STOCHASTIC VOLATILITY: UNIVARIATE AND MULTIVARIATE EXTENSIONS

by

Eric Jacquier
Nicholas G. Polson
Peter Rossi

19-95

RODNEY L. WHITE CENTER FOR FINANCIAL RESEARCH
The Wharton School
University of Pennsylvania
3254 Steinberg Hall-Dietrich Hall
Philadelphia, PA 19104-6367
(215) 898-7616

Stochastic Volatility: Univariate and Multivariate Extensions

Eric Jacquier

Nicholas G. Polson

Peter Rossi *

December 1993

Revised May 1995

*Polson and Rossi: Graduate School of Business, University of Chicago. Jacquier is the corresponding author, on leave from Johnson School of Management, Cornell University, address: Wharton School - Finance Dept., 2323 Steinberg Hall - Dietrich Hall, University of Pennsylvania, Philadelphia PA 19107. Ph: (215) 898-4239, email: jacquier@jacquier.wharton.upenn.edu. Support from the Institute for Quantitative Research in Finance (Q GROUP) and the CIRANO (Centre Interuniversitaire d'Analyse des Organisations, Montreal) is gratefully acknowledged. We received helpful comments from Torben Andersen, Tim Bollerslev, John Geweke, Eric Ghysels, Dan Nelson, Esther Ruiz, George Tauchen, and the participants of the CIRANO / CRDE conference on stochastic volatility in Montreal, October 94.

Stochastic Volatility: Univariate and Multivariate Extensions

Abstract

Discrete time stochastic volatility models (hereafter SVOL) are noticeably harder to estimate than the successful ARCH family of models. Recent advances in the literature now make it possible to produce efficient estimation and prediction for a basic univariate SVOL model. However, the basic model may be insufficient for numerous economics and finance applications. In this paper, we develop methods for finite sample inference, smoothing, and prediction for a number of univariate and multivariate SVOL models. Specifically, we model fat-tailed and skewed conditional distributions, correlated errors distributions (leverage effect), and two multivariate models, a stochastic factor structure model and a stochastic discount dynamic model. We apply some of these extensions to financial series. We find (1) strong evidence of non-normal conditional distributions for stock returns and exchange rates, and (2) evidence of correlated errors for stock returns. These departures from the basic model affect the measured persistence and hence the prediction of volatility. This result as a policy implication on decisions and models such as asset allocation and option pricing which inputs are prediction of volatility.

We specify the models as a hierarchy of conditional distributions: $p(data \mid volatilities)$, $p(volatilities \mid parameters)$ and $p(parameters)$. Given a model and the data, inference and prediction are based on the joint posterior distribution of the volatilities and the parameters which we simulate via Markov chain Monte Carlo (herafter MCMC) methods. This approach also provides a sensitivity analysis for parameter inference and an outlier diagnostic. The hierarchical framework is a natural environment for the construction of SVOL models departing from standard distributional assumptions.

1 Introduction

Time varying volatility is a characteristic of many financial series. The SVOL model, a natural alternative to the popular ARCH framework, allows both the conditional mean and variance to be driven by separate stochastic processes. Empirical evidence implies that the additional flexibility of the SVOL over the ARCH model is necessary, e.g., Hsieh (1991), Danielsson (1994) and Geweke (1994). Likelihood-based analysis of SVOL models is notoriously difficult and, until recently, there have been few implementations of such models. Early research relied on non likelihood-based estimation techniques like the Method of Moments (MM) or Quasi Maximum Likelihood (QML), e.g., Melino and Turnbull (1990) and Harvey, Ruiz, and Shephard (1993). Jacquier, Polson and Rossi (1994), (JPR, henceforth), document the efficiency gains of the MCMC technique over the QML and the method of moments, casting doubt on their reliability at conventional sample sizes. Danielsson (1994) shows how to compute simulated maximum likelihood estimates of the parameters of a basic SVOL model. This procedure relies on asymptotic approximations and does not solve the smoothing and prediction problem. JPR introduced MCMC techniques for the analysis, estimation, smoothing, and prediction, of a univariate stochastic volatility model with a normal conditional distribution for the observations.

First, following the ARCH literature, see Bollerslev, Chou, and Kroner (1992), it is natural to investigate the normality of the conditional distribution. Indeed, Gallant, Hsieh, and Tauchen (1994) show that the conditional distribution is non normal for stock returns. Second, one may also need to extend the SVOL model to allow for the so-called leverage effect where changes in volatility are asymmetrically related to the sign and magnitude of price changes, see the EGARCH model of Nelson (1991). Introducing a negative correlation between the errors of the variance and the observables, e.g., stock returns, produces this leverage effect. These two extensions have empirical and economic importance because they modify volatility predictions, and in turn, economic models based upon volatility predictions such as option pricing models, see Hull and White (1987), Heston (1993), and Jacquier, Polson and Rossi (1995). Finally, there is a clear need for multivariate models, for example in portfolio and asset pricing applications. Mueller and Pole (1994) discuss inference issues for multivariate ARCH-style models. In this paper we develop MCMC algorithms for the implementation of stochastic volatility models that allow for: fat tailed and skewed conditional errors, correlated errors, and multivariate volatility strucutres. Our methodology allows the implementation of dynamic time series models with a stochastic discount factor (see West and Harrison (1989) for a number of applications) or factor structures allowing for time varying correlations. One attractive feature of our modelling approach is the substantial commonality among the algorithms for different model specifications. We show how this commonality arises and how the hierarchical structure can be used to develop further generalizations.

The need for implementing SVOL models with these generalizations is clear. It is confirmed by our empirical findings in section 4. The relatively poor performance of approximate methods documented in JPR and Andersen and Sørensen (1994) for the basic model casts doubt on the ability of approximate methods to handle these extensions well. A problem with the QML algorithm is that it approximates the logarithm of the squared conditional error, i.e., of a χ^2 in the basic model, to a normal distribution. Kim and Shephard (1994) show that the $\log(\chi^2)$ can be well approximated by a discrete mixture of normals. Mahieu and Schotman (1994) consider a discrete mixture of normals with unknown parameters which allows for a flexible distribution for the log of the squared conditional error. One disadvantage of the mixture of normals framework is that it does not extend naturally to provide approximations in the correlated errors or multivariate cases. Our inference is based on the exact joint posterior distribution of the unobserved volatility sequence and parameters of the volatility process. MCMC algorithms allow us to simulate from this joint posterior distribution without resorting to model approximations. Moreover, this approach simultaneously addresses the problems of smoothing, prediction and parameter estimation under the non-standard distributional assumptions described above. This flexibility is crucial for finance applications which require not only parameter inference but prediction of future volatilities as option pricing.

We employ a three stage hierarchical model. The hierarchical model is composed of the conditional probability distributions: $p(\text{data}|\text{volatilities})$, $p(\text{volatilities}|\text{parameters})$ and $p(\text{parameters})$. The first stage distribution, $p(\text{data}|\text{volatilities})$, models the distribution of the data given the volatilities. In the basic SVOL model, this latter distribution is normal, but not in the extensions discussed above. We will show that non-normality and correlations are fairly straightforward to implement in our framework. For all but one of our financial series, we find evidence that the data favour a fat-tailed distribution over the generally assumed normal distribution. We also find some evidence of negative correlation. The second stage distribution, p(volatilities|parameters), models prior beliefs about the stochastic evolution of the volatility sequence. For example, it is generally assumed that the logarithm of conditional volatility follows an AR (1) process. See Jacquier, Polson, and Rossi (1995) for a discussion of models in the level of volatility. We find that the degree of persistence of the AR(1) changes when non-normality is allowed in the first stage distribution. Therefore, the proper modelling of the first stage distribution affects the quality of prediction, and is of paramount concern for many uses of the model. We extend the second stage distribution to a multivariate setup. We describe two multivariate hierarchical models for p(volatilities|parameters): a stochastic factor structure model and a stochastic discount dynamic model. The third stage distribution, p(parameters) reflects beliefs about the parameters of the volatility process. Typically, this stage employs a diffuse distribution that constrains the parameters of the volatility process to the region of stationarity. But tighter priors are easily incorporated. We also examine the effects of distributional assumptions on the posterior distributions of the parameters of the volatility process. Thus, hierarchical models will prove extremely useful for researchers concerned with the influence of distributional assumptions on resulting inferences.

The advantage of the MCMC estimation methodology is that it simultaneously produces the exact posterior distributions of the parameters and of each of the unobservable variances, as well as the predictive density of future volatilities. The explicit joint estimation of both parameters and volatilities is more appealing than the ad hoc combinations of approaches found in other procedures. For example, point estimates from the method of moments have to be substituted into an approximate Kalman filter to obtain smoothed estimates of volatilities. The problems due to the inefficiency of the parameter estimation is compounded by the approximation in the Kalman filter. Also, further approximation is required to produce inference for nonlinear functions of the parameters. Exact inference for nonlinear functions of the parameters or variances is straightforward for our simulation based estimator because a draw of the joint distribution of the parameters produces a draw of the desired function by direct computation. MCMC procedures use a Markov chain to simulate draws from the joint posterior $p(\text{volatilities}, \text{parameters}|\text{data})$. Marginal distributions, for example $p(\text{volatilities}|\text{data})$, are computed by simply averaging the appropriate conditional distributions over the simulated draws.

The paper is organized as follows. Section 2 describes our hierarchical modelling framework for the proposed models. Section 3 develops the specific MCMC algorithms to implement the models. Finally, Section 4 considers empirical applications.

2 Extending the Basic Univariate SVOL Model

JPR develop a Metropolis algorithm for computing the joint posterior distribution in a univariate stochastic volatility model with normal errors. Although not presented in JPR as a hierarchical model, it is a particular example of the general framework discussed here. In that model, the first stage assumes normality, the second stage distribution specifies a smooth mean reversion for the volatility sequence where $p(\text{volatilities}|\text{parameters})$ is specified by a log AR(1) process and the third stage assumes a diffuse distribution for $p(\omega)$ restricted to the region of stationarity of the volatility process. The model for the observations y_t and the volatilities h_t is given by

$$y_t = h_t^{\frac{1}{2}} \epsilon_t, \tag{1}$$

$$\log h_t = \alpha + \delta \log h_{t-1} + \sigma_v v_t, \quad t = 1, \ldots, T$$

$$(\epsilon_t, v_t) \sim N_2(0, I) \ , \ (\alpha, \delta, \sigma_v) \sim p(\alpha, \delta, \sigma_v)$$

Here the parameter vector $\omega = (\alpha, \delta, \sigma_v)$ consists of a location α, a volatility persistence δ and a volatility of the volatility σ_v. We denote this standard model by $SVOL(\mathbf{h}, \omega)$.

The major assumption relaxed in this paper is that of an uncorrelated bivariate normal distribution for ϵ_t and v_t. The natural directions for an extension of the model are for example; skewed and fat-tailed error distributions; correlated errors. Straightforward alternative extensions not considered here include regressors in the mean/volatility equations; 'jumps' in the volatility process as in McCulloch and Tsay (1993). They can be accomodated by the hierachical framework which we use. We also describe two multivariate hierarchical models for stochastic volatility, a stochastic discount factor model and a stochastic factor structure model. Finally, we address the issue of sensitivity analysis available in our approach.

2.1 Fat-tailed Departures from Normality

One of the conclusions of the literature on the estimation of ARCH models for financial time series (see Bollerslev, Chou, and Kroner (1992)) is that the conditional errors are non-normally distributed with fat tails. To perform a similar diagnostic with the stochastic volatility model, it is necessary to extend the model to allow for fat tails in the error of the mean equation. A fat tail in the error of the variance equation can be modelled in a fashion nearly identical to tha shown below.

In the simple stochastic volatility model defined in equation (**??**) ϵ_t is modelled as a standard normal. In order to fatten the tail of this distribution whilst keeping symmetry, we model the distribution $p(\epsilon_t)$ as a scale mixture of normals, a special case of which is the student-t distribution. Consider the stochastic volatility model defined by the conditionals

$$y_t = h_t^{\frac{1}{2}}\epsilon_t, \quad \epsilon_t \sim \int N(0, \lambda_t)p(\lambda_t)d\lambda_t \tag{2}$$

$$\log h_t = \alpha + \delta \log h_{t-1} + \sigma_v v_t, \quad t = 1, \ldots, T$$

$$(\alpha, \delta, \sigma_v) \sim p(\alpha, \delta, \sigma_v)$$

where the distribution $p(\lambda_t)$ is chosen to reflect the appropriate fat-tailness required in the distribution $p(\epsilon_t)$. By varying $p(\lambda_t)$ the distribution $p(\epsilon_t)$ can exhibit a wide range of fat-tailed behaviour ranging from the double exponential, exponential power, stable, logistic, or t-family of distributions (see Carlin and Polson (1991), for details). Within the hierachical framework it is also straightforward to model heavy-tailed errors for v_t in the variance equation. Apply the same scale mixture of normals idea and adapt the modelling of section 3.3.

A natural choice is to model the distribution $p(\epsilon_t)$ as a t_ν-distribution with ν degrees of freedom. The appropriate choice of $p(\lambda_t)$ for this specification is an inverse gamma distribution, specifically $\nu/\lambda_t \sim \chi^2_\nu$ for a hyperparameter ν. Given a fixed ν, this extends the basic model and can be thought of as an outlier robustification. Fixing ν at small values will downweight outliers in the estimation of stochastic volatility parameters. Posterior inferences about λ_t can be used as an outlier diagnostic. The scale mixture model accomodates an outlier at time t by increasing the scale. Observations with marginal posteriors centered on large values of λ_t are candidates for outliers.

The hierarchical structure allows us to go one step further and model ν with a distribution $p(\nu)$. It is then possible to let the data infer about the severity of the departure from normality through the marginal posterior distribution $p(\nu|\mathbf{y})$.

