Local scale models

State space alternative to integrated GARCH processes

Neil Shephard*

Nuffield College, Oxford OX1 1NF, UK

State space alternative to autoregressive conditional heteroskedasticity models are proposed. The initial model, which is labelled the Gaussian local scale model, has a measurement density which is Gaussian, conditional on the unobservable precision. The precision is assumed to be a gamma variable which evolves by being scaled by a beta variable. The resulting forecast is a student's t random variable, with a scale which is approximately an exponentially weighted moving average (EWMA) of the squares of the past observations. The degrees of freedom of the student's t distribution is controlled by the size of the discount parameter of the EWMA procedure. The Gaussianity of the measurement density is potentially inadequate when the model is applied to heavy tailed finance data. Instead, this assumption can be replaced by an exponential power density which allows the observed excess kurtosis to be modelled. The choice of the exponential power means that the model still maintains conjugacy, so allowing the derivation of an exact filter and likelihood function. This model is called the generalised local scale model. It has been used to model two exchange rate series.

1. Introduction

The recent development of time series models for heteroskedastic processes has had a profound influence on the way many economists specify their time series models. Engle's (1982) seminal work on autoregressive conditional hetero-skedastic (ARCH) processes has set the tone for this literature. If Y_{t-1} denotes the information set available at time $t - 1$, then a Gaussian pth-order ARCH

process specifies that

$$y_t \mid Y_{t-1} \sim N(0, \alpha_0 + \alpha_1 y_{t-1}^2 + \cdots + \alpha_p y_{t-p}^2). \qquad (1)$$

In practice, a more useful model is the generalised ARCH (GARCH) process. This model is usually credited to Bollerslev (1986), although the GARCH(1, 1) was discovered at the same time by Taylor (1986) who called it an ARMACH process. The Gaussian GARCH(1, 1) process is given by

$$y_t \mid Y_{t-1} \sim N(0, \sigma_t^2), \qquad (2a)$$

where

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 \sigma_{t-1}^2. \qquad (2b)$$

When $\alpha_1 + \alpha_2 = 1$, the process is called integrated GARCH (IGARCH). If α_0 is strictly positive, then the IGARCH model is no longer covariance stationary, but it is strictly stationary. Its properties have recently been studied by Nelson (1990a). Although at first sight the integration restriction seems very strong, Bollerslev and Engle (1993) suggest that there is, for high frequency financial data, a ' . . . *pronounced empirical regularity [providing] a strong motivation for the IGARCH class of models*'. Further, Nelson (1992) has provided a theoretical rationale for the tendency to estimate IGARCH models in high frequency financial time series.

The IGARCH recursion for σ_t^2 is reminiscent of an exponentially weighted moving average (EWMA) updating equation for the squares of the observations, with α_1 playing the role of a smoothing constant, while α_0 is a slope term. Although this analogy is informative, we can see that knowledge (perhaps obtained by consistent estimation) of α_0 and α_1 will (asymptotically) reveal the value of the evolving variance. This is quite an unusual use of the EWMA technique as it directly forms a stochastic element, rather than estimating the level of an unobserved component.

To develop this analogy a little further, recall that Muth (1960) showed that the Gaussian local level model,

$$y_t = \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2), \qquad (3a)$$

$$\mu_t = \mu_{t-1} + \eta_t, \qquad \eta_t \sim NID(0, \sigma_\eta^2), \qquad (3b)$$

$$\mu_0 \mid Y_0 \sim N(m_0, p_0), \qquad (3c)$$

where (ε_t), (η_t), and $\mu_0 \mid Y_0$ are totally independent, provides a rationale for taking an EWMA of the observations. By this we mean that the unobservable

level μ_{t-1} is estimated by $m_{t-1} = E\mu_{t-1} \mid Y_{t-1}$, which is an EWMA of the observations if t is reasonably large. This result, which gives the filtered estimate of the level, implies that for the local level model,

$$y_t \mid Y_{t-1} \sim N(m_{t-1}, p_{t-1} + \sigma_\eta^2 + \sigma_\varepsilon^2) = N(m_{t-1}, f_t). \qquad (4)$$

This one-step density is the natural location model analog to the IGARCH model.

Of course, $E\mu_t \mid Y_t$ is not the only object of interest, for if we wanted to estimate the level of the series at time t, we may wish to use the whole data set to carry this out. This operation, which is called smoothing, delivers the quantity $E\mu_t \mid Y_T$ and is more precise than the corresponding filtered estimate unless σ_ε^2 is zero; see, for example, de Jong (1989).

In this paper we develop local scale models which have unobservable variances. Typically the filtered estimate of the variance is approximately an EWMA of the squares of the observations, so providing some rationale for the IGARCH mechanism with a nonnormal one-step-ahead forecast density. A smoothing algorithm for the unobservable variance is derived. Empirical work indicates that it offers a large efficiency gain over the corresponding filter.

The notion of an evolving unobservable variance has much appeal in economics. The IGARCH mechanism allows the variance to depend only on past observations which are observed by the econometrician. This is a perfectly reasonable way of constructing the one-step-ahead forecast density, but it tells us only a small amount about the actual volatility on a specific day. This is because the IGARCH estimate of the variance at time t ignores y_t^2, the most informative of all the observations, as well as all the future observations. This will lead the econometrician to lose over half of the available information if he uses the IGARCH mechanism for this purpose.

Evolving, unobservable variance models are not new, they have been frequently suggested in the finance literature. A review article on this topic has recently been written by Taylor (1991). He concentrates on the autoregressive stochastic variance, or stochastic volatility, model. This has been analysed by Hull and White (1987), Scott (1987), Wiggins (1987), Chesney and Scott (1989), and Melino and Turnbull (1990) in order to solve the Black–Scholes option pricing problem. It turns out that the local scale model is very close to these models, as we will see in section 2.

The rest of this paper is organised in the following manner. In section 2, a Gaussian local scale model is proposed. The model has a Gaussian measurement density, with a gamma transition equation. The third section develops an exact filter for the model allowing the tracking of the evolving variance. Further, the one-step-ahead forecast is shown to be a scaled student's t random variable. The model is capable of dealing with irregularly spaced observations, missing values, and can provide analytic multi-step prediction densities, unlike the

ARCH-based models. In the fourth section, the corresponding smoother is developed. In section 5, the model is combined with the Gaussian state space form. The role of the Kalman filter in this new setup is outlined. In section 6, the measurement density is generalised from a normal to an exponential power density [sometimes called the generalised error distribution; see Box and Tiao (1973)]. The new model allows much more flexible behaviour and can deal with the extreme observations which frequently occur in financial series. The corresponding filter, smoother, and likelihood are derived for this model. The local scale models are applied to two exchange rate series in section 7. Various nonlinear generalisations of the model are discussed in section 8. All the main themes of the paper are pulled together in section 9, where the conclusions are put forward.

2. A Gaussian local scale model

In this paper the precision θ_t, which is the inverse of the variance, will be directly modelled. The measurement density of the state space model takes on the form

$$y_t | \theta_t \sim \text{N}(0, \theta_t^{-1}), \qquad t = 1, \ldots, T. \tag{5}$$

The normal density mixes exactly with the gamma density and so we must become interested in a transition equation for gamma random variables.

