Journal of Econometrics 168 (2012) 189–206

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

The HESSIAN method: Highly efficient simulation smoothing, in a nutshell

William J. McCausland *

Département de sciences économiques, CIREQ, CIRANO, C.P. 6128, succursale Centre-ville, Montréal QC H3C 3J7, Canada

ARTICLE INFO

Article history:
Received 14 March 2008
Received in revised form
11 August 2011
Accepted 14 December 2011
Available online 2 January 2012

Keywords:
State space models
Simulation smoothing
Importance sampling
MCMC
Stochastic volatility
Count models
Duration models

ABSTRACT

I introduce the HESSIAN (highly efficient simulation smoothing in a nutshell) method for numerically efficient simulation smoothing in state space models with univariate states. Given a vector θ of parameters, the vector of states $\alpha = (\alpha_1, \ldots, \alpha_n)$ is Gaussian and the observed vector $y = ({y_1}^\top, \ldots, {y_n}^\top)^\top$ need not be. I describe a procedure to construct a close approximation $q(\alpha|\theta, y)$ to the target density $p(\alpha|\theta, y)$. It requires code to compute five derivatives of $\log p(y_t|\theta, \alpha_t)$ with respect to α_t, $t = 1, \ldots, n$, and is not otherwise model specific. Since $q(\alpha|\theta, y)$ is proper, fully normalised and simulable, it can be used as an importance density for importance sampling (IS) or as a proposal density for Markov chain Monte Carlo (MCMC). HESSIAN is an acronym but it also refers to the (sparse) Hessian matrix of $\log p(\alpha|\theta, y)$ with respect to α—the HESSIAN method is based on sparse matrix operations rather than the Kalman filter. I construct $q(\alpha|\theta, y)$ and a related approximation $q(\theta, \alpha|y)$ of $p(\theta, \alpha|y)$ for two stochastic volatility models, two stochastic count models and a stochastic duration model. I illustrate their use for numerical approximation of likelihood function values and marginal likelihoods, using IS, and for posterior inference, using IS and MCMC. Compared with other simulation smoothing methods, the HESSIAN method is highly numerically efficient. In an IS application featuring a Student's t stochastic volatility model and $n = 8851$ daily log returns, the efficiency of IS for numerical approximation of the elements of the posterior mean $E[\theta|y]$ is between 80% and 100%.

1. Introduction

I introduce new methods for inference in state space models of the form

$$\alpha_1 = d_0 + u_0, \qquad \alpha_{t+1} = d_t + \phi_t \alpha_t + u_t, \quad t = 1, \ldots, n, \tag{1}$$

$$p(y_1, \ldots, y_n|\alpha_1, \ldots, \alpha_n) = \prod_{t=1}^{n} p(y_t|\alpha_t), \tag{2}$$

where the α_t are univariate latent states, the u_t are independent Gaussian random variables with mean 0 and variance σ_t^2, the y_t are observable random scalars or vectors, and the $p(y_t|\alpha_t)$ are probability density or mass functions. I define $\alpha \doteq (\alpha_1, \ldots, \alpha_n)^\top$ and $y \doteq (y_1^\top, \ldots, y_n^\top)^\top$. Until further notice, I condition on the coefficients d_t and ϕ_t, the variances σ_t^2, and any parameters on which the $p(y_t|\alpha_t)$ might depend, and suppress the notation for this conditioning.

Inference in state space models often proceeds through simulation smoothing, the simulation of the conditional distribution of α given y. I will call this distribution the target distribution.

The importance of the HESSIAN (highly efficient simulation smoothing in a nutshell) method lies not only in its high numerical efficiency for simulation smoothing. It also leads to highly efficient joint simulation of states and parameters, for both IS and MCMC, demonstrated here for several difficult empirical problems, even compared with model specific methods.

The only model specific elements required to implement the HESSIAN method are routines to evaluate $\log p(y_t|\alpha_t)$ and its first five derivatives with respect to α_t at a given point. For the five state space models considered in this paper, it is reasonably easy to compute analytic derivatives in closed form. These are tabulated in Appendix I. It may be more difficult to do this for other models, but there are two reasons for optimism. First, it is possible to compute exact derivatives at a point without closed form analytic expressions. For example, we can use automatic routines to combine derivatives of primitive functions according to Leibniz's rule (for multiple derivatives of products) and Faà di Bruno's rule (for multiple derivatives of composite functions). These rules can be nested. A second reason for optimism is that one can use numerical derivatives or other approximations. There is a cost in numerical efficiency, but as long as the evaluation of $\log p(y_t|\alpha_t)$ is exact, simulation consistency is not compromised. Code for implementing the HESSIAN method is available on request.

Relationship with other methods. In the special case where the distributions $y_t|\alpha_t$ are linear and Gaussian, $\alpha|y$ is multivariate Gaussian and there are efficient methods to evaluate the likelihood function and draw α given y. Carter and Kohn (1994), Frühwirth-Schnatter (1994), de Jong and Shephard (1995) and Durbin and Koopman (2002) offer methods based on the Kalman filter. Rue (2001), Chan and Jeliazkov (2009) and McCausland et al. (2011)

* Tel.: +1 514 343 7281; fax: +1 514 343 7221.
E-mail address: william.j.mccausland@umontreal.ca.
URL: http://web.mac.com/mccauslw/Public_Web_Site/Home_Accueil.html.

0304-4076/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2011.12.003

describe methods based on the sparse Hessian matrix of the log target density.

When they are not all linear and Gaussian, the target distribution is no longer multivariate Gaussian, which makes simulation smoothing much more difficult. I identify three broad approaches to this problem. The direct approach is to use a Metropolis–Hastings update for blocks $(\alpha_t, \alpha_{t+1}, \ldots, \alpha_{t+T-1})$ of length T, where $1 \leq T \leq n$. The auxiliary mixture model approach involves transforming the model into a linear Gaussian model, approximating any non-Gaussian distributions in the transformed model by finite Gaussian mixtures. A third approach is sequential Monte Carlo, which includes particle filtering.

The HESSIAN method falls squarely in the category of direct approaches, and is based on operations on the Hessian matrix of $\log p(\alpha|y)$ rather than the Kalman filter. It updates the entire state sequence α as a single block ($T = n$) using a non-Gaussian proposal density $q(\alpha|y)$. The proposal density is a very close approximation of the target density—when used as a independence Metropolis–Hastings proposal density for a Student's t SV model and a data set with $n = 8851$ observations, the acceptance probability was 0.9985.

I now review in some detail the three approaches to simulation smoothing in non-Gaussian or non-linear models, starting with the direct approach. Jacquier et al. (1994) introduce a posterior simulator with $T = 1$ for a Gaussian SV model. An advantage of such "single-move" simulators is that it is relatively easy to choose a suitable non-Gaussian proposal distribution to closely fit the univariate target distribution. The main disadvantage is low numerical efficiency due to the high posterior autocorrelation of log volatility.

Shephard and Pitt (1997), Watanabe and Omori (2004), Strickland et al. (2006), Jungbacker and Koopman (2008) and Omori and Watanabe (2008) use block lengths with $1 < T < n$ and multivariate Gaussian proposal distributions. Optimal "block" simulators are typically considerably more efficient than single-move simulators. For small enough T, putting highly correlated random variables in the same block more than compensates for the inflexibility of the multivariate Gaussian distribution. Omori and Watanabe (2008) analyse numerical efficiency for a Gaussian SV model with asymmetric volatility and artificial data. They compare a random block sampler with a single move sampler similar to that of Jacquier et al. (2004), but adapted to a slightly different model. For computing posterior means of parameters, the former is about 2.7–15.1 times more efficient.

Shephard and Pitt (1997) and Durbin and Koopman (1997) consider a multivariate Gaussian approximation of the full target distribution. Its mean is the mode of the target distribution and its variance is the negative inverse Hessian of the log target density at the mode. Shephard and Pitt (1997) ultimately reject this approximation as perhaps "too ambitious" for SV models, citing rapid deterioration in the numerical efficiency of block simulators as T increases. Durbin and Koopman (1997) use the multivariate Gaussian approximation for IS. While numerical efficiency is low, they are able to increase it using antithetic sampling and control variates. However, their empirical application involves only $n = 192$ observations, much less than the numbers of observations found in SV applications. See also Gamerman (1998) and Pitt (2000) on the difficulties of using a multivariate Gaussian approximation as a proposal or importance distribution.

Feasibility and numerical efficiency depend very much on how close the approximation $q(\alpha|y)$ is to the target density $p(\alpha|y)$, and there exist closer approximations than the multivariate Gaussian. Richard and Zhang (2007) describe one, which they use for Efficient Importance Sampling (EIS). Richard and Zhang (2007) and Liesenfeld and Richard (2003, 2006) consider applications of EIS to SV and other models. For state space models, the analyst supplies an *auxiliary parametric importance sampler* based on parametric approximations to the conditional densities $p(\alpha_t|\alpha_{t+1}, y)$. The state space model is not required to be univariate or Gaussian, but the parametric approximations have to be well tailored to the target distribution for tolerable numerical precision with large data sets. Parameters of the approximation are determined using an iterative procedure which depends on the Common Random Numbers (CRNs) used to draw from the approximate distribution. The approximate density and the draws are thus jointly random, and EIS is not, strictly speaking, importance sampling. As Richard and Zhang (2007) acknowledge, EIS estimators of integrals with respect to the target distribution, such as values of the likelihood function, may be biased and the usual formula for the variance of IS estimators of integrals does not apply. To estimate standard errors, they use replications of sample means, with CRNs that are independent across replications.

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

The second broad approach to simulation smoothing in non-linear and non-Gaussian state space models is auxiliary mixture sampling. Using transformations, data augmentation and the approximation of non-Gaussian error distributions by finite Gaussian mixtures, one can sometimes obtain an *auxiliary mixture model* with the property that conditioning on auxiliary variables, including latent variables used as mixture component indicators, yields a Gaussian linear state space model. Using Gibbs sampling and standard methods for Gaussian linear state space models, one can simulate the joint distribution of parameters, states and auxiliary variables. Kim et al. (1998), Chib et al. (2002) and Omori et al. (2007) use auxiliary mixture sampling for various SV models. Stroud et al. (2003) use it for Gaussian, but non-linear, state space models with state dependent variances; Frühwirth-Schnatter and Wagner (2006) for state space models with Poisson counts; and Frühwirth-Schnatter and Frühwirth (2007) for logit and multinomial logit models. Frühwirth-Schnatter et al. (2009) offer efficiency improvements for some of these models.

Drawing all states as a block is numerically efficient because states tend to have high posterior autocorrelation. Combining states and parameters in a compound block can lead to further numerical efficiency improvements if the parameters and states have strong posterior dependence. Kim et al. (1998) obtain considerably higher numerical efficiency for a basic Gaussian SV model when they draw parameters and states together in an approximate transformed model. Simulation experiments in Omori et al. (2007) (auxiliary mixture sampling) and Omori and Watanabe (2008) (direct approach, with $1 < T < n$) show something similar for an SV model with asymmetric volatility. Since this is true despite the fact that Kim et al. (1998) and Omori et al. (2007) use data augmentation in order to draw parameters and the full state vector as a block suggests that the reported efficiency improvements understate the advantage of drawing states and parameters as a block.

The approximation of distributions by finite Gaussian mixtures introduces error, which, if not compensated, compromises the simulation consistency of MCMC sample means. Simulation consistency can often be obtained through IS reweighting, as in Kim et al. (1998) and Omori et al. (2007), or through a Metropolis–Hastings accept/reject, as in Stroud et al. (2003), Frühwirth-Schnatter and Wagner (2006) and Frühwirth-Schnatter and Frühwirth (2007). The cost is typically a reduction in numerical precision and simplicity.

The third broad approach to simulation smoothing relies on sequential Monte Carlo methods. Doucet and Johansen (2010) review the literature on particle filtering and smoothing. They

survey three approaches to particle smoothing and describe how to approximate likelihood values using the bi-products of particle filtering. These techniques are computationally demanding, but of course they deliver more, including numerical approximations of filtering distributions.

Rue et al. (2009) is also relevant to the problem of simulation smoothing. They approximate marginal densities of latent variables for latent Gaussian Markov models, including state space models with Gaussian states. The computational cost is $O(n^2)$ for the special case of state space models, but since there is no simulation whatsoever, computations are very rapid. The paper uses a Laplace approximation of the marginal posterior density $p(\theta|y)$, where θ is a vector of parameters. The approximation of $p(\theta|y)$ is proportional to $p(\theta)L_G(\theta; y)$, where $L_G(\theta; y)$ is the Laplace approximation, defined in Section 3, of the likelihood function $L(\theta; y)$. We will see that for stochastic volatility models, the approximation error can be considerable and depend on θ. The approach of Rue et al. (2009) does not rely on any particular approximation, however, and Section 4 shows that a simulation consistent stochastic alternative $\hat{L}_H(\theta; y)$, based on the HESSIAN approximation $q(\alpha|\theta, y)$, has very high numerical precision for the example considered. Since the cost of computing $\hat{L}_H(\theta; y)$ is $O(n)$, this would not be particularly onerous. Importantly, $\hat{L}_H(\theta; y)$ has attractive and better understood properties. Using $\hat{L}_H(\theta; y)$ instead of $L(\theta; y)$ would allay some of the concerns about Laplace approximations put forward by Robert (2009) in his discussion of Rue et al. (2009), at least for the special case of state space models.

Outline. I describe the construction of $q(\alpha|y)$ in Section 2. Since one can evaluate and sample $q(\alpha|y)$ exactly, standard results on simulation error for IS and MCMC hold. I also show that the ratio $p(\alpha|y)/q(\alpha|y)$ is bounded. In the context of IS, this guarantees the boundedness of importance weights, an important condition for the existence of the variance of importance sample means: see Geweke (1989). In the context of independence Metropolis–Hastings Markov chains, this guarantees the geometric ergodicity of the Markov chain: see Roberts and Rosenthal (1998).

The term "HESSIAN method" refers to the use of the approximation $q(\alpha|y)$ for simulation smoothing and numerical integration. Section 3 shows how to do this. We are not limited to using $q(\alpha|y)$ directly as an importance density or a proposal density. We can combine it with other densities to obtain joint importance and proposal densities for α and other unknown variables such as parameters. Much of the efficiency advantage of the HESSIAN method lies in doing this.

In Section 4, I discuss the results of simulation experiments and empirical applications using data. In the simulation experiments, I subject my computer code to stringent tests of correctness of the kind described in Geweke (2004). I use the empirical applications to compare the HESSIAN method with competing methods, sometimes using the same data that were used to illustrate those methods. I illustrate the high numerical efficiency that can be obtained using the HESSIAN method by comparing the numerical precision it achieves with that of other methods in the literature. In Section 5, I conclude and mention some possible extensions.

2. The HESSIAN approximation $q(\alpha|y)$

Here I show how to construct the HESSIAN approximation $q(\alpha|y)$ of the target density $p(\alpha|y)$ for a given state space model. I describe how to evaluate and sample $q(\alpha|y)$ exactly using $O(n)$ operations. I establish that the density $q(\alpha|y)$ is proper and fully normalised, and that the ratio $p(\alpha|y)/q(\alpha|y)$ is bounded.

I take as input a state space model of the form in (1) and (2). It is convenient to express the state dynamics concisely in terms of the precision (inverse of variance) $\bar{\Omega}$ and covector (precision times mean) $\bar{c}$ of α. Appendix A shows how to compute the $O(n)$ nonzero elements of $\bar{\Omega}$ and $\bar{c}$ as a function of the values d_t, ϕ_t and σ_t^2 in (1). As for the specification of (2), I require routines to compute the value and first five derivatives of $\psi_t(\alpha_t) \doteq \log p(y_t|\alpha_t)$ with respect to α_t. Appendix I gives analytic expressions for five different state space models.

Notation introduced in this section is summarised in Table 1.

2.1. Overview of the approximation

Like the target density $p(\alpha|y)$, $q(\alpha|y)$ has the Markov property, which allows us to factor it as

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

To draw from or evaluate $q(\alpha|y)$, we draw from or evaluate each factor, beginning with $q(\alpha_n|y)$ and ending with $q(\alpha_1|\alpha_2, y)$. At each iteration $t < n$, the value of α_{t+1} is available.

Approximations rely heavily on Taylor series expansions, some exact and some approximate, of various functions. Most of these expansions are computed during a forward pass, around a static point of expansion: for functions of α_{t+1}, the point of expansion is a_{t+1}, where $(a_1, \ldots, a_n)$ is the mode of the target distribution. Examples of these functions are $b_{t|t+1}(\alpha_{t+1})$ and $\mu_{t|t+1}(\alpha_{t+1})$, which denote the mode and mean of α_t given α_{t+1} and y.

During the backward pass, we compute an approximate Taylor series expansion of $h_n(\alpha_n) \doteq \log p(\alpha_n|y)$ and each $h_t(\alpha_t; \alpha_{t+1}) \doteq \log p(\alpha_t|\alpha_{t+1}, y)$. The notation for $h_t(\alpha_t; \alpha_{t+1})$ indicates that we treat it as a univariate function of α_t with parameter α_{t+1}. For $h_t(\alpha_t; \alpha_{t+1})$, the point of expansion is a moving target: it depends on α_{t+1}. The expansion is fifth order, which allows us to go well beyond a quadratic (i.e. Gaussian) approximation.

I use a parametric family of perturbed Gaussian densities $p_{\text{pert}}(\cdot; \vartheta)$, described in Appendix G, for the $q(\alpha_t|\alpha_{t+1}, y)$ and $q(\alpha_n|y)$. The parameter vector ϑ includes b, which gives the mode of $p_{\text{pert}}(\cdot; \vartheta)$, and h_2, h_3, h_4 and h_5, which give the second through fifth derivatives of $\log p_{\text{pert}}(\cdot; \vartheta)$ at b. The family $p_{\text{pert}}(\cdot; \vartheta)$ is particularly suitable: if we have a fifth order Taylor series expansion of some log density around its mode, we can choose ϑ so that the fifth order Taylor series expansion of $\log p_{\text{pert}}(\cdot; \vartheta)$ matches exactly.

Constructing $q(\alpha_n)$ and the $q(\alpha_t|\alpha_{t+1}, y)$ amounts to choosing appropriate values of ϑ. The mode of $q(\alpha_n|y)$ approximates the mode of $p(\alpha_n|y)$, denoted b_n, and the mode of $q(\alpha_t|\alpha_{t+1}, y)$ approximates the mode of $p(\alpha_t|\alpha_{t+1}, y)$, denoted $b_{t|t+1}(\alpha_{t+1})$. The derivative parameters are set to approximations of the second through fifth derivatives of $h_n(\alpha_n)$ and $h_t(\alpha_t; \alpha_{t+1})$.

