Web Supplementary Materials

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

Dacheng Liu, Tao Lu, Xu-Feng Niu, and Hulin Wu

S.1 Parameter estimation in MESSM by EM algorithm

If the state variables $\boldsymbol{x}_{it}$ and the random effects $\boldsymbol{b}_i$ and $\boldsymbol{r}_i$ 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 $\boldsymbol{D}$, $\boldsymbol{Q}$ and $\boldsymbol{R}$ would be easily derived. These estimators are constructed as follows.

Notice that $\boldsymbol{F}_i$ can be decomposed as $\boldsymbol{F}_i = \tilde{\boldsymbol{F}} + \boldsymbol{F}_i^{(f)} + \boldsymbol{F}_i^{(r)}$ where $\tilde{\boldsymbol{F}}$ is a matrix containing known entries in $\boldsymbol{F}_i$; $\boldsymbol{F}_i^{(f)}$ and $\boldsymbol{F}_i^{(r)}$ is the fixed and random effect component of $\boldsymbol{F}_i$, respectively. $\boldsymbol{G}_i$ is handled the same way. We first rewrite the MESSM (3)-(4) as

$$\boldsymbol{x}_{it} - (\tilde{\boldsymbol{F}} + \boldsymbol{F}_i{}^{(r)})\boldsymbol{x}_{i,t-1} = (\boldsymbol{x}'_{i,t-1} \otimes \boldsymbol{I}_p)\boldsymbol{U_1}\boldsymbol{\theta} + \boldsymbol{v}_{it}, \tag{S.1}$$

$$\boldsymbol{y}_{it} - (\tilde{\boldsymbol{G}} + \boldsymbol{G}_i{}^{(r)})\boldsymbol{x}_{it} = (\boldsymbol{x}'_{it} \otimes \boldsymbol{I}_q)\boldsymbol{U_2}\boldsymbol{\theta} + \boldsymbol{w}_{it}, \tag{S.2}$$

where $\boldsymbol{U_1}$ and $\boldsymbol{U_2}$ are defined before. Let

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

$$
\boldsymbol{M}_{\boldsymbol{i}} = \left(\boldsymbol{M}_i{}^{(1)'} \middle| \boldsymbol{M}_i{}^{(2)'}\right)' = \left(\boldsymbol{U_1}'(\boldsymbol{x}_{i0} \otimes \boldsymbol{I}_{n_i}), \cdots, \boldsymbol{U_1}'(\boldsymbol{x}_{i,n_i-1} \otimes \boldsymbol{I}_{n_i}) \middle| \boldsymbol{U_2}'(\boldsymbol{x}_{i1} \otimes \boldsymbol{I}_p), \cdots, \boldsymbol{U_2}'(\boldsymbol{x}_{in_i} \otimes \boldsymbol{I}_p)\right)'.
$$

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

$$
\hat{\boldsymbol{\theta}} = \left[\sum_{i=1}^{m} \boldsymbol{M}_{\boldsymbol{i}}{}' \hat{\boldsymbol{\Pi}}_i^{-1} \boldsymbol{M}_{\boldsymbol{i}}\right]^{-1} \left[\sum_{i=1}^{m} \boldsymbol{M}_{\boldsymbol{i}}{}' \hat{\boldsymbol{\Pi}}_i^{-1} \tilde{\boldsymbol{y}}_{\boldsymbol{i}}\right], \tag{S.3}
$$

$$
\hat{\boldsymbol{Q}} = \frac{1}{s} \sum_{i=1}^{m} \sum_{t=1}^{n_i} \hat{\boldsymbol{v}}_{it} \hat{\boldsymbol{v}}_{it}', \tag{S.4}
$$

$$
\hat{\boldsymbol{R}} = \frac{1}{s} \sum_{i=1}^{m} \sum_{t=1}^{n_i} \hat{\boldsymbol{w}}_{it} \hat{\boldsymbol{w}}_{it}', \tag{S.5}
$$

