Likelihood-based estimation of latent generalised ARCH structures*

Gabriele Fiorentini
Università di Firenze, Viale Morgagni 59, I-50134 Firenze, Italy
fiorentini@ds.unifi.it

Enrique Sentana
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
sentana@cemfi.es

Neil Shephard
Nuffield College, Oxford OX1 1NF, U.K.
neil.shephard@nuf.ox.ac.uk

June 2003

Abstract

GARCH models are commonly used as latent processes in econometrics, financial economics and macroeconomics. Yet no exact likelihood analysis of these models has been provided so far. In this paper we outline the issues and suggest a Markov chain Monte Carlo algorithm which allows the calculation of a classical estimator via the simulated EM algorithm or a Bayesian solution in $O(T)$ computational operations, where T denotes the sample size. We assess the performance of our proposed algorithm in the context of both artificial examples and an empirical application to 26 UK sectorial stock returns, and compare it to existing approximate solutions.

Keywords: Bayesian inference; Dynamic Heteroskedasticity; Factor models; Markov chain Monte Carlo; Simulated EM algorithm; Volatility.

*A previous draft of this paper was circulated under the title "Exact likelihood-based estimation of conditionally heteroskedastic factor models". We are grateful to Manuel Arellano, Sid Chib, Antonis Demos, Lars Hansen, Brendan McCabe, Francisco Peñaranda, Eric Renault, Tina Rydberg, Harald Uhlig, and the referees, as well as seminar audiences at Ente L. Einaudi, Universitat Autonoma de Barcelona, University of Basle, Universitat Pompeu Fabra, the European Meeting of the Econometric Society (Berlin, September 1998), the Isaac Newton Institute Workshop on Econometrics and Finance (Cambridge, October, 1998), and the tenth and eleventh $(EC)^2$ Meetings (Madrid, December 1999, and Dublin, December 2000) for very helpful comments and suggestions. Of course, the usual caveat applies. Financial support from IVIE (Fiorentini) and the UK's ESRC grants "Econometrics of trade-by-trade price dynamics" and "High frequency financial econometrics based on power variation" (Shephard) is gratefully acknowledged.

1 Introduction

A key feature of the parametric ARCH models pioneered by Engle (1982) and Bollerslev (1986)

$$f_t = \varepsilon_t \lambda_t^{1/2}, \qquad \varepsilon_t | \mathcal{F}_{t-1}^f \sim N(0,1),$$

is that the volatility process $\{\lambda_t\}$ is measurable by construction with respect to the sequence of natural filtrations generated by the past values of the $\{f_t\}$ ARCH process, $\mathcal{F}_{t-1}^f$, at least up to a finite dimensional vector of unknown parameters φ. An extensive review of the econometric literature on this topic is given in Bollerslev, Engle, and Nelson (1994). An immediate consequence of this setup is that the usual prediction decomposition delivers the likelihood function of the sample conditional on initial values $\mathcal{F}_0$ as

$$p(f|\varphi, \mathcal{F}_0) = \prod_{t=1}^{T} p(f_t | \mathcal{F}_{t-1}^f, \varphi) = \prod_{t=1}^{T} \frac{1}{\sqrt{\lambda_t(\varphi)}} \phi\left(\frac{f_t}{\sqrt{\lambda_t(\varphi)}} \right),$$

where T is the sample size, $f = (f_1, ..., f_T)'$, and $\phi(.)$ denotes a standard normal density. This means that inference on unknown parameters can be carried out relatively easily. This argument continues to hold when the normality assumption is replaced by another parametric distribution, such as the Student t (e.g. Bollerslev (1987)) or the normal inverse Gaussian (e.g. Jensen and Lunde (2001)). Further, it does not really get any more complicated when we deal with the fractional ARCH models of Baillie, Bollerslev, and Mikkelsen (1996), although some care has to be taken with dealing with $\mathcal{F}_0$ appropriately (see Diebold and Schuermann (2000)). The same prediction decomposition has been used in the multivariate ARCH context by Engle and Kroner (1995) among many others (see again Bollerslev, Engle, and Nelson (1994)). The fact that the conditional likelihood function for ARCH models is easily computed is a major reason for the rapid adoption of this class of models by economists. However, the analysis becomes substantially more complicated if an ARCH model is used as a latent process, for the log-likelihood function of the observed variables can no longer be written in closed form.

The main modern way of carrying out likelihood inference on latent models is via a Markov chain Monte Carlo (MCMC) algorithm (see Chib (2001) for an extensive review). The MCMC algorithm generates simulations from the distribution of the latent process $\{f_t\}$ conditional upon the data $\{x_t\}$ and parameters φ. This simulation procedure can be used either to carry out Bayesian inference on the unknown φ, or to classically estimate φ by means of a simulated EM algorithm. In either case, a crucial feature of these methods is the continual conditional simulation of the latent vector f given $x = (x_1', ..., x_T')'$ and φ, either one element at a time or in blocks. Unfortunately, the non-Markovian nature of the GARCH process implies that each time we simulate a single f_t, we implicitly change all future conditional variances. As pointed

out by Shephard (1996), a regrettable consequence of this path-dependence in volatility is that standard MCMC algorithms will involve a $O(T^2)$ computational load, unlike what happens in the ARCH(1) case (see Giakoumatos, Dellaportas, and Politis (1999)). Since this cost has to be borne for each value of φ, such procedures are generally infeasible for the types of large financial datasets that we see in practice, even with the massive computational power economists have available to them. To some degree this has prompted interest in models that replace the GARCH assumption with discrete-time versions of stochastic volatility (SV) processes, whether lognormal or not, because their first-order Markovian structure is particularly convenient for conducting inference via MCMC methods (see the reviews of SV models by Taylor (1994), Ghysels, Harvey, and Renault (1996) and Shephard (1996), as well as the papers by Pitt and Shephard (1999b), Aguilar and West (2000), Doz and Renault (2001), and Meddahi and Renault (2002). A related line of research that also relies on MCMC methods concentrates on models in which the volatility dynamics of the latent variables is characterised by a discrete-state first-order Markov chain (see Albert and Chib (1993), Carter and Kohn (1994), Shephard (1994), Kim and Nelson (1999), and the references therein).

Despite the attractive features of these alternative models, it does not necessarily follow from the foregoing discussion that one should automatically replace GARCH processes by either SV ones or empirically unrealistic ARCH(1) formulations in models with partially unobserved variables, especially when the degree of unobservability is small. In this respect, it is important to remember that many macro and finance theories are often specified using one-step-ahead moments, although those moments are defined with respect to the economic agents' information set, not the econometrician's. Hence, it is perhaps not surprising that economists have built, and continue to build, many models that involve latent GARCH processes in order to tackle a number of important empirical problems. Here we discuss two main categories of models that have been extensively used, together with some illustrative examples:

1. **Asset pricing models**

   (a) **Aggregate models**. Chou, Engle, and Kane (1992) proposed a time-varying parameter GARCH in mean model in which the slope coefficient in the linear relationship between the mean excess return on a stock market index and its variance changes over time according to a random walk. Specifically,

$$\begin{aligned} x_t &= b_t \lambda_t + f_t \\ b_t &= b_{t-1} + v_t \\ \lambda_t &= \theta + \beta \lambda_{t-1} + \alpha f_{t-1}^2. \end{aligned}$$

The problem is that f_t is not measurable with respect to the econometrician's information set $\mathcal{F}_{t-1}^x$ as long as the variance of v_t is positive, which means that the conditional variance of the observable process $\{x_t\}$ given its past $\mathcal{F}_{t-1}^x$ cannot be written in closed form. Both Chou, Engle, and Kane (1992) and Harvey, Ruiz, and Sentana (1992) proposed approximate maximum likelihood estimators to this problem, but the quality of their approximations remains unknown.

(b) **Multiasset models.** Starting with Diebold and Nerlove (1989) and King, Sentana, and Wadhwani (1994), N-dimensional multivariate dynamic heteroskedasticity models for asset returns have been developed on the basis of traditional factor structures. In this paper we will discuss in detail the case where we observe the return vector

$$x_t = cr_{ft} + w_t,$$

where the risk awarded factor r_{ft} follows a univariate GARCH-M process with conditional mean $\tau\lambda_t$ and variance λ_t, c is a vector of N factor loadings, and w_t is N-variate Gaussian white noise with diagonal covariance matrix Γ. These assumptions imply that the distribution of x_t conditional on $\mathcal{F}_{t-1}^{x,f}$ is $N(c\lambda_t\tau, \Sigma_t)$, where the conditional covariance matrix $\Sigma_t = cc'\lambda_t + \Gamma$ has the usual exact, single factor structure. For this reason, this model is called a *conditionally heteroskedastic factor* model. Although the common factor could be consistently recovered from asset return data as $N \to \infty$ (see Sentana (2002)), for a finite collection of assets r_{ft} is a latent GARCH-M process cloaked in the noise w_t. As a result, we cannot directly write down the likelihood function of the observable process $\{x_t\}$ regardless of the strength of the signal relative to the noise, except in the limiting cases in which $\{r_{ft}\}$ is fully revealed by $\{x_t\}$, or there are no ARCH effects. Papers which propose estimators of this model include Diebold and Nerlove (1989), Harvey, Ruiz, and Sentana (1992), Gourieroux, Monfort, and Renault (1993), Dungey, Martin, and Pagan (2000) and Calzolari, Fiorentini, and Sentana (2003). Nevertheless, none of them works out how to perform exact likelihood-based inference, which remains an unsolved problem. Further, it is previously not known how to estimate the density of $f|x, \varphi$ or its moments.

2. **Hidden equilibrium prices**

(a) **Bid/ask prices.** Many economic models assume an evolving equilibrium price that cannot be observed due to market imperfections. A classic example of this is the market microstructure work of Hasbrouck (1999), who models the bid and ask price

on the New York stock exchange (observed, say, every thirty minutes) as being the outcome of three continuous random variables: the equilibrium price (say F_t) and the costs of quote exposure on the bid (B_t) and ask sides (A_t). The bid and ask prices in the market are then formed by rounding, with the observed bid and ask prices being

$$b_t = Floor(F_t - B_t) \qquad \text{and} \qquad a_t = Ceiling(F_t + A_t).$$

In a simple model, the logs of F_t, B_t and A_t are assumed to be Gaussian autoregressions with i.i.d. innovations, although Hasbrouck would have liked to allow for GARCH effects in f_t ($=\ln F_t$), which he was unable to handle econometrically.

(b) **Target interest rates:** There are also examples in macroeconomics in which prices can only change in discrete amounts. For instance, Eichengreen, Watson, and Grossman (1985) and Dueker (1999) describe the interest rate setting behaviour of monetary policy authorities by means of ordered probits. In their models, the desired change in a target interest rate is given by a single equation of the form

$$\Delta x_t^* = \Delta z_{t-1}' \beta + f_t,$$

where z_t is a set of observable conditioning variables, and f_t follows a possibly dynamic heteroskedastic process. In this framework, the change in observed interest rates depends on the interval on which Δx_t^* lies. Although Dueker (1999) motivates his work by reference to ARCH processes, he only derives MCMC estimation methods for the case in which the volatility of f_t follows a discrete-state Markov chain. An elegant approach to likelihood inference for these types of situations is put forward by Lee (1999), who employs importance sampling to estimate the likelihood function. But apart from the difficulties posed by large sample sizes, the reliable use of importance sampling requires that the variance of the sampler is finite, a condition that has not yet been checked in this context (see Geweke (1989) and Koopman and Shephard (2003)). In addition, unlike MCMC procedures, importance sampling does not simultaneously generate estimates of the density of the latent variables.

(c) **Futures contracts.** Some cash markets and many futures markets often shut when prices move in one day beyond a prespecified maximum. This is discussed in theory by Brennan (1986) and econometrically by McCurdy and Morgan (1987), Kodres (1988), Kodres (1993), Morgan and Trevor (1999) and Wei (2002). We might think about this problem as being one where $\{x_t\}$ are the observed, potentially truncated, geometric returns while we think of $\{f_t\}$ as the corresponding daily (unobserved)

geometric return which we would have observed had there being no constraints. In effect, f_t measures the notional equilibrium return. Similar situations arise in target zone exchange rate agreements and commodity price support mechanisms. Given the empirical behaviour of future prices, it makes sense to have f_t following a GARCH process. But since in regulated markets f_t is not always observed, especially in volatile periods, ignoring the censoring will typically underestimate the true level of volatility in the market. The extent to which censoring occurs, however, varies greatly across markets and periods. For instance, until December 1997 price limits existed for coffee, cocoa and sugar futures traded on what is now the New York Board of Trade. But while between 1994 and 1997 the coffee limits were hit on 193 occasions, there were only 25 limit days for sugar, and 3 for cocoa (see Hall, Kofman, and Manaster (2001)). In addition, those limits were lifted in January 1998, and have not been reintroduced.

Since it would seem odd to change the model specification depending on the sample period and/or the asset at hand, it is clear that from an econometric viewpoint, it is important to study how to efficiently estimate models with partially unobserved GARCH processes. In this context, our main contribution is to level the playing field by showing that MCMC likelihood-based estimation of latent GARCH models can in fact be handled by means of feasible $O(T)$ algorithms. The crucial idea will be a novel transformation of the GARCH processes that makes them first-order Markovian. For the sake of concreteness, this will be developed within the context of the conditionally heteroskedastic factor model in which the common factors follow Quadratic GARCH processes (see Engle (1990) and Sentana (1995)), but applies much more widely. In fact, an important characteristic of our proposed algorithm is that it can be easily adapted to deal with other GARCH functional forms. In addition, the existence of leverage effects and/or risk premia components adds no substantive complications to our MCMC simulation scheme.

The structure of our paper is as follows. In section 2 we outline both classical and Bayesian likelihood approaches to inference for the conditionally heteroskedastic factor model. We show in both cases that the key task is to be able to produce simulators for $\{f_t\} \,|\, \{x_t\}, \varphi$, that is the factors given the data and the parameters. Section 3 explains how we exploit our Markov inducing transformation to design fast, and yet simple, algorithms to carry this out. We also assess the properties of our solution by use of a Monte Carlo experiment. This section is written in a self-contained manner, so that readers who are not interested in factor models can read its contribution directly in order to be able to apply it to other problems. An illustrative empirical application of the factor model to UK sectorial stock market returns is presented in section 4. Finally, our conclusions can be found in section 5.

