THE HESSIAN METHOD FOR MODELS WITH LEVERAGE-LIKE EFFECTS

BARNABÉ DJEGNÉNÉ AND WILLIAM J. MCCAUSLAND

Abstract. We propose a new method for simulation smoothing in state space models with univariate states and leverage-like effects. Given a vector θ of parameters, the state sequence $\alpha = (\alpha_1, \ldots, \alpha_n)^\top$ is Gaussian and the sequence $y = (y_1^\top, \ldots, y_n^\top)^\top$ of observed vectors may be conditionally non-Gaussian. By leverage-like effect, we mean conditional dependence between the observation y_t and the contemporaneous *innovation* of the state equation, not just the contemporaneous state α_t. We use this term since stochastic volatility models with the leverage effect are a leading example.

Our method is an extension of the HESSIAN method described in McCausland [2012], which only works for models without leverage-like effects, models in which the density $f(y_t|\theta, \alpha)$ depends only on θ and α_t. Like that method, ours is based on a close approximation $g(\alpha|\theta, y)$ of the conditional density $f(\alpha|\theta, y)$. One can use $g(\alpha|\theta, y)$ for importance sampling or Markov chain Monte Carlo (MCMC). With a suitable approximation $g(\theta|y)$ of $f(\theta|y)$, we can use $g(\theta, \alpha|y) = g(\theta|y)g(\alpha|\theta, y)$ as an importance or proposal density for the joint posterior distribution of parameters and states. Applications include the approximation of likelihood function values and the marginal likelihood, and Bayesian posterior simulation. We construct the approximation $g(\alpha|\theta, y)$ for Gaussian and Student's t stochastic volatility models with leverage. For both models, we make a joint proposal of the state and parameter vectors. Unlike Omori et al. [2007] and Nakajima and Omori [2009], we do not augment the data by adding mixture indicators or heavy tail scaling factors. Our generic procedure is more numerically efficient than the model specific procedures of those papers — using randomised pseudo-Monte Carlo importance sampling, we obtain relative numerical efficiencies close to 100%, at least 4 times higher than those obtained using the method of Omori et al. [2007]. The Highest value of numerical efficiency reported by Nakajima and Omori [2009] is 29.1% for the ASV-Student model. The lowest efficiency factor reported using the Hessian with randomised pseudo-Monte Carlo importance sampling is 82.61%. Comparing these two figures suggest that the Hessian procedure is numerically efficient than the one described in Nakajima and Omori [2009].

Date: Current version: March 18, 2012.

Mailing address: Département de sciences économiques, CIREQ, C.P. 6128, succursale Centre-ville, Montréal QC H3C 3J7, Canada. e-mail: barnabe.djegnene@umontreal.ca.

1. Introduction

State space models govern the interaction of observable data $y = (y_1^\top, \ldots, y_t^\top, \ldots, y_n^\top)^\top$ and latent states $\alpha = (\alpha_1, \ldots, \alpha_t, \ldots, \alpha_n)^\top$, 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 Bayesian posterior simulation. They involve simulating the conditional distribution of states given data and parameters. We will call this distribution the *target* distribution. Simulation typically entails importance sampling or Markov chain Monte Carlo (MCMC). We show examples of both in Section 4.

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 conditional independence a leverage effect.

Others have extended this model. Jacquier et al. [2004] and Omori et al. [2007] consider inference in stochastic volatility models with leverage and heavy-tailed conditional return distributions. This and other empirical work has shown convincingly that stochastic volatility models with leverage are more realistic descriptions of stock returns 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 et al. [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 et al. [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 does simulation smoothing for 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 a similar standard of approximation that McCausland [2012] does, 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}. Using these derivatives to construct an approximation of the target density requires more effort, largely because when there are leverage-like effects, 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:

(1) 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 Gaussian model, and then any non-Gaussian distributions in the transformed model are approximated by finite Gaussian mixtures. Kim et al. [1998], Chib et al. [2002], Omori et al. [2007] use this auxiliary mixture model approach for stochastic volatility models; Stroud et al. [2003], Frühwirth-Schnatter and Wagner [2006] and Frühwirth-Schnatter et al. [2009] use it for other non-linear 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.

(2) It involves drawing the entire state sequence as a single MCMC block. This leads to efficiency improvements when there is posterior serial dependence. While drawing the entire state sequence using a multivariate Gaussian proposal distribution is impractical, we make it possible by constructing a much closer approximation of the target distribution. Many articles have used multivariate Gaussian proposal distributions to update the state vector, but usually only for about 10–50 observations at a time, not the entire sample. These include Shephard and Pitt [1997], Watanabe and Omori [2004], Strickland et al. [2006], Jungbacker and Koopman [2008] and Omori and Watanabe [2008]. The Efficient Importance Sampling (EIS) method of Richard and Zhang [2007] features draws of the entire state sequence as a block, but since

their approximate 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 do. See the discussion in McCausland [2012] for more details.

(3) Since the approximation is so close, we can draw parameters and states together as a single block. We do this using a joint 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. In this way, we achieve numerical efficiencies comparable to model-specific auxiliary mixture model approaches, which also often feature joint draws of parameter and states. The examples of Section 4 suggest that our method is even more efficient than these approaches, 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 — it allows for importance sampling, variance reduction using randomised pseudo Monte Carlo, and very efficient approximations of the marginal likelihood, as we see in Section 4.

(4) 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 the routines for computing model-specific derivatives 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 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.

(5) 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 et al. [2011], for linear Gaussian state space models, and Rue et al. [2009] for non-linear 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 non-linear or non-Gaussian models tend to use the Kalman filter, but this is not an essential feature of auxiliary mixture model methods.

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

$$\begin{aligned} \alpha_1 &= d_0 + u_0, \qquad \alpha_{t+1} = d_t + \phi_t \alpha_t + u_t, \\ f(y|\alpha) &= \left[\prod_{t=1}^{n-1} f(y_t|\alpha_t, \alpha_{t+1})\right] f(y_n|\alpha_n), \end{aligned} \tag{4}$$

where $\alpha \equiv (\alpha_1, \ldots, \alpha_n)$ is a vector of univariate latent states α_t, the u_t are independent Gaussian random variables with mean 0 and precision (inverse of variance) ω_t, the y_t are observable random vectors, and the $f(y_t|\alpha_t, \alpha_{t+1})$ are measurement density or mass functions. We do not require them to be Gaussian, linear or univariate. 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 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 the contemporaneous state α_t.

Throughout most of the paper, we condition on d_t, ϕ_t, ω_t and any other parameters on which the $f(y_t|\alpha_t, \alpha_{t+1})$ might depend, and suppress notation for this conditioning. In Section 4, where we consider joint inference for parameters and states, we are explicit about this conditioning.

It is easy to see that the model in equations (1), (2) and (3) is of the form given by (4). We use (1) to write

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

then use the standard formula for conditional Gaussian distributions to obtain

$$y_t|\alpha \sim \mathrm{N}\left((\rho/\sigma)\exp(\alpha_t/2)(\alpha_{t+1} - (1-\phi)\bar{\alpha} - \phi\alpha_t), (1-\rho^2)\exp(\alpha_t)\right). \tag{5}$$

In Section 2 we describe our approximation $g(\alpha|y)$ of the target density $f(\alpha|y)$. We show how to evaluate it and how to draw from the distribution with density $g(\alpha|y)$. In Section 3 we apply tests of program correctness to the code we use to compute $g(\alpha|y)$ and draw from the approximate distribution. These tests are similar to those described in Geweke [2004]. Section 4 illustrates our methods using stochastic volatility models with leverage, with Gaussian and Student's t measurement innovations. Section 5 concludes.

2. An approximation of the target density

In this section we define our approximation $g(\alpha|y)$ of the target density $f(\alpha|y)$. We do not provide a closed form expression for $g(\alpha|y)$, but instead show how to evaluate and sample from $g(\alpha|y)$ using $O(n)$ operations. The density $g(\alpha|y)$ is proper and fully normalized.

Our approximation is not model specific. We construct $g(\alpha|y)$ for a particular state space model using a suitable description of the model, consisting of the following quantities and computational routines.

We specify the state dynamics by providing $\bar{\Omega}$ and $\bar{c}$, the precision and covector of the marginal distribution of α, the state sequence. This gives the distribution of α as $\alpha \sim N(\bar{\Omega}^{-1}\bar{c}, \bar{\Omega}^{-1})$. The precision, unlike the variance, is a tri-diagonal matrix, with $O(n)$ elements. Appendix A describes how to compute $\bar{\Omega}$ and $\bar{c}$ in terms of the d_t, ϕ_t and ω_t.

We specify the measurement distributions by supplying routines to compute, for $t = 1, \ldots, n-1$, the functions

$$\psi_t(\alpha_t, \alpha_{t+1}) \doteq \log f(y_t|\alpha_t, \alpha_{t+1}), \quad \psi_n(\alpha_n) = \log f(y_n|\alpha_n), \tag{6}$$

and the partial derivatives

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

for orders p and q up to certain values P and Q. For convenience, Table 1 summarizes this and other important notation.

The routines to compute the $\psi_t(\alpha_t, \alpha_{t+1})$ and $\psi_n(\alpha_n)$ must give exact results, as they are used to evaluate $f(\alpha|y)$ up to a normalization factor. The partial derivatives, however, may be numerical derivatives or other approximations. Approximation error may make $g(\alpha|y)$ a cruder approximation of $f(\alpha|y)$ and thus diminish the numerical precision of IS or MCMC. But we will still be able to evaluate and simulate $g(\alpha|y)$ without error, and so it does not compromise simulation consistency.

Like the target density, the approximation $g(\alpha|y)$ has the Markov property, allowing us to decompose it as

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

Each factor is a proper fully normalized density function closely approximating the corresponding factor of $f(\alpha|y)$. Whether we need to evaluate $g(\alpha|y)$, simulate it or both, the decomposition allows us to do so sequentially, for t descending from n to 1.

Approximations rely on Taylor series expansions, some exact and some approximate, of various functions, including $b_{t|t+1}(\alpha_{t+1})$ and $\mu_{t|t+1}(\alpha_{t+1})$, the mode and mean of the conditional distribution of α_t given α_{t+1} and y. Some expansions are computed during a forward pass, around the mode $(a_1, \ldots, a_n)$ of the target distribution, a static point of expansion. So for example, we compute $B_{t|t+1}(\alpha_{t+1})$ and $M_{t|t+1}(\alpha_{t+1})$ as approximate Taylor series expansions of $b_{t|t+1}(\alpha_{t+1})$ and $\mu_{t|t+1}(\alpha_{t+1})$ around a_{t+1}.

