

# Martingale unobserved component models

NEIL SHEPHARD\*

*Nuffield College, New Road, Oxford OX1 1NF, UK and  
Department of Economics, University of Oxford  
neil.shephard@economics.ox.ac.uk*

10th February 2013

## Abstract

I discuss models which allow the local level model, which rationalised exponentially weighted moving averages, to have a time-varying signal/noise ratio. I call this a martingale component model. This makes the rate of discounting of data local. I show how to handle such models effectively using an auxiliary particle filter which deploys  $M$  Kalman filters run in parallel competing against one another. Here one thinks of  $M$  as being 1,000 or more. The model is applied to inflation forecasting. The model generalises to unobserved component models where Gaussian shocks are replaced by martingale difference sequences.

Keywords: auxiliary particle filter; EM algorithm; EWMA; forecasting; Kalman filter; likelihood; martingale unobserved component model; particle filter; stochastic volatility.

JEL codes: C01; C14; C58; D53; D81

## 1 Introduction

When I went to the LSE in 1986 as a new graduate student I wanted to study time series. My first supervisor was Jim Durbin, who was excited about his new paper Harvey and Durbin (1986) which used time series unobserved component models to estimate the size of a public policy intervention. Jim was about to retire and so when Andrew Harvey returned from sabbatical I asked if I could work with him as a Ph.D. student.

The central unobserved component model is the Gaussian “random walk plus noise model” or “local level model”. This is a profoundly important model for it (1) rationalises exponentially weighted moving average forecasting; (2) is the simplest state space model which can be handled through the Kalman filter and (3) through its analysis lead to the rational expectations school of forward looking expectations in macroeconomics (for good and for bad). I learnt time series modelling from Andrew through thinking about this model and its extensions and what it implies for the degree of “discounting the data.” Such models should be flexible but sensible and importantly

---

\*This paper was written in honour of Andrew C. Harvey’s 65 birthday. I am grateful to James W. Taylor for introducing me to the non-model based EWMA literature which allows the discount parameter to change through time as a function of past data. The detailed comments of S.J. Koopman and two referees were also very helpful, as well as suggestions from Mike Pitt.

should fit the data. Much of the flavour of this approach can be seen in my still favourite time series book Harvey (1981), the exhaustive Harvey (1989) and was elegantly broadcast in Durbin and Koopman (2001, Ch. 2). The question answered here is how should one allow this rate of discounting to change through time?

I advocate the following solution. Replace the random walk with a martingale and the independent and identically distribution (i.i.d.) noise with a martingale difference sequence. I call this class of models “Martingale unobserved component models”. These martingales are parameterised through stochastic volatility (SV) processes.

I started working on stochastic volatility before I left the LSE in 1991. Charles Goodhart asked for thoughts on how to remove the time varying diurnality seen in the volatility in exchange rate markets, which had been clearly revealed by the work of Richard Olsen and his colleagues in Zurich. Andrew and Esther Ruiz were working on seasonality and so we discussed this challenge. To put it into their framework we came up with a “stochastic variance model”, where the returns could be transformed into a linear state space form and so handled using their methods. Once we had that we went back to the simplest model as being interesting in its own right (and forgot about the diurnality). This model is now most accurately called a log-normal stochastic volatility model. At the time we thought “stochastic variance models” were new but we found from Stephen Taylor the existing work on the topic. Our initial multivariate work was published in Harvey, Ruiz, and Shephard (1994). A discussion of the history of SV models is given in Shephard (2005, Ch. 1). The linkage of SV with both realised volatility (e.g. Barndorff-Nielsen and Shephard (2002) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008)) and simulation based inference has meant that SV models are now extremely popular in econometrics and have been a common theme to much of my research in the last 20 years.

Martingale unobserved component models parameterised through stochastic volatility innovations are related to Harvey, Ruiz, and Sentana (1992) and Fiorentini, Sentana, and Shephard (2004), but my direct past connections to it include Shephard (1994) and, for example, Bos and Shephard (2006). The latter paper has an extensive discussion of the literature on this topic. I thought about writing this paper after reading Stock and Watson (2007). Intellectually, one can think of the contribution of this paper as arguing for a different parameterisation from that used by Stock and Watson (2007), as well as employing a different computational device. I think there are also some attractions in thinking about the models as martingale unobserved component models, rather than starting with the default Gaussian model associated with the Kalman filter.

Computationally I handle the model using an auxiliary particle filter, implementing it by running thousands of Kalman filters in parallel, allowing the data to select which ones blossom as time

evolves. I will use the particular structure of the model to do this statistically efficiently, the most related work to this is Chen and Liu (2000) as well as Fearnhead and Clifford (2003). I should also note the work of Koopman and Bos (2004) and Creal (2012) around this topic, while Stock and Watson (2007) use the Kim, Shephard, and Chib (1998) approach to SV. The related work in macroeconomics includes Cogley, Primiceri, and Sargent (2010), D'Agostino, Gambetti, and Giannone (2013), Fernandez-Villaverde, Guerron, Rudio-Ramirez, and Uribe (2010) and Caldara, Fernandez-Villaverde, Guerron, Rudio-Ramirez, and Wen (2012).

The rest of this paper has the following form. In Section 2 a martingale unobserved component model is defined and various special cases are discussed. A key feature of this model is that it has a simple conditional probabilistic structure. This is discussed in Section 3, where the relations to the Kalman filter are brought out. Section 4 focuses on how the model can be handled using a particular type of particle filter, which allows both state and parameter estimation. In Section 5 the model is used to analyse a time series of quarterly inflation from the United States. In Section 6 some conclusions are made.

## 2 Martingale unobserved component models

### 2.1 A first example

I start by considering a local level version of the martingale unobserved component model

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t^*, \\ \mu_{t+1} &= \mu_t + \eta_t^*, \quad t = 1, 2, \dots, n, \end{aligned}$$

with

$$\mathbb{E} \begin{pmatrix} \varepsilon_t^* \\ \eta_t^* \end{pmatrix} \Big| \mathcal{F}_{t-1}^{\varepsilon^*, \eta^*, \mu_0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

where  $\mathcal{F}_t^x$  generically denotes the past and current information of an arbitrary  $x$  process, that is  $\mathcal{F}_0^x, x_1, \dots, x_t$ , where  $\mathcal{F}_0^x$  is some prior. So here  $\varepsilon_t^*$  and  $\eta_t^*$  are martingale difference sequences with respect to their joint natural filtration  $\mathcal{F}^{\varepsilon^*, \eta^*, \mu_0}$ , while  $\mu_t$  is a  $\mathcal{F}^{\varepsilon^*, \eta^*, \mu_0}$ -martingale. The key idea in martingale unobserved component models is that the filtration is not with respect to the observables but with respect to the components.

### 2.2 Parameterising the model

An elegant and mainstream way of parameterising martingales is through stochastic volatility, e.g. Harvey, Ruiz, and Shephard (1994), Ghysels, Harvey, and Renault (1996) and Shephard (2005). Although at first sight the use of SV looks at ad hoc it is well known that large classes of martingales (i.e. basically continuous sample path martingales with absolutely continuous quadratic variation) can be represented in this way, e.g. Shephard (2005, Ch. 1).

Here two “discrete time stochastic volatility” models are used, with

$$\begin{aligned} y_t &= \mu_t + \sigma_{\varepsilon,t}\varepsilon_t, & (\varepsilon_t, \eta_t)' &\stackrel{iid}{\sim} N(0, I_2), \\ \mu_{t+1} &= \mu_t + \sigma_{\eta,t}\eta_t, \end{aligned} \tag{1}$$

where  $t = 1, 2, \dots, n$ . I will call  $\sigma_{\varepsilon,t}$  and  $\sigma_{\eta,t}$  transitory and permanent component volatilities. Throughout I will assume that the non-negative  $(\sigma_{\varepsilon,t}, \sigma_{\eta,t})$  as a process is probabilistically independent from  $(\varepsilon_t, \eta_t)$  as a process. More sophisticated dependence is possible, allowing features such as statistical leverage, but that will not be attempted here.

The model is completed by a prior on the initial location. Here I decided to put this on  $\mu_2$  and use a simple prior. This is built by making inference conditional on an initial datapoint  $y_1$  and so take the prior

$$\mu_2 | \mathcal{F}_1^{y, \sigma_\varepsilon, \sigma_\eta} \sim N(y_1, \sigma_{\varepsilon,1}^2 + \sigma_{\eta,1}^2). \tag{2}$$

This prior can be motivated by noting that  $\mu_2 = y_1 - \sigma_{\varepsilon,1}\varepsilon_1 + \sigma_{\eta,1}\eta_1$ . Of course more rigorous treatment of initial conditions can be provided, see for example, de Jong (1991) and Durbin and Koopman (2001, Ch. 2).

The following are important special cases:

- *Random walk plus strict white noise model.* This is where  $\sigma_{\varepsilon,t}$  and  $\sigma_{\eta,t}$  are both constant, meaning the model is Gaussian. This is the traditional model which can be computationally efficiently handled by the Kalman filter. It is sometimes called the local level model.
- *Martingale plus strict white noise model.* This is where  $\sigma_{\varepsilon,t} = \sigma_\varepsilon$  but  $\sigma_{\eta,t}$  can vary through time. This means that the location is a martingale, rather than a random walk.
- *Random walk plus SV scale.* This is where  $\sigma_{\eta,t} = q^{1/2}\sigma_{\varepsilon,t}$ , but  $\sigma_{\varepsilon,t}$  can vary through time allowing the scale of the volatilities of change through time but the model has a fixed signal to noise ratio  $q$ . This model is studied in, for example, Koopman and Bos (2004) who develop special computational methods to handle it.

Various local models for the volatilities will be defined and analysed in a moment.

This martingale unobserved component model class extends to cover all Gaussian state space models replacing all i.i.d. noise terms with martingale difference sequences. Here we detail two classic examples.

## 2.3 Other martingale unobserved component models

### 2.3.1 Local linear trend plus cycle

To be explicit, the univariate local linear trend plus cycle version of a martingale unobserved component model would have

$$\begin{aligned} y_t &= \mu_t + \sqrt{1 - \rho^2} \psi_t + \varepsilon_t^*, \\ \mu_{t+1} &= \mu_t + \beta_t + \eta_t^*, \\ \beta_{t+1} &= \beta_t + \zeta_t^*, \\ \begin{pmatrix} \psi_{t+1} \\ \dot{\psi}_{t+1} \end{pmatrix} &= \rho \begin{pmatrix} \cos(\lambda_c) & \sin(\lambda_c) \\ -\sin(\lambda_c) & \cos(\lambda_c) \end{pmatrix} \begin{pmatrix} \psi_t \\ \dot{\psi}_t \end{pmatrix} + \begin{pmatrix} \kappa_t^* \\ \dot{\kappa}_t^* \end{pmatrix}. \end{aligned} \quad (3)$$

In this model  $\mu_t$  is interpreted as the trend,  $\beta_t$  as a time varying slope and  $\sqrt{1 - \rho^2} \psi_t$  as the cycle. In some applications researchers have preferred to use a “smooth trend” component, which imposes a priori  $\eta_t^* = 0$  (which connects to the literature on cubic splines). In the martingale unobserved component model the vector  $(\varepsilon_t^*, \eta_t^*, \zeta_t^*, \kappa_t^*, \dot{\kappa}_t^*)'$  is a martingale difference sequence with respect to its own natural filtration conditional on  $\beta_0, \mu_0, \psi_0, \dot{\psi}_0$ . Here  $\rho \in (0, 1)$ , which means that if  $(\kappa_t^*, \dot{\kappa}_t^*)'$  is weak white noise then  $\psi_t$  has a linear ARMA(2,1) representation with the complex autoregressive roots. Further  $\lambda_c$  is a frequency in radians. I have parameterised it so that  $\text{Var}(\sqrt{1 - \rho^2} \psi_t) = \text{Var}(\kappa_t^*) = \text{Var}(\dot{\kappa}_t^*)$ , and thus does not vary with  $\rho$  or  $\lambda_c$ . The martingale difference sequences can be parameterised as SV processes

$$\varepsilon_t^* = \sigma_{\varepsilon,t} \varepsilon_t, \quad \eta_t^* = \sigma_{\eta,t} \eta_t, \quad \zeta_t^* = \sigma_{\zeta,t} \zeta_t, \quad \kappa_t^* = \sigma_{\kappa,t} \kappa_t, \quad \dot{\kappa}_t^* = \sigma_{\dot{\kappa},t} \dot{\kappa}_t, \quad (4)$$

where  $(\varepsilon_t, \eta_t, \zeta_t, \kappa_t, \dot{\kappa}_t)' \stackrel{iid}{\sim} N(0, I_5)$  and stochastically independent from  $(\sigma_{\varepsilon,t}, \sigma_{\eta,t}, \sigma_{\zeta,t}, \sigma_{\kappa,t})'$ .

Of course a related simpler model writes

$$\begin{aligned} y_t &= \mu_t + \psi_t, \\ \psi_{t+1} &= \rho \psi_t + \sqrt{1 - \rho^2} \kappa_t^*, \quad \rho \in (-1, 1), \end{aligned} \quad (5)$$

which has an autoregressive measurement error (whose marginal distribution does not depend upon  $\rho$ ), while a moving average measurement error version can be achieved by writing

$$\begin{aligned} y_t &= \mu_t + (1 \ 0) \psi_t, \\ \psi_{t+1} &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \psi_t + \begin{pmatrix} 1 \\ \lambda \end{pmatrix} \frac{1}{\sqrt{1 + \lambda^2}} \kappa_t^*, \quad \lambda \in (-1, 1). \end{aligned} \quad (6)$$

Again  $(1 \ 0) \psi_t$  is setup so that the implied marginal distribution does not depend upon  $\lambda$ .

