Web-based supporting material for “Bayesian Inference for Dynamic Cointegration Models: a case study on the Soybean Crush Spread”

Maciej Marowka

Department of Mathematics, Imperial College London, UK.

E-mail: maciej.marowka14@imperial.ac.uk

Gareth W. Peters

Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, UK.

Nikolas Kantas

Department of Mathematics, Imperial College London, UK.

Guillaume Bagnarosa

ESC Rennes School of Business, France.

Summary. In this document we provide the derivations for the posterior conditional distributions and numerical results for the Markov Chain Monte Carlo (MCMC) sampler on simulated data.

1. Introduction

Recall we interested to simulate from the following posterior:

$$\begin{aligned} p(\alpha, \beta, H, R, B, Q, \xi, X|Y) \propto\ & p(Y|\alpha, \beta, H, R, \xi, X)p(X|B, Q) \\ & \times p(\alpha|\beta)p(\beta)p(\xi)p(B)p(Q)p(H)p(R) \end{aligned} \tag{1}$$

where the likelihood is from

$$Y_t = Y_{t-1} + \alpha\beta^T Y_{t-1} + \mu_t + g_t + \epsilon_t, \quad \epsilon_t \sim N(0, R), \tag{2}$$

with

$$\mu_t = HX_t \tag{3}$$

$$X_t = BX_{t-1} + \delta_t, \quad \delta_t \sim N(0, Q), \quad x_0 = 0 \tag{4}$$

and

$$g_t = \xi_1 + \sum_{i=2}^{m} \xi_i \mathbb{I}_{\{t \in \mathcal{I}_i\}}, \tag{5}$$

where $\mathcal{I}_i$ denotes a specific interval in time such as the i-th week or month. For the priors we use

$$\begin{aligned} B &\sim N_{k,k}(0, \sigma_B^2 I_k, \sigma_B^2 I_k), \\ Q &\sim \mathcal{W}^{-1}(\nu_Q, \sigma_Q^2 I_k), \\ H &\sim N_{n,k}(0, \sigma_H^2 I_n, \sigma_H^2 I_k), \\ R &\sim \mathcal{W}^{-1}(\nu_R, \sigma_R^2 I_n), \\ \xi &\sim N_{n,m}(0, \sigma_\xi^2 I_n, \sigma_\xi^2 I_m), \end{aligned}$$

and for the hyper parameters:

$$\begin{aligned} \sigma_H^2 &\sim IG(\alpha_H, \beta_H), \\ \sigma_B^2 &\sim IG(\alpha_B, \beta_B), \\ \sigma_R^2 &\sim G(\alpha_R, \beta_R). \end{aligned}$$

1.1. Notation

We will use the following notation: $N(\mu, \Sigma)$ will be multivariate normal with mean vector $\mu \in \mathbb{R}^{n\times 1}$ and covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$. We denote a Gaussian random $(n \times T)$ matrix by $Y \sim N_{n,T}(\mu, \Sigma, \Psi)$ with row dependence in $(n \times n)$ covariance matrix Σ and column dependence in $(T \times T)$ matrix Ψ. Matrix variate Inverse Wishart distribution with the support on symmetric positive definite matrices $(n \times n)$ with $\nu > n - 1$ degrees of freedom and scale matrix Ψ is $\mathcal{W}^{-1}(\nu, \Psi)$. Gamma and Inverse Gamma distributions will be denoted as $G(\alpha_\gamma, \beta_\gamma)$ and $IG(\alpha_\gamma, \beta_\gamma)$ respectively with scale and shape parameters α_γ and β_γ. By $I_n \in \mathbb{R}^{n\times n}$ we will refer to the square identity matrix; $Vec(M)$ denotes the matrix vectorization operator which transform a matrix M into column vector in which columns of M are successively stacked. Furthermore we denote the Kronecker product or tensor product between two matrices by $\otimes$ and Kronecker sum as $\oplus$. For a sequence $y_1, y_2, ..., y_T$, we will also use the concise notation $y_{1:T}$.

2. Expressions for the posterior conditionals

For the convenience of the reader we state again Algorithm 1 used in the paper. The expressions of all distributions used in this algorithm are stated below in the Proposition 1. The derivations are presented in Section 3.3

Algorithm 1 A Partially Collapsed Gibbs Sampler to simulate from (1)

Initialization: Draw the parameters from their respective prior distributions For $i = 1, 2, ..., N$**:** Sample

