 & \theta_{2i} \end{pmatrix}$  in which reparameterization would not help. However, instead of sampling from the conditional joint distribution of  $(\theta_{1i} \ \theta_{2i})$  as in (19), we could formulate the conditional distributions of  $\theta_{1i}$  and  $\theta_{2i}$  given the other, and then apply sampling methods such as the accept-reject algorithm (Robert and Casella, 2004) or embed a Metropolis-Hastings step to sample from the conditional distributions. The drawback of such approach is that the Gibbs samples could have higher correlation and thus may be less efficient. However, this sampling scheme

is very flexible in dealing with complicated  $\mathbf{F}_i(\boldsymbol{\theta}_i)$ , and  $\mathbf{G}_i(\boldsymbol{\theta}_i)$ . Accordingly, for the EM algorithm approach, the Gibbs sampling in the E-step should be modified. In the M-step, while the estimates of  $\mathbf{Q}$ ,  $\mathbf{R}$  and  $\mathbf{D}$  satisfy equations (27)-(29), the equation for the estimate of  $\boldsymbol{\theta}$  would not have the clean form of (26). A nonlinear optimization procedure might be considered to update the estimate of  $\boldsymbol{\theta}$ .

Another extension of the proposed MESSM may include covariates in both the state and the observation equations. The full conditional distributions for the parameters and the state variables may be derived without much difficulty, and the construction of the EM algorithm is also straightforward. We want to mention that practitioners should be cautious of the identifiability problems as the models get more complicated. To our knowledge, for general state space models, theoretical results for detecting identifiability problems are limited. We are currently pursuing a solution using numerical methods.

The MESSM may be further extended to nonlinear and non-Gaussian cases. It is well-known that for nonlinear and non-Gaussian state space models the Kalman estimates of states are not optimal in terms of unbiasedness and mean squared errors. Other filters, such as the extended Kalman filter (Anderson and Moore, 1979), usually involve linear approximation of the nonlinear functions. Although lacking optimality and often awkward when handling complicated models, these filters have been applied extensively simply because there were no other alternative methods. In the past decade, the Bayesian approach and the Gibbs sampler techniques have provided good solutions for smoothing problems in nonlinear non-Gaussian state space models (Carlin, Polson and Stoffer, 1992; Carter and Kohn, 1994 and 1996). Such approach could be adapted to nonlinear non-Gaussian MESSM. A major limitation for the Gibbs sampler approach lies in its incapability of online filtering. Recently sequential Monte Carlo methods (Liu and Chen, 1998; Doucet, Freitas and Gordon, 2001), such as the particle filter (Gordon, 1993; Kitagawa 1996), have provided a general framework for online filtering for nonlinear and non-Gaussian state space models. However, when the particle filter approach is extended to MESSM, the extra layer of the mixed effects in the system matrices of MESSM adds significant computational difficulties. We believe this is a worthy topic for the future research.

In summary, we have introduced the MESSM which brings together the techniques of

mathematical modelling, time series analysis, and longitudinal data analysis to deal with longitudinal dynamic systems with many potential applications in biomedical research. Although we only provide a simple example of modeling HIV dynamics in AIDS research, some other potential applications may include hepatitis virus dynamics (Nowak *et al.* 1996), tumor dynamics in cancer research (Martin and Teo, 1994; Swan, 1984), and genetic regulatory network modeling (Holter *et al.* 2001; Chen, He and Church, 1999; De Hoon *et al.* 2003). However, many theoretical and methodological problems for MESSM remain unsolved. We expect that our paper can attract more attention and stimulate more research in this promising research area.

