Efficient Bayesian Estimation of Multivariate State Space Models

Chris M. Strickland*, Ian. W. Turner, Robert Denham, Kerrie L. Mengersen.

February 14, 2008

Abstract

A Bayesian Markov chain Monte Carlo methodology is developed for the estimation of multivariate linear Gaussian state space models. In particular, an efficient simulation smoothing algorithm is proposed that makes use of the univariate representation of the state space model. Substantial gains over existing algorithms in computational efficiency are achieved using the new simulation smoother for the analysis of high dimensional multivariate time series.

The methodology is used to analyse a multivariate timeseries dataset of the Normalised Difference Vegetation Index (NDVI), which is a proxy for the level of live vegetation, for a particular grazing property located in Queensland, Australia.

Key Words: Multivariate; State space model; Markov chain Monte Carlo; Kalman filter; Simulation smoother; Univiarate representation; MODIS; Stochastic cycle.

*Corresponding Author: School of Mathematics, GPO Box 2434, Queensland University of Technology, Queensland, 4001, Australia. Email: christopher.strickland@qut.edu.au. Phone: 61 7 3138 8313. Fax: 61 7 3138 2310.

1 Introduction

Structural time series models (STSMs) provide an attractive framework for the analysis of both univariate and multivariate time series data. The state space model (SSM) provides a generic representation for STSMs that facilitates the development and use of general estimation algorithms. This paper presents a simple Bayesian Markov chain Monte Carlo (MCMC) framework for the estimation of multivariate SSMs that delivers substantial gains in computational efficiency over the standard Bayesian approach. A transformation is used that enables the use of the univariate representation of the SSM. A simulation smoothing algorithm is presented for this representation, extending the algorithms of Anderson and Moore (1979), Koopman and Durbin (2000) and Durbin and Koopman (2002). A comparison is made between the proposed simulation smoothing algorithm and the standard approach of Durbin and Koopman (2002). Large gains in computational efficiency are shown to be produced by the proposed method in high dimensional multivariate analysis.

This method is illustrated through an empirical analysis of Normalized Difference Vegetation Index (NDVI) images collected from the Moderate Resolution Imaging Spectroradiometer (MODIS). NDVI is a common vegetation index providing information on the amount of live green vegetation in an area. It is commonly used to infer information about properties of vegetation such as biomass, cover or leaf area index. A STSM is constructed to characterise the trend and seasonal components that are inherent in the time series. The trend is modelled as a random walk and a stochastic cycle is used to characterise the seasonal pattern.

The structure of the paper is as follows. In Section 2 the multivariate Gaussian SSM is introduced and a possible MCMC sampling scheme is outlined in Section 3, with details on how to sample the state vector as well as a computational comparison between the standard and the proposed approaches. In Section 4 an empirical study of NDVI for a grazing property in Queensland, Australia is conducted. For the empirical analysis, the use of the new algorithm means that the computational burden as measured by the number of multiplications, for the simulation smoothing aspect of the MCMC scheme, is approximately 90% lower than if the standard approach had been used. Some conclusions and extensions are provided in Section 5.

2 Multivariate Linear Gaussian State Space Model

The observation equation for the linear Gaussian state space model (SSM) for the $(p \times 1)$ vector of observations $\mathbf{y}_t = (y_{t,1}, y_{t,2}, \ldots, y_{t,p})'$, is given by

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim N\left(0, \mathbf{H}^{-1}\right), \tag{1}$$

for $t = 1, 2, \ldots, n$, where the $(p \times m)$ matrix $\mathbf{Z}_t$ is assumed to be known, the state vector, $\boldsymbol{\alpha}_t$, is an $(m \times 1)$ vector of latent variables and $\boldsymbol{\varepsilon}_t$ is a $(p \times 1)$ vector of disturbances that are serially uncorrelated with a mean of $\mathbf{0}$ and an unknown $(p \times p)$ covariance matrix $\mathbf{H}^{-1}$. The state equation is defined as

$$\boldsymbol{\alpha}_{t+1} = \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{G}_t \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N\left(0, \mathbf{Q}^{-1}\right), \tag{2}$$

for $t = 1, 2, \ldots, n-1$, where $\mathbf{T}_t$ is a $(p \times p)$ transition matrix that is assumed to contain both known and unknown parameters. The known parameters are denoted by $\boldsymbol{\theta}$. Further, $\mathbf{G}_t$ is an $(m \times r)$ selection matrix that is assumed to be known and $\boldsymbol{\eta}_t$ is an $(r \times 1)$ vector of normally distributed disturbances that are serially uncorrelated with a mean of $\mathbf{0}$ and an unknown covariance matrix $\mathbf{Q}^{-1}$. Furthermore, the disturbance vectors $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ are assumed to be mutually uncorrelated. For notational convenience define $\boldsymbol{\Psi} = \{\boldsymbol{\theta}, \mathbf{Q}\}$.

The initial state, $\boldsymbol{\alpha}_1$, is assumed to be normally distributed such that

$$\boldsymbol{\alpha}_1 | \boldsymbol{\Psi} \sim N\left(\mathbf{a}_1, \mathbf{P}_1\right), \tag{3}$$

where $\mathbf{a}_1$ and $\mathbf{P}_1$ are both assumed to be known conditional on $\boldsymbol{\Psi}$. Whilst diffuse initial conditions are not explicitly dealt with in this paper, the relevant extensions can be deduced from the algorithms presented in Koopman and Durbin (2000) and Durbin and Koopman (2002).

For the proceeding sections, it is convenient to define, for $s < t$, $\mathbf{y}_{s:t} = \left(\mathbf{y}_s', \mathbf{y}_{s+1}', \ldots, \mathbf{y}_t'\right)'$ and $\mathbf{y} = \mathbf{y}_{1:n}$. Throughout, this notation extends to any vector for example $\boldsymbol{\alpha}_{s:t} = \left(\boldsymbol{\alpha}_s', \boldsymbol{\alpha}_{s+1}', \ldots, \boldsymbol{\alpha}_t'\right)'$ and $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{1:n}$.

The form of the SSM in (1) and (2) has been used in a wide range of Bayesian and non Bayesian applications including Fernandez and Harvey (1990), Koopman and Lucas (2005) and Harvey *et al.* (2007), amongst others.