There has been considerable interest in gamma transition equations in the statistical modelling literature [see, in a different context, Lewis, McKenzie, and Hugus (1989)]. Most of the work can be traced back to Bather (1965), although it was Smith and Miller (1986) who first wrote down gamma-based state space models. Their measurement equation was exponential, allowing the time series analysis of extreme values. This work was picked up by Harvey and Fernandes (1989a, b) in their time series analysis of Poisson observations and insurance claims; see also Harvey (1989, pp. 348–363).

The basic transition equation of Harvey and Fernandes (1989a, b) was

$$\theta_t = \omega^{-1} \theta_{t-1} \eta_t, \qquad t = 1, \ldots, T, \tag{6a}$$

where $\omega \in (0, 1]$ and

$$\eta_t \sim \text{Beta}(\omega a_{t-1}, (1-\omega)a_{t-1}). \tag{6b}$$

ω controls the speed at which the precision moves. In the limit, as $\omega \to 1$, $\theta_t \to \theta_{t-1}$ almost surely, while as ω reduces, the movements in θ_t become progressively more exaggerated. The rather curious a_{t-1} term is needed so that

the densities are conjugate; its role will become clearer in section 3. It comes from the prior distribution for θ_{t-1}, which in turn is driven by the prior for θ_0. To be more precise, we assume that $\theta_0 \mid Y_0 \sim G(a_0, b_0)$. The notation $X \sim G(\alpha, \beta)$ is written when X is a gamma random variable. The form of the density is taken to be

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1} \beta^\alpha \exp(-x\beta)}{\Gamma(\alpha)}, \qquad x, \alpha, \beta > 0. \tag{7}$$

This setup of a measurement density, transition equation, and prior density precisely analogs the usual Gaussian state space form for the local level model – see the Introduction.

In the next section, we will see that we can derive a filter for this local scale model so that $\theta_1 \mid Y_0$ and $\theta_1 \mid Y_1$ are both gamma random variables. This means that conjugacy is maintained as we go through time. This result exactly matches the Kalman filter for the Gaussian state space form, where the conditional distributions for the state are all Gaussian.

Before we progress we will have to slightly alter our transition equation. Unfortunately, using a similar argument to that of Nelson (1990a) on integrated GARCH processes, it is possible to demonstrate that if $\omega < 1$, $\theta_t \to 0$ almost surely as $t \to \infty$. A simple explanation for this is that $\log \theta_t / \theta_{t-1}$ approximates the growth rate in the level of the precision, and since $\log \theta_t / \theta_{t-1} = \log \eta_t - \log \omega$, its growth rate is negative on average by Jensen's inequality.

The transition equation which allows the elimination of this problem is

$$\theta_t = e^{r_t} \theta_{t-1} \eta_t, \qquad t = 1, \dots, T, \tag{8a}$$

where

$$r_t = - E \log \eta_t = \Psi(a_{t-1}) - \Psi(\omega a_{t-1}), \tag{8b}$$

since this equation has a zero expected growth rate. Eq. (8) is the transition equation which is adopted throughout the remaining parts of this paper. Here $\Psi(\cdot)$ denotes the Euler Psi, or digamma, function. This function can be evaluated using the approximations developed in Abramowitz and Stegun (1970, p. 258). An alternative expression for the computation of r_t, based on Weierstrass's infinite sum representation of the log of the gamma function, is discussed in Shephard (1990).

This new transition equation is a random walk in its logarithm. We follow Nelson's (1990a) arguments by writing

$$\log \theta_t = \log \theta_0 + \sum_{t=1}^{T} (\log \eta_t - r_t). \tag{9}$$

If we start a_0 in a steady state (this will be defined in the next section and will occur if $\omega < 1$), then $(\log \eta_t - r_t)$ are zero mean, independent and identically distributed, random variables. As a result, we can see that $\limsup \log \theta_t = \infty$ and $\liminf \log \theta_t = -\infty$ as $t \to \infty$, with probability one. Hence, (θ_t) is nonstationary and consequently (y_t) inherits the same property. Again, this is the property that we would expect, for the local level model is nonstationary.

The local scale model is very close to the unit root autoregressive stochastic variance models discussed in Harvey, Ruiz, and Shephard (1992). There a rather similar state space model is used with

$$y_t \mid \theta_t \sim N(0, \theta_t^{-1}), \tag{10a}$$

and

$$\log \theta_t = \log \theta_{t-1} + \omega_t, \qquad \omega_t \sim NID(0, \sigma_\omega^2). \tag{10b}$$

The measurement density can be written as

$$y_t = \varepsilon_t \theta_t^{-1/2}, \qquad \varepsilon_t \sim NID(0, 1), \tag{11a}$$

implying that

$$z_t = \log y_t^2 = \log \varepsilon_t^2 - \log \theta_t. \tag{11b}$$

So, z_t has a non-Gaussian, but linear state space representation. Hence the Kalman filter can be used to provide the best linear (using $\log y_t^2$) estimator of the evolving precision, while a pseudo-likelihood function allows σ_ω^2 to be estimated; see Harvey, Ruiz, and Shephard (1992).

The only difference between the Gaussian local scale model and this unit root AR stochastic variance model is that the distribution of ω_t is changed, with it being

$$\omega_t = \log \eta_t - E \log \eta_t \tag{12}$$

in the local scale model case, rather than Gaussian. This small change will allow an exact filter to be derived, rather than relying on a best linear one. As a result the exact likelihood can be found for the local scale model. The price that is paid for this exactness is the lack of modelling flexibility induced by being forced, in order to maintain conjugacy, to have a random walk in the transition equation rather than some more general autoregression.

3. Filtering the Gaussian local scale model

The complete Gaussian local scale model is given by

$$y_t \,|\, \theta_t \sim N(0, \theta_t^{-1}), \qquad t = 1, \ldots, T, \tag{13a}$$

$$\theta_t = e^{r_t} \theta_{t-1} \eta_t, \qquad t = 1, \ldots, T, \tag{13b}$$

$$\theta_0 \,|\, Y_0 \sim G(a_0, b_0). \tag{13c}$$

If we combine the transition equation with the prior, then we have that

$$\theta_1 \,|\, Y_0 \sim G(\omega a_0, e^{-r_1} b_0) = G(a_{1|0}, b_{1|0}). \tag{14}$$

Bayes' theorem then delivers

$$\theta_1 \,|\, Y_1 \sim G(a_{1|0} + \tfrac{1}{2}, b_{1|0} + \tfrac{1}{2} y_1^2) = G(a_1, b_1). \tag{15}$$

As these results give a conjugate form, the model is recursive by induction, not losing its form as we progress through time.