During the backward pass, we compute approximate Taylor series expansions of $h_n(\alpha_n) \doteq \log f(\alpha_n|y)$ and $h_t(\alpha_t; \alpha_{t+1}) \doteq \log f(\alpha_t|\alpha_{t+1}, y)$, $t = n-1, \ldots, 1$, which we will treat as univariate functions of α_t with parameter α_{t+1}. Here, the point of expansion is a moving target, depending on α_{t+1}. The expansion is fifth order, allowing a much better than Gaussian (second order) approximation.

The densities $g(\alpha_t|\alpha_{t+1}, y)$ are members of the class of perturbed Gaussian distributions described in Appendix G of McCausland [2012]. Parameters of the perturbed Gaussian distribution give a mode of the distribution and the second through fifth derivatives of $\log g(\alpha_t|\alpha_{t+1}, y)$ at that mode. Choosing parameters amounts to approximating $b_{t|t+1}(\alpha_{t+1})$, the mode of $f(\alpha_t|\alpha_{t+1}, y)$, and the second through fifth derivatives of $\log f(\alpha_t|\alpha_{t+1}, y)$ there.

In Appendix C.1, we derive this exact result for the first derivative of h_t:

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

where $\mu_{t-1|t}(\alpha_t) \doteq E[\alpha_{t-1}|\alpha_t, y]$ and $x_{t-1|t}(\alpha_t) \doteq E[\psi_{t-1}^{(0,1)}(\alpha_{t-1}, \alpha_t)|\alpha_t, y]$. We also give analogous results for the cases $t = 1$ and $t = n$.

We cannot evaluate $\mu_{t-1|t}(\alpha_t)$, $x_{t-1|t}(\alpha_t)$ or their derivatives exactly. Nor can we evaluate the mode $b_{t|t+1}(\alpha_{t+1})$ exactly. Instead, we provide polynomial approximations $M_{t-1|t}(\alpha_t)$, $B_{t|t+1}(\alpha_{t+1})$ and $X_{t-1|t}(\alpha_t)$ of $\mu_{t-1|t}(\alpha_t)$, $b_{t|t+1}(\alpha_{t+1})$ and $x_{t-1|t}(\alpha_t)$. We use these to approximate the value $b_{t|t+1}(\alpha_{t+1})$ and the derivatives $h_t^{(r)}(\alpha_t; \alpha_{t+1})$, $r = 1, \ldots, 5$. $M_{t-1|t}(\alpha_t)$ and $X_{t-1|t}(\alpha_t)$ are approximate Taylor expansions of $\mu_{t-1|t}(\alpha_t)$ and $x_{t-1|t}(\alpha_t)$ around a_t. $B_{t|t+1}(\alpha_{t+1})$ is an approximate Taylor expansion of $b_{t|t+1}(\alpha_{t+1})$ around a_{t+1}.

We draw α, evaluate $g(\alpha|y)$, or both using the following steps. We first compute the mode $a = (a_1, \ldots, a_n)$ of the target distribution using the method described in Appendix B of McCausland [2012]. In a forward pass we compute the coefficients of the polynomials $B_{t|t+1}(\alpha_{t+1})$, $M_{t-1|t}(\alpha_t)$, and $X_{t-1|t}(\alpha_t)$, for $t = 1, \ldots, n-1$. Finally, we compute, for $t = n, \ldots, 1$, $B_{t|t+1}(\alpha_{t+1})$ and $H_t^{(r)}(B_{t|t+1}(\alpha_{t+1}); \alpha_{t+1})$, using these values as the parameters of the perturbed Gaussian distribution. With these values sets, we can draw α_t, evaluate $g(\alpha_t|\alpha_{t+1}, y)$ or both. In the rest of this section, we describe these steps in more detail. Full detail is left to various appendices.

2.1. **Precomputation.** We first compute the precision $\bar{\Omega}$ and covector $\bar{c}$ of the Gaussian prior distribution of states as a function of d_t, ϕ_t and ω_t in (4). We then compute the mode a of the target distribution. This gives, as bi-products, several quantities used later. This includes the precision $\bar{\bar{\Omega}}$ and covector $\bar{\bar{c}}$ of a Gaussian approximation $\mathrm{N}(\bar{\bar{\Omega}}^{-1}\bar{\bar{c}}, \bar{\bar{\Omega}}^{-1})$ of the target density. It also gives the conditional variances $\Sigma_t \doteq \mathrm{Var}[\alpha_t|\alpha_{t+1}]$, $t = 1, \ldots, n-1$, and $\Sigma_n \doteq \mathrm{Var}[\alpha_n]$ implied by this Gaussian approximation.

This precomputation is similar to that described in Appendix B of McCausland [2012]. Little modification is required, and we give details in Appendix A.

2.2. **A Forward Pass.** In order to describe the forward pass, it will be helpful to introduce a sequence of multivariate Gaussian conditional distributions. We define, for $t = 1, \ldots, n-1$, $(a_{1|t+1}(\alpha_{t+1}), \ldots, a_{t|t+1}(\alpha_{t+1}))$ as the conditional mode of $(\alpha_1, \ldots, \alpha_t)$ given α_{t+1} and y, and $\bar{\bar{\Omega}}_{1:t|t+1}$ as the negative Hessian matrix of $\log f(\alpha_1, \ldots, \alpha_t|\alpha_{t+1}, y)$ with respect to $(\alpha_1, \ldots, \alpha_t)$, evaluated at $(a_{1|t+1}, \ldots, a_{t|t+1})$. Thus we can view the distribution $N((a_{1|t+1}, \ldots, a_{t|t+1}), \bar{\bar{\Omega}}^{-1}_{1:t|t+1})$ as an approximation of the conditional distribution of $(\alpha_1, \ldots, \alpha_t)$ given α_{t+1} and y. Result 2.1 of McCausland et al. [2011] implies that if $\tilde{\alpha} \sim N((a_{1|t+1}, \ldots, a_{t|t+1}), \bar{\bar{\Omega}}^{-1}_{1:t|t+1})$, then $\tilde{\alpha}_t|\tilde{\alpha}_{t+1} \sim N(a_{t|t+1}, \Sigma_{t|t+1})$, where $\Sigma_{t|t+1}$ is the final value in the following forward recursion:

$$\Sigma_{1|t+1} \doteq \bar{\bar{\Omega}}_{11}^{-1}, \qquad \Sigma_{\tau|t+1} \doteq (\bar{\bar{\Omega}}_{\tau\tau} - \bar{\bar{\Omega}}^2_{\tau,\tau-1}\Sigma_{\tau-1|t+1})^{-1}, \quad \tau = 2, \ldots, t. \tag{10}$$

We also define, for $t = 1, \ldots, n-1$, $s_{t|t+1}(\alpha_{t+1}) \doteq \log \Sigma_{t|t+1}(\alpha_{t+1})$.

The forward pass consists of performing the following steps, for $t = 1, \ldots, n-1$:

(1) Compute

$$\begin{aligned} a_t^{(r)} &\doteq \left.\frac{\partial^r a_{t|t+1}(\alpha_{t+1})}{\partial \alpha_{t+1}^r}\right|_{\alpha_{t+1}=a_{t+1}}, \quad r = 1, \ldots, R, \\ s_t^{(r)} &\doteq \left.\frac{\partial^r s_{t|t+1}(\alpha_{t+1})}{\partial \alpha_{t+1}^r}\right|_{\alpha_{t+1}=a_{t+1}}, \quad r = 1, \ldots, R-1. \end{aligned} \tag{11}$$

The choice of R determines how closely we can approximate the functions $a_{t|t+1}(\alpha_{t+1})$ and $s_{t|t+1}(\alpha_{t+1})$ using Taylor expansions. For our empirical illustration, we use $R = 5$.

Appendix B gives details. Equation (28) gives $a_1^{(r)}$ and for $t > 1$, (38) gives $a_t^{(r)}$ as a function of $a_{t-1}^{(i)}$, $i = 1, \ldots, r$, and $a_t^{(i)}$, $i = 1, \ldots, r-1$. Equations (35), (40), (43), (46) and (48) give simplified expressions for $r = 1, \ldots, 5$ and $t > 1$.

Equation (30) gives $s_1^{(r)}$ and equations (38) and (39) give $s_t^{(r)}$. Equations (42), (45), (47) and (49) give simplified expressions for $s_t^{(r)}$, $r = 1, \ldots, 4$ and $t = 2, \ldots, n-1$.

Appendix B includes a proof that these computations are exact. The proof uses a first order necessary condition for $(a_{1|t+1}, \ldots, a_{t|t+1})$ to maximize $f(\alpha_1, \ldots, \alpha_t|\alpha_{t+1}, y)$, the identity $a_{t-1|t+1}(\alpha_{t+1}) = a_{t-1|t}(\alpha_{t+1}(\alpha_{t+1}))$ and the difference equation (10) defining $\Sigma_{t|t+1}(\alpha_{t+1})$.

(2) Compute approximations B_t, $B_t^{(1)}$, $B_t^{(2)}$, $B_t^{(3)}$ and $B_t^{(4)}$ of the value and first four derivatives of $b_{t|t+1}(\alpha_{t+1})$ at a_{t+1}. Recall that $b_{t|t+1}(\alpha_{t+1})$ is the conditional mode of α_t given α_{t+1} and y. For $t = n$, we only compute an approximation B_n of the value b_n, the conditional mode of α_n given y. Appendix C.3 defines these approximations and shows how to compute them. Specifically, equation (62) defines $B_t^{(r)}$ as a function of the $a_t^{(i)}$ and $s_t^{(i)}$. The approximations are based on an approximation of $b_{t|t+1}(\alpha_{t+1}) - a_{t|t+1}(\alpha_{t+1})$ using a first order necessary condition for $b_{t|t+1}(\alpha_{t+1})$ to maximize $f(\alpha_t|\alpha_{t+1}, y)$.

(3) Compute approximations M_t, $M_t^{(1)}$, $M_t^{(2)}$, $M_t^{(3)}$ and $M_t^{(4)}$ of the value and first four derivatives of $\mu_{t|t+1}(\alpha_{t+1})$ at a_{t+1}. Recall that $\mu_{t|t+1}(\alpha_{t+1})$ is the conditional mean of α_t given α_{t+1} and y. Appendix (C.4) defines these approximations. We compute $M_t^{(r)}$, from equation (70), as a function of the $B_t^{(i)}$, $a_t^{(i)}$ and $s_t^{(i)}$.

