99s-26

Stochastic Volatility: Univariate and Multivariate Extensions

Éric Jacquier, Nicholas G. Polson,
Peter E. Rossi

Série Scientifique
Scientific Series

CIRANO
Centre interuniversitaire de recherche en analyse des organisations

Montréal
Juillet 1999

CIRANO

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du ministère de l'Industrie, du Commerce, de la Science et de la Technologie, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère de l'Industrie, du Commerce, de la Science et de la Technologie, and grants and research mandates obtained by its research teams.

Les organisations-partenaires / The Partner Organizations

- École des Hautes Études Commerciales
- École Polytechnique
- Université Concordia
- Université de Montréal
- Université du Québec à Montréal
- Université Laval
- Université McGill
- MEQ
- MICST
- Alcan Aluminium Ltée
- Banque Nationale du Canada
- Bell Québec
- Développement des ressources humaines Canada (DRHC)
- Egis
- Fédération des caisses populaires Desjardins de Montréal et de l'Ouest-du-Québec
- Hydro-Québec
- Imasco
- Industrie Canada
- Microcell Labs inc.
- Raymond Chabot Grant Thornton
- Téléglobe Canada
- Ville de Montréal

ISSN 1198-8177

Stochastic Volatility: Univariate and Multivariate Extensions*

Éric Jacquier†, Nicholas G. Polson‡, Peter E. Rossi§

Résumé / Abstract

Les modèles de volatilité stochastique, alias SVOL, sont plus durs à estimer que les modèles traditionnels de type ARCH. La littérature récente offre des estimateurs éprouvés pour un modèle SVOL univarié de base. Ce modèle est trop contraignant pour une utilisation en économie financière. Les prévisions de volatilité qu'il produit peuvent etre biaisées, particulièrement quand la volatilité est élevée. Nous généralisons le modèle de base en y ajoutant des effets de levier par le biais d'une corrélation entre les chocs observables et de variance, et la possibilité de distributions conditionnelles à queues épaisses. Nous développons un algorithme bayésien à chaînes markoviennes de Monte Carlo. Nous développons aussi un algorithme pour l'analyse d'un modèle SVOL multivarié à facteurs. Ces estimateurs permettent une inférence en échantillon fini pour les paramètres et les volatilités. Nous documentons les performances de l'estimateur et montrons que les extensions sont nécessaires. Nous testons la normalité des distributions conditionnelles. Cette méthode est mise en œuvre sur plusieurs séries financières. Il y a une forte évidence (1) de distributions conditionnelles à queues épaisses, et (2) d'effets de levier pour les actifs financiers. Les résultats sont robustes et ont d'importantes implications sur les décisions fondées sur les prédictions de volatilité, particulièrement pour la gestion de risques.

Stochastic volatility models, aka SVOL, are more difficult to estimate than standard time-varying volatility models (ARCH). Advances in the literature now offer well tested estimators for a basic univariate SVOL model. However, the basic model is too restrictive for many economic and finance applications. The use of the basic model can lead to biased volatility forecasts especially around crucial periods of high volatility. We extend the basic SVOL needs to allow for the

* Corresponding Author: Éric Jacquier, Finance Department, Fulton Hall 224B, Boston College, Chestnut Hill, MA 02167, USA Tel: (617) 552-2943 Fax: (617) 552-0431 email: jacquier@jacquier.bd.edu
Essential support from the Institute for Quantitative Research in Finance (Q GROUP) and CIRANO is gratefully acknowledged. We also thank the Physics Department at Boston College for their hospitality on the Theoretical Physics Computing Facility. We received helpful comments from Torben Andersen, Tim Bollerslev, John Geweke, Eric Ghysels, Dan Nelson, Éric Renault, Esther Ruiz, George Tauchen, Harald Uhlig, the participants of the Econometrics seminar at Harvard, and the 1994 CIRANO conference on stochastic volatility. We are especially grateful for detailed comments from two anonymous referees and the associate editor.

† Boston College and CIRANO

‡ University of Chicago

§ University of Chicago

leverage effect, through a correlation between observable and variance errors, and fat-tails in the conditional distribution. We develop a Bayesian Markov Chain Monte Carlo algorithm for this extended model. We also provide an algorithm to analyze a multivariate factor SVOL model. The method simultaneously performs finite sample inference and smoothing. We document the performance of the estimator and show why the extensions are warranted. We provide the researcher with a range of model diagnostics, such as the identification of outliers for stochastic volatility models or the assessment of the normality of the conditional distribution. We implement this methodology on a number of univariate financial time series. There is strong evidence of (1) non-normal conditional distributions for most series, and (2) a leverage effect for stock returns. We illustrate the robustness of the results to the choice of the prior distributions. These results have policy implications on decisions based upon prediction of volatility, especially when dealing with tail prediction as in risk management.

Mots Clés : Volatilité stochastique, ARCH, algorithme MCMC, effets de levier, gestion de risque, distributions à queues épaisses

Keywords: Stochastic volatility, ARCH, MCMC algorithm, leverage effect, risk management, fat-tailed distributions

JEL: C1, C11, C15, G1

1 Introduction

Time varying volatility is a characteristic of many financial series. The discrete time SVOL model is a natural alternative to standard time-varying volatility models of the ARCH family. It allows the conditional mean and variance to be driven by separate stochastic processes. There is evidence implying that the additional flexibility of the SVOL is warranted, e.g., Geweke (1994b), Fridman and Harris (1998). This is especially true when the current level of volatility is high and the ARCH model may not provide enough mixing. The SVOL model provides more mixing than the ARCH because the volatility and the observable shocks are imperfectly correlated with one another. However, the basic SVOL featuring normal errors and zero correlation between mean and volatility errors is still too restrictive for economic and financial series.

The basic SVOL model requires extendions for several reasons. The results in Gallant, Hsieh, and Tauchen (1994) show that the conditional distribution of financial series in a SVOL model may be skewed and fat-tailed. First, Geweke (1994c) shows that the basic SVOL model may inadequately model the large outliers. It is when volatility is high that the cost of a misspecified model may be the highest. This is the cause for the frequent critiques of the standard value-at-risk methods. This is remedied by allowing the conditional distribution to have fat-tails. Second, the basic SVOL model does not address the *leverage* effect, whereby changes in volatility are asymmetrically related to the sign and magnitude of price changes, e.g., Black (1976), Nelson (1991), and Glosten, Jagannathan and Runkle (1993). A negative correlation between the errors of the variance and the observable, e.g., stock returns, can produce this effect. This approach is favored in stochastic volatility based option pricing models.

We show that these two extensions have empirical and economic importance because they significantly modify volatility predictions. This in turn affects economic modeling based upon these predictions such as option pricing, e.g., Hull and White (1987), risk management and asset allocation. There is also a clear need for a multivariate extension of the basic model, for example in portfolio management and asset pricing applications.

