

Discretised Non-Linear Filtering for Dynamic Latent Variable Models: with Application to Stochastic Volatility

A.E. Clements, A.S. Hurn, S.I. White.

School of Economics and Finance, Queensland University of Technology

February 5, 2004

Abstract

Filtering techniques are often applied to the estimation of dynamic latent variable models. However, these techniques are often based on a set of assumptions which restrict models to be specified in a linear state-space form. Numerical filtering techniques have been proposed that avoid invoking such restrictive assumptions, thus permitting a wider class of latent variable models to be considered. This paper proposes an accurate yet computationally efficient numerical filtering algorithm (based on a discretisation of the state space) for estimating the general class of dynamic latent variable models. The empirical performance of this algorithm is considered within the context of the stochastic volatility model. It is found that the proposed algorithm outperforms a number of accepted procedures in terms of volatility forecasting.

Keywords

Non-linear filtering, latent variable models, stochastic volatility, volatility forecasting.

JEL Classification C22, C51, C53.

Corresponding author

Adam Clements
School of Economics and Finance
Queensland University of Technology
Brisbane, 4001
Qld, Australia
email a.clements@qut.edu.au

1 Introduction

The Kalman filter has proved extremely useful in estimating the parameters of dynamic latent factor models in econometrics where the model is either linear or can be transformed and treated as linear (see, for example, Harvey, 1981, 1989). Kitagawa (1987) demonstrated how the same basic filtering cycle of prediction and update could be used in a general nonlinear framework. Despite the potential for being a powerful tool at the disposal of the applied practitioner, the prohibitive computational burden imposed by the numerical techniques required to implement Kitagawa's algorithm has meant that it has not been widely used. The contribution of this paper is to introduce a computationally efficient method for implementing the nonlinear filter. The proposed method is based upon a fixed discretisation of the state space of the latent factor that permits the necessary nonlinear filtering equations to be solved by means of a simple yet accurate numerical integration scheme. The performance of the proposed method, which is called the discrete nonlinear filter (DNF), is illustrated in Monte Carlo simulation experiments and in an application based on an important problem in financial econometrics, namely that of estimating the parameters of stochastic-volatility models.

The paper is structured as follows. Section 2 outlines the general non-linear filtering framework, and describes estimation method employed by Kitigawa (1987). The proposed DNF method is described in Section 3 where the differences between the current and previous approaches are also discussed. Section 4 defines the general stochastic volatility framework and Section 5 reports the results of a number of Monte-carlo experiments to ascertain the relative performance of the DNF method when applied to the parameter-estimation problem in the stochastic volatility problem. Section 6 consider the application of the DNF to estimating SV models. Results are based on FX futures data, with the performance of the DNF method compared to two competing volatility models. Section 7 provides concluding remarks and suggests avenues for future research.

2 Kitagawa's Algorithm

Consider a system described by the state-space model

$$y_t = H(x_t, u_t | \theta), \quad x_t = F(x_{t-1}, w_t | \theta) \quad (1)$$

where:

y_t is an observed data series conditional on the value of the (unobserved) state variable x ;

u_t and w_t are (possibly correlated) observation and system noise terms; and

θ is an unknown the parameter vector to be estimated.

In the event of $H(\cdot)$ and $F(\cdot)$ being linear and with $u_t \sim N(0, \sigma_u^2)$ and $w_t \sim N(0, \sigma_w^2)$, standard linear Kalman-filtering techniques may be used to generate maximum likelihood estimates of the unknown parameters, θ , by means of the prediction-error decomposition of the likelihood. In the more general case where linearity does not apply, the maximum likelihood estimates of θ are computed as

$$\begin{aligned} \hat{\theta}_{ML} &= \arg \max_{\theta} \left[f \left(\{y_t\}_{t=1}^T | \theta \right) \right] \\ &= \arg \max_{\theta} \left[\int \cdots \int f \left(\{y_t\}_{t=1}^T | \{x_t\}_{t=1}^T \right) f \left(\{x_t\}_{t=1}^T | \theta \right) dx_1 \dots dx_T \right]. \end{aligned} \quad (2)$$

which is a T -fold integration problem that cannot generally be solved by analytical means.

Kitagawa (1987) suggests that the evaluating the integral in equation (2) be accomplished in terms of a recursive prediction-update algorithm. For this purpose it is useful to express the state-space model in equation (1) in terms of two conditional distributions

$$y_t \sim r(\cdot | x_t, \theta) \quad x_t \sim q(\cdot | x_{t-1}, \theta) \quad (3)$$

where $r(\cdot | x_t)$ is the conditional distribution of y on x , and $q(\cdot | x_{t-1})$ is the conditional distribution of x_t on x_{t-1} . The prediction-update algorithm proceeds as follows.

Prediction step

The one-step ahead prediction of the distribution of x_t conditional on y_{t-1} , $f(x_t|y_{t-1}, \theta)$, is given by

$$f(x_t|y_{t-1}, \theta) = \int_{-\infty}^{\infty} q(x_t|x_{t-1}, \theta) f(x_{t-1}|y_{t-1}, \theta) dx_{t-1}. \quad (4)$$

Update step

The form of the probability distribution of the state variable at time t , conditional on information at time t is then given by

$$f(x_t|y_t, \theta) = \frac{r(y_t|x_t, y_{t-1}, \theta) f(x_t|y_{t-1}, \theta)}{f(y_t|y_{t-1}, \theta)}. \quad (5)$$

where the denominator of equation (5) is the likelihood of observing y_t given y_{t-1} and θ

$$f(y_t|y_{t-1}, \theta) = \int_{-\infty}^{\infty} r(y_t|x_t, \theta) f(x_t|y_{t-1}, \theta) dx_t. \quad (6)$$