In Section 3 we describe how to construct a MCMC algorithm to implement this model. It will be useful to re-write the model as

$$y_t = h_t^{\frac{1}{2}} \lambda_t^{\frac{1}{2}} z_t, \tag{3}$$

$$\log h_t = \alpha + \delta \log h_{t-1} + \sigma_v v_t, \quad t = 1, \ldots, T$$

$$(\alpha, \delta, \sigma_v, \lambda) \sim p(\alpha, \delta, \sigma_v) p(\lambda)$$

$$(z_t, v_t) \sim N_2(0, I)$$

where $p(\lambda) = \prod_{t=1}^{T} p(\lambda_t)$. We will use the vector of latent variables λ to construct an algorithm to simulate from the joint posterior $\pi(\mathbf{h}, \omega, \lambda|\mathbf{y})$. The posterior distribution of interest $\pi(\mathbf{h}, \omega|\mathbf{y})$ is then a marginal distribution from this joint distribution. See Section 3.2 for further details.

2.2 Skewed Departures from Normality

One simple way of introducing an asymmetric skewed distribution for ϵ_t is to consider a contaminated normal mixture of the form $\epsilon_t|\epsilon, \mu, \tau \sim (1-\epsilon)N(-\eta\mu, 1) + \epsilon N(\mu, \tau^2)$ for parameters (μ, τ, ϵ). We chose $\eta = \epsilon/(1-\epsilon)$ to force the mean of the conditional error ϵ_t to be zero. Note that the framework here can extend beyond the modelling of conditional skewness. For example, when $\mu = 0$ we have a symmetrical fat-tailed mixture distribution. The conditional variance of returns is multiplied by τ^2 with probability ϵ. This is similar to the regime switching model used by Hamilton and Susmel (1994) for ARCH models. The difference is in the formulation of transition probabilities. Our main goal is to model skewness, so we introduce one probability of being in state 1 or 2, ϵ and $1-\epsilon$. Hamilton and Susmel think of regimes and model two transition probabilities from state 1 to state 2, and from state 2 to state 1. With $\mu = 0$, our model is easily extended to accommodate this regime switching framework which appears to be a successful extension of the basic ARCH framework.

The parameter μ is used to reflect asymmetrical departures in the error distribution. The parameter ϵ helps model a probability of each observation being an "outlier". Heuristically, μ governs skewness and τ influences the kurtosis of the error distribution. In our hierarchical framework, these parameters can themselves be modelled with a distribution $p(\epsilon, \mu, \tau)$. The model is specified as

$$y_t = h_t^{\frac{1}{2}} \epsilon_t, \quad \epsilon_t \sim (1-\epsilon)N(-\eta\mu, 1) + \epsilon N(\mu, \tau^2) \tag{4}$$

$$\log h_t = \alpha + \delta \log h_{t-1} + \sigma_v v_t, \quad t = 1, \ldots, T$$

$$(\alpha, \delta, \sigma_v, \epsilon, \mu, \tau) \sim p(\alpha, \delta, \sigma_v) p(\epsilon) p(\mu, \tau)$$

A standard approach to re-writing the first stage of the hierarchy that allows easy implementation of MCMC methods is to use a set of unobserved indicator variables $\mathbf{I} = \{I_t\}$. Each error term ϵ_t has an indicator variable I_t with conditional distribution $\epsilon_t | I_t = 0, \epsilon, \mu, \tau \sim N(-\eta\mu, 1)$ and $\epsilon_t | I_t = 1, \epsilon, \mu, \tau \sim N(\mu, \tau^2)$ where $p(I_t = 1|\epsilon) = \epsilon$. Under these conditionals we get the appropriate marginal mixture distribution for $\epsilon_t | \epsilon, \mu, \tau \sim (1-\epsilon)N(-\eta\mu, 1) + \epsilon N(\mu, \tau^2)$. The third stage prior distribution for the vector of indicator variables is given by $p(\mathbf{I}|\epsilon) = \prod_{t=1}^{T} p(I_t|\epsilon)$. This allows the first stage of the model in (??) to be re-written as

$$\frac{h_t^{-\frac{1}{2}} y_t - (-\eta)^{1-I_t}\mu}{\tau^{I_t}} = z_t \ , \ z_t \sim N(0,1)$$

$$p(\mathbf{I}|\epsilon) = \prod_{t=1}^{T} p(I_t|\epsilon)$$

One of the important features of our analysis is that we can use the data to learn about the parameters (τ, μ, ϵ). In particular, given the data, we can compute marginal posteriors for the (μ, ϵ, τ) parameters. For example; the posterior distribution $p(\mu|\mathbf{y})$ can be used to assess the severity of the departure from symmetry whilst $p(\epsilon|\mathbf{y})$ and $p(\tau|\mathbf{y})$ can be used to assess the proportion of outliers and fat-tailedness, respectively.

2.3 Correlated Errors Model

The final extension of the univariate SVOL model which we consider is a correlation between the two errors ϵ_t and v_t. This may be particularly important for stock returns. For example, with a negative correlation, a decrease in price, i.e. a negative return ϵ_t is likely to be associated with a positive variance shock v_t, therefore producing the behavior referred to as the leverage effect. As a hierarchical model we have

$$y_t = h_t^{\frac{1}{2}} \epsilon_t,$$

$$\log h_t = \alpha + \delta \log h_{t-1} + \sigma_v v_t, \quad t = 1, \ldots, T$$

$$(\alpha, \delta, \sigma_v, \rho) \sim p(\alpha, \delta, \sigma_v, \rho)$$

$$(\epsilon_t, v_t) \sim N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

The ex post analysis of this model now requires simulation from the joint posterior distribution $\pi(\mathbf{h}, \omega, \rho|\mathbf{y})$. Two difficulties arise. First, one needs to specify an appropriate joint prior distribution for the correlation ρ and the volatility of volatility σ_v. Second, the algorithm needs to be modified to accommodate the correlation ρ. We address these problems by reparameterising the model in terms of $(\alpha, \delta, \Sigma^*)$, where Σ^* and the joint distribution of the data and the volatilities are given by

$$p(\mathbf{y}, \mathbf{h}|\alpha, \delta, \Sigma^*) = \prod_{t=1}^{T} h_t^{-\frac{3}{2}} p(h_t^{-\frac{1}{2}} y_t, \log h_t | h_{t-1}, \alpha, \delta, \Sigma^*)$$

$$\Sigma^* = \begin{pmatrix} 1 & \rho\sigma_v \\ \rho\sigma_v & \sigma_v^2 \end{pmatrix}$$

It is algebraically easier to rewrite this joint distribution in terms of the residuals (r_t, u_t) where $\mathbf{r}_t = (h_t^{-\frac{1}{2}} y_t, u_t)$ and $u_t = \log h_t - \alpha - \delta \log h_{t-1}$ for $1 \leq t \leq T$. The joint distribution of the data and the volatilities is now

$$p(\mathbf{y}, \mathbf{h}|A, \Sigma^*) = \prod_{t=1}^{T} h_t^{-\frac{3}{2}} |\Sigma^*|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathrm{tr}(\Sigma^{*-1} A)\right)$$

where $A = \sum_t \mathbf{r}_t \mathbf{r}_t'$ is the residual matrix. In Section 3 we use this likelihood function to develop an MCMC algorithm for the analysis of the posterior distribution.

The difficulty in specifying a prior $p(\Sigma^*)$ on Σ^* is that the top left element is equal to one. Unfortunately, we cannot use the standard conjugate the inverse Wishart family as it is impossible to model beliefs where some elements of the matrix are known and others are not. We rewrite the covariance matrix with a hierarchical structure to the probability distribution. Let $\psi = \rho\sigma_v, \Omega = \sigma_v^2(1-\rho^2)$ with inverse transformation $\sigma_v^2 = \Omega + \psi^2$ and $\rho = \psi/\sqrt{\Omega+\psi^2}$. We elicit the prior distribution on Σ^* in a conditional fashion via $p(\Sigma^*) = p(\psi|\Omega)p(\Omega)$. The intuition of this reparameterization is straightforward: we reformulate the covariance matrix of r_t and u_t in terms of the linear regression of u_t on r_t. The new parameters are the slope coefficient and the variance of the noise for which a normal-gamma prior is natural. We report quantiles of the exact posterior distribution of ρ to document its uncertainty. Our empirical results will show that the posterior for ρ is relatively diffuse although its mass is concentrated primarily on negative values. Given this and the bounded nature of the correlation parameter, it is all the more important to make finite sample inferences without resorting to asymptotic methods.

2.4 A Stochastic Discount Factor Model

In the multivariate case the data vector $\mathbf{y}_t$ is a $q \times 1$ vector of (excess) asset returns. The stochastic discount factor model is defined by

$$\mathbf{y}_t = h_t^{\frac{1}{2}} \Sigma^{\frac{1}{2}} \epsilon_t,$$

$$\log h_t = \alpha + \delta \log h_{t-1} + \sigma_v v_t, \quad t = 1, \ldots, T$$

$$(\alpha, \delta, \sigma_v, \Sigma) \sim p(\alpha, \delta, \sigma_v)\, p(\Sigma)$$

where (ϵ_t, v_t) is typically modelled as a standard $q+1$ multivariate normal. The indeterminacy in the level of variance is resolved by the use of priors or by simply fixing α. Extensions to fat-tailed distributions or a vector of correlations between ϵ_t and v_t can be made easily by adapting the models in Sections 2.1 and 2.3. This model is a stochastic generalisation of the discount dynamic model using extensively in Bayesian forecasting (see Harrison and Stevens (1976), and West and Harrison (1989)). Quintana and West (1989) provide an application of discount dynamic models to exchange rate data. An implication of this model however is that asset returns cross-correlations are non stochastic. On the other hand the covariance matrix Σ is not constrained by any simplifying assumption.

2.5 Factor Structure Multivariate Stochastic Volatility Model

An alternative to the previous specification is a factor structure for the variance covariance matrix, in which the factors possess stochastic volatility. This allows the cross-correlations to be stochastic. The factor structure, widely hypothesized in asset pricing, helps reduce the dimensionality of the parameter space. This type of specification is needed in the study of cross-sections of assets when dynamic factors are hypothesized to be priced. See Engle, Ng, and Rothschild (1990) for an incorporation of the factor structure into the ARCH framework. See also Harvey, Ruiz, and Shephard (1994). Consider q assets with returns $\mathbf{y}_t = (y_{1t}, \ldots, y_{qt})'$ where $1 \leq t \leq T$. Suppose that there are k underlying time varying factors $\mathbf{F}_t = (F_{1t}, \ldots, F_{kt})'$ that generate the returns according to

$$\mathbf{y}_t = \Delta \mathbf{F}_t + \Omega^{-\frac{1}{2}} \epsilon_t,$$

$$\mathbf{F}_t \sim N_k(\mathbf{0}, \mathbf{H}_t)$$

$$\log h_{it} = \alpha_i + \delta_i \log h_{i,t-1} + \sigma_{iv} \mathbf{v}_t,$$

$$p(\Delta, \Omega, \omega) = p(\Omega) \prod_{i=1}^{k} p(\Delta_i) p(\omega_i)$$

where $\mathbf{H}_t = diag(h_{it})$, $\Delta = (\Delta_1, \ldots, \Delta_k)$ is a matrix of factor loadings, and $\epsilon_t \sim N(0, I_q)$. Let $\delta = (\delta_1, \ldots, \delta_k)$ be the vector of persistence parameters for the volatilities of each factor and $\Sigma_v = diag(\sigma_{iv})$ the matrix of volatilities. Then, this model assumes that the k underlying factors are independent of each other and follow a simple univariate stochastic volatility model as in JPR. Under these assumptions we have a time-varying factor structure to the evolution of the variance of the observations $Var(\mathbf{y}_t) = \Delta H_t \Delta' + \Omega$. We do not necessarily assume that Ω is diagonal. That is, the k factors are introduced to model the stochastic variation in the covariance matrix of $\mathbf{y}_t$. If one wants to assume that the factors also explain the non-diagonality of this covariance matrix, then Ω can be assumed to be diagonal.

The advantage of this specification of the multivariate evolution of the covariance matrices is that the posterior distribution $\pi(\mathbf{F}, \Delta, \omega | \mathbf{y})$ can be broken into conditionals $\pi(F_j | F_{(-j)}, \Delta, \omega, \mathbf{y})$ $(1 \leq j \leq k)$ and $\pi(\Delta, \omega | \mathbf{F}, \mathbf{y})$ where the first set of conditionals can be simulated via a simple transformation of the JPR Metropolis algorithm used in the univariate model (see section 3.7 for details).

2.6 Parameter Estimation and Sensitivity Analysis

Given a model and the data $\mathbf{y}$, inference about the volatilities $\mathbf{h}$ and parameters ω is given by the joint distribution $\pi(\mathbf{h}, \omega|\mathbf{y})$. In the univariate volatility case, the posterior distribution $\pi(\mathbf{h}, \omega|\mathbf{y})$ is known up to proportionality and is given by

$$\pi(\mathbf{h}, \omega|\mathbf{y}) \propto \prod_{t=1}^{T} p(y_t|h_t) \prod_{t=1}^{T} p(h_t|h_{t-1}, \omega) p(\omega)$$

Parameter inference is based on the marginal posterior distribution $p(\omega|\mathbf{y})$. JPR (1994) provide a number of examples for stocks, stock indices and exchange rates. The empirical evidence shows that the posterior distribution for the persistence parameter $p(\delta \mid \mathbf{y})$ is concentrated away from the random walk case ($\delta = 1$) but exhibits strong skewness. This renders standard classical asymptotics inappropriate. For asymptotic methods, the use of asymptotic standard errors and the normality assumption may result in significant probabilities of explosive behavior or negative standard deviation. A similar problem is reported by Harvey and Shephard (1993) with respect to the coefficient ρ. The asymptotic normal approximation to the distribution of the QML estimate of ρ puts mass outside [-1,1]. The simulation based estimator allows us to obtain quantiles of the exact posterior distribution of the parameters or variances.

In the fat-tailed error model, we can address a number of inference problems as well as sensitivity issues. For example, consider the model in section 2.1. Sensitivity of inference for δ to the normal error distribution is conducted using the conditional posterior distribution $p(\delta|\nu, \mathbf{y})$ for which we have draws. Outliers are pinpointed using the draws of the posterior distribution of the latent scale variable λ_t. In our empirical examples, we compare the posterior mean $E(\lambda_t|\mathbf{y})$ with the 95 % quantiles of the prior distribution $p(\lambda_t)$ for each observation. In the hierarchical framework it is also possible to conduct inference about the size of departure from normality in the error distribution. The degrees of freedom parameter ν is modelled with a distribution $p(\nu)$ and the data infer about the severity of the departure from normality through the marginal posterior distribution $p(\nu|\mathbf{y})$. In all but one of our series we find evidence for non-normality.