Earlier estimation and smoothing strategies were based upon a normality approximation in order to use the Kalman filter, e.g., Melino and Turnbull (1990), Harvey, Ruiz, and Shephard (1993). This approximation is not necessary. Jacquier, Polson, and Rossi (1994), hereafter JPR, introduced Markov Chain Monte Carlo techniques (MCMC), for the estimation of basic SVOL models with normally distributed conditional errors. This paper extends this MCMC framework and shows that the method performs well and the extensions are warranted.

We develop models to allow for (1) the leverage effect through a correlation between observable and variance errors and (2) fat-tails in the conditional error distribution. The fat-tails are modeled via scale mixtures of normals. This allows for an outlier diagnostic for each observation. First, we show that the different models produce different volatility forecasts. Second, we document the sampling performance of the estimator and smoother for the extended models. In particular, we show that the performance amply justifies an a priori decision to use an extended rather than a basic model. Finally, we estimate the extended models for a number of financial series and conclude that the extensions are supported by the data. We show how to use the model as an outlier diagnostic or conduct posterior analysis for the relevant parameters. The posterior analysis is the basis for a diagnostic of non-normality of the conditional distribution. We assess the robustness of the results to the prior specification. An algorithm to analyze a multivariate factor SVOL model

is also provided.

Our methodology allows the researcher to perform a comprehensive sensitivity analysis. For example, the persistence of the AR(1) volatility process increases when fat-tails are allowed in the first stage distribution. The extensions can yield very different predictive densities. The proper modeling of the first two stages affects the quality of prediction, and is of paramount concern for many uses of the model. Our hierarchical modeling approach will therefore prove extremely useful for researchers concerned with the influence of distributional assumptions on resulting inferences. We assess the robustness of the results to the third stage parameter distribution.

The paper is organized as follows. Section 2 describes the hierarchical modeling framework and develops specific MCMC algorithms for implementation of the model extensions. It also shows that the extended models depart from the basic SVOL in ways which affect the user, e.g., inference and marginal density. Section 3 discusses inference, outlier and conditional normality diagnostics. It documents the performance of the estimators. Section 4 applies the extended models to equity indices and exchange rates. We conduct posterior analysis, and apply the outlier and conditional normality diagnostics. The sensitivity of the results to the choice of priors is discussed. Section 5 concludes.

2 Extensions of SVOL Models

We view SVOL modeling as a three stage hierarchical process. The three conditional distributions are $p(\mathbf{y}|\mathbf{h})$, $p(\mathbf{h}|\omega)$ and $p(\omega)$ where $\mathbf{y}$, $\mathbf{h}$, and ω are the data, the volatilities, and the parameters. The first stage distribution $p(\mathbf{y}|\mathbf{h})$, reflects probability beliefs about the data given the volatilities. The basic SVOL assumes conditional normality. In this paper we allow for fat-tailed and correlated errors. The second stage $p(\mathbf{h}|\omega)$, models beliefs about the stochastic evolution of the volatility sequence, e.g., a log-AR(1) process. The third stage $p(\omega)$ reflects beliefs about the parameters of the volatility process. This three stage hierarchical model allows the implementation of MCMC procedures that allow the researcher to make inferences concerning the joint posterior $p(\mathbf{h}, \omega|\mathbf{y})$. Marginal distributions such as $p(\mathbf{h}|\mathbf{y})$ and $p(\omega|\mathbf{y})$, are obtained by simply averaging the appropriate conditional distributions over the simulated draws.

2.1 Basic Model and MCMC

The basic SVOL model extended in this paper is

$$\begin{aligned} y_t &= \sqrt{h_t}\,\epsilon_t, \\ \log h_t &= \alpha + \delta \log h_{t-1} + \sigma_v v_t, \quad t = 1, \ldots, T \\ (\epsilon_t, v_t) &\sim N_2(0, I_2) \\ (\alpha, \delta, \sigma_v) &\sim p(\alpha, \delta, \sigma_v), \end{aligned} \tag{1}$$

The parameter vector $\omega = (\alpha, \delta, \sigma_v)$ consists of a location α, a volatility persistence δ and a volatility of volatility σ_v. Although not presented as a hierarchical model in JPR (1994), it is a particular example of the general framework discussed here. The first stage assumes normality. The second stage specifies a stationary volatility sequence where $p(\log \mathbf{h}|\omega)$ follows an AR(1) process. The third

stage assumes a diffuse distribution for $p(\omega)$ restricted to the region of stationarity of the volatility process.

MCMC methods, such as the *Gibbs* and *Metropolis* algorithms, construct Markov chains with equilibrium distribution equal to the joint posterior distribution of the volatilities and parameters given the data. The parameter space Θ is partitioned into r blocks. The Gibbs sampler consists in the iterative sampling of $\Theta_i^{(n+1)}$ from the conditional $p(\Theta_i^{(n+1)}|\Theta_1^{(n+1)},\ldots,\Theta_{i-1}^{(n+1)},\Theta_{i+1}^{(n)},\ldots,\Theta_r^{(n)})$ or $p(\Theta_i^{(n+1)}|\Theta_{-i}^{(n)})$ for $i=1,\ldots,r$. For some blocks, the conditional may not be easily sampled directly. One can then use a Metropolis step whereby $\Theta_i^{(n+1)}$ is sampled from a Metropolis kernel, i.e., blanketing density, $Q(\Theta_i^{(n+1)},\Theta_i^{(n)})$. The new draw from Q is accepted with probability

$$\min\left(\frac{\pi(\Theta^{(n+1)})Q(\Theta^{(n+1)},\Theta^{(n)})}{\pi(\Theta^{(n)})Q(\Theta^{(n)},\Theta^{(n+1)})},1\right). \tag{2}$$

Otherwise, the previous draw is repeated, and one moves to the next block. Ones starts from an initial state, possibly picked at random. The draws of the simulated distribution converge to draws of the stationary distribution namely the required joint posterior.

