

POSTERIOR MODE ESTIMATION FOR NONLINEAR AND NON-GAUSSIAN STATE SPACE MODELS

Mike K. P. So

Hong Kong University of Science and Technology

Abstract: In this paper, we develop a posterior mode estimation method for nonlinear and non-Gaussian state space models. By exploiting special structures of the state space models, we derive a modified quadratic hill-climbing procedure which can be implemented efficiently in $O(n)$ operations. The method can be used for estimating the state variable, performing Bayesian inference and carrying out Monte Carlo likelihood inference. Numerical illustrations using simulated and real data demonstrate that our procedure is much more efficient than a common gradient method. It is also evident that our method works very well in a new stochastic volatility model which contains a nonlinear state equation.

Key words and phrases: Filtering, Kalman filter, quadratic hill-climbing, stochastic volatility model, time series.

1. Introduction

Let y_t be a $p \times 1$ observable time series and α_t be a $m \times 1$ state vector. Define also $\alpha = (\alpha'_1, \dots, \alpha'_n)'$ and $y = (y'_1, \dots, y'_n)'$. The general nonlinear and non-Gaussian state space model considered in this paper is

$$p(y_t|\alpha_t) \tag{1}$$

$$\alpha_{t+1} = g_t(\alpha_t) + \eta_t, \quad t = 1, \dots, n, \tag{2}$$

where $\log p(y_t|\alpha_t)$ is twice continuously differentiable and $g_t(\cdot)$ is a known function of α_t which can be time dependent. The specification in (1) allows the modeling of non-Gaussian measurement time series as in Shephard and Pitt (1997). The innovation η_t can also be non-Gaussian. In addition, the Markov dependence in α_t and α_{t+1} defined by $g_t(\cdot)$ encompasses a wide variety of nonlinear structure in α_t . In the following discussion, we assume that

- (A1) Given α , y_t and y_s are independent for all t and s .
- (A2) $p(y_t|\alpha) = p(y_t|\alpha_t)$ for all t .
- (A3) $g_t(\alpha_t)$ is differentiable with respect to α_t .

These assumptions are standard in nonlinear state space models.

An important topic in state space modeling is the estimation of the unknown state variable α . To approach the problem, it seems natural to consider the marginal posterior density $p(\alpha_t|y)$. Then a point estimate can be derived from the marginal posterior mean and marginal posterior mode. In standard Gaussian linear state space models where $p(y_t|\alpha_t)$ is Gaussian and $g_t(\alpha_t)$ is a linear function of α_t , $p(\alpha_t|y)$ is also Gaussian with mean and variance that can be computed easily by the Kalman filter. However, in the general nonlinear and non-Gaussian model written in (1) and (2), computing $p(\alpha_t|y)$ can be very difficult as it involves high-dimension integration. Kitagawa (1987) introduced a numerical method to evaluate the marginal density but it is still very computer intensive when the dimension is high. Therefore, it is not computationally efficient to find the posterior mean and mode directly from $p(\alpha_t|y)$ when n is large; see Fahrmeir (1992, p.501) and Harvey (1989, p.165).

An alternative way of extracting the state variable α from the observed data y is to consider $p(\alpha|y)$. The estimate obtained by maximizing $p(\alpha|y)$ with respect to α is usually called *the maximum a posteriori estimate*; see Anderson and Moore (1979, p.26) and Sage and Melsa (1971, p.441). Fahrmeir (1992) demonstrated that optimizing $p(\alpha|y)$ by a scoring method works well in exponential family distributions with linear transition equation even though n is large. Besides the state estimation, the posterior mode searching from $p(\alpha|y)$ happens to be a crucial step in performing the posterior sampling by Markov chain Monte Carlo methods (Shephard and Pitt (1997)) and in approximating the likelihood function (Durbin and Koopman (1997), (2000)). In the above two applications, the posterior mode serves as a good reference point to construct an artificial model for sampling in a Metropolis-Hasting step and for computing the likelihood function by importance sampling. Obviously, from the posterior sample generated by the Markov chain Monte Carlo methods, one can also estimate α by the posterior mean. In summary, the posterior mode from $p(\alpha|y)$, which facilitates the Bayesian and likelihood inference, is important for state space modeling.

In the majority of results that are related to posterior mode searching in state space models, the existence of a linear transition equation is assumed for α_t (Fahrmeir (1992); Durbin and Koopman (1997), (2000); Shephard and Pitt (1997)). Provided $l''(\alpha_t)$ is negative semi-definite, where $l(\alpha_t) = \log p(y_t|\alpha_t)$, the posterior mode can be evaluated easily by Newton's method via the Kalman filter. The condition holds if $p(y_t|\alpha_t)$ is from the exponential family. However, the specification in (1) does not guarantee that $l''(\alpha_t)$ is negative semi-definite. Furthermore, (2) may not define a linear transition equation. In this paper, we introduce a new method to maximize $p(\alpha|y)$ for α under the nonlinear and non-Gaussian state space model in (1) and (2) via the Kalman filter. Our method

can be applied even though $l''(\alpha_t)$ is not always negative definite or $g_t(\alpha_t)$ is nonlinear. More importantly, the new method is very efficient as compared with common gradient methods because it can be implemented with $O(n)$ operations. Our proposed method is very useful, especially when we want to estimate the states with known or estimated parameter values, to carry out the posterior sampling as in Shephard and Pitt (1997) and to perform the likelihood inference as in Durbin and Koopman (1997, 2000).

The rest of this paper is as follows. Section 2 motivates use of quadratic hill-climbing for finding the posterior mode. Section 3 reviews the properties and the main features of the quadratic hill-climbing method. In Section 4, the main results of the paper are presented. We discuss how to apply the Kalman filter for implementing the quadratic hill-climbing method, two modifications for computational enhancement, ways to handle nonlinear transition equation and our modified quadratic hill-climbing method for nonlinear models. We consider some extensions in Section 5. Simulated examples for examining the performance of our method are given in Section 6. Section 7 presents a real data illustration. An appendix shows how the quadratic hill-climbing method can be implemented via the Kalman filter.

2. Problems in Existing Methods

The estimation of posterior mode is to maximize $\log p(\alpha|y)$, or

$$q(\alpha) = \log p(\alpha, y) = \log p(\alpha_1) + \sum_{t=1}^{n-1} \log p(\alpha_{t+1}|\alpha_t) + \sum_{t=1}^n l(\alpha_t) \quad (3)$$

with respect to α , where $l(\alpha_t) = \log p(y_t|\alpha_t)$. This problem has long been studied for the dynamic model

$$p(y_t|\alpha_t), \quad \alpha_{t+1} = c_t + T_t\alpha_t + \eta_t, \quad \eta_t \sim N(0, Q_t). \quad (4)$$

For example, Fahrmeir (1992) focused on observations generated from exponential family distributions. Recently, Shephard and Pitt (1997) and Durbin and Koopman ((1997), (2000)) have considered this problem for Bayesian inference and Monte Carlo likelihood evaluation. Most of the existing works applied Newton's method with the typical iterative step

$$\alpha^{(1)} = \alpha^{(0)} - q''(\alpha^{(0)})^{-1} q'(\alpha^{(0)}), \quad (5)$$

where $\alpha^{(0)}$ and $\alpha^{(1)}$ are the current and the updated iterates of α , q' and q'' are the first and second derivatives of q with respect to α respectively. The basic

idea of Newton's method is to find the maximum of a quadratic approximation of $q(\alpha)$ around $\alpha^{(0)}$. As shown in the Appendix,

$$q''(\alpha) = \begin{pmatrix} l''(\alpha_1) & & 0 \\ & \ddots & \\ 0 & & l''(\alpha_n) \end{pmatrix} - \Gamma^{-1}$$

under (4), where $\Gamma = \text{cov}(\alpha)$. To use Newton's method, it is required to have $q''(\alpha)$ negative definite. This is satisfied when all the second derivatives $l''(\alpha_t)$ are negative definite. Otherwise, there is no guarantee that the quadratic approximation to $q(\alpha)$ has a maximum.