### 2.3.2 Multivariate martingale unobserved component local level model

Likewise a  $d$ -dimensional vector martingale unobserved component version of the local level model would have the form

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t^*, \\ \mu_{t+1} &= \mu_t + \eta_t^*, \end{aligned} \quad (7)$$

which is parameterise as  $\varepsilon_t^* = \sigma_{\varepsilon,t}\varepsilon_t$  and  $\eta_t^* = \sigma_{\eta,t}\eta_t$ , with  $(\varepsilon_t', \eta_t')' \stackrel{iid}{\sim} N(0, I_{2d})$  which are stochastically independent from  $(\sigma_{\varepsilon,t}, \sigma_{\eta,t})'$ . Here  $\sigma_{\varepsilon,t}$  and  $\sigma_{\eta,t}$  are  $d \times d$  dimensional time-varying volatility matrices.

### 3 Conditional properties

#### 3.1 General case

This stochastic volatility parameterisation of the martingale unobserved component model is a special case of the partially or conditionally Gaussian time series, a class of state space models defined in Shephard (1994). The core feature of this model is that it can be placed into a Gaussian state space form by conditioning on, in this special case, the volatilities (note there is a significant and influential literature on the case where we condition on variables with discrete support to deliver a Gaussian state space form, e.g. Ackerson and Fu (1970), Akashi and Kumamoto (1977) and Carter and Kohn (1994)). To start the discussion, focus on the local level martingale unobserved component model (1) and assume that  $\sigma_{\varepsilon,t}, \sigma_{\eta,t} > 0$  for all  $t$ .

A key characteristic of this model is the “signal/noise ratio process”

$$q_t = \sigma_{\eta,t}^2 / \sigma_{\varepsilon,t}^2. \quad (8)$$

Throughout  $q_{1:t} = (q_1, \dots, q_t)'$ .

**Remark 1** *In the univariate martingale local linear trend plus cycle model (3) there are three signal/noise processes  $q_{\eta,t} = \sigma_{\eta,t}^2 / \sigma_{\varepsilon,t}^2$ ,  $q_{\zeta,t} = \sigma_{\zeta,t}^2 / \sigma_{\varepsilon,t}^2$ ,  $q_{\kappa,t} = \sigma_{\kappa,t}^2 / \sigma_{\varepsilon,t}^2$ . This has the virtue that the signal/noise processes are invariant to rescaling the data. In the smooth trend plus cycle model there are only two signal/noise processes  $q_{\zeta,t}$  and  $q_{\kappa,t}$ .*

**Remark 2** *In the multivariate martingale local level unobserved component model (7) there is a matrix signal/noise process  $Q_t = \Sigma_{\varepsilon,t}^{-1/2} \Sigma_{\eta,t} \Sigma_{\varepsilon,t}^{-1/2}$ , where  $\Sigma_{\varepsilon,t} = \sigma_{\varepsilon,t} \sigma'_{\varepsilon,t}$  and  $\Sigma_{\eta,t} = \sigma_{\eta,t} \sigma'_{\eta,t} = \Sigma_{\varepsilon,t}^{1/2} Q_t \Sigma_{\varepsilon,t}^{1/2}$ . This signal/noise process is invariant to rescaling and rotation of the data.*

#### 3.2 Gaussian state space models

Now take a step backwards and recall the general  $d$ -dimensional Gaussian state space form (e.g. Durbin and Koopman (2001, p. 67)), where  $\mu_t$  is the general state. This takes on the form of observing

$$y_t = Z_t \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, H_t), \quad (9)$$

$$\mu_{t+1} = T_t \mu_t + R_t \eta_t, \quad \eta_t \sim N(0, Q_t), \quad (10)$$

where the sequences  $Z_t$ ,  $H_t$ ,  $Q_t$ ,  $T_t$  and  $R_t$  are assumed to be non-stochastic and the  $\varepsilon_1, \dots, \varepsilon_n$ ,  $\eta_1, \dots, \eta_n$  are stochastically independent. The conditional model is completed by the prior  $\mu_1 | \mathcal{F}_0^y \sim N(m_{1|0}, P_{1|0})$ .

Then  $\mu_{t+1} | \mathcal{F}_t^y \sim N(m_{t+1|t}, P_{t+1|t})$ , where the mean and variance updates according to the Kalman filter recursions

$$m_{t+1|t} = T_t m_{t|t-1} + T_t P_{t|t-1} Z_t' F_t^{-1} (y_t - Z_t m_{t|t-1}) \quad (11)$$

$$= T_t m_{t|t-1} + K_t v_t, \quad K_t = T_t P_{t|t-1} Z_t' F_t^{-1} \quad (12)$$

$$= K_t y_t + L_t m_{t|t-1}, \quad L_t = T_t - K_t Z_t \quad (13)$$

$$= T_t \{ W_t y_t + (I - W_t Z_t) m_{t|t-1} \}, \quad W_t = P_{t|t-1} Z_t' F_t^{-1}, \quad (14)$$

$$P_{t+1|t} = T_t P_{t|t-1} T_t' - T_t P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1} T_t' + R_t Q_t R_t' \quad (15)$$

$$= T_t P_{t|t-1} L_t' + R_t Q_t R_t', \quad L_t = T_t (I - W_t Z_t) \quad (16)$$

$$= T_t P_{t|t-1} (I - W_t Z_t)' T_t' + R_t Q_t R_t', \quad (17)$$

$$F_t = Z_t P_{t|t-1} Z_t' + H_t, \quad (18)$$

and as a side product

$$y_t | \mathcal{F}_{t-1}^y \sim N(Z_t m_{t|t-1}, F_t). \quad (19)$$

**Remark 3** In the important special case were  $H_t = \sigma_{\varepsilon,t}^2 I$ ,  $R_t = \sigma_{\varepsilon,t}$  and I write  $P_{t|t-1} = \sigma_{\varepsilon,t}^2 P_{t|t-1}^*$ , where  $\sigma_{\varepsilon,t}$  is a non-stochastic sequence, then

$$F_t = \sigma_{\varepsilon,t}^2 \left( Z_t P_{t|t-1}^* Z_t' + I \right), \quad W_t = P_{t|t-1}^* Z_t' \left( Z_t P_{t|t-1}^* Z_t' + I \right)^{-1} \quad (20)$$

$$P_{t+1|t} = \sigma_{\varepsilon,t}^2 \left\{ T_t P_{t|t-1}^* (I - W_t Z_t)' T_t' + Q_t \right\}, \quad (21)$$

and so  $P_{t+1|t}^* = \frac{\sigma_{\varepsilon,t}^2}{\sigma_{\varepsilon,t+1}^2} \left\{ T_t P_{t|t-1}^* (I - W_t Z_t)' T_t' + Q_t \right\}$ . Hence the  $P_{t+1|t}^*$  process, and so  $W_t$  and  $m_{t|t-1}$ , depends not on the level of volatility but on the change in the level of volatility. In the special univariate case with  $Z_t = T_t = I$ , then this can be more compactly written as

$$m_{t+1|t} = m_{t|t-1} + \frac{p_{t|t-1}}{p_{t|t-1} + h_t} (y_t - m_{t|t-1}), \quad p_{t+1|t} = \frac{p_{t|t-1} \sigma_{\varepsilon,t}^2}{p_{t|t-1} + \sigma_{\varepsilon,t}^2} + \sigma_{\varepsilon,t}^2 q_t. \quad (22)$$

### 3.3 Conditioning on the volatilities

For the local level version of the martingale unobserved component model it is useful to condition on the stochastic volatility processes  $(\sigma_{\varepsilon,t}, \sigma_{\eta,t})$ . This conditional model can be handled computationally efficiently using the Kalman filter which becomes

$$m_{t+1|t} = (1 - \omega_t) m_{t|t-1} + \omega_t y_t, \quad (23)$$

where  $E(\mu_{t+1} | \mathcal{F}_t^{y, \sigma_\varepsilon, \sigma_\eta}) = m_{t+1|t}$ . This is driven off the simple recursions

$$\omega_t = \frac{p_{t|t-1}}{p_{t|t-1} + \sigma_{\varepsilon,t}^2}, \quad p_{t+1|t} = \sigma_{\varepsilon,t}^2 (\omega_t + q_t). \quad (24)$$

Of course here  $p_{t+1|t} = E\left\{(\mu_{t+1} - m_{t+1|t})^2 | \mathcal{F}_t^{y, \sigma_\varepsilon, \sigma_\eta}\right\}$ . The conditional likelihood is, via the prediction decomposition

$$\log f(y_2, \dots, y_n | \mathcal{F}_1^y, \mathcal{F}_n^{\sigma_\varepsilon, \sigma_\eta}) \quad (25)$$

$$= -\frac{1}{2} \sum_{t=2}^n \log(p_{t|t-1} + \sigma_{\varepsilon,t}^2) - \frac{1}{2} \sum_{t=2}^n \frac{(y_t - m_{t|t-1})^2}{p_{t|t-1} + \sigma_{\varepsilon,t}^2} \quad (26)$$

$$= -\frac{1}{2} \sum_{t=2}^n \log \sigma_{\varepsilon,t}^2 - \frac{1}{2} \sum_{t=2}^n \log(p_{t|t-1}^* + 1) - \frac{1}{2} \sum_{t=2}^n \frac{(y_t - m_{t|t-1})^2}{\sigma_{\varepsilon,t}^2 (p_{t|t-1}^* + 1)}. \quad (27)$$

**Remark 4** For the local linear trend plus cycle model in section 2.3.1 and Remark 1 the conditional state space form becomes

$$Z_t = (1, 0, \sqrt{1 - \rho^2}, 0), \quad h_t = \sigma_{\varepsilon,t}^2, \quad r_t = \sigma_{\varepsilon,t}^2, \quad Q_t = \text{diag}(q_{\eta,t}, q_{\zeta,t}, q_{\kappa,t}, q_{\kappa,t}),$$

$$T_t = \begin{pmatrix} 1 & 1 & & 0 \\ 0 & 1 & & 0 \\ 0 & 0 & \rho \begin{pmatrix} \cos(\lambda_c) & \sin(\lambda_c) \\ -\sin(\lambda_c) & \cos(\lambda_c) \end{pmatrix} & \end{pmatrix},$$

and hence (20) and (21) applies.

### 3.4 Gaussian case, no volatility clustering

Suppose there is no volatility clustering, so  $\sigma_{\varepsilon,t}^2 = \sigma_\varepsilon^2$  and  $\sigma_{\eta,t}^2 = \sigma_\eta^2$ . This reproduces the celebrated Gaussian random walk plus noise model which can be handled efficiently by the Kalman filter. All the following is well known.

If  $t$  is large and then  $\sigma_\varepsilon^2, \sigma_\eta^2 > 0$  then the updating recursion converges to the solution to the Riccati equation

$$\lim_{t \rightarrow \infty} p_{t|t-1}^* = p^* = q + \frac{p^*}{p^* + 1} = \frac{q + \sqrt{q^2 + 4q}}{2} \in [0, \infty), \quad (28)$$

where, again,  $p_{t|t-1} = \sigma_{\varepsilon,t}^2 p_{t|t-1}^*$ , which means that the updating equation for the conditional mean has the form of

$$m_{t+1|t} = (1 - \omega) m_{t|t-1} + \omega y_t, \quad \omega = \frac{p^*}{p^* + 1} = \frac{q + \sqrt{q^2 + 4q}}{2 + q + \sqrt{q^2 + 4q}} \in [0, 1], \quad (29)$$

a simple discount of the past data. Hence the local level model rationalises the EWMA updating recursion (Muth (1960))

$$m_{t+1|t} = \omega \sum_{j=0}^{\infty} (1 - \omega)^j y_{t-j}. \quad (30)$$



Of course the first difference of the local level model can be written as a first order moving average model  $\Delta y_t = \varsigma_t + \psi \varsigma_{t-1}$  where  $\varsigma_t$  is strict white noise and

$$\psi = \omega - 1 = \frac{-2}{2 + q + \sqrt{q^2 + 4q}} \in [-1, 0]. \quad (31)$$

Breaking away from the time-invariant model the

$$\psi_t = \frac{-2}{2 + q_t + \sqrt{q_t^2 + 4q_t}} \in [-1, 0], \quad (32)$$

is defined here as a “psi process”. One way of directly measuring the the relative weight of a data point at time  $t$  compared to at time  $t - j$  is through  $(1 - \omega)^j$  in (30). I determine when  $j$  is large enough for  $(1 - \omega)^j$  to be around 0.1, which means this is long enough for the data not to have much impact on the forecast. I will call this the “life span process”, which may vary through time as

$$\begin{aligned} s_t &= \log 0.1 / \log(1 - \omega_t) \\ &= \log 0.1 / \log(-\psi_t) \in [0, \infty). \end{aligned} \quad (33)$$

When  $s_t$  is close to zero the forecast is approximately a martingale, while when it is for example 20, it roughly averages the data over the last 20 periods.