The MCMC algorithm for computing the joint posterior distribution of $p(\mathbf{h},\omega)$ developed in JPR (1994) uses an independence Hastings-Metropolis algorithm for the underlying Markov chain. It can be defined with only *local* movements in the parameter space and is particularly useful for SVOL models. See Hastings (1970) and Tierney (1994), and Chib and Greenberg (1995) for an exposition. For the model in (1), JPR (1994) show that the Gibbs block can be written as

$$p(\alpha,\delta,\sigma_v \mid \mathbf{y},\mathbf{h}) \propto \ell(\alpha,\delta,\sigma_v \mid \mathbf{y},\mathbf{h}) \times p(\alpha,\delta,\sigma_v),$$

and T conditionals $p(h_t \mid h_{-t},y_t,\alpha,\delta,\sigma_v)$ for $t=1,\ldots T$.

$$p(h_t|h_{t-1},h_{t+1},\alpha,\delta,\sigma_v,\mathbf{y}) \propto \frac{1}{h_t^{0.5}}\exp\frac{-y_t^2}{2h_t} \times \frac{1}{h_t}\exp\frac{-(\log h_t-\mu_t)^2}{2\sigma^2}, \tag{3}$$

where $\mu_t=(\alpha(1-\delta)+\delta(\log h_{t+1}+\log h_{t-1}))/(1+\delta^2)$ and $\sigma^2=\sigma_v^2/(1+\delta^2)$. Given the vector of $\mathbf{h}$'s, the first conditional distribution is similar to a linear regression. Direct draws from p in (3) are not feasible, so an accept/reject Metropolis step is used. The Metropolis kernel, or blanketing density, is obtained by best approximating p. For this, the log-normal kernel in (3) is approximated by an inverse gamma with same mean and variance. The inverse gamma kernel can then be multiplied by the first kernel, itself an inverse gamma. This yields an inverse gamma distribution from which direct draws are made. The blanketing density is $q(h_t \mid .) \propto h_t^{-(\phi+1)}e^{-\lambda/h_t}$, where $\phi=0.5+(1-2\exp\sigma^2)/(1-\exp\sigma^2)$, and $\lambda=0.5y_t^2+(\phi-1)\exp(\mu_t+0.5\sigma^2)$. In JPR, the $(n+1)^{st}$ draw of h_t first goes through an accept/reject step. That is, the draw is accepted with probability $p(h)/cq(h)$, where p is unnormalized. The constant c is chosen as the median value of p/q for three values of h around the mode of q. If the draw is rejected, another draw is made. Otherwise the draw enters the Metropolis step. There, $h_t^{(n+1)}$ is accepted with the probability in (2). Otherwise, the previous draw $h_t^{(n)}$ is repeated and one moves to the block h_{t+1}. This choice of c balances the rejections (below 20%) and repeats (below 5%).

Note that (2) is implemented with q as Q, and p from equation (3) as π. So the kernel $Q(\Theta_i^{(n+1)},\Theta_i^{(n)})$ is in fact $Q(\Theta_i^{(n+1)})$, an *independence* Metropolis kernel. Also, a pure Metropolis,

without the extra accept/reject step described above, would have the same theoretical convergence properties. We found that eliminating the computation of c and the extra accept/reject step cuts the CPU time in half. However, we noticed that, for the fat-tail model and when starting from extreme parameter values, the initial conditions sometimes dissipate faster with the accept/reject step. The CPU time needed to run the algorithm is quite short. So we keep the accept/reject step since it does not impose any measurable burden. In this paper we compute c as the value of p/q at the mode of q.

Since JPR (1994)'s first implemented this MCMC method, alternatives have been suggested to address issues of computational efficiency with respect to the posterior, $p(\mathbf{h}|\omega, \mathbf{y})$. For example, Geweke (1994c) recognizes that the density in (3) is log-concave. He proposes a pure accept/reject step based on the algorithm in Wild and Gilks (1993). The two approaches have similar theoretical convergence properties. Geweke's approach extends to the fat-tail that does not modify $p(h \mid .)$, as we will see below. It may also extend to the correlated case if one uses the algorithm of Gilks et al. (1995) instead. This would be necessary because the conditional posterior of h_t is not log-concave in the correlated case, see equation (11) below. Also, a pure accept/reject step requires the normalization constant of p in (3), a possibly significant computational burden. It is unclear then, which algorithm would be faster.

Another estimation approach for the basic SVOL is suggested by Carter and Kohn (1994), Mahieu and Schotman (1994), and Kim and Shephard (1994). It allows a joint draw of the vector of volatilities by approximating the distribution of $\log \epsilon_t^2$ with a discrete mixture of normals. Kim and Shephard note that the sampler based on (3) fails to converge at the limiting case of a unit root coupled with $\sigma_v = 0$. First, this argument, though correct, is not very relevant since the limiting case invoked is that of constant variance. Second, JPR (1994) document the convergence of their algorithm by several means, most noticeably by simulating the sampling behavior of location estimates of relevant posterior distributions. Here we see no reason to use an approximation of the model when we can estimate the exact one. Some alternative strategies may offer a computing time advantage. But we have not found computing time to limit in any way our analysis. More importantly, the mixture of normal approximation does not extend readily to the correlated case developed below. Finally, recall that we model the extensions so as to nest the basic SVOL model via latent variables. If desired, feasible alternatives for $p(\mathbf{h}|.)$, can be incorporated into our approach. The goal of this paper is to document the need for, and provide models and a reliable algorithm for the extensions of the basic SVOL. We document the convergence and performance of our MCMC strategy.

The major assumption relaxed here is that of an uncorrelated bivariate normal distribution for ϵ_t and v_t. The natural extensions of the model are skewed and fat-tailed error distributions, and correlated errors. We introduce a correlation ρ between ϵ_t and v_t, a scale variable λ_t where $y_t = \sqrt{\lambda_t h_t}\epsilon_t$, and a fat-tailness parameter ν. Straightforward extensions not considered here include regressors in the mean/volatility equations. We also model multivariate time series with stochastic volatility by a hierarchical factor model. The state space Θ is extended to include the needed additional parameters and state variables. For example, in the full correlated and fat-tailed model, $\Theta = \omega \times \lambda \times \mathbf{h}$ where $\omega = (\alpha, \delta, \sigma_v, \rho, \nu)$. The algorithms described below sample from the joint posterior $\pi(\mathbf{h}, \omega, \lambda|\mathbf{y})$. This modular approach is convenient because a feature of the model can be *turned-off* or modified without affecting the rest of the algorithm. For clarity of exposition, we describe separately the fat-tailness and the correlation extensions. We then show that the

combination of these two extensions is straightforward.

2.2 Fat-tailed Departures from Normality

2.2.1 The Model

For most financial time series, the conditional distribution resulting from ARCH models is strongly fat-tailed, see Bollerslev, Chou, and Kroner (1994). For SVOL models one suspects that, albeit possibly to a lesser extent, this could also be the case, e.g., Gallant, Hsieh, and Tauchen (1994). Geweke (1994c) shows that the basic model may be inadequate with respect to outliers. To fatten the tail of $p(\epsilon_t)$ whilst keeping it symmetric, we model it as a scale mixture of normals. Consider the model

$$\begin{aligned} y_t &= \sqrt{h_t}\epsilon_t = \sqrt{h_t}\sqrt{\lambda_t}z_t \\ \log h_t &= \alpha + \delta \log h_{t-1} + \sigma_v v_t, \quad t = 1, \ldots, T \\ (z_t, v_t) &\sim N_2(0, I) \\ \lambda_t &\sim p(\lambda_t \mid \nu) \\ p(\alpha, \delta, \sigma_v, \nu) &= p(\alpha, \delta, \sigma_v)\, p(\nu). \end{aligned} \tag{4}$$

Within the hierarchical framework, fat tails in the error v_t would be modeled in an identical fashion to that shown below. The distributions $p(\lambda_t \mid \nu)$ and $p(\nu)$ are chosen to allow flexibility in the modeling of the fat-tailness of the distribution $p(\epsilon_t)$. $p(\lambda_t \mid \nu)$ can be specified so as to allow for a wide range of fat-tailed behavior for $p(\epsilon_t)$, such as a double exponential, exponential power, stable, logistic, or Student-t distributions. Carlin and Polson (1991), with a fixed ν, and Geweke (1993), who estimates ν apply this idea in the context of linear regressions.