$$
\hat{\boldsymbol{D}} = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{b}_i \boldsymbol{b}_i', \tag{S.6}
$$

where $\hat{\boldsymbol{v}}_{it} = \boldsymbol{x}_{it}^{*} - (\boldsymbol{x}_{i,t-1}{}' \otimes \boldsymbol{I}_p)\boldsymbol{U_1}\hat{\boldsymbol{\theta}}$, $\quad \hat{\boldsymbol{w}}_{it} = \boldsymbol{y}_{it}^{*} - (\boldsymbol{x}_{it}{}' \otimes \boldsymbol{I}_q)\boldsymbol{U_2}\hat{\boldsymbol{\theta}}$. By iterating (S.3)-(S.5) one could find the estimates of $\boldsymbol{\theta}$, $\boldsymbol{Q}$ and $\boldsymbol{R}$ that satisfy these equations. Equations (S.3)- (S.6) define the M-step in the EM algorithm. To establish the E-step we need to 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 $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{b})$ can be written as

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

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

$$
\boldsymbol{U}_1'(\boldsymbol{x}_{\boldsymbol{i}}^{-}{}'\boldsymbol{x}_{\boldsymbol{i}}^{-} \otimes \hat{\boldsymbol{Q}}^{-1})\boldsymbol{U}_1 + \boldsymbol{U}_2'(\boldsymbol{x_i}'\boldsymbol{x_i} \otimes \hat{\boldsymbol{R}}^{-1})\boldsymbol{U}_2 \tag{S.7}
$$

with $\boldsymbol{x_i}' = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{in_i})$, and $\boldsymbol{x}_{\boldsymbol{i}}^{-}{}' = (\boldsymbol{x}_{i0}, \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{i,n_i-1})$. Similarly, $\boldsymbol{M_i}'\hat{\boldsymbol{\Pi}}_i^{-1}\tilde{\boldsymbol{y}}_{\boldsymbol{i}}$ can be written as

$$
\boldsymbol{U}'{}_1(\boldsymbol{I}_p \otimes \hat{\boldsymbol{Q}}^{-1})\text{vec}(\boldsymbol{x_i}^{*\prime}\boldsymbol{x}_{\boldsymbol{i}}^{-}) + \boldsymbol{U}_2'(\boldsymbol{I}_q \otimes \hat{\boldsymbol{R}}^{-1})\text{vec}(\boldsymbol{y_i}^{*\prime}\boldsymbol{x_i}) \tag{S.8}
$$

with $\boldsymbol{x_i}^{*\prime} = (\boldsymbol{x}_{i1}^*, \boldsymbol{x}_{i2}^*, \ldots, \boldsymbol{x}_{in_i}^*)$ and $\boldsymbol{y_i}^{*\prime} = (\boldsymbol{y}_{i1}^*, \boldsymbol{y}_{i2}^*, \ldots, \boldsymbol{y}_{in_i}^*)$. Plugging (S.7) and (S.8) into the log-likelihood function, it can be seen that we need to evaluate the conditional expectations for the following matrices:

