NIH Public Access
Author Manuscript
Biometrics. Author manuscript; available in PMC 2012 November 27.

Published in final edited form as:
Biometrics. 2011 June ; 67(2): 476–485. doi:10.1111/j.1541-0420.2010.01485.x.

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

Dacheng Liu[1], **Tao Lu**[2], **Xu-Feng Niu**[3], and **Hulin Wu**[2,*]

[1]Boehringer-Ingelheim Pharmaceuticals, Ridgefield, Connecticut, U.S.A

[2]Department of Biostatistics and Computational Biology, University of Rochester School of Medicine and Dentistry, Rochester, New York, U.S.A

[3]Department of Statistics, Florida State University, Tallahassee, Florida, U.S.A

Summary

The rapid development of new biotechnologies allows us to deeply understand biomedical dynamic systems in more detail 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 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 method for standard mixed-effects models and state space models are modified and investigated for estimating unknown parameters in the proposed models. In the Bayesian approach, 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 studies are conducted to compare the two proposed methods. 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.

Keywords

EM algorithm; Gibbs sampler; Kalman filter; Mixed-effects models; Parameter estimation; State space models

1. Introduction

State space modelling has been used mainly in time series data analysis. It has found application 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

*hwu@bst.rochester.edu.

Supplementary Materials
Web Appendix 1 is available under the Paper Information link at the Biometrics website http://www.tibs.org/biometrics.

$watermark-text
$watermark-text
$watermark-text

$watermark-text

$watermark-text

$watermark-text

$$\boldsymbol{x}_t = \boldsymbol{F}_t \boldsymbol{x}_{t-1} + \boldsymbol{v}_t, \quad (1)$$

$$\boldsymbol{y}_t = \boldsymbol{G}_t \boldsymbol{x}_t + \boldsymbol{w}_t, \quad (2)$$

where $\boldsymbol{x}_t$ is the state vector, $\boldsymbol{y}_t$ is the observation vector, $\boldsymbol{F}_t$ is the state transition matrix, and $\boldsymbol{G}_t$ is the observation matrix. It is assumed that vectors $\boldsymbol{v}_t$ and $\boldsymbol{w}_t$ are independent and identically distributed with $\boldsymbol{v}_t \sim \mathrm{N}(0, \boldsymbol{Q})$ and $\boldsymbol{w}_t \sim \mathrm{N}(0, \boldsymbol{R})$. The system matrices ($\boldsymbol{F}_t$, $\boldsymbol{G}_t$) and the covariance matrices ($\boldsymbol{Q}$, $\boldsymbol{R}$) 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 system matrices contain random effects, and the dimension and structure of the system matrices 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 reasonable 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, 1980; Jensen and Petersen, 1999).

The state space model provides rich covariance structures for the observations $\boldsymbol{y}_t$. It indeed covers the covariance structure of ARMA (Autoregressive Moving Average) models as the latter can be written into a state space form. Another interesting aspect of the state space model is its flexibility in modeling the underlying 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. In the very popular linear random-effects models (Laird and Ware, 1982)

$$\boldsymbol{y}_i = \boldsymbol{x}_i \beta + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{e}_i, \quad \boldsymbol{b}_i \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{D}), \quad \boldsymbol{e}_i \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Omega}_i),$$

the random effects $\boldsymbol{b}_i$ and errors $\boldsymbol{e}_i = [e_{i1}, \cdots, e_{in_i}]'$ account for the between-subject and within-subject variation, respectively. Covariance matrix $\boldsymbol{\Omega}_i$ may be diagonal or has a complicated structure (Jones, 1993; Diggle, Liang and Zeger, 1994). In particular, Jones (1993) considered the ARMA covariance structure for $\boldsymbol{\Omega}_i$. In this work ARMA models were transformed into state space forms, and the Kalman filter was used to compute the likelihood.

Stochastic models have been used to model longitudinal data. For example, Rahiala (1999) proposed a random-effects autoregressive (AR) model for a growth curve study in which the

$watermark-text

$watermark-text

$watermark-text

AR coefficients were treated as random effects. The state space approach was briefly discussed in Rahiala's paper for the computation of the likelihood with missing data.

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 $\boldsymbol{F}_t$ and $\boldsymbol{G}_t$ in (1)–(2) to account for between-individual variations 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 (Huang, Liu and Wu, 2006).

This paper is organized as follows. In Section 2, we introduce the MESSM, state some basic assumptions on the system matrices, and give an illustration example based on an HIV dynamic model. Section 3 presents estimation methods for both the individual and the population state variables. The Bayesian approach and the maximum likelihood (ML) method are proposed to estimate unknown parameters in the proposed models in Section 4. In Section 5 we present 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 of the MESSM and other related issues.

2. Model Specification

A linear MESSM can be written as

$$\boldsymbol{x}_{i,t+1}=\boldsymbol{F}(\theta_i)\boldsymbol{x}_{it}+\boldsymbol{v}_{it}, \quad \boldsymbol{v}_{it}\sim \mathrm{N}(0,\boldsymbol{Q}), \quad (3)$$

$$\boldsymbol{y}_{it}=\boldsymbol{G}(\theta_i)\boldsymbol{x}_{it}+\boldsymbol{w}_{it}, \quad \boldsymbol{w}_{it}\sim \mathrm{N}(0,\boldsymbol{R}), \quad (4)$$

where $\boldsymbol{y}_{it}$ is the $q \times 1$ vector of observations for the ith ($i = 1, \ldots, m$) subject at time t ($t = 1, \ldots, n_i$), $\boldsymbol{x}_{it}$ is the $p \times 1$ state vector, $\boldsymbol{v}_{it}$ is the $p \times 1$ dynamic disturbance vector, and $\boldsymbol{w}_{it}$ is the $q \times 1$ vector of the observation errors. $\{\boldsymbol{v}_{it}\}$ and $\{\boldsymbol{w}_{it}\}$ are assumed to be mutually independent within themselves and independent to each other. We assume that the initial values $\boldsymbol{x}_{i0}$'s are independent and $\boldsymbol{x}_{i0} \sim \mathrm{N}(\boldsymbol{\tau}, \boldsymbol{A})$. 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 $\boldsymbol{F}(\boldsymbol{\theta}_i)$ and the $q \times p$ observation-system matrix $\boldsymbol{G}(\boldsymbol{\theta}_i)$ are parameterized with the $r \times 1$ parameter vector $\boldsymbol{\theta}_i$, and the $p \times p$ matrix $\boldsymbol{Q}$ and the $q \times q$ matrix $\boldsymbol{R}$ with the $d \times 1$ vector of parameters $\boldsymbol{\omega}$. Note that the parameters for $\boldsymbol{F}$ and $\boldsymbol{G}$ are not necessarily same. They may share some common parameters or may have different parameters. However, for presentation convenience, we use $\boldsymbol{\theta}_i$ to denote all the unknown parameters in both $\boldsymbol{F}$ and $\boldsymbol{G}$. Similar handling is applied for $\boldsymbol{\omega}$. Recognized as a vector of individual parameters, $\boldsymbol{\theta}_i$ can be modelled as

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

where $\boldsymbol{\theta}$ is the fixed effect or the population parameter vector ($r \times 1$), and the vectors $\boldsymbol{b}_i$ are the random effects, which are assumed to be i.i.d. random vectors. 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).