When there is no transitory component volatility clustering, that is  $\sigma_{\varepsilon,t}^2 = \sigma_\varepsilon^2$ , then this model is called the martingale plus strict white noise model. This model has a some flexibility for forecasting and is still relatively easy to handle. Now, recalling  $q_t = \sigma_{\eta,t}^2 / \sigma_\varepsilon^2$ ,

$$m_{t+1|t} = (1 - \omega_t) m_{t|t-1} + \omega_t y_t, \quad p_{t+1|t}^* = q_t + \omega_t, \quad \omega_t = \frac{p_{t|t-1}^*}{p_{t|t-1}^* + 1}, \quad (34)$$

where  $m_{t+1|t}$  and  $p_{t+1|t}^*$  are invariant to  $\sigma_\varepsilon^2$ . Clearly  $\mu_{t+1} | \mathcal{F}_t^{y,q}, \sigma_\varepsilon^2 \sim N(m_{t+1|t}, \sigma_\varepsilon^2 p_{t+1|t}^*)$ .

### 3.5 Volatility models

For the martingale unobserved component models the signal/noise ( $q_t$ ) and volatility processes ( $\sigma_t = \sigma_{\varepsilon,t}$ ) can evolve. I focus on the local level type model and initially make these two processes stochastically independently with each following log-Gaussian processes

$$\log q_{t+1} = \log q_t + \theta_q \xi_{q,t}, \quad \begin{pmatrix} \xi_{q,t} \\ \xi_{\sigma,t} \end{pmatrix} \stackrel{iid}{\sim} N(0, I_2), \quad (35)$$

$$\log \sigma_{t+1}^2 = \log \sigma_t^2 + \theta_\sigma \xi_{\sigma,t}. \quad (36)$$

This means that  $\frac{q_{t+1}}{q_t} = \exp(\theta_q \xi_{q,t})$ . Hence for moderate  $\theta_q$  the proportional change is roughly  $(q_{t+1} - q_t) / q_t \simeq \theta_q \xi_{q,t}$ , which makes sense whatever the current value of the volatility due to its attractive scale invariance. In particular then the expected percentage change at any time point is roughly  $E |(q_{t+1} - q_t) / q_t| \simeq \theta_q \sqrt{\frac{\pi}{2}}$ .

This model implies that the level

$$\mu_{t+1} = \mu_t + \sigma_t q_t \eta_t = \mu_t + \sigma_{\mu,t} \eta_t, \quad (37)$$

is like an integrated stochastic volatility model from financial econometrics, although these models are typically setup with a small degree of mean reversion in  $\sigma_{\mu,t}$  — e.g. Harvey, Ruiz, and Shephard (1994), Ghysels, Harvey, and Renault (1996) and Shephard (2005).

Now this model implies that

$$\log \sigma_{\mu,t+1}^2 = \log \sigma_{\mu,t}^2 + \theta_{\mu} \xi_{\mu,t}, \quad \xi_{\mu,t} \sim N(0, 1), \quad (38)$$

and  $\theta_{\mu} \xi_{\mu,t} = \theta_{\sigma} \xi_{\sigma,t} + \theta_q \xi_{q,t}$ . This means that  $\theta_{\mu} = \sqrt{\theta_{\sigma}^2 + \theta_q^2}$ . Of course, this model implies  $\text{Cov}(\xi_{\mu,t}, \xi_{\sigma,t}) = \theta_{\sigma}/\theta_{\mu}$ ,  $\text{Cov}(\xi_{\mu,t}, \xi_{q,t}) = \theta_q/\theta_{\mu}$ .

### 3.6 Encompasses

Stock and Watson (2007) parameterise their changing volatilities using independent random walks

$$\log \sigma_{\mu,t+1}^2 = \log \sigma_{\mu,t}^2 + \theta_{\mu} \xi_{\mu,t}, \quad \log \sigma_{\sigma,t+1}^2 = \log \sigma_{\sigma,t}^2 + \theta_{\sigma} \xi_{\sigma,t}, \quad (\xi_{\mu,t}, \xi_{\sigma,t})' \stackrel{iid}{\sim} N(0, I_2), \quad (39)$$

This means that

$$\log q_{t+1} = \log \sigma_{\mu,t+1}^2 - \log \sigma_{\sigma,t+1}^2 = \log q_t + \theta_{\mu} \xi_{\mu,t} - \theta_{\sigma} \xi_{\sigma,t} = \log q_t + \theta_q \xi_{q,t}, \quad (40)$$

where  $\xi_{q,t} \stackrel{iid}{\sim} N(0, 1)$ , while

$$\text{Var} \begin{pmatrix} \theta_q \xi_{q,t} \\ \theta_{\sigma} \xi_{\sigma,t} \end{pmatrix} = \begin{pmatrix} \theta_{\mu}^2 + \theta_{\sigma}^2 & -\theta_{\sigma}^2 \\ -\theta_{\sigma}^2 & \theta_{\sigma}^2 \end{pmatrix}. \quad (41)$$

This means that a priori  $\theta_q^2 \geq \theta_{\sigma}^2$ , while the shocks are negatively correlated. This is a little concerning for if there is a great deal of volatility clustering then this model implies the signal/noise process must also move around a great deal.

To reconcile the two parameterisation an extended SW parameterisation can be introduced which writes

$$\begin{pmatrix} \xi_{\mu,t} \\ \xi_{\sigma,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} = \begin{pmatrix} u_{1,t} \\ \rho u_{1,t} + \sqrt{1-\rho^2} u_{2,t} \end{pmatrix}, \quad (42)$$

for this then encompasses our alternative. We will see later that allowing a correlation between these two innovations will be supported by the data. Then

$$\begin{aligned} \theta_q \xi_{q,t} &= (\theta_{\mu} - \rho \theta_{\sigma}) u_{1,t} - \theta_{\sigma} \sqrt{1-\rho^2} u_{2,t}, \\ \theta_q &= \sqrt{(\theta_{\mu} - \rho \theta_{\sigma})^2 + \theta_{\sigma}^2 (1-\rho^2)} = \sqrt{\theta_{\mu}^2 + \theta_{\sigma}^2 - 2\rho \theta_{\mu} \theta_{\sigma}}, \\ \text{Cov}(\theta_q \xi_{q,t}, \theta_{\sigma} \xi_{\sigma,t}) &= \theta_{\sigma} \rho (\theta_{\mu} - \rho \theta_{\sigma}) - \theta_{\sigma}^2 (1-\rho^2) = \theta_{\sigma} (\rho \theta_{\mu} - \theta_{\sigma}) \end{aligned}$$

$$\text{Cor}(\theta_q \xi_{q,t}, \theta_\sigma \xi_{\sigma,t}) = \frac{\rho \theta_\mu - \theta_\sigma}{\theta_q} = \frac{\rho \theta_\mu - \theta_\sigma}{\sqrt{\theta_\mu^2 + \theta_\sigma^2 - 2\rho \theta_\mu \theta_\sigma}}.$$

Hence if  $\rho$  is incorrectly imposed to be zero then the implied  $\theta_q$  will be too high if the true  $\rho > 0$ . This is crucial as this parameter determines the degree of discounting used in forecasting. We will see this is exactly what happens empirically.

**Remark 5** *Suppose the Gaussian local level model is parameterised as*

$$y_t = \mu_t + \sigma \varepsilon_t, \quad \mu_{t+1} = \mu_t + q^{1/2} \sigma \eta_t, \quad \mu_t \sim N(m_0, P_0), \quad (\varepsilon_t, \eta_t)' \stackrel{iid}{\sim} N(0, I_2). \quad (43)$$

*It is helpful to write this time series as the vector using the notation  $y \sim LLM(\sigma, q, m_0, P_0)$ . Then if  $a$  and  $b$  are scalar constants  $a + by \sim LLM(b\sigma, q, a + bm_0, b^2 P_0)$ , so the model is closed under location shifts and rescaling and the impact of rescaling is just to scale up and down  $\sigma$  (as well as some impact on the prior). It has no impact on the key time series properties of the model, which are governed through  $q$ . If the model is setup as*

$$y_t = \mu_t + \sigma_\varepsilon \varepsilon_t, \quad \mu_{t+1} = \mu_t + \sigma_\eta \eta_t, \quad \mu_t \sim N(m_0, P_0), \quad (\varepsilon_t, \eta_t)' \stackrel{iid}{\sim} N(0, I_2). \quad (44)$$

*then  $a + by \sim LLM(b\sigma_\varepsilon, b\sigma_\eta, a + bm_0, b^2 P_0)$ . This shows that under a scale shift the two volatilities  $\sigma_\varepsilon$  and  $\sigma_\eta$  must move together, for the implied  $q$  will otherwise change. This suggests it is not so attractive to place independent priors on  $\sigma_\varepsilon$  and  $\sigma_\eta$  (i.e. these two parameters should move together to reflect the scaling of the data), while it may make some sense to make  $q$  and  $\sigma$  independent. This observation is also key in the dynamic case, where scaling of economic data can vary through time. This persuades me to prefer the parameterisation (43) to (44). Of course in practice it is an empirical question as to which model is empirically more successful in terms of fit and parsimony.*

### 3.7 EWMA and adaptive discounting

The Gaussian local level model provides a statistical rationale for EWMA forecasting

$$m_{t+1|t} = (1 - \omega) m_{t|t-1} + \omega y_t. \quad (45)$$

The above models allow the discount parameter to change through time as an unobservable process which can be efficiently learnt using data and Bayes theory.

There is another literature which has directly allowed the discount factor to move in response to past data. The most well known approach is Trigg and Leach (1967) who specify

$$\omega_t = \frac{|A_t|}{|M_t|}, \quad A_t = \phi e_{t-1} + (1 - \phi) A_{t-1}, \quad M_t = \phi |e_{t-1}| + (1 - \phi) M_{t-1}, \quad (46)$$

where  $e_t = y_t - m_{t|t-1}$ . Typically  $\phi = 0.2$ , but this does not enforce  $\omega_t \in (0, 1)$ . Discussions of alternatives are given in, for example, Taylor (2004).

Related to this work is the so-called generalised autoregressive score based models of Creal, Koopman, and Lucas (2008, Section 4.4) which builds an observation driven model based upon a Stock and Watson (2007) type parameterisation.

## 4 Particle filter based analysis

### 4.1 Basics of filtering

In this paper I have decided to use a particle filter as the basis of sequential Bayesian inference on  $\mu_t$ ,  $\sigma_t$  and  $q_t$ . Particle filters are now the established extension of the Kalman filter to deal with non-Gaussian and non-linear state space models. They use simulation to provide filtered estimates of the states and an unbiased estimator of the likelihood. Particle filters can be implemented in various ways, which effect their Monte Carlo statistical efficiency. Early contributions to this include, for example, Gordon, Salmond, and Smith (1993), Liu and Chen (1998), Pitt and Shephard (1999) and Doucet, de Freitas, and Gordon (2001). Modern surveys include, for example, Doucet and Johansen (2011) and Creal (2012).

In this section I assume a Markov model  $q_{t+1}, \sigma_{t+1}^2 | q_t, \sigma_t^2, \theta$ , from which I can simulate. The simplest case is (35)-(36). For now assume that the value of  $\theta$  is known and return to that issue in the next subsection.

The conditional Gaussian structure of the model can be used to improve upon the bootstrap particle filter of Gordon, Salmond, and Smith (1993), integrating out the conditionally Gaussian  $\mu_t$  for the complete model

$$\begin{aligned} y_t &= \mu_t + \sigma_t \varepsilon_t, & \mu_{t+1} &= \mu_t + q_t^{1/2} \sigma_t \eta_t, & (\varepsilon_t, \eta_t, \xi_{q,t}, \xi_{\sigma,t})' &\sim N(0, I_4), \\ \log q_{t+1} &= \log q_t + \theta_q \xi_{q,t}, & \log \sigma_{t+1}^2 &= \log \sigma_t^2 + \theta_\sigma \xi_{\sigma,t}. \end{aligned}$$

In this approach the only “particle state variable” is

$$\alpha_t = (\sigma_t^2, q_t)'. \tag{47}$$

The genesis of this appears in the auxiliary particle filter of Pitt and Shephard (1999), but it is dealt with more extensively by Doucet, Godsill, and Andrieu (2000) and in particular Chen and Liu (2000) and Fearnhead and Clifford (2003). It has also been used by Creal (2012) in this context and by Creal, Koopman, and Zivot (2010). This approach is sometimes called a mixture Kalman filter and is also related to Rao-Blackwellisation.

In effect this auxiliary particle filter simply runs  $M$  Kalman filters in parallel, each with an individual signal/noise ratio and volatility process which nudges the signal/noise rate and volatility in random directions at each time increment. Some of these Kalman filters will run up pretty high likelihoods through time, others will perform poorly — as they have uncompetitive signal/noise

ratios or volatilities. When there is sufficient imbalance in the likelihoods between the different filters, the Kalman filters are resampled. Resampling samples with probability in proportion to the likelihood to produce a new set of  $M$  Kalman filters from their parents. Thus Kalman filters which generate large likelihoods survive and replicate, those which have poor fit die. The particles  $\alpha_t^{(j)}$  are indexed by  $j = 1, 2, \dots, M$ . Like the Kalman filter, the particle filter is run conditioned on  $\theta = (\theta_q, \theta_\sigma)'$ .

Algorithmically it takes on the following form.

