BAYESIAN ESTIMATION OF AN EXTENDED LOCAL SCALE STOCHASTIC VOLATILITY MODEL

PHILIPPE J. DESCHAMPS

Université de Fribourg

This version: December 2009

ABSTRACT. A new version of the local scale model of Shephard (1994) is presented. Its features are identically distributed evolution equation disturbances, the incorporation of in-the-mean effects, and the incorporation of variance regressors. A Bayesian posterior simulator and an exact simulation smoother are presented. The model is applied to simulated data and to publicly available exchange rate and asset return data. Simulation smoothing turns out to be essential for the accurate interval estimation of volatilities. Bayes factors show that the new model is competitive with GARCH and Lognormal stochastic volatility formulations. Its forecasting performance is comparable to GARCH.

SÉMINAIRE D'ÉCONOMÉTRIE, BOULEVARD DE PÉROLLES 90, CH-1700 FRIBOURG, SWITZERLAND. TELEPHONE: +41-26-300-8252. TELEFAX: +41-26-300-9781.
   E-mail address: philippe.deschamps@unifr.ch

1. Introduction

Since the seminal paper by Engle (1982), models that attempt to explain the conditional heteroskedasticity of asset returns have become essential tools for financial analysts. These models can be subdivided into two broad classes. The first one is the class of ARCH, GARCH, and EGARCH models with their many variants; see, e. g., Engle (1982), Bollerslev (1986), Engle *et al.* (1987), Nelson (1991), and the excellent survey in Bollerslev *et al.* (1994). In this first class, the variance equation is deterministic, and much effort has been devoted to its specification. The second one is the class of stochastic volatility (SV) models, where the variance equation is stochastic and can be considered as the evolution equation in a state space model.

The evolution equation in SV models usually has a rather simple autoregressive form. Nevertheless, it may be argued that SV models are potentially more flexible than models in the GARCH class, since they involve two random shocks rather than one. Furthermore, since an inference on the conditional variances in SV models can be based on the entire sample rather than on past observations only, a sophisticated modeling of the variance equation is perhaps less necessary in SV models than in GARCH models: the evolution equation in SV models can be viewed as a hierarchical Bayesian prior on the volatilities, which is updated by the information present in the entire sample. As a consequence, the smoothed estimates of the latent variables can exhibit systematic nonlinearities that were not implied by the prior specification. The paper by Deschamps (2003) illustrates this fact in a multivariate state space model of consumer demand.

Stochastic volatility models are non-linear or non-Gaussian state space models, and this makes their estimation difficult. It is perhaps for this reason that their use has been less frequent than that of models in the GARCH class. Their treatment seems to follow three main approaches. The first approach is based on a linear evolution equation implying Lognormal volatilities. This formulation appears to be the most common one: see, e. g., Jacquier *et al.* (1994, 2004), Kim *et al.* (1998), Chib *et al.* (2002), and Omori *et al.* (2007). The second one is based on arbitrary differentiable, but Gaussian, observation and/or evolution

equations; an example of this approach is in Stroud *et al.* (2003). The third approach to SV models assumes a multiplicative evolution equation with Beta innovations. Such an equation appears to have been first used by Smith and Miller (1986) with an Exponential observation density, for the time series analysis of extreme values; it was subsequently used by Shephard (1994) in the context of SV models with observation equation disturbances having the Generalized Error Distribution, which includes the Normal as a special case. The model of Shephard (1994) is known as the local scale model. Uhlig (1997) presents a multivariate generalization of the local scale model but his analysis is limited to multivariate Normal observation equation errors. The same limitation is present in the closely related paper by Philipov and Glickman (2006).

The local scale model has two main advantages over the first two approaches. First, the one step ahead prediction densities are known analytically. This means that forecast intervals are easy to obtain, and that the hyperparameter likelihood (marginalized over the volatilities) can be computed exactly from the prediction error decomposition. In contrast, prediction in the SV models that use the first two approaches must be done by particle filtering. This implies choosing suitable model-dependent importance sampling densities, a problem that has only been solved in specific cases (Kim *et al.*, 1998; Chib *et al.*, 2002; Omori *et al.*, 2007). The second advantage involves posterior simulation. In the first two classes of SV models, this requires either single-move sampling, or the use of auxiliary models based on mixtures of Normals. Single-move sampling was found to be inefficient by Omori *et al.* (2007), and using the auxiliary model proposed in their paper causes misspecification that must be corrected by reweighting the posterior replications obtained by Markov chain Monte Carlo (MCMC). Such reweighting is not necessary when the local scale model is estimated, since the posterior density kernel of the hyperparameters is known analytically.

The present paper follows the tradition of the local scale model. It presents a new version of this model, and a new method of estimation. Its main contributions are the following. First, contrary to the original formulation, the disturbances in the evolution equation are identically distributed, with a distribution that depends on two fundamental parameters in a natural way. This new formula-

tion will be called the steady-state version of the local scale model, and has some advantages over the original version. Second, ARCH-M effects are introduced and found to be very significant in an application to asset return data. This may be due, at least in part, to the fact that these effects are formulated in terms of the scale parameter of the one step ahead prediction density, which is available in closed form for the local scale model. Indeed, French *et al.* (1987) show that it is important in this context to distinguish between anticipated and unanticipated components of volatility. Third, variance regressors are introduced in the evolution equation; they allow for irregularly spaced data, and for a form of leverage. Fourth, a fully Bayesian treatment of the estimation problem is presented and illustrated. In particular, the present paper derives an exact simulation smoother for the stochastic volatilities, whereas Shephard (1994) used a quasi-smoother based on a Gaussian local level approximation. The exact simulation smoother is particularly easy to implement, since it only involves generating the inverse volatilities from a linear stochastic difference equation with Gamma innovations. The advantages of using a Bayesian approach in latent variable models have been stated elsewhere in the literature; see, e. g., Harvey *et al.* (2007).

An outline of the paper follows. Section 2 presents and motivates our version of the local scale model. Section 3 presents the simulation smoother; its validity is proved in Appendix A. Section 4 presents a posterior simulator of the hyperparameters, with further details given in Appendix B; a posterior sample of the volatilities can be obtained by applying the simulation smoother to each hyperparameter replication. Section 5 presents an extensive Monte Carlo experiment, where the model of this paper and two competing models are estimated on sixty samples generated under six different assumptions: namely, two versions of the local scale model, two versions of the Lognormal SV model in Omori *et al.* (2007), and two versions of the GARCH model with Student errors (t-GARCH for short). Section 6 applies the local scale model to daily asset return and exchange rate series. Section 7 compares the results of Section 6 with estimates obtained from the model in Omori *et al.* (2007) and from the t-GARCH model. Section 8 presents in-sample and out-of-sample misspecification diagnostics for those models that were found, in Sections 6 and 7, to have the best marginal

likelihoods and information criteria. Section 9 concludes.

2. A STEADY-STATE VERSION OF THE LOCAL SCALE MODEL

2.1 Notation.

In what follows, we will denote a variable X having a standardized Generalized Error Distribution with parameter $r \geq 1$ as $X \sim \text{GED}(r)$. This distribution is discussed in Box and Tiao (1992, pp. 156–160), who call it the Exponential Power Distribution, and was used by Nelson (1991) in an important article. We write the density of X as:

$$f_{GED}(x; r) = \frac{r}{2} \frac{\Gamma(3/r)^{1/2}}{\Gamma(1/r)^{3/2}} \exp\left[-|x|^r \psi(r)\right], \quad \text{with} \quad \psi(r) = \left[\frac{\Gamma(3/r)}{\Gamma(1/r)}\right]^{r/2}.$$

Note that since $\psi(2) = 1/2$, $f_{GED}(x; 2)$ is the standard Normal density. The GED density has heavier tails than the Normal if $r < 2$ and thinner tails if $r > 2$. A Beta variable Y with parameters α and β is noted as $Y \sim \text{Be}(\alpha, \beta)$, and its density as $f_B(y; \alpha, \beta)$. Finally, a Gamma variable Z with density $f_G(z; a, b) \propto z^{a-1} \exp(-bz)$ is noted as $Z \sim \text{Ga}(a, b)$.

2.2 The model.

A general version of the local scale model is, for $t = 1, \ldots, T$:

$$y_t = \mu_t + (\lambda \theta_t)^{-1/r} u_t \tag{2.1}$$

$$\ln \theta_t = \ln \theta_{t-1} + [\ln \eta_t - E(\ln \eta_t)] + \sum_{j=1}^{k} \beta_j D_{jt} \tag{2.2}$$

$$\theta_0 \sim \text{Ga}(a_0, b_0) \tag{2.3}$$

where u_t and η_t are independently distributed with:

$$\eta_t \sim \text{Be}(\omega a_{t-1}, (1 - \omega)a_{t-1})$$

$$u_t \sim \text{GED}(r)$$

and where $0 < \omega < 1$. In the Beta distribution of η_t, a_{t-1} is a parameter of the filter density of θ_{t-1}, which will be stated in Section 2.3. Smith and

4

Miller (1986) show that the marginal distribution of θ_t in (2.2) is Gamma, and Shephard (1994) shows that this Gamma distribution is conjugate with the GED distribution implied by (2.1).

There are two important differences between our version of the local scale model and the original one in Shephard (1994). First, in this paper, a_{t-1} will be a constant that depends on ω and r, so that the evolution equation disturbances η_t are identically distributed. Secondly, the empirical applications presented in Shephard (1994) were limited to $\lambda = 1$, $\mu_t = 0$, and $\beta_j = 0$ in (2.1) and (2.2), whereas we do not impose these restrictions.

In equation (2.1), the presence of λ ensures that the θ_t do not depend on units of measurement, and the choice of the conditional expectation μ_t will be motivated by the results of LeBaron (1992), who finds that the first-order auto-correlation in asset return series is a decreasing function of anticipated volatility; see also Bollerslev *et al.* (1994). We define the anticipated square root volatility as:

$$s_t = (\lambda E[\theta_t \mid y_1, \ldots, y_{t-1}])^{-1/2} \qquad (2.4)$$

and model y_t by an autoregression with coefficients that depend on s_t, so that:

$$\mu_t = \alpha_{00} + \alpha_{10}F(s_t) + \sum_{i=1}^{p}\big[\alpha_{0i} + \alpha_{1i}F(s_t)\big]y_{t-i}, \text{ with:} \qquad (2.5)$$

$$F(s_t) = \frac{1}{1 + \exp(-s_t)} - \frac{1}{2}. \qquad (2.6)$$

It will be seen in Section 2.3 that s_t turns out to be the scale parameter of the one step ahead predictive distribution of y_t. The logistic function $F(s_t)$ lies between 0 and 1/2 and is almost linear for s_t close to 0; its boundedness is important since s_t can take very large values.

Equation (2.2) states that the latent variables θ_t follow a logarithmic random walk, with covariates D_{jt}. The random walk implies important restrictions on these covariates. The term $\sum_{j=1}^{k}\beta_j D_{jt}$ can be interpreted as a (non constant) drift, which should be bounded and should average out to zero for the model to

remain stable. Unfortunately, this excludes an asymmetric news impact function of the EGARCH type. Suitable choices for the variance regressors D_{jt}, which can depend on past observations on y_t, will be discussed in the empirical part of this paper.

It should also be noted that the random walk (2.2) could not be replaced by a more general autoregression without losing the conjugacy of (2.1) and (2.2).

2.3 Filtering and prediction.

We now state the filtering equations associated with (2.1) and (2.2). These equations will enable us to obtain the one step ahead prediction densities, from which the hyperparameter likelihood follows by the prediction error decomposition. Upon defining, for convenience:

$$\phi_t = \exp\left[\sum_{j=1}^{k}\beta_j D_{jt} - E(\ln\eta_t)\right], \tag{2.7}$$

it can be shown that:

$$f(\theta_t \mid y_1, \ldots, y_{t-1}) = f_G(\theta_t; a_{t|t-1}, b_{t|t-1}) \tag{2.8}$$

$$f(\theta_t \mid y_1, \ldots, y_t) = f_G(\theta_t; a_t, b_t) \tag{2.9}$$

with:

$$a_{t|t-1} = \omega a_{t-1} \quad \text{and} \quad b_{t|t-1} = \frac{b_{t-1}}{\phi_t} \tag{2.10}$$

$$a_t = a_{t|t-1} + \frac{1}{r} \quad \text{and} \quad b_t = b_{t|t-1} + \lambda\psi(r)|y_t - \mu_t|^r. \tag{2.11}$$

Finding the one step ahead prediction density is straightforward. The simplest case occurs when the observation equation disturbances in (2.1) are normally distributed, which occurs when $r = 2$. In this case, we have:

$$f(y_t \mid y_1, \ldots, y_{t-1}) = \int_0^{\infty} f(y_t \mid \mu_t, \theta_t, \lambda)f(\theta_t \mid y_1, \ldots, y_{t-1})d\theta_t \tag{2.12}$$

which is a scale mixture of Normals with a Gamma weighting density; from (2.1) and (2.8), this is a Student-t with expectation μ_t, $2a_{t|t-1}$ degrees of freedom, and scale parameter:

$$s_t^2 = b_{t|t-1}/(\lambda a_{t|t-1}) = (\lambda E[\theta_t \mid y_1, \ldots, y_{t-1}])^{-1}; \tag{2.13}$$

6

see, e.g., Bernardo and Smith (2000, p. 123).

As noted by Shephard (1994), when r is arbitrary, the predictive becomes:

$$f(y_t \mid y_1, \ldots, y_{t-1}) = f_{GST}(y_t; \mu_t, s_t^2, a_{t|t-1}, r)$$

$$= \frac{r}{2} \frac{\Gamma\left(a_{t|t-1} + 1/r\right)}{\Gamma\left(a_{t|t-1}\right) \Gamma(1/r)} \left[\frac{\psi(r)}{a_{t|t-1} s_t^2}\right]^{1/r} \times$$

$$\left[1 + \frac{\psi(r)|y_t - \mu_t|^r}{a_{t|t-1} s_t^2}\right]^{-a_{t|t-1} - 1/r} \tag{2.14}$$

where s_t^2 is again given by (2.13).

The density (2.14) is known in the literature as a Generalized Student-t (see McDonald and Newey, 1988) and was used by Bollerslev *et al.* (1994) as error distribution in an EGARCH-type model of asset returns. It is easy to see that (2.14) specializes to the usual Student distribution with $2a_{t|t-1}$ degrees of freedom when $r = 2$.