- 1. $Vec(B^i) \sim p(\cdot|Q^{i-1}, x^{i-1}, \sigma_B^{2\ i-1})$,
- 2. $\sigma_B^{2\ i} \sim p(\cdot|B^i)$,
- 3. $Q^i \sim p(\cdot|B^i, x^{i-1})$,
- 4. $Vec(H^i) \sim p(\cdot|y, \alpha^{i-1}, \beta^{i-1}, R^{i-1}, x^{i-1}, \xi^{i-1}, \sigma_H^{2\ i-1})$,
- 5. $\sigma_H^{2\ i} \sim p(\cdot|H^i)$,
- 6. $R^i \sim p(\cdot|y, \alpha^{i-1}, \beta^{i-1}, H^i, x^{i-1}, \xi^{i-1}, \sigma_R^{2\ i-1})$,
- 7. $\sigma_R^{2\ i} \sim p(\cdot|R^i)$,
- 8. $Vec(\alpha^*) \sim p(\cdot|y, \beta^{i-1}, H^i, R^i, B^i, Q^i, \xi^{i-1})$. Compute $\mathcal{A}^* = \alpha^*(\alpha^{*T}\alpha^*)^{-1/2}$,
- 9. $Vec((\mathcal{B}^*)^T) \sim p(\cdot|y, \mathcal{A}^*, H^i, R^i, B^i, Q^i, \xi^{i-1})$. Compute $\beta^i = \mathcal{B}^*(\mathcal{B}^{*T}\mathcal{B}^*)^{-1/2}$ and $\alpha^i = \mathcal{A}^*(\mathcal{B}^{*T}\mathcal{B}^*)^{1/2}$
- 10. $Vec(\xi^i) \sim p(\cdot|y, \alpha^i, \beta^i, H^i, R^i, B^i, Q^i)$,
- 11. Sample $x^i \sim p(\cdot|y, \alpha^i, \beta^i, \xi^i, H^i, R^i, B^i, Q^i)$

Proposition 1. *The conditional distributions used in Algorithm 1 are as follows:*

$$\begin{aligned}
p_B(Vec(B)|Q, X) =& N(\mu^B_{post}, \Sigma^B_{post}),\\
p_Q(Q|B, X) =& \mathcal{W}^{-1}(\nu_Q + T, \tilde{X}\tilde{X}^T + \sigma_Q^2 I_K),\\
p_H(Vec(H)|\alpha, \beta, Y, \xi) =& N(\mu^H_{post}, \Sigma^H_{post}),\\
p_R(R|Y, \alpha, \beta, \xi) =& \mathcal{W}^{-1}(\nu + T, \tilde{Y}\tilde{Y}^T + \sigma^R I_n),\\
p_\alpha(Vec(\alpha)|Y, \beta, H, R, B, Q, \xi) =& N(\mu^\alpha_{post}, \Sigma^\alpha_{post}), && (6)\\
p_{\mathcal{B}}(Vec(\mathcal{B}^T)|Y, \mathcal{A}, H, R, B, Q, \xi) =& N(\mu^{\mathcal{B}}_{post}, \Sigma^{\mathcal{B}}_{post}), && (7)\\
p_\xi(Vec(\xi)|Y, \alpha, \beta, H, R, B, Q) =& N(\mu^\xi_{post}, \Sigma^\xi_{post}),\\
p_{\sigma_R^2}(\sigma_R^2|R) =& G(\frac{n\nu_R}{2} + \alpha_{\sigma_R^2}, \beta_{\sigma_R^2} + \frac{1}{2}tr(R^{-1})),\\
p_{\sigma_H^2}(\sigma_H^2|H) =& IG(\alpha_{\sigma_H^2} + \frac{nK}{2}, \beta_{\sigma_H^2} + \frac{1}{2}Vec(H)^T Vec(H)),\\
p_{\sigma_B^2}(\sigma_B^2|B) =& IG(\alpha_{\sigma_B^2} + \frac{K^2}{2}, \beta_{\sigma_B^2} + \frac{1}{2}Vec(B)^T Vec(B)).
\end{aligned}$$

The parameters of each distribution are given by

$$
\begin{aligned}
\left(\Sigma^B_{post}\right)^{-1} &= (X_0^T \otimes I_K)^T(\oplus_{t=1}^T Q^{-1})(X_0^T \otimes I_K)^T + (\sigma_B)^{-2} I_{K^2},\\
\mu^B_{post} &= \Sigma^B_{post}((X_0^T \otimes I_K)^T(\oplus_{t=1}^T Q^{-1}) vec(X)),\\
\left(\Sigma^H_{post}\right)^{-1} &= (X^T \otimes I_K)^T(\oplus_{t=1}^T R^{-1})(X^T \otimes I_K)^T + (\sigma_H)^{-2} I_{K^2},\\
\mu^H_{post} &= \Sigma^H_{post}((X^T \otimes I_K)^T(\oplus_{t=1}^T R^{-1}) Vec(\tilde{Y})),\\
\left(\Sigma^{\cdot}_{post}\right)^{-1} &= M_{\cdot}^T \widetilde{V}^{-1} M_{\cdot} + (\Sigma_{\cdot})^{-1},\\
\mu^{\cdot}_{post} &= \Sigma^{\cdot}_{post}(M_{\cdot}^T \widetilde{V}^{-1} \widetilde{y}),
\end{aligned}
$$

where we use $\cdot$ *for either* $\alpha, \mathcal{B}, \xi$ *and define* $\widetilde{V} = \oplus_{t=1}^T R + \widetilde{H} V \widetilde{H}^T$, $\widetilde{y} = Vec(Y) - (\Upsilon^T \otimes I_n)\xi$, $M_\alpha = Y_0^T \beta \otimes I_n$, $M_{\mathcal{B}} = Y_0^T \otimes \alpha$, $\tilde{X} = (x_1 - Bx_0, \ldots, x_T - Bx_{T-1})$, $X_0 = (x_0, \ldots, x_{T-1})$, $\widetilde{H} = (I_T \otimes H)$, $V = (\widetilde{A}^T \oplus_{t=1}^T Q^{-1} \widetilde{A})^{-1}$ *and* $\widetilde{A} = I_{TK} - (I_T \otimes B) P M$, *with*