### Auxiliary particle filter

1. Set  $t = 1$ , draw  $\alpha_1^{(j)}$  from a prior  $f(\alpha_1)$ . Here I have taken

$$q_1^{(j)} \stackrel{L}{=} 0.3\chi_1^2, \quad \sigma_1^{2(j)} \stackrel{L}{=} 0.25\chi_1^2, \quad (48)$$

where  $\stackrel{L}{=}$  denotes being equal in law (i.e. distribution). Set

$$m_{2|1}^{(j)} = y_1, \quad p_{2|1}^{(j)} = \sigma_1^{2(j)} (1 + q_1^{(j)}), \quad w_{2|1}^{(j)} = 0 \quad (49)$$

and sample  $\alpha_2^{(j)} \sim \alpha_2 | \alpha_1^{(j)}, \theta$ .

2. Set  $t = t + 1$ . Compute for each  $j = 1, 2, \dots, M$  in parallel the conditional Kalman filter (23) and (24), recording

$$l_t^{(j)} = -\frac{1}{2} \log \left( p_{t|t-1}^{(j)} + \sigma_t^{2(j)} \right) - \frac{1}{2} \frac{\left( y_t - m_{t|t-1}^{(j)} \right)^2}{\left( p_{t|t-1}^{(j)} + \sigma_t^{2(j)} \right)}. \quad (50)$$

updating  $w_{t+1|t}^{(j)} = w_{t|t-1}^{(j)} + l_t^{(j)}$  and computing

$$\omega_t^{(j)} = \frac{p_{t|t-1}^{(j)}}{p_{t|t-1}^{(j)} + \sigma_t^{2(j)}}, \quad (51)$$

then

$$m_{t+1|t}^{(j)} = \left( 1 - \omega_t^{(j)} \right) m_{t|t-1}^{(j)} + \omega_t^{(j)} y_t, \quad p_{t+1|t}^{(j)} = \sigma_t^{2(j)} \left( q_t^{(j)} + \omega_t^{(j)} \right). \quad (52)$$

Then simulate forward the particle state

$$\alpha_{t+1}^{(j)} \sim \alpha_{t+1} | \alpha_t^{(j)}, \theta. \quad (53)$$

3. Record some summary results if desired. This will be discussed in a moment.
4. Resample every 3 (an ad hoc choice) increments in time, by resampling with replacement from the

$$\left\{ \alpha_t^{(1)}, m_{t+1|t}^{(1)}, p_{t+1|t}^{(1)} \right\}, \dots, \left\{ \alpha_t^{(M)}, m_{t+1|t}^{(M)}, p_{t+1|t}^{(M)} \right\} \quad (54)$$

with probability proportional to  $w_{t+1|t}^{*(1)}, \dots, w_{t+1|t}^{*(M)}$ , where  $w_{t|t-1}^{*(j)}$  is given in (55) below. Having done this set  $w_t^{(j)} = 0$  for  $j = 1, 2, \dots, M$ .

5. Goto 2.

The fact that the key step 2 is entirely parallel in  $j$  means it can be implemented using matrix computations. This is important as  $M$  may be quite large.

Finally, we note that for the Stock-Watson parameterisation discussed in Section 3.6 the only change would be that  $\alpha_t = (\sigma_{\varepsilon,t}^2, \sigma_{\eta,t}^2)'$ ,  $\sigma_{\eta,1}^{2(j)} = \sigma_1^{2(j)} q_1^{(j)}$ ,  $p_{t+1|t}^{(j)} = \sigma_{\eta,t}^{2(j)} + \sigma_{\varepsilon,t}^{2(j)} \omega_t^{(j)}$  and replacing  $\sigma_t^{2(j)}$  by  $\sigma_{\varepsilon,t}^{2(j)}$  everywhere.

## 4.2 Recording some output

As the particle filter steps through time the output needs to be recorded. Here I highlight some important quantities which can be saved, the extension to many others is straightforward.

3. Record some output.

(a) Compute

$$w_{t|t-1}^{*(j)} = \frac{\exp(w_{t|t-1}^{(j)})}{\sum_{i=1}^M \exp(w_{t|t-1}^{(i)})}. \quad (55)$$

(b) Then a variety of things can be estimated, such as

i. Particle estimates of the log-likelihood contribution

$$\log \hat{f}(y_t | \mathcal{F}_{t-1}^y, \theta) = \log \left[ \sum_{j=1}^M w_{t|t-1}^{*(j)} \exp \left\{ l_t^{(j)} \right\} \right], \quad (56)$$

ii. Particle estimates of the level  $\hat{E}(\mu_t | \mathcal{F}_{t-1}^y, \theta) = \sum_{j=1}^M w_{t|t-1}^{*(j)} m_{t|t-1}^{(j)}$ .

iii. Particle estimates of the state process  $\hat{E}(\alpha_t | \mathcal{F}_{t-1}^y, \theta) = \sum_{j=1}^M w_{t|t-1}^{*(j)} \alpha_t^{(j)}$ .

In practice it is often better to use quantiles as summaries of  $\alpha_t | \mathcal{F}_{t-1}^y, \theta$  and  $\mu_t | \mathcal{F}_{t-1}^y, \theta$ . To do this recall that the distribution of the posterior  $\mu_t | \mathcal{F}_{t-1}^y, \theta$  can be estimated as  $\hat{F}(\mu_t | \mathcal{F}_{t-1}^y, \theta) = \sum_{j=1}^M w_{t|t-1}^{*(j)} F_N(\mu_t; m_{t|t-1}^{(j)}, P_{t|t-1}^{(j)})$ , where  $F_N$  is the normal distribution function. The composition method can be used to simulate from this and then the corresponding empirical quantiles can be calculated. Alternatively this distribution can be first analytically marginalised (i.e. using the properties of the normal distribution) and then numerically inverted to compute the quantiles. To estimate the quantiles of  $\alpha_t | \mathcal{F}_{t-1}^y, \theta$  the weighted particles  $w_{t|t-1}^{*(j)}, \alpha_t^{(j)}$  can be resampled and the corresponding empirical quantiles computed.

### 4.3 Particle MCMC

The particle filter delivers an unbiased estimator of the likelihood. This property is important and so it is worthwhile reflecting upon it for a moment. Here I draw upon Andrieu, Doucet, and Holenstein (2010) and the exposition in Flury and Shephard (2011). See also Pitt, Silva, Giordani, and Kohn (2012).

Think of all the uniform random numbers behind a run of a particle filter as  $u$ , then the particle filter based estimated likelihood can be thought of as

$$\widehat{L}_u(\theta|y) = f(y|\theta, u), \quad (57)$$

an artificial conditional density, which has the property that

$$E_U \left\{ \widehat{L}_U(\theta|y) \right\} = \int f(y|\theta, u) f(u) du, \quad (58)$$

where  $f(u) \propto 1$  is the density of the uniforms.

This means that an artificial joint posterior can be constructed as

$$f(u, \theta|y) \propto \widehat{L}_u(\theta|y) f(u) f(\theta), \quad (59)$$

by thinking of  $u$  as a set of auxiliary variables. Hence inference can be carried out by simulating from  $u, \theta|y$  and discarding the draws from  $u$ . This approach is called particle MCMC and is due to Andrieu, Doucet, and Holenstein (2010), while outside the particle context it can be traced back to Beaumont (2003). It simply makes a proposal to move  $\theta$  by drawing a proposed move to  $\theta'$  but where the likelihood is estimated using a particle filter based upon some draws  $u'$ . If the proposal is accepted both  $\theta', u'$  are accepted. The resulting chain is a correct draw from  $\theta|y$  even though an estimated likelihood is being used. This contrasts markedly with the maximum simulated likelihood literature discussed by, for example, Gouriéroux and Monfort (1996).

## 5 Illustration using inflation data

### 5.1 The complete model

I follow an example given in Harvey (1981) on a time series of inflation, which he modelled by a Gaussian random walk plus noise model. Here the martingale unobserved component model

$$\begin{aligned} y_t &= \mu_t + \sigma_t \varepsilon_t, & \mu_{t+1} &= \mu_t + \sigma_t q_t^{1/2} \eta_t, \\ \log \sigma_{t+1}^2 &= \log \sigma_t^2 + \theta_\sigma \zeta_{\sigma,t}, & \log q_{t+1} &= \log q_t + \theta_q \zeta_{q,t}, \end{aligned} \quad (60)$$

will be used where  $\sigma_t$  is called ‘‘SV scaling’’, as it effects all aspect of the scale of movements in the series, and assume  $(\varepsilon_t, \eta_t, \zeta_{\sigma,t}, \zeta_{q,t})' \stackrel{iid}{\sim} N(0, I_4)$ . The virtue of the martingale unobserved component model here is that it produces rather simple forecasts of inflation a few steps out, say

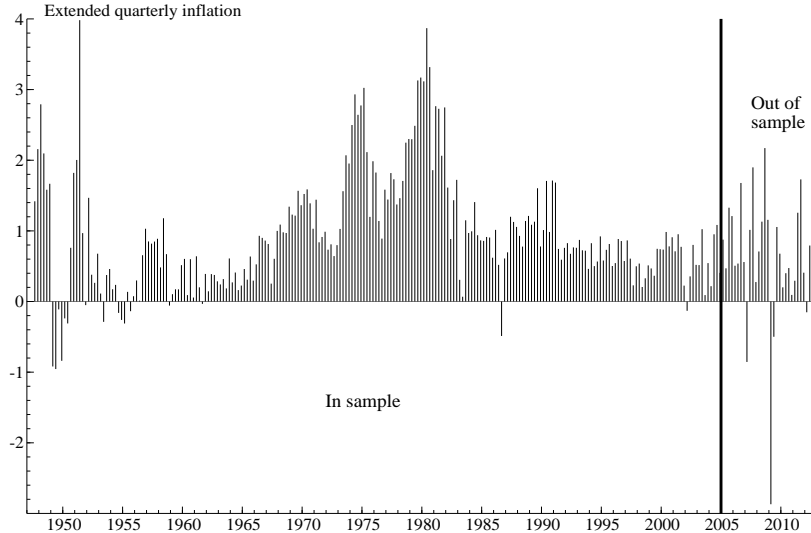


Figure 1: Computed quarterly US inflation series, measured through CPI-U. This is constructed as 100 times the first difference of the log of the average of the price index during the quarter. The vertical line indicates when the out of sample new data starts, which is the first quarter of 2005.

1 year, just extrapolating the current estimate of the level  $\mu_t$ . At that time horizon it is likely to be somewhat robust to any moderate unmodelled serially correlation in  $\varepsilon_t$ .

Here the addition of the time varying  $q$  allows the EWMA type forecasts implied by the local level model to adapt the rate of discounting to the recent data.

This model is indexed by three “parameters” with priors which are integrated directly over:  $\mu_2|y_1, q_1, \sigma_1^2 \sim N(m_1, p_{2|1})$ , where  $m_1 = y_1$  and  $p_{2|1} = \sigma_1^2(1 + q_1)$  and the independent initial conditions

$$q_1 \stackrel{L}{=} 0.3\chi_1^2, \quad \sigma_1^2 \stackrel{L}{=} 0.25\chi_1^{-2}. \quad (61)$$

The focus here will be on the last parameters  $\theta_\sigma$  and  $\theta_q$  and they will be assumed to be a priori independent and that

$$\theta_\sigma \stackrel{L}{=} 0.15 \left( \frac{1}{3}\chi_3^2 \right), \quad \theta_q \stackrel{L}{=} 0.3 \left( \frac{1}{3}\chi_3^2 \right). \quad (62)$$

These are ad hoc choices, but I selected them to ensure the mode is away from zero, centred on roughly plausible values and the tails of these prior are quite thin to crush implausible values.

Stock and Watson (2007) also focused on this problem but they used a different parameterisation (the results for their model will be given at the end of Section 5.3). However, I will use the same data as them. The series is the main US inflation series CPI-U and starts in January 1947. The Stock and Watson version of the data finishes December 2004. I follow them in computing quarterly price levels as the average monthly price level in each quarter. In our graphs of the raw quarterly



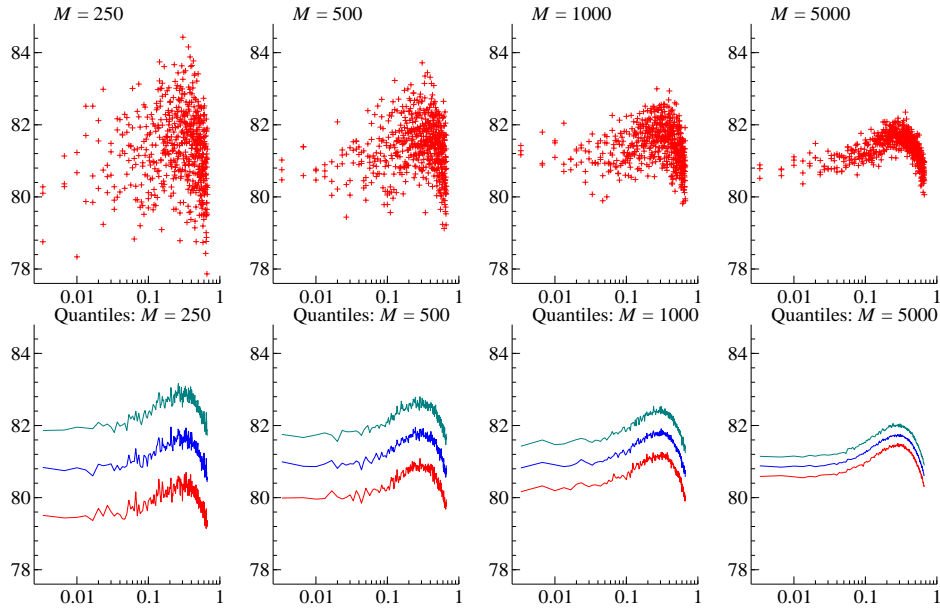


Figure 2: Estimated log-likelihood function for the inflation series for  $\theta_q$  fixing throughout  $\theta_\sigma = 0.25$ . Top row of graphs shows the particle filter estimator of the log-likelihood for various values of  $M$  using 3 replications evaluating  $\theta_q$  at 200 different values. Bottom graphs are the pointwise 0.1, 0.5 and 0.9 quantiles of the estimates. 400 replications are used for each parameter point in estimating the quantiles. The likelihood is quite flat for this parameter.

data, estimated volatilities and diagnostics will be extended to August 2012, keeping the estimated parameters  $\theta_\sigma$  and  $\theta_q$  fixed at the December 2004 estimates.