Given a prior $p(\nu)$, the posterior distribution $p(\nu|\mathbf{y})$ is an output of the algorithm. Through this posterior we can infer from the data the severity of the departure from normality. In this paper, we focus on a t_ν fat-tailed model. That is, $p(\lambda_t \mid \nu)$ is an inverse gamma distribution, i.e., $\nu/\lambda_t \sim \chi^2_\nu$. The model has noteworthy features even if ν is fixed. $p(\lambda_t \mid \nu)$ is an inverse gamma distribution, even if ν is not updated. This still nests the basic model and allows for an outlier robustification. The model transfers some of the outliers from the autocorrelated h_t to the non-autocorrelated λ_t. This down-weighs the outliers in the estimation of the stochastic volatility parameters and of h_t itself.

The importance of the tail assumption is illustrated in figure 1 with an example of the British Pound to US Dollar exchange rate. It shows that this effect has major consequences. The top left plot shows the time series of the daily percentage change in exchange rate. The bottom left plot shows the exchange rate, divided by its sample standard deviation, for the subsequence from January to November 1985. We now concentrate on this period. We use the estimated parameters of both basic and fat tail models for this example (see table 4 discussed later). The top right plot contrasts the volatility estimates of the two models. The thick and dashed lines show the in-sample posterior means of $\sqrt{h_t}$ for the basic and fat-tail models respectively. The basic SVOL produces a less persistent volatility series because it tries to *meet* the outliers. The thin line is the time series of posterior means of $s_t = \sqrt{h_t\lambda_t}$. Compare the thin and dotted lines. It reveals the robustness of the fat-tail model to the outliers. Their effect on h_t is reduced because they are incorporated in λ_t.

Now consider having observed the series until September 23^{rd}, i.e., the thick part of the bottom left plot. We now want to make out-of sample predictions of the time varying standard deviation for the upcoming days. For the basic model, the estimated AR(1) is used to produce forecasts of $\sqrt{h_t}$, starting from the last estimated h_t. For the fat-tail model, the (smoother) estimated AR(1) is used. It it also inflated by the (non-time varying) posterior mean of $\sqrt{\lambda}$, to produce forecasts of $\sqrt{\lambda h_t}$. Formally, each joint draw of $\mathbf{h}, \omega, \nu$ is used to generate a draw of the future vector of standard deviations, see Geweke (1994c) and Geweke (1989). The draws are averaged to produce predictive means. This is more efficient than the approach suggested in JPR (1994).

The bottom right plot in figure 1 shows the results. For the basic model, the forecast starts at 0.012 and reverts to the long term mean of $\sqrt{h_t}$. For the fat-tail model, the forecast starts from 0.0088. This large relative difference in forecast profiles between normal and fat-tail models stems from differences in the estimates of h_t as well as δ. At any time at which the two models make different inferences on h_t, their forecast profiles will diverge. Looking back at the top right plot, one sees that the two models often disagree markedly on h_t. So the difference between the basic and the fat-tail models has important practical implications.

In this framework, the identification of outliers can be modeled as follows. Posterior inference about the scale λ_t can be used as an outlier diagnostic. Observations with a marginal posterior centered on a large λ_t are possible outliers. A simple diagnostic is to plot the posterior means $E(\lambda_t|\mathbf{y})$ and compare them with the quantiles of the prior distribution $p(\lambda_t)$.

2.2.2 The Algorithm

We now describe an MCMC algorithm to implement the model. The parameter space is augmented with the vector of latent variables (λ, ν). The algorithm simulates from the joint posterior $\pi(\mathbf{h}, \omega, \lambda, \nu|\mathbf{y})$ by cycling from the conditional posterior distributions $p(\mathbf{h} \mid \omega, \lambda, \nu, \mathbf{y})$, $p(\omega \mid \mathbf{h}, \lambda, \nu, \mathbf{y})$, $p(\lambda \mid \mathbf{h}, \omega, \nu, \mathbf{y})$, and $p(\nu \mid \mathbf{h}, \omega, \lambda, \mathbf{y})$. Each block is simulated as follows:

- $p(\mathbf{h} \mid \omega, \lambda, \nu, \mathbf{y})$: The model is rewritten

$$\begin{aligned} y_t/\sqrt{\lambda_t} &= \sqrt{h_t} z_t, \quad z_t \sim N(0,1) \\ \log h_t &= \alpha + \delta \log h_{t-1} + \sigma_v v_t, \quad t = 1, \ldots, T \end{aligned}$$

  Since $y_t/\sqrt{\lambda_t}$ is observable, this is exactly the basic model in (1). We cycle from the conditional distributions $p(h_t \mid y_t, \omega, h_{t-1}, h_{t+1})$ in equation (3). We only need to replace y_t with $y_t/\sqrt{\lambda_t}$. The blanket q used is that discussed in section 2.1.

- $p(\alpha, \delta, \sigma_v \mid \mathbf{h}, \lambda, \nu, \mathbf{y})$: The $\mathbf{h}$'s are a sufficient statistics for $(\alpha, \delta, \sigma_v)$. Therefore, as in the basic model in (1), this step is a regression. The draws are the same as in the original algorithm.