$$
\begin{aligned}
P &= \begin{bmatrix} 0_{K\times K(T-1)} & I_{K\times K} \\ I_{K(T-1)\times K(T-1)} & 0_{K(T-1)\times K} \end{bmatrix}\\
M &= \begin{bmatrix} I_{K(T-1)\times K(T-1)} & 0_{K\times K(T-1)} \\ 0_{K(T-1)\times K} & 0_{K\times K} \end{bmatrix}.
\end{aligned}
$$

3. Numerical examples using Simulated Data

In this section we demonstrate the performance of Algorithm 1 for a simulated data-set using $n = 4$, $R = diag(1, 1, 1, 1)$ and $Q = diag(1, 1)$. The numerical examples presented in this section aim to assess the accuracy of the estimation and the improvement of Algorithm 1 over a standard Gibbs sampler based on data augmentation; Carter and Kohn [1994], Frühwirth-Schnatter [1994], De Jong and Shephard [1995]. In these experiments we ran Algorithm 1 for $N = 10^5$ and discarded 10^4 samples for burn-in. Furthermore, throughout this section we will resort to using the posterior means as point estimates.

Posterior credible intervals, estimated path and true data generating μ are presented in Figure 2, posterior densities for $\Pi = \alpha\beta^T$ in Figure 3 and box plots for the seasonal components ξ in Figure 4. The estimation seems to be very accurate. Furthermore, for a sample output $\Pi^{(i)} = \alpha^{(i)}\beta^{(i)T}$ we compare Algorithm 1 with a standard Gibbs sampler in terms of autocorrelation functions in Figure 1 and Effective Sample Size (ESS) presented in Table 3. It is clear that in terms of mixing and efficiency Algorithm 1 shows superior performance.

3.1. Derivations for the posterior conditional distributions in Algorithm 1

We begin with some auxiliary results and then proceed withe the proof of Proposition 1

	$\Pi_{\cdot,1}$	$\Pi_{\cdot,2}$	$\Pi_{\cdot,3}$	$\Pi_{\cdot,4}$
$\Pi_{1,\cdot}$	0.40	0.63	0.44	0.31
$\Pi_{2,\cdot}$	0.74	0.70	0.48	0.53
$\Pi_{3,\cdot}$	0.62	0.51	0.49	0.39
$\Pi_{4,\cdot}$	0.40	0.46	0.37	0.35

	$\Pi_{\cdot,1}$	$\Pi_{\cdot,2}$	$\Pi_{\cdot,3}$	$\Pi_{\cdot,4}$
$\Pi_{1,\cdot}$	0.34	0.31	0.24	0.17
$\Pi_{2,\cdot}$	0.33	0.14	0.22	0.15
$\Pi_{3,\cdot}$	0.20	0.09	0.10	0.10
$\Pi_{4,\cdot}$	0.19	0.08	0.09	0.10

Table 1. ESS for each of sixteen entries of $\Pi = \alpha\beta^T$ with Algorithm 1 (left table) and standard Gib sampler (right table). The values suggest that the proposed algorithm is consistenly better than a standa systematic scan Gibbs.

3.2. Auxiliary results

Proposition 2. *For $X = (x_1, \ldots, x_T)$ given by $x_t = Bx_{t-1} + \delta_t$, (with $x_0 = 0$,) we have that $Vec(X) \sim N(0, V)$, with $V = (\widetilde{A}^T \oplus_{t=1}^T Q^{-1}\widetilde{A})^{-1}$ and $\widetilde{A} = I_{TK} - (I_T \otimes B)PM$, with $P = \begin{bmatrix} 0_{K\times K(T-1)} & I_K \\ I_{K(T-1)} & 0_{K(T-1)\times K} \end{bmatrix}$ and $M = \begin{bmatrix} I_{K(T-1)} & 0_{K\times K(T-1)} \\ 0_{K(T-1)\times K} & 0_{K\times K} \end{bmatrix}$.*

Proof. We can rewrite $x_t = Bx_{t-1} + \delta_t$ in a matrix regression format by stacking vectors of the series realizations x_t in the columns of matrix X,

$$X = BX_0 + \delta_{1:T},$$

where $X_0 = (x_0, \ldots, x_{T-1})$. After applying the vectorization we get

$$Vec(X) = (I_T \otimes B)Vec(X_0) + W$$

where $W \sim N(0, \oplus_{i=1}^T Q)$. Recall the permutation matrix P is invertible and hence we can write

$$\begin{aligned} Vec(X) &= (I_T \otimes B)PP^{-1}Vec(X_0) + W \\ Vec(X) &= \widetilde{B}Vec(X_0') + W \\ Vec(X) &= \widetilde{C}Vec(X) + W \\ \widetilde{A}Vec(X) &= W \end{aligned} \tag{8}$$