2 Likelihood inference: EM and Bayesian approaches

2.1 A working example

In this section we discuss the basic likelihood framework for a conditionally heteroskedastic (CH) exact k-factor model with GARCH conditional variances. The complete model is

$$x_t = Cr_{ft} + w_t, \tag{1}$$

$$r_{ft} = \Lambda_t \tau + f_t, \tag{2}$$

and

$$\begin{pmatrix} f_t \\ w_t \end{pmatrix} | \mathcal{F}_{t-1}^{x,f} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Lambda_t & 0 \\ 0 & \Gamma \end{pmatrix} \right], \tag{3}$$

where f_t is a $k \times 1$ vector of unobserved common factors, C is the $N \times k$ matrix of factor loadings, with $N \geq k$ and $\text{rank}\,(C) = k$, Λ_t is a diagonal positive definite matrix of k time-varying factor variances, τ is a $k \times 1$ vector of factor risk prices,[1] and Γ is the diagonal matrix of the N constant conditional variances of the idiosyncratic terms w_t. In addition, we assume that each element of f_t, f_{jt} say, follows some univariate GARCH process whose conditional variance λ_{jt} depends on a set of unknown parameters ψ_j. In this respect, since (1) can be equivalently written as

$$x_t = \left(C\Lambda^{1/2}\right)\left(\Lambda^{-1/2} r_{ft}\right) + w_t,$$

where Λ is any $k \times k$ diagonal positive definite (p.d.) matrix of constant factor scales, we shall initially impose restrictions on ψ_j which guarantee that

$$\lambda_j = E(\lambda_{jt}) = V(f_{jt}) = 1 \quad (j = 1, \ldots, k)$$

in order to eliminate such a scale indeterminacy. We will return to this issue in section 2.4.2.[2] Throughout the parameters of interest are $\varphi' = (c', \gamma', \delta')$, where $c = vec(C') = (c_1', \ldots, c_N')'$, $c_i' = (c_{i1}, \ldots, c_{ik})$, $\gamma = vecd(\Gamma) = (\gamma_1, \ldots, \gamma_N)'$, $\delta = (\delta_1', \ldots, \delta_k')'$, and $\delta_j = (\tau_j, \psi_j)$. In the next subsections we discuss the basics of the simulated EM algorithm and develop the corresponding Bayesian approach. But before, we briefly discuss the role of some of our assumptions.

First of all, the normality assumption is less crucial than it may seem, and could be replaced by general location-scale mixtures of normals at the cost of introducing additional state variables (see e.g. Chib, Nardari, and Shephard (2002))). Likewise, the mutual independence of

[1] See King, Sentana, and Wadhwani (1994) for conditions that justify the specific functional form of expected returns.

[2] If the unconditional variance is unbounded, as in integrated GARCH-type models, other symmetric scaling assumptions can be made. For instance, we could choose $\inf_t \lambda_{jt} = 1$, or simply fix to 1 the constant element of λ_{jt}. Alternatively, we can use the asymmetric scaling approach described in section 2.4.2 below, which fixes c_{ii} to 1 for $i = 1, \ldots, k$ on the assumption that $c_{ii} \neq 0$. In any case, note that in principle there is no need to set to zero the strict upper triangle of the factor loading matrix C in view of the identification results in Sentana and Fiorentini (2001).

the stochastic processes for the common factors could in principle be replaced by some k-variate GARCH formulation, although the introduction of correlation between the factors would not only increase the computational burden, but may also affect the identifiability of the matrix of factor loadings C, as discussed by Sentana and Fiorentini (2001). As for the diagonality of the idiosyncratic covariance matrix, it would certainly be more attractive to consider the approximate factor structures introduced by Chamberlain and Rothschild (1983) because of their robustness to portfolio formation. However, proposing plausible parametrisations for such structures is beyond the scope of this paper. Moreover, the alternative assumption of a fully unrestricted but constant Γ would not only alter the usual interpretation of the factors as common risk influences, but it would also seriously impair the identifiability of the model parameters (see Sentana and Fiorentini (2001) and Doz and Renault (2001)). Finally, the most restrictive assumption is arguably the constancy of the idiosyncratic variances. As we shall see in section 2.4.1, though, the main purpose of this assumption is to allow us to cast the multivariate CH model (1), (2) and (3) in the univariate framework of section 3, with the added benefit that the signal to noise ratio is identified from the cross-sectional dimension.

2.2 Simulated EM algorithm

Suppose for a moment that both the returns x, and the risk awarded factors $r_f = \left(r'_{f1}, ..., r'_{fT}\right)'$ are observed. Then, we can write down the log-likelihood function of $x, r_f|\varphi, \mathcal{F}_0$, as

$$-\frac{TN}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\ln|\gamma_i| + \frac{(x_{it} - c_i' r_{ft})^2}{\gamma_i}\right\} \tag{4}$$

$$-\frac{Tk}{2}\ln 2\pi - \frac{1}{2}\sum_{j=1}^{k}\left\{\sum_{t=1}^{T}\left[\ln|\lambda_{jt}(\delta_j)| + \frac{[r_{fjt} - \tau_j\lambda_{jt}(\delta_j)]^2}{\lambda_{jt}(\delta_j)}\right]\right\}. \tag{5}$$

Given our parametrisation of factors as univariate GARCH processes, such a factorisation performs a sequential cut on the joint log-likelihood function of x_t, r_{ft}, which makes r_{ft} strongly exogenous for c and γ (see Engle, Hendry, and Richard (1983)). As a result, since c and γ only enter through (4), their unrestricted MLE's could be obtained from N univariate OLS regressions of each x_{it} $(i = 1, \ldots, N)$ on r_{ft}. In addition, ML estimates of the conditional variance parameters ψ_j and price of risk coefficients τ_j could be obtained from k univariate dynamic heteroskedasticity in mean models for r_{fjt} $(j = 1, \ldots, k)$.

Unfortunately, the $r'_{ft}s$ are generally unobserved. Nevertheless, the previous discussion suggests using the EM algorithm of Dempster, Laird, and Rubin (1977) as a convenient way to obtain estimates of φ as close to the optimum as desired. An elegant review of this method is given by Ruud (1991). At each iteration, the EM algorithm obtains $\varphi^{(n+1)}$ by maximising

the expectation of the complete log-likelihood given by equations (4) and (5) conditional on the data and the current parameter values, i.e. $E\left\{\ln p(r_f, x|\varphi, \mathcal{F}_0)|x, \varphi^{(n)}, \mathcal{F}_0\right\}$, with respect to φ keeping $\varphi^{(n)}$ fixed. When we do this it is helpful to write $r^{(n)}_{ft|x} = E\left(r_{ft}|x, \varphi^{(n)}\right)$ and $\Omega^{(n)}_{t|x} = V\left(r_{ft}|x, \varphi^{(n)}\right)$. For the moment, we abstract from the fact that such quantities are not easy to calculate. Then, the expectation of the complete likelihood is

$$-\frac{TN}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\ln|\gamma_i| + \frac{\left(x_{it} - c_i' r^{(n)}_{ft|x}\right)^2 + c_i'\Omega^{(n)}_{t|x}c_i}{\gamma_i}\right\} \tag{6}$$

$$-\frac{Tk}{2}\ln 2\pi - \frac{1}{2}\sum_{j=1}^{k}\sum_{t=1}^{T} E\left\{\ln|\lambda_{jt}(\delta_j)| + \frac{[r_{fjt} - \tau_j\lambda_{jt}(\delta_j)]^2}{\lambda_{jt}(\delta_j)}\middle| x, \varphi^{(n)}\right\}, \tag{7}$$

where $^{(n)}$ refers to expressions evaluated at $\varphi^{(n)}$. Such a factorisation is particularly useful for the Maximisation step of the EM algorithm, for C and Γ, which are typically high dimensional objects, only enter through (6). Specifically:

$$c_i^{(n+1)} = \left\{\sum_{t=1}^{T}\left(r^{(n)}_{ft|x}r'^{(n)}_{ft|x} + \Omega^{(n)}_{t|x}\right)\right\}^{-1}\left\{\sum_{t=1}^{T} r^{(n)}_{ft|x}x_{it}\right\}, \tag{8}$$

$$\gamma_i^{(n+1)} = \frac{1}{T}\sum_{t=1}^{T}\left\{(x_{it} - c_i'^{(n+1)}r^{(n)}_{ft|x})^2 + c_i'^{(n+1)}\Omega^{(n)}_{t|x}c_i^{(n+1)}\right\}. \tag{9}$$

This is very helpful for it means the M-step is extremely easy to compute for the vast majority of the parameter space. Similarly, $\delta_j^{(n+1)}$can be obtained by maximizing (7) for $j = 1, \ldots, k$. But since no closed expression exists in general, one has to resort to numerical optimization methods. This is the price to be paid for modelling the time variation in the $\lambda'_{jt}s$, but can be regarded as a sunk cost which is completely independent of the number of series under consideration. In fact, a very large number of series constitutes a computational blessing in this framework, because for large N the unobservable factors can be consistently estimated (see Sentana (2002)), and the model effectively becomes a multivariate regression model, plus k univariate conditionally heteroskedastic ones. Furthermore, it is not really necessary to maximize (7) at each EM iteration if it is too costly relative to maximizing (6). In practice, we can do a few iterations over (8) and (9) alone, before maximizing (7) (see Demos and Sentana (1998)).

Nevertheless, it is well known (e.g. Tanner (1996, p. 84-85)) that the EM algorithm slows down significantly in the neighbourhood of the optimum. As a result, after some initial EM iterations it is tempting to switch to a derivative based optimisation routine, which is more likely to quickly converge to the maximum. EM type arguments can be used to facilitate this switch by allowing the computation of the score. In particular, it is easy to see that

$$E\left\{\frac{\partial \ln p(r_f|x, \varphi, \mathcal{F}_0)}{\partial\varphi}|x, \varphi, \mathcal{F}_0\right\} = 0,$$

so it is clear that the score can be obtained as the expected value (given x, φ and $\mathcal{F}_0$) of the sum of the unobservable scores corresponding to $\ln p(x|r^f, \varphi, \mathcal{F}_0)$ and $\ln p(r^f|\varphi, \mathcal{F}_0)$. This result was first noted by Louis (1982); see also Ruud (1991) and Tanner (1996, p. 84).

All the above algorithms are infeasible for we do not know how to analytically compute the required conditional expectations. The approach we follow here is to replace each of the expectations by averages of simulations drawn from $r^f|x, \varphi^{(n)}, \mathcal{F}_0$. In section 3 we will show how to carry out this simulation, the major contribution of the paper. The replacement of expectations by averages of simulations means that the algorithms described in the previous two subsections will become a simulated EM algorithm and a simulated score one, respectively. These approaches to computing the maximum likelihood estimator are discussed in the statistics literature by Celeux and Diebolt (1985) and Tanner (1996). They have also appeared in econometrics on a number of occasions. Prominent references include Bresnahan (1981), Hajivassiliou and McFadden (1998), Nielsen (2000), Ruud (1991) and Shephard (1993).

2.3 Simulation-based Bayesian inference

The task of simulating from $r_f|x, \varphi$ also appears in the Bayesian analysis of this model. Recall that in our problem the key issue is that the likelihood function of the sample

$$p(x|\varphi, \mathcal{F}_0) = \int p(x|r_f, \varphi, \mathcal{F}_0) p(r_f|\varphi, \mathcal{F}_0) \mathrm{d}r_f$$

is intractable, which precludes the direct analysis of the posterior density $p(\varphi|x, \mathcal{F}_0)$. This problem can be overcome by focusing instead on the density

$$p(\varphi, r_f|x, \mathcal{F}_0) \propto p(x|r_f, \varphi, \mathcal{F}_0) \cdot p(r_f|\varphi, \mathcal{F}_0) \cdot p(\varphi|\mathcal{F}_0).$$

We will aim to sample from this joint density, for we can then discard the r_f draws yielding a sample from the posterior $p(\varphi|x, \mathcal{F}_0)$. Assuming independent priors between static and dynamic variance parameters, our suggestion is to carry this out in some natural blocks.

1. Initialise φ.
2. Update draw from $p(r_f|\varphi, x, \mathcal{F}_0)$.
3. Update draw from $p(\varphi|r_f, x, \mathcal{F}_0)$ in the following blocks:
   (a) Update $p(c, \gamma|x, r_f)$. This is updated in a block.
   (b) Update $p(\delta_j|\{r_{fjt}\}, \mathcal{F}_0)$ for $j = 1, ..., k$.
4. Goto 2.

Step 3b is the task of simulating from the posterior of the parameters of k univariate ARCH-M processes. This has already been briefly addressed by Kim, Shephard, and Chib (1998), and at more length later by Bauwens and Lubrano (1998) and Nakatsuma (2000). Similarly, the tasks in 3a are those which appear in standard regression models. All that remains therefore is Step 2, sampling from $r_f|\varphi, x, \mathcal{F}_0$.

The algorithm above is a special case of a MCMC algorithm, which converges, as it iterates, to draws from the required density $p(\varphi, r_f|x, \mathcal{F}_0)$. It should be kept in mind that sample variates from a MCMC algorithm are a high-dimensional (correlated) sample from the target density of interest. These draws can be used as the basis for making inferences by appealing to suitable ergodic theorems for Markov chains. For example, posterior moments and marginal densities can be estimated (simulation consistently) by averaging the relevant function of interest over the sampled variates. The posterior mean of φ is simply estimated by the sample mean of the simulated φ values. These estimates can be made arbitrarily accurate by increasing the simulation sample size. The accuracy of the resulting estimates (the so called numerical standard error) can be assessed by standard time series methods that correct for the serial correlation in the draws. Unfortunately, the serial correlation can be quite high for badly behaved algorithms.