If we run the filter through time, we get the conditional densities

$$\theta_t \,|\, Y_t \sim G(a_t, b_t) \quad \text{and} \quad \theta_t \,|\, Y_{t-1} \sim G(a_{t|t-1}, b_{t|t-1}), \tag{16a}$$

where

$$a_t = a_{t|t-1} + \tfrac{1}{2}, \qquad a_{t|t-1} = \omega a_{t-1} \tag{16b}$$

and

$$b_t = b_{t|t-1} + \tfrac{1}{2} y_t^2, \qquad b_{t|t-1} = e^{-r_t} b_{t-1}. \tag{16c}$$

This implies that if $a_1 = \tfrac{1}{2}$ and $b_1 = \tfrac{1}{2} y_1^2$ [see Harvey (1989, p. 351) for some justification for this], then

$$a_T = \tfrac{1}{2} \sum_{i=1}^{T} \omega^{i-1} \tag{17a}$$

and

$$b_T = \tfrac{1}{2} \sum_{i=1}^{T} \gamma_{T-i+1} y_{T-i+1}^2, \tag{17b}$$

where $\gamma_T = 1$ and

$$\gamma_{T-i} = \prod_{j=1}^{i} e^{-r_{T-j+1}}. \tag{17c}$$

If ω is reasonably large, for example greater than 0.8, then e^{-r_t} will be very close to ω, implying a_T/b_T is roughly the inverse of the exponentially weighted moving average (EWMA) of the squares of the observations. If the Harvey and Fernandes (1989a, b) transition equation had been used, then the EWMA result would be exact. These results are important for $E\theta_t \mid Y_t = a_t/b_t$, so the filtered estimate of θ_t is the inverse of an EWMA.

A similarly simple result also holds for the one-step-ahead forecast density. It takes the form

$$
\begin{aligned}
f(y_t \mid Y_{t-1}) &= \int_0^\infty f(y_t, \theta_t \mid Y_{t-1}) \, \mathrm{d}\theta_t \\[2mm]
&= \int_0^\infty f(y_t \mid \theta_t) f(\theta_t \mid Y_{t-1}) \, \mathrm{d}\theta_t \\[2mm]
&= \frac{1}{\sqrt{2\pi}} \left(\frac{b_{t|t-1}}{b_t} \right)^{a_t} \frac{1}{\sqrt{b_{t|t-1}}} \frac{\Gamma(a_t)}{\Gamma(a_{t|t-1})} \\[2mm]
&= \frac{1}{\sqrt{2\pi a_{t|t-1}}} \frac{\Gamma(a_t)}{\Gamma(a_{t|t-1})} \left(1 + \frac{y_t^2}{2b_{t|t-1}} \right)^{-a_t} \frac{1}{\sqrt{b_{t|t-1}/a_{t|t-1}}}.
\end{aligned}
\tag{18}
$$

This implies that $y_t \mid Y_{t-1}$ is distributed as a student's t distribution with $2a_{t|t-1}$ degrees of freedom, with location 0, and scale $b_{t|t-1}/a_{t|t-1}$. We will write this as

$$T_{2a_{t|t-1}}(0, \, b_{t|t-1}/a_{t|t-1}). \tag{19}$$

The scale in this distribution is approximately an EWMA of the squares of the observations, with the data only entering (19) through the $b_{t|t-1}$ term. As such it provides a sound theoretical foundation for estimating scale by using the EWMA of the squares – see Taylor (1986, p. 104) who uses this technique to predict volatility and Taylor and Kingsman (1979) who suggested ways of estimating the discount parameter of the EWMA. In their work they do not impose the connection between this parameter and the degrees of freedom of the forecast density.

A somewhat similar distributional result has been established for the continuous time limit (as the space between observations goes to zero) of the stationary

GARCH(1, 1) process by Nelson (1990b, theorem 2.3). He proves the interesting result that the unconditional distribution of the precision for the GARCH(1, 1) process asymptotically converges to a gamma random variable, implying the unconditional distribution of the observations is student's t. Although the genesis of our results are very different, the stylised facts are similar for they both indicate heavy tailed behaviour for our models caused by changing variances.

Eq. (18) implies that the prediction density can be used to enable us to write down the likelihood function, for unknown ω, by using the prediction decomposition. In an obvious notation,

$$\log L(\omega; y \mid Y_0) = \sum_{t=1}^{T} \log L(\omega; y_t \mid Y_{t-1})$$

$$= \text{const} + \sum_{t=1}^{T} a_t \log(b_{t|t-1}/b_t) + \log \Gamma(a_t) \qquad (20)$$

$$- \log \Gamma(a_{t|t-1}) - \tfrac{1}{2} \log b_{t|t-1}.$$

The sequence a_t rapidly converges to a steady state as t increases if $\omega < 1$. This is important for the one-step-ahead forecast density is indexed by $2a_{t|t-1}$, in its degrees of freedom. It is easy to see that

$$2a_{t|t-1} \cong \omega/(1 - \omega). \qquad (21)$$

Low values of ω mean that the forecast density will have very thick tails, while as $\omega \to 1$ the density will converge to normality. The rth moment of the forecast will exist if $\omega > r/(1 + r)$. If $\omega = \tfrac{1}{2}$, the forecast density is Cauchy and none of its moments will exist. ω has to be greater than $\tfrac{4}{5}$ for the fourth moment to exist.

The local scale model can be extended to include fixed location shifts in the measurement equation. A straightforward example of this is the linear regression model $y_t \mid \theta_t \sim \text{N}(x_t'\beta, \theta_t^{-1})$. The above analysis can be carried over to this case with the only change being that $b_{t+1} = b_{t+1|t} + (y_{t+1} - x_{t+1}'\beta)^2/2$. Of course, the unknown β vector could then be estimated by maximum likelihood methods.

The local scale model can also be generalised to deal with the possibilities of irregularly spaced or missing observations. If we write the measurement density as $y_{\tau+1} \mid \theta_{\tau+1} \sim \text{N}(0, \theta_{\tau+1}^{-1})$ and $\theta_{\tau+1} e^{-r_{\tau+1}} \theta_\tau^{-1} \mid \theta_\tau \sim \text{Beta}(\omega^{\Delta_{\tau+1}} a_\tau, (1 - \omega^{\Delta_{\tau+1}}) a_\tau)$, where the τth observation occurs at time t_τ and $\Delta_\tau = t_\tau - t_{\tau-1}$,

then the previous arguments carry over to deliver the updating equations

$$a_{\tau+1|\tau} = \omega^{\Delta_{\tau+1}} a_\tau, \tag{22a}$$

$$a_{\tau+1} = a_{\tau+1|\tau} + \tfrac{1}{2}, \tag{22b}$$

$$b_{\tau+1|\tau} = e^{-r_{\tau+1}} b_\tau, \tag{22c}$$

$$b_{t+1} = b_{\tau+1|\tau} + \tfrac{1}{2} y_{\tau+1}^2. \tag{22d}$$

The only complication of this formulation is that $a_{\tau+1}$ does not in general converge to a steady state. The expected value of $\theta_\tau | Y_\tau$ can be interpreted as the inverse of an EWMA of past squared observations in which the power to which each of the observations' weights are raised is equal to the time elapsed since it occurred.

For high frequency data, Δ_τ will vary considerably, meaning in turn that ω^{Δ_τ} will change substantially. As Δ_τ lengthens, ω^{Δ_τ} will increase, allowing θ_τ to move more rapidly implying an increase in kurtosis in the data. Conversely, small Δ_τ will reduce the nonnormality implied by the model. Hence, in the context of financial data, one consequence of the model is that periods which exhibit heavy trading will be modelled more like Brownian motion than thin periods which will display fatter tails.