where we let $\widetilde{B} = (I_T \otimes B)P$, $X_0' = (x_1, \ldots, x_{T-1}, x_0)$ $\widetilde{C} = \widetilde{B}\begin{bmatrix} I_{K(T-1)} & 0_{K(T-1)\times K} \\ 0_{K\times K(T-1)} & 0_{K\times K} \end{bmatrix}$ and $\widetilde{A} = I_{TK} - \widetilde{C}$. From (8) we get

$$\begin{aligned} p(Vec(X)) &\propto \exp(-\frac{1}{2}(\widetilde{A}Vec(X))^T \oplus_{t=2}^T Q^{-1}((\widetilde{A}Vec(X)) \\ &\propto \exp(-\frac{1}{2}Vec(X)^T(\widetilde{A}^T \oplus_{t=1}^T Q^{-1}\widetilde{A})Vec(X)). \end{aligned}$$

and so $Vec(X) \sim N(0, V)$, where $V = (\widetilde{A}^T \oplus_{t=1}^T Q^{-1}\widetilde{A})^{-1}$. □

Note that in Proposition 2 we use partial collapsing and sample the cointegration parameters in (6), (7) without conditioning on the hidden states X,. This significantly reduces dimensionality in Gibbs updates and improves the mixing of the sampler.

We will make extensive use of the lemma below, which allows for analytic representation of the full and marginalized conditionals in Gaussian matrix variate likelihood model.

Lemma 1. *Consider a Gaussian vector regression model with output $\mathcal{Y} \in \mathbb{R}^{n\times 1}$ and input $\mathcal{X} \in \mathbb{R}^{n\times n}$ where*

$$\mathcal{Y} = \mathcal{X}\theta + \epsilon$$

and $\theta \in \mathbb{R}^{n\times 1}$ and $\epsilon \in \mathbb{R}^{n\times 1}$ is an innovation error vector such that $\epsilon \sim N(0, \Sigma_{\mathcal{Y}})$. Suppose that a conjugate prior of θ is a Gaussian vector $p(\theta) \sim N(m_\theta, \Sigma_\theta)$. Then $p(\theta|\mathcal{Y}) \sim N(\widetilde{m_\theta}, \widetilde{\Sigma_\theta})$, where

$$\widetilde{\Sigma_\theta} = (\mathcal{X}^T \Sigma_{\mathcal{Y}}^{-1} \mathcal{X} + \Sigma_\theta^{-1})^{-1},$$
$$\widetilde{m_\theta} = M_\theta(\mathcal{X}^T \Sigma_{\mathcal{Y}}^{-1} \mathcal{Y} + \Sigma_\theta^{-1} m_\theta).$$

The proof is omitted as it follows by straightforward linear algebra.

Lemma 2. *Suppose $\boldsymbol{\mathcal{Y}} := [\mathcal{Y}_1, \mathcal{Y}_2, ..., \mathcal{Y}_n]$ is a matrix whose columns are i.i.d samples from $N(0, \Sigma)$. If $\Sigma \sim \mathcal{W}^{-1}(\nu, \Psi)$, then $p(\Sigma|\boldsymbol{\mathcal{Y}}) \sim \mathcal{W}^{-1}(\nu + n, \Psi + \boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{Y}}^T)$.*

See Gupta and Nagar [1999] for a proof.

3.3. Proof of Proposition 1

We consider each posterior conditional separately.

$p_B(Vec(B)|Q, x_{1:T})$: As in the proof of Proposition 2 above,

$$x_t = Bx_{t-1} + \delta_t \ \forall_{t=1,...,T}$$

can be rewritten in matrix regression format as:

$$X = BX_0 + W$$

with $W = (\delta_1, \delta_2, \dots, \delta_T)$. Applying matrix vectorization operator:

$$Vec(X) = (X_0^T \otimes I_K)Vec(B) + w, \tag{9}$$

where $w \sim N(0, \oplus_{t=1}^T Q)$. Hence by multivariate normal conjugacy implied by Gaussian likelihood in (9) and the results of Lemma 1, the conditional posterior of $Vec(B)$ is also multivariate normal

$$Vec(B)|Q, X \sim N(\mu_{post}^B, \Sigma_{post}^B), \tag{10}$$

where

$$\Sigma_{post}^B = ((X_0^T \otimes I_K)^T (\oplus_{t=1}^T Q^{-1})(X_0^T \otimes I_K)^T + (\sigma_B)^{-2} I_{K^2})^{-1}$$
$$\mu_{post}^B = \Sigma_{post}^B((X_0^T \otimes I_K)^T (\oplus_{t=1}^T Q^{-1}) vec(X)).$$

We proceed with $p_Q(Q|B, x_{1:T})$. We can simulate from full conditional update by exploiting conjugacy class relations of normal and inverse Wishart distributions as well as conditional independence of Q and observed time series realizations given latent path. The hidden model likelihood can be written as

$$\widetilde{x}_t = Q^{1/2} r_t, \quad \widetilde{x}_t = x_t - Bx_{t-1},\ x_0 = 0,$$

with r_t i.i.d. multivariate standard random normal. Since the prior for Q is $\mathcal{W}^{-1}(\nu_Q, \sigma_Q^2 I_K)$, then by the Lemma 2 we have

$$Q|B, X \sim \mathcal{W}^{-1}(\nu_Q + T, \widetilde{X}\widetilde{X}^T + \sigma_Q^2 I_K), \tag{11}$$

with $\widetilde{X} = X - BX_0$.