2.4 Filter initialization.

We now address the choice of a_0 and b_0 in (2.3). Following Harvey (1989), Shephard (1994) proposes to initialize the filter (2.10)–(2.11) with the values $a_0 = b_0 = 0$, corresponding to a diffuse prior on θ_0. However, there are several reasons why $a_0 > 0$ and $b_0 > 0$ might be preferred. First, if $a_0 = b_0 = 0$, the marginal density of y_1 is not defined. Second, we may view (2.2) and (2.3) as defining a joint prior on $(\theta_0, \ldots, \theta_T)$; the marginal prior of θ_t is then $\mathrm{Ga}(\omega^t a_0, b_0 [\prod_{i=1}^t \phi_i]^{-1})$, which becomes improper as a_0 and b_0 tend to 0. This would exclude any prior simulation of the volatilities. Third, it will be seen in Section 3 that the full conditional posterior of θ_0 becomes improper as $b_0 \to 0$. This would exclude any posterior simulation involving θ_0. Last, if $a_0 = 0$, a_t will not reach a steady-state value until some time after the start of the sample; since this start usually depends on data availability rather than on more fundamental considerations, this is difficult to justify.

We will adopt a flexible hierarchical prior on θ_0 based on regularization concepts. Similar treatments of initial conditions have appeared in the literature; see, e. g., Schotman and van Dijk (1991). Upon solving the recurrence equations

7

(2.10) and (2.11), we obtain for $t = 1, \ldots, T$:

$$a_t = \omega^t a_0 + \frac{1}{r} \sum_{i=1}^{t} \omega^{i-1} = \omega^t a_0 + \frac{1 - \omega^t}{r(1 - \omega)} \qquad (2.15)$$

$$b_t = \frac{b_0}{\prod_{i=1}^{t} \phi_i} + \sum_{j=1}^{t} \lambda \psi(r) \gamma_j |y_j - \mu_j|^r \qquad (2.16)$$

with $\gamma_j = \prod_{i=j+1}^{t} \phi_i^{-1}$ for $j < t$ and $\gamma_t = 1$.

Equation (2.15) reveals that an obvious choice for a_0 is the steady-state value:

$$a_0 = \frac{1}{r(1 - w)} \qquad (2.17)$$

since this implies $a_t = a_0 > 0$ for all t.

An appropriate choice for b_0 is less obvious. The solution proposed in this paper treats the first sample observation as one that is preceded by a large number of missing data covering a time interval of length Δt. In this case, Shephard (1994, pp. 189-190) shows the following generalizations of the prediction equation and evolution density:

$$b_{t|t - \Delta t} = b_{t - \Delta t} / \phi_t$$
$$\eta_t \sim \text{Be}(\omega^{\Delta t} a_{t - \Delta t}, (1 - \omega^{\Delta t}) a_{t - \Delta t}).$$

If $\Delta t \to \infty$, η_t converges in probability to 0 since $\omega < 1$. In this case, ϕ_t in (2.7) diverges to $+\infty$, and $b_{t|t - \Delta t} \to 0$. This means that the second part of (2.11) must be modified as follows for $t = 0$:

$$b_0 = b_{0|-\Delta t} + \lambda \psi(r) |y_0 - \mu_0|^r$$
$$\approx \lambda \psi(r) |y_0 - \mu_0|^r.$$

We will treat the pre-sample value y_0 as generated by a steady-state version of the model, obtained from the density (2.14) with $s_0^2 = 1/\lambda$ and $a_{0|-1} = \omega a_0$, and set b_0 equal to the expected value of the previous approximation. The choice

of s_0^2 turns out to be the simplest one yielding a solution that depends only on ω and r, and is therefore motivated by considerations of parsimony. It can be verified using standard analytic integration software that:

$$\int_{-\infty}^{+\infty} |y_t - \mu_t|^r f_{GST}\left(y_t; \mu_t, \frac{1}{\lambda}, \frac{\omega}{r(1-\omega)}, r\right) dy_t = \frac{b_0}{\lambda\psi(r)}$$

with:

$$b_0 = \frac{\omega}{r\omega + r^2(\omega - 1)}. \tag{2.18}$$

The denominator of (2.18) will be positive if $\omega > r/(1+r)$. When the observation equation disturbances are Normal ($r = 2$), this inequality merely constrains the degrees of freedom in the Student-t predictive to be greater than two, a constraint that ensures the existence of the first two moments. Since $E(\theta_0) = a_0/b_0$, (2.17) and (2.18) allow the full range of expectations $0 < E(\theta_0) < +\infty$ and do not, therefore, seriously restrict the flexibility of the model.

Choosing a_0 as in (2.17) and b_0 as in (2.18) leads to the steady-state version of the local scale model, which is written below for the sake of easy reference:

$$y_t = \mu_t + (\lambda\theta_t)^{-1/r} u_t, \quad u_t \sim \text{GED}(r) \quad \text{i.i.d} \tag{2.19}$$

$$\theta_t = \phi_t \theta_{t-1} \eta_t \tag{2.20}$$

$$\phi_t = \exp\left[\sum_{j=1}^{k} \beta_j D_{jt} + \Psi\left(\frac{1}{r(1-\omega)}\right) - \Psi\left(\frac{\omega}{r(1-\omega)}\right)\right] \tag{2.21}$$

$$\eta_t \sim \text{Be}\left(\frac{\omega}{r(1-\omega)}, \frac{1}{r}\right) \quad \text{i.i.d} \tag{2.22}$$

$$\theta_0 \sim \text{Ga}\left(\frac{1}{r(1-\omega)}, \frac{\omega}{r\omega + r^2(\omega - 1)}\right) \tag{2.23}$$

where $\Psi(z) = d\ln\Gamma(z)/dz$ is the digamma function. Equations (2.19) – (2.23) can be viewed as a restatement of (2.1) – (2.3) where the choice of a_0 and b_0 is made endogenous; equation (2.21) follows from (2.7) and from a well-known result on the expected logarithm of a Beta variate.

2.5 Comparing the diffuse and steady-state likelihoods.

This subsection will illustrate some differences between the steady-state version of the local scale model and the Shephard (1994) version, using simulated data. Figure 1 presents, for the three sample sizes of $T = 100$, 500, and 2000, the loglikelihoods of ω obtained with the diffuse filter initialization (dashed lines) and with the steady-state initialization (solid lines). Since the density of y_1 is not defined with the diffuse filter, both loglikelihoods are conditional on the first observation in order to ensure comparability: they are obtained by summing the logarithms of the densities (2.14) for $t = 2, \ldots, T$. The data were simulated from the steady-state version of the model, assuming $\mu_t = 0$, $\beta_j = 0$, $\lambda = 1$, $r = 2$, and $\omega = 0.93$. The DGP values of r and ω are close to maximum likelihood estimates obtained by Shephard (1994). In computing the loglikelihoods, all parameters except ω were set equal to their DGP values.

It is apparent in the top panel of Figure 1, and to a lesser extent in the other panels, that the steady-state likelihood penalizes those values of ω that are close to 2/3 or close to unity. In our model, when ω tends to $r/(1 + r) = 2/3$, $E(\theta_0)$ tends to zero, and when ω tends to one, $E(\theta_0)$ tends to $+\infty$. Since neither assumption is supported by the data, the steady-state likelihood is lower than the diffuse one. When T becomes large, the two likelihoods of course become increasingly similar, but the differences remain apparent in this example.

Since the evolution density becomes singular as ω tends to one, it may be argued that the steady-state likelihood, which bounds ω away from unity, is preferable to the diffuse one.

3. Simulation smoothing

This section addresses the simulation of $(\theta_0, \ldots, \theta_T)$ from their distribution conditional on the observables $(y_1, \ldots, y_T)$ and on all the other unobservables in the model, including ω and r. We denote by $\mathbf{y}_{t_1:t_2}$ the vector $(y_{t_1}, \ldots, y_{t_2})$ and similarly for $\boldsymbol{\theta}_{t_1:t_2}$. The full conditional posterior of $\boldsymbol{\theta}_{0:T}$ can be written as:

$$f(\boldsymbol{\theta}_{0:T} \mid \mathbf{y}_{1:T}) = f(\theta_T \mid \mathbf{y}_{1:T}) \prod_{t=0}^{T-1} f(\theta_t \mid \mathbf{y}_{1:T}, \boldsymbol{\theta}_{t+1:T}). \tag{3.1}$$

A simulation smoother can be based on the decomposition (3.1), using an argument similar to the ones given by Carter and Kohn (1994) and Chib (1996) in different contexts. Each of the last T terms in the product on the right-hand side of (3.1) can be written as:

$$f(\theta_t \mid \mathbf{y}_{1:T}, \boldsymbol{\theta}_{t+1:T}) \propto f(\theta_t \mid \mathbf{y}_{1:t}) f(\theta_{t+1} \mid \theta_t, \mathbf{y}_{1:t}) \tag{3.2}$$

since we have, from Bayes' theorem and from the decomposition of a joint density:

$$f(\theta_t \mid \mathbf{y}_{1:T}, \boldsymbol{\theta}_{t+1:T}) \propto f(\theta_t \mid \mathbf{y}_{1:t}) f(\mathbf{y}_{t+1:T}, \boldsymbol{\theta}_{t+1:T} \mid \mathbf{y}_{1:t}, \theta_t)$$
$$\propto f(\theta_t \mid \mathbf{y}_{1:t}) f(\theta_{t+1} \mid \theta_t, \mathbf{y}_{1:t}) f(\mathbf{y}_{t+1:T}, \boldsymbol{\theta}_{t+2:T} \mid \mathbf{y}_{1:t}, \theta_{t+1}, \theta_t) \tag{3.3}$$

and since the last term in (3.3) does not depend on θ_t. The first term on the right-hand side of (3.2) is of course the filter density of θ_t, given by (2.9); the second term can be viewed as the likelihood of θ_t implied by equation (2.20). It is shown in Appendix A that applying Bayes' theorem to the right-hand side of (3.2) yields the following translated Gamma posterior:

$$f(\theta_t \mid \mathbf{y}_{1:t}, \theta_{t+1}) = f_G\left(\theta_t - \phi_{t+1}^{-1}\theta_{t+1}; \frac{1}{r}, b_t\right).$$

In order to simulate $\boldsymbol{\theta}_{0:T}$ from its full conditional posterior distribution, we may then use the following forward-filtering backward-sampling algorithm based on (3.1) and (3.2):

(1) Compute and store b_t from the filter (2.10)–(2.11) for $t = 0, \ldots, T$.
(2) Sample:

$$\theta_T \sim \text{Ga}\left(\frac{1}{r(1-\omega)}, b_T\right). \tag{3.4}$$

(3) Generate $\theta_{T-1}, \ldots, \theta_0$ from the stochastic difference equation:

$$\theta_t = \phi_{t+1}^{-1}\theta_{t+1} + \epsilon_t \tag{3.5}$$

where the ϵ_t are independent draws from $\text{Ga}(1/r, b_t)$ distributions.

11

The posterior expectations $\bar{\theta}_t$ of the θ_t can be obtained by replacing, in (3.4) and (3.5), the random variables by their expected values. This yields the recursion:

$$\bar{\theta}_T = \frac{1}{rb_T(1-\omega)} \tag{3.6}$$

$$\bar{\theta}_t = \phi_{t+1}^{-1}\bar{\theta}_{t+1} + \frac{1}{rb_t} \quad \text{for } t = T-1, \ldots, 0. \tag{3.7}$$

4. POSTERIOR SIMULATION

Let, for convenience:

$$\boldsymbol{\alpha}' = \begin{pmatrix} \alpha_{00} & \alpha_{10} & \alpha_{01} & \alpha_{11} & \ldots & \alpha_{0p} & \alpha_{1p} \end{pmatrix}$$

and:

$$\boldsymbol{\beta}' = \begin{pmatrix} \beta_1 & \ldots & \beta_k \end{pmatrix}.$$

We will adopt the following prior:

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda, \omega, r) \propto$$
$$f_N(\boldsymbol{\alpha}; \mathbf{m}_\alpha, \Sigma_\alpha) f_N(\boldsymbol{\beta}; \mathbf{m}_\beta, \Sigma_\beta) f_G(\lambda; a_\lambda, b_\lambda) \mathcal{I}_{(r_{\min}, r_{\max})}(r) \mathcal{I}_{(\omega_{\min}, 1)}(\omega) \tag{4.1}$$

where f_N denotes the Multinormal density and $\mathcal{I}$ denotes an indicator function. From the definition of a GED density, we must have $r_{\min} \geq 1$; and since typical asset return distributions exhibit heavy tails, it is sensible to choose $r_{\max}$ not much greater than 2. For (2.18) to remain well-defined, we must have:

$$\omega_{\min} > \frac{r_{\max}}{1 + r_{\max}}$$

which is an innocuous constraint in view of the remark made after the statement of (2.18).

Our posterior sampler for the extended local scale model can be summarized as follows. Define $\boldsymbol{\xi} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda, \omega, r)$. First sample $\boldsymbol{\xi}^{(0)}$ in the support of its prior distribution. A Markov chain $\{\boldsymbol{\xi}^{(i)}\}_{i=1}^N$ is generated as follows:

(1) Sample $r^{(i)}$ conditional on the data, $\boldsymbol{\alpha}^{(i-1)}$, $\boldsymbol{\beta}^{(i-1)}$, $\omega^{(i-1)}$, and $\lambda^{(i-1)}$.

(2) Sample $\omega^{(i)}$ conditional on the data, $\boldsymbol{\alpha}^{(i-1)}$, $\boldsymbol{\beta}^{(i-1)}$, $r^{(i)}$, and $\lambda^{(i-1)}$.

(3) Sample $\boldsymbol{\alpha}^{(i)}$ conditional on the data, $\boldsymbol{\beta}^{(i-1)}$, $\lambda^{(i-1)}$, $r^{(i)}$, and $\omega^{(i)}$.

(4) Sample $\boldsymbol{\beta}^{(i)}$ conditional on the data, $\boldsymbol{\alpha}^{(i)}$, $\lambda^{(i-1)}$, $r^{(i)}$, and $\omega^{(i)}$.

(5) Sample $\lambda^{(i)}$ conditional on the data, $\boldsymbol{\alpha}^{(i)}$, $\boldsymbol{\beta}^{(i)}$, $r^{(i)}$, and $\omega^{(i)}$.