There are two important by-products obtained by recursing through equations (4) and (5) for all observations T . In the first instance the the log-likelihood function to be maximised to obtain the ML estimates of θ is obtained directly from equation (5) and is given by

$$\ln L = \sum_{t=1}^T \ln[f(y_t|y_{t-1}, \theta)]. \quad (7)$$

In addition to parameter estimation, the recursions of the filter also allow the expected value of the state variable, conditional on the parameters θ and all information up to and including T , to be constructed. Note that the conditional distribution of x_t conditional on y_T and θ is constructed as

$$f(x_t|y_T, \theta) = f(x_t|y_t, \theta) \int_{-\infty}^{\infty} \frac{f(x_{t+1}|y_T, \theta) q(x_{t+1}|x_t, \theta)}{f(x_{t+1}|y_t, \theta)} dx_{t+1}, \quad (8)$$

with expected value

$$E[x_t|y_T, \theta] = \int_{-\infty}^{\infty} x_t \cdot f(x_t|y_T, \theta) dx_t. \quad (9)$$

From the perspective of parameter estimation it is clear that the intractable high-dimensional integral in equation (2) has been replaced with the relatively

straightforward summation in equation (7), the problem now becomes that of providing a numerical technique to evaluate the integrals in the prediction and update equations, (4) and (5) respectively. Kitagawa (1987) suggested that the probability distributions in the relevant integrals be approximated by linear splines. This requires the specification of the number of linear segments in the spline, the location of the spline knots and consequently the value of the functions (heights of the densities, $f(x_t|y_{t-1}, \theta)$ and $f(x_t|y_t, \theta)$) at the knots¹. Based on this linear approximation, the trapezoidal rule is then used to compute the required integrals.

In their comment on Kitagawa's original paper, Martin and Raferty (1987) point out that computing complexity is likely to be a deterrent for all but simple problems. Indeed, they argued that the computational burden of the proposed numerical integration procedure was so great that it was unlikely to be of practical use and suggested that the provision of accurate computationally attractive alternatives was an important area for future research. The next section is devoted to the description of such an approximation which delivers significant computational gain without any deterioration in numerical accuracy.

3 The Discrete Nonlinear Filter

At the heart of the nonlinear filtering problem is the approximation of the relevant probabilities in expression (3) and their numerical integration. A practical complication with the implementation of this approach, therefore, stems from that fact its efficacy relies on two quite different procedures (approximation of the probability density and numerical integration), both of which suffer from error. Clearly a compromise must be struck between the control of error due to the numerical integration and error arising in the construction of estimates of probability density. In the approach suggested by Kitagawa (1987) the linear spline approximation to the density functions requires a cumbersome numerical

¹Kitagawa (1987) proposed a very simple scheme for knot placement with knots equally spaced over the finite interval taken to be the domain of the state variable. Watanabe (1999) following the suggestion of Tanizaki (1993) uses a more elaborate algorithm for knot placement where the knots are placed in regions where the state variable is most likely to occur.

integration procedure. By contrast, the central contribution of this paper is to suggest an approximation to these distributions based on discretising the state space of the latent variable, x , and computing the probability of observing x within a set of discrete intervals in a manner similar to a histogram. The major advantage of this approximation is that the integration problem is now reduced to the simple sum of the product of probabilities. Thus the DNF requires far fewer computations per integration than previous approaches thereby dramatically reducing computational time and the associated scope for numerical error. The details of the DNF are now outlined.

Consider N adjacent intervals in x space bounded by $w^1 \dots w^{N+1}$ and centered on the points $z^1 \dots z^N$ where

$$z^i = \frac{w^i + w^{i+1}}{2}. \quad (10)$$

In general terms, the probability of observing x within the interval centered on z^i , i.e. $x \in (w^i, w^{i+1}]$ is given by

$$p(x \in (w^i, w^{i+1}]) = \int_{w^i}^{w^{i+1}} f(x) dx \approx p(z^i) \quad (11)$$

where $f(x)$ is the probability distribution of the of the unobserved state variable x .

Based on this discretisation the first task is to generate the approximations to the conditional distributions in (3).

Transitional density

The transitional density of x may be discretised into a set of transitional probabilities. Given that the state space is defined over N adjacent intervals of width δ , it is possible to compute an $N \times N$ transitional probability matrix, \hat{q} . The elements of this matrix, $\hat{q}^{i,j} \forall i, j = 1, \dots, N$, represent the probability of x migrating from the interval centred on z^j to the interval centred on z^i defined by

$$\hat{q}^{i,j} = p(x \in (w^i, w^{i+1}] | x \in (w^j, w^{j+1}], \theta) \approx \delta q(z^i | z^j, \theta). \quad (12)$$

Conditional likelihood

Similarly, the likelihood of observing y_t conditional on x_t being in each discrete interval may be defined as the $T \times N$ likelihood matrix containing elements, \hat{r}_t^i $\forall i = 1, \dots, N$, defined by

$$\hat{r}_t^i = r(y_t | x \in (w^i, w^{i+1}], \theta) \approx r(y_t | z^i, \theta). \quad (13)$$

Clearly the approximation of these conditional distributions will depend on the exact nature of the discretisation. In this regard there are two separate questions that need to be addressed, namely, whether the intervals $(w^i, w^{i+1}]$ are equal for all i , and whether these intervals are fixed for each time step or allowed to be time-varying. With respect to the first issue, by concentrating the intervals in the vicinity of the mode of the distribution of x , greater resolution is gained in the area of greatest probability at the expense of accuracy in the tail. In the case of fixed- or time-varying intervals the central issue is one of the stationarity of the latent variable. In any instance where the latent variable is nonstationary, the use of time varying intervals is necessary. It may be noted, however, that use of fixed intervals allows even greater reduction in computational cost as the two matrices $\hat{q}^{i,j}$ and \hat{r}_t^i may be pre-computed and held fixed for any given set of parameters. Ultimately the question of interval definition is one which must be settled empirically. A small Monte Carlo exercise is presented in Section 5 which explores these questions in more detail.