The extension to irregularly spaced observations also allows us to construct s-step-ahead predictions for regularly spaced data. This is because the multi-step forecast could be thought of as being a single-step forecast for irregularly spaced data. An interesting feature of the resulting forecast density is that the degrees of freedom fall as s increases, changing the shape of the density, increasing its kurtosis, and reducing the number of moments which exist.

This density will only approximate the actual multi-step prediction density because it ignores the intermediate observations which affect the stochastic path of the actual precision. To assess the degree of error induced by this approximation we will compare the moments of the true multi-step precision,

$$\theta_{T+s} = \theta_T \left[\prod_{i=1}^{s} e^{r_{T+i}} \eta_{T+i} \right], \tag{23}$$

with the one-step approximation, $\theta_{T+1}^* = \theta_T \eta_{T+1}^* e^{r_{T+1}^*}$, conditional on the value of θ_T. Using the properties of beta random variables, we determine that

$$E\theta_{T+s}^c | \theta_T = \theta_T^c \left[\prod_{i=1}^{s} e^{r_{T+i}} \right]^c \left[\prod_{j=0}^{c-1} \frac{\omega a_T + j}{a_T + j} \right]^s, \tag{24a}$$

Table 1.
Ratio of multi-step-ahead to one-step-ahead forecasts.

Moments	Steps				
	1	2	3	4	5
1	1.000	1.000	1.001	1.001	1.002
2	1.000	1.000	1.000	1.001	1.001
3	1.000	1.000	0.999	0.999	0.998
4	1.000	0.999	0.998	0.994	0.994

while

$$
E\theta^{*c}_{T+1} \mid \theta_T = \theta^c_T e^{r^*_{T+1}} \prod_{j=0}^{c-1} \frac{\omega^s a_T + j}{a_T + j}.
\tag{24b}
$$

If ω is large $\prod e^{r_{T+i}} \cong e^{r^*_{T+1}}$ if s is small. Likewise if ω is large, a_T will be large implying the products will be similar if s is not too big. These approximations deteriorate as c increases.

To strengthen these general remarks, we give in table 1 results for the case $\omega = 0.93$. The reported numbers are the ratio of the two moments, with $E\theta^c_{T+s} \mid \theta_T$ in the numerator. The table indicates a remarkable agreement between the two sets of figures, which suggests that the degree of approximation incurred by using the multi-step predictor is extremely mild.

4. Smoothing the Gaussian local scale model

Potentially there are large efficiency gains to be achieved from developing a smoother for tracking the unobservable precision. The filter's density $\theta_t \mid Y_t$ uses only contemporaneously available data, the IGARCH mechanism looks at $\theta_t \mid Y_{t-1}$ and so holds even less, while the smoothing density exploits all the historical data and so should be more precise. The likelihood of substantial efficiency gains means that smoothing has attracted a great deal of attention in the literature on Gaussian state space models; see, for example, Anderson and Moore (1979) and de Jong (1989). In this section we develop summaries of the $\theta_t \mid Y_T$ density.

Clearly the exact density of $\theta_t \mid Y_T$ could be computed by using a numerical integration procedure such as that suggested by Kitagawa (1987). However, using this routine on such a lengthy series is likely to be both expensive and perhaps numerically unstable. Consequently, Harvey's (1989, p. 359) suggestion of a quasi-smoother for $E\theta_t \mid Y_T$ is adopted. The quasi-smoother uses the property that the filtered estimate of θ_t is approximately the inverse of the

EWMA of the squared observations and that the local level model provides a statistical rationale for the EWMA process. Hence the corresponding smoother for the local level model is used on the squares of the observations to obtain approximately a double EWMA [see Whittle (1991)], using a value for the signal-to-noise ratio, $\sigma_\eta^2/\sigma_\varepsilon^2$, of $(1 + \omega^2 - 2\omega)/\omega$.

5. State space location models

The Gaussian local scale model can be incorporated into autoregressive or even mixed autoregressive integrated moving average models. However, the unobserved component nature of the model fits more naturally within the structural time series models of Harvey (1989) or the dynamic linear models of West and Harrison (1989). These models are formulated directly in terms of interpretable components of interest, such as trends, seasonals, and cycles. We will now examine how the endogenous heteroskedasticity of the local scale model can be mixed into these models.

The basic model will be the state space form with the gamma transition equation for the precision

$$y_t \mid \theta_t, \mu_t \sim \mathrm{N}(Z_t\mu_t, \theta_t^{-1}H_t), \tag{25a}$$

$$\mu_t \mid \theta_t, \mu_{t-1} \sim \mathrm{N}(T_t\mu_{t-1}, \theta_t^{-1}Q_t), \tag{25b}$$

$$\theta_t \mathrm{e}^{-r_t}\theta_{t-1}^{-1} \mid \theta_{t-1} \sim \mathrm{Beta}(\omega a_{t-1}, (1-\omega)a_{t-1}), \tag{25c}$$

where Z_t, H_t, T_t, and Q_t are known matrices. If $S_t = (\theta_t, \theta_{t-1}, \ldots, \theta_1)$, then the Kalman filter provides the densities

$$y_{t+1} \mid Y_t, S_{t+1} \sim \mathrm{N}(Z_{t+1}m_{t+1|t}, \theta_{t+1}^{-1}F_{t+1}), \tag{26a}$$

$$\mu_{t+1} \mid Y_t, S_{t+1} \sim \mathrm{N}(m_{t+1|t}, \theta_{t+1}^{-1}P_{t+1|t}), \tag{26b}$$

$$\mu_t \mid Y_t, S_t \sim \mathrm{N}(m_t, \theta_t^{-1}P_t), \tag{26c}$$

where

$$m_{t+1|t} = T_{t+1}m_t, \quad m_t = m_{t|t-1} + P_{t|t-1}Z_t'F_t^{-1}(y_t - Z_tm_{t|t-1}), \tag{26d}$$

$$F_t = Z_tP_{t|t-1}Z_t' + H_t, \quad P_t = P_{t|t-1} - P_{t|t-1}Z_t'F_t^{-1}Z_tP_{t|t-1}, \tag{26e}$$

$$P_{t+1|t} = (\theta_{t+1}/\theta_t)T_{t+1}P_tT_{t+1}' + Q_{t+1}. \tag{26f}$$

The presence in the filter of the unknown ratio θ_{t+1} to θ_t will mean that numerical integration rules would have to be used to compute the exact density of $y_{t+1} \mid Y_t$. Following the suggestion of West and Harrison (1989, p. 370), we will avoid the need for these calculations by replacing θ_{t+1}/θ_t with an estimate derived from the information available up to time t. The value selected is $c_{t+1} = (a_t b_{t-1})/(b_t a_{t-1})$, which results in $P_{t+1 \mid t}$ taking the form $c_{t+1} T_{t+1} P_t T'_{t+1} + Q_{t+1}$. Thus, c_t is roughly the ratio of the discounted estimates of the variance in the current and previous time periods. If ω is one, then c_t will converge to one. If we were to think of $P_{t \mid t-1} Z'_t F_t^{-1}$ as the discount parameter, then smaller values of ω will allow c_t to move around unity, opening and closing the filter, responding to the heteroskedasticity in the data.