Note that this algorithm does not depend on the simulation smoother of Section 3, but uses the likelihood implied by the prediction error decomposition. This use is partly due to the fact that the joint likelihood of $\boldsymbol{\xi}$ and $\boldsymbol{\theta}_{0:T}$ is only defined for those values of the volatilities which are consistent with (2.20): it would not be practical to maintain this consistency across random draws of the remaining parameters. However, the deterministic smoother (3.6)–(3.7) will be used in the construction of candidate proposal densities, and this turns out to be important for a good performance of the algorithm.[1]

Steps (1)–(5) are "Metropolis within Gibbs" steps having the following generic expression. Let $\boldsymbol{\vartheta}$ be the subvector of $\boldsymbol{\xi}$ being simulated, and let $\boldsymbol{\varphi}$ be that subvector of $\boldsymbol{\xi}$ which does not contain $\boldsymbol{\vartheta}$. Let $k(\boldsymbol{\vartheta} \mid \boldsymbol{\varphi})$ be the kernel of the conditional posterior of $\boldsymbol{\vartheta}$. Let $\boldsymbol{\vartheta}_{\text{old}}$ be the previous draw of $\boldsymbol{\vartheta}$, and let $q(\boldsymbol{\vartheta} \mid \boldsymbol{\vartheta}_{\text{old}}, \boldsymbol{\varphi})$ be a normalized proposal density. The proposal densities used in this paper are described in Appendix B. One draws a candidate $\boldsymbol{\vartheta}$ from $q(\boldsymbol{\vartheta} \mid \boldsymbol{\vartheta}_{\text{old}}, \boldsymbol{\varphi})$, and sets $\boldsymbol{\vartheta}^{(i)} = \boldsymbol{\vartheta}$ with probability:

$$\min\left[1, \frac{k(\boldsymbol{\vartheta} \mid \boldsymbol{\varphi})}{k(\boldsymbol{\vartheta}_{\text{old}} \mid \boldsymbol{\varphi})} \frac{q(\boldsymbol{\vartheta}_{\text{old}}, \mid \boldsymbol{\vartheta}, \boldsymbol{\varphi})}{q(\boldsymbol{\vartheta} \mid \boldsymbol{\vartheta}_{\text{old}}, \boldsymbol{\varphi})}\right]$$

and $\boldsymbol{\vartheta}^{(i)} = \boldsymbol{\vartheta}_{\text{old}}$ otherwise.

In order to compute the posterior kernel $k(\boldsymbol{\vartheta} \mid \boldsymbol{\varphi})$ above, one multiplies by the prior density (4.1) the likelihood implied by (2.14):

$$\mathcal{L}(\boldsymbol{\xi}) \equiv \mathcal{L}(\boldsymbol{\vartheta}; \boldsymbol{\varphi}) = \prod_{t=1}^{T} f_{GST}\left[y_t; \mu_t(\boldsymbol{\xi}), s_t^2(\boldsymbol{\xi}), \frac{\omega}{r(1-\omega)}, r\right] \tag{4.2}$$

where the $\mu_t(\boldsymbol{\xi})$ and $s_t^2(\boldsymbol{\xi})$ are obtained by the filter of Section 2.3. For convenience, we restate this filter by incorporating into (2.10)–(2.11) the conditions

---

[1]An alternative way of generating candidates would be to use random draws of the volatilities, generated at the beginning of each pass. This did not result in an improved sampler.

(2.5), (2.13), (2.17), (2.18), and (2.21). We assume that p initial observations $\mathbf{y}_{1-p:0}$ are available. For given $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, λ, r, and ω, the sequences $\mathbf{b}_{0:T}$, $\boldsymbol{\phi}_{1:T}$, $\boldsymbol{\mu}_{1:T}$ and $\mathbf{s}_{1:T}$ can be generated by the following recursion:

$$\mu_t = \alpha_{00} + \alpha_{10}F(s_t) + \sum_{i=1}^{p}\left[\alpha_{0i} + \alpha_{1i}F(s_t)\right]y_{t-i} \tag{4.3}$$

$$b_t = \frac{b_{t-1}}{\phi_t} + \lambda\psi(r)|y_t - \mu_t|^r \tag{4.4}$$

$$\ln\phi_{t+1} = \sum_{i=1}^{k}\beta_i D_{i,t+1} + \Psi\left(\frac{1}{r(1-\omega)}\right) - \Psi\left(\frac{\omega}{r(1-\omega)}\right) \tag{4.5}$$

$$s_{t+1}^2 = \frac{b_t}{\lambda\phi_{t+1}}\frac{r(1-\omega)}{\omega} \tag{4.6}$$

with the initial conditions:

$$b_0 = \frac{\omega}{r\omega + r^2(\omega-1)} \tag{4.7}$$

$$\ln\phi_1 = \Psi\left(\frac{1}{r(1-\omega)}\right) - \Psi\left(\frac{\omega}{r(1-\omega)}\right) \tag{4.8}$$

$$s_1^2 = \frac{1-\omega}{\lambda\phi_1[\omega + r(\omega-1)]}. \tag{4.9}$$

Using (2.14), it is easy to show that the likelihood (4.2) is equal to:

$$\left[\frac{\Gamma\left(\frac{1}{r(1-\omega)}\right)[r^{r+1}\psi(r)(1-\omega)]^{1/r}}{2\,\Gamma\left(\frac{\omega}{r(1-\omega)}\right)\Gamma\left(\frac{1}{r}\right)\omega^{1/r}}\right]^T \times$$

$$\prod_{t=1}^{T}s_t^2(\boldsymbol{\xi})^{-1/r}\left[1 + \frac{r\psi(r)(1-\omega)}{\omega s_t^2(\boldsymbol{\xi})}|y_t - \mu_t(\boldsymbol{\xi})|^r\right]^{\frac{1}{r(\omega-1)}}. \tag{4.10}$$

5. Monte Carlo experiment

In this section, we will investigate the ability of the extended local scale model to provide accurate point and interval estimates of the volatilities generated by six data generating processes (DGPs).

DGP1 and DGP2 are versions of the Lognormal SV model in Omori *et al.* (2007), with and without leverage. This model reads as:

Model 1.

$$y_t = \epsilon_t \exp(h_t/2) \tag{5.1}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t \tag{5.2}$$

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix} \right] \quad \text{i.i.d} \tag{5.3}$$

with $\rho = 0$ for DGP2.

DGP3 and DGP4 are AR(1) and AR(0) t-GARCH models with an asymmetric variance equation:

Model 2.

$$y_t = \gamma_0 + \gamma_1 y_{t-1} + u_t \tag{5.4}$$

$$u_t = \sigma_t \sqrt{\frac{\nu - 2}{\nu}} \epsilon_t, \quad \epsilon_t \sim \text{Student}(\nu) \quad \text{i.i.d} \tag{5.5}$$

$$\sigma_t^2 = \alpha_0 + \left[\alpha_1 \mathcal{I}_{[0,+\infty)}(u_{t-1}) + \alpha_2 \mathcal{I}_{(-\infty,0)}(u_{t-1}) \right] u_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{5.6}$$

with $\gamma_0 = \gamma_1 = 0$ for DGP4.

DGP5 and DGP6 are extended and basic versions of the local scale model:

Model 3.

$$y_t = \alpha_{00} + \alpha_{10} F(s_t) + [\alpha_{01} + \alpha_{11} F(s_t)] y_{t-1} + (\lambda \theta_t)^{-1/r} u_t \tag{5.7}$$

$$s_t = (\lambda E[\theta_t \mid y_1, \ldots, y_{t-1}])^{-1/2} \tag{5.8}$$

$$\ln \theta_t = \ln \theta_{t-1} + [\ln \eta_t - E(\ln \eta_t)] + \beta_1 F(y_{t-1}) \tag{5.9}$$

$$u_t \sim \text{GED}(r) \quad \text{i.i.d} \tag{5.10}$$

$$\eta_t \sim \text{Be}\left(\frac{\omega}{r(1-\omega)}, \frac{1}{r} \right) \quad \text{i.i.d} \tag{5.11}$$

$$\theta_0 \sim \text{Ga}\left(\frac{1}{r(1-\omega)}, \frac{\omega}{r\omega + r^2(\omega - 1)} \right) \tag{5.12}$$

15

where:

$$F(x) = \frac{1}{1 + \exp(-x)} - \frac{1}{2} \tag{5.13}$$

and where all parameters are nonzero in DGP5, whereas DGP6 sets $\alpha_{00} = \alpha_{10} = \alpha_{01} = \alpha_{11} = \beta_1 = 0$. So, DGP6 is the local scale model with GED errors, but without ARCH-M effects and variance regressors. It will be seen in Section 6 that the logistic transform $F(y_{t-1})$ in equation (5.9) turns out to be a good approximation of y_{t-1}, while ensuring model stability.

All DGP parameter values can be found in Table 1. As will be seen in Sections 6 and 7, the values for DGPs 1, 3, and 5 are typical of asset return series and those for DGPs 2, 4, and 6 are typical of exchange rate series.

Ten samples of 2000 observations each were generated under each DGP, for a total of 60 samples. Table 2 presents the ranges of the unconditional moments, of the first autocorrelations, and of the first autocorrelations of the squares of the observations generated by all six DGPs. Unrestricted versions of Models 1, 2, and 3 were then estimated on all 60 samples, and posterior samples of the volatilities were generated. The volatilities were defined as $\exp(h_t/2)$ for Model 1, σ_t for Model 2, and $(\lambda\theta_t)^{-1/r}$ for Model 3; they were estimated by simulation smoothing for Models 1 and 3. The MCMC algorithm used in GARCH estimation is fully described in Ardia (2008, pp. 59–64); the algorithm used in estimating the Lognormal SV model is described in Omori *et al.* (2007). The validity of the computer codes was tested by the joint distribution method of Geweke (2004). In all cases, convergence was tested by the method of Gelman and Rubin (1992). The priors will be described in Sections 6 and 7.

Table 3 presents, for each model and DGP combination, the mean absolute errors (MAE) of the volatility point estimates (posterior means), the root mean squared errors (RMSE) of these estimates, and the outlier percentages (proportions of true volatilities outside of 95% confidence intervals). It also presents t-statistics on the deviations between these percentages and the 5% nominal value, computed using heteroskedasticity and autocorrelation consistent standard errors. All the figures in Table 3 are based on the 20000 volatilities estimated for each DGP/Model combination.

16

We first discuss the MAE and RMSE in Table 3. The point estimates in both SV models outperform GARCH on DGP1, DGP2, DGP5, and DGP6. It is noteworthy that the MAE and RMSE of both SV models are quite similar; actually, in the case of DGP3 and DGP6, the differences between the MAE for Models 1 and 3 are statistically insignificant.

A striking fact in Table 3 is the consistent failure of the GARCH model to provide accurate interval volatility estimates, as can be seen from the high values of the coverage t-statistics for this model. Indeed, from a coverage standpoint, both the Lognormal and the local scale stochastic volatility model actually outperform GARCH on DGP4, where 13% of the true volatilities go beyond the 95% confidence bounds estimated by Model 2, even though the samples were generated by a GARCH process. In fact, the local scale model passes the coverage tests for both GARCH DGPs. In contrast, the Lognormal SV model fails these two tests, but correctly identifies the volatility outliers for DGP6 (the basic local scale model). This occurs in spite of the fact that the Lognormal SV model in this paper has a Gaussian observation equation, whereas the corresponding distribution in DGP6 is GED.

The failure of the GARCH model to provide adequate volatility coverage appears to be due to its deterministic formulation of the variance equation. This is illustrated in Figure 2, where the observations in the first sample generated with DGP1 are plotted, together with the true volatilities and 95% confidence bands generated by the three models. For Models 1 and 3, the widths of these bands reflects two types of uncertainty: parameter uncertainty, and the uncertainty caused by the stochastic nature of the variance process. Since the confidence bands in the two right-hand panels of Figure 2 were estimated by simulation smoothing, their widths reflect both types of uncertainty. In contrast, confidence bands on the GARCH variance function only reflect parameter uncertainty; this results in narrower bands, which prove inadequate to accurately identify the volatility outliers.

17

6. Estimating the local scale model on real data

6.1 The data.

In this section, we will estimate versions of the local scale model on one asset return and two exchange rate series. The first series is constructed from the Standard and Poor 500 (S&P500) daily asset price data, available on finance.yahoo.com, and ranges from January 6, 1970 to April 17, 2009 (9916 observations). The second and third series range from January 5, 1982 to June 19, 2009 (6905 observations) and are constructed from the Swiss Franc/US dollar and US dollar/Pound Sterling exchange rates, available on www.federalreserve.gov. All three series are defined as $y_t = 100 \ln(P_t/P_{t-1})$, where P_t is either the closing price index (in the case of the S&P500 data) or the exchange rate (in the two other cases). The three series are irregularly dated; whereas the missing data in the S&P500 series only cover Saturdays, Sundays, holidays, and the September 2001 event, the two exchange rate series include occasional missing observations that are not due to market closure.

Two predetermined variables D_{1t} and D_{2t} will be considered for use in the variance equation (2.2). D_{1t} is the logistic transform $F(y_{t-1})$, where $F(x)$ is given by (5.13). Since $D_{1t} < 0$ corresponds to a negative past asset return, one would expect its coefficient β_1 to be positive in the S&P500 case (bad news increase volatility), and this can be interpreted as a form of leverage. D_{2t} is defined as the first difference of the number of consecutive non-trading days prior to date t. Its presence allows for effects due to market closure. In view of the remark made at the end of the previous paragraph, it will be included in some local scale asset return models but omitted from the exchange rate models.

Figure 3 presents the histograms of both variance regressors for the asset return data, as well as a scatter plot of y_t on its logistic transform. The distribution of both variance regressors is clearly centered on zero; as mentioned at the end of Section 2.2, this ensures model stability. The bottom panel of Figure 3 shows that $F(y_t)$ is approximately linear for values of y_t close to zero, but does not exhibit the extreme outliers present in the asset return series.

6.2 Specification search.

Estimating the local scale model involves deciding whether to include ARCH-M effects and variance regressors, and whether to choose a GED rather than a Normal observation density. In addition, the autoregressive order p must be specified. We will choose the models with the highest information criteria. The first benchmark will be the Schwarz information criterion:

$$\text{BIC}(\hat{\boldsymbol{\xi}}) = 2\ln[\mathcal{L}(\hat{\boldsymbol{\xi}})] - q\ln(T) \tag{6.1}$$

where $\hat{\boldsymbol{\xi}}$ is the posterior mean of the parameter vector defined in Section 4, q is the number of parameters in $\boldsymbol{\xi}$, and $\mathcal{L}(\boldsymbol{\xi})$ is given by (4.2).

