Département de sciences économiques

2008-03

The Hessian Method (Highly Efficient State Smoothing, In a Nutshell)

McCAUSLAND, William

Université de Montréal

Département de sciences économiques
Université de Montréal
Faculté des arts et des sciences
C.P. 6128, succursale Centre-Ville
Montréal (Québec) H3C 3J7
Canada
http://www.sceco.umontreal.ca
SCECO-information@UMontreal.CA
Téléphone : (514) 343-6539
Télécopieur : (514) 343-7221

Ce cahier a également été publié par le Centre interuniversitaire de recherche en économie quantitative (CIREQ) sous le numéro 03-2008.

This working paper was also published by the Center for Interuniversity Research in Quantitative Economics (CIREQ), under number 03-2008.

ISSN 0709-9231

THE HESSIAN METHOD
(HIGHLY EFFICIENT STATE SMOOTHING, IN A NUTSHELL)

WILLIAM J. MCCAUSLAND

Abstract. I introduce the HESSIAN method for semi-Gaussian state space models with univariate states. The vector of states $\alpha \equiv (\alpha_1, \ldots, \alpha_n)$ is Gaussian and the observed vector $y \equiv (y_1^\top, \ldots, y_n^\top)^\top$ need not be. I describe a close approximation $g(\alpha)$ to the density $f(\alpha|y)$. It is easy and fast to evaluate $g(\alpha)$ and draw from the approximate distribution. In particular, no simulation is required to approximate normalization constants. Applications include likelihood approximation using importance sampling and posterior simulation using Markov chain Monte Carlo (MCMC). HESSIAN is an acronym but it also refers to the Hessian of $\log f(\alpha|y)$, which figures prominently in the derivation. I compute my approximation for a basic stochastic volatility model and compare it with the multivariate Gaussian approximation described in Durbin and Koopman (1997) and Shephard and Pitt (1997). For a wide range of plausible parameter values, I estimate the variance of $\log f(\alpha|y) - \log g(\alpha)$ with respect to the approximate density $g(\alpha)$. For my approximation, this variance ranges from 330 to 39000 times smaller.

1. Introduction

I introduce a close approximation of the conditional distribution of states given observations in state space models of the form

$$\alpha_1 = x_0^\top \beta + u_0, \qquad \alpha_{t+1} = x_t^\top \beta + \phi_t \alpha_t + u_t, \quad t = 1, \ldots, n,$$

$$f(y_1, \ldots, y_n | \alpha_1, \ldots, \alpha_n) = \prod_{t=1}^{n} f_t(y_t|\alpha_t),$$

where the α_t are univariate latent states, the x_t are vectors of covariates, β is a vector of coefficients, the u_t are independant Gaussian random variables with mean 0 and precision (inverse of variance) ω_t, the y_t are observable random vectors, and the f_t are probability density or mass functions. I define $\alpha \equiv (\alpha_1, \ldots, \alpha_n)^\top$ and $y \equiv (y_1^\top, \ldots, y_n^\top)^\top$.

The distribution of y_t given α_t is quite flexible. It can be univariate or multivarite and the length of y_t can change with t. The f_t may be density functions or mass functions or a combination, and this too can vary over time. I require $\log f_t(y_t|\alpha_t)$ to be five times differentiable in α_t. I also assume that the distribution of α given y has a unique mode.

Date: Current version: March 17, 2008.

Mailing address: Département de sciences économiques, CIREQ, CIRANO, C.P. 6128, succursale Centre-ville, Montréal QC H3C 3J7, Canada. e-mail: william.j.mccausland@umontreal.ca. Web site: www.cirano.qc.ca/∼mccauslw.

In the special case where the densities $f_t(y_t|\alpha_t)$ are Gaussian, $\alpha|y$ is multivariate Gaussian and there are efficient methods to evaluate the likelihood function and draw α given y. Carter and Kohn (1994), Frühwirth-Schnatter (1994), de Jong and Shephard (1995), and Durbin and Koopman (2002) offer methods based on the Kalman filter. Rue (2001) and McCausland, Miller, and Pelletier (2007) introduce methods based on the precision of the conditional distribution of α given y.

When the f_t are not Gaussian, we can make use of a suitably close approximation $g(\alpha)$ of the target distribution $f(\alpha|y)$. We can estimate the likelihood function $f(y)$ using $g(\alpha)$ as an importance density. One can write $f(y) = E_g[f(\alpha)f(y|\alpha)/g(\alpha)]$, where E_g denotes expectation with respect to the approximate distribution. If regularity conditions for a strong law of large numbers hold, then

$$\frac{1}{M}\sum_{m=1}^{M}\frac{f(\alpha^{(m)})f(y|\alpha^{(m)})}{g(\alpha^{(m)})} \to f(y) \tag{1}$$

almost surely, where $\{\alpha^{(m)}\}_{m=1}^{M}$ is a random sample from the approximate distribution. We can do posterior inference for α using the approximation as a proposal distribution in a Metropolis-Hastings update of the target distribution. Providing that the approximation is close enough, drawing α as a block in this way may be much more efficient than one-at-a-time updates of the α_t, due to posterior autocorrelation in α. One can also use a good approximation of $f(\alpha|y)$ as part of a joint proposal distribution for parameters and states: one can draw a proposal for the parameters followed by a conditional proposal of α given the proposed parameter value then accept or reject the parameters and states together. This is particularly efficient when parameters and states are strongly *a posteriori* independent.

The closer the approximation, the better it is. In the importance sampling case, the variance of the sample mean in (1) decreases (or becomes finite) as $g(\alpha)$ more closely approximates $f(\alpha|y)$. In the MCMC case, the acceptance probability increases. Often the multivariate Gaussian approximation is not suitable as a proposal or importance distribution because it is not a close enough approximation of the target distribution. See Shephard and Pitt (1997), Gamerman (1998) and Pitt (2000) and the examples in Section 6 of the present paper.

There exist better approximations than the multivariate Gaussian. Richard and Zhang (forthcoming) describe one. The user supplies an "auxiliary parametric importance sampler", which consists of parametric approximations to the conditional densities $f(\alpha_t|\alpha_{t+1}, y)$. Computing parameters for the auxiliary sampler relies on computationally intensive simulation. Richard and Zhang (forthcoming), Liesenfeld and Richard (2003) and Liesenfeld and Richard (2008) consider applications of this approximation to stochastic volatility models. They suggest a Gaussian parametric approximation, which leads to particularly simple simulations but does not capture deviations from Gaussianity of the conditional densities $f(\alpha_t|\alpha_{t+1}, y)$. A more flexible choice of approximation would impose a considerable implementation burden on the user and require more costly simulations.

Rue, Steinsland, and Erland (2004) also describe improvements to the multivariate Gaussian approximation. Their approximation applies in the more general context of hidden Gaussian random fields, but also requires costly simulations to compute.

I will describe three approximations of $f(\alpha|y)$, the conditional density of α given y. The first is a multivariate normal approximation introduced by Durbin and Koopman (1997) and Shephard and Pitt (1997), and I use this as a benchmark. The second and third approximations introduce refinements to the multivariate Gaussian approximation. One can quickly and easily evaluate the fully normalized approximate densities and draw random vectors from the corresponding distributions.

In Section 2, I discuss the first, multivariate Gaussian, approximate distribution. I describe efficient methods to draw from the approximate distribution and evaluate its density. As in Durbin and Koopman (1997) and Shephard and Pitt (1997), the approximation is based on the quadratic Taylor series approximation of $\log f(y|\alpha)$ at α°, the conditional mode of α given y. However, I use very different techniques to draw α from the approximate distribution. I draw the α_t sequentially backwards. To compute $\mathrm{E}_g[\alpha_t|\alpha_{t+1},\ldots,\alpha_n]$ and $\mathrm{Var}_g[\alpha_t|\alpha_{t+1},\ldots,\alpha_n]$, I use the procedure by McCausland, Miller, and Pelletier (2007), based on an algorithm by Vandebril, Mastronardi, and Van Barel (2007) for solving band diagonal symmetric systems.

Sections 3 and 4 both introduce refinements to the multivariate Gaussian approximation. For the first refinement, the approximate conditional distributions $\alpha_t|\alpha_{t+1},\ldots,\alpha_n$ are still Gaussian, but α is no longer multivariate Gaussian. The second refinement replaces the conditional Gaussian distributions with skewed distributions capturing some of the departure from Gaussianity of the target distributions $\alpha_t|\alpha_{t+1},\ldots,\alpha_n,y$.

In Section 5, I briefly address problems that arise when the ratio $f(\alpha|y)/g(\alpha)$ is unbounded. I mention a standard solution and offer a variation on it.

In Section 6, I compare the three approximations of the target density for a basic stochastic volatility model. I first estimate the variance (with respect to the approximate distribution) of $\log f(\alpha|y) - \log g(\alpha)$ for the three approximations $g(\alpha)$. The first refinement leads to a moderate reduction in variance. Together with the first refinement, the second refinement leads to a dramatic reduction in variance for a wide range of plausible parameter values. I also demonstrate the efficacy of the final approximation as an importance distribution and as a proposal distribution. The HESSIAN method refers to the use of the second refinement for state smoothing.

In Section 7, I conclude and mention some possible extensions.

2. A Basic Multivariate Gaussian Proposal

Here I consider a first approximation of the target distribution. It is a multivariate Gaussian distribution based on a quadratic approximation of $\log f(\alpha|y)$ at the mode α° of the target distribution. One can write

$$\log f(\alpha|y) = \log f(\alpha) + \log f(y|\alpha) + k.$$

Here and throughout the paper I use k to denote an unimportant term not depending on α. This is a slight abuse of notation since the term is not always the same.

The first term is already quadratic in α. It is convenient to write it in the following form:

$$\log f(\alpha) = \frac{1}{2}\left[\log|\bar{H}| - n\log 2\pi - \alpha^\top \bar{H}\alpha + 2\bar{c}^\top\alpha + \bar{c}\bar{H}^{-1}\bar{c}\right], \tag{2}$$

where the precision (inverse of variance) $\bar{H}$ and covector (precision times mean) $\bar{c}$ of the marginal distribution of α are given by

$$\bar{H} \equiv \begin{bmatrix} \omega_0 + \omega_1\phi_1^2 & -\omega_1\phi_1 & 0 & 0 & \dots & 0 \\ -\omega_1\phi_1 & \omega_1 + \omega_2\phi_2^2 & -\omega_2\phi_2 & 0 & \dots & 0 \\ 0 & -\omega_2\phi_2 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \omega_{n-2} + \omega_{n-1}\phi_{n-1}^2 & -\omega_{n-1}\phi_{n-1} \\ 0 & 0 & \dots & 0 & -\omega_{n-1}\phi_{n-1} & \omega_{n-1} \end{bmatrix}$$

and

$$\bar{c} \equiv \begin{bmatrix} \omega_0 x_0\beta - \omega_1\phi_1 x_1\beta \\ \vdots \\ \omega_{n-2}x_{n-2}\beta - \omega_{n-1}\phi_{n-1}x_{n-1}\beta \\ \omega_{n-1}x_{n-1}\beta \end{bmatrix}.$$

A second order Taylor expansion of $\log f(y|\alpha)$ around the posterior mode α° gives the approximation

$$\log f(y|\alpha) \approx -\frac{1}{2}\left[\alpha^\top H\alpha - 2c^\top\alpha\right] + k, \tag{3}$$

where $H \equiv \operatorname{diag}(h_1, \dots, h_n)$, $c \equiv (c_1, \dots, c_n)^\top$, and for $t = 1, \dots, n$,

$$h_t \equiv -\left.\frac{\partial^2 f(y_t|\alpha_t)}{\partial\alpha_t^2}\right|_{\alpha_t=\alpha_t^\circ} \quad \text{and} \quad c_t \equiv \left.\frac{\partial f(y_t|\alpha_t)}{\partial\alpha_t}\right|_{\alpha_t=\alpha_t^\circ} - h_t\alpha_t^\circ. \tag{4}$$

The sum of the right hand sides of (2) and (3) gives the second order Taylor series expansion of $\log f(\alpha|y)$ around α°:

$$\log f(\alpha|y) \approx \frac{1}{2}\left[\log|\bar{\bar{H}}| - n\log 2\pi - \alpha^\top\bar{\bar{H}}\alpha + 2\bar{\bar{c}}^\top\alpha + \bar{\bar{c}}\bar{\bar{H}}^{-1}\bar{\bar{c}}\right] + k,$$

where $\bar{\bar{H}} \equiv \bar{H} + H$ and $\bar{\bar{c}} \equiv \bar{c} + c$. Since by assumption α° is the unique mode of α given y and $f(\alpha|y)$ is more than twice differentiable, $\bar{\bar{H}}$ must be positive definite. This in turn implies that the approximation is multivariate Gaussian.

Although framed differently, the multivariate Gaussian approximation is identical to the approximation introduced by Durbin and Koopman (1997) as an importance distribution. They constructed, as a device, a fully Gaussian state space model where the conditional distribution of the latent states given y approximates that of the semi-Gaussian model.

2.1. **Implementation.** Drawing from the approximate distribution involves finding α°, $\bar{\bar{H}}$ and $\bar{\bar{c}}$ and then drawing $\alpha^* \sim N(\bar{\bar{H}}^{-1}\bar{\bar{c}}, \bar{\bar{H}}^{-1})$.

α°, $\bar{\bar{H}}$ and $\bar{\bar{c}}$ satisfy $\alpha^\circ = \bar{\bar{H}}^{-1}\bar{\bar{c}}$. One can find these values by iterating the computation $\alpha' := \bar{\bar{H}}(\alpha)^{-1}\bar{\bar{c}}(\alpha)$ until numerical convergence, where $\bar{\bar{H}}(\alpha)^{-1}$ and $\bar{\bar{c}}(\alpha)$ are defined in the same way as $\bar{\bar{H}}$ and $\bar{\bar{c}}$ except that the derivatives in (4) are evaluated at the value α rather than at α°. I use the algorithm of Vandebril, Mastronardi, and Van Barel (2007) to compute α'. We will refer to this algorithm as the solver, since it solves a band diagonal system. There is a forward pass and a backward pass. The foward pass is:

(1) Compute $\Sigma_1 := 1/\bar{\bar{H}}_{11}$ and $m_1 := \Sigma_1\bar{\bar{c}}_1$.
(2) For $t = 2, \ldots, n$, compute $\Sigma_t := (\bar{\bar{H}}_{tt} - \bar{\bar{H}}^2_{t,t-1}\Sigma_{t-1})^{-1}$, and $m_t = \Sigma_t(\bar{\bar{c}}_t - \bar{\bar{H}}_{t,t-1}m_{t-1})$.