There is a number of other advantages to using our approach to inference. Aside from the unifying perspective on implementing these models, inference for nonlinear functions of the parameters or variances is straightforward. Prediction for unobservables like future average volatility, needed in option pricing is easy (see Jacquier, Polson, and Rossi (1994a)). Typical nonlinear parameters of interest can be the half-life of a shock to volatility $\log(.5)/\log\delta$, or the coefficient of variation of the volatility sequence $\exp\left(\sigma_v^2/1-\delta^2\right)-1$. JPR provide a number of examples for stocks, stock indices and exchange rates. Our simulation based approach therefore provides an exact finite sample inference with no need to appeal to delta-method asymptotics.

3 Implementation of SVOL models via MCMC Algorithms

To construct a Markov chain Monte Carlo algorithm for simulating draws from a stochastic volatility model, it is usual to consider the following conditional posterior distributions $\pi(\mathbf{h}|\omega,\mathbf{y})$ and $\pi(\omega|\mathbf{h},\mathbf{y})$. While the latter distribution is available for direct simulation, the former is not. Under the normal assumption, a number of solutions have been proposed. JPR (1994) decompose $\pi(\mathbf{h}|\omega,\mathbf{y})$ further by considering the one dimensional conditionals $\pi(h_t|h_{(-t)},\omega,\mathbf{y})$ where $h_{-t}=(h_1,\ldots,h_{t-1},h_{t+1},\ldots,h_T)$. A Metropolis algorithm is used to perform this step of the algorithm.

For the models that depart from the normal assumptions, we simulate from the joint posterior distribution of all the volatilities, the parameters and additional state variables. For example, in the fat-tailed model the state space $\mathbf{x}=(\mathbf{h},\omega,\lambda)$ where λ is the vector of scale parameters, and in the skewed departures $\mathbf{x}=(\mathbf{h},\omega,\mathbf{I})$ where $\mathbf{I}$ is the vector of indicators. A MCMC algorithm specifies an irreducible and aperiodic Markov chain, see Tierney (1991), with stationary distribution given by the desired joint distribution. There is a variety of choices for the underlying Markov chain. The Hastings-Metropolis algorithms (Hastings (1970)) provide a family of algorithms that can be defined using only *local* movements in the parameter space and knowledge of the joint posterior only up to proportionality. To simulate elements from the posterior distribution we proceed by picking an initial state, possibly at random, and then simulating transitions of the chain. The draws of the simulated distribution converge to draws of the stationary distribution namely the required joint posterior.

3.1 Metropolis algorithm

A Metropolis algorithm requires the specification of a Markov transition kernel, Q, density ratio $\pi(\mathbf{y})/\pi(\mathbf{x})$ and state space for the chain. The researcher then runs the Markov chain based on Q and 'adjusts' its movements using the density ratio as follows: suppose the chain is at $\mathbf{x}$, generate a candidate point $\mathbf{y}$ from $Q(\mathbf{x},\mathbf{y})$. If $\pi(\mathbf{y})Q(\mathbf{y},\mathbf{x}) \geq \pi(\mathbf{x})Q(\mathbf{x},\mathbf{y})$, the chain goes to $\mathbf{y}$. If $\pi(\mathbf{y})Q(\mathbf{y},\mathbf{x}) < \pi(\mathbf{x})Q(\mathbf{x},\mathbf{y})$, accept the new draw $\mathbf{y}$ with probability $\pi(\mathbf{y})Q(\mathbf{y},\mathbf{x})/\pi(\mathbf{x})Q(\mathbf{x},\mathbf{y})$, otherwise repeat the last draw $\mathbf{x}$. This new process is a Markov chain that is time reversible with respect to π and therefore π is its stationary distribution. The acceptance probability of a new draw is related to the transition kernel and the density ratio by:

$$\alpha(\mathbf{x},\mathbf{y}) = \min\left(\frac{\pi(\mathbf{y})Q(\mathbf{y},\mathbf{x})}{\pi(\mathbf{x})Q(\mathbf{x},\mathbf{y})}, 1\right)$$

Tierney (1991) and Müller (1991) describe extensions and implementations of Metropolis-type algorithms. Note how this algorithm nests the cases when we can make direct draws from π. Then, $Q(\mathbf{x},\mathbf{y}) \propto \pi(\mathbf{y})$, the acceptance probability is one, and we are back to a standard Gibbs sampling framework.

3.2 JPR univariate Metropolis algorithm

In the univariate stochastic volatility model with posterior $\pi(\mathbf{h},\omega|\mathbf{y})$, JPR (1994) use a product form for the Metropolis algorithm of the form

$$Q(\mathbf{x},\mathbf{y}) = Q_\omega(\mathbf{x},\mathbf{y}) \prod_{t=1}^{T} Q_{h_t}(\mathbf{x},\mathbf{y})$$

Here the suffix is used to indicate which component of $\mathbf{x} = (\mathbf{h},\omega)$ gets updated conditional on the rest of the variables. The one dimensional blanket $Q_{h_t}(\mathbf{x},\mathbf{y})$ is specified to mimic the conditional $\pi(h_t|h_{(-t)},\omega,\mathbf{y})$. Specifically $Q_{h_t}(\cdot,\cdot)$ is an inverse gamma density $h_t \sim 1/Z$ where $Z \sim Gamma(\phi, 1/\psi_t)$ with parameters

$$\psi_t = (\phi - 1)e^{\mu_t + .5\sigma^2} + .5y_t^2 \tag{5}$$

$$\phi = \frac{1-2e^{\sigma^2}}{1-e^{\sigma^2}} + .5$$

with μ_t and σ are given in JPR. The blanket $Q_\omega(\mathbf{x}, \mathbf{y})$ which updates the parameter vector given the volatility sequence can be set equal to the conditional posterior $\pi(\omega|\mathbf{h}, \mathbf{y})$ since the latter is available from standard linear models methodology.

3.3 Fat-tailed Departures from Normality

For latent variable modelling we augment the posterior distribution with the vector of unobserved latent variables λ. A MCMC algorithm needs to be constructed for the full posterior $\pi(\mathbf{h}, \omega, \lambda \mid \mathbf{y})$. This is done by decomposing the distribution in terms of the conditionals

$$\pi(\mathbf{h}, \omega \mid \lambda, \mathbf{y})$$

$$\pi(\lambda \mid \mathbf{h}, \omega, \mathbf{y})$$

We can now construct an algorithm to simulate from the joint posterior $\pi(\mathbf{h}, \omega, \lambda|\mathbf{y})$. Notice that we can simulate from $\pi(\mathbf{h}, \omega|\lambda, \mathbf{y})$ using the original algorithm with y_t replaced by $\lambda_t^{-\frac{1}{2}} y_t$. As for $\pi(\lambda \mid \mathbf{h}, \omega, \mathbf{y})$, it can be simulated using results from Carlin and Polson (1991).

To draw from the conditional $\pi(\lambda \mid \mathbf{h}, \omega, \mathbf{y}, \nu)$ notice that $\pi(\lambda \mid \mathbf{h}, \omega, \mathbf{y}, \nu) = \prod p(\lambda_t|\mathbf{h}, \omega, \mathbf{y}, \nu)$ where

$$p(\lambda_t|y_t, h_t, \nu) \equiv p(\lambda_t \mid h_t^{-1/2} y_t, \nu) \propto p(h_t^{-1/2} y_t \mid \lambda_t, \nu) p(\lambda_t \mid \nu) \tag{6}$$

The form of the conditional posterior therefore depends on the choice of prior $p(\lambda_t \mid \nu)$. Two useful fat-tailed distributions are given by the following priors and conditional posteriors

- If $p(\epsilon_t) \sim t_\nu$ then $\nu/\lambda_t \sim \chi^2_\nu$ and $\lambda_t|y_t, h_t \sim IG\left(\frac{1}{2}(\nu+1), 2\left(\frac{y_t^2}{h_t} + \nu\right)^{-1}\right)$
- If $p(\epsilon_t) \sim DE(0,1)$ then $\lambda_t \sim \text{Exp}(2)$ and $\lambda_t|y_t, h_t \sim GIG\left(\frac{1}{2}, 1, \frac{y_t^2}{h_t}\right)$

where DE denotes the double exponential distribution and GIG the generalized inverse gaussian distribution. To let the data gauge the severity of the departure from normality we can extend the hierarchical model by letting ν have a prior distribution $p(\nu)$. This adds a conditional distribution to our simulation procedure, namely $p(\nu|\lambda, \mathbf{h}, \omega, \mathbf{y}) \sim p(\nu|\lambda)$. By Bayes theorem,

$$p(\nu|\lambda, \mathbf{h}, \omega, \mathbf{y}) \propto p(\nu) \prod_{t=1}^{T} p(\lambda_t|\nu)$$

Therefore,

$$p(\nu|\lambda, \mathbf{h}, \omega, \mathbf{y}) \propto \frac{1}{(\prod_t \lambda_t)^{\nu+1}} \exp\left(-\frac{\nu}{2}\sum_t \frac{1}{\lambda_t^2}\right) \frac{\nu^{\nu/2}}{\Gamma(\nu/2)} p(\nu)$$

The prior distribution of ν is a simple discrete distribution on an integer interval such as [1,60] or [1,30]. Simulations from this one dimensional distribution can be handled directly.

3.4 Skewed Departures from Normality

Instead of exploring possible fat-tailed departures from the standard normal assumption we can model skewness and fat-tails directly by taking ϵ_t to be a contaminated normal mixture. Specifically, we take $\epsilon_t|\epsilon, \mu, \tau \sim (1-\epsilon)N(-\eta\mu, 1) + \epsilon N(\mu, \tau^2)$ for parameters (μ, τ, ϵ). These hyperparameters can themselves be modelled with a distribution $p(\epsilon, \mu, \tau)$. The hierarchical specification of this model is

$$\frac{h_t^{-\frac{1}{2}} y_t - (-\eta)^{1-I_t}\mu}{\tau^{I_t}} = z_t \ , \ z_t \sim N(0,1)$$

$$\log h_t = \alpha + \delta \log h_{t-1} + \sigma_v v_t, \quad t = 1, \ldots, T$$

$$(\alpha, \delta, \sigma_v, \mathbf{I}, \epsilon, \mu, \tau) \sim p(\omega)\, p(\mathbf{I}|\epsilon)\, p(\epsilon)\, p(\mu, \tau)$$

where $p(\mathbf{I}|\epsilon) = \prod_{t=1}^{T} p(I_t|\epsilon)$ with $p(I_t = 1|\epsilon) = \epsilon$. We now construct a MCMC algorithm to simulate from the joint posterior $p(\mathbf{h}, \omega, \mathbf{I}, \tau, \mu, \epsilon \mid \mathbf{y})$ where we have augmented the state space with the unobserved indicator variables. We can simulate from this joint distribution by cycling through the conditionals $p(\mathbf{h}, \omega \mid \mathbf{I}, \tau, \mu, \epsilon, \mathbf{y})$, $p(\mathbf{I}|\mathbf{h}, \tau, \mu, \epsilon, \mathbf{y})$, $p(\tau, \mu \mid \mathbf{h}, \mathbf{I}, \epsilon, \mathbf{y})$ and $p(\epsilon \mid \tau, \mu\mathbf{h}, \mathbf{I}, \mathbf{y})$. Let us now consider each of these conditionals. To simulate from the conditional $p(\mathbf{h}, \omega|\mathbf{I}, \tau, \mu, \epsilon, \mathbf{y})$ notice that given $\mathbf{I}, \tau, \mu, \epsilon$, the random variables $\{z_t\}$ are $N(0,1)$. As in the basic algorithm, this conditional is simulated by cycling through $p(\omega \mid \mathbf{h}, \cdot)$ and $p(\mathbf{h} \mid \omega, \cdot)$. The distribution of ω given $\mathbf{h}$ is identical to that of the simple model. The distribution of $\mathbf{h}$ given ω is broken down into univariate distributions $p(h_t \mid h_{-t}, \cdot)$ given by

$$p(h_t|h_{t-1}, h_{t+1}, \omega, \mathbf{I}, \tau, \mu, \epsilon, \mathbf{y}) \propto h_t^{-\frac{3}{2}} \exp\left\{-\frac{1}{2}\left(\frac{h_t^{-\frac{1}{2}} y_t - (-\eta)^{1-I_t}\mu}{\tau^{I_t}}\right)^2\right\} \exp\left\{-\frac{(\log h_t - \mu_t)^2}{2\sigma^2}\right\}$$

where (μ_t, σ) are given in JPR. We can use the same blanketing density as in (??) but with y_t^2 transformed to y_t^2/τ^{2I_t}. This accounts for the extra variation τ in the skewed distribution. The remaining term $\exp((-\eta)^{1-I_t}\mu y_t/h_t^{\frac{1}{2}}\tau)$ accounts for the location shift μ. We partially incorporate it into the blanketing density by expanding it around the $\sqrt{1/h_t}$. The remainder term in the expansion appears is accomodated via the acceptance probability. As we iterate between this step and the indicators and hyperparameters we have a different data sequence for the volatilies each time. In fact, one can see how the hyperparameters affect the weighting of each observation in the conditional posterior of the volatility sequence.