$$
\begin{aligned}
&\boldsymbol{x_i}', \;\; \boldsymbol{x}_{\boldsymbol{i}}^{-}{}', \;\; \boldsymbol{x_i}'\boldsymbol{x_i}, \;\; \boldsymbol{x}_{\boldsymbol{i}}^{-}{}'\boldsymbol{x}_{\boldsymbol{i}}^{-}, \boldsymbol{x_i}'\boldsymbol{x}_{\boldsymbol{i}}^{-}, \\
&\boldsymbol{F}_i^{(r)}\boldsymbol{x}_{\boldsymbol{i}}^{-}{}', \;\; \boldsymbol{F}_i^{(r)}\boldsymbol{x}_{\boldsymbol{i}}^{-}{}'\boldsymbol{x}_{\boldsymbol{i}}^{-}, \;\; \boldsymbol{F}_i^{(r)}\boldsymbol{x}_{\boldsymbol{i}}^{-}{}'\boldsymbol{x_i}, \;\; \boldsymbol{F}_i^{(r)}\boldsymbol{x}_{\boldsymbol{i}}^{-}{}'\boldsymbol{x}_{\boldsymbol{i}}^{-}\boldsymbol{F}_i^{(r)\prime}, \\
&\boldsymbol{G}_i^{(r)}\boldsymbol{x_i}', \;\; \boldsymbol{G}_i^{(r)}\boldsymbol{x_i}'\boldsymbol{x_i}, \;\; \boldsymbol{G}_i^{(r)}\boldsymbol{x_i}'\boldsymbol{x_i}\boldsymbol{G}_i^{(r)\prime}, \text{and } \boldsymbol{b}_i\boldsymbol{b}_i'.
\end{aligned} \tag{S.9}
$$

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. 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 $\boldsymbol{\Lambda} = (\boldsymbol{\theta}, \boldsymbol{Q}, \boldsymbol{R}, \boldsymbol{D})$ be the collection of the parameters. The Gibbs sampler discussed in Section 4.1. can be modified to draw samples from the distributions of $\boldsymbol{x}_{it}$ and $\boldsymbol{b}_i$ conditional on the observations $\boldsymbol{y}$ and the current estimate $\hat{\boldsymbol{\Lambda}}^{(k)}$. Specifically, one can derive the full conditional distributions $p(\boldsymbol{x}_{it}|\boldsymbol{b}_i, \boldsymbol{y}, \hat{\boldsymbol{\Lambda}}^{(k)})$ and $p(\boldsymbol{b}_i|\boldsymbol{x}_{it}, \boldsymbol{y}, \hat{\boldsymbol{\Lambda}}^{(k)})$, and draw Gibbs samples $\boldsymbol{x}_{it}^{(j)}$ and $\boldsymbol{b}_i^{(j)}$ $(j = 1, \ldots, J)$. Then the conditional expectations can be evaluated based on these samples. Taking the matrix $\boldsymbol{b}_i\boldsymbol{b}_i{}'$ in (S.9) as an example, $E(\boldsymbol{b}_i\boldsymbol{b}_i{}'|\boldsymbol{y}, \hat{\boldsymbol{\Lambda}}^{(k)}) \approx \frac{1}{J}\sum_{j=1}^{J} \boldsymbol{b}_i^{(j)}\boldsymbol{b}_i'^{(j)}$. Note that at the final iteration of the EM algorithm the Gibbs samples for $\boldsymbol{b}_i$ generated in the E-step can be used to estimate the random effects $\boldsymbol{b}_i$.

The EM algorithm for the mixed-effects state space models is now fully defined. In the E-step the samples for $\boldsymbol{x}_{it}$ and $\boldsymbol{b}_i$ are drawn from the Gibbs sampler, and the conditional expectations of the matrices in (S.9) are computed. In the M-step the estimates of $\boldsymbol{\theta}$, $\boldsymbol{Q}$ and $\boldsymbol{R}$ are updated by solving Equations (S.3)-(S.5) and the estimate of $\boldsymbol{D}$ is directly updated from the E-step outputs, i.e. $\hat{\boldsymbol{D}}^{(k+1)} = E(\frac{1}{m}\sum_{i=1}^{m} \boldsymbol{b}_i\boldsymbol{b}_i'|\boldsymbol{y}, \hat{\boldsymbol{\Lambda}}^{(k)})$. As suggested in Watson and Engle (1983) it is not necessary to iterate between (S.3)-(S.5) during each M-step. Instead, one can use $\hat{\boldsymbol{Q}}^{(k)}$ and $\hat{\boldsymbol{R}}^{(k)}$ to construct $\hat{\boldsymbol{\theta}}^{(k+1)}$ in (S.3), and then use $\hat{\boldsymbol{\theta}}^{(k+1)}$ to create $\hat{\boldsymbol{Q}}^{(k+1)}$ and $\hat{\boldsymbol{R}}^{(k+1)}$. Upon convergence of the EM algorithm, the final parameter estimates will satisfy (S.3)-(S.6) by construction.