MCMC methods, namely the Metropolis-Hastings and Gibbs sampling algorithms, have had a widespread influence on the theory and practice of Bayesian inference. Early work on these methods appears in Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953), Hastings (1970), Ripley (1977) and Geman and Geman (1984), while some of the more recent developments, spurred by Tanner and Wong (1987) and Gelfand and Smith (1990), are included in Gilks, Richardson, and Spiegelhalter (1996), Tanner (1996, Ch. 6) and Chib (2001).

2.4 Implementation details

2.4.1 The sufficient statistics for $\{r_{ft}\}\,|\,\{x_t\}, \varphi, \mathcal{F}_0$

As we have already seen, whether we follow a classical or a Bayesian approach to estimation, our main task is to simulate from $r_f|\varphi, x, \mathcal{F}_0$. This is not straightforward. One of the challenges of working with the CH factor model is that the cross-sectional dimension of vector of returns x_t can be rather large. Nevertheless, enormous computational savings can be made by realizing that there are k stochastic processes which suffice to represent $\{x_t\}$ in the simulation steps. Specifically, assume for simplicity that Γ is non-singular, and define the following full rank transformation of the observed series x_t:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} = \begin{pmatrix} \Upsilon C'\Gamma^{-1} \\ U_1'\left[I_N - C\Upsilon C'\Gamma^{-1}\right] \end{pmatrix} x_t = \begin{pmatrix} I_k \\ 0 \end{pmatrix} r_{ft} + \begin{pmatrix} \Upsilon C'\Gamma^{-1} \\ U_1'\left[I - C\Upsilon C'\Gamma^{-1}\right] \end{pmatrix} w_t,$$

where $\Upsilon = \left(C'\Gamma^{-1}C\right)^{-1}$, and $U_1\Delta_1U_1'$ provides the spectral decomposition of the rank $N-k$ matrix $\Gamma - C(C'\Gamma^{-1}C)^{-1}C'$. Note that if we think of the CH model as a cross-sectional heteroskedastic regression of x_{it} on the "regressors" c_i' with regression parameters r_{ft} and residual variances γ_i, then y_t corresponds to the generalised least squares (GLS) estimator of r_{ft},[3] while z_t contains the first $N-k$ principal components of the GLS residuals $x_t - Cy_t$. Given that

$$\begin{aligned} y_t &= r_{ft} + \Upsilon C'\Gamma^{-1}w_t = r_{ft} + \eta_t, \\ \begin{pmatrix} r_{ft} \\ \eta_t \\ z_t \end{pmatrix} &\mid \mathcal{F}_{t-1}^{x,f}, \varphi \sim N\left[\begin{pmatrix} \Lambda_t\tau \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Lambda_t & 0 & 0 \\ 0 & \Upsilon & 0 \\ 0 & 0 & \Delta_1 \end{pmatrix}\right], \end{aligned}$$

(see Gourieroux, Monfort, and Renault (1991)), and that Λ_t is a function of lag values of r_{ft} only, it is clear that $\{y_t\}$ contains the same information about $\{r_{ft}\}$ as $\{x_t\}$. In addition, this means that the degree of unobservability of the common factors depends exclusively on the magnitude of Υ relative to Λ.

Therefore, the key remaining issue in MCMC is to how deal with this "compressed" model structure. In particular, we focus on simulating from

$$p(r_f|x,\varphi,\mathcal{F}_0) = p(r_f|y,\varphi,\mathcal{F}_0) \propto p(y|r_f,\varphi)p(r_f|\varphi,\mathcal{F}_0).$$

For simplicity of exposition we will assume a single factor hereinafter, as the extension to the multifactor case is tedious but straightforward. Thus the model we study will be

$$y_t = r_{ft} + \eta_t \qquad r_{ft} = \lambda_t\tau + f_t \qquad f_t = \varepsilon_t\lambda_t^{1/2}.$$

For the same reason, we will assume a Generalised Quadratic ARCH structure of orders 1,1 - or GQARCH(1,1) for short - for the conditional variance

$$\lambda_t = \theta + \beta\lambda_{t-1} + \alpha\left(f_{t-1} - \mu\right)^2, \tag{10}$$

where the dynamic asymmetry parameter μ is usually different from 0, allowing for the possibility of a leverage effect (see Engle (1990) and Sentana (1995)). The special case of $\mu = 0$ gives the GARCH(1,1) model, while the additional assumption of $\beta = 0$ gives ARCH(1). Given that this process is covariance stationary if $\alpha + \beta < 1$, we can define the unconditional signal to noise ratio of the innovations in y_t as λ/υ, where

$$\lambda = \frac{\theta + \alpha\mu^2}{1-\beta-\alpha}, \quad \upsilon = \left(c'\Gamma^{-1}c\right)^{-1}.$$

In section 3 we study different ways of generating non-independent draws from the conditional distribution of $r_f|y,\varphi$. As we shall see, the value of λ/υ will be one of the most important determinants of the performance of the different simulators.

[3] These factor mimicking portfolios are usually known as Barlett scores in the multivariate statistical analysis literature (see e.g. Sentana (2002)).

2.4.2 Scaling and related parametrisation issues

We have assumed so far that the scale indeterminacy of the common factors is resolved by restricting their unconditional variance to be 1. In particular, in the single factor GQARCH(1,1) model we can set $\theta = (1-\beta-\alpha) - \alpha\mu^2$, which implies $\lambda = 1$ and $\theta \geq 0$ as long as $\beta + \alpha < 1$ and μ is not too large. More specifically, we use the re-parametrisation $\alpha + \beta = \psi_1 = \sin^2(\psi_1^*)$ and $\beta(\alpha+\beta)^{-1} = \psi_2 = \sin^2(\psi_2^*)$ to guarantee $0 \leq \beta \leq 1-\alpha \leq 1$. In addition, we set $\mu = [(1-\alpha-\beta)/\alpha]^{1/2}\sin(\psi_3^*)$ to ensure $\theta \geq 0$ and $\lambda = 1$. However, the performance of the Gibbs sampler can be very sensitive to the chosen normalisation (for further discussion, see Pitt and Shephard (1999a) and section 4 below). For that reason, we have also considered an alternative asymmetric scaling assumption that sets the value of a particular factor loading, c_1 say, to 1, and unrestricts the unconditional variance parameter λ (as in Aguilar and West (2000), Chib, Nardari, and Shephard (2002), or Pitt and Shephard (1999b)).[4] Although such a scaling assumption does not constitute a proper, symmetric normalisation, as it requires the additional assumption that $c_1 \neq 0$, it solves the sign indeterminacy that would otherwise exist (cf. Geweke and Zhou (1996)). Once the drawings from this alternative parametrisation have been obtained, though, we transform them back to make them comparable with the earlier one.

From a practical point of view, fixing one factor loading to 1 introduces two main differences. First, the log-likelihood function of the limiting factor representing portfolio r_{ft} will depend on an extra parameter, λ. And second, the posterior distributions of the factor loading parameter and idiosyncratic variance corresponding to the first element of the vector x_t have to be modified appropriately to reflect the fact that the prior distribution of c_1 is now degenerate. It turns out, though, that in the single factor case all we have to do is to apply standard Bayesian inference procedures to the variance of $(x_{1t} - r_{ft})$ under the assumption that its mean is zero. It is also straightforward to modify the EM algorithm so as to impose the restriction $c_1^{(n+1)} = 1\ \forall n$. First, note that (9) is unaffected. Second, since Γ is diagonal, expression (8) remains valid for rows $2, \ldots, N$. Finally, the unconditional variance parameter λ will enter in (7) multiplying $\lambda_t(\delta)$.

3 MCMC simulation of $\{f_t\} \,|\, \{y_t\}, \varphi$

It is clear from the discussion in section 2 that the key econometric issue in carrying out likelihood-based inference for CH factor models and other latent GARCH structures is to be able to effectively simulate from $\{f_t\} \,|\, \{y_t\}, \varphi$, where the relationship between the latent stochastic

[4] In this respect, note that while γ_i, α and β are invariant to scale, the same is not true of c_i, μ or τ, which must be multiplied, divided and multiplied by $\sqrt{\lambda}$, respectively.

process $\{f_t\}$ and the observed process $\{y_t\}$ is described by a dynamic model of the form:

$$y_t = r_{ft} + \eta_t, \quad r_{ft} = \lambda_t \tau + f_t, \quad f_t = \varepsilon_t \lambda_t^{1/2},$$
$$\lambda_t = \theta + \beta \lambda_{t-1} + \alpha \left(f_{t-1} - \mu\right)^2,$$

where

$$\begin{pmatrix} \eta_t \\ \varepsilon_t \end{pmatrix} \overset{i.i.d.}{\sim} N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \upsilon & 0 \\ 0 & 1 \end{pmatrix} \right\}.$$

This section deals with this issue in detail, and represents the main contribution of the paper. The first subsection discusses a simple but computationally expensive approach, which is in fact infeasible in many cases of practical interest, while the other subsections develop computationally feasible algorithms.[5]

3.1 A simple but extremely expensive $O(T^2)$ algorithm

The task is to simulate from $p(f|y,\varphi)$. Now sampling from

$$\begin{aligned} p(f|y,\varphi) &\propto p(y|f,\gamma)p(f|\varphi) \\ &= \prod_{t=1}^{T} \left\{ \frac{1}{\sqrt{\upsilon}} \phi\left(\frac{y_t - \tau\lambda_t - f_t}{\sqrt{\upsilon}} \right) \right\} \cdot \left\{ \frac{1}{\sqrt{\lambda_t}} \phi\left(\frac{f_t}{\sqrt{\lambda_t}} \right) \right\}, \end{aligned}$$

is entirely feasible by using a Metropolis-Hastings algorithm. In particular, let us write the r^{th} iteration of a Markov chain as f^r. Then we generate a potential new value of the Markov chain f^{new} by proposing it from some candidate density $g(f|f^r, y, \varphi)$, which we accept with probability

$$\min\left[1, \frac{p(f^{new}|y,\varphi)}{p(f^r|y,\varphi)} \frac{g(f^r|f^{new},y,\varphi)}{g(f^{new}|f^r,y,\varphi)}\right].$$

If it is accepted then we set $f^{r+1} = f^{new}$, otherwise we keep $f^{r+1} = f^r$. This procedure is iterated and will generate an ergodic Markov chain with equilibrium distribution $p(f|y,\varphi)$ so long as $g(f|f^r,y,\varphi) > 0$ for all f (e.g. Tierney (1994) and Chib (2001)).

There are many potential candidate choices for $g(f|f^r,y,\varphi)$. For instance, we could use an independent wholemove sampler which proposes f from the Kalman filter approximation to $p(f|y,\varphi)$ put forward by Harvey, Ruiz, and Sentana (1992) (HRS). Unfortunately, the dimension of the state vector f is so huge that it is likely that each choice will be rather poor, unless, of course, the correct distribution $p(f|y,\varphi)$ were known, as in the conditionally homoskedastic Gaussian case discussed by Geweke and Zhou (1996). As a result, the MCMC algorithm will

[5] In this respect, it is important to note that for a given set of parameter values and initial conditions, it is generally simpler to simulate f_t for $t = 1, \ldots, T$ and then compute $r_{ft} = \lambda_t\tau + f_t$, than to simulate r_f directly. For that reason, we concentrate on simulators of f_t given y and φ. Importantly, we systematically set the mean and variance of the initial factor f_1 to their unconditional values of 0 and λ respectively. Given that λ_t is a sufficient statistic for $\mathcal{F}_{t-1}^f$, and that λ is a deterministic function of φ, $\mathcal{F}_0$ can thus be eliminated from the conditioning set without information loss.

hardly ever accept a proposal, generating unacceptably large dependence in the chain. A conventional MCMC strategy for overcoming this problem is to update only a subset of the required elements. In particular, Shephard (1996) suggested updating only one f_t at a time, leaving all the others unchanged (see also Wei (2002)). In this context, if we propose from $g(f_t|f^r_{\backslash t}, y, \varphi)$, where $f^r_{\backslash t} = \{f_1^{r+1}, \ldots, f_{t-1}^{r+1}, f_{t+1}^r, \ldots, f_T^r\}$, the acceptance rate will be

$$\min\left[1, \frac{p(f_t^{new}|f^r_{\backslash t}, y, \varphi)}{p(f_t^r|f^r_{\backslash t}, y, \varphi)} \frac{g(f_t^r|f^r_{\backslash t}, y, \varphi)}{g(f_t^{new}|f^r_{\backslash t}, y, \varphi)}\right],$$

since

$$p(f|y, \varphi) = p\left(f_{\backslash t}|y, \varphi\right) p(f_t|f_{\backslash t}, y, \varphi).$$

The proposal is now much better as it is only in a single dimension, but unfortunately, each time we consider modifying a single factor we have to compute

$$\frac{p(f_t^{new}|f^r_{\backslash t}, y, \varphi)}{p(f_t^r|f^r_{\backslash t}, y, \varphi)} = \frac{p(y_t|f_t^{new}, \lambda_t^{new,t}, \varphi)p(f_t^{new}|\lambda_t^{new,t}, \varphi)}{p(y_t|f_t^r, \lambda_t^{r,t}, \varphi)p(f_t^r|\lambda_t^{r,t}, \varphi)} \prod_{s=t+1}^{T} \frac{p(y_s|f_s^r, \lambda_s^{new,t}, \varphi)p(f_s^r|\lambda_s^{new,t}, \varphi)}{p(y_s|f_s^r, \lambda_s^{r,t}, \varphi)p(f_s^r|\lambda_s^{r,t}, \varphi)},$$

where for $s = t+1, \ldots, T$

$$\begin{aligned} \lambda_s^{new,t} &= V(f_s|f_{s-1}^r, f_{s-2}^r, \ldots, f_{t+1}^r, f_t^{new}, f_{t-1}^{r+1}, \ldots, f_1^{r+1}), \\ \lambda_s^{r,t} &= V(f_s|f_{s-1}^r, f_{s-2}^r, \ldots, f_{t+1}^r, f_t^r, f_{t-1}^{r+1}, \ldots, f_1^{r+1}), \end{aligned}$$

while $\lambda_t^{new,t} = \lambda_t^{r,t}$. So even if we propose f_t^{new} from the conditional distribution of f_t given y_t, $\lambda_t^{r,t}$ and φ to simplify the acceptance rate slightly, unless $\beta = 0$, or indeed $\alpha = 0$, doing so requires $O(T-t)$ univariate normal density evaluations in view of the recursive nature of the GQARCH process in (10), which makes $\lambda_s^{new,t}$ to depend upon f_t^{new} for $s = t+1, \ldots, T$. As this cost would have to be borne T times in an MCMC sweep through all the elements of f_t, this algorithm is $O(T^2)$ for each value of φ, and so is generally impractical for the sort of large values of T characteristic of most financial time series applications.