The backward pass is:

(1) Compute $\alpha'_n = m_n$.
(2) For $t = n-1, \ldots, 1$, compute $\alpha'_t = m_t - \Sigma_t\bar{\bar{H}}_{t,t+1}\alpha'_{t+1}$.

Once we have α°, we can draw α^* from the approximate distribution using the algorithm described in McCausland, Miller, and Pelletier (2007). Here, too, there is a forward and a backward pass. The forward pass is the same as that of the solver. The backward pass is:

(1) Draw $\alpha^*_n \sim N(m_n, \Sigma_n)$.
(2) For $t = n-1, \ldots, 1$, draw $\alpha^*_t \sim N(m_t - \Sigma_t\bar{\bar{H}}_{t,t+1}\alpha^*_{t+1}, \Sigma_t)$.

Evaluation of the density $g(\cdot)$ corresponding to the approximate distribution is straightforward. We have

$$\log g(\alpha) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\left[\log(\Sigma_t) + \Sigma_t^{-1}(\alpha_t - m_t - \Sigma_t\bar{\bar{H}}_{tt}\alpha_{t+1})^2\right]$$

While the multivariate Gaussian approximation is the same as the distribution Durbin and Koopman (1997) and Shephard and Pitt (1997) introduce, the method described here to draw variates is quite different, since it does not use the Kalman filter. McCausland, Miller, and Pelletier (2007) compare their method with methods based on the Kalman filter and find that their method is somewhat more efficient when one draw is required for each value of the parameters and considerably more efficient when repeated draws are required. MCMC is typically an example of the former. Importance sampling is an example of the latter.

3. A First Refinement

In this section I propose a first refinement to the multivariate Gaussian approximation. For the multivariate Gaussian approximation, $E_g[\alpha_t|\alpha_{t+1}, y]$ depends only affinely on α_{t+1}, but for the first refinement there are also quadratic and cubic terms. For the basic approximation, $\mathrm{Var}_g[\alpha_t|\alpha_{t+1}]$ does not depend on α_{t+1} at all, but for the first refinement $\log\mathrm{Var}_g[\alpha_t|\alpha_{t+1}, y]$ depends on α_{t+1} through linear and quadratic terms. The conditional distribution of α_t given α_{t+1} and y is still Gaussian, but the distribution of α given y is no longer multivariate Gaussian.

To provide some intuition for the first refinement, I first describe the following impractical approximate distribution. It is closer to the target distribution but I do not know of any computationally efficient procedure for drawing from it. The approximate marginal distribution of α_n is the same as for the multivariate Gaussian approximation. For $t = n-1, \ldots, 1$, the approximate conditional distribution of α_t given $\alpha_{t+1}, \ldots, \alpha_n$ is based on the quadratic Taylor series expansion of $\log f(\alpha_1, \ldots, \alpha_t | \alpha_{t+1}, \ldots, \alpha_n, y)$ at $(\alpha^\bullet_{1|t+1}, \ldots, \alpha^\bullet_{t|t+1})$, the conditional mode of $(\alpha_1, \ldots, \alpha_t)$ given $(\alpha_{t+1}, \ldots, \alpha_n)$ and y. There is a multivariate Gaussian distribution whose log density equals this expansion up to an additive constant. We know that its mean is $(\alpha^\bullet_{1|t+1}, \ldots, \alpha^\bullet_{t|t+1})$, since that it its mode. For $\tau = 1, \ldots, t$, I define $\Sigma^\bullet_{\tau|t+1}$ as the implied conditional variance of α_τ given α_{t+1}. The (impractical) approximate distribution is defined by

$$\alpha_n | y \sim N(\alpha^\circ_n, \Sigma_n), \qquad \alpha_t | \alpha_{t+1}, \ldots, \alpha_n, y \sim N(\alpha^\bullet_{t|t+1}, \Sigma^\bullet_{t|t+1}), \quad t = n-1, \ldots, 1.$$

In the way previous draws are used to update the quadratic approximations of the $\log f(y_t|\alpha_t)$, this should be a better approximation of the target distribution than the multivariate Gaussian approximation, which is based on the static quadratic approximations of the $\log f(y_t|\alpha_t)$ at α°_t. Our simulation results bear this out.

Computing the $\alpha^\bullet_{t|t+1}$ and $\Sigma^\bullet_{t|t+1}$ is feasible: for each t, we can compute $\alpha^\bullet_{1:t|t+1} \equiv (\alpha^\bullet_{1|t+1}, \ldots, \alpha^\bullet_{t|t+1})$ and $(\Sigma^\bullet_{1|t+1}, \ldots, \Sigma^\bullet_{t|t+1})$ in much the same way as we compute α° and the Σ_t. However, drawing the whole vector α in this way requires $O(n^2)$ operations and this is impractical. I do not know of any method of computing these quantities exactly using only $O(n)$ operations.

However, we can do something almost as good using $O(n)$ operations. Instead of computing the $\alpha^\bullet_{t|t+1}$ and $\Sigma^\bullet_{t|t+1}$ exactly, I use the following Taylor series approximations:

$$\alpha^\bullet_{t|t+1} \approx \hat{\alpha}^\bullet_{t|t+1} \equiv \alpha^\circ_t + \dot{\alpha}_t(\alpha_{t+1} - \alpha^\circ_{t+1}) + \frac{1}{2}\ddot{\alpha}_t(\alpha_{t+1} - \alpha^\circ_{t+1})^2 + \frac{1}{6}\dddot{\alpha}_t(\alpha_{t+1} - \alpha^\circ_{t+1})^3, \tag{5}$$

$$\log \Sigma^\bullet_{t|t+1} \approx \log \hat{\Sigma}^\bullet_{t|t+1} \equiv \log \Sigma^\circ_t + \dot{s}_t(\alpha_{t+1} - \alpha^\circ_{t+1}) + \frac{1}{2}\ddot{s}_t(\alpha_{t+1} - \alpha^\circ_{t+1})^2, \tag{6}$$

where I define

$$\dot{\alpha}_t \equiv \left.\frac{\partial \alpha^\bullet_{t|t+1}}{\partial \alpha_{t+1}}\right|_{\alpha_{t+1} = \alpha^\circ_{t+1}}, \quad \ddot{\alpha}_t \equiv \left.\frac{\partial^2 \alpha^\bullet_{t|t+1}}{\partial \alpha^2_{t+1}}\right|_{\alpha_{t+1} = \alpha^\circ_{t+1}}, \quad \dddot{\alpha}_t \equiv \left.\frac{\partial^3 \alpha^\bullet_{t|t+1}}{\partial \alpha^3_{t+1}}\right|_{\alpha_{t+1} = \alpha^\circ_{t+1}},$$

$$\dot{s}_t \equiv \left.\frac{\partial \log \Sigma^\bullet_{t|t+1}}{\partial \alpha_{t+1}}\right|_{\alpha_{t+1} = \alpha^\circ_{t+1}} \quad \text{and} \quad \ddot{s}_t \equiv \left.\frac{\partial^2 \log \Sigma^\bullet_{t|t+1}}{\partial \alpha^2_{t+1}}\right|_{\alpha_{t+1} = \alpha^\circ_{t+1}}.$$

I compute these five derivatives exactly, for $t = 1, \ldots, n-1$, using $O(n)$ operations.

I derive difference equations for $\dot{\alpha}_t$, $\ddot{\alpha}_t$ and $\dddot{\alpha}_t$ in the following way. Let $t \in \{1, \ldots, n-1\}$ be arbitrary. The log conditional density of $\alpha_1, \ldots, \alpha_t$ given $\alpha_{t+1}, \ldots, \alpha_n$ and y is

$$\log f(\alpha_{1:t}|\alpha_{t+1}, y) = -\frac{1}{2}(\alpha_{1:t} - \bar{\alpha}_{1:t})^\top \bar{H}_{1:t}(\alpha_{1:t} - \bar{\alpha}_{1:t}) - \bar{H}_{t,t+1}\alpha_t\alpha_{t+1} + \log f(y_{1:t}|\alpha_{1:t}) + k,$$

where $\bar{H}_{1:t}$ is the leading $t \times t$ submatrix of $\bar{H}$, $y_{1:t} \equiv (y_1, \ldots, y_t)^\top$, $\alpha_{1:t} \equiv (\alpha_1, \ldots, \alpha_t)$ and $\bar{\alpha}_{1:t} = \mathrm{E}[\alpha_{1:t}]$. A first order necessary condition for $\alpha^\bullet_{1:t|t+1}$ to be the conditional mode is

$$\bar{c}_{1:t} - \bar{H}_{1:t}\alpha^\bullet_{1:t|t+1} - (0, \ldots, 0, \bar{H}_{t,t+1}\alpha_{t+1})^\top + \frac{\partial \log f(y_{1:t}|\alpha^\bullet_{1:t|t+1})}{\partial \alpha^\top_{1:t}} = 0, \tag{7}$$

where $\bar{c}_{1:t} \equiv (\bar{c}_1, \ldots, \bar{c}_t)^\top$. We can rewrite this as

$$\bar{\bar{H}}^\bullet_{1:t|t+1}\alpha^\bullet_{1:t|t+1} = \bar{\bar{c}}^\bullet_{1:t|t+1}, \tag{8}$$

where

$$\bar{\bar{H}}^\bullet_{1:t|t+1} \equiv \bar{H}_{1:t} + H^\bullet_{1:t|t+1}, \quad \bar{\bar{c}}^\bullet_{1:t|t+1} \equiv \bar{c}_{1:t} + c^\bullet_{1:t|t+1} + (0, \ldots, 0, -\bar{H}_{t,t+1}\alpha_{t+1})^\top,$$

$$H^\bullet_{1:t|t+1} \equiv \mathrm{diag}(h^\bullet_{1|t+1}, \ldots, h^\bullet_{t|t+1}), \quad c^\bullet_{1:t|t+1} \equiv (c^\bullet_{1|t+1}, \ldots, c^\bullet_{t|t+1})^\top,$$

$$h^\bullet_{\tau|t+1} \equiv -\frac{\partial^2 \log f(y_\tau|\alpha^\bullet_{\tau|t+1})}{\partial \alpha^2_\tau}, \quad c^\bullet_{\tau|t+1} \equiv \frac{\partial \log f(y_\tau|\alpha^\bullet_{\tau|t+1})}{\partial \alpha_\tau} - h^\bullet_{\tau|t+1}\alpha^\bullet_{\tau|t+1}, \quad \tau = 1, \ldots, t.$$

Deriving difference equations for $\dot{\alpha}_t$, $\ddot{\alpha}_t$ and $\dddot{\alpha}_t$ involves differentiating (7) once, twice and three times with respect to α_{t+1} and applying results based on the solver algorithm. I show in Appendix A that $\dot{\alpha}_t$ is given by

$$\dot{\alpha}_t = -\Sigma_t \bar{H}_{t,t+1}, \qquad t = 1, \ldots, n-1, \tag{9}$$

that the $\ddot{\alpha}_t$ are given by the difference equation

$$\ddot{\alpha}_1 = \Sigma_1\psi_1\dot{\alpha}_1^2, \qquad \ddot{\alpha}_t = \Sigma_t\psi_t\dot{\alpha}_t^2 + \gamma_t\dot{\alpha}_t^2\ddot{\alpha}_{t-1}, \quad t = 2, \ldots, n-1, \tag{10}$$

where I define, for $t = 1, \ldots, n$,

$$\gamma_t \equiv \bar{H}_{t-1,t}\Sigma^\circ_t, \quad \psi_t \equiv \frac{\partial^3 \log f(y_t|\alpha^\circ_t)}{\partial \alpha^3_t} \quad \text{and} \quad \psi'_t \equiv \frac{\partial^4 \log f(y_t|\alpha^\circ_t)}{\partial \alpha^4_t},$$

and that the $\dddot{\alpha}_t$ are given by the difference equation

$$\dddot{\alpha}_1 = \Sigma^\circ_1(\psi'_1\dot{\alpha}_1^3 + 3\psi_1\dot{\alpha}_1\ddot{\alpha}_1),$$

$$\dddot{\alpha}_t = \Sigma^\circ_t(\psi'_t\dot{\alpha}_t^3 + 3\psi_t\dot{\alpha}_t\ddot{\alpha}_t) + \gamma_t(\dddot{\alpha}_{t-1}\dot{\alpha}_t^3 + 3\ddot{\alpha}_{t-1}\dot{\alpha}_t\ddot{\alpha}_t), \quad t = 2, \ldots, n. \tag{11}$$

I now describe how to compute $\dot{s}_t$ and $\ddot{s}_t$. Recall that the solver, applied to the unconditional precision $\bar{\bar{H}}$ and unconditional covector $\bar{\bar{c}}$, generates the conditional variances $\Sigma_1, \ldots, \Sigma_n$ for the multivariate Gaussian approximation. In the same way, it gives the conditional variances $\Sigma^\bullet_{1|t+1}$ and $\Sigma^\bullet_{t|t+1}$ as

$$\Sigma^\bullet_{1|t+1} = (\bar{\bar{H}}^\bullet_{11|t+1})^{-1}, \quad \Sigma^\bullet_{\tau|t+1} = \left[\bar{\bar{H}}^\bullet_{\tau\tau|t+1} - (\bar{\bar{H}}^\bullet_{\tau,\tau-1})^2\Sigma^\bullet_{\tau-1|t+1}\right]^{-1}, \quad \tau = 1, \ldots, t. \tag{12}$$

Deriving difference equations for $\dot{s}_t$ and $\ddot{s}_t$ involves differentiating (12) once and twice with respect to α_{t+1} and applying results based on the solver algorithm. I show in Appendix B that $\dot{s}_t$ is given by the difference equation

$$\dot{s}_1 = \Sigma^\circ_1\psi_1\dot{\alpha}_1, \qquad \dot{s}_t = \Sigma^\circ_t\psi_t\dot{\alpha}_t + \dot{\alpha}_{t-1}\dot{\alpha}_t\gamma_t\dot{s}_{t-1}, \quad t = 2, \ldots, n, \tag{13}$$

and that $\ddot{s}_t$ is given by the difference equation

$$\ddot{s}_1 = \dot{s}_1^2 + \Sigma_1^\circ(\psi_1'\dot{\alpha}_1^2 + \psi_1\ddot{\alpha}_1),$$

$$\ddot{s}_t = \dot{s}_t^2 + \Sigma_t(\psi_t'\dot{\alpha}_t^2 + \psi_t\ddot{\alpha}_t) + \gamma_t\dot{\alpha}_{t-1}(\dot{\alpha}_t^2\ddot{s}_{t-1} + \dot{s}_{t-1}\ddot{\alpha}_t + \dot{s}_{t-1}^2\dot{\alpha}_t^2), \quad t = 2, \ldots, n. \tag{14}$$

4. A Second Refinement

In the first refinement, the approximate conditional distributions $\alpha_t|\alpha_{t+1}, y$ adjust to changes in the quadratic approximation of $\sum_{\tau=1}^t \log f(y_\tau|\alpha_\tau)$ as α_{t+1} deviates from α_{t+1}°. The conditional distributions are still Gaussian, but the approximate distribution $\alpha|y$ is no longer multivariate Gaussian. The second refinement captures departures from Gaussianity of the target conditional distributions $\alpha_t|\alpha_{t+1}, y$.