$watermark-text

$watermark-text

$watermark-text

We give an illustration example for the proposed MESSM based on an HIV dynamic model proposed by Perelson *et al.* (1997). This model consists of three differential equations describing the interaction between the target cells and HIV infection. It was assumed that the antiviral therapy completely blocked the reproduction of infectious HIV virions, thus the newly produced virions were all noninfectious. The model was written as

$$\frac{dT^*}{dt}=kV_IT-\delta T^*,\quad \frac{dV_I}{dt}=-cV_I,\quad \frac{dV_{NI}}{dt}=N\delta T^*-cV_{NI},\quad (6)$$

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 a 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 to replace the differential equations (6). If the total plasma HIV concentration ($V = V_{NI} + V_I$) and T^* were observed, a mixed-effects state space model for system (6) could be formulated as

$$\boldsymbol{x}_{it}=\boldsymbol{F}_i\boldsymbol{x}_{i,t-1}+\boldsymbol{v}_{it},\quad (7)$$

$$\boldsymbol{y}_{it}=\boldsymbol{G}\boldsymbol{x}_{it}+\boldsymbol{w}_{it},\quad (8)$$

where $\boldsymbol{y}_{it}$ is the observed (T^* V) for the ith patient at time t,

$$\boldsymbol{x}_{it}=(\ T^*_{it}\quad V_{I_{it}}\quad V_{NI_{it}}\ )',\quad \boldsymbol{F}_i=\begin{bmatrix}1-\delta_i & k_iT_i & 0\\ 0 & 1-c_i & 0\\ N_i\delta_i & 0 & 1-c_i\end{bmatrix},\quad \boldsymbol{G}=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 1\end{bmatrix}.$$

The mixed effects $\theta_i=(\ 1-\delta_i\quad k_iT_i\quad N_i\delta_i\quad 1-c_i\ )'$ are assumed to follow

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

where $\theta=(\ 1-\delta\quad kT\quad N\delta\quad 1-c\ )'=(\ \theta_1\quad \theta_2\quad \theta_3\quad \theta_4\ )'$ are population parameters.

The proposed MESSM in this study is a hybrid of mixed-effects models and the state space models. It inherits the strengths of both types of 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 for 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. Moreover, state space models can flexibly incorporate various correlation structures for serially correlated longitudinal data.

$watermark-text

$watermark-text

$watermark-text

3. Estimation of State Variables

In this section we consider the estimation of both individual state variables and population state variables in a mixed-effects state space model for given parameters θ and ω.

3.1 The Kalman filter and the estimation of individual states

The estimation of an individual state x_{it} can be classified into three types based on the amount of information that is available. The estimation of x_{it} is called a one-step-ahead prediction if $\{y_{i1}, \cdots, y_{i,t-1}\}$ are observed; it is called a filtered estimation if $\{y_{i1}, \cdots, y_{it}\}$ are observed; and a smoothed estimation if $\{y_{i1}, \cdots, y_{in_i} (n_i > t)\}$ are observed. If the parameters θ_i and ω 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. We review the iterative algorithm of the Kalman filter in the following discussion. Details of the Kalman filter can be found in Harvey (1989), Brockwell and Davis (1991), and West and Harrison (1997).

Let $x_{it|k}$ denote the estimation of the individual state x_{it} given the observations $\{y_{i1}, \ldots, y_{ik}\}$. Let $P_{it|k}$ be the covariance matrix of the estimation error $x_{it|k} - x_{it}$. Use F_i and G_i to represent $F(\theta_i)$ and $G_i(\theta_i)$, respectively. The Kalman recursion proceeds as follows,

$$\begin{aligned} x_{it|t-1} &= F_i x_{i,t-1|t-1}, & P_{it|t-1} &= F_i P_{i,t-1|t-1} F_i' + Q, \\ f_{it} &= y_{it} - G_i x_{it|t-1}, & O_{it} &= G_i P_{it|t-1} G_i' + R, \\ x_{it|t} &= x_{it|t-1} + K_{it} f_{it}, & P_{it|t} &= P_{it|t-1} - K_{it} G_{it} P_{it|it-1}, \end{aligned}$$

where $K_{it} = P_{it|t-1} G_i' O_{it}^{-1}$ is the Kalman gain matrix, f_{it} is the innovation with covariance matrix O_{it}, $x_{it|t-1}$ and $x_{it|t}$ are the one-step-ahead prediction and the filtered estimation of x_{it}, respectively.

Let $P_{it,n} = E\{(x_{it} - x_{it|t-1})(x_{in} - x_{in|n-1}')\}$, $\Theta_{in} = F_i P_{in|n-1} G_i'$, and $\Delta_{in} = G_i P_{in|n-1} G_i' + R$. The fixed-point smoothing algorithm proceeds as follows:

$$x_{it|n} = x_{it|n-1} + P_{it,n} G_i' \Delta_{in}^{-1} f_{in}, \quad (9)$$

$$\begin{aligned} P_{it,n+1} &= P_{it,n} (F_i - \Theta_{in} \Delta_{in}^{-1} G_i)', \\ P_{it|n} &= P_{it|n-1} - P_{it,n} G_i' \Delta_{in}^{-1} G_i' P_{it,n}', \end{aligned} \quad (10)$$

with the initial condition, $P_{it,t} = P_{it|t-1}$.

Under the normality assumption, the Kalman estimator $x_{it|k}$ is the mean of x_{it} conditional on the data $\{y_{i1}, \ldots, y_{ik}\}$. 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 (Section 4.1).

3.2 Population state variable and the EM algorithm enhanced smoothing

From (3), we can write x_{it} in terms of x_{i0} and $\{v_{is}, s \le t\}$:

$$\boldsymbol{x}_{it}=\boldsymbol{F}_i^t\boldsymbol{x}_{i0}+\sum_{k=0}^{t-1}\boldsymbol{F}_i^k\boldsymbol{v}_{i,t-k}. \quad (11)$$

Here $\boldsymbol{F}_i = \boldsymbol{F}(\boldsymbol{\theta}_i)$ is a function of $\boldsymbol{\theta}_i$. Therefore, under the assumptions that the initial values $\boldsymbol{x}_{i0}$'s are i.i.d. and $\boldsymbol{\theta}_i$'s are i.i.d., Equation (11) implies that $\boldsymbol{x}_{it}$'s are i.i.d. at a given time point t. Thus the individual state variables $\boldsymbol{x}_{it}$ ($i = 1, \cdots, m$) follow a common mean process $\boldsymbol{x}_t$ plus a mean zero process, and can be modelled as

$$\boldsymbol{x}_{it}=\boldsymbol{x}_t+\boldsymbol{z}_{it}, \quad \boldsymbol{z}_{it}\sim(0,\boldsymbol{B}_t). \quad (12)$$

We call $\boldsymbol{x}_t$ the population state variable at time t, which represents the mean response of all individuals at time t. Vector $\boldsymbol{z}_{it}$ represents the deviation of the individual state $\boldsymbol{x}_{it}$ from the population variable $\boldsymbol{x}_t$ with $\boldsymbol{B}_t$ being the covariance matrix.