S.2 Simulation Study: Univariate Model

We carried out simulation studies using the two estimation methods discussed in Section 4 for the following univariate MESSM

$$x_{it} = \theta_i x_{i,t-1} + v_{it}, \quad v_{it} \sim \text{i.i.d. N}(0, Q), \tag{S.10}$$

$$y_{it} = x_{it} + w_{it}, \quad w_{it} \sim \text{i.i.d. N}(0, R), \tag{S.11}$$

$$\theta_i = \theta + b_i, \quad b_i \sim \text{N}(0, D), \quad i = 1, \ldots, m. \tag{S.12}$$

Note that model (S.10)-(S.12) 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 $\theta = 0.8057$, $D = 0.04$, $Q = 1.44$ and $R = 1$ for the simulation studies. θ was chosen based on the results of Wu and Ding (1999) for the late stage of HIV dynamics. D was close to the between-subject variation in the real data example blow. 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. Data y_{it} were generated from the model (S.10)-(S.12) with $m = 20$ or 60, and $n = 10$ or 30 ($n_1 = n_2 = \ldots = n_m = n$) . We assumed the following priors for the Bayesian approach

$$\theta \sim \text{N}(\eta, \Delta), \; D^{-1} \sim \text{G}(\beta_0, \beta_1), \; Q^{-1} \sim \text{G}(\nu_0, \nu_1), \; R^{-1} \sim \text{G}(\omega_0, \omega_1), \quad x_{i0} \sim \text{N}(\tau, A) \tag{S.13}$$

Here $G(.,.)$ stands for the Gamma distribution. The full conditional distributions can be derived as

$$p(\theta_i|\boldsymbol{x},\boldsymbol{y},\theta,D,\boldsymbol{\Theta}_{-i},Q,R)\sim \mathrm{N}\left(\frac{D\sum_{t=1}^{n_i}x_{i,t-1}x_{it}+\theta Q}{D\sum_{t=1}^{n_i}X_{i,t-1}^2+Q},\frac{QR}{D\sum_{t=1}^{n_i}X_{i,t-1}^2+Q}\right),$$

$$p(\theta|\boldsymbol{x},\boldsymbol{y},D,\boldsymbol{\Theta},Q,R)\sim \mathrm{N}\left(\frac{\Delta\sum_{i=1}^{m}\theta_i+\eta D}{D+m\Delta},\frac{D\Delta}{D+m\Delta}\right),$$

$$p(x_{it}|\boldsymbol{y},\boldsymbol{x}_{i,k\neq t},\theta,D,\boldsymbol{\Theta},Q,R)\sim \mathrm{N}\left(\frac{\theta_i R(x_{i,t-1}+x_{i,t+1})+Qy_{it}}{(1+\theta_i^2)R+Q},\frac{QR}{(1+\theta_i^2)R+Q}\right),$$

$$p(x_{i0}|\boldsymbol{y},\boldsymbol{x}_{i,k\neq 0},\theta,D,\boldsymbol{\Theta},Q,R)\sim \mathrm{N}\left(\frac{Q\tau+A\theta_i x_{i1}}{Q+\theta_i^2 A},\frac{QA}{Q+\theta_i^2 A}\right),$$

$$p(x_{in_i}|\boldsymbol{y},\boldsymbol{x}_{i,k\neq n_i},\theta,D,\boldsymbol{\Theta},Q,R)\sim \mathrm{N}\left(\frac{\theta_i R x_{i,n_i-1}+Qy_{in_i}}{R+Q},\frac{QR}{R+Q}\right),$$