## References

1. Anderson, B.D.O and Moore, J.B (1979), *Optimal Filtering*, Prentice-Hall.
2. Brockwell, P.J. and Davis, R.A. (1991), *Time Series: Theory and Methods*, Springer-Verlag.
3. Carlin, B.P., Polson, N.G., and Stoffer, D.S. (1992), "A Monte Carlo Approach to Nonnormal and Nonlinear State-space Modeling," *Journal of American Statistical Association*, 87, 493-500.
4. Carter, C.K. and Kohn, R. (1994), "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541-553.
5. Carter, C.K. and Kohn, R. (1996). "Markov Chain Monte Carlo in Conditionally Gaussian State Space Models," *Biometrika*, 83, 589-601.
6. Chen R. and Liu J. (2000), "Mixture Kalman Filters," *Journal of the Royal Statistical Society, Series B*, 62, 493-508.
7. Chen, T., He, H.L., and Church, G.M. (1999), "Modeling Gene Expression with Differential Equations," *Pacific Symposium of Biocomputing*, 1999, 29-40.
8. De Hoon, M.J.L., Imoto, S., Kobayashi, K., *et al.* (2003), "Inferring Gene Regulatory Networks from Time-Ordered Gene Expression Data of *Bacillus Subtilis* Using Differential Equations," *Pacific Symposium of Biocomputing*, 2003, 17-28.
9. Davidian, M. and Giltinan, D.M. (1995), *Nonlinear Models for Repeated Measurements Data*. Chapman & Hall.
10. Dempster, A.P., Laird, N.M and Rubin, D.B. (1977), "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, 39, 1-38.
11. De Jong, P. and Shephard, N. (1995), "The Simulation Smoother for Time Series Models," *Biometrika*, 82, 339-350.

12. Delyon, B., Lavielle, M., and Moulines, E. (1999), "Convergence of a stochastic approximation version of the EM algorithm," *Annals of Statistics*, 27, No.1, 94-128.
13. Diggle, P., Liang, K.Y., and Zeger S. (1994), "Analysis of Longitudinal Data," Oxford Science.
14. Doucet, A., de Freitas, J.F.G., and Gordon, N. (2001), *Sequential Monte Carlo Methods in Practice*, Springer, New York.
15. Engle, R.F. and Watson, M.W., (1981), "A one factor multivariate time series models of metropolitan wage rates," *Journal of American Statistical Association*, 76, 774-781.
16. Geman, S. and Geman D., (1984), "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, No.6, 721-741.
17. Gordon, N.J., Salmond, D.J. and Smith, A.F.M. (1993), "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEEE Proceedings-F*, 140, 107-113.
18. Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.
19. Holter, N.S., Maritan, A., Cieplak, M., *et al.* (2001), "Dynamic Modeling of Gene Expression Data," *PNAS*, 98, 1693-1698.
20. Ho D.D., Neumann A.U., Perelson A.S., Chen W, Leonard J.M. and Markowitz M., (1995), "Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection," *Nature*, 373, 123-126.
21. Jensen J.L. and Petersen, N.V., (1999), "Asymptotic normality of the maximum likelihood estimator in state space models," *The Annals of Statistics*, 27, 514-535.
22. Jones, R.H. (1993), *Longitudinal Data with Serial Correlation: A State-space Approach*, New York, Chapman & Hall.
23. Kalman, R.E. (1960), "A new approach to linear filtering and prediction theory," *Journal of Basic Engineering, Transactions ASME, Series D*, 82, 35-45.