Given this approximation, if we write $d_t = b_t/a_t$ and $d_{t+1 \mid t} = b_{t+1 \mid t}/a_{t+1 \mid t}$, then we find

$$y_{t+1} \mid Y_t \sim T_{2a_{t+1 \mid t}}(Z_{t+1} m_{t+1 \mid t}, d_{t+1 \mid t} F_{t+1}), \tag{27a}$$

$$\mu_{t+1} \mid Y_t \sim T_{2a_{t+1 \mid t}}(m_{t+1 \mid t}, d_{t+1 \mid t} P_{t+1 \mid t}), \tag{27b}$$

$$\mu_t \mid Y_t \sim T_{2a_t}(m_t, d_t P_t). \tag{27c}$$

The density of $\mu_t \mid Y_t$ has r moments if $\omega > 1 - (1/r)$. Therefore, if $\omega > \frac{3}{4}$, the fourth moment will exist.

To complete the location and scale filter we have to deal with the scales' recursions. Only the equation for b changes to become

$$b_{t+1} = b_{t+1 \mid t} + \frac{v'_{t+1} F_{t+1}^{-1} v_{t+1}}{2}, \tag{28a}$$

where

$$v_t = y_t - Z_t m_{t \mid t-1}. \tag{28b}$$

A simple example of this procedure is the local level model, which is given by $Z_t = H_t = T_t = 1$, $Q_t = q$. This model, one of the most fundamental in time series, provides a rationale for the EWMA forecasting scheme. The location filter becomes

$$m_{t+1 \mid t} = m_{t \mid t-1} + \frac{p_{t \mid t-1}}{p_{t \mid t-1} + 1}(y_t - m_{t \mid t-1}), \tag{29a}$$

$$p_{t+1 \mid t} = c_{t+1} \frac{p_{t \mid t-1}}{p_{t \mid t-1} + 1} + q, \tag{29b}$$

while the local scale's b recursion becomes

$$b_{t+1|t} = \left[b_{t|t-1} + \frac{(y_{t+1} - m_{t+1|t})^2}{p_{t+1|t} + 1} \right] e^{-r_{t+1}}. \tag{29c}$$

6. A generalised local scale model

The Gaussian local scale model implies the one-step-ahead forecasts are independent scaled student's t random variables with degrees of freedom which are approximately $\omega/(1 - \omega)$. It maybe necessary to replace the Gaussian measurement density with an exponential power density in order to fatten the tails of the one-step-ahead forecast distribution even more so that the model is consistent with the kind of data found in financial applications. This observation is motivated by the ARCH literature; see, for example, Nelson (1991), Bollerslev (1987), Baillie and DeGennaro (1990), Engle and Bollerslev (1986), and Baillie and Bollerslev (1989). The exponential power density is extensively discussed in Box and Tiao (1973, pp. 156–243). Thus, if X is distributed as an exponential power variable, then we write $X \sim EP(\theta, \beta)$, where the density is

$$f(x; \theta, \beta) = k\theta^{\frac{1}{2}(1+\beta)}\exp(-\tfrac{1}{2}\theta|x|^{2/(1+\beta)}), \tag{30a}$$

where

$$k^{-1} = 2^{1+\frac{1}{2}(1+\beta)}\Gamma(1 + \tfrac{1}{2}(1 + \beta)), \qquad \theta > 0, \quad \beta \in (-1, 1], \tag{30b}$$

and

$$EX = 0, \qquad VX = 2^{(1+\beta)}\theta^{-(1+\beta)}\frac{\Gamma(3(1 + \beta)/2)}{\Gamma((1 + \beta)/2)}. \tag{31}$$

The density for X is symmetric. As β moves from -1 to 1 the density is transformed from the uniform, which is platykurtic, into the double exponential, which is leptokurtic. When β is zero, the density is Gaussian. Hence, when we apply the exponential power density to finance data we should expect β to be nonnegative.

Our generalised local scale model will be given by

$$y_t | \theta_t, \quad \beta \sim EP(\theta_t, \beta), \qquad t = 1, \dots, T, \tag{32a}$$

$$\theta_t = e^{r_t}\theta_{t-1}\eta_t, \qquad\qquad t = 1, \dots, T, \tag{32b}$$

$$\theta_0 | Y_0 \sim G(a_0, b_0). \tag{32c}$$

It is then relatively straightforward exercise to see that this model is recursive and that

$$\theta_t \mid Y_{t-1} \sim G(\omega a_{t-1}, e^{-r_t} b_{t-1}) = G(a_{t\mid t-1}, b_{t\mid t-1}), \tag{33a}$$

$$\theta_t \mid Y_t \sim G(a_{t\mid t-1} + \tfrac{1}{2}(1 + \beta), b_{t\mid t-1} + \tfrac{1}{2}|y_t|^{2/(1+\beta)}) = G(a_t, b_t). \tag{33b}$$

This means that

$$a_T = \frac{(1 + \beta)}{2} \sum_{i=1}^{T} \omega^{i-1}, \tag{34a}$$

$$b_T \cong \tfrac{1}{2} \sum_{i=1}^{T} \omega^{i-1} |y_{T-i+1}|^{2/(1+\beta)}, \tag{34b}$$

if ω is reasonably large. So b_T/a_T is approximately a scaled version of the EWMA of $|y_t|^{2/(1+\beta)}$. So if β approaches unity, which will be the case for the double exponential, then the EWMA will be built out of the absolute values of the observations.

The one-step-ahead forecast density is

$$f(y_t \mid Y_{t-1}) = k \frac{\Gamma(a_t)}{\Gamma(a_{t\mid t-1})} \frac{(b_{t\mid t-1})^{a_t}}{(b_t)^{a_{t\mid t-1}}} = k \frac{\Gamma(a_t)}{\Gamma(a_{t\mid t-1})} \left[\frac{b_{t\mid t-1}}{b_t} \right]^{a_t} \frac{1}{b_{t\mid t-1}^{\frac{1}{2}(1+\beta)}}. \tag{35}$$

This density is a scaled version of what might be called a *generalised student's distribution*. To reveal its structure write

$$x_t = y_t (a_{t\mid t-1}/b_{t\mid t-1})^{\frac{1}{2}(1+\beta)}, \tag{36a}$$

then

$$f(x_t \mid Y_{t-1}) = k^* \left(1 + \frac{|x_t|^{2/(1+\beta)}}{2a_{t\mid t-1}} \right)^{-a_t}, \tag{36b}$$

where

$$k^* = k \frac{\Gamma(a_t)}{(\Gamma(a_{t\mid t-1}) a_{t\mid t-1}^{\frac{1}{2}(1+\beta)})}. \tag{36c}$$