The second refinement is based on an approximation of an approximation of the log conditional densities $\log f(\alpha_t|\alpha_{t+1}, y)$. I first approximate $\log f(\alpha_t|\alpha_{t+1}, y)$ directly as a third order polynomial in α_t of the following form:

$$\log f(\alpha_t|\alpha_{t+1}, y) \approx -\frac{1}{2}(\Sigma_{t|t+1}^\star)^{-1}(\alpha_t - \alpha_{t|t+1}^\star)^2 + \lambda_{t|t+1}^\star(\alpha_t - \alpha_{t|t+1}^\star)^3, \tag{15}$$

where $\alpha_{t|t+1}^\star$, $\Sigma_{t|t+1}^\star$ and $\lambda_{t|t+1}^\star$ are determined below. While the approximation is quite good near $\alpha_{t|t+1}^\bullet$, it is not the log of a proper density, due to the unbounded cubic term. I next approximate this approximation, up to an additive constant, by the log of a density that is proper, fully normalized, easy to evaluate, and simple to draw from. The collection of these indirect approximations of $\log f(\alpha_t|\alpha_{t+1}, y)$, for $t = 1, \ldots, n$, defines the second refinement.

4.1. Finding $\alpha_{t|t+1}^\star$, $\Sigma_{t|t+1}^\star$ and $\lambda_{t|t+1}^\star$.

We now proceed to find suitable values of the $\alpha_{t|t+1}^\star$, $\Sigma_{t|t+1}^\star$ and $\lambda_{t|t+1}^\star$ in (15). The case $t = 1$ is easy, since we know $\log f(\alpha_1|\alpha_2, y)$ up to an additive constant. We can write down its third order Taylor series approximation around $\alpha_{1|2}^\bullet$ as

$$\log f_2(\alpha_1|\alpha_2) \approx -\frac{1}{2}(\Sigma_{1|2}^\bullet)^{-1}(\alpha_1 - \alpha_{1|2}^\bullet)^2 + \frac{1}{6}\psi_{1|2}^\bullet(\alpha_1 - \alpha_{1|2}^\bullet)^3 + k.$$

This leads to the following obvious choices:

$$\alpha_{1|2}^\star \equiv \alpha_{1|2}^\bullet, \quad \Sigma_{1|2}^\star \equiv \Sigma_{1|2}^\bullet, \quad \lambda_{1|2}^\star \equiv \frac{1}{6}\psi_{1|2}^\bullet. \tag{16}$$

Finding suitable values of $\alpha_{t|t+1}^\star$, $\Sigma_{t|t+1}^\star$ and $\lambda_{t|t+1}^\star$ for $t > 1$ is more difficult, since we do not know $\log f(\alpha_t|\alpha_{t+1}, y)$ up to an additive constant. I assume for now that we have available a good approximation of $E[\alpha_{t-1} - \alpha_{t-1|t}^\bullet|\alpha_t, y]$ as a second order polynomial in $\alpha_t - \alpha_t^\circ$. That is, I assume we know the constant (not depending on α) coefficients A_{t-1}, B_{t-1} and C_{t-1} in the following expression:

$$E[\epsilon_{t-1}|\alpha_t] \approx A_{t-1} + B_{t-1}a_t + \frac{1}{2}C_{t-1}a_t^2, \tag{17}$$

where we define, for $t = 1, \ldots, n-1$, $\epsilon_t \equiv \alpha_t - \alpha^{\bullet}_{t|t+1}$ and for $t = 2, \ldots, n$, $a_t \equiv \alpha_t - \alpha^{\circ}_t$. Later in the section I will derive the approximations in (17), computing the A_t, B_t and C_t.

We are now ready to find an approximation for $\log f(\alpha_t|\alpha_{t+1}, y)$ of the form given in (15), for $t > 0$. We can write

$$\log f(\alpha_t|\alpha_{t+1}, y) = \log c(\alpha_t) + \log f(y_t|\alpha_t) + \log f(\alpha_{t+1}|\alpha_t) + k, \tag{18}$$

where the integration constant $c(\alpha_t)$ is given by

$$c(\alpha_t) \equiv \int f(\alpha_1) f(y_1|\alpha_1) \left\{ \prod_{\tau=2}^{t-1} f(\alpha_\tau|\alpha_{\tau-1}) f(y_\tau|\alpha_\tau) \right\} f(\alpha_t|\alpha_{t-1}) \prod_{\tau=1}^{t-1} d\alpha_\tau.$$

We now proceed to approximate each of the non-constant terms of (18) as a third order polynomial in ϵ_t. Later, we use the coefficients of this polynomial to determine $\alpha^{\star}_{t|t+1}$, $\Sigma^{\star}_{t|t+1}$ and $\lambda^{\star}_{t|t+1}$.

We begin with the first term of (18), $\log c(\alpha_t)$. Taking the derivative of $c(\alpha_t)$ with respect to α_t, we obtain

$$\begin{aligned}
\frac{\partial c(\alpha_t)}{\partial \alpha_t} &= \int f(\alpha_1) f(y_1|\alpha_1) \left\{ \prod_{\tau=1}^{t-1} f(\alpha_\tau|\alpha_{\tau-1}) f(y_\tau|\alpha_\tau) \right\} \frac{\partial f(\alpha_t|\alpha_{t-1})}{\partial \alpha_t} \prod_{\tau=1}^{t-1} d\alpha_\tau \\
&= \int f(\alpha_1) f(y_1|\alpha_1) \left\{ \prod_{\tau=1}^{t-1} f(\alpha_\tau|\alpha_{\tau-1}) f(y_\tau|\alpha_\tau) \right\} f(\alpha_t|\alpha_{t-1}) \frac{\partial \log f(\alpha_t|\alpha_{t-1})}{\partial \alpha_t} \prod_{\tau=1}^{t-1} d\alpha_\tau.
\end{aligned}$$

We can write

$$\begin{aligned}
\frac{\partial \log c(\alpha_t)}{\partial \alpha_t} &= \frac{1}{c(\alpha_t)} \frac{\partial c(\alpha_t)}{\partial \alpha_t} = E\left[\left. \frac{\partial \log f(\alpha_t|\alpha_{t-1})}{\partial \alpha_t} \right| \alpha_t \right] \\
&= -\omega_{t-1}\alpha_t + \omega_{t-1} x_{t-1}\beta + \omega_{t-1}\phi_{t-1} E[\alpha_{t-1}|\alpha_t] \\
&\approx -\omega_{t-1}\epsilon_t - \omega_{t-1}\alpha^{\bullet}_{t|t+1} + \omega_{t-1}x_{t-1}\beta - \bar{H}_{t-1,t}(\alpha^{\bullet}_{t-1|t} + A_{t-1} + B_{t-1}a_t + \frac{1}{2}C_{t-1}a_t^2).
\end{aligned}$$

We can approximate $\alpha^{\bullet}_{t-1|t}$ as follows:

$$\begin{aligned}
\alpha^{\bullet}_{t-1|t} &\approx \alpha^{\circ}_{t-1} + \dot{\alpha}_{t-1}a_t + \frac{1}{2}\ddot{\alpha}_{t-1}a_t^2 + \frac{1}{6}\dddot{\alpha}_{t-1}a_t^3 \\
&= \left(\alpha^{\circ}_{t-1} + \dot{\alpha}_{t-1}\delta_t + \frac{1}{2}\ddot{\alpha}_{t-1}\delta_t^2 + \frac{1}{6}\dddot{\alpha}_{t-1}\delta_t^3 \right) + \left(\dot{\alpha}_{t-1} + \ddot{\alpha}_{t-1}\delta_t + \frac{1}{2}\dddot{\alpha}_{t-1}\delta_t^2 \right)\epsilon_t \\
&\quad + \frac{1}{2}(\ddot{\alpha}_{t-1} + \dddot{\alpha}_{t-1}\delta_t)\epsilon_t^2 + \frac{1}{6}\dddot{\alpha}_{t-1}\epsilon_t^3 \\
&\approx \alpha^{\bullet}_{t-1|t+1} + \left(\dot{\alpha}_{t-1} + \ddot{\alpha}_{t-1}\delta_t + \frac{1}{2}\dddot{\alpha}_{t-1}\delta_t^2 \right)\epsilon_t + \frac{1}{2}(\ddot{\alpha}_{t-1} + \dddot{\alpha}_{t-1}\delta_t)\epsilon_t^2 + \frac{1}{6}\dddot{\alpha}_{t-1}\epsilon_t^3.
\end{aligned}$$

Similarly, we have

$$A_{t-1} + B_{t-1}a_t + \frac{1}{2}C_{t-1}a_t^2 = (A_{t-1} + B_{t-1}\delta_t + \frac{1}{2}C_{t-1}\delta_t^2) + (B_{t-1} + C_{t-1}\delta_t)\epsilon_t + \frac{1}{2}C_{t-1}\epsilon_t^2.$$

Gathering coefficients of 1, ϵ_t and ϵ_t^2 and dropping the cubic term, we write

$$\begin{aligned}
\frac{\partial \log c(\alpha_t)}{\partial \alpha_t} &\approx \left[-\omega_{t-1}\alpha^{\bullet}_{t|t+1} + \omega_{t-1}x_{t-1}\beta - \bar{H}_{t-1,t}\left(\alpha^{\bullet}_{t-1|t+1} + A_{t-1} + B_{t-1}\delta_t + \frac{1}{2}C_{t-1}\delta_t^2\right)\right] \\
&+ \left[-\omega_{t-1} - \bar{H}_{t-1,t}\left(\dot{\alpha}_{t-1} + \ddot{\alpha}_{t-1}\delta_t + \frac{1}{2}\dddot{\alpha}_{t-1}\delta_t^2 + B_{t-1} + C_{t-1}\delta_t\right)\right]\epsilon_t \\
&+ \frac{1}{2}\left[-\bar{H}_{t-1,t}(\ddot{\alpha}_{t-1} + \dddot{\alpha}_{t-1}\delta_t + C_{t-1})\right]\epsilon_t^2.
\end{aligned}$$

Integrating the right hand side, we obtain

$$\begin{aligned}
\log c(\alpha_t) &\approx \left[-\omega_{t-1}\alpha^{\bullet}_{t|t+1} + \omega_{t-1}x_{t-1}\beta - \bar{H}_{t-1,t}\left(\alpha^{\bullet}_{t-1|t+1} + A_{t-1} + B_{t-1}\delta_t + \frac{1}{2}C_{t-1}\delta_t^2\right)\right]\epsilon_t \\
&+ \frac{1}{2}\left[-\omega_{t-1} - \bar{H}_{t-1,t}\left(\dot{\alpha}_{t-1} + \ddot{\alpha}_{t-1}\delta_t + \frac{1}{2}\dddot{\alpha}_{t-1}\delta_t^2 + B_{t-1} + C_{t-1}\delta_t\right)\right]\epsilon_t^2 \\
&+ \frac{1}{6}\left[-\bar{H}_{t-1,t}(\ddot{\alpha}_{t-1} + \dddot{\alpha}_{t-1}\delta_t + C_{t-1})\right]\epsilon_t^3 + k.
\end{aligned}$$

I now approximate the second term of (18), $\log f(y_t|\alpha_t)$, by its third order Taylor series expansion around $\alpha^{\bullet}_{t|t+1}$:

$$\begin{aligned}
\log f(y_t|\alpha_t) &\approx \log f(y_t|\alpha^{\bullet}_{t|t+1}) + \frac{\partial \log f(y_t|\alpha^{\bullet}_{t|t+1})}{\partial \alpha_t}\epsilon_t + \frac{1}{2}\frac{\partial^2 \log f(y_t|\alpha^{\bullet}_{t|t+1})}{\partial \alpha_t^2}\epsilon_t^2 + \frac{1}{6}\frac{\partial^3 \log f(y_t|\alpha^{\bullet}_{t|t+1})}{\partial \alpha_t^3}\epsilon_t^3 \\
&= \log f(y_t|\alpha^{\bullet}_{t|t+1}) + (c^{\bullet}_{t|t+1} - h^{\bullet}_{t|t+1}\alpha^{\bullet}_{t|t+1})\epsilon_t - \frac{1}{2}h^{\bullet}_{t|t+1}\epsilon_t^2 + \frac{1}{6}\psi^{\bullet}_{t|t+1}\epsilon_t^3.
\end{aligned}$$

The third term of (18), $\log f(\alpha_{t+1}|\alpha_t)$, is already quadratic in α_t, so we can write it exactly as the following polynomial in ϵ_t:

$$\begin{aligned}
\log f(\alpha_{t+1}|\alpha_t) &= -\frac{1}{2}\omega_t(\alpha_{t+1} - x_t\beta - \phi_t\alpha_t)^2 \\
&= k + \left(\omega_t\phi_t\alpha_{t+1} - \omega_t\phi_t x_t\beta - \omega_t\phi_t^2\alpha^{\bullet}_{t|t+1}\right)\epsilon_t - \frac{1}{2}\omega_t\phi_t^2\epsilon_t^2.
\end{aligned}$$

Now we have all non-constant terms of (18) approximated by explicit third order polynomials in ϵ_t. We ignore the constant coefficient, which corresponds to the normalization constant of $f(\alpha_t|\alpha_{t+1}, y)$. Using the definitions $\bar{H}_{tt} = \omega_{t-1} + \omega_t\phi_t^2$ and $\bar{c}_t = \omega_{t-1}x_{t-1}\beta - \omega_t\phi_t x_t\beta$, we can write the coefficient of ϵ_t as

$$\begin{aligned}
&(\bar{c}_t + c^{\bullet}_{t|t+1}) - (\bar{H}_{tt} + h^{\bullet}_{t|t+1})\alpha^{\bullet}_{t|t+1} - \bar{H}_{t,t-1}\alpha^{\bullet}_{t-1|t+1} - \bar{H}_{t,t+1}\alpha_{t+1} - \bar{H}_{t,t-1}\left(A_{t-1} + B_{t-1}\delta_t + \frac{1}{2}C_{t-1}\delta_t^2\right) \\
&= \bar{\bar{c}}^{\bullet}_{t|t+1} - \bar{\bar{H}}^{\bullet}_{tt|t+1}\alpha^{\bullet}_{t|t+1} - \bar{H}_{t,t-1}\alpha^{\bullet}_{t-1|t+1} - \bar{H}_{t,t+1}\alpha_{t+1} - \bar{H}_{t,t-1}\left(A_{t-1} + B_{t-1}\delta_t + \frac{1}{2}C_{t-1}\delta_t^2\right) \\
&= -\bar{H}_{t,t-1}\left(A_{t-1} + B_{t-1}\delta_t + \frac{1}{2}C_{t-1}\delta_t^2\right).
\end{aligned}$$

To compute the final right hand side of this equation, I use the last element of the vector equation in (8).

