Journal of Financial Econometrics Advance Access published January 30, 2014

Journal of Financial Econometrics, 2014, Vol. 0, No. 0, 1–34

The HESSIAN Method for Models with Leverage-like Effects

BARNABÉ DJEGNÉNÉ and WILLIAM J. MCCAUSLAND

Université de Montréal, Département de sciences économiques and CIREQ

ABSTRACT

We propose a new method for simulation smoothing in state space models with univariate states and conditional dependence between the observation y_t and the contemporaneous innovation of the state equation. Stochastic volatility models with the leverage effect are a leading example. Our method extends the HESSIAN method of McCausland (2012, *Journal of Econometrics*, 168, 189–206), which required conditional independence between y_t and the state innovation. Our generic method is more numerically efficient than the model-specific methods of Omori *et al*. (2007, *J. Fin. Econ.*, 140, 425–449)—for a stochastic volatility model with Gaussian innovations—and Nakajima and Omori (2009, *Comput. Stat. Data Anal.*, 53, 2335–2353)—for a model with Student's *t* innovations. (*JEL*: C11, C15, C58, C63)

KEYWORDS: state space models, MCMC, numerical efficiency, stochastic volatility, leverage effect

State space models govern the joint evolution of latent states $\alpha=(\alpha_1,\ldots,\alpha_t,\ldots,\alpha_n)^\top$ and observable data $y=(y_1^\top,\ldots,y_t^\top,\ldots,y_n^\top)^\top$ given a vector θ of parameters. They are very useful in capturing dynamic relationships, especially where there are changing, but latent, economic conditions: the states may be unobserved state variables in macroeconomic models, log volatility in asset markets or time varying model parameters.

Simulation smoothing methods have proven useful for approximating likelihood function values and for Bayesian posterior simulation. They involve simulating the conditional distribution of states given data and parameters, which we will call the *target* distribution. Simulation typically entails importance sampling (IS) or Markov chain Monte Carlo (MCMC). We show examples of both in Section 2.

We thank our reviewers and John Geweke for helpful comments.

doi:10.1093/jjfinec/nbt027

State space models with conditional dependence between the observed value y_t and the contemporaneous *innovation* of the state equation, not just the contemporaneous state α_t, are of particular interest. The best known examples are stochastic volatility models with an asymmetric volatility effect known as the leverage effect. In the model introduced by Harvey and Shephard (1996), the latent states α_t are log volatilities, given by

$$\alpha_1 = \bar{\alpha} + \frac{\sigma}{\sqrt{1-\phi^2}} u_0, \qquad \alpha_{t+1} = (1-\phi)\bar{\alpha} + \phi\alpha_t + \sigma u_t, \tag{1}$$

and observed returns y_t are given by

$$y_t = \exp(\alpha_t/2) v_t, \tag{2}$$

where the (u_t, v_t) are serially independent with

$$u_0 \sim N(0,1), \qquad \begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim \text{i.i.d. N}\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right), \tag{3}$$

and $(\sigma, \phi, \rho, \bar{\alpha})$ is a vector of parameters. If $\rho = 0$, y_t and the contemporaneous innovation σu_t are conditionally independent given α_t. When $\rho \neq 0$, they are conditionally dependent and we call this dependence a leverage effect.

Others have extended this model. Jacquier, Polson, and Rossi (2004) and Omori *et al*. (2007) consider inference in stochastic volatility models with leverage and heavy-tailed conditional return distributions. This and other works have shown convincingly that stochastic volatility models with leverage are more realistic than models without. Leverage-like effects may be useful in other models as well. There is little reason beyond computational convenience to rule them out. Feng, Jiang, and Song (2004) show that conditional dependence is more realistic in stochastic conditional duration models.

Designing inferential methods for such models has proven difficult, however, and methods with high numerical efficiency have been model-specific. Nine years passed between Kim, Shephard, and Chib (1998), introducing the auxiliary mixture model approach for stochastic volatility models without leverage, and Omori *et al*. (2007), extending it to models with leverage.

We extend the HESSIAN method of McCausland (2012), which required models without leverage-like effects. That method used multiple derivatives of $\log f(y_t|\theta, \alpha_t)$ with respect to α_t to construct a close approximation to the target distribution. In models with leverage-like effects, the conditional distribution of y_t given α depends not only on α_t but also α_{t+1}. To obtain as good an approximation, we need multiple partial derivatives of $\log f(y_t|\theta, \alpha_t, \alpha_{t+1})$ with respect to α_t and α_{t+1}. Another difficulty is that all non-zero elements of the Hessian of the log target density depend on α, not just the diagonal elements.

Our method inherits the following features of the original method: first, it involves direct simulation of states from their posterior distribution using a proposal or importance distribution approximating the target distribution. This is unlike auxiliary mixture model approaches, in which a model is first transformed into a linear model, and then any non-Gaussian distributions in the transformed model are approximated by finite Gaussian mixtures. Kim, Shephard, and Chib (1998), Chib, Nardari, and Shephard (2002), Omori *et al*. (2007) use this auxiliary mixture model approach for stochastic volatility models; Stroud, Müller, and Polson (2003), Frühwirth-Schnatter and Wagner (2006), and Frühwirth-Schnatter *et al*. (2009) use it for other nonlinear non-Gaussian state space models. Using the direct approach we avoid model-specific transformations, data augmentation and the need to re-weight or apply additional accept-reject steps to correct for approximation error.

Second, it involves drawing the entire state sequence as a single MCMC block. This leads to efficiency improvements when there is posterior serial dependence. We make this possible by constructing a non-Gaussian approximation of the target distribution that is much closer than any multivariate Gaussian distribution. Many articles have used multivariate Gaussian proposals to update the state vector, but usually only for about 10–50 observations at a time. These include Shephard and Pitt (1997), Watanabe and Omori (2004), Strickland, Forbes, and Martin (2006), Jungbacker and Koopman (2008), and Omori and Watanabe (2008). The Efficient Importance Sampling (EIS) method of Richard and Zhang (2007) features a non-Gaussian draw of the entire state sequence, but since their approximation of the target distribution is constructed using the random numbers used to draw from it, EIS estimators of likelihood function values do not have the simulation consistency or lack of simulation bias that true importance sampling estimators have. See the discussion in McCausland (2012) for more details.

Third, since the approximation is so close, we can draw parameters and states jointly using a proposal distribution combining our approximation of the conditional posterior distribution of states given parameters with an approximation of the marginal posterior distribution of parameters. Drawing states and parameters in a single block leads to further efficiency improvements because of posterior dependence between states and parameters. Thus we achieve numerical efficiencies comparable to model-specific auxiliary mixture model approaches featuring joint draws. The examples of Section 2 suggest that our method is even more efficient, partly because we avoid data augmentation and the need to correct for approximation error. Being able to draw all parameters and states jointly in an untransformed model also opens up new opportunities—for importance sampling, variance reduction using randomised pseudo Monte Carlo, and very efficient marginal likelihood approximations, as we see in Section 2.

Fourth, we construct our approximation of the target distribution in a generic way. The only model-specific computation is the evaluation of derivatives of the log measurement density. Existing, well tested, and publicly available generic code uses output of the model-specific computation in order to do simulation smoothing for that model. Exact evaluation of derivatives does not require finding analytic expressions—we can use generic routines to combine derivative values according to Leibniz' rule for multiple derivatives of products and Faà di Bruno's rule for multiple derivatives of composite functions. Although we do not do so here, we could also resort to numerical derivatives—there would be a cost in numerical efficiency, but simulation consistency would not be compromised. The Student's t distribution and other scale mixtures of normals are often used in stochastic volatility models, partly because they work well in auxiliary mixture model approaches using data augmentation for the mixing random variables. A generic approach allows for other, possibly skewed, measurement distributions.

Fifth, it is based on operations using the sparse Hessian matrix of the log target density rather than on the Kalman filter. Articles using the former approach include Rue (2001), for linear Gaussian Markov random fields, Chan and Jeliazkov (2009) and McCausland, Miller, and Pelletier (2011), for linear Gaussian state space models, and Rue, Martino, and Chopin (2009) for nonlinear non-Gaussian Markov random fields. The Integrated Nested Laplace Approximation (INLA) method described in the last article has spawned a large applied literature. Articles using the Kalman filter include Carter and Kohn (1994), Frühwirth-Schnatter (1994), de Jong and Shephard (1995), and Durbin and Koopman (2002) for linear Gaussian state space models. Auxiliary mixture model methods for nonlinear or non-Gaussian models tend to use the Kalman filter, but this is not an essential feature of these methods.

We will now be more precise about the class of state space models we consider. The state and measurement equations are

$$\alpha_1 = d_0 + \omega_0^{-1/2} u_0, \qquad \alpha_{t+1} = d_t + \phi_t \alpha_t + \omega_t^{-1/2} u_t, \qquad u_t \sim \mathrm{iid}\, \mathrm{N}(0,1), \tag{4}$$

$$f(y|\alpha) = \left[\prod_{t=1}^{n-1} f(y_t|\alpha_t, \alpha_{t+1})\right] f(y_n|\alpha_n), \tag{5}$$

where $\alpha \equiv (\alpha_1, \ldots, \alpha_n)$ is a vector of univariate latent states α_t, the y_t are observable random vectors and the $f(y_t|\alpha_t, \alpha_{t+1})$ are density or mass functions. The d_t, ϕ_t and ω_t are parameters of the state dynamics and do not depend on y. We say that models of this form exhibit a leverage-like effect whenever $f(y_t|\alpha_t, \alpha_{t+1})$ depends on α_{t+1}. This will be the case when the observable vector y_t and the contemporaneous state innovation $u_t = \alpha_{t+1} - d_t - \phi_t \alpha_t$ are conditionally dependent given α_t.

Appendix A gives an equivalent specification of (4) in terms of a tridiagonal precision matrix $\bar{\Omega}$ and a covector $\bar{c}$, giving the marginal distribution of α as $\alpha \sim N(\bar{\Omega}^{-1}\bar{c}, \bar{\Omega}^{-1})$. Formulas for $\bar{\Omega}$ and $\bar{c}$ in terms of the h_t, ϕ_t and ω_t are provided there. Throughout most of the study, we condition on $\bar{\Omega}$ and $\bar{c}$ and any parameters on which the $f(y_t|\alpha_t, \alpha_{t+1})$ might depend and suppress notation for this conditioning. In Section 2, where we consider joint inference for parameters and states, we are explicit about this conditioning.

The model in equations (1), (2), and (3) is of the form given by (4) and (5). Take $d_0=\bar{\alpha}$, $\omega_0=\sigma^{-2}(1-\phi^2)$, and for $t>0$, $d_t=(1-\phi)\bar{\alpha}$, $\phi_t=\phi$, $\omega_t=\sigma^{-2}$. Then use (1) to write $u_t=[\alpha_{t+1}-(1-\phi)\bar{\alpha}-\phi\alpha_t]/\sigma$ and then the standard formula for conditional Gaussian distributions to obtain

$$y_t|\alpha \sim \mathrm{N}\Big((\rho/\sigma)\exp(\alpha_t/2)(\alpha_{t+1}-(1-\phi)\bar{\alpha}-\phi\alpha_t),(1-\rho^2)\exp(\alpha_t)\Big). \tag{6}$$

In Section 1 we describe our approximation $g(\alpha|y)$ of the target density $f(\alpha|y)$. We show how to evaluate it and draw from the distribution with density $g(\alpha|y)$. Section 2 illustrates our methods using stochastic volatility models with leverage, both with Gaussian and Student's t measurement innovations. Section 3 concludes.

1 AN APPROXIMATION OF THE TARGET DENSITY

Here we show how to construct our approximation $g(\alpha|y)$ of the target density $f(\alpha|y)$. To use $g(\alpha|y)$ as a proposal or importance density, we must be able to evaluate the fully normalized $g(\alpha|y)$ and make exact draws from the corresponding distribution. Typically, we do both at once, computing $g(\alpha|y)$ at the value we draw. Table 1 summarizes the notation of this section.

We take as input a complete specification of the state space model in (4) and (5), for example (1), (2), and (3). The state dynamics in (4) are determined by specifying values for $\bar{\Omega}$ and $\bar{c}$, which Appendix A shows how to compute. We specify the measurement densities in (5) by supplying code to evaluate, at arbitrary values of α_t, α_{t+1} and α_n, the functions $\psi_t(\alpha_t,\alpha_{t+1}) \doteq \log f(y_t|\alpha_t,\alpha_{t+1})$, $t=1,\dots,n-1$, and $\psi_n(\alpha_n) \doteq \log f(y_n|\alpha_n)$, together with partial derivatives

$$\psi_t^{(p,q)}(\alpha_t,\alpha_{t+1}) \doteq \frac{\partial^{p+q}\psi_t(\alpha_t,\alpha_{t+1})}{\partial\alpha_t^p\,\partial\alpha_{t+1}^q}, \quad \psi_n^{(p)}(\alpha_n) \doteq \frac{\partial^p\psi(\alpha_n)}{\partial\alpha_n^p}, \tag{7}$$

at various orders p and q. Generic code calls these model-specific routines from time to time, passing them points of evaluation (α_t,α_{t+1}) or α_n. In this article, we compute derivatives up to orders $p=6$ and $q=6$, enough for a very close approximation.

The approximation $g(\alpha|y)$ has the Markov property, just as the target $f(\alpha|y)$ does, so we can decompose it as

$$g(\alpha|y)=g(\alpha_n|y)\prod_{t=n-1}^{1} g(\alpha_t|\alpha_{t+1},y). \tag{8}$$

We restrict attention here to the construction of $g(\alpha_t|\alpha_{t+1},y)$ as an approximation of $f(\alpha_t|\alpha_{t+1},y)$ for $t=2,\dots,n-1$. Similar constructions of $g(\alpha_1|\alpha_2,y)$ and $g(\alpha_n|y)$ are given in the appendices.