So $x_t \mid Y_{t-1} \sim T_{2a_{t\mid t-1}}(0, 1)$ if β is zero. Otherwise, β will move the density away from the student's t, with $\beta > 0$ having fatter tails and $\beta < 0$ thinner, ceteris

paribus. As $a_{t|t-1} \to \infty$, the density converges to an exponential power distribution. Thus this forecast density provides a rather interesting cross between student's t and the exponential power distributions which are usually used for speculative prices. Clearly all the odd moments of $x_t \mid Y_{t-1}$ are zero if they exist, while the even moments can be derived analytically by using the complete beta integral; as they do not have a simple analytic form we will not reproduce them here. Likewise the distribution function can be calculated by using a call to a routine which computes the incomplete beta integral, for

$$\int_0^r f(x_t \mid Y_{t-1}) \, dt = \tfrac{1}{2} \frac{1}{B(\tfrac{1}{2}(1+\beta), a_{t|t-1})} \int_0^s z^{\frac{1}{2}(1+\beta)-1}(1-z)^{a_{t|t-1}-1} \, dz,$$

(37a)

where

$$s = r^{2/(1+\beta)} / (r^{2/(1+\beta)} + 2a_{t|t-1}).$$

(37b)

7. Empirical application of the local scale model

The logarithmic transformation of the UK sterling and German DM exchange rates against the US dollar are often thought to approximately follow random walks with heteroskedastic disturbances. Here we model the centred first difference of the logs of the weekday close of these rates from 1/10/81 to 28/6/85, giving $T = 946$, although when we come to report our estimated models it should be noted that they were computed using only the last 846 observations as 100 observations were used to startup the procedures [this data was also used in the empirical application reported in Harvey, Ruiz, and Shephard (1992)]. The inherent irregularity of the data, such as the lack of observations over the weekend and bank holidays, will be ignored.

The Box–Ljung statistics for ten lags on the series is 11.19 and 10.03 for the sterling/dollar and DM/dollar rate, respectively, suggesting the random walk hypothesis is maintainable. If the transformed observations are squared, the Box–Ljung statistics become 128.25 and 67.79, indicating strong evidence for nonlinearity. When the Gaussian local scale model was fitted to the sterling/dollar and DM/dollar, the estimator of ω took on the values 0.916 and 0.914, respectively, giving 10.9 and 10.6 degrees of freedom for the one-step-ahead forecast densities. The resulting smoothed and filtered estimates of the precision are displayed in fig. 1 for the sterling/dollar rate and in fig. 2 for the DM/dollar rate. They indicate that there are substantial gains to be made from exploiting historically available information to pin down the precision.

To allow for the possibility of heavy tailed behaviour, generalised local scale models were estimated for these series. The estimated values of ω and β were
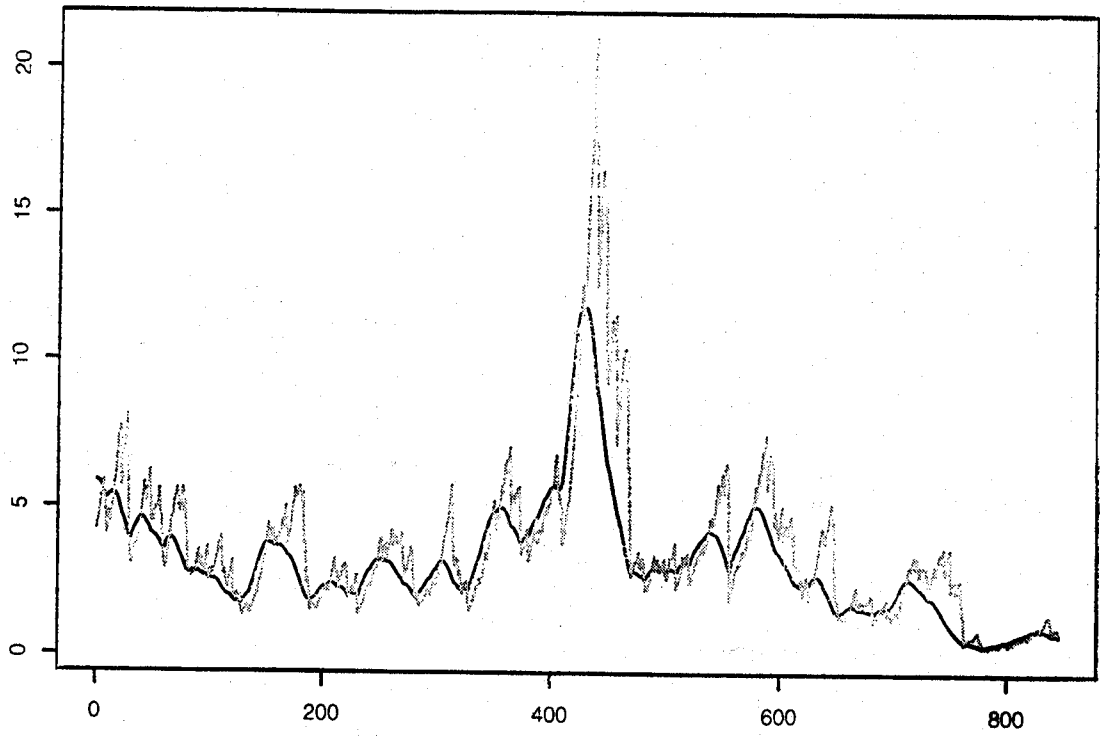
Fig. 1. Sterling/dollar rate: Smoothed (full line) and filtered (dotted line) precision.
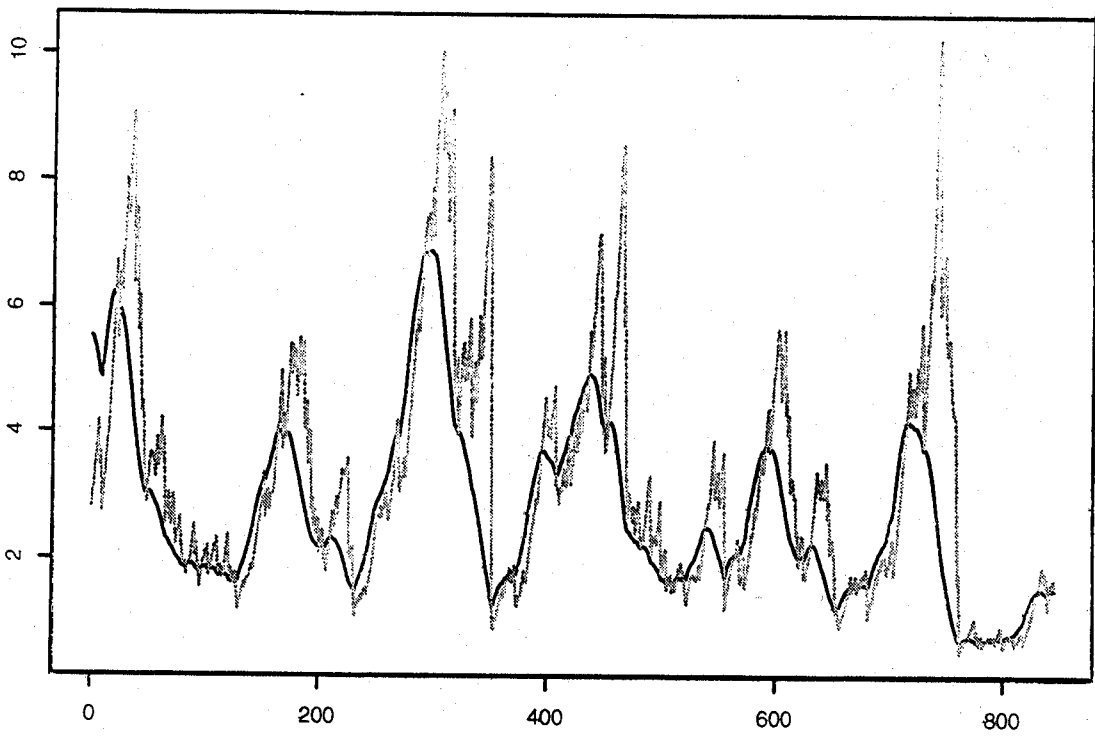

Fig. 2. DM/dollar rate: Smoothed (full line) and filtered (dotted line) precision.