When the negative definiteness of $l''(\alpha_t)$ does not hold, Durbin and Koopman (2000, p.13) suggest using an iterative method, called method 2 here, for posterior mode searching. However, we have to assume the following linear measurement equation:

$$y_t = Z_t \alpha_t + \epsilon_t, \quad \epsilon_t \sim p(\epsilon_t), \quad -\frac{1}{\epsilon_t} \frac{\partial \log p(\epsilon_t)}{\partial \epsilon_t} > 0. \quad (6)$$

Although we can apply Newton's method or method 2 to a variety of models, there exist many unsolved cases.

An example where all the existing methods fail is

$$y_t = \alpha_t^2 + \epsilon_t, \quad \epsilon_t \sim N(0, 1). \quad (7)$$

Similar measurement equations were considered in Alspach and Sorenson (1972) and Carlin, Polson and Stoffer (1992). The former discussed a quadratic scalar example and the latter studied an nonstationary growth model. It can be shown easily that $l''(\alpha_t) = 2y_t - 6\alpha_t^2$, which is not necessarily negative. Moreover, (7) cannot be written in the linear form (6). Hence, the posterior mode estimation problem cannot be dealt by Newton's method or method 2.

Another interesting example is the stochastic volatility in the mean model:

$$y_t | \alpha_t \sim N\left(\delta \exp\left(\frac{\alpha_t}{2}\right), \exp(\alpha_t)\right), \quad \alpha_{t+1} = c + \phi \alpha_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2). \quad (8)$$

It is an important model for financial market volatility. Its basic version, that is when $\delta = 0$, has been studied widely in the literature (Harvey, Ruiz and Shephard (1994); Kim, Shephard and Chib (1998)). The additional conditional mean component $\delta \exp(\frac{\alpha_t}{2})$ is used to capture the potential risk premium (Engle, Lilien and Robins (1987)). In this case, $l''(\alpha_t) = -\frac{y_t}{2\beta_t}(y_t - \frac{1}{2}\mu_t)$, where $\mu_t = \delta \exp(\frac{\alpha_t}{2})$ and $\beta_t = \exp(\alpha_t)$ are the conditional mean and variance of y_t respectively. Since $l''(\alpha_t)$ is not always negative and the conditional mean in (8) is not a linear function of α_t , Newton's method and method 2 are not applicable.

Here we introduce the modified quadratic hill-climbing method to solve the posterior mode estimation problem in these alternative situations. More importantly, the hill-climbing method admits extension to the nonlinear transition as in (2): $\alpha_{t+1} = g(\alpha_t) + \eta_t$. To the best of our knowledge, this is the first time the posterior mode estimation is considered in nonlinear state space models.

3. Review of Quadratic Hill-Climbing Method

To solve the non-negativity problem of $q''(\alpha)$ encountered in Newton's method, we propose using Goldfeld, Quandt and Trotter (1966)'s quadratic hill-climbing method. This optimization algorithm has two attractive features. It does not require $q''(\alpha)$ to be everywhere negative definite. Furthermore, the step size in each iteration is controlled in such a way as to yield a good quadratic approximation of the target function $q(\alpha)$ in a neighborhood of the current iterate $\alpha^{(0)}$. Let δ be a constant that $q''(\alpha^{(0)}) - \delta I$ is negative definite and S_δ be the region consisting of α with $\|\alpha - \alpha^{(0)}\| \leq \|(q''(\alpha^{(0)}) - \delta I)^{-1} q'(\alpha^{(0)})\|$, where $\|\cdot\|$ is the standard Euclidean norm. According to the theorem in Goldfeld, Quandt and Trotter (1966, p.545), the second order expansion of $q(\alpha)$ at $\alpha^{(0)}$,

$$q(\alpha^{(0)}) + (\alpha - \alpha^{(0)})' q'(\alpha^{(0)}) + \frac{1}{2} (\alpha - \alpha^{(0)})' q''(\alpha^{(0)}) (\alpha - \alpha^{(0)}), \tag{9}$$

will attain the maximum in S_δ at $\alpha^{(0)} - (q''(\alpha^{(0)}) - \delta I)^{-1} q'(\alpha^{(0)})$ if $\delta \geq 0$, and at $\alpha^{(0)} - q''(\alpha^{(0)})^{-1} q'(\alpha^{(0)})$ if $\delta < 0$. In view of (9), consider the following iterative scheme:

- (a) Initialize a positive parameter R at a suitable value.
- (b) Calculate $\delta = \lambda + R \|q'(\alpha^{(0)})\|$ and

$$\alpha^{(1)} = \begin{cases} \alpha^{(0)} - (q''(\alpha^{(0)}) - \delta I)^{-1} q'(\alpha^{(0)}), & \delta \geq 0, \\ \alpha^{(0)} - q''(\alpha^{(0)})^{-1} q'(\alpha^{(0)}), & \delta < 0. \end{cases} \tag{10}$$

- (c) Check to see whether $q(\alpha^{(1)}) > q(\alpha^{(0)})$.
- (d) If yes, then stop and one iteration is completed. If not, increase R , say by setting $R = 2R$, and goto (b).

The parameter λ is taken as the largest eigenvalue of $q''(\alpha^{(0)})$ so that $q''(\alpha^{(0)}) - \delta I$ is negative definite. The updating scheme in (10) defines $\alpha^{(1)}$ as the maximum of the quadratic approximation of $q(\alpha)$ in S_δ with center $\alpha^{(0)}$ and radius $\|(q''(\alpha^{(0)}) - \delta I)^{-1} q'(\alpha^{(0)})\|$. Goldfeld, Quandt and Trotter (1966) showed that the upper bound of the radius $\|(q''(\alpha^{(0)}) - \delta I)^{-1} q'(\alpha^{(0)})\|$ is R^{-1} . Therefore, one can limit the step size by controlling R . At each iteration, a positive R is selected. If the quadratic approximation of $q(\alpha)$ within S_δ is good, it is likely that $\alpha^{(1)}$ determined by (10) is close to the maximum of $q(\alpha)$ in the same region,

thus producing a functional increment in step (c). Otherwise, the value of R is gradually increased in step (d) to reduce the radius of the region until a move to $\alpha^{(1)}$ with $q(\alpha^{(1)}) > q(\alpha^{(0)})$ is generated. In cases where $q(\alpha)$ is concave, the iterative scheme in (10) reduces to Newton's method.