The following key result, proved in Appendix C, gives exact expressions for the first derivatives of the $h_t(\alpha_t; \alpha_{t+1})$ and $h_n(\alpha_n)$:

$$h_t'(\alpha_t; \alpha_{t+1}) = \begin{cases} -\bar{\Omega}_{11}\alpha_1 - \bar{\Omega}_{12}\alpha_2 + \bar{c}_1 + \psi_1'(\alpha_1), & t = 1, \\ -\bar{\Omega}_{t-1,t}\mu_{t-1|t}(\alpha_t) - \bar{\Omega}_{tt}\alpha_t - \bar{\Omega}_{t,t+1}\alpha_{t+1} & \\ \quad + \bar{c}_t + \psi_t'(\alpha_t), \quad 1 < t < n, & \end{cases} \tag{3}$$

$$h_n'(\alpha_n) = -\bar{\Omega}_{n-1,n}\mu_{n-1|n}(\alpha_n) - \bar{\Omega}_{nn}\alpha_n + \bar{c}_n + \psi_n'(\alpha_n), \tag{4}$$

where $\mu_{t-1|t}(\alpha_t) \doteq E[\alpha_{t-1}|\alpha_t, y]$. One can compute the value and four derivatives of everything in these expressions except $\mu_{t-1|t}(\alpha_t)$, $t = 2, \ldots, n$.

If we knew b_n and the functions $b_{t|t+1}(\alpha_{t+1})$ and $\mu_{t-1|t}(\alpha_t)$, we could compute exact fifth order Taylor series expansions of $h_t(\alpha_t; \alpha_{t+1})$ and $h_n(\alpha_n)$ at $b_{t|t+1}(\alpha_{t+1})$ and b_n. Unfortunately, we do not. Instead, I will approximate the $b_{t|t+1}(\alpha_{t+1})$ and $\mu_{t|t+1}(\alpha_{t+1})$ by polynomials $B_{t|t+1}(\alpha_{t+1})$ and $M_{t|t+1}(\alpha_{t+1})$. These polynomials are approximate Taylor series expansions of $b_{t|t+1}(\alpha_{t+1})$ and $\mu_{t|t+1}(\alpha_{t+1})$ around a_{t+1}.

Table 1
Table of definitions.

Notation	Description
$a = (a_1, \ldots, a_n)$	Conditional mode of α given y
Σ_t	$\mathrm{Var}[\alpha_t \vert \alpha_{t+1}, y]$ for 1st reference distribution. See (7)
$(a_{1\vert t+1}(\alpha_{t+1}), \ldots, a_{t\vert t+1}(\alpha_{t+1}))$	Conditional mode of $\alpha_1, \ldots, \alpha_t$ given α_{t+1} and y
$\dot{a}_t, \ddot{a}_t, \dddot{a}_t, \ddddot{a}_t$	Four derivatives of $a_{t\vert t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$
$\Sigma_{t\vert t+1}(\alpha_{t+1})$	$\mathrm{Var}[\alpha_t \vert \alpha_{t+1}, y]$ for 2nd reference distribution. See (8)
$s_{t\vert t+1}(\alpha_{t+1})$	$\log \Sigma_{t\vert t+1}(\alpha_{t+1})$
$s_t, \dot{s}_t, \ddot{s}_t, \dddot{s}_t$	Value and three derivatives of $s_{t\vert t+1}(\alpha_{t+1})$ at a_{t+1}
$b_{t\vert t+1}(\alpha_{t+1})$	Conditional mode of α_t given α_{t+1} and y
$b_t, \dot{b}_t, \ddot{b}_t, \dddot{b}_t, \ddddot{b}_t$	Value and four derivatives of $b_{t\vert t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$
b_n	Conditional mode of α_n given y
$B_{t\vert t+1}(\alpha_{t+1})$	Polynomial approximation of $b_{t\vert t+1}(\alpha_{t+1})$ given by (9)
$B_t, \dot{B}_t, \ddot{B}_t, \dddot{B}_t, \ddddot{B}_t, B_n$	Approximations of $b_t, \dot{b}_t, \ddot{b}_t, \dddot{b}_t, \ddddot{b}_t, b_n$, see Appendix E
$\mu_{t\vert t+1}(\alpha_{t+1})$	$E[\alpha_t \vert \alpha_{t+1}, y]$
$\mu_t, \dot{\mu}_t, \ddot{\mu}_t, \dddot{\mu}_t, \ddddot{\mu}_t$	Value and four derivatives of $\mu_{t\vert t+1}(\alpha_{t+1})$ at a_{t+1}
$M_{t\vert t+1}(\alpha_{t+1})$	Polynomial approximation of $\mu_{t\vert t+1}(\alpha_{t+1})$ given by (10)
$M_t, \dot{M}_t, \ddot{M}_t, \dddot{M}_t, \ddddot{M}_t$	Approximations of $\mu_t, \dot{\mu}_t, \ddot{\mu}_t, \dddot{\mu}_t, \ddddot{\mu}_t$, see Appendix F
$\psi_t(\alpha_t)$	$\log p(y_t \vert \alpha_t)$
$h_t(\alpha_t; \alpha_{t+1}), h'_t(\alpha_t; \alpha_{t+1})$	$\log p(\alpha_t \vert \alpha_{t+1}, y)$ and first derivative with respect to α_t
$h_t^{(j)}(\alpha_t), j \geq 2$	j'th derivative of $h_t(\alpha_t; \alpha_{t+1})$ with respect to α_t
$h_n(\alpha_n)$	$\log p(\alpha_n \vert y)$
$h_n^{(j)}(\alpha_n), j \geq 1$	j'th derivative of $h_n(\alpha_n)$ with respect to α_n
$H'_t(\alpha_t; \alpha_{t+1})$	Approximation of $h'_t(\alpha_t; \alpha_{t+1})$ given by (5)
$H'_n(\alpha_n)$	Approximation of $h'_n(\alpha_n)$ given by (6)
$H_t^{(j)}(\alpha_t), j \geq 2$	Approximation of $h_t^{(j)}(\alpha_t)$ implied by (5) and (6)

The approximation of $\mu_{t-1\vert t}(\alpha_t)$ by $M_{t-1\vert t}(\alpha_t)$ leads to the following approximations of $h_t(\alpha_t; \alpha_{t+1})$ and $h_n(\alpha_n)$:

$$H'_t(\alpha_t; \alpha_{t+1}) \doteq \begin{cases} -\bar{\Omega}_{11}\alpha_1 - \bar{\Omega}_{12}\alpha_2 + \bar{c}_1 + \psi'_1(\alpha_1), & t = 1, \\ -\bar{\Omega}_{t-1,t}M_{t-1\vert t}(\alpha_t) - \bar{\Omega}_{tt}\alpha_t - \bar{\Omega}_{t,t+1}\alpha_{t+1} & \\ \quad + \bar{c}_t + \psi'_t(\alpha_t), & 1 < t < n, \end{cases} \tag{5}$$

$$H'_n(\alpha_n) \doteq -\bar{\Omega}_{n-1,n}M_{n-1\vert n}(\alpha_n) - \bar{\Omega}_{nn}\alpha_n + \bar{c}_n + \psi'_n(\alpha_n). \tag{6}$$

One can compute the value and four derivatives of $H'_n(\alpha_n)$ and the $H'_t(\alpha_t; \alpha_{t+1})$ at any point. Choosing a suitable point of expansion and evaluating the second through fifth derivatives of $H_n(\alpha_n)$ or $H_t(\alpha_t; \alpha_{t+1})$ there give the parameter values required to specify $q(\alpha_n \vert y)$ or $q(\alpha_t \vert \alpha_{t+1}, y)$.

In the remainder of this section I give more detail on how to evaluate $q(\alpha \vert y)$, sample it, or both. There are three main steps. The first step is to find the conditional mode $a = (a_1, \ldots, a_n)$ of α given y, and related quantities. The second step is a forward pass to construct polynomials $B_{t\vert t+1}(\alpha_{t+1})$ and $M_{t\vert t+1}(\alpha_{t+1})$, $t = 1, \ldots, n-1$. The third step is a backward pass. At period $t < n$ of the backward pass, we can evaluate $q(\alpha_t \vert \alpha_{t+1}, y)$, sample it, or both. The next three sections describe the three main steps. Full technical detail is left to various appendices.

2.2. Preliminary computation

The first step is to find the mode $a = (a_1, \ldots, a_n)$ of the target distribution and compute the bi-products $\bar{\bar{\Omega}}$ and $\Sigma_1, \ldots, \Sigma_n$, described below. The elements of a serve as points of expansion for the computations of the forward pass. I use an efficient implementation of the Newton–Raphson algorithm, described in Appendix B, to compute a and the bi-products.

$\bar{\bar{\Omega}}$ is the Hessian matrix of $-\log(\alpha \vert y)$ with respect to α, evaluated at a. Since $p(\alpha \vert y) \propto p(\alpha)p(y \vert \alpha)$, we can write

$$\bar{\bar{\Omega}} = \bar{\Omega} - \frac{\partial^2 \log p(y \vert \alpha)}{\partial \alpha \partial \alpha^\top}.$$

The second term is diagonal, so there are only $O(n)$ elements to store. Computations described in technical appendices rely on operations involving this sparse Hessian matrix.

The variances $\Sigma_1, \ldots, \Sigma_n$ are based on the distribution $N(a, \bar{\bar{\Omega}}^{-1})$, a multivariate Gaussian approximation of the target distribution: both distributions have the same mode a and both log densities have the same Hessian matrix $-\bar{\Omega}$ there. Result 2.1 of McCausland et al. (2011) implies that if $\tilde{\alpha} \sim N(a, \bar{\bar{\Omega}}^{-1})$, then $\tilde{\alpha}_n \sim N(a_n, \Sigma_n)$ and $\tilde{\alpha}_t \vert \tilde{\alpha}_{t+1} \sim N(a_t - \bar{\bar{\Omega}}_{t,t+1}\Sigma_t\tilde{\alpha}_{t+1}, \Sigma_t)$ for $t = n-1, \ldots, 1$, where Σ_t is given by the forward recursion

$$\Sigma_1 \doteq \bar{\bar{\Omega}}_{11}^{-1}, \qquad \Sigma_t \doteq (\bar{\bar{\Omega}}_{tt} - \bar{\bar{\Omega}}_{t,t-1}^2\Sigma_{t-1})^{-1}, \quad t = 2, \ldots, n. \tag{7}$$

2.3. A forward pass

The purpose of the forward pass is to construct polynomial approximations $B_{t\vert t+1}(\alpha_{t+1})$ and $M_{t\vert t+1}(\alpha_{t+1})$ of $b_{t\vert t+1}(\alpha_{t+1})$ and $\mu_{t\vert t+1}(\alpha_{t+1})$. A very useful intermediate step is to construct a polynomial approximation of the function $a_{t\vert t+1}(\alpha_{t+1})$, where $(a_{1\vert t+1}(\alpha_{t+1}), \ldots, a_{t\vert t+1}(\alpha_{t+1}))$ is the conditional mode of $\alpha_1, \ldots, \alpha_t$ given α_{t+1} and y. Then direct approximations of $b_{t\vert t+1}(\alpha_{t+1}) - a_{t\vert t+1}(\alpha_{t+1})$ and $\mu_{t\vert t+1}(\alpha_{t+1}) - b_{t\vert t+1}(\alpha_{t+1})$ give indirect approximations of $b_{t\vert t+1}(\alpha_{t+1})$ and $\mu_{t\vert t+1}(\alpha_{t+1})$.

The main advantage of passing through $a_{t\vert t+1}(\alpha_{t+1})$ is that we can compute exact derivatives of $a_{t\vert t+1}(\alpha_{t+1})$ at a_{t+1} and therefore an exact Taylor series expansion of $a_{t\vert t+1}(\alpha_{t+1})$ around a_{t+1}. Not surprisingly, $a_{t\vert t+1}(\alpha_{t+1})$ is typically much more sensitive to α_{t+1} than are $b_{t\vert t+1}(\alpha_{t+1}) - a_{t\vert t+1}(\alpha_{t+1})$ and $\mu_{t\vert t+1}(\alpha_{t+1}) - b_{t\vert t+1}(\alpha_{t+1})$. In this sense, the most important terms of the expansions of $b_{t\vert t+1}(\alpha_{t+1})$ and $\mu_{t\vert t+1}(\alpha_{t+1})$ are exact.

As another intermediate step, I compute exact derivatives of $s_{t\vert t+1}(\alpha_{t+1}) \doteq \log \Sigma_{t\vert t+1}(\alpha_{t+1})$ at α_{t+1}, where $\Sigma_{t\vert t+1}(\alpha_{t+1})$ is the final value in the recursion

$$\Sigma_{1\vert t+1} \doteq \bar{\bar{\Omega}}_{11}^{-1}, \qquad \Sigma_{\tau\vert t+1} \doteq (\bar{\bar{\Omega}}_{\tau\tau} - \bar{\bar{\Omega}}_{\tau,\tau-1}^2\Sigma_{\tau-1\vert t+1})^{-1}, \quad \tau = 2, \ldots, t. \tag{8}$$

Here, $\bar{\bar{\Omega}}_{1:t\vert t+1}$ is the Hessian of $-\log p(\alpha_1, \ldots, \alpha_t \vert \alpha_{t+1}, y)$ with respect to $(\alpha_1, \ldots, \alpha_t)$ at $(a_{1\vert t+1}(\alpha_{t+1}), \ldots, a_{t\vert t+1}(\alpha_{t+1}))$. Result 2.1 of McCausland et al. (2011) implies that if $\tilde{\alpha} \vert \alpha_{t+1} \sim N((a_{1\vert t+1}(\alpha_{t+1}), \ldots, a_{t\vert t+1}(\alpha_{t+1})), \bar{\bar{\Omega}}_{1:t\vert t+1}^{-1})$, then $\tilde{\alpha}_t \vert \alpha_{t+1} \sim N(a_{t\vert t+1}(\alpha_{t+1}), \Sigma_{t\vert t+1}(\alpha_{t+1}))$.

In the forward pass, I compute the coefficients of polynomial approximations of the functions $a_{t|t+1}(\alpha_{t+1})$, $s_{t|t+1}(\alpha_{t+1})$, $b_{t|t+1}(\alpha_{t+1})$ and $\mu_{t|t+1}(\alpha_{t+1})$. I perform these steps for $t = 1, \ldots, n-1$:

(1) Compute the following derivatives of $a_{t|t+1}(\alpha_{t+1})$ and $s_{t|t+1}(\alpha_{t+1})$ at a_{t+1}:

$$\dot{a}_t \doteq a'_{t|t+1}(a_{t+1}), \qquad \ddot{a}_t \doteq a''_{t|t+1}(a_{t+1}),$$
$$\dddot{a}_t \doteq a'''_{t|t+1}(a_{t+1}), \qquad \ddddot{a}_t \doteq a^{(4)}_{t|t+1}(a_{t+1}),$$
$$\dot{s}_t \doteq s'_{t|t+1}(a_{t+1}), \qquad \ddot{s}_t \doteq s''_{t|t+1}(a_{t+1}),$$
$$\dddot{s}_t \doteq s'''_{t|t+1}(a_{t+1}).$$

Appendix D shows how to compute these exactly in terms of $\dot{a}_{t-1}$ through $\dddot{a}_{t-1}$ and $\psi'_t(a_t)$ through $\psi^{(5)}_t(a_t)$. Eqs. (41) and (44)–(46) give expressions for $\dot{a}_t$, $\ddot{a}_t$, $\dddot{a}_t$ and $\ddddot{a}_t$. Eqs. (48)–(50) give expressions for $\dot{s}_t$, $\ddot{s}_t$ and $\dddot{s}_t$.

(2) Compute approximations B_t, $\dot{B}_t$, $\ddot{B}_t$, $\dddot{B}_t$ and $\ddddot{B}_t$ to the value and first four derivatives of $b_{t|t+1}(\alpha_{t+1})$ at a_{t+1}, and an approximation B_n to the value of b_n. Appendix E defines these approximations and shows how to compute them in terms of $\dot{a}_t$, $\ddot{a}_t$, $\dddot{a}_t$, $\ddddot{a}_t$, $\dot{s}_t$, $\ddot{s}_t$, $\dddot{s}_t$, M_{t-1}, $\dot{M}_{t-1}$ and $\ddot{M}_{t-1}$. While M_{t-1}, $\dot{M}_{t-1}$ and $\ddot{M}_{t-1}$ are computed at step 3 below, they are lagged values, available at iteration t. Eqs. (52)–(57) give expressions for B_t, $\dot{B}_t$, $\ddot{B}_t$, $\dddot{B}_t$, $\ddddot{B}_t$ and B_n.

(3) Compute approximations M_t, $\dot{M}_t$, $\ddot{M}_t$, $\dddot{M}_t$ and $\ddddot{M}_t$ to the value and first two derivatives of $\mu_{t|t+1}(\alpha_{t+1})$ at a_{t+1}. Appendix F defines these approximations and shows how to compute them in terms of B_t, $\dot{B}_t$, $\ddot{B}_t$, $\dddot{B}_t$ and $\ddddot{B}_t$. Eqs. (59)–(62) give expressions for M_t, $\dot{M}_t$, $\ddot{M}_t$, $\dddot{M}_t$ and $\ddddot{M}_t$.

The polynomial approximations $B_{t|t+1}(\alpha_{t+1})$ and $M_{t|t+1}(\alpha_{t+1})$ of $b_{t|t+1}(\alpha_{t+1})$ and $\mu_{t|t+1}(\alpha_{t+1})$ are approximate Taylor series expansions, defined by

$$B_{t|t+1}(\alpha_{t+1}) \doteq B_t + \dot{B}_t(\alpha_{t+1} - a_{t+1}) + \frac{1}{2}\ddot{B}_t(\alpha_{t+1} - a_{t+1})^2 + \frac{1}{6}\dddot{B}_t(\alpha_{t+1} - a_{t+1})^3 + \frac{1}{24}\ddddot{B}_t(\alpha_{t+1} - a_{t+1})^4, \quad (9)$$

$$M_{t|t+1}(\alpha_{t+1}) \doteq M_t + \dot{M}_t(\alpha_{t+1} - a_{t+1}) + \frac{1}{2}\ddot{M}_t(\alpha_{t+1} - a_{t+1})^2 + \frac{1}{6}\dddot{M}_t(\alpha_{t+1} - a_{t+1})^3 + \frac{1}{24}\ddddot{M}_t(\alpha_{t+1} - a_{t+1})^4. \quad (10)$$

2.4. Backward pass

The next step is a backward pass to evaluate $q(\alpha|y)$, draw from it or both. Using the coefficients computed during the forward pass, I construct fully normalised densities $q(\alpha_n|y)$ and $q(\alpha_t|\alpha_{t+1}, y)$ approximating $p(\alpha_n|y)$ and $p(\alpha_t|\alpha_{t+1}, y)$.

We perform the following steps at iteration t, $t = n, n-1, \ldots, 1$. For $t = n$, we take $B_{t|t+1}(\alpha_{t+1})$ to mean B_n. For $t < n$, α_{t+1} is known.

(1) Evaluate $B_{t|t+1}(\alpha_{t+1})$, a first approximation of $b_{t|t+1}(\alpha_{t+1})$, using (9).