For computing $p_H(Vec(H)|\alpha, \beta, Y, x_{1:T,}\xi)$: Note that, conditioning on the realization of latent path, observation model parameters are independent of hidden static parameters. Recall, $\tilde{Y} = Y - \alpha\beta^T Y_0 - \xi\Upsilon$ and noting vectorized regression format as

$$Vec(\widetilde{Y}) = (X^T \otimes I_K)vec(H) + \widetilde{E},$$

where $\widetilde{E} \sim N(0, \oplus_{t=1}^T R)$, we can exploiting the same conjugacy properties and Lemma 1 and 2 to obtain

$$Vec(H)|Y, \alpha, \beta, X, \xi \sim N(\mu_{post}^H, \Sigma_{post}^H),$$

where

$$\Sigma_{post}^H = ((X^T \otimes I_K)^T (\oplus_{t=1}^T R^{-1})(X^T \otimes I_K)^T + (\sigma_H)^{-2} I_{K^2})^{-1},$$

$$\mu_{post}^H = \Sigma_{post}^B((X^T \otimes I_K)^T (\oplus_{t=1}^T R^{-1})vec(\widetilde{Y})).$$

$p_R(R|Y, \alpha, \beta, x_{1:T}, \xi)$: Conditional on latent path and all remaining parameters, $\widetilde{Y} - HX$ can be regarded as T vector variate normal samples with zero mean and variance R. Hence the full conditional of R follows directly from Lemma 3:

$$R|\alpha, \beta, Y, X, \xi \sim \mathcal{W}^{-1}(\nu + T, (\widetilde{Y} - HX)(\widetilde{Y} - HX)^T + \sigma_R^2 I_n). \tag{12}$$

$p_\alpha(Vec(\alpha)|Y, \beta, H, R, B, Q, \xi)$, $p_{\mathcal{B}}(Vec(\mathcal{B}^T)|Y, \mathcal{A}, H, R, B, Q, \xi)$: To obtain cointegration parameters' marginalized conditionals of partially collapsed Gibbs algorithm, we utilize convolution property of normal distribution, Proposition 2 and Lemma 1 together with the following representation:

$$\begin{aligned} Vec(Y) &= (Y_0^T \beta \otimes I_n)Vec(\alpha) + Vec(\mu) + (\Upsilon^T \otimes I_n)Vec(\xi) + \widetilde{E} \\ &= (Y_0^T \otimes \mathcal{A})Vec(\mathcal{B}^T) + Vec(\mu) + (\Upsilon^T \otimes I_n)Vec(\xi) + \widetilde{E} \\ &= (Y_0^T \beta \otimes I_n)Vec(\alpha) + (\Upsilon^T \otimes I_n)Vec(\xi) + \widetilde{\boldsymbol{E}} \\ &= (Y_0^T \otimes \mathcal{A})Vec(\mathcal{B}^T) + (\Upsilon^T \otimes I_n)Vec(\xi) + \widetilde{\boldsymbol{E}}, \end{aligned} \tag{13}$$

where $\widetilde{\boldsymbol{E}} \sim N(0, \widetilde{V})$; $\widetilde{V} = \oplus_{t=1}^{T} R + \widetilde{H} V \widetilde{H}^T$ with $\widetilde{H}$ and V defined in proof of Proposition 2 above. If we define further:

$$\begin{aligned}
\widetilde{y} &= Vec(Y) - (\varUpsilon^T \otimes I_n)\xi; \\
M_\alpha &= (Y_0^T \beta \otimes I_n); \\
M_{\mathcal{B}} &= (Y_0^T \otimes \alpha).
\end{aligned}$$

Then, it follows from Lemma 1 that collapsed conditionals (with marginalized latent paths) for vectorized components of cointegration model, $Vec(\alpha)$ and $Vec(\mathcal{B}^T)$ are distributed as vector variate normals, p_α, p_β respectively, with variances and means as provided below:

$$p_\alpha \sim N(\mu_{post}^{\alpha}, \Sigma_{post}^{\alpha}), \tag{14}$$

$$p_{\mathcal{B}} \sim N(\mu_{post}^{\mathcal{B}}, \Sigma_{post}^{\mathcal{B}}). \tag{15}$$

$$\begin{aligned}
\Sigma_{post}^{\alpha} &= (M_\alpha^T \widetilde{V}^{-1} M_\alpha + (\Sigma_\alpha)^{-1})^{-1}, \\
\mu_{post}^{\alpha} &= \Sigma_{post}^{\alpha} (M_\alpha^T \widetilde{V}^{-1} \widetilde{y}).
\end{aligned} \tag{16}$$

$$\begin{aligned}
\Sigma_{post}^{\mathcal{B}} &= (M_{\mathcal{B}}^T \widetilde{V}^{-1} M_{\mathcal{B}} + (\Sigma_{\mathcal{B}})^{-1})^{-1}, \\
\mu_{post}^{\mathcal{B}} &= \Sigma_{post}^{\mathcal{B}} (M_{\mathcal{B}}^T \widetilde{V}^{-1} \widetilde{y}).
\end{aligned} \tag{17}$$