Although the hill-climbing approach can be applied to any conditional distribution $p(y_t|\alpha_t)$, the crude application of (10), which involves the inversion of $q''(\alpha^{(0)}) - \delta I$, is highly inefficient as $\alpha^{(0)}$ is typically of high dimension. A fast algorithm is needed.

4. Main Results

We begin this section with a fast algorithm for applying the quadratic hill-climbing method. We assume here a linear transition equation for α_t and discuss some ways to handle the nonlinear transition equation in Section 4.2. In other words, the assumed underlying model is that in (4). The typical step of quadratic hill-climbing involves the updating of $\alpha^{(0)}$ to $\alpha^{(1)}$ from $\alpha^{(1)} = \alpha^{(0)} - (q''(\alpha^{(0)}) - \delta I)^{-1} q'(\alpha^{(0)})$. Computing this quantity directly using standard matrix inversion methods is computer intensive when the sample size n is large. To facilitate fast calculation of the posterior mode, which is important in block sampling (Shephard and Pitt (1997)) and Monte Carlo likelihood evaluation (Durbin and Koopman (1997), (2000)), artificial observations $\hat{y}_t = \alpha_t^{(0)} + H_t l'(\alpha_t^{(0)})$ and noise variance $H_t = -(l''(\alpha_t^{(0)}) - \delta I)^{-1}$ are formed. To get $\alpha^{(1)}$ from $\alpha^{(0)} - (q''(\alpha^{(0)}) - \delta I)^{-1} q'(\alpha^{(0)})$, we propose using the following Gaussian state space model:

$$\hat{y}_t = \alpha_t + \epsilon_t, \quad \epsilon_t \sim N(0, H_t), \quad \alpha_{t+1} = c_t + T_t \alpha_t + \eta_t, \quad \eta_t \sim N(0, Q_t).$$

The desired iterate $\alpha^{(1)}$ can then be obtained as the Kalman filter smoothed value of α under the above model. A proof for the equivalence of $\alpha^{(1)}$ and the smoothed value is given in the Appendix. It should be noted that updating $\alpha^{(0)}$ to $\alpha^{(1)}$ with the use of the Kalman filter is very important for estimating the posterior mode in $O(n)$ operations. Otherwise, we need to invert the $mn \times mn$ matrix $q''(\alpha^{(0)}) - \delta I$ by some standard matrix inversion algorithms which are typically of order $O(n^2)$ or even $O(n^3)$. Clearly, standard methods that do not exploit the special dynamic structure of $\{y_t\}$ and $\{\alpha_t\}$ are much less efficient.

4.1. Two modifications

We introduce two modifications of quadratic hill-climbing procedure to enhance computational efficiency. The first modification is about the selection of λ . Goldfeld, Quandt and Trotter (1966) suggested setting λ at the largest eigenvalue of $q''(\alpha^{(0)})$. Evaluating the largest eigenvalue is likely very time consuming because $q''(\alpha^{(0)})$ is of dimension mn . Moreover, evaluation involves the inverse

of the $mn \times mn$ matrix Γ . Therefore, this choice of λ may result in a highly computer-intensive algorithm. According to Lemma 1 of Goldfeld, Quandt and Trotter (1966), the purpose of selecting λ as the largest eigenvalue is to guarantee that $q''(\alpha^{(0)}) - \delta I$ is negative definite. To avoid dealing with eigenvalues of high dimension matrices, we propose another choice of λ that serves the same purpose. Let β_1, \dots, β_n be the largest eigenvalue of the $m \times m$ matrices $l''(\alpha_1), \dots, l''(\alpha_n)$ respectively. With $\lambda = \max\{\beta_1, \dots, \beta_n\}$,

$$q''(\alpha) - \delta I = \begin{pmatrix} l''(\alpha_1) - \lambda I & & 0 \\ & \ddots & \\ 0 & & l''(\alpha_n) - \lambda I \end{pmatrix} - \Gamma^{-1} - R\|q'(\alpha^{(0)})\|I$$

is negative definite. Computing β_t is easy because $l''(\alpha_t)$ is $m \times m$ only. In terms of computation time, the suggested method is obviously preferable to Goldfeld, Quandt and Trotter (1966)'s approach, especially when n is large.

The second modification is about the value of δ . In the original proposal of Goldfeld, Quandt and Trotter (1966), $\delta = \lambda + R\|q'(\alpha^{(0)})\|$ and so we have to compute the norm $\|q'(\alpha^{(0)})\|$ at each iteration. Again, this is not desirable because evaluating the norm of $q'(\alpha)$ involves Γ^{-1} . We propose setting $\delta = \lambda + R$ rather than $\delta = \lambda + R\|q'(\alpha^{(0)})\|$. It can be shown that the maximum of the radius $\|(q''(\alpha^{(0)}) - \delta I)^{-1}q'(\alpha^{(0)})\|$ is $R^{-1}\|q'(\alpha^{(0)})\|$ and we have bypassed the calculation of $\|q'(\alpha^{(0)})\|$. In practice, we need to select the starting value of R carefully. If R is initialized at a 'large' value, say 1, the maximum radius will be $\|q'(\alpha^{(0)})\|$. In this case, the iteration can be easily trapped in a local optimum because when $\alpha^{(0)}$ is close to a local optimum, $\|q'(\alpha^{(0)})\|$ will be small, implying that the next move to $\alpha^{(1)}$, limited by the maximum distance $\|q'(\alpha^{(0)})\|$, will be a small step. Convergence might then be to a local optimum. On the other hand, if R is started up with a 'small' value, say 10^{-10} , it may take many trials to increase R so as to have an increase in $q(\alpha)$. Our experience is that an initial value of $R = 0.001$, implying the maximum radius of $1000 \|q'(\alpha^{(0)})\|$, usually does a good job.

4.2. Nonlinear transition with non-Gaussian errors

So far we have assumed that a transition equation exists for $\{\alpha_t\}$. In the general specification stated in (2), $g_t(\alpha_t)$ can be a nonlinear function of α_t and η_t can be non-Gaussian. There are standard approaches for estimating the state variable α in nonlinear and non-Gaussian state space models. Two well-known examples are the extended Kalman filter and the Gaussian sum filter; see Andersen and Moore (1979) and Harvey (1989). The former applies the first order approximation to $g_t(\alpha_t)$, while the latter uses sums of Gaussian densities to approximate the filtering density $p(\alpha_t|y_1, \dots, y_t)$. We adopt another approach

for nonlinear filtering, a posterior mode estimation method under the nonlinear model in (1) and (2) that can be used for filtering the state variable, implementing the block sampling of Shephard and Pitt (1997) and carrying out the Monte Carlo likelihood inference in Durbin and Koopman (1997, 2000).