(2) Compute $\tilde{B}_t$, a refined approximation of $b_{t|t+1}(\alpha_{t+1})$, using an approximate Newton–Raphson step:

$$\tilde{B}_t \doteq B_{t|t+1}(\alpha_{t+1}) - \frac{H'_t(B_{t|t+1}(\alpha_{t+1}); \alpha_{t+1})}{H''_t(B_{t|t+1}(\alpha_{t+1}))}.$$

(3) Compute $H''_t(\tilde{B}_t)$, $H'''_t(\tilde{B}_t)$, $H^{(4)}_t(\tilde{B}_t)$ and $H^{(5)}_t(\tilde{B}_t)$ using (5).

(4) Set $b = \tilde{B}_t$, $h_2 = H''_t(\tilde{B}_t)$, $h_3 = H'''_t(\tilde{B}_t)$, $h_4 = H^{(4)}_t(\tilde{B}_t)$ and $h_5 = H^{(5)}_t(\tilde{B}_t)$. Set σ^2_{tail} to a value greater than the prior variance $\text{Var}[\alpha_t|\alpha_{t+1}]$. In Section 4, I use $\sigma^2_{\text{tail}} = 1.01\text{Var}[\alpha_t|\alpha_{t+1}]$. Set other parameters of ϑ as recommended in Appendix G.

(5) Draw from and/or evaluate $q(\alpha_t|\alpha_{t+1}, y) \doteq p_{\text{pert}}(\alpha_t|\vartheta)$.

As shown in Appendix G, setting σ^2_{tail} to a value greater than $\text{Var}[\alpha_t|\alpha_{t+1}]$ guarantees that $p(\alpha_t|\alpha_{t+1})/q(\alpha_t|\alpha_{t+1}, y)$ is bounded. This in turn guarantees that the ratio $p(\alpha_t|\alpha_{t+1}, y)/q(\alpha_t|\alpha_{t+1}, y)$ is bounded. Since this is true for all t, $p(\alpha|y)/q(\alpha|y)$ is bounded.

3. Applications of the HESSIAN approximation

Starting in this section, I will be explicit about conditioning on parameters of the state space model, denoted θ. The HESSIAN approximation $q(\alpha|\theta, y)$ can be used as an importance density for IS, a proposal density for MCMC, and for Laplace-like approximations. An important theme of this section is that $q(\alpha|\theta, y)$ can be combined with other densities to obtain joint importance and proposal densities.

3.1. Importance sampling

We can use $q(\alpha|\theta, y)$ as an importance density to compute approximations of the likelihood function $L(\theta; y) = p(y|\theta)$ and other integrals with respect to $p(\alpha|\theta, y)$. If $\alpha^{(1)}, \ldots, \alpha^{(M)}$ is a random sample from the HESSIAN approximation then the sample mean of the importance weights

$$w_m \doteq \frac{p(\alpha^{(m)}|\theta)p(y|\theta, \alpha^{(m)})}{q(\alpha^{(m)}|\theta, y)}, \quad m = 1, \ldots, M,$$

converges to $L(\theta; y)$. Since $p(\alpha|\theta, y)/q(\alpha|\theta, y)$ is bounded, the variance of the weights is finite. If the posterior expectation $E[g(\alpha)|\theta, y]$ of a function of interest g exists, then $\sum_{m=1}^M w_m g(\alpha^{(m)})/\sum_{m=1}^M w_m$ is a consistent estimator of $E[g(\alpha)|\theta, y]$. Geweke (1989) describes how to compute numerical standard errors and numerical efficiency.

The HESSIAN approximation can be combined with an approximation $q(\theta|y)$ of the marginal posterior $p(\theta|y)$ to give an importance density for integration with respect to $p(\theta, \alpha|y)$. For example, we can compute the following approximation of the marginal likelihood $p(y)$ using a random sample $\{\theta^{(m)}, \alpha^{(m)}\}_{m=1}^M$ from $q(\theta|y)q(\alpha|\theta, y)$:

$$\hat{p}(y) \doteq \frac{1}{M}\sum_{m=1}^M \frac{p(\theta^{(m)}, \alpha^{(m)}, y)}{q(\theta^{(m)}|y)q(\alpha^{(m)}|\theta^{(m)}, y)} \to E_q\left[\frac{p(\theta, \alpha, y)}{q(\theta|y)q(\alpha|\theta, y)}\right] = p(y), \quad (11)$$

where E_q denotes expectation with respect to the approximate density $q(\alpha|\theta, y)$. If $p(\theta|y)/q(\theta|y)$ is bounded, then the importance weights have bounded variance and we can estimate the variance $\hat{\sigma}^2_p$ of the estimator $\hat{p}(y)$ as the sample variance of the importance weights $p(\theta^{(m)}, \alpha^{(m)}, y)/[q(\theta^{(m)}|y)q(\alpha^{(m)}|\theta^{(m)}, y)]$ divided by M. Then the delta method approximation of the numerical standard error for $\log \hat{p}(y)$ is $\hat{\sigma}_p/\hat{p}(y)$.

3.2. Markov chain Monte Carlo

We can also use $q(\alpha|\theta, y)$ as a proposal density for an MCMC block updating the conditional distribution of α given θ and y. Drawing α as a single block is more numerically efficient than single-move and block samplers when the posterior autocorrelation of states is high.

If the posterior dependence between θ and α is also high, further gains in numerical efficiency are possible by updating θ and α jointly in a way that preserves the conditional distribution of θ and α given y.

We can do this as a standard Metropolis–Hastings update if we have a proposal distribution $q(\theta^*|\theta, y)$ for θ. The joint proposal is

then $q(\theta^*, \alpha^*|\theta, y) \doteq q(\theta^*|\theta, y)q(\alpha^*|\theta^*, y)$. We first draw θ^* from $q(\theta^*|\theta, y)$, then draw α^* from $q(\alpha^*|\theta^*, y)$, then accept (θ^*, α^*) with probability

$$\pi(\theta^*, \alpha^*, \theta, \alpha) \doteq \min\left[1, \frac{p(\theta^*)p(\alpha^*|\theta^*)p(y|\theta^*, \alpha^*)}{p(\theta)p(\alpha|\theta)p(y|\theta, \alpha)} \cdot \frac{q(\theta|\theta^*, y)q(\alpha|\theta, y)}{q(\theta^*|\theta, y)q(\alpha^*|\theta^*, y)}\right].$$

The boundedness of $p(\alpha|\theta, y)/q(\alpha|\theta, y)$ guarantees the geometric ergodicity of a Metropolis–Hastings chain for the target density $p(\alpha|\theta, y)$ using $q(\alpha|\theta, y)$ as a proposal. If we choose an independence proposal $q(\theta^*|\theta, y) = q(\theta^*|y)$ such that $p(\theta|y)/q(\theta|y)$ is bounded, then the resulting chain is geometrically ergodic.

3.3. Laplace-like approximations

We obtain a Laplace approximation of the likelihood function $L(\theta; y) = p(y|\theta)$ by replacing $p(\alpha|\theta, y)$ in the right hand side of the identity

$$p(y|\theta) \equiv \frac{p(\alpha|\theta)p(y|\theta, \alpha)}{p(\alpha|\theta, y)} \tag{12}$$

with a Gaussian approximation $q_G(\alpha|\theta, y)$ and evaluating the right hand side at the posterior mode a. The mean of $q_G(\alpha|\theta, y)$ is a and the variance is such that $\log q_G(\alpha|\theta, y)$ and $\log p(\alpha|\theta, y)$ have the same Hessian at a.

Replacing $p(\alpha|\theta, y)$ with the HESSIAN approximation $q(\alpha|\theta, y)$ instead gives the approximation

$$L_H(\theta; y) \doteq \left.\frac{p(\alpha|\theta)p(y|\theta, \alpha)}{q(\alpha|\theta, y)}\right|_{\alpha=a}. \tag{13}$$

It is difficult to assess the error of Laplace and similar approximations, except by direct comparison with methods known to work, and I so would advise caution. The results of Section 4 suggest that $L_H(\theta; y)$ is a much better approximation than the Laplace approximation. Since $L_H(\theta; y)$ is easily evaluated, it is useful for computing approximate derivatives of $L(\theta; y)$—we see an example in the next section.

4. Simulations and empirical illustrations

In this section, I illustrate the use of the HESSIAN approximation $q(\alpha|\theta, y)$ for simulation smoothing, both as an importance density for IS and as a proposal density for MCMC. Simulations described in this section run on a Macintosh mini, mid 2010, with an Intel Core 2 Duo 2.4 GHz processor. Source code is written in *C* and is available on request.

4.1. Getting it right

The first illustration is an artificial data experiment intended as a test for the correctness of computer code for the evaluation of $q(\alpha|\theta, y)$ and the simulation of random variates from $q(\alpha|\theta, y)$. The tests described here are similar to tests of posterior simulator correctness in Geweke (2004), and the title of this section comes from the title of that paper.

I illustrate using a basic stochastic volatility model with Student's t innovations. The log volatility state is governed by

$$\alpha_1 = \bar{\alpha} + u_0, \qquad \alpha_{t+1} = (1-\phi)\bar{\alpha} + \phi\,\alpha_t + u_t, \tag{14}$$

where the u_t are independent, with $u_0 \sim N(0, \sigma^2/(1-\phi^2))$ and $u_t \sim N(0, \sigma^2)$. I give the elements of $\bar{\Omega}$ and $\bar{c}$ for this stationary state model in Appendix A. Log asset returns are given by

$$y_t|\theta, \alpha_t \sim t(0, \exp(\alpha_t), \nu), \tag{15}$$

where ν is an unknown degrees of freedom parameter. I give expressions for the $\psi_t(\alpha_t) \doteq \log p(y_t|\alpha_t, \theta)$ and their first five derivatives in Appendix I.

I fixed the parameter values $\mu = -9.0$, $\phi = 0.97$, $\sigma = 0.20$, and $\nu = 12.0$. I chose a number $n = 20$ of observations. I generated an MCMC sample $\{\alpha^{(m)}, y^{(m)}\}_{m=1}^{M}$, with $M = 10^8$, from the conditional distribution of α and y given θ, using a Gibbs sampler with two blocks. The first block updates α, and the independence proposal density is the HESSIAN approximation $q(\alpha|\theta, y)$. The second block updates y by direct simulation from (15). The initial draw comes directly from the conditional distribution of α and y given θ.

Under the hypothesis that the code works correctly, the marginal distribution of each $\alpha^{(m)}$ is $N(\bar{\Omega}^{-1}\bar{c}, \bar{\Omega}^{-1})$. This implies that for all $m = 1, \ldots, M$ and all $q \in (0, 1)$, the following indicators are Bernoulli with probability parameter q:

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

$$I_{t|t-1,q}^{(m)} \doteq 1\left(\frac{\alpha_t^{(m)} - (1-\phi)\bar{\alpha} - \phi\alpha_{t-1}^{(m)}}{\sigma} \le \Phi^{-1}(q)\right), \quad t = 2, \ldots, n,$$

where Φ is the cumulative distribution function of the standard Gaussian.

I used 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. I computed time series numerical standard errors with the *R* package coda, then constructed symmetric 95% and 99% intervals based on an asymptotic Gaussian approximation. The 95% interval did not include q in 11 cases out of 351 (3.13%) and the 99% interval included q in every case. The sample mean always lay in the interval $[q - 0.00013, q + 0.00013]$. These results fail to cast doubt on the correctness of the implementation.

4.2. Exchange rates

I now illustrate the numerical estimation of values of the likelihood function for exchange rate data. The data consist of log returns for the Deutschmark in US dollars. The series runs from January 2, 1980 to May 31, 1990 and consists of 2613 daily log returns. The data are similar to data used by Kim et al. (1998), Shephard and Pitt (1997) and Liesenfeld and Richard (2006) and originate from the Board of Governors of the US Federal Reserve System. They were kindly supplied by Éric Jacquier and Nicholas Polson.

I use the Gaussian SV model in Liesenfeld and Richard (2006). State dynamics are given by (14) and log returns y_t by

$$y_t|\theta, \alpha_t \sim N(0, \exp(\alpha_t)).$$

I performed the following computational experiment. I used four different methods to approximate a cross section of the likelihood function $L(\theta; y)$ on a grid. I kept $\bar{\alpha}$ and ϕ fixed at their maximum likelihood values and used the following grid of 17 values of σ: 0.12, 0.12375, 0.1275, …, 0.18. The maximum likelihood estimate for σ, denoted $\hat{\sigma}$, is approximately 0.147. I use the notation $L(\sigma; y)$ for the restriction of $L(\theta; y)$ to the grid.

Let $\hat{L}_H(\sigma; y)$ be the IS estimator of $L(\sigma; y)$ using the HESSIAN approximation as an importance density. Let $\hat{L}_{EIS}(\sigma; y)$ be the Efficient Importance Sampling estimator of $L(\sigma; y)$—as in Liesenfeld and Richard (2006), I used a Gaussian auxiliary parametric importance sampler, three EIS iterations to compute its parameters and $M = 30$ trajectories. Finally, let $L_G(\sigma; y)$ be the Laplace approximation of $L(\sigma; y)$ and $L_H(\sigma; y)$ be the Laplace-like approximation

using the HESSIAN approximation. Both $L_G(\sigma; y)$ and $L_H(\sigma; y)$ are described in Section 3.

Fig. 1 show the results. The approximation errors $\log L_G(\sigma; y) - \log L(\sigma; y)$ (left panel) and $\log L_H(\sigma; y) - \log L(\sigma; y)$ (right panel) are plotted as dot-dashed lines. The left panel also shows the sample mean and quantiles 0.01, 0.05, 0.95 and 0.99 of the error $\log \hat{L}_{EIS}(\sigma; y) - \log L(\sigma; y)$, based on $R = 5000$ replications of $\hat{L}_{EIS}(\sigma; y)$. As Liesenfeld and Richard (2006) recommend, Common Random Numbers (CRNs) were used for all values of σ within a replication, and CRNs are independent across replications. The right panel shows the mean and quantiles of the error $\log \hat{L}_H(\sigma; y) - \log L(\sigma; y)$, also for $M = 30$. These are based on an asymptotic Gaussian approximation obtained using a larger sample, with $M = 20,000$. The larger sample also gives very precise and unbiased numerical estimates of $L(\sigma; y)$ used to compute the errors of other methods.

The EIS estimator $\log \hat{L}_{EIS}(\sigma; y)$ is biased, and the bias depends on σ. For $M = 30$, the bias has roughly the same order of magnitude as the numerical standard error. For the same number of draws, the variance of $\log \hat{L}_{EIS}(\sigma; y)$ is well over four orders of magnitude higher than that of $\log \hat{L}_H(\sigma; y)$. According to timing experiments, the latter takes about 3.40 times as much processor time, so the overall efficiency ratio is about four orders of magnitude in favour of $\hat{L}_H(\sigma; y)$.

Since I use rejection sampling to draw from $q(\alpha|\theta, y)$, $\hat{L}_H(\theta; y)$ is not continuous in θ for fixed primitive random variates, as $\hat{L}_{EIS}(\sigma; y)$ is. This makes the CRN technique less effective for computing $\hat{L}_H(\theta; y)$. However, $L_H(\theta; y)$ is continuous in θ. For the example above, the error $L_H(\sigma; y) - L(\sigma; y)$ is much smaller than both the bias and the numerical standard error of $\hat{L}_{EIS}(\sigma; y)$. $L_H(\sigma; y)$ requires less than one fifth the time to compute.

The fact that the bias of $\log \hat{L}_{EIS}(\sigma; y)$ depends on σ suggests that numerical approximations of the maximum likelihood estimate are also biased. For each of the $R = 5000$ trajectories, and with $\bar{\alpha}$ and ϕ set to their maximum likelihood values, I approximate the restricted maximum likelihood estimate $\hat{\sigma}_{EIS}$ using a cubic spline interpolation of $\log \hat{L}_{EIS}(\sigma; y)$ within the grid points. I use the sample mean and standard deviation of $R = 5000$ independent draws of $\hat{\sigma}_{EIS} - \hat{\sigma}$ to estimate the bias and numerical standard error of $\hat{\sigma}_{EIS}$. I obtain a bias of -4.64×10^{-4} and a numerical standard error of 4.84×10^{-4}. Both are fairly small compared with the sampling variability of the full maximum likelihood estimator of σ: for a similar sample, Liesenfeld and Richard (2006) report an asymptotic standard error for the MLE of σ of 0.032.

4.3. Stock market index returns

I now illustrate the performance of the HESSIAN method for posterior simulation using a data set with $n = 8851$ observations of S&P 500 daily log returns, without dividends, from July 3, 1962 to August 26, 1997, inclusive. I obtained this data set at Yahoo finance, and it corresponds to the data set used in Chib et al. (2002).

I use the following Student's t SV model, described in Chib et al. (2002), a variant of the model in (15):

$$y_1|\theta, \alpha_1 \sim t(a, \exp(\alpha_1), \nu),$$
$$y_t|\theta, \alpha_t \sim t(a + by_{t-1}, \exp(\alpha_t), \nu),$$

where a and b are additional parameters. This model had the highest marginal likelihood of the models compared in that paper. I use the same expressions for the $\psi_t(\alpha_t)$ and their derivatives as for the model in (15), tabulated in Appendix I, except that I evaluate them at $y_1 - a$ and $y_t - a - by_{t-1}$ instead of y_1 and y_t.

I complete the model with a prior distribution on the parameters $\bar{\alpha}$, ϕ, σ, ν, a and b. Using prior predictive analysis, I constructed a prior distribution over the parameters that covers reasonable values of nine functions of interest. These are the six parameters themselves, the unconditional standard deviation of α_t, given by $\sigma/\sqrt{1-\phi^2}$, and two functions discussed in Jacquier et al. (1994), namely the square of the coefficient of variation of volatility (CV^2) and the half life of shocks to volatility (HL). The two functions are defined as

$$CV^2 \doteq \frac{V[\exp(\alpha_t)|\theta]}{E[\exp(\alpha_t)|\theta]^2} = \exp\left[\frac{\sigma^2}{1-\phi^2}\right] - 1,$$
$$HL \doteq -\log 2/\log\phi.$$

The result was the following multivariate Gaussian prior distribution on the transformed parameter $\theta \doteq (\bar{\alpha}, \tanh^{-1}\phi, \log\sigma, \log\nu, a, b)$:

$$\theta \doteq \begin{bmatrix} \bar{\alpha} \\ \tanh^{-1}\phi \\ \log\sigma \\ \log\nu \\ a \\ b \end{bmatrix} \sim N\left(\begin{bmatrix} -11.0 \\ 2.1 \\ -1.8 \\ 2.5 \\ 0.0 \\ 0.0 \end{bmatrix}, \begin{bmatrix} 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & -0.05 & 0 & 0 & 0 \\ 0 & -0.05 & 0.125 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4\times10^{-6} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.04 \end{bmatrix}\right).$$

Table 2 gives prior sample quantiles of the nine functions of interest, for a random sample of size 10^5. The marginal distributions for CV^2 and HL are quite diffuse compared with the range of values obtained in the studies reviewed by Jacquier et al. (1994).

I now compare various methods for joint posterior simulation of θ and α, for the Student's t stochastic volatility model and the S&P 500 data set.