Throughout the series analysed will be 100 times the first difference of the log of the series, which is roughly the percentage quarterly inflation series. This raw series is given in Figure 1, where the updating of the series to reflect new data is indicated by the vertical line.

## 5.2 Estimated likelihood

The first step to understanding the empirical content of the model will be to graph the particle filter based estimates of the log-likelihood as a function of the two parameters in  $\theta$ . For the plot as a function of  $\theta_q$  I fixed  $\theta_\sigma = 0.25$  and when the function in  $\theta_\sigma$  is drawn I took  $\theta_q = 0.1$ . The particle filter is run using  $M$  as 250, 500, 1000 and 5000. Figures 2 and 3 shows the results and indicate, as expected, that the filter becomes more precise as  $M$  increases. The top row of graphs shows the particle filter estimates of the log-likelihood, throughout using a log10 scale for  $\theta_q$  and  $\theta_\sigma$ . It indicates quite a flat likelihood for  $\theta_q$  but with some support for a value of away from  $\theta_q = 0$ , which is the random walk plus SV scale model special case. It shows quite a heavily peaked likelihood for  $\theta_\sigma$  away from  $\theta_\sigma = 0$ . Taken together this suggests a full martingale unobserved component model is needed, but volatility clustering in the scale is the main feature to be added to the Gaussian

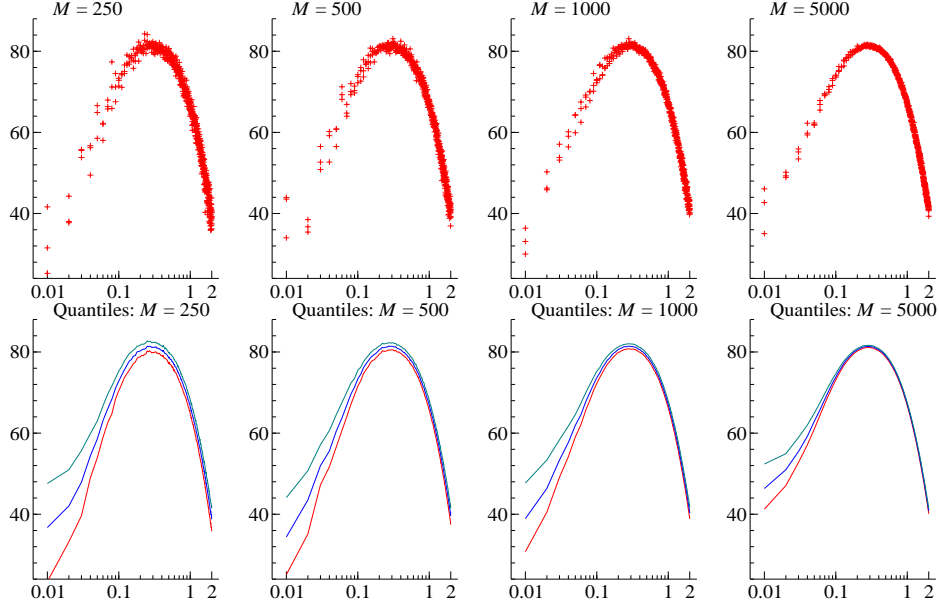


Figure 3: Estimated log-likelihood function for the inflation series for  $\theta_\sigma$  fixing throughout  $\theta_q = 0.1$ . Top row of graphs shows the particle filter estimator of the log-likelihood for various values of  $M$  using 3 replications evaluating  $\theta_\sigma$  at 200 different values. Bottom graphs are the pointwise 0.1, 0.5 and 0.9 quantiles of the estimates. 400 replications are used for each parameter point. The parameter is shown on a log10 scale. This shows a highly informative likelihood.

model.

### 5.3 Particle MCMC output

The pictures in Figures 2 and 3 shows that the likelihood is quite tightly estimated and so the mixing properties of the particle MCMC should not change very much as  $M$  varies. The output from an MCMC chain is placed in Figure 4, which is based on the results from running a MCMC chain for 20,000 iterations. This uses a random walk proposals for  $\log \theta_q$  and  $\log \theta_\sigma$  with independent proposals with standard deviations of 0.25 — which are large moves. Writing the proposals as  $\theta_q, \theta_\sigma, \hat{f}$  (recalling  $\hat{f}(y|\theta_q, \theta_\sigma)$  is the particle filter estimate of  $f(y|\theta_q, \theta_\sigma)$ ) and the existing values in the MCMC chain as  $\theta'_q, \theta'_\sigma, \hat{f}'$ , then the probability of accepting these values is

$$\begin{aligned} & \min \left\{ 1, \frac{\hat{f}(y|\theta_q, \theta_\sigma) f(\theta_q, \theta_\sigma) \theta_q e^{-\frac{1}{2 \times 0.25^2} (\log \theta_q - \log \theta'_q)^2} \theta_\sigma e^{-\frac{1}{2 \times 0.25^2} (\log \theta_\sigma - \log \theta'_\sigma)^2}}{\hat{f}'(y|\theta'_q, \theta'_\sigma) f(\theta'_q, \theta'_\sigma) \theta'_q e^{-\frac{1}{2 \times 0.25^2} (\log \theta_q - \log \theta'_q)^2} \theta'_\sigma e^{-\frac{1}{2 \times 0.25^2} (\log \theta_\sigma - \log \theta'_\sigma)^2}} \right\} \\ & = \min \left\{ 1, \frac{\hat{f}(y|\theta_q, \theta_\sigma) f(\theta_q, \theta_\sigma) \theta_q \theta_\sigma}{\hat{f}'(y|\theta'_q, \theta'_\sigma) f(\theta'_q, \theta'_\sigma) \theta'_q \theta'_\sigma} \right\}. \end{aligned}$$

The key feature here is that the value of the estimated likelihood is accepted or rejected, as well as the proposed parameter values. Further  $f(\theta_q, \theta_\sigma)$  is the joint prior.

In all 8 such chains are run independently in parallel, using a multicore processor. The output is then thought of as a cross-section of long independent time series with the same marginal distribution. The Figure shows the path from a single chain, while the autocorrelation function is the average autocorrelation function from the 8 independent chains.

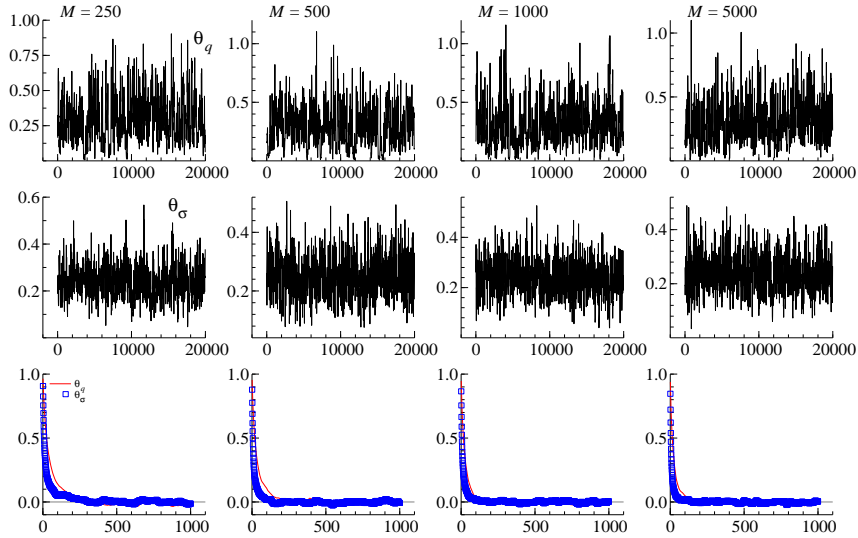


Figure 4: Particle MCMC inference for the  $\theta_q$  and  $\theta_\sigma$  parameters for the inflation series. Resulting MCMC chain for a variety of values of  $M$  being 250, 500, 1000 and 5000. Suggests the mixing of the chain is relatively fast.

The posteriors are typically summarised using quantiles. This is carried out by computing the quantile for each chain separately and then cross-sectionally averaging the resulting quantiles. The uncertainty of this estimate can be measured by using the standard error of this arithmetic mean, but when I did this the errors are so small there is little utility in reporting them here.

Table 1 shows quantile summaries of  $\theta|y$  for different values of  $M$ , showing that the results are entirely comparable. There is an improvement in precision as  $M$  increases, but the computational cost of running the algorithm is proportional to  $M$ . The results suggest that using a small value of  $M$  maybe the most computationally effective for this problem.

The posterior  $\theta_q|y$  shows quite a high degree of spread and non-symmetry, with 80% of the mass roughly between 0.09 and 0.55. The posterior  $\theta_\sigma|y$  is much tighter and symmetric, with 80% of the mass roughly between 0.14 and 0.33. The posterior means are  $\theta_q = 0.37$  and  $\theta_\sigma = 0.23$  respectively, while the standard deviation measures reflect the results from the quantiles. The posterior correlation between the parameters is around  $-0.4$ .

Table 1 also shows summaries for the fitted Stock-Watson parameterisation (note the corresponding likelihood profiles are given in Figure 5 and 6). In their empirical work they impose a

$M$	$Q_{.1}(\theta_q y)$	$Q_{.5}(\theta_q y)$	$Q_{.9}(\theta_q y)$	$Q_{.1}(\theta_\sigma y)$	$Q_{.5}(\theta_\sigma y)$	$Q_{.9}(\theta_\sigma y)$	
250	0.097	0.279	0.558	0.141	0.230	0.333	
500	0.093	0.281	0.551	0.146	0.232	0.333	
1,000	0.103	0.285	0.560	0.143	0.230	0.334	
5,000	0.096	0.281	0.548	0.145	0.232	0.333	
$M$	$E(\theta_q y)$	$sd(\theta_q y)$	$IAC_{\theta_q}$	$E(\theta_\sigma y)$	$sd(\theta_\sigma y)$	$IAC_{\theta_\sigma}$	$Cor(\theta_q, \theta_\sigma y)$
250	0.307	0.171	90	0.234	0.074	54	-0.382
500	0.306	0.173	81	0.237	0.073	26	-0.339
1,000	0.313	0.187	54	0.235	0.075	25	-0.416
5,000	0.306	0.174	51	0.236	0.074	21	-0.384
Stock-Watson parameterisation							
$M$	$Q_{.1}(\theta_{\sigma_\eta} y)$	$Q_{.5}(\theta_{\sigma_\eta} y)$	$Q_{.9}(\theta_{\sigma_\eta} y)$	$Q_{.1}(\theta_{\sigma_\varepsilon} y)$	$Q_{.5}(\theta_{\sigma_\varepsilon} y)$	$Q_{.9}(\theta_{\sigma_\varepsilon} y)$	
250	0.294	0.478	0.711	0.112	0.202	0.335	
500	0.299	0.480	0.713	0.110	0.200	0.332	
1,000	0.301	0.482	0.718	0.109	0.199	0.330	
5,000	0.299	0.482	0.713	0.111	0.199	0.332	
$M$	$E(\theta_{\sigma_\eta} y)$	$sd(\theta_{\sigma_\eta} y)$	$IAC_{\theta_{\sigma_\eta}}$	$E(\theta_{\sigma_\varepsilon} y)$	$sd(\theta_{\sigma_\varepsilon} y)$	$IAC_{\theta_{\sigma_\varepsilon}}$	$Cor(\theta_{\sigma_\eta}, \theta_{\sigma_\varepsilon} y)$
250	0.494	0.165	35	0.215	0.089	52	-0.344
500	0.496	0.157	23	0.213	0.092	37	-0.321
1,000	0.499	0.162	19	0.211	0.087	28	-0.316
5,000	0.497	0.164	17	0.213	0.090	30	-0.372

Table 1: Results from a random walk based particle MCMC algorithm analysis of the posterior  $\theta|y$  for the inflation series. Generically  $Q_q(X)$  denotes the  $q$ -th quantile of the random variable  $X$ , computed by cross-sectionally averaging the individual quantiles from each chain.  $IAC_X$  denotes the integrated autocorrelation  $1 + 2 \sum_{j=1}^{1500} \hat{\rho}_j$ , where  $\hat{\rho}_j$  is the cross-sectional average of the autocorrelation of the MCMC chain for  $X$  using this level of  $M$ .

priori that  $\theta_{\sigma_\varepsilon} = \theta_{\sigma_\eta} = \sqrt{0.2} \simeq 0.447$ , which is not a million miles away from the values supported by the data, although the data would prefer a lower value for  $\theta_{\sigma_\varepsilon}$ . What is most striking about these estimated values is that under (39) the implied  $\theta_q \simeq \sqrt{0.47^2 + 0.18^2} \simeq 0.50$ , which is high compared to the estimated directly parameterised model.