3 Bayesian Estimation

Bayesian inference summarises uncertainty about the unknown parameters of interest through the joint posterior density function. For the multivariate SSM in (1), (2) and (3), the joint posterior density is defined by

$$p\left(\boldsymbol{\alpha}, \mathbf{H}, \boldsymbol{\Psi}|\mathbf{y}\right) \propto p\left(\mathbf{y}|\boldsymbol{\alpha}, \mathbf{H}\right) p\left(\boldsymbol{\alpha}|\boldsymbol{\Psi}\right) p\left(\mathbf{H}\right) p\left(\boldsymbol{\Psi}\right), \tag{4}$$

where $p\left(\mathbf{y}|\boldsymbol{\alpha}, \mathbf{H}\right)$ denotes the joint probability density function (pdf) of $\mathbf{y}$ conditional on $\boldsymbol{\alpha}$ and $\mathbf{H}$, $p\left(\boldsymbol{\alpha}|\boldsymbol{\Psi}\right)$ is the joint pdf of $\boldsymbol{\alpha}$ conditional on $\boldsymbol{\Psi}$ and $p\left(\mathbf{H}\right)$ and $p\left(\boldsymbol{\Psi}\right)$ are respectively the prior pdfs for $\mathbf{H}$ and $\boldsymbol{\Psi}$. Throughout, $\mathbf{H}$ and $\boldsymbol{\Psi}$ are assumed to be *a priori* independent. Given the specification of the measurement equation in (1), it is evident that

$$p\left(\mathbf{y}|\boldsymbol{\alpha}, \mathbf{H}\right) \propto \exp\left\{-\frac{1}{2}\sum_{t=1}^{n}\left(\mathbf{y}_t - \mathbf{Z}_t\boldsymbol{\alpha}_t\right)'\mathbf{H}\left(\mathbf{y}_t - \mathbf{Z}_t\boldsymbol{\alpha}_t\right)\right\}. \tag{5}$$

From the definition of the state equation in (2) and the initial state in (3), it follows that

$$p\left(\boldsymbol{\alpha}|\boldsymbol{\Psi}\right) = p\left(\boldsymbol{\alpha}_1|\boldsymbol{\Psi}\right)\prod_{i=1}^{n-1} p\left(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t, \boldsymbol{\Psi}\right), \tag{6}$$

where

$$p\left(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t, \boldsymbol{\Psi}\right) \propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\alpha}_{t+1} - \mathbf{T}_t\boldsymbol{\alpha}_t\right)'\mathbf{Q}\left(\alpha_{t+1} - \mathbf{T}_t\boldsymbol{\alpha}_t\right)\right\}, \tag{7}$$

for $t = 1, 2, \ldots, n-1$.

3.1 A General Gibbs Sampling Scheme

Bayesian estimation of the linear Gaussian SSM is most commonly undertaken using MCMC methods based on the Gibbs sampler. A general sampling scheme for the estimation of the posterior distribution in (4) is defined at iteration j as follows:

1. Sample $\mathbf{H}^j$ from $p\left(\mathbf{H}|\mathbf{y}, \boldsymbol{\alpha}^{j-1}, \boldsymbol{\Psi}^{j-1}\right)$.
2. Sample $\boldsymbol{\Psi}^j$ from $p\left(\boldsymbol{\Psi}|\mathbf{y}, \boldsymbol{\alpha}^{j-1}, \mathbf{H}^j\right)$.
3. Sample $\boldsymbol{\alpha}^j$ from $p\left(\boldsymbol{\alpha}|\mathbf{y}, \mathbf{H}^j, \boldsymbol{\Psi}^j\right)$.

The conditional posteriors for $\mathbf{H}$ and $\boldsymbol{\alpha}$, including the algorithms required to draw from each conditional, are specified in Sections 3.1.1 and 3.1.2, respectively. The sampling scheme for $\boldsymbol{\Psi}$ is model specific and subsequently defined in Section 4 in the context of the empirical study.

3.1.1 Sampling H

Given (4), the full conditional posterior distribution for $\mathbf{H}$ is given by

$$p\left(\mathbf{H}|\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\Psi}\right) \propto p\left(\mathbf{y}|\boldsymbol{\alpha},\mathbf{H}\right) p\left(\mathbf{H}\right),$$

where $p\left(\mathbf{y}|\boldsymbol{\alpha},\mathbf{H}\right)$ is specified in (5). The conjugacy of the Wishart distribution makes it a natural specification for the prior on $\mathbf{H}$, such that

$$p\left(\mathbf{H}\right) \sim W\left(\underline{\mathbf{V}}_H^{-1},\underline{\nu}_H\right), \tag{8}$$

where the hyperparameters $\underline{\mathbf{V}}_H^{-1}$ and $\underline{\nu}_H$ are a scale matrix and degrees of freedom parameter, respectively. Whilst the conjugate form of the Wishart prior makes it attractive, it has been noted by many in the literature to be too restrictive and various alternative priors have been pursued; see Koop(2003) and Gelman *et al.* (2004) for details. For the purposes of this paper, attention is restricted to a flat prior over the positive domain of $\mathbb{R}^p$. This corresponds to setting $\underline{\mathbf{V}}_H^{-1} = \mathbf{0}$ and $\underline{\nu}_H = -\left(p+1\right)$ in (8). This particular prior specification is employed following the justification given in Harvey *et al.* (2007).

Given the prior distributional assumption in (8), the full conditional posterior distribution for $\mathbf{H}$ has a Wishart form, such that

$$p\left(\mathbf{H}|\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\Psi}\right) \sim W\left(\overline{\mathbf{V}}_H^{-1},\overline{\nu}_H\right), \tag{9}$$

where

$$\overline{\mathbf{V}}_H^{-1} = \underline{\mathbf{V}}_H + \sum_{t=1}^{n}\left(\mathbf{y}_t - \mathbf{Z}_t\boldsymbol{\alpha}_t\right)\left(\mathbf{y}_t - \mathbf{Z}_t\boldsymbol{\alpha}_t\right)'$$

and

$$\overline{\nu} = \underline{\nu} + T.$$

Odell and Feiveson (1966) present an efficient algorithm to sample from the Wishart distribution. This algorithm is concisely summarised in Everson and Morris (2000) and is used here to sample from (9). The algorithm comprises the following steps:

1. Using the Cholesky decomposition, calculate an upper triangular matrix $\mathbf{C}$ that satisfies $\overline{\mathbf{V}}_H^{-1} = \mathbf{C}'\mathbf{C}$.

2. Construct an upper triangular matrix $\mathbf{U}$, where

$$\mathbf{U} = \begin{bmatrix} \sqrt{X_1} & N_{1,2} & \cdots & N_{1,p} \\ 0 & \sqrt{X_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & N_{p-1,p} \\ 0 & \cdots & 0 & \sqrt{X_p} \end{bmatrix},$$

with

$$N_{ij} \sim i.i.d.N\left(0,1\right), \qquad i = 1,2,\ldots,p, j = 1,2,\ldots,p, i < j$$

and

$$X_i \sim \chi^2\left(\overline{\nu} - i + 1\right), \qquad i = 1,2,\ldots,p.$$

3. Calculate the upper triangular matrix $\mathbf{R}$, where

$$\mathbf{R} = \mathbf{CU}. \tag{10}$$

4. Take $\mathbf{H}^j$ as $\mathbf{R}'\mathbf{R} \sim W\left(\overline{\mathbf{V}}_H^{-1}, \overline{\nu}_H\right)$.

3.1.2 Sampling $\boldsymbol{\alpha}$

The standard approach for sampling the state vector, $\boldsymbol{\alpha}$, from its full conditional posterior distribution for a linear Gaussian state space model is to use a simulation smoothing algorithm. Früwirth-Schnatter(1994), Carter and Kohn (1994), de Jong and Shephard (1995) and Durbin and Koopman (2002) provide alternative simulation smoothing algorithms. Of the aforementioned algorithms, the method of Durbin and Koopman is the simplest and in general the most computationally efficient; see Durbin and Koopman for further details.