The second criterion is the marginal likelihood:

$$p(\mathbf{y}) = \int f(\mathbf{y} \mid \boldsymbol{\xi})p(\boldsymbol{\xi})d\boldsymbol{\xi} \tag{6.2}$$

where $\mathbf{y} = \mathbf{y}_{1:T}$, $f(\mathbf{y} \mid \boldsymbol{\xi})$ is the right-hand side of (4.2), and $p(\boldsymbol{\xi})$ is the prior (4.1).

The two criteria (6.1) and (6.2) are complementary: (6.1) has the advantage of being insensitive to the prior, but the magnitude of its difference across models is difficult to interpret. The marginal likelihood (6.2) does depend on the prior, even asymptotically, but the Bayes factors that can be derived from (6.2) have a clear probabilistic interpretation.

A number of methods are available for estimating (6.2); see, e. g., Gelfand and Dey (1994), Meng and Wong (1996) and Chib and Jeliazkov (2001). The author chose the bridge sampling method of Meng and Wong (1996) for its ease of implementation and numerical efficiency. In the present context, the bridge sampling identity reads as:

$$p(\mathbf{y}) = \frac{\int \big[p(\boldsymbol{\xi})f(\mathbf{y} \mid \boldsymbol{\xi})\alpha(\boldsymbol{\xi})\big]q(\boldsymbol{\xi})d\boldsymbol{\xi}}{\int \big[q(\boldsymbol{\xi})\alpha(\boldsymbol{\xi})\big]p(\boldsymbol{\xi} \mid \mathbf{y})d\boldsymbol{\xi}} \tag{6.3}$$

where $q(\boldsymbol{\xi})$ is a normalized importance sampling density and $\alpha(\boldsymbol{\xi})$ is a "bridge function" to be defined shortly. So, the numerator in (6.3) can be estimated by an average of n replications of $p(\boldsymbol{\xi})f(\mathbf{y} \mid \boldsymbol{\xi})\alpha(\boldsymbol{\xi})$, where $\boldsymbol{\xi}$ is drawn from $q(\boldsymbol{\xi})$, and

the denominator by an average of m replications of $q(\boldsymbol{\xi})\alpha(\boldsymbol{\xi})$, where $\boldsymbol{\xi}$ is drawn from the posterior. The bridge function is obtained by an iterative procedure, as:

$$\alpha(\boldsymbol{\xi}) = \frac{1}{\left[nq(\boldsymbol{\xi}) + m\dfrac{p(\boldsymbol{\xi})f(\mathbf{y} \mid \boldsymbol{\xi})}{p(\mathbf{y})}\right]}. \tag{6.4}$$

Frühwirth-Schnatter (2004) provides many useful implementation details, and uses theoretical arguments to show that this method is an improvement over earlier importance sampling methods, such as that of Gelfand and Dey (1994). Bridge sampling has been used successfully with several other models (Frühwirth-Schnatter, 2004; Deschamps, 2008; Ardia, 2008, 2009).

The importance sampling density was chosen as:

$$q(\boldsymbol{\xi}) = q_1(r)q_2(\omega)q_3(\boldsymbol{\alpha})q_4(\boldsymbol{\beta})q_5(\lambda)$$

where q_1 and q_2 are the densities of linear functions of Beta variates with ranges equal to the prior supports of r and ω; q_3 and q_4 are Normal; and q_5 is Lognormal. The first two moments of the q_i were chosen to match the empirical posterior moments obtained by MCMC.

The priors on the elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ were independent $N(0, 10)$; the prior on λ was $Ga(10^{-6}, 10^{-6})$; and that for ω was Uniform $U(0.8, 1)$. The prior on r was $U(1, 2)$ for the S&P500 data and $U(1, 2.5)$ for the exchange rate series (in each case, these choices ensured prior supports much larger than the range of the posterior replications).

Tables 4 and 5 present the decimal logarithms $\log_{10}(BF)$ of the estimated Bayes factors against the preferred models, the numerical standard errors (NSE) of these logarithms, and the differences $\Delta(BIC)$ between Schwarz information criteria. The Bayes factor estimates are indeed very precise, confirming that our choice of the method and importance sampling densities is appropriate. For the S&P500 data, both criteria favor an AR(1) model with ARCH-M effects, two variance regressors, and GED errors ($\beta_1 \neq 0, \beta_2 \neq 0, p = 1, r \neq 2$). The Bayes factor evidence against all the other models in Table 4 is decisive (Jeffreys, 1961, Appendix B). For both exchange rate series, ARCH-M effects are irrelevant and

20

serial correlation is negligible, so that $\mu_t = 0$ was imposed, and it can be seen in Table 5 that models with GED disturbances but without the variance regressor D_{1t} are preferred. The evidence against the inclusion of D_{1t} is strong in the Swiss Franc case, and very strong in the Sterling case. The evidence in favor of GED disturbances is decisive in both cases.

6.3 Estimates.

Tables 6, 7, and 8 present posterior replication summaries for the specifications having the largest posterior probabilities. These results are based on 10000 replications obtained by running the posterior simulator twice for 30000 passes, of which the first 5000 were discarded. Convergence was checked by heteroskedasticity and autocorrelation consistent Wald equality tests on the expected values of the two chains, and by the method of Gelman and Rubin (1992). The final posterior sample was then obtained by combining the two chains and selecting every fifth replication. The autocorrelations in the final posterior sample decay quickly, as indicated by ρ_1 and ρ_5 in Tables 6 to 8. The priors are the same as in Section 6.2.

For the S&P500 data (Table 6), the credible set for α_{01} implies the presence of autocorrelation, and that for α_{11} only contains negative values, confirming the findings of LeBaron (1992): autocorrelation is a decreasing function of anticipated volatility. The credible set for β_1 confirms the intuition in Section 6.1 of a positive sign. The credible set for β_2 confirms the stylized fact that market closure causes a subsequent increase in volatility (recall the definition of D_{2t} in Section 6.1, and that θ_t is an inverse volatility). The GED parameter r is clearly less than 2, confirming the leptokurticity of the observation distribution. This is also the case for the exchange rate series (Tables 7 and 8). The parameter ω is particularly well identified in all cases.

The burn-in period of 5000 passes (which was also used in the other simulations of Sections 5 and 6.2) appears to be very conservative. Indeed, Figure 4 presents the sample paths of 1000 replications for the local scale model used on the S&P500 data, obtained without discarding intermediate and burn-in passes; it suggests convergence after a few hundred sweeps of the Metropolis-Hastings algorithm. The 1000 replications took about 5 minutes of processor time on a 3.2 Ghz

workstation, using compiled code.

The top panel of Figure 5 is a time series of point estimates (posterior means) of the anticipated volatilities $s_t = s_t(\boldsymbol{\xi})$ for the S&P500 data, obtained by the filter (4.3)–(4.9) from a posterior sample. The middle panel is a time series of point estimates of the conditional expectations:

$$E(y_t \mid s_t) = \frac{\alpha_{00} + \alpha_{10} F(s_t)}{1 - \alpha_{01} - \alpha_{11} F(s_t)} \tag{6.5}$$

where $F(s_t)$ is the logistic function (2.6). The bottom panel is a scatter plot of s_t against $E(y_t \mid s_t)$. These graphs clearly indicate that a significant risk premium is present.

Figure 6 presents, for all three data sets, line graphs of the observations, together with median volatilities and 95% confidence bands, obtained by the simulation smoother of Section 3 from a posterior sample of 1000 replications of $\boldsymbol{\xi}$. The volatility graphs closely reflect intuition; however, for the S&P500 data, it is interesting to note that the peak volatility estimate occurs during the recent financial crisis rather than during the 1987 crash.

7. COMPARING THE LOCAL SCALE AND OTHER MODELS

In this section, we will compare the local scale models yielding the best information criteria and marginal likelihoods in Tables 4 and 5 with the competitors defined in Section 5 (Models 1 and 2 in that section).

The prior used in the Lognormal SV model was the same as in Omori *et al.* (2007). For the GARCH model, the author used:

$$p(\gamma_i) = f_N(\gamma_i; 0, 10)$$
$$p(\alpha_0) \propto f_N(\alpha_0; 0, 10) \, \mathcal{I}_{(0,\infty)}(\alpha_0)$$
$$p(\alpha_i) \propto f_N(\alpha_i; 0.5, 10) \, \mathcal{I}_{(0,\infty)}(\alpha_i) \quad (i = 1, 2)$$
$$p(\beta) \propto f_N(\beta; 0.5, 10) \, \mathcal{I}_{(0,1)}(\beta)$$
$$\nu = 2 + \nu^*, \quad \nu^* \sim \text{Exponential}(0.1).$$

Table 9 presents logarithmic Bayes factors and differences of information criteria against the extended local scale model ($p = 1$, $r \neq 2$, $\beta_1 \neq 0$, $\beta_2 \neq 0$) in

22

the asset return case, and against the basic local scale model ($\mu_t = 0$, $r \neq 2$, $\beta_1 = \beta_2 = 0$) for the exchange rate series. The GARCH marginal likelihoods were estimated by bridge sampling, and those for the Lognormal SV model were estimated by the method of Chib (1995), as implemented in Chib and Greenberg (1998). The evidence in favor of leverage in the Lognormal SV model ($\rho \neq 0$) can be obtained as a difference of the relevant logarithmic Bayes factors. It is not surprising that this evidence is decisive for asset returns, whereas the absence of leverage ($\rho = 0$) is favored for the exchange rates (though not very strongly). For the GARCH models, the evidence decisively favors AR(1) ($\gamma_1 \neq 0$, $\gamma_2 \neq 0$) for the S&P500 series, and AR(0) ($\gamma_1 = \gamma_2 = 0$) for the exchange rates.

The local scale models chosen in Section 6 are preferred to all Lognormal SV and GARCH formulations according to the marginal likelihood criterion (6.2), and the Schwarz criterion (6.1) agrees in all cases except one. It should be noted, however, that the evidence against the preferred GARCH models is very weak in all instances, the odds against GARCH and in favor of local scale never exceeding three to one.

Comparing Tables 4 and 9 reveals that for the S&P500 data, the odds against Lognormal SV are due to the inclusion of D_{1t} and to $\mu_t \neq 0$ in the local scale model, rather than to GED errors. Comparing Tables 5 and 9 reveals that for the exchange rate series, the superiority of the local scale to the Lognormal SV model is due to the GED observation density.

The MCMC estimates of the GARCH and Lognormal SV models are in Tables 10 and 11. Table 10 shows that the variance equations in the Lognormal SV models are close to logarithmic random walks. Table 11 suggests that some asymmetry is present in the GARCH variance equations for the exchange rate series, even though this can obviously not be ascribed to leverage; as expected, strong asymmetry is present in the S&P500 case.

Figure 7 presents the differences in the volatility point estimates (posterior means, estimated by simulation smoothing in the SV cases) obtained with the local scale and competing models. As expected in the light of the Monte Carlo results of Section 5, these differences have a much higher amplitude when the local scale model is compared to GARCH. Another noticeable fact is that they

23

are more pronounced in periods of high volatility (compare with Figure 6).

8. MISSPECIFICATION DIAGNOSTICS

8.1. Generalized residuals and density forecasting.

Given $\boldsymbol{\xi} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda, \omega, r)$, generalized residuals can be defined as:

$$v_t(\mathbf{y}, \boldsymbol{\xi}) = \Phi^{-1}\left(P[Y_t \leq y_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\xi}]\right) \tag{8.1}$$

where $\Phi(.)$ is the standard Normal cumulative distribution function (cdf). If the predictive probabilities in (8.1) are indeed those implied by the process generating $\mathbf{y}$, the generalized residuals are independent standard Normal; for a proof, see, e. g., Diebold *et al.* (1998). These predictive probabilities are obtained by integrating (2.14). It can be checked that:

$$F_{GST}(y_t; \mu_t, s_t^2, a_{t|t-1}, r) = 0.5 \left[1 - F_B\left(\tau_t(y_t); \frac{1}{r}, a_{t|t-1}\right)\right] \quad \text{if } y_t < \mu_t$$

$$= 0.5 \left[1 + F_B\left(\tau_t(y_t); \frac{1}{r}, a_{t|t-1}\right)\right] \quad \text{if } y_t \geq \mu_t \tag{8.2}$$

where:

$$\tau_t(y_t) = \frac{\dfrac{\psi(r)}{s_t^2}|y_t - \mu_t|^r}{\dfrac{\psi(r)}{s_t^2}|y_t - \mu_t|^r + a_{t|t-1}} \tag{8.3}$$

and where $F_B(\tau; a, b)$ is the cdf of a Beta variate with parameters a and b. With the steady-state filter initialization defined in Section 2, we have $a_{t|t-1} = \omega/(r - r\omega)$.

Inverting the cdf (8.2) yields the following useful algorithm for generating simulated observations from the conditional predictive (2.14):

(1) Draw $x \sim \text{Be}(1/r, a_{t|t-1})$.
(2) Compute:
$$z = \left(a_{t|t-1}\frac{x}{1-x}\right)^{1/r}.$$
(3) Draw $u \sim U(0, 1)$.

24

(4) If $u \geq 0.5$ let $z^* = z$; otherwise let $z^* = -z$.

(5) Return:

$$y_t = \mu_t + \left[\frac{s_t^2}{\psi(r)} \right]^{1/r} z^*.$$

It is of course easy to compute generalized residuals analogous to (8.1) for the GARCH model of Section 5, and to generate simulated observations from the one step ahead GARCH predictive. For the Lognormal SV model, the counterpart of (8.2) must be estimated by particle filtering, and it is impractical to do so for all the parameter replications in a posterior sample. For this reason, simulating from the Lognormal SV marginal predictive is not straightforward.

8.2 In-sample diagnostics.

This subsection will present posterior predictive p-values of diagnostics computed from the residuals v_t defined in Section 8.1. These diagnostics are:

(1) The F-statistic for testing the nullity of the autoregression coefficients in an AR(6) model of the v_t. This is indicated in Table 12 by AC.

(2) The F-statistic for testing the nullity of the autoregression coefficients in an AR(6) model of the v_t^2. This is indicated by ARCH.

(3) The likelihood ratio statistic for testing $E(v_t) = 0$ and $E(v_t^2) = 1$, assuming Normality. This is indicated by LR.

(4) The Bera-Jarque statistic (indicated by BJ).

(5) The Kolmogorov-Smirnov statistic corresponding to the null hypothesis that $v_t \sim N(0, 1)$, indicated by KS.

(6) The sample skewness of the v_t, indicated by M3.

(7) The sample excess kurtosis of the v_t, indicated by M4.