If the $\boldsymbol{x}_{it}$'s were observable, the maximum pseudo-likelihood estimation of $\boldsymbol{x}_t$ would then be the sample mean $\sum_{i=1}^{m}\boldsymbol{x}_{it}/m$. However, because in practice the $\boldsymbol{x}_{it}$'s are usually not directly observable, we instead use the Kalman smoothing estimate $\boldsymbol{x}_{it|n_i}(t < n_i)$, denoted by $\tilde{\boldsymbol{x}}_{it}$, to replace $\boldsymbol{x}_{it}$ in (12). Let $\boldsymbol{\varsigma}_{it} = \tilde{\boldsymbol{x}}_{it} - \boldsymbol{x}_{it}$ be the estimation error of $\tilde{\boldsymbol{x}}_{it}$, then $\boldsymbol{\varsigma}_{it}$ has mean 0 and covariance matrix $\boldsymbol{\Sigma}_{it} = \boldsymbol{P}_{it|n_i}$ as defined in Section 3.1. One can establish a variance-components model for $\tilde{\boldsymbol{x}}_{it}$

$$\tilde{\boldsymbol{x}}_{it}=\boldsymbol{x}_t+\boldsymbol{z}_{it}+\boldsymbol{\varsigma}_{it}. \quad (13)$$

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

$$l(\textstyle\sum_t,\boldsymbol{B}_t|\tilde{\boldsymbol{x}}_{it})=-\sum_{i=1}^{m}\{\log|\textstyle\sum_i+\boldsymbol{B}_t|+(\tilde{\boldsymbol{x}}_{it}-\boldsymbol{x}_t)'(\textstyle\sum_i+\boldsymbol{B}_t)^{-1}(\tilde{\boldsymbol{x}}_{it}-\boldsymbol{x}_t)\}.$$

Note that if we assume normality in model (13), the resulting likelihood function is equivalent to the pseudo-likelihood function. Thus, the EM algorithm developed by Steimer, Mallet, Golmard and Boisvieux (1984) to maximize the likelihood function under the normality assumption can be adapted as follows to maximize the pseudo-likelihood function for the estimation of $\boldsymbol{x}_t$ and $\boldsymbol{B}_t$:

1. Initialization: $\widehat{\boldsymbol{x}}_t^{(0)}=m^{-1}\sum_{i=1}^{m}\tilde{\boldsymbol{x}}_{it}$, $\widehat{\boldsymbol{B}}_t^{(0)}=m^{-1}\sum_{i=1}^{m}(\tilde{\boldsymbol{x}}_{it}-\widehat{\boldsymbol{x}}_t^{(0)})(\tilde{\boldsymbol{x}}_{it}-\widehat{\boldsymbol{x}}_t^{(0)})'$
2. E-step: Update the individual estimates of state variables:

$$\widehat{\boldsymbol{x}}_{it}^{(k+1)}=[(\widehat{\boldsymbol{B}}_t^{(k)})^{-1}+\textstyle\sum_{it}^{-1}]^{-1}[(\widehat{\boldsymbol{B}}_t^{(k)})^{-1}\widehat{\boldsymbol{x}}_t^{(k)}+\textstyle\sum_{it}^{-1}\tilde{\boldsymbol{x}}_{it}].$$

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

$watermark-text $watermark-text $watermark-text

$$\widehat{\boldsymbol{x}}_t^{(k+1)} = m^{-1}\sum_{i=1}^{m}\widehat{\boldsymbol{x}}_{it}^{(k+1)},$$
$$\widehat{\boldsymbol{B}}_t^{(k+1)} = m^{-1}\left\{\sum_{i=1}^{m}(\widehat{\boldsymbol{x}}_{it}^{(k+1)}-\widehat{\boldsymbol{x}}_t^{(k+1)})(\widehat{\boldsymbol{x}}_{it}^{(k+1)}-\widehat{\boldsymbol{x}}_t^{(k+1)})' + \sum_{i=1}^{m}((\widehat{\boldsymbol{B}}_t^{(k)})^{-1}+\Sigma_{it}^{-1})^{-1}\right\}.$$

4. Repeat Steps 2 and 3 until convergence is reached. Estimates of both individual states and population states are collected at the final step.

We have proposed to estimate the population state variable based on the smoothing estimates of the individual state variables. In a similar way, 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. Note that if the assumption that $\boldsymbol{x}_{i0}$'s are i.i.d. does not hold, then the population state variable $\boldsymbol{x}_t$ is no longer the mean of $\boldsymbol{x}_{it}$. In this case, the estimated $\hat{\boldsymbol{x}}_t$ using the EM algorithm may be considered as an average of the estimated individual state variables after adjusting for the estimation errors.

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 for mixed-effects models and the state space models have been well established (Harvey, 1989; Davidian and Giltinan, 1995; Vonesh and Chinchilli, 1996; 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). Nonlinear non-Gaussian state space models have been studied in, for example, Kitagawa (1987, 1996), Carlin, Polson, and Stoffer (1992), De Jong and Shephard (1995), and Chen and Liu (2000). Most of these studies focus on the estimation of state variables, while the estimation of unknown parameters in the models has not been paid enough attention. For MESSM we adopt the method proposed by Carlin, Polson and Stoffer (1992), in which the Gibbs sampler was used to draw samples from the full conditional distributions of state variables $\boldsymbol{x}_t$ and parameters conditional on the observations $\boldsymbol{y}_t$. This approach is straightforward and easy to implement, and it has produced good results in our simulation studies. However, 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($\boldsymbol{\theta}$, $\boldsymbol{D}$) is assigned to $\boldsymbol{\theta}_i$, then conceptually $\boldsymbol{\theta}$ and $\boldsymbol{D}$ 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 interpretations at the population level.

When the conjugate prior distributions are carefully chosen, 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(Υ, ν) denote a Wishart distribution with the scale matrix Υ and the degrees of freedom parameter ν. We use the following prior distributions:

$$\theta \sim \mathrm{N}(\eta, \Delta), \quad \boldsymbol{D}^{-1} \sim \mathrm{W}((\boldsymbol{D}^{*})^{-1}, \rho),$$
$$\boldsymbol{Q}^{-1} \sim \mathrm{W}(\boldsymbol{Q}^{*-1}, \rho_1), \quad \boldsymbol{R}^{-1} \sim \mathrm{W}(\boldsymbol{R}^{*-1}, \rho_2), \quad \boldsymbol{x}_{i0} \sim \mathrm{N}(\boldsymbol{\tau}, \boldsymbol{A}). \tag{14}$$

As one of the referees pointed out, for dynamic systems, vague and proper priors may lead to improper posteriors. In order to avoid improper posterior densities, Bates and Watts (1988) suggested to put a locally uniform prior on the expectation surface to represent a non-informative prior, instead of putting a locally uniform prior on the parameter space. Under our model setting, it is possible that vague proper priors on the parameter space can be translated into informative priors in the state space and can lead to improper posteriors.