- $p(\lambda|\mathbf{h}, \omega, \nu, \mathbf{y})$: Note that $p(\lambda \mid \mathbf{h}, \omega, \nu, \mathbf{y}) = \prod_{t=1}^{T} p(\lambda_t \mid y_t, h_t, \nu, \omega)$. The fact that λ_t is not autocorrelated, unlike h_t, is an essential identifying feature of the model. Then

$$p(\lambda_t|y_t, h_t, \nu, \omega) \equiv p(\lambda_t \mid \frac{y_t}{\sqrt{h_t}}, \nu) \propto p(\frac{y_t}{\sqrt{h_t}} \mid \lambda_t, \nu)\ p(\lambda_t \mid \nu)$$

For Student-t$_\nu$ errors, the prior and posterior of λ_t are

$$\begin{aligned} p(\lambda_t \mid \nu) &\propto \frac{1}{\lambda_t^{\frac{\nu}{2}+1}} e^{\frac{-\nu}{2\lambda_t}} \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{2}{\nu}\right), \\ p(\lambda_t \mid y_t, h_t, \nu) &\propto \frac{1}{\lambda_t^{\frac{\nu+1}{2}+1}} \exp\left(-\frac{(y_t^2/h_t)+\nu}{2\lambda_t}\right) \sim \mathcal{IG}\left(\frac{\nu+1}{2}, \frac{2}{\frac{y_t^2}{h_t}+\nu}\right). \end{aligned} \tag{5}$$

- $p(\nu|\lambda, \mathbf{h}, \omega, \mathbf{y})$: Consider a prior distribution $p(\nu)$. Note that $p(\nu|\lambda, \mathbf{h}, \omega, \mathbf{y}) \sim p(\nu|\lambda)$ because λ is a sufficient statistic for ν. One can make direct draws from the posterior distribution of ν which is

$$\begin{aligned} p(\nu|\lambda, \mathbf{h}, \omega, \mathbf{y}) &\propto p(\nu) \prod_{t=1}^{T} p(\lambda_t|\nu) \\ &\propto p(\nu) \times \left(\frac{\nu^{\nu/2}}{\Gamma(\nu/2)}\right)^T \exp\left(-\frac{\nu}{2}\sum_t (\frac{1}{\lambda_t} + \log\lambda_t)\right) \end{aligned} \tag{6}$$

We now discuss the choice of priors and further document the behavior of the fat-tail model.

2.2.3 The Priors

First, we use the same specification of $p(\alpha, \delta, \sigma_v)$ as in the basic model, i.e., a Normal-Gamma prior as in a standard regression model. We truncate the prior for δ to force the stationarity of $\log h_t$. In a simulation framework, this truncation is effected by rejecting *posterior* draws that violate stationarity. Alternative priors for δ are possible. Kim et al. (1998) use a Beta prior for δ. One must be careful here not to unwittingly force a very tight prior around, say 0.9. This could produce a false impression of convergence or precision of the algorithm. Our method of truncating the posterior could be feared to be ineffective. We track the number of times we encounter and reject draws of δ above 1. For all the estimations performed on ten financial series, each with often above 50000 draws, we encountered draws above 1 only twice after burn-in periods of 5000 draws. This is independent of where the algorithm starts. The data are tremendously informative about persistence but equally so in favor of stationarity.

An alternative view is that this is due to a phenomenon similar to that noted for near non-stationarity in the mean, i.e., the likelihood goes to $-\infty$ as one approaches the unit root. On the one hand the location estimates we obtain seem quite far from one for this to be the reason. No matter what model we use, most location estimates are below 0.98, sometimes around 0.9. On the other hand the non-stationarity of an unobserved latent variable has a more complex effect on the likelihood than that of the observed mean process. Geweke (1994) discusses the plausibility of flat priors as a function of the sampling interval. He proposes an alternate prior to allow the formulation of odds ratios for non stationarity. Here we also believe for economic reasons that a non-stationary stochastic volatility may be undesirable as well as unlikely. For example it would mean that portfolio managers should permanently re-balance their portfolios, or traders permanently re-adjust their option price, after a volatility shock. Also, Christoffersen and Diebold (1997) find evidence in favor of stationarity in a model free context.

For simplicity we take the prior distribution $p(\nu)$ to be uniform discrete on $[5, 30]$. An upper bound as low as 30 may seem surprising. The right plot of figure 2 shows the prior distribution of $\sqrt{\lambda}$ for $\nu \in [5, 80]$. The difference between $p(\lambda|\nu = 30)$ and $p(\lambda|\nu = 80)$ is not great. So, excluding the higher values of ν does not seem costly. Further, one may argue that a high upper bound, say 60, implies an undesirably high prior probability on *normality*. For example, assume that for practical purpose, one considers values of ν above 20 to produce essentially normal errors. Then a prior domain $[5, 60]$ implies a 73% prior belief on normality, whereas a prior domain $[5, 30]$ implies a 40% prior belief on normality. One may then worry that an upper bound of 30 affects the behavior of the posterior in small sample, especially its upper quantiles. We will conduct a sensitivity analysis to our choice of upper bound by reporting some results obtained with an upper bound of 60.

The left plot of figure 2 illustrates the ability of the fat-tail model to generate marginal distributions very different from the basic SVOL. It shows the quantile-quantile plot of densities generated by a basic and a fat-tail. Even a moderately low value of ν, here 10, produces data with markedly more fat-tailness than the basic SVOL data. The other parameters are, $\delta = 0.95, Var(h)/E(h)^2 = 1$. So the prior does not require an extremely low bound for ν. The right plot of figure 2 shows that for $\nu = 5$, there is a 5% prior probability of a $\sqrt{\lambda}$ larger than 2.14. Values of ν below 5 imply that the first four moments of ϵ_t are undefined. We believe this to be extreme since there is already a time varying volatility. Also, Eraker, Jacquier, and Polson (1998) mention possible convergence problems for very low values of ν.

We note here an alternative way to decrease the implicit prior odds on normality; $p(\nu) \propto e^{-g\nu}$ discussed in Geweke (1993). A point of convenience in favor of the uniform prior is that the resulting posterior $p(\nu|\mathbf{y})$ is proportional to the marginal likelihood function for ν namely $p(\mathbf{y}|\nu)$. This provides a model selection criterion for conditional normality analogous to a Bayes factor. We can then interpret the marginal posterior as the marginal likelihood for inference about ν.

2.3 Correlated Errors and the Leverage Effect

2.3.1 The Model

We now discuss the correlation between ϵ_t and v_t. Melino and Turnbull (1991) estimate this model on exchange rates with the method of moments, see JPR (1994) and Andersen (1994) for discussions on the sampling performance of the method of moments. The model may be particularly important for stock returns. 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. This produces the leverage effect. We will show that it can also generate a significant amount of skewness in the marginal distribution of returns. This addresses the skewness of the conditional distribution of returns mentioned in Gallant, Hsieh, and Tauchen (1994). The model is

$$\begin{aligned} y_t &= \sqrt{h_t}\epsilon_t, \\ \log h_t &= \alpha + \delta \log h_{t-1} + \sigma_v v_t, \quad t = 1, \ldots, T \\ (\epsilon_t, v_t) &\sim N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) \end{aligned} \tag{7}$$

Figure 3 documents the behavior of the model. It shows that the marginal distributions

generated are very different from those generated by the basic SVOL. It also shows that the model produces a degree of leverage effect and skewness as observed in stock returns. We simulated 8000 observations of a basic SVOL and a correlated model with $\rho = -0.6$. We choose a negative value to explore the ability of the model to produce a leverage effect and the skewness observed in the data.