Using (13) again, we can define

$$\begin{aligned}
\mathcal{Y} &= Vec(Y) - (Y_0^T \otimes \mathcal{A}) Vec(\mathcal{B}^T), \\
M_\xi &= (\Upsilon^T \otimes I_n);
\end{aligned}$$

Then analogously as in (16), (17) and by Lemma 1,

$$p_\xi(Vec(\xi)|Y, \alpha, \beta, H, R, B, Q) = N(\mu_{post}^{\xi}, \Sigma_{post}^{\xi})$$

where

$$\begin{aligned}
\Sigma_{post}^{\xi} &= (M_\xi^T \widetilde{V}^{-1} M_\xi + (\Sigma_\xi)^{-1})^{-1} \\
\mu_{post}^{\xi} &= \Sigma_{post}^{\xi} (M_\xi^T \widetilde{V}^{-1} \tilde{y}),
\end{aligned}$$

and $\Sigma_\xi = \sigma_\xi^2 I_{mn}$.

$p_{\sigma_R^2}(\sigma_R^2|R)$: Assuming Inverse Wishart prior of $R \sim \mathcal{W}^{-1}(\nu_R, \sigma_R^2 I_n)$ with hyperparameter σ_R^2 following gamma distribution $\sigma_R^2 \sim Gamma(\alpha_R, \beta_R)$, then by straightforward algebra:

$$\begin{aligned}
p(\sigma_R^2|R, y_{1:T}) = p(\sigma_R^2|R) & \\
&\propto (\sigma_R^2)^{\frac{\nu_R n}{2}} \exp\left(-\frac{1}{2}\sigma_R^2 tr(R^{-1})\right) \cdot (\sigma_R^2)^{\alpha_R - 1} \exp(-\beta_R \sigma_R^2) \\
&= (\sigma_R^2)^{\frac{\nu_R n}{2} + \alpha_R - 1} \exp\left(-\sigma_R^2 \left(\beta_R + \frac{1}{2} tr(R^{-1})\right)\right)
\end{aligned}$$

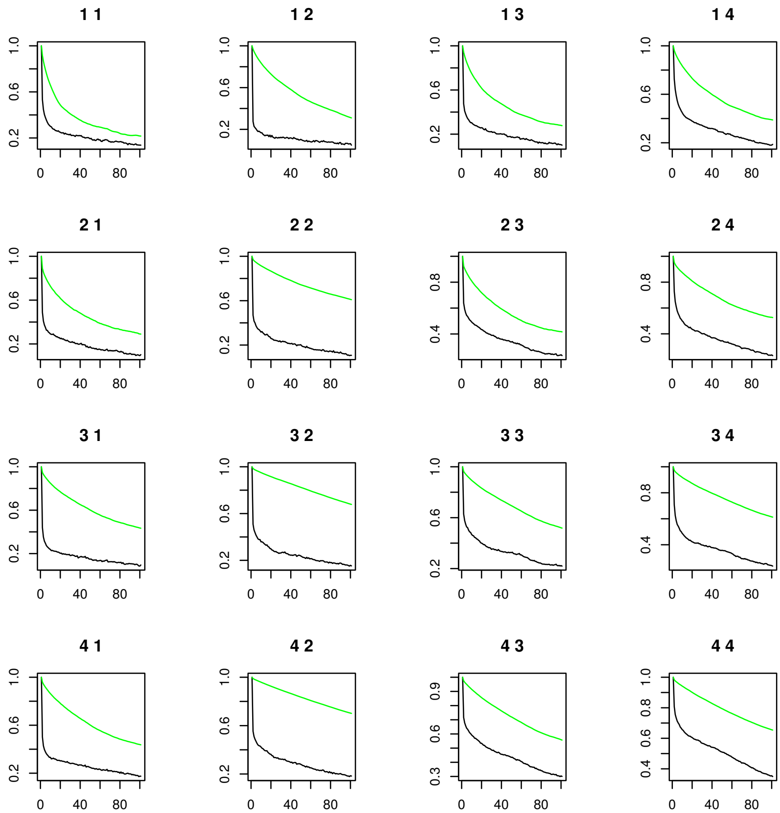

Fig. 1. Autocorrelation of $\alpha^{(i)}\beta^{(i)T}$ against lag: black is for Algorithm 1 and green for standard standard Gibbs sampler.

so it follows clearly that $p_{\sigma^2_R}(\sigma^2_R|R)$ is a $Gamma\left(\frac{n\nu_R}{2} + \alpha_{\sigma^2_R}, \beta_{\sigma^2_R} + \frac{1}{2}tr(R^{-1})\right)$.