The full conditional distributions can be derived as follows:

1. The conditional distributions of the state vectors are

$$p(\boldsymbol{x}_{it}, |\boldsymbol{y}, \boldsymbol{x}_{i,k\neq t}, \theta, \boldsymbol{D}, \boldsymbol{\Theta}, \boldsymbol{Q}, \boldsymbol{R}) \sim \mathrm{N}(\boldsymbol{H}_{it}^{-1}\lambda_{it}, \boldsymbol{H}_{it}^{-1}), \tag{15}$$

where for $t \neq 0$ or n_i

$$\boldsymbol{H}_{it} = \boldsymbol{Q}^{-1} + \boldsymbol{F}_i' \boldsymbol{Q}^{-1} \boldsymbol{F}_i + \boldsymbol{G}_i' \boldsymbol{R}^{-1} \boldsymbol{G}_i, \tag{16}$$

$$\lambda_{it} = \boldsymbol{Q}^{-1} \boldsymbol{F}_i \boldsymbol{x}_{i,t-1} + \boldsymbol{F}_i' \boldsymbol{Q}^{-1} \boldsymbol{x}_{i,t+1} + \boldsymbol{G}_i' \boldsymbol{R}^{-1} \boldsymbol{y}_{it}; \tag{17}$$

for $t = 0$, $\boldsymbol{H}_{i0} = \boldsymbol{F}_i' \boldsymbol{Q}^{-1} \boldsymbol{F}_i + \boldsymbol{A}^{-1}$, $\boldsymbol{\lambda}_{i0} = \boldsymbol{F}_i' \boldsymbol{Q}^{-1} \boldsymbol{x}_{i1} + \boldsymbol{A}^{-1} \boldsymbol{\tau}$, and for $t = n_i$, $\boldsymbol{H}_{in_i} = \boldsymbol{Q}^{-1} + \boldsymbol{G}_i' \boldsymbol{R}^{-1} \boldsymbol{G}_i$, $\boldsymbol{\lambda}_{in_i} = \boldsymbol{Q}^{-1} \boldsymbol{F}_i \boldsymbol{x}_{i,n_i-1} + \boldsymbol{G}_i' \boldsymbol{R}^{-1} \boldsymbol{y}_{in_i}$.

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

$$p(\theta_i | \boldsymbol{x}, \boldsymbol{y}, \theta, \boldsymbol{D}, \boldsymbol{\Theta}_{-i}, \boldsymbol{Q}, \boldsymbol{R}) \sim \mathrm{N}(\boldsymbol{J}_i^{-1}\boldsymbol{\zeta}_i, \boldsymbol{J}_i^{-1}), \tag{18}$$

$$p(\theta | \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{D}, \boldsymbol{\Theta}, \boldsymbol{Q}, \boldsymbol{R}) \sim \mathrm{N}\left\{(m\boldsymbol{D}^{-1} + \boldsymbol{\Delta}^{-1})^{-1} (\textstyle\sum_i^m \boldsymbol{D}^{-1}\theta_i + \boldsymbol{\Delta}^{-1}\eta), \ (m\boldsymbol{D}^{-1} + \boldsymbol{\Delta}^{-1})^{-1}\right\},$$
$$p(\boldsymbol{D}^{-1} | \boldsymbol{x}, \boldsymbol{y}, \theta, \boldsymbol{\Theta}, \boldsymbol{Q}, \boldsymbol{R}) \sim \mathrm{W}[\{\textstyle\sum_i^m (\theta_i - \theta)(\theta_i - \theta)' + \boldsymbol{D}^{*}\}^{-1}, m + \rho], \tag{19}$$

where $\boldsymbol{J}_i = \sum_{t=1}^{n_i} (\boldsymbol{T}_{i,t-1}' \boldsymbol{Q}^{-1} \boldsymbol{T}_{i,t-1} + \boldsymbol{L}_{it}' \boldsymbol{R}^{-1} \boldsymbol{L}_{it}) + \boldsymbol{D}^{-1}$,

$$\boldsymbol{\zeta}_i = \sum_{t=1}^{n_i} \left\{\boldsymbol{T}_{i,t-1}' \boldsymbol{Q}^{-1} (\boldsymbol{x}_{it} - \tilde{\boldsymbol{F}} \boldsymbol{x}_{i,t-1}) + \boldsymbol{L}_{it}' \boldsymbol{R}^{-1} (\boldsymbol{y}_{it} - \tilde{\boldsymbol{G}} \boldsymbol{x}_{it})\right\} + \boldsymbol{D}^{-1}\theta,$$
$$\boldsymbol{T}_{i,t-1} = (\boldsymbol{x}_{i,t-1}' \otimes \boldsymbol{I}_p)\boldsymbol{U}_1, \quad \text{and } \boldsymbol{L}_{it} = (\boldsymbol{X}_{it}' \otimes \boldsymbol{I}_q)\boldsymbol{U}_2.$$

'⊗' stands for kronecker product. $\boldsymbol{U}_1$ or $\boldsymbol{U}_2$ has dimension of $p^2 \times r$ or $q^2 \times r$ (r is length of $\boldsymbol{\theta}$) and is defined as follows. Remember that each element of $\boldsymbol{F}_i$ or $\boldsymbol{G}_i$ can be represented by a linear combination of all elements in $\boldsymbol{\theta}_i$. If we strip out $\boldsymbol{F}_i$ or $\boldsymbol{G}_i$ by column and stack them, the formalized vector with length p^2 or q^2 which we denote as vec($\boldsymbol{F}_i$) or vec($\boldsymbol{G}_i$) can be represented by the product of $\boldsymbol{U}_1$ or $\boldsymbol{U}_2$ with $\boldsymbol{\theta}_i$, i.e., vec($\boldsymbol{F}_i$) = $\boldsymbol{U}_1\boldsymbol{\theta}_i$ and vec($\boldsymbol{G}_i$) = $\boldsymbol{U}_2\boldsymbol{\theta}_i$.

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

$$p(\boldsymbol{Q}^{-1}|\boldsymbol{x},\boldsymbol{y},\theta,\boldsymbol{\Theta},\boldsymbol{Q},\boldsymbol{D})\sim \mathrm{W}[\{\boldsymbol{Q}^{*}+\frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{n_i}(\boldsymbol{x}_{it}-\boldsymbol{F}_i\boldsymbol{x}_{i,t-1})'(\boldsymbol{x}_{it}-\boldsymbol{F}_i\boldsymbol{x}_{i,t-1})\}^{-1}, s+\rho_1], \quad (20)$$

$$p(\boldsymbol{R}^{-1}|\boldsymbol{x},\boldsymbol{y},\theta,\boldsymbol{\Theta},\boldsymbol{R},\boldsymbol{D})\sim \mathrm{W}[\{\boldsymbol{R}^{*}+\frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{n_i}(\boldsymbol{y}_{it}-\boldsymbol{G}_i\boldsymbol{x}_{it})'(\boldsymbol{y}_{it}-\boldsymbol{G}_i\boldsymbol{x}_{it})\}^{-1}, s+\rho_2], \quad (21)$$

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 $\boldsymbol{y}_{it}$ is missing, the conditional distribution $p(\boldsymbol{x}_{it}|\boldsymbol{y}, \boldsymbol{x}_{i,k\neq t}, \cdots)$ in (15) depends only on $\boldsymbol{x}_{i,t-1}$, $\boldsymbol{x}_{i,t+1}$ and the parameters, but not on $\boldsymbol{y}_{it}$. Thus, one needs to drop the terms $\boldsymbol{G}_i'\boldsymbol{R}^{-1}\boldsymbol{G}_i$ in (16) and $\boldsymbol{G}_i'\boldsymbol{R}^{-1}\boldsymbol{y}_{it}$ in (17), which are related to the observation. With the large number of missing data, the Bayesian approach ran reasonably fast. However, a large number of missing data can slow down the EM algorithm. It is well-known that the EM algorithm has a linear convergence rate, which is determined by the information ratio of the missing data to the complete data. For sparse data, a large amount of information, including the random effects, the state vectors and many observations, is missing, and therefore the convergence of the EM algorithm can be quite slow.