To deal with the possible nonlinear transition, we propose linearizing $g_t(\alpha_t)$ by expanding it about a point $\hat{\alpha}_t$ using the first-order Taylor expansion, $g_t(\alpha_t) \approx g_t(\hat{\alpha}_t) + \left. \frac{\partial g_t(\alpha_t)}{\partial \alpha_t} \right|_{\alpha_t = \hat{\alpha}_t} (\alpha_t - \hat{\alpha}_t)$. If η_t is non-Gaussian, we suggest the second order approximation to $l_{\eta_t}(x) = \log p_{\eta_t}(x)$, where p_{η_t} denotes the distribution of η_t . In other words, p_{η_t} is approximated by the normal distribution with mean $\hat{\eta}_t - l''_{\eta_t}(\hat{\eta}_t)^{-1} l'_{\eta_t}(\hat{\eta}_t)$ and variance $-l''_{\eta_t}(\hat{\eta}_t)^{-1}$, where $\hat{\eta}_t$ is the point of expansion, l'_{η_t} and l''_{η_t} are the first and second derivatives of l_{η_t} with respect to η_t . If $l''_{\eta_t}(\hat{\eta}_t)$ is not negative definite, we can use $-l'_{\eta_t}(\hat{\eta}_t) l'_{\eta_t}(\hat{\eta}_t)'$ instead. The rationale is that $E(-l''_{\eta_t}(\hat{\eta}_t)) = E(l'_{\eta_t}(\hat{\eta}_t) l'_{\eta_t}(\hat{\eta}_t)')$ under some regularity assumptions on p_{η_t} . In summary, we approximate the model in (1) and (2) by a model in (4) with

$$\begin{aligned} T_t &= \left. \frac{\partial g_t(\alpha_t)}{\partial \alpha_t} \right|_{\alpha_t = \hat{\alpha}_t}, \\ Q_t &= \begin{cases} -l''_{\eta_t}(\hat{\eta}_t)^{-1} & \text{if } l''_{\eta_t}(\hat{\eta}_t) \text{ is n.d.} \\ (l'_{\eta_t}(\hat{\eta}_t) l'_{\eta_t}(\hat{\eta}_t)')^{-1} & \text{otherwise,} \end{cases} \quad (11) \\ c_t &= g_t(\hat{\alpha}_t) - T_t \hat{\alpha}_t + \hat{\eta}_t + Q_t l'_{\eta_t}(\hat{\eta}_t), \end{aligned}$$

so that we can use the Kalman filter smoother of the Appendix.

To apply the quadratic hill-climbing procedure, we choose the points of expansion as $\hat{\alpha}_t = \alpha_t^{(0)}$ and $\hat{\eta}_t = \alpha_{t+1}^{(0)} - g_t(\alpha_t^{(0)})$. The main purpose of constructing the linear model in (11) is to approximate $p(\alpha_{t+1}|\alpha_t)$ locally around $\alpha^{(0)}$ as the log-posterior in (3) involves $p(\alpha_{t+1}|\alpha_t)$. By increasing R to a certain level that the radius of the region S_δ is sufficiently small, we expect that (11) is a good proxy for the model in (1) and (2) in the sense that if α lies within a small S_δ , the conditional distribution of α_{t+1} given α_t under the model in (1) and (2) satisfies

$$\begin{aligned} & p(\alpha_{t+1}|\alpha_t) \\ &= p_{\eta_t}(\alpha_{t+1} - g_t(\alpha_t)) \approx p_{\eta_t}(\alpha_{t+1}^{(0)} - g_t(\alpha_t^{(0)})) \\ &\propto \exp\left(-\frac{1}{2}(\alpha_{t+1}^{(0)} - g_t(\alpha_t^{(0)}) - \hat{\eta}_t - Q_t l'_{\eta_t}(\hat{\eta}_t))' Q_t^{-1} (\alpha_{t+1}^{(0)} - g_t(\alpha_t^{(0)}) - \hat{\eta}_t - Q_t l'_{\eta_t}(\hat{\eta}_t))\right) \\ &= \exp\left(-\frac{1}{2}(\alpha_{t+1}^{(0)} - c_t - T_t \alpha_t^{(0)})' Q_t^{-1} (\alpha_{t+1}^{(0)} - c_t - T_t \alpha_t^{(0)})\right) \\ &\approx \exp\left(-\frac{1}{2}(\alpha_{t+1} - c_t - T_t \alpha_t)' Q_t^{-1} (\alpha_{t+1} - c_t - T_t \alpha_t)\right). \end{aligned}$$

This is, ignoring a multiplicative constant, the distribution of α_{t+1} conditional on α_t under the model in (11). The approximations hold because α is close to $\alpha^{(0)}$; The proportionality follows as $\alpha_{t+1}^{(0)} - g_t(\alpha_t^{(0)})$ is the point of expansion of the quadratic approximation to l_{η_t} ; The last equality is true because we expand $g_t(\alpha_t)$ about $\alpha_t^{(0)}$. According to the main theorem of Goldfeld, Quandt and Trotter (1966, p.545), for fixed R in (10) the $\alpha^{(1)}$ obtained from applying the quadratic hill-climbing with the transition equation in (11) gives the maximum of the quadratic approximation to $q(\alpha)$ within the region S_δ under the model in (11). Hence, a functional increment, that is $q(\alpha^{(1)}) > q(\alpha^{(0)})$, is expected if the radius of S_δ is small enough or R is large enough.

4.3. A modified quadratic hill-climbing method

The modified quadratic hill-climbing method proposed here for posterior mode estimation under the nonlinear state space model in (1) and (2) is summarized below.

- (a) Initialize R at a suitable value, such as 0.001.
- (b) Set $\delta = \lambda + R$ with λ as modified, calculate $\alpha^{(1)}$ as the Kalman filter smoothed values based on the state space model

$$\hat{y}_t = \alpha_t + \epsilon_t, \quad \epsilon_t \sim N(0, H_t), \quad \alpha_{t+1} = c_t + T_t \alpha_t + \eta_t, \quad \eta_t \sim N(0, Q_t),$$

where

$$\begin{aligned}
 \hat{y}_t &= \alpha_t^{(0)} + H_t l'(\alpha_t^{(0)}), \quad H_t = \begin{cases} -(l''(\alpha_t^{(0)}) - \delta I)^{-1}, & \delta \geq 0, \\ -l''(\alpha_t^{(0)})^{-1}, & \delta < 0, \end{cases} \\
 T_t &= \left. \frac{\partial g_t(\alpha_t)}{\partial \alpha_t} \right|_{\alpha_t = \alpha_t^{(0)}}, \\
 Q_t &= \begin{cases} -l''_{\eta_t}(\hat{\eta}_t)^{-1} & \text{if } l''_{\eta_t}(\hat{\eta}_t) \text{ is n.d.} \\ (l'_{\eta_t}(\hat{\eta}_t) l'_{\eta_t}(\hat{\eta}_t)')^{-1} & \text{otherwise,} \end{cases} \\
 c_t &= g_t(\alpha_t^{(0)}) - T_t \alpha_t^{(0)} + \hat{\eta}_t + Q_t l'_{\eta_t}(\hat{\eta}_t), \\
 \hat{\eta}_t &= \alpha_{t+1}^{(0)} - g_t(\alpha_t^{(0)}).
 \end{aligned} \tag{12}$$

- (c) Check to see whether $q(\alpha^{(1)}) > q(\alpha^{(0)})$.
- (d) If yes, then stop and one iteration is completed. If not, increase R , say by setting $R = 2R$, and goto (b).

A nice feature of the above hill-climbing method is that the step size δ is self-adjusted according to the quality of the quadratic approximation: when the approximation is good, the step size can be larger; when the approximation is

poor in high dimensional problems, the step size is automatically tuned down to cater for an increase in $q(\alpha)$.