$p_{\sigma^2_H}(\sigma^2_H|H)$: Assuming multivariate normal prior of $Vec(H) \sim N(0, \sigma^2_H I_{nK})$ with hyperparameter σ^2_H following inverse gamma distribution $\sigma^2_H \sim IG(\alpha_H, \beta_H)$, it follows that

$$
\begin{aligned}
p(\sigma^2_H|H, y_{1:T}) &= p(\sigma^2_H|H) \\
&\propto (\sigma^2_H)^{-nK/2} \exp\left(-\frac{1}{2\sigma^2_H} Vec(H)^T Vec(H)\right) \cdot \left(\sigma^2_H\right)^{-\alpha_H - 1} \exp(-\frac{\beta_H}{\sigma^2_H}),
\end{aligned}
$$

so $p_{\sigma^2_H}(\sigma^2_H|H)$ is $IG\left(\alpha_{\sigma^2_H} + \frac{nK}{2}, \beta_{\sigma^2_H} + \frac{1}{2}Vec(H)^T Vec(H)\right)$

$p_{\sigma^2_B}(\sigma^2_B|B)$: By analogous argument as above we have $p_{\sigma^2_B}(\sigma^2_B|B) = IG(\alpha_{\sigma^2_B} + \frac{K^2}{2}, \beta_{\sigma^2_B} + \frac{1}{2}Vec(B)^T Vec(B))$.

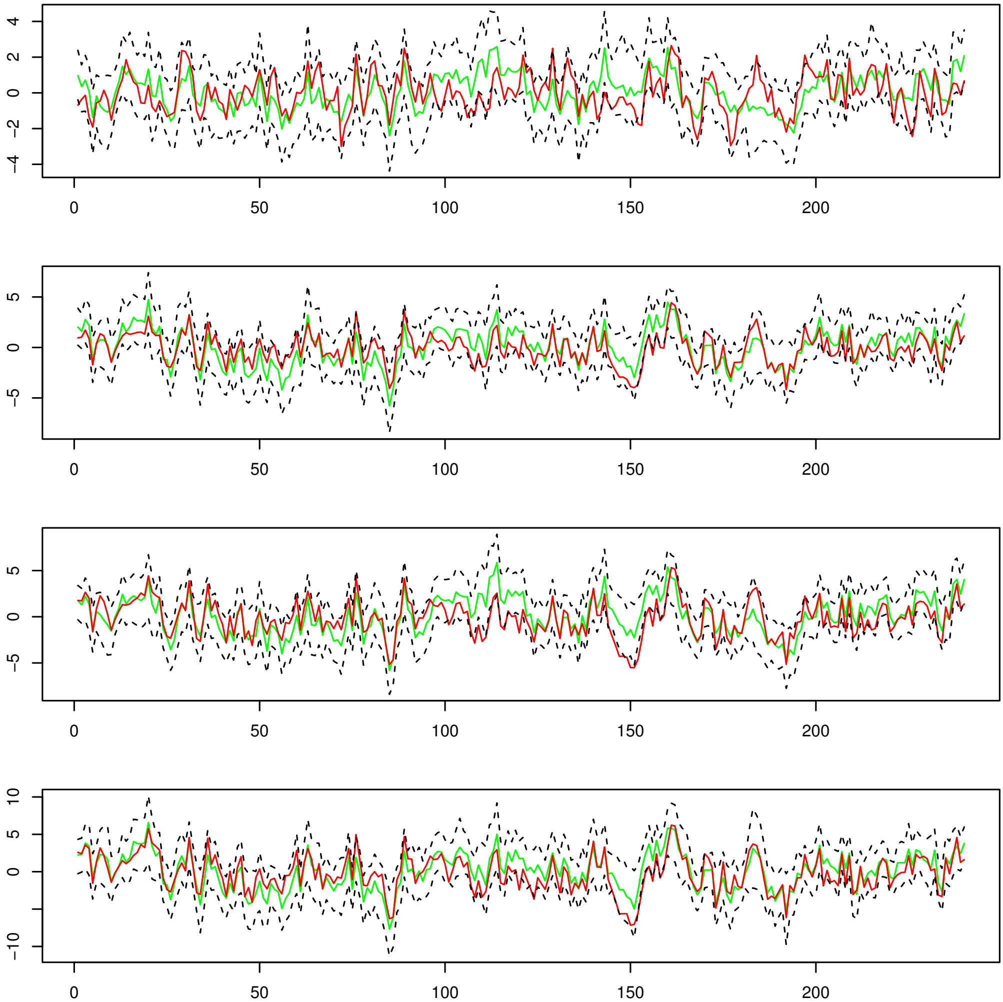

Fig. 2. Estimation of $\mu_t(l)$ for $l = 1, \ldots, 4$: in each panel we show the true $\mu_t(l)$ used to simulate the data (red line), the posterior mean estimate from the Algorithm 1 (green line) and 95% posterior confidence intervals (dashed lines).