The conditional posterior $p(\mathbf{I} \mid \mathbf{h}, \tau, \mu, \epsilon, \mathbf{y})$ is a product of independent univariate distributions $p(I_t \mid h_t, \tau, \mu, \epsilon, y_t)$, each of which is a simple binary draw where

$$p(I_t = 1 \mid h_t, \tau, \mu, \epsilon, y_t) \propto \frac{\epsilon}{\tau^{I_t}} \exp\left(-\frac{1}{2}\frac{(h_t^{-\frac{1}{2}}y_t - (-\eta)^{1-I_t}\mu)^2}{\tau^{2I_t}}\right) \tag{7}$$

Now consider the conditional posterior distribution of the parameters $p(\mu, \tau|\mathbf{h}, \mathbf{I}, \epsilon, \mathbf{y})$. It is known up to porportionality as

$$p(\mu, \tau \mid \mathbf{I}, \mathbf{h}, \epsilon, \mathbf{y}) \propto \frac{1}{\tau^{\sum_t I_t}} \exp\left(-\frac{1}{2}\sum_t \frac{(h_t^{-\frac{1}{2}}y_t - (-\eta)^{1-I_t}\mu)^2}{\tau^{2I_t}}\right) p(\mu, \tau)$$

where $p(\mu, \tau)$ is the third stage prior for the parameter. If we assume a normal/inverse gamma prior $p(\mu, \tau)$ we can simulate directly from the conditional posteriors

$$p(\tau \mid \mu, \mathbf{I}, \mathbf{h}, \epsilon, \mathbf{y}) \propto \frac{1}{\tau^{\sum_t I_t}} \exp\left(-\frac{1}{2}\sum_{t:I_t=1} \frac{(h_t^{-\frac{1}{2}}y_t - (-\eta)^{1-I_t}\mu)^2}{\tau^{2I_t}}\right) p(\tau)$$

$$p(\mu|\tau, \mathbf{I}, \mathbf{h}, \epsilon, \mathbf{y}) \propto \exp\left(-\frac{1}{2}\sum_t \frac{(h_t^{-\frac{1}{2}}y_t - (-\eta)^{1-I_t}\mu)^2}{\tau^{2I_t}}\right) p(\mu|\tau)$$

Notice that this differs from the standard conjugate implementation of the normal/inverse gamma prior where the joint posterior is directly available.

Finally, we need the conditional distribution of the outlier probability ϵ. We take ϵ to have a Beta prior distribution with hyperparameters ϵ_1 and ϵ_2. The Beta distribution is very flexible and can include a flat prior as a special case. However a flat prior is not appropriate for, presumably, we do not expect a large amount of outliers. For example, a prior mean of ϵ equal to 0.05, and $P(\epsilon < 0.5) = 0.99$, can be modelled by the hyperparameters $\epsilon_1 = 0.2$ and $\epsilon_2 = 3.5$. The conditional posterior distribution of ϵ is

$$\begin{aligned} p(\epsilon \mid \mu, \tau, \mathbf{I}, \mathbf{h}, \mathbf{y}) &\propto p(\epsilon)\, p(\mu, \tau, \mathbf{I}, \mathbf{h}, \mathbf{y} \mid \epsilon) \\ &\propto p(\epsilon)\, p(\mathbf{I} \mid \epsilon)\, p(\mathbf{y} \mid \mu, \tau, \mathbf{I}, \mathbf{h}, \epsilon) \end{aligned}$$

where we have used the fact that $p(\mathbf{h}, \mu, \tau) = p(\mathbf{h}, \mu, \tau | \epsilon, \mathbf{I})$. This is because ϵ and $\mathbf{I}$ carry no information on $\mathbf{h}, \mu, \tau$ in the absence of the data. Under the *Beta* prior assumption, we can combine the product $p(\epsilon)\, p(\mathrm{I} \mid \epsilon)$ by noting that $p(\mathrm{I} \mid \epsilon)$ corresponds to a Binomial experiment with t_1 successes, where t_1 is the number of I_t's equal to one. We have

$$p(\epsilon \mid \mu, \tau, \mathbf{I}, \mathbf{h}, \mathbf{y}) \propto Beta(\epsilon_1 + t_1, \epsilon_2 + T - t_1)\, p(\mathbf{y} \mid \mu, \tau, \mathbf{I}, \mathbf{h}, \epsilon) \tag{8}$$

The second term in this conditional is given by

$$p(\mathbf{y} \mid \mu, \tau, \mathbf{I}, \mathbf{h}, \epsilon) \propto \exp\left(-\frac{1}{2} \sum_t \frac{(h_t^{-\frac{1}{2}} y_t - (-\eta)^{1-I_t} \mu)^2}{\tau^{2I_t}} \right) \tag{9}$$

where $\eta = \epsilon/(1-\epsilon)$. Although simulation from the conditional distribution of ϵ in (??) is not direct it is only one dimensional. We use accept/reject by simulating from the $Beta(\epsilon_1 + t_1, \epsilon_2 + T - t_1)$ distribution and using (??) as an acceptance probability.

3.5 Correlated Errors Model

When the observation equation and the variance equation errors are correlated the joint posterior $\pi(\mathbf{h}, \Sigma^*, \alpha, \delta | \mathbf{y})$ is given by

$$\pi(\mathbf{h}, \Sigma^*, \alpha, \delta | \mathbf{y}) \propto p(\Sigma^*)\, p(\alpha, \delta) \prod_{t=1}^{T} h_t^{-\frac{3}{2}} |\Sigma^*|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathrm{tr}(\Sigma^{*-1} A) \right)$$

where $A = \sum_t \mathbf{r}_t \mathbf{r}_t'$ is the residual matrix with $\mathbf{r}_t = (h_t^{-\frac{1}{2}} y_t, \log h_t - \alpha - \delta \log h_{t-1})$. Under the reparameterisation of Σ^* in terms of $\psi = \rho \sigma_v, \Omega = \sigma_v^2(1 - \rho^2)$ we have a prior distribution on Σ^* given by $p(\Sigma^*) = p(\psi | \Omega) p(\Omega)$. It is natural to take a conditionally conjugate prior where $\psi | \Omega \sim N(\psi_0, \Omega^{-1} \tau_0^{-1})$ and $\Omega \sim IG(\alpha_0, \beta_0)$ for specified hyperparameters $(\alpha_0, \beta_0, \psi_0, \tau_0)$.

The analysis of the posterior distribution $\pi(\mathbf{h}, \alpha, \delta, \Sigma^* | \mathbf{y})$ is performed by cycling through the three conditionals

$$\pi(\mathbf{h} | \psi, \Omega, \alpha, \delta, \mathbf{y})$$

$$p(\psi, \Omega | \alpha, \delta, \mathbf{h}, \mathbf{y})$$

$$p(\alpha, \delta, | \psi, \Omega, \mathbf{h}, \mathbf{y})$$

The latter conditional is a normal distribution. The conditional $p(\psi, \Omega|\alpha, \delta, \mathbf{h}, \mathbf{y})$ is dealt with as follows:

$$p(\psi, \Omega|\alpha, \delta, \mathrm{h}, \mathbf{y}) \propto \frac{1}{\Omega^{T/2}} \exp\left(\Omega^{-1}\mathrm{tr}(CA)\right) p(\psi, \Omega)$$

$$C = \begin{pmatrix} \psi^2 & -\psi \\ -\psi & 1 \end{pmatrix}$$

Expanding the term $\mathrm{tr}(CA)$ we have $\mathrm{tr}(CA) = A_{22.1} + (\psi - \hat{\psi})' A_{11} (\psi - \hat{\psi})$ where $\hat{\psi} = A_{11}^{-1} A_{12}$. By conjugacy of the prior $p(\psi, \Omega)$ we can use standard linear bayes methodology to simulate the joint conditional posterior of (ψ, Ω) with the conditionals

$$p(\psi|\Omega, \alpha, \delta, \mathbf{h}, \mathbf{y}) \sim N\left(\tilde{\psi}, (A_{11} + \tau_0)^{-1}\right)$$

$$p(\Omega|\alpha, \delta, \mathbf{h}, \mathbf{y}) \sim IG\left(\alpha_0 + \frac{T}{2}, \beta_0 + \frac{1}{2}A_{22.1}\right)$$

where $\tilde{\psi} = (A_{11} + \tau_0)^{-1}(A_{11}\hat{\psi} + \tau_0\psi_0)$. Finally, we need to consider the conditional distribution $\pi(\mathbf{h}|\psi, \Omega, \alpha, \delta, \mathbf{y})$. As usual we cycle through the one dimensional conditional distributions $\pi(h_t|h_{t-1}, h_{t+1}, \Sigma^*, \alpha, \delta, \mathbf{y})$. By Bayes theorem,

$$\pi(h_t|h_{t-1}, h_{t+1}, \Sigma^*, \alpha, \delta, \mathbf{y}) \propto h_t^{-\frac{3}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma^{*-1}\mathbf{r}_t\mathbf{r}_t') - \frac{1}{2}\mathrm{tr}(\Sigma^{*-1}\mathbf{r}_{t+1}\mathbf{r}_{t+1}')\right)$$

We can now rewrite this in terms of (ψ, Ω) to get a conditional posterior

$$\pi(h_t|h_{t-1}, h_{t+1}, \psi, \Omega, \mathbf{y}) \propto h_t^{-\frac{3}{2}} \exp\left(-\frac{y_t^2}{2h_t}\left(1 + \frac{\psi^2}{\Omega^2}\right) - \frac{u_t^2}{2\Omega} - \frac{u_{t+1}^2}{2\Omega} + \frac{\psi y_t u_t}{\Omega\sqrt{h_t}} + \frac{\psi y_{t+1} u_{t+1}}{\Omega\sqrt{h_{t+1}}}\right)$$

As in JPR we use an inverse gamma blanketing density which is calculated by collecting together the terms in h_t^{-1} and the log-normal terms (see the uncorrelated case in section 3.1). The blanket $Q_{h_t}(\cdot, \cdot)$ is given by a $IG(\phi_t, \nu_t)$ distribution where

$$\phi_t = \frac{1 - 2\exp(\sigma^2)}{1 - \exp(\sigma^2)} + .5 + \frac{\psi\delta y_{t+1}}{\sqrt{h_{t+1}}} \tag{10}$$

$$\nu_t = (\phi_t - 1)\exp(\mu_t + .5\sigma^2) + \frac{1}{2}\left(1 + \frac{\psi^2}{\Omega^2}\right) y_t^2$$

and $\sigma^2 = \Omega/(1+\delta^2)$. The parameters of the inverse gamma have therefore been modified to account for the correlation parameter. The remaining term $\exp(\psi y_t u_t/\Omega\sqrt{h_t})$ is left in the acceptance probability of the Metropolis step.

3.6 Stochastic Discount Model

In the multivariate stochastic discount model we have a time varying covariance matrix of the form $h_t\Sigma$ for the $q \times 1$ observation vector $\mathbf{y}_t$. The discount factor h_t follows a standard log AR(1) volatility model and it is natural to model Σ with a standard inverse Wishart distribution. The analysis of the posterior distribution $\pi(\mathbf{h}, \omega, \Sigma|\mathbf{y})$ is performed by breaking it into the three conditionals

$$\pi(\mathbf{h}|\omega, \Sigma, \mathbf{y})$$

$$p(\omega|\Sigma, \mathbf{h}, \mathbf{y})$$

$$p(\Sigma|\omega, \mathbf{h}, \mathbf{y})$$

The last two conditionals follow directly from standard linear Bayes methodology: the posterior $p(\omega|\Sigma, \mathbf{h}, \mathbf{y}) \sim p(\omega|\mathbf{h})$ is the same as in the basic univariate case. $p(\Sigma|\omega, \mathbf{h}, \mathbf{y})$ is an inverse Wishart with updated parameters due to the conjugacy. We now show that the first conditional can be dealt with by using a simple transformation of the univariate JPR algorithm. The desired conditional $\pi(\mathbf{h}|\omega, \Sigma, \mathbf{y})$ is given up to proportionality as

$$\pi(\mathbf{h}|\omega, \Sigma, \mathbf{y}) \propto \prod_{t=1}^{T} h_t^{-\frac{1}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{y}_t'\Sigma^{-1}\mathbf{y}_t\right) p(\mathbf{h}|\omega)$$

It is equivalent to the posterior of the univariate case but with y_t^2 replaced by $\mathbf{y}_t'\Sigma^{-1}\mathbf{y}_t$. Conditional on Σ we can therefore use the univariate JPR algorithm to generate the $\mathbf{h}$ sequence.

3.7 Stochastic Factor Structure Model

The stochastic factor structure model is specified by the sequence of conditionals

$$\mathbf{y}_t = \Delta\mathbf{F}_t + \Omega^{-\frac{1}{2}}\epsilon_t,$$

$$\mathbf{F}_t \sim N_k(\mathbf{0}, \mathbf{H}_t)$$

$$\log h_{it} = \alpha_i + \delta_i \log h_{i,t-1} + \sigma_{iv} \mathrm{v}_t, \quad 1 \le i \le k$$

$$p(\Delta, \Omega, \omega) = p(\Omega) \prod_{i=1}^{k} p(\Delta_j) p(\omega_j)$$

where $\mathbf{H}_t = diag(h_{it})$ and $\Sigma_v = diag(\sigma_{iv})$. The joint posterior $\pi(\mathbf{H}, \omega, \Delta, \Omega | \mathbf{F}, \mathbf{y})$ is then given by

$$\pi(\mathbf{H}, \omega, \Delta, \Omega | \mathbf{F}, \mathbf{y}) \propto |\Omega|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Omega^{-1} \sum_t (\mathbf{y}_t - \Delta \mathbf{F}_t)'(\mathbf{y}_t - \Delta \mathbf{F}_t)\right)\right) p(\mathbf{H}|\omega) p(\omega) p(\Delta, \Omega)$$

where $p(\mathbf{H}|\omega)$ is the product of independent log AR(1) volatilies and Δ, Ω has a standard matrix normal/inverse Wishart prior distribution given by

$$p(\Delta|\Omega) \propto |\Omega|^{-\frac{k}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Omega^{-1}(\Delta - \Delta_0) A_0 (\Delta - \Delta_0)\right)\right)$$

$$p(\Omega) \propto |\Omega|^{-\frac{\nu_0}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Omega^{-1} B_0\right)\right)$$

for specified hyperparameters (A_0, B_0, ν_0).