I write the coefficient of ϵ_t^2 as

$$\begin{aligned} & -\frac{1}{2}\left[\bar{H}_{tt} + h^{\bullet}_{t|t+1} + \bar{H}_{t-1,t}\left(\dot{\alpha}_{t-1} + \ddot{\alpha}_{t+1}\delta_t + \frac{1}{2}\dddot{\alpha}_t\delta_t^2\right) + \bar{H}_{t-1,t}(B_{t-1} + C_{t-1})\delta_t\right] \\ = & -\frac{1}{2}\left[\bar{\bar{H}}^{\bullet}_{tt|t+1} - \bar{\bar{H}}^2_{t-1,t}\Sigma^{\bullet}_{t-1|t} + \bar{\bar{H}}_{t-1,t}(B_{t-1} + C_{t-1}\delta_t)\right] \\ = & -\frac{1}{2}\left[(\Sigma^{\bullet}_{t|t+1})^{-1} + \bar{\bar{H}}_{t-1,t}(B_{t-1} + C_{t-1}\delta_t)\right]. \end{aligned}$$

For the final line of this equation, I use equation (12), with $\tau = t$.

The coefficient of ϵ_t^3 is

$$\frac{1}{6}\left[\psi^{\bullet}_{t|t+1} - \bar{H}_{t-1,t}(\ddot{\alpha}_{t-1} + \dddot{\alpha}_{t-1}\delta_t + C_{t-1})\right].$$

Gathering terms, we now have the following approximation $g(\epsilon_t)$ of the log conditional target density $\log f(\epsilon_t|\alpha_{t+1})$:

$$\begin{aligned} g(\epsilon_t) \equiv & -\frac{1}{2}(\Sigma^{\bullet}_{t|t+1})^{-1}\epsilon_t^2 - \bar{H}_{t,t-1}\left(A_{t-1} + B_{t-1}\delta_t + \frac{1}{2}C_{t-1}\delta_t^2\right)\epsilon_t \\ & -\frac{1}{2}\bar{H}_{t,t-1}(B_{t-1} + C_{t-1}\delta_t)\epsilon_t^2 + \frac{1}{6}\left[\psi^{\bullet}_{t|t+1} - \bar{H}_{t,t-1}(\ddot{\alpha}_{t-1} + \dddot{\alpha}_{t-1}\delta_t) - \bar{H}_{t,t-1}C_{t-1}\right]\epsilon_t^3 \end{aligned} \tag{19}$$

A good choice for $\alpha^{\star}_{t|t+1}$ would be $\alpha^{\bullet}_{t|t+1} + \epsilon^{\star}$, where $\epsilon^{\star}$ is the local maximum of $g(\epsilon_t)$ near zero. This local maximum $\epsilon^{\star}$ must satisfy

$$\begin{aligned} \frac{\partial g(\epsilon^{\star})}{\partial \alpha_t} = & -(\Sigma^{\bullet}_{t|t+1})^{-1}\epsilon^{\star} - \bar{H}_{t,t-1}\left(A_{t-1} + B_{t-1}\delta_t + \frac{1}{2}C_{t-1}\delta_t^2\right) \\ & -\bar{H}_{t,t-1}(B_{t-1} + C_{t-1}\delta_t)\epsilon^{\star} + \frac{1}{2}\left[\psi^{\bullet}_{t|t+1} - \bar{H}_{t,t-1}(\ddot{\alpha}_{t-1} + \dddot{\alpha}_{t-1}\delta_t + C_{t-1})\right](\epsilon^{\star})^2 = 0. \end{aligned} \tag{20}$$

Finding the roots of this equation is feasible, but computationally intensive since a floating point division is required. Instead we will find a first order approximation of the root near zero. It is easy to see that for $A_{t-1} = B_{t-1} = C_{t-1} = 0$, we have a root at zero. We first find the first partial derivatives of (20) with respect to A_{t-1}, B_{t-1} and C_{t-1}, evaluated at $A_{t-1} = B_{t-1} = C_{t-1} = 0$. We obtain

$$\left.\frac{\partial^2 g(\epsilon^{\star})}{\partial A_{t-1}\partial\epsilon_t}\right|_{A_{t-1}=B_{t-1}=C_{t-1}=0} = -(\Sigma^{\bullet}_{t|t+1})^{-1}\left.\frac{\partial\epsilon^{\star}}{\partial A_{t-1}}\right|_{A_{t-1}=B_{t-1}=C_{t-1}=0} - \bar{H}_{t,t-1} = 0,$$

$$\left.\frac{\partial^2 g(\epsilon^{\star})}{\partial B_{t-1}\partial\epsilon_t}\right|_{A_{t-1}=B_{t-1}=C_{t-1}=0} = -(\Sigma^{\bullet}_{t|t+1})^{-1}\left.\frac{\partial\epsilon^{\star}}{\partial B_{t-1}}\right|_{A_{t-1}=B_{t-1}=C_{t-1}=0} - \bar{H}_{t,t-1}\delta_t = 0,$$

$$\left.\frac{\partial^2 g(\epsilon^{\star})}{\partial C_{t-1}\partial\epsilon_t}\right|_{A_{t-1}=B_{t-1}=C_{t-1}=0} = -(\Sigma^{\bullet}_{t|t+1})^{-1}\left.\frac{\partial\epsilon^{\star}}{\partial C_{t-1}}\right|_{A_{t-1}=B_{t-1}=C_{t-1}=0} - \frac{1}{2}\bar{H}_{t,t-1}\delta_t^2 = 0.$$

Next, we solve these equations for the first partial derivatives of $\epsilon^\star$ with respect to A_{t-1}, B_{t-1} and C_{t-1}, evaluated at $A_{t-1} = B_{t-1} = C_{t-1} = 0$, to obtain

$$\left.\frac{\partial \epsilon^\star}{\partial A_{t-1}}\right|_{A_{t-1}=B_{t-1}=C_{t-1}=0} = -\Sigma^\bullet_{t|t+1}\bar{H}_{t,t-1}, \quad \left.\frac{\partial \epsilon^\star}{\partial B_{t-1}}\right|_{A_{t-1}=B_{t-1}=C_{t-1}=0} = -\Sigma^\bullet_{t|t+1}\bar{H}_{t,t-1}\delta_t,$$

$$\left.\frac{\partial \epsilon^\star}{\partial C_{t-1}}\right|_{A_{t-1}=B_{t-1}=C_{t-1}=0} = -\frac{1}{2}\Sigma^\bullet_{t|t+1}\bar{H}_{t,t-1}\delta_t^2.$$

We now define

$$\alpha^\star_{t|t+1} \equiv \alpha^\bullet_{t|t+1} + \epsilon^\star_{t|t+1}, \tag{21}$$

where $\epsilon^\star_{t|t+1}$ is the following first order approximation of the local maximum of (19) near zero:

$$\epsilon^\star_{t|t+1} \equiv -\Sigma^\bullet_{t|t+1}\bar{H}_{t,t-1}\left(A_{t-1} + B_{t-1}\delta_t + \frac{1}{2}C_{t-1}\delta_t^2\right). \tag{22}$$

Ideally, I would choose $\Sigma^\star_{t|t+1}$ so that $-(\Sigma^\star_{t|t+1})^{-1}$, the second derivative of (15) with respect to α_t at $\alpha^\star_{t|t+1}$ matches the second derivative of (19) with respect to ϵ_t at $\epsilon^\star_{t|t+1}$. The latter is given by:

$$\frac{\partial^2 g(\epsilon^\star_{t|t+1})}{\partial \epsilon_t^2} \;=\; -(\Sigma^\bullet_{t|t+1})^{-1} - \bar{H}_{t,t-1}(B_{t-1} + C_{t-1}\delta_t) + \left[\psi^\bullet_{t|t+1} - \bar{H}_{t,t-1}(\ddot{\alpha}_{t-1} + \dddot{\alpha}_{t-1}\delta_t + C_{t-1})\right]\epsilon^\star_{t|t+1}.$$

Using the approximation $\log(1+x) \approx 1+x$, reasonable for small x, I define $\Sigma^\star_{t|t+1}$ indirectly by defining its logarithm as follows:

$$\begin{aligned}\log \Sigma^\star_{t|t+1} \;&\equiv\; \log \Sigma^\bullet_{t|t+1} \\ +\;& \Sigma^\bullet_{t|t+1}\left\{-\bar{H}_{t,t-1}(B_{t-1} + C_{t-1}\delta_t) + \left[\psi^\bullet_{t|t+1} - \bar{H}_{t,t-1}(\ddot{\alpha}_{t-1} + \dddot{\alpha}_{t-1}\delta_t + C_{t-1})\right]\epsilon^\star_{t|t+1}.\right\}\end{aligned} \tag{23}$$

I define $\lambda^\star_{t|t+1}$ as 1/6 (the 3rd order Taylor series coefficient) times the (constant) third derivative of (19) with respect to ϵ_t:

$$\lambda^\star_{t|t+1} \equiv \frac{1}{6}\frac{\partial^3 g(\epsilon^\star_{t|t+1})}{\partial \epsilon_t^3} = \frac{1}{6}\left[\psi^\bullet_{t|t+1} - \bar{H}_{t,t-1}(\ddot{\alpha}_{t-1} + \dddot{\alpha}_{t-1}\delta_t + C_{t-1})\right]. \tag{24}$$

I have now defined $\alpha^\star_{t|t+1}$, $\Sigma^\star_{t|t+1}$ and $\lambda^\star_{t|t+1}$ in terms of A_{t-1}, B_{t-1} and C_{t-1}.

4.2. **Computation of A_t, B_t, C_t.** I now need to compute values of A_t, B_t and C_t to approximate $E[\epsilon_t|\alpha_{t+1}, y]$ as

$$E[\epsilon_t|\alpha_{t+1}, y] = A_t + B_t a_{t+1} + \frac{1}{2}C_t a_{t+1}^2. \tag{25}$$

I begin by computing A_1, B_1 and C_1. We can write

$$\begin{aligned} E[\epsilon_1|\alpha_2] &\approx \int f_N(\epsilon_1; 0, \Sigma^\bullet_{1|2}) \left(1 + \frac{1}{6}\psi^\bullet_{1|2}\epsilon_1^3\right) \epsilon_1 \, d\epsilon_1 = \frac{1}{6}\psi^\bullet_{1|2} E_1[\epsilon_1^4] = \frac{1}{2}\psi^\bullet_{1|2}(\Sigma^\bullet_{1|2})^2 \\ &\approx \frac{1}{2}\left[\psi_1 + \psi_1'\dot{\alpha}_1 a_2 + \frac{1}{2}(\psi_1'\ddot{\alpha}_1 + \psi_1''\dot{\alpha}_1^2)a_2^2\right](\Sigma_1^\circ)^2 \exp(2\dot{s}_1 a_2 + \ddot{s}_1 a_2^2), \end{aligned}$$

where $f_N(\cdot;\mu,\Sigma)$ denotes the density of a Gaussian random variable with mean μ and variance Σ. The second order Taylor series expansion of this expression around $a_2 = 0$ gives the approximation $E[\epsilon_1|\alpha_2] \approx A_1 + B_1 a_2 + \frac{1}{2}C_1 a_2^2$, where

$$A_1 \equiv \frac{1}{2}(\Sigma_1^\circ)^2 \psi_1, \tag{26}$$

$$B_1 \equiv \frac{1}{2}(\Sigma_1^\circ)^2 (2\psi_1 \dot{s}_1 + \psi_1'\dot{\alpha}_1), \tag{27}$$

$$C_1 \equiv \frac{1}{2}(\Sigma_1^\circ)^2 \left[(4\dot{s}_1^2 + 2\ddot{s}_1)\psi_1 + (4\dot{s}_1\dot{\alpha}_1 + \ddot{\alpha}_1)\psi_1' + \dot{\alpha}_1^2\psi_1''\right]. \tag{28}$$

I now compute A_t, B_t and C_t for $t > 1$. I decompose $E[\epsilon_t|\alpha_{t+1}, y]$ as

$$E[\epsilon_t|\alpha_{t+1}, y] = \epsilon^\star_{t|t+1} + E[\epsilon_t - \epsilon^\star_{t|t+1}|\alpha_{t+1}, y]. \tag{29}$$