Comparing with our proposed method in (12), the extended Kalman filter also applies the linearization of $g_t(\alpha_t)$, but the point of expansion is the filtered α_t rather than the previous smoothed estimate $\alpha_t^{(0)}$ representing the mode of the posterior density $p(\alpha|y)$. While both methods can be used for estimating α , the extended Kalman filter is not designed to calculate the posterior mode which is found to be important for Bayesian and likelihood inference of state space models. In contrast to (12), the extended Kalman filter does not require iteration. However, its performance in nonlinear filtering can be seriously affected by the precision of the filtered α_t as a predicted value of α_t . In practice, we can combine the extended Kalman filter and modified quadratic hill-climbing as follows:

- (a) Apply the extended Kalman filter to generate $\alpha^{(0)}$.
- (b) Iterate (12) once to obtain $\alpha^{(1)}$.
- (c) Use $\alpha^{(1)}$ as the point of expansion to form the linear approximated model for applying the extended Kalman filter and one iteration is completed.

Step (c) is similar to the iterated extended Kalman filter (Anderson and Moore (1979), p.204) in that the points of expansion are revised in each iteration. As far as state estimation is concerned, the combined method may perform better than the extended Kalman filter in some applications.

5. Extensions

The previous methodology for posterior mode estimation can also be applied to the more general model:

$$p(y_t|\alpha_t), \quad \alpha_{t+1} = f_t(\alpha_t, \xi_t), \quad t = 1, \dots, n,$$

where ξ_t is a noise in the state equation. For example, in multiplicative models $f_t(\alpha_t, \xi_t) = g_t(\alpha_t)\xi_t$. By linearizing $f_t(\alpha_t, \xi_t)$ about $(\hat{\alpha}_t, \hat{\xi}_t)$,

$$\begin{aligned} & f_t(\alpha_t, \xi_t) \\ \approx & f_t(\hat{\alpha}_t, \hat{\xi}_t) + \frac{\partial f_t(\alpha_t, \xi_t)}{\partial \alpha_t} \Big|_{(\alpha_t, \xi_t) = (\hat{\alpha}_t, \hat{\xi}_t)} (\alpha_t - \hat{\alpha}_t) + \frac{\partial f_t(\alpha_t, \xi_t)}{\partial \xi_t} \Big|_{(\alpha_t, \xi_t) = (\hat{\alpha}_t, \hat{\xi}_t)} (\xi_t - \hat{\xi}_t) \\ = & g_t(\alpha_t) + \eta_t, \end{aligned} \quad (13)$$

$$\begin{aligned} g_t(\alpha_t) &= f_t(\hat{\alpha}_t, \hat{\xi}_t) + T_t(\alpha_t - \hat{\alpha}_t) - R_t \hat{\xi}_t, & \eta_t &= R_t \xi_t \\ T_t &= \frac{\partial f_t(\alpha_t, \xi_t)}{\partial \alpha_t} \Big|_{(\alpha_t, \xi_t) = (\hat{\alpha}_t, \hat{\xi}_t)}, & R_t &= \frac{\partial f_t(\alpha_t, \xi_t)}{\partial \xi_t} \Big|_{(\alpha_t, \xi_t) = (\hat{\alpha}_t, \hat{\xi}_t)}, \end{aligned}$$

we have an additive model form as in (1) and (2). Then the approximated state space model adopted in (12) is constructed based on (13) and the modified quadratic hill-climbing method can be used accordingly.

6. Simulation Studies

6.1. An additive model example

We illustrate the computational advantage of the modified quadratic hill-climbing in (12) by a numerical example. In this experiment, we simulated from the nonlinear state space model

$$y_t = 0.5(\alpha_t + 40) - \frac{10(\alpha_t + 40)}{1 + (\alpha_t + 40)^2} + \epsilon_t, \quad \epsilon_t \sim N(0, 1),$$

$$\alpha_{t+1} = 0.9\alpha_t - \frac{10\alpha_t}{1 + \alpha_t^2} - 0.1t + \eta_t, \quad \eta_t \sim N(0, 1),$$

in Andrade Netto, Gimeno and Mendes (1978). It has the additive model specification as in (2) that α_{t+1} is a function of α_t with an additive noise. To maximize $\log p(\alpha|y)$ with respect to α , we may apply some common gradient methods such as Newton’s method and the quadratic hill-climbing in Goldfeld, Quandt and Trotter (1966). Alternatively, we can use the modified quadratic hill-climbing method in (12) which is applicable to the nonlinear state space model in (1) and (2). In our experiment, we simulated time series of length $n = 10, 20, 50, 100, 200$ and 500 . We performed both (12) and the crude method in Goldfeld, Quandt and Trotter (1966) and followed the convergence criterion that iterations were stopped when the increment in $q(\alpha)$ was less than 10^{-8} . To facilitate comparisons we initialized the iterations at $2y_t - 40$, obtained from the equation $y_t = 0.5(\alpha_t + 40)$ by ignoring a nonlinear term in the measurement equation. Table 1 presents the number of iterations for convergence, the CPU time (in seconds) required and the optimal $q(\alpha)$ attained upon convergence. We can see that the crude hill-climbing generally requires less iterations. However, our modified

Table 1. Posterior mode estimation under the Andrade Netto, Gimeno and Mendes (1978) model.

n	number of			number of		
	iterations	CPU seconds	$\log p(\alpha y)$	iterations	CPU seconds	$\log p(\alpha y)$
	<i>modified quadratic hill-climbing in (12)</i>			<i>crude quadratic hill-climbing</i>		
10	69	0.2	-10.3	10	0.04	-10.6
20	28	0.1	-23.0	10	0.1	-22.1
50	28	0.3	-32.6	13	1.6	-35.5
100	140	4.2	-146.9	16	14.3	-149.2
200	163	14.4	-180.9	17	107.7	-194.6
500	283	118.7	-389.0	15	1509.0	-391.6

hill-climbing approach is computationally more efficient when n is as large as 200. The situation is more pronounced when n is 500. The time required for the crude method is in agreement with the assertion that the computation time is of order n^3 . On the other hand, the time required per iteration for our method roughly grows linearly with n . This is not surprising as we have shown in the Appendix that our algorithm in (12) is able to carry out the $n \times n$ matrix inversion as required in the hill-climbing method via Kalman filter type recursions. Therefore, it is expected that the benefit from using our method is very significant when the sample size n is large. Even though there is discrepancy in the maximum $q(\alpha)$ attained by the two methods, due mainly to the existence of multiple modes, the difference in the posterior modes obtained by the two methods is negligible in practice.

6.2. Logistic model

We present simulation results for the logistic model

$$y_t = \frac{\exp(\alpha_t)}{\exp(\alpha_t) + \exp(\epsilon_t)}, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2),$$

$$\alpha_{t+1} = \frac{\exp(\alpha_t)}{\exp(\alpha_t) + \exp(\eta_t)}, \quad \eta_t \sim N(0, 1),$$

in Tanizaki and Mariano (1994). Since the transition equation is not in the additive model form of (2), we are not able to apply (12) directly. One possibility is to use the technique in Section 5 to approximate the logistic model by an additive model. However, it seems nontrivial to impose the bounds of 0 and 1 on α_t in the optimization.