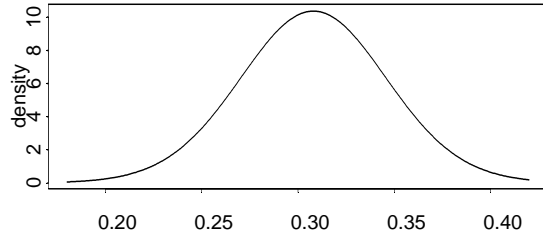
24. Kitagawa, G. (1987), "Non-Gaussian state space modelling of nonstationary time series (with discussion)," *Journal of American Statistical Association*, 82, 1032-63.
  25. Kitagawa, G. (1996), "Monte Carlo filter and smoother for nonlinear non-Gaussian state space models," *Journal of Computational and Graphical Statistics*, 5, 1-25.
  26. Laird, N.M. and Ware, J.H. (1982), "Random effects models for longitudinal data," *Biometrics*, 38, 963-974.
  27. Liu, J.S., Chen, R. (1998), " Sequential Monte Carlo methods for dynamic systems," *Journal of the American Statistical Association*, 93, 1032-1044.
  28. Martin, R., and Teo, K.L. (1994), *Optimal Control of Drug Administration in Cancer Chemotherapy*, World Scientific, Singapore.
  29. McLachlan, G.J. and Krishnan, T. (1997), *The EM Algorithm and Extensions*, Wiley, New York.
  30. Nowak, M. A., Bonhoeffer, S., Hill, A.M. , Boehme, R., Thomas, H., and McDade, H. (1996), "Viral dynamics in hepatitis B virus infection," *Proc. Natl. Acad. Sci. USA*, 93,4398-4402.
  31. Pagan, A. (1980), "Some Identification and Estimation Results for Regression Models with Stochastically Varying Parameters," *Journal of Econometrics*, 13, 341-363.
- inter-block information when
32. Perelson, A.S., Neumann A.U., Markowitz, M., Leonard, J.M., and Ho, D.D. (1996), "HIV-1 Dynamics in Vivo: virion clearance rate, infected cell life-span, and viral generation time," *Science*, Vol 271, 1582-1586.
  33. Perelson, A.S., Essunger P., and Cao, Y. (1997), "Decay Characteristics of HIV-1 Infected Compartments during Combination Therapy," *Nature*, Vol 387, 188-191.
  34. Rahiala, M. (1999), "Random Coefficient Autoregressive Models for Longitudinal Data," *Biometrika*, 86, 718-722.

35. Robert, C.P. and Casella G. (2004), *Monte Carlo Statistical Methods*, Springer.
36. Schweppe, F. (1965), "Evaluation of likelihood functions for Gaussian signals," *IEEE Transactions on Information Theory*, 11, 61-70.
37. Staudenmayer, J. and Buonaccorsi, J. (2004), "Measurement Error in Linear Autoregressive Models," *Journal of the American Statistical Association*, to appear.
38. Steimer, J.L., Mallet, A., Golmard, J. and Boiesieux, J. (1984), "Alternative Approaches to Estimation of Population Pharmacokinetic Parameters: Comparison with the Non-linear Mixed Effects Model," *Drug Metabolism Reviews*, 15, 265-292.
39. Stengel, R.F. (1993), *Optimal Control and Estimation*, Dover, New York.
40. Swan, G.W. (1984). *Applications of Optimal Control Theory in Biomedicine*, Marcel Dekker, Inc., New York.
41. Vonesh, E.F. and Chinchilli, V.M. (1996), *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, Marcel Dekker, New York.
42. Watson, M.W. and Engle, R.F. (1983), "Alternative Algorithms for the Estimation of Dynamic Factor, Mimic and Varying Coefficient Regression Models," *Journal of Econometrics*, 23, 385-400.
43. West, M. and Harrison, J. (1997), *Bayesian Forecasting and Dynamic models*, New York, Springer.
44. Wu, H. and Ding, A. (1999), "Population HIV-1 Dynamics in Vivo: Applicable Models and Inferential Tools for Virological Data from AIDS Clinical Trials," *Biometrics*, 55, 410-418. The data are available online at <http://stat.tamu.edu/biometrics/datasets/980622.txt>.
45. Wu, H., Ding, A. and DeGruttola. V. (1998), "Estimation of HIV Dynamic Parameters," *Statistics in Medicine*, 17, 2463-2485.

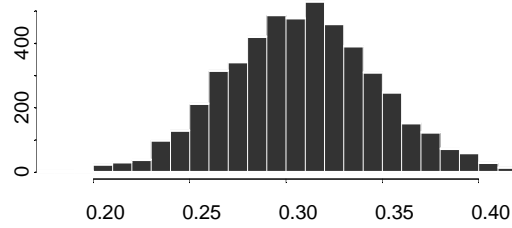
## Appendix

For the univariate MESSM (33)-(35) and the priors (36) the full conditional distributions are given as follows.