The top right plot is a Q-Q plot of the two series. The strong skewness is apparent in the asymmetric tails generated by the correlated model; fat on the left and thin on the right. That is, over and above a basic SVOL model. The correlation should be modeled because its omission may result in biases in volatility estimates and forecasts. For a given h_{t-1}, a negative shock ϵ_t likely results in a higher h_t than a positive ϵ_t of the same size. This higher h_t propagates into the future via the persistence δ. The larger ϵ_t is, the larger this effect is. Consider the 8000 data points. The top left plot shows the boxplots of $\sqrt{h_{t+1}}$ after a positive and a negative ϵ_t. Volatilities are clearly higher after a negative shocks. The asymmetry gets more pronounced as $|\epsilon_t|$ get larger. This is shown in the last four boxplots where we concentrate on the larger $|\epsilon_t|$'s. The bottom left plot shows that the phenomenon is apparent in the data themselves. It shows the boxplots for $|x_{t+1}|$. The differences clearly occur in the bulk of the distribution and are not due to outliers. In fact, in this sample the most extreme observations did not occur after very large shocks.

For comparison, the bottom right plot shows these boxplots for the daily value weighted CRSP index return from 1962 to 1987. To avoid crushing the bodies of the distributions, we removed the -20% return of October 19, 1987. Each pair of boxplots reveals the difference in volatility on a day after negative versus positive shocks. The volatility also increases with the size of x_t while, in the bottom left plot it does not. This is because, for the simulated sample, we allocate an observation to a boxplot on the basis of ϵ_t which we know since we know h_t. This lets us isolate somewhat the leverage effect. For the CRSP, we allocate on the basis of x_t, the return standardized by its unconditional variance. So we also see the SVOL effect: the size of x_t is directly related to h_t. The leverage effect persists a long time. For the CRSP index, the standard deviation of observations ten days after a $x_t < -2$ is 24% larger than after $x_t > 2$.

The basic model can not fit this phenomenon and most likely produces smoothed estimates of h_t that are too variable. This is because the larger h_t's coming along with negative shocks can only be incorporated in the model with a higher σ_v. This may lead to the observation made that *"negative outliers are more frequent than positive outliers"*. The fat-tailed model is also *tricked*. It mistakes the negative observations for outliers. This results in a lower estimate of ν, larger λ_t's and too low h_t's for the negative observations. This is because a fraction of h_t goes into λ_t which has no persistence. This in turn biases downward the predictions following these observations.

2.3.2 The Algorithm

In the correlated case, one needs to specify a prior in the space of covariance matrices where one element is fixed, see below. Our approach allows us to make direct draws of (transformed) parameters, see McCulloch et al. (1998) for another application. Let $u_t = \sigma_v v_t$, then $\mathbf{r}_t \equiv (\epsilon_t, u_t)'$ is the vector of noises for the correlated model. Consider its covariance matrix

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

The joint distribution of the data and the volatilities is then

$$p(\mathbf{y}, \mathbf{h}|\alpha, \delta, \Sigma^*) \propto \prod_{t=1}^{T} h_t^{-\frac{3}{2}} p(\frac{y_t}{\sqrt{h_t}}, \log h_t | h_{t-1}, \alpha, \delta, \Sigma^*) = \prod_{t=1}^{T} h_t^{-\frac{3}{2}} |\Sigma^*|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma^{*-1}A)\right), \quad (8)$$

where $A = \sum_t \mathbf{r}_t \mathbf{r}_t'$ is the residual matrix. The joint posterior $p(\mathbf{h}, \Sigma^*, \alpha, \delta \mid \mathbf{y})$ is then

$$p(\mathbf{h}, \Sigma^*, \alpha, \delta|\mathbf{y}) \propto p(\Sigma^*)\, p(\alpha, \delta) \prod_{t=1}^{T} h_t^{-\frac{3}{2}} |\Sigma^*|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma^{*-1}A)\right) \quad (9)$$

Specifying the prior $p(\Sigma^*)$ is difficult because one element is fixed. One cannot use the standard conjugate family, i.e., inverse Wishart, as it does not allow a subset of the elements to be fixed. We model (σ_v, ρ) jointly by considering the regression of u_t on ϵ_t:

$$u_t = \psi \epsilon_t + \eta_t, \quad \eta_t \sim N(0, \Omega).$$

ψ and Ω are the slope and noise variance of the regression, so $\sigma_v^2 = \psi^2 + \Omega$ and $\rho = \psi/\sqrt{\psi^2 + \Omega}$. An additional intuition on this parameterization is gained by noticing the alternative formulation for equation (7):

$$\log h_t = \alpha + \delta \log h_{t-1} + \sigma_v \rho \epsilon_t + \sigma_v \sqrt{1-\rho^2}\eta_t,$$

where η_t is a unit normal shock uncorrelated with ϵ_t. The parameters $\psi = \rho\sigma_v$ and $\Omega = \sigma_v^2(1-\rho^2)$ are apparent in this formulation. The prior on Σ^* is now straightforward: $p(\Sigma^*) = p(\psi|\Omega)p(\Omega)$. In this regression setup, a normal-gamma prior is natural for $(\psi \mid \Omega)$, and Ω. A draw of (ρ, σ_v) is computed directly from the draw of (ψ, Ω). We can now write the various blocks of the MCMC algorithm.

- $p(\psi, \Omega|\alpha, \delta, \mathbf{h}, \mathbf{y})$: We note that $|\Sigma^*| = \Omega$, and rewrite Σ^{*-1} as a function of ψ and Ω:

$$\Sigma^{*-1} = \frac{1}{\Omega}\begin{pmatrix} \psi^2 & -\psi \\ -\psi & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \equiv \frac{C}{\Omega} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

  Then it follows from (9) that

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

  This is because a_{11} does not involve ψ or Ω. We can now write $\mathrm{tr}(CA) = a_{22.1} + (\psi - \hat{\psi})^2 a_{11}$ where $\hat{\psi} = a_{12}/a_{11}$ and $a_{22.1} = a_{22} - a_{12}^2/a_{11}$, to complete the square. Recall that A is a function of α and δ. We specify a normal / inverse gamma prior $p(\Omega) \sim \mathcal{IG}(\nu_0, \nu_0 t_0^2)$ and $p(\psi \mid \Omega) \sim \mathcal{N}(\psi_0, \Omega/p_0)$. The joint posterior of this regression step follows by conjugacy of the prior:

$$\begin{aligned} p(\psi|\Omega, \alpha, \delta, \mathbf{h}, \mathbf{y}) &\sim \mathcal{N}\left(\tilde{\psi}, \Omega/(a_{11} + p_0)\right) \\ p(\Omega|\alpha, \delta, \mathbf{h}, \mathbf{y}) &\sim \mathcal{IG}\left(\nu_0 + T - 1, \nu_0 t_0^2 + a_{22.1}\right) \end{aligned} \quad (10)$$

  where $\tilde{\psi} = (a_{11}\hat{\psi} + p_0\psi_0)/(a_{11} + p_0)$. σ_v and ρ are recovered through the inverse transformation.