Table 1 Main notation used in the study

Notation	Description
$\psi_t(\alpha_t,\alpha_{t+1})$	$\log f(y_t\|\alpha_t,\alpha_{t+1})$
$\psi_t^{(p,q)}(\alpha_t,\alpha_{t+1})$	order (p,q) derivative of $\psi_t(\alpha_t,\alpha_{t+1})$ with respect to α_t and α_{t+1}.
$\psi_n(\alpha_n)$	$\log f(y_n\|\alpha_n)$
$\psi_n^{(p)}(\alpha_n)$	order p derivative of $\psi_n(\alpha_n)$ with respect to α_n
$a=(a_1,\ldots,a_n)$	mode of $\log f(\alpha\|y)$
Σ_t	$\mathrm{Var}(\alpha_t\|\alpha_{t+1},y)$ according to Gaussian approximation of $\alpha\|y$
$(a_{1\|t+1}(\alpha_{t+1}),\ldots,a_{t\|t+1}(\alpha_{t+1}))$	mode of $f(\alpha_1,\ldots,\alpha_t\|\alpha_{t+1},y)$
$\Sigma_{t\|t+1}(\alpha_{t+1})$	$\mathrm{Var}(\alpha_t\|\alpha_{t+1},y)$ according to Gaussian approximation of $\alpha_1,\ldots,\alpha_t\|\alpha_{t+1},y$
$A_{t\|t+1}(\alpha_{t+1})$	polynomial approximation of $a_{t\|t+1}(\alpha_{t+1})$
$s_{t\|t+1}(\alpha_{t+1})$	$\log\Sigma_{t\|t+1}(\alpha_{t+1})$
$a_t^{(r)}$	order r derivative of $a_{t\|t+1}(\alpha_{t+1})$ at $\alpha_{t+1}=a_{t+1}$
$s_t^{(r)}$	order r derivative of $s_{t\|t+1}(\alpha_{t+1})$ at $\alpha_{t+1}=a_{t+1}$.
$b_{t\|t+1}(\alpha_{t+1})$	mode of $f(\alpha_t\|\alpha_{t+1},y)$
$b_t,b_t^{(r)}$	value, order r derivative of $b_{t\|t+1}(\alpha_{t+1})$ at $\alpha_{t+1}=a_{t+1}$
b_n	mode of $f(\alpha_n\|y)$
$B_{t\|t+1}(\alpha_{t+1})$	polynomial approximation of $b_{t\|t+1}(\alpha_{t+1})$
$B_t,B_t^{(r)}$	value, order r derivative of $B_{t\|t+1}(\alpha_{t+1})$ at $\alpha_{t+1}=a_{t+1}$
$\mu_{t\|t+1}(\alpha_{t+1})$	$E[\alpha_t\|\alpha_{t+1},y]$
$\mu_t,\mu_t^{(r)}$	value, order r derivative of $\mu_{t\|t+1}(\alpha_{t+1})$ at $\alpha_{t+1}=a_{t+1}$
$M_{t\|t+1}(\alpha_{t+1})$	polynomial approximation of $\mu_{t\|t+1}(\alpha_{t+1})$
$M_t,M_t^{(r)}$	value, order r derivative of $M_{t\|t+1}(\alpha_{t+1})$ at $\alpha_{t+1}=a_{t+1}$
$h_t^{(r)}(\alpha_t;\alpha_{t+1})$	order r derivative of $\log f(\alpha_t\|\alpha_{t+1},y)$ with respect to α_t
$H_t^{(r)}(\alpha_t;\alpha_{t+1})$	approximation of $h_t^{(r)}(\alpha_t;\alpha_{t+1})$
$h_t^{(r)}(\alpha_n)$	order r derivative of $\log f(\alpha_n\|y)$
$H_n^{(r)}(\alpha_n)$	approximation of $h_t^{(r)}(\alpha_n)$

Each factor $g(\alpha_t|\alpha_{t+1},y)$ is a member of a parametric family of univariate distributions described in McCausland (2012). One of the parameters gives the mode; others give the second through fifth derivatives of the log density there. For a particular value of α_{t+1}, we specify $g(\alpha_t|\alpha_{t+1},y)$ by selecting values for these parameters. For the parameter giving the mode, we use an approximation $B_{t|t+1}(\alpha_{t+1})$ of the conditional mode of α_t given α_{t+1} and y; we denote the exact conditional mode by $b_{t|t+1}(\alpha_{t+1})$. For the derivative parameters, we use approximations of the second through fifth derivatives of $h_t(\alpha_t;\alpha_{t+1})\doteq\log f(\alpha_t|\alpha_{t+1},y)$ with respect to α_t, evaluated at $B_{t|t+1}(\alpha_{t+1})$.

We still need suitable approximations of the derivatives of $h_t(\alpha_t;\alpha_{t+1})$. These are based on the following exact expression for the first derivative:

$$\begin{aligned} h_t^{(1)}(\alpha_t;\alpha_{t+1}) = &\bar{c}_t - \bar{\Omega}_{t-1,t}\mu_{t-1|t}(\alpha_t) - \bar{\Omega}_{t,t}\alpha_t - \bar{\Omega}_{t,t+1}\alpha_{t+1} \\ &+ \delta_{t-1|t}(\alpha_t) + \psi_t^{(1,0)}(\alpha_t,\alpha_{t+1}), \quad t=2,\ldots,n-1, \end{aligned} \tag{9}$$

where $\mu_{t-1|t}(\alpha_t) \doteq E[\alpha_{t-1}|\alpha_t, y]$, $\delta_{t-1|t}(\alpha_t) \doteq E[\psi_{t-1}^{(0,1)}(\alpha_{t-1},\alpha_t)|\alpha_t, y]$, $\bar{c}_t$ and $\bar{\Omega}_{t,s}$ are respectively elements of the covector $\bar{c}$ and covariance matrix $\bar{\Omega}$. This result and analogous results for $t=1$ and $t=n$ are derived in Appendix C.1.

We cannot evaluate $h_t^{(1)}(\alpha_t;\alpha_{t+1})$ and its derivatives exactly because we cannot evaluate $\mu_{t-1|t}(\alpha_t)$ or $\delta_{t-1|t}(\alpha_t)$ exactly. Nor can we evaluate the conditional mode $b_{t|t+1}(\alpha_{t+1})$ exactly. We will construct polynomial approximations $B_{t|t+1}(\alpha_{t+1})$, $M_{t-1|t}(\alpha_t)$, and $\Delta_{t-1|t}(\alpha_t)$ of $b_{t|t+1}(\alpha_{t+1})$, $\mu_{t-1|t}(\alpha_t)$ and $\delta_{t-1|t}(\alpha_t)$, all based on approximate Taylor expansions. Replacing the functions $\delta_{t-1|t}(\alpha_t)$ and $\mu_{t-1|t}(\alpha_t)$ in Equation (9) by $\Delta_{t-1|t}(\alpha_t)$ and $M_{t-1|t}(\alpha_t)$ gives an approximation of $h_t^{(1)}(\alpha_t;\alpha_{t+1})$ which we denote $H_t^{(1)}(\alpha_t;\alpha_{t+1})$. $H_t^{(1)}(\alpha_t;\alpha_{t+1})$ and higher order derivatives are easy to evaluate.

Computation proceeds according to the following steps. We first find the mode $a=(a_1,\ldots,a_n)$ of the target density $f(\alpha|y)$. Then we compute coefficients of the polynomial approximations $B_{t|t+1}(\alpha_{t+1})$, $M_{t-1|t}(\alpha_t)$ and $\Delta_{t-1|t}(\alpha_t)$ in a forward pass. These functions are approximate Taylor expansions, with the elements a_t and a_{t+1} of the mode as points of expansion. Finally, we evaluate and/or draw from $g(\alpha|y)$ using a backward pass. At step t a value for α_{t+1} is available. We first evaluate $B_{t|t+1}(\alpha_{t+1})$, the approximate conditional mode. Next we evaluate approximate derivatives $H_t^{(r)}(B_{t|t+1}(\alpha_{t+1});\alpha_{t+1})$ of orders $r=2,3,4,5$. The approximate mode and derivatives are used as parameter values for the parametric family of univariate distributions in McCausland (2012). Appendix G of that paper describes how to draw variates and evaluate the density.

The generic part of the computation is documented in detail in various appendices. Appendix A shows how to compute the mode of α given y; some of the computational details are in Appendix B of McCausland (2012). McCausland, Miller, and Pelletier (2011) comment on the higher numerical efficiency of such "precision based methods," relative to the Kalman filter. This numerical efficiency is an advantage, but a small one here, since computing the mode is not a large fraction of the computational cost. Appendix B describes how to compute the coefficients of the Taylor expansions of $a_{t|t+1}(\alpha_{t+1})$ and $\Sigma_{t|t+1}(\alpha_{t+1})$. These functions, which are used to construct $B_{t|t+1}(\alpha_{t+1})$, are the mean and variance of α_t according to a Gaussian approximation whose log density has the same gradient and Hessian as $\log f(\alpha_1,\ldots,\alpha_t|\alpha_{t+1},y)$ at the mode of $f(\alpha_1,\ldots,\alpha_t|\alpha_{t+1},y)$. Appendix C describes how to compute the coefficients of $\Delta_{t-1|t}(\alpha_t)$, $B_{t|t+1}(\alpha_{t+1})$ and $M_{t-1|t}(\alpha_t)$.

On request we will provide code for the generic computation, as well as code computing the required partial derivatives for the two models of the next section. Appendix D documents the computation of these partial derivatives. Code for the two models and Appendix D should be useful as an example for anyone wishing to construct an approximation $g(\alpha|y)$ for a new model.

2 EMPIRICAL EXAMPLE

2.1 Models

We consider two stochastic volatility models with leverage or asymmetric stochastic volatility (ASV) models. The first, ASV-Gaussian, is the basic model of equations (1), (2), and (3). The second, ASV-Student, replaces observation equation (2) with

$$y_t = \exp(\alpha_t/2)\frac{v_t}{\sqrt{\lambda_t/\nu}}, \tag{10}$$

where $\lambda_t \sim \chi^2(\nu)$ and the λ_t and (u_t, v_t) are mutually independent.

We integrate out λ_t to obtain the conditional distribution of y_t given α_t and α_{t+1}. This is a scaled noncentral Student's t; to see this, write $y_t = \exp(\alpha_t/2)\sqrt{1-\rho^2}X$, where

$$X \doteq \frac{u_t/\sqrt{1-\rho^2}}{\sqrt{\lambda_t/\nu}}.$$

Now condition on α_t and α_{t+1}. The numerator and denominator are independent; the numerator is Gaussian with mean $\mu \doteq \rho(1-\rho^2)^{-1/2}\sigma^{-1}(\alpha_{t+1} - (1-\phi)\bar{\alpha} - \phi\alpha_t)$ and unit variance; and λ_t is chi-squared with ν degrees of freedom. Therefore X is noncentral Student's t with noncentrality parameter μ and ν degrees of freedom. Its density is

$$\begin{aligned} f_X(x;\nu,\mu) &= \frac{\nu^{\nu/2}}{2^\nu}\frac{\Gamma(\nu+1)}{\Gamma(\nu/2)}\exp(-\mu^2/2)(\nu+x^2)^{-\nu/2} \\ &\quad \times \left[\frac{\sqrt{2}\mu x}{\nu+x^2}\frac{M\left(\frac{\nu}{2}+1;\frac{3}{2};\frac{\mu^2x^2}{2(\nu+x^2)}\right)}{\Gamma(\frac{\nu+1}{2})} + \frac{1}{\sqrt{\nu+x^2}}\frac{M\left(\frac{\nu+1}{2};\frac{1}{2};\frac{\mu^2x^2}{2(\nu+x^2)}\right)}{\Gamma(\nu/2+1)} \right], \end{aligned} \tag{11}$$

where $\Gamma(\nu)$ is the gamma function and $M(a;b;z)$ is Kummer's function of the first kind, a confluent hypergeometric function. See Scharf (1991). We obtain $f(y_t|\alpha_t,\alpha_{t+1})$ using the change of variables $y_t = \exp(\alpha_t/2)\sqrt{1-\rho^2}X$. The log conditional density $\psi_t(\alpha_t,\alpha_{t+1}) \equiv \log f(y_t|\alpha_t,\alpha_{t+1})$ and its derivatives are given in Appendix D.

For both models, the state equation parameters are $\omega_t = \sigma^{-2}$, $\phi_t = \phi$ and $d_t = (1-\phi)\bar{\alpha}$ for all $t > 1$. The marginal distribution of the initial state α_1 is the stationary distribution, so that $\omega_0 = (1-\phi^2)\sigma^{-2}$ and $d_0 = \bar{\alpha}$.

Our prior is a multivariate Gaussian distribution over the transformed parameter vector θ:

$$\theta \doteq \begin{bmatrix} \log\sigma \\ \tanh^{-1}\phi \\ \bar{\alpha} \\ \tanh^{-1}\rho \\ \log\nu \end{bmatrix} \sim N\left(\begin{bmatrix} -1.8 \\ 2.1 \\ -11.0 \\ -0.4 \\ 2.5 \end{bmatrix}, \begin{bmatrix} 0.125 & -0.05 & 0 & 0 & 0 \\ -0.05 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0 & 0.25 \end{bmatrix} \right).$$

The marginal distribution of $(\log\sigma, \tanh^{-1}\phi, \bar{\alpha}, \log\nu)$ is the same as the prior in McCausland (2012) for a Student's t stochastic volatility model without leverage, and is based on a prior predictive analysis. The parameter $\tanh^{-1}\rho$ is Gaussian and *a priori* independent with mean –0.4 and standard deviation 0.5. This implies prior quantiles 0.1, 0.5, and 0.9 for ρ approximately equal to –0.78, –0.38, and 0.23.

2.2 MCMC and IS Methods for Posterior Simulation

We illustrate the performance of the HESSIAN approximation, using Markov chain Monte Carlo (MCMC) and importance sampling, (IS) comparing with Omori *et al.* (2007). In both cases, we draw θ and α jointly. We use as proposal density (resp. importance density) $g(\alpha,\theta|y) = g(\alpha|\theta,y)g(\theta|y)$, based on an approximation $g(\theta|y)$ of $f(\theta|y)$ described next and the HESSIAN approximation $g(\alpha|\theta,y)$ of $f(\alpha|\theta,y)$.

We construct $g(\theta|y)$ as follows. Just as $g(\alpha|\theta,y)$ is a close approximation of $f(\alpha|\theta,y)$, $\tilde{g}(\theta|y) \doteq f(\alpha,\theta,y)/g(\alpha|\theta,y)$ is a good unnormalized approximation of $f(\theta|y)$. Let θ° be the maximizer of $\tilde{g}(\theta|y)$, Σ° be the inverse of the negative Hessian of $\log\tilde{g}(\theta|y)$ at θ°, and n_θ be the dimension of θ, equal to 4 for the Gaussian model and 5 for the Student's t model. We choose $g(\theta|y)$ to be a n_θ-variate Student's t density with location θ°, scale matrix Σ°, and degrees of freedom equal to 30.