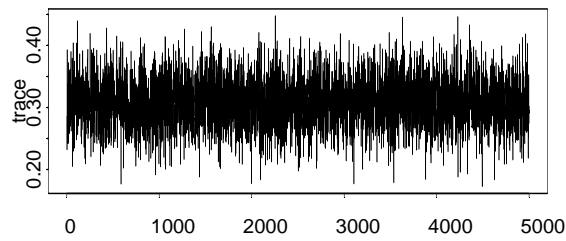
$$\begin{aligned}
p(\theta_i | \mathbf{X}, \mathbf{Y}, \theta, D, \Theta_{-i}, Q, R) &\sim N \left( \frac{D \sum_{t=1}^{n_i} X_{i,t-1} X_{it} + \theta Q}{D \sum_{t=1}^{n_i} X_{i,t-1}^2 + Q}, \frac{QR}{D \sum_{t=1}^{n_i} X_{i,t-1}^2 + Q} \right), \\
p(\theta | \mathbf{X}, \mathbf{Y}, D, \Theta, Q, R) &\sim N \left( \frac{\Delta \sum_{i=1}^m \theta_i + \eta D}{D + m\Delta}, \frac{D\Delta}{D + m\Delta} \right), \\
p(X_{it} | \mathbf{Y}, X_{i,k \neq t}, \theta, D, \Theta, Q, R) &\sim N \left( \frac{\theta_i R (X_{i,t-1} + X_{i,t+1}) + Q Y_{it}}{(1 + \theta_i^2) R + Q}, \frac{QR}{(1 + \theta_i^2) R + Q} \right), \\
p(X_{i0} | \mathbf{Y}, X_{i,k \neq 0}, \theta, D, \Theta, Q, R) &\sim N \left( \frac{Q\tau + A\theta_i X_{i1}}{Q + \theta_i^2 A}, \frac{QA}{Q + \theta_i^2 A} \right), \\
p(X_{in_i} | \mathbf{Y}, X_{i, \neq n_i}, \theta, D, \Theta, Q, R) &\sim N \left( \frac{\theta_i R X_{i,n_i-1} + Q Y_{in_i}}{R + Q}, \frac{QR}{R + Q} \right), \\
p(D^{-1} | \mathbf{X}, \mathbf{Y}, \theta, \Theta, Q, R) &\sim G \left( \beta_0 + \frac{m}{2}, \beta_1 + \frac{\sum_{i=1}^m (\theta_i - \theta)^2}{2} \right), \\
p(Q^{-1} | \mathbf{X}, \mathbf{Y}, \theta, \Theta, R, D) &\sim G \left( \nu_0 + \frac{s}{2}, \frac{\nu_1 + \sum_{i=1}^m \sum_{t=1}^{n_i} (X_{it} - \theta_i X_{i,t-1})^2}{2} \right), \\
p(R^{-1} | \mathbf{X}, \mathbf{Y}, \theta, \Theta, Q, D) &\sim G \left( \omega_0 + \frac{s}{2}, \frac{\omega_1 + \sum_{i=1}^m \sum_{t=0}^{n_i} (Y_{it} - X_{it})^2}{2} \right).
\end{aligned}$$



(a)



(b)



(c)

Figure 1: Parameter Estimation for  $\theta_1$  (Bivariate Model). (a) Estimated posterior density of  $\theta_1$ . (b) Histogram of the Gibbs  $\theta_1^{(j)}$ . (c) Trace plot for the last 5000 samples.