0.926 and 0.14 for sterling and 0.912 and -0.01 for the DM. The log-likelihoods (ignoring constants) for these models were -35.3 and -60.8, respectively, against -36.8 and -60.8 for the Gaussian local scale model. This implies the estimated value of β is not significantly different from zero in both cases, although there is some evidence that the series for Sterling has heavy tails. These results suggest that the fat tails observed in exchange rates are almost completely due to the estimation of the unobservable precision.

To benchmark these results, a variety of GARCH(1, 1) models were estimated for these two series. To allow for a fair comparison with the estimated local scale models, the normal assumption of the GARCH model in eq. (2) will be generalised. Fat tails will be allowed by introducing the student's t GARCH(1, 1) model analysed by, for instance, Bollerslev (1987). This has that

$$f(y_t \mid Y_{t-1}; v, \alpha_0, \alpha_1, \alpha_2) = \frac{1}{\sqrt{2\pi v \sigma_t^2}} \frac{\Gamma\left(\dfrac{v+1}{2}\right)}{\Gamma\left(\dfrac{v}{2}\right)} \left(1 + \frac{y_t^2}{\sigma_t^2}\right)^{-(v+1)/2},$$

(38a)

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 \sigma_{t-1}^2.$$

(38b)

Table 2 reports the results from an unconstrained estimation of (38), the case where $v = \infty$ [the normal case, the model (2)], and the IGARCH model, where α_2 is constrained to be $1 - \alpha_1$.

The estimates indicate that the DM has less fat tails than sterling, which agrees with the local scale results. The IGARCH constraint that $\alpha_1 + \alpha_2 = 1$ is rejected using a conventional likelihood ratio test for both exchange rates, even though $\alpha_1 + \alpha_2$ is 0.96 and 0.97 for the DM and sterling, respectively.

Table 2 also gives the *AIC* and *BIC* criteria for the fit of the models [see Tong (1990, pp. 285–292) for an interesting discussion of these techniques]; they take the form $AIC = -2\log(\text{maximised likelihood}) + 2(\text{number of independently adjusted parameters})$ and $BIC = -2\log(\text{maximised likelihood}) + \log(T)(\text{number of independently adjusted parameters})$. These statistics attempt to penalise the fall in the likelihood by a factor which reflects the movement to a more involved model – in this case ones with heavy tails and more complicated dynamics. Its application here is standard for each possible model is fully parametric. Once again, these measures support the adoption of the general, unconstrained, fat tailed GARCH(1, 1) model, eq. (38).

Table 2 also reports the *AIC* and *BIC* for the local scale models. For both exchange rates fat tailed GARCH(1, 1) beats both of the local scale models using *AIC*, but the parsimonious local scale models overturn this on the *BIC* measure. Again, these statistics suggest no benefit from the use of the

Table 2.

Estimates of GARCH(1, 1) type models.

Series[a]	ν	α_0	α_1	α_2	*Lik.*	*AIC*	*BIC*
(1) DM	∞	0.0175	0.10	0.87	-62.3	130.7	144.9
(2) DM	16	0.0021	0.10	0.90	-60.9	127.9	142.1
(3) DM	12	0.0101	0.07	0.90	-54.7	117.4	136.4
(4) DM					-60.8	125.6	135.1
(5) DM					-60.8	123.6	128.4
(1) Sterling	∞	0.01	0.1	0.89	-43.7	93.4	107.6
(2) Sterling	12	0.0001	0.07	0.93	-37.3	80.5	94.8
(3) Sterling	8	0.0072	0.06	0.90	-31.8	71.5	90.5
(4) Sterling					-35.3	74.6	84.1
(5) Sterling					-36.7	75.3	80.1

[a](1)–(5) indicate: (1) GARCH (1, 1) with Gaussian errors, (2) IGARCH (1, 1) with student's *t* errors, (3) GARCH (1, 1) with student's *t* errors, (4) generalised local scale model, (5) Gaussian local scale model.

generalized local scale model over the simpler Gaussian local scale model. Interestingly, the Gaussian local scale model outperforms its natural competitor, the fat tailed IGARCH model, on all measures of fit: likelihood, *AIC*, and *BIC*.

As the *BIC* measure is a consistent model selection criteria and *AIC* is well-known for being biased towards overparameterised models, the results presented in table 2 seems to indicate that the tightly parameterised Gaussian local scale model has certain discernible empirical advantages over the four-parameter fat tailed GARCH(1, 1) model.

8. Conditional Gaussian local scale model

A rich class of interesting extensions to the Gaussian local scale model can be generated by allowing the measurement and transition equations to be conditional on Y_{t-1}, so that

$$y_t \,|\, (\theta_t, Y_{t-1}) \sim N(z_t(Y_{t-1}), h_t(Y_{t-1})\theta_t^{-1}), \tag{39a}$$

$$\exp(-r_t(Y_{t-1}))\theta_{t-1}^{-1}\theta_t \,|\, (\theta_{t-1}, Y_{t-1})$$

$$\sim \text{Beta}(\omega a_{t-1}, (1-\omega)a_{t-1}), \tag{39b}$$

$$\theta_0 \,|\, Y_0 \sim G(a_0, b_0). \tag{39c}$$

This model, which is labelled the conditionally Gaussian local scale model, is an analog of the conditionally Gaussian state space models studied in Harvey (1989, ch. 3) and Lipster and Shiryayev (1978). Clearly $\theta_t \mid Y_{t-1}$ and $\theta_t \mid Y_t$ are still gamma variates, but now the updating equations are

$$a_t = a_{t\mid t-1} + \tfrac{1}{2}, \qquad\qquad a_{t\mid t-1} = \omega a_{t-1}, \qquad\qquad (40a)$$

$$b_t = b_{t\mid t-1} + \frac{(y_t - z_t(Y_{t-1}))^2}{2h_t(Y_{t-1})}, \quad b_{t\mid t-1} = \exp(-r_t(Y_{t-1}))b_{t-1}. \quad (40b)$$

The density for $y_t \mid Y_{t-1}$ maintains its generic form, but different $a_{t\mid t-1}$ and $b_{t\mid t-1}$ now feed into it.