2.3. **A Backward Pass.** We use the backward pass to draw a random variate α^* from the distribution with density $g(\alpha|y)$ and evaluate $g(\alpha^*|y)$. One can also evaluate $g(\alpha|y)$ at an arbitrary value α^* without drawing.

To implement the backward pass, we use the following approximation of the derivative of $\log f(\alpha_t|\alpha_{t+1}, y)$, based on (9) and the approximations $M_{t-1|t}(\alpha_t)$ of $\mu_{t-1|t}(\alpha_t)$, $X_{t-1|t}(\alpha_t)$ of $x_{t-1|t}(\alpha_t)$ and $\Psi_t^{(1,0)}(\alpha_t, \alpha_{t+1})$ of $\psi_t^{(1,0)}(\alpha_t, \alpha_{t+1})$.

$$
\begin{aligned}
H_t^{(1)}(\alpha_t; \alpha_{t+1}) \doteq \bar{c}_t - \bar{\Omega}_{t-1,t} M_{t-1|t}(\alpha_t) - \bar{\Omega}_{t,t}\alpha_t - \bar{\Omega}_{t,t+1}\alpha_{t+1} \\
+ X_{t-1|t}(\alpha_t) + \Psi_t^{(1,0)}(\alpha_t, \alpha_{t+1}).
\end{aligned} \tag{12}
$$

We define the approximation $X_{t-1|t}(\alpha_t)$ and show how to compute it in Appendix E. $M_{t-1|t}(\alpha_t)$ is the polynomial

$$
M_{t-1|t}(\alpha_t) = \sum_{r=0}^{4} \frac{M_{t-1}^{(r)}}{r!}(\alpha_t - a_t)^r, \tag{13}
$$

We require routines to evaluate $\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$ for several orders p, q, so in principle it is not necessary to approximate $\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$. However, we find the computational costs high relative to the benefits. We already have $\psi_t^{(p,q)} = \psi_t^{(p,q)}(a_t, a_{t+1})$ from the forward pass, and we choose to approximate $\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$ by

$$\Psi_t^{(p,q)}(\alpha_t, \alpha_{t+1}) \doteq \sum_{r=0}^{P-p} \sum_{s=0}^{Q-q} \psi_t^{(p+r,q+s)} \frac{(\alpha_t - a_t)^r}{r!} \frac{(\alpha_{t+1} - a_{t+1})^s}{s!}. \tag{14}$$

The backward pass consists of performing the following steps, for $t = n, \ldots, 1$.

(1) Evaluate $B_{t|t+1}(\alpha_{t+1}^*)$, where $B_{t|t+1}(\alpha_{t+1})$ is the polynomial given by

$$B_{t|t+1}(\alpha_{t+1}) = \sum_{r=0}^{5} \frac{B_t^{(r)}}{r!} (\alpha_{t+1} - a_{t+1})^r \tag{15}$$

(2) Compute $H_t^{(r)}(B_{t|t+1}(\alpha_{t+1}^*); \alpha_{t+1}^*), r = 2, \ldots, 5$, using (12).
(3) Draw α_t^* and evaluate $g(\alpha_t|\alpha_{t+1}, y)$ at α_t^* and α_{t+1}^*. The density $g(\alpha_t|\alpha_{t+1}, y)$ is a member of the five-parameter perturbed Gaussian distribution described in Appendix G of McCausland [2012]. The mode parameter is given by $b = B_{t|t+1}(\alpha_{t+1}^*)$, and the derivative parameters are given by $h_r = H_t^{(r)}(B_{t|t+1}(\alpha_{t+1}^*); \alpha_{t+1}^*)$, $r = 2, \ldots, 5$. These give the desired mode $B_{t|t+1}(\alpha_{t+1}^*)$ and desired derivatives of $\log g(\alpha_t|\alpha_{t+1}, y)$ at this mode.

3. Getting it right

In posterior simulation, analytical or coding errors can lead to reasonable but inaccurate results. Geweke [2004] develops tests for the correctness of posterior simulations, based on two different methods for simulating the joint distribution of a model's observable and unobservable variables. Correctness tests take the form of tests of the hypothesis that the two samples come from the same distribution. Since the two methods have little in common, the tests have power against a wide array of conceptual and coding errors. We apply these ideas to build tests for the correctness of the independence Metropolis-Hastings update of the target distribution using the HESSIAN approximation $g(\alpha|y, \theta)$ as a proposal distribution.

We do this for the the ASV-Student model described in the next section. We choose a fixed value of θ of the parameter vector. Then we generate a large sample from the conditional distribution of α and y given θ. We initialize with a draw $\alpha^{(0)}$ from the conditional distribution of α given θ, then draw $\{\alpha^{(m)}, y^{(m)}\}_{m=1}^M$ as follows. For $m = 1, \ldots, M$,

(1) Draw $y^{(m)}$ from the conditional distribution of y given θ and α, with α set to $\alpha^{(m-1)}$.

(2) Update from $\alpha^{(m-1)}$ to $\alpha^{(m)}$ using an independence Metropolis-Hastings step, with $g(\alpha|y,\theta)$ as a proposal distribution and $y = y^{(m)}$.

This is a Gibbs sampler for the conditional distribution of α and y given θ. The initial and stationary distributions of this chain are both equal to this distribution. By induction, so are the distributions of all the $(\alpha^{(m)}, y^{(m)})$. In particular, $\alpha^{(m)} \sim \mathrm{N}(\bar{\alpha}\imath, \bar{\Omega}^{-1})$ for all m, where $\imath$ is the n-vector with all elements equal to one. This implies that for all $m = 1, \ldots, M$ and $q \in (0,1)$, the following indicators are Bernoulli with probability parameter q:

$$I_{t,q}^{(m)} \doteq 1\left(\frac{\alpha_t^{(m)} - \bar{\alpha}}{\sigma/\sqrt{1-\phi^2}} \leq \Phi^{-1}(q)\right), t = 1, \ldots, n, \tag{16}$$

$$I_{t|t-1,q}^{(m)} \doteq 1\left(\frac{\alpha_t^{(m)} - (1-\phi)\bar{\alpha} - \phi\alpha_{t-1}^{(m)}}{\sigma} \leq \Phi^{-1}(q)\right), t = 1, \ldots, n, \tag{17}$$

where $\Phi(x)$ is the cumulative distribution function of the univariate standard Gaussian distribution.

We use sample means of the $I_{t,q}^{(m)}$ and $I_{t|t-1,q}^{(m)}$ to test the hypotheses that the corresponding population means are equal to q. We report results for the ASV-Student model. The parameter values are fixed to $\bar{\alpha} = -9.0$, $\phi = 0.97$, $\sigma = 0.15$, $\rho = -0.3$ and $\nu = 10.0$. We use a vector of length $n = 20$ and a sample size of $M = 10^7$. We use the **R** package coda to compute time series numerical standard errors and use Gaussian asymptotic approximations to construct symmetric 95% and 99% intervals. The 95% confidence interval does not include q in 7 cases out of 360 (1.94%). The 99% confidence interval does not include q in a single case (0.28%). The sample mean always lies well within the interval $[q - 0.001, q + 0.001]$. These results fail to cast doubt on the correctness of the implementation.

4. Empirical example

4.1. **Models.** We consider two different stochastic volatility models with asymmetric volatility. The first model, which we will call ASV-Gaussian, is the basic asymmetric volatility model given in equations (1), (2) and (3).

The second model, which we will call ASV-Student, replaces the observation equation in (2) with

$$y_t = \exp(\alpha_t/2)\frac{v_t}{\sqrt{\lambda_t/\nu}}, \tag{18}$$

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

In order to allow us to draw parameters and states together in a single block, we will now integrate out λ_t to obtain the conditional distribution of y_t given α_t

and α_{t+1}. This distribution is a scaled non-central 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

$$\mu \doteq \rho\sqrt{\frac{\omega}{1-\rho^2}}[\alpha_{t+1} - d_t - \phi_t\alpha_t]$$

and unit variance; and λ_t is chi-squared with ν degrees of freedom. Therefore X is non-central Student's t with non-centrality parameter μ and ν degrees of freedom. The density of X is given by

$$\begin{aligned} f_X(x;\nu,\mu) &= \frac{\nu^{\nu/2}}{2^\nu}\frac{\Gamma(\nu+1)}{\Gamma(\nu/2)}\exp(-\mu^2/2)(\nu+x^2)^{-\nu/2} \\ &\times \left[\frac{\sqrt{2}\mu x}{\nu+x^2}\frac{M\left(\frac{\nu}{2}+1;\frac{3}{2};\frac{\mu^2x^2}{2(\nu+x^2)}\right)}{\Gamma(\frac{\nu+1}{2})} + \frac{1}{\sqrt{\nu+x^2}}\frac{M\left(\frac{\nu+1}{2};\frac{1}{2};\frac{\mu^2x^2}{2(\nu+x^2)}\right)}{\Gamma(\nu/2+1)}\right], \end{aligned} \tag{19}$$

where $\Gamma(\nu)$ is the gamma function and $M(a;b;z)$ is Kummer's function of the first kind, a confluent hypergeometric function given by

$$M(a;b;z) = \sum_{k=0}^{+\infty}\frac{(a)_k}{(b)_k}\frac{z^k}{k!}, \tag{20}$$

where $(a)_k = a(a+1)\dots(a+k-1)$. See Scharf (1991). We obtain the conditional density $f(y_t|\alpha_t,\alpha_{t+1})$ 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 $\omega_t = \omega$, $\phi_t = \phi$ 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 $\omega_0 = (1-\phi^2)\omega$ and $d_0 = \bar{\alpha}$.