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 $\boldsymbol{\Phi}$ be the collection of all the components involved in the full conditional distributions, i.e., $\boldsymbol{\Phi} = (\boldsymbol{x}_{10}, \cdots, \boldsymbol{x}_{1n_1}, \cdots, \boldsymbol{x}_{m0}, \cdots, \boldsymbol{x}_{mn_m}, \boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_m, \boldsymbol{\theta}, \boldsymbol{D}, \boldsymbol{Q}, \boldsymbol{R})$. Let $\boldsymbol{\Phi} - \{\boldsymbol{u}\}$ be the set of all the components except $\{\boldsymbol{u}\}$. Here $\{\boldsymbol{u}\}$ can be any component or a set of components in $\boldsymbol{\Phi}$. Starting from the initial value $\boldsymbol{\Phi}^{(0)}$, the Gibbs sampler draws a sample $\boldsymbol{x}_{10}^{(1)}$ from $p(\boldsymbol{x}_{10}|\{\boldsymbol{\Phi} - \{\boldsymbol{x}_{10}\}\}^{(0)}, \boldsymbol{y})$, then draws $\boldsymbol{x}_{11}^{(1)}$ from $p(\boldsymbol{x}_{11}|\boldsymbol{x}_{10}^{(1)}, \{\boldsymbol{\Phi}-\{\boldsymbol{x}_{10}, \boldsymbol{x}_{11}\}\}^{(0)}, \boldsymbol{y})$, and so on up to $\boldsymbol{R}^{(1)}$ from $p(\boldsymbol{R}|\{\boldsymbol{\Phi} - \{\boldsymbol{R}\}\}^{(1)}, \boldsymbol{y})$. Suppose that the Jth iteration produces a sample $\boldsymbol{\Phi}^{(J)}$. When J is large, $\boldsymbol{\Phi}^{(J)}$ is a sample from the posterior distribution $p(\boldsymbol{\Phi}|\boldsymbol{y})$ (Geman and Geman, 1984). The posterior distribution of the kth component of $\boldsymbol{\Phi}$ can be approximated by (Carlin, Polson and Stoffer, 1992)

$$p(\Phi_k|\boldsymbol{y}) \approx \frac{1}{J}\sum_{j=1}^{J} p\{\Phi_k|(\boldsymbol{\Phi}-\Phi_k)^{(j)}, \boldsymbol{y}\}. \quad (22)$$

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 that the observations are available up to time n_i, the Gibbs sampler constructed above can produce samples from the posterior distribution $p(\boldsymbol{x}_{it}|\boldsymbol{y})$ for $t \le n_i$. For $t < n_i$ this is a solution to the smoothing problem defined in Section 3.1. For $t = n_i$ the posterior distribution $p(\boldsymbol{x}_{in_i}|\boldsymbol{y})$ 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 $\{\boldsymbol{y}_{i,n_i+1}, \boldsymbol{y}_{i,n_i+2}, \cdots, \boldsymbol{y}_{it}\}$ as missing and derive the conditional distributions for the state vectors $\{\boldsymbol{x}_{i,n_i+1}, \boldsymbol{x}_{i,n_i+2}, \cdots, \boldsymbol{x}_{it}\}$, then include these state vectors in the Gibbs sampler. 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), which also provides a unified framework for the state estimation

$watermark-text $watermark-text $watermark-text

$watermark-text

$watermark-text

$watermark-text

and parameter estimation. However, the Gibbs sampler approach is computationally 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 often applied. We can extend the EM algorithm proposed by Watson and Engle (1983) to find the MLE of unknown parameters for the MESSM (3)–(5). The variance estimate can be obtained by inversing the observed Fisher information matrix, which can be used to get the interval estimates for the unknown parameters under a normal distribution assumption. To save space, we provide the details on implementation of EM algorithm in the Web Supplementary Appendix.

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 developed for the likelihood approach, thus the parameters ($\boldsymbol{\theta}$, $\boldsymbol{Q}$, $\boldsymbol{R}$, $\boldsymbol{D}$) are not treated as random and no prior distributions are needed. Computationally, in the Bayesian approach the parameters ($\boldsymbol{\theta}$, $\boldsymbol{Q}$, $\boldsymbol{R}$, $\boldsymbol{D}$) are constantly updated by sampling from their conditional distributions, while 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 extra computational costs is needed 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. However, an alternative approach, the stochastic approximation EM (SAEM) approach (Delyon, Lavielle and Moulines 1999) is very efficient and generally applicable to obtain the maximum likelihood estimates for the unknown parameters in nonlinear mixed-effects models from longitudinal data, which is worthy more exploring for potential online estimation.

5. Numerical Examples

To illustrate the usefulness of the proposed MESSM and estimation methods, we carried out two simulations studies: one is a simple univariate MESSM and another is a more complicated bivariate MESSM. To save space, we only report the simulation results from the bivariate model in this section and report the results from the univariate MESSM in the Web Supplementary Appendix. We also applied the proposed MESSM and estimation methods to an AIDS clinical data set in this section.

Simulation Study—Consider a MESSM model (3)–(4) with $p = 2$, $q = 1$, $r = 2$, $\boldsymbol{F}_i = \begin{pmatrix} 0.7 & 0 \\ \theta_{1i} & \theta_{2i} \end{pmatrix}$ and $G = (1 \quad 1)$. Let $\boldsymbol{D} = \sigma_D^2 \boldsymbol{I}$, $\boldsymbol{Q} = \sigma_Q^2 \boldsymbol{I}$. Set $\theta = (\theta_1 \quad \theta_2) = (0.3 \quad 0.6)$, $\sigma_D^2 = 0.0064$, $\sigma_Q^2 = 1.44$, and $R = 1$. The variances Q and R were selected to make the Filter Input Signal to Noise Ratio (FISNR) (Anderson and Moore, 1979) close to 3:2. We assumed the following priors for the Bayesian approach

$$\sigma_D^{-2} \sim \mathrm{G}(\beta_0, \beta_1),\ \ \sigma_Q^{-2} \sim \mathrm{G}(\nu_0, \nu_1),\ \ R^{-1} \sim \mathrm{G}(\omega_0, \omega_1),\ \ x_{i0} \sim \mathrm{N}(\tau, A). \quad (23)$$

Here $G(.,.)$ stands for the Gamma distribution. We specified the priors by setting $\beta_0 = 0.5$, $\beta_1 = 0.0001$, $\nu_0 = 0.5$, $\nu_1 = 0.5$, $\omega_0 = 0.5$, $\omega_1 = 0.5$, $\tau = 20$, and $A = 100$. Note that the true value of σ_D^2 is very small (close to zero). If a prior for σ_D^2 is selected to be skewed toward zero such as the one we used in our simulation, it is easier to identify this variance parameter as suggested by one of the referees. Otherwise, the estimate of σ_D^2 is poor. We assumed a normal prior distribution $N(\boldsymbol{\eta}, \boldsymbol{\Delta})$ for $\boldsymbol{\theta}$ with $\boldsymbol{\eta} = (0.5, 0.5)'$ and $\boldsymbol{\Delta} = \sigma_\Delta^2 \boldsymbol{I} = 2\boldsymbol{I}$. For this bivariate MESSM, $\tilde{\boldsymbol{F}} = \begin{pmatrix} 0.7 & 0 \\ 0 & 0 \end{pmatrix}$ contained the known parameters in the system matrix $\boldsymbol{F}_i$. 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 1. In this table $RE = \sqrt{MSE}/(True\ Parameter)$ represents the relative error. For both the Bayesian approach and the EM algorithm the estimates for θ_1 and θ_2 were close to the true values. The Bayesian estimates of σ_Q^2 and R were better than the EM estimates. However the EM algorithm provided better estimation for σ_D^2.