A simple example of the use of this setup is the generation of an ARCH-M type model; see Engle, Lilien, and Robins (1987). This model has its location dependent on the expected one-step-ahead conditional volatility and so, in a simple example, it could take on the form

$$z_t(Y_{t-1}) = \alpha + \beta(b_{t\mid t-1}/a_{t\mid t-1})^{1/2}, \qquad\qquad (41a)$$

or perhaps

$$= \alpha' + \beta' \log(b_{t\mid t-1}/a_{t\mid t-1}). \qquad\qquad (41b)$$

A second example, which may prove useful in the modelling of asset pricing, is where we allow $r_t(Y_{t-1})$ to depend on the sign of the previous observations. This type of model is very powerfully motivated by Nelson (1991) when he introduces his exponential GARCH (EGARCH) model. The suggestion we have is that

$$\log \theta_t = \log \theta_{t-1} + \log \eta_t - r_t(Y_{t-1}), \qquad\qquad (42a)$$

where

$$r_t(Y_{t-1}) = \mathrm{E} \log \eta_t + \theta v_{t-1} + \gamma(|v_{t-1}| - \mathrm{E}|v_{t-1}|), \qquad\qquad (42b)$$

with

$$v_t = y_t(a_{t\mid t-1}/b_{t\mid t-1})^{1/2}. \qquad\qquad (42c)$$

For positive values of v_{t-1}, $\partial r_t(Y_{t-1})/\partial v_{t-1} = \theta + \gamma$, while the corresponding term is $\theta - \gamma$ for negative values. More general models for r_t can be produced by the mechanism of eq. (2.3) in Nelson (1991).

9. Conclusion

The Gaussian local scale model provides a parsimonious alternative to integrated GARCH models and provides a rationale for estimating the underlying variance at a given point in time by an EWMA of the squares of the past observations. As such it can be regarded as a natural analog of the basic Gaussian local level time series model. Generalisations to allow for the capture of fat tailed behaviour can be dealt with in this framework.

The local scale model involves an unobservable precision and so fits within the framework of structural time series models. Further, this particular model dovetails into the state space formulation which is so central to the modelling of Gaussian components such as level, slope, seasonal, and cycles.

In the empirical applications of this paper the Gaussian local scale model performs well against its fat tailed GARCH(1, 1) rival. As the Gaussian local scale model is so much easier to fit than GARCH models, this suggests that it might be a model which can be usefully applied when modelling the changing second moment of a time series.

References

Abramowitz, M. and I.A. Stegun, 1970, Handbook of mathematical functions (Dover, New York, NY).

Anderson, B.D.O. and J.B. Moore, 1979, Optimal filtering (Prentice Hall, Englewood Cliffs, NJ).

Baillie, R.T. and T. Bollerslev, 1989, The message in daily exchange rates: A conditional variance tale, Journal of Business and Economic Statistics 7, 297–305.

Baillie, R.T. and R.P. DeGennaro, 1990, Stock returns and volatility, Journal of Financial and Quantitative Analysis 25, 203–214.

Bather, J.A., 1965, Invariant conditional distributions, Annals of Mathematical Statistics 36, 829–846.

Bollerslev, T., 1986, Generalised autoregressive conditional heteroskedasticity, Journal of Econometrics 51, 307–327.

Bollerslev, T., 1987, A conditionally heteroskedastic time series model for speculative prices and the rates of return, Review of Economics and Statistics LXIX, 542–547.

Bollerslev, T. and R.F. Engle, 1993, Common persistence in conditional variance, Econometrica 61, 167–186.

Box, G.E.P. and G.C. Tiao, 1973, Bayesian inference in statistical analysis (Addison-Wesley, reprinted by Wiley, New York, NY, 1992).

Chesney, M. and L.O. Scott, 1989, Pricing European currency options: A comparison of the modified Black–Scholes model and a random variance model, Journal of Financial and Quantitative Analysis 24, 267–284.

de Jong, P., 1989, Smoothing and interpolation with the state space model, Journal of the American Statistical Association 84, 1085–1088.

Engle, R.F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation, Econometrica 50, 987–1007.

Engle, R.F. and T. Bollerslev, 1986, Modelling the persistence of conditional variances, Econometric Reviews 5, 1–50, 81–87.

Engle, R.F., D.M. Lilien, and R.P. Robins, 1987, Estimating time varying risk premia in the term structure: The ARCH-M model, Econometrica 55, 391–407.

Harvey, A.C., 1989, Forecasting, structural time series models and the Kalman filter (Cambridge University Press, Cambridge).

Harvey, A.C. and C. Fernandes, 1989a, Time series models for count or qualitative observations (with discussion), Journal of Business and Economic Statistics 7, 407–422.

Harvey, A.C. and C. Fernandes, 1989b, Time series models for insurance claims, Journal of the Institute of Actuaries 116, 513–528.

Harvey, A.C., E. Ruiz, and N. Shephard, 1992, Multivariate stochastic variance models, Financial Markets Group discussion paper (LSE, London).

Hull, J. and A. White, 1987, Hedging the risks from writing foreign currency options, Journal of International Money and Finance 6, 131–152.

Kitagawa, G., 1987, Non-Gaussian state-space modelling of nonstationary time series, Journal of the American Statistical Association 82, 1032–1041.

Lewis, P.A.W., E. McKenzie, and D.K. Hugus, 1989, Gamma processes, Communications in Statistics–Stochastic Models 5, 1–30.

Lipster, R.S. and A.N. Shiryayev, 1978, Statistics of random processes II: Applications (Springer-Verlag, New York, NY).

Melino, A. and S.M. Turnbull, 1990, Pricing foreign options with stochastic volatility, Journal of Econometrics 45, 239–265.

Muth, J.F., 1960, Optimal properties of exponentially weighted forecasts, Journal of the American Statistical Association 55, 299–305.

Nelson, D.B., 1990a, Stationarity and persistence in the GARCH(1, 1) model, Econometric Theory 6, 318–334.

Nelson, D.B., 1990b, ARCH models as diffusion approximations, Journal of Econometrics 45, 7–39.

Nelson, D.B., 1991, Conditional heteroskedasticity in asset returns: A new approach, Econometrica 59, 347–370.

Nelson, D.B., 1992, Filtering and forecasting with misspecified ARCH models I: Getting the right variance with the wrong model, Journal of Econometrics 52, 61–90.

Scott, L.O., 1987, Option pricing when the variance changes randomly: Theory, estimation and an application, Journal of Financial and Quantitative Analysis 22, 417–438.

Shephard, N., 1990, A local scale model: An unobserved component alternative to integrated GARCH processes, STICERD discussion paper EM/90/220 (LSE, London).

Smith, R.L. and J.E. Miller, 1986, A non-Gaussian state space model and application to prediction of records, Journal of the Royal Statistical Society B 48, 79–88.

Taylor, S.J., 1986, Modelling financial time series (Wiley, Chichester).

Taylor, S.J., 1991, Modelling stochastic volatility, Unpublished paper (University of Lancaster, Lancaster).

Taylor, S.J. and B.G. Kingsman, 1989, An analysis of the variance and distribution of commodity price-changes, Australian Journal of Management 4, 135–149.

Tong, H., 1990, Non-linear time series (Oxford University Press, Oxford).

West, M. and P.J. Harrison, 1989, Bayesian forecasting and dynamic models, Springer series in statistics (Springer-Verlag, Berlin).

Whittle, P., 1991, Likelihood as path integrals (with discussion), Journal of the Royal Statistical Society B 53, 505–538.

Wiggins, J.B., 1987, Option values under stochastic volatility: Theory and empirical estimates, Journal of Financial Economics 19, 351–372.