I use the same approximation $q(\theta, \alpha|y) = q(\theta|y)q(\alpha|\theta, y)$ as an importance density for IS and as a proposal density for MCMC. The marginal proposal density $q(\theta|y)$ is the following perturbed multivariate Student's t density:

$$q(\theta|y) \doteq \frac{\Gamma((\nu+K)/2)}{\Gamma(\nu/2)}(\nu\pi)^{-K/2}|\Omega|^{1/2} \times \left[1 + \frac{(\theta-\bar{\theta})^\top\Omega(\theta-\bar{\theta})}{\nu}\right]^{-(\nu+K)/2}(1+g(\theta)), \quad (16)$$

where $K = 6$, the length of θ, $\bar{\theta}$ maximises $p(\theta)L_H(\theta; y)$, Ω is the Hessian of $-\log[p(\theta)L_H(\theta; y)]$ at $\bar{\theta}$ and

$$g(\theta) \doteq \min\left(0.9, \max\left(-0.9, \sum_{k=1}^{K}\frac{1}{6}\frac{\partial^3\log[p(\theta)L_H(\theta; y)]}{\partial\theta_k^3}\cdot\theta_k^3\right)\right),$$

an odd function with range $[-0.9, 0.9]$. Recall the definition of the approximate likelihood $L_H(\theta; y)$ in (13). I simulate from $q(\theta|y)$ using reflection sampling, described in Appendix H, together with standard methods for drawing multivariate Student's t variates.

I denote the use of $q(\theta, \alpha|y)$ for IS as the H-IS method and its use for MCMC as the H-MCMC method. I compare these methods with that described in Chib et al. (2002) (denoted CNS-7) and a variant (CNS-10) where the mixture approximation with seven components is replaced by the one with ten components described in Omori et al. (2007). I use the same prior for all methods for comparability. Note that this is not the same prior as used in Chib et al. (2002). In all cases, I generated posterior samples of size 10^5. I used the same realised draws from $q(\theta, \alpha|y)$ for both H-IS and H-MCMC.

Table 3 shows the numerical efficiency of the posterior sample mean as an estimator of the posterior mean, for each of six

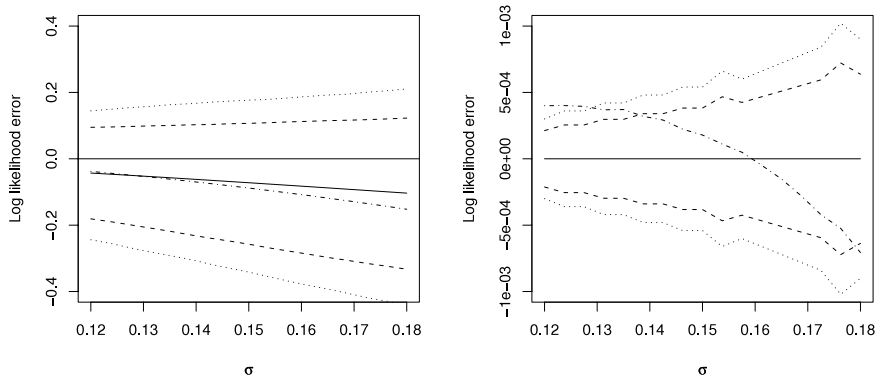

Fig. 1. Left panel: mean (solid), quantiles 0.01 (dotted), 0.05 (dashed), 0.95 (dashed) and 0.99 (dotted) of log likelihood error for $\hat{L}_{EIS}(\sigma; y)$ with $M = 30$, and likelihood error for $L_G(\sigma; y)$ approximation (dot-dashed). Right panel: mean (solid) and same quantiles of likelihood error for $\hat{L}_H(\sigma; y)$ approximation with $M = 30$, and likelihood error for $L_H(\theta; y)$ approximation (dot-dashed).

Table 2
Prior quantiles.

	$\bar{\alpha}$	ϕ	σ	ν	a	b	$\sigma/\sqrt{1-\phi^2}$	CV^2	HL
0.01	−15.65	0.8773	0.073	3.80	−0.00468	−0.468	0.424	0.097	5.29
0.05	−14.30	0.9185	0.093	5.35	−0.00329	−0.328	0.542	0.163	8.15
0.25	−12.35	0.9550	0.131	8.70	−0.00133	−0.135	0.760	0.343	15.04
0.5	−11.01	0.9704	0.165	12.18	0.00001	−0.001	0.963	0.603	23.08
0.75	−9.66	0.9806	0.210	17.12	0.00136	0.133	1.222	1.132	35.35
0.95	−7.71	0.9895	0.296	27.74	0.00328	0.328	1.721	3.466	65.43
0.99	−6.34	0.9931	0.378	39.04	0.00464	0.461	2.200	10.530	100.57

Table 3
Multiple comparison.

	Mean	Std	NSE-B	RNE-B	NSE-TS	RNE-TS
H-IS, $\bar{\alpha}$	−10.07966	0.12337	4.28e−04	0.83		
H-MCMC, $\bar{\alpha}$	−10.07957	0.12305	4.83e−04	0.65	4.67e−04	0.70
CNS-7, $\bar{\alpha}$	−10.08400	0.12155	5.88e−04	0.43	5.74e−04	0.45
CNS-10, $\bar{\alpha}$	−10.07957	0.12288	5.37e−04	0.52	5.28e−04	0.54
H-IS, ϕ	0.99019	0.00192	6.68e−06	0.83		
H-MCMC, ϕ	0.99019	0.00192	7.76e−06	0.61	8.20e−06	0.55
CNS-7, ϕ	0.98990	0.00197	1.35e−05	0.21	1.43e−05	0.19
CNS-10, ϕ	0.99017	0.00191	1.57e−05	0.15	1.47e−05	0.17
H-IS, σ	0.10794	0.00853	2.74e−05	0.97		
H-MCMC, σ	0.10794	0.00854	2.89e−05	0.87	2.65e−05	1.04
CNS-7, σ	0.10981	0.00870	8.55e−05	0.10	9.10e−05	0.09
CNS-10, σ	0.10812	0.00842	1.03e−04	0.07	9.67e−05	0.08
H-IS, ν	12.79220	1.77886	5.85e−03	0.92		
H-MCMC, ν	12.79172	1.78922	6.69e−03	0.72	6.03e−03	0.88
CNS-7, ν	12.60477	1.72429	2.24e−02	0.06	2.16e−02	0.06
CNS-10, ν	12.79387	1.77099	2.33e−02	0.06	2.51e−02	0.05
H-IS, a	0.00041	0.00007	2.08e−07	0.98		
H-MCMC, a	0.00041	0.00007	2.46e−07	0.71	2.56e−07	0.65
CNS-7, a	0.00041	0.00007	2.67e−07	0.60	3.21e−07	0.41
CNS-10, a	0.00041	0.00007	2.86e−07	0.52	2.62e−07	0.62
H-IS, b	0.13806	0.01076	3.44e−05	0.98		
H-MCMC, b	0.13807	0.01078	4.39e−05	0.60	5.07e−05	0.45
CNS-7, b	0.13800	0.01074	4.90e−05	0.48	4.16e−05	0.67
CNS-10, b	0.13812	0.01073	4.83e−05	0.49	4.95e−05	0.47

parameters and four methods. For the methods H-MCMC, CNS-7 and CNS-10, I compute numerical standard errors and relative numerical efficiency using two methods implemented in the *R* package coda. The first is the batch mean method, for which I chose a batch length of 500. This gives the numerical standard error NSE and the relative numerical efficiency RNE. The second is a time-series method which uses an estimate of the spectral density at frequency zero. This gives NSE-TS and RNE-TS. For the method H-IS, I compute the numerical standard error and relative numerical efficiency defined in Geweke (1989) and tabulate them in the columns NSE and RNE. Computational time was 2056 s for H-IS and H-MCMC, 3013 s for CNS-7 and 3325 s for CNS-10.

For some parameters, the numerical efficiency of the CNS-7 and CNS-10 sample means is greater than 50%. For these parameters, the numerical efficiency of H-IS and H-MCMC is comparable to that of CNS-7 and CNS-10. For the other parameters, the numerical efficiency of CNS-7 and CNS-10 sample means is much lower, but the efficiency of H-IS and H-MCMC sample means remains higher

than 50%. The numerical efficiency of H-IS dominates all other methods for all parameters, and H-IS takes about 1/3 less time as CNS-7 and CNS-10.

For CNS-7 and CNS-10, the reported numerical standard errors do not fully account for the error of the sample means. Standard MCMC theory does not guarantee the simulation consistency of the CNS-7 and CNS-10 sample means: Gibbs draws of ν and its associated latent variables λ_t are based on their true conditional distribution, but other variables are drawn from approximate conditional distributions. There is strong evidence against the simulation consistency of CNS-7 sample means: for σ and ν, the bias is much larger than the numerical standard error. For CNS-10, the approximate conditional distributions are closer to the true ones and any bias there may be is very small—it does not clearly dominate the numerical standard error. But we only know this through external comparison with a simulation consistent method.

Numerical efficiency of the H-IS and H-MCMC sample means is considerably higher than the efficiency reported for block sampling methods in Shephard and Pitt (1997), Watanabe and Omori (2004), Strickland et al. (2006), Jungbacker and Koopman (2008) and Omori and Watanabe (2008). These papers analyse similar (some simpler, some richer) stochastic volatility models. In all cases, data sample sizes were considerably smaller.

We obtain $\log \hat{p}(y) = 31,092.96$ and $\hat{\sigma}_p/\hat{p}(y) = 0.0011$ for the Student's t SV example. The addition computation takes a negligible amount of time. Chib et al. (2002) approximate marginal likelihoods using a method involving supplementary simulation and particle filtering. They do not report standard errors, but results they report in Table 10 for different simulation sample sizes suggest that numerical efficiency is several orders of magnitude lower.

In a separate computational experiment, I alternated between an MCMC block updating θ and α and a block updating α only, to measure the relative contributions to numerical inefficiency of $q(\alpha|\theta, y)$ and $q(\theta|y)$. The acceptance probability for α alone, averaged over the posterior distribution of θ, was 0.9985, compared to 0.906 for a joint proposal of θ and α. This suggests that the approximation $q(\theta|y)$ is the weak link for this empirical example.

4.4. Transaction counts

The next example illustrates the use of the HESSIAN method for dynamic count data. I use a data set with 4914 observations of transaction counts for IBM stock. There are 78 observations for each day, corresponding to 5 min intervals from 8:30 am to 4:30 pm. There are 63 trading days, from November 1, 1990 to January 31, 1991. Details are in Chapter 5 of Tsay (2002), and the raw data are kindly provided by Ruey Tsay at the website for his book.

Dynamic count models have been successfully used in many applications where count intensity varies over time. Harvey and Durbin (1986) and Durbin and Koopman (1997) study counts of motorist, passenger, cyclist and pedestrian casualties in Britain; Frühwirth-Schnatter and Wagner (2006), counts of pedestrian casualties in Linz. Harvey and Fernandes (1989) analyse time series of counts of goals scored in soccer games, reported purse snatchings in Chicago, and deaths of van drivers in Britain. Zeger (1988) use data on numbers of reported incidents of polio in the United States and Jung et al. (2006) consider counts of daily admissions for asthma to a hospital in Sydney. Rydberg and Shephard (2003) and Liesenfeld et al. (2006) look at the absolute value of asset price changes as multiples of the tick size.

Most of these studies use a conditional Poisson distribution. Some are parameter-driven, with the Poisson intensity depending on a latent process. Others are observation-driven, with the intensity given as a deterministic function of lagged observations. A simple parameter-driven model, without covariates, is given by the state dynamics in (14) and

$$y_t|\theta, \alpha_t \sim \text{Poisson}(\exp(\alpha_t)).$$

Just as the Student's t SV model features overdispersion of the conditional log return distribution relative to the Gaussian SV model, the following Gamma–Poisson count model features overdispersion of the conditional count distribution relative to the Poisson count model:

$$y_t|\theta, \alpha_t \sim \text{Gamma Poisson}(r, \exp(\alpha_t)).$$

The Gamma–Poisson distribution is a Gamma mixture of Poisson distributions: if we draw λ from a Gamma distribution with shape r and scale β, then y from the Poisson distribution with mean λ, the marginal distribution of y is Gamma Poisson(r, β). For integer values of r, this is the negative binomial distribution with r failures and success probability $\beta/(1+\beta)$. Rydberg and Shephard (2003) and Liesenfeld et al. (2006) use a negative binomial distribution to capture conditional overdispersion in price change data.

Here, I use a more diffuse prior for the parameters. For the state parameters $\bar{\alpha}$, ϕ and σ, I use the following prior.

$$\begin{bmatrix} \bar{\alpha} \\ \tanh^{-1}\phi \\ \log\sigma \end{bmatrix} \sim N\left(\begin{bmatrix} 0.0 \\ 1.5 \\ -1.5 \end{bmatrix}, \begin{bmatrix} 25.0 & 0 & 0 \\ 0 & 0.625 & -0.25 \\ 0 & -0.25 & 0.5 \end{bmatrix} \right). \quad (17)$$

For the Gamma–Poisson model, r is *a priori* independent of the other parameters, with $\log r \sim N(2.5, 1)$. Values and derivatives of the $\psi_t(\alpha_t)$ for both the Gamma and the Gamma–Poisson model are tabulated in Appendix I.

Table 4 shows posterior results for the Poisson (Po) and Gamma–Poisson (Ga–Po) models. Posterior simulation is by independence Metropolis–Hastings, using a joint proposal for θ and α based on the HESSIAN approximation and an approximation $q(\theta|y)$ analogous to that in (16). Results are based on a posterior sample size of $M = 25,000$. This takes 320 s for the Poisson model and 295 s for the Gamma–Poisson model.

Numerical efficiency is quite high despite considerable non-Gaussianity of the count data: the mean count is 12.18, and the numbers of periods with 0, 1, and 2 transactions are 83, 84 and 119, respectively.

For the Poisson and Gamma–Poisson examples, I obtain marginal likelihoods -15372.94 (0.0023) and -15279.15 (0.0033), respectively. The Bayes factor decisively favours the more flexible Gamma–Poisson model.

4.5. Transaction durations

The next example is a model of durations between financial transactions, and illustrates the performance of the HESSIAN method with a much larger sample, of high frequency financial data. I use a data set with adjusted durations between trades, in seconds, also for IBM stock, and during the same period as the count data. There are 59,838 adjusted durations in the sample, available at the same website. The adjustments, which remove diurnal patterns, are described in Tsay (2002).

Dynamic models of durations between transactions are useful for understanding market microstructure. Engle and Russell (1998) introduce the autoregressive conditional duration model (ACD), an observation-driven model for durations. Bauwens and Veredas (2004) introduce the stochastic conditional duration model (SCD), a similar, but parameter-driven model, and describe a quasi-maximum likelihood method for estimation. Strickland et al. (2006, 2008) also consider SCD models, using MCMC for Bayesian inference. They update blocks of latent states that are smaller than the entire sequence.

Table 4
Results for count and duration models.

	$q_{0.01}$	$q_{0.05}$	$q_{0.5}$	$q_{0.95}$	$q_{0.99}$	Mean	Std	NSE	RNE
Po, $\bar{\alpha}$	2.227	2.248	2.299	2.348	2.369	2.2986	0.0303	0.00020	0.883
Po, ϕ	0.792	0.800	0.818	0.835	0.843	0.8179	0.0108	0.00007	0.828
Po, σ	0.358	0.363	0.375	0.388	0.393	0.3755	0.0075	0.00006	0.608
Ga–Po, $\bar{\alpha}$	−0.367	−0.302	−0.158	−0.018	0.041	−0.1586	0.0869	0.00081	0.460
Ga–Po, ϕ	0.908	0.914	0.928	0.941	0.946	0.9279	0.0081	0.00006	0.636
Ga–Po, σ	0.196	0.202	0.219	0.237	0.245	0.2196	0.0106	0.00010	0.496
Ga–Po, r	10.377	10.821	12.115	13.747	14.572	12.1800	0.8952	0.00801	0.500
Exp, $\bar{\alpha}$	0.558	0.570	0.599	0.629	0.642	0.5992	0.0180	0.00010	0.693
Exp, ϕ	0.906	0.910	0.919	0.927	0.930	0.9187	0.0050	0.00003	0.566
Exp, σ	0.311	0.319	0.338	0.359	0.367	0.3382	0.0121	0.00007	0.549

I use a simple stochastic duration model, where state dynamics are given by (14) and durations are conditionally exponentially distributed:

$$y_t|\theta, \alpha_t \sim \text{Exponential}(\exp(\alpha_t)).$$

Table 4 shows posterior results for this (Exp) model, based on a posterior independence Metropolis–Hastings sample of size $M = 50,000$. Simulation takes 7437 s. Despite a very long sample, numerical efficiency is quite high.

4.6. Other comparisons

I now make some indirect comparisons of the HESSIAN method with other methods for approximating values of the likelihood function for state space models of the kind described in the introduction.

It is typically trivial to approximate values of the likelihood function using the computational bi-products of particle filtering. See the examples in the survey by Doucet and Johansen (2010). Numerical precision, however, is much lower. To see why, consider an ideal case where we have a fully adapted auxiliary particle filter, as described in Pitt and Shephard (1999). Even if we had an iid particle sample $\{\alpha_t^{(m)}\}_{m=1}^M$ from $p(\alpha_t|y_1, \ldots, y_t)$, then the usual estimate $\hat{p}(y_{t+1}|y_1, \ldots, y_t) \doteq M^{-1}\sum_{m=1}^M p(y_{t+1}|\alpha_t^{(m)})$ of the likelihood factor $p(y_{t+1}|y_1, \ldots, y_t)$ is the sample mean of a quantity that varies considerably in the population. This is for a *single* factor of the likelihood function. In contrast, it is clear from Fig. 1 that $p(\alpha, y|\theta)/q(\alpha|\theta, y)$ has very small variance with respect to $q(\alpha|\theta, y)$. For this reason, one would require an extremely large number of particles to achieve the same numerical precision as $\hat{L}_H(\theta)$ for the full likelihood.

Durbin and Koopman (1997) do importance sampling using a multivariate Gaussian approximation. Their model is a stochastic Poisson count model of road accident deaths, and their data consist of 192 observations. Their most numerically efficient approximations of values of the likelihood function are for the case where they use antithetic variables for both location and scale and a control variable. They achieve a numerical variance of 4.61×10^{-5} using $M = 50$ trajectories. In the example above, there are $M = 30$ trajectories and $n = 2613$ observations, and all numerical variances are less than 2.0×10^{-7}. This is without using variation reduction techniques to improve the numerical efficiency of the HESSIAN method.

5. Conclusions

I have constructed an approximation $q(\alpha|\theta, y)$ to the target density $p(\alpha|\theta, y)$ and shown that it can be used as an importance density for IS or as a proposal density for an independence Metropolis–Hastings update of the target distribution.

I have shown how to evaluate and sample $q(\alpha|\theta, y)$ exactly, and have tested the correctness of my implementation. This means we can appeal to standard results characterising IS and MCMC numerical errors. The same is not true for some competing methods. The EIS estimator for values of the likelihood function can be biased, and since this bias can depend on parameter values, numerical approximations of maximum likelihood estimates can also be biased. Auxiliary mixture sampling methods that do not compensate for approximation error do not necessarily give simulation consistent sample means: i have given evidence that the simulator in Chib et al. (2002) generates posterior samples whose means are not simulation consistent. Errors associated with approximate IS or MCMC are often small, but we only know this by comparing results with those obtained using exact IS or MCMC—ultimately, measurement of numerical error is limited by the precision of the exact method.