Analyzing predictive p-values based on these diagnostics (apart from KS) can be viewed as a Bayesian counterpart of conditional moment testing, a procedure that is commonly used for evaluating models in the GARCH class; see, e. g., Nelson (1991) and Bollerslev *et al.* (1994).

The notion of predictive p-value is discussed in Gelman and Meng (1996), Carlin and Louis (2000, p. 48), and Koop (2003, pp. 100–104). We write as $s(\mathbf{x}, \boldsymbol{\xi})$ any of the seven statistics described above, computed from a residual

series $\{v_t(\mathbf{x}, \boldsymbol{\xi})\}_{t=1}^{T}$. The predictive distribution of $s(\mathbf{x}, \boldsymbol{\xi})$ can be simulated by recursively generating $\mathbf{x} = (x_1, \ldots, x_T)$ with the algorithm of Section 8.1, for each replication of $\boldsymbol{\xi}$ taken from a posterior sample. A one-sided posterior predictive p-value is then defined as:

$$p_1 = P\big[s(\mathbf{x}, \boldsymbol{\xi}) > E[s(\mathbf{y}, \boldsymbol{\xi}) \mid \mathbf{y}]\big] \tag{8.4}$$

and is estimated by the percentage of replications of $s(\mathbf{x}, \boldsymbol{\xi})$ that exceed the posterior average of the values computed from the actual data. A two-sided p-value is defined as $p_2 = \min(p_1, 1 - p_1)$. The one-sided p-value will be used in all cases except for M3 and M4, where the two-sided value is used.

Carlin and Louis (2000, p. 48) stress that predictive p-values should not be used for model choice, but rather as a validation tool after model selection has been done. It is partly for this reason that the analysis of this section is limited to those models that yielded the best selection criteria in Sections 6 and 7, namely, the extended local scale and AR(1)-GARCH models for the S&P500 data and the basic local scale and AR(0)-GARCH for the exchange rate series. The fact that these models yielded posterior odds close to unity lends additional motivation to the analysis of this section.

The p-values can be found in the first part of Table 12. With several thousand observations, a significance level of 1% is appropriate. It can be seen that neither model fully accounts for residual skewness, as indicated by the results for BJ and M3. On the other hand, KS, which is less sensitive to outliers, is never significant. This provides some evidence that the residual quantiles (as opposed to moments) are in line with the values expected under the maintained hypothesis. Some residual autocorrelation is present in both models for the USD-Sterling series. Apart from the slight significance of M4 for the S&P500 local scale estimates, the in-sample diagnostics of the two models are very similar.

8.3 Out-of-sample diagnostics.

In this subsection, out-of-sample forecast evaluation will be done on the same models as in Section 8.2, using five rolling and non-overlapping two-year forecasting windows, each preceded by an eight-year estimation window. This means that the posteriors are updated every two years. Combining the five forecasting

windows yields 2515 one step ahead density forecasts for the S&P500 data, and 2517 such forecasts for the exchange rate data, covering the period January 4, 1999 to December 31, 2008 in each case.

Our first forecast evaluation tool will be the probability integral transform, defined analogously to (8.1):

$$PIT(y_t^f) = \Phi^{-1}\big(P[Y_t \le y_t^f \mid \mathbf{y}_{1:t-1}]\big) \tag{8.5}$$

where y_t^f is the observation being predicted; see Rosenblatt (1952), Diebold *et al.* (1998) and Berkowitz (2001). The probability in (8.5) can be estimated by an average of the predictive probabilities (8.2), or of their GARCH counterparts, over the parameter replications in the most recent posterior sample. The first five diagnostics in Section 8.2 were used for testing the hypothesis that the series (8.5) is Gaussian white noise; the p-values in the second part of Table 12 are based on their asymptotic distributions.

The second tool is based on the methodology of Christoffersen (1998). In evaluating forecasting performance on the S&P500 series, we use, for each t in the combined forecasting window:

$$
\begin{aligned}
\mathcal{I}_t &= 1 \quad \text{if } y_t^f < q_t \\
&= 0 \quad \text{otherwise}
\end{aligned}
\tag{8.6}
$$

where q_t is the value at risk, defined as the 5th percentile of the estimated one step ahead predictive. For the exchange rate series, a two-sided counterpart of (8.6) is used:

$$
\begin{aligned}
\mathcal{J}_t &= 1 \quad \text{if } y_t^f < q_{1t} \text{ or } y_t^f > q_{2t} \\
&= 0 \quad \text{otherwise}
\end{aligned}
\tag{8.7}
$$

where q_{1t} and q_{2t} are the lower and upper bounds of a 95% one step ahead prediction interval. The tests proposed by Christoffersen are:

(1) UC : unconditional coverage: test of $p = 0.05$, where p is the probability that $\mathcal{I}_t$ or $\mathcal{J}_t$ equals one, assuming that $\mathcal{I}_t$ or $\mathcal{J}_t$ have independent Bernoulli distributions.

27

(2) IND : test of independence, versus the alternative that $\mathcal{I}_t$ or $\mathcal{J}_t$ follows a Markov chain.

(3) CC : joint test of unconditional coverage and independence.

The out-of-sample performance of the local scale model is better than its in-sample counterpart. In the bottom part of Table 12, it is seen that the only significant diagnostic for the local scale model is the BJ statistic for the S&P500 data. The diagnostics of both models are again generally quite similar; in particular, both models estimate adequate out-of-sample coverage, which is perhaps the most important objective for financial analysts.

To conclude, there is little in Table 12 that would lead one to prefer either the GARCH or the local scale model, so that this section confirms the nearly equal rankings implied by Table 9.

9. Discussion and conclusions

This paper has attempted to develop a full Bayesian treatment of an extended version of the local scale model in Shephard (1994). An extensive Monte Carlo experiment and applications to three publicly available data sets have shown that posterior simulation is straightforward, and that exact, efficient simulation smoothing is possible. Simulation smoothing, which is not available in GARCH models, turned out to be essential for an accurate interval estimation of the volatilities.

This paper has also shown that introducing ARCH-M effects and variance regressors significantly improves the marginal likelihood when the model is estimated on S&P500 data. The ARCH-M function has two features that would be difficult to introduce in Lognormal SV models. First, it is formulated in terms of the scale parameter s_t of the one step ahead predictive of the dependent variable. This is arguably more natural than a formulation using the expectation of the volatility conditional on its past history, since volatility is unobservable. Second, the autoregression coefficients depend on s_t, and this dependence is found to be very significant for the S&P500 data.

The predictive performance of the extended local scale model turned out to be comparable to that of an asymmetric t-GARCH model, both in sample and out

of sample. Information criteria were nearly equal or better. The latter conclusion confirms the findings obtained by Shephard (1994), using maximum likelihood and much smaller samples of exchange rate data.

The extended local scale model yielded marginal likelihoods and information criteria that were significantly better than those obtained with the Lognormal SV model. It has the additional advantage of enabling the straightforward simulation of one step ahead Bayesian predictives. This is impractical in the Lognormal SV model, which requires particle filtering to achieve that objective.

A potential disadvantage of the local scale model, when compared to Lognormal SV, is the difficulty of introducing correlation between the observation and evolution disturbances (Yu, 2005; Omori et $al.$, 2007). It is, however, straightforward to introduce dependence between current volatilities and past observations. Another limitation is the inability to specify an asymmetric news impact function of the EGARCH type: all the attempts to generalize the model in that direction led to serious numerical problems for large values of T, including exponent overflows. This limitation does not appear to have been mentioned in the previous literature. However, the fact that the model remains competitive with asymmetric t-GARCH formulations suggests that the empirical implications of this shortcoming are not too serious.

APPENDIX A. VALIDITY OF THE SIMULATION SMOOTHER

In this Appendix, we will show that the two densities on the right-hand side of (3.2), though not conjugate, yield a translated Gamma posterior.

Proposition 1. *Bayes' theorem implies:*

$$f(\theta_t \mid \mathbf{y}_{1:t}, \theta_{t+1}) = f_G\left(\theta_t - \phi_{t+1}^{-1}\theta_{t+1}; \frac{1}{r}, b_t\right) \quad for\ t = T-1, \ldots, 0$$

where ϕ_{t+1} is given by (4.5) and b_t is given by (4.4).

Proof. First note that if $Y = cX$ with $X \sim Be(\alpha, \beta)$ and $c > 0$, then:

$$
\begin{aligned}
f_Y(y) &= f_B\left(\frac{y}{c}; \alpha, \beta\right)\frac{1}{c}\mathcal{I}_{(0,c)}(y) \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}y^{\alpha-1}(c-y)^{\beta-1}c^{1-\alpha-\beta}\mathcal{I}_{(0,c)}(y).
\end{aligned}
$$

Applying the preceding result to $\theta_{t+1} = \phi_{t+1}\theta_t\eta_{t+1}$ and recalling that $a_t = a = 1/[r(1-\omega)]$ yields:

$$f(\theta_{t+1} \mid \theta_t, \mathbf{y}_{1:t}) \propto \theta_{t+1}^{\omega a-1}(\phi_{t+1}\theta_t - \theta_{t+1})^{1/r-1}(\phi_{t+1}\theta_t)^{1-\omega a-1/r}\mathcal{I}_{(0,\phi_{t+1}\theta_t)}(\theta_{t+1})$$

$$\propto \theta_{t+1}^{\omega a-1}\phi_{t+1}^{1/r-1}(\theta_t - \phi_{t+1}^{-1}\theta_{t+1})^{1/r-1} \quad \times$$

$$(\phi_{t+1}\theta_t)^{1-\omega a-1/r}\mathcal{I}_{(0,\phi_{t+1}\theta_t)}(\theta_{t+1}). \tag{A.1}$$

Viewing (A.1) as the likelihood of θ_t and omitting all those terms that do not depend on θ_t yields:

$$\mathcal{L}(\theta_t; \theta_{t+1}, \mathbf{y}_{1:t}) \propto (\theta_t - \phi_{t+1}^{-1}\theta_{t+1})^{1/r-1}\theta_t^{1-\omega a-1/r}\mathcal{I}_{(\phi_{t+1}^{-1}\theta_{t+1},\infty)}(\theta_t). \tag{A.2}$$

On the other hand, (2.9) implies:

$$f(\theta_t \mid \mathbf{y}_{1:t}) \propto \theta_t^{a-1}\exp(-b_t\theta_t)\mathcal{I}_{(0,\infty)}(\theta_t). \tag{A.3}$$

Multiplying (A.2) and (A.3) yields:

$$f(\theta_t \mid \mathbf{y}_{1:t}, \theta_{t+1}) \propto (\theta_t - \phi_{t+1}^{-1}\theta_{t+1})^{1/r-1}\exp(-\phi_{t+1}^{-1}\theta_{t+1}b_t) \quad \times$$

$$\exp[-b_t(\theta_t - \phi_{t+1}^{-1}\theta_{t+1})]\mathcal{I}_{(\phi_{t+1}^{-1}\theta_{t+1},\infty)}(\theta_t)$$

$$\propto (\theta_t - \phi_{t+1}^{-1}\theta_{t+1})^{1/r-1}\exp[-b_t(\theta_t - \phi_{t+1}^{-1}\theta_{t+1})]\mathcal{I}_{(\phi_{t+1}^{-1}\theta_{t+1},\infty)}(\theta_t)$$

where we have used $a - \omega a = 1/r$. This is the kernel of a $\mathrm{Ga}(1/r, b_t)$ density on $\theta_t - \phi_{t+1}^{-1}\theta_{t+1}$, proving Proposition 1. $\quad\square$

Appendix B. Proposal densities

B.1 Sampling ω and r.

Since the method used will be the same, we will use the generic expression x to denote a draw of either parameter. The prior of both parameters is Uniform $U(x_{\min}, x_{\max})$.

The candidate we propose is a linear function $x_{\min} + (x_{\max} - x_{\min})X$ of a Beta variable X with parameters a_x and b_x:

$$q(x \mid \boldsymbol{\varphi}) \propto (x - x_{\min})^{a_x-1}(x_{\max} - x)^{b_x-1}\mathcal{I}_{(x_{\min}, x_{\max})}(x). \tag{B.1}$$

In (B.1), $\boldsymbol{\varphi}$ contains all the parameters in $\boldsymbol{\xi}$ except x, and a_x, b_x are chosen in such a way that the location of the candidate density corresponds to an approximate expected value $\bar{x}$ of the full conditional posterior of x. This value can be computed as:

$$\bar{x} = \sum_{i=1}^{L} p_i x_i$$

where:

$$p_i = \frac{\mathcal{L}(x_i; \boldsymbol{\varphi})}{\sum_{j=1}^{L} \mathcal{L}(x_j; \boldsymbol{\varphi})}$$

and where the x_i form a grid of equally spaced values $x_1, \ldots, x_L$ covering the prior support. The values for a_x and b_x solve the equation:

$$\frac{a_x}{a_x + b_x} = \frac{\bar{x} - x_{\min}}{x_{\max} - x_{\min}} \tag{B.2}$$

subject to $a_x = \delta_x$ (if the right-hand side of (B.2) is less than 0.5) or $b_x = \delta_x$ (if the right-hand side of (B.2) is larger than 0.5), where $\delta_x \geq 1$ is a tuning parameter. This ensures that both a_x and b_x are larger than δ_x. Decreasing δ_x will increase the variance of the candidate while leaving its location unchanged. In all the simulations of this paper, the same values of $\delta_x = 20$, $L = 10$ were chosen after some experimentation and led to well-mixing MCMC chains.

B.2 Sampling $\boldsymbol{\alpha}$.