We express our prior uncertainty about the parameters in terms of a multivariate Gaussian distribution over the transformed parameter vector

$$\theta = (\log\sigma, \tanh^{-1}\phi, \bar{\alpha}, \tanh^{-1}\rho, \log\nu).$$

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. The result is the following prior:

$$\theta \sim N\left(\begin{bmatrix} -1.8 \\ 2.1 \\ -11.0 \\ -0.4 \\ 2.5 \end{bmatrix}, \begin{bmatrix} 0.125 & -0.05 & 0 & 0 & 0 \\ -0.05 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0 & 0.25 \end{bmatrix}\right).$$

4.2. **MCMC and IS methods for posterior simulation.** To illustrate the performance of the HESSIAN approximation, we use Markov chain Monte Carlo (MCMC) and importance sampling posterior simulations and compare with Omori et al. [2007]. For both posterior simulations, we draw jointly θ and α. 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 below, 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 unnormalised approximation of $f(\theta|y)$. Let θ° be the maximiser of $\tilde{g}(\theta|y)$ and Σ° be the inverse of the negative Hessian of $\log \tilde{g}(\theta|y)$ at θ°. Also let 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 parameter 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 density $g(\theta|y)g(\alpha|\theta, y)$ is accepted with probability

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

The fact that we can approximate the entire posterior distribution opens up the possibility of doing importance sampling. Unlike proposals in MCMC, importance draws do not need to be independent and this presents opportunities for variance reduction. We exploit this fact to do importance sampling using a combination of quasi-random and pseudo-random sequences for draws of θ. We construct M blocks of length S each, for a total of MS draws. S should be a power of two, which is 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 importance sampling 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

$$w^{(m,s)} = \frac{f(\theta^{(m,s)}, \alpha^{(m,s)}, y)}{g(\theta^{(m,s)}, \alpha^{(m,s)}|y)}.$$

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 standard error for R using the delta method. This gives the following approximation of the numerical variance of the ratio R:

$$\hat{\sigma}_R^2 \doteq (\hat{\sigma}_N^2 - 2R\hat{\sigma}_{ND} + R^2\hat{\sigma}_D^2)(MS/D)^2,$$

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.

4.3. **Results.** For the ASV-Gaussian model, we report results of the HESSIAN independence Metropolis-Hastings and importance sampling posterior simulations. We implement the procedure of Omori et al. [2007], denoted OCSN, and compare results. We apply the three methods to two real data sets. The first consists of daily returns of the S&P 500 index from January 1980 to December 1987, for a total of 2022 observations. This matches a sample used by Yu [2005]. The second 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.

In the MCMC posterior simulation, the initial 10 draws are discarded and the independence Metropolis-Hastings chain is of length 12,800. We choose this chain size

to match the total draws of the importance sampling chain where we use $M = 100$ and $S = 128$. In our replication of the OCSN chain, the initial 500 values are discarded and we retain the 12,800 subsequent values. Table 2 gives the computational time by dataset and estimation procedure. For all three methods, the code is written in C++. We used a Windows PC with an Intel Core i5 2.90GHz processor.

Table 3 summarizes estimation results of the ASV-Gaussian model. The labels *HIS*, *HIM* and *OCSN* indicate the HESSIAN importance sample, the HESSIAN independence Metropolis-Hastings chain, and the chain obtained using the Omori et al. [2007] procedure. 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 numerical approximations of the posterior mean. The RNE measures numerical efficiency relative to that of the mean of a random sample from the posterior. We use the results of Section 4.2 to compute the NSE and RNE of the importance sampling 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.

The HIS and HIM methods produce numerical estimates of the same posterior mean. We implement the procedure of Omori et al. [2007] using the prior described in their article, which is different from our own. As a result, reported values are different not only because of numerical sample variance but also because the posterior mean is slightly different.

The HESSIAN importance sampler outperforms the OCSN method in all cases. Its numerical efficiency is higher compared to OCSN, and apart from the unconditional mean $\bar{\alpha}$ of log volatility, at least four times higher. The efficiency of the importance sample means are 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, measured by $(1/(\text{Time}\times \text{NSE}^2))$. Except for the unconditional mean of the log volatility, the HESSIAN independence Metropolis Hastings methods outperforms the OCSN procedure, with regard to the relative numerical efficiency and precision per unit time.

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 4 summarizes the results of both datasets. The estimates of the parameters

$\bar{\alpha}$, ϕ, σ and ρ, for the real data, are close to those obtained with the ASV-Gaussian. The numerical efficiency is also substantially higher.

Nakajima and Omori [2009] proposed a procedure to estimate the ASV-Student models. It is an extension of Omori et al. [2007]. This procedure is illustrated using S&P500(January 1, 1970 to December 31, 2003) and Topix (from Janury 6, 1992 to December 30, 2004). Table 4 and Table 5 in Nakajima and Omori [2009] report respectively results for S&P500 dataset and Topix dataset. Their reported results for the ASV-Student model (SVLt in their paper) are similar to ours, except the leverage coefficient of S&P500 dataset. The inefficiency factor is the inverse of the numerical efficiency factor. The inverse of the reported values of their inefficiency factors spanned from 0.006 (ν) to 0.291 (μ) for the S&P500 dataset. For the Topix data, the highest value of efficiency reported is 0.0893. Although we do not used the same length of data, comparing these figures to those reported in Table 4 suggests that our procedure is more numerically efficient than the procedure of Nakajima and Omori [2009].

5. 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 the HESSIAN posterior simulators.

Simulations on artificial and real data suggest 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 et al. [2004] and Omori and Watanabe [2008]. The high numerical efficiency relies on $g(\alpha|\theta, y)$ being extremely close to the target density $f(\alpha|\theta, y)$. Constructing a joint proposal of θ and α not only solves the problem of numerical inefficiencies due to posterior autocorrelation of α but also those due to posterior dependence between θ and α.

The scope of applications goes beyond the ASV-Gaussian and ASV-Student models. Application to a new model of the form (4) only requires routines to compute partial derivatives of the log conditional densities $\log f(y_t|\alpha_t, \alpha_{t+1})$ with respect to α_t and α_{t+1}. This requirement 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 of these derivatives. 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 routines, which do not depend on the particular functions involved.

We now require the state vector, α, to be Gaussian. We are currently trying to extend the HESSIAN method to models where the state vector is Markov, but not necessary Gaussian. We are also working on approximations to filtering densities, useful for sequential learning.

Appendix A. Precomputation

Here we compute the precision $\bar{\Omega}$ and covector $\bar{c}$ of the marginal distribution of α, and the mode $a = (a_1, \ldots, a_n)$ of the target distribution. Bi-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 $\Sigma_1, \ldots, \Sigma_t, \ldots, \Sigma_n$.

As the state dynamics are no different, we compute $\bar{\Omega}$ and $\bar{c}$ exactly as in McCausland (2010):

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

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

$$\bar{c}_t = \begin{cases} \omega_{t-1} d_{t-1} - \omega_t \phi_t d_t & t = 1, \ldots, n-1, \\ \omega_{n-1} d_{n-1} & t = n. \end{cases} \tag{21}$$

As in McCausland (2010), 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 non-zero upper triangular elements given by

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

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

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

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

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

and its elements are

(22)
$$\bar{\bar{c}}_t(\alpha) = \begin{cases} \bar{c}_t + \bar{\bar{\Omega}}_{t,t}\alpha_t + \bar{\bar{\Omega}}_{t,t+1}\alpha_{t+1} + \psi_t^{(1,0)}(\alpha_t, \alpha_{t+1}) & t = 1 \\ \bar{c}_t + \bar{\bar{\Omega}}_{t,t-1}\alpha_{t-1} + \bar{\bar{\Omega}}_{t,t}\alpha_t + \bar{\bar{\Omega}}_{t,t+1}\alpha_{t+1} + \psi_t^{(1,0)}(\alpha_t, \alpha_{t+1}) + \psi_{t-1}^{(0,1)}(\alpha_{t-1}, \alpha_t) & t = 2, \ldots, n-1 \\ \bar{c}_n + \bar{\bar{\Omega}}_{n,n-1}\alpha_{n-1} + \bar{\bar{\Omega}}_{nn}\alpha_n + \psi_n^{(1)}(\alpha_n) + \psi_{n-1}^{(0,1)}(\alpha_{n-1}, \alpha_n)(\alpha_{n-1}, \alpha_n) & t = n. \end{cases}$$

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 (2010), 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 non-convergence.

Appendix B. Polynomial approximations of $a_{t|t+1}$ and $s_{t|t+1}$

Here we compute coefficients of polynomial approximations of $a_{t|t+1}(\alpha_{t+1})$ and $s_{t|t+1}(\alpha_{t+1})$. Recall that these are the conditional mean and log variance of α_t given α_{t+1} according to a Gaussian approximation of the conditional distribution of $\alpha_1, \ldots, \alpha_t$ given α_{t+1} and y. The approximations are exact Taylor series expansions around a_{t+1} and so the coefficients are based on the derivatives of these functions at a_{t+1}.

We derive recursive expressions for these derivatives that are correct for any order r. In practice, the computational cost rises quickly and the benefits diminish quickly in r. We provide simplified expressions for $a_t^{(r)} \doteq a_{t|t+1}^{(r)}(a_{t+1})$ up to order $r = 5$ and $s_t^{(r)} \doteq s_{t|t+1}^{(r)}(a_{t+1})$ up to order $r = 4$.

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

$$\log f(\alpha_1|\alpha_2, y) = -\frac{1}{2}\bar{\Omega}_{1,1}\alpha_1^2 - \bar{\Omega}_{1,2}\alpha_1\alpha_2 + \bar{c}_1\alpha_1 + \log f(y_1|\alpha_1, \alpha_2) + k. \tag{23}$$

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

$$-\bar{\Omega}_{1,1}a_{1|2}(\alpha_2) - \bar{\Omega}_{1,1}\alpha_2 + \bar{c}_1 + \psi_1^{(1,0)}(a_{1|2}(\alpha_2), \alpha_2) = 0. \tag{24}$$

Taking the derivative of (24) 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

$$\bar{\bar{\Omega}}_{1,1|2}(\alpha_2)a_{1|2}^{(1)}(\alpha_2) = -\bar{\bar{\Omega}}_{1,2|2}(\alpha_2). \tag{25}$$

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

$$a_{1|2}^{(1)}(\alpha_2) = -\Sigma_{1|2}(\alpha_2)\bar{\bar{\Omega}}_{1,2|2}(\alpha_2), \tag{26}$$

where $\Sigma_{1|2}(\alpha_2) = \bar{\bar{\Omega}}_{1,1|2}^{-1}(\alpha_2)$ from equation (10). 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 (25) $(r-1)$ 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

$$a_{1|2}^{(r)}(\alpha_2) = -\Sigma_{1|2}(\alpha_2)\left[\sum_{i=0}^{r-2}\binom{r-1}{i}\bar{\bar{\Omega}}_{1,1|2}^{(r-1-i)}(\alpha_2)a_{1|2}^{(i+1)}(\alpha_2) + \bar{\bar{\Omega}}_{1,2|2}^{(r-1)}(\alpha_2)\right]. \tag{27}$$