We consider two cases. First, the underlying factors are observed as in the case of factor mimicking portfolios or economic variables. Second, the factors are unobserved. The first case is straightforward under independence of the factors as the joint posterior of the volatilities and parameters, $\pi(\Delta, \mathbf{H}, \omega \mid \mathbf{F}, \mathbf{y})$ decomposes as a product

$$\pi(\Delta, \Omega, \mathbf{H}, \omega | \mathbf{F}, \mathbf{y}) = \pi(\Delta, \Omega | \mathbf{F}, \mathbf{y}) \prod_{i=1}^{k} \pi(h_i, \omega_i | \mathbf{F})$$

The individual terms $\pi(h_i, \omega_i | \mathbf{F})$ can be generated using the univariate algorithm of JPR (1994) and the term $\pi(\Delta, \Omega | \mathbf{F}, \mathbf{y})$ follows from standard conjugate multivariate analysis.

When F is unobserved we need to simulate from the joint posterior $\pi(\Delta, \Omega, \mathbf{H}, \omega, \mathbf{F} | \mathbf{y})$. This can be broken into into conditionals

$$\pi(\Delta, \Omega, \mathbf{H}, \omega | \mathbf{F}, \mathbf{y})$$

$$\pi(\mathbf{F} | \Delta, \Omega, \mathbf{H}, \omega, \mathbf{y})$$

The first distribution is conditional on $\mathbf{F}$. It reduces to the case described above where $\mathbf{F}$ is known. The second conditional can be decomposed as

$$\pi(\mathbf{F} | \Delta, \Omega, \mathbf{H}, \omega, \mathbf{y}) \propto p(\mathbf{F} \mid \mathbf{H})\, p(\mathbf{y} \mid \mathbf{F}, \Delta, \Omega) = p(\mathbf{F}|\mathbf{H}) \prod_{t=1}^{T} p(\mathbf{y}_t \mid F_t, \Delta, \Omega)$$

That is

$$\pi(\mathbf{F}|\Delta,\Omega,\mathbf{H},\omega,\mathbf{y}) \propto |\Omega|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Omega^{-1}(\mathbf{y}-\Delta\mathbf{F})'(\mathbf{y}-\Delta\mathbf{F})\right)\right) \exp\left(-\frac{1}{2}\mathrm{tr}\left(\mathbf{F}'\mathbf{H}\mathbf{F}\right)\right)$$

where $\mathbf{F} = \{\mathbf{F}_t\}$. This is a standard matrix normal multivariate distribution from which we can make direct draws.

4 Empirical Applications

4.1 Fat-tailed Errors

In this section, we report the results of our analysis of weekly and daily series using a heavy-tailed t_ν-distribution for the mean equation as outlined in section 3.3. ν is also modelled as random which provides the benefit of inferring about the severity of the departure from normality. We also provide a set of outlier diagnostics by computing the posterior distribution of the scale parameters λ_t for each observation.

Table 1 reports summary statistics for the posterior distribution of the model parameters for weekly returns on the equal-weighted and value-weighted CRSP stock indices, while table 2 provides information on the posterior of the daily S&P500 series for the 80s, and three exchange rate series, Canadian \$, UK £, and Deutsch Mark. The posteriors of the persistence parameters and volatility equation variances resemble those found in JPR. The most striking finding is that the posterior of ν – the scale mixture parameter – is centered at fairly low values (low teens for most series). Figure 1 shows the posterior distribution of ν for each of the six series. We employ a uniform prior for ν on the range [1,60]. With the sole exception of the Canadian \$ exchange rate, all series put substantial posterior mass on values of ν which represent a departure from normality.

A closer examination of the posterior distributions of ν shows that while there is considerable evidence of non-normality, it is difficult to infer precisely about the values of ν. The relatively spread-out distribution of ν, especially for the weekly series, might cause estimation difficulties for a maximization approach to estimation (see Gallant, Hsieh and Tauchen (1994) for a discussion of these problems). However, the MCMC estimator does not get stuck on a flat area of the likelihood surface. Instead, it simply navigates across the flat and provides an accurate representation of the posterior uncertainty.

While the data provide fairly strong evidence against normality, they do not support a very low degree of freedom t distribution such as those found necessary in the ARCH literature. This is because stochastic volatility models possess an extra level of mixing through the variance equation that is not present in the ARCH-style models. Our results show that this mixing does result in a somewhat less fat-tailed conditional distribution than the ARCH models, but in no ways eliminates the need to model the fat-tailness of this conditional distribution.

Some would argue that the real problem with these time series is the presence of extreme outliers which cannot be accommodated by t-distribution with 10 or 15 degrees of freedom. We address this question by conducting two methods for outlier analysis. The first method is an outlier sensitivity analysis. As ν varies, the relative weight accorded to outlying observations is changed. For example, for very large values of ν we are very close to the normal case and outliers receive very high weight in posterior inference about the persistence and σ_v parameters. On the other hand if ν is very low, outliers receive little or no weight. We fix ν at 5, 10, 15 and 30 and compute the posterior distribution of other model parameters conditional on ν. Figures 2 and 3 present the results of this outlier sensitivity analysis. For both the equal-weighted weekly return (figure 2) and daily S&P 500 index (figure 3), the weighting method has a strong effect on the posterior distributions of δ and σ_v. As outliers are downweighted (ν declines), the posterior distribution of delta shifts to higher values representing greater persistence. As expected, σ_v declines so that the marginal variance of $\log(h_t)$ remains roughly the same $\frac{\sigma_v^2}{1-\delta^2}$. This finding is consistent with a view that the bulk of the data are generated from a model with higher persistence than the outlying observations. This result has important implications for prediction as the basic model will produce predictions with an underestimated persistence.

A unique feature of our analysis is the ability to infer about the draws of the scale-mixing parameter λ_t, for each observation in the time series. This is the second method of outlier analysis which we document. If a particular observation is outlying, our model will attempt to accommodate it with a posterior of λ_t centered at a large value. This will reweight the outlier and bring it more in line with the rest of the observations. We can use the joint posterior distribution of $(\lambda_1, ..., \lambda_T)$ to identify outliers by looking at each marginal posteriors $p(\lambda_t|\mathbf{y})$. One attractive feature of this approach is that it avoids the masking effect that plagues standard outlier diagnostic procedures. The time-varying volatility feature of our model complicates the detection of outliers. Obviously, October 17, 1987 is easily detected by a mere time series plot of the series. Other observations which are outlying in terms of the pattern of returns around a given time period may be more difficult to detect.

The top panel of Figure 4 shows the daily S&P 500 index returns (filtered to remove calendar effects as in Gallant, Rossi, and Tauchen (1992)), from 1/2/80-12/30/87, T = 2023. In the bottom panel, we plot the posterior means of each of the λ_t's. This estimation is carried out conditional on a fixed ν equal to 5. This implies a very diffuse prior for each of the scale parameters, $\lambda_t \sim$Inverted Gamma(ν,1). We plot the 75^{th} and 95^{th} quantiles of this prior distribution as horizontal lines on the bottom panel. A number of observations are flagged as outlying according to this diagnostic. These are observations that even a t distribution with 5 degrees of freedoms has difficulty accommodating. For reference there are 12 outliers. The posterior means of λ_t, observation number and date are given in the table below.

$E(\lambda_t \mid D)$	Observation	Date
5.44	1972	87/10/19 *
3.5	1693	87/09/11 *
2.5	1257	84/12/18
2.47	1974	87/10/21
2.36	664	82/08/17 *
2.35	1279	85/01/21
2.28	1522	86/01/08 *
2.18	1977	87/10/26
2.17	78	80/04/22
2.15	1215	84/10/18
2.14	525	82/01/28
2.10	1646	86/07/07 *

It is interesting to compare this diagnostic with that of Nelson (1991). Nelson uses an expected frequency criterion, in essence computing the expected number of appearances of a given (absolute) value at least as large as the one observed in the sample. The basis for his computation is the actual return R_t standardized by the MLE point estimate of σ_t. The largest outliers he identifies are by order of decreasing size on the dates: 87/10/19 , 87/9/11 , 82/8/17 , 86/1/8, and 86/7/7. The *'s in the above table are the Nelson outliers which we also identify as outliers. However, we also locate other outliers at least as severe which are not spotted by the Nelson analysis.

We then provide a diagnostic of the ability of the SVOL model to capture the marginal distribution of the data for the S&P 500 daily series. We simulate from the model with parameters equal to the estimated posterior means. We then make a Quantile-Quantile plot of the simulated data against the actual data distribution. Alternatively, one could have used the predictive density of the data by simulating multiple series for each draw of the parameters. We have omitted the day or week of Oct 17, 1987 in order to make the Q-Q plot more informative. Granted that none of these models fits that week very well, a Q-Q plot including it prevents us from seeing how well the rest of the distribution is fitted. Note that the discrete mixture model of section 2.2 can easily be combined with the fat-tailed model if one wanted to see if the two models together can accomodate even an exceptional outlier like October 17. The top panel of figure 5 shows the Q-Q plot of the actual returns with data simulated for ν=30 (normal case), the middle for ν=posterior mean, and the bottom panel for ν=5. As might be expected, the middle panel shows the closest correspondence in the distributions. There appears to be a very close correspondence between the marginal distribution implied by the SVOL model and the marginal distribution of the data.

4.2 Correlated Fat-tailed Errors

We now implement a SVOL model that simultaneously allows for correlation and fat tails. In our error specification we have $\epsilon_t = \lambda_t^{1/2} z_t$ with (v_t, z_t) distributed as a correlated bivariate normal and λ_t as an independent inverted gamma distribution. Clearly there is going to be a posterior correlation between the correlation parameter ρ and the heavy-tail parameter ν. The large negative returns will have greatest influence in determining the posterior for ρ.

As in Section 3.5, we model the parameters (ρ, σ_v) with a normal, inverted gamma natural conjugate prior on (ψ, Ω). The advantage of this prior is that it facilitates a direct draw of both ψ and Ω given the other model parameters and the data. We now show that we can elicit a prior that is reasonable for the more interpretable (ρ, σ_v) parameterization.

Figure 6 shows the implied prior distribution of ρ and σ_v for a particular set of prior hyperparameters, chosen so that the implied priors are very diffuse. The left panel shows that the prior on ρ is very nearly uniform in the interval (-.95, .95). The right panel shows a very diffuse prior on σ_v which puts most mass in the (0, .5) range encountered in most data analysis. We should note that it is possible to place an improper diffuse prior on these parameters as well. We believe that proper but very diffuse priors are more reasonable in applied work.

In addition to introducing correlation into the basic model, we also use the scale mixture strategy outlined in section 3.3 to simultaneously introduce fat-tailed error densities. Table 3 provides summary statistics of the posterior calculated for the correlated case for three CRSP series - weekly equal and value weighted returns and daily value-weighted returns. The daily CRSP value-weighted is exactly the same series used in Nelson's seminal (1991) paper in which he introduces the EGARCH model with asymmetric conditional variance functions.

As in the uncorrelated case, table 3 shows a high level of persistence in the volatility equation (.97 to .99 for the CRSP daily index). The distribution of δ is not affected by the introduction of the correlation parameter. It is important to note that while the posterior distribution of delta is massed at a high value near 1, there is little evidence of unit roots in the volatility equation. The posteriors damp down near 1 so that no appreciable mass is put on the region over .99. This occurs in spite of a prior which is locally uniform around 1.0 (a diffuse normal prior truncated at 1.) The introduction of the correlation does not appreciably affect the level of persistence.

There are two striking features of the posteriors shown in table 3. First, the posterior distribution of ρ is massed around fairly low values. Figure 7 graphs the posteriors for ρ for each of the three datasets. For the EW and VW series, the posterior distribution is massed around -.3 with about .9 posterior probability of negative correlation. Even with 6000 plus observations in the Nelson/CRSP daily dataset, the posterior of ρ still has a large standard deviation of .16

and there is a 5% probability of $\rho > 0$. Second, table 3 shows that the posterior mean of ν, the scale mixture parameter is relatively high. For all three series, it hovers around 25 degrees of freedom. This is in contrast to the results reported in section 4.1 for the uncorrelated case, where the posteriors of ν put substantial mass on values from 10 to 15. This results suggest a rather subtle interaction between the correlation and scale mixture parameters in the model. It is possible that the fat-tail results with no correlation are simply an artefact of model misspecification. Nelson (1991) estimates an exponential power density with a coefficient of 1.5. This is far from the low degree of freedom t found in our analysis without correlation and it is more consistent with the higher degrees of freedom found when we introduce correlation into the model. However, it is difficult to make exact parallels between the fat-tailness of the errors in the EGARCH and the SVOL models. The magnitude of the conditional variance asymmetry found here deserves further comment. It is difficult to compare the asymmetry parameter of Nelson's EGARCH model with the SVOL correlation parameter. We need a measure of asymmetry that is independent of model parameters. We consider the differential effect of a one σ rise or fall in the mean equation error, i.e. ϵ_t, on log-volatility. Consider for the EGARCH model the quantity

$$\Delta_e = (\ln h_{t+1} | \ln h_t, \epsilon_t = -1) - (\ln h_{t+1} | \ln h_t, \epsilon_t = +1)$$

Given Nelson's (1991) notation, we have $\Delta = -2\theta$. For the SVOL model, the quantity

$$\Delta_s = (\ln h_t | \ln h_{t-1}, \epsilon_{t+1} = -1) - (\ln h_t | \ln h_{t-1}, \epsilon_t = +1)$$

is equal to $-2\rho \times \sigma_v$. We are interested in standard deviations $\sqrt{h_t}$ rather than log volatilities. We translate Δ into a percentage differential between the response to negative and positive shocks by considering the quantity $\exp \frac{\Delta}{2} - 1$. Using Nelson's parameter estimates $\theta = -0.12$, for the CRSP daily series we find that the EGARCH implies a 13 per cent larger increase in volatility for a negative than for a positive shock. Using the SVOL estimates from the same dataset and period, we find a 3 per cent asymmetry. With the QML method, Harvey and Shephard (1993) find a ρ value of -.66 which implies a 7% asymmetry. So, even with the largest SVOL correlation estimates, the asymmetries in the conditional variance function are smaller than those reported by Nelson.