This section describes a method that achieves substantial gains in computational efficiency over the standard implementation of simulation smoothing that was presented in Durbin and Koopman (2002). In particular, the simulation smoothing algorithm of Durbin and Koopman is modified by using the univariate filtering algorithms in Anderson and Moore (1979) and Koopman

and Durbin (2000) to calculate the necessary moments, instead of using the standard Kalman filter and smoothing algorithms.

The proposed approach makes use of the univariate representation of the SSM. The univariate representation requires the error term in the measurement equation to be contemporaneously uncorrelated, motivating the following transformation that makes use of the upper triangular matrix $\mathbf{R}$ in (10), such that

$$\mathbf{y}_t^* = \mathbf{Z}_t^* \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t^*, \qquad \boldsymbol{\varepsilon}_t^* \sim N(0, \mathbf{I}), \tag{11}$$

where $\mathbf{y}_t^* = \mathbf{R}\mathbf{y}_t$, $\mathbf{Z}_t^* = \mathbf{R}\mathbf{Z}_t$ and $\boldsymbol{\varepsilon}_t^* = \mathbf{R}\boldsymbol{\varepsilon}_t$. The transformed measurement equation in (11) has the following univariate representation.

$$y_{t,i}^* = \mathbf{z}_{t,i}^{*\prime} \boldsymbol{\alpha}_t + \varepsilon_{t,i}^*, \qquad t = 1, 2, \ldots, n, i = 1, 2, \ldots p, \tag{12}$$

where $y_{t,i}^*$ is the i^{th} element of $\mathbf{y}_t^*$, $\mathbf{z}_{t,i}^{*\prime}$ is the i^{th} row of $\mathbf{Z}_t^*$ and $\varepsilon_{t,i}^*$ is the i^{th} element in $\boldsymbol{\varepsilon}_t^*$. Note that the formulation of the state equation remains unchanged from (2).

Given the univariate representation of the multivariate SSM in (11), (2) and the initial state in (3), the filtering equations are initialised by setting $\mathbf{a}_{11} = \mathbf{a}_1$ and $\mathbf{P}_{1,1} = \mathbf{P}_1$ then for $i = 1, 2, \ldots, p$ and for $t = 1, 2, \ldots, n$ are defined by

$$\begin{aligned}
\nu_{t,i} &= y_{t,i}^* - \mathbf{z}_{t,i}^{*\prime} \mathbf{a}_{t,i} \\
\mathbf{m}_{t,i} &= \mathbf{P}_{t,i} \mathbf{z}_{t,i}^* \\
f_{t,i} &= \mathbf{z}_{t,i}^{*\prime} \mathbf{m}_{t,i} + 1 \\
\mathbf{k}_{t,i} &= \mathbf{m}_{t,i} / f_{t,i} \\
\mathbf{a}_{t,i+1} &= \mathbf{a}_{t,i} + \mathbf{k}_{t,i} \nu_{t,i} \\
\mathbf{P}_{t,i+1} &= \mathbf{P}_{t,i} - \mathbf{k}_{t,i} \mathbf{m}_{t,i}^{\prime},
\end{aligned} \tag{13}$$

where $\nu_{t,i}$ is a scalar, $\mathbf{m}_{t,i}$ is an $(m \times 1)$ vector, $f_{t,i}$ is a scalar, $\mathbf{k}_{t,i}$ is an $(m \times 1)$ vector, $\mathbf{a}_{t,i}$ is an $(m \times 1)$ vector and $\mathbf{P}_{t,i}$ is an $(m \times m)$ matrix. Further, $\mathbf{a}_{t,i} = E(\boldsymbol{\alpha}_t | \mathbf{y}_{1:t-1}, y_{t,1}, y_{t,2}, \ldots, y_{t,i-1})$ and $\mathbf{P}_{t,i} = Var(\boldsymbol{\alpha}_t | \mathbf{y}_{1:t-1}, y_{t,1}, y_{t,2}, \ldots, y_{t,i-1})$. To obtain $\mathbf{a}_{t+1,1} = E(\boldsymbol{\alpha}_{t+1} | \mathbf{y}_{1:t})$ and $\mathbf{P}_{t+1,1} = Var(\boldsymbol{\alpha}_{t+1} | \mathbf{y}_{1:t})$ for $t = 1, 2, \ldots, n$ calculate

$$\begin{aligned}
\mathbf{a}_{t+1,1} &= \mathbf{T}_t \mathbf{a}_{t,p+1} \\
\mathbf{P}_{t+1,1} &= \mathbf{T}_t \mathbf{P}_{t,p+1} \mathbf{T}_t^{\prime} + \mathbf{G}_t \mathbf{Q}^{-1} \mathbf{G}_t^{\prime}.
\end{aligned} \tag{14}$$

The corresponding smoothing equations require $\nu_{t,i}$, $f_{t,i}$, $\mathbf{k}_{t,i}, \mathbf{a}_{t,1}$ and $\mathbf{P}_{t,1}$ to be stored during the filtering pass. For the multivariate SSM in (12),

(2) and (3), the smoothing equations are defined for $i = p, p-1, \ldots, 1$ and $t = n, n-1, \ldots, 1$ as

$$\begin{aligned} \mathbf{L}_{t,i} &= \mathbf{I}_m - \mathbf{k}_{t,i}\mathbf{z}_{t,i}^{*\prime} \\ \mathbf{r}_{t,i} &= \mathbf{z}_{t,i}/f_{t,i}\nu_{t,i} + \mathbf{L}_{t,i}^{\prime}\mathbf{r}_{t,i}, \end{aligned} \tag{15}$$

where $\mathbf{L}_{t,i}$ is an $(m \times m)$ matrix and $\mathbf{r}_{t,i}$ is an $(m \times 1)$ vector. The transition through time is given by

$$\begin{aligned} \widehat{\mathbf{a}}_t &= \mathbf{a}_{t,1} + \mathbf{P}_{t,1}\mathbf{r}_{t,0} \\ \mathbf{r}_{t-1,p} &= \mathbf{T}_{t-1}^{\prime}\mathbf{r}_{t,0}, \end{aligned} \tag{16}$$

where $\widehat{\mathbf{a}}_t = E\left(\alpha_t|\mathbf{y}^*\right)$ for $t = n, n-1, \ldots, 1$. The smoothing equations are initialised with $\mathbf{r}_{n,p} = \mathbf{0}$. For more details on the univariate representation, including how to modify the equations for diffuse or partially diffuse initial conditions, refer to Koopman and Durbin (2000) and Durbin and Koopman (2001). Given the filtering and smoothing equations in (13), (14), (15) and (16), the simulation smoothing algorithm of Durbin and Koopman (2002) can be modified to draw $\boldsymbol{\alpha}$ from its conditional distribution at iteration j of the Gibbs scheme in Section 3.1 as follows:

1. Sample $\boldsymbol{\varepsilon}_t^{*+}$ from $p\left(\boldsymbol{\varepsilon}_t^*|\mathbf{H}^j\right)$ and sample $\boldsymbol{\eta}_t^+$ from $p\left(\boldsymbol{\eta}_t|\mathbf{Q}^j\right)$ for $t = 1, 2, \ldots, n$.