- $p(\alpha, \delta \mid \psi, \Omega, \mathbf{h}, \mathbf{y})$: $\mathbf{h}$ and σ_v are sufficient statistics for this step. It is a regression where the error variance σ_v^2 is known. Given the normal prior $p(\alpha, \delta) \sim N(P_0, A\sigma_v^2)$, a direct draw of the posterior is made as for the basic model.

- $p(\mathbf{h}|\psi, \Omega, \alpha, \delta, \mathbf{y})$: As previously, we can break $\mathbf{h}$ into T components $h_t \mid h_{t-1}, h_{t+1}$ due to the AR(1) specification. It follows from the likelihood (8) that

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

Rewriting this in terms of (ψ, Ω) and after simplifications, we obtain

$$p(h_t|h_{t-1}, h_{t+1}, \psi, \Omega, \mathbf{y}) \propto \frac{1}{h_t^{\frac{3}{2}+\frac{\delta\psi y_{t+1}}{\Omega\sqrt{h_{t+1}}}}} \exp\left(\frac{-y_t^2}{2h_t}(1+\frac{\psi^2}{\Omega}) - \frac{(\log h_t - \mu_t)^2}{2\Omega/(1+\delta^2)} + \frac{\psi y_t u_t}{\Omega\sqrt{h_t}}\right), \quad (11)$$

where μ_t is the same as in equation (3), and we recall that $u_t = \log h_t - \alpha - \delta \log h_{t-1}$. We now consider the choice of the blanket q. The power term and the first two terms in the exponent can be treated as in the basic model, albeit with changes in the parameters. We approximate the log-normal kernel by an inverse gamma and combine it to the inverse gamma kernel. The resulting combined inverse gamma and its parameters are

$$\begin{aligned}
\phi_t \;=\; \phi_{LN} + 0.5 + \frac{\psi\delta y_{t+1}}{\Omega\sqrt{h_{t+1}}} \;&=\; \frac{1 - 2e^{\frac{\Omega}{1+\delta^2}}}{1 - e^{\frac{\Omega}{1+\delta^2}}} + 0.5 + \frac{\psi\delta y_{t+1}}{\Omega\sqrt{h_{t+1}}} \\
\Lambda_t \;&=\; (\phi_{LN} - 1)e^{\mu_t + .5\frac{\Omega}{1+\delta^2}} \;+\; \frac{y_t}{2}(1+\frac{\psi^2}{\Omega}) \\
q(h_t \mid h_{t-1}, h_{t+1}, \psi, \Omega, y_t) \;&\sim\; \mathcal{IG}(\phi_t, \Lambda_t) \propto \frac{1}{h_t^{\phi_t+1}}\, e^{\Lambda_t/h_t}, \qquad (12)
\end{aligned}$$

The question remains of the third term in the exponent of (11). Theoretical convergence results offer little guidance on the effect of ignoring this term in the design of the blanket. If we use the inverse gamma in (12), the Metropolis steps should adjust for the shape difference via the acceptance probability in (2). Our experimentation showed that this convergence does not occur for a practical number of draws. Figure 4 illustrates this. Consider typical values of the parameters $\delta = 0.95, V_h/E_h^2 = 1$, and a possibly large $\rho = -0.6$. We simulated samples of this model and plotted p and q for various observations. For the observation used in figure 4, $y_t/\sqrt{h_t} = 2.02$. The top left plot shows the (exact) p and the approximate p ignoring the third term. The bottom left plot shows the first two terms in the exponent and the third term, dotted line. Here ignoring the downward sloping third term leads to an overestimation of the location of p. The top right plot shows p again, and q computed as in (12), dotted line. The densities are normalized to sum to 1. This was done by simple univariate integration. The vertical bars are the 5^{th}, 50^{th}, and 95^{th} quantiles of p. One should find a better blanket than (12).

The third term involves $\log h_t$ and $\sqrt{h_t}$. Its approximation by an inverse gamma was unstable. This is because it sometimes implies a mode at a very unrealistic value of h_t. Instead, we implement a very simple approximation of the third term as a linear function of

$1/h_t$. To do this, we computing its value in two points on either side of the mode, and then its slope s. We then rewrite the kernel as $\frac{\psi y_t}{\Omega h_t} s$. We just subtract $\psi y_t s/\Omega$ from Λ_t in (12). This simple modification which hardly adds CPU time, has a dramatic effect on performance. In the top right plot in figure 4, the modified q in thick dashes now matches p very well. Another way to assess the potential performance of an algorithm is to plot the ratio p/q. This ratio drives acceptance and repeat probabilities. The flatter it is, the better the algorithm will perform. The bottom right plot in figure 4 shows the dramatic improvement in p/q resulting from the incorporation of the third term.

Without accounting for the third term, simulations showed that the algorithm systematically underestimated ρ. Typical posterior means would be as high as -0.2 for a true $\rho = -0.6$ and several thousand observations. We also computed a measure of (lack of) flatness for p/q, as the squared relative differences for p/q at the mode versus one point on each side. This measure was then averaged across all draws and observations. The results confirmed that the incorporation of the third term in q dramatically improves the flatness of p/q.

2.3.3 The Priors

Consider the parameters $\psi = \sigma_v \rho$ and $\Omega = \sigma_v^2(1-\rho^2)$. The priors for ψ, and Ω allow flexibility in the implied priors for σ_v and ρ. The joint density of ρ, σ_v follows from the normal gamma prior for Ω, ψ by a change of variables. It can be shown to be

$$p(\rho, \sigma_v) \propto \frac{\sigma_v^{-\nu_0}}{(1-\rho^2)^{\frac{\nu_0+3}{2}}} \exp\left(-\frac{\frac{p_0}{2}(\sigma_v\rho - \psi_0)^2 + \nu_0 t_0^2}{\sigma_v^2(1-\rho^2)}\right) \tag{13}$$

Even for the most general values of the hyper-parameters, the moments and shape of the distribution are quickly implied by bivariate integration or simulation. When $\psi_0 = 0$, or if the prior on Ω is diffuse, (13) simplifies considerably.