In the MCMC posterior simulation we use an independence Metropolis–Hastings chain. The joint proposal $(\alpha^\star,\theta^\star)$ from $g(\theta|y)g(\alpha|\theta,y)$ is accepted with probability

$$\pi(\theta^\star,\alpha^\star,\theta,\alpha) = \min\left[1, \frac{f(\theta^\star)f(\alpha^\star|\theta^\star)f(y|\theta^\star,\alpha^\star)}{f(\theta)f(\alpha|\theta)f(y|\theta,\alpha)} \frac{g(\theta|y)g(\alpha|\theta,y)}{g(\theta^\star|y)g(\alpha^\star|\theta^\star,y)}\right].$$

Since we can approximate well the full posterior density, we can use IS. Unlike MCMC proposals, importance draws do not need to be independent. We exploit this to reduce variance, using a combination of quasi-random and pseudo-random sequences to draw θ. We form M blocks of length S each, for a total of MS draws. S is a power of two, convenient for Sobol quasi-random sequences.

We draw $U^{(m)}$, $m = 1,\ldots,M$, independently from the uniform distribution on the hypercube $(0,1)^{n_\theta}$. For $s = 1,\ldots,S$, $V^{(s)}$ is the s'th element of the n_θ-dimensional Sobol sequence. For $m = 1,\ldots,M$ and $s = 1,\ldots,S$, we compute $U^{(m,s)}$, defined as the modulo 1 sum of $U^{(m)}$ and $V^{(s)}$. Thus $U^{(m,s)}$ is uniformly distributed on $(0,1)^{n_\theta}$ and the M blocks of length S are independent. We use $U^{(m,s)}$ to draw $\theta^{(m,s)}$ from $g(\theta|y)$: use $U^{(m,s)}$ to construct a 6-vector of independent standard Gaussian variates using the inverse cdf method then construct $\theta^{(m,s)}$ by pre-multiplying by the Cholesky decomposition of the scale matrix times $\sqrt{\nu/\omega^2}$, where $\omega^2 \sim \chi^2(\nu)$.

Let $h(\theta,\alpha)$ be any function of interest. The IS estimator for $E[h(\theta,\alpha)|y]$ is N/D, where

$$N \doteq \sum_{m=1}^{M}\sum_{s=1}^{S} w^{(m,s)} h(\theta^{(m,s)},\alpha^{(m,s)}), \quad D \doteq \sum_{m=1}^{M}\sum_{s=1}^{S} w^{(m,s)},$$

and $w^{(m,s)} = f(\theta^{(m,s)},\alpha^{(m,s)},y)/g(\theta^{(m,s)},\alpha^{(m,s)}|y)$. If the posterior mean of $h(\theta,\alpha)$ exists, then the ratio $R = N/D$ is a simulation convergent estimator of $E[h(\theta,\alpha)|y]$.

Following Geweke (1989), we approximate the posterior variance of $h(\theta,\alpha)$ by

$$\hat{\sigma}_h^2 \doteq \frac{\sum_{m=1}^{M}\sum_{s=1}^{S}[w^{(m,s)}(h(\theta^{(m,s)},\alpha^{(m,s)}) - R)]^2}{D^2}.$$

We compute a numerical variance $\hat{\sigma}_R^2$ for R using the delta method: $\hat{\sigma}_R^2 \doteq (\hat{\sigma}_N^2 - 2R\hat{\sigma}_{ND} + R^2\hat{\sigma}_D^2)(MS/D)^2$, where $\hat{\sigma}_N^2$ and $\hat{\sigma}_D^2$ are estimates of the variances of N and D and $\hat{\sigma}_{ND}$ is an estimate of the covariance. Specifically, $\hat{\sigma}_N^2$ is $(1/M)$ times the sample variance of the M independent terms

$$N_m = \frac{1}{S}\sum_{s=1}^{S} w^{(m,s)} h(\theta^{(m,s)},\alpha^{(m,s)}), \quad m = 1,\ldots,M,$$

and analogously for $\hat{\sigma}_D^2$ and $\hat{\sigma}_{ND}$. Then $\hat{\sigma}_h^2/MS\hat{\sigma}_R^2$ is an estimate of the relative numerical efficiency.

2.3 Marginal Likelihood Approximation

Using $g(\theta|y)g(\alpha|\theta,y)$ as an importance density, the mean of the (independent) importance weights is a simulation-consistent and -unbiased estimator of the marginal likelihood. Our close approximation makes the variation in weights extremely small, which implies highly numerically efficient marginal likelihood estimation.

2.4 Getting it Right

We performed extensive testing of our posterior simulators using tests of correctness similar to those described in Geweke (2004). These tests have power against a wide array of conceptual and programming errors. Our results fail to cast doubt on the correctness of our implementation. Interested readers can consult a previous version of this article, available by request from the authors.

2.5 Results

For the ASV-Gaussian model, we report results for our HESSIAN independence Metropolis-Hastings and IS methods and the method of Omori *et al*. (2007). We use the labels *HIM*, *HIS*, and *OCSN* to identify them. We analyze three data sets. The first consists of daily returns of the S&P 500 index from January 2000 to December 2012, a period including the recent global financial crisis. It consists of 3268 daily returns. The second is a sample, used by Yu (2005), of S&P returns from January 1980 to December 1987, for a total of 2022 observations. The third data set consists of 1232 daily returns of the TOPIX index. This data set, used by Omori *et al*. (2007), is available at Nakajima's website http://sites.google.com/site/jnakajimaweb/sv.

For the HIM chain, we discard the first 10 draws and retain the next 12,800. We use $M=100$ and $S=128$ for HIS, giving the same total number of draws. For the OCSN chain, we discard the first 500 values and retain the next 12,800. All three methods are coded in C++. We used a Windows PC with an Intel Core i5 2.90GHz processor. We report execution times for the ASV-Gaussian model. For the 2000-2012 S&P 500 data, they are 134*s*, 138*s*, and 172*s* for HIS, HIM, and OCSN, respectively; for the 1980–1987 S&P 500 data, they are 81*s*, 82*s*, and 111*s*; for the TOPIX data, 45*s*, 45, and 67*s*.

Table 2 summarizes estimation results for the ASV-Gaussian model. The first two columns show numerical estimates of the posterior mean and standard deviation for the various parameters. The third and fourth columns give the numerical standard error (NSE) and the relative numerical efficiency (RNE) of the posterior sample mean. The RNE measures numerical efficiency relative to that of the mean of an iid posterior sample, see Geweke (1989) for more details. We use the results of Section 2.2 to compute the NSE and RNE of the IS chain and the OCSN chain. We use the contributed coda library of the **R** software to compute those of the HESSIAN independence Metropolis–Hastings method. This uses a time series method based on the estimated spectral density at frequency zero. Since we implement the procedure of Omori *et al*. (2007) using the prior described in their article, the actual posterior means for OCSN are slightly different from those of our methods.

The HIS method outperforms the OCSN method in all cases. Its numerical efficiency is higher than that of OCSN, and apart from the unconditional mean parameter $\bar{\alpha}$ of log volatility, at least four times higher. The efficiency of the IS sample means is sometimes greater than 1. This is possible because of the variance reduction achieved by using quasi-random numbers. In addition, the HIS procedure has a lower execution time and thus higher numerical precision per unit time. Except for $\bar{\alpha}$, the HESSIAN independence Metropolis–Hastings method outperforms the OCSN procedure, as measured by relative numerical efficiency.

Table 2 ASV-Gaussian parameter estimation using the HESSIAN method and the OCSN procedure on S&P500s′ data and TOPIX data

Parameters	Mean	Std	NSE	RNE
S& P500:2000-2012				
$\bar{\alpha}$:HIS	−9.1865	0.1092	1.1519e-3	0.9725
$\bar{\alpha}$:HIM	−9.1821	0.1107	1.7398e-3	0.3160
$\bar{\alpha}$:OCSN	−9.1583	0.1424	1.6240e-3	0.7842
ϕ:HIS	0.9819	0.0029	3.0280e-5	0.9589
ϕ:HIM	0.9820	0.0029	4.0109e-5	0.4070
ϕ:OCSN	0.9816	0.0030	3.5407e-5	0.2907
σ:HIS	0.1777	0.0138	1.3714e-4	1.0234
σ:HIM	0.1773	0.0138	2.2144e-4	0.3029
σ:OCSN	0.1830	0.0146	1.7230e-4	0.2139
ρ:HIS	−0.7986	0.0345	2.9533e-4	1.4057
ρ:HIM	−0.7985	0.0343	4.7501e-4	0.4071
ρ:OCSN	−0.8140	0.0353	5.3463e-4	0.3146
S& P500:1980-1987				
$\bar{\alpha}$:HIS	−9.5167	0.1573	2.0113e-3	0.9082
$\bar{\alpha}$:HIM	−9.5181	0.1583	3.1266e-3	0.2002
$\bar{\alpha}$:OCSN	−9.5029	0.3378	3.4767e-3	0.7428
ϕ:HIS	0.9751	0.0080	8.9356e-5	0.9000
ϕ:HIM	0.9752	0.0081	1.3592e-4	0.2765
ϕ:OCSN	0.9776	0.0083	1.8947e-4	0.1506
σ:HIS	0.1524	0.0200	1.9681e-4	0.9871
σ:HIM	0.1521	0.0201	3.2814e-4	0.2919
σ:OCSN	0.1394	0.0203	5.8443e-4	0.0945
ρ:HIS	−0.2032	0.0957	9.2493e-4	1.0647
ρ:HIM	−0.2044	0.0950	1.3265e-3	0.4005
ρ:OCSN	−0.2007	0.1005	1.8453e-3	0.2374
TOPIX				
$\bar{\alpha}$:HIS	−8.8545	0.1080	1.1533e-3	1.2014
$\bar{\alpha}$:HIM	−8.8545	0.1083	1.5951e-3	0.4609
$\bar{\alpha}$:OCSN	−8.8426	0.2172	2.0867e-3	0.8574
ϕ:HIS	0.9574	0.0156	1.5893e-4	0.9537
ϕ:HIM	0.9576	0.0160	2.0428e-4	0.4769
ϕ:OCSN	0.9520	0.0185	3.9992e-4	0.1664
σ:HIS	0.1408	0.0254	2.5871e-4	0.8657
σ:HIM	0.1414	0.0258	2.8818e-4	0.6277
σ:OCSN	0.1387	0.0266	5.9850e-4	0.1556
ρ:HIS	−0.3833	0.1188	1.2561e-3	0.8503
ρ:HIM	−0.3833	0.1195	1.7136e-3	0.3801
ρ:OCSN	−0.3715	0.1231	2.6536e-3	0.1792

The reported posterior means of the parameters ϕ, σ and ρ are similar to the values reported by Omori *et al*. (2007) for the TOPIX index. The difference in the posterior means $\bar{\alpha}$ is due to the fact that these authors measure daily returns in percentages. The same is true for Yu (2005) in the case of the S&P500.

For the ASV-Student model we only report results for the HESSIAN procedures. Table 3 summarizes the results of the three datasets. The estimates of the parameters $\bar{\alpha}$, ϕ, σ, and ρ are close to those obtained with the ASV-Gaussian. The numerical efficiency is also substantially higher.

Comparing parameter estimates using the two S&P500 data sets reveals that the dynamics have changed. Recent data exhibits more turmoil in markets in terms of volatility of volatility and leverage—estimates of σ and ρ obtained with the 2000–2012 sample are higher in absolute values than those obtained with the 1980–1987 sample.

Nakajima and Omori (2009) proposed an extension of the procedure in Omori *et al*. (2007) for ASV-Student and other models. They illustrate the procedure using S&P500 (nominally January 1, 1970 to December 31, 2003) and Topix (January 6, 1992 to December 30, 2004) data. Table 4 and Table 5 in Nakajima and Omori (2009) report results for S&P500 and Topix data, respectively. Numerical efficiency for the ASV-Student model (SVLt in their paper) ranges from 0.006 (ν) to 0.291 (μ) for the S&P500 data set. For the Topix data, the highest value of efficiency reported is 0.0893. To compare efficiency, we measured the numerical efficiency of the HESSIAN method, with randomised pseudo-Monte Carlo IS, on S&P500 data from January 1, 1970 to December 31, 2003. Our sample size is 8586 rather than 8869 reported in Nakajima and Omori (2009). Numerical efficiency ranges from 0.91 (ϕ) to 1.01 (μ).

For the S&P500 data from 1970 to 2003, the log marginal likelihoods are 6595.91 for ASV-Gaussian and 6609.67 for ASV-Student, with numerical standard errors of 0.043 and 0.055. The Bayes factor of exp(13.76) decisively favours the ASV-Student model. Similarly, the more recent S&P dataset, from 2000 to 2012, gives a Bayes factor of exp(3.434) in favor of the ASV-Student model. The log marginal likelihoods of the ASV-Student and ASV-Gaussian models are respectively 10270.01 and 10266.57, with numerical standard errors 0.0096 and 0.0083.

3 CONCLUSION

We have derived an approximation $g(\alpha|\theta, y)$ of the target density $f(\alpha|\theta, y)$ that can be used as a proposal density for MCMC or as an importance density for importance sampling. We have tested the correctness of our posterior simulators. Applications suggests that the HESSIAN method, which is not model-specific, is more numerically efficient than the model-specific method of Omori *et al*. (2007), which is in turn more efficient than the methods of Jacquier, Polson, and Rossi (2004) and Omori and Watanabe (2008). High numerical efficiency relies on $g(\alpha|\theta, y)$ being extremely close to the target $f(\alpha|\theta, y)$. A joint proposal of (θ, α) improves

Table 3 ASV-Student parameter estimation using the HESSIAN method, Independence Metropolis–Hastings, and Importance Sampling, on S&P500s and TOPIX data

Parameters	Mean	Std	NSE	RNE
S& P 500:2000-2012				
$\bar{\alpha}$:HIS	−9.2493	0.1092	1.0716e-3	1.1632
$\bar{\alpha}$:HIM	−9.2469	0.1102	1.8512e-3	0.2769
ϕ:his	0.9826	0.0028	3.1016e-5	0.9160
ϕ:him	0.9826	0.0028	4.0811e-5	0.3687
σ:his	0.1752	0.014	1.5122e-4	0.9369
σ:him	0.1751	0.014	2.5636e-4	0.2298
ρ:his	−0.8255	0.0329	3.3562e-4	1.0204
ρ:him	−0.8245	0.0334	5.4797e-4	0.2899
ν:his	21.1773	6.1814	5.7534e-2	1.2809
ν:him	21.1636	6.1116	1.0145e-1	0.2835
S& P 500: 1980-1987				
$\bar{\alpha}$:HIS	−9.7230	0.1865	2.8719e-3	1.0496
$\bar{\alpha}$:HIM	−9.7224	0.1806	3.1769e-3	0.2525
ϕ:HIS	0.9851	0.0054	6.8752e-5	0.9663
ϕ:HIM	0.9850	0.0053	7.9290e-5	0.3513
σ:HIS	0.1061	0.0164	1.7719e-4	1.1002
σ:HIM	0.1065	0.0164	3.0925e-4	0.2204
ρ:HIS	−0.2440	0.1224	1.6006e-4	0.8261
ρ:HIM	−0.2493	0.1222	2.2437e-3	0.2318
ν:HIS	9.8647	2.1622	2.4734e-2	0.9722
ν:HIM	9.9128	2.1828	3.6789e-2	0.2750
TOPIX				
$\bar{\alpha}$:HIS	−8.9488	0.1156	1.5983e-3	0.9672
$\bar{\alpha}$:HIM	−8.9506	0.1115	1.9474e-3	0.2560
ϕ:HIS	0.9624	0.0142	1.7252e-4	0.8727
ϕ:HIM	0.9621	0.0144	2.2029e-4	0.3336
σ:HIS	0.1261	0.0242	2.6775e-4	0.9570
σ:HIM	0.1266	0.0240	3.7636e-4	0.3188
ρ:HIS	−0.4194	0.1285	1.3790e-4	1.1266
ρ:HIM	−0.4191	0.1236	2.2023e-3	0.2461
ν:HIS	20.6041	7.6904	8.6997e-2	0.9573
ν:HIM	20.4777	7.7394	1.4048e-1	0.2371

efficiency beyond what is achieved by drawing α as a single block and allows for importance sampling with variance reduction and very numerically efficient marginal likelihood approximations.