### 5.3.1 Results for the extended model

Figure 7 reconciles the results from the two different models, by allowing a correlation between the two SV innovations in (42), and showing the resulting estimated likelihood profile as a function of the correlation. This shows the log-likelihood is substantially more sensitive to  $\rho$  for the SW parameterisation, while for both parameterisations the graphs support a correlation which is very close to one, say around 0.8 (when we subsequently look at the Bayesian posteriors the median and expected values are substantially lower than this as I have used a flat prior over the interval

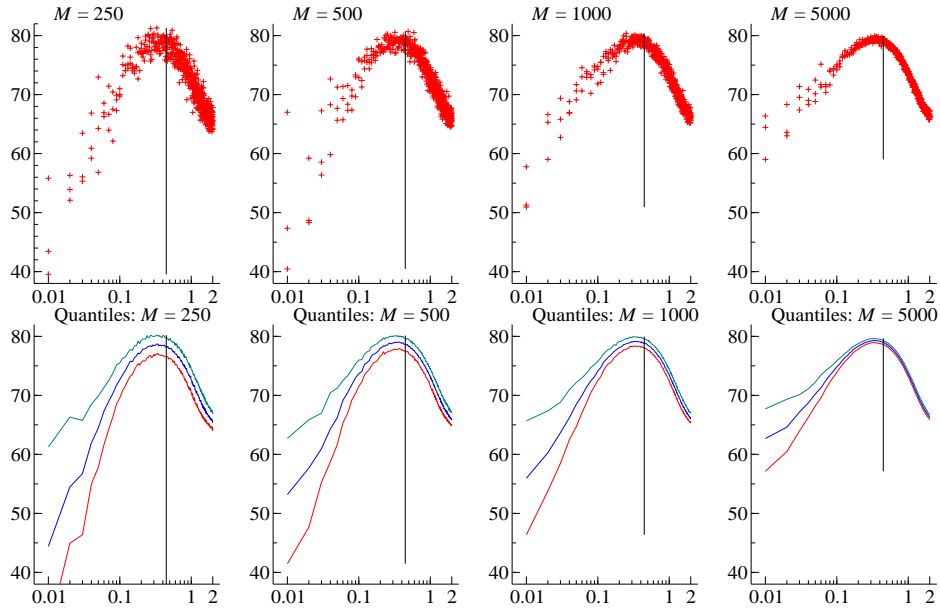


Figure 5: Estimated log-likelihood function for the inflation series using the SW model for  $\theta_{\sigma_\eta}$  fixing  $\theta_{\sigma_\varepsilon} = 0.25$ . Top row: particle filter estimator of the log-likelihood for various values of  $M$  using 3 replications evaluating  $\theta_{\sigma_\eta}$  at 200 different values. Bottom: pointwise 0.1, 0.5 and 0.9 quantiles of the estimates. 400 replications are used for each parameter point in estimating the quantiles.

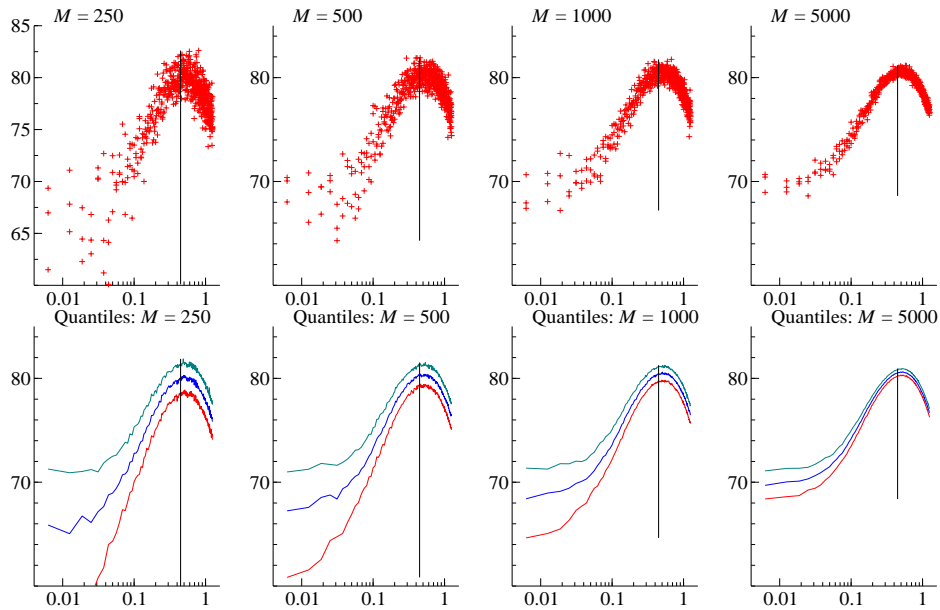


Figure 6: Estimated log-likelihood function for the inflation series using the SW model for  $\theta_{\sigma_\varepsilon}$  fixing throughout  $\theta_{\sigma_\eta} = 0.25$ . Top row: particle filter estimator of the log-likelihood for various values of  $M$  using 3 replications evaluating  $\theta_{\sigma_\varepsilon}$  at 200 different values. Bottom: pointwise 0.1, 0.5 and 0.9 quantiles of the estimates. 400 replications are used for each parameter point in estimating the quantiles.

-1,1 and the likelihood is not tightly peaked, but the posterior summaries are higher for the extended SW as the corresponding score is larger). For the extended model the change in log-likelihood by setting  $\rho = 0$  looks to be around 0.5, while for the extended SW model it is around 1.5. When we focus on the SW parameterisation and take this  $\rho = 0.8$  value then the implied  $\theta_q \simeq \sqrt{(0.48 - 0.8 \times 0.20)^2 + 0.36 \times 0.20^2} \simeq 0.34$ , which is not a great deal from the estimate value for the original model.

$M$	$Q_p(\theta_q y)$				$Q_p(\theta_\sigma y)$				$Q_p(\rho y)$			
	$p = .1$	$p = .5$	$p = .9$	IAC	$p = .1$	$p = .5$	$p = .9$	IAC	$p = .1$	$p = .5$	$p = .9$	IAC
250	0.080	0.260	0.532	106	0.141	0.228	0.332	47	-0.368	0.278	0.775	41
500	0.087	0.266	0.528	74	0.142	0.228	0.334	31	-0.364	0.273	0.776	29
1,000	0.082	0.261	0.526	63	0.142	0.229	0.334	26	-0.359	0.276	0.780	23
5,000	0.081	0.254	0.528	45	0.144	0.230	0.335	18	-0.366	0.276	0.785	22

---

Extended SW parameterisation												
$M$	$Q_p(\theta_{\sigma_\eta} y)$				$Q_p(\theta_{\sigma_\varepsilon} y)$				$Q_p(\rho y)$			
	$p = .1$	$p = .5$	$p = .9$	IAC	$p = .1$	$p = .5$	$p = .9$	IAC	$p = .1$	$p = .5$	$p = .9$	IAC
250	0.272	0.443	0.674	43	0.121	0.206	0.323	39	-0.117	0.487	0.861	48
500	0.270	0.437	0.665	26	0.119	0.206	0.327	31	-0.113	0.487	0.860	29
1,000	0.271	0.435	0.665	17	0.121	0.207	0.324	26	-0.104	0.499	0.862	24
5,000	0.270	0.436	0.668	16	0.119	0.205	0.321	22	-0.125	0.486	0.855	24

Table 2: Results from a random walk based particle MCMC algorithm analysis of the posterior  $\theta|y$  for the inflation series. Generically  $Q_q(X)$  denotes the  $q$ -th quantile of the random variable  $X$ , computed by cross-sectionally averaging the individual quantiles from each chain.  $IAC_X$  denotes the integrated autocorrelation  $1 + 2 \sum_{j=1}^{1500} \hat{\rho}_j$ , where  $\hat{\rho}_j$  is the cross-sectional average of the autocorrelation of the MCMC chain for  $X$  using this level of  $M$ .

#### 5.4 Headline features of fitted martingale unobserved component model

Figure 8 shows the results from the fit of the martingale unobserved component model for the quarterly inflation series. This was computed using the parameter values  $\theta_q = 0.27$  and  $\theta_\sigma = 0.23$ .

The first graph is the time series of the raw quarterly inflation data (denoted by a cross) together with the one-step ahead forecast from the model (shown by a time series line).

The second graph shows the time series of the 0.1, 0.5 and 0.9 quantiles of

$$\sigma_t \sqrt{q_t} | \mathcal{F}_t, \theta \tag{63}$$

which represents the standard deviation of the quarterly change in the level of underlying inflation. The results show that the volatility changes by an order of magnitude through time, sometimes

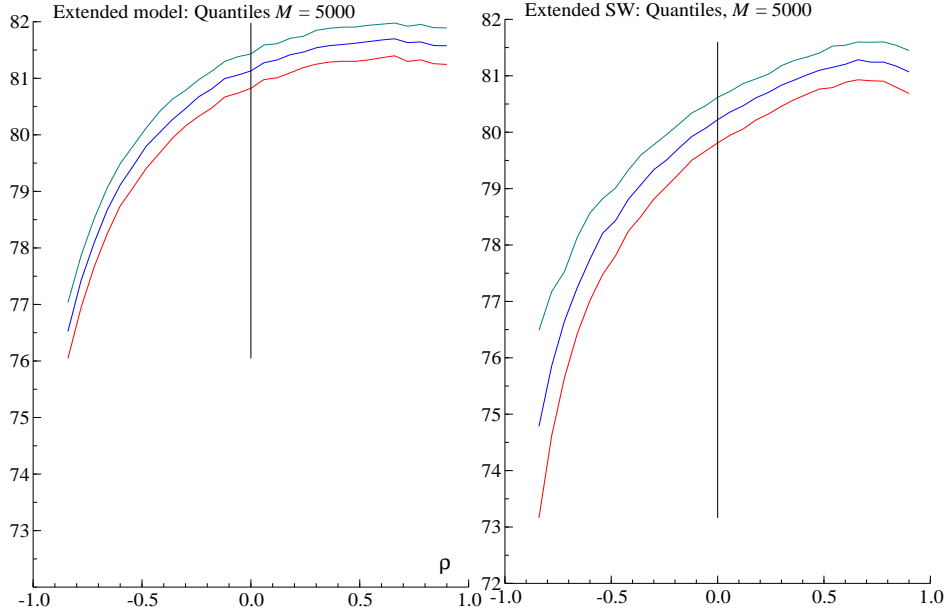


Figure 7: Estimated log-likelihood function for the inflation series for  $\rho \in (-1, 1)$  fixing the other parameters for the extended model (fixing  $\theta_q = 0.18$ ,  $\theta_\sigma = 0.22$ ) and the extended SW parameterisation (fixing  $\theta_{\sigma_\eta} = 0.46$ ,  $\theta_{\sigma_\varepsilon} = 0.16$ ). Shows pointwise 0.1, 0.5 and 0.9 quantiles of the estimates based on 400 replications.

being around 0.05 while reaching a peak which is close to 0.9. The quantiles show the precision with which this volatility is estimated changes through time.

The third graph shows the time series of the 0.1, 0.5 and 0.9 quantiles of

$$\sigma_t | \mathcal{F}_t, \theta \quad (64)$$

which represents the time series of the volatility of the temporary component of inflation. This has a minimum of around 0.2 and maximises at around 1 in the early to mid 1950s and again around 2010. The volatility of the temporary component did not rise during the 1970s and 1980s, which significantly differentiates it from the volatility of the permanent component. Stock and Watson (2007) report that  $\sigma_t$  is relatively constant between the 1960s and 1990s, which is true but it is importantly not the case outside this period and enforcing homogeneity would be highly dangerous to forecasting and understanding. A further note is that during periods of high transitory volatility it is hard to precisely estimate the permanent component and so the permanent component's volatility. Thus high periods of transitory volatility, such as the recent times, are associated with wide confidence intervals for the permanent volatility.