Finally, we evaluate (27) at $\alpha_2 = a_2$ to obtain

$$a_1^{(r)} = -\Sigma_1\left[\sum_{i=0}^{r-2}\binom{r-1}{i}\bar{\bar{\Omega}}_{1,1}^{(r-1-i)}a_1^{(i+1)} + \bar{\bar{\Omega}}_{1,2}^{(r-1)}\right]. \tag{28}$$

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 $i \geq 1$,

$$\begin{aligned}\Sigma_{1|2}^{(i)}(\alpha_2) &= \sum_{j=1}^{i}\exp(s_{1|2}(\alpha_2))B_{i,j}(s_{1|2}^{(1)}(\alpha_2), \ldots, s_{1|2}^{(i-j+1)}(\alpha_2)) \\ &= \Sigma_{1|2}(\alpha_2)B_i(s_{1|2}^{(1)}(\alpha_2), \ldots, s_{1|2}^{(i)}(\alpha_2)),\end{aligned} \tag{29}$$

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 (26) $(r-1)$ 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:

$$a_1^{(r)} = -\Sigma_1\sum_{i=0}^{r-1}\binom{r-1}{i}B_i(s_1^{(1)},\ldots,s_1^{(i)})\bar{\bar{\Omega}}_{1,2}^{(r-1-i)}. \tag{30}$$

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

$$\begin{aligned} 0 =& \bar{c}_t - \bar{\Omega}_{t-1,t}a_{t-1|t+1}(\alpha_{t+1}) - \bar{\Omega}_{t,t}a_{t|t+1}(\alpha_{t+1}) - \bar{\Omega}_{t,t+1}\alpha_{t+1} \\ &+ \psi_{t-1}^{(0,1)}(a_{t-1|t}(a_{t|t+1}),a_{t|t+1}) + \psi_t^{(1,0)}(a_{t|t+1},\alpha_{t+1}). \end{aligned} \tag{31}$$

Taking the derivative of (31) with respect to α_{t+1} gives

$$\bar{\bar{\Omega}}_{t,t-1}(\alpha_{t+1})a_{t-1|t+1}^{(1)}(\alpha_{t+1}) + \bar{\bar{\Omega}}_{t,t}(\alpha_{t+1})a_{t|t+1}^{(1)}(\alpha_{t+1}) + \bar{\bar{\Omega}}_{t,t+1}(\alpha_{t+1}) = 0. \tag{32}$$

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

$$a_{t-1|t+1}^{(1)}(\alpha_{t+1}) = a_{t-1|t}^{(1)}(a_{t|t+1}(\alpha_{t+1}))a_{t|t+1}^{(1)}(\alpha_{t+1}). \tag{33}$$

Substituting (33) in (32), 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

$$a_{t|t+1}^{(1)}(\alpha_{t+1}) = -\Sigma_{t|t+1}(\alpha_{t+1})\bar{\bar{\Omega}}_{t,t+1}(\alpha_{t+1}), \quad t = 2,\ldots,n-1, \tag{34}$$

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 (34) gives

$$a_t^{(1)} = -\Sigma_t\bar{\bar{\Omega}}_{t,t+1}. \tag{35}$$

For $r \geq 2$, we use Leibniz's rule to differentiate (32) $(r-1)$ times with respect to α_{t+1} and obtain

(36)
$$\sum_{i=0}^{r-1} \binom{r-1}{i} \left(\bar{\bar{\Omega}}_{t,t-1}^{(i)}(\alpha_{t+1}) a_{t-1|t+1}^{(r-i)}(\alpha_{t+1}) + \bar{\bar{\Omega}}_{t,t}^{(i)}(\alpha_{t+1}) a_{t|t+1}^{(r-i)}(\alpha_{t+1}) \right) = -\bar{\bar{\Omega}}_{t,t+1}^{(r-1)}(\alpha_{t+1}).$$

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

$$a_{t-1|t+1}^{(i)}(\alpha_{t+1}) = \sum_{j=1}^{i} a_{t-1|t}^{(j)}(a_{t|t+1}) B_{i,j}(a_{t|t+1}^{(1)}(\alpha_{t+1}), \ldots, a_{t|t+1}^{(i-j+1)}(\alpha_{t+1})). \tag{37}$$

If we substitute $a_{t-1|t+1}^{(i)}(\alpha_{t+1})$ of (37) in (36) and set $\alpha_{t+1} = a_{t+1}$, we obtain

(38)
$$\sum_{i=0}^{r-1} \binom{r-1}{i} \left\{ \bar{\bar{\Omega}}_{t,t-1}^{(i)} \left[\sum_{j=1}^{r-i} a_{t-1}^{(j)} B_{r-i,j}(a_t^{(1)}, \ldots, a_t^{(r-i-j+1)}) \right] + \bar{\bar{\Omega}}_{t,t}^{(i)} a_t^{(r-i)} \right\} = -\bar{\bar{\Omega}}_{t,t+1}^{(r-1)}.$$

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 (29), 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 (34) with respect to α_{t+1}, and evaluating at $\alpha_{t+1} = a_{t+1}$, we obtain

$$a_t^{(r)} = \sum_{i=0}^{r-1} \binom{r-1}{i} B_i(s_t^{(1)}, \ldots, s_t^{(i)}) \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(r-1-i)}. \tag{39}$$

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 (83) and (84) 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 (35). Taking $r = 2$ in (38) 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

$$a_t^{(2)} = \left(\gamma_t a_t^{(1)}a_{t-1}^{(2)} - \Sigma_t\bar{\bar{\mathbf{\Omega}}}_t^{(1)}\right)a_t^{(1)} - \Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(1)}, \tag{40}$$

where $\gamma_t = -\Sigma_t\bar{\bar{\Omega}}_{t,t-1}$ and $\bar{\bar{\mathbf{\Omega}}}_t^{(i)} = \bar{\bar{\Omega}}_{t,t-1}^{(i)}a_t^{(1)} + \bar{\bar{\Omega}}_{t,t}^{(i)}$. Setting $r = 2$ in (39) gives

$$a_t^{(2)} = s_t^{(1)}a_t^{(1)} - \Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(1)}. \tag{41}$$

Equating the right hand sides of (40) and (41) and solving for $s_t^{(1)}$ gives

$$s_t^{(1)} = \gamma_t a_t^{(1)}a_{t-1}^{(2)} - \Sigma_t\bar{\bar{\mathbf{\Omega}}}_t^{(1)}. \tag{42}$$

Setting $r = 3$ in (38) 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

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

We use (42) to simplify this to

$$\begin{aligned}a_t^{(3)} =&2s_t^{(1)}a_t^{(2)} + \left(\gamma_t a_t^{(1)}a_{t-1}^{(3)} - 2\Sigma_t\bar{\bar{\Omega}}_{t,t-1}^{(1)}a_{t-1}^{(2)}\right)\left(a_t^{(1)}\right)^2\\ &+ \left(\gamma_t a_t^{(2)}a_{t-1}^{(2)} - \Sigma_t\bar{\bar{\mathbf{\Omega}}}_t^{(2)}\right)a_t^{(1)} - \Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(2)}.\end{aligned} \tag{43}$$

Setting $r = 3$ in (39) gives an alternative expression for $a_t^{(3)}$:

$$\begin{aligned} a_t^{(3)} &= \left(s_t^{(2)} + \left(s_t^{(1)}\right)^2\right) a_t^{(1)} - \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(2)} - 2s_t^{(1)} \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(1)} \\ &= \left(s_t^{(2)} + \left(s_t^{(1)}\right)^2\right) a_t^{(1)} - \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(2)} + 2s_t^{(1)} \left(a_t^{(2)} - s_t^{(1)} a_t^{(1)}\right) \\ &= \left(s_t^{(2)} - \left(s_t^{(1)}\right)^2\right) a_t^{(1)} + 2s_t^{(1)} a_t^{(2)} - \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(2)}. \end{aligned} \tag{44}$$

Equating the right hand sides of (43) and (44) and solving for $s_t^{(2)}$ gives

$$s_t^{(2)} = \left(s_t^{(1)}\right)^2 + \left(\gamma_t a_t^{(1)} a_{t-1}^{(3)} - 2\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(2)}\right) a_t^{(1)} + \left(\gamma_t a_t^{(2)} a_{t-1}^{(2)} - \Sigma_t \bar{\bar{\mathbf{\Omega}}}_t^{(2)}\right). \tag{45}$$

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

$$\begin{aligned} a_t^{(4)} = &\left(\gamma_t a_t^{(1)} a_{t-1}^{(4)} - 3\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(3)}\right) \left(a_t^{(1)}\right)^3 + 3\left(\gamma_t a_t^{(2)} a_{t-1}^{(3)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(2)}\right) \left(a_t^{(1)}\right)^2 \\ &+ \left(\gamma_t a_t^{(3)} a_{t-1}^{(2)} - 3\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_t^{(2)} a_{t-1}^{(2)} - \Sigma_t \bar{\bar{\mathbf{\Omega}}}_t^{(3)}\right) a_t^{(1)} - \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(3)} \\ &+ 3\left(s_t^{(2)} - \left(s_t^{(1)}\right)^2\right) a_t^{(2)} + 3s_t^{(1)} a_t^{(3)}, \end{aligned} \tag{46}$$

$$\begin{aligned} s_t^{(3)} = &- \left(s_t^{(1)}\right)^3 + 3s_t^{(1)} s_t^{(2)} + \left(\gamma_t a_t^{(1)} a_{t-1}^{(4)} - 3\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(3)}\right) \left(a_t^{(1)}\right)^2 \\ &+ 3\left(\gamma_t a_t^{(2)} a_{t-1}^{(3)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(2)}\right) a_t^{(1)} + \left(\gamma_t a_t^{(3)} - 3\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_t^{(2)}\right) a_{t-1}^{(2)} - \Sigma_t \bar{\bar{\mathbf{\Omega}}}_t^{(3)} \end{aligned} \tag{47}$$

$$\begin{aligned} a_t^{(5)} = &- \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(4)} + \left(\gamma_t a_{t-1}^{(5)} a_t^{(1)} - 4\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(4)}\right) \left(a_t^{(1)}\right)^4 \\ &+ 6\left(\gamma_t a_{t-1}^{(4)} a_t^{(2)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(3)}\right) \left(a_t^{(1)}\right)^3 \\ &+ 4\left(\gamma_t a_{t-1}^{(3)} a_t^{(3)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(3)} a_{t-1}^{(2)} - 2\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(3)} a_t^{(2)}\right) \left(a_t^{(1)}\right)^2 \\ &+ \left(\gamma_t \left(a_{t-1}^{(2)} a_t^{(4)} + 3a_{t-1}^{(3)} \left(a_t^{(2)}\right)^2\right) - \Sigma_t \bar{\bar{\mathbf{\Omega}}}_t^{(4)} - 6\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(2)} a_t^{(2)} - 4\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(2)} a_t^{(3)}\right) a_t^{(1)} \\ &+ 4s_t^{(1)} a_t^{(4)} + 6\left(s_t^{(2)} - \left(s_t^{(1)}\right)^2\right) a_t^{(3)} + 4\left(s_t^{(3)} + \left(s_t^{(1)}\right)^3 - 3s_t^{(1)} s_t^{(2)}\right) a_t^{(2)}, \end{aligned} \tag{48}$$