2. Draw $\boldsymbol{\alpha}_1^+$ from $p\left(\boldsymbol{\alpha}_1|\boldsymbol{\Psi}^j\right)$ and then generate $\boldsymbol{\alpha}_{2:n}^+$ and $\mathbf{y}^{*+}$, using (11) and the state equation in (2).

3. Calculate $\widetilde{\mathbf{y}} = \mathbf{y}^{**} - \mathbf{y}^+$.

4. Use the filter and smoother in (13),(14), (15) and (16) substituting $\widetilde{\mathbf{y}}_{t,i}$ for $y_{t,i}^{**}$ to calculate $\widehat{\boldsymbol{\alpha}}_{1:n} = E\left(\boldsymbol{\alpha}|\widetilde{\mathbf{y}}\right)$.

5. Take $\boldsymbol{\alpha}^j = \widehat{\boldsymbol{\alpha}} + \boldsymbol{\alpha}^+$ as a draw from $p\left(\boldsymbol{\alpha}|\mathbf{y}, \mathbf{H}^j, \boldsymbol{\Psi}^j\right)$.

Computational Comparison This section contains a comparison of the computational efficiency of the standard simulation smoothing algorithm of Durbin and Koopman (2002) and the univariate filtering and simulation smoothing approach presented in Section 3.1.2. In the comparison the initial cost of transforming the model for the univariate approach and the cost of

inverting $\mathbf{H}^{-1}$ for the standard approach is ignored.[1] For the comparison the standard Kalman filter is formulated in an efficient fashion comparable to that of the univariate filtering algorithm in (13) and (14). For the standard simulation smoothing approach it is further assumed that $\mathbf{F}_t^{-1}$ is stored for the smoothing algorithm, rather than storing $\mathbf{F}_t$. It is also assumed that all inversions are calculated using the Cholesky decomposition. Following Koopman and Durbin (2000) calculations involving $\mathbf{T}_t$ are not considered as the transition matrix typically has a sparse structure with most elements equal to ones and zeros. Furthermore, calculations involving $\mathbf{G}_t$ are also not included as it typically has a sparse structure made up of ones and zeros and is often simply the identity matrix.[2]

Table 1: presents the percentage savings associated with the univariate filtering and simulation smoothing approach over that of the standard Kalman filtering and simulation smoothing approach. The top row reports the number of timeseries, p, and the first column presents the dimension of the state, m. The centre of the table contains the percentage savings in multiplications associated with the univariate approach.

Percentage Savings for the Univariate Filtering and Simulation Smoothing

	$p=1$	$p=2$	$p=5$	$p=10$	$p=15$	$p=20$	$p=25$
$m=1$	15	52	83	94	97	98	99
$m=2$	5	29	75	86	93	95	97
$m=5$	1	5	31	61	76	84	88
$m=10$	0	-4	8	33	51	63	72
$m=15$	0	-7	-2	17	34	47	56
$m=20$	0	-8	-7	7	22	34	44
$m=25$	0	-9	-10	1	13	25	34

In Table 1 the percentage savings of the univariate approach to filtering

[1] It is worth noting that in general the initial transform for the univariate approach is far cheaper than the inverse for the standard Kalman filtering approach.

[2] Note that in cases that $\mathbf{T}_t$ or $\mathbf{G}_t$ contains elements that aren't ones or zeros then the same number of additional multiplications are required by both filters.

and simulation smoothing presented in Section 3.1.2 over that of the standard simulation smoother approach are presented. It is interesting that a small gain in computational efficiency is made when $p = 1$. Examination of both approaches reveals that the gain is made in the simulation smoothing algorithm in sampling $\boldsymbol{\varepsilon}_t^+$ as the transformed model, for $p = 1$, requires one less multiplication at each time period, t. Given that the filtering and smoothing algorithms are so cheap when $p = 1$, and in particular when m is also equal to one, even this small saving is apparent. It is also evident that when m is much larger than p the standard approach is slightly more computationally efficient. The small reduction in computational efficiency arrises as the smoothing recursions that are required to calculate the moments of interest are more expensive that the standard equations when m is much greater than p. Importantly, for the majority of cases, the univariate simulation smoothing approach is substantially more efficient that the standard approach. In particular, when p is larger than m the gains in computational efficiency are substantial. This is arguably an important case in practice as the underlying dynamics are often governed by one or a few common underlying components. Further, it is particularly promising that the gains of the new approach are most substantial when p is large and that is when the computationally costs are particularly great and efficiency is most important.

3.2 Integration Sampler

The likelihood for the state space model in (1), (2) and (3) is defined using

$$p\left(\mathbf{y}|\mathbf{H},\boldsymbol{\Psi}\right) = \int p\left(\mathbf{y}|\boldsymbol{\alpha},\mathbf{H}\right) p\left(\boldsymbol{\alpha}|\boldsymbol{\Psi}\right) d\boldsymbol{\alpha}, \tag{17}$$

where $p\left(\mathbf{y}|\boldsymbol{\alpha},\mathbf{H}\right)$ is defined in (5) and $p\left(\boldsymbol{\alpha}|\boldsymbol{\Psi}\right)$ is defined by (6) and (7). Koopman and Durbin (2000) show that the likelihood can be calculated analytically, using (13) and (14), as

$$p\left(\mathbf{y}|\mathbf{H},\boldsymbol{\Psi}\right) \propto -0.5\left(\sum_{t=1}^{n}\sum_{i=1}^{p}\ln F_{t,i} + \nu_{t,i}^2 F_{t,i}^{-1}\right). \tag{18}$$

Using (18), the posterior for any or all of the parameters of interest can be expressed up until an integrating constant marginally of the state vector $\boldsymbol{\alpha}$. For example,

$$p\left(\mathbf{H},\boldsymbol{\Psi}|\mathbf{y}\right) \propto p\left(\mathbf{y}|\mathbf{H},\boldsymbol{\Psi}\right) p\left(\mathbf{H}\right) p\left(\boldsymbol{\Psi}\right). \tag{19}$$