$$p(D^{-1}|\boldsymbol{x},\boldsymbol{y},\theta,\boldsymbol{\Theta},Q,R)\sim \mathrm{G}\left(\beta_0+\frac{m}{2},\beta_1+\frac{\sum_{i=1}^{m}(\theta_i-\theta)^2}{2}\right),$$

$$p(Q^{-1}|\boldsymbol{x},\boldsymbol{y},\theta,\boldsymbol{\Theta},R,D)\sim \mathrm{G}\left(\nu_0+\frac{s}{2},\frac{\nu_1+\sum_{i=1}^{m}\sum_{t=1}^{n_i}(x_{it}-\theta_i x_{i,t-1})^2}{2}\right),$$

$$p(R^{-1}|\boldsymbol{x},\boldsymbol{y},\theta,\boldsymbol{\Theta},Q,D)\sim \mathrm{G}\left(\omega_0+\frac{s}{2},\frac{\omega_1+\sum_{i=1}^{m}\sum_{t=0}^{n_i}(y_{it}-x_{it})^2}{2}\right).$$

We specified the priors by setting $\eta = 0.5$, $\Delta = 4$, $\beta_0 = 0.5$, $\beta_1 = 0.0001$, $\nu_0 = 0.5$, $\nu_1 = 0.5$, $\omega_0 = 0.5$, $\omega_1 = 0.5$, $\tau = 20$, and $A = 100$. Under this set-up the prior distributions for D, Q and R were improper priors.

The simulation studies were performed on a Dell Dimension 4600 with 3.0G Hz CPU and 1GB memory. All programs were written in C++. 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 S1. 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			
	n	10		30		10		30	
		B	EM	B	EM	B	EM	B	EM
$\hat{\theta}$	Mean	0.8023	0.7908	0.8051	0.7893	0.8067	0.7923	0.8044	0.79
	Bias	-0.0034	-0.0149	-0.0006	-0.0164	0.001	-0.0134	-0.0013	-0.0157
	MSE	0.0016	0.0004	0.0014	0.0003	0.0004	0.0002	0.0005	0.0002
	RE	0.0496	0.0236	0.0464	0.0217	0.0248	0.0174	0.0278	0.0195
$\hat{D}$	Mean	0.0512	0.0387	0.045	0.0376	0.0476	0.0354	0.0443	0.0345
	Bias	0.0112	-0.0013	0.005	-0.0024	0.0076	-0.0046	0.0043	-0.0055
	MSE	0.0009	0.0002	0.0007	0.0001	0.0004	0.00005	0.0001	0.00006
	RE	0.75	0.3381	0.6614	0.2817	0.5	0.1739	0.25	0.1937
$\hat{Q}$	Mean	1.4456	0.0801	1.4819	1.1278	1.4513	0.019	1.4553	1.1245
	Bias	0.0056	-1.3599	0.0419	-0.3122	0.0113	-1.421	0.0153	-0.3155
	MSE	0.2172	1.8913	0.0592	0.1321	0.0758	2.0197	0.0376	0.1139
	RE	0.3236	0.955	0.1689	0.2524	0.1912	0.9869	0.1347	0.2343
$\hat{R}$	Mean	1.0474	2.273	0.9871	1.1425	1.0148	2.3428	0.9914	1.1845
	Bias	0.0474	1.273	-0.0129	0.1425	0.0148	1.3428	-0.0086	0.1845
	MSE	0.1832	1.733	0.0583	0.0478	0.0695	1.8329	0.0185	0.0474
	RE	0.428	1.3164	0.2415	0.2186	0.2636	1.3538	0.136	0.2177