3.2 Sampling the factors in $O(T)$ operations

3.2.1 Markov transformations of GQARCH

An alternative approach is to work with a transformation of the latent GQARCH process that becomes first-order Markov. The simplest way to achieve this is to augment the factor with the conditional variance λ_{t+1}, and then sample the joint Markov process $\{f_t, \lambda_{t+1}\}$ given y and φ. To see the validity of this argument, we recall that

$$\lambda_{t+1} = \theta + \beta\lambda_t + \alpha\left(f_t - \mu\right)^2, \quad f_t = \sqrt{\lambda_t}\varepsilon_t.$$

Notice that the time-shift between both variables is only apparent, as $f_t, \lambda_{t+1} \in \mathcal{F}_t^f$. A technical difficulty with this approach is that the joint distribution of $\{f_t, \lambda_{t+1}\}$ is singular, which makes sampling slightly complicated as various Jacobian terms enter the acceptance rate of the Metropolis-Hastings algorithm (see Spivak (1965, chap. 5)). This type of algorithm is discussed in some detail in section 3.2.3. But an equivalent approach which avoids the singularities is to map from the factors $\{f_t\}$ into a different set of Markov variables: the conditional variances λ_{t+1} and their "signs" s_t, where $s_t = sign\,(f_t - \mu)$, so that $s_t = \pm 1$ with probability 1, because $s_t = 0$ is a zero-probability event. Although the resulting transformation involves working with a partly discrete distribution, the mapping is one-to-one with no singularities. Specifically, if we know $\{\lambda_{t+1}\}$ and φ (and consequently λ), then we know the value of

$$(f_t - \mu)^2 = \frac{\lambda_{t+1} - \theta - \beta\lambda_t}{\alpha}, \quad \forall t \geq 1.$$

Hence the additional knowledge of the signs of $(f_t - \mu)$ would reveal the entire path of $\{f_t\}$ so long as $\lambda_1 (= \lambda$ in our case) is known. Given that this second transformation also shares the Markov nature of $\{f_t, \lambda_{t+1}\}$, we can easily design MCMC algorithms which sample $\{s_t, \lambda_{t+1}\}$, and hence the factors, in $O(T)$ flops.

The MCMC algorithm samples from the conditional distribution of $\{s_t, \lambda_{t+1}\}$ given y and φ:

$$p(\{s_t, \lambda_{t+1}\}|y;\varphi) \propto \prod_{t=1}^{T} p(\lambda_{t+1}|\lambda_t;\varphi)p(s_t|\lambda_{t+1},\lambda_t;\varphi)p(y_t|s_t,\lambda_{t+1},\lambda_t;\varphi),$$

where the first and third terms come straight from the model, and the second one from the fact that while $f_t|\mathcal{F}_{t-1}^f \sim N(0,\lambda_t)$, $f_t|\lambda_{t+1},\mathcal{F}_{t-1}^f$ can only take the values $\mu \pm d_t$, with

$$d_t = \sqrt{(\lambda_{t+1} - \theta - \beta\lambda_t)/\alpha}.$$

Such a factorisation shows that we do not alter the volatility process when we flip from $s_t = -1$ to $s_t = 1$ (implying that signs do not cause the volatility process), but we do alter f_t. Hence, we can first simulate the Markovian process $\{\lambda_{t+1}\}$ given y and φ, and then go on to simulate $\{s_t\}$ from its distribution conditional on $\{\lambda_{t+1}\}$, $\{y_t\}$ and φ. Note that the second step is effectively a Gibbs sampling scheme whose acceptance rate is always 1. In addition, conditional on $\{\lambda_{t+1}\}$, $\{y_t\}$ and φ, the elements of $\{s_t\}$ are independent, which further simplifies the calculations.

In the next subsections, we shall discuss simulators of $\{\lambda_{t+1}\}$, but first, we shall explain in detail how to simulate s_t. To obtain the required conditionally Bernoulli distribution, it is helpful to establish some notation. Let us write

$$\begin{aligned} c_{t+1} &= \frac{1}{\sqrt{\omega_{t|y_t,\lambda_t}}}\left[\phi\left(\frac{\mu + d_t - f_{t|y_t,\lambda_t}}{\sqrt{\omega_{t|y_t,\lambda_t}}}\right) + \phi\left(\frac{\mu - d_t - f_{t|y_t,\lambda_t}}{\sqrt{\omega_{t|y_t,\lambda_t}}}\right)\right], \\ f_{t|y_t,\lambda_t} &= E\left(f_t|y_t,\lambda_t,\varphi\right) = \frac{\lambda_t}{\lambda_t + \upsilon}(y_t - \tau\lambda_t) = \frac{\omega_{t|y_t,\lambda_t}}{\upsilon}(y_t - \tau\lambda_t), \end{aligned}$$

and

$$\omega_{t|y_t,\lambda_t} = V\left(f_t|y_t,\lambda_t,\varphi\right) = \left(\lambda_t^{-1} + \upsilon^{-1}\right)^{-1}.$$

Since, conditional on $\{\lambda_{t+1}\}$, the probability that s_t is 1 is the same as the probability that $f_t = d_t + \mu$, then

$$p\left(s_t = 1|\left\{\lambda_{t+1}\right\},\left\{y_t\right\},\varphi,\lambda_1\right) = p\left(f_t = d_t + \mu|\lambda_{t+1},\lambda_t,y_t,\varphi\right) = \frac{1}{c_t\sqrt{\omega_{t|y_t,\lambda_t}}}\phi\left(\frac{\mu + d_t - f_{t|y_t,\lambda_t}}{\sqrt{\omega_{t|y_t,\lambda_t}}}\right).$$

This is extremely cheap to compute. Further, in the special case of $\mu = 0$, we can exploit the fact that

$$p(f_t|\lambda_t,y_t,\varphi) = \frac{p(y_t|f_t,\lambda_t,\varphi)p(f_t|\lambda_t,\varphi)}{p(y_t|\lambda_t,\varphi)} = \frac{1}{\sqrt{\omega_{t|y_t,\lambda_t}}}\phi\left(\frac{f_t - f_{t|y_t,\lambda_t}}{\sqrt{\omega_{t|y_t,\lambda_t}}}\right), \tag{11}$$

and the symmetry of the normal distribution to prove that we can alternatively draw $s_t = +1$ with probability

$$\frac{\phi\left[\upsilon^{-1/2}\left(y_t - \tau\lambda_t - d_t\right)\right]}{\phi\left[\upsilon^{-1/2}\left(y_t - \tau\lambda_t - d_t\right)\right] + \phi\left[\upsilon^{-1/2}\left(y_t - \tau\lambda_t + d_t\right)\right]}$$

without affecting the results.

3.2.2 Single move samplers of $\{\lambda_{t+1}\}$

In this section, we concentrate on updating one single λ_{t+1} at a time, leaving all the others unchanged. In general, if we draw λ_{t+1}^{new} from an arbitrary proposal density $g(\lambda_{t+1}|\lambda^r_{\backslash t+1},y,\varphi)$, where $\lambda^r_{\backslash t+1} = \{\lambda_2^{r+1},\ldots,\lambda_t^{r+1},\lambda_{t+2}^r,\ldots,\lambda_{T+1}^r\}$, then the MH acceptance rate would be

$$\min\left[1, \frac{p(\lambda_{t+1}^{new}|\lambda^r_{\backslash t+1},y,\varphi)}{p(\lambda_{t+1}^r|\lambda^r_{\backslash t+1},y,\varphi)}\frac{g(\lambda_{t+1}^r|\lambda^r_{\backslash t+1},y,\varphi)}{g(\lambda_{t+1}^{new}|\lambda^r_{\backslash t+1},y,\varphi)}\right].$$

Such an acceptance rate turns out to be exactly the same as in the popular discrete time version of the log-normal stochastic volatility process (see e.g. Kim, Shephard, and Chib (1998)). In fact, the similarity is not merely coincidental, because latent GARCH models are effectively parameter driven processes, as opposed to standard GARCH models, which are observation driven ones (see Andersen (1994) and Shephard (1996)). Nevertheless, there are important differences in the distribution of λ_t between the two cases. In particular, since the support of the conditional distribution of λ_{t+1} given λ_t is bounded from below by $\theta + \beta\lambda_t$, and the same applies to the distribution of λ_{t+2} given λ_{t+1}, the range of values of λ_{t+1} compatible with λ_{t+2} and λ_t in the GQARCH case is bounded from above and below. Specifically, the lower limit corresponds to $d_t = 0$, and the upper limit to $d_{t+1} = 0$. Therefore, in practice it makes sense to make the proposal to obey the support of the density. The crucial thing, though, is that we can

quickly evaluate

$$\begin{aligned} p(\lambda_{t+1}|\lambda_{\setminus t+1}, y, \varphi) &\propto p(\lambda_{t+2}|\lambda_{t+1}, \varphi)p(\lambda_{t+1}|\lambda_t, \varphi)p(y_{t+1}|\lambda_{t+2}, \lambda_{t+1}, \varphi)p(y_t|\lambda_{t+1}, \lambda_t, \varphi), \\ \lambda_{t+1} &\in [\theta + \beta\lambda_t, \beta^{-1}(\lambda_{t+2} - \theta)], \end{aligned}$$

where $p(y_t|\lambda_{t+1}, \lambda_t, \varphi)$ is a mixture of two univariate normal densities, and $p(\lambda_{t+1}|\lambda_t, \varphi)$ is a scaled non-central chi-squared distribution with a single degree of freedom.[6]

There are many ways in which we can carry out MCMC on $p(\lambda_{t+1}|\lambda_{\setminus t+1}, y, \varphi)$. At first sight, it is tempting to simplify the acceptance rate by proposing λ_{t+1} from $p(\lambda_{t+1}|\lambda_t, \varphi)$, appropriately truncated from above, since the lower truncation will be automatically satisfied. However, such a proposal would be ignoring the information in y_{t+1}. Our experience suggests that the conditional distribution of λ_{t+1} can be radically modified by incorporating the information in the observed series, especially when the signal to noise ratio λ/v is high. For that reason, we can achieve a substantially higher acceptance rate by proposing λ_{t+1} from $p(\lambda_{t+1}|y_t, \lambda_t, \varphi)$. A numerically efficient way to simulate λ_{t+1} from this distribution is to sample an underlying Gaussian random variable $f_t^{new} \sim N(f_{t|y_t,\lambda_t}, \omega_{t|y_t,\lambda_t})$ doubly truncated so that it remains within the interval $\mu \pm l_t$, where

$$l_t = \sqrt{\frac{\lambda_{t+2} - \theta(1+\beta) - \beta^2\lambda_t}{\alpha\beta}},$$

and then compute

$$\lambda_{t+1}^{new} = \theta + \beta\lambda_t + \alpha\left(f_t^c - \mu\right)^2.$$

Note that the truncation implicitly guarantees a real value for

$$d_{t+1}^{new} = \sqrt{\frac{\lambda_{t+2} - \theta - \beta\lambda_{t+1}^{new}}{\alpha}},$$

which in turn implies that λ_{t+1}^{new} lies within the acceptable bounds. Simulating from the truncated normal distribution of f_t (given y_t, λ_t and φ) can be done by using a simple accept/reject algorithm, or by exploiting the probability integral transform.[7] In any case, since the conditional

[6] The density of a scaled non-central chi-squared variable $z = \sigma(x + \mu)^2$, where $x \sim N(0, 1)$, is

$$p_{NC}(z; \delta, \sigma) = \frac{1}{2\sigma}\exp\left\{-\frac{(\delta + z/\sigma)}{2}\right\}\left(\frac{z}{\sigma\delta}\right)^{-1/4}\sqrt{\frac{2\sigma}{\pi z}}\cosh(\sqrt{\delta z/\sigma}), \quad z \geq \delta,$$

where $\delta = \mu^2$ and $2\cosh(u) = \exp(u) + \exp(-u)$. This result is due to Fisher (1928) and is discussed by Johnson, Kotz, and Balakrishnan (1995, Ch. 29). In addition, if we wish to compute, $\Pr(z < c^2)$ for $c > 0$, then we have

$$\Pr\left\{x \in (-c/\sigma^{1/2} - \mu, c/\sigma^{1/2} - \mu)\right\} = \Phi\left(c/\sigma^{1/2} - \mu\right) - \Phi\left(-c/\sigma^{1/2} - \mu\right).$$

[7] Specifically, we could follow standard practice and compute the value of the conditionally Gaussian distribution function at both truncation limits, draw a uniform random number in the intermediate range, and use the inverse Gaussian distribution function to obtain the truncated normal variate. If the degree of truncation is small, the extra computations involved make this method unnecesarily slow. On the other hand, a simple accept/reject

density of f_t^{new} will be given by the following expression

$$p(f_t^{new} \,||f_t^{new} - \mu| \leq l_t, y_t, \lambda_t, \varphi) = \frac{1}{\sqrt{\omega_{t|y_t,\lambda_t}}} \phi\left(\frac{f_t^{new} - f_{t|y_t,\lambda_t}}{\sqrt{\omega_{t|y_t,\lambda_t}}}\right)$$
$$\times \left[\Phi\left(\frac{\mu + l_t - f_{t|y_t,\lambda_t}}{\sqrt{\omega_{t|y_t,\lambda_t}}}\right) - \Phi\left(\frac{\mu - l_t - f_{t|y_t,\lambda_t}}{\sqrt{\omega_{t|y_t,\lambda_t}}}\right)\right]^{-1},$$

where $\Phi(\cdot)$ is the distribution function of a standard normal, the density of λ_{t+1}^{new} will be

$$p(\lambda_{t+1}^{new} \,|\lambda_{t+1}^{new} \in [\theta + \beta\lambda_t, \beta^{-1}(\lambda_{t+2} - \theta)], y_t, \lambda_t, \varphi)$$
$$= \frac{c_t^{new}}{|2\alpha d_t^{new}|} \times \left[\Phi\left(\frac{\mu + l_t - f_{t|y_t,\lambda_t}}{\sqrt{\omega_{t|y_t,\lambda_t}}}\right) - \Phi\left(\frac{\mu - l_t - f_{t|y_t,\lambda_t}}{\sqrt{\omega_{t|y_t,\lambda_t}}}\right)\right]^{-1}$$

by virtue of the usual change of variable formula. But since the degree of truncation is the same for old and new, the acceptance probability will be the minimum of 1 and

$$\frac{p(y_{t+1}|\lambda_{t+1}^{new})}{p(y_{t+1}|\lambda_{t+1}^{r})} \frac{p(\lambda_{t+2}|y_{t+1}, \lambda_{t+1}^{new})}{p(\lambda_{t+2}|y_{t+1}, \lambda_{t+1}^{r})} = \frac{p(y_{t+1}|\lambda_{t+1}^{new})}{p(y_{t+1}|\lambda_{t+1}^{r})} \cdot \frac{c_{t+1}^{new}}{c_{t+1}^{r}} \cdot \frac{d_{t+1}^{r}}{d_{t+1}^{new}},$$

where we have again used (11).