(49)
$$
\begin{aligned}
s_t^{(4)} = &\left(\gamma_t a_{t-1}^{(5)} a_t^{(1)} - 4\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(4)}\right)\left(a_t^{(1)}\right)^3 \\
&+ 6\left(\gamma_t a_{t-1}^{(4)} a_t^{(2)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(3)}\right)\left(a_t^{(1)}\right)^2 \\
&+ 4\left(\gamma_t a_{t-1}^{(3)} a_t^{(3)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(3)} a_{t-1}^{(2)} - 2\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(3)} a_t^{(2)}\right) a_t^{(1)} \\
&+ \left(\gamma_t\left(a_{t-1}^{(2)} a_t^{(4)} + 3a_{t-1}^{(3)}\left(a_t^{(2)}\right)^2\right) - \Sigma_t \bar{\bar{\mathbf{\Omega}}}_t^{(4)} - 6\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(2)} a_t^{(2)} - 4\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(2)} a_t^{(3)}\right) \\
&+ \left(s_t^{(1)}\right)^4 + 4 s_t^{(1)} s_t^{(3)} + 3\left(s_t^{(2)} - 2\left(s_t^{(1)}\right)^2\right) s_t^{(2)}.
\end{aligned}
$$

Appendix C. Polynomial approximations of $b_t^{(r)}$ and $\mu_t^{(r)}$

C.1. **First derivative of $\log f(\alpha_t|\alpha_{t+1}, y)$.** In this subsection, 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 $h_1(\alpha_1; \alpha_2) = \log f(y_1|\alpha_1, \alpha_2)$ is written

$$
h_1^{(1)}(\alpha_1; \alpha_2) = \psi_1^{(1,0)}(\alpha_1, \alpha_2) + \bar{c}_1 - \bar{\Omega}_{1,2}\alpha_2 - \bar{\Omega}_{1,1}\alpha_1. \tag{50}
$$

For $t = 2, \ldots, n-1$, we compute $f(\alpha_t|\alpha_{t+1}, y)$ as a marginal density of $f(\alpha_{1:t}|\alpha_{t+1}, y)$. Thus, we have

$$
\begin{aligned}
f(\alpha_t|\alpha_{t+1}, y) &= \int f(\alpha_{1:t-1}, \alpha_t|\alpha_{t+1}, y)\, d\alpha_{1:t-1} \\
&\propto f(\alpha_{t+1}|\alpha_t) f(y_t|\alpha_t, \alpha_{t+1}) c(\alpha_t),
\end{aligned} \tag{51}
$$

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 (51) and differentiating with respect to α_t gives

$$
h_t^{(1)}(\alpha_t; \alpha_{t+1}) = \frac{\partial \log c(\alpha_t)}{\partial \alpha_t} + \frac{\partial \log f(\alpha_{t+1}|\alpha_t)}{\partial \alpha_t} + \frac{\partial \log f(y_t|\alpha_t, \alpha_{t+1})}{\partial \alpha_t}. \tag{52}
$$

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 derivatives $h_t(\alpha_t;\alpha_{t+1})$ then becomes

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

The first term above simplifies as in Appendix C of McCausland (2010). We use (7) to finally derive

$$\begin{aligned} h_t^{(1)}(\alpha_t;\alpha_{t+1}) =& \bar{c}_t - \bar{\Omega}_{t,t}\alpha_t - \bar{\Omega}_{t,t+1}\alpha_{t+1} + \psi_t^{(1,0)}(\alpha_t,\alpha_{t+1}) \\ &- \bar{\Omega}_{t-1,t}\mu_{t-1|t}(\alpha_t) + x_{t-1|t}(\alpha_t), \end{aligned} \tag{53}$$

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

$$h_n^{(1)}(\alpha_n) = \bar{c}_n - \bar{\Omega}_{n,n}\alpha_n + \psi_n^{(1)}(\alpha_n) - \bar{\Omega}_{n-1,n}\mu_{n-1|n}(\alpha_n) + x_{n-1|n}(\alpha_n). \tag{54}$$

C.2. **Approximation of $h_t^{(1)}(\alpha_t;\alpha_{t+1})$.** Since we do not know the conditional expectations $\mu_{t-1|t}(\alpha_t)$ and $x_{t-1|t}(\alpha_t)$, we cannot compute $h_t(\alpha_t;\alpha_{t+1})$ exactly. We propose an approximation $H_t^{(1)}(\alpha_t;\alpha_{t+1})$ of $h_t^{(1)}(\alpha_t;\alpha_{t+1})$. For $t = 2,\ldots,n-1$, we have

$$H_t^{(1)}(\alpha_t;\alpha_{t+1}) \doteq \bar{c}_t - \bar{\Omega}_{t,t}\alpha_t - \bar{\Omega}_{t,t+1}\alpha_{t+1} + \Psi_t^{(1,0)}(\alpha_t,\alpha_{t+1}) - \bar{\Omega}_{t-1,t}M_{t-1|t}(\alpha_t) + X_{t-1|t}(\alpha_t) \tag{55}$$

where $M_{t-1|t}(\alpha_t)$ is an approximation of $\mu_{t-1|t}(\alpha_t)$, $X_{t-1|t}(\alpha_t)$ is an approximation of $x_{t-1|t}(\alpha_t)$ and $\Psi_t^{(1,0)}(\alpha_t,\alpha_{t+1})$ is an approximation of $\psi_t^{(1,0)}(\alpha_t,\alpha_{t+1})$.[1] The polynomials $M_{t-1|t}(\alpha_t)$ and $\Psi_t^{(p,q)}(\alpha_t,\alpha_{t+1})$ are defined in (13) and (14).

[1]For $t = n$, we need just to replace $\Psi_t^{(1,0)}(\alpha_t,\alpha_{t+1})$ by $\Psi_n^{(1)}(\alpha_n)$ in (55) to obtain $H_n^{(1)}(\alpha_n)$, the approximation of $h_n^{(1)}(\alpha_n)$.

We construct $X_{t-1|t}(\alpha_t)$ in two steps. First, we 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 $a_{t-1|t}(\alpha_t)$:

$$\begin{aligned}\psi_{t-1}^{(0,1)}(\alpha_{t-1}, \alpha_t) \approx \psi_{t-1}^{(0,1)}(a_{t-1|t}(\alpha_t), \alpha_t) + \psi_{t-1}^{(1,1)}(a_{t-1|t}(\alpha_t), \alpha_t)(\alpha_{t-1} - a_{t-1|t}(\alpha_t)) \\ + \tfrac{1}{2}\psi_{t-1}^{(2,1)}(a_{t-1|t}(\alpha_t), \alpha_t)(\alpha_{t-1} - a_{t-1|t}(\alpha_t))^2.\end{aligned} \tag{56}$$

Taking the conditional expectation of both sides of (56), 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

$$\begin{aligned}x_{t-1|t}(\alpha_t) \approx &\psi_{t-1}^{(0,1)}(a_{t-1|t}(\alpha_t), \alpha_t) + \psi_{t-1}^{(1,1)}(a_{t-1|t}(\alpha_t), \alpha_t)(\mu_{t-1|t}(\alpha_t) - a_{t-1|t}(\alpha_t)) \\ &+ \frac{1}{2}\psi_{t-1}^{(2,1)}(a_{t-1|t}(\alpha_t), \alpha_t)\Sigma_{t-1|t}(\alpha_t).\end{aligned} \tag{57}$$

Now we define the polynomial $X_{t-1|t}(\alpha_t)$ as the R'th order Taylor series expansion of the right hand side of (57):

$$X_{t-1|t}(\alpha_t) \doteq \sum_{r=0}^{R} \frac{X_{t-1}^{(r)}}{r!}(\alpha_t - a_t)^r, \tag{58}$$

where $X_{t-1}^{(r)}$ is the r'th derivative of the RHS of (57) with respect to α_t, evaluated at a_t. We evaluate these derivatives bottom up using Faà Di Bruno's formula, equations (81) and (82), and Leibniz's rule, equation (77).

C.3. **Approximation of the conditional mode $b_{t|t+1}(\alpha_{t+1})$.** Recall that $b_{t|t+1}(\alpha_{t+1})$ is the conditional mode of α_t given α_{t+1} and y. We provide an approximation $B_{t|t+1}(\alpha_{t+1})$ of the Taylor expansion of $b_{t|t+1}(\alpha_{t+1})$ around $\alpha_{t+1} = a_{t+1}$. We show in this subsection how to compute the coefficients of the resulting polynomial. The degree of this polynomial is $R - 1 = 4$.

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 $a_{t|t+1}(\alpha_{t+1})$:

$$b_{t|t+1}(\alpha_{t+1}) \approx a_{t|t+1}(\alpha_{t+1}) - \frac{h_t^{(1)}(a_{t|t+1}(\alpha_{t+1}); \alpha_{t+1})}{h_t^{(2)}(a_{t|t+1}(\alpha_{t+1}); \alpha_{t+1})}. \tag{59}$$

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 $\alpha_{t+1} = a_{t+1}$.