The scope of applications goes beyond the ASV-Gaussian and ASV-Student models. Application to a new model requires code to compute partial derivatives of the $\log f(y_t|\alpha_t, \alpha_{t+1})$ with respect to α_t and α_{t+1}. This is not as demanding as it might first appear, for two reasons. First, we can use numerical derivatives or other approximations. Second, we do not require analytic expressions; if $\log f(y_t|\alpha_t, \alpha_{t+1})$ is a composition of primitive functions, we can combine evaluations of the derivatives of the primitive functions using routines applying Fàa Di Bruno's rule for multiple derivatives of compound functions. We have already coded these generic routines.

We now require states to be Gaussian. We plan to extend the HESSIAN method to models where they are non-Gaussian but still Markov. We are also working on approximations to filtering densities for sequential learning.

A COMPUTATION OF THE CONDITIONAL MODE OF α GIVEN y.

Here we compute the precision $\bar{\Omega}$ and covector $\bar{c}$ of the marginal distribution of α, and the mode $a=(a_1,\dots,a_n)$ of the target distribution. By-products of the computation of a include several quantities used elsewhere, including $\bar{\bar{\Omega}}$ and $\bar{\bar{c}}$, the precision and covector of a Gaussian approximation $\mathrm{N}(\bar{\bar{\Omega}}^{-1}\bar{\bar{c}}, \bar{\bar{\Omega}}^{-1})$ of the target distribution, and the conditional variances $\Sigma_1,\dots,\Sigma_t,\dots,\Sigma_n$. The precision matrices $\bar{\Omega}$ and $\bar{\bar{\Omega}}$ are both tridiagonal.

As the state dynamics are not different, we compute $\bar{\Omega}$ and $\bar{c}$ exactly as in McCausland (2012):

$$\bar{\Omega}_{t,t} = \omega_{t-1} + \omega_t \phi_t^2, \quad \bar{\Omega}_{t,t+1} = -\omega_t \phi_t, \quad t=1,\dots,n-1,$$

$$\bar{\Omega}_{n,n} = \omega_{n-1},$$

$$\bar{c}_t = \begin{cases} \omega_{t-1} d_{t-1} - \omega_t \phi_t d_t & t=1,\dots,n-1, \\ \omega_{n-1} d_{n-1} & t=n. \end{cases} \tag{A1}$$

As in McCausland (2012), we use a Newton–Raphson method to find the mode of the target distribution. At each iteration, we compute a precision $\bar{\bar{\Omega}}(\alpha)$ and covector $\bar{\bar{c}}(\alpha)$ of a Gaussian approximation to the target distribution based on a second-order Taylor series expansion of the log target density around the current value of α. Specifically, $\bar{\bar{\Omega}}(\alpha)$ is the negative Hessian matrix of $\log f(\alpha|y)$ with respect to α at the current value of α. It is a symmetric tri-diagonal matrix with nonzero upper triangular elements given by

$$\bar{\bar{\Omega}}_{t,t}(\alpha) = \bar{\Omega}_{t,t} - \left(\psi_t^{(2,0)}(\alpha_t, \alpha_{t+1}) + \psi_{t-1}^{(0,2)}(\alpha_{t-1}, \alpha_t) \right), \quad t=2,\dots,n-1,$$

$$\bar{\bar{\Omega}}_{1,1}(\alpha) = \bar{\Omega}_{1,1} - \psi_t^{(2,0)}(\alpha_t, \alpha_{t+1}), \quad \bar{\bar{\Omega}}_{nn}(\alpha) = \bar{\Omega}_{n,n} - \left(\psi_n^{(2)}(\alpha_n) + \psi_{n-1}^{(0,2)}(\alpha_{n-1}, \alpha_n) \right),$$

$$\bar{\bar{\Omega}}_{t,t+1}(\alpha) = \bar{\Omega}_{t,t+1} - \psi_t^{(1,1)}(\alpha_t, \alpha_{t+1}), \quad t=1,\dots,n-1.$$

The covector $\bar{\bar{c}}(\alpha)$ is

$$\bar{\bar{c}}(\alpha) \doteq \left(\bar{\bar{\Omega}} - \bar{\Omega}\right)\alpha + \bar{c} + \frac{\partial \log f(y|\alpha)}{\partial \alpha^\top},$$

and its elements are

$$\bar{\bar{c}}_t(\alpha) = \begin{cases} \bar{c}_t + \left(\bar{\bar{\Omega}}_{t,t} - \bar{\Omega}_{t,t}\right)\alpha_t + \left(\bar{\bar{\Omega}}_{t,t+1} - \bar{\Omega}_{t,t+1}\right)\alpha_{t+1} + \psi_t^{(1,0)}(\alpha_t,\alpha_{t+1}) & t=1 \\ \bar{c}_t + \left(\bar{\bar{\Omega}}_{t,t-1} - \bar{\Omega}_{t,t-1}\right)\alpha_{t-1} + \left(\bar{\bar{\Omega}}_{t,t} - \bar{\Omega}_{t,t}\right)\alpha_t & \\ \quad + \left(\bar{\bar{\Omega}}_{t,t+1} - \bar{\Omega}_{t,t+1}\right)\alpha_{t+1} + \psi_t^{(1,0)}(\alpha_t,\alpha_{t+1}) + \psi_{t-1}^{(0,1)}(\alpha_{t-1},\alpha_t) & t=2,\ldots,n-1 \\ \bar{c}_t + (\bar{\bar{\Omega}}_{t,t-1} - \bar{\Omega}_{t,t-1})\alpha_{t-1} + (\bar{\bar{\Omega}}_{t,t} - \bar{\Omega}_{t,t})\alpha_t + \psi_t^{(1)}(\alpha_t) + \psi_{t-1}^{(0,1)}(\alpha_{t-1},\alpha_t) & t=n. \end{cases} \tag{A2}$$

Let $\bar{\bar{\Omega}} \doteq \bar{\bar{\Omega}}(a)$ and $\bar{\bar{c}} \doteq \bar{\bar{c}}(a)$. Then the mean (and mode) of the Gaussian approximation $\mathrm{N}(\bar{\bar{\Omega}}^{-1}\bar{\bar{c}}, \bar{\bar{\Omega}}^{-1})$ is a, the mode of the target distribution, and its log density has the same Hessian matrix as the log target density at a.

While these expressions for $\bar{\bar{\Omega}}$ and $\bar{\bar{c}}$ are more complicated than those in McCausland (2012), once we have them we compute the mode a in the same way. Roughly speaking, we iterate the computation $\alpha' = \bar{\bar{\Omega}}(\alpha)^{-1}\bar{\bar{c}}(\alpha)$ until numerical convergence. We use two modifications to this procedure, one to accelerate convergence using higher order derivatives and the other to resort to line searches in the rare cases of nonconvergence.

B POLYNOMIAL APPROXIMATIONS OF $\alpha_{t|t+1}$ AND $s_{t|t+1}$

Here we compute coefficients of exact Taylor expansions of $a_{t|t+1}(\alpha_{t+1})$ and $s_{t|t+1}(\alpha_{t+1})$. These are the conditional mean and log variance of α_t given α_{t+1} for a Gaussian approximation of the conditional distribution of $\alpha_1,\ldots,\alpha_t$ given α_{t+1} and y.

We derive recursive expressions giving coefficients at all orders r. In practice, computational costs rises quickly and benefits diminish quickly in r. We give simplified expressions for $a_t^{(r)} \doteq a_{t|t+1}^{(r)}(a_{t+1})$ up to $r=5$ and $s_t^{(r)} \doteq s_{t|t+1}^{(r)}(a_{t+1})$ up to $r=4$.

The basic strategy involves taking derivatives of two identities. The first is a first-order necessary condition on $a_{t-1|t+1}(\alpha_{t+1})$ and $a_{t|t+1}(\alpha_{t+1})$ for $(a_{1|t+1}(\alpha_{t+1}),\ldots,a_{t|t+1}(\alpha_{t+1}))$ to be the conditional mode of $(\alpha_1,\ldots,\alpha_t)$ given α_{t+1} and y. The second is the identity $a_{t-1|t+1}(\alpha_{t+1}) = a_{t-1|t}(a_{t|t+1}(\alpha_{t+1}))$.

B.1 General Formula

We begin with the case $t=1$. Since $f(\alpha_1|\alpha_2,y) \propto f(\alpha_1,\alpha_2) f(y_1|\alpha_1,\alpha_2)$, we can write

$$\log f(\alpha_1|\alpha_2,y) = -\frac{1}{2}\bar{\Omega}_{1,1}\alpha_1^2 - \bar{\Omega}_{1,2}\alpha_1\alpha_2 + \bar{c}_1\alpha_1 + \log f(y_1|\alpha_1,\alpha_2) + k. \tag{A3}$$

where k does not depend on α_1. The conditional mode $a_{1|2}(\alpha_2)$ maximizes $\log f(\alpha_1|\alpha_2,y)$ and must therefore satisfy

$$-\bar{\Omega}_{1,1}a_{1|2}(\alpha_2)-\bar{\Omega}_{1,2}\alpha_2+\bar{c}_1+\psi_1^{(1,0)}(a_{1|2}(\alpha_2),\alpha_2)=0. \tag{A4}$$

Taking the derivative of (A4) with respect to α_2, and using the definitions $\bar{\bar{\Omega}}_{1,1|2}(\alpha_2)=(\bar{\Omega}_{1,1}-\psi_1^{(2,0)}(a_{1|2}(\alpha_2),\alpha_2))$ and $\bar{\bar{\Omega}}_{1,2|2}(\alpha_2)=\bar{\Omega}_{1,2}-\psi_1^{(1,1)}(a_{1|2}(\alpha_2),\alpha_2)$ gives

$$\bar{\bar{\Omega}}_{1,1|2}(\alpha_2)a_{1|2}^{(1)}(\alpha_2)=-\bar{\bar{\Omega}}_{1,2|2}(\alpha_2). \tag{A5}$$

Solving for $a_{1|2}^{(1)}(\alpha_2)$, we obtain

$$a_{1|2}^{(1)}(\alpha_2)=-\Sigma_{1|2}(\alpha_2)\bar{\bar{\Omega}}_{1,2|2}(\alpha_2), \tag{A6}$$

where $\Sigma_{1|2}(\alpha_2)=\bar{\bar{\Omega}}_{1,1|2}^{-1}(\alpha_2)$. Setting $\alpha_2=a_2$ gives $a_1^{(1)}=-\Sigma_1\bar{\bar{\Omega}}_{1,2}$.

We now derive an expression allowing us to compute $a_1^{(r)}$ in terms of $a_1^{(i)}$, $i<r$. First differentiate (A5) $(r-1)$ times with respect to α_2. Using Leibniz's rule, we obtain

$$\sum_{i=0}^{r-1}\binom{r-1}{i}\bar{\bar{\Omega}}_{1,1|2}^{(r-1-i)}(\alpha_2)a_{1|2}^{(i+1)}(\alpha_2)=-\bar{\bar{\Omega}}_{1,2|2}^{(r-1)}(\alpha_2).$$

Then solving for $a_{1|2}^{(r)}(\alpha_2)$ gives

$$a_{1|2}^{(r)}(\alpha_2)=-\Sigma_{1|2}(\alpha_2)\left[\sum_{i=0}^{r-2}\binom{r-1}{i}\bar{\bar{\Omega}}_{1,1|2}^{(r-1-i)}(\alpha_2)a_{1|2}^{(i+1)}(\alpha_2)+\bar{\bar{\Omega}}_{1,2|2}^{(r-1)}(\alpha_2)\right]. \tag{A7}$$

Finally, we evaluate (A7) at $\alpha_2=a_2$ to obtain

$$a_1^{(r)}=-\Sigma_1\left[\sum_{i=0}^{r-2}\binom{r-1}{i}\bar{\bar{\Omega}}_{1,1}^{(r-1-i)}a_1^{(i+1)}+\bar{\bar{\Omega}}_{1,2}^{(r-1)}\right]. \tag{A8}$$

We now derive an expression relating the $a_1^{(r)}$ and the $s_1^{(r)}$, which we will use to obtain the latter from the former. First recall the definition $\Sigma_{1|2}(\alpha_2)=\exp(s_{1|2}(\alpha_2))$. Using Faà Di Bruno's formula for derivatives of compound functions, we obtain, for $i\geq 1$,

$$\begin{aligned}\Sigma_{1|2}^{(i)}(\alpha_2)&=\sum_{j=1}^{i}\exp(s_{1|2}(\alpha_2))B_{i,j}(s_{1|2}^{(1)}(\alpha_2),\ldots,s_{1|2}^{(i-j+1)}(\alpha_2))\\&=\Sigma_{1|2}(\alpha_2)B_i(s_{1|2}^{(1)}(\alpha_2),\ldots,s_{1|2}^{(i)}(\alpha_2)),\end{aligned} \tag{A9}$$

where the $B_{i,j}$ are Bell polynomials and B_i is the i-th complete Bell polynomial. Appendix E shows how to compute these polynomials. We now differentiate (A6) $(r-1)$ times with respect to α_2, to obtain

$$\begin{aligned} a_{1|2}^{(r)}(\alpha_2) &= -\sum_{i=0}^{r-1} \binom{r-1}{i} \Sigma_{1|2}^{(i)}(\alpha_2) \bar{\bar{\Omega}}_{1,2|2}^{(r-1-i)}(\alpha_2) \\ &= -\Sigma_{1|2}(\alpha_2) \sum_{i=0}^{r-1} \binom{r-1}{i} B_i(s_{1|2}^{(1)}(\alpha_2), \ldots, s_{1|2}^{(i)}(\alpha_2)) \bar{\bar{\Omega}}_{1,2|2}^{(r-1-i)}(\alpha_2). \end{aligned}$$