Figure 1 shows the histogram of the Gibbs samples for θ_1 and θ_2 and the estimated marginal posterior densities for both parameters. Considering θ_1, one notices that $p(\theta_1 | \{\boldsymbol{\Phi} - \theta_1\}, \boldsymbol{y}) = p(\theta_1 | \theta_{11}, \cdots, \theta_{1m}, \sigma_D^2)$ with $\boldsymbol{\Phi}$ being the collection of all the parameters and the state variables. Using the formula (22) the estimated posterior density can be derived as

$$\widehat{p}(\theta_1 | \boldsymbol{y}) \approx \frac{1}{J} \sum_{j=1}^{J} \mathrm{N}\left(\theta_1 | \frac{\sigma_\Delta^2}{m\sigma_\Delta^2 + \sigma_D^{2(j)}} \sum_{i=1}^{m} \theta_{1i}^{(j)} + \frac{\sigma_D^{2(j)}}{m\sigma_\Delta^2 + \sigma_D^{2(j)}} \eta_1, \frac{\sigma_\Delta^2 \sigma_D^{2(j)}}{m\sigma_\Delta^2 + \sigma_D^{2(j)}} \right)$$

where η_1 is the first element of $\boldsymbol{\eta}$, and $\mathrm{N}(\theta_1 | a, b)$ stands for the normal density with mean a and variance b. The histogram shows that the Gibbs samples for θ_1 are tightly centered around 0.3, and agrees very well with the estimated posterior density. The similar plots for θ_2 are also displayed in Figure 1. It can be seen that the Gibbs samples for θ_2 are concentrated around 0.6, the true parameter value of θ_2. To monitor the convergence of the Gibbs samples the trace plots for both parameters are also shown in Figure 1. The trace plots do not display any signs of patterns, suggesting the convergence to the stationary distribution. Similarly the Gibbs samples for the other parameters D, Q and R (not shown) also suggest the convergence of the Gibbs samplers.

HIV Dynamic Application—We applied the MESSM to a data set from the AIDS Clinical Trial Group (ACTG) Study 315 (Wu and Ding, 1999). In this trial, 53 HIV infected patients were treated with potent antiviral drugs (ritonavir, 3TC, and AZT). Among those 5 dropped out of the study. 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. Perelson *et al.* (1997) suggested a two-phase clearance of viral load. The first phase featured a rapid drop in viral load due to the exponential death rate of the productively infected cells during the first four weeks of treatment. The decline of viral load was slower in the second phase due to the slow clearance rate of the long lived infected T-cells or latently infected T-cells. Here we considered a time-varying MESSM for the ACTG 315 data assuming the rate of clearance changed at the mid-point of week 4 (i.e.

$watermark-text $watermark-text $watermark-text

$watermark-text

$watermark-text

$watermark-text

day 24). This model is closely related to the simple HIV dynamic model with a constant rate (Ho *et al.* 1995). The MESSM is written as

$$x_{it}=\theta_{it}x_{i,t-1}+v_{it}, \text{ and } y_{it}=x_{it}+w_{it},$$

where 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. θ_{it} indicates the viral clearance rate for patient i at day t ($\theta_{it}=\theta_i^{(1)}$ for $t \le 23$; $\theta_{it}=\theta_i^{(2)}$ for $t>23$). Here $\theta_i^{(k)} \sim N(\theta^{(k)}, D_k)$ for $k = 1, 2$, and $\theta_i^{(k)}$ are independent. Note that we may use the log-transformation to guarantee θ_{it} to be positive. For other complicated constraints of parameters, there is no general method to deal with and we may need to deal with this problem case-by-case. Although this is an simplified HIV dynamic model, it serves to illustrate the proposed methodologies.

The similar prior distributions as in the simulation studies were utilized for the Bayesian approach (see the Web Supplementary Appendix). Table 2 gives the estimation values of both the population parameters and the individual parameters $\theta_i^{(k)}$ following the Bayesian approach. The EM estimation was similar and omitted.

Figure 2 shows the one-step-ahead Kalman forecasts (dotted curves) of the individual states for the first four patients using the estimated individual parameters. The estimated posterior means for the individual states using the Gibbs sampler approach are also shown in solid curves. As expected the posterior mean curves are smooth and close to the observed values. A Kalman forecast jumps one day after the time point of an observation, indicating an adjustment based on the observation. The model checking results are shown in Figure S1 in the Web Supplementary Appendix. The fitted viral loads from both methods match the data quite well. The normal assumption on measurement errors is validated by QQ plots. The density estimates of individual parameters show that the normality assumption on random effects is satisfied for both methods.

6. Discussion

In many biomedical longitudinal studies, the underlying biological mechanisms are well studied and mathematical representations, usually a set of differential or difference equations, are available. 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 studies. 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.

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 decades, 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

$watermark-text

$watermark-text

$watermark-text

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), 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 that need to be overcome in practice.

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, such as hepatitis virus dynamics (Nowak *et al.* 1996), tumor dynamics in cancer research (Swan, 1984; Martin and Teo, 1994), and genetic regulatory network modeling (Chen, He and Church, 1999; Holter *et al.* 2001). 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was partially supported by NIAID/NIH research grants AI078842, AI078498, AI50020 and AI087135. The authors appreciate the two referees for their constructive and insightful comments and suggestions.

References

Anderson, BDO.; Moore, JB. Optimal Filtering. Prentice-Hall; 1979.

Bates, DM.; Watts, DB. Nonlinear Regression Analysis and Its Applications. New York: Wiley; 1988.

Brockwell, PJ.; Davis, RA. Time Series: Theory and Methods. Springer-Verlag; 1991.

Carlin BP, Polson NG, Stoffer DS. A Monte Carlo Approach to Nonnormal and Nonlinear State-space Modeling. Journal of American Statistical Association. 1992; 87:493–500.

Carter CK, Kohn R. On Gibbs Sampling for State Space Models. Biometrika. 1994; 81:541–553.

Carter CK, Kohn R. Markov Chain Monte Carlo in Conditionally Gaussian State Space Models. Biometrika. 1996; 83:589–601.

Chen R, Liu J. Mixture Kalman Filters. Journal of the Royal Statistical Society, Series B. 2000; 62:493–508.

Chen T, He HL, Church GM. Modeling Gene Expression with Differential Equations. Pacific Symposium of Biocomputing. 1999; 1999:29–40.