Before stating the recursions required to implement the DNF, it is convenient to simplify the notation for the one-step-ahead prediction of the distribution of x , represented in terms of the probability of observing $x \in (w^i, w^{i+1}]$ at time t , as follows

$$p_t(x \in (w^i, w^{i+1}] | y_{t-1}, \theta) = p_{t|t-1}^i.$$

Note that to be initialised the DNF requires an estimate of $p_{1|0}^i$. These initial probabilities are obtained by discretising the unconditional distribution of the state variable so that

$$p_{1|0}^i = \int_{w^i}^{w^{i+1}} f(x | \theta) dx.$$

for all intervals.

Similarly, the updated probability of observing $x \in (w^i, w^{i+1}]$ at time t , may be written as

$$p_t(x \in (w^i, w^{i+1}] | y_t, \theta) = p_{t|t}^i$$

The prediction and update steps of the DNF are now very simply given by

$$p_{t|t-1}^i = \sum_{j=1}^N \hat{q}^{i,j} \cdot p_{t-1|t-1}^j, \quad (14)$$

and

$$p_{t|t}^i = \frac{\hat{r}_t^i \cdot p_{t|t-1}^i}{\sum_{i=1}^N \hat{r}_t^i \cdot p_{t|t-1}^i} \quad (15)$$

respectively. The denominator of equation (15) is the likelihood of observing y_t , integrated across possible states of x and is used to evaluate the log-likelihood in equation 7. Initial estimates are that the computation time required for an evaluation of the likelihood using this method is, on average, five times faster than that of Kitigawa (1987).

Following from equations (8) and (9), a method for extracting estimates of the expected value of the state variable based on the DNF numerical scheme is also proposed. Given estimated values for the elements of θ , the smoothed distribution of x , the probability of observing $x \in (w^j, w^{j+1}]$ at time t , conditional on information up to and including time T , $p_t(x \in (w^j, w^{j+1}] | y_T, \theta)$ is determined by

$$p_t(x \in (w^i, w^{i+1}] | y_T, \theta) = p_{t|t}^i \cdot \sum_{j=1}^N \frac{p_{t+1}^j \cdot \hat{q}^{i,j}}{p_{t+1}(x \in (w^j, w^{j+1}] | y_T, \theta)}. \quad (16)$$

Based on the estimate of the *pdf* of the state variable, the expected value of x_t is obtained from

$$E[x_t | y_T, \theta] = \sum_{i=1}^N z^i \cdot p_t(x \in (w^i, w^{i+1}] | y_T, \theta). \quad (17)$$

4 Stochastic Volatility

The operation of the DNF is now illustrated with reference to the stochastic volatility (SV) model. Let returns (the observed variable) $\{y_t\}_{t=1}^T$ be generated by,

$$y_t = \sigma_t u_t \quad u_t \sim N(0, 1) \quad (18)$$

where σ_t is the time t conditional standard deviation of y_t . SV models treat σ_t as an unobserved (latent) stochastic variable, whose dynamics need to be specified. The simplest model for σ_t is the AR(1) process,

$$\ln(\sigma_t^2) = \alpha + \beta \ln(\sigma_{t-1}^2) + w_t \quad w_t \sim N(0, \sigma_w^2) \quad (19)$$

where the errors u_t and w_t are assumed to be independent².

Numerous approaches have been devised for estimating the parameters of the SV model (see Ghysels *et al.* (1996) and Shephard (1996) for surveys). Perhaps the most popular method is to express equations (18) and (19) in a linear state-space form by using an appropriate transformation and then apply standard Kalman filtering methods (Harvey *et al.*, 1994, Ruiz, 1994). This is a quasi-maximum likelihood (QML) approach because, in the simple implementation of this approach, the transformation of equation (18) results in a non-normal error term and thus a violation of the strict conditions required for Kalman filter to yield maximum likelihood estimates. This shortcoming is easily addressed in the current context by recognising that the nonlinear filter as proposed by Kitigawa (1987) may be used to provide maximum likelihood estimates without the need for any prior transformation of the model.

To apply the filtering approach outlined in Section 3 to the estimation of SV parameters, the relevant probability densities, $q(\cdot|x_{t-1})$ and $r(\cdot|x_t)$ must first be defined. Given the observed returns series $\{y_t\}_{t=1}^T$ and defining $x_t = \ln(\sigma_t^2)$, $q(\cdot|x_{t-1})$ and $r(\cdot|x_t)$ given by

$$q(x_t | x_{t-1}, y_{t-1}, \theta) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp \left[-\frac{(x_t - \alpha - \beta x_{t-1})^2}{2\sigma_w^2} \right], \quad (20)$$

²As the focus of this paper is the performance of the proposed filter only this uncorrelated case will be examined. Correlation between these two error terms can also be dealt with quite easily.

$$r(y_t | x_t, y_{t-1}, \theta) = \frac{1}{\sqrt{2\pi \exp(x_t)}} \exp \left[-\frac{y_t^2}{2 \exp(x_t)} \right]. \quad (21)$$

respectively.