However, it is typically difficult to sample $\mathbf{H}$ and $\boldsymbol{\Psi}$ directly from (19), whereas conditional on the state vector the posterior densities for the parameters of interest often have a closed form. This issue is discussed in Pitt and Shephard (1999) and it provides motivation to use Gibbs based sampling schemes like the one described in Section 3.1. However, as part of a Gibbs scheme, an analytical form for the likelihood may still be useful. In particular, even when conditioning on $\boldsymbol{\alpha}$ certain parameters may not be easy to sample from, in which case an algorithm such as the random walk Metropolis Hastings (RWMH) is often used; for details on the RWMH and other more advanced algorithms suitable for such problems see Robert and Cassella (1999) and Lui (2001). As (18) allows us to define the posterior for any parameter marginally of the state vector, we can always use the RWMH to sample indirectly from the resultant density. If there is a high degree of correlation in the Gibbs scheme between this particular parameter and the state then this approach will lead to gains in the simulation efficiency of the MCMC sampling scheme. If the correlation in the Gibbs scheme is particularly high between certain parameters and the state vector, sampling these parameters of interest marginally of the state vector may be the preferred approach using a RWMH type algorithm, even if a closed form is available when conditioning on the state. This was in fact the case for Kim, Shephard and Chib (1998) who use this idea for their stochastic volatility algorithm. They, however, calculate the likelihood using the augmented Kalman filter of de Jong (1991). The algorithm of de Jong is more complicated and is computationally expensive in comparison to the proposed algorithm, however it can provide an exact solution in the case when diffuse initial conditions are specified. Given the focus on Bayesian methods in this paper, this is arguably not that important as it is usually possible in practice to specify a sensible informative prior, which can be vague or strict depending on the available information. If it is necessary to produce an exact solution for diffuse initial conditions then the filtering algorithms in this paper can be extended using the exact initial filtering approach, as described by Koopman and Durbin (2000).

In Table 2 the percentage savings in multiplications made using the univariate filter in (13) and (14) to calculate the likelihood over that of the Kalman filter are reported. As in Table 1, it is assumed that all inversions are calculated using the Cholesky decomposition and the calculations involving $\mathbf{T}_t$ and $\mathbf{G}_t$ are not considered, because both matrices typically have a sparse structure with most elements equal to ones and zeros. It is evident

Table 2: presents the percentage savings that are made when using the univariate filter to calculate the likelihood instead of the Kalman filter. The top row reports the number of timeseries, p, and the first column presents the dimension of the state, m. The centre of the table contains the percentage savings in multiplications associated with the univariate approach.

Percentage Savings for Likelihood Calculation Using the Univariate Filter

	$p=1$	$p=2$	$p=5$	$p=10$	$p=15$	$p=20$	$p=25$
$m=1$	0	59	88	96	98	99	99
$m=2$	0	40	76	91	95	97	98
$m=5$	0	19	51	76	86	91	93
$m=10$	0	10	32	56	70	78	84
$m=15$	0	7	23	44	58	67	74
$m=20$	0	5	18	36	49	59	66
$m=25$	0	4	15	30	42	52	59

that a large reduction in the computational burden in calculating (18) is made when p is greater than one.

Denote $\boldsymbol{\Psi}_i$ as the parameters of $\boldsymbol{\Psi}$ that are to be sampled marginally of $\boldsymbol{\alpha}$, and denote $\boldsymbol{\Psi}_{\backslash i}$ to be the remaining parameters in $\boldsymbol{\Psi}$, then an alternative Gibbs based sampling scheme, to the one presented in Section 3.1, for iteration j is defined as follows:

1. Sample $\mathbf{H}^j$ from $p\left(\mathbf{H}|\mathbf{y}, \boldsymbol{\alpha}^{j-1}, \boldsymbol{\Psi}^{j-1}\right)$.
2. Sample $\boldsymbol{\Psi}^j_{\backslash i}$ from $p\left(\boldsymbol{\Psi}_{\backslash i}|\mathbf{y}, \mathbf{H}^j, \boldsymbol{\Psi}^{j-1}_i, \boldsymbol{\alpha}^{j-1}\right)$.
3. Sample $\boldsymbol{\Psi}^j_i$ and $\boldsymbol{\alpha}^j$ jointly from $p\left(\boldsymbol{\Psi}_i, \boldsymbol{\alpha}|\mathbf{y}, \mathbf{H}^j, \boldsymbol{\Psi}^j_{\backslash i}\right)$.

As in the Gibbs scheme described in Section 3.1, the sampling of $\mathbf{H}$ is described in Section 3.1.1. Sampling $\boldsymbol{\Psi}^j_{\backslash i}$ is model specific and is defined in the empirical study in Section 4. Sampling $\boldsymbol{\Psi}^j_i$ and $\boldsymbol{\alpha}^j$ is completed in two steps. The first step is to sample $\boldsymbol{\Psi}^j_i$ from

$$p\left(\boldsymbol{\Psi}_i|\mathbf{y}, \mathbf{H}^j, \boldsymbol{\Psi}^{j-1}_{\backslash i}\right),$$

where

$$p\left(\boldsymbol{\Psi}_i|\mathbf{y},\mathbf{H},\boldsymbol{\Psi}_{\setminus i}\right) \propto p\left(\mathbf{y}|\mathbf{H},\boldsymbol{\Psi}\right)p\left(\boldsymbol{\Psi}\right)$$

and $p\left(\mathbf{y}|\mathbf{H},\boldsymbol{\Psi}\right)$ is calculated using (18). One possibility is to use the RWMH. The second step is to sample $\boldsymbol{\alpha}^j$ from $p\left(\boldsymbol{\alpha}|\mathbf{y},\mathbf{H}^j,\boldsymbol{\Psi}^j\right)$ as described in Section 3.1.2.

4 Empirical Illustration

4.1 Data Description

The Bayesian methodology for estimating multivariate state space models is used to analyse MODIS satellite image data. In particular, a multivariate time series dataset, of NDVI from the 250m pixel MOD13Q1 product, is constructed from the image data. The NDVI provides a measure of the amount of live vegetation, whereby a higher reading signifies a higher level of live vegetation.[3] The region of interest is a grazing property located in Queensland, Australia. The dataset is comprised of 25 timeseries, with 171 observations that occur every 16 days extending from February, 2000 to August, 2007. Each time series maps a pixel of the image through time.

4.2 Structural Time Series Model

For the analysis, a STSM is constructed that incorporates both a trend and a seasonal component. The STSM provides a mapping between the dynamic components and the $(p \times 1)$ vector of observations, $\mathbf{y}_t$, through the following equation

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\psi}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim N\left(\mathbf{0},\mathbf{H}^{-1}\right), \tag{20}$$

for $t = 1, 2, \ldots, n$, $\boldsymbol{\mu}_t$ is a $(p \times 1)$ vector of trends, $\boldsymbol{\psi}_t$ is a $(p \times 1)$ vector of stochastic cycles and $\boldsymbol{\varepsilon}_t$ is a $(p \times 1)$ vector of disturbances that is serially uncorrelated and has a mean of $\mathbf{0}$ and a covariance matrix $\mathbf{H}^{-1}$. Given that the timeseries are constructed using pixels from a localised region with similar soil types, the same land management and common weather patterns, it is assumed that the dynamics of the model are governed by a common underlying process. Consequently, it is assumed that $\mu_{t,1} = \mu_{t,2} = \cdots = \mu_{t,p}$,

[3]Note that the MODIS scaled NDVI ranges from 0-10000. For the analysis it is scaled down by a factor of 1000.

and $\psi_{t,1} = \psi_{t,2} = \cdots = \psi_{t,p}$ where $\mu_{t,i}$ is the i^{th} trend component at time t and $\psi_{t,i}$ represents the i^{th} stochastic cycle.