Davidian, M.; Giltinan, DM. Nonlinear Models for Repeated Measurements Data. Chapman & Hall; 1995.

De Jong P, Shephard N. The Simulation Smoother for Time Series Models. Biometrika. 1995; 82:339–350.

Delyon B, Lavielle M, Moulines E. Convergence of a stochastic approximation version of the EM algorithm. Annals of Statistics. 1999; 27(1):94–128.

Diggle, P.; Liang, KY.; Zeger, S. Analysis of Longitudinal Data. Oxford Science; 1994.

Engle RF, Watson MW. A one factor multivariate time series model of metropolitan wage rates. Journal of American Statistical Association. 1981; 76:774–781.

Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell. 1984; (6):721–741. [PubMed: 22499653]

Gordon NJ, Salmond DJ, Smith AFM. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEEE Proceedings-F. 1993; 140:107–113.

$watermark-text

$watermark-text

$watermark-text

Harvey, AC. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press; 1989.

Holter NS, Maritan A, Cieplak M, et al. Dynamic Modeling of Gene Expression Data. PNAS. 2001; 98:1693–1698. [PubMed: 11172013]

Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature. 1995; 373:123–126. [PubMed: 7816094]

Huang Y, Liu D, Wu H. Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. Biometrics. 2006; 62:413–423. [PubMed: 16918905]

Jensen JL, Petersen NV. Asymptotic normality of the maximum likelihood estimator in state space models. The Annals of Statistics. 1999; 27:514–535.

Jones, RH. Longitudinal Data with Serial Correlation: A State-space Approach. New York: Chapman & Hall; 1993.

Kalman RE. A new approach to linear filtering and prediction theory. Journal of Basic Engineering, Transactions ASME, Series D. 1960; 82:35–45.

Kitagawa G. Non-Gaussian state space modelling of nonstationary time series (with discussion). Journal of American Statistical Association. 1987; 82:1032–1063.

Kitagawa G. Monte Carlo filter and smoother for nonlinear non-Gaussian state space models. Journal of Computational and Graphical Statistics. 1996; 5:1–25.

Laird NM, Ware JH. Random effects models for longitudinal data. Biometrics. 1982; 38:963–974. [PubMed: 7168798]

Liu JS, Chen R. Sequential Monte Carlo methods for dynamic systems. Journal of the American Statistical Association. 1998; 93:1032–1044.

Martin, R.; Teo, KL. Optimal Control of Drug Administration in Cancer Chemotherapy. World Scientific; Singapore: 1994.

Nowak MA, Bonhoeffer S, Hill AM, Boehme R, Thomas H, McDade H. Viral dynamics in hepatitis B virus infection. Proc Natl Acad Sci USA. 1996; 93:4398–4402. [PubMed: 8633078]

Pagan A. Some Identification and Estimation Results for Regression Models with Stochastically Varying Parameters. Journal of Econometrics. 1980; 13:341–363.

Perelson AS, Essunger P, Cao Y. Decay Characteristics of HIV-1 Infected Compartments during Combination Therapy. Nature. 1997; 387:188–191. [PubMed: 9144290]

Rahiala M. Random Coefficient Autoregressive Models for Longitudinal Data. Biometrika. 1999; 86:718–722.

Schweppe F. Evaluation of likelihood functions for Gaussian signals. IEEE Transactions on Information Theory. 1965; 11:61–70.

Steimer JL, Mallet A, Golmard J, Boisvieux J. Alternative Approaches to Estimation of Population Pharmacokinetic Parameters: Comparison with the Nonlinear Mixed Effects Model. Drug Metabolism Reviews. 1984; 15:265–292. [PubMed: 6745083]

Swan, GW. Applications of Optimal Control Theory in Biomedicine. Marcel Dekker, Inc; New York: 1984.

Vonesh, EF.; Chinchilli, VM. Linear and Nonlinear Models for the Analysis of Repeated Measurements. Marcel Dekker; New York: 1996.

Watson MW, Engle RF. Alternative Algorithms for the Estimation of Dynamic Factor, Mimic and Varying Coefficient Regression Models. Journal of Econometrics. 1983; 23:385–400.

West, M.; Harrison, J. Bayesian Forecasting and Dynamic models. New York: Springer; 1997.

Wu H, Ding A. Population HIV-1 Dynamics in Vivo: Applicable Models and Inferential Tools for Virological Data from AIDS Clinical Trials. Biometrics. 1999; 55:410–418. The data are available online at http://stat.tamu.edu/biometrics/datasets/980622.txt. [PubMed: 11318194]

Wu H, Ding A, DeGruttola V. Estimation of HIV Dynamic Parameters. Statistics in Medicine. 1998; 17:2463–2485. [PubMed: 9819839]