The HESSIAN method is highly numerically efficient for the examples in Section 4. These feature a variety of models and large data sets. Efficiency relies on $q(\alpha|\theta, y)$ being an extremely close approximation of $p(\alpha|\theta, y)$ even for large n, and $p(\alpha|\theta, y)/q(\alpha|\theta, y)$ being bounded.

Further advantages of the HESSIAN approximation are realised by combining it with an approximation $q(\theta|y)$ of $p(\theta|y)$ such that $p(\theta|y)/q(\theta|y)$ is bounded. The result is an approximation $q(\theta, \alpha|y) = q(\theta|y)q(\alpha|\theta, y)$ of the full posterior density $p(\theta, \alpha|y)$ such that $p(\theta, \alpha|y)/q(\theta, \alpha|y)$ is bounded. When $q(\theta, \alpha|y)$ is used for IS to compute integrals with respect to $p(\theta, \alpha|y)$, the variance of the importance sample weights is finite. When it is used as a proposal density for independence Metropolis–Hastings simulation of $p(\alpha|\theta, y)$, the chain is geometrically ergodic.

Drawing (θ, α) as a single block, with no data augmentation, has many important advantages. As with drawing α as a block, it solves the problem of high posterior correlation within α. In addition, it also avoids inefficiencies due to posterior dependence between α and θ. In the examples in this paper it is highly numerically efficient and no burn-in draws need to be removed.

It also enables IS for the full posterior distribution, which has some advantages over MCMC. In the stock market index example, IS is more efficient than MCMC for all six parameters. One can approximate the marginal likelihood $p(y)$ as the mean of a random sample with low variance—this is trivial to implement, takes negligible time to compute, and gives much higher numerical precision than particle filtering methods. Importance draws do not need to be independent, which facilitates variance reduction techniques such as antithetic sampling and randomised quasi-Monte Carlo.

The HESSIAN method is reasonably fast, requiring $O(n)$ operations for evaluating $q(\alpha|\theta, y)$ and drawing from it. No simulation is required. It takes about 1/3 less time to generate a posterior sample than the method of Chib et al. (2002), for the S&P example. Simulation smoothing is computationally intensive, and for long data sets, the amount of time it takes to generate a sample whose numerical standard error is a small fraction of the posterior

standard deviation is far from negligible. Such improvements in computational efficiency are valuable.

Since the construction of $q(\alpha|\theta, y)$ does not require special distributional assumptions on observations, data augmentation is easily avoided: in the Student's t SV model, it is not necessary to use data augmentation to make log returns conditionally Gaussian.

Compared with $q(\alpha|\theta, y)$, the approximation $q(\theta|y)$ used in this paper seems relatively crude: in the S&P 500 SV example, the acceptance probability for an update of α alone was 0.9985 and for a joint update of θ and α, 0.906. For this model and data, the $q(\theta|y)$ approximation seems to be the weak link. This observation, as well as the very high efficiency obtained in the stochastic duration example suggest that applying the HESSIAN method to very large high frequency financial data sets is promising.

Computing the $q(\alpha|\theta, y)$ for a new model requires code to evaluate the first five derivatives of $\log p(y_t|\theta, \alpha_t)$. This is not trivial, but one can use numerical derivatives or other approximations: the approximation $q(\alpha|\theta, y)$ may be worse as a result, but one can still evaluate and sample it exactly. Furthermore, we do not actually need analytic expressions for the derivatives to compute exact values at a point. If the expression for $\log p(y_t|\theta, \alpha_t)$ involves products or compositions of primitive functions, we can evaluate analytic derivatives of the primitive functions and then use general purpose routines to combine them according to Leibniz' rule or Faà di Bruno's rule to give the derivatives of $\log p(y_t|\theta, \alpha_t)$.

The perturbed Gaussian distribution has some secondary parameters controlling tail thickness, and one needs to choose their values. However, all empirical examples in Section 4 used the same recommended values in Appendix G. Furthermore, experimentation not reported suggests that results are not sensitive to their values within a large region.

There are important limitations to the HESSIAN method, and some of these are the impetus for ongoing work. I now require the state α_t to be univariate, but the HESSIAN method might be extended to models with multivariate states in two different ways. The first and simplest way would be to sample multivariate states as a single block in the temporal dimension but one-at-a-time in the cross section dimension. The HESSIAN approximation is computed for each target conditional distribution in turn. A true multivariate extension of the HESSIAN method, giving a close enough approximation of the full target distribution to allow a single draw of multivariate states, would be more difficult. One key result, however, generalises easily to multivariate states: the derivation in Appendix C of Eq. (3) for $h'_t(\alpha_t, \alpha_{t+1})$ does not require univariate states.

Another limitation is that the conditional distribution of the observed vector y_t given α can only depend on the contemporaneous state α_t. This rules out any conditional dependence between the state *innovation* at time t and the observed vector at t. SV models with the asymmetric volatility effect often known as the leverage effect feature this conditional dependence and are more realistic than those without. Another limitation is that α must be Gaussian. I am currently working with a coauthor to extend the HESSIAN method to models with conditional dependence and states that are non-Gaussian but still Markov.

Acknowledgements

I thank Barnabé Djegnéné, Éric Jacquier, Shirley Miller, Nick Polson, Håvard Rue and several reviewers for helpful comments and corrections.

Appendix A. Computing $\bar{\Omega}$ and $\bar{c}$

The marginal distribution of states is multivariate Gaussian. We can write $\alpha \sim N(\bar{\Omega}^{-1}\bar{c}, \bar{\Omega}^{-1})$, where $\bar{\Omega}$ is the prior precision of α and $\bar{c}$ is the prior covector. We can compute $\bar{\Omega}$ and $\bar{c}$ by writing out the log density

$$\log p(\alpha) = \log p(\alpha_1) + \sum_{t=2}^{n} \log p(\alpha_t|\alpha_{t-1})$$

using (1), writing the same density in terms of $\bar{\Omega}$ and $\bar{c}$,

$$\log p(\alpha) = \frac{1}{2}\left[\log|\bar{\Omega}| - n\log 2\pi - \alpha^\top\bar{\Omega}\alpha + 2\bar{c}^\top\alpha - \bar{c}^\top\bar{\Omega}^{-1}\bar{c}\right], \quad (18)$$

and then matching linear and quadratic terms. We obtain, for $t = 1, \ldots, n-1$,

$$\begin{aligned}&\bar{\Omega}_{tt} = \sigma_{t-1}^{-2} + \sigma_t^{-2}\phi_t^2, \qquad \bar{\Omega}_{t,t+1} = \bar{\Omega}_{t+1,t} = -\sigma_t^{-2}\phi_t,\\ &\bar{c}_t = \sigma_{t-1}^{-2}d_{t-1} - \sigma_t^{-2}\phi_t d_t,\end{aligned} \quad (19)$$

and

$$\bar{\Omega}_{nn} = \sigma_{n-1}^{-2}, \qquad \bar{c}_n = \sigma_{n-1}^{-2}d_{n-1}, \quad (20)$$

with all other elements equal to zero.

Eq. (14) in Section 4 describes a stationary special case with parameters $\bar{\alpha}$, ϕ and σ. Here, $\sigma_0 = \sigma/\sqrt{1-\sigma^2}$, $d_0 = \bar{\alpha}$, and for $t > 0$, $\sigma_t = \sigma$, $d_t = (1-\phi)\bar{\alpha}$ and $\phi_t = \phi$. Thus

$$\begin{aligned}&\bar{\Omega}_{11} = \bar{\Omega}_{nn} = \sigma^{-2}, \qquad \bar{\Omega}_{tt} = \sigma^{-2}(1+\phi^2),\\ &\quad t = 2, \ldots, n-1,\\ &\bar{\Omega}_{t,t+1} = -\sigma^{-2}\phi, \quad t = 1, \ldots, n-1,\\ &\bar{c}_1 = \bar{c}_n = \sigma^{-2}(1-\phi)\bar{\alpha}, \qquad \bar{c}_t = \sigma^{-2}(1-\phi)^2\bar{\alpha},\\ &\quad t = 2, \ldots, n-1.\end{aligned}$$

Appendix B. Computing the mode of the target distribution

In this appendix we show how to compute a, the conditional mode of α given θ and y. Bi-products of the computation include $\tilde{\bar{\Omega}}$ and $\tilde{\bar{c}}$, the precision and covector of the multivariate Gaussian distribution whose mode is a and whose log density has the same Hessian (with respect to α) as $\log p(\alpha|y)$ at a. Another bi-product is the sequence of conditional variances $\Sigma_1, \ldots, \Sigma_n$ defined in Section 2.2.

Previous authors have used the Kalman filter (Durbin and Koopman, 1997) or Extended Kalman filter (Fahrmeir, 1992) to compute the conditional mode a. As in Durbin and Koopman (1997), I use the Newton–Raphson method. However, the implementation of each step is quite different, as mine is not based on the Kalman filter.

One can write $\log p(\alpha|y) = \log p(\alpha) + \log p(y|\alpha) + k$, where k does not depend on α. We now define, for all $\tilde{a} \in \mathbb{R}^n$,

$$\begin{aligned}&\tilde{\bar{\Omega}}(\tilde{a}) \doteq \bar{\Omega} + \operatorname{diag}(-\psi_1''(\tilde{a}_1), \ldots, -\psi_n''(\tilde{a}_n)),\\ &\tilde{\bar{c}}(\tilde{a}) \doteq \bar{c} + (\psi_1'(\tilde{a}_1) - \psi_1''(\tilde{a}_1)\tilde{a}_1, \ldots, \psi_n'(\tilde{a}_n) - \psi_n''(\tilde{a}_n)\tilde{a}_n).\end{aligned} \quad (21)$$

Thus $\tilde{\bar{\Omega}} = \tilde{\bar{\Omega}}(a)$ and $\tilde{\bar{c}} = \tilde{\bar{c}}(a)$. It is straightforward to check that the multivariate Gaussian log density with the same gradient and Hessian as $\log p(\alpha|y)$ at $\tilde{a}$ is given by

$$\frac{1}{2}\log|\tilde{\bar{\Omega}}(\tilde{a})| - \frac{n}{2}\log 2\pi - \frac{1}{2}\left[\alpha^\top\tilde{\bar{\Omega}}(\tilde{a})\alpha - 2\tilde{\bar{c}}(\tilde{a})^\top\alpha + \tilde{\bar{c}}[\tilde{\bar{\Omega}}(\tilde{a})]^{-1}\tilde{\bar{c}}\right].$$

The precision and co-vector of this distribution are $\tilde{\bar{\Omega}}(\tilde{a})$ and $\tilde{\bar{c}}(\tilde{a})$, so the mean is $\mu(\tilde{a}) \doteq [\tilde{\bar{\Omega}}(\tilde{a})]^{-1}\tilde{\bar{c}}(\tilde{a})$, which can be computed using the following procedure, based on an algorithm by Vandebril et al. (2007) for solving the symmetric band diagonal system $\Omega\mu = c$.

The forward pass is

$$\Sigma_1 = \Omega_{11}^{-1}, \qquad m_1 = \Sigma_1 c_1,$$
$$\Sigma_t = (\Omega_{tt} - \Omega_{t,t-1}^2 \Sigma_{t-1})^{-1}, \qquad m_t = \Sigma_t (c_t - \Omega_{t,t-1} m_{t-1}),$$

and the backwards pass is

$$\mu_n = m_n, \qquad \mu_t = m_t - \Omega_{t,t+1} \Sigma_t \mu_{t+1}.$$

To compute the mode a, I use a Newton–Raphson method. The initial value of a is the value computed for the previous value of θ, if it is available, otherwise the prior mean. The computation $\tilde{a}' = [\bar{\bar{\Omega}}(\tilde{a})]^{-1} \bar{\bar{c}}(\tilde{a})$ described above is repeated until numerical convergence. The final values of $\bar{\bar{\Omega}}(\tilde{a})$, $\bar{\bar{c}}(\tilde{a})$ and $\Sigma_1, \ldots, \Sigma_n$ correspond to $\bar{\bar{\Omega}}$, $\bar{\bar{c}}$ and $\Sigma_1, \ldots, \Sigma_n$, defined in Section 2.2.

I make two modifications of this algorithm to reduce the number of iterations required to find the mode and to handle cases of non-convergence. First, I add an approximation of $\ddot{a}_t(\tilde{a}'_{t+1} - \tilde{a}_{t+1})^2/2$ to the value of $\tilde{a}'_t$ to take advantage of third derivative information. Second, in rare cases of non-convergence I restart, beginning with a sequence of line search steps.

Appendix C. Exact derivatives of $\log p(\alpha_t|\alpha_{t+1}, y)$

In this appendix, I derive an exact expression for all derivatives of $\log p(\alpha_t|\alpha_{t+1}, y)$ with respect to α_t, in terms of $\mu_{t-1|t}(\alpha_t)$ and its derivatives.

We first write, for $t = 2, \ldots, n-1$,

$$\begin{aligned} p(\alpha_t|\alpha_{t+1}, y) &= \int p(\alpha_1, \ldots, \alpha_t|\alpha_{t+1}, y) \prod_{\tau=1}^{t-1} d\alpha_\tau \\ &\propto \int p(\alpha_1, \ldots, \alpha_{t+1}, y) \prod_{\tau=1}^{t-1} d\alpha_\tau \\ &\propto \int p(\alpha_1, \ldots, \alpha_{t-1}, y_1, \ldots, y_{t-1}) \\ &\quad \times p(\alpha_t|\alpha_{t-1}) p(\alpha_{t+1}|\alpha_t) p(y_t|\alpha_t) \prod_{\tau=1}^{t-1} d\alpha_\tau \\ &= c(\alpha_t) p(\alpha_{t+1}|\alpha_t) p(y_t|\alpha_t), \end{aligned}$$

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

$$c(\alpha_t) \doteq \int p(\alpha_1, \ldots, \alpha_{t-1}, y_1, \ldots, y_{t-1}) p(\alpha_t|\alpha_{t-1}) \prod_{\tau=1}^{t-1} d\alpha_\tau.$$

Taking the logarithm yields

$$\log p(\alpha_t|\alpha_{t+1}, y) = \log c(\alpha_t) + \log p(\alpha_{t+1}|\alpha_t) + \psi_t(\alpha_t) + k, \quad (22)$$

where k is the logarithm of the unknown normalisation constant. Differentiating (22) with respect to α_t gives

$$\frac{\partial \log p(\alpha_t|\alpha_{t+1}, y)}{\partial \alpha_t} = \frac{1}{c(\alpha_t)} \frac{\partial c(\alpha_t)}{\partial \alpha_t} + \frac{\partial \log p(\alpha_{t+1}|\alpha_t)}{\partial \alpha_t} + \psi'_t(\alpha_t). \quad (23)$$

We now proceed to write the first term of (23) in a convenient form. Taking the derivative of $c(\alpha_t)$ with respect to α_t, we obtain

$$\begin{aligned} \frac{\partial c(\alpha_t)}{\partial \alpha_t} &= \int p(\alpha_1, \ldots, \alpha_{t-1}, y_1, \ldots, y_{t-1}) \\ &\quad \times \frac{\partial p(\alpha_t|\alpha_{t-1})}{\partial \alpha_t} \prod_{\tau=1}^{t-1} d\alpha_\tau \\ &= \int p(\alpha_1, \ldots, \alpha_{t-1}, y_1, \ldots, y_{t-1}) p(\alpha_t|\alpha_{t-1}) \\ &\quad \times \frac{\partial \log p(\alpha_t|\alpha_{t-1})}{\partial \alpha_t} \prod_{\tau=1}^{t-1} d\alpha_\tau. \end{aligned}$$

Dividing both sides by $c(\alpha_t)$ yields

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

This is the integral of the function $\partial \log p(\alpha_t|\alpha_{t-1})/\partial \alpha_t$ with respect to the normalised density $p(\alpha_1, \ldots, \alpha_{t-1}|\alpha_t, y_1, \ldots, y_{t-1})$. We can therefore write

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

Insert this expression into (23) to obtain

$$\begin{aligned} \frac{\partial \log p(\alpha_t|\alpha_{t+1}, y)}{\partial \alpha_t} &= E\left[\frac{\partial \log p(\alpha_t|\alpha_{t-1})}{\partial \alpha_t} \middle| \alpha_t, y \right] \\ &\quad + \frac{\partial \log p(\alpha_{t+1}|\alpha_t)}{\partial \alpha_t} + \psi'_t(\alpha_t) \\ &= E\left[\frac{\partial \log p(\alpha_t|\alpha_{t-1})}{\partial \alpha_t} \right. \\ &\quad \left. + \frac{\partial \log p(\alpha_{t+1}|\alpha_t)}{\partial \alpha_t} \middle| \alpha_t, \alpha_{t+1}, y \right] + \psi'_t(\alpha_t) \\ &= E\left[\frac{\partial \log p(\alpha_t|\alpha_{t-1}, \alpha_{t+1})}{\partial \alpha_t} \middle| \alpha_t, \alpha_{t+1}, y \right] \\ &\quad + \psi'_t(\alpha_t). \end{aligned}$$

Since

$$\begin{aligned} \frac{\partial \log p(\alpha_t|\alpha_{t-1}, \alpha_{t+1})}{\partial \alpha_t} &= -\bar{\Omega}_{t-1,t} \alpha_{t-1} - \bar{\Omega}_{tt} \alpha_t \\ &\quad - \bar{\Omega}_{t,t+1} \alpha_{t+1} + \bar{c}_t, \\ h'_t(\alpha_t; \alpha_{t+1}) &\doteq \frac{\partial \log p(\alpha_t|\alpha_{t+1}, y)}{\partial \alpha_t} \\ &= -\bar{\Omega}_{t-1,t} \mu_{t-1|t}(\alpha_t) - \bar{\Omega}_{tt} \alpha_t \\ &\quad - \bar{\Omega}_{t,t+1} \alpha_{t+1} + \bar{c}_t + \psi'_t(\alpha_t). \end{aligned} \quad (24)$$

We now have an exact expression for the first derivative of $\log p(\alpha_t|\alpha_{t+1}, y)$ with respect to α_t, in terms of $\mu_{t-1|t}(\alpha_t)$. Similar derivations give results for $t = 1$ and $t = n$:

$$\begin{aligned} h'_1(\alpha_1; \alpha_2) &= \frac{\partial \log p(\alpha_1|\alpha_2, y)}{\partial \alpha_1} \\ &= -\bar{\Omega}_{11} \alpha_1 - \bar{\Omega}_{12} \alpha_2 + \bar{c}_1 + \psi'_1(\alpha_1), \end{aligned} \quad (25)$$

$$\begin{aligned} h'_n(\alpha_n) &= \frac{\partial \log p(\alpha_n|y)}{\partial \alpha_n} \\ &= -\bar{\Omega}_{n-1,n} \mu_{n-1|n}(\alpha_n) - \bar{\Omega}_{nn} \alpha_n + \bar{c}_n + \psi'_n(\alpha_n). \end{aligned} \quad (26)$$

Higher order derivatives are given by

$$h''_t(\alpha_t) = \begin{cases} -\bar{\Omega}_{11} + \psi''_1(\alpha_1), & t = 1, \\ -\bar{\Omega}_{t-1,t} \mu'_{t-1|t}(\alpha_t) - \bar{\Omega}_{tt} + \psi''_t(\alpha_t), & t = 2, \ldots, n. \end{cases} \quad (27)$$

$$h^{(j)}_t(\alpha_t) = \begin{cases} \psi^{(j)}_1(\alpha_1), & t = 1, \\ -\bar{\Omega}_{t-1,t} \mu^{(j-1)}_{t-1|t}(\alpha_t) + \psi^{(j)}_t(\alpha_t), & t = 2, \ldots, n, \end{cases}$$
$$j = 3, 4, \ldots. \quad (28)$$

Appendix D. Derivatives of $a_{t|t+1}(\alpha_{t+1})$ and $s_{t|t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$

In this appendix I derive, for $t = 1, \ldots, n-1$, the first four derivatives of $a_{t|t+1}(\alpha_{t+1})$ and the first three derivatives of $s_{t|t+1}(\alpha_{t+1}) \doteq \log \Sigma_{t|t+1}(\alpha_{t+1})$ at the point $\alpha_{t+1} = a_{t+1}$. It should be clear how to compute higher order derivatives if these are desired.