The trend component is assumed to evolve according to

$$\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t + \boldsymbol{\zeta}_t, \tag{21}$$

where the $(p \times 1)$ vector of disturbances $\boldsymbol{\zeta}_t$ is defined such that

$$\boldsymbol{\zeta}_t = \mathbf{i} \otimes \zeta_t,$$

where $\mathbf{i}$ is a $(p \times 1)$ vector of ones, $\otimes$ denotes the Kronecker product and $\zeta_t \sim N\left(0, \sigma_\zeta^2\right)$. The stochastic cycle is assumed to evolve according to

$$\begin{bmatrix} \boldsymbol{\psi}_{t+1} \\ \boldsymbol{\psi}^*_{t+1} \end{bmatrix} = \boldsymbol{\Gamma} \begin{bmatrix} \boldsymbol{\psi}_t \\ \boldsymbol{\psi}^*_t \end{bmatrix} + \boldsymbol{\Omega}_t, \tag{22}$$

where $\boldsymbol{\psi}^*_{t+1}$ is a $(p \times 1)$ vector of auxiliary variables, $\boldsymbol{\Omega}_t = \mathbf{i} \otimes \boldsymbol{\kappa}_t$, with $\boldsymbol{\kappa}_t \sim N\left(0, \sigma_\kappa^2 \mathbf{I}_2\right)$ and $\boldsymbol{\Gamma} = \mathbf{I}_p \otimes \boldsymbol{\Gamma}_\psi$, with

$$\boldsymbol{\Gamma}_\psi = \rho \begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix},$$

where $\mathbf{I}_s$ generically denotes the identity matrix of order s. It is assumed that the cycle is stationary implying that $|\rho| < 1$ and λ denotes the frequency of the cycle in radians.

The STSM in (20), (21) and (22) can be compactly expressed using the state space form in (1) and (2), with

$$\begin{aligned} \mathbf{Z}_t &= \mathbf{i} \otimes \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}, \quad \boldsymbol{\alpha}_t = \begin{bmatrix} \mu_t & \psi_t & \psi^*_t \end{bmatrix}', \\ \mathbf{T}_t &= \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Gamma}_\psi \end{bmatrix}, \quad \mathbf{G}_t = \mathbf{I}_3, \\ \mathbf{Q}^{-1} &= \begin{bmatrix} \sigma_\zeta^2 & 0 & 0 \\ 0 & \sigma_\kappa^2 & 0 \\ 0 & 0 & \sigma_\kappa^2 \end{bmatrix}. \end{aligned} \tag{23}$$

The SSM is completed using prior expectations about the state space process that are specified through (3). The trend is initialised with a mean of five, which centres it on the range of the NDVI. The variance is set equal to nine to ensure positive prior probability over the range of the data. To complete

the specification of the initial state the marginal distribution of the stochastic cycle is used, such that

$$\mathbf{a}_1 = \begin{bmatrix} 5 & 0 & 0 \end{bmatrix}' \text{ and } \mathbf{P}_1 = \begin{bmatrix} 9 & \mathbf{0}' \\ \mathbf{0} & \frac{\sigma_k^2}{1-\rho^2}\mathbf{I}_2 \end{bmatrix}.$$

For the empirical analysis a flat prior is specified over the $(-1, 1)$ interval for ρ. For the parameter λ a flat prior is specified over the $\left(\frac{\pi}{18}, \frac{3\pi}{18}\right)$ interval. This implies a prior mean of $\frac{2\pi}{18}$, which corresponds to a period of approximately one year. For both σ_ζ and σ_κ a flat prior with positive support is specified following the rationale in Harvey *et al.* (2007).

4.3 Estimation

Whilst the sampling schemes for $\mathbf{H}$ and $\boldsymbol{\alpha}$ are described in Sections 3.1.1 and 3.1.2 respectively, the sampling scheme for $\boldsymbol{\Psi}$ still needs to be defined. For the model specified in Section 4.2 sampling from $\boldsymbol{\Psi}$ equates to sampling the parameters of the trend and stochastic cycle. That is sampling ρ, λ, σ_ζ and σ_κ. One possible approach is to follow Harvey *et al.* (2007), in which $\boldsymbol{\Psi}$ is sampled conditional on $\boldsymbol{\alpha}$. An alternative approach is proposed here, which fits into the Gibbs scheme presented in Section 3.2. Both methods are described below and a comparison is given in the analysis in Section 4.4.

4.3.1 Sampling ρ, λ σ_ζ and σ_κ

Harvey *et al.* (2007) sample ρ, λ, σ_ζ and σ_κ from

$$p\left(\rho, \lambda, \sigma_\zeta, \sigma_\kappa | \mathbf{y}, \boldsymbol{\alpha}, \mathbf{H}\right) \propto p\left(\boldsymbol{\alpha} | \rho, \lambda, \sigma_\zeta, \sigma_\kappa\right) p\left(\rho\right) p\left(\lambda\right) p\left(\sigma_\zeta\right) p\left(\sigma_\kappa\right),$$

where $p\left(\boldsymbol{\alpha} | \rho, \lambda, \sigma_\zeta, \sigma_\kappa\right) \equiv p\left(\boldsymbol{\alpha} | \boldsymbol{\Psi}\right)$ is defined in (6) and the pdfs $p\left(\rho\right), p\left(\lambda\right), p\left(\sigma_\zeta\right)$ and $p\left(\sigma_\kappa\right)$ were informally defined in Section 4.2. Harvey *et al.* suggest a Gibbs scheme that is defined at iteration j as follows:

1. Sample ρ^j from $p\left(\rho | \mathbf{y}, \mathbf{H}^j, \boldsymbol{\alpha}^{j-1}, \lambda^{j-1}, \sigma_\zeta^{j-1}, \sigma_\kappa^{j-1}\right)$.
2. Sample λ^j from $p\left(\lambda | \mathbf{y}, \mathbf{H}^j, \boldsymbol{\alpha}^{j-1}, \rho^j, \sigma_\zeta^{j-1}, \sigma_\kappa^{j-1}\right)$.
3. Sample σ_ζ^j from $p\left(\sigma_\zeta | \mathbf{y}, \mathbf{H}^j, \boldsymbol{\alpha}^{j-1}, \rho^{j-1}, \lambda^{j-1}, \sigma_\kappa^{j-1}\right)$.
4. Sample σ_κ^j from $p\left(\sigma_\kappa | \mathbf{y}, \mathbf{H}^j, \boldsymbol{\alpha}^{j-1}, \rho^j, \lambda^j, \sigma_\zeta^j\right)$.