I approximate the first term of (29) as

$$\epsilon^\star_{t|t+1} = \gamma_t \exp\left(\dot{s}_t a_{t+1} + \frac{1}{2}\ddot{s}_t a_{t+1}^2\right)\left[A_{t-1} + B_{t-1}\dot{\alpha}_t a_{t+1} + \frac{1}{2}(B_{t-1}\ddot{\alpha}_t + C_{t-1}\dot{\alpha}_t^2)a_{t+1}^2\right].$$

I can approximate this in turn by its second order Taylor series expansion around a_{t+1}:

$$\begin{aligned} \epsilon^\star_{t|t+1} &\approx \gamma_t A_{t-1} + (\gamma_t A_{t-1}\dot{s}_t + \gamma_t B_{t-1}\dot{\alpha}_t)a_{t+1} \\ &+ \frac{1}{2}\left[(\gamma_t A_{t-1}(\dot{s}_t^2 + \ddot{s}_t) + \gamma_t B_{t-1}(2\dot{s}_t\dot{\alpha}_t + \ddot{\alpha}_t) + \gamma_t C_{t-1}\dot{\alpha}_t^2\right] a_{t+1}^2. \end{aligned}$$

Following the example of computing $E[\epsilon_1|\alpha_2]$ in (25), I approximate the second term of (29) as $E[\epsilon_t|\alpha_{t+1}, y] - \epsilon^\star_{t|t+1} \approx 3(\Sigma^\star_{t|t+1})^2\lambda^\star_{t|t+1}$, then $(\Sigma^\star_{t|t+1})^2$ as

$$(\Sigma^\star_{t|t+1})^2 \approx (\Sigma^\bullet_{t|t+1})^2 \approx (\Sigma_t^\circ)^2 \exp(2\dot{s}_t a_{t+1} + \ddot{s}_t a_{t+1}^2) \approx (\Sigma_t^\circ)^2(1 + 2\dot{s}_t a_{t+1} + (\dot{s}_t^2 + \ddot{s}_t)a_{t+1}^2),$$

and $6\lambda^\star_{t|t+1}$ as

$$\begin{aligned} 6\lambda^\star_{t|t+1} &\approx \psi^\bullet_{t|t+1} - \bar{H}_{t,t-1}(\ddot{\alpha}_{t-1} + \dddot{\alpha}_{t-1}\delta_t + C_{t-1}) \\ &= \left[\psi_t - \bar{H}_{t,t-1}(\ddot{\alpha}_{t-1} + \dddot{\alpha}_{t-1}\delta_t + C_{t-1})\right] + (\psi_t' - \bar{H}_{t,t-1}\dddot{\alpha}_{t-1})\delta_t + \frac{1}{2}\psi_t''\delta_t^2 \\ &= \bar{\psi}_t + \bar{\psi}_t'\dot{\alpha}_t a_{t+1} + \frac{1}{2}(\bar{\psi}_t'\ddot{\alpha}_t + \psi_t''\dot{\alpha}_t^2)a_{t+1}^2, \end{aligned}$$

where $\bar{\psi}_t \equiv \psi_t - \bar{H}_{t,t-1}(\ddot{\alpha}_{t-1} + C_{t-1})$ and $\bar{\psi}'_t \equiv \psi'_t - \bar{H}_{t,t-1}\dddot{\alpha}_{t-1}$. Multiplying 3 times the approximation for $(\Sigma^\star_{t|t+1})^2$ times the approximation for $\lambda^\star_{t|t+1}$ and retaining terms up to order 2 in a_{t+1}, we obtain

(30)
$$E[\epsilon_t|\alpha_{t+1}, y] \approx \frac{1}{2}(\Sigma^\circ_t)^2 \left\{\bar{\psi}_t + (2\bar{\psi}_t\dot{s}_t + \bar{\psi}'_t\dot{\alpha}_t)a_{t+1} + \left[(4\dot{s}_t^2 + 2\ddot{s}_t)\bar{\psi}_t + (4\dot{s}_t\dot{\alpha}_t + \ddot{\alpha}_t)\bar{\psi}'_t + \dot{\alpha}_t^2\psi''_t\right] a_{t+1}^2\right\}$$

Adding the two terms (30) and (30) approximating the two terms of $E[\epsilon_t|\alpha_{t+1}, y]$ in (29), we obtain

$$E[\epsilon_t|\alpha_{t+1}, y] \approx A_t + B_t a_{t+1} + \frac{1}{2}C_t a_{t+1}^2,$$

where

$$A_t \equiv \frac{1}{2}(\Sigma^\circ_t)^2\bar{\psi}_t + \gamma_t A_{t-1}, \tag{31}$$

$$B_t \equiv \frac{1}{2}(\Sigma^\circ_t)^2(2\bar{\psi}_t\dot{s}_t + \bar{\psi}'_t\dot{\alpha}_t) + \gamma_t A_{t-1}\dot{s}_t + \gamma_t B_{t-1}\dot{\alpha}_t, \tag{32}$$

$$\begin{aligned} C_t \quad \equiv \quad & \frac{1}{2}(\Sigma^\circ_t)^2 \left[(4\dot{s}_t^2 + 2\ddot{s}_t)\bar{\psi}_t + (4\dot{s}_t\dot{\alpha}_t + \ddot{\alpha}_t)\bar{\psi}'_t + \dot{\alpha}_t^2\psi''_t\right] \\ & + \gamma_t A_{t-1}(\dot{s}_t^2 + \ddot{s}_t) + \gamma_t B_{t-1}\left(2\dot{\alpha}_t\dot{s}_t + \ddot{\alpha}_t\right) + \gamma_t C_{t-1}\dot{\alpha}_t^2 \end{aligned} \tag{33}$$

4.3. **A Skewed Approximate Distribution for $\alpha_t|\alpha_{t+1}, y$.** We can now compute all the quantities in the expression for the approximation of $\log f(\alpha_t|\alpha_{t+1}, y)$ in (15). Since the approximation is based on Taylor series expansions around $\alpha^\bullet_{t|t+1}$ and α°_t, we expect it to be good near those values. Unfortunately, it is not the log of a proper density, due to the unbounded cubic term. I now give an approximation of this approximation, up to an additive constant, by the logarithm of a density that is proper and fully normalized. It is easy to evaluate the density and to draw random variates from the distribution it represents.

Using the approximation $e^x \approx 1 + x$, we approximate (15) near $\alpha^\star_{t|t+1}$ up to a positive multiplicative constant by the following function.

$$f^\star(\alpha_t) = \left(2\pi\Sigma^\star_{t|t+1}\right)^{-1/2} \exp\left[-\frac{1}{2}(\Sigma^\star_{t|t+1})^{-1}(\alpha_t - \alpha^\star_{t|t+1})^2\right]\left[1 + u(\lambda^\star_{t|t+1}(\alpha_t - \alpha^\star_{t|t+1})^3\right],$$

where

$$u(x) = \begin{cases} x & |x| < 1 \\ \text{sign}(x) & |x| \geq 1. \end{cases}$$

In practice, $\lambda^\star_{t|t+1}$ is small enough that in the region of high probability, $\lambda^\star_{t|t+1}(\alpha_t - \alpha^\star_{t|t+1})^3$ is very close to 0 and thus $u(\lambda(\alpha_t - \alpha^\star_{t|t+1})^3) = \lambda(\alpha_t - \alpha^\star_{t|t+1})^3$. We use u to ensure that $f^\star$ is non-negative everywhere. Choosing a function that is odd around $\alpha^\star_{t|t+1}$ makes the normalization constant easy to compute.

We take the density $f^\star(\alpha_t)$ as defining the conditional second refinement distribution. Since $u(\lambda^\star_{t|t+1}(\alpha_t - \alpha^\star_{t|t+1})^3)$ is an odd function of α_t around $\alpha^\star_t$, $f^\star$ is a proper and fully normalized density.

We can draw a random variate in the following way. We first draw a α_t from the Gaussian distribution with mean $\alpha^\star_{t|t+1}$ and variance $\Sigma^\star_{t|t+1}$. If $(\alpha_t - \alpha^\star_{t|t+1})$ and $\lambda^\star_{t|t+1}$ have opposite signs, then with probability $\lambda^\star_{t|t+1}(\alpha_t - \alpha^\star_{t|t+1})^3$ we replace α_t with $\alpha^\star_{t|t+1} - (\alpha_t - \alpha^\star_{t|t+1})$.

4.4. **An algorithm for drawing α.** I now outline the algorithm for drawing α from the second refinement approximate distribution.

(1) Compute $\Sigma_1, \ldots, \Sigma_n$ using the forward pass of the solver.
(2) Compute $\dot\alpha_t$, $\ddot\alpha_t$, $\dddot\alpha_t$, $\dot s_t$ and $\ddot s_t$, $t = 1, \ldots, n$, using equations (9), (10), (11), (13) and (14).
(3) For $t = 1, \ldots, n-1$, compute A_t, B_t and C_t using equations (26), (27), (28), (31), (32) and (33).
(4) For $t = n, \ldots, 1$
   (a) Compute $\hat\alpha^\bullet_{t|t+1}$ and $\hat\Sigma^\bullet_{t|t+1}$ using equations (5) and (6).
   (b) Compute $\hat\alpha^\star_{t|t+1}$, $\hat\Sigma^\star_{t|t+1}$ and $\hat\lambda^\star_{t|t+1}$ using equations (16), (21), (23) and (24).
   (c) Draw $\alpha_t \sim N(\hat\alpha^\star_{t|t+1}, \hat\Sigma^\star_{t|t+1})$.
   (d) If $\alpha_t - \hat\alpha^\star_{t|t+1}$ and $\lambda^\star_{t|t+1}$ have opposite signs, then with probability $\lambda^\star_{t|t+1}(\alpha_t - \alpha^\star_{t|t+1})^3$ replace α_t with $\alpha^\star_{t|t+1} - (\alpha_t - \alpha^\star_{t|t+1})$.

5. Tail Behavior

It is desirable that an approximation $g(\alpha)$ have the property that the ratio $f(\alpha|y)/g(\alpha)$ be bounded. In the context of importance sampling, this guarantees the boundedness of the importance weights, an important condition for the existence of the variance of importance sample means: see Geweke (1989) for example. In the context of independence Metropolis Hastings Markov chains, this guarantees the geometric ergodicity of the Markov chain: see Roberts and Rosenthal (1998).

There is no reason to suppose that the approximation $g(\alpha)$ introduced here has this property. In particular, the simulations of the next section do *not* demonstrate this, even for the examples considered. They only suggest that the region, if any, where the ratio $f(\alpha|y)/g(\alpha)$ becomes much larger than its mean value has extremely low probability.

Fortunately, it is quite simple to modify any approximation $g(\alpha)$ so that the ratio $f(\alpha|y)/g(\alpha)$ *is* bounded. The simplest solution is to mix in the marginal density $f(\alpha)$ into $g(\alpha)$, replacing $g(\alpha)$ with $(1-\pi)g(\alpha) + \pi f(\alpha)$, where $\pi \in (0, 1)$. A small but non-zero value of π is sufficient and does not overly distort the approximation in the region of high probability.

This solution may be less than satisfying as a practical device. For state space models with tens of thousands of observations, the prior will typically be extremely diffuse relative to the posterior. A better alternative might be to mix in the density $f(\alpha_t|\alpha_{t+1})$ with the conditional density $g(\alpha_t|\alpha_{t+1})$ for all (or a random number of) periods t.

6. An Empirical Example

As an example, I consider a simple stochastic volatility model which, together with variants differing only in parameterization, is widely used. The log-volatility equation is

$$\alpha_t = (1-\phi)\bar{\alpha} + \phi\, \alpha_{t-1} + u_t,$$

and the return equation is

$$y_t = e^{\alpha_t/2} v_t.$$

The error sequences $\{u_t\}$ and $\{v_t\}$ are Gaussian white noise and mutually independent. The precision of u_t is ω and the precision of v_t is 1. We observe the return y_t for $t = 1, \ldots, n$.

Jacquier, Polson, and Rossi (1994) propose a posterior simulator that draws volatility proposals one observation at a time. The high posterior autocorrelation of volatility, especially for daily returns, leads to highly autocorrelated posterior draws.

Kim, Shephard, and Chib (1998) (KSC) transform the model into a linear one, and approximate the random component of the transformed model as a mixture of Gaussian random variates. They employ a data augmentation scheme where mixture component indicators are included in the vector of unknown quantities. Given the component indicators, the approximate model is linear and Gaussian, allowing them to draw all volatilities and parameters as a block. Since volatilities are highly correlated with each other and with parameters, this improves numerical efficiency.

The chain's stationary distribution is only an approximation of the posterior distribution and to obtain simulation-consistent sample moments they re-weight draws. Re-weighting and data augmentation imply lower numerical efficiency than for independent draws.

The simple stochastic volatility model is a special case of the state space model described in the introduction, with

$$\bar{H}_{11} = \bar{H}_{nn} = \omega, \quad \bar{H}_{tt} = \omega(1+\phi^2) \quad t = 2, \ldots, n-1,$$

$$\bar{H}_{t,t+1} = -\omega\phi, \quad t = 1, \ldots, n-1$$

$$\bar{c}_1 = \bar{c}_n = (1-\phi)\bar{\alpha}, \quad \bar{c}_t = (1-\phi)^2\bar{\alpha}, \quad t = 2, \ldots, n-1,$$

$$h_t = \frac{1}{2} y_t^2 e^{-\alpha_t^\circ}, \quad c_t = \frac{1}{2}\left[y_t^2 e^{-\alpha_t^\circ}(1+\alpha_t^\circ) - 1\right], \quad t = 1, \ldots, n,$$

$$\psi_t = h_t, \quad \psi_t' = -h_t, \quad \psi_t'' = h_t, \quad t = 1, \ldots, n.$$

In the rest of this section, I use simulation exercises to evaluate how well the 2nd refinement works for this stochastic volatility example. In all simulations, we set $\bar{\alpha} = -9.0$, which implies that when the log volatility is at its mean, the standard deviation of returns is around 0.01, which is reasonable for stock returns. Since $\exp(\bar{\alpha}/2)$ is just a scale parameter for returns, the value of $\bar{\alpha}$ has no bearing on our results.