Table S1 shows the results of parameter estimation by the Bayesian approach and the EM algorithm. In this table $RE = \sqrt{MSE}/(True\ Parameter)$ represents the relative error. The Bayesian estimates of θ, D, Q and R were very good in all cases. For a fixed number of subjects, increasing the number of observations from 10 to 30 improves the estimation of D in the Bayesian approach: for $m = 20$, the relative errors of $\hat{D}$ for $n = 10$ and $n = 30$ were relatively close at 0.75 and 0.66 respectively; for $m = 60$ the relative errors were significantly reduced by larger n (0.5 and 0.25 for $n = 10$ and $n = 30$ respectively). Furthermore, the relative error of $\hat{D}$ improved noticeably as the number of subjects increased from 20 to 60. On the other hand, the EM algorithm gave good estimates for θ 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 example, for $m = 60$ the relative errors for $\hat{R}$ were 1.35 and 0.218 for $n = 10$

and $n = 30$, respectively.

References

Anderson, B.D.O and Moore, J.B (1979). *Optimal Filtering*, Prentice-Hall.

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.

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.

McLachlan, G.J. and Krishnan, T. (1997). *The EM Algorithm and Extensions* Wiley, New York.

Staudenmayer, J. and Buonaccorsi, J. (2005). Measurement Error in Linear Autoregressive Models. *Journal of the American Statistical Association* **471**, 841–852.

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.

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.

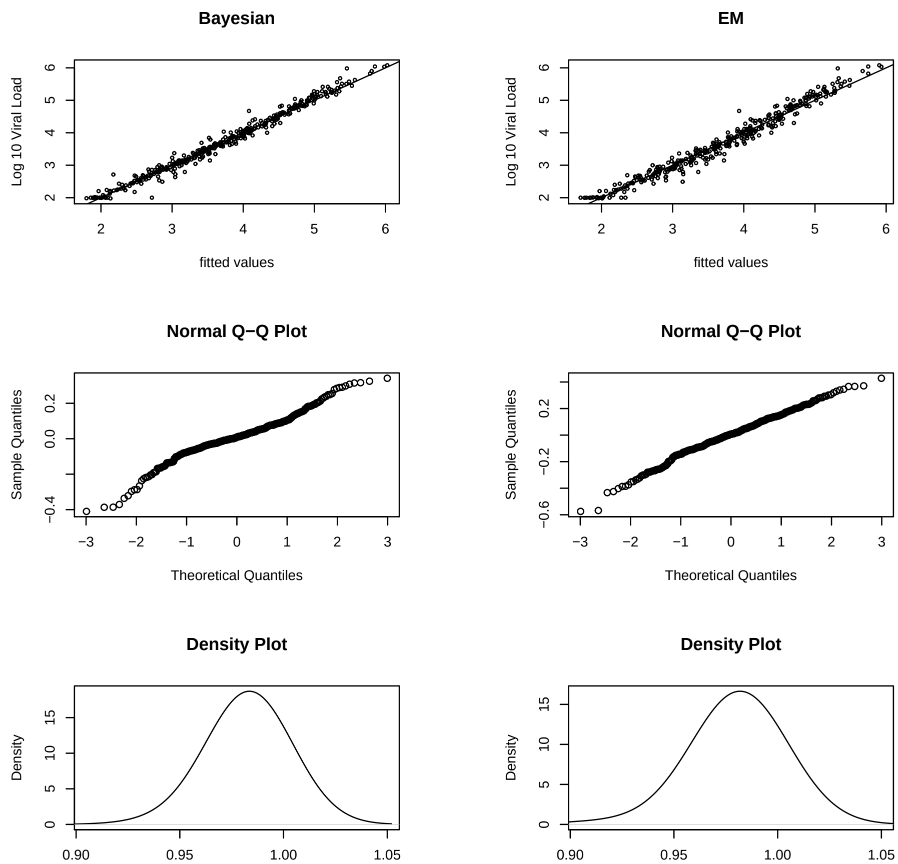

Figure S1. Model checking plots for the HIV dynamic application study. The plots of fitted value vs data, QQ plots of residual and density plots of individual estimates from the Bayes method (left panel) and the EM algorithm (right panel) are given, respectively.