The fourth plot shows the cumulative difference in the log-likelihood between the martingale unobserved component model and the Gaussian random walk plus noise model. This shows a consistent improvement from the martingale model, except during the late 1970s to around 1990,

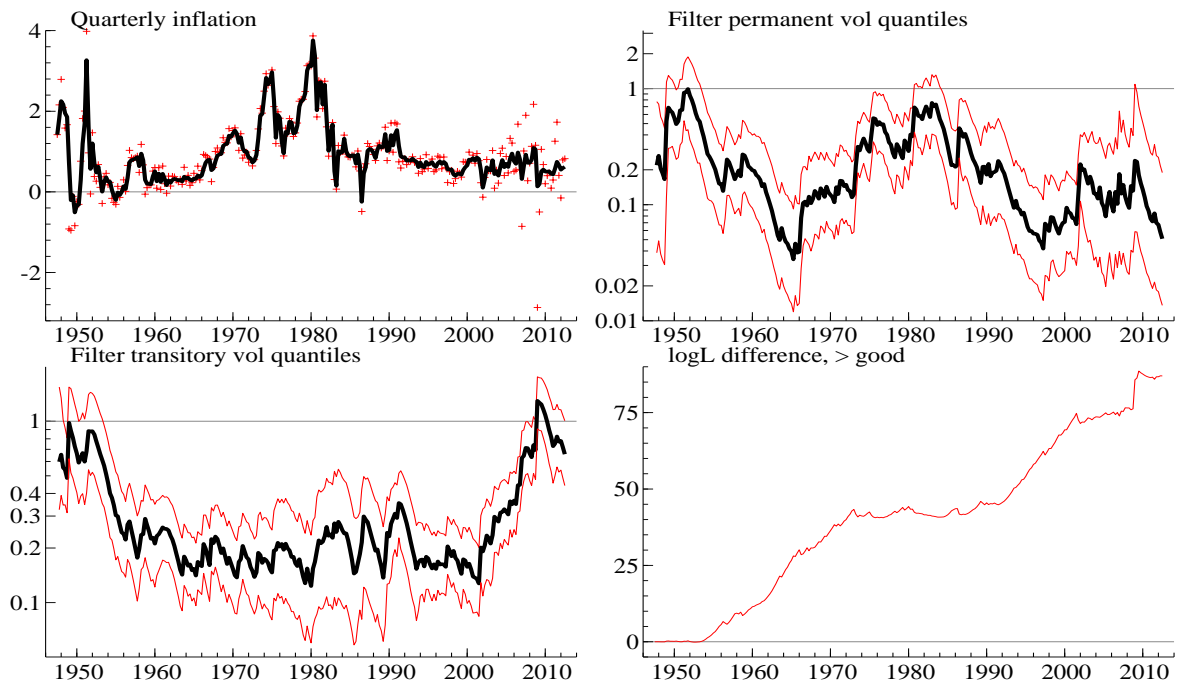


Figure 8: Main fit from the inflation series. Top left shows the raw quarterly inflation with the 1-step ahead forecasts. Top right shows the 0.1, 0.5 and 0.9 quantiles of the filtered volatility of the permanent component, and the bottom left shows the volatility of the transitory component. The filtering looks at the distributions  $\sigma_t|\mathcal{F}_t$  and  $\sigma_t q_t^{1/2}|\mathcal{F}_t$ . They are plotted on the same log10 scale. Bottom right shows the cumulative difference in the log-likelihood compared to the linear model.



where the models fitted roughly similarly.

## 5.5 Analysis

Figure 9 looks underneath some of the time series properties of the fitted model. The spot moving average process  $\psi_t$  is shown on the top left, plotted as  $-\psi_t$ . Recall that as  $\psi_t \rightarrow 0$  the model becomes a martingale, while when it is close to one it is using a lot of recent data to average away the noise. The movements during the sample are considerable and important, largely replicating the observation of Stock and Watson (2007). At the start of the 1980s the process went close to zero, which is suggesting a martingale forecast. At the end of the sample the process is close to one.

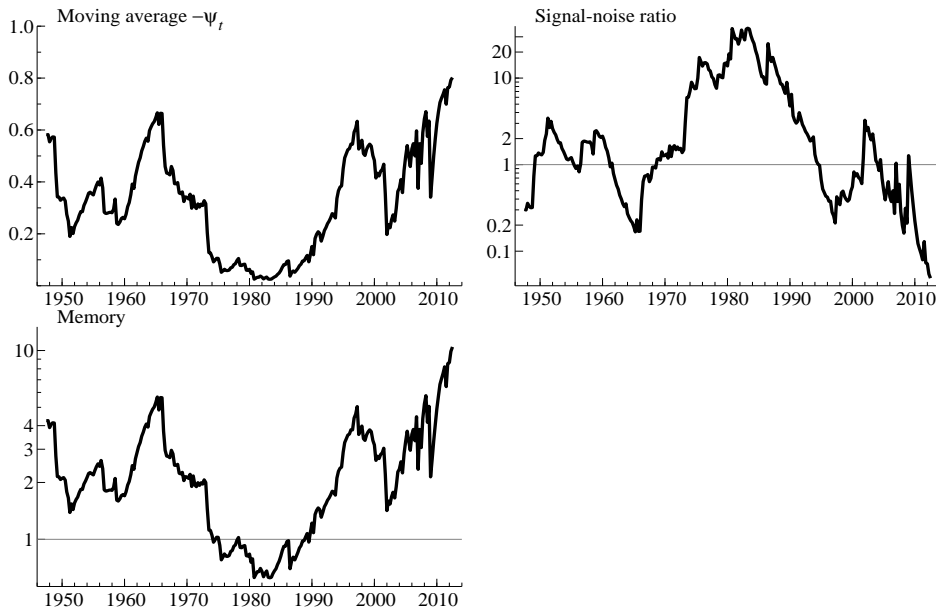


Figure 9: Analysis of a martingale unobserved component model applied to quarterly inflation. Top left: filtered estimate of the moving average coefficient. Top right: filtered estimate of the signal/noise ratio. Bottom left: filtered estimate of the memory of the time series.

The top right shows the estimate of the signal-noise ratio

$$E(q_t | \mathcal{F}_t, \theta) \tag{65}$$

which shows very large moves, with changes of roughly two orders of magnitude. The current value of the signal-noise ratio is very low by historical standards.

The bottom left picture is the time series of the memory measure  $s_t$  through time. Recall this works out how long it takes before the weight on a past data point is under 10% and is computed via (33). This picture is drawn on the log10 y-scale but has little extra information in it compared to the plot of the moving average process.

## 5.6 Diagnostic statistics

Figure 10 shows some plots to diagnose the empirical effectiveness of the models. The first column of figures looks at the innovations from the fitted model

$$v_t = \text{E} \left( \frac{y_t - m_{t|t-1}}{\sigma_t \sqrt{p_{t|t-1}^* + 1}} \middle| \mathcal{F}_{t-1} \right), \quad (66)$$

where the posterior mean is averaged over  $q_{1:t}$  and  $\sigma_{1:t}$ . This should be roughly i.i.d. standard normal if the model fits well. In the top plot  $v_t$  is drawn against  $t$  and there is little apparent structure visible to the eye. The bottom plot shows the correlogram for  $v_t$ , drawn using an index plot, together with the correlogram for  $|v_t|$  which is shown by a series of dots. The latter is designed to pick up any missing volatility clustering in the innovations, the former looks for linear dependence in the innovations. The results are somewhat encouraging, there seems very little volatility clustering. There are some quite large autocorrelations appearing for  $v_t$  in lags 2 and 3, while the results for  $|v_t|$  look very strong. This indicates the model could be improved for very short-term forecasting by allowing some very short memory into  $\varepsilon_t$ .

The second column looks at the plot of the probability integral transforms

$$U_t = \text{E} \left[ \Phi \left( \frac{y_t - m_{t|t-1}}{\sigma_t \sqrt{p_{t|t-1}^* + 1}} \right) \middle| \mathcal{F}_{t-1} \right] \quad (67)$$

against time, where  $\Phi$  is the standard normal distribution function. The  $U_t$  should be an i.i.d.  $U(0, 1)$  sequence if the model is correctly specified and so the plot should appear like the scatter of a two dimensional homogenous Poisson process on  $[0, 1] \times [1, n]$ . These types of transforms have been used by many researchers, e.g. Rosenblatt (1952), Shephard (1994) and Diebold, Gunther, and Tay (1998). The correlogram of the  $U_t$  and the so-called reflected uniforms  $2|U_t - 1/2|$  (see Shephard (1994) who introduced them to check for volatility clustering, realising they should also be i.i.d.  $U(0, 1)$  if the model is well fitting) are given in the plot below. These correlograms should be less effected by outliers than the corresponding ones for  $v_t$ , but in this case there is little difference.

The same diagnostics are also reported for the Gaussian random walk plus noise model in the third and fourth columns. Here there are some obvious failings, most dramatically due to volatility clustering. The correlogram in the third column also shows a very high correlation at lag 2 for the  $v_t$  (much higher than for the martingale plus noise model, and this feature also appears in the corresponding results for  $U_t$ ). Overall the fit of the linear model is quite poor.

All of these results are summarised in Table 3, which shows the fitted martingale unobserved component model's parameters, the corresponding log-likelihood and Box-Pierce summarises from the above correlograms. These were computed using 20 lags. Also given is the result from the

linear model (which is where  $\theta_q = \theta_\sigma = 0$ ), which has a poorer likelihood and worse dependence measures. Most of the gain is made by dealing with the heteroskedasticity in the data. The Table also reports results from the random walk plus SV noise model ( $\theta_q = 0$ ) and the martingale plus strict white noise model ( $\theta_\sigma = 0$ ). These models have parameters which need to be estimated and this is carried out using particle MCMC in the usual way. Note that these two constrained models

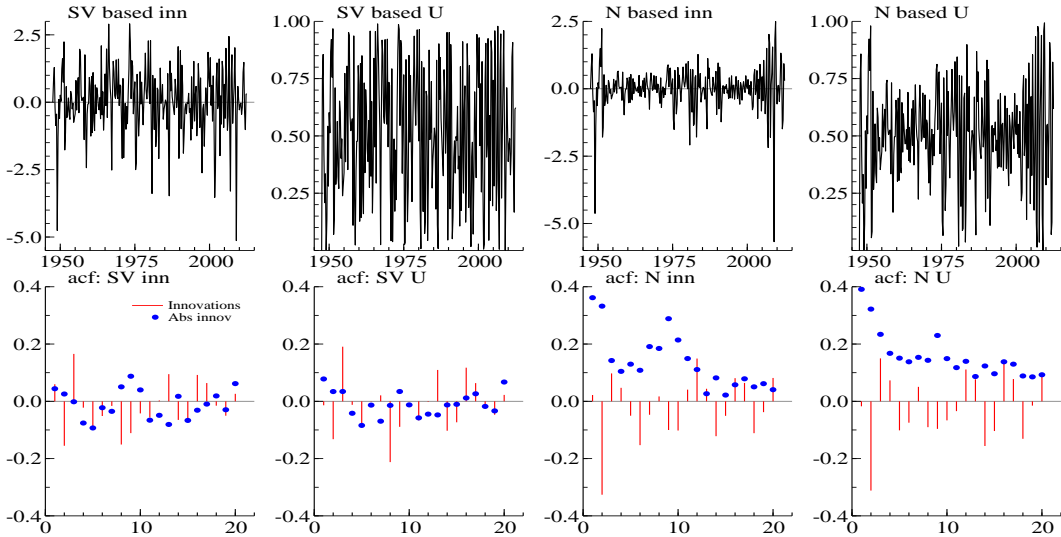


Figure 10: Diagnostics from the model for the extended inflation series. Top are raw innovations and probability integral transforms (U), for two models: martingale plus noise and random walk plus noise. Bottom denotes correlograms either of the raw series or the absolute value or reflected version  $2|U - 1/2|$ . Estimation is based on data up until 2004, log-likelihood is computed at these estimated parameter points and includes all data up to 2012.

deliver parameter estimates which are considerably higher than for the martingale unobserved component model, as they try to use their flexibility to deal with the heteroskedasticity in the data. That the parameters jump upwards is not surprising given the posterior for the general model is negatively correlated and these model simplifications wrongly impose one of the parameters as zero. Overall the martingale unobserved component model is slightly better than the random walk plus SV noise model. This is in turn better than the martingale plus strict white noise model, but the degree of difference is surprisingly small. This latter model boosts up  $\theta_q$  to such a large degree that it can deal with some of the heteroskedasticity in the data.