I vary ϕ and ω across simulations. I use the following values of ϕ: 0.8, 0.9, 0.95, 0.98, 0.99. Jacquier, Polson, and Rossi (1994) examine several financial returns, including those of major currencies; market indices; portfolios of first, fifth, and tenth decile stocks, by market capitalization, listed on the New York Stock Exchange; and individual stocks varying in market capitalization over a factor of about 370. They also report results obtained by other researchers. All estimates of ϕ for daily returns exceed 0.95. All estimates of ϕ for weekly returns exceed 0.8.

For each value of ϕ, I use three values of ω. The three values are chosen such that the corresponding values of the coefficient of variation of volatility are 0.25, 0.75 and 2.5. The coefficient of variation is the following function of the parameters ϕ and ω:

$$\frac{\mathrm{Var}[\exp(\alpha_t)]}{\mathrm{E}[\exp(\alpha_t)]^2} = \exp\left(\frac{1}{\omega(1-\phi^2)}\right) - 1.$$

Most of the estimates of the coefficient of variation in Jacquier, Polson, and Rossi (1994) lie between 0.5 and 1.0. The highest reported estimate is 1.61, with a posterior standard deviation of 0.38. Thus, the high value of the coefficient of variation is quite extreme.

Intuitively, lower values of ω, the precision of α_t given α_{t-1}, mean lower relative weight for the quadratic part of $\log f(\alpha|y)$ relative to the non-quadratic part. We can write, for example,

$$\frac{\partial^2}{\partial \alpha_t^2} \log f(\alpha_t|\alpha_{t-1}, y_t) = -\omega - \frac{1}{2} y_t^2 \exp(-\alpha_t).$$

The first term, which comes from the marginal density of α_t, is constant. The second term, which comes from the conditional density of y_t given α_t, depends on α_t but its expected value, conditional or not on α_t, is -0.5. Thus we can interpret ω as a measure of the relative weight of the quadratic to non-quadratic parts of $\log f(\alpha|y)$.

In the first simulation exercise, I do the following for all pairs (ϕ, ω). I first simulate $n = 10000$ returns using the stochastic volatility model and the given values of ϕ and ω. For daily returns in markets open five days a week, $n = 10000$ corresponds to about 40 years of data; for weekly returns, 200 years. Thus the worst case I consider, with a low value of ϕ even for weekly data, extremely high variation of volatility and $n = 10000$ is quite severe. I draw a sample of size $M = 10000$ from each of the three approximations and report the sample standard deviation of the log ratio between the target and approximate densities. Table 1 shows the results.

In the second simulation exercise, I assess the performance of the second refinement as an importance distribution for the stochastic volatility example. For each pair (ϕ, ω), I simulate $n = 10000$ returns, then draw a sample of size $M = 100$ from the second refinement distribution. I use the sample as an importance sample to estimate the log likelihood evaluated at the true value of the parameters. I report, in Table 2, numerical standard errors for the log likelihood estimate.

In the next two simulation exercises, I assess the second refinement as a proposal distribution in a Metropolis Hastings chain whose target distribution is the conditional distribution of α given y, $\bar{\alpha}$, ϕ and ω.

The first Metropolis-Hastings simulation exercise measures the numerical efficiency of estimating the conditional mean of α given y, $\bar{\alpha}$, ϕ and ω. For each pair (ϕ, ω), I simulate $n = 10000$ returns, then draw a Metropolis-Hastings chain of length $M = 5 \times 10^6$. The proposal distribution is the second refinement and the target distribution is the conditional distribution of α given y and the true values of $\bar{\alpha}$, ϕ and ω. For all $t = 1, \ldots, n$, I compute

TABLE 1. Standard deviation of $\log f(\alpha^*|y)/g(\alpha^*)$ by ϕ and ω

ϕ	ω	MVN proposal	1st refinement	2nd refinement
0.80	12.45	4.370	2.841	0.107
0.80	4.96	10.085	6.624	0.365
0.80	2.22	18.822	12.739	1.035
0.90	23.59	4.118	2.568	0.049
0.90	9.40	8.226	5.153	0.154
0.90	4.20	13.946	8.623	0.468
0.95	45.96	3.378	2.103	0.027
0.95	18.33	6.165	3.796	0.069
0.95	8.19	9.896	6.046	0.186
0.98	113.17	2.428	1.463	0.014
0.98	45.12	4.056	2.438	0.034
0.98	20.16	6.303	3.820	0.062
0.99	225.20	1.781	1.070	0.009
0.99	89.80	2.927	1.771	0.021
0.99	40.11	4.422	2.687	0.034

TABLE 2. Numerical standard error for the importance sampling log likelihood estimate, by ϕ and ω

ϕ	ω	numerical standard error
0.80	12.45	0.0109
0.80	4.96	0.0782
0.80	2.22	0.1336
0.90	23.59	0.0052
0.90	9.40	0.0152
0.90	4.20	0.0524
0.95	45.96	0.0029
0.95	18.33	0.0070
0.95	8.19	0.0157
0.98	113.17	0.0013
0.98	45.12	0.0027
0.98	20.16	0.0061
0.99	225.20	0.0008
0.99	89.80	0.0019
0.99	40.11	0.0039

the relative numerical efficiency[1] (RNE) of the sample mean of α_t, using the batch mean

[1]The relative numerical efficiency is the ratio of the squared numerical standard error for a hypothetical random sample to the squared numerical standard error for the Markov chain.

method. The number of batches is 500 and the batch length is 10000. In Table 3, I report the average (over t) relative numerical efficiencies for all pairs (ϕ, ω).

TABLE 3. Relative numerical efficiency for α draws, by ϕ and ω

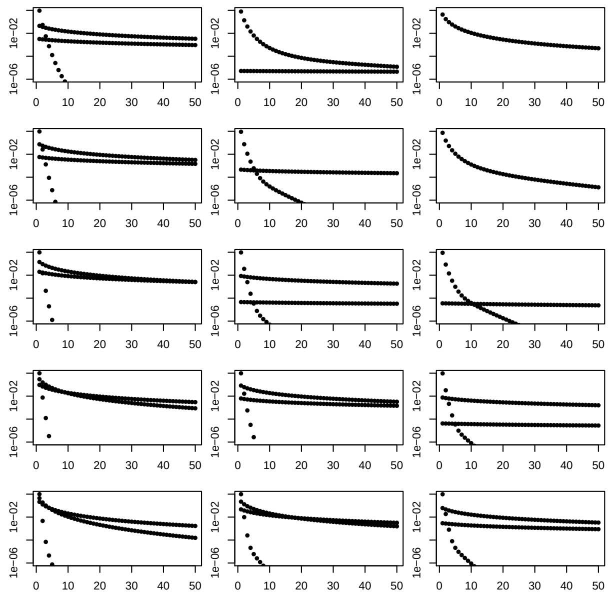
ϕ	ω	average RNE
0.80	12.45	0.88
0.80	4.96	0.39
0.80	2.22	0.06
0.90	23.59	0.95
0.90	9.40	0.83
0.90	4.20	0.42
0.95	45.96	0.97
0.95	18.33	0.93
0.95	8.19	0.78
0.98	113.17	0.99
0.98	45.12	0.97
0.98	20.16	0.93
0.99	225.20	0.99
0.99	89.80	0.98
0.99	40.11	0.96

The numerical inefficiency of independence Metropolis-Hastings chains arises from the rejection of proposals. I examine this issue in more detail in the last simulation exercise. For each pair (ϕ, ω) and all three proposal distributions, I estimate the unconditional distribution of the number of proposals that will be rejected before the next acceptance.

Figure 1 shows probability mass functions up to a run length of 50. Each plot is for a different pair (ϕ, ω). From top to bottom, the values of ϕ for all the plots in a row are 0.8, 0.9, 0.95, 0.98, and 0.99. From left to right, the values of the coefficient of variation of volatility for all the plots in a column are 0.25, 0.75 and 2.5. Each plot shows the log probability mass function for the multivariate Gaussian proposal, the first refinement and the second refinement. In all cases, the second refinement has the thinnest tails, followed by the first refinement. The multivariate Gaussian proposal always has the thickest tails. Where it seems that one or two curves are missing, all probabilities are below 10^{-6}.

For a given pair (ϕ, ω), and given y, I estimate the probability mass functions in the following manner. I draw samples of size $M = 20000$ from each of the three approximations. For each draw $\alpha^{(m)}$ from the second refinement sample, I compute the conditional probability that, given $\alpha^{(m)}$ is the chain's current state, of accepting the next proposal drawn from a given approximate distribution by averaging Metropolis-Hastings acceptance probabilities over all draws in the sample associated with that distribution. This yields a geometric conditional probability mass function for run lengths given that $\alpha^{(m)}$ is the chain's current state. I obtain the unconditional probability mass function by averaging over the $\alpha^{(m)}$, $1, \ldots, M$, using importance sampling weights.

Figure 1. Probabilities of numbers of rejections on a log scale. Values of ϕ are, from top to bottom, 0.8, 0.9, 0.95, 0.98 and 0.99. Values of the coefficient of variation of volatility are, from left to right, 0.25, 0.75 and 2.5.



7. Conclusions

The two refinements described in this paper define a new approximation of the conditional distribution of α given y in univariate semi-Gaussian state space models. For the stochastic volatility examples I consider, the approximation is dramatically closer to the target distribution than is a multivariate Gaussian approximation. When the relative precision of the Gaussian marginal distribution of α, compared with the precision that y contributes through $f(y_t|\alpha_t)$, is high, as it is for daily data and low or moderate variation in volatility, the variance of the log ratio of target to approximate densities is reduced by a factor in the thousands. Even when the relative precision is low, as it is for weekly data and volatility with low persistence and high variability, the variance reduction is in the hundreds. Applied as an importance distribution or a proposal distribution, the new approximation works well for problems where the multivariate Gaussian approximation is infeasible.

If these improvements carry over to other semi-Gaussian state space models, importance sampling and independence Metropolis-Hastings should become feasible for many problems where they are currently impractical. The approximation is not model specific, and one can implement it for a new model by providing routines to compute derivatives of $\log f(y_t|\alpha_t)$. This is not trivial, especially since we use derivatives up to the fifth. We note, however, that one can use numerical derivatives or other approximations. The approximate distribution may deteriorate as a result, but we will still be able to evaluate and draw exactly from an approximate density.

While it does not seem worth the trouble for the stochastic volatility model, there remains scope for further refinements. In particular, we could add higher order terms to the approximations $\hat{\alpha}^{\bullet}_{t|t+1}$ and $\log \hat{\Sigma}^{\bullet}_{t|t+1}$ of $\alpha^{\bullet}_{t|t+1}$ and $\log \Sigma^{\bullet}_{t|t+1}$. We could add a quartic term to the log of the approximate conditional density, although it would be difficult to construct a fully normalized approximation and to draw variates.

Possible future work includes applying the ideas of this paper to more general models, such as state space models with multivariate states or models with hidden Gaussian random fields.

Ackowledgements

I thank Nick Polson for some useful comments and Shirley Miller for pointing out some errors in a previous version.

Appendix A. Conditional Mean Derivatives

In this appendix we derive, for $t = 1, \ldots, n-1$, the first, second and third derivatives $\dot{\alpha}_t$, $\ddot{\alpha}_t$ and $\dddot{\alpha}_t$ of $\alpha^{\bullet}_{t|t+1}$ with respect to α_{t+1}, where $(\alpha^{\bullet}_{1|t+1}, \ldots, \alpha^{\bullet}_{t|t+1})$ is the conditional mode of $\alpha_1, \ldots, \alpha_t$ given $\alpha_{t+1}, \ldots, \alpha_n$ and y.

A.1. **First derivative.** Taking derivatives of both sides of (7) with respect to α_{t+1} yields

$$-\bar{H}_{1:t} \frac{\partial \alpha^{\bullet}_{1:t|t+1}}{\partial \alpha_{t+1}} - \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \bar{H}_{t,t+1} \end{bmatrix} - H^{\bullet}_{1:t|t+1} \frac{\partial \alpha^{\bullet}_{1:t|t+1}}{\partial \alpha_{t+1}} = 0, \tag{34}$$

Rearranging and setting $\alpha_{t+1} = \alpha^{\circ}_{t+1}$ gives:

$$\bar{\bar{H}}_{1:t} \left. \frac{\partial \alpha^{\bullet}_{1:t|t+1}}{\partial \alpha_{t+1}} \right|_{\alpha_{t+1} = \alpha^{\circ}_{t+1}} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ -\bar{H}_{t,t+1} \end{bmatrix}. \tag{35}$$

We can use the solver to solve equation (35) for the vector of first derivatives on the left hand side. This gives the solution in terms of the following difference equation:

$$\left. \frac{\partial \alpha^{\bullet}_{t|t+1}}{\partial \alpha_{t+1}} \right|_{\alpha_{t+1} = \alpha^{\circ}_{t+1}} = -\Sigma^{\circ}_t \bar{H}_{t,t+1},$$

$$\left. \frac{\partial \alpha^{\bullet}_{\tau|t+1}}{\partial \alpha_{t+1}} \right|_{\alpha_{t+1} = \alpha^{\circ}_{t+1}} = -\Sigma^{\circ}_{\tau} \bar{H}_{\tau,\tau+1} \left. \frac{\partial \alpha^{\bullet}_{\tau+1|t+1}}{\partial \alpha_{t+1}} \right|_{\alpha_{t+1} = \alpha^{\circ}_{t+1}} \qquad \tau = t-1, \ldots, 1.$$

Since t is abitrary, we have from the first equation that $\dot{\alpha}_t = -\Sigma_t^\circ \bar{H}_{t,t+1}$ for $t = 1, \ldots, n-1$, which is the result in equation (9).