Although the sign of $(f_t^{new} - \mu)$ does not affect the acceptance rate, note that if we retain f_t^{new}, then we will not need to simulate s_t at a later stage. Taken together this implies that we sweep through all T conditional variances, signs and factors in $O(T)$ operations. In addition, if $\beta = 0$, the upper truncation disappears, and this sampler coincides with the single move sampler over $\{f_t\}$ described in section 3.1.

3.2.3 Equivalent double-move samplers for $\{f_t, \lambda_{t+1}\}$

In fact, it is possible to arrive at the same sampling procedure by jointly sampling (f_t, λ_{t+1}, f_{t+1}, λ_{t+2}) conditional on $f_{\backslash t,t+1}^r$, $\lambda_{\backslash t+1,t+2}^r$, y and φ, where $f_{\backslash t,t+1}^r = \{f_1^{r+1}, \ldots, f_{t-1}^{r+1}, f_{t+2}^r, \ldots, f_T^r\}$ and $\lambda_{\backslash t+1,t+2}^r = \{\lambda_2^{r+1}, \ldots, \lambda_t^{r+1}, \lambda_{t+3}^r, \ldots, \lambda_{T+1}^r\}$. To see why, note that given that λ_{t+2} will be fully revealed by λ_{t+3} and f_{t+2} when $\beta \neq 0$, the candidate random vector must satisfy $\lambda_{t+2}^{new} = \lambda_{t+2}^r$. In addition, since

$$\lambda_{t+2} = \theta(1 + \beta) + \alpha(f_{t+1} - \mu)^2 + \alpha\beta(f_t - \mu)^2 + \beta^2\lambda_t,$$

any admissible proposal for f_t and f_{t+1} must actually lie on the two-dimensional ellipse

$$\beta(f_t - \mu)^2 + (f_{t+1} - \mu)^2 = [\lambda_{t+2} - \theta(1 + \beta) - \beta^2\lambda_t]/\alpha \tag{12}$$

The implication is twofold. First, any candidate f_t must be restricted to the interval $[\mu - l_t, \mu + l_t]$. Second, for any given f_t then we can solve out $f_{t+1} = \mu \pm d_{t+1}$.

method can be very inefficient if the double truncation of f_t^{new} is in the tails of the distribution. Hence, it may be worth assessing the degree of truncation first, and depending on its tightness, choose one simulation method or the other.

Let $g(f_{t+1}, f_t)$ denote any proposed distribution whose drawings $(f_{t+1}^{new}, f_t^{new})$ satisfy equation (12) - otherwise, we simply discard them. Then, the acceptance rate will be the minimum of 1 and[8]

$$\frac{p(y_t|f_t^{new},\lambda_t)p(y_{t+1}|f_{t+1}^{new},\lambda_{t+1}^{new})p(f_{t+1}^{new},\lambda_{t+2}|\lambda_{t+1}^{new})p(f_t^{new},\lambda_{t+1}^{new}|\lambda_t^r)}{p(y_t|f_t^r,\lambda_t)p(y_{t+1}|f_{t+1}^r,\lambda_{t+1}^r)p(f_{t+1}^r,\lambda_{t+2}|\lambda_{t+1}^r)p(f_t^r,\lambda_{t+1}^r|\lambda_t^r)}\cdot\frac{g(f_{t+1}^r,f_t^r)}{g(f_{t+1}^{new},f_t^{new})}.$$

If we then simulate f_t^{new} from its distribution given y_t, λ_t^{r+1}, and φ, but taking into account the truncation bounds on f_t, it is clear that associated with this proposal, there is a single candidate λ_{t+1}^{new}, and two possible values for f_{t+1} on the ellipse. If we finally choose between those two values by means of the conditional distribution of f_{t+1} given y_{t+1}, λ_{t+1}^{new} and φ, we can prove that we obtain exactly the same acceptance rate as before.[9] The only difference is that in this case we actually propose a value for f_{t+1}, while in the previous subsection we only implicitly proposed a value for $|f_{t+1}-\mu|$. In terms of an algorithm then this works as follows:

0. Set $t=1$.

1. Current values are f_t^r, f_{t+1}^r.

2. Sample $f_t^{new} \sim N(f_{t|y_t,\lambda_t}, \omega_{t|y_t,\lambda_t})$ doubly truncated so that it remains within the interval $\mu \pm l_t$, which implies the permissible values of f_{t+1}^{new} are $\mu \pm d_{t+1}^{new}$.

3. Draw

$$f_{t+1}^{new} = \begin{cases} \mu + d_{t+1}^{new}, & \text{with probability} \quad P(f_{t+1} = \mu + d_{t+1}^{new}|\lambda_{t+2}, \lambda_{t+1}^{new}, y_{t+1}) \\ \mu - d_{t+1}^{new}, & \text{with probability} \quad 1 - P(f_{t+1} = \mu + d_{t+1}^{new}|\lambda_{t+2}, \lambda_{t+1}^{new}, y_{t+1}) \end{cases}$$

   and accept proposal $\left\{f_t^{new}, f_{t+1}^{new}\right\}$ with probability

$$\min\left\{1, \frac{p(y_{t+1}|\lambda_{t+1}^{new})}{p(y_{t+1}|\lambda_{t+1}^r)}\frac{p(\lambda_{t+2}|y_{t+1},\lambda_{t+1}^{new})}{p(\lambda_{t+2}|y_{t+1},\lambda_{t+1}^r)}\right\}$$

4. Increase t by 1 and Goto 2.

3.2.4 Block-samplers

Unfortunately, the degree of truncation of the distribution of λ_{t+1} conditional with λ_{t+2} and λ_t can be severe. For that reason, it is convenient to consider double, triple, and in general h-tuple move samplers for λ_{t+1}. In addition, the mixing properties of the resulting chain are

[8] As we mentioned before, a technical complication with this approach is that since f_{t+1}, f_t satisfy (12), we have to be particularly careful in evaluating $g(f_{t+1}, f_t)$ as a function of the marginal density of f_t and the conditional probability of f_{t+1} given f_t, because there are extra Jacobian terms involved, which reflect the differentials along the ellipse (12) at the points f_{t+1}^{new}, f_t^{new} and f_{t+1}^r, f_t^r. For analogous reasons, we also have to be careful in evaluating $p(f_t, \lambda_{t+1}|\lambda_t)$, since f_t, λ_{t+1} lie on a parabola.

[9] Note that although the support of the distribution of f given λ, y and φ is highly restricted, the support of the distribution of f given y and φ is fully unrestricted. Thus, the EM algorithm can be safely applied.

usually much better as a result (see e.g. Liu, Wong, and Kong (1994)). For $h = 2$ for instance, a numerically efficient procedure is to simulate $f_t^{new} \sim N(f_{t|y_t,\lambda_t}, \omega_{t|y_t,\lambda_t})$, but imposing that

$$|f_t - \mu| \leq \sqrt{\frac{\lambda_{t+3} - \theta(1+\beta+\beta^2) - \beta^3\lambda_t}{\alpha\beta^2}}.$$

Associated with this proposal, there is a candidate λ_{t+1}^{new}, which we can combine with y_{t+1} to simulate f_t^{new}, but taking into account again that

$$|f_{t+1} - \mu| \leq \sqrt{\frac{\lambda_{t+3} - \theta(1+\beta) - \beta^2\lambda_{t+1}^{new}}{\alpha\beta}},$$

which in turn implies a candidate λ_{t+2}^{new} that is fully compatible with λ_{t+3}. Tedious algebraic manipulations show that for a given set of parameter values φ, the acceptance probability will be the minimum of 1 and

$$\frac{p(\lambda_{t+3}|y_{t+2},\lambda_{t+2}^{new}) \cdot p(\lambda_{t+2} \leq (\lambda_{t+3}-\theta)/\beta|y_{t+1},\lambda_{t+1}^{new}) \cdot p(y_{t+2}|\lambda_{t+2}^{new}) \cdot p(y_{t+1}|\lambda_{t+1}^{new})}{p(\lambda_{t+3}|y_{t+2},\lambda_{t+2}^{r}) \cdot p(\lambda_{t+2} \leq (\lambda_{t+3}-\theta)|y_{t+1},\lambda_{t+1}^{r}) \cdot p(y_{t+2}|\lambda_{t+2}^{r}) \cdot p(y_{t+1}|\lambda_{t+1}^{r})},$$

where the terms $p(\lambda_{t+2} \leq (\lambda_{t+3}-\theta)/\beta|y_{t+1},\lambda_{t+1})$ reflect the degree of truncation of the conditional distribution of λ_{t+2}. Again, since we can use f_{t+1}^{new} and f_t^{new} to sample s_{t+1} and s_t, we can reinterpret this double move sampler over $\lambda_{t+2}, \lambda_{t+1}$ as a triple move sampler over $f_t, f_{t+1}, f_{t+2}, \lambda_{t+1}, \lambda_{t+2}$ and λ_{t+3}, which effectively becomes a sampler of f_t, f_{t+1}, f_{t+2} on the tridimensional ellipsoid:

$$\beta^2(f_t-\mu)^2 + \beta(f_{t+1}-\mu)^2 + (f_{t+2}-\mu)^2 = [\lambda_{t+3} - \theta(1+\beta+\beta^2) - \beta^3\lambda_t]/\alpha.$$

As we increase h, this procedure converges to the following independent wholemove sampler: starting with $\lambda_1 = \lambda$, we recursively propose f_t^{new} from the unrestricted univariate normal density $N(f_{t|y_t,\lambda_t^{new}}, \omega_{t|y_t,\lambda_t^{new}})$, where $\lambda_t^{new} = V(f_t|f_1^{new}, \ldots, f_{t-1}^{new})$. The acceptance probability in this case is the minimum of 1 and

$$\frac{\prod_{t=1}^{T} p(y_t|\lambda_t^{new},\varphi)}{\prod_{t=1}^{T} p(y_t|\lambda_t^{r},\varphi)}.$$

Since this T-tuple method suffers from the problem discussed at the beginning of section 3.1, it is clear that as far the choice of h is concerned, there is an implicit trade-off between alleviating the degree of truncation, and increasing the dimension of the proposal. For that reason, our preferred method is a mixed procedure, in which the value of h is randomly chosen from a uniform distribution in the range $1 \leq h \leq H \leq T$.

3.3 A comparison of the different simulators

In order to compare the performance of the different MCMC samplers introduced in the previous subsections, we have generated realizations of size $T = 240$ of the GLS factor representing

portfolios that would correspond to the trivariate single factor models analyzed by Monte Carlo methods in Sentana and Fiorentini (2001). These authors set $\lambda = 1$, c=$(1,1,1)'$, and Γ=γI_3, with $\gamma = 2$ or $1/2$, corresponding to relatively low ($\upsilon = 2/3$) and high ($\upsilon = 1/6$) signal to noise ratios.[10] They also set (α,β) = $(.2, .6)$ or $(.4, .4)$, to represent persistent but smooth GARCH behaviour, and persistent but volatile conditional variances respectively ($\upsilon = 1/6, \alpha = .2, \beta = .6$ matches roughly what we tend to see in the empirical literature for monthly data). In addition, for each of the four combinations, we have considered not only $\mu = 0$, but also $\mu = 1/2$, to allow for leverage effects in the conditional variance of the factors. Finally, we have interacted the resulting eight combinations with $\tau = 0$ and $\tau = 1/2$, the second of which reflects the fact that conditioning information often plays a crucial role in deriving asset risk premia (see King, Sentana, and Wadhwani (1994)).

We analyze four different samplers: inefficient, single move ($h = 1$), block move ($h = 9$), and random length block move ($h \sim U(1, 19)$). We first examine the increase in the variance of the sample mean of f_t across 500,000 MCMC simulations due to the autocorrelation in the drawings relative to an ideal but infeasible independent sampler. We do so by estimating the autocorrelation generating function at the origin for observations $t = 80$ and $t = 160$ using standard spectral density estimation techniques (see e.g. Priestley (1981)). In addition, we record the mean acceptance probabilities over all observations, the average CPU time needed to simulate one complete drawing of $f|y, \varphi$, as well as the CPU effort required to obtain MCMC samples that match the numerical standard error of the random h simulator at observation $t = 160$. The behaviour of the different simulators, which is summarised in Table 1 and Figures 1a-1b, is very much as one would expect. The computationally inefficient sampler shows relatively little serial correlation, and a high acceptance rate for each individual t, but it is extremely time consuming to compute even though our sample size is fairly small. In fact, when we increase T from 240 to $2,400$ and $24,000$, the average CPU time increases by a factor of 100 and 10,000, respectively, as opposed to 10 and 100 for the other three simulators, which makes it impossible to implement in most cases of practical interest. On the other hand, the single and 9-tuple move samplers over λ_{t+1} produce results much faster, with a reasonably high acceptance rate but more autocorrelation in the drawings. As expected, the best overall performance seems to be achieved by the simulator with random block size. Importantly, the distributions of f_t generated by the four samplers are indistinguishable from one another.[11]

[10] More specifically, when $\tau = 0$ the coefficient of determination in the regression of y_t on f_t will be $R^2 = 3/5$ or $6/7$ for $\upsilon = 2/3$ or $1/6$ respectively. Note that $R^2 = \lambda/(\lambda + \upsilon)$ is a monotonic transformation of the innovations' signal to noise ratio λ/υ.