Our approximations of numerator and denominator are, using (55) and its derivative, both evaluated at $\alpha_t = a_{t|t+1}(\alpha_{t+1})$, are

$$\begin{aligned} H_t^{(1)}(a_{t|t+1};\alpha_{t+1}) &= \bar{c}_t - \bar{\Omega}_{t,t}a_{t|t+1} - \bar{\Omega}_{t,t+1}\alpha_{t+1} + \Psi_t^{(1,0)}(a_{t|t+1},\alpha_{t+1}) \\ &\quad - \bar{\Omega}_{t-1,t}M_{t-1|t}(a_{t|t+1}) + X_{t-1|t}(a_{t|t+1}) \end{aligned} \tag{60}$$

$$\begin{aligned} H_t^{(2)}(a_{t|t+1};\alpha_{t+1}) &= -\bar{\Omega}_{t,t} + \Psi_t^{(2,0)}(a_{t|t+1},\alpha_{t+1}) \\ &\quad - \bar{\Omega}_{t-1,t}M_{t-1|t}^{(1)}(a_{t|t+1}) + X_{t-1|t}^{(1)}(a_{t|t+1}), \end{aligned} \tag{61}$$

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 $X_{t-1|t}(a_{t|t+1}(\alpha_{t+1}))$ with respect to α_{t+1}, at $\alpha_{t+1} = a_{t+1}$.

Based on equation (59), we define the following approximations $B_t^{(r)}$ of $b_t^{(r)}$, $r = 0,1,2,3$:

$$B_t^{(r)} \doteq a_t^{(r)} - \frac{\partial^r}{\partial \alpha_{t+1}^r}\left(\frac{H_t^{(1)}(a_{t|t+1}(\alpha_{t+1});\alpha_{t+1})}{H_t^{(2)}(a_{t|t+1}(\alpha_{t+1});\alpha_{t+1})}\right)\Bigg|_{\alpha_{t+1}=a_{t+1}}. \tag{62}$$

The second term on the right hand side of (62) is the r'th order derivative of a quotient, which we compute using the quotient rule for derivatives, equation (78) in Appendix E.

In practice, we find that 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:

$$B_n \doteq a_n - \frac{H_n^{(1)}(a_n)}{H_n^{(2)}(a_n)}. \tag{63}$$

C.4. **Coefficients of the polynomial approximation of $\mu_{t|t+1}(\alpha_{t+1})$.** 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(2011) suggests the following approximation for $\mu_{t|t+1} - b_{t|t+1}$:

$$\mu_{t|t+1} - b_{t|t+1} \approx \frac{1}{2}h_t^{(3)}(b_{t|t+1};\alpha_{t+1})\left[h_t^{(2)}(b_{t|t+1};\alpha_{t+1})\right]^{-2} \tag{64}$$

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

$$h_t^{(1)}(b_{t|t+1};\alpha_{t+1}) = 0 \tag{65}$$

Taking the derivative of (65) two times with respect to α_{t+1} gives

$$h_t^{(2)}(b_{t|t+1};\alpha_{t+1})b_{t|t+1}^{(1)} = \bar{\Omega}_{t,t+1} - \psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1}) \tag{66}$$

and

$$\begin{aligned} h_t^{(3)}(b_{t|t+1};\alpha_{t+1})\left(b_{t|t+1}^{(1)}\right)^2 + h_t^{(2)}(b_{t|t+1};\alpha_{t+1})b_{t|t+1}^{(2)} = &-2\frac{d\psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1})}{d\alpha_{t+1}} \\ &+ \psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1}) \end{aligned} \tag{67}$$

Solve for $h_t^{(3)}(b_{t|t+1};\alpha_{t+1})$ in equation (67) and divide by the square of $h_t^{(2)}(b_{t|t+1};\alpha_{t+1})$ to obtain

$$\begin{aligned} \frac{h_t^{(3)}(b_{t|t+1};\alpha_{t+1})}{\left(h_t^{(2)}(b_{t|t+1};\alpha_{t+1})\right)^2} = &-\frac{b_{t|t+1}^{(2)}/b_{t|t+1}^{(1)}}{h_t^{(2)}(b_{t|t+1};\alpha_{t+1})b_{t|t+1}^{(1)}} \\ &-\frac{2d\psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1})/d\alpha_{t+1} - \psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1})}{\left(h_t^{(2)}(b_{t|t+1};\alpha_{t+1})b_{t|t+1}^{(1)}\right)^2} \end{aligned} \tag{68}$$

Substitute the right hand side of equation (66) in (68) to obtain

$$\begin{aligned} \mu_{t|t+1} - b_{t|t+1} \approx &-\frac{1}{2}\frac{b_{t|t+1}^{(2)}/b_{t|t+1}^{(1)}}{\bar{\Omega}_{t,t+1} - \psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1})} \\ &-\frac{1}{2}\frac{2d\psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1})/d\alpha_{t+1} - \psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1})}{\left(\bar{\Omega}_{t,t+1} - \psi_t^{(1,1)}(b_{t|t+1},\alpha_{t+1})\right)^2} \end{aligned} \tag{69}$$

Based on equation (69), we define our approximation $M_{t|t+1}$ of $\mu_{t|t+1}$ as the Taylor series expansion of:

(70)

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

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 (78). 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 (81) and (82). Derivatives of the two main ratios in (70) are computed using the quotient rule in equation (78). We compute $M_t^{(r)} = M_{t|t+1}^{(r)}(a_{t+1})$, $r = 0, 1, 2$ using (70).

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

Appendix 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 (5), we can write

$$\psi_t(\alpha_t, \alpha_{t+1}) = -\frac{1}{2}\left[\log(2\pi/\beta) + \alpha_t + \beta(\varphi_t - \theta u_t)^2\right], \qquad t = 1, \dots, n-1, \tag{71}$$

$$\psi_n(\alpha_n) = -\frac{1}{2}\left[\log(2\pi) + \alpha_n + \varphi_n^2\right], \tag{72}$$

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

$$\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1}) = \begin{cases} -\frac{1}{2} - \frac{\beta}{2}\left(\tilde{\varphi}_{t,p} - 2\theta^2\phi u_t\right) & q = 0,\ p = 1 \\ -\frac{\beta}{2}\left(\tilde{\varphi}_{t,p} + 2\theta^2\phi^2\right) & q = 0,\ p = 2 \\ -\frac{\beta}{2}\tilde{\varphi}_{t,p} & q = 0, p \geq 3 \\ \beta\theta\left(\varphi_t - \theta u_t\right) & q = 1,\ p = 0 \\ \beta\theta\left(-\frac{1}{2}\varphi_t + \theta\phi\right) & q = 1,\ p = 1 \\ \beta\theta\left(-\frac{1}{2}\right)^p \varphi_t & q = 1,\ p \geq 2 \\ -\beta\theta^2 & q = 2,\ p = 0 \\ 0 & \text{otherwise,} \end{cases} \tag{73}$$

where

$$\tilde{\varphi}_{t,p} \doteq (-1)^p\varphi_t^2 - \left(-\frac{1}{2}\right)^{p-2}\theta\varphi_t\left(p\phi + \frac{1}{2}u_t\right), \qquad t = 1, \dots, n-1. \tag{74}$$

For $t = n$,

$$\psi_n^{(p)}(\alpha_n)(\alpha_n) = \begin{cases} -\frac{1}{2} - \frac{1}{2}\tilde{\varphi}_{n,p} & p = 1 \\ -\frac{1}{2}\tilde{\varphi}_{n,p} & p \geq 2, \end{cases} \tag{75}$$

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 (19) we can write $\psi_t(\alpha_t, \alpha_{t+1})$, for $t = 1, \ldots, n-1$, as

$$\psi_t(\alpha_t, \alpha_{t+1}) = k + \psi_{1,t}(\alpha_t, \alpha_{t+1}) + \psi_{2,t}(\alpha_t) + \psi_{3,t}(\alpha_t, \alpha_{t+1}), \tag{76}$$

where k does not depend on α_t or 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 (81) and (82).

(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 (78).

(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 non-zero) of $x = z^2$ with respect to z, evaluated at z, combining them using the Faà Di Bruno's rule, equations (81) and (82). 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

$$m^{(p)}(z) = 2\frac{\Gamma\left(\frac{\nu}{2}+1\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\left(z m_1^{(p)}(z) + r m_1^{(p-1)}(z)\right) + m_2^{(p)}(z), \quad p = 1, \ldots, P.$$

(8) Evaluate P derivatives of $\log m(z)$ with respect to z using the derivatives evaluated at step 7, and the logarithm rule, equations (79) and (80).

(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 (85) and (86).

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 (79) and (80). 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.

Appendix E. Rules for derivatives of compound functions

In this paper, we make extensive use of automatic rules for evaluating multiple derivatives of compound functions at a point. These rules combine multiple derivatives of component functions, also evaluated at points. This Appendix gathers these rules in one place.

For univariate functions f and g, we give well known rules for multiple derivatives of the product fg, the quotient f/g, and the composition $f \circ g$. We give a rule for multiple derivatives of $\log g$, a special case where we exploit the properties of the logarithmic function to simplify computations. We also give derivatives of $f \circ g$ for $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R}^2 \to \mathbb{R}$ and partial derivatives of $f \circ g$ for $f\colon \mathbb{R}^2 \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}^2$.

We have coded all of these rules as computer routines. Values passed to these 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 function 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.** For the first three rules, let x be a point in $\mathbb{R}$ and let 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

$$(fg)^{(p)}(x) = \sum_{r=0}^{p} \binom{p}{r} f^{(r)}(x) g^{(p-r)}(x), \quad p = 1, \ldots, P. \tag{77}$$

We have a routine taking the first P derivatives of f at x and the first P derivatives of g at x and returning the first P derivatives of fg at x.

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

$$(f/g)^{(p)}(x) = \frac{1}{g(x)} \left[f^{(p)}(x) - \sum_{r=0}^{p-1} \binom{p}{r} (f/g)^{(r)}(x) g^{(p-r)}(x) \right], \quad p = 1, \ldots, P. \tag{78}$$

We have a routine taking the first P derivatives of f at x and the first P derivatives of g at x and returning the first P derivatives of f/g at x.

Log rule. We consider the compound function $h = \log f$ and suppose that $f(x) > 0$. Then the function h is differentiable up to order P. Applying the quotient rule to

$$h^{(1)}(x) = \frac{f^{(1)}(x)}{f(x)} \tag{79}$$

gives

$$h^{(p)}(x) = \frac{1}{f(x)} \left[f^{(p)}(x) - \sum_{r=1}^{p-1} \binom{p-1}{r-1} h^{(p)}(x) f^{(p-r)}(x) \right], \quad p = 2, \ldots, P. \tag{80}$$

Together, equations (79) and (80) give the first P derivatives of $\log(f(x))$. We have a routine taking the first P derivatives of f at x and returning the first P derivatives of $\log f$ at x.