I compute $\dot{a}_t$, $\ddot{a}_t$, $\dddot{a}_t$, $\ddddot{a}_t$, $\dot{s}_t$, $\ddot{s}_t$ and $\dddot{s}_t$ exactly using linear first order difference equations, where the forcing term at t depends on the derivatives of $\psi_t(\alpha_t)$ at a_t. The difference equations are based on derivatives of an equation giving a first order necessary condition for $(a_{1|t+1}, \ldots, a_{t|t+1})$ to be the conditional mode of $(\alpha_1, \ldots, \alpha_t)$ given $(\alpha_{t+1}, \ldots, \alpha_n)$ and y.

Derivatives $\dot{a}_t$, $\ddot{a}_t$, $\dddot{a}_t$ and $\ddddot{a}_t$. For $t = 1, \ldots, n-1$, the log conditional density of $(\alpha_1, \ldots, \alpha_t)$ given $(\alpha_{t+1}, \ldots, \alpha_n)$ and y is

$$\log p(\alpha_1, \ldots, \alpha_t | \alpha_{t+1}, y) = -\frac{1}{2}\alpha^\top \bar{\Omega}\alpha + \bar{c}^\top \alpha + \sum_{\tau=1}^{t} \psi_\tau(\alpha_\tau) + k,$$

where k does not depend on $(\alpha_1, \ldots, \alpha_t)$. A first order necessary condition for $(a_{1|t+1}(\alpha_{t+1}), \ldots, a_{t|t+1}(\alpha_{t+1}))$ to be the conditional mode is that the derivative of this log density with respect to α_t be equal to zero there. For $t = 1$, we have the condition

$$-\bar{\Omega}_{11} a_{1|2}(\alpha_2) - \bar{\Omega}_{12}\alpha_2 + \bar{c}_1 + \psi_1'(a_{1|2}(\alpha_2)) = 0, \tag{29}$$

and for $t = 2, \ldots, n-1$, we have

$$-\bar{\Omega}_{t-1,t} a_{t-1|t+1}(\alpha_{t+1}) - \bar{\Omega}_{tt} a_{t|t+1}(\alpha_{t+1}) - \bar{\Omega}_{t,t+1}\alpha_{t+1} + \bar{c}_t + \psi_t'(a_{t|t+1}(\alpha_{t+1})) = 0. \tag{30}$$

Eqs. (29) and (30) feature the function $a_{t|t+1}(\alpha_{t+1})$ whose derivatives we are trying to evaluate at a_{t+1}. Our strategy will be to take multiple derivatives of these two equations, set α_{t+1} to a_{t+1}, and rearrange to obtain linear first order difference equations giving $\dot{a}_t$ through $\dddot{a}_t$ and $\dot{s}_t$ through $\dddot{s}_t$ in terms of $\dot{a}_{t-1}$ through $\dddot{a}_{t-1}$.

I will first establish the identity

$$a_{t-1|t+1}(\alpha_{t+1}) \equiv a_{t-1|t}(a_{t|t+1}(\alpha_{t+1})), \tag{31}$$

which is useful because it expresses a term in Eq. (30) as the composition of two functions whose derivatives we are trying to evaluate. This link between time periods t and $t-1$ turns out to be particularly useful for generating difference equations.

By definition, $(a_{1|t+1}(\alpha_{t+1}), \ldots, a_{t|t+1}(\alpha_{t+1}))$ is the mode of $(\alpha_1, \ldots, \alpha_t)$ given α_{t+1} and y. This implies that $(a_{1|t+1}(\alpha_{t+1}), \ldots, a_{t-1|t+1}(\alpha_{t+1}))$ is the mode of $(\alpha_1, \ldots, \alpha_{t-1})$ given $\alpha_t = a_{t|t+1}(\alpha_{t+1})$ and y. The identity in (31) follows.

I now introduce some notation to make the ensuing derivation more concise and to avoid having to treat $t = 1$ as a special case. Define

$$\Psi_t(\alpha_t) = \begin{cases} \psi_t'(\alpha_t) & t = 1, \\ \psi_t'(\alpha_t) - \bar{\Omega}_{t,t-1} a_{t-1|t}(\alpha_t) & t = 2, \ldots, n. \end{cases}$$

We can now write the first order conditions in Eqs. (29) and (30), for all $t = 1, \ldots, n$, as

$$\Psi_t(a_{t|t+1}(\alpha_{t+1})) + \bar{c}_t = \bar{\Omega}_{t,t+1}\alpha_{t+1} + \bar{\Omega}_{tt} a_{t|t+1}(\alpha_{t+1}). \tag{32}$$

The first derivative of these first order conditions with respect to α_{t+1} gives

$$\frac{d\Psi_t(a_{t|t+1}(\alpha_{t+1}))}{d\alpha_{t+1}} = \bar{\Omega}_{t,t+1} + \bar{\Omega}_{tt} a_{t|t+1}'(\alpha_{t+1}). \tag{33}$$

Higher order derivatives are

$$\frac{d^j \Psi_t(a_{t|t+1}(\alpha_{t+1}))}{d\alpha_{t+1}^j} = \bar{\Omega}_{tt} a_{t|t+1}^{(j)}(\alpha_{t+1}), \quad j = 2, 3, \ldots. \tag{34}$$

I now have the derivatives I wish to evaluate at α_{t+1} on the right hand sides of Eqs. (33) and (34). The left hand sides consist of derivatives of $\Psi_t(a_{t|t+1}(\alpha_{t+1}))$ with respect to α_{t+1}. I now proceed to compute the first four and evaluate them at α_{t+1}.

The first derivative of $\Psi_t(a_{t|t+1}(\alpha_{t+1}))$ is

$$\frac{d\Psi_t(a_{t|t+1}(\alpha_{t+1}))}{d\alpha_{t+1}} = \Psi_t'(a_{t|t+1}(\alpha_{t+1})) a_{t|t+1}'(\alpha_{t+1}). \tag{35}$$

Evaluating (35) at $\alpha_{t+1} = a_{t+1}$, I obtain

$$\frac{d\Psi_t(a_{t|t+1}(a_{t+1}))}{d\alpha_{t+1}} = \Psi_t'(a_t)\dot{a}_t. \tag{36}$$

I do the same for the second through fourth derivatives, using Faà di Bruno's formula for derivatives of composite functions to avoid error and tedium. This gives

$$\frac{d^2 \Psi_t(a_{t|t+1}(a_{t+1}))}{d\alpha_{t+1}^2} = \Psi_t''(a_t)\dot{a}_t^2 + \Psi_t'(a_t)\ddot{a}_t, \tag{37}$$

$$\frac{d^3 \Psi_t(a_{t|t+1}(a_{t+1}))}{d\alpha_{t+1}^3} = \Psi_t'''(a_t)\dot{a}_t^3 + 3\Psi_t''(a_t)\dot{a}_t\ddot{a}_t + \Psi_t'(a_t)\dddot{a}_t, \tag{38}$$

$$\frac{d^4 \Psi_t(a_{t|t+1}(a_{t+1}))}{d\alpha_{t+1}^4} = \Psi_t^{(4)}(a_t)\dot{a}_t^4 + 6\Psi_t'''(a_t)\ddot{a}_t\dot{a}_t^2 + \Psi_t''(a_t)(3\ddot{a}_t^2 + 4\dddot{a}_t\dot{a}_t) + \Psi_t'(a_t)\ddddot{a}_t. \tag{39}$$

I am now ready to give the $\dot{a}_t$ in terms of the Σ_t, which are by-products of the computation of a. I now show by induction that for all $t = 1, \ldots, n-1$,

$$a_{t|t+1}'(\alpha_{t+1}) = -\bar{\bar{\Omega}}_{t,t+1}\Sigma_{t|t+1}(\alpha_{t+1}), \tag{40}$$

from which follows

$$\dot{a}_t = -\bar{\bar{\Omega}}_{t,t+1}\Sigma_t. \tag{41}$$

Using (33) and (35) to eliminate the two left hand sides of these equations, and noting that $\bar{\bar{\Omega}}_{t,t+1} = \bar{\Omega}_{t,t+1}$ we can write

$$\left[\bar{\Omega}_{tt} - \Psi_t'(a_{t|t+1}(\alpha_{t+1}))\right] a_{t|t+1}'(\alpha_{t+1}) = -\bar{\Omega}_{t,t+1}. \tag{42}$$

For $t = 1$, the left hand side of (42) is

$$\left[\bar{\Omega}_{11} - \psi_1''(a_{1|2}(\alpha_2))\right] a_{1|2}'(\alpha_2) = \bar{\bar{\Omega}}_{11}(\alpha_2) a_{1|2}'(\alpha_2) = \Sigma_{1|2}^{-1} a_{1|2}'(\alpha_2).$$

I use (8) and (21) to obtain the first and second equalities. This establishes the result in (40) for $t = 1$.

I now assume that $a_{t-1|t}'(\alpha_t) = -\bar{\bar{\Omega}}_{t-1,t}\Sigma_{t-1|t}(\alpha_t)$ and show that $a_{t|t+1}'(\alpha_{t+1}) = -\bar{\bar{\Omega}}_{t,t+1}\Sigma_{t|t+1}(\alpha_{t+1})$ follows. We can write the left hand side of (42) as

$$\begin{aligned} &\left[\bar{\Omega}_{tt} - \psi_t''(a_{t|t+1}(\alpha_{t+1})) + \bar{\Omega}_{t,t-1} a_{t-1|t}'(a_{t|t+1}(\alpha_{t+1}))\right] a_{t|t+1}'(\alpha_{t+1}) \\ &= \left[\bar{\bar{\Omega}}_{tt} - \bar{\Omega}_{t,t-1}^2 \Sigma_{t-1|t}(a_{t|t+1}(\alpha_{t+1}))\right] a_{t|t+1}'(\alpha_{t+1}) \\ &= \left[\bar{\bar{\Omega}}_{tt} - \bar{\Omega}_{t,t-1}^2 \Sigma_{t-1|t+1}(\alpha_{t+1})\right] a_{t|t+1}'(\alpha_{t+1}) \\ &= \Sigma_{t|t+1}^{-1} a_{t|t+1}'(\alpha_{t+1}). \end{aligned}$$

I use (21) and $a_{t-1|t}'(\alpha_t) = -\bar{\bar{\Omega}}_{t-1,t}\Sigma_{t-1|t}(\alpha_t)$ to obtain the first equation and (8) to obtain the third. By induction, this establishes the result for all t.

I pause to note a simple corollary of (40) which will be useful later. Together, (40) and (42) give

$$\bar{\Omega}_{tt} - \Psi_t'(a_{t|t+1}(\alpha_{t+1})) = \Sigma_{t|t+1}^{-1}(\alpha_{t+1}). \tag{43}$$

We are now ready to compute $\ddot{a}_t$, $\dddot{a}_t$ and $\ddddot{a}_t$. Take (34), with $j = 2$ and $\alpha_{t+1} = a_{t+1}$, and (37) to write

$$(\bar{\Omega}_{tt} - \Psi_t'(a_t))\ddot{a}_t = \Psi_t''(a_t)\dot{a}_t^2.$$

Eq. (43), with $\alpha_{t+1} = a_{t+1}$ gives $\bar{\Omega}_{tt} - \Psi_t'(a_t) = \Sigma_t^{-1}$, which we combine with the previous result to give

$$\ddot{a}_t = \Sigma_t \Psi_t''(a_t)\dot{a}_t^2. \tag{44}$$

In a similar way, but using $j = 3, 4$ in (34), we obtain:

$$\begin{aligned}
\dddot{a}_t &= \Sigma_t[\Psi_t'''(a_t)\dot{a}_t^3 + 3\Psi_t''(a_t)\dot{a}_t\ddot{a}_t], \\
\ddddot{a}_t &= \Sigma_t[\Psi_t^{(4)}(a_t)\dot{a}_t^4 + 6\Psi_t'''(a_t)\ddot{a}_t\dot{a}_t^2 \\
&\quad + \Psi_t''(a_t)(3\ddot{a}_t^2 + 4\dddot{a}_t\dot{a}_t)].
\end{aligned}$$

Simplifying the last two equations yields

$$\dddot{a}_t = \left\{\left[\Sigma_t\Psi_t'''(a_t)\right] + 3\left[\Sigma_t\Psi_t''(a_t)\right]^2\right\}\dot{a}_t^3, \tag{45}$$

$$\begin{aligned}
\ddddot{a}_t &= \left\{\left[\Sigma_t\Psi_t^{(4)}(a_t)\right] + 10\left[\Sigma_t\Psi_t''(a_t)\right]\left[\Sigma_t\Psi_t'''(a_t)\right]\right. \\
&\quad \left. + 15\left[\Sigma_t\Psi_t''(a_t)\right]^3\right\}\dot{a}_t^4.
\end{aligned} \tag{46}$$

The square brackets suggest how to compute these expressions efficiently.

Derivatives $\dot{s}_t$, $\ddot{s}_t$ and $\dddot{s}_t$. Differentiating both sides of (40) with respect to α_{t+1} and using the definition $s_{t|t+1}(\alpha_{t+1}) \doteq \log \Sigma_{t|t+1}(\alpha_{t+1})$ gives

$$\begin{aligned}
a_{t|t+1}''(\alpha_{t+1}) &= -\bar{\Omega}_{t,t+1}\Sigma_{t|t+1}'(\alpha_{t+1}) \\
&= -\bar{\Omega}_{t,t+1}\Sigma_{t|t+1}(\alpha_{t+1})s_{t|t+1}'(\alpha_{t+1}) \\
&= a_{t|t+1}'(\alpha_{t+1})s_{t|t+1}'(\alpha_{t+1}).
\end{aligned} \tag{47}$$

Evaluating the extreme left- and right-hand sides of (47) at $\alpha_{t+1} = a_{t+1}$ yields

$$\ddot{a}_t = \dot{a}_t\dot{s}_t.$$

Using Leibniz' rule for the derivatives of a product to differentiate (47) two times with respect to α_{t+1} and then setting $\alpha_{t+1} = a_{t+1}$ yields

$$\dddot{a}_t = \ddot{a}_t\dot{s}_t + \dot{a}_t\ddot{s}_t, \qquad \ddddot{a}_t = \dddot{a}_t\dot{s}_t + 2\ddot{a}_t\ddot{s}_t + \dot{a}_t\dddot{s}_t.$$

Solving for $\dot{s}_t, \ddot{s}_t$ and $\dddot{s}_t$ then simplifying gives

$$\dot{s}_t = \left[\Sigma_t\Psi_t''(a_t)\right]\dot{a}_t, \tag{48}$$

$$\ddot{s}_t = \left\{\left[\Sigma_t\Psi_t'''(a_t)\right] + 2\left[\Sigma_t\Psi_t''(a_t)\right]^2\right\}\dot{a}_t^2, \tag{49}$$

$$\begin{aligned}
\dddot{s}_t &= \left\{\left[\Sigma_t\Psi_t^{(4)}(a_t)\right] + 7\left[\Sigma_t\Psi_t''(a_t)\right]\left[\Sigma_t\Psi_t'''(a_t)\right]\right. \\
&\quad \left. + \mathbf{8}\left[\Sigma_t\Psi_t''(a_t)\right]^3\right\}\dot{a}_t^3.
\end{aligned} \tag{50}$$

Appendix E. Approximate value and derivatives of $b_{t|t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$

In this appendix, I derive an approximation of the value and first three derivatives of $b_{t|t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$. To simplify notation I will usually suppress the arguments of $a_{t|t+1}(\alpha_{t+1})$ and $b_{t|t+1}(\alpha_{t+1})$ to write $a_{t|t+1}$ and $b_{t|t+1}$.

The case $t = 1$ is trivial since $a_{1|2}(\alpha_2)$ and $b_{1|2}(\alpha_2)$ are the same function by definition. I now consider the case $2 \le t \le n-1$, leaving the case $t = n$ for last. I evaluate $h_t'(\alpha_t; \alpha_{t+1})$, given by (24), at $\alpha_t = a_{t|t+1}(\alpha_{t+1})$ to obtain

$$\begin{aligned}
h_t'(a_{t|t+1}; \alpha_{t+1}) &= -\bar{\Omega}_{t-1,t}\mu_{t-1|t}(a_{t|t+1}) - \bar{\Omega}_{tt}a_{t|t+1} \\
&\quad - \bar{\Omega}_{t,t+1}\alpha_{t+1} + \bar{c}_t + \psi_t'(a_{t|t+1}).
\end{aligned} \tag{51}$$

Subtracting the first order necessary condition for $a_{t|t+1}(\alpha_{t+1})$ in Eq. (30) from Eq. (51) yields

$$h_t'(a_{t|t+1}; \alpha_{t+1}) = -\bar{\Omega}_{t-1,t}\left[\mu_{t-1|t}(a_{t|t+1}) - a_{t-1|t}(a_{t|t+1})\right].$$

Evaluating $h_t''(\alpha_t)$, given by (27), at $\alpha_t = a_{t|t+1}$ gives

$$\begin{aligned}
h_t''(a_{t|t+1}) &= -\bar{\Omega}_{t-1,t}\mu_{t-1|t}'(a_{t|t+1}) - \bar{\Omega}_{tt} + \psi_t''(a_{t|t+1}) \\
&= -(\Sigma_{t|t+1})^{-1} \\
&\quad - \bar{\Omega}_{t-1,t}[\mu_{t-1|t}'(a_{t|t+1}) - a_{t-1|t}'(a_{t|t+1})],
\end{aligned}$$

where I use (43) to obtain the second equality.