To implement the algorithm, the set of intervals bounded by $w^1 \dots w^{N+1}$ must be chosen. For SV estimation purposes, points distributed uniformly in x ($\log \sigma^2$) space with range given by

$$\frac{\alpha}{(1-\beta)} \pm 6 \frac{\sigma_w}{\sqrt{(1-\beta^2)}}, \quad (22)$$

from which N discrete intervals centred on $z^1 \dots z^N$ are defined. Given a value for θ , the transition probabilities between each interval, $\hat{q}^{i,j}$, and conditional likelihoods, \hat{r}_t^i , may be pre-computed using equations 12 and 13 respectively. To initialise the filter, the initial profile of the state variable must be defined and discretised to obtain the prediction of the state variable at $t = 1$, ie, p_1^i . The initial profile of the state variable (assuming normality) is taken to be its unconditional distribution, which for equation 19 leads to

$$f(x | \theta) \sim N \left(\frac{\alpha}{(1-\beta)}, \frac{\sigma_w^2}{(1-\beta^2)} \right) \quad (23)$$

from which the initialisation of p_1^i follows equation 11. Based on $\hat{q}^{i,j}$, \hat{r}_t^i , and p_1^i , an estimate of θ , $\hat{\theta}_{ML}$ is obtained by maximising the log-likelihood from equation 7, an approximation to the *log* of equation 2, which is a by-product of recursing between equations 14 and 15.

Based on $\hat{\theta}_{ML}$, the expected value of unobserved volatility may be extracted from the smoothed distribution, $p_t(x \in (w^i, w^{i+1}] | y_T, \hat{\theta}_{ML})$ constructed using equation 16,

$$E(\sigma_t^2 | y_T) = \sum_{i=1}^N \exp(z_i) p_t(x \in (w^i, w^{i+1}] | y_T, \hat{\theta}_{ML}). \quad (24)$$

The performance of the DNF model as applied to the stochastic volatility problem will be now be examined in both theoretical and applied settings.

5 Simulation Experiments

The purpose of this section is to replicate the Monte Carlo study undertaken by Jacquier, Polson and Rossi (1994) to evaluate the performance of parameter estimation methods in the context of SV models. There are three aspects of the performance of the DNF which require thorough investigation.

1. The accuracy with which the DNF integration scheme evaluates the likelihood function given by equation (2).
2. The overall performance of the DNF in terms of parameter estimation.
3. The accuracy with which the latent variable (volatility) is extracted.

In performing the experiments the three parameter combinations used by Jacquier *et al.* (1994), namely,

$$\begin{aligned}(\alpha, \beta, \sigma_w) &= (-0.736, 0.90, 0.363) \\(\alpha, \beta, \sigma_w) &= (-0.368, 0.95, 0.260) \\(\alpha, \beta, \sigma_w) &= (-0.147, 0.98, 0.166)\end{aligned}$$

are used.

5.1 Accuracy of likelihood evaluation

This section considers the relative numerical accuracy of the integration scheme employed by the DNF, within the context of the SV model. Consider equations (18) and (19) applied to a situation where there is only one time step, namely, time $t-1$ to time t . In this instance, the t -fold integration problem encountered in the construction of the likelihood function, equation (2), is now reduced to the more manageable double integral

$$\begin{aligned}f(y_t|y_{t-1}, \theta) &= \int_{-\infty}^{\infty} r(y_t|x_t, \theta) f(x_t|y_{t-1}, \theta) dx_t \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} r(y_t|x_t, \theta) q(x_t|x_{t-1}, \theta) f(x_{t-1}|y_{t-1}, \theta) dx_t dx_{t-1}\end{aligned}$$

for given parameters θ and where $f(x_{t-1}|y_{t-1}, \theta)$ and y_t are specified. This reduction in the dimension of the integral required to compute the likelihood is important as it allows a full double quadrature numerical routine to be utilized as a benchmark with which to compare the approximations suggested by

Kitigawa (1987), Watanabe (1999) and the discrete binning method suggested in this paper.

The experiment proceeds as follows:

1. Values for the elements of the parameter vector $\theta = (\alpha, \beta, \sigma_w)$ are set.
2. The density function $f(x_{t-1} | y_{t-1}, \theta)$ set equal to the starting distribution given by equation (23).
3. A value of y_t is randomly drawn and the likelihood evaluated at this point using four methods
 - (a) A full numerical double quadrature method as implemented by MATLAB.
 - (b) The trapezoidal implementation suggested by Kitagawa (1987) using evenly spaced nodes (NFML(E)).
 - (c) The trapezoidal implementation used by Watanabe (1999) in which the nodes are normally distributed around the expected value of the state variable (NFML).
 - (d) The method central to the DNF where the state-space is discretised into 50 bins of equal width set up as specified in expression (22).
4. The process is repeated 2000 times for the purpose of computing average errors.

Table 1 reports the root mean square errors obtained by comparing the values for the likelihood obtained in each of methods (b)-(d) above relative to the full double quadrature in (a).

The results of this simple Monte-Carlo experiment yield two important conclusions. First, a comparison of the *NFML(E)* and *NFML* methods suggests that trapezoidal integration based on evenly-spaced nodes is more accurate. It appears that the benefit of concentrating nodes in the vicinity of the expected value of x (and thereby obtaining greater resolution of the distribution of x in this region), is dominated by the cost in terms of lack of accuracy in the tails of

Table 1:
RMSE for likelihood evaluation relative to double quadrature

Parameters	RMSEs		
	<i>DNF</i>	<i>NFML(E)</i>	<i>NFML</i>
$(\alpha, \beta, \sigma_w)$ (-0.736, 0.90, 0.363)	0.0032	0.0003	0.3754
$(\alpha, \beta, \sigma_w)$ (-0.368, 0.95, 0.260)	0.0037	0.0037	0.4082
$(\alpha, \beta, \sigma_w)$ (-0.147, 0.98, 0.166)	0.0041	0.0028	0.4294

the distribution. Second, relative to trapezoidal-based integration, employing the discrete binning approach upon which the DNF is based exhibits comparable levels of accuracy achieved at significantly less computational expense.