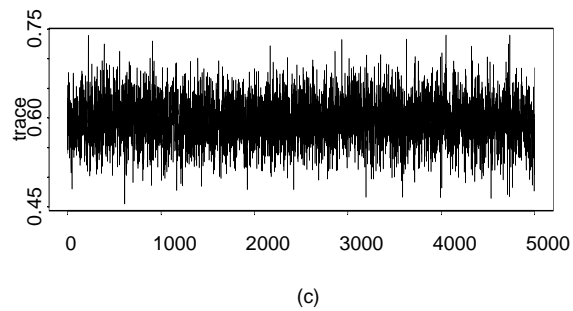
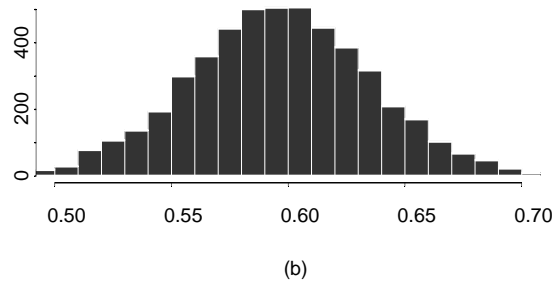
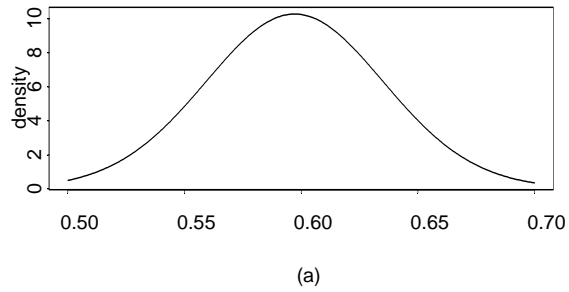


Figure 2: Parameter Estimation for  $\theta_2$  (Bivariate Model). (a) Estimated posterior density of  $\theta_2$ . (b) Histogram of the Gibbs  $\theta_2^{(j)}$ . (c) Trace plot for the last 5000 samples.

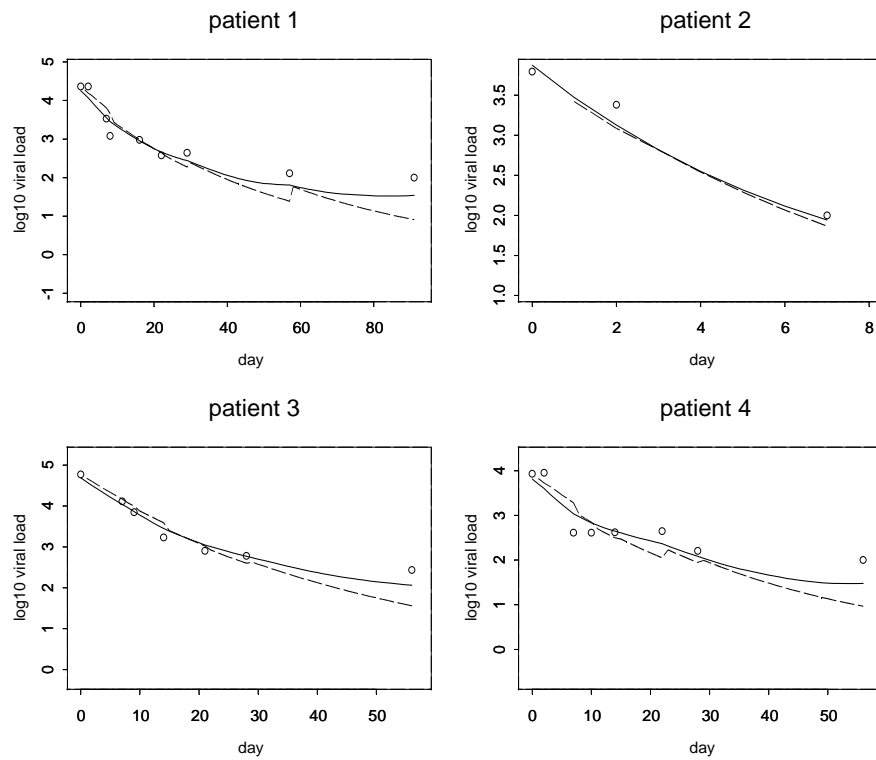


Figure 3: Estimation of Viral Load for Four Patients. Base 10 logarithm of the viral loads are shown in circles. Solid curves show the smoothing estimate using the Gibbs sampler approach. Dashed curves show the one-step-ahead Kalman forecasts.