Harvey *et al.* suggest sampling both ρ and λ separately using the RWMH algorithm. They suggest sampling both σ_ζ and σ_κ separately directly from the inverted gamma distribution.

An alternative sampling scheme, along the lines of the generic Gibbs sampling scheme in Section 3.2, is to sample σ_ζ, σ_κ and $\boldsymbol{\alpha}$ jointly from

$$p\left(\sigma_\zeta, \sigma_\kappa, \boldsymbol{\alpha}|\mathbf{y}, \mathbf{H}, \rho, \lambda\right) \propto p\left(\mathbf{y}|\boldsymbol{\alpha}, \mathbf{H}\right) p\left(\boldsymbol{\alpha}|\rho, \lambda, \sigma_\zeta, \sigma_\kappa\right) p\left(\sigma_\zeta\right) p\left(\sigma_\kappa\right), \qquad (24)$$

where $p\left(\mathbf{y}|\boldsymbol{\alpha}, \mathbf{H}\right)$ is defined in (5). A modified Gibbs sampling scheme where σ_ζ, σ_κ and $\boldsymbol{\alpha}$ are sampled jointly from (24) is defined at iteration j as follows:

1. Sample ρ^j from $p\left(\rho|\mathbf{y}, \mathbf{H}^j, \boldsymbol{\alpha}^{j-1}, \lambda^{j-1}, \sigma_\zeta^{j-1}, \sigma_\kappa^{j-1}\right)$.
2. Sample λ^j from $p\left(\lambda|\mathbf{y}, \mathbf{H}^j, \boldsymbol{\alpha}^{j-1}, \rho^j, \sigma_\zeta^{j-1}, \sigma_\kappa^{j-1}\right)$.
3. Sample σ_ζ^j and σ_κ^j jointly from $p\left(\sigma_\zeta, \sigma_\kappa|\mathbf{y}, \mathbf{H}^j, \rho^j, \lambda^j\right)$.
4. Sample $\boldsymbol{\alpha}^j$ from $p\left(\boldsymbol{\alpha}|\mathbf{y}, \mathbf{H}^j, \rho^j, \lambda^j, \sigma_\zeta^j, \sigma_\kappa^j\right)$.

The parameters ρ and λ are sampled separately, as before, using the RWMH. To sample σ_ζ and σ_κ jointly from $p\left(\sigma_\zeta, \sigma_\kappa|\mathbf{y}, \rho, \lambda\right)$ the RWMH algorithm is again implemented. To evaluate the RWMH acceptance ratio, given the relevant argument, it is noted that

$$p\left(\sigma_\zeta, \sigma_\kappa|\mathbf{y}, \mathbf{H}, \rho, \lambda\right) \propto p\left(\mathbf{y}|\mathbf{H}, \rho, \lambda, \sigma_\zeta, \sigma_\kappa\right) p\left(\sigma_\zeta\right) p\left(\sigma_\kappa\right),$$

where $p\left(\mathbf{y}|\mathbf{H}, \rho, \lambda, \sigma_\zeta, \sigma_\kappa\right) \equiv p\left(\mathbf{y}|\mathbf{H}, \boldsymbol{\Psi}\right)$ is evaluated using (18).

4.4 Analysis

In Table 3 estimation output for the MODIS data set using the MCMC sampling scheme is reported, whereby the parameters of the trend and stochastic cycle are estimated following Harvey *et al.* (2007) and the proposed sampling scheme described in Section 4.3.1. For both estimation schemes 100000 iterations of the MCMC sampler are produced and the first 10000 iterations are discarded. The table shows the marginal posterior mean, the marginal posterior standard deviation, the inefficiency factor (IF) calculated based on the sampling scheme described in Harvey *et al.* (IF1), and the IF for the sampling scheme for the stochastic cycle proposed in this paper (IF2), respectively. The IF is used as a measure of the simulation efficiency

Table 3: Estimation results for the MODIS data set. From the MCMC sampling scheme 100000 iterations are obtained, of which the first 10000 are discarded. Column 1 contains the parameters to which the statistics (reported in the subsequent columns) refer, Column 2 reports the MCMC estimates of the marginal posterior mean, Column 3 reports the marginal posterior standard deviation, Column 4 reports the inefficiency factors (IF)s for the estimates obtained using the sampling scheme Harvey et al. and Column 5 contains IFs from estimates using the sampling scheme proposed in this paper. The bottom row of the table reports the average time taken in seconds for 1000 iterations of the MCMC sampler on a Pentium Core 2 Duo with a 2.4 GHz processor and 1 GB of RAM.

Parameter	Mean	SD	IF1	IF2
σ_ζ	1.2×10^{-1}	4×10^{-2}	134	52
ρ	8.9×10^{-1}	4×10^{-2}	35	24
λ	2.9×10^{-1}	4×10^{-2}	16	17
σ_κ	2.1×10^{-1}	3×10^{-2}	99	43
Time			3	5

of MCMC samplers and it has a prominent role in the literature. See for example Chib and Greenberg (1996), Shephard and Pitt (1997), Bos and Shephard (2006) and Strickland, Martin and Forbes (2007), amongst many others. The IF is calculated as

$$IF = 1 + 2\frac{B}{B-1}\sum_{i=1}^{B} K_{QS}\left(\frac{i}{B}\right)\widehat{\rho}_i,$$

where $\widehat{\rho}_i$ is the estimate of the correlation at lag i of the MCMC iterates, K_{QS} is the Quadratic Spectral (QS) kernel defined as

$$K_{QS}(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right)$$

and B is the bandwidth, which is selected automatically following Andrews (1991). Further details on the implementation of this procedure can be found

in Strickland, Forbes and Martin (2006). The IF can be interpreted as the ratio of the simulation variance of the marginal posterior mean from an MCMC sampler to the simulation variance to that of a hypothetical sampler that has M independent draws. The average time taken for 1000 iterations of both MCMC sampling schemes is reported at the bottom of the table.

The marginal posterior mean estimates indicate that there is a large amount of persistence in the stochastic cycle and that the period of the cycle is approximately one year, which is sensible given that the cycle is used to model seasonality.

Examination of the largest IF, where the largest is compared as it provides a minimum level of accuracy, shows that the sampling scheme for the stochastic cycle in Section 4.3.1 is nearly three times more simulation efficient than the benchmark sampling scheme of Harvey *et al.*, despite being less than two times more computationally expensive. For both schemes, the savings in the computational cost of the simulation smoothing algorithm is approximately 90%. For the second sampling scheme, the use of the univariate representation also leads to a reduction in the computational cost of around 90% in the calculation of the likelihood.

Figure 1 contains plots the marginal posterior mean estimates of both the trend and the stochastic cycle. The upper plot superimposes the estimated trend over the observed data, whilst the lower plot contains the estimated stochastic cycle.