## 5.7 Multistep out of sample forecasting

I finish by looking at how the models perform out of sample, forecasting a year ahead, that is four steps ahead. The results are given in Figure 11 for the martingale unobserved component model and the Gaussian random walk plus noise model. This out of sample period includes the start of

	$\theta_q$	$\theta_\sigma$	$\rho$	$\theta_{\sigma_\eta}$	$\theta_{\sigma_\varepsilon}$	$\rho$	$\log L$	$\Delta \log L$	$BP_v$	$BP_{ v }$	$BP_U$	$BP_{2 U-1/2 }$
Insample												
$E(\theta_q, \theta_\sigma   y)$	0.31	0.23					81.6		44.7	14.9	43.5	10.3
$E(\theta_\sigma   y, \theta_q = 0)$	0.00	0.27					80.7		45.9	14.3	42.2	10.4
$E(\theta_q   y, \theta_\sigma = 0)$	0.61	0.00					76.1		40.9	11.6	43.1	9.4
$E(\theta_{\sigma_\varepsilon}, \theta_{\sigma_\eta}   y)$				0.49	0.21		80.6		43.6	13.9	43.9	10.0
$E(\theta_{\sigma_\varepsilon}   y, \theta_{\sigma_\eta} = 0)$				0.00	0.39		72.4		40.8	13.9	42.1	12.4
$E(\theta_{\sigma_\eta}   y, \theta_{\sigma_\varepsilon} = 0)$				0.61	0.00		75.9		42.1	11.7	43.9	10.0
$E(\theta_q, \theta_\sigma, \rho   y)$	0.26	0.23	0.28				81.8		45.1	14.9	43.2	9.6
$E(\theta_{\sigma_\varepsilon}, \theta_{\sigma_\eta}, \rho   y)$				0.44	0.21	0.49	81.4		44.2	14.4	43.1	9.8
$\theta_q = 0, \theta_\sigma = 0$	0.00	0.00					14.5		55.1	57.7	52.2	62.4
Extended dataset — out of sample												
$E(\theta_q, \theta_\sigma   y)$	0.31	0.23					66.0	-15.6	36.0	14.0	43.7	9.7
$E(\theta_\sigma   y, \theta_q = 0)$	0.00	0.27					62.2	-18.5	37.9	14.8	41.6	10.9
$E(\theta_q   y, \theta_\sigma = 0)$	0.61	0.00					54.4	-21.7	39.9	24.8	41.6	14.4
$E(\theta_{\sigma_\varepsilon}, \theta_{\sigma_\eta}   y)$				0.49	0.21		65.8	-14.8	35.6	12.9	44.3	9.3
$E(\theta_{\sigma_\varepsilon}   y, \theta_{\sigma_\eta} = 0)$				0.00	0.39		57.4	-15.0	34.9	13.3	43.3	12.2
$E(\theta_{\sigma_\eta}   y, \theta_{\sigma_\varepsilon} = 0)$				0.61	0.00		54.1	-21.8	40.1	25.2	41.7	14.6
$E(\theta_q, \theta_\sigma, \rho   y)$	0.26	0.23	0.28				65.6	-16.2	35.4	13.8	43.1	9.1
$E(\theta_{\sigma_\varepsilon}, \theta_{\sigma_\eta}, \rho   y)$				0.44	0.21	0.49	65.4	-16.0	35.7	12.9	43.5	8.7
$\theta_q = 0, \theta_\sigma = 0$	0.00	0.00					-14.8	-29.3	59.2	115	64.5	144

Table 3: Analysis based upon  $M=25,000$ .  $\log L$  is the estimated log-likelihood evaluated at that parameter point, while  $\Delta \log L$  is the change in the log-likelihood caused by the extended sample size — so is the out of sample log-likelihood (parameter values are unchanged with the extended data). Box-Pierce (BP) with 20 degrees of freedom. Here  $v$  denotes the innovations from the model and  $U$  denotes the probability integral transforms from the model. First model is the martingale unobserved component model. Second is the SV scaled random walk plus noise. Third is the martingale plus strict white noise. Fourth is the Stock-Watson model. Fifth is the martingale plus the white noise model. Sixth is the random walk plus SV noise. Seventh is the Gaussian random walk plus noise model.

the great recession and so considerably stresses these forecasting devices. Table 3 also shows the out of sample likelihood for the fitted models.

Throughout the parameters  $\theta_\sigma$  and  $\theta_q$  are fixed at their December 2004 values. Then running through the data, the estimated level  $\mu_t$  is extrapolated four periods into the future as well as computing the forecast's standard error under the various models.

Top left shows the data together with the two sets of forecasts. The results are dominated by the large negative inflation shock at the start of the recession. Of course the forecasts take four periods to react to this innovation. The Gaussian random walk plus noise model overreacts and this causes massive forecast errors later in the period. The martingale unobserved component model reacts more appropriately. The difference in the cumulative absolute forecast errors compared to the same quantity for other models is given in the bottom left. Numbers below zero suggest some support for the martingale unobserved component model. This suggests the random walk plus noise with SV scaling is the most competitive alternative to the martingale unobserved component model, followed by the random walk plus noise model. The martingale plus strict white noise

model is poor in comparison.

The associated forecast standard errors are drawn in the bottom right, with the results for the Gaussian random walk plus noise obviously being time invariant. The martingale unobserved component model ramps up the standard error before the start of the crisis and ever more so at the onset. The standard errors have tailed off considerably in recent times, down to roughly their long run average values.

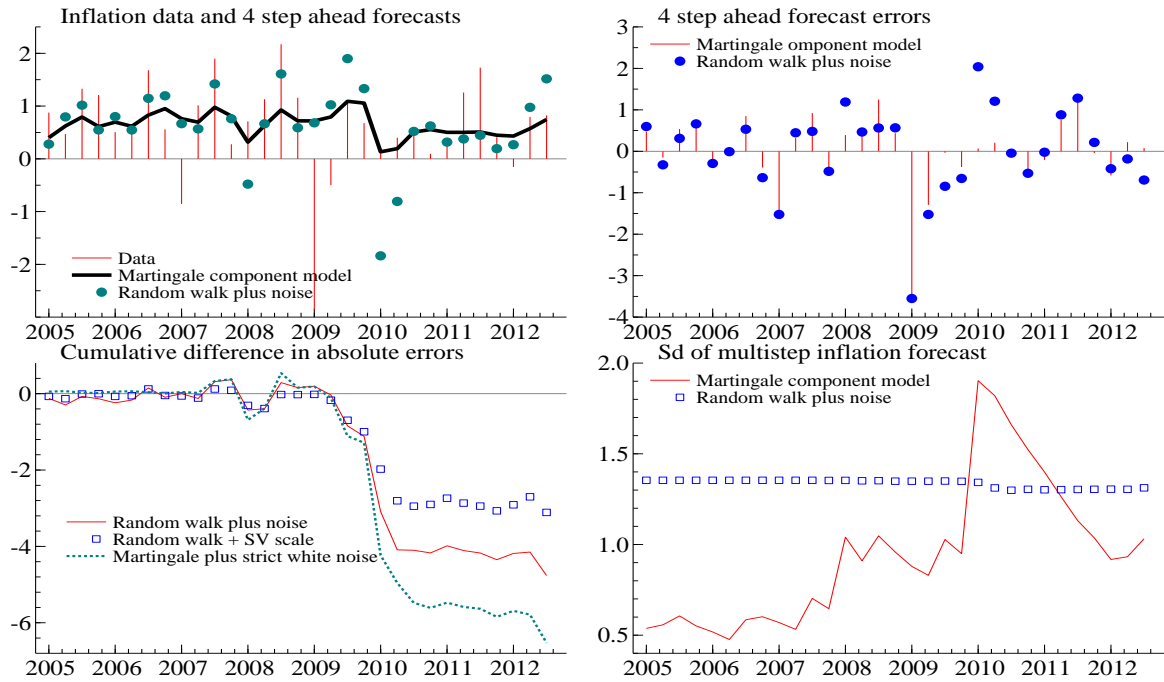


Figure 11: Out of sample multistep (year ahead) ahead forecasts for the inflation series. Top left are the raw quarterly data and the multistep ahead forecasts of those series using the martingale unobserved component model and random walk plus noise model. Top right are the forecast errors from the models. Bottom left is the cumulate sum of the difference of the absolute forecast errors from the three models compared to the martingale unobserved component model. If the cumulative sum is below zero this is some evidence in favour of the forecasts from the martingale unobserved component model. Bottom right is the standard deviation of the 4 step ahead forecast of the inflation series, shown for the two models. The results here are very different.

## 6 Conclusion

This paper looks at the Gaussian local level model and inflation forecasting, generalises it by allowing for stochastic volatility scaling and allowing the signal/noise ratio to change through time. I call this type of model a martingale unobserved component model. The model is relatively easy to handle and extends to a wide class of important time series model. The diagnostics of the model are calculated and compared to the results for various special cases.

The particle filter is used to handle the model, which extends the Kalman filter to allow for non-Gaussianity and non-linearity. The particle filter is used to generate an estimate of the log-likelihood which is used inside a MCMC algorithm, in order to make inference on the parameters of the model. The MCMC chains are quite well behaved and are simple to parallelise to exploit multicore computers or indeed GPUs.

The martingale unobserved component model generalises in various ways, to allow for trends, cycles and for seasonal components. The methods developed here can be extended in the same way. The martingale unobserved component model can also be set in continuous time and used to look at high frequency financial data, where stochastic volatility scaling is clearly very important. Likewise non-parametric regression can be analysed using this kind of model, as noted by for example Wecker and Ansley (1983), Kohn, Ansley, and Wong (1992) and Harvey and Koopman (2000). By allowing for the signal/noise ratio to change through time, the approach discussed here allows the smoothing to be carried out with in effect a local bandwidth which might be important for some applications.

## References

- Ackerson, G. A. and K. S. Fu (1970). On state estimation in switching environments. *IEEE Transactions on Automatic Control* 15, 10–17.
- Akashi, H. and H. Kumamoto (1977). Random sampling approach to state estimation in switching environments. *Automatica* 13, 429–434.
- Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society, Series B* 72, 1–33.
- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* 76, 1481–1536.
- Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B* 64, 253–280.
- Beaumont, M. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics* 164, 1139.
- Bos, C. and N. Shephard (2006). Inference for adaptive time series models: stochastic volatility and conditionally Gaussian state space form. *Econometric Reviews* 25, 219–244.
- Caldara, D., J. Fernandez-Villaverde, P. Guerron, J. F. Rubio-Ramirez, and Y. Wen (2012). Computing dsge models with recursive preferences and stochastic volatility. *Review of Economic Dynamics* 15, 188–206.
- Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika* 81, 541–553.
- Chen, R. and J. S. Liu (2000). Mixture Kalman filters. *Journal of the Royal Statistical Society, Series B* 62, 493–508.
- Cogley, T., G. Primiceri, and T. J. Sargent (2010). Inflation-gap persistence in the U.S. *American Economic Journal: Macroeconomics* 2, 43–69.
- Creal, D. (2012). A survey of sequential Monte Carlo methods for economics and finance. *Econometric Reviews* 31, 245–296.
- Creal, D., S. J. Koopman, and A. Lucas (2008). A general framework for observation driven time-varying parameter models. Unpublished paper: Tinbergen Institute Discussion Paper 108.
- Creal, D., S. J. Koopman, and E. Zivot (2010). Extracting a robust U.S. business cycle using a time-varying multivariate model-based bandpass filter. *Journal of Applied Econometrics* 25, 695–719.

- D'Agostino, A., L. Gambetti, and D. Giannone (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics* 28, 82–101.
- de Jong, P. (1991). The diffuse Kalman filter. *Annals of Statistics* 19, 1073–1083.
- Diebold, F. X., T. A. Gunther, and T. S. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39, 863–883.
- Doucet, A., N. de Freitas, and N. J. Gordon (Eds.) (2001). *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag.
- Doucet, A., S. J. Godsill, and C. Andrieu (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing* 10, 197–208.
- Doucet, A. and A. Johansen (2011). A tutorial on particle filtering and smoothing: fifteen years later. In D. Crisan and B. Rozovsky (Eds.), *The Oxford Handbook of Nonlinear filtering*. Oxford University Press.
- Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.
- Fearnhead, P. and P. Clifford (2003). On-line inference for hidden Markov models via particle filters. *Journal of the Royal Statistical Society, Series B* 65, 887–899.
- Fernandez-Villaverde, J., P. Guerron, J. F. Rudio-Ramirez, and M. Uribe (2010). Risk matters: The real effects of volatility shocks. *American Economic Review* 101, 2530–2561.
- Fiorentini, G., E. Sentana, and N. Shephard (2004). Likelihood-based estimation of latent generalised ARCH structures. *Econometrica* 72, 1481–1517.
- Flury, T. and N. Shephard (2011). Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models. *Econometric Theory* 27, 933–956.
- Ghysels, E., A. C. Harvey, and E. Renault (1996). Stochastic volatility. In C. R. Rao and G. S. Maddala (Eds.), *Statistical Methods in Finance*, pp. 119–191. Amsterdam: North-Holland.
- Gordon, N. J., D. J. Salmond, and A. F. M. Smith (1993). A novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE-Proceedings F* 140, 107–113.
- Gourieroux, C. and A. Monfort (1996). *Simulation Based Econometric Methods*. Oxford: Oxford University Press.
- Harvey, A. C. (1981). *Time Series Models* (1 ed.). New York: Philip Allan.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Harvey, A. C. and J. Durbin (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 B* 49, 187–227.
- Harvey, A. C. and S. J. Koopman (2000). Signal extraction and the formulation of unobserved components models. *Econometrics Journal* 3, 84–107.
- Harvey, A. C., E. Ruiz, and E. Sentana (1992). Unobserved component time series models with ARCH disturbances. *Journal of Econometrics* 52, 129–158.
- Harvey, A. C., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies* 61, 247–264.
- Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361–393.
- Kohn, R., C. F. Ansley, and C.-M. Wong (1992). Nonparametric spline regression with autoregressive moving average errors. *Biometrika* 79, 335–346.
- Koopman, S. J. and C. Bos (2004). State space models with a common stochastic variance. *Journal of Business and Economic Statistics* 22, 346–357.
- Liu, J. S. and R. Chen (1998). Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association* 93, 1032–1044.
- Muth, J. F. (1960). Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association* 55, 299–306.

- Pitt, M. K. and N. Shephard (1999). Filtering via simulation: auxiliary particle filter. *Journal of the American Statistical Association* 94, 590–599.
- Pitt, M. K., R. Silva, P. Giordani, and R. Kohn (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics*. Forthcoming.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics* 23, 470–472.
- Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika* 81, 115–131.
- Shephard, N. (Ed.) (2005). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.
- Stock, J. H. and M. W. Watson (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit, and Banking* 39, 3–34.
- Taylor, J. W. (2004). Smooth transition exponential smoothing. *Journal of Forecasting* 23, 385–394.
- Trigg, D. W. and A. G. Leach (1967). Exponential smoothing with an adaptive response rate. *Operational Research Quarterly* 18, 53–59.
- Wecker, W. E. and C. F. Ansley (1983). The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association* 78, 81–89.