[11] For the sake of brevity, we only present results for the parameter configuration $\alpha = .2$, $\beta = .6$, $\mu = .5$, $\tau = .5$ and $\upsilon = 2/3$. However, the relative performance of the simulators is not affected much by changes in the parameter values. Their mean acceptance rates, though, are sensitive to the parameter values, being uniformly

We also assess the quality of the approximate Kalman filter procedure put forward by HRS by comparing the Gaussian distributions for f_t given data and parameters that their method implies with the ones we have obtained by our exact MCMC approach. As can be seen from Table 2 and Figures 2a-2h, the results crucially depend on the parameter values. In particular, the approximate Kalman filter provides more reliable results the closer the unconditional distribution of the latent factors is to the normal ($\alpha = .2$, $\beta = .6$), and the larger the signal to noise ratio ($\upsilon = 1/6$). In contrast, the degree of approximation is significantly worse when there is substantial variation in the conditional variance of the factors ($\alpha = .4$, $\beta = .4$), and the signal to noise ratio is smaller ($\upsilon = 2/3$). Nevertheless, while increasing the variability of the conditional variances keeping everything else constant directly leads to a deterioration of the HRS approximation, ceteris paribus increases in the idiosyncratic variances mostly seem to magnify the existing differences in proportion to the reciprocal of the root mean square error $\omega_{t|x}^{1/2}$. Increases in the price of risk parameter τ also affect negatively the quality of the approximation. In contrast, variations in the dynamic asymmetry parameter μ (not reported here) have a rather small effect, with an ambiguous sign. Similar results are obtained when we compare the distributions of the risk rewarded factors, r_{ft}. Therefore, given that in many empirical applications it is likely that the signal to noise ratio will be high, and the conditional variance a fairly smooth process, we would expect the HRS simulators to be fairly accurate in practice.

4 Empirical Application to UK Sectorial Stock Returns

In this section, we investigate the practical performance of the procedures discussed above. To do so, we revisit the empirical application in Sentana (1995), who analyzed the relationship between first and second conditional moments for monthly excess stock returns on 26 U.K. sectors for the period 1971:2 to 1990:10 (237 observations). On the basis of the approximate Kalman filter-based Gaussian pseudo-likelihood function, he estimated a conditionally heteroskedastic *in mean* latent factor model, in which the common unobservable factor follows a GQARCH (1,1) process. Therefore, the total number of parameters is $2 \times 26 + 4 = 56$.

4.1 Scaling choice

Before applying either the classical or Bayesian estimation procedures, though, we must decide between the alternative normalisations $c_1 = 1$ and $\lambda = 1$ discussed in section 2. In this respect, our results clearly favour the former over the latter, which is line with the theoretical and

higher the higher the signal to noise ratio, and the lower the variability in λ_t. But while they are lower for $\tau = .5$ than for $\tau = 0$, they are hardly sensitive to μ.

empirical findings obtained by Pitt and Shephard (1999a) for univariate stochastic volatility models. In particular, if we set $c_1 = 1$ for estimation purposes, but then re-scale the results so that $\lambda = 1$, we obtain significantly less serial correlation in the MCMC drawings from the posterior distributions of the factor loadings than if we directly set $\lambda = 1$ (see Table 3). For that reason, in what follows the reported results correspond to the first approach.

4.2 Simulated EM algorithm

Although in this example we knew the approximate ML estimates produced by the HRS method, in practice no such initial values are necessarily available. For that reason, we decided to set all parameters to plausible but arbitrary starting values. Specifically, we set all factor loadings to 1, all idiosyncratic variances to 0.1, the ARCH and GARCH parameters to 0.2 and 0.6 respectively, and both the price of risk, τ, and the dynamic asymmetry parameter, μ, to 0. Importantly, the numerical maximisation of (7) with respect to the underlying parameters ψ_1^*, ψ_2^*, ψ_3^*, τ and λ described in section 2.4.2 was performed every single iteration.[12]

After just a few iterations, the EM algorithm takes the parameters fairly close to their ML estimates. However, it slows down considerably in the neighbourhood of the optimum. For that reason, we decided to stop it after 1250 iterations, at which point the Euclidean norm of the changes in the parameter vector between iterations was around 10^{-4}. As can be seen from Tables 4 and 5, the differences with respect to the approximate ML estimates are fairly small, which is not surprising in view of the results in section 3.3 because the innovations' signal to noise ratio, which measures the degree of observability of the latent factor, is around 100. The most noticeable discrepancies appear in the conditional variance parameters α, β and μ, the price of risk coefficient τ, and the first factor loading c_1, which remember coincides with the unconditional standard deviation of the common factor in the normalisation used for estimation purposes. In this respect, it is important to note that the discrepancies in the remaining factor loadings are simply the result of the fact that the sample average of the estimated λ_t's obtained by the HRS method is approximately 5% smaller than the average factor variance generated by our proposed MCMC procedure. If we look at the ratio of any two factor loading estimates (excluding the first), the differences are as small as in the idiosyncratic variances.

4.3 Simulated Bayesian inference

Since the model for $r_{ft}|\delta$ corresponds to a univariate GQARCH in mean process, whose log-likelihood function can be easily evaluated, we use Metropolis-Hastings MCMC methods to

[12] Note that in order to maximise (7) at each EM iteration, we must maintain the latest simulated values of r_f constant. We must also maintain the underlying random drawings constant across EM iterations in oder to allow the algorithm to converge (cf. Nielsen (2000)).

simulate from the posterior distribution of δ given the "observed" r_{ft}.[13] In particular, we employed the Adaptive Rejection Metropolis Sampling (ARMS) algorithm of Gilks, Best, and Tan (1995), and Gilks, Neal, Best, and Tan (1997).[14] In this case, we use independent beta priors on $\psi_1(=\alpha+\beta)$ and $\psi_2(=\beta/(\alpha+\beta))$, with mean 3/4 and standard deviation .1443, which are centred around the typical values estimated by previous studies with monthly return data. We also use a shifted and scaled independent beta prior on ψ_3^*, so that it fluctuates between $\pm\pi/2$ with zero mean and standard deviation $\pi/4$. Please note that such a distribution implies that we do not take any ex-ante view on the sign of the dynamic asymmetry effect. Similarly, we assume a normal prior for the price of risk coefficient τ, with zero mean and standard deviation .1, which is also neutral about its possible sign. Finally, we use a standard inverted gamma prior for the unconditional variance of the common factor, λ, with mean 1 and variance 1/2.

As for the static variance parameters, given that we chose to normalise with $c_1 = 1$ for estimation purposes, and that previous studies suggest that the dispersion of the factor loadings across different industrial sectors during our sample period is likely to be small, we chose informative normal priors for the remaining factor loadings, with unit mean and variance $\gamma_i/5$. In addition, we specified the usual marginal inverted gamma prior for γ_i $(i = 1, \ldots, 26)$, with mean and standard deviation equal to 1/4. If we recall the prior distribution for λ, such values imply that the theoretical R^2 in the regression of each sectorial return on the common factor would be on average approximately equal to .8, which seems plausible for the sectorial return data that we are analyzing.

If we use the posterior means of the parameters reported in Tables 4 and 5 as point estimates, our results suggests that the Bayesian and classical procedures are largely in agreement. Again, the discrepancies in the second and successive factor loadings are almost entirely driven by the differences in the sample means of the estimated conditional variances of the factors across the three methods. The most noticeable difference, in fact, corresponds to the dynamic asymmetry parameter μ, which is the least precisely estimated. Nevertheless, our findings confirm the main result in Sentana (1995), namely that there appears to be a significant leverage effect in the sectorial returns through the common factor. They also confirm that the price of risk is estimated as being positive.

We have also performed an analysis of the sensitivities of the results in Tables 4 and 5 to our

[13] In this respect, it is important to note that we must recompute λ_t every time the conditional variance parameters are updated, before proceeding to the next round of simulation of the common factors given observed data and parameters (cf. Nakatsuma (2000)).

[14] We also considered the procedure suggested by Chib and Greenberg (1994), which makes proposals from a multivariate normal prior on ψ_1^*, ψ_2^*, ψ_3^*, τ and λ, with mean given by their ML estimators obtained on the current r_f, and variance given by the estimated inverse information matrix. But since both pocedures yield almost identical results, we only report the ARMS ones.

choice of priors. In particular, we have halved and doubled the dispersion of the prior distributions of the factor loadings, idiosyncratic variances, price of risk coefficient τ, and unconditional variance of the common factor, λ, around their respective means. As for the three parameters with beta priors, we increased and reduced their prior variances as much as possible, but without changing the mean or the concavity of the distribution. Specifically, we could only increase the prior variances of ψ_1 and ψ_2 by approximately 50% each, and the prior variance of ψ_3^* by 30%. The results, which are reported in Figures 3a-d and 4a-4d, indicate that the choice of priors does not unduly influence our conclusions. In particular, the positivity of the dynamic asymmetry parameter μ and the price of risk coefficient τ seems to be robust.

5 Conclusions

We derive exact likelihood-based estimators of latent variable models in which the variances of the unobservable processes are functions of their past values. Since in general the expression for the likelihood function is unknown, we resort to simulation methods. In this context, we show that MCMC likelihood-based estimation of latent GARCH models can in fact be handled by means of feasible $O(T)$ algorithms.

Our samplers of the latent variables given data and parameters involve two main steps. First, we augment the state vector to achieve a first-order Markovian process in an analogous manner to the way in which GARCH models are simulated in practice. Then, we discuss alternative procedures for dealing with the dynamic singularity implicit in the GARCH model that results from the fact that there is only one shock driving innovations to the level of the process and its variance. In this sense, the situation is radically different from the one existing in stochastic volatility models, which often assume that the shocks driving the level and volatility of the process are independent.

A numerical comparison of our proposed procedures suggests that a random size block sampler yields the best trade-off between serial dependence of the drawings, and speed. It also shows that the Kalman filter-based Gaussian approximation introduced by HRS produces reasonable results when the signal to noise ratio is high, and the unconditional coefficient of variation of the volatility of the common factor is low, but that it may lead to substantial differences in other cases.

We can then use several recent proposals from the simulation-based direct inference literature to estimate the model parameters. In particular, from a classical perspective, we consider both a simulated EM algorithm, and the related method of simulated scores. We also develop simulation-based Bayesian inference procedures by combining within a Gibbs sampler the

MCMC simulators with the posterior distributions of the parameters given observed series and latent variables. In this respect, we find that the parametrisations induced by two alternative scaling assumptions can have a substantial effect on the efficiency of the Gibbs sampler.

In order to investigate the practical performance of our proposed procedures, we fully re-assess the empirical application in Sentana (1995), who analyzed the relationship between first and second conditional moments for monthly excess stock returns on 26 U.K. sectors for the period 1971 to 1990. Given the extremely high signal to noise ratio, our exact-likelihood based results are not very different from his. In particular, we confirm his main findings that there is a dynamic leverage effect in sectorial returns through the common factor, and that the price of risk coefficient is positive. Nevertheless, we also show that there are some differences in the estimation of the conditional variance parameters.

Although we have developed our results within the context of a CH factor model with common factors that follow Quadratic GARCH in mean processes, it applies much more widely. For instance, if we assume that λ_t follows the log-GARCH model of Geweke (1986) and Pantula (1986), most of our expressions go through largely unchanged by simply replacing λ_t by its log as the additional element of the state. Similarly, it is relatively easy to modify our proposed procedures to deal with GARCH models in situations in which there is partial observability due to missing prices. To keep the discussion as simple as possible, imagine that an asset price is recorded every day for the first half of the sample, but it is only recorded every other day during the second half. Such a mixed observation timing interval implies that for the second subsample we only observe the sum of pairs of successive daily returns (cf. Wei (2002)). In this context, if we assume that a strong GQARCH(1,1) model applies at the daily frequency, we could again "simulate" the daily returns given the observed prices in $O(T)$ operations by augmenting the state to include the daily conditional volatilities. Specifically, we could jointly propose overlapping blocks of daily returns and volatilities (conditional on the observed returns, and the remaining daily returns and volatilities) along the lines described in sections 3.2.3 and 3.2.4. The main difference is that the block sizes would become 4, 6, 8, ... as opposed to 2, 3, 4 ... because in this case we would know the sum of consecutive returns.

Finally, it is important to mention that a relevant issue that we have not thoroughly discussed in this paper is the identification of the model parameters in latent variable processes. Although the exact factor model that we consider in detail is indeed identified, as discussed in section 2, and the same is generally true of models with partial observability resulting from missing prices, there are other models in which this is not always necessarily the case. The reader is referred to Sentana and Fiorentini (2001) and Doz and Renault (2001) for details.

References

Aguilar, O. and M. West (2000). Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *Journal of Business and Economic Statistics 18*, 338–357.

Albert, J. H. and S. Chib (1993). Bayesian inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business and Economic Statistics 11*, 1–15.

Andersen, T. G. (1994). Volatility. Unpublished paper: Finance Department, Northwestern University.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 74*, 3–30.

Bauwens, L. and M. Lubrano (1998). Bayesian inference on GARCH models using the Gibbs sampler. *Econometrics Journal 1*, C23–C46.

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

Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics 69*, 542–47.

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994). ARCH models. In R. F. Engle and D. McFadden (Eds.), *The Handbook of Econometrics, Volume 4*, pp. 2959–3038. Amsterdam: North-Holland.

Brennan, M. J. (1986). A theory of price limits in futures markets. *Journal of Financial Economics 16*, 213–233.

Bresnahan, T. F. (1981). Departures from marginal-cost pricing in the American automobile industry, estimates from 1977-1978. *Journal of Econometrics 17*, 201–227.