In this simulation experiment, we consider the transformed variables $\alpha'_t = \log\left(\frac{\alpha_t}{1-\alpha_t}\right)$ and $y'_t = \log\left(\frac{y_t}{1-y_t}\right)$. Then the logistic model is reduced to the following additive model

$$y'_t = \frac{\exp(\alpha'_t)}{1 + \exp(\alpha'_t)} + \epsilon'_t, \quad \alpha'_{t+1} = \frac{\exp(\alpha'_t)}{1 + \exp(\alpha'_t)} + \eta'_t,$$

where $\epsilon'_t = -\epsilon_t \sim N(0, \sigma_\epsilon^2)$ and $\eta'_t = -\eta_t \sim N(0, 1)$. With $\sigma_\epsilon^2 = 0.01$, we simulated $n = 100$ observations and applied the modified quadratic hill-climbing to compute the posterior mode. With the initial value of $\alpha'_t = 0$, implying that $\alpha_t = 0.5$, we attained convergence in six iterations. Figure 1 shows the time trend of α_t and the path of the posterior mode. As an estimate of α_t , the posterior mode matches closely with the real time trend. This indicates that our method works well in this model even though the dimension $n = 100$ is pretty large.

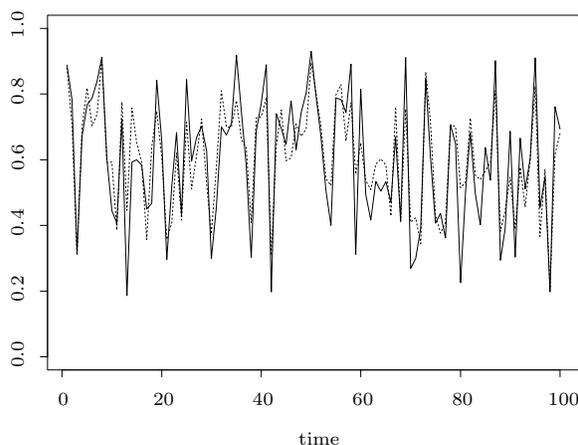


Figure 1. Time series plot of α_t in the logistic model is given by the solid line. The dotted line joins the posterior mode estimates.

To investigate whether the posterior mode solution derived from (12) is robust to the choice of the initial value of α , we performed the following sensitivity analysis. For the same simulated data set adopted above, we selected 500 sets of initial values and performed the optimization 500 times. In each trial, we initialized α_t randomly by drawing from $\text{Uniform}[0,1]$ and iterated (12) until convergence. Even though the above procedure produces very different starting values, we end up with almost identical results. The standard deviation of the 500 $q(\alpha)$ evaluated at the convergence value of α is less than 10^{-8} . Practically speaking, all 500 starting values yield the same solution. The results indicate that the modified quadratic hill-climbing is highly robust to the choice of starting values even for dimension as high as 100. To see the impact of σ_ϵ^2 on the robustness result, we repeat the above experiment by increasing σ_ϵ^2 to 0.1. We got similar results in that the standard deviation of the 500 $q(\alpha)$ upon convergence is less than 10^{-3} . Although the posterior modes from 500 random starting values are slightly more dispersed than that in the previous case of $\sigma_\epsilon^2 = 0.01$, the difference in the posterior mode solution due to the variation in the starting values is quite minor.

6.3. ARCH(1) + noise model

Consider the following ARCH(1) + noise model:

$$\begin{aligned} y_t &= \alpha_t + \epsilon_t, & \epsilon_t &\sim N(0, \sigma_\epsilon^2), \\ \alpha_{t+1} &= \sqrt{a + b\alpha_t^2} \xi_t, & \xi_t &\sim N(0, 1). \end{aligned}$$

The state variable α_t follows an ARCH(1) process and the observed variable y_t is simply α_t with an additive noise. This model provides an example of multi-

plicative transition equation. Following the notation in Section 5, $f_t(\alpha_t, \xi_t) = \sqrt{a + b\alpha_t^2} \xi_t$. To apply the modified quadratic hill-climbing method, we form the approximated transition equation as in (13). By expanding $f_t(\alpha_t, \xi_t)$ at $\alpha_t = \hat{\alpha}_t$ and $\xi_t = 0$, we get $T_t = 0$ and $R_t = \sqrt{a + b\hat{\alpha}_t^2}$, where $\hat{\alpha}_t$ is the current iterate. In the simulation, we fix $\sigma_\epsilon^2 = 0.1$, $a = 0.1$ and $b = 0.9$. With standard normal α_0 , we simulated y_t , $t = 1, \dots, 100$. Figure 2 shows the time series plot of α_t and its posterior mode estimate. We observe that the posterior mode follows closely the time trend of the state variable. This experiment indicates that the posterior mode can be a good estimate of α and our method works well in the above non-additive model.

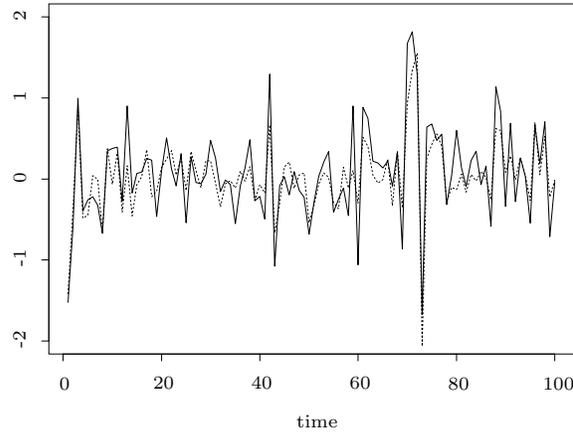


Figure 2. Time series plot of α_t in the ARCH(1) + noise model is given by the solid line. The dotted line joins the posterior mode estimates.

7. Real Data Illustration

Although stochastic volatility models are important alternatives to ARCH models for market volatility, the original form

$$y_t = \exp\left(\frac{\alpha_t}{2}\right)\epsilon_t, \quad \epsilon_t \sim N(0, 1), \quad \alpha_{t+1} = c + \phi\alpha_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2)$$

is too simple to capture some of the stylized facts observed in real data (Harvey, Ruiz and Shephard (1994); Ghysels, Harvey and Renault (1996); Kim, Shephard and Chib (1998)). In this section, we introduce a stochastic volatility in the mean model and demonstrate that volatility estimation by the posterior mode of α_t is feasible with our modified quadratic hill-climbing method. Let y_t be the financial return at time t . The stochastic volatility in the mean model proposed in this paper is defined as

$$y_t = \delta \exp\left(\frac{\alpha_t}{2}\right) + \exp\left(\frac{\alpha_t}{2}\right)\epsilon_t, \quad \epsilon_t \sim \sqrt{\frac{\nu - 2}{\nu}}t_\nu,$$

$$\alpha_{t+1} = c + \phi\alpha_t + \gamma_1 z_t + \gamma_2 |z_t| + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \quad (14)$$

where $z_t = \exp(-\frac{\alpha_t}{2}) y_t$ is the standardized return, that is, y_t standardized by the conditional standard deviation of $\sqrt{\text{Var}(y_t|\alpha_t)} = \exp(\alpha_t/2)$. When $\delta = \gamma_1 = \gamma_2 = 0$ and $\nu = \infty$, this reduces to the original model.