A.2. **Second derivative.** Taking derivatives of both sides of (34) with respect to α_{t+1} yields

$$-\bar{H}_{1:t} \frac{\partial^2 \alpha^\bullet_{1:t|t+1}}{\partial \alpha^2_{t+1}} - \frac{\partial H^\bullet_{1:t|t+1}}{\partial \alpha_{t+1}} \frac{\partial \alpha^\bullet_{1:t|t+1}}{\partial \alpha_{t+1}} - H^\bullet_{1:t|t+1} \frac{\partial^2 \alpha^\bullet_{1:t|t+1}}{\partial \alpha^2_{t+1}} = 0. \tag{36}$$

Rearranging and setting $\alpha_{t+1} = \alpha^\circ_{t+1}$ gives:

$$\bar{\bar{H}}_{1:t} \left. \frac{\partial^2 \alpha^\bullet_{1:t|t+1}}{\partial \alpha^2_{t+1}} \right|_{\alpha_{t+1}=\alpha^\circ_{t+1}} = \left. \begin{bmatrix} \psi_1 \left(\frac{\partial \alpha^\bullet_{1|t+1}}{\partial \alpha_{t+1}} \right)^2 \\ \vdots \\ \psi_t \left(\frac{\partial \alpha^\bullet_{t|t+1}}{\partial \alpha_{t+1}} \right)^2 \end{bmatrix} \right|_{\alpha_{t+1}=\alpha^\circ_{t+1}} \tag{37}$$

We use the solver again to solve equation (37) for the vector of second derivatives on the left hand side. This gives the solution in terms of the following difference equation:

$$\ddot{\alpha}_t \equiv \left. \frac{\partial^2 \alpha^\bullet_{t|t+1}}{\partial \alpha^2_{t+1}} \right|_{\alpha_{t+1}=\alpha^\circ_{t+1}} = m_{t,t+1}, \tag{38}$$

$$\left. \frac{\partial^2 \alpha^\bullet_{\tau|t+1}}{\partial \alpha^2_{t+1}} \right|_{\alpha_{t+1}=\alpha^\circ_{t+1}} = m_{\tau,t+1} + \dot{\alpha}_\tau \left. \frac{\partial^2 \alpha^\bullet_{\tau+1|t+1}}{\partial \alpha^2_{t+1}} \right|_{\alpha_{t+1}=\alpha^\circ_{t+1}}, \quad \tau = t-1, \ldots, 1, \tag{39}$$

where the $m_{\tau,t+1}$ are given by the forward pass

$$m_{1,t+1} = \Sigma_1^\circ \psi_1 \left(\left. \frac{\partial \alpha^\bullet_{1|t+1}}{\partial \alpha_{t+1}} \right|_{\alpha_{t+1}=\alpha^\circ_{t+1}} \right)^2, \tag{40}$$

$$m_{\tau,t+1} = \Sigma_\tau^\circ \psi_\tau \left(\left. \frac{\partial \alpha^\bullet_{\tau|t+1}}{\partial \alpha_{t+1}} \right|_{\alpha_{t+1}=\alpha^\circ_{t+1}} \right)^2 + \gamma_\tau m_{\tau-1,t+1}, \quad \tau = 2, \ldots, t, \tag{41}$$

where, recall, for all τ, $\gamma_\tau \equiv \Sigma_\tau^\circ \bar{\bar{H}}_{\tau,\tau-1}$.

If we take $\alpha^\bullet_{\tau|t+1}(\alpha_{t+1}) = \alpha^\bullet_{\tau|t}(\alpha^\bullet_{t|t+1}(\alpha_{t+1}))$ we can write

$$\frac{\partial \alpha^\bullet_{\tau|t+1}}{\partial \alpha_{t+1}} = \frac{\partial \alpha^\bullet_{\tau|t}}{\partial \alpha_t} \frac{\partial \alpha^\bullet_{t|t+1}}{\partial \alpha_{t+1}}.$$

using the chain rule. Taking another partial derivative with respect to α_{t+1}, again chaining through α_t, we obtain

$$\frac{\partial^2 \alpha^\bullet_{\tau|t+1}}{\partial \alpha^2_{t+1}} = \frac{\partial^2 \alpha^\bullet_{\tau|t}}{\partial \alpha^2_t} \left(\frac{\partial \alpha^\bullet_{t|t+1}}{\partial \alpha_{t+1}} \right)^2 + \frac{\partial \alpha^\bullet_{\tau|t}}{\partial \alpha_t} \frac{\partial^2 \alpha^\bullet_{t|t+1}}{\partial \alpha^2_{t+1}}. \tag{42}$$

Taking $\tau = t-1$ and setting $\alpha_{t+1} = \alpha^\circ_{t+1}$ we obtain

$$\left. \frac{\partial^2 \alpha^\bullet_{t-1|t+1}}{\partial \alpha^2_{t+1}} \right|_{\alpha_{t+1} = \alpha^\circ_{t+1}} = \ddot{\alpha}_{t-1} \dot{\alpha}^2_t + \dot{\alpha}_{t-1} \ddot{\alpha}_t.$$

We now write equation (39), with τ set to $t-1$:

$$\left. \frac{\partial^2 \alpha^\bullet_{t-1|t+1}}{\partial \alpha^2_{t+1}} \right|_{\alpha_{t+1} = \alpha^\circ_{t+1}} = m_{t-1,t+1} + \dot{\alpha}_{t-1} \ddot{\alpha}_t.$$

Substituting this equation in the previous equation gives

$$m_{t-1,t+1} = \dot{\alpha}^2_t \ddot{\alpha}_{t-1}. \tag{43}$$

We now write equation (41), with $\tau = t$:

$$m_{t,t+1} = \Sigma^\circ_t \psi_t \dot{\alpha}^2_t + \gamma_t m_{t-1,t+1}.$$

Substituting equations (38) and (43) in this expression gives:

$$\ddot{\alpha}_t = \Sigma^\circ_t \psi_t \dot{\alpha}^2_t + \gamma_t \dot{\alpha}^2_t \ddot{\alpha}_{t-1}.$$

Equations (38) and (40), with $t = 1$, give

$$\ddot{\alpha}_1 = \Sigma^\circ_1 \psi_1 \dot{\alpha}^2_1.$$

Since t is arbitrary, the two previous equations give the result in equation (10).

A.3. **Third derivative.** Taking derivatives of both sides of (36) with respect to α_{t+1} yields

$$-\bar{H}_{1:t} \frac{\partial^3 \alpha^\bullet_{1:t|t+1}}{\partial \alpha^3_{t+1}} + \begin{bmatrix} \psi'^\bullet_{1|t+1} \left(\frac{\partial \alpha^\bullet_{1|t+1}}{\partial \alpha_{t+1}} \right)^3 + 3 \psi^\bullet_{1|t+1} \frac{\partial \alpha^\bullet_{1|t+1}}{\partial \alpha_{t+1}} \frac{\partial^2 \alpha^\bullet_{1|t+1}}{\partial \alpha^2_{t+1}} \\ \vdots \\ \psi'^\bullet_{t|t+1} \left(\frac{\partial \alpha^\bullet_{t|t+1}}{\partial \alpha_{t+1}} \right)^3 + 3 \psi^\bullet_{t|t+1} \frac{\partial \alpha^\bullet_{t|t+1}}{\partial \alpha_{t+1}} \frac{\partial^2 \alpha^\bullet_{t|t+1}}{\partial \alpha^2_{t+1}} \end{bmatrix} - H^\bullet_{1:t} \frac{\partial^3 \alpha^\bullet_{1:t|t+1}}{\partial \alpha^3_{t+1}} = 0,$$

where

$$\psi^\bullet_{\tau|t+1} \equiv \left. \frac{\partial^3 \log f(y_\tau | \alpha_\tau)}{\partial \alpha^3_\tau} \right|_{\alpha_\tau = \alpha^\bullet_{\tau|t+1}} \qquad \psi'^\bullet_{\tau|t+1} \equiv \left. \frac{\partial^4 \log f(y_\tau | \alpha_\tau)}{\partial \alpha^4_\tau} \right|_{\alpha_\tau = \alpha^\bullet_{\tau|t+1}} \qquad \tau = 1, \ldots, t$$

Rearranging and setting $\alpha_{t+1} = \alpha^\circ_{t+1}$ gives:

$$\bar{\bar{H}}_{1:t} \left. \frac{\partial^3 \alpha^\bullet_{1:t|t+1}}{\partial \alpha^3_{t+1}} \right|_{\alpha_{t+1}=\alpha^\circ_{t+1}} = \left. \begin{bmatrix} \psi'_1 \left(\frac{\partial \alpha^\bullet_{1|t+1}}{\partial \alpha_{t+1}} \right)^3 + 3\psi_1 \frac{\partial \alpha^\bullet_{1|t+1}}{\partial \alpha_{t+1}} \frac{\partial^2 \alpha^\bullet_{1|t+1}}{\partial \alpha^2_{t+1}} \\ \vdots \\ \psi'_t \left(\frac{\partial \alpha^\bullet_{t|t+1}}{\partial \alpha_{t+1}} \right)^3 + 3\psi_t \frac{\partial \alpha^\bullet_{t|t+1}}{\partial \alpha_{t+1}} \frac{\partial^2 \alpha^\bullet_{t|t+1}}{\partial \alpha^2_{t+1}} \end{bmatrix} \right|_{\alpha_{t+1}=\alpha^\circ_{t+1}} \tag{44}$$

We use the solver again to solve equation (44) for the vector of third derivatives on the left hand side. This gives the solution in terms of the following difference equation:

$$\dddot{\alpha}_t \equiv \left. \frac{\partial^3 \alpha^\bullet_{t|t+1}}{\partial \alpha^3_{t+1}} \right|_{\alpha_{t+1}=\alpha^\circ_{t+1}} = \bar{m}_{t,t+1}, \tag{45}$$

$$\left. \frac{\partial^3 \alpha^\bullet_{\tau|t+1}}{\partial \alpha^3_{t+1}} \right|_{\alpha_{t+1}=\alpha^\circ_{t+1}} = \bar{m}_{\tau,t+1} + \dot{\alpha}_\tau \left. \frac{\partial^3 \alpha^\bullet_{\tau+1|t+1}}{\partial \alpha^3_{t+1}} \right|_{\alpha_{t+1}=\alpha^\circ_{t+1}}, \quad \tau = t-1, \ldots, 1, \tag{46}$$

where the $m_{\tau,t+1}$ are given by the forward pass

$$\bar{m}_{1,t+1} = \Sigma^\circ_1 \left. \left[\psi'_1 \left(\frac{\partial \alpha^\bullet_{1|t+1}}{\partial \alpha_{t+1}} \right)^3 + 3\psi_1 \frac{\partial \alpha^\bullet_{1|t+1}}{\partial \alpha_{t+1}} \frac{\partial^2 \alpha^\bullet_{1|t+1}}{\partial \alpha^2_{t+1}} \right] \right|_{\alpha_{t+1}=\alpha^\circ_{t+1}}, \tag{47}$$

(48)

$$\bar{m}_{\tau,t+1} = \Sigma^\circ_\tau \left. \left[\psi'_\tau \left(\frac{\partial \alpha^\bullet_{\tau|t+1}}{\partial \alpha_{t+1}} \right)^3 + 3\psi_\tau \frac{\partial \alpha^\bullet_{\tau|t+1}}{\partial \alpha_{t+1}} \frac{\partial^2 \alpha^\bullet_{\tau|t+1}}{\partial \alpha^2_{t+1}} \right] \right|_{\alpha_{t+1}=\alpha^\circ_{t+1}} + \gamma_\tau \bar{m}_{\tau-1,t+1}, \quad \tau = 2, \ldots, t.$$

Taking the partial derivative with respect to α_{t+1} of both sides of equation (42), again chaining through α_t, we obtain

$$\frac{\partial^3 \alpha^\bullet_{\tau|t+1}}{\partial \alpha^3_{t+1}} = \frac{\partial^3 \alpha^\bullet_{\tau|t}}{\partial \alpha^3_t} \left(\frac{\partial \alpha^\bullet_{t|t+1}}{\partial \alpha_{t+1}} \right)^3 + 3 \frac{\partial^2 \alpha^\bullet_{\tau|t}}{\partial \alpha^2_t} \frac{\partial \alpha^\bullet_{t|t+1}}{\partial \alpha_{t+1}} \frac{\partial^2 \alpha^\bullet_{t|t+1}}{\partial \alpha^2_{t+1}} + \frac{\partial \alpha^\bullet_{\tau|t}}{\partial \alpha_t} \frac{\partial^3 \alpha^\bullet_{t|t+1}}{\partial \alpha^3_{t+1}}.$$

Taking $\tau = t-1$ and setting $\alpha_{t+1} = \alpha^\circ_{t+1}$ we obtain

$$\left. \frac{\partial^3 \alpha^\bullet_{t-1|t+1}}{\partial \alpha^3_{t+1}} \right|_{\alpha_{t+1}=\alpha^\circ_{t+1}} = \dddot{\alpha}_{t-1} \dot{\alpha}^3_t + 3\ddot{\alpha}_{t-1} \dot{\alpha}_t \ddot{\alpha}_t + \dot{\alpha}_{t-1} \dddot{\alpha}_t.$$

We now write equation (46), with τ set to $t-1$:

$$\left. \frac{\partial^3 \alpha^\bullet_{t-1|t+1}}{\partial \alpha^3_{t+1}} \right|_{\alpha_{t+1}=\alpha^\circ_{t+1}} = \bar{m}_{t-1,t+1} + \dot{\alpha}_{t-1} \dddot{\alpha}_t.$$

Substituting this equation in the previous equation gives

$$\bar{m}_{t-1,t+1} = \dddot{\alpha}_{t-1}\dot{\alpha}_t^3 + 3\ddot{\alpha}_{t-1}\dot{\alpha}_t\ddot{\alpha}_t. \tag{49}$$

We now write equation (48), with $\tau = t$:

$$\bar{m}_{t,t+1} = \Sigma_t^\circ(\psi_t'\dot{\alpha}_t^3 + 3\psi_t\dot{\alpha}_t\ddot{\alpha}_t) + \gamma_t\bar{m}_{t-1,t+1}.$$

Substituting equations (45) and (49) in this expression gives:

$$\dddot{\alpha}_t = \Sigma_t^\circ(\psi_t'\dot{\alpha}_t^3 + 3\psi_t\dot{\alpha}_t\ddot{\alpha}_t) + \gamma_t(\dddot{\alpha}_{t-1}\dot{\alpha}_t^3 + 3\ddot{\alpha}_{t-1}\dot{\alpha}_t\ddot{\alpha}_t).$$

Equations (45) and (47), with $t = 1$, give

$$\dddot{\alpha}_1 = \Sigma_1^\circ(\psi_1'\dot{\alpha}_1^3 + 3\psi_1\dot{\alpha}_1\ddot{\alpha}_1)$$

Since t is arbitrary, the two previous equations give the result in equation(11).