I now approximate $b_{t|t+1}(\alpha_{t+1}) - a_{t|t+1}(\alpha_{t+1})$ using a linear approximation of $h_t'(\alpha_t; \alpha_{t+1})$ around $\alpha_{t+1} = a_{t|t+1}(\alpha_{t+1})$:

$$\begin{aligned}
&b_{t|t+1} - a_{t|t+1} \\
&\quad \approx -\frac{h_t'(a_{t|t+1}; \alpha_{t+1})}{h_t''(a_{t|t+1})} \\
&\quad = \frac{-\bar{\Omega}_{t-1,t}\left[\mu_{t-1|t}(a_{t|t+1}) - a_{t-1|t}(a_{t|t+1})\right]}{\exp(-s_{t|t+1}) + \bar{\Omega}_{t-1,t}[\mu_{t-1|t}'(a_{t|t+1}) - a_{t-1|t}'(a_{t|t+1})]}.
\end{aligned}$$

The right hand side of this approximation features the unknown function $\mu_{t-1|t}$. Replacing $\mu_{t-1|t}$ with its polynomial approximation $M_{t-1|t}$ gives the approximation

$$R(\alpha_{t+1}) \doteq N(\alpha_{t+1})/D(\alpha_{t+1}) = N(\alpha_{t+1})V(\alpha_{t+1}),$$

with approximations of the numerator, denominator and reciprocal of the denominator given by

$$\begin{aligned}
N(\alpha_{t+1}) &\doteq -\bar{\Omega}_{t-1,t}\left[M_{t-1|t}(a_{t|t+1}(\alpha_{t+1})) - a_{t-1|t}(a_{t|t+1}(\alpha_{t+1}))\right], \\
D(\alpha_{t+1}) &\doteq \exp(-s_{t|t+1}(\alpha_{t+1})) + \bar{\Omega}_{t-1,t} \\
&\quad \times \left[M_{t-1|t}'(a_{t|t+1}(\alpha_{t+1})) - a_{t-1|t}'(a_{t|t+1}(\alpha_{t+1}))\right], \\
V(\alpha_{t+1}) &\doteq \frac{1}{D(\alpha_{t+1})}.
\end{aligned}$$

Now let $N \doteq N(a_{t+1})$, $\dot{N} \doteq N'(a_{t+1})$, $\ddot{N} = N''(a_{t+1})$, and $\dddot{N} \doteq N'''(a_{t+1})$, and similarly for $D(\alpha_{t+1})$, $V(\alpha_{t+1})$ and $R(\alpha_{t+1})$.

I now approximate the value and three derivatives of $b_{t|t+1}(\alpha_{t+1})$ at α_{t+1} by the corresponding value and derivatives of $a_{t|t+1}(\alpha_{t+1}) + R(\alpha_{t+1})$. I use Leibniz' rule to compute the derivatives of $R(\alpha_{t+1})$ at a_{t+1} in terms of the derivatives of $N(\alpha_{t+1})$ and $V(\alpha_{t+1})$ at α_{t+1}. This gives

$$b_t \approx B_t \doteq a_t + R = a_t + NV, \tag{52}$$

$$\dot{b}_t \approx \dot{B}_t \doteq \dot{a}_t + \dot{R} = \dot{a}_t + \dot{N}V + N\dot{V}, \tag{53}$$

$$\ddot{b}_t \approx \ddot{B}_t \doteq \ddot{a}_t + \ddot{R} = \ddot{a}_t + \ddot{N}V + 2\dot{N}\dot{V} + N\ddot{V}, \tag{54}$$

$$\dddot{b}_t \approx \dddot{B}_t \doteq \dddot{a}_t + \dddot{R} = \dddot{a}_t + \dddot{N}V + 3\ddot{N}\dot{V} + 3\dot{N}\ddot{V} + N\dddot{V}. \tag{55}$$

The quotient rule gives the following derivatives of $V(\alpha_{t+1})$ at a_{t+1}:

$$\begin{aligned}
\dot{V} &= \frac{-\dot{D}}{D^2}, \qquad \ddot{V} = \frac{-\ddot{D}}{D^2} + 2\frac{(\dot{D})^2}{D^3}, \\
\dddot{V} &= \frac{-\dddot{D}}{D^2} + 6\frac{\dot{D}\ddot{D}}{D^3} - 6\frac{(\dot{D})^3}{D^4}.
\end{aligned}$$

Evaluating $N(\alpha_{t+1})$ and its derivatives at a_{t+1} using Faà di Bruno's rule gives

$$\begin{aligned}
N &= -\bar{\Omega}_{t-1,t}(M_{t-1} - a_{t-1}), \\
\dot{N} &= -\bar{\Omega}_{t-1,t}(\dot{M}_{t-1} - \dot{a}_{t-1})\dot{a}_t, \\
\ddot{N} &= -\bar{\Omega}_{t-1,t}[(\dot{M}_{t-1} - \dot{a}_{t-1})\ddot{a}_t + (\ddot{M}_{t-1} - \ddot{a}_{t-1})\dot{a}_t^2], \\
\dddot{N} &= -\bar{\Omega}_{t-1,t}[(\dot{M}_{t-1} - \dot{a}_{t-1})\dddot{a}_t + 3(\ddot{M}_{t-1} - \ddot{a}_{t-1})\dot{a}_t\ddot{a}_t \\
&\quad + (\dddot{M}_{t-1} - \dddot{a}_{t-1})\dot{a}_t^3],
\end{aligned}$$

and doing the same for $D(\alpha_{t+1})$ gives

$$\begin{aligned}
D &= \Sigma_t + \bar{\Omega}_{t-1,t}(\dot{M}_{t-1} - \dot{a}_{t-1}),\\
\dot{D} &= -\dot{s}_t\Sigma_t + \bar{\Omega}_{t-1,t}(\ddot{M}_{t-1} - \ddot{a}_{t-1})\dot{a}_t,\\
\ddot{D} &= (\dot{s}_t^2 - \ddot{s}_t)\Sigma_t + \bar{\Omega}_{t-1,t}[(\ddot{M}_{t-1} - \ddot{a}_{t-1})\ddot{a}_t\\
&\quad + (\dddot{M}_{t-1} - \dddot{a}_{t-1})\dot{a}_t^2],\\
\dddot{D} &= (-\dot{s}_t^3 + 3\dot{s}_t\ddot{s}_t - \dddot{s}_t)\Sigma_t + \bar{\Omega}_{t-1,t}[(\ddot{M}_{t-1} - \ddot{a}_{t-1})\dddot{a}_t\\
&\quad + 3(\dddot{M}_{t-1} - \dddot{a}_{t-1})\dot{a}_t\ddot{a}_t + (\ddddot{M}_{t-1} - \ddddot{a}_{t-1})\dot{a}_t^3].
\end{aligned}$$

We are now ready to compute approximations B_t, $\dot{B}_t$, $\ddot{B}_t$ and $\dddot{B}_t$ for b_t, $\dot{b}_t$, $\ddot{b}_t$ and $\dddot{b}_t$. We first compute N through $\ddot{N}$ and D through $\ddot{D}$, then V through $\ddot{V}$, then B_t through $\ddot{B}_t$ using (52)–(55).

The approximation of $b_{t|t+1}(\alpha_{t+1}) - a_{t|t+1}(\alpha_{t+1})$ above is a third order polynomial. A higher order approximation is not worth the computational cost and I set

$$\ddddot{B}_t = \ddddot{a}_t. \tag{56}$$

I now need to take care of the case $t = n$. Let b_n be the mode of the conditional distribution of α_n given y. This is a number, not a function, and we only need to approximate its value. A similar argument gives

$$b_n - a_n \approx \frac{-\bar{\Omega}_{n-1,n}(\mu_{n-1} - a_{n-1})}{\Sigma_n + \bar{\Omega}_{n-1,n}(\dot{\mu}_{n-1} - \dot{a}_{n-1})}.$$

Analogously, I define

$$B_n \doteq a_n + \frac{-\bar{\Omega}_{n-1,n}(M_{n-1} - a_{n-1})}{\Sigma_n + \bar{\Omega}_{n-1,n}(\dot{M}_{n-1} - \dot{a}_{n-1})}. \tag{57}$$

Appendix F. Approximate value and derivatives of $\mu_t(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$

In this appendix, I derive an approximation of the value and first two derivatives of $\mu_{t|t+1}(\alpha_{t+1}) \doteq E[\alpha_t|\alpha_{t+1}, y]$ at $\alpha_{t+1} = a_{t+1}$. To simply notation, I suppress the argument of $b_{t|t+1}(\alpha_{t+1})$ to write $b_{t|t+1}$.

The third order Taylor series expansion of $\log p(\alpha_t|\alpha_{t+1}, y)$ around $b_{t|t+1}$ gives

$$\begin{aligned}
\log p(\alpha_t|\alpha_{t+1}, y) &\approx k + \frac{1}{2}h_t''(b_{t|t+1})(\alpha_t - b_{t|t+1})^2\\
&\quad + \frac{1}{6}h_t'''(b_{t|t+1})(\alpha_t - b_{t|t+1})^3,
\end{aligned}$$

since $h_t'(b_{t|t+1}; \alpha_{t+1}) = 0$ by the definition of $b_{t|t+1}$. This leads to the approximation

$$\begin{aligned}
p(\alpha_t|\alpha_{t+1}, y) &\approx p_G(\alpha_t; b_{t|t+1}, -1/h_t''(b_{t|t+1}))\\
&\quad \times \exp\left[\frac{1}{6}h_t'''(b_{t|t+1})(\alpha_t - b_{t|t+1})^3\right]\\
&\approx p_G(\alpha_t; b_{t|t+1}, -1/h_t''(b_{t|t+1}))\\
&\quad \times \left[1 + \frac{1}{6}h_t'''(b_{t|t+1})(\alpha_t - b_{t|t+1})^3\right],
\end{aligned}$$

where $p_G(\cdot; \mu, \sigma^2)$ denotes the density of a Gaussian random variable with mean μ and variance σ^2. The second approximation is not a proper density, since it becomes negative, but it does integrate to one because of the odd symmetry of $(\alpha_t - b_{t|t+1})^3$ around $b_{t|t+1}$.

This leads in turn to the approximation of $\mu_{t|t+1} - b_{t|t+1}$ as the integral of $(\alpha_t - b_{t|t+1})$ with respect to this 'density'. Thus

$$\begin{aligned}
\mu_{t|t+1} - b_{t|t+1} &\approx \int p_G(\alpha_t; b_{t|t+1}, -1/h_t''(b_{t|t+1}))\\
&\quad \times \left[1 + \frac{1}{6}h_t'''(b_{t|t+1})(\alpha_t - b_{t|t+1})^3\right]\\
&\quad \times (\alpha_t - b_{t|t+1})\, d\alpha_t\\
&= \frac{1}{6}h_t'''(b_{t|t+1}) \int p_G(\alpha_t; b_{t|t+1}, -1/h_t''(b_{t|t+1}))\\
&\quad \times (\alpha_t - b_{t|t+1})^4\, d\alpha_t\\
&= \frac{1}{6}h_t'''(b_{t|t+1}) \cdot \frac{3}{[-h_t''(b_{t|t+1})]^2}\\
&= \frac{1}{2}\frac{h_t'''(b_{t|t+1})}{[-h_t''(b_{t|t+1})]^2}.
\end{aligned}$$

In the final line, I use the fact that the kurtosis of a Gaussian random variable is equal to 3.

I will now find an equivalent expression for this approximation in terms of $b_{t|t+1}$ and its derivatives. A first order necessary condition for $b_{t|t+1}$ to be the conditional mode of α_t given α_{t+1} and y is $h_t'(b_{t|t+1}; \alpha_{t+1}) = 0$. Evaluating (24) at $b_{t|t+1}$ gives

$$\begin{aligned}
h_t'(b_{t|t+1}; \alpha_{t+1}) &= -\bar{\Omega}_{t-1,t}\mu_{t-1|t}(b_{t|t+1}) - \bar{\Omega}_{tt}b_{t|t+1}\\
&\quad - \bar{\Omega}_{t,t+1}\alpha_{t+1} + \bar{c}_t + \psi_t'(b_{t|t+1}) = 0.
\end{aligned}$$

Taking the derivative with respect to α_{t+1} gives

$$\left[-\bar{\Omega}_{t-1,t}\mu_{t-1|t}'(b_{t|t+1}) - \bar{\Omega}_{tt} + \psi_t''(b_{t|t+1})\right]b_{t|t+1}' - \bar{\Omega}_{t,t+1} = 0.$$

Eq. (27) shows that the term in brackets is $h_t''(b_{t|t+1})$, so

$$h_t''(b_{t|t+1})b_{t|t+1}' = \bar{\Omega}_{t,t+1}. \tag{58}$$

Taking the first derivative of (58) with respect to α_{t+1} gives

$$h_t'''(b_{t|t+1})(b_{t|t+1}')^2 + h_t''(b_{t|t+1})b_{t|t+1}'' = 0.$$

Thus we can write

$$\mu_{t|t+1} - b_{t|t+1} \approx \frac{1}{2}\frac{h_t'''(b_{t|t+1})}{[-h_t''(b_{t|t+1})]^2} = \frac{1}{-2\bar{\Omega}_{t,t+1}}\frac{b_{t|t+1}''}{b_{t|t+1}'}.$$

Replacing the unknown $b_{t|t+1}$ with its polynomial approximation $B_{t|t+1}$ gives

$$\mu_{t|t+1} \approx B_{t|t+1} + \frac{1}{-2\bar{\Omega}_{t,t+1}}\frac{B_{t|t+1}''}{B_{t|t+1}'}.$$

Evaluating this approximation and its first two derivatives at $\alpha_{t+1} = a_{t+1}$ gives the following approximations for μ_t, $\dot{\mu}_t$ and $\ddot{\mu}_t$:

$$\mu_t \approx M_t \doteq B_t + \frac{1}{-2\bar{\Omega}_{t,t+1}} \cdot \frac{\ddot{B}_t}{\dot{B}_t}, \tag{59}$$

$$\dot{\mu}_t \approx \dot{M}_t \doteq \dot{B}_t + \frac{1}{-2\bar{\Omega}_{t,t+1}} \cdot \left[\frac{\dddot{B}_t}{\dot{B}_t} - \left(\frac{\ddot{B}_t}{\dot{B}_t}\right)^2\right], \tag{60}$$

$$\ddot{\mu}_t \approx \ddot{M}_t \doteq \ddot{B}_t + \frac{1}{-2\bar{\Omega}_{t,t+1}} \cdot \left[\frac{\ddddot{B}_t}{\dot{B}_t} - 3\frac{\ddot{B}_t\dddot{B}_t}{\dot{B}_t^2} + 2\left(\frac{\ddot{B}_t}{\dot{B}_t}\right)^3\right]. \tag{61}$$

Going beyond a second order polynomial approximation of $\mu_{t|t+1}(\alpha_{t+1}) - b_{t|t+1}(\alpha_{t+1})$ is not worth the computational cost and I set

$$\dddot{M}_t \doteq \dddot{B}_t, \qquad \ddddot{M}_t \doteq \ddddot{B}_t. \tag{62}$$

Appendix G. A perturbed Gaussian density

I introduce here a perturbed univariate Gaussian density $p_{\text{pert}}(x;\vartheta)$, with parameter vector $\vartheta \doteq (b, h_2, h_3, h_4, h_5, \sigma^2_{\text{tail}}, \pi_{\text{tail}}, \bar{x}, K_1, K_2)$. The location parameter b is the mode. The derivative parameters $h_2 < 0$, h_3, h_4, and h_5 give the second through fifth derivatives of $\log p_{\text{pert}}(x;\vartheta)$ at b. The parameters $\sigma^2_{\text{tail}} > 0$, $\pi_{\text{tail}} \in (0, 1)$ and $\bar{x} > 0$ describe tail behaviour, and are useful when one wants to ensure that $p_{\text{pert}}(x;\vartheta)$ dominates a target density in the tails. The integer order parameters $K_1 \geq 1$ and $K_2 \geq 1$ determine how well $\log p_{\text{pert}}(x;\vartheta)$ approximates $\sum_{i=2}^{5} h_i(x-b)^i/i!$ away from zero. If $h_4 < 0$, then K_2 must be even.

I first define the density, then demonstrate some of its useful properties, then describe how to draw variates. Finally, I discuss the choice of the secondary parameters σ^2_{tail}, π_{tail}, $\bar{x}$, K_1 and K_2. I now suppress notation for conditioning on ϑ and assume $b = 0$. A simple translation gives the density for other values of b.

G.1. Definition of $p_{\text{pert}}(x)$

I define

$$p_{\text{pert}}(x) \doteq \frac{\exp(g(x))}{\cosh(g(x))}\left[(1-\pi_{\text{tail}})p_{\text{main}}(x) + \pi_{\text{tail}}p_{\text{tail}}(x)\right],$$

where the function $g(x)$ and the densities $p_{\text{main}}(x)$ and $p_{\text{tail}}(x)$ are defined below.

I define

$$g(x) \doteq \left[\frac{h_3}{6}\min\left(x^2, \bar{x}^2\right) + \frac{h_5}{120}\min\left(x^4, \bar{x}^4\right)\right]x.$$

The function $g(x)$ is odd, and equal to $h_3x^3/6 + h_5x^5/120$ in the interval $[-\bar{x}, \bar{x}]$ where most of the probability mass lies.

I define $p_{\text{main}}(x)$ as

$$p_{\text{main}}(x) \doteq C^{-1}\exp(h_2x^2/2)\sum_{i=0}^{5K_1+2K_2} c_ix^{2i},$$

where

$$\begin{aligned} C &\doteq \int_{-\infty}^{\infty}\exp(h_2x^2/2)\left[\sum_{i=0}^{5K_1+2K_2} c_ix^{2i}\right]dx \\ &= \sum_{i=0}^{5K_1+2K_2} c_i\left(\frac{2}{-h_2}\right)^{i+\frac{1}{2}}\Gamma\left(i+\frac{1}{2}\right), \end{aligned}$$

and the polynomial coefficients c_i are defined implicitly by the equation

$$\begin{aligned} \sum_{i=0}^{5K_1+2K_2} c_ix^{2i} \doteq \left[\sum_{i=0}^{K_1}\frac{1}{(2i)!}\left(\frac{h_3}{6}x^3 + \frac{h_5}{120}x^5\right)^{2i}\right] \\ \times\left[\sum_{i=0}^{K_2}\frac{1}{i!}\left(\frac{h_4}{24}x^4\right)^i\right]. \end{aligned} \tag{63}$$

The first factor on the right of (63) is a Taylor series expansion of $\cosh(y)$ around $y = 0$, evaluated at $y = h_3x^3/6 + h_5x^5/120$. It is obviously greater than or equal to one. The second factor is a Taylor series expansion of $\exp(y)$ at $y = 0$, evaluated at $y = h_4x^4/24$. If $h_4 > 0$, the second factor is obviously greater than one. If $h_4 \leq 0$, I require K_2 to be even and then the second factor is an even-order expansion. An even order Taylor series expansion of $\exp(y)$ is continuously differentiable, has a limit of ∞ at $-\infty$ and ∞, and is equal to its own derivative plus a positive term. Therefore it is positive at any critical value, and therefore positive everywhere. We conclude that $p_{\text{main}}(x)$ is positive. Note that

$$\begin{aligned} p_{\text{main}}(x) &= C^{-1}\exp(h_2x^2/2 + h_4x^4/24) \\ &\quad\times\cosh(h_3x^3/6 + h_5x^5/120) + o(x^6). \end{aligned}$$

I now define the density $p_{\text{tail}}(x)$ as

$$p_{\text{tail}}(x) \doteq \begin{cases} \dfrac{1}{\sqrt{2\pi\sigma^2_{\text{tail}}}}\dfrac{(|x|-\bar{x})^2}{\sigma^2_{\text{tail}}}\exp\left(-\dfrac{(|x|-\bar{x})^2}{2\sigma^2_{\text{tail}}}\right) & |x| \geq \bar{x}, \\ 0 & \text{otherwise}, \end{cases}$$

which is non-negative and integrates to one.