$watermark-text
$watermark-text
$watermark-text

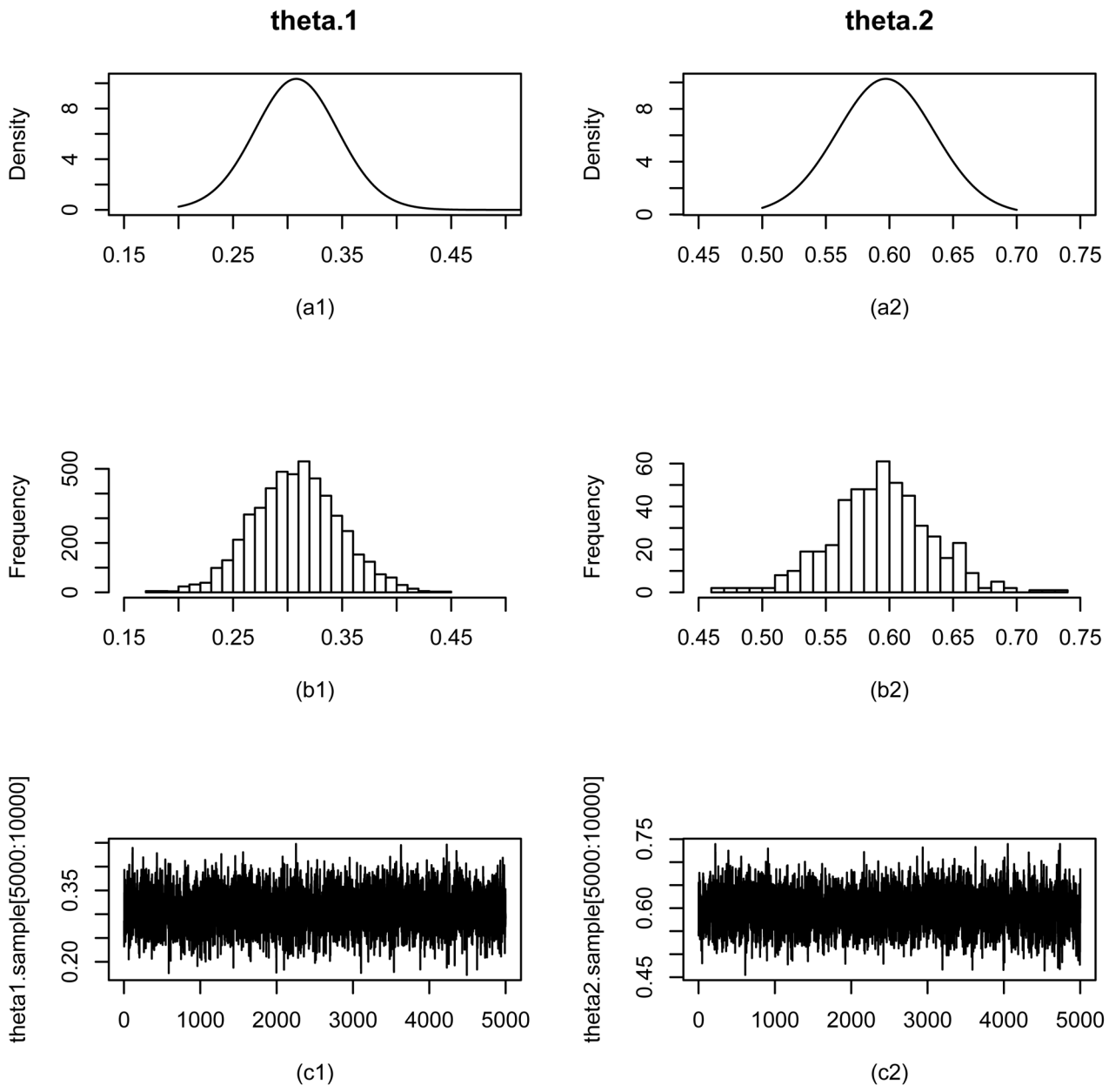

Figure 1.
Parameter estimation results for θ_1 and θ_2 from the Bayes method for the simulation study (the bivariate MESSM model). (a1) and (a2) are estimated posterior densities, (b1) and (b2) are histograms of the Gibbs sampler, and (c1) and (c2) are trace plots for the last 5000 samples for θ_1 and θ_2, respectively.

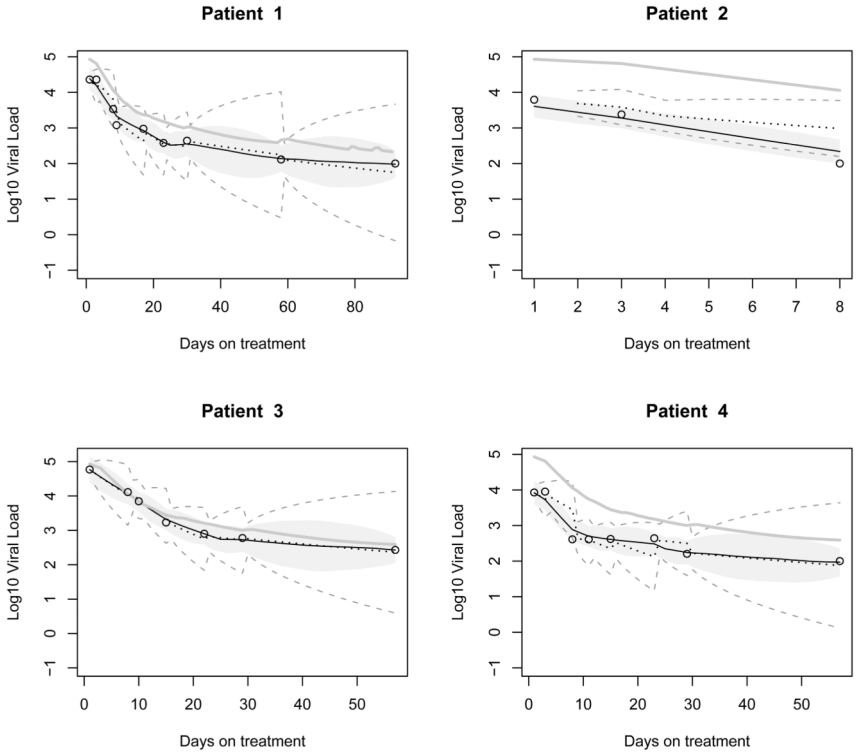

Figure 2.
Estimation and prediction of viral load for four patients in the HIV dynamic study (four subjects with different number of measurements and different response patterns are selected for illustration). Base 10 logarithm of the viral loads are shown in circles. The smoothing estimates using the Gibbs sampler approach are shown in black solid curves, and the 95% error bounds are shown by the shaded grey region. The one-step-ahead Kalman forecasts are shown in black dotted curves, and the prediction error bounds are in grey dashed curves. The estimates of population state variable are shown as the gray solid curve.

$watermark-text
$watermark-text
$watermark-text

$watermark-text $watermark-text $watermark-text

Table 1

Parameter estimation for the bivariate MESSM model with 50 replicates, $m = 20$ and $n = 20$.

	$\hat{\theta}_1$		$\hat{\theta}_2$		$\widehat{\sigma}_D^2$		$\widehat{\sigma}_Q^2$		$\hat{R}$	
	B	**EM**	**B**	**EM**	**B**	**EM**	**B**	**EM**	**B**	**EM**
Mean	0.305	0.293	0.594	0.606	0.009	0.006	1.357	1.122	1.04	1.164
Bias	0.005	−0.007	−0.006	0.006	0.0026	−0.0004	−0.083	−0.318	0.04	0.164
MSE	0.0014	0.0012	0.0018	0.0014	7×10^{-6}	3×10^{-6}	0.142	0.345	0.09	0.215
RE	0.125	0.115	0.071	0.062	0.413	0.271	0.262	0.401	0.300	0.463

Biometrics. Author manuscript; available in PMC 2012 November 27.

$watermark-text $watermark-text $watermark-text

Table 2

Parameter estimation and forecasts of viral load for the HIV dynamic model. Confidence interval for population estimates are included in brackets.

Estimation of population parameters							
$\hat{\theta}_1$	$\hat{\theta}_2$		$\hat{D}_1$	$\hat{D}_2$		$\hat{Q}$	$\hat{R}$
$0.976_{(0.974,0.979)}$	$0.994_{(0.991,0.997)}$		$0.0052_{(0.004,0.0075)}$	$0.0046_{(0.003,0.0065)}$		$0.119_{(0.104,0.134)}$	$0.204_{(0.174,0.238)}$
Estimation of the individual parameters for 48 patients							
$\theta_i^{(1)}$							
0.975753	0.972227	0.976278	0.975794	0.97177	0.976373	0.98232	0.976072
0.975554	0.979314	0.976282	0.978256	0.973686	0.98045	0.976808	0.976119
0.972679	0.976762	0.97706	0.974961	0.975959	0.974782	0.97605	0.977631
0.976737	0.977898	0.979602	0.978431	0.974501	0.975863	0.977751	0.976239
0.977426	0.977808	0.978213	0.976879	0.973311	0.979874	0.97188	0.978342
0.974063	0.972705	0.974047	0.975449	0.975593	0.976236	0.979845	0.9746
$\theta_i^{(2)}$							
0.994389	-	0.994317	0.993837	0.993597	0.996628	0.993837	-
-	0.994033	0.993657	0.994675	0.992734	0.994795	0.994546	0.994494
0.994527	0.994921	0.994156	0.993475	0.9951	0.994282	-	0.992501
0.994217	0.994472	0.991974	-	-	0.994325	0.993253	0.99328
-	0.993176	0.993242	0.993993	0.993556	0.994226	0.993714	0.9946
0.994275	0.993367	0.99472	0.994184	0.993569	0.996103	0.993211	0.994178