Evaluating at $\alpha_2 = a_2$ gives us the desired expression:

$$a_1^{(r)} = -\Sigma_1 \sum_{i=0}^{r-1} \binom{r-1}{i} B_i(s_1^{(1)}, \ldots, s_1^{(i)}) \bar{\bar{\Omega}}_{1,2}^{(r-1-i)}. \tag{A10}$$

We now move on to the case $1 < t < n$. The conditional mode $a_{1:t|t+1}(\alpha_{t+1}) = (a_{1|t+1}(\alpha_{t+1}), \ldots, a_{t|t+1}(\alpha_{t+1}))$ must satisfy the first-order necessary condition

$$\begin{aligned} 0 &= \bar{c}_t - \bar{\Omega}_{t-1,t} a_{t-1|t+1}(\alpha_{t+1}) - \bar{\Omega}_{t,t} a_{t|t+1}(\alpha_{t+1}) - \bar{\Omega}_{t,t+1}\alpha_{t+1} \\ &\quad + \psi_{t-1}^{(0,1)}(a_{t-1|t}(a_{t|t+1}), a_{t|t+1}) + \psi_t^{(1,0)}(a_{t|t+1}, \alpha_{t+1}). \end{aligned} \tag{A11}$$

Taking the derivative of (A11) with respect to α_{t+1} gives

$$\bar{\bar{\Omega}}_{t,t-1}(\alpha_{t+1}) a_{t-1|t+1}^{(1)}(\alpha_{t+1}) + \bar{\bar{\Omega}}_{t,t}(\alpha_{t+1}) a_{t|t+1}^{(1)}(\alpha_{t+1}) + \bar{\bar{\Omega}}_{t,t+1}(\alpha_{t+1}) = 0. \tag{A12}$$

Using the identity $a_{t-1|t+1}(\alpha_{t+1}) = a_{t-1|t}\left(a_{t|t+1}(\alpha_{t+1})\right)$ and the chain rule gives

$$a_{t-1|t+1}^{(1)}(\alpha_{t+1}) = a_{t-1|t}^{(1)}(a_{t|t+1}(\alpha_{t+1})) a_{t|t+1}^{(1)}(\alpha_{t+1}). \tag{A13}$$

Substituting (A13) in (A12), we obtain

$$\left(\bar{\bar{\Omega}}_{t,t-1}(\alpha_{t+1}) a_{t-1|t}^{(1)}(a_{t|t+1}(\alpha_{t+1})) + \bar{\bar{\Omega}}_{t,t}(\alpha_{t+1})\right) a_{t|t+1}^{(1)}(\alpha_{t+1}) = -\bar{\bar{\Omega}}_{t,t+1}(\alpha_{t+1}).$$

Then, following an analogous development in McCausland (2012), we can show by induction that

$$a_{t|t+1}^{(1)}(\alpha_{t+1}) = -\Sigma_{t|t+1}(\alpha_{t+1}) \bar{\bar{\Omega}}_{t,t+1}(\alpha_{t+1}), \quad t = 2, \ldots, n-1, \tag{A14}$$

where $\left[\Sigma_{t|t+1}(\alpha_{t+1})\right]^{-1} = \bar{\bar{\Omega}}_{t,t-1}(\alpha_{t+1}) a_{t-1|t}^{(1)}(a_{t|t+1}(\alpha_{t+1})) + \bar{\bar{\Omega}}_{t,t}(\alpha_{t+1})$. Taking $\alpha_{t+1} = a_{t+1}$ in (A14) gives

$$a_t^{(1)} = -\Sigma_t \bar{\bar{\Omega}}_{t,t+1}. \tag{A15}$$

For $r \geq 2$, we use Leibniz's rule to differentiate (A12) $(r-1)$ times with respect to α_{t+1} and obtain

$$\sum_{i=0}^{r-1} \binom{r-1}{i} \left(\bar{\bar{\Omega}}_{t,t-1}^{(i)}(\alpha_{t+1}) a_{t-1|t+1}^{(r-i)}(\alpha_{t+1}) + \bar{\bar{\Omega}}_{t,t}^{(i)}(\alpha_{t+1}) a_{t|t+1}^{(r-i)}(\alpha_{t+1}) \right) = -\bar{\bar{\Omega}}_{t,t+1}^{(r-1)}(\alpha_{t+1}). \tag{A16}$$

Using Faà di Bruno's formula for arbitrary order derivatives of compound functions, we compute the i-th derivative of $a_{t-1|t+1}(\alpha_{t+1})$ with respect to α_{t+1} as

$$a_{t-1|t+1}^{(i)}(\alpha_{t+1}) = \sum_{j=1}^{i} a_{t-1|t}^{(j)}(a_{t|t+1}) B_{i,j}(a_{t|t+1}^{(1)}(\alpha_{t+1}), \ldots, a_{t|t+1}^{(i-j+1)}(\alpha_{t+1})). \tag{A17}$$

If we substitute $a_{t-1|t+1}^{(i)}(\alpha_{t+1})$ of (A17) in (A16) and set $\alpha_{t+1} = a_{t+1}$, we obtain

$$\sum_{i=0}^{r-1} \binom{r-1}{i} \left\{ \bar{\bar{\Omega}}_{t,t-1}^{(i)} \left[\sum_{j=1}^{r-i} a_{t-1}^{(j)} B_{r-i,j}(a_t^{(1)}, \ldots, a_t^{(r-i-j+1)}) \right] + \bar{\bar{\Omega}}_{t,t}^{(i)} a_t^{(r-i)} \right\} = -\bar{\bar{\Omega}}_{t,t+1}^{(r-1)}. \tag{A18}$$

This gives an expression for $a_t^{(r)}$ in terms of $a_t^{(i)}$, $i = 0, \ldots, r-1$; $a_{t-1}^{(i)}$, $i = 0, \ldots, r$; $\bar{\bar{\Omega}}_{t,t-1}^{(i)}$ and $\bar{\bar{\Omega}}_{t,t}^{(i)}$, $i = 1, \ldots, r-1$; and $\bar{\bar{\Omega}}_{t,t+1}^{(r-1)}$.

We now derive a result that will give us $s_t^{(r)}$ in terms of $a_t^{(i)}$ and $s_t^{(i)}$, $i = 1, \ldots, r-1$ and $a_{t-1}^{(i)}$, $i = 1, \ldots, r+1$. Analogously with equation (A9), we have

$$\Sigma_{t|t+1}^{(r)}(\alpha_{t+1}) = \Sigma_{t|t+1}(\alpha_{t+1}) B_r(s_{t|t+1}^{(1)}(\alpha_{t+1}), \ldots, s_{t|t+1}^{(r)}(\alpha_{t+1})).$$

Using Leibniz's rule to take derivatives of (A14) with respect to α_{t+1}, and evaluating at $\alpha_{t+1} = a_{t+1}$, we obtain

$$a_t^{(r)} = \sum_{i=0}^{r-1} \binom{r-1}{i} B_i(s_t^{(1)}, \ldots, s_t^{(i)}) \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(r-1-i)}. \tag{A19}$$

The quantities $\bar{\bar{\Omega}}_{t,s}^{(r)}$ involved in the computation of $a_t^{(r)}$ and $s_t^{(r)}$ are functions of derivatives of $\psi_t^{(p,q)}(a_{t|t+1}, \alpha_{t+1})$ with respect to α_{t+1}, evaluated at a_{t+1}. Equations (A62) and (A63) of Appendix E show how to compute these derivatives as functions of derivatives of $\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$, supplied as part of the model specification.

B.2 Explicit Formula for $R = 5$

We now derive simplified expressions for $a_t^{(r)}$, $r = 1, \ldots, 5$ and $s_t^{(r)}$, $r = 1, \ldots, 4$, for $t = 1, \ldots, n-1$. We give details of the computation for $t = 2, \ldots, n-1$. For the special case $t=1$, we can obtain analogous results simply by setting any terms with a time index of zero to zero.

We have already have an expression for $a_t^{(1)}$, $t=1,\ldots,n-1$, in (A15). Taking $r=2$ in (A18) gives

$$\bar{\bar{\Omega}}_{t,t-1}\left(a_{t-1}^{(1)}a_t^{(2)}+a_{t-1}^{(2)}\left(a_t^{(1)}\right)^2\right)+\bar{\bar{\Omega}}_{t,t}a_t^{(2)}+\bar{\bar{\Omega}}_{t,t-1}^{(1)}a_{t-1}^{(1)}a_t^{(1)}+\bar{\bar{\Omega}}_{t,t}^{(1)}a_t^{(1)}=\bar{\bar{\Omega}}_{t,t+1}^{(1)},$$

which simplifies to

$$a_t^{(2)}=\left(\gamma_t a_t^{(1)}a_{t-1}^{(2)}-\Sigma_t\bar{\bar{\boldsymbol{\Omega}}}_t^{(1)}\right)a_t^{(1)}-\Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(1)}, \tag{A20}$$

where $\gamma_t=-\Sigma_t\bar{\bar{\Omega}}_{t,t-1}$ and $\bar{\bar{\boldsymbol{\Omega}}}_t^{(i)}=\bar{\bar{\Omega}}_{t,t-1}^{(i)}a_t^{(1)}+\bar{\bar{\Omega}}_{t,t}^{(i)}$. Setting $r=2$ in (A19) gives

$$a_t^{(2)}=s_t^{(1)}a_t^{(1)}-\Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(1)}. \tag{A21}$$

Equating the right hand sides of (A20) and (A21) and solving for $s_t^{(1)}$ gives

$$s_t^{(1)}=\gamma_t a_t^{(1)}a_{t-1}^{(2)}-\Sigma_t\bar{\bar{\boldsymbol{\Omega}}}_t^{(1)}. \tag{A22}$$

Setting $r=3$ in (A18) gives

$$\begin{aligned}-\bar{\bar{\Omega}}_{t,t+1}^{(2)}=&\bar{\bar{\Omega}}_{t,t-1}\left(a_{t-1}^{(1)}a_t^{(3)}+3a_{t-1}^{(2)}a_t^{(1)}a_t^{(2)}+a_{t-1}^{(3)}\left(a_t^{(1)}\right)^3\right)+\bar{\bar{\Omega}}_{t,t}a_t^{(3)}\\&+2\left(\bar{\bar{\Omega}}_{t,t-1}^{(1)}\left(a_{t-1}^{(1)}a_t^{(2)}+a_{t-1}^{(2)}\left(a_t^{(1)}\right)^2\right)+\bar{\bar{\Omega}}_{t,t}^{(1)}a_t^{(1)}\right)+\bar{\bar{\Omega}}_{t,t-1}^{(2)}a_{t-1}^{(1)}a_t^{(1)}+\bar{\bar{\Omega}}_{t,t}^{(2)}a_t^{(1)}.\end{aligned}$$

Solving for $a_t^{(3)}$, we obtain

$$\begin{aligned}a_t^{(3)}=&\gamma_t\left(3a_t^{(1)}a_t^{(2)}a_{t-1}^{(2)}+\left(a_t^{(1)}\right)^3a_{t-1}^{(3)}\right)-2\Sigma_t\left(\bar{\bar{\Omega}}_{t,t-1}^{(1)}\left(a_t^{(1)}\right)^2a_{t-1}^{(2)}+\bar{\bar{\boldsymbol{\Omega}}}_t^{(1)}a_t^{(2)}\right)\\&-\Sigma_t\bar{\bar{\boldsymbol{\Omega}}}_t^{(2)}a_t^{(1)}-\Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(2)}\\=&2\left(\gamma_t a_t^{(1)}a_{t-1}^{(2)}-\Sigma_t\bar{\bar{\boldsymbol{\Omega}}}_t^{(1)}\right)a_t^{(2)}+\left(\gamma_t a_t^{(1)}a_{t-1}^{(3)}-2\Sigma_t\bar{\bar{\Omega}}_{t,t-1}^{(1)}a_{t-1}^{(2)}\right)\left(a_t^{(1)}\right)^2\\&+\left(\gamma_t a_t^{(2)}a_{t-1}^{(2)}-\Sigma_t\bar{\bar{\boldsymbol{\Omega}}}_t^{(2)}\right)a_t^{(1)}-\Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(2)}.\end{aligned}$$

We use (A22) to simplify this to

$$\begin{aligned}a_t^{(3)}=&2s_t^{(1)}a_t^{(2)}+\left(\gamma_t a_t^{(1)}a_{t-1}^{(3)}-2\Sigma_t\bar{\bar{\Omega}}_{t,t-1}^{(1)}a_{t-1}^{(2)}\right)\left(a_t^{(1)}\right)^2\\&+\left(\gamma_t a_t^{(2)}a_{t-1}^{(2)}-\Sigma_t\bar{\bar{\boldsymbol{\Omega}}}_t^{(2)}\right)a_t^{(1)}-\Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(2)}.\end{aligned} \tag{A23}$$

Setting $r=3$ in (A19) gives an alternative expression for $a_t^{(3)}$:

$$\begin{aligned}
a_t^{(3)} &= \left(s_t^{(2)} + \left(s_t^{(1)}\right)^2\right) a_t^{(1)} - \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(2)} - 2s_t^{(1)} \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(1)} \\
&= \left(s_t^{(2)} + \left(s_t^{(1)}\right)^2\right) a_t^{(1)} - \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(2)} + 2s_t^{(1)} \left(a_t^{(2)} - s_t^{(1)} a_t^{(1)}\right) \qquad \text{(A24)} \\
&= \left(s_t^{(2)} - \left(s_t^{(1)}\right)^2\right) a_t^{(1)} + 2s_t^{(1)} a_t^{(2)} - \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(2)}.
\end{aligned}$$

Equating the right hand sides of (A23) and (A24) and solving for $s_t^{(2)}$ gives

$$s_t^{(2)} = \left(s_t^{(1)}\right)^2 + \left(\gamma_t a_t^{(1)} a_{t-1}^{(3)} - 2\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(2)}\right) a_t^{(1)} + \left(\gamma_t a_t^{(2)} a_{t-1}^{(2)} - \Sigma_t \bar{\bar{\boldsymbol{\Omega}}}_t^{(2)}\right). \qquad \text{(A25)}$$