I now show that $p_{\text{pert}}(x)$ integrates to one. First observe that $g(x)$ is an odd function and $p_{\text{main}}(x)$ and $p_{\text{tail}}(x)$ are both even. Also, $\exp(g(x))/\cosh(g(x)) = 1 + \sinh(g(x))/\cosh(g(x))$, the sum of one and an odd function. The product of an odd and an even function is odd, and odd functions integrate to zero. Therefore $p_{\text{pert}}(x)$ integrates to the same value as $(1-\pi_{\text{tail}})p_{\text{main}}(x)+\pi_{\text{tail}}p_{\text{tail}}(x)$ does, which is one.

I now show that the first five derivatives of $\log p_{\text{pert}}(x)$ at 0 are $0, h_2, h_3, h_4$ and h_5. The odd part of $\log p_{\text{pert}}(x)$ is $g(x)$, so both have the same odd derivatives. The first, third and fifth derivatives of $g(x)$ at 0 are 0, h_3 and h_5. In the interval $[-\bar{x}, \bar{x}]$, the even part of $\log p_{\text{pert}}(x)$ is

$$\begin{aligned} &\log\left[\cosh\left(\frac{h_3}{6}x^3 + \frac{h_5}{120}x^5\right)\exp\left(\frac{h_2}{2}x^2 + \frac{h_4}{24}x^4\right) + o(x^6)\right] \\ &\quad - \log\cosh\left(\frac{h_3}{6}x^3 + \frac{h_5}{120}x^5\right). \end{aligned}$$

Its second and fourth derivatives at $x = 0$ are h_2 and h_4.

The density $p_{\text{pert}}(x)$ is strictly positive, and for $|x| > \bar{x}$, it satisfies

$$p_{\text{pert}}(x) \geq \exp(-2|g(x)|)\pi_{\text{tail}}p_{\text{tail}}(x).$$

Note that $|g(x)|$ is affine for $x > \bar{x}$ and for $x < -\bar{x}$. For all $\sigma^2 < \sigma^2_{\text{tail}}$, there exists a $D > 0$ and an $x^* > 0$ such that for all x such that $|x| > x^*$, $p_{\text{pert}}(x) \geq D\exp(-x^2/2\sigma^2)$. The significance of this is that one can choose σ^2_{tail} so that $p_{\text{pert}}(x)$ dominates a given Gaussian density in the tails. For any μ and $\sigma^2 < \sigma^2_{\text{tail}}$, $(1/\sigma)\phi((x-\mu)/\sigma)/p_{\text{pert}}(x)$ is bounded.

G.2. Drawing variates from $p_{\text{pert}}(x)$

I use reflection sampling, described in Appendix H, to obtain a draw from $p_{\text{pert}}(x)$: I first draw from its even part, $p^e_{\text{pert}}(x) = (1-\pi_{\text{tail}})p_{\text{main}}(x) + \pi_{\text{tail}}p_{\text{tail}}(x)$, then replace x with $-x$ with probability $\max(0, -p^o_{\text{pert}}/p^e_{\text{pert}})$, where $p^o_{\text{pert}}(x)$ is the odd part of $p_{\text{pert}}(x)$. Here, $p^o_{\text{pert}}(x)/p^e_{\text{pert}}(x) = \sinh(g(x))/\cosh(g(x))$.

To draw from $p^e_{\text{pert}}(x)$, I first select which of the mixture components p_{main} and p_{tail} to draw from, with probabilities $(1-\pi_{\text{tail}})$ and π_{tail}.

I use the rejection sampling method described in Gallant and Tauchen (1992) to draw from the component $p_{\text{main}}(x)$. The proposal density is obtained by replacing the c_i in $p_{\text{main}}(x)$ by $\tilde{c}_i \doteq \max(c_i, 0)$ and C by $\tilde{C} \doteq \sum_{i=0}^{5K_1+2K_2}\tilde{c}_i$. The proposal is a mixture, with component probabilities $\tilde{c}_i/\tilde{C}$. A component i with non-zero probability has density $p_i(x) = c_i^{-1}x^{2i}\exp(h_2x^2/2)$, and a simple transformation gives the density of $z \doteq x^2$ as $p_i(z) = c_i^{-1}z^{i-1/2}\exp(h_2z/2)$. This is a Gamma density with shape parameter $i + 1/2$ and rate parameter $-h_2/2$. We draw z, then set x to $\sqrt{z}$ or $-\sqrt{z}$, each with probability 1/2.

I draw from $p_{\text{tail}}(x)$ is a similar way. First draw $z = (|x| - \bar{x})^2$ from a Gamma density with shape parameter 3/2 and rate parameter $1/2\sigma^2_{\text{tail}}$, then set $x = \bar{x} + \sqrt{z}$ or $x = -\bar{x} - \sqrt{z}$, each with probability 1/2.

G.3. Choice of the secondary parameters

The following values work well in all applications. I use $\bar{x} = 5\sigma_x$, where $\sigma_x \doteq (-h_2)^{-1/2}$, and $\pi_{\text{tail}} = 10^{-9}$. I set σ^2_{tail} as described in

Section 2. I choose K_1 and K_2 by adding a new term to an expansion whenever its absolute value at $\bar{x}$ exceeds the constant $\delta = 0.1$, up to a maximum of $K_1 = 2$ and $K_2 = 5$. Precisely,

$$K_1 = \begin{cases} 1 & (h_3\bar{x}^3 + h_5\bar{x}^5)^2 < 24\delta, \\ 2 & \text{otherwise.} \end{cases}$$

$$K_2 = \begin{cases} 1 & 0 \le (h_4\bar{x}^4)^2 < 2\delta, \\ 2 & (h_4\bar{x}^4)^2 \ge 2\delta \text{ and } |h_4\bar{x}^4|^3 < 6\delta, \\ 3 & |h_4\bar{x}^4|^3 \ge 6\delta \text{ and } (h_4\bar{x}^4)^4 < 24\delta, \\ 4 & (h_4\bar{x}^4)^4 \ge 24\delta \text{ and } |h_4\bar{x}^4|^5 < 120\delta, \\ 5 & |h_4\bar{x}^4|^5 \ge 120\delta. \end{cases}$$

I then add one to an odd value of K_2 whenever $h_4 \le 0$.

Appendix H. Reflection sampling

In this appendix I introduce *reflection sampling* and show that it works as described. Let $p(x)$ be any density on $\mathbb{R}^N$ and $p^e(x)$ and $p^o(x)$ be its even and odd parts:

$$p^e(x) \doteq \frac{p(x) + p(-x)}{2}, \qquad p^o(x) \doteq \frac{p(x) - p(-x)}{2}.$$

Note that $p^e(x)$ is a density function.

Reflection sampling is a method for drawing a random variable X from $p(x)$. One first draws Z from $p^e(x)$ and U independently from the uniform distribution on [0, 1], then sets

$$X = \begin{cases} -Z, & U \le \max(-p^o(Z)/p^e(Z), 0), \\ Z, & \text{otherwise.} \end{cases}$$

I now show that this gives a draw from $p(x)$. We can write

$$\begin{aligned} \Pr[X \le x] &= \Pr[X \le x, X = Z] + \Pr[X > -x, X = -Z] \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_N} \int_{\max\left(-\frac{p^o(\xi)}{p^e(\xi)}, 0\right)}^{1} p^e(\xi)\, du\, d\xi \\ &\quad + \int_{-x_1}^{\infty} \cdots \int_{-x_N}^{\infty} \int_{0}^{\max\left(-\frac{p^o(\xi)}{p^e(\xi)}, 0\right)} p^e(\xi)\, du\, d\xi \\ &= \left[\int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_N} 1_A(\xi)(p^e(\xi) + p^o(\xi))\, d\xi \right. \\ &\quad \left. + \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_N} 1_{A^c}(\xi) p^e(\xi)\, d\xi \right] \\ &\quad - \int_{-x_1}^{\infty} \cdots \int_{-x_N}^{\infty} 1_A(\xi) p^o(\xi)\, d\xi, \end{aligned}$$

where $A = \{\xi \in \mathbb{R} : p^o(\xi) < 0\}$, A^c is the complement of A in $\mathbb{R}^N$, and $1_A(x)$ and $1_{A^c}(x)$ are indicator functions for A and A^c. Since $p^o(x)$ is odd, we can write the last term as

$$\begin{aligned} &- \int_{-x_1}^{\infty} \cdots \int_{-x_N}^{\infty} 1_A(\xi) p^o(\xi)\, d\xi \\ &= \int_{-x_1}^{\infty} \cdots \int_{-x_N}^{\infty} 1_A(\xi) p^o(-\xi)\, d\xi \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_N} 1_{A^c}(\xi) p^o(\xi)\, d\xi. \end{aligned}$$

Thus $\Pr[X \le x] = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_N} (p^e(\xi) + p^o(\xi))\, d\xi = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_N} p(\xi)\, d\xi$, the cumulative distribution function corresponding to the density $p(x)$. Therefore X is a draw from $p(x)$.

Appendix I. Model specific derivatives

In this appendix I give, for several state space models, the first five derivatives of $\log p(y_t|\alpha_t)$ with respect to α_t.

I.1. A Gaussian stochastic volatility model

In this model, the conditional distribution of y_t given α_t is Gaussian with variance $\exp(\alpha_t)$. We can write

$$\begin{aligned} p(y_t|\alpha_t) &= \frac{\exp(-\alpha_t/2)}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\exp(-\alpha_t)y_t^2\right], \\ \psi_t(\alpha_t) &\doteq \log p(y_t|\alpha_t) = -\frac{1}{2}[\log 2\pi + \alpha_t + \exp(-\alpha_t)y_t^2], \\ \psi_t'(\alpha_t) &= -\frac{1}{2}[1 - \exp(-\alpha_t)y_t^2], \\ \psi_t''(\alpha_t) &= -\frac{1}{2}\exp(-\alpha_t)y_t^2, \quad \psi_t'''(\alpha_t) = \frac{1}{2}\exp(-\alpha_t)y_t^2, \\ \psi_t^{(4)}(\alpha_t) &= -\frac{1}{2}\exp(-\alpha_t)y_t^2, \qquad \psi_t^{(5)}(\alpha_t) = \frac{1}{2}\exp(-\alpha_t)y_t^2. \end{aligned}$$

I.2. A Student's t stochastic volatility model

Here, the conditional distribution of y_t given α_t is Student's t with mean zero, scale $\exp(\alpha_t)$ and degrees of freedom ν. The conditional density is

$$p(y_t|\alpha_t, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \exp(-\alpha_t/2)\left(1 + \frac{\exp(-\alpha_t)y_t^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

Letting $z_t \doteq \exp(-\alpha_t)y_t^2/\nu$, we can write

$$\begin{aligned} \psi_t(\alpha_t) &\doteq \log p(y_t|\alpha_t, \nu) = \log\left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\right] \\ &\quad - \frac{1}{2}[\alpha_t + (\nu+1)\log(1+z_t)]. \end{aligned}$$

Noting $\partial z_t/\partial\alpha_t = -z_t$, we compute

$$\begin{aligned} \psi_t'(\alpha_t) &= -\frac{1}{2} + \frac{\nu+1}{2}\frac{z_t}{1+z_t}, \qquad \psi_t''(\alpha_t) = -\frac{\nu+1}{2}\frac{z_t}{(1+z_t)^2}, \\ \psi_t'''(\alpha_t) &= \frac{\nu+1}{2}\frac{z_t(1-z_t)}{(1+z_t)^3}, \\ \psi_t^{(4)}(\alpha_t) &= -\frac{\nu+1}{2}\frac{z_t(1-4z_t+z_t^2)}{(1+z_t)^4}, \\ \psi_t^{(5)}(\alpha_t) &= \frac{\nu+1}{2}\frac{z_t(1-11z_t+11z_t^2-z_t^3)}{(1+z_t)^5}. \end{aligned}$$

I.3. A Poisson model

In this model, the conditional distribution of y_t given α_t is Poisson with mean $\exp(\alpha_t)$. Thus

$$\begin{aligned} p(y_t|\alpha_t) &= \frac{\exp(-\exp(\alpha_t))\exp(\alpha_t)^{y_t}}{y_t!}, \\ \psi_t(\alpha_t) &\doteq \log p(y_t|\alpha_t) = -\exp(\alpha_t) + y_t\alpha_t - \log(y_t!), \\ \psi_t'(\alpha_t) &= -\exp(\alpha_t) + y_t, \\ \psi_t''(\alpha_t) &= \psi_t'''(\alpha_t) = \psi_t^{(4)}(\alpha_t) = \psi_t^{(5)}(\alpha_t) = -\exp(\alpha_t). \end{aligned}$$

I.4. A Gamma–Poisson model

Here, the conditional distribution of y_t given α_t is Gamma–Poisson with shape parameter r and mean $r\exp(\alpha_t)$. Thus

$$p(y_t|r,\alpha_t) = \frac{\Gamma(r+y_t)}{y_t!\Gamma(r)}\frac{(\exp(\alpha_t))^{y_t}}{(1+\exp(\alpha_t))^{r+y_t}},$$
$$\psi_t(\alpha_t) \doteq \log p(y_t|r,\alpha_t) = k + y_t\alpha_t - (r+y_t)\log(1+\exp(\alpha_t)),$$
$$\psi_t'(\alpha_t) = y_t - (r+y_t)\frac{\exp(\alpha_t)}{1+\exp(\alpha_t)},$$
$$\psi_t''(\alpha_t) = -(r+y_t)\frac{\exp(\alpha_t)}{(1+\exp(\alpha_t))^2},$$
$$\psi_t'''(\alpha_t) = -(r+y_t)\frac{\exp(\alpha_t)-\exp(2\alpha_t)}{(1+\exp(\alpha_t))^3},$$
$$\psi_t^{(4)}(\alpha_t) = -(r+y_t)\frac{\exp(\alpha_t)-4\exp(2\alpha_t)+\exp(3\alpha_t)}{(1+\exp(\alpha_t))^4},$$
$$\psi_t^{(5)}(\alpha_t) = -(r+y_t)\frac{\exp(\alpha_t)-11\exp(2\alpha_t)+11\exp(3\alpha_t)-\exp(4\alpha_t)}{(1+\exp(\alpha_t))^5}.$$

I.5. An exponential duration model

In this model, the conditional distribution of y_t given α_t is exponential with mean $\exp(\alpha_t)$. Thus

$$p(y_t|\alpha_t) = \exp(-\alpha_t - y_t\exp(-\alpha_t)),$$
$$\psi_t(\alpha_t) \doteq \log p(y_t|\alpha_t) = -\alpha_t - y_t\exp(-\alpha_t),$$
$$\psi_t'(\alpha_t) = -1 + y_t\exp(-\alpha_t),$$
$$\psi_t''(\alpha_t) = -y_t\exp(-\alpha_t), \qquad \psi_t'''(\alpha_t) = y_t\exp(-\alpha_t),$$
$$\psi_t^{(4)}(\alpha_t) = -y_t\exp(-\alpha_t), \qquad \psi_t^{(5)}(\alpha_t) = y_t\exp(-\alpha_t).$$

References

Bauwens, L., Veredas, D., 2004. The stochastic conditional duration model: a latent variable model for the analysis of financial durations. Journal of Econometrics 119, 381–412.

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

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

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

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

Doucet, A., Johansen, A.M., 2010. A tutorial on particle filtering and smoothing: Fifteen years later. In: Crisan, D., Rozovsky, B. (Eds.), Oxford Handbook of Nonlinear Filtering. Oxford University Press, Oxford.

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

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

Engle, R., Russell, J.R., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 66, 1127–1162.

Fahrmeir, L., 1992. Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models. Journal of the American Statistical Association 87, 501–509.

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

Frühwirth-Schnatter, S., Frühwirth, R., 2007. Auxiliary mixture sampling with applications to logistic models. Computational Statistics & Data Analysis 51, 3509–3528.

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

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

Gallant, A.R., Tauchen, G., 1992. A nonparametric approach to nonlinear time series analysis: estimation and simulation. In: Brillinger, D., Caines, P., Geweke, J., Parzen, E., Rosenblatt, M., Taqqu, M.S. (Eds.), New Directions in Time Series Analysis, Part II. Springer, New York, pp. 71–92.

Gamerman, D., 1998. Markov chain Monte Carlo for dynamic generalised linear models. Biometrika 85, 215–227.

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

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

Harvey, A.C., Durbin, J., 1986. The effects of seat belt legislation on British road casualties: a case study in structural time series modelling. Journal of the Royal Statistical Society Series A 149, 187–227.

Harvey, A.C., Fernandes, C., 1989. Time series models for count or qualitative observations. Journal of Business and Economic Statistics 7, 407–417.

Jacquier, E., Polson, N., Rossi, P., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12 (4), 371–388.

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

Jung, R.C., Kukuk, M., Liesenfeld, R., 2006. Time series of count data: modeling, estimation and diagnostics. Computational Statistics & Data Analysis 51, 2350–2364.

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

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

Liesenfeld, R., Nolte, I., Pohlmeier, W., 2006. Modelling financial transaction price movements: a dynamic integer count data model. Empirical Economics 30, 795–825.

Liesenfeld, R., Richard, J., 2003. Univariate and multivariate stochastic volatility models: estimation and diagnostics. Journal of Empirical Finance 10, 505–531.

Liesenfeld, R., Richard, J., 2006. Classical and Bayesian analysis of univariate and multivariate stochastic volatility models. Econometric Reviews 25, 335–360.

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

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

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

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

Pitt, M.K., Shephard, N., 1999. Filtering via simulation: auxiliary particle filters. Journal of the American Statistical Association 94, 590–599.

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

Robert, C.P., 2009. Discussion on the paper by Rue, Martino and Chopin. Journal of the Royal Statistical Society Series B 71, 355–357.

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

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

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

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

Rydberg, T., Shephard, N., 2003. Dynamics of trade-by-trade price movements: decompositions and models. Journal of Financial Econometrics 1, 2–25.

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

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

Strickland, C.M., Forbes, C.S., Martin, G.M., 2008. Parametrisation and efficient MCMC estimation of non-Gaussian state space models. Computational Statistics & Data Analysis 52, 2911–2930.

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

Tsay, R., 2002. Analysis of Financial Time Series. Wiley.

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

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

Zeger, S.-L., 1988. A regression model for time series of counts. Biometrika 75, 621–629.