We obtain the proposal density by combining the Multinormal prior in (4.1) with a likelihood of $\boldsymbol{\alpha}$ suggested by (2.1) and (2.5). However, considerations suggested by the choice of an appropriate importance sampling density indicate that heavier tails than the Normal might be required. Let $\boldsymbol{\varphi}_\alpha$ denote all the parameters in $\boldsymbol{\xi}$ except $\boldsymbol{\alpha}$. We first run the filter (4.3)–(4.9) to obtain $\mathbf{b}_{0:T}$, $\boldsymbol{\phi}_{1:T}$, and $\mathbf{s}_{1:T}$, then run the smoother (3.6)–(3.7) to obtain the conditional posterior expectations $\bar{\theta}_t$. Note that all these quantities depend on $\boldsymbol{\alpha}$ and $\boldsymbol{\varphi}_\alpha$. Define $\mathbf{y} = \mathbf{y}_{1:T}$,

$$\Theta(\boldsymbol{\alpha}, \boldsymbol{\varphi}_\alpha) = \mathrm{diag}[(\lambda\bar{\theta}_1)^{2/r}, \ldots, (\lambda\bar{\theta}_T)^{2/r}]$$

and $X_\alpha(\boldsymbol{\alpha}, \boldsymbol{\varphi}_\alpha)$ as the T by $2p+2$ matrix with row t equal to:

$$\begin{pmatrix} 1 & F(s_t) & y_{t-1} & F(s_t)y_{t-1} & \cdots & y_{t-p} & F(s_t)y_{t-p} \end{pmatrix}.$$

The proposal density is multivariate Student with ν_α degrees of freedom, location vector $\boldsymbol{\alpha}^*$, and scale matrix $s_\alpha^2 \Sigma_\alpha^*$:

$$q(\boldsymbol{\alpha} \mid \boldsymbol{\alpha}_{\text{old}}, \boldsymbol{\varphi}_\alpha) \propto (\det s_\alpha^2 \Sigma_\alpha^*)^{-1/2} \quad \times$$
$$\left[1 + \frac{(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)'(s_\alpha^2 \Sigma_\alpha^*)^{-1}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)}{\nu_\alpha} \right]^{-(\nu_\alpha + 2p + 2)/2} \tag{B.3}$$

where:

$$(\Sigma_\alpha^*)^{-1} = \Sigma_\alpha^{-1} + X_\alpha'(\boldsymbol{\alpha}_{\text{old}}, \boldsymbol{\varphi}_\alpha) \Theta(\boldsymbol{\alpha}_{\text{old}}, \boldsymbol{\varphi}_\alpha) X_\alpha(\boldsymbol{\alpha}_{\text{old}}, \boldsymbol{\varphi}_\alpha)$$
$$\boldsymbol{\alpha}^* = \Sigma_\alpha^* \left[\Sigma_\alpha^{-1} \mathbf{m}_\alpha + X_\alpha'(\boldsymbol{\alpha}_{\text{old}}, \boldsymbol{\varphi}_\alpha) \Theta(\boldsymbol{\alpha}_{\text{old}}, \boldsymbol{\varphi}_\alpha) \mathbf{y} \right] .$$

and where ν_α and s_α^2 are tuning parameters. In all the simulations of this paper, identical values of $\nu_\alpha = 3$, $s_\alpha^2 = 1$ were chosen after some experimentation and led to well-mixing MCMC chains.

B.3 Sampling $\boldsymbol{\beta}$.

We obtain the proposal density by combining the Multinormal prior in (4.1) with a likelihood suggested by (2.2). It is easy to show that:

$$\sigma_\eta^2 = V(\ln \eta_{t+1}) = \Psi'\left(\frac{\omega}{r(1 - \omega)}\right) - \Psi'\left(\frac{1}{r(1 - \omega)}\right) \tag{B.4}$$

where $\Psi'(z) = d^2 \ln \Gamma(z)/dz^2$ is the trigamma function, which can be computed using the algorithm in Bowman (1984). Let $\boldsymbol{\varphi}_\beta$ denote all the parameters in $\boldsymbol{\xi}$ except $\boldsymbol{\beta}$. We run the filter and smoother to obtain the $\bar{\theta}_t$, define $\mathbf{y}_\beta(\boldsymbol{\beta}, \boldsymbol{\varphi}_\beta)$ as the $(T-1) \times 1$ vector with element t equal to $\ln \bar{\theta}_{t+1} - \ln \bar{\theta}_t$ for $1 \leq t \leq T - 1$, and X_β as the $(T-1) \times k$ matrix with row t equal to:

$$(D_{1,t+1} \quad \ldots \quad D_{k,t+1}) .$$

The proposal density is multivariate Student with ν_β degrees of freedom, location vector $\boldsymbol{\beta}^*$, and scale matrix $s_\beta^2 \Sigma_\beta^*$:

$$q(\boldsymbol{\beta} \mid \boldsymbol{\beta}_{\text{old}}, \boldsymbol{\varphi}_\beta) \propto (\det s_\beta^2 \Sigma_\beta^*)^{-1/2} \left[1 + \frac{(\boldsymbol{\beta} - \boldsymbol{\beta}^*)'(s_\beta^2 \Sigma_\beta^*)^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}^*)}{\nu_\beta} \right]^{-(\nu_\beta + k)/2}$$

(B.5)

where:

$$(\Sigma_\beta^*)^{-1} = \Sigma_\beta^{-1} + \frac{X_\beta' X_\beta}{\sigma_\eta^2}$$

$$\boldsymbol{\beta}^* = \Sigma_\beta^* \left[\Sigma_\beta^{-1} \mathbf{m}_\beta + \frac{X_\beta' \mathbf{y}_\beta(\boldsymbol{\beta}_{\text{old}}, \boldsymbol{\varphi}_\beta)}{\sigma_\eta^2} \right].$$

and where ν_β and s_β^2 are again tuning parameters. In all the simulations of this paper, identical values of $\nu_\beta = 3$, $s_\beta^2 = 400$ were chosen after some experimentation and led to well-mixing MCMC chains. Setting $s_\beta^2 = 1$ led to low rejection rates but high autocorrelations.

B.4 Sampling λ.

We obtain the candidate density by combining the Gamma prior in (4.1) with a likelihood suggested by (2.1). Let $\boldsymbol{\varphi}_\lambda$ denote all the parameters in $\boldsymbol{\xi}$ except λ. We run the filter and smoother to obtain the μ_t and $\bar{\theta}_t$, and define:

$$\Lambda_t(\lambda, \boldsymbol{\varphi}_\lambda) = \left| y_t - \mu_t(\lambda, \boldsymbol{\varphi}_\lambda) \right|^r \bar{\theta}_t(\lambda, \boldsymbol{\varphi}_\lambda).$$

The candidate density is:

$$q(\lambda \mid \lambda_{\text{old}}, \boldsymbol{\varphi}_\lambda) = f_G(\lambda; \delta_\lambda a_\lambda^*, \delta_\lambda b_\lambda^*)$$

(B.6)

with:

$$a_\lambda^* = a_\lambda + \frac{T}{r}$$

$$b_\lambda^* = b_\lambda + \psi(r) \sum_{t=1}^T \Lambda_t(\lambda_{\text{old}}, \boldsymbol{\varphi}_\lambda)$$

and where the tuning parameter δ_λ is chosen by experimentation. Lowering δ_λ increases the variance of the candidate but leaves its expectation unchanged. In all the simulations of this paper, an identical value of $\delta_\lambda = 1/625$ was chosen after some experimentation and led to well-mixing MCMC chains. As before, high values of δ_λ led to low rejection rates but high autocorrelations.

References

Ardia, D., 2008, Financial Risk Management with Bayesian Estimation of GARCH Models: Theory and Applications. Springer-Verlag, Berlin.

Ardia, D., 2009, Bayesian estimation of a Markov-switching threshold asymmetric GARCH model with Student-t innovations. Econometrics Journal 12, 105–126.

Berkowitz, J., 2001, Testing density forecasts, with applications to risk management. Journal of Business and Economic Statistics 19, 465–474.

Bernardo, J.M. and A.F.M. Smith, 2000, Bayesian theory. Wiley, Chichester.

Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Bollerslev, T., Engle, R.F, and D.B. Nelson, 1994, ARCH models, in: R.F. Engle and D.L. McFadden (Eds.), Handbook of econometrics, vol. 4. Elsevier, Amsterdam, pp. 2959–3038.

Box, G.E.P., and G.C. Tiao, 1992, Bayesian inference in statistical analysis. Wiley, New York.

Bowman, K.O., 1984, Computation of the polygamma functions. Communications in Statistics – Simulation and Computation 13, 409–415.

Carlin, B.P., and T.A. Louis, 2000, Bayes and empirical Bayes methods for data analysis, 2nd edition. Chapman and Hall/CRC, Boca Raton, Florida.

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

Chib, S., 1995, Marginal likelihood from the Gibbs output. Journal of the American Statistical Association 90, 1313–1321.

Chib, S., 1996, Calculating posterior distributions and modal estimates in Markov mixture models. Journal of Econometrics 75, 79–97.

Chib, S., and E. Greenberg, 1998, Analysis of multivariate probit models. Biometrika 85, 347–361.

Chib, S., and I. Jeliazkov, 2001, Marginal likelihood from the Metropolis-Hastings output. Journal of the American Statistical Association 96, 270–281.

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

Christoffersen, P.F., 1998, Evaluating interval forecasts. International Economic Review 39, 841–862.

Deschamps, P.J., 2003, Time-varying intercepts and equilibrium analysis: an extension of the dynamic almost ideal demand model. Journal of Applied Econometrics 18, 209–236.

Deschamps, P.J., 2008, Comparing smooth transition and Markov switching autoregressive models of US unemployment. Journal of Applied Econometrics 23, 435–462.

Diebold, F.X., Gunther, T.A., and A.S. Tay, 1998, Evaluating density forecasts with applications to financial risk management. International Economic Review 39, 863–883.

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

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

French, K.R., Schwert, G.W., and R.F. Stambaugh, 1987, Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.

Frühwirth-Schnatter, S., 2004, Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. Econometrics Journal 7, 143–167.

Gelfand, A.E., and D.K. Dey, 1994, Bayesian model choice: asymptotics and exact calculations. Journal of the Royal Statistical Society (series B) 56, 501–514.

Gelman, A., and X.L. Meng, 1996, Model checking and model improvement, in: W.R. Gilks, S. Richardson, and D.J. Spiegelhalter (Eds.), Markov chain Monte Carlo in practice. Chapman and Hall, London, pp. 189–201.

Gelman, A., and D. Rubin, 1992, Inference from iterative simulation using multiple sequences. Statistical Science 7, 457–511.

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

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

Harvey, A.C, Trimbur, T.M., and H.K. van Dijk, 2007, Trends and cycles in economic time series: A Bayesian approach. Journal of Econometrics 140, 618–649.

Jacquier, E., Polson, N.G., and P.E. Rossi, 1994, Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371–389.

Jacquier, E., Polson, N.G., and P.E. Rossi, 2004, Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. Journal of Econometrics 122, 185–212.

Jeffreys, H., 1961, Theory of Probability (3d. edition). Oxford University Press, Oxford.

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

Koop, G., 2003, Bayesian econometrics. Wiley, Chichester.

LeBaron, B., 1992, Some relations between volatility and serial correlation in stock market returns. Journal of Business 65, 199–220.

McDonald, J.B., and W.K. Newey, 1988, Partially adaptive estimation of regression models via the generalized t distribution. Econometric Theory 4, 428–457.

Meng, X.L., and W.H. Wong, 1996, Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. Statistica Sinica 6, 831–860.

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

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

Philipov, A., and M.E. Glickman, 2006, Multivariate stochastic volatility via Wishart processes. Journal of Business and Economic Statistics 24, 313–328.

Rosenblatt, M., 1952, Remarks on a multivariate transformation. Annals of Mathematical Statistics 23, 470–472.

Schotman, P. C., and H.K. van Dijk, 1991, A Bayesian analysis of the unit root in real exchange rates. Journal of Econometrics 49, 195–238.

Shephard, N., 1994, Local scale models: state space alternative to integrated GARCH processes. Journal of Econometrics 60, 181–202.

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 (series B) 48, 79–88.

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

Uhlig, H., 1997, Bayesian vector autoregressions with stochastic volatility. Econometrica 65, 59–73.

Yu, J., 2005, On leverage in a stochastic volatility model. Journal of Econometrics 127, 165–178.

TABLE 1. *DGP parameter values (Monte Carlo experiment)*

Model	Param.	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6
Lognormal SV	ϕ	0.986	0.984				
	σ^2	0.018	0.016				
	ρ	-0.558	0.000				
	μ	-0.171	-1.150				
t-GARCH	γ_0			0.029	0.000		
	γ_1			0.074	0.000		
	α_0			0.009	0.004		
	α_1			0.020	0.045		
	α_2			0.097	0.056		
	β			0.933	0.942		
	ν			8.610	6.866		
Local Scale	α_{00}					0.019	0.000
	α_{10}					0.034	0.000
	α_{01}					0.202	0.000
	α_{11}					-0.645	0.000
	β_1					0.232	0.000
	λ					0.151	0.095
	r					1.680	1.520
	ω					0.951	0.959

TABLE 2. *Statistics of simulated data (Monte Carlo experiment)*

Statistic		DGP1	DGP2	DGP3	DGP4	DGP5	DGP6
Mean	Min.	−0.028	−0.020	−0.039	−0.021	0.023	−0.031
	Max.	0.029	0.025	0.043	0.024	0.090	0.068
Std. dev.	Min.	0.982	0.545	0.836	0.590	0.057	0.132
	Max.	1.294	0.699	1.477	0.851	1.554	3.476
Skewness	Min.	−0.374	−0.224	−0.209	−0.221	−0.005	−0.615
	Max.	0.400	0.245	0.285	0.412	0.648	0.174
Excess kurtosis	Min.	1.182	0.988	1.132	1.281	2.285	2.220
	Max.	6.276	2.604	9.833	9.233	67.914	16.991
Autocorrelation	Min.	−0.024	−0.037	0.036	−0.049	−0.136	−0.078
	Max.	0.020	0.035	0.122	0.043	0.215	0.031
Autocorrelation of squares	Min.	0.091	0.108	0.055	0.031	0.109	0.097
	Max.	0.314	0.199	0.205	0.194	0.323	0.384

TABLE 3. *Results of Monte Carlo experiment*

Estimated model	Statistic	Data Generating Process					
		DGP1	DGP2	DGP3	DGP4	DGP5	DGP6
Model 1:	MAE	0.114	0.076	0.105	0.080	0.055	0.077
Lognormal SV	RMSE	0.155	0.100	0.148	0.108	0.133	0.152
with leverage	Coverage	0.049	0.055	0.029	0.034	0.188	0.049
	Coverage t-stat.	-0.163	1.666	-7.708	-5.286	14.345	-0.215
Model 2:	MAE	0.162	0.103	0.032	0.030	0.075	0.106
AR(1)	RMSE	0.218	0.135	0.051	0.049	0.180	0.208
t-GARCH	Coverage	0.743	0.757	0.068	0.130	0.717	0.685
	Coverage t-stat.	132.739	143.149	3.405	10.841	102.284	84.930
Model 3:	MAE	0.117	0.077	0.105	0.072	0.049	0.074
Extended	RMSE	0.158	0.101	0.152	0.097	0.125	0.143
Local Scale	Coverage	0.069	0.094	0.044	0.044	0.054	0.061
	Coverage t-stat.	4.709	9.831	-1.784	-1.722	0.987	2.470

MAE : Mean absolute error of volatility point estimates.