Finally, we report some diagnostics on the convergence and information content of our chain. Kim and Shephard (1994) in a discussion of JPR (see also Carter and Kohn (1994) and Shephard (1994)) make a case for the use of an approximate "multi-move" sampler as opposed to the "single-move" sampler used in JPR; it should be noted that, as proposed, the Kim and Shephard algorithm can not be extended to the correlated case. They argue that the convergence rate of the single-move samplers may be slow, especially near the limiting case of a unit root coupled with $\sigma_v = 0$. If we were near this limiting case, we should see slow dissipation of initial conditions and high autocorrelation in the chain draws. For the CRSP daily return, we made 15000 draws of ρ. Figure 8 shows the acf of the last 10000 draws of ρ, and one boxplot for the distribution of each subsequence of 3000 draws. For daily CRSP data the posterior distribution of δ is high. It is primarily the correlation in the draws of the volatility sequence which will induce correlation in the ρ draws. The acf shown in figure 8 does show reasonably high dependence ($\text{ACF}_1 = .7$) but this does not suggest that we are at all close to the singularity which would occur if we approached a unit root as σ_v declines to zero. This limiting case is the constant variance model so often rejected in the extensive ARCH/stochastic volatility literature and therefore of limited practical relevance. The bottom plot of figure 8 shows that, after the first 3000 draws which exhibit higher variance, the distribution of the draw sequence stabilizes and remains unchanged.

5 Summary

A hierarchical approach to the modelling of stochastic volatility models is presented. Typical extensions of the simple model result in the introduction of additional latent variables. These new latent variables are then incorporated into a Markov Chain Monte Carlo algorithm through data augmentation. This allows for efficient parameter estimation and variance smoothing and prediction, for a large class of stochastic volatility models. This approach also provides a sensitivity analysis for parameter inference and an outlier diagnostic. This framework is implemented for major extensions of the basic stochastic volatility model such as non normality (skewness and fat-tails) of the conditional mean, correlation between conditional mean and variance shocks (leverage effect). This hierarchical framework is also applied to two multivariate stochastic volatility models, with constant (stochastic discount factor) and time varying (factor structure) cross-correlations. The empirical analysis performed for various financial and economic series shows that the distributional extensions modelled are validated by the data. These extensions modify the prediction of variance. Economic models which use variance forecasts as inputs may therefore benefit from the incorporation of these extensions.

REFERENCES

ANDERSEN, T., and SØRENSEN, B. (1994). "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study", Manuscript KGSM Northwestern University.

BOLLERSLEV, T., CHOU, R. and KRONER, K. (1992), "ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence", *Journal of Econometrics*, 52, 5-59.

CARTER, C.K., and KOHN, R. (1994), "On Gibbs Sampling for State Space Models" *Biometrika*, 61.

CARLIN, B.P., and POLSON, N.G. (1991), "Inference for Nonconjugate Bayesian Models Using the Gibbs Sampler," to appear *Canadian Journal of Statistics.*

DANIELSSON, J. (1994), "Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood" *Journal of Econometrics*, 54.

ENGLE, R.,, NG, V., and ROTHSCHILD, M. (1990), "Asset Pricing with a Factor Arch Covariance Structure: Empirical Estimates for Treasury Bills ", *Journal of Econometrics*, 45, 213-237

GALLANT, A.R,, HSIEH, D., and TAUCHEN, G. (1994), "Estimation of Stochastic Volatility Models with Diagnostics", *Working Paper, Duke university,*

GEWEKE, J. (1989). "Exact predictive densities for linear models with ARCH disturbances". *Journal of Econometrics*, 40, 63-86.

GEWEKE, J. (1994). "Bayesian Comparison of Econometric Models". Working Paper, Federal Reserve Bank of Minneapolis Research Department.

HAMILTON, J.D., and SUSMEL, R. (1994), "Autoregressive Conditional Heteroskedasticity and Changes in Regime", *Journal of Econometrics*, 64, 307-333.

HARRISON, P.J., and STEVENS, C.F. (1976), "Bayesian Forecasting" (with discussion), *J. R. Statist. Soc., Ser. B*, **38**, 205-247.

HARVEY, A.C., RUIZ, E., and SHEPHARD, N. (1994), "Multivariate Stochastic Variance Models" *Review of Economic Studies*, 61, 247-264.

HARVEY, A.C., and SHEPHARD, N. (1993), "The Econometrics of Stochastic Volatility" *London School of Economics, Discussion Paper*, 166.

HESTON, S., (1993), "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bonds and Currency Options" *Review of Financial Studies*, 6, 327-343.

HSIEH, D. (1991), "Chaos and Non-linear Dynamics: Application to Financial Markets", *Journal of Finance*, 46, 5, 1839-1877.

HULL, J., and WHITE, A. (1987), "The Pricing of Options on Assets with Stochastic Volatility" *Journal of Finance*, 42, 281-300.

JACQUIER, E., POLSON, N. and ROSSI, P. (1994), "Bayesian analysis of Stochastic Volatility Models", (with discussion). *J. Business and Economic Statistics.*, 12, 4.

JACQUIER, E., POLSON, N. and ROSSI, P. (1995), "Efficient Option Pricing under Stochastic Volatility", Working Paper, Cornell University.

KIM, S., and SHEPHARD, N. (1994), "Discussion of Bayesian Analysis of Stochastic Volatility by Jacquier, Polson, and Rossi, " *J. Business and Economics Statistics*, 12 4 .

Mahieu, R., and Schotman, P. (1994), "Stochastic Volatility and the Distribution of Exchange Rate News", Manuscript, LIFE, University of Limburg.

MELINO, A. and TURNBULL, S.M. (1990), "Pricing options with stochastic volatility", *Journal of Econometrics*, 45, 239-265.

METROPOLIS, N., ROSENBLUTH, A.W., ROSENBLUTH, M.N., TELLER, A.H., and TELLER, E. (1953), "Equations of State Calculations by Fast Computing Machines," *J. Chemical Physics*, **21**, 1087-1091.

MÜLLER, P. (1991), "A Dynamic Vector ARCH Model for Exchange Rate Data", Manuscript, Duke University, *Institute for Statistics and Decision Sciences.*

MÜLLER, P. and POLE, A. (1994). "Monte Carlo Posterior Integration in GARCH Models" Manuscript, Duke University, *Institute for Statistics and Decision Sciences.*

NELSON, D. (1991). "Conditional Heteroskedasticity in Asset Pricing: A New Approach". *Econometrica*, **59**, 347-370

QUINTANA, J.M. and WEST, M. (1987). "An analysis of international exchange rates using multivariate DLM's". *The Statistician*, 36, 275-281.

SHEPHARD, N. (1994), "Partial Non-Gaussian State Space", *Biometrika*, 81, 115-131.

TIERNEY, L. (1991). Markov Chains for Exploring Posterior Distributions. Technical Report 560 University of Minnesota, School of Statistics, forthcoming *Annals of Statistics.*

WEST, M. and HARRISON, J. (1990). *Bayesian Forecasting and Dynamic Models.* Springer Series in Statistics.

Table 1

Posterior Analysis for Selected Weekly Series

Parameter	EW	VW
α	-0.33	-0.31
	(0.12)	(0.11)
	[-0.63, -0.15]	[-0.56,0.14]
δ	0.960	0.961
	(0.015)	(0.013)
	[0.92,0.98]	[0.93,0.98]
σ_v	0.211	0.197
	(0.038)	(0.036)
	[0.15, 0.30]	[0.14,0.28]
$\frac{V_h}{/E_h^2}$	0.86	0.74
	(0.30)	(0.29)
	[0.46,1.56]	[0.4,1.35]
ν	15	25
	(7.5)	(13.7)
	[8,13,37]	[10,20,57]

[a]The first number is the posterior mean. The number between parentheses is the posterior standard deviation. The two numbers between brackets indicate the 95% posterior credibility interval. For ν we also give the posterior median. ρ is fixed at 0. EW ≡ Equal-weighted NYSE; VW ≡ Value-weighted NYSE; weekly returns , 7/62-12/91. T = 1539. Returns are prefiltered to remove AR(1) and monthly seasonals from the mean equation.

Table 2

Posterior Analysis for Selected Daily Financial Series

Parameter	SP500	U.K.£	DM	CD$
α	-0.0034	-0.21	-0.32	-0.53
	(0.003)	(0.066)	(0.08)	0.11
	[-0.01,0.0015]	[-0.35,-0.09]	[-0.50,-0.18]	[-0.75,-0.34]
δ	0.986	0.98	0.969	0.958
	(0.007)	(0.006)	(0.008)	0.009
	[0.969, 0.996]	[0.966,0.990]	[0.952,0.983]	[0.94,0.972]
σ_v	0.099	0.11	0.157	0.24
	(0.021)	(0.017)	(0.023)	0.023
	[0.07, 0.15]	[0.086,15.1]	[0.12,0.20]	[0.20,0.29]
$\frac{V_h}{E_h^2}$	0.60	0.42	0.52	1.07
	(2.5)	(0.16)	(0.14)	0.24
	[0.24, 1.49]	[0.24,0.78]	[0.30,0.86]	[0.72,1.63]
ν	11	10	11	47
	(2.6)	(2)	(3)	9
	[7,10,17]	[7,10,15]	[8,10,19]	[28,49,60]

[a]The first number is the posterior mean. The number between parentheses is the posterior standard deviation. The two numbers between brackets indicate the 95% posterior credibility interval. For ν we also give the posterior median. ρ is fixed at 0. The S&P500-daily change in log of the index, filtered to remove calendar effects as in Gallant, Rossi, and Tauchen (1992);1/2/80-12/30/87; T=2023. UK £ and DM/$ daily noon spot rates (log change) from the board of Governors of the Federal Reserve System, supplied by David Hsieh; 1/2/80-5/31/90;T=2614. CD$ daily noon interbank market spot rates from Bank of Canada supplied by Melino and Turnbull (1990); 1/2/75-12/10/86; T=3001

Table 3

Posterior Analysis for Selected Financial Series - ρ estimated

Parameter	EW	VW	CRSP Daily
α	-.217	-.215	-.071
	(.064)	(.065)	(.019)
δ	.971	.973	.99
	(.0082)	(.0084)	(.0019)
	[.954,.987]	[.954,.987]	[.989,.996]
σ_v	.17	.16	.10
	(.020)	(.020)	(.010)
$\frac{V_h}{E_h^2}$	.86	.76	1.2
	(1.63)	(1.34)	(1.3)
Half-life	27.2	27.7	103.0
	(10.3)	(10.8)	(33.0)
ν	24.5	25.5	28
	(4.4)	(3.8)	(1.9)
	[15,30]	[16,30]	[23,30]
ρ	-.33	-.31	-.30
	(.20)	(.22)	(.16)
	[-.66,.13]	[-.66,17]	[-.58,.06]
Pr(ρ <0)	.93	.90	.95

[a]The first number is the posterior mean. The number between parentheses is the posterior standard deviation. When applicable, the two numbers between brackets indicate the 95% posterior credibility interval. For ρ, we also indicate the probability of ρ being negative. EW ≡ Equal-weighted NYSE; VW ≡ Value-weighted NYSE; weekly returns , 7/62-12/91. T = 1539. EW and VW Returns are prefiltered to remove AR(1) and monthly seasonals from the mean equation. The CRSP Daily returns are from 07/03/62 to 87/12/31; T=6409.

Figure 1. Histograms of the Posterior Distribution of NU
6 Weekly and Daily Financial Series
NU: Uniform Prior on [1,60]