We follow a similar procedure to compute the following formulas for $a_t^{(4)}, s_t^{(3)}$, and $a_t^{(5)}, s_t^{(4)}$:

$$\begin{aligned}
a_t^{(4)} &= \left(\gamma_t a_t^{(1)} a_{t-1}^{(4)} - 3\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(3)}\right) \left(a_t^{(1)}\right)^3 + 3\left(\gamma_t a_t^{(2)} a_{t-1}^{(3)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(2)}\right) \left(a_t^{(1)}\right)^2 \\
&\quad + \left(\gamma_t a_t^{(3)} a_{t-1}^{(2)} - 3\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_t^{(2)} a_{t-1}^{(2)} - \Sigma_t \bar{\bar{\boldsymbol{\Omega}}}_t^{(3)}\right) a_t^{(1)} - \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(3)} \\
&\quad + 3\left(s_t^{(2)} - \left(s_t^{(1)}\right)^2\right) a_t^{(2)} + 3s_t^{(1)} a_t^{(3)}, \qquad \text{(A26)}
\end{aligned}$$

$$\begin{aligned}
s_t^{(3)} &= -\left(s_t^{(1)}\right)^3 + 3s_t^{(1)} s_t^{(2)} + \left(\gamma_t a_t^{(1)} a_{t-1}^{(4)} - 3\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(3)}\right) \left(a_t^{(1)}\right)^2 \\
&\quad + 3\left(\gamma_t a_t^{(2)} a_{t-1}^{(3)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(2)}\right) a_t^{(1)} + \left(\gamma_t a_t^{(3)} - 3\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_t^{(2)}\right) a_{t-1}^{(2)} - \Sigma_t \bar{\bar{\boldsymbol{\Omega}}}_t^{(3)} \qquad \text{(A27)}
\end{aligned}$$

$$\begin{aligned}
a_t^{(5)} &= -\Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(4)} + \left(\gamma_t a_{t-1}^{(5)} a_t^{(1)} - 4\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(4)}\right) \left(a_t^{(1)}\right)^4 \\
&\quad + 6\left(\gamma_t a_{t-1}^{(4)} a_t^{(2)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(3)}\right) \left(a_t^{(1)}\right)^3 \\
&\quad + 4\left(\gamma_t a_{t-1}^{(3)} a_t^{(3)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(3)} a_{t-1}^{(2)} - 2\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(3)} a_t^{(2)}\right) \left(a_t^{(1)}\right)^2 \\
&\quad + \left(\gamma_t \left(a_{t-1}^{(2)} a_t^{(4)} + 3a_{t-1}^{(3)} \left(a_t^{(2)}\right)^2\right) - \Sigma_t \bar{\bar{\boldsymbol{\Omega}}}_t^{(4)} - 6\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(2)} a_t^{(2)} - 4\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(2)} a_t^{(3)}\right) a_t^{(1)} \\
&\quad + 4s_t^{(1)} a_t^{(4)} + 6\left(s_t^{(2)} - \left(s_t^{(1)}\right)^2\right) a_t^{(3)} + 4\left(s_t^{(3)} + \left(s_t^{(1)}\right)^3 - 3s_t^{(1)} s_t^{(2)}\right) a_t^{(2)}, \qquad \text{(A28)}
\end{aligned}$$

$$\begin{aligned}
s_t^{(4)} &= \left(\gamma_t a_{t-1}^{(5)} a_t^{(1)} - 4\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(4)}\right) \left(a_t^{(1)}\right)^3 \\
&\quad + 6\left(\gamma_t a_{t-1}^{(4)} a_t^{(2)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(3)}\right) \left(a_t^{(1)}\right)^2
\end{aligned}$$

$$+4\left(\gamma_t a_{t-1}^{(3)} a_t^{(3)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(3)} a_{t-1}^{(2)} - 2\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(3)} a_t^{(2)}\right) a_t^{(1)}$$
$$+\left(\gamma_t \left(a_{t-1}^{(2)} a_t^{(4)} + 3a_{t-1}^{(3)} \left(a_t^{(2)}\right)^2\right) - \Sigma_t \bar{\bar{\boldsymbol{\Omega}}}_t^{(4)} - 6\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(2)} a_t^{(2)} - 4\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(2)} a_t^{(3)}\right)$$
$$+\left(s_t^{(1)}\right)^4 + 4s_t^{(1)} s_t^{(3)} + 3\left(s_t^{(2)} - 2\left(s_t^{(1)}\right)^2\right) s_t^{(2)}. \tag{A29}$$

C POLYNOMIAL APPROXIMATIONS OF $b_t^{(r)}$ AND $\mu_t^{(r)}$

C.1 First derivative of $\log f(\alpha_t|\alpha_{t+1}, y)$

Here we derive an exact expression for $h_t^{(1)}(\alpha_t;\alpha_{t+1})$, the first derivative of $\log f(\alpha_t|\alpha_{t+1}, y)$ with respect to α_t.

The case $t=1$ is straightforward using Bayes' rule. We have

$$\frac{\partial \log f(\alpha_1|\alpha_2, y)}{\partial \alpha_1} = \frac{\partial \log f(y_1|\alpha_1, \alpha_2)}{\partial \alpha_1} + \frac{\partial \log f(\alpha_2, \alpha_1)}{\partial \alpha_1}$$

Recalling the definition of $\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$ in (7), the first derivative of $h_1(\alpha_1;\alpha_2)$ is

$$h_1^{(1)}(\alpha_1;\alpha_2) = \psi_1^{(1,0)}(\alpha_1, \alpha_2) + \bar{c}_1 - \bar{\Omega}_{1,2}\alpha_2 - \bar{\Omega}_{1,1}\alpha_1. \tag{A30}$$

For $t=2,\ldots,n-1$, we compute $f(\alpha_t|\alpha_{t+1}, y)$ by marginalizing $f(\alpha_{1:t}|\alpha_{t+1}, y)$:

$$\begin{aligned} f(\alpha_t|\alpha_{t+1}, y) &= \int f(\alpha_{1:t-1}, \alpha_t|\alpha_{t+1}, y)\, d\alpha_{1:t-1} \\ &\propto f(\alpha_{t+1}|\alpha_t) f(y_t|\alpha_t, \alpha_{t+1}) c(\alpha_t), \end{aligned} \tag{A31}$$

where

$$c(\alpha_t) = \int f(\alpha_t|\alpha_{t-1}) f(y_{t-1}|\alpha_{t-1}, \alpha_t) f(y_{1:t-2}, \alpha_{1:t-1})\, d\alpha_{1:t-1}.$$

Taking the logarithm of (A31) and differentiating with respect to α_t gives

$$h_t^{(1)}(\alpha_t;\alpha_{t+1}) = \frac{\partial \log c(\alpha_t)}{\partial \alpha_t} + \frac{\partial \log f(\alpha_{t+1}|\alpha_t)}{\partial \alpha_t} + \frac{\partial \log f(y_t|\alpha_t, \alpha_{t+1})}{\partial \alpha_t}. \tag{A32}$$

We use a development similar to Appendix C of McCausland (2012) to show that

$$\frac{\partial \log c(\alpha_t)}{\partial \alpha_t} = E\left[\frac{\partial \log f(\alpha_t|\alpha_{t-1})}{\partial \alpha_t} + \frac{\partial \log f(y_{t-1}|\alpha_{t-1}, \alpha_t)}{\partial \alpha_t}\middle|\alpha_t, y\right].$$

The first derivative $h_t(\alpha_t;\alpha_{t+1})$ then becomes

$$\begin{aligned} h_t^{(1)}(\alpha_t;\alpha_{t+1}) &= E\left[\frac{\log f(\alpha_t|\alpha_{t-1})}{\partial\alpha_t}+\frac{\log f(y_{t-1}|\alpha_{t-1},\alpha_t)}{\partial\alpha_t}\middle|\alpha_t,\alpha_{t+1},y\right] \\ &\quad +\frac{\partial\log f(\alpha_{t+1}|\alpha_t)}{\partial\alpha_t}+\frac{\partial\log f(y_t|\alpha_t,\alpha_{t+1})}{\partial\alpha_t} \\ &= E\left[\frac{\log f(\alpha_t|\alpha_{t-1})}{\partial\alpha_t}+\frac{\log f(\alpha_{t+1}|\alpha_t)}{\partial\alpha_t}\middle|\alpha_t,\alpha_{t+1},y\right] \\ &\quad +E\left[\frac{\log f(y_{t-1}|\alpha_{t-1},\alpha_t)}{\partial\alpha_t}\middle|\alpha_t,\alpha_{t+1},y\right]+\frac{\partial\log f(y_t|\alpha_t,\alpha_{t+1})}{\partial\alpha_t}. \end{aligned}$$

The first term above simplifies as in Appendix C of McCausland (2012). We use (7) to finally derive

$$\begin{aligned} h_t^{(1)}(\alpha_t;\alpha_{t+1}) &= \bar{c}_t-\bar{\Omega}_{t,t}\alpha_t-\bar{\Omega}_{t,t+1}\alpha_{t+1}+\psi_t^{(1,0)}(\alpha_t,\alpha_{t+1}) \\ &\quad -\bar{\Omega}_{t-1,t}\mu_{t-1|t}(\alpha_t)+\delta_{t-1|t}(\alpha_t), \end{aligned} \tag{A33}$$

where $\mu_{t-1|t}(\alpha_t)=E[\alpha_{t-1}|\alpha_t,y]$ and $\delta_{t-1|t}(\alpha_t)=E\left[\psi_{t-1}^{(0,1)}(\alpha_{t-1},\alpha_t)\middle|\alpha_t,y\right]$. The case $t=n$ is similar, and we obtain

$$h_n^{(1)}(\alpha_n)=\bar{c}_n-\bar{\Omega}_{n,n}\alpha_n+\psi_n^{(1)}(\alpha_n)-\bar{\Omega}_{n-1,n}\mu_{n-1|n}(\alpha_n)+\delta_{n-1|n}(\alpha_n). \tag{A34}$$

C.2 Coefficients of the polynomial $\Delta_{t-1|t}(\alpha_t)$

We construct $\Delta_{t-1|t}(\alpha_t)$, $t=2,\dots,n$ in two steps. We first approximate $\psi_{t-1}^{(0,1)}(\alpha_{t-1},\alpha_t)$, as a function of α_{t-1}, by its second order Taylor series expansion around $a_{t-1|t}(\alpha_t)$:

$$\begin{aligned} \psi_{t-1}^{(0,1)}(\alpha_{t-1},\alpha_t) &\approx \psi_{t-1}^{(0,1)}(a_{t-1|t}(\alpha_t),\alpha_t)+\psi_{t-1}^{(1,1)}(a_{t-1|t}(\alpha_t),\alpha_t)(\alpha_{t-1}-a_{t-1|t}(\alpha_t)) \\ &\quad +\tfrac{1}{2}\psi_{t-1}^{(2,1)}(a_{t-1|t}(\alpha_t),\alpha_t)(\alpha_{t-1}-a_{t-1|t}(\alpha_t))^2. \end{aligned} \tag{A35}$$

Taking conditional expectations of both sides of (A35), given α_t and y, and using $\Sigma_{t-1|t}(\alpha_t)$ as an approximation of $E\left[(\alpha_{t-1}-a_{t-1|t}(\alpha_t))^2|\alpha_t,y\right]$ gives the approximation

$$\begin{aligned} \delta_{t-1|t}(\alpha_t) &\approx \psi_{t-1}^{(0,1)}(a_{t-1|t}(\alpha_t),\alpha_t)+\psi_{t-1}^{(1,1)}(a_{t-1|t}(\alpha_t),\alpha_t)(\mu_{t-1|t}(\alpha_t)-a_{t-1|t}(\alpha_t)) \\ &\quad +\frac{1}{2}\psi_{t-1}^{(2,1)}(a_{t-1|t}(\alpha_t),\alpha_t)\Sigma_{t-1|t}(\alpha_t). \end{aligned} \tag{A36}$$

Now we define the polynomial $\Delta_{t-1|t}(\alpha_t)$ as the R'th order Taylor series expansion of the right-hand side of (A36):

$$\Delta_{t-1|t}(\alpha_t) \doteq \sum_{r=0}^{R} \frac{\Delta_{t-1}^{(r)}}{r!}(\alpha_t - a_t)^r, \tag{A37}$$

where $\Delta_{t-1}^{(r)}$ is the r-th derivative of the RHS of (A36) with respect to α_t, evaluated at a_t. We evaluate these derivatives bottom up using Faà Di Bruno's formula, equations (A60) and (A61), and Leibniz's rule, equation (A56).

C.3 Coefficients of the Polynomial $B_{t|t+1}(\alpha_{t+1})$

For $t=1$, $b_{t|t+1}(\alpha_{t+1})$ equals exactly $a_{t|t+1}(\alpha_{t+1})$. Thus, we have $B_t^{(r)} = a_t^{(r)}$, $r=0,\ldots,R$.

For $t=2,\ldots,n-1$, by definition, $b_{t|t+1}(\alpha_{t+1})$ is the root of $h_t^{(1)}(\alpha_t;\alpha_{t+1})=0$. We can approximate this root, as a function of α_{t+1}, using one iteration of the Newton–Raphson algorithm for root finding, from the starting point $a_{t|t+1}(\alpha_{t+1})$:

$$b_{t|t+1}(\alpha_{t+1}) \approx a_{t|t+1}(\alpha_{t+1}) - \frac{h_t^{(1)}(a_{t|t+1}(\alpha_{t+1});\alpha_{t+1})}{h_t^{(2)}(a_{t|t+1}(\alpha_{t+1});\alpha_{t+1})}. \tag{A38}$$

We want to approximate the function $b_{t|t+1}(\alpha_{t+1})$, not just perform the Newton–Raphson step for a particular value of $a_{t|t+1}$. Our strategy will be to find an approximate Taylor expansion of the second term of the right-hand side around $\alpha_{t+1} = a_{t+1}$.

We approximate the numerator and denominator, using $H_t^{(1)}(\alpha_t;\alpha_{t+1})$ and its derivative, both evaluated at $\alpha_t = a_{t|t+1}(\alpha_{t+1})$. These are

$$\begin{aligned} H_t^{(1)}(a_{t|t+1};\alpha_{t+1}) = \bar{c}_t - \bar{\Omega}_{t,t} a_{t|t+1} - \bar{\Omega}_{t,t+1}\alpha_{t+1} + \psi_t^{(1,0)}(a_{t|t+1},\alpha_{t+1}) \\ - \bar{\Omega}_{t-1,t} M_{t-1|t}(a_{t|t+1}) + \Delta_{t-1|t}(a_{t|t+1}) \end{aligned} \tag{A39}$$

$$\begin{aligned} H_t^{(2)}(a_{t|t+1};\alpha_{t+1}) = -\bar{\Omega}_{t,t} + \psi_t^{(2,0)}(a_{t|t+1},\alpha_{t+1}) \\ - \bar{\Omega}_{t-1,t} M_{t-1|t}^{(1)}(a_{t|t+1}) + \Delta_{t-1|t}^{(1)}(a_{t|t+1}), \end{aligned} \tag{A40}$$

where we suppress the argument of $a_{t|t+1}(\alpha_{t+1})$ to write $a_{t|t+1}$.