The above stochastic volatility in the mean model is constructed by making three generalizations to the basic model. First, a scalar multiple of the conditional standard deviation $\exp(\alpha_t/2)$ is added to the mean part. As in the ARCH-M model of Engle, Lilien and Robins (1987), the additional component can be interpreted as the risk premium. An alternative mean specification was also considered in Koopman and Hol Uspensky (2002). The second generalization is to allow ϵ_t to have a fat-tailed distribution. Here, we assume that ϵ_t follows a standardized t distribution with mean 0 and variance 1. This extension is important to capture the leptokurtosis observed in many market returns. The last generalization is based on some empirical findings that volatility responds asymmetrically to positive and negative returns. The specification in the variance equation is set up similarly to Nelson (1991) and Li and Li (1996) in capturing the volatility asymmetry. Under the model in (14), the conditional mean of α_{t+1} given α_t is

$$c + \phi\alpha_t + \gamma_1 z_t + \gamma_2 |z_t| = \begin{cases} c + \phi\alpha_t + (\gamma_1 + \gamma_2)|z_t|, & \text{if } z_t > 0, \\ c + \phi\alpha_t + (-\gamma_1 + \gamma_2)|z_t|, & \text{if } z_t < 0, \end{cases}$$

which is dependent on both the sign and magnitude of z_t . With all the three new features included, we believe that the new stochastic volatility in (14) can provide a better fit to real data.

It is not difficult to see that the stochastic volatility in the mean model in (14) belongs to the class of nonlinear models in (1) and (2) with

$$\begin{aligned} l(\alpha_t) &= \log p(y_t|\alpha_t) \\ &= \log \Gamma\left(\frac{\nu+1}{2}\right) - \log \Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2} \log(\pi(\nu-2)) - \frac{\alpha_t}{2} - \left(\frac{\nu+1}{2}\right) \log \left[1 + \frac{(z_t - \delta)^2}{\nu - 2}\right], \\ l'(\alpha_t) &= -\frac{1}{2} + \frac{\nu+1}{2(\nu-2)} z_t(z_t - \delta) \left[1 + \frac{(z_t - \delta)^2}{\nu - 2}\right]^{-1}, \\ l''(\alpha_t) &= \frac{(\nu+1)z_t}{2(\nu-2)} \left[1 + \frac{(z_t - \delta)^2}{\nu - 2}\right]^{-2} \left\{ \frac{1}{2} \delta \left[1 + \frac{(z_t - \delta)^2}{\nu - 2}\right] - z_t \right\}, \\ g_t(\alpha_t) &= c + \phi\alpha_t + \gamma_1 z_t + \gamma_2 |z_t|. \end{aligned}$$

To estimate the posterior mode by optimizing $p(\alpha|y)$, existing methods are not applicable because $l''(\alpha_t)$ is not necessarily negative and we have nonlinear transition in the variance part.

S & P 500 returns are used for illustration. Daily returns (in %) from 1988 to 1998 were centered by the sample mean to produce the data series y_t . Applying the method in Durbin and Koopman (1997) with some modifications to account for the nonlinear transition equation, unknown parameters were estimated by standard maximum likelihood. Technical details are available from the author upon request. Given the maximum likelihood estimates in Table 2, we used the posterior mode generated by the modified quadratic hill-climbing method in (12) to provide a smoothed estimate for α_t . Figure 3 shows the time series plot of the returns and the smoothed estimate of the volatility given by $\exp(\hat{\alpha}_t/2)$, where $\hat{\alpha}_t$ is the posterior mode. This volatility estimate is useful for derivatives pricing and financial risk management.

Table 2. Maximum likelihood estimates from fitting the model in (14).

Parameter	MLE
δ	0.00706
ν	6.574
c	0.00873
σ_η^2	0.00730
ϕ	0.990
γ_1	-0.048
γ_2	-0.019

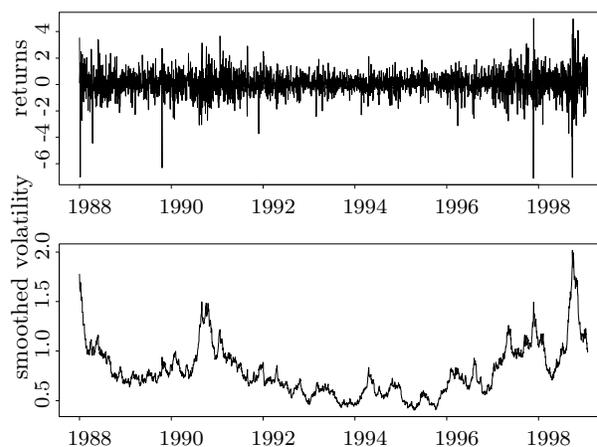


Figure 3. Time series plots of the S and P 500 returns and the smoothed volatility estimate $\exp(\hat{\alpha}_t/2)$.

Following Section 6.1, we compare some numerical properties of our posterior mode estimation method in (12) and the crude hill-climbing method under the

stochastic volatility in the mean model. Assuming the model parameters in Table 2, we considered $n = 10, 20, 50, 100, 200$ and 500 . The first day of 1998 was taken as $t = 1$ in all cases. We adopted the same convergence criterion as in Section 6.1 and initialized all α_t at 0. Summary results are reported in Table 3. In contrast to the first experiment's results in Table 1, our method uses less iterations than the crude method and the maximum $\log p(\alpha|y)$ attained agrees in both methods even for large n . The CPU time taken illustrates again the $O(n)$ and $O(n^3)$ operations for the modified and crude quadratic hill-climbing methods respectively. According to the figures in Table 3, if we want to obtain the posterior mode of $n = 2781$ data points from the period 1988 to 1998, the crude method will take more than 3000 CPU hours, while our method requires 16 CPU seconds.

Table 3. Posterior mode estimation under the stochastic volatility in the mean model in (14).

n	number of			number of		
	iterations	CPU seconds	$\log p(\alpha y)$	iterations	CPU seconds	$\log p(\alpha y)$
	<i>modified quadratic hill-climbing in (12)</i>			<i>crude quadratic hill-climbing</i>		
10	4	0.03	-10.4	12	0.7	-10.4
20	5	0.02	-13.2	13	4.4	-13.2
50	5	0.3	-7.7	13	57.1	-7.6
100	5	0.6	-2.6	14	466.3	-2.6
200	5	1.3	22.0	18	4694.6	22.0
500	5	2.7	150.9	19	76261.6	150.9

Acknowledgements

The author would like to thank the Editor, an associate editor and a referee for very constructive comments.