5.2 Parameter Estimates

For each parameter set 1000 series of length $T = 2000$ are simulated from equations (18) and (19). For each simulated series the parameter vector is estimated using the DNF. Furthermore, for each simulated series, smoothed estimates of daily volatility are generated from equation 24. Table 2 shows the mean and root mean squared error (RMSE) for both the DNF and NFML methods as applied to the parameter estimation problem.

Examining the results of the DNF with 25 intervals and the NFML with 25 nodes, it is seen that the DNF has lower RMSE in all cases and lower bias in all but two cases ($\alpha = -0.147, \beta = 0.98$). In addition to this, the DNF with 25 intervals outperforms (both bias and RMSE) the NFML with 50 nodes in the first two parameter sets. Examining the results of the DNF method using 50 intervals reveals that it outperforms the NFML with both 25 and 50 nodes for all parameter sets.

<i>Method</i>	$(\alpha, \beta, \sigma_w)$			$(\alpha, \beta, \sigma_w)$			$(\alpha, \beta, \sigma_w)$		
	(-0.736, 0.90, 0.363)			(-0.368, 0.95, 0.260)			(-0.147, 0.98, 0.166)		
<i>NFML</i>									
$N = 25$	-0.812	0.890	0.406	-0.426	0.942	0.294	-0.194	0.974	0.197
	(0.199)	(0.027)	(0.068)	(0.124)	(0.017)	(0.052)	(0.083)	(0.011)	(0.043)
$N = 50$	-0.776	0.895	0.368	-0.406	0.945	0.264	-0.178	0.976	0.169
	(0.168)	(0.023)	(0.041)	(0.106)	(0.014)	(0.032)	(0.067)	(0.009)	(0.024)
<i>DNF</i>									
$N = 25$	-0.765	0.896	0.365	-0.397	0.946	0.264	-0.201	0.973	0.189
	(0.163)	(0.022)	(0.040)	(0.098)	(0.013)	(0.030)	(0.068)	(0.009)	(0.030)
$N = 50$	-0.765	0.896	0.364	-0.395	0.946	0.263	-0.169	0.977	0.169
	(0.159)	(0.021)	(0.041)	(0.100)	(0.013)	(0.031)	(0.058)	(0.008)	(0.022)

Table 2: Mean and RMSE for NFML and DNF methods.

5.3 Volatility Estimates

To evaluate the ability of the DNF and NFML methods to extract estimates of latent volatility the grand RMSE proposed by Jacquier et al (1994) is employed. For each simulated series, the volatility at each time step is estimated using parameter estimates from section 5.2, with the grand RMSE is given by

$$RMSE_G = \sqrt{\frac{1}{1000(T-199)} \sum_{i=1}^{1000} \sum_{t=100}^{T-100} (\sigma_{i,t}^2 - \hat{\sigma}_{i,t}^2)^2}. \quad (25)$$

Where, $T = 2000$, $\sigma_{i,t}^2$ is the volatility simulated at period t on the i^{th} simulation of equations 18 and 19, and $\hat{\sigma}_{i,t}^2$ is the estimate of the volatility.

Table 3 contains the $RMSE_G$ for the NFML method with 25 and 50 nodes as well as the RMSE for the DNF with 25 and 50 intervals. Results in Table 3 shows that the performance of the DNF method is comparable to the NFML. The DNF method seems to generate marginally more accurate estimates of volatility irrespective of the number of intervals used. This result is consistent with the Monte-Carlo results reported in Section 5.2 in that reducing the number of intervals used in the filtering procedure from 50 to 25 has little impact on the DNF performance.

	$(\alpha, \beta, \sigma_w)$	$(\alpha, \beta, \sigma_w)$	$(\alpha, \beta, \sigma_w)$
	$(-0.736, 0.90, 0.363)$	$(-0.368, 0.95, 0.260)$	$(-0.147, 0.98, 0.166)$
<i>NFML</i>			
$N = 25$	6.08	5.24	4.45
$N = 50$	6.03	5.20	4.39
<i>DNF</i>			
$N = 25$	5.98	5.22	4.33
$N = 50$	5.98	5.22	4.33

Table 3: Volatility RMSE for NFML and DNF methods.

6 Empirical Application

This section considers an empirical application of the DNF method, whereby the performance of the DNF method will be compared with standard QML and realised volatility (RV) estimates. Relative performance will be assessed in terms of in sample volatility estimation and out of sample volatility forecasting.

The dataset upon which this comparison is based, consists of intra-day data on Japanese yen (JPY) and US dollar (USD) futures obtained from Tick Data Inc (www.tickdata.com). The dataset covers the time period of 2 January 1990 to 31 March 2000, a total of 2586 trading days. Data is recorded tick by tick giving a total of 3,494,384 observations. To overcome issues relating to market microstructure bias, the tick data was collated into 40 minute intervals³, or 10 intra-day trading periods. From these intra-day trading periods, a dataset of 2586 daily returns and RV estimates are calculated.