5 Conclusion

In this paper, a Bayesian estimation methodology for multivariate SSMs is introduced. In particular, the simulation smoothing algorithm of Durbin and Koopman (2002) is modified using the filtering and smoothing algorithms of the univariate representation of the SSM that is facilitated through a transformation of the SSM. The new algorithm provides substantial gains in computational efficiency over the standard simulation smoothing algorithm of Durbin and Koopman.

An illustrative empirical analysis of MODIS data is used to demonstrate the Bayesian methodology. In particular, a SSM is constructed to characterise the seasonal and trend components present in the NDVI dataset for a particular property that is located in Queensland, Australia. The SSM is specified such that the trend component is modelled as a random walk and a

This figure contains plots of the estimated marginal posterior mean for both the trend and stochastic cycle. In the upper plot the estimated trend is superimposed over the data, whilst the lower plot contains the estimated stochastic cycle.

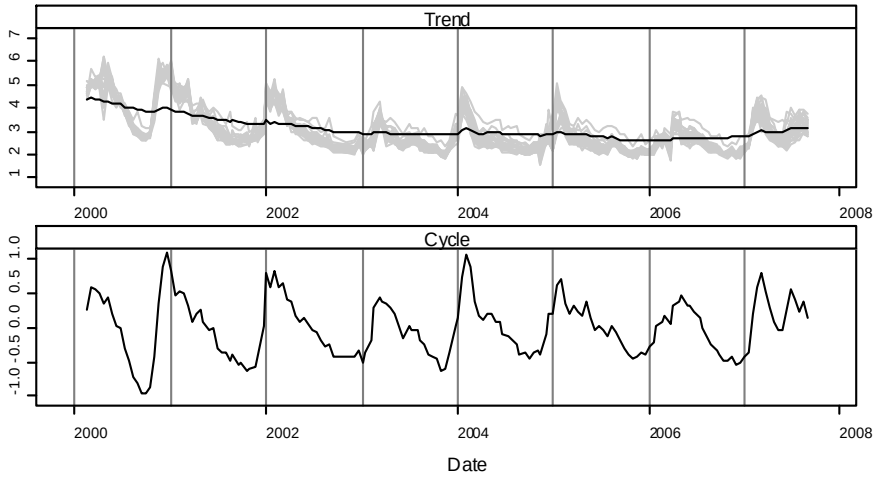

Figure 1:

stochastic cycle is incorporated to capture the dynamic seasonal component that is present in the timeseries. For the analysis, the computational burden of the proposed method is approximately 90% less than that of the standard simulation smoothing approach of Durbin and Koopman (2002). Further, a new sampling scheme is presented for the trend and stochastic cycle that is simulation efficient relative to the existing approach of Harvey *et al.* The analysis reveals that there is strong evidence that the level of NDVI is slowly decreasing through time.

An important extension to the methodology is to utilise more parsimonious methods for the covariance matrix estimation. Such an extension would allow for the analysis of higher dimensional multivariate time series and hence the importance of the computational gains presented in this paper would be even greater.

Acknowledgements

This research has been supported by an Australian Research Council Linkage Grant, LP668185. Comments from participants of the 2007 Spring

Bayes conference and the 2007 International Biometric Society Australasian Region conference are gratefully acknowledged. The numerical results in the paper are produced using C++ code that makes use of the Template Numerical Toolkit, which can be found at http://math.nist.gov/tnt/index.html. Software for the random number generator used can be found at http://www.agner.org /random/randomc.htm.

References

[1] Anderson, B. D. O and Moore, J. B. (1979), "Optimal Filtering," *Prentice Hall*, Englewood Cliffs.

[2] Andrews, D. (1991), "Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.

[3] Bos, C. and Shephard, N. (2006), "Inference for Adaptive Time Series Models: Stochastic Volatility and Conditionally Gaussian State Space Form," *Econometric Reviews*, 25, 279-244.

[4] Carter, C. and Kohn, R. (1994), "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541-553.

[5] Chib, S. and Greenberg, E. (1996), "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory*, 12, 407-431.

[6] de Jong, P. (1991), "The Diffuse Kalman Filter," *The Annals of Statistics*, 19, 1073-1083.

[7] de Jong, P. and Shephard, N. (1995), "The Simulation Smoother for Time Series Models," *Biometrika,* 82, 339-350.

[8] Durbin, J. and Koopman, S. (2001), *Time Series Analysis by State Space Methods*, Oxford University Press.

[9] Durbin, J. and Koopman, S.J. (2002), "A Simple and Efficient Simulation Smoother for Time Series Models," *Biometrika*, 81, 341-353.

[10] Fernandez, F. J. and Harvey, A. C. (1990), "Seemingly Unrelated Time Series Equations and a Test for Homogeneity," *Journal of Business and Economic Statistics*, 8, 71-81.

[11] Frühwirth-Schnatter, S. (1994), "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis*, 15, 183-202.

[12] Harvey, A. C., Trimbur, T. M. and Van Djik, H. K. (2007), "Trends and Cycles in Economic Time Series: A Bayesian Approach," *Journal of Econometrics*, 140, 618-649.

[13] Gelman, A. Carlin, H. S., Stern, H. S. and Rubin, D. B. (2004), "Bayesian Data Analysis, Second Edition," *Chapman & Hall*, London.

[14] Kim, S., Shephard, N. and Chib, S. (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361-393.

[15] Koop, G. (2003), "Bayesian Econometrics," *John Wiley & Sons*, England.

[16] Koopman, S.J.and Durbin, J. (2000), "Fast Filtering and Smoothing for Multivariate State Space Models," *Journal of Time Series Analysis*, 21, 281-96.

[17] Koopman, S. J. and Lucas, A. (2005), "Business and default cycles for credit risk," *Journal of Applied Econometrics*, 20, 311-323.

[18] Liu, J. S. (2001), "Monte Carlo Strategies in Scientific Computing," *Springer-Verlag*, New York.

[19] Pitt, M. K. and Shephard, N. (1999), "Analytic Convergence Rates and Parameterisation Issues for the Gibbs Sampler Applied to State Space Models," *Journal of Time Series Analysis*, 20, 63-85.

[20] Robert, C. P. and Casella, G. (1999), "Monte Carlo Statistical Methods," *Springer-Verlag*, New York.

[21] Shephard, N. and Pitt, M. (1997), "Likelihood Analysis of Non-Gaussian Measurement Time Series," *Biometrika*, 84, 653-667.

[22] Strickland, C. M., Forbes, C. S. and Martin, G. M. (2006), "Bayesian Analysis of the Stochastic Conditional Duration Model," *Computational Statistics and Data Analysis*, 50, 2247-2267.

[23] Strickland, C. M., Martin, G. M. and Forbes, C. S. (2007), "Parameterisation and Efficient MCMC Estimation of Non-Gaussian State Space Models," *Computational Statistics and Data Analysis*, doi:10.1016/j.csda.2007.10.010.