Appendix B. Conditional Variance Derivatives

In this appendix we derive, for $t = 1, \ldots, n-1$, the first and second derivatives $\dot{s}_t$ and $\ddot{s}_t$ of $\log \Sigma_{t|t+1}^\bullet$ with respect to α_{t+1}.

We begin with some useful preliminary results. Differentiating $\log \Sigma_{t-1|t+1}^\bullet$ with respect to α_{t+1}, chaining through α_t, yields

$$\frac{\partial \log \Sigma_{t-1|t+1}^\bullet}{\partial \alpha_{t+1}} = \frac{\partial \log \Sigma_{t-1|t}^\bullet}{\partial \alpha_t}\frac{\partial \alpha_{t|t+1}^\bullet}{\partial \alpha_{t+1}}, \tag{50}$$

Differentiating again with respect to α_{t+1}, we obtain

$$\frac{\partial^2 \log \Sigma_{t-1|t+1}^\bullet}{\partial \alpha_{t+1}^2} = \frac{\partial^2 \log \Sigma_{t-1|t}^\bullet}{\partial \alpha_t^2}\left(\frac{\partial \alpha_{t|t+1}^\bullet}{\partial \alpha_{t+1}}\right)^2 + \frac{\partial \log \Sigma_{t-1|t}^\bullet}{\partial \alpha_t}\frac{\partial^2 \alpha_{t|t+1}^\bullet}{\partial \alpha_{t+1}^2}. \tag{51}$$

Setting $\alpha_{t+1} = \alpha_{t+1}^\circ$ in (50) gives

$$\left.\frac{\partial \log \Sigma_{t-1|t+1}^\bullet}{\partial \alpha_{t+1}}\right|_{\alpha_{t+1}=\alpha_{t+1}^\circ} = \dot{s}_{t-1}\dot{\alpha}_t, \tag{52}$$

and setting $\alpha_{t+1} = \alpha_{t+1}^\circ$ in (51) gives

$$\left.\frac{\partial^2 \log \Sigma_{t-1|t+1}^\bullet}{\partial \alpha_{t+1}^2}\right|_{\alpha_{t+1}=\alpha_{t+1}^\circ} = \ddot{s}_{t-1}\dot{\alpha}_t^2 + \dot{s}_{t-1}\ddot{\alpha}_t. \tag{53}$$

Differentiating $h_{t|t+1}^\bullet$ with respect to α_{t+1}, chaining through α_t, yields

$$\frac{\partial h_{t|t+1}^\bullet}{\partial \alpha_{t+1}} = -\psi_{t|t+1}^\bullet\frac{\partial \alpha_{t|t+1}^\bullet}{\partial \alpha_{t+1}}. \tag{54}$$

Differentiating again with respect to α_{t+1}, we obtain

$$\frac{\partial^2 h_{t|t+1}^\bullet}{\partial \alpha_{t+1}^2} = -\psi_{t|t+1}'^\bullet\left(\frac{\partial \alpha_{t|t+1}^\bullet}{\partial \alpha_{t+1}}\right)^2 - \psi_{t|t+1}^\bullet\frac{\partial^2 \alpha_{t|t+1}^\bullet}{\partial \alpha_{t+1}^2}. \tag{55}$$

Setting $\alpha_{t+1} = \alpha^\circ_{t+1}$ in (54), we obtain

$$\left.\frac{\partial h^\bullet_{t|t+1}}{\partial \alpha_{t+1}}\right|_{\alpha_{t+1}=\alpha^\circ_{t+1}} = -\psi_t \dot{\alpha}_t \tag{56}$$

Setting $\alpha_{t+1} = \alpha^\circ_{t+1}$ in (55), we obtain

$$\left.\frac{\partial^2 h^\bullet_{t|t+1}}{\partial \alpha^2_{t+1}}\right|_{\alpha_{t+1}=\alpha^\circ_{t+1}} = -\psi'_t \dot{\alpha}^2_t - \psi_t \ddot{\alpha}_t. \tag{57}$$

B.1. **First derivative.** For clarity, we rewrite the equations in (12) as

$$\Sigma^\bullet_{1|t+1} = \left(\bar{H}_{11} + h^\bullet_{1|t+1}\right)^{-1} \tag{58}$$

$$\Sigma^\bullet_{\tau|t+1} = \left(\bar{H}_{\tau\tau} + h^\bullet_{\tau|t+1} - \bar{H}^2_{\tau-1,\tau}\Sigma^\bullet_{\tau-1|t+1}\right)^{-1}, \quad \tau = 2, \ldots, t. \tag{59}$$

Taking the logarithm of equations (58) and (59) then the derivative with respect to α_{t+1} gives

$$\frac{\partial \log \Sigma^\bullet_{1|t+1}}{\partial \alpha_{t+1}} = -\Sigma^\bullet_{1|t+1} \frac{\partial h^\bullet_{1|t+1}}{\partial \alpha_{t+1}}, \tag{60}$$

and for $\tau = 2, \ldots, t$,

$$\begin{aligned}
\frac{\partial \log \Sigma^\bullet_{\tau|t+1}}{\partial \alpha_{t+1}} &= -\Sigma^\bullet_{\tau|t+1}\left[\frac{\partial h^\bullet_{\tau|t+1}}{\partial \alpha_{t+1}} - \bar{H}^2_{\tau-1,\tau}\frac{\partial \Sigma^\bullet_{\tau-1|t+1}}{\partial \alpha_{t+1}}\right] \\
&= -\Sigma^\bullet_{\tau|t+1}\frac{\partial h^\bullet_{\tau|t+1}}{\alpha_{t+1}} + \Sigma^\bullet_{\tau|t+1}\bar{H}^2_{\tau-1,\tau}\Sigma^\bullet_{\tau-1|t+1}\frac{1}{\Sigma^\bullet_{\tau-1|t+1}}\frac{\partial \Sigma^\bullet_{\tau-1|t+1}}{\partial \alpha_{t+1,i}} \\
&= -\Sigma^\bullet_{\tau|t+1}\frac{\partial h^\bullet_{\tau|t+1}}{\partial \alpha_{t+1}} + (-\Sigma^\bullet_{\tau|t+1}\bar{H}_{\tau-1,\tau})(-\Sigma^\bullet_{\tau-1|t+1}\bar{H}_{\tau-1,\tau})\frac{\partial \log \Sigma^\bullet_{\tau-1|t+1}}{\partial \alpha_{t+1}}.
\end{aligned} \tag{61}$$

Taking $t = 1$ and $\alpha_{t+1} = \alpha^\circ_{t+1}$ in (60) and using (56) gives

$$\dot{s}_1 = \Sigma_1 \psi_1 \dot{\alpha}_1$$

Taking $\tau = t$ and $\alpha_{t+1} = \alpha^\circ_{t+1}$ in (61) and using (52) and (56) gives

$$\dot{s}_t = \Sigma_t \psi_t \dot{\alpha}_t + \gamma_t \dot{\alpha}_{t-1}\dot{\alpha}_t \dot{s}_{t-1}.$$

Since t is arbitrary, the two previous equations give the result of equation (13).

B.2. **Second derivatives.** Taking derivatives of equation (60) and the first equation of (61) with respect to α_{t+1} gives

$$\begin{aligned}\frac{\partial^2 \log \Sigma^\bullet_{1|t+1}}{\partial \alpha^2_{t+1}} &= -\frac{\partial \Sigma^\bullet_{1|t+1}}{\partial \alpha_{t+1}} \frac{\partial h^\bullet_{1|t+1}}{\partial \alpha_{t+1}} - \Sigma^\bullet_{1|t+1} \frac{\partial^2 h^\bullet_{1|t+1}}{\alpha^2_{t+1}} \\ &= \left(\frac{\partial \log \Sigma^\bullet_{1|t+1}}{\partial \alpha_{t+1}} \right)^2 - \Sigma^\bullet_{1|t+1} \frac{\partial^2 h^\bullet_{1|t+1}}{\alpha^2_{t+1}},\end{aligned} \tag{62}$$

$$\begin{aligned}\frac{\partial^2 \log \Sigma^\bullet_{\tau|t+1}}{\partial \alpha^2_{t+1}} &= -\frac{\partial \Sigma^\bullet_{\tau|t+1}}{\partial \alpha_{t+1}} \left[\frac{\partial h^\bullet_{\tau|t+1}}{\alpha_{t+1}} - \bar{H}^2_{\tau-1,\tau} \cdot \frac{\partial \Sigma^\bullet_{\tau-1|t+1}}{\partial \alpha_{t+1}} \right] \\ &\quad -\Sigma^\bullet_{\tau|t+1} \left[\frac{\partial^2 h^\bullet_{\tau|t+1}}{\alpha^2_{t+1}} - \bar{H}^2_{\tau-1,\tau} \cdot \frac{\partial^2 \Sigma^\bullet_{\tau-1|t+1}}{\partial \alpha^2_{t+1}} \right] \\ &= \left(\frac{\partial \log \Sigma^\bullet_{\tau|t+1}}{\partial \alpha_{t+1}} \right)^2 - \Sigma^\bullet_{\tau|t+1} \frac{\partial^2 h^\bullet_{\tau|t+1}}{\alpha^2_{t+1}} \\ &\quad +(-\Sigma^\bullet_{\tau,t+1} \bar{H}_{\tau-1,\tau})(-\Sigma^\bullet_{\tau-1,t+1} \bar{H}_{\tau-1,\tau}) \left[\frac{\partial^2 \log \Sigma^\bullet_{\tau-1|t+1}}{\partial \alpha^2_{t+1}} + \left(\frac{\partial \log \Sigma^\bullet_{\tau-1|t+1}}{\partial \alpha_{t+1}} \right)^2 \right],\end{aligned} \tag{63}$$

where we used the following identity to compute the second equality of equation (63):

$$\frac{\partial^2 y}{\partial x^2} = y \left[\frac{\partial^2 \log y}{\partial x^2} + \left(\frac{\partial \log y}{\partial x} \right)^2 \right].$$

Taking $t = 1$ and $\alpha_{t+1} = \alpha^\circ_{t+1}$ in (62) and using (57) gives

$$\ddot{s}_1 = \dot{s}^2_1 + \Sigma_1 (\psi'_1 \dot{\alpha}^2_1 + \psi_1 \ddot{\alpha}_1).$$

Taking $\tau = t$ and $\alpha_{t+1} = \alpha^\circ_{t+1}$ in (63) and using (53) and (57) gives

$$\ddot{s}_t = \dot{s}^2_t + \Sigma_t (\psi'_t \dot{\alpha}^2_t + \psi_t \ddot{\alpha}_t) + \gamma_t \dot{\alpha}_{t-1} (\dot{\alpha}^2_t \ddot{s}_{t-1} + \dot{s}_{t-1} \ddot{\alpha}_t + \dot{s}^2_{t-1} \dot{\alpha}^2_t).$$

Since t is arbitrary, the two previous equations give the result of equation (14).

References

Carter, C. K., and Kohn, R. (1994). 'On Gibbs Sampling for State Space Models', *Biometrika*, 81(3): 541–553.

de Jong, P., and Shephard, N. (1995). 'The Simulation Smoother for Time Series Models', *Biometrika*, 82(1): 339–350.

Durbin, J., and Koopman, S. J. (1997). 'Monte Carlo maximum likelihood estimation for non-Gaussian state space models', *Biometrika*, 84(3): 669–684.

——— (2002). 'A Simple and Efficient Simulation Smoother for State Space Time Series Analysis', *Biometrika*, 89(3): 603–615.

Frühwirth-Schnatter, S. (1994). 'Data augmentation and Dynamic Linear Models', *Journal of Time Series Analysis*, 15: 183–202.

Gamerman, D. (1998). 'Markov chain Monte Carlo for dynamic generalised linear models', *Biometrika*, 85: 215–227.

Geweke, J. (1989). 'Bayesian Inference in Econometric Models Using Monte Carlo Integration', *Econometrica*, 57: 1317–1340.

Jacquier, E., Polson, N., and Rossi, P. (1994). 'Bayesian Analysis of Stochastic Volatility Models', *Journal of Business and Economic Statistics*, 12(4): 371–388.

Kim, S., Shephard, N., and Chib, S. (1998). 'Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models', *Review of Economic Studies*, 65(3): 361–393.

Liesenfeld, R., and Richard, J. (2003). 'Univariate and Multivariate Stochastic Volatility Models: Estimation and Diagnostics', *Journal of Empirical Finance*, 10: 505–531.

——— (2008). 'Classical and Bayesian Analysis of Univariate and Multivariate Stochastic Volatility Models', *Working paper*.

McCausland, W. J., Miller, S., and Pelletier, D. (2007). 'A New Approach to Drawing States in State Space Models', Cahiers de recherche du Département de sciences économiques, Université de Montréal, no. 2007-06.

Pitt, M. K. (2000). 'Discussion on "Time series analysis on non-Guassian observations based on state space models from both classical and Bayesian perspectives," (by J. Durbin and S. J. Koopman)', *Journal of the Royal Statistical Society Series B*, pp. 38–39.

Richard, J., and Zhang, W. (forthcoming). 'Efficient High-Dimensional Importance Sampling', *Journal of Econometrics*.

Roberts, G. O., and Rosenthal, J. S. (1998). 'Markov-chain Monte Carlo: some practical implications of theoretical results (with discussion)', *Canadian Journal of Statistics*, 26: 5–31.

Rue, H. (2001). 'Fast Sampling of Gaussian Markov Random Fields with Applications', *Journal of the Royal Statistical Society Series B*, 63: 325–338.

Rue, H., Steinsland, I., and Erland, S. (2004). 'Approximating hidden Gaussian Markov random fields', *Journal of the Royal Statistical Society Series B*, 66.

Shephard, N., and Pitt, M. K. (1997). 'Likelihood Analysis of Non-Gaussian Measurement Time Series', *Biometrika*, 84(3): 653–667.

Vandebril, R., Mastronardi, N., and Van Barel, M. (2007). 'A Levinson-like algorithm for symmetric strongly nonsingular higher order semiseparable plus band matrices', *Journal of Computational and Applied Mathematics*, 198(1): 74–97.