Calzolari, G., G. Fiorentini, and E. Sentana (2003). Indirect inference estimation of conditionally heteroskedastic factor models. mimeo, CEMFI, Madrid, Spain.

Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 81*, 541–53.

Celeux, G. and J. Diebolt (1985). The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly 2*, 73–82.

Chamberlain, G. and M. Rothschild (1983). Arbitrage and mean-variance analysis of large asset markets. *Econometrica 51*, 1281–1301.

Chib, S. (2001). Markov chain Monte Carlo methods: Computation and inference. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, pp. 3569–3649. Amsterdam: North-Holland.

Chib, S. and E. Greenberg (1994). Bayes inference for regression models with ARMA(p,q) errors. *Journal of Econometrics 64*, 183–206.

Chib, S., F. Nardari, and N. Shephard (2002). Analysis of high dimensional multivariate stochastic volatility models. Unpublished paper: Nuffield College, Oxford.

Chou, R., R. F. Engle, and A. Kane (1992). Measuring risk aversion from excess returns on a stock index. *Journal of Econometrics 52*, 201–224.

Demos, A. and E. Sentana (1998). An EM algorithm for conditionally heteroskedastic factor models. *Journal of Business and Economic Statistics 16*, 357–361.

Dempster, A. P., N. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B 39*, 1–38.

Diebold, F. X. and M. Nerlove (1989). The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *Journal of Applied Econometrics 4*, 1–21.

Diebold, F. X. and T. Schuermann (2000). Exact maximum likelihood estimation of observation-driven econometric models. In R. S. Mariano, M. Weeks, and T. Schuermann (Eds.), *Simulation-Based Inference in Econometrics: Methods and Applications*, pp. 205–217. Cambridge: Cambridge University Press.

Doz, C. and E. Renault (2001). Conditionally heteroskedastic factor models: identification and instrumental variable estimation. Unpublished paper: University of Cergy-Pontoise.

Dueker, M. (1999). Conditional heteroscedasticity in qualitative response models of time series: a Gibbs-sampling approach to the bank prime rate. *Journal of Business and Economic Statistics 17*, 466–472.

Dungey, M., V. L. Martin, and A. R. Pagan (2000). A multivariate latent factor decomposition of international bond yield spreads. *Journal of Applied Econometrics 15*, 697–715.

Eichengreen, B., M. W. Watson, and R. S. Grossman (1985). Bank rate policy under the interwar gold standard: a dynamic probit model. *Economic Journal 95*, 725–745.

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

Engle, R. F. (1990). Discussion: Stock market volatility and the crash of 87. *Review of Financial Studies 3*, 103–106.

Engle, R. F., D. F. Hendry, and J. F. Richard (1983). Exogeneity. *Econometrica 51*, 277–304.

Engle, R. F. and K. F. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory 11*, 122–150.

Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Transactions of the Royal Society of London, Series A 121*, 654–673.

Gelfand, A. E. and A. F. M. Smith (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*, 398–409.

Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions, PAMI 6*, 721–41.

Geweke, J. (1986). Modelling the persistence of conditional variances: a comment. *Econometric Reviews 5*, 57–61.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica 57*, 1317–1339.

Geweke, J. F. and G. Zhou (1996). Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies 9*, 557–87.

Ghysels, E., A. C. Harvey, and E. Renault (1996). Stochastic volatility. In C. R. Rao and G. S. Maddala (Eds.), *Statistical Methods in Finance*, pp. 119–191. Amsterdam: North-Holland.

Giakoumatos, S. G., P. Dellaportas, and D. M. Politis (1999). Bayesian analysis of the unobserved ARCH model. Department of Statistics, Athens University of Economics and Business.

Gilks, W. K., S. Richardson, and D. J. Spiegelhalter (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

Gilks, W. R., N. G. Best, and K. K. C. Tan (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics 44*, 155–73.

Gilks, W. R., R. M. Neal, N. G. Best, and K. K. C. Tan (1997). Corrigendum: adaptive rejection metropolis sampling. *Applied Statistics 46*, 541–542.

Gourieroux, C., A. Monfort, and E. Renault (1991). A general framework for factor models. mimeo, INSEE.

Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics 8*, S85–S118.

Hajivassiliou, V. and D. McFadden (1998). The method of simulated scores for the estimation of LDV models. *Econometrica 66*, 863–896.

Hall, A. D., P. Kofman, and S. Manaster (2001). Migration of price discovery with constrained futures markets. Unpublished paper: University of Technology, Sidney.

Harvey, A. C., E. Ruiz, and E. Sentana (1992). Unobserved component time series models with ARCH disturbances. *Journal of Econometrics 52*, 129–158.

Hasbrouck, J. (1999). The dynamics of discrete bid and ask quotes. *Journal of Finance 54*, 2109–2142.

Hastings, W. K. (1970). Monte-Carlo sampling methods using Markov chains and their applications. *Biometrika 57*, 97–109.

Jensen, M. B. and A. Lunde (2001). The NIG-S&ARCH model: A fat tailed, stochastic, and autoregressive conditional heteroskedastic volatility model. *Econometrics Journal 4*, 319–342.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1995). *Continuous Univariate Distributions* (2 ed.), Volume 2. New York: John Wiley.

Kim, C.-J. and C. R. Nelson (1999). *State-Space Models with Regime Switching. Classical and Gibbs-Sampling Approaches with Applications*. Cambridge: MIT.

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

King, M., E. Sentana, and S. Wadhwani (1994). Volatility and links between national stock markets. *Econometrica 62*, 901–933.

Kodres, L. E. (1988). Tests of unbiasedness in foreign exchange futures markets: the effects of price limits. *Review of Futures Markets 7*, 138–166.

Kodres, L. E. (1993). Tests of unbiasedness in foreign exchange futures markets: an explanation of price limits and conditional heteroskedasticity. *Journal of Business 66*, 463–490.

Koopman, S. J. and N. Shephard (2003). Testing the assumptions behind the use of importance sampling. Unpublished paper: Nuffield College, Oxford.

Lee, L.-f. (1999). Estimation of dynamic and ARCH Tobit models. *Journal of Econometrics 92*, 355–390.

Liu, J. S., W. H. Wong, and A. Kong (1994). Covariance structure of the Gibbs sampler with applications to the comparison of estimators and augmentation schemes. *Biometrika 81*, 27–40.

Louis, T. A. (1982). Finding observed information using the EM algorithm. *Journal of the Royal Statistical Society, Series B 44*, 98–103.

McCurdy, T. H. and I. G. Morgan (1987). Tests of the Martingale hypothesis for foreign currency futures with time-varying volatility. *International Journal of Forecasting 3*, 131–148.

Meddahi, N. and E. Renault (2002). Temporal aggregation of volatility models. *Journal of Econometrics*. Forthcoming.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics 21*, 1087–92.

Morgan, I. G. and R. G. Trevor (1999). Limit moves as censored observations of equilibrium future prices in GARCH processes. *Journal of Business and Economic Statistics 14*, 397–408.

Nakatsuma, T. (2000). Bayesian analysis of ARMA-GARCH models: A Markov chain sampling approach. *Journal of Econometrics 95*, 57–69.

Nielsen, S. F. (2000). On simulated EM algorithms. *Journal of Econometrics 96*, 267–292.

Pantula, S. G. (1986). Modelling the persistence of conditional variances: a comment. *Econometric Reviews 5*, 71–74.

Pitt, M. K. and N. Shephard (1999a). Analytic convergence rates and parameterisation issues for the Gibbs sampler applied to state space models. *Journal of Time Series Analysis 21*, 63–85.

Pitt, M. K. and N. Shephard (1999b). Time varying covariances: a factor stochastic volatility approach (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 6*, pp. 547–570. Oxford: Oxford University Press.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*. London: Academic Press.

Ripley, B. D. (1977). Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society, Series B 39*, 172–212.

Ruud, P. (1991). Extensions of estimation methods using the EM algorithm. *Journal of Econometrics 49*, 305–341.

Sentana, E. (1995). Quadratic ARCH models. *Review of Economic Studies 62*, 639–661.

Sentana, E. (2002). Factor representing portfolios in large asset markets. *Journal of Econometrics*. Forthcoming.

Sentana, E. and G. Fiorentini (2001). Identification, estimation and testing of conditionally heteroskedastic factor models. *Journal of Econometrics 102*, 143–164.

Shephard, N. (1993). Fitting non-linear time series models, with applications to stochastic variance models. *Journal of Applied Econometrics 8*, S135–52.

Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika 81*, 115–31.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen (Eds.), *Time Series Models in Econometrics, Finance and Other Fields*, pp. 1–67. London: Chapman & Hall.

Spivak, M. (1965). *Calculus on Manifolds. A Modern Approach to Classical Theorems of Advanced Calculus*. New York: W A Benjamin.

Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for Exploration of Posterior Distributions and Likelihood Functions* (3 ed.). New York: Springer-Verlag.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association 82*, 528–50.

Taylor, S. J. (1994). Modelling stochastic volatility. *Mathematical Finance 4*, 183–204.

Tierney, L. (1994). Markov Chains for exploring posterior distributions (with discussion). *Annals of Statistics 21*, 1701–62.

Wei, S. X. (2002). A censored-GARCH model of asset returns with price limits. *Journal of Empirical Finance 9*, 197–223.

Table 1

Comparison between alternative MCMC simulators of $f|y, \varphi$

	MAP	IR		Time (CPU Effort)		
		f_{80}	f_{160}	$T = 240$	$T = 2,400$	$T = 24,000$
Inefficient	.901	1.42	3.39	7.12 (3.4)	692.9 (32)	71,761 (342)
$h = 1$	.690	8.10	68.7	*1* (9.7)	10.1 (9.6)	100.4 (9.7)
$h = 9$	.518	19.0	16.8	.942 (2.2)	9.56 (2.2)	95.7 (2.2)
random h	.602	3.42	7.91	.895 (*1*)	9.13 (*1*)	89.9 (*1*)

Notes: MAP denotes mean acceptance probability over the whole sample, while IR refers to the inefficiency ratio of the MCMC drawings of the latent factor at observations 80 and 160. Time refers to the total CPU time taken to simulate one complete drawing of $f|y, \varphi$ relative to the $h = 1, T = 240$ simulator. Finally, CPU Effort is the time required to obtain an MCMC sample of $f|y, \varphi$ that matches the numerical standard error of the random h simulator at observation 160.

Parameter values: $\alpha = .2, \beta = .6, \mu = .5, \tau = .5, \upsilon = 2/3$.

Table 2

Comparison between the simulated distribution generated by our proposed MCMC method, and the Kalman-filter based Gaussian approximation. The focus of attention is on the latent factor at observations 80 and 160 for different parameter configurations.

	f_{80}		f_{160}	
	MCMC	HRS	MCMC	HRS
	$\alpha = .4, \beta = .4, \mu = .5, \tau = .5, \upsilon = 2/3$			
mean	.818	.896	-2.450	-2.009
variance	.412	.471	.465	.471
skewness	-.139	*0*	-.075	*0*
kurtosis	3.079	*3*	2.813	*3*
	$\alpha = .2, \beta = .6, \mu = .5, \tau = .5, \upsilon = 2/3$			
mean	.973	1.006	-2.066	-1.883
variance	.426	.444	.434	.432
skewness	-.048	*0*	-.117	*0*
kurtosis	3.031	*3*	3.118	*3*
	$\alpha = .2, \beta = .6, \mu = .5, \tau = .5, \upsilon = 1/6$			
mean	.884	.850	-2.349	-2.304
variance	.156	.149	.144	.143
skewness	-.041	*0*	-.062	*0*
kurtosis	2.989	*3*	2.976	*3*
	$\alpha = .2, \beta = .6, \mu = .5, \tau = 0, \upsilon = 1/6$			
mean	.911	.901	-2.206	-2.178
variance	.148	.148	.135	.143
skewness	-.011	*0*	-.018	*0*
kurtosis	3.006	*3*	3.009	*3*

Note: HRS denotes the Harvey, Ruiz, and Sentana (1992) approximation.

Table 3

Inefficiency ratios for posterior drawings of static variance parameters under alternative scaling assumptions

Parameter	c_1	c_{26}	γ_1	γ_{26}
IR ($c_1 = 1$)	27.11	5.52	1.42	1.79
IR ($\lambda = 1$)	210.20	280.43	1.32	1.31

Table 4

Estimates of static variance parameters

	Factor Loadings				Idiosyncratic Variances			
Sector	HRS	SEM	Bayesian		HRS	SEM	Bayesian	
			PM	PSD			PM	PSD
1	.805	.836	.852	.092	.349	.349	.348	.033
2	.764	.784	.806	.090	.198	.198	.199	.019
3	.969	.994	1.016	.110	.123	.123	.127	.013
4	.740	.759	.782	.087	.182	.182	.183	.017
5	1.027	1.053	1.075	.117	.182	.182	.185	.018
6	.816	.837	.860	.096	.230	.230	.230	.022
7	.816	.837	.859	.093	.115	.115	.118	.012
8	.771	.791	.813	.088	.091	.091	.094	.009
9	.812	.833	.856	.095	.219	.219	.220	.021
10	.826	.847	.870	.101	.410	.410	.407	.038
11	.801	.822	.844	.095	.258	.258	.258	.024
12	.876	.899	.921	.102	.255	.255	.255	.024
13	.797	.818	.840	.092	.149	.149	.151	.015
14	.884	.907	.930	.100	.118	.118	.121	.012
15	.970	.996	1.018	.113	.329	.329	.329	.031
16	.810	.832	.854	.095	.220	.220	.221	.021
17	.835	.857	.879	.095	.096	.096	.099	.010
18	.767	.787	.810	.092	.285	.285	.284	.027
19	.783	.804	.827	.094	.280	.280	.280	.026
20	.834	.855	.878	.098	.269	.269	.268	.025
21	.644	.661	.683	.085	.526	.526	.523	.048
22	.829	.850	.873	.095	.131	.131	.134	.013
23	.862	.885	.907	.103	.365	.365	.363	.034
24	.656	.673	.696	.080	.254	.254	.255	.024
25	.872	.894	.917	.101	.197	.197	.198	.019
26	.808	.830	.852	.095	.211	.211	.211	.020

Notes: HRS denotes the Harvey, Ruiz, and Sentana (1992) approximation, SEM simulated EM, while PM and PSD represent the posterior mean and standard deviation respectively.