Estimates of daily RV may be constructed from the cumulation of the cross-products of intra-day returns. Thus an ex-post estimate of one day volatility is given by the sum of the squares of intra day returns sampled at an appropriate frequency,

$$RV_t = \sum_{j=1}^{1/\Delta} r_{t-1+j\Delta}^2 \quad (26)$$

where RV_t is daily volatility at time t , $\Delta = 1/n$ where n is the number of intra-day periods, and $r_{t-1+j\Delta}$ is the return realised during each of the n periods within the trading day. Andersen, Bollerslev, Diebold and Labys (2003)

³The intra-day interval of 40 minutes was chosen using the volatility signature plot method of Anderson, Bollerslev, Diebold, and Labys (1999).

show that RV, in an ex-post sense precisely captures the evolution of daily FX volatility in that daily spot FX returns standardised by RV are approximately Gaussian. Standardising returns by volatility estimates generated by competing approaches such as GARCH models invariably lead to distributions that are leptokurtotic, albeit to a lesser degree than the raw data. Andersen *et. al.* (2003) suggest that the superior performance of RV is due to its ability to generate volatility estimates that quickly adapt to changes in the prevailing level of volatility. This section will reveal that SV models based on the DNF method (in comparison to QML) also have the capability to adapt to abrupt changes in the level of volatility. Given daily data, the DNF approach can generate volatility estimates and predictions comparable to RV based on intra-day returns.

To construct the daily intra-day volatility the intra-day data requires deseasonalising. This deseasonalising takes the form put forward in Andersen *et. al.* (2003) whereby seasonal factors are estimated by averaging individual squared returns in the 10 intra-day intervals such that:

$$s_i^2 = \frac{1}{T} \sum_{t=1}^T r_{i,t}^2 \quad i = 1, \dots, 10$$

where $r_{i,t}^2$ is the return in the i^{th} period of day t . Based on these seasonal factors, each intra-day return can be deseasonalised as:

$$\tilde{r}_{i,t} = \frac{r_{i,t}}{s_i} \quad i = 1, \dots, 10; \quad t = 1, \dots, T$$

Where s is the sample standard deviation of the entire set of intra-day returns. From these deseasonalised intra-day returns the daily realised volatility is calculated using equation 26.

Since realised volatility is an ex-post estimate of the volatility that prevailed on a particular day, a time series model is fitted to the series of daily realised volatility. This serves to denoise the realised volatility series such that inference may be made about the overall level of volatility. Furthermore, fitting a model to realised volatility allows for the prediction of volatility in subsequent periods. Andersen *et. al.* (2003) suggest fitting a fractionally integrated ARMA

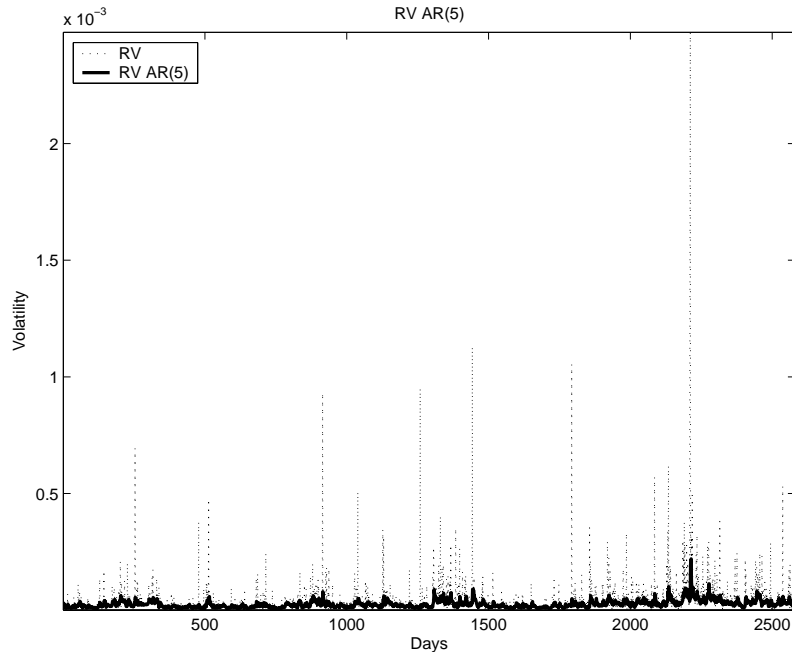


Figure 1: Plot of daily RV and AR(5) smoothed RV.

(ARIMA(p,d,q)) to the logarithm of daily RV. For this comparative exercise, this paper follows Andersen, Bollerslev and Meddahi (2002) and estimates an AR(5) directly from the RV series.

Given the entire dataset, Figures 1, 2 and 3 plot in-sample daily RV along with smoothed estimates of daily volatility obtained from an AR(5) fit directly to daily RV, and smoothed estimates from the QML and New Filter (both utilising daily returns). Figures 2 and 3 reveal that in comparison to standard QML estimates, volatility estimates based on the proposed filtering method are quickly adapting to abrupt changes in the level of volatility. This pattern of changing volatility is similar to that observed in Figure 1 when an AR(5) process is directly fitted to the RV series. The ability of the proposed filter to quickly incorporate changes in the level of volatility is important as it circumvents the need to incorporate intra-day returns in the volatility calculation.