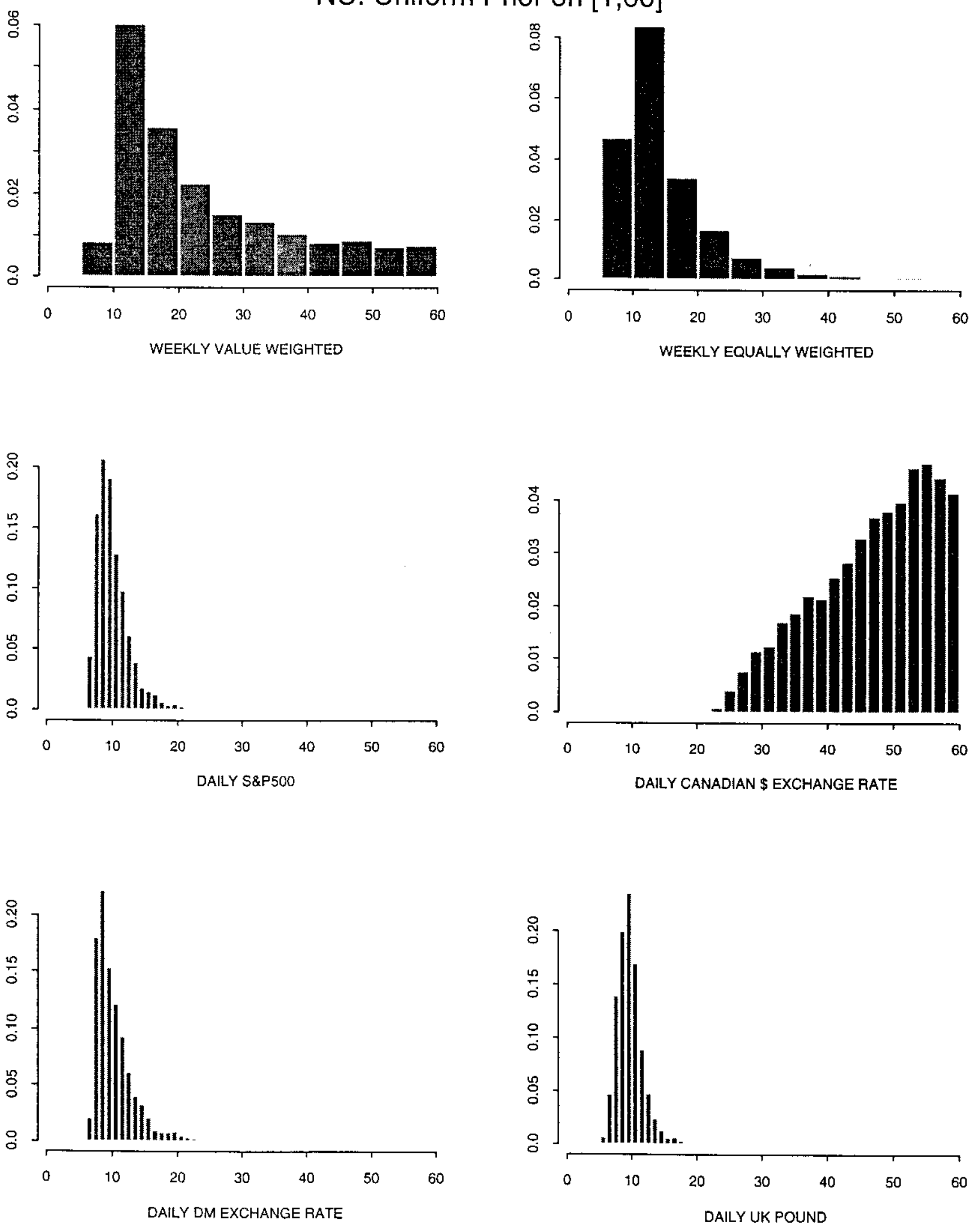


Figure 2. EFFECT of NU on DELTA and SIGMA_V
Weekly Equally Weighted Index Return

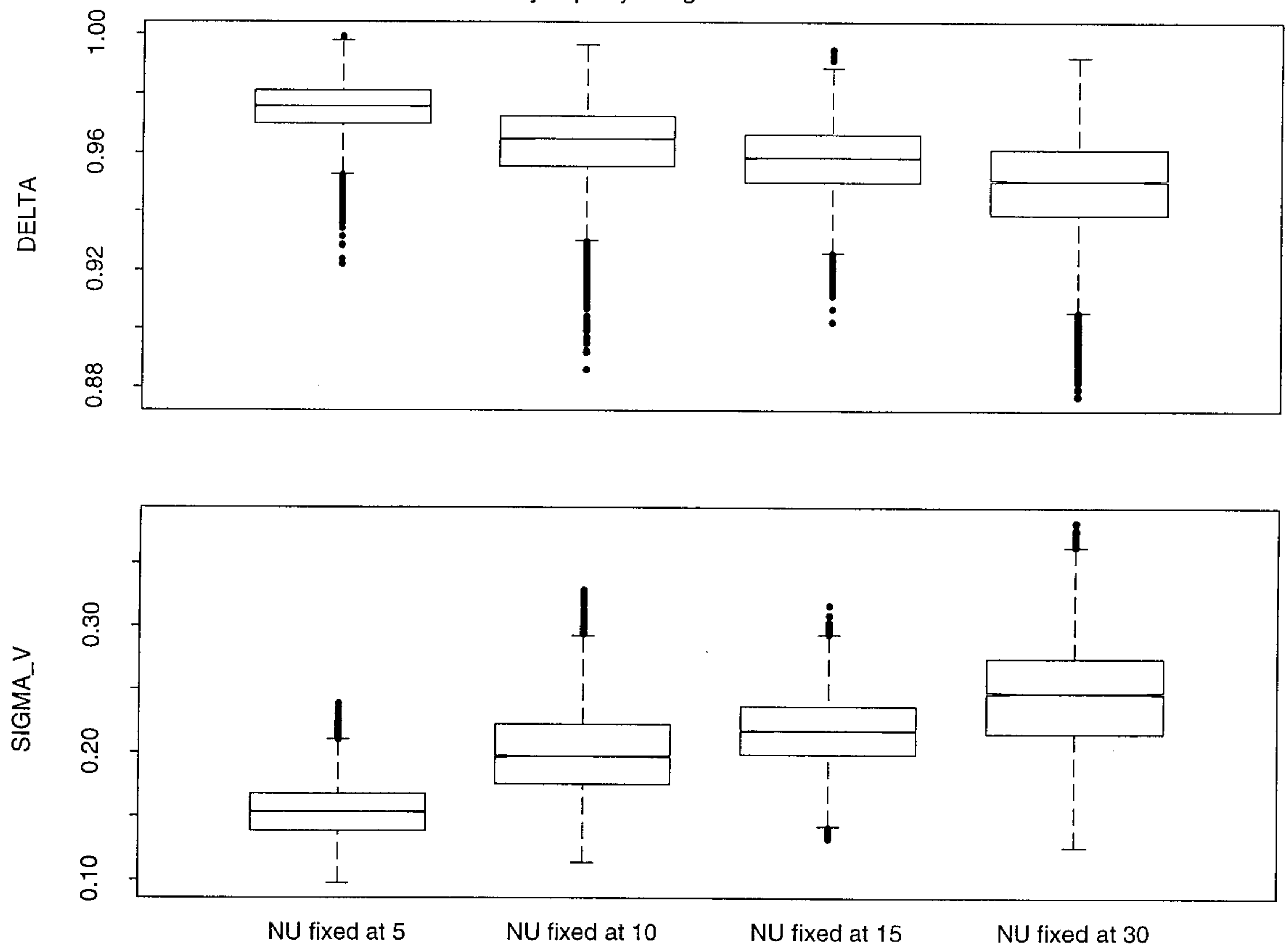


Figure 3. EFFECT of NU on DELTA and SIGMA_V
Daily S&P500 Index Return

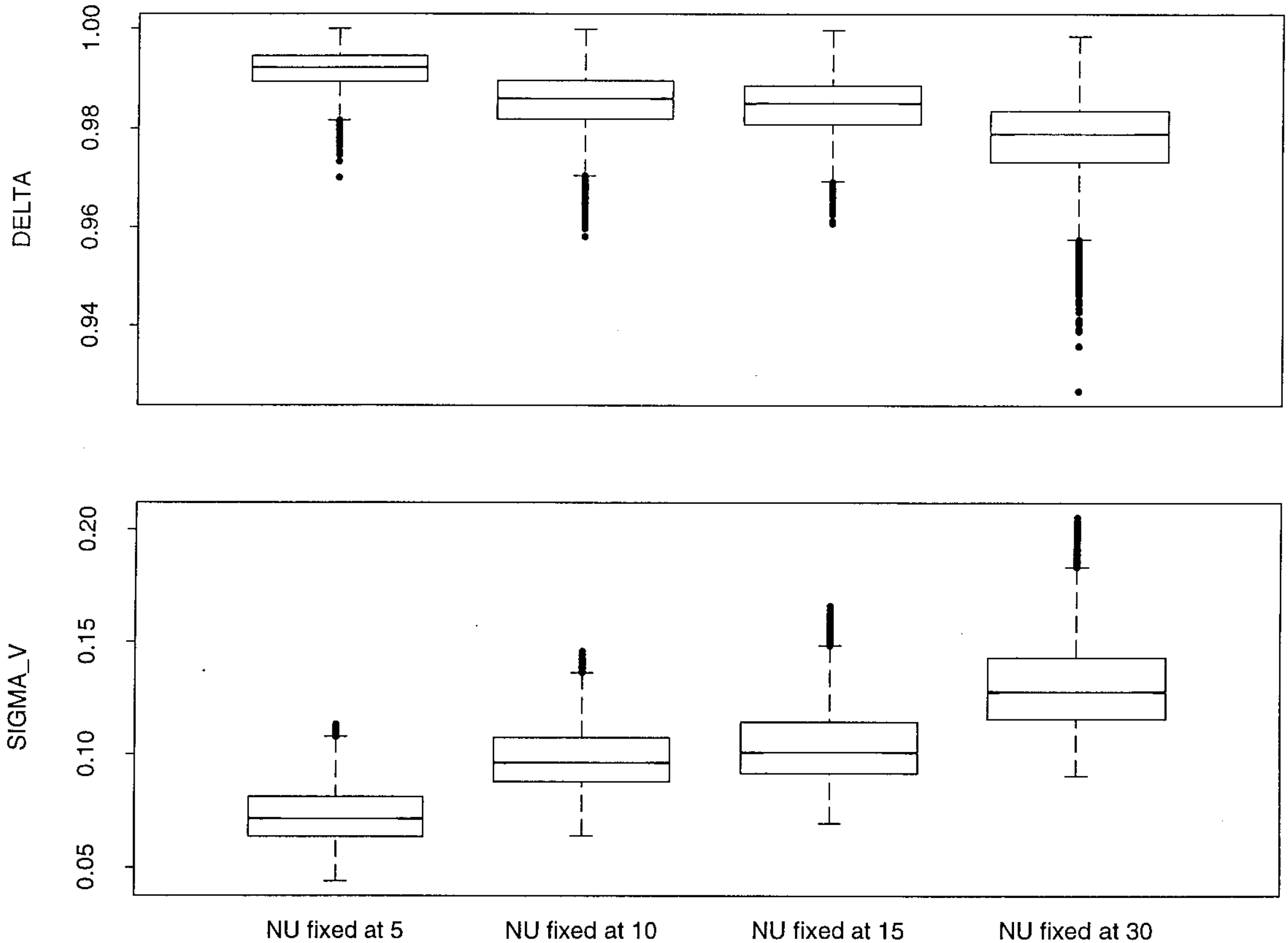


Figure 4. S&P500: Daily Returns and
Posterior Means of lambda_t, nu = 5

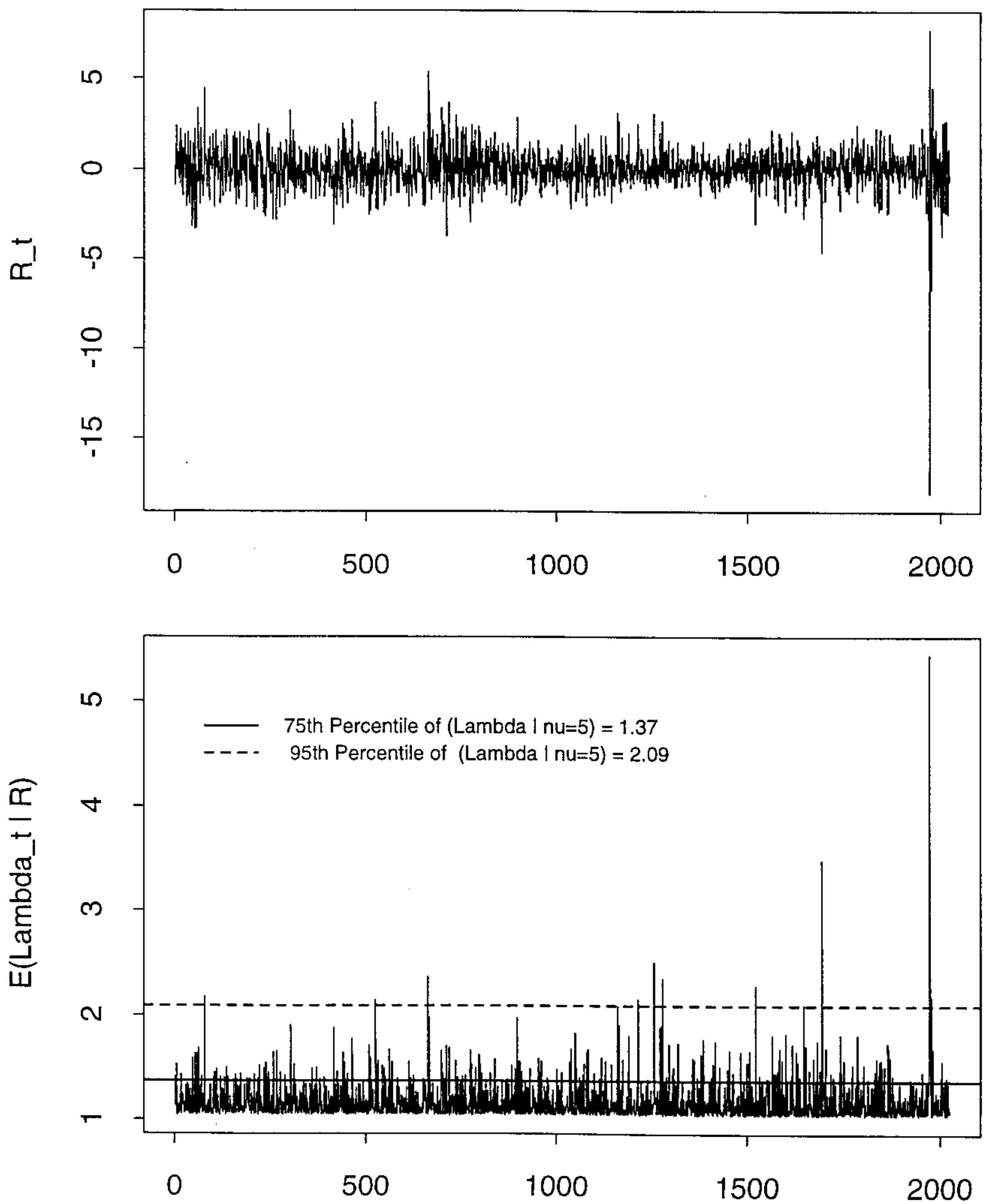


Figure 5. Quantile-Quantile Plots: SP&500 Daily Returns vs Data Simulated from SVOL Model