Appendix

Consider the non-Gaussian measurement time series model in (4). Let $\mu = E(\alpha)$ and $\Gamma = \text{Cov}(\alpha)$, so $\alpha \sim N(\mu, \Gamma)$. The logarithm of the posterior density $p(\alpha|y)$ is

$$\begin{aligned} \log p(\alpha|y) &= \text{constant} + \log p(y|\alpha) + \log p(\alpha) \\ &= \text{constant} + \sum_{t=1}^n \log p(y_t|\alpha_t) - \frac{1}{2}(\alpha - \mu)' \Gamma^{-1}(\alpha - \mu) = \text{constant} + q(\alpha), \end{aligned}$$

where $q(\alpha) = \sum_{t=1}^n l(\alpha_t) - \frac{1}{2}(\alpha - \mu)' \Gamma^{-1}(\alpha - \mu)$. Direct differentiation gives the

first and second derivatives of $q(\alpha)$ with respect to α as,

$$q'(\alpha) = \begin{pmatrix} l'(\alpha_1) \\ \vdots \\ l'(\alpha_n) \end{pmatrix} - \Gamma^{-1}(\alpha - \mu) \quad \text{and} \quad q''(\alpha) = \begin{pmatrix} l''(\alpha_1) & & 0 \\ & \ddots & \\ 0 & & l''(\alpha_n) \end{pmatrix} - \Gamma^{-1}.$$

To facilitate the implementation of the quadratic hill-climbing procedure, we need to calculate $\alpha^{(1)}$ from

$$\begin{aligned} \alpha^{(1)} &= \alpha^{(0)} - (q''(\alpha^{(0)}) - \delta I)^{-1} q'(\alpha^{(0)}) \\ &= -(q''(\alpha^{(0)}) - \delta I)^{-1} \{q'(\alpha^{(0)}) - (q''(\alpha^{(0)}) - \delta I)\alpha^{(0)}\} \\ &= -(q''(\alpha^{(0)}) - \delta I)^{-1} \{q'(\alpha^{(0)}) - q''(\alpha^{(0)})\alpha^{(0)} + \delta\alpha^{(0)}\} \\ &= -(q''(\alpha^{(0)}) - \delta I)^{-1} \left\{ \begin{pmatrix} l'(\alpha_1^{(0)}) - l''(\alpha_1^{(0)})\alpha_1^{(0)} + \delta\alpha_1^{(0)} \\ \vdots \\ l'(\alpha_n^{(0)}) - l''(\alpha_n^{(0)})\alpha_n^{(0)} + \delta\alpha_n^{(0)} \end{pmatrix} + \Gamma^{-1}\mu \right\}. \end{aligned} \quad (15)$$

Recall that in linear Gaussian state space models, we have $y|\alpha \sim N(\alpha, H)$ and $\alpha \sim N(\mu, \Gamma)$. The logarithm of the posterior density and its derivative are

$$\begin{aligned} \log p(\alpha|y) &= \text{constant} - \frac{1}{2}(y - \alpha)'H^{-1}(y - \alpha) - \frac{1}{2}(\alpha - \mu)'\Gamma^{-1}(\alpha - \mu), \\ \frac{\partial \log p(\alpha|y)}{\partial \alpha} &= H^{-1}(y - \alpha) - \Gamma^{-1}(\alpha - \mu). \end{aligned}$$

Setting the first derivative to zero yields the posterior mode under the Gaussian state space model as

$$\hat{\alpha} = (H^{-1} + \Gamma^{-1})^{-1} \{H^{-1}y + \Gamma^{-1}\mu\}. \quad (16)$$

Since α is normally distributed conditional on y , the posterior mode is the posterior mean, which can be evaluated by one pass of standard Kalman filter smoothers; see Harvey (1989, p.149-155).

To show that (15) can be computed from (16) via the Kalman filter, we form the following artificial observations and noise variance: $y_t = \alpha_t^{(0)} + (-l''(\alpha_t^{(0)}) + \delta I)^{-1}l'(\alpha_t^{(0)})$ and

$$H = \begin{pmatrix} -(l''(\alpha_1^{(0)}) - \delta I)^{-1} & & 0 \\ & \ddots & \\ 0 & & -(l''(\alpha_n^{(0)}) - \delta I)^{-1} \end{pmatrix}.$$

Then,

$$H^{-1} + \Gamma^{-1} = \begin{pmatrix} -(l''(\alpha_1^{(0)}) - \delta I) & & 0 \\ & \ddots & \\ 0 & & -(l''(\alpha_n^{(0)}) - \delta I) \end{pmatrix} + \Gamma^{-1} = -(q''(\alpha^{(0)}) - \delta I),$$

$$H^{-1}y = \begin{pmatrix} -(l''(\alpha_1^{(0)}) - \delta I) & & 0 \\ & \ddots & \\ 0 & & -(l''(\alpha_n^{(0)}) - \delta I) \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

with the typical element $-(l''(\alpha_t^{(0)}) - \delta I)y_t = -(l''(\alpha_t^{(0)}) - \delta I)\{\alpha_t^{(0)} + (-l''(\alpha_t^{(0)}) + \delta I)^{-1}l'(\alpha_t^{(0)})\} = -l''(\alpha_t^{(0)})\alpha_t^{(0)} + \delta\alpha_t^{(0)} + l'(\alpha_t^{(0)})$. Hence, $\hat{\alpha}$ in (16) is identical to $\alpha^{(1)}$ in (15) and the result follows.

References

- Alspach, D. L. and Sorenson, H. W. (1972). Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Trans. Automat. Control* **17**, 439-48.
- Anderson, B. D. O. and Moore, J. B. (1979). *Optimal Filtering*. Prentice-Hall, Englewood Cliffs.
- Andrade Netto, M. L., Gimeno, L. and Mendes, M. J. (1978). On the optimal and suboptimal nonlinear filtering problem for discrete-time systems. *IEEE Trans. Automat. Control* **23**, 1062-67.
- Carlin, B. P., Polson, N. G. and Stoffer, D. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *J. Amer. Statist. Assoc.* **87**, 493-500.
- Durbin, J. and Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika* **84**, 669-84.
- Durbin, J. and Koopman, S. J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *J. Roy. Statist. Soc. Ser. B* **62**, 3-29.
- Engle, R. F., Lilien, D. M. and Robins, R. P. (1987). Estimating time-varying risk premia in the term structure: the ARCH-M model. *Econometrica* **55**, 391-407.
- Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models. *J. Amer. Statist. Assoc.* **87**, 501-9.
- Ghysels, E., Harvey, A. C. and Renault, E. (1996). Stochastic volatility. In *Handbook of Statistics* Vol. 14 *Statistical Methods in Finance* (Edited by C. R. Rao and G. S. Maddala), 119-191. Elsevier, Amsterdam.
- Goldfeld, S. M., Quandt, R. E. and Trotter, H. F. (1966). Maximization by quadratic hill-climbing. *Econometrica* **34**, 541-51.
- Harvey, A. C. (1989). *Forecasting Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Harvey, A. C., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models. *Rev. Econom. Stud.* **61**, 247-264.
- Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Rev. Econom. Stud.* **65**, 361-93.
- Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series. *J. Amer. Statist. Assoc.* **82**, 1032-63.
- Koopman, S. J. and Hol Uspensky, E. (2002). The stochastic volatility in mean model: empirical evidence from international stock markets. *J. Appl. Econom.* **17**, in press.
- Li, C. W. and Li, W. K. (1996). On a double threshold autoregressive heteroskedastic autoregressive time series model. *J. Appl. Econom.* **11**, 253-274.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**, 347-70.

- Sage, A. P. and Melsa, J. L. (1971) *Estimation Theory With Applications to Communications and Control*. McGraw-Hill Book Company, New York.
- Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* **84**, 653-67.
- Tanizaki H. and Mariano, R. S. (1994). Prediction, filtering and smoothing in non-linear and non-normal cases using Monte Carlo integration. *J. Appl. Econom.* **9**, 163-79.

Department of Information and Systems Management, School of Business and Management,
The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong.

E-mail: immkps@ust.hk

(Received December 2000; accepted June 2002)