To assess the forecasting performance of each model the sample is split into an estimation period of the initial 1586 trading days along with a sample for forecast evaluation consisting of the final 1000 trading days. Parameter es-

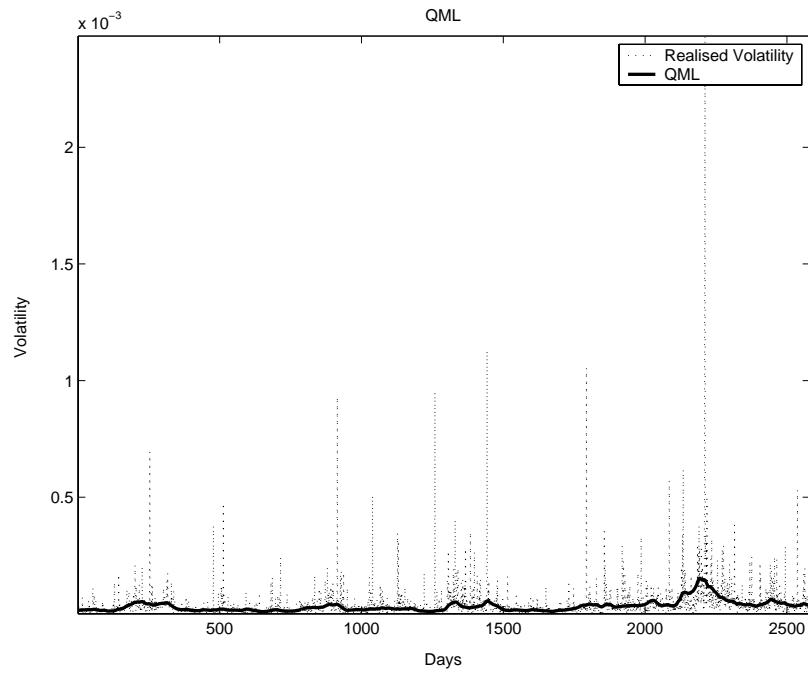


Figure 2: Plot of daily RV and smoothed daily volatility estimates based on QML.

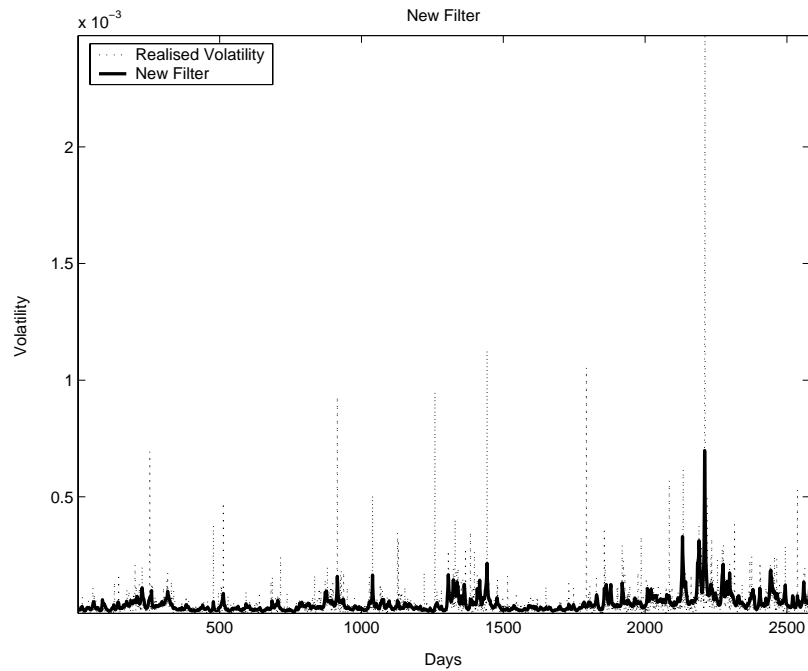


Figure 3: Plot of daily RV and smoothed estimates of volatility based on the proposed filter.

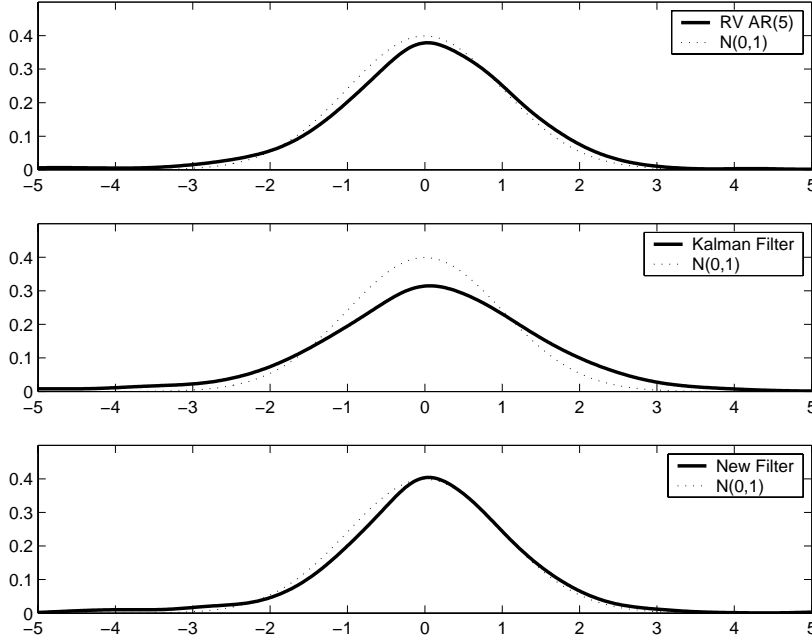


Figure 4: Kernel density estimates of daily returns from the forecast period standardised by one day ahead volatility forecasts (solid lines). $N(0, 1)$ reference is also shown (dotted lines).

timination is carried out within the estimation period, and given these parameters, one-day ahead forecasts are generated from the forecast evaluation sample. Volatility predictions from the proposed filter are taken to be the expected value of volatility from a one-step ahead prediction of the distribution of the state variable,

$$E(\sigma_t^2 | y_{t-1}) = \sum_{i=1}^N \exp(z_i) p_t(x \in (w^i, w^{i+1}] | y_{t-1}, \theta). \quad (27)$$

While there is no single accepted method to evaluate volatility forecasts, Figure 4 reports the density of daily returns standardised by the appropriate one-day ahead volatility forecast. The forecasting performance of each model is given by its ability to generate standardised returns that are $N(0, 1)$. Figure 4 reveals that the DNF produces volatility forecasts that are superior to those of the QML method. Mean squared error (MSE) estimates obtained from comparing the standardised return distributions of the DNF and QML to the $N(0, 1)$ distribution are 2.22×10^{-4} and 1.4×10^{-3} respectively. Volatility predictions

based on the DNF method also outperform those given the direct AR(5) fit to RV which lead to a MSE of 2.94×10^{-4} .