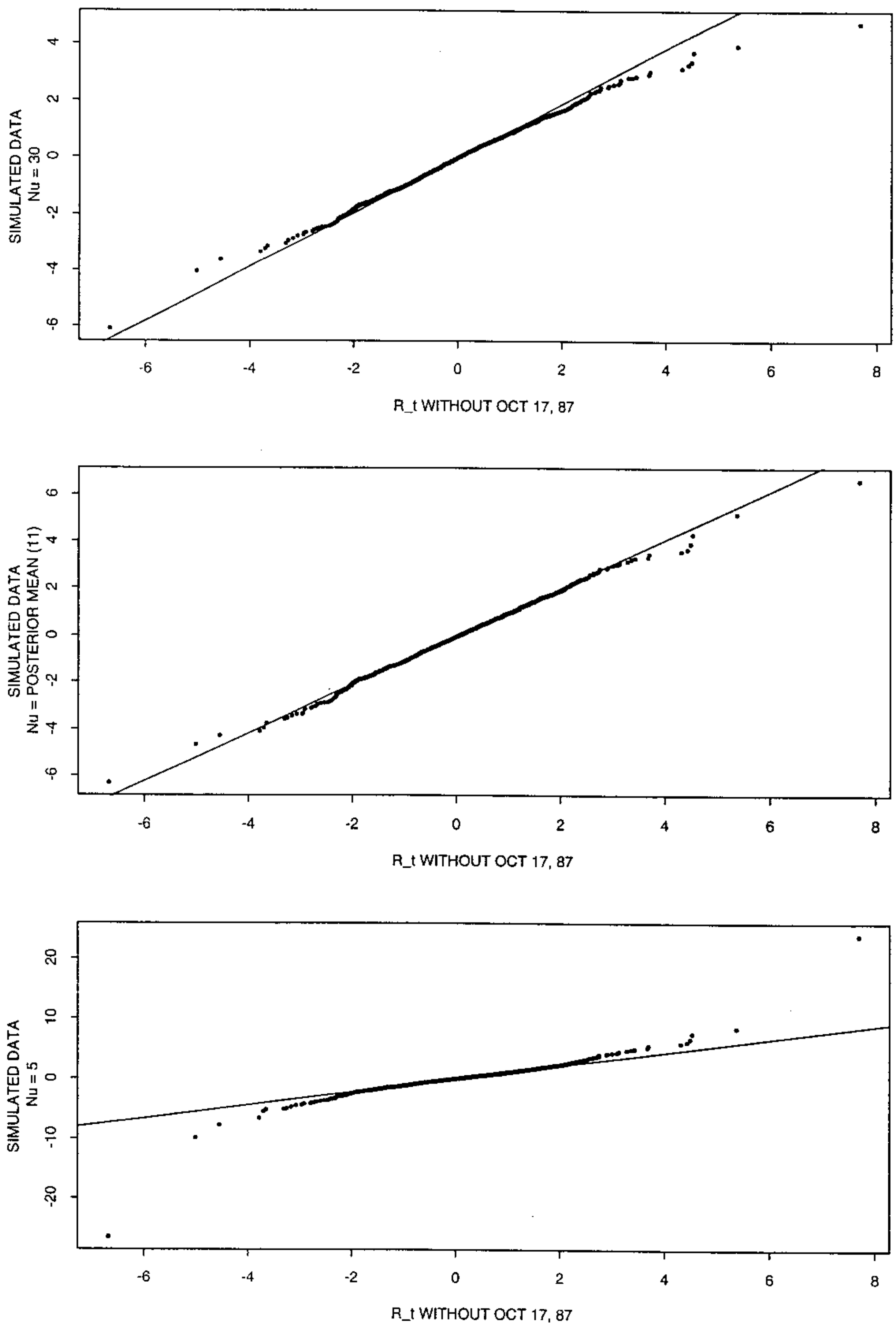


Figure 6. Implied Prior Distribution for RHO and SIGMA_V

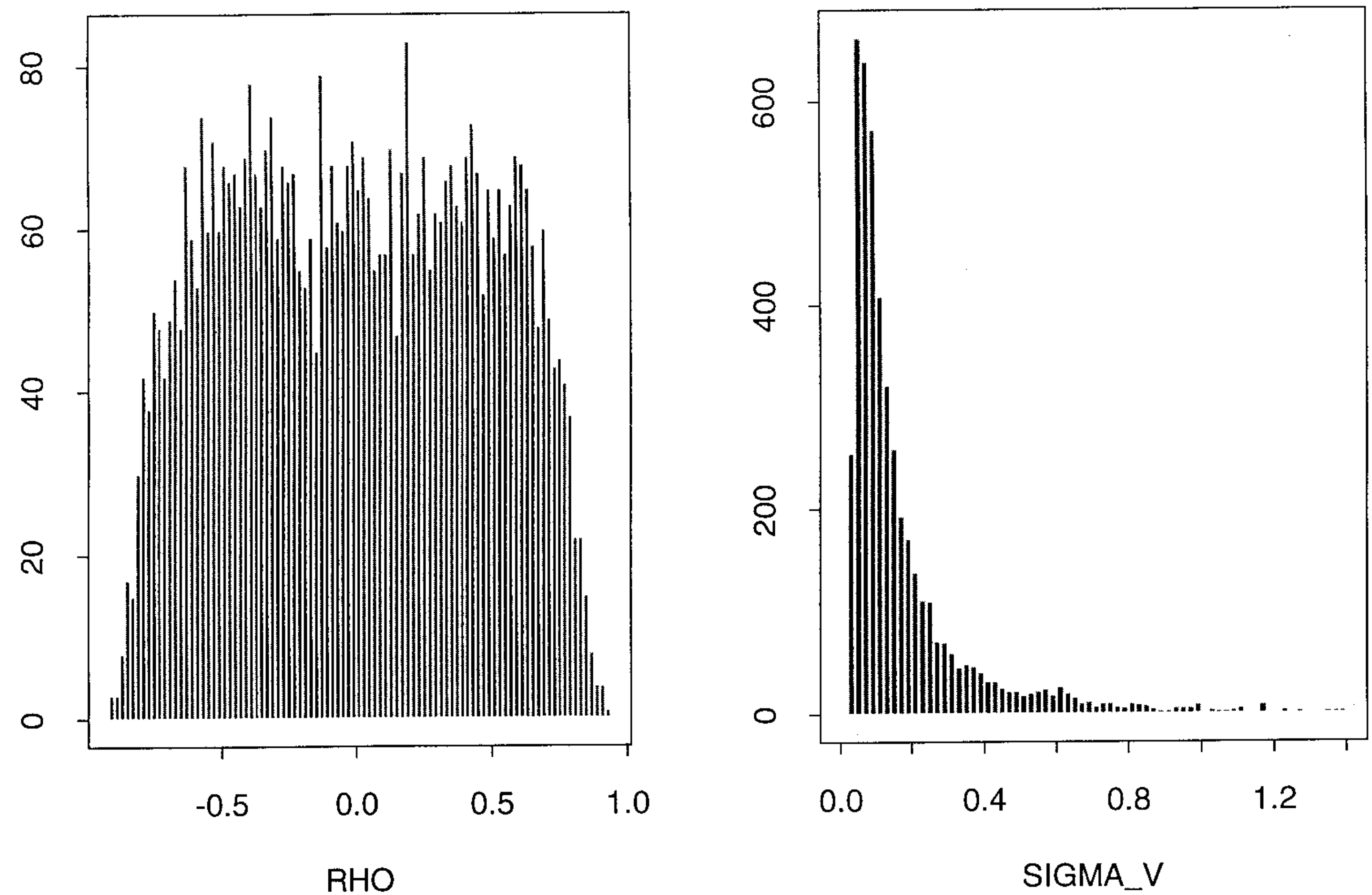


Figure 7. Posterior Distribution of RHO for 3 Series
(Vertical line: 90% quantile)

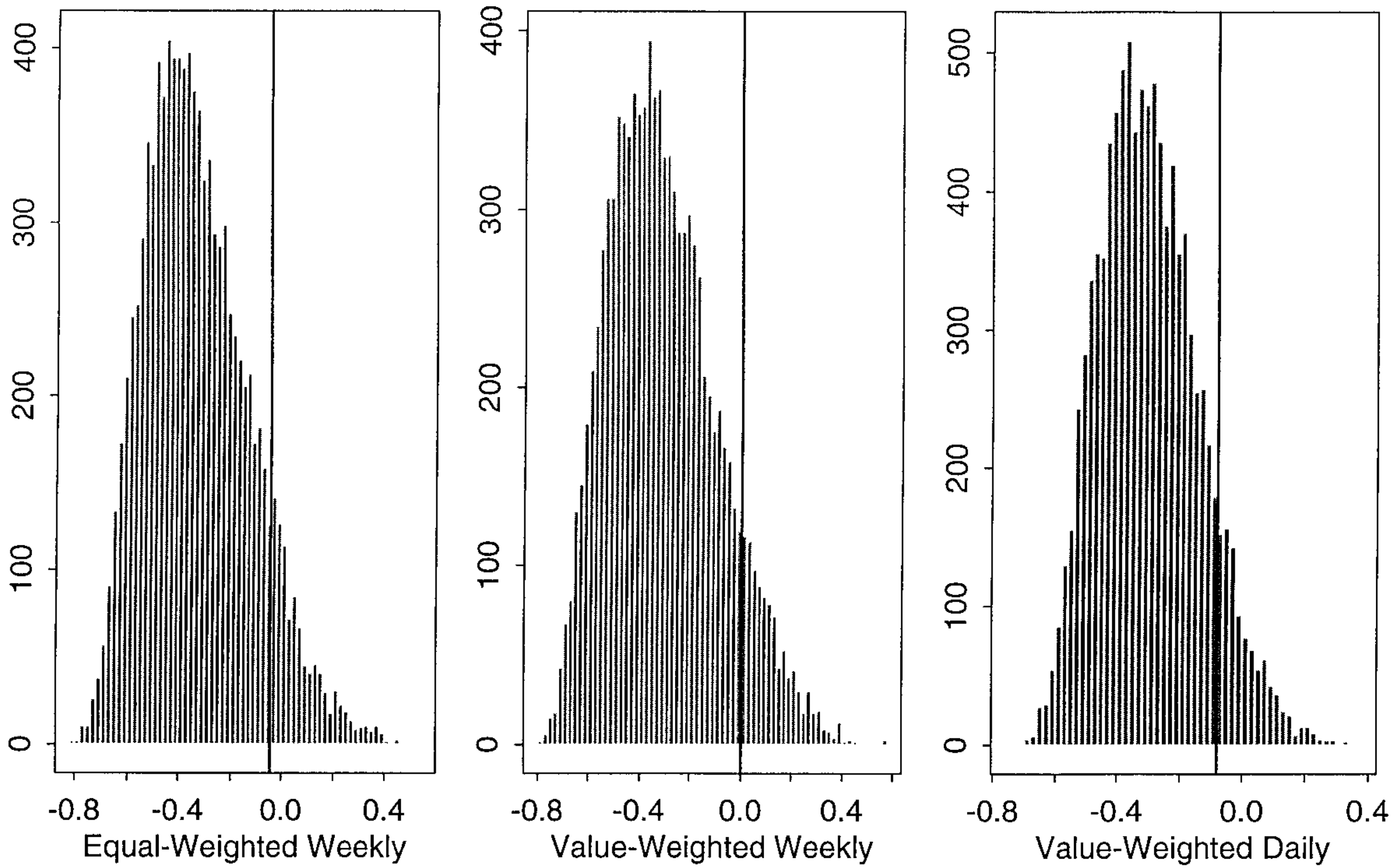
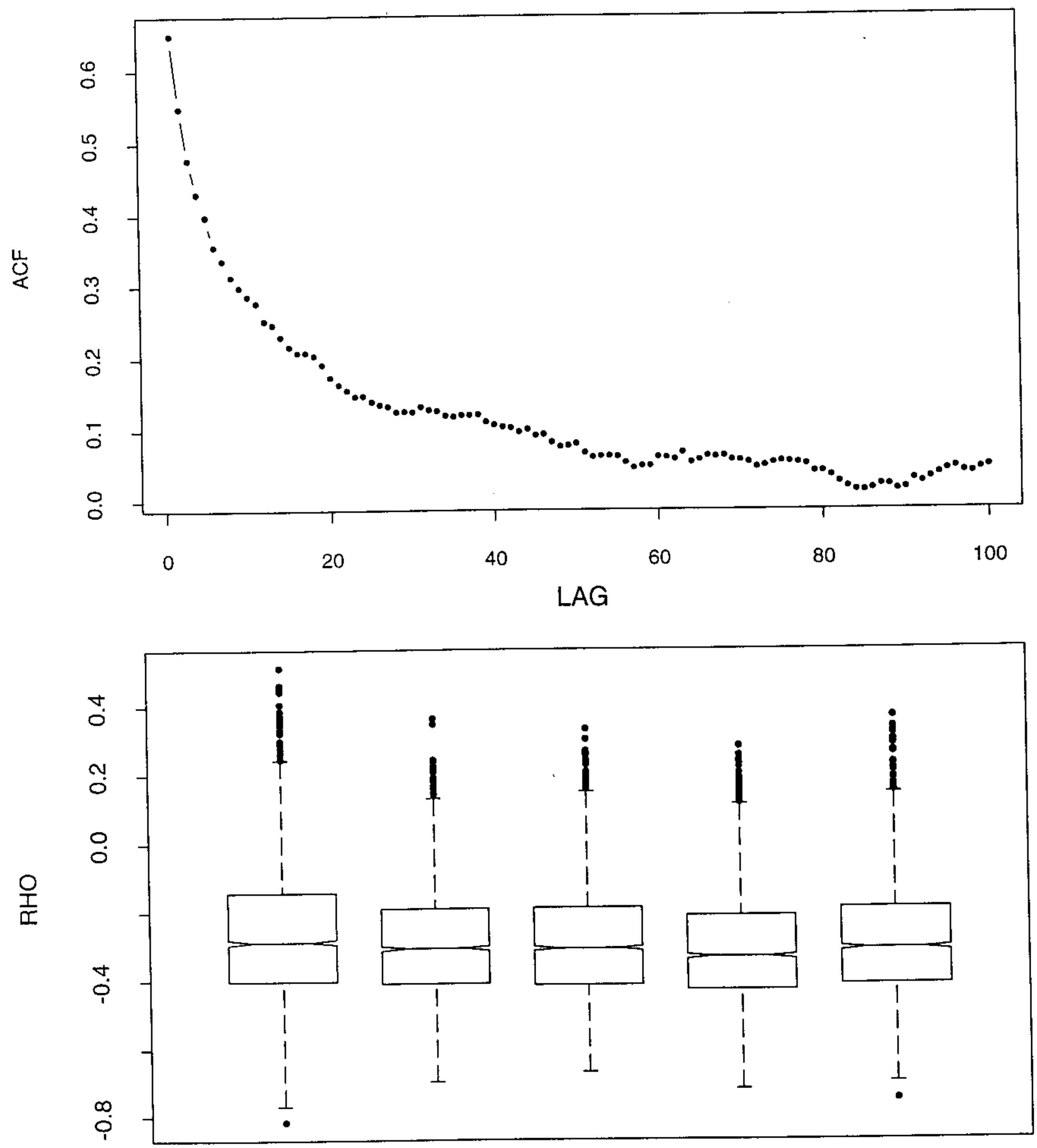


Figure 8. ACF and Distribution of RHO draws



1 Boxplot for each sub-sequence of 3000 draws