RMSE : Root mean squared error of volatility point estimates.

Coverage : Proportion of true volatilities outside of 95% confidence intervals.

TABLE 4. *Bayes factors and differences of information criteria (S&P500 data)*

Observation density	Variance regressors		No ARCH-M $\mu_t = 0$	ARCH-M effects $p = 0$	$p = 1$	$p = 2$
Normal	D_{1t}, D_{2t}	$\log_{10}(BF)$	-22.843	-23.396	-10.692	-13.185
		NSE	0.003	0.003	0.003	0.003
		$\Delta(BIC)$	-100.862	-103.634	-45.523	-57.462
GED	D_{1t}, D_{2t}	$\log_{10}(BF)$	-9.574	-9.395	0.000	-2.061
		NSE	0.003	0.003		0.003
		$\Delta(BIC)$	-43.676	-42.425	0.000	-10.135
Normal	D_{2t}	$\log_{10}(BF)$	-53.267	-48.541	-35.506	-37.620
		NSE	0.003	0.003	0.003	0.003
		$\Delta(BIC)$	-241.732	-220.266	-160.569	-170.789
GED	D_{2t}	$\log_{10}(BF)$	-36.729	-32.139	-22.531	-24.179
		NSE	0.003	0.003	0.003	0.003
		$\Delta(BIC)$	-168.884	-148.308	-104.857	-113.235
Normal	None	$\log_{10}(BF)$	-58.728	-54.381	-39.350	-41.480
		NSE	0.002	0.003	0.003	0.003
		$\Delta(BIC)$	-268.156	-248.424	-179.567	-189.927
GED	None	$\log_{10}(BF)$	-40.855	-36.584	-25.560	-27.220
		NSE	0.003	0.003	0.003	0.003
		$\Delta(BIC)$	-189.374	-170.324	-120.346	-128.804

TABLE 5. *Bayes factors and differences of information criteria (exchange rate data, $\mu_t = 0$)*

Data	Observation density	Variance regressors	$\log_{10}(BF)$	NSE	$\Delta(BIC)$
US dollar/Swiss Franc	Normal	None	-19.849	0.002	-88.999
US dollar/Swiss Franc	Normal	D_{1t}	-21.254	0.002	-95.210
US dollar/Swiss Franc	GED	None	0.000		0.000
US dollar/Swiss Franc	GED	D_{1t}	-1.111	0.002	-4.128
US dollar/Sterling	Normal	None	-21.090	0.001	-94.878
US dollar/Sterling	Normal	D_{1t}	-22.917	0.002	-103.506
US dollar/Sterling	GED	None	0.000		0.000
US dollar/Sterling	GED	D_{1t}	-1.945	0.002	-8.540

TABLE 6. *Posterior replication summaries (S&P500 data)*

θ	$\theta_{0.025}$	$\theta_{0.5}$	$\theta_{0.975}$	RNE	ρ_1	ρ_5	Reject. rate
α_{00}	-0.038	0.019	0.075	0.80	0.11	-0.00	0.58
α_{10}	-0.288	0.035	0.359	0.80	0.11	-0.01	0.58
α_{01}	0.138	0.202	0.266	0.70	0.18	0.01	0.58
α_{11}	-0.924	-0.646	-0.365	0.75	0.14	-0.00	0.58
β_1	0.186	0.232	0.280	0.52	0.32	0.01	0.73
β_2	-0.106	-0.076	-0.047	0.14	0.76	0.23	0.73
r	1.602	1.676	1.755	0.82	0.10	0.01	0.33
λ	0.061	0.135	0.350	0.36	0.47	0.02	0.24
ω	0.945	0.951	0.957	0.71	0.17	0.03	0.63

θ_α: quantile at probability α; RNE: relative numerical efficiency;

ρ_i: autocorrelation at lag i.

TABLE 7. *Posterior replication summaries (US Dollar/Swiss Franc exchange rate)*

θ	$\theta_{0.025}$	$\theta_{0.5}$	$\theta_{0.975}$	RNE	ρ_1	ρ_5	Reject. rate
r	1.466	1.542	1.623	0.90	0.05	-0.01	0.53
λ	0.032	0.060	0.126	0.50	0.34	0.01	0.34
ω	0.959	0.966	0.973	0.78	0.12	0.02	0.54

θ_α: quantile at probability α; RNE: relative numerical efficiency;

ρ_i: autocorrelation at lag i.

TABLE 8. *Posterior replication summaries (US Dollar/Sterling exchange rate)*

θ	$\theta_{0.025}$	$\theta_{0.5}$	$\theta_{0.975}$	RNE	ρ_1	ρ_5	Reject. rate
r	1.443	1.519	1.601	0.91	0.05	-0.01	0.52
λ	0.045	0.087	0.188	0.51	0.33	0.03	0.32
ω	0.951	0.959	0.966	0.78	0.13	0.01	0.60

θ_α: quantile at probability α; RNE: relative numerical efficiency;

ρ_i: autocorrelation at lag i.

TABLE 9. *Bayes factors and differences of information criteria (competing models against best local scale models)*

Data	Model	$\log_{10}(BF)$	NSE	$\Delta(BIC)$
S&P500	AR(0) t-GARCH	-10.941	0.009	-54.564
S&P500	AR(1) t-GARCH	-0.057	0.008	0.697
S&P500	Lognormal SV (leverage)	-16.808	0.031	-104.252
S&P500	Lognormal SV (no leverage)	-53.680	0.019	-270.446
USD-CHF	AR(0) t-GARCH	-0.469	0.009	-6.698
USD-CHF	AR(1) t-GARCH	-5.454	0.014	-23.863
USD-CHF	Lognormal SV (leverage)	-6.100	0.021	-56.577
USD-CHF	Lognormal SV (no leverage)	-5.131	0.017	-48.037
USD-Sterling	AR(0) t-GARCH	-0.155	0.013	-3.687
USD-Sterling	AR(1) t-GARCH	-3.991	0.009	-15.201
USD-Sterling	Lognormal SV (leverage)	-11.285	0.025	-79.103
USD-Sterling	Lognormal SV (no leverage)	-10.777	0.020	-72.950

Table 10. *MCMC estimates of the best Lognormal SV models*

θ	S&P500			USD-CHF			USD-Sterling		
	$\theta_{0.025}$	$\theta_{0.5}$	$\theta_{0.975}$	$\theta_{0.025}$	$\theta_{0.5}$	$\theta_{0.975}$	$\theta_{0.025}$	$\theta_{0.5}$	$\theta_{0.975}$
ϕ	0.981	0.986	0.989	0.968	0.979	0.987	0.977	0.984	0.990
σ^2	0.015	0.018	0.023	0.008	0.012	0.018	0.010	0.015	0.022
ρ	-0.618	-0.558	-0.481	NA	NA	NA	NA	NA	NA
μ	-0.280	-0.193	0.008	-0.897	-0.770	-0.637	-1.335	-1.147	-0.949

θ_α: quantile at probability α; NA: not applicable.

Table 11. *MCMC estimates of the best GARCH models*

θ	S&P500			USD-CHF			USD-Sterling		
	$\theta_{0.025}$	$\theta_{0.5}$	$\theta_{0.975}$	$\theta_{0.025}$	$\theta_{0.5}$	$\theta_{0.975}$	$\theta_{0.025}$	$\theta_{0.5}$	$\theta_{0.975}$
γ_0	0.014	0.029	0.044	NA	NA	NA	NA	NA	NA
γ_1	0.054	0.074	0.094	NA	NA	NA	NA	NA	NA
α_0	0.006	0.008	0.012	0.004	0.007	0.011	0.002	0.003	0.006
α_1	0.012	0.020	0.029	0.027	0.038	0.052	0.033	0.045	0.059
α_2	0.083	0.097	0.113	0.034	0.044	0.057	0.042	0.056	0.072
β	0.922	0.933	0.942	0.931	0.946	0.959	0.926	0.942	0.956
ν	7.415	8.570	10.072	6.559	7.822	9.537	5.855	6.820	8.133

θ_α: quantile at probability α; NA: not applicable.

TABLE 12. *P-values of misspecification diagnostics*

| | | S&P500 | | USD-CHF | | USD-Sterling | |
		Local scale	GARCH	Local scale	GARCH	Local scale	GARCH
	AC	0.036	0.114	0.357	0.446	0.004	0.004
	ARCH	0.390	0.095	0.609	0.402	0.413	0.607
	LR	0.039	0.306	0.814	0.510	0.873	0.543
In sample	BJ	0.003	0.002	0.000	0.002	0.017	0.034
	KS	0.240	0.183	0.127	0.135	0.230	0.025
	M3	0.000	0.001	0.000	0.000	0.006	0.008
	M4	0.004	0.195	0.259	0.361	0.114	0.357
	AC	0.114	0.003	0.074	0.070	0.183	0.181
	ARCH	0.051	0.033	0.268	0.052	0.404	0.202
	LR	0.030	0.025	0.491	0.636	0.777	0.146
Out of	BJ	0.000	0.003	0.023	0.167	0.052	0.016
sample	KS	0.127	0.074	0.556	0.234	0.469	0.458
	UC	0.144	0.031	0.702	0.659	0.942	0.059
	IND	0.697	0.268	0.235	0.715	0.598	0.324
	CC	0.318	0.052	0.459	0.849	0.868	0.103

AC: autocorrelation; ARCH: autocorrelation of squares; LR: LR test for $N(0, 1)$; BJ: Bera-Jarque stat.;
KS : Kolmogorov-Smirnov stat.; M3: skewness; M4: kurtosis; UC: unconditional coverage;
IND: coverage independence; CC: conditional coverage.

# Figure 1. Loglikelihood comparisons
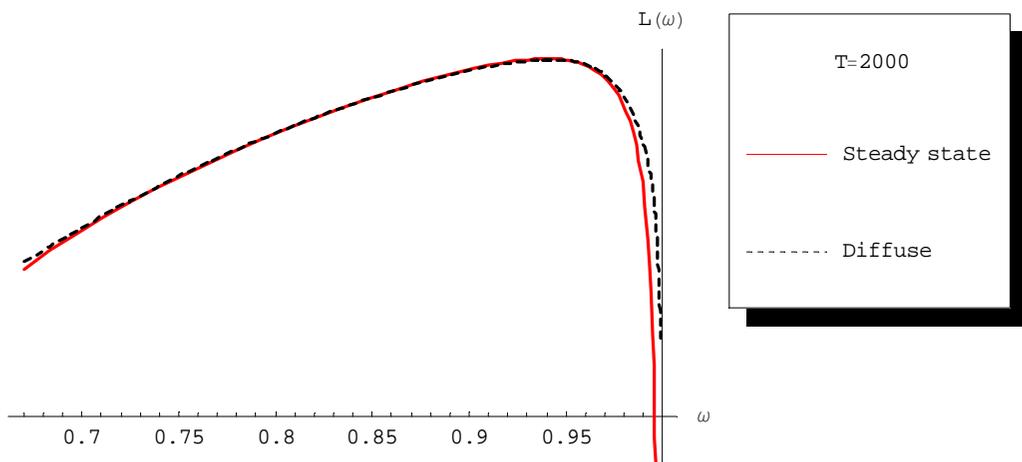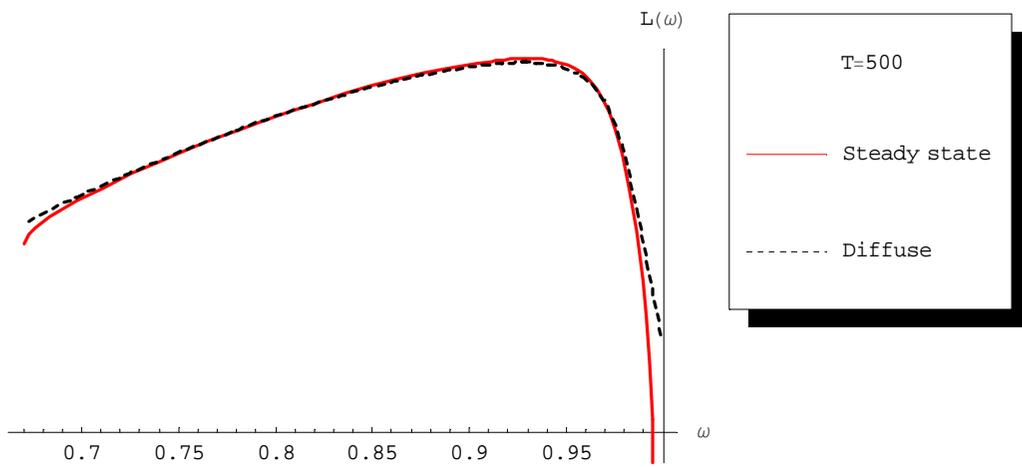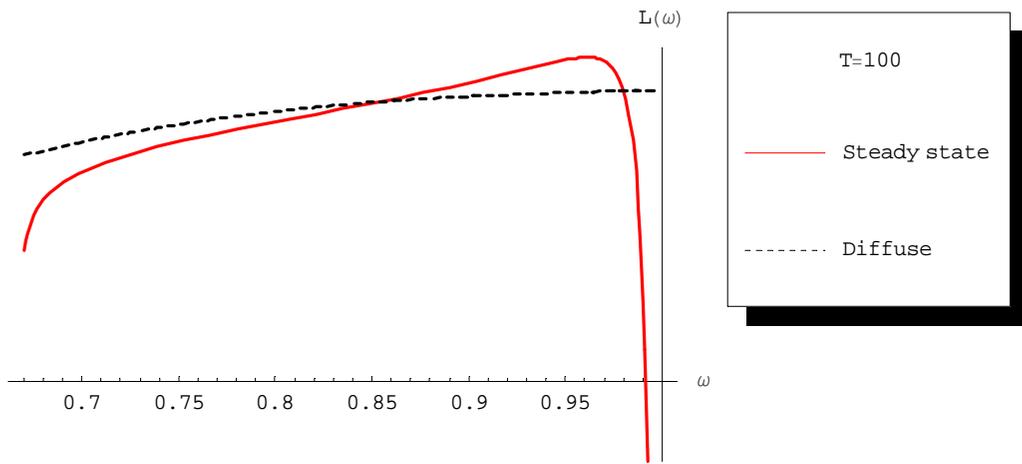
# Figure 2. True volatilities (DGP1) and volatility confidence bands