We compute total derivatives of $H_t^{(1)}(a_{t|t+1}(\alpha_{t+1});\alpha_{t+1})$ and $H_t^{(2)}(a_{t|t+1}(\alpha_{t+1});\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$ using Faà di Bruno's formula to compute the derivatives of $M_{t-1|t}(a_{t|t+1}(\alpha_{t+1}))$, $a_{t-1|t}(a_{t|t+1}(\alpha_{t+1}))$ and $\Delta_{t-1|t}(a_{t|t+1}(\alpha_{t+1}))$ with respect to α_{t+1}, at $\alpha_{t+1} = a_{t+1}$.

Based on equation (A38), we define the following approximations $B_t^{(r)}$ of $b_t^{(r)}$, $r=0,1,2,3$:

$$B_t^{(r)} \doteq a_t^{(r)} - \frac{\partial^r}{\partial \alpha_{t+1}^r} \left(\frac{H_t^{(1)}(a_{t|t+1}(\alpha_{t+1});\alpha_{t+1})}{H_t^{(2)}(a_{t|t+1}(\alpha_{t+1});\alpha_{t+1})} \right)\Bigg|_{\alpha_{t+1}=a_{t+1}}. \tag{A41}$$

The second term on the right-hand side of (A41) is the r-th order derivative of a quotient, which we compute using the quotient rule for derivatives, equation (A57).

In practice, going beyond a third-order approximation of $b_{t|t+1}(\alpha_{t+1})-a_{t|t+1}(\alpha_{t+1})$ does not justify the computational cost and so we set $B_t^{(4)}=a_t^{(4)}$.

For $t=n$, we approximate a value b_n, not a function. We define, analogously, the following approximation of b_n:

$$B_n \doteq a_n - \frac{H_n^{(1)}(a_n)}{H_n^{(2)}(a_n)}. \tag{A42}$$

C.4 Coefficients of the polynomial $M_{t|t+1}(\alpha_{t+1})$

Recall that $\mu_{t|t+1}(\alpha_{t+1})=E[\alpha_t|\alpha_{t+1},y]$. We provide an approximation $M_{t|t+1}(\alpha_{t+1})$ of a Taylor expansion of $\mu_{t|t+1}(\alpha_{t+1})$ around $\alpha_{t+1}=a_{t+1}$. We show in this subsection how to compute the coefficients of the resulting fourth-order polynomial.

McCausland (2012) suggests the following approximation for $\mu_{t|t+1}-b_{t|t+1}$:

$$\mu_{t|t+1}-b_{t|t+1} \approx \frac{1}{2} h_t^{(3)}(b_{t|t+1};\alpha_{t+1}) \left[h_t^{(2)}(b_{t|t+1};\alpha_{t+1}) \right]^{-2} \tag{A43}$$

As the mode $b_{t|t+1}$ is the root of $h_t^{(1)}(\alpha_t;\alpha_{t+1})$, we have

$$h_t^{(1)}(b_{t|t+1};\alpha_{t+1})=0 \tag{A44}$$

Taking the derivative of (A44) two times with respect to α_{t+1} gives

$$h_t^{(2)}(b_{t|t+1};\alpha_{t+1}) b_{t|t+1}^{(1)} = \bar{\Omega}_{t,t+1} - \psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1}) \tag{A45}$$

and

$$\begin{aligned} h_t^{(3)}(b_{t|t+1};\alpha_{t+1}) \left(b_{t|t+1}^{(1)} \right)^2 + h_t^{(2)}(b_{t|t+1};\alpha_{t+1}) b_{t|t+1}^{(2)} = -2 \frac{d\psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1})}{d\alpha_{t+1}} \\ + \psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1}) \end{aligned} \tag{A46}$$

Solve for $h_t^{(3)}(b_{t|t+1};\alpha_{t+1})$ in equation (A46) and divide by the square of $h_t^{(2)}(b_{t|t+1};\alpha_{t+1})$ to obtain

$$\begin{aligned}\frac{h_t^{(3)}(b_{t|t+1};\alpha_{t+1})}{\left(h_t^{(2)}(b_{t|t+1};\alpha_{t+1})\right)^2} = &-\frac{b_{t|t+1}^{(2)}/b_{t|t+1}^{(1)}}{h_t^{(2)}(b_{t|t+1};\alpha_{t+1})b_{t|t+1}^{(1)}} \\ &-\frac{2d\psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1})/d\alpha_{t+1}-\psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1})}{\left(h_t^{(2)}(b_{t|t+1};\alpha_{t+1})b_{t|t+1}^{(1)}\right)^2}\end{aligned} \tag{A47}$$

Substitute the right-hand side of equation (A45) in (A47) to obtain

$$\begin{aligned}\mu_{t|t+1}-b_{t|t+1} \approx &-\frac{1}{2}\frac{b_{t|t+1}^{(2)}/b_{t|t+1}^{(1)}}{\bar{\Omega}_{t,t+1}-\psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1})} \\ &-\frac{1}{2}\frac{2d\psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1})/d\alpha_{t+1}-\psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1})}{\left(\bar{\Omega}_{t,t+1}-\psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1})\right)^2}\end{aligned} \tag{A48}$$

Based on equation (A48), we define our approximation $M_{t|t+1}$ of $\mu_{t|t+1}$ as the Taylor series expansion of:

$$-\frac{1}{2}\frac{B_{t|t+1}^{(2)}/B_{t|t+1}^{(1)}}{\bar{\Omega}_{t,t+1}-\psi_t^{(1,1)}(B_{t|t+1},\alpha_{t+1})}-\frac{1}{2}\frac{2d\psi_t^{(1,1)}(B_{t|t+1},\alpha_{t+1})/d\alpha_{t+1}-\psi_t^{(1,1)}(B_{t|t+1},\alpha_{t+1})}{\left(\bar{\Omega}_{t,t+1}-\psi_t^{(1,1)}(B_{t|t+1},\alpha_{t+1})\right)^2} \tag{A49}$$

The derivatives of $B_{t|t+1}^{(2)}/B_{t|t+1}^{(1)}$ with respect to α_{t+1} are computed using the quotient rule for derivatives, equation (A57). Those of $\psi_t^{(1,1)}(B_{t|t+1},\alpha_{t+1})$ and $d\psi_t^{(1,1)}(B_{t|t+1},\alpha_{t+1})/d\alpha_{t+1}$ are computed using the Faà-Di-Bruno formula, equations (A60) and (A61). Derivatives of the two main ratios in (A49) are computed using the quotient rule in equation (A57). We compute $M_t^{(r)}=M_{t|t+1}^{(r)}(a_{t+1})$, $r=0,1,2$ using (A49).

In practice, going beyond a second-order approximation of $\mu_{t|t+1}(\alpha_{t+1})-b_{t|t+1}(\alpha_{t+1})$ does not repay the computational cost and so we set $M_t^{(3)}=B_t^{(3)}$ and $M_t^{(4)}=a_t^{(4)}$.

D MODEL DERIVATIVES

Here we show how to compute partial derivatives of $\psi_t(\alpha_t,\alpha_{t+1})$ and derivatives $\psi_n(\alpha_n)$, for the ASV-Gaussian and ASV-Student models. In our empirical applications, we compute $\psi_t^{(p,q)}(\alpha_t,\alpha_{t+1})$ up to orders $P=7$ and $Q=7$ and $\psi_n^{(p)}(\alpha_n)$ up to order $P=7$.

D.1 ASV-Gaussian

Using (6), we can write

$$\psi_t(\alpha_t,\alpha_{t+1}) = -\frac{1}{2}\left[\log(2\pi/\beta)+\alpha_t+\beta(\varphi_t-\theta u_t)^2\right], \qquad t=1,\ldots,n-1, \tag{A50}$$

$$\psi_n(\alpha_n) = -\frac{1}{2}\left[\log(2\pi)+\alpha_n+\varphi_n^2\right], \tag{A51}$$

where $\beta \doteq (1-\rho^2)^{-1}$, $\theta \doteq \rho/\sigma$, $u_t \doteq \alpha_{t+1}-d_t-\phi\alpha_t$ and $\varphi_t \doteq y_t\exp(-\alpha_t/2)$.

For $t=1,\ldots,n-1$ and $(p,q)\neq(0,0)$ we have

$$\psi_t^{(p,q)}(\alpha_t,\alpha_{t+1}) = \begin{cases} -\frac{1}{2}-\frac{\beta}{2}\left(\tilde{\varphi}_{t,p}-2\theta^2\phi u_t\right) & q=0, p=1 \\ -\frac{\beta}{2}\left(\tilde{\varphi}_{t,p}+2\theta^2\phi^2\right) & q=0, p=2 \\ -\frac{\beta}{2}\tilde{\varphi}_{t,p} & q=0, p\geq 3 \\ \beta\theta\left(\varphi_t-\theta u_t\right) & q=1, p=0 \\ \beta\theta\left(-\frac{1}{2}\varphi_t+\theta\phi\right) & q=1, p=1 \\ \beta\theta\left(-\frac{1}{2}\right)^p\varphi_t & q=1, p\geq 2 \\ -\beta\theta^2 & q=2, p=0 \\ 0 & \text{otherwise}, \end{cases} \tag{A52}$$

where

$$\tilde{\varphi}_{t,p} \doteq (-1)^p\varphi_t^2-\left(-\frac{1}{2}\right)^{p-2}\theta\varphi_t\left(p\phi+\frac{1}{2}u_t\right), \qquad t=1,\ldots,n-1. \tag{A53}$$

For $t=n$,

$$\psi_n^{(p)}(\alpha_n)(\alpha_n) = \begin{cases} -\frac{1}{2}-\frac{1}{2}\tilde{\varphi}_{n,p} & p=1 \\ -\frac{1}{2}\tilde{\varphi}_{n,p} & p\geq 2, \end{cases} \tag{A54}$$

where $\tilde{\varphi}_{n,p}=(-1)^p\varphi_n^2$.

D.2 ASV-Student

We use the definitions of β,θ,u_t, and φ_t from D.1. Using (11) we can write $\psi_t(\alpha_t,\alpha_{t+1})$, for $t=1,\ldots,n-1$, as

$$\psi_t(\alpha_t,\alpha_{t+1}) = k+\psi_{1,t}(\alpha_t,\alpha_{t+1})+\psi_{2,t}(\alpha_t)+\psi_{3,t}(\alpha_t,\alpha_{t+1}), \tag{A55}$$

where k does not depend on α_t and α_{t+1},

$$\psi_{1,t}(\alpha_t,\alpha_{t+1}) \doteq -\frac{1}{2}(\theta^2\beta u_t^2+\alpha_t), \qquad \psi_{2,t}(\alpha_t) \doteq -(\nu+1)\log d(\alpha_t),$$

$$\psi_{3,t}(\alpha_t,\alpha_{t+1}) \doteq \log m(z(\alpha_t,\alpha_{t+1})), \qquad m(z) = 2\frac{\Gamma\left(\frac{\nu}{2}+1\right)}{\Gamma\left(\frac{\nu+1}{2}\right)} z m_1(z) + m_2(z),$$

$$m_1(z) = M\left(\frac{\nu}{2}+1;\frac{3}{2};z^2\right), \qquad m_2(z) = M\left(\frac{\nu+1}{2};\frac{1}{2};z^2\right),$$

$$z(\alpha_t,\alpha_{t+1}) = \frac{n(\alpha_t,\alpha_{t+1})}{d(\alpha_t)}, \qquad n(\alpha_t,\alpha_{t+1}) = \frac{\theta\beta}{\sqrt{2\nu}} u_t\varphi_t,\ d(\alpha_t) = \sqrt{1+\frac{\beta}{\nu}\varphi_t^2}.$$

Computing analytical expressions for high-order partial derivatives of $\psi_t(\alpha_t,\alpha_{t+1})$ is daunting, but fortunately we can avoid it. All we need to do is evaluate the derivatives at a given point (α_t,α_{t+1}), and for this, we can use general purpose routines to combine derivatives of products, quotients, and composite functions.

We first compute the derivatives of the third component $\psi_{3,t}(\alpha_t,\alpha_{t+1})$ of the log-density of the ASV-Student model. We do it bottom up using the following steps:

1. Evaluate $n(\alpha_t,\alpha_{t+1})$ and its derivatives with respect to α_t and α_{t+1} up to orders P and Q:

$$n^{(p,q)}(\alpha_t,\alpha_{t+1}) = \begin{cases} \frac{\beta\theta}{\sqrt{2\nu}}\left(-\frac{1}{2}\right)^p (2p\phi+u_t)\varphi_t & p\geq 0, q=0 \\ \frac{\beta\theta}{\sqrt{2\nu}}\left(-\frac{1}{2}\right)^p \varphi_t & p\geq 0, q=1 \\ 0 & p\geq 0, q\geq 2. \end{cases}$$

2. Evaluate derivatives of $(1+\beta/\nu\varphi_t^2(\alpha_t))$ with respect to α_t up to order P:

$$\frac{d^p}{d\alpha_t}\left(1+\frac{\beta}{\nu}\varphi_t^2(\alpha_t)\right) = (-1)^p\frac{\beta}{\nu}\varphi_t^2(\alpha_t), \quad p=0,\ldots,P.$$

3. Evaluate $d(\alpha_t)$ and its derivatives with respect to α_t, up to order P. Use derivatives of the square root function, evaluated at $(1+\beta/\nu\varphi_t^2(\alpha_t))$ and the derivatives evaluated in step 2, combining them using Faà Di Bruno's formula, equations (A60) and (A61).
4. Evaluate $z=n/d$ and partial derivatives $z^{(p,q)}(\alpha_t,\alpha_{t+1})$ up to order P and Q. Use the value n and partial derivatives $n^{(p,q)}(\alpha_t,\alpha_{t+1})$ computed at step (1), as well as the value d and derivatives $d^{(p)}(\alpha_t)$ computed at step (3). For each $p=1,\ldots,P$, compute $z^{(p,q)}(\alpha_t,\alpha_{t+1})$ using the quotient rule, equation (A57).
5. Evaluate $M(\nu/2+1,3/2,x)$ and partial derivatives $M^{(0,0,p)}(\nu/2,3/2,x)$ up to order P. We use the property $M^{(0,0,p)}(a,b,x)=(a)_k/(b)_k M(a+k,b+k,x)$ and compute values of $M(a,b,x)$ using the routine `gsl_sf_hyperg_1F1` in the GNU scientific

library. Similarly, compute $M((\nu+1)/2,1/2,x)$ and partial derivatives $M^{(0,0,p)}((\nu+1)/2;1/2;x)$ up to order P.