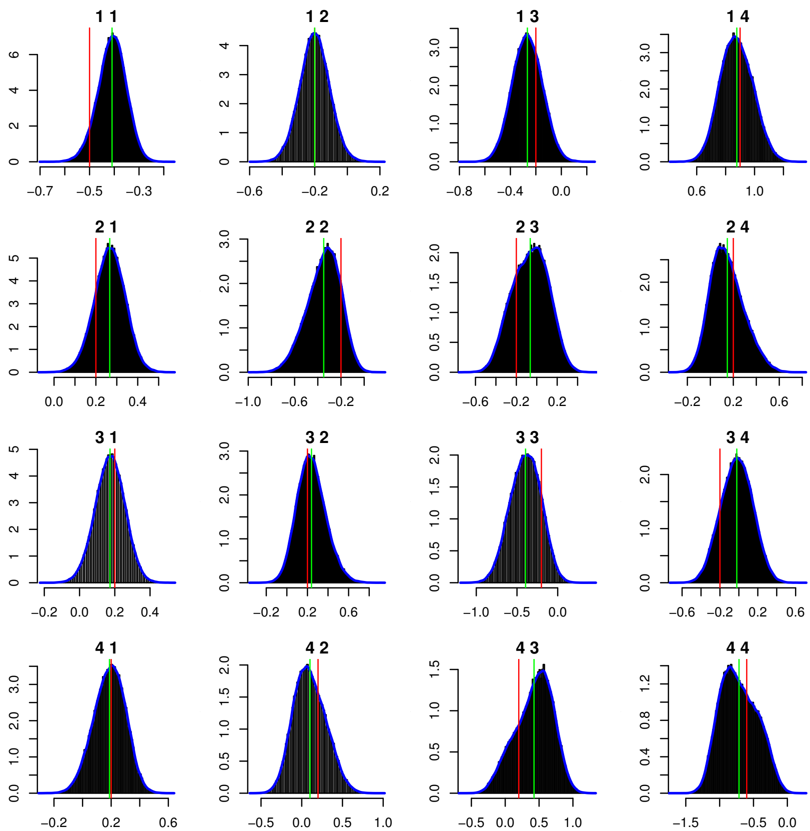

Fig. 3. Histograms with estimated density curves of all entries of the long-run multiplier matrix $\alpha\beta^T$. The red horizontal lines mark the true value and the green ones the estimated posterior mean. Blue line is the fitted posterior density curve.

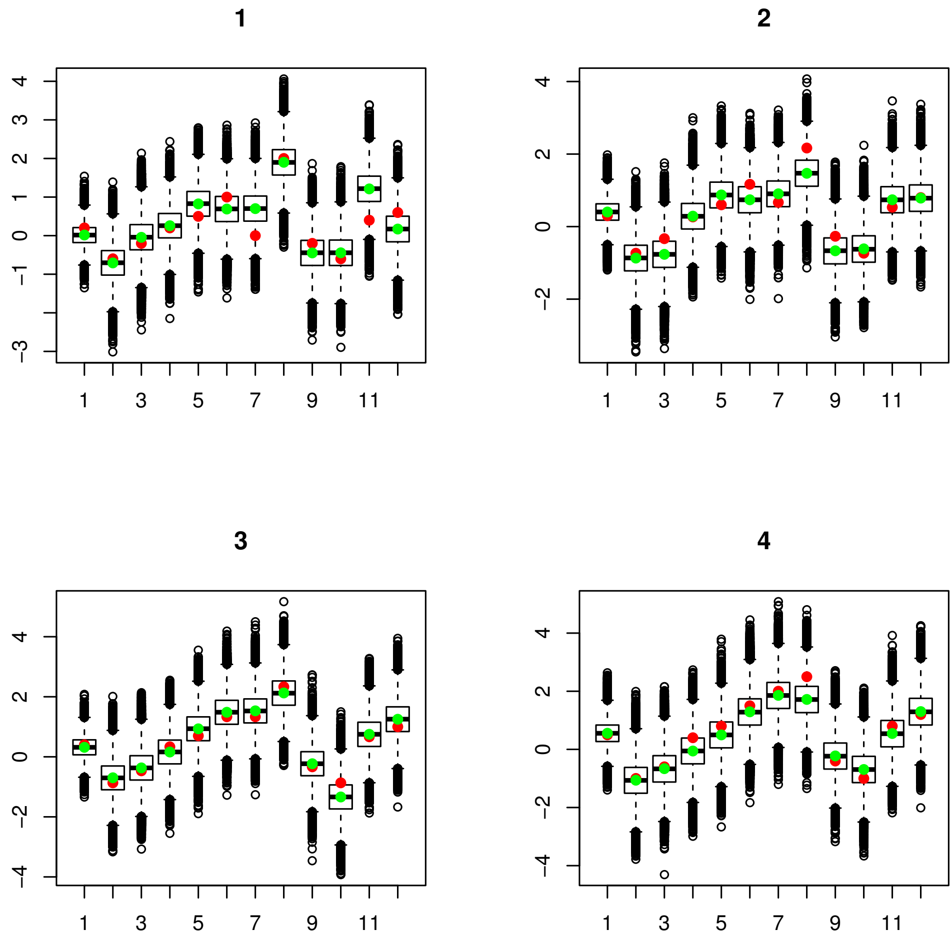

Fig. 4. Boxplots for posterior of ξ: the green dot indicates the true value and the red one the posterior mean.

References

Chris K Carter and Robert Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

Piet De Jong and Neil Shephard. The simulation smoother for time series models. *Biometrika*, 82(2):339–350, 1995.

Sylvia Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of time series analysis*, 15(2):183–202, 1994.

Arjun K Gupta and Daya K Nagar. *Matrix variate distributions*, volume 104. CRC Press, 1999.