Simulated observations

Model 1: Lognormal SV with leverage

Model 2: AR(1) t-GARCH

Model 3: Extended local scale
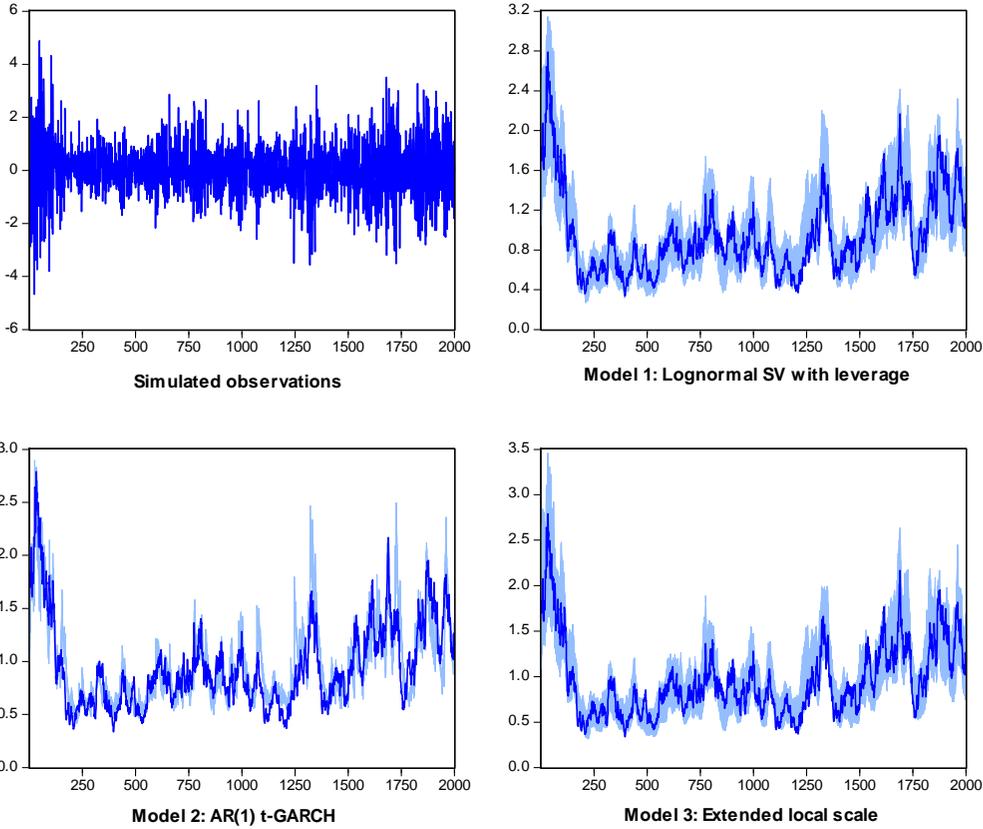
Figure 3. Variance regressors (S&P500 data)

Histogram of D(1t)

Histogram of D(2t)
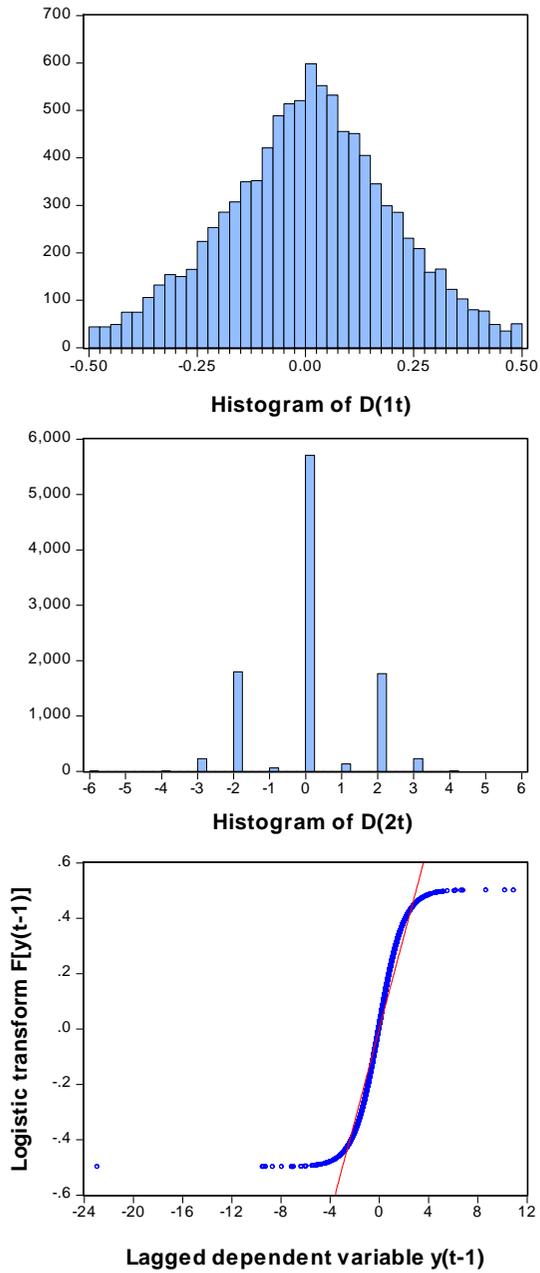
Figure 4. MCMC sample paths (S&P500 data)

α(00) α(10) α(01)

α(11) β(1) β(2)

GED parameter (r) λ ω
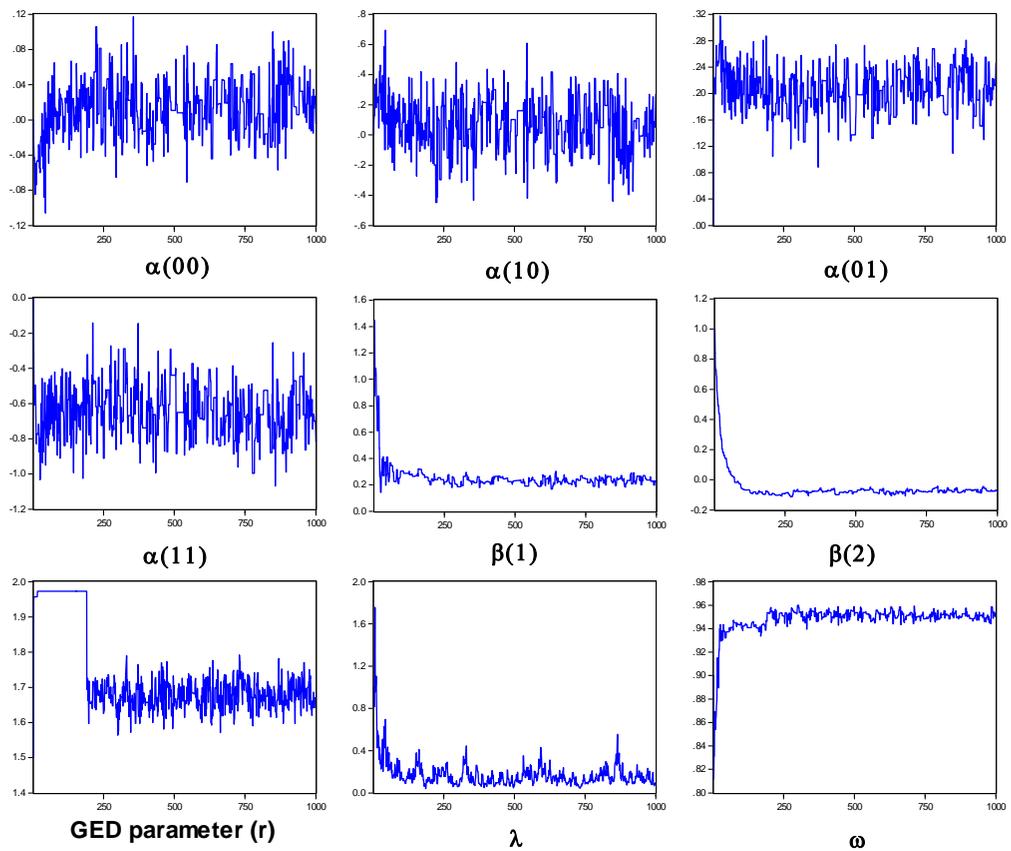
Figure 5. Anticipated volatilities and ARCH-M effects (S&P500 data)

Estimates of anticipated volatilities s(t)

Estimates of E[y(t) | s(t)]
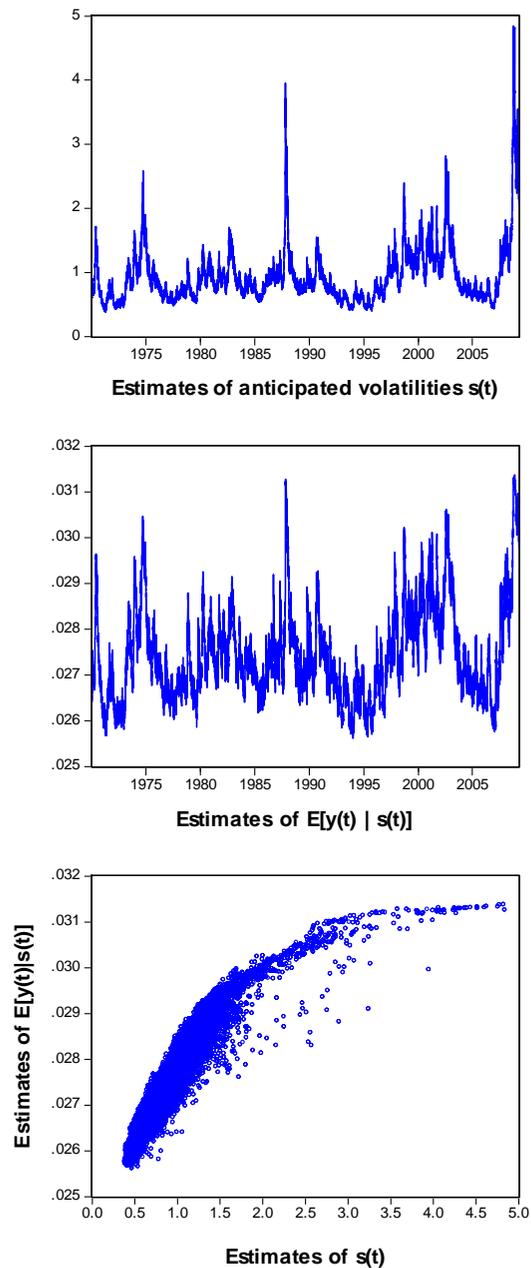
Estimates of s(t)

Figure 6. Observations and volatility confidence bands


S&P500 DATA

US DOLLAR/SWISS FRANC

US DOLLAR/STERLING
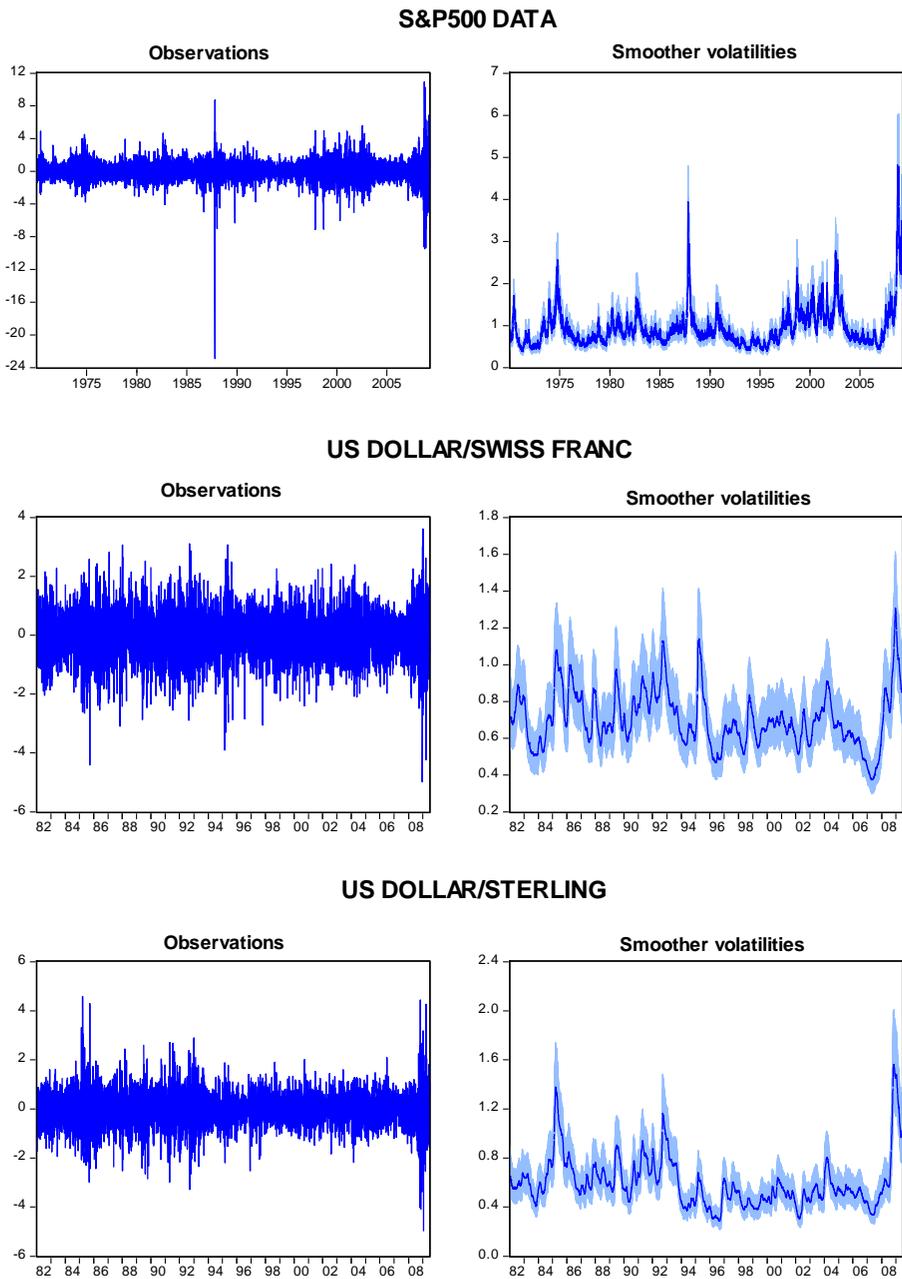
Figure 7. Comparison of volatility point estimates

Local scale minus Lognormal SV   **Local scale minus GARCH**

S&P500 data

US dollar - Swiss Franc

US dollar - Sterling

53