6. Set $m_1(z)=M(\nu/2+1,3/2,z^2)$ and compute P derivatives of $m_1(z)$ with respect to z. Use P derivatives of $M(\nu/2+1,3/2,x)$ with respect to x, computed in step 5 and P derivatives (only 2 are nonzero) of $x=z^2$ with respect to z, evaluated at z, combining them using the Faà Di Bruno's rule, equations (A60) and (A61). Similarly, set $m_2(z)=M((\nu+1)/2,1/2,z^2)$ and evaluate P derivatives of $m_2(z)$ with respect to z.
7. Evaluate P derivatives of $m(z)$ with respect to z using the derivatives evaluated at step 6, combining them according to

$$m^{(p)}(z)=2\frac{\Gamma\left(\frac{\nu}{2}+1\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\left(zm_1^{(p)}(z)+pm_1^{(p-1)}(z)\right)+m_2^{(p)}(z),\quad p=1,\dots,P.$$

8. Evaluate P derivatives of $\log m(z)$ with respect to z using the derivatives evaluated at step 7, and the logarithm rule, equations (A58) and (A59).
9. Evaluate partial derivatives of $\psi_{3,t}(\alpha_t,\alpha_{t+1})$ up to orders P and Q. Use derivatives of $\log m(z)$ with respect to z computed in step 8 and partial derivatives of $z(\alpha_t,\alpha_{t+1})$ computed in step 4, combining them according to the multivariate Faa-Di-Bruno rule defined in equations (A64) and (A65).

The first component, $\psi_{1,t}(\alpha_t,\alpha_{t+1})$, is a quadratic function of α_t and α_{t+1}. Its derivatives, for $(p,q)\neq(0,0)$, are

$$\psi_{1,t}^{(p,q)}(\alpha_t,\alpha_{t+1})=\begin{cases}-\frac{1}{2}\theta^2\beta u_t & p=0,q=1,\\ -\frac{1}{2}\theta^2\beta & p=0,q=2,\\ -\frac{1}{2}(-\phi\theta^2\beta u_t+1) & p=1,q=0,\\ \frac{1}{2}\phi\theta^2\beta & p=1,q=1,\\ -\frac{1}{2}\phi^2\theta^2\beta & p=2,q=1,\\ 0 & \text{otherwise.}\end{cases}$$

Recall that $\psi_{2,t}(\alpha_t)=-(\nu+1)\log d(\alpha_t)$. We compute derivatives of $\log d(\alpha_t)$ using the log rule in equations (A58) and (A59). Derivatives of $\psi_{2,t}(\alpha_t)$ are simply $-(\nu+1)$ times the derivatives of $\log d(\alpha_t)$.

The special case of $t=n$ is easily handled. We have

$$\psi_n(\alpha_n)=\log\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}}-\frac{1}{2}\left[\alpha_n+(\nu+1)\log\left(1+\frac{\varphi_n^2}{\nu}\right)\right],$$

whose derivatives are the same as those of $\psi_{2,t}$ except for β replaced by 1.

E RULES FOR DERIVATIVES OF COMPOUND FUNCTIONS

In this article, we often use automatic rules for evaluating multiple derivatives of compound functions at a point. The rules combine multiple derivatives of component functions, also evaluated at points. This Appendix gathers these rules together.

For univariate functions f and g, we list rules for derivatives of the product fg, the quotient f/g, the composition $f\circ g$ and $\log g$. We also give derivatives of $f\circ g$ for $f:\mathbb{R}\to\mathbb{R}$ and $g:\mathbb{R}^2\to\mathbb{R}$ and partial derivatives of $f\circ g$ for $f:\mathbb{R}^2\to\mathbb{R}$ and $g:\mathbb{R}\to\mathbb{R}^2$.

We have coded all of these rules as routines. Values passed to the routines are vectors (or matrices) giving multiple derivatives (or partial derivatives) of f and g, evaluated at particular points. The routines return a vector (or a matrix) giving multiple derivatives (or partial derivatives) of a compound function, evaluated at a point. For example, the routine computing P derivatives of the product fg at a point x takes as input the integer P, a P-vector with the first P derivatives of f at x and a P-vector with the first P derivatives of g at x. It returns a P-vector with the first P derivatives of fg at x.

E.1 Univariate functions

Let x be a point in $\mathbb{R}$ and f and g be two univariate functions, continuously differentiable at x up to order P.

Leibniz rule for products The product fg is differentiable up to order P at x and

$$(fg)^{(p)}(x)=\sum_{r=0}^{p}\binom{p}{r}f^{(r)}(x)g^{(p-r)}(x),\quad p=1,\dots,P. \tag{A56}$$

Quotient rule Applying Leibniz' rule to the product of f/g and g gives the recursive rule

$$(f/g)^{(p)}(x)=\frac{1}{g(x)}\left[f^{(p)}(x)-\sum_{r=0}^{p-1}\binom{p}{r}(f/g)^{(r)}(x)g^{(p-r)}(x)\right],\quad p=1,\dots,P. \tag{A57}$$

Log rule Let $h=\log f$ and suppose that $f>0$. Then h is differentiable up to order P. Applying the quotient rule to

$$h^{(1)}(x)=\frac{f^{(1)}(x)}{f(x)} \tag{A58}$$

gives

$$h^{(p)}(x) = \frac{1}{f(x)}\left[f^{(p)}(x) - \sum_{r=1}^{p-1}\binom{p-1}{r-1}h^{(r)}(x)f^{(p-r)}(x)\right], \quad p=2,\dots,P. \tag{A59}$$

Together, equations (A58) and (A59) give the first P derivatives of $\log(f(x))$.

Faà di Bruno's rule for composite functions Now let x be a point in $\mathbb{R}$, g be a univariate function, P times differentiable at x, and f be a univariate function, P times differentiable at $g(x)$. Faà di Bruno's rule gives the p-th derivative of $f \circ g$ at x as

$$(f\circ g)^{(p)}(x) = \sum_{r=1}^{p} f^{(r)}(g(x))B_{p,r}(g^{(1)}(x),\dots,g^{(p-r+1)}(x)), \tag{A60}$$

where the $B_{p,r}(z_1,\dots,z_{p-r+1})$ are Bell polynomials. The Bell polynomials are a triangular array of polynomials that can be computed using the boundary conditions $B_{0,0}(z_1)=1$ and $B_{p,0}(z_1,\dots,z_{p+1})=0$, $p>0$, and the recursion

$$B_{p,r}(z_1,\dots,z_{p-r+1}) = \sum_{i=r-1}^{p-1}\binom{p-1}{i} z_{p-i}B_{i,r-1}(z_1,\dots,z_{i-r}), \quad r=1,\dots,p. \tag{A61}$$

For example, we have $B_{1,1}(z_1)=z_1B_{0,0}(z_1)=z_1$, which gives $(f\circ g)^{(1)}(x) = f^{(1)}(g(x))g^{(1)}(x)$, the chain rule. For the second derivative, we compute $B_{2,1}(z_1,z_2) = z_2B_{0,0}(z_1)+z_1B_{1,0}(z_1,z_2)=z_2$ and $B_{2,2}(z_1)=z_1B_{1,1}(z_1)=z_1^2$, which gives

$$(f\circ g)^{(2)}(x) = f^{(1)}(g(x))g^{(2)}(x) + f^{(2)}(g(x))\left(g^{(1)}(x)\right)^2.$$

E.2 Multivariate functions

Savits (2006) generalizes Faà di Bruno's rule to multivariate functions. Equations (3.1) and (3.5) in that paper give multiple partial derivatives of $f\circ g$, where $f: \mathbb{R}^m \to \mathbb{R}$ and $g: \mathbb{R}^d \to \mathbb{R}^m$. We show how to compute partial derivatives for two special cases.

Case $d=1$ and $m=2$ Here $(f\circ g)(x) = f(g_1(x), g_2(x))$, where f is a scalar valued function with continuous partial derivatives up to orders P and P, and g_1 and g_2 are scalar-valued functions, continuously differentiable up to order P. The value of the p-th derivative of $f\circ g$ at is

$$(f\circ g)^{(p)}(x) = \sum_{r=0}^{p}\sum_{s=\max\{0,1-r\}}^{p-r} f^{(r,s)}(g_1(x), g_2(x))v_{p,(r,s)}, \tag{A62}$$

where the values $v_{p,(r,s)}$ are defined by the boundary conditions $v_{0,(0,0)}=1$ and $v_{p,(0,0)}=0$ for $p>0$, and the recursion

$$v_{p,(r,s)} = \sum_{i=r+s-1}^{p-1} \binom{p-1}{i} \left[g_1^{(p-i)}(x) v_{i,(r-1,s)} + g_2^{(p-i)}(x) v_{i,(r,s-1)} \right]. \tag{A63}$$

We have a routine taking as input the first P derivatives of g_1 at x, the first P derivatives of g_2 at x, and the partial derivatives $f^{(p,q)}$ at $(g_1(x), g_2(x))$ up to orders P and P, returning the first P derivatives of $f(g_1(x), g_2(x))$ at x.

Case $d=2$, $m=1$ Here $(f \circ g)(x) = f(g(x_1, x_2))$, where x_1 and x_2 are scalars, f is continuously differentiable up to order $P+Q$, and g is a scalar-valued function with continuous partial derivatives up to orders P and Q. The values of the derivatives of $f \circ g$ at (x_1, x_2) are computed using

$$(f \circ g)^{(p,q)}(x_1, x_2) = \sum_{r=1}^{p+q} f^{(r)}(g(x_1, x_2)) v_{(p,q),r}, \tag{A64}$$

where the values $v_{(p,q),r}$ are defined by the conditions $v_{(0,0),0}=1$ and $v_{(p,q),0}(x_1,x_2)=0$ for $(p,q) \neq (0,0)$, $v_{(p,q),r}=0$ for $r<0$ or $p+q<r$ and the recursion

$$v_{(p,q),r} = \begin{cases} \sum_{i=r-1}^{p-1} \binom{p-1}{i} g^{(p-i,0)}(x_1,x_2) v_{(i,0),r-1} & q=0, p\geq 1 \\ \sum_{i=0}^{p} \sum_{j=0}^{q-1} \binom{p}{i} \binom{q-1}{j} g^{(p-i,q-j)}(x_1,x_2) v_{(i,j),r-1} & q \geq 1, p \geq 0. \end{cases} \tag{A65}$$

We have a routine taking as input the partial derivatives $g^{(p,q)}$ at (x_1, x_2), up to orders P and Q and the first $P+Q$ derivatives of f at $g(x_1, x_2)$, returning the partial derivatives $(f \circ g)^{(p,q)}$ at (x_1, x_2), up to orders P and Q.

Received May 20, 2012; revised December 4, 2013; accepted December 12, 2013.

REFERENCES

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

Chan, J. C. C., and I. Jeliazkov. 2009. "Efficient Simulation and Integrated Likelihood Estimation in State Space Models". Working paper.

Chib, S., F. Nardari, and N. Shephard. 2002. Markov Chain Monte Carlo Methods for Stochastic Volatility Models. *Journal of Econometrics* 108: 281–316.

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

Durbin, J., and S. J. Koopman. 2002. A Simple and Efficient Simulation Smoother for State Space Time Series Analysis. *Biometrika* 89: 603–615.

Feng, D., J. J. Jiang, and P. Song. 2004. Stochastic Conditional Durations Models with "Leverage Effect" for Financial Transaction Data. *Journal of Financial Econometrics* 2: 390–421.

Frühwirth-Schnatter, S., and H. Wagner. 2006. Auxiliary Mixture Sampling for Parameter-driven Models of Time Series of Counts with Applications to State Space Modelling. *Biometrika* 93: 827–841.

Frühwirth-Schnatter, S., R. Früwirth, L. Held, and H. Rue. 2009. Improved Auxiliary Mixture Sampling for Hierarchical Models of non-Gaussian Data. *Statistics and Computing* 19: 479–492.

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

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

Geweke, J. 2004. Getting it Right: Joint Distribution Tests of Posterior Simulators. *Journal of the American Statistical Association* 99: 799–804.

Harvey, A. C. and N. Shephard. 1996. The Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns. *Journal of Business and Economic Statistics* 14: 429–434.

Jacquier, E., N. Polson, and P. Rossi. 2004. Bayesian Analysis of Stochastic Volatility Models with Leverage Effect and Fat Tails. *Journal of Econometrics*, 122: 185–212.

Jungbacker, B., and S. J. Koopman. 2008. Monte Carlo Estimation for Nonlinear non-Gaussian State Space Models. *Biometrika* 94: 827–839.

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

McCausland, W. J., S. Miller, and D. Pelletier. 2011. Simulation Smoothing for State-Space Models: A Computational Efficiency Analysis. *Computational Statistics and Data Analysis* 55: 199–212.

McCausland, W. J. 2012. The HESSIAN Method: Highly Efficient Simulation Smoothing, In A Nutshell. *Journal of Econometrics* 168: 189–206.

Nakajima, J., and Y. Omori. 2009. Leverage, Heavy-tails and Correlated Jumps in Stochastic Volatility Models. *Computational Statistics and Data Analysis* 53: 2335–2353.

Omori, Y., and T. Watanabe. 2008. Block Sampler and Posterior Mode Estimation for Asymmetric Stochastic Volatility Models. *Computational Statistics and Data Analysis* 52: 2892–2910.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima. 2007. Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference. *Journal of Econometrics* 140: 425–449.

Richard, J. and W. Zhang. 2007. Efficient High-Dimensional Importance Sampling. *Journal of Econometrics* 141: 1385–1411.

Rue, H., S. Martino, and N. Chopin. 2009. Approximate Bayesian Inference for Latent Gaussian Models using Integrated nested Laplace Approximations. *Journal of the Royal Statistical Society Series B* 71: 319–392.

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

Savits, T. H. 2006. Some Statistical Applications of Faa Di Bruno. *Journal of Multivariate Analysis* 97: 2131–2140.

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

Strickland, C. M., C. S. Forbes, and G. M. Martin. 2006. Bayesian Analysis of the Stochastic Conditional Duration Model. *Computational Statistics and Data Analysis* 50: 2247–2267.

Stroud, J. R., P. Müller, and N. G. Polson. 2003. Nonlinear State-space Models with State-dependent Variances. *Journal of the American Statistical Association* 98: 377–386.

Watanabe, T., and Y. Omori. 2004. A Multi-move sampler for Estimating non-Gaussian Time Series Models: Comments on Shephard and Pitt (1997). *Biometrika* 91: 246–248.

Yu, J. 2005. On Leverage in a Stochastic Volatility Model. *Journal of Econometrics* 127: 165–178.