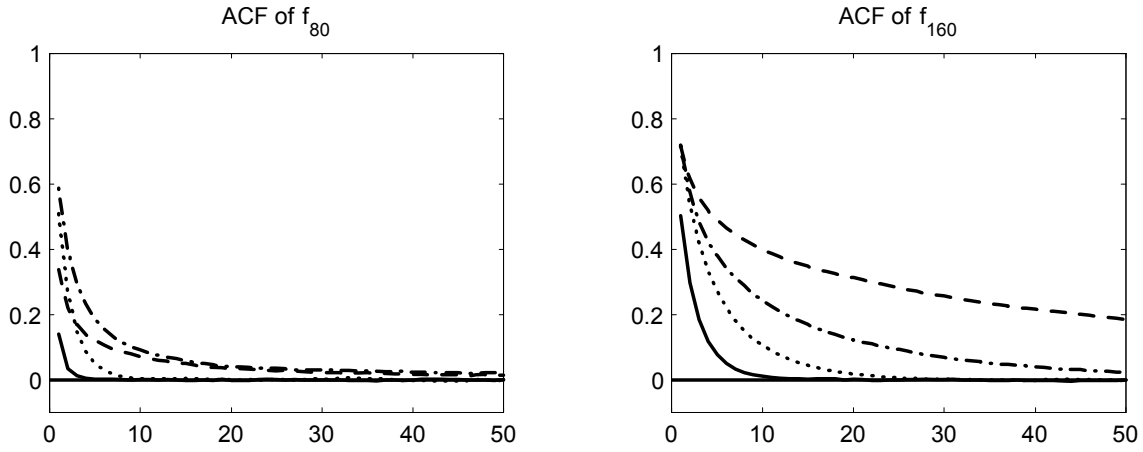
Table 5

Estimates of dynamic variance parameters

Parameter	HRS	SEM	Bayesian	
			PM	PSD
α	.143	.159	.173	.064
β	.639	.591	.627	.080
μ	.892	.944	.785	.400
τ	.145	.142	.140	.061

Notes: HRS denotes the Harvey, Ruiz, and Sentana (1992) approximation, SEM simulated EM, while PM and PSD represent the posterior mean and standard deviation respectively.

Figure 1: Comparison of the alternative simulators of the latent factors given observables and parameters by means of the autocorrelation functions (ACF) of the drawings.



— Inefficient – – Block length = 1 – · Block length = 9 ⋯ Random block length

$(\alpha = .2, \quad \beta = .6, \quad \mu = .5, \quad \tau = .5, \quad \upsilon = 2/3).$

Figure 2: Comparison of the p.d.f. of the simulated latent factors given observables and parameters with its Kalman filter-based Gaussian approximation for different parameter configurations.

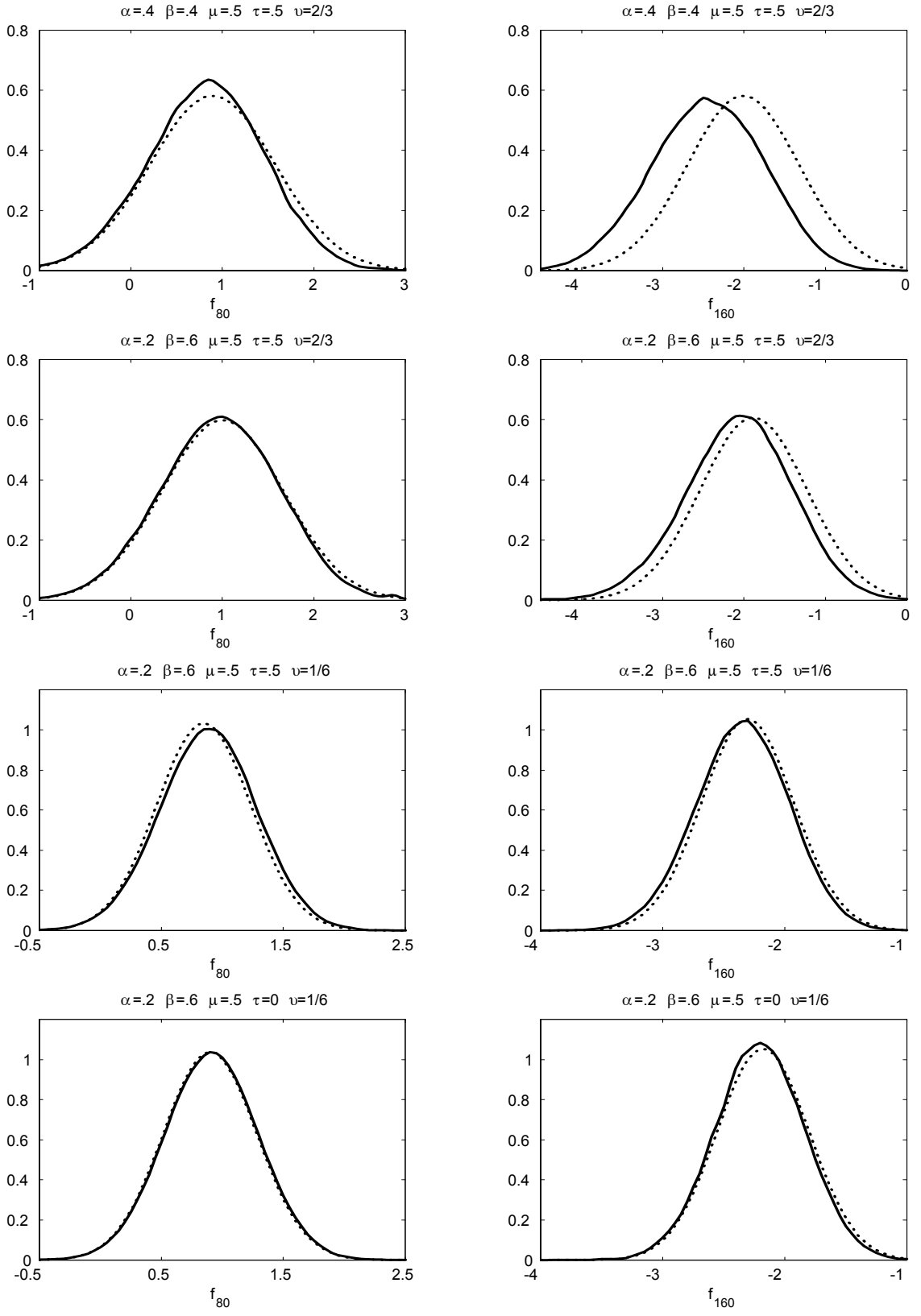
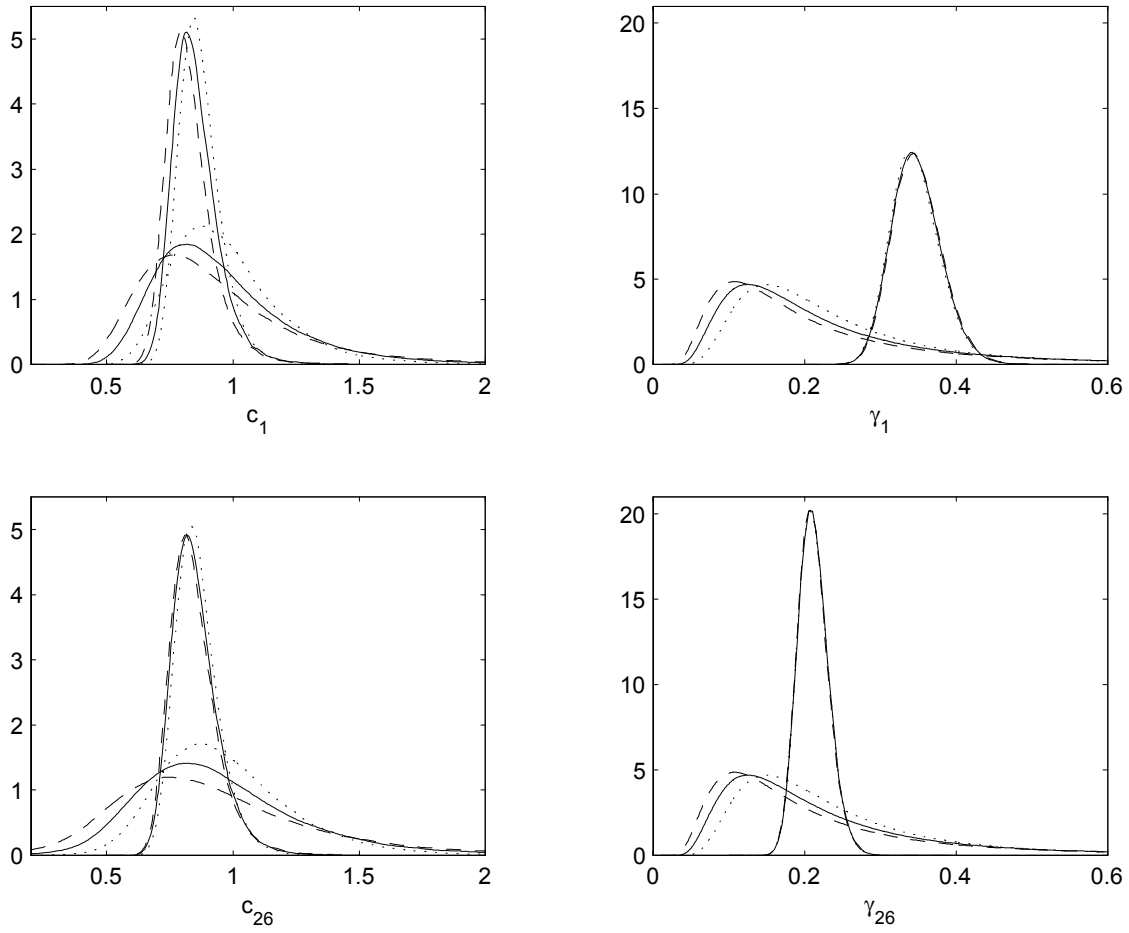


Figure 3: Sensitivity of the simulated posterior distributions of the static variance parameters c_1, γ_1, c_{26}, and γ_{26} to increases and decreases in the variance of the prior distributions.

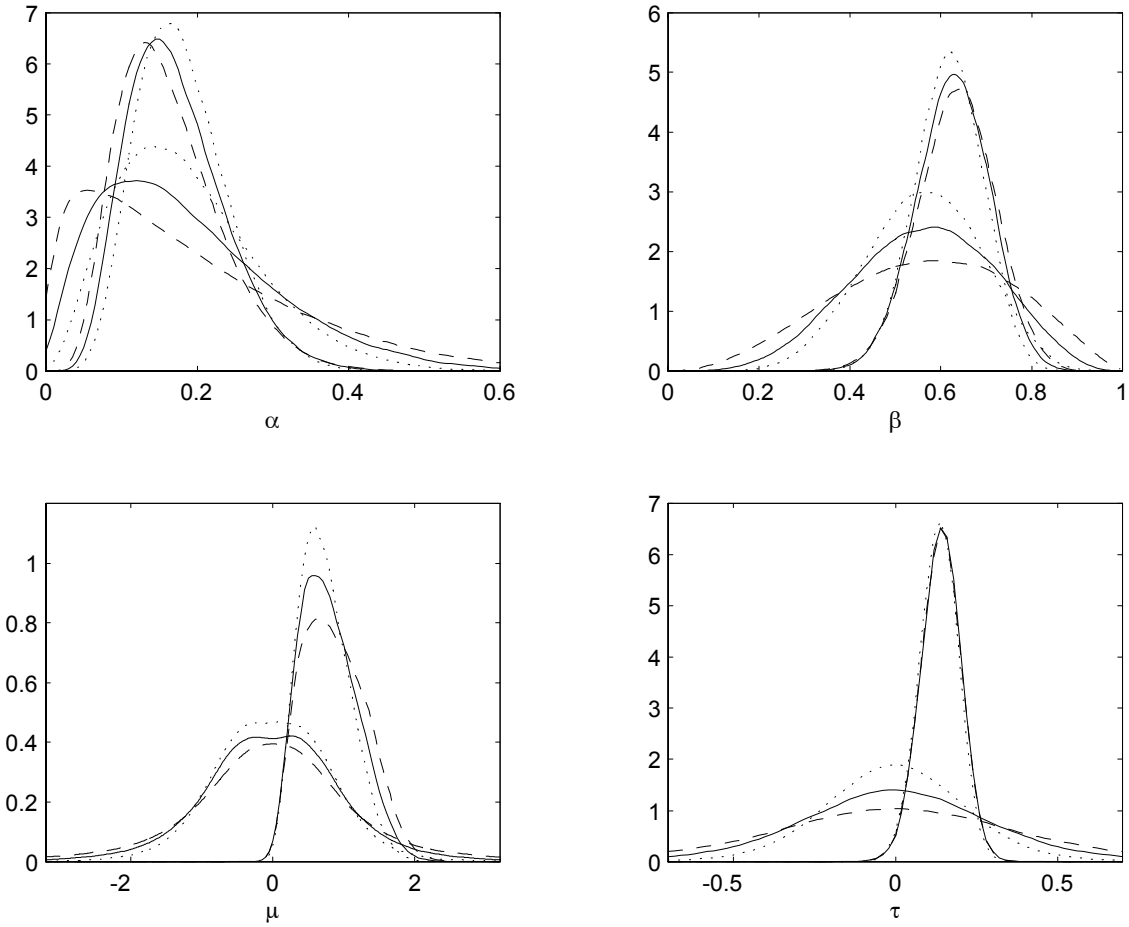


— Baseline prior and posterior

– – Large variance prior and posterior

· · · Small variance prior and posterior

Figure 4: Sensitivity of the simulated posterior distributions of the dynamic variance parameters α, β, μ, and τ to increases and decreases in the variance of the prior distributions.



— Baseline prior and posterior

– – Large variance prior and posterior

· · · Small variance prior and posterior