Results from this section reveal that the proposed filter has the ability to quickly adapt to changes in the level of the state variable. This is a feature that distinguishes non-linear filtering methods from standard Kalman Filter based methods. Within the context of SV models, applying the proposed filtering method to daily returns captures characteristics that are otherwise only revealed when intra-day returns are considered.

7 Conclusion

Estimation of latent variable models is often problematic, while simple approaches are available, the conditions under these are applicable are restrictive. Even though methods for estimating these models avoid these problems have been proposed, they are often complex and computationally burdensome to implement. The central contribution of this paper is that it suggests an alternative non-linear filtering (DNF) method, one that avoids the computational burden of competing approaches. While being simpler to implement, it still retains the accuracy of more complex methods.

In this paper, the performance of the DNF method is analysed, from three perspectives in the context of SV models, a common latent variable process dealt with in financial econometrics. The accuracy of the DNF integration scheme was compared with that of Kitagawa (1987) an exercise that highlighted two important points. Integration using a discretisation based on equally spaced points is superior to that using many points placed near the expected value of the latent variable. In comparison to the integration scheme of Kitagawa (1987) the DNF method is more computationally efficient. In terms of parameter and volatility estimation, in relation to the NFML (of Watanabe 1999), the DNF method proves to be both more accurate, and robust to the number of intervals chosen.

The DNF method was applied to the estimation of SV models using daily FX futures data. It was shown that this method outperforms the Kalman-filter

based QML SV method in terms of its ability to rapidly adjust to changes in the level of volatility, and thus forecast the distribution of daily returns. Furthermore, the distributional forecasts obtained from the DNF perform admirably well when compared to those based on realised volatility (constructed from intra-day data).

While a univariate application have been considered in this paper, the DNF method may be extended to higher dimensional problems. Given the computational efficiencies of the DNF integration scheme, it is more amenable to multivariate problems in comparison to the methods suggested by Kitagawa (1987). Extensions to such problems are certainly an avenue for future research.

References

- Andersen, T.G., 1996, "Return volatility and trading volume: an information flow interpretation of stochastic volatility", *Journal of Finance*, **51**, 169-204.
- Andersen, T.G., and Bollerslev, T., and Diebold, F.X., and Labys, P., 1999, "(Understanding, optimizing, using and forecasting) realized volatility and correlation", Working Paper, University of Pennsylvania.
- Andersen, T.G., and Bollerslev, T., and Diebold, F.X., and Labys, P., 2003, "Modeling and Forecasting realized volatility", *Econometrica*, **71**, 579-625.
- Andersen, T.G., and Bollerslev, T., and Meddahi, N., 2002, "Correcting the errors: a note on volatility forecast evaluation based on high-frequency data and realized volatilities", Working Paper 322, Northwestern.
- Clark, P., 1973, "A subordinated stochastic process model with finite variance for speculative process", *Econometrica*, **41**, 135-155.
- Ghysels, E. and Harvey, A.C. and Renault, E., 1996, "Stochastic volatility", in G.S. Maddala and C.R. Rao (eds), *Statistical Methods in Finance*, North Holland, Amsterdam.
- Harvey, A.C. (1981). *The Econometric Analysis of Time Series*. Phillip Allan: Oxford.

- Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge university Press, Cambridge.
- Harvey, A.C. and Ruiz, E. and Shephard, N., 1994, “Multivariate stochastic variance models”, *Review of Economic Studies*, **61**, 247-264.
- Jacquier, E. and Polson, G. and Rossi, P.E., 1994, “Bayesian analysis of stochastic volatility models”, *Journal of Business and Economic Statistics*, **12**, 371-389.
- Kalman, R.E., “A New Approach to linear filtering and prediction problems”, *Transactions of ASME, Journal of Basic Engineering*, **80D**, 35-45.
- Kitagawa, G., 1987, “Non-gaussian state-space modeling of non-stationary time series”, *Journal of the American Statistical Association*, **82**, 1032-1062.
- Martin, D.R. and Raferty, A.E. (1987). “Robustness, computation and non-Euclidean models” *Journal of the American Statistical Association*, **82**, 1044-1050.
- Ruiz, E., 1994, “Quasi-maximum likelihood estimation of stochastic volatility models”, *Journal of Econometrics*, **63**, 289-306.
- Shephard, N., 1996, “Statistical aspects of ARCH and stochastic volatility”, in D. R. Cox, D. V. Hinkley and O. E. Barndorf-Nielsen (eds), *Time Series Models in Econometrics, Finance and Other Fields*, Chapman & Hall, London, 1-67.
- Tanizaki, H., 1993, “Nonlinear filters: Estimation and applications (Lecture notes in economics and mathematical systems)”, Springer-Verlag, Berlin.
- Tauchen, G. and Pitts, M., 1983, “The price variability-volume relationship on speculative markets”, *Econometrica*, **51**, 485-505.
- Watanabe, T., 1999, “A non-linear filtering approach to stochastic volatility models with an application to daily stock returns”, *Journal of Applied Econometrics*, **14**, 101-121.