Faà di Bruno's rule for composite functions. Now suppose that x is a point in $\mathbb{R}$, g is a univariate function, P times differentiable at x, and f is 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

$$(f \circ g)^{(p)}(x) = \sum_{r=1}^{p} f^{(r)}(g(x)) B_{p,r}(g^{(1)}(x), \ldots, g^{(p-r+1)}(x)), \tag{81}$$

where the $B_{p,r}(z_1, \ldots, z_{p-r+1})$ 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 $B_{p,0}(z_1, \ldots, z_{p+1}) = 0$, $p > 0$, and the recursion

$$B_{p,r}(z_1, \ldots, z_{p-r+1}) = \sum_{i=r-1}^{p-1} \binom{p-1}{i} z_{p-i} B_{i,r-1}(z_1, \ldots, z_{i-r}), \quad r = 1, \ldots, p. \tag{82}$$

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

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

We have a routine taking the first P derivatives of g at x and the first P derivatives of f at $g(x)$, returning the first P derivatives of $f \circ g$ at x.

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 $f \colon \mathbb{R}^m \to \mathbb{R}$ and $g \colon \mathbb{R}^d \to \mathbb{R}^m$. We are only concerned with two special cases here, and we describe below how to compute partial derivatives for these 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

$$(f \circ g)^{(p)}(x) = \sum_{r=0}^{p} \sum_{s=\max\{0,1-r\}}^{p-r} f^{(r,s)}(g_1(x), g_2(x)) v_{p,(r,s)}, \tag{83}$$

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

$$v_{p,(r,s)} = \sum_{i=r+s-1}^{p-1} \binom{p-1}{i} \left[g_1^{(p-i)}(x) v_{i,(r-1,s)} + g_2^{(p-i)}(x) v_{i,(r,s-1)} \right]. \tag{84}$$

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

$$(f \circ g)^{(p,q)}(x_1, x_2) = \sum_{r=1}^{p+q} f^{(r)}(g(x_1, x_2)) v_{(p,q),r}, \tag{85}$$

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

$$v_{(p,q),r} = \begin{cases} \sum_{i=r-1}^{p-1} \binom{p-1}{i} g^{(p-i,0)}(x_1, x_2) v_{(i,0),r-1} & q = 0,\ p \geq 1 \\ \sum_{i=0}^{p} \sum_{j=0}^{q-1} \binom{p}{i} \binom{q-1}{j} g^{(p-i,q-j)}(x_1, x_2) v_{(i,j),r-1} & q \geq 1,\ p \geq 0. \end{cases} \tag{86}$$

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.

References

C. K. Carter and R. Kohn. On Gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

J. C. C. Chan and I. Jeliazkov. Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1:101–120, 2009.

S. Chib, F. Nardari, and N. Shephard. Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108:281–316, 2002.

P. de Jong and N. Shephard. The simulation smoother for time series models. *Biometrika*, 82(1):339–350, 1995.

J. Durbin and S. J. Koopman. A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–615, 2002.

D. Feng, J. J. Jiang, and P. Song. Stochastic conditional durations models with "Leverage Effect" for financial transaction data. *Journal of Financial Econometrics*, 2:390–421, 2004.

S. Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15:183–202, 1994.

S. Frühwirth-Schnatter and H. Wagner. Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. *Biometrika*, 93:827–841, 2006.

S. Frühwirth-Schnatter, R. Früwirth, L. Held, and H. Rue. Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. *Statistics and Computing*, 19:479–492, 2009.

J. Geweke. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, 57:1317–1340, 1989.

J. Geweke. Getting it right: Joint distribution tests of posterior simulators. *Journal of the American Statistical Association*, 99:799–804, 2004.

A. C. Harvey and N. Shephard. The estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics*, 14:429–434, 1996.

E. Jacquier, N. Polson, and P. Rossi. Bayesian analysis of stochastic volatility models with leverage effect and fat tails. *Journal of Econometrics*, pages 185–212, 2004.

B. Jungbacker and S. J. Koopman. Monte Carlo estimation for nonlinear non-Gaussian state space models. *Biometrika*, 94:827–839, 2008.

S. Kim, N. Shephard, and S. Chib. Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65(3):361–393, 1998.

W. J. McCausland. The HESSIAN method: Highly efficient state smoothing, in a nutshell. *Journal of Econometrics*, 2012. Forthcoming.

W. J. McCausland, S. Miller, and D. Pelletier. Simulation smoothing for state-space models: A computational efficiency analysis. *Computational Statistics and Data Analysis*, 55:199–212, 2011.

J. Nakajima and Y. Omori. Leverage, heavy-tails and correlated jumps in stochastic volatility models. *Computational Statistics and Data Analysis*, 53:2335–2353, 2009.

Y. Omori and T. Watanabe. Block sampler and posterior mode estimation for asymmetric stochastic volatility models. *Computational Statistics and Data Analysis*, 52:2892–2910, 2008.

Y. Omori, S. Chib, N. Shephard, and J. Nakajima. Stochastic volatility with leverage: fast and efficient likelihood inference. *Journal of Econometrics*, 140:425–449, 2007.

J. Richard and W. Zhang. Efficient high-dimensional importance sampling. *Journal of Econometrics*, 141:1385–1411, 2007.

H. Rue. Fast sampling of Gaussian Markov random fields with applications. *Journal of the Royal Statistical Society Series B*, 63:325–338, 2001.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B*, 71:319–392, 2009.

T. H. Savits. Some statistical applications of Faa di Bruno. *Journal of Multivariate Analysis*, 97(10):2131 – 2140, 2006.

N. Shephard and M. K. Pitt. Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, 84(3):653–667, 1997.

C. M. Strickland, C. S. Forbes, and G. M. Martin. Bayesian analysis of the stochastic conditional duration model. *Computational Statistics and Data Analysis*, 50:2247–2267, 2006.

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

T. Watanabe and Y. Omori. A multi-move sampler for estimating non-Gaussian time series models: Comments on Shephard and Pitt (1997). *Biometrika*, 91:246–248, 2004.

J. Yu. On leverage in a stochastic volatility model. *Journal of Econometrics*, 127: 165–178, 2005.

Département de Sciences Économiques, Université de Montréal

Département de Sciences Économiques, Université de Montréal

Notation	Description
$\psi_t(\alpha_t, \alpha_{t+1})$	$\log f(y_t\|\alpha_t, \alpha_{t+1})$
$\psi_t^{p,q}(\alpha_t, \alpha_{t+1})$	derivative of $\psi_t^{p,q}(\alpha_t, \alpha_{t+1})$ with respect to α_t and α_{t+1} of orders p and q.
$\psi_n(\alpha_n)$	$\log f(y_n\|\alpha_n)$
$\psi_n^p(\alpha_n)$	p'th 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	$\mathrm{Var}(\alpha_t\|\alpha_{t+1}, y)$ for the 1st reference distribution
$(a_{1\|t+1}(\alpha_{t+1}), \ldots, a_{t\|t+1}(\alpha_{t+1}))$	mode of the conditional density $f(\alpha_1, \ldots, \alpha_t\|\alpha_{t+1}, y)$
$\Sigma_{t\|t+1}(\alpha_{t+1})$	$\mathrm{Var}(\alpha_t\|\alpha_{t+1}, y)$ for the 2nd reference distribution
$A_{t\|t+1}(\alpha_{t+1})$	polynomial approximation of $a_{t\|t+1}(\alpha_{t+1})$
$s_{t\|t+1}(\alpha_{t+1})$	$\log \Sigma_{t\|t+1}(\alpha_{t+1})$
$a_t^{(r)}$, $r = 1, \ldots, R$	r'th derivative of $a_{t\|t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$
$s_t^{(r)}$, $r = 1, \ldots, R-1$	r'th derivatives of $s_{t\|t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$.
$b_{t\|t+1}(\alpha_{t+1})$	mode of the conditional density $f(\alpha_t\|\alpha_{t+1}, y)$
$b_t, b_t^{(r)}, r = 1, \ldots, R$	value and derivatives of $b_{t\|t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$
b_n	mode of the conditional density $f(\alpha_n\|y)$
$B_{t\|t+1}(\alpha_{t+1})$	polynomial approximation of $b_{t\|t+1}(\alpha_{t+1})$
$B_t, B_t^{(r)}, r = 1, \ldots, R$	value and derivatives of $B_{t\|t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$
$\mu_{t\|t+1}(\alpha_{t+1})$	$E[\alpha_t\|\alpha_{t+1}, y]$
$\mu_t, \mu_t^{(r)}, r = 1, 2$	value and two derivatives 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 $\mu_{t\|t+1}(\alpha_{t+1})$
$M_t, M_t^{(r)}$, $r = 1, 2$	value and two derivatives $M_{t\|t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$
$h_t(\alpha_t; \alpha_{t+1})$	first derivative of $\log f(\alpha_t\|\alpha_{t+1}, y)$ with respect to α_t
$H_t^{(p)}(\alpha_t; \alpha_{t+1})$, $p \geq 1$	approximation of $h_t^{(p)}(\alpha_t; \alpha_{t+1})$, p'th derivatives of $h_t(\alpha_t; \alpha_{t+1})$ with respect to α_t
$h_t(\alpha_n)$	first derivative of $\log f(\alpha_n\|y)$ with respect to α_n
$H_n^{(p)}(\alpha_n)$, $p \geq 1$	approximation of the p'th derivatives of $h_t(\alpha_n)$ with respect to α_n

TABLE 1. Main notations used in the paper

	SP500	TOPIX
HIS	70	43
HIM	85	45
OCSN	108	67

TABLE 2. Computational time in seconds by dataset and estimation procedure for the ASV-Gaussian model. For all procedures, we draw a chain of size 12800. The S&P500 dataset is of size 2022 and the Topix dataset is of size 1232.

Parameters	Mean	Std	NSE	RNE
		S& P500		
$\bar{\alpha}$: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

TABLE 3. ASV-Gaussian parameters estimation using the HESSIAN method and the OCSN's procedure on S&P500 and TOPIX.

Parameters	Mean	Std	NSE	RNE
		S& P 500		
$\bar{\alpha}$: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

TABLE 4. ASV-Student model parameters estimation using the HESSIAN with Independent Metropolis-Hastings and Importance Sampling on artificial data, S&P500 and TOPIX.