If $\psi = 0$, then $\rho = 0$ as well. If $\psi_0 = 0$, the prior of ρ is symmetric around zero. p_0 is the precision of ψ. It has an intuitive and direct effect on ρ. When p_0 is very large, this implies a tight prior for ρ. When p_0 gets small, this results in a bimodal distribution for ρ since it is bounded. This allows the specification of a prior in favor of a non zero correlation of either sign. For the estimation, we chose $\psi_0 = 0$, and $p_0 = 2.5$. This value of p_0 was low enough to allow large values of ρ of either sign. For Ω, we chose a very diffuse prior with 1 degree of freedom and a very small sum of squares 0.005. This is consistent with the prior used for σ_v for the uncorrelated models. Figure 5 shows the joint and marginal distribution of this prior. A close inspection of the marginal for ρ shows a very small bimodality. One could be worried that the prior seems to taper off quickly beyond 0.8. We will address this concern by overlaying our posterior histograms with the prior. This will show that the data information overwhelms our prior for the series studied. If desired, large values of ρ of either sign can be allowed by further decreasing the precision. This will be at the cost of more bimodality, which is logical. As for Ω, the degree to which the data may swamp the prior can be anticipated by putting reasonable data values in the posterior in (10). ν_0 is much smaller than T. The expectation of $a_{22.1}$ can be shown to be $T\sigma_v^2(1-\rho^2)$, i.e., $960\sigma_v^2$ for $\rho = -0.6$ and $T = 1500$. This is orders of magnitude larger than our choice $\nu_0 t_0^2 = 0.005$ for realistic values of σ_v.

2.4 Algorithm for Fat-tails and Correlation

The full model is

$$\begin{aligned} y_t &= \sqrt{h_t \lambda_t} z_t, \\ \log h_t &= \alpha + \delta \log h_{t-1} + \sigma_v v_t, \quad t = 1, \ldots, T \\ \nu/\lambda_t &\sim \chi^2_\nu \\ (z_t, v_t) &\sim \mathcal{N}\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \end{aligned} \tag{14}$$

with prior $p(\alpha, \delta, \sigma_v, \nu, \rho)$ as before. Consider extending the correlated model to allow for fat tails. The extension is simple because of the hierarchical modeling. First, given λ, replace y_t with $y_t/\sqrt{\lambda_t}$ and the algorithm described in section 2.3 applies exactly. Second, $\mathbf{h}$ and $\mathbf{y}$ are sufficient statistics to draw from the posteriors in (5) and (6). The iteration of these two steps constitutes the algorithm of the full model.

2.5 Factor Model for Multivariate SVOL

2.5.1 The Model

A very popular approach to the modeling of homoskedastic covariance matrices in financial application is a factor structure. The factor structure, widely used in asset pricing, reduces the dimensionality of the parameter space. A MCMC algorithm for this model with constant variance is implemented by Geweke and Zhou (1996). Here we allow for stochastic volatility in the factors. This in turns allows for stochastic cross-correlations. This type of specification is needed in the study of cross-sections of assets when dynamic factors are hypothesized to be priced. See for example Uhlig (1997), and Jacquier, Polson, and Rossi (1995) for alternative multivariate specification. See Engle, Ng, and Rothschild (1990) for an incorporation of the factor structure into the ARCH framework. Harvey, Ruiz, and Shephard (1994) approximate the model in order to use the Kalman filter. This unnecessary approximation, however exhibits poor performance even in univariate setups. Finally Aguilar and West (1998) implement a model similar to the one below.

Consider q assets with returns $\mathbf{y}_t = (y_{1t}, \ldots, y_{qt})\prime$ where $1 \leq t \leq T$. Suppose that there are k underlying time varying factors $\mathbf{F}_t = (F_{1t}, \ldots, F_{kt})\prime$ with variance covariance matrix $\mathrm{H}_t = diag(h_{it})$. These factors generate the returns according to

$$\begin{aligned} \mathbf{y}_t &= \mathrm{B}\mathbf{F}_t + \Omega^{-\frac{1}{2}} \epsilon_t, \\ \mathbf{F}_t &\sim \mathcal{N}_k(\mathbf{0}, \mathbf{H}_t) \\ \log h_{it} &= \alpha_i + \delta_i \log h_{i,t-1} + \sigma_{iv} \mathbf{v}_t, \quad i = 1, \ldots, k \\ p(\mathrm{B}, \Omega, \omega) &= p(\Omega) \prod_{i=1}^{k} p(\mathrm{B}_i) p(\omega_i), \end{aligned} \tag{15}$$

where $\mathrm{B} = (\mathrm{B}_1, \ldots, \mathrm{B}_k)$ is a matrix of factor loadings, and $\epsilon_t \sim \mathcal{N}(0, I_q)$. Let $\omega = (\ldots, \alpha_i, \delta_i, \sigma_{iv} \ldots)$ be the vector of volatility parameters 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. Under these assumptions we have a time-varying factor structure to the evolution of the variance of the observations $\Sigma_{y_t} = \mathrm{B}H_t\mathrm{B}' + \Omega$. We do not necessarily assume that Ω is diagonal, only that it is constant. That is, the k factors are introduced to model the stochastic variation in the covariance matrix of Σ_{y_t}. The model will be simplified if one assumes that Ω is diagonal.

The product $\mathrm{B}\mathbf{F}_t$ does not identify B and F uniquely when the factors are not observable. A common approach in the i.i.d. factor literature is to set the factor variance to 1. For example this can be achieved here by setting $\alpha_i = -0.5\sigma_{vi}^2/(1+\delta_i)$. This is however not sufficient to identify the parameters in B. See Geweke and Zhou (1996) and Aguilar and West (1998) for discussions of the further restrictions on the number of parameters in B and the maximum number of factors possible.

2.5.2 The Algorithm

With this specification of the covariance matrix, the posterior distribution $\pi(\mathbf{F}, \mathrm{B}, \omega|\mathbf{y})$ can be broken into k conditionals $\pi(F_j|F_{-j}, \mathrm{B}, \omega, \mathbf{y})$, and the conditional $\pi(\mathrm{B}, \omega|\mathbf{F}, \mathbf{y})$. The first set of conditionals is simulated via a simple transformation of the basic univariate algorithm. For the model in (15), the joint posterior distribution conditional on the factors is

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

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

$$\begin{aligned} p(\mathrm{B}|\Omega) &\propto |\Omega|^{-\frac{k}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Omega^{-1}(\mathrm{B} - \mathrm{B}_0)A_0(\mathrm{B} - \mathrm{B}_0)\right)\right) \\ p(\Omega) &\propto |\Omega|^{-\frac{\nu_0}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Omega^{-1}B_0\right)\right) \end{aligned}$$

for specified hyper-parameters (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. In the first case, by independence of the factors, the joint posterior of the volatilities and parameters, $\pi(\mathrm{B}, \mathbf{H}, \omega \mid \mathbf{F}, \mathbf{y})$ decomposes as a product

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

The individual terms $\pi(h_i, \omega_i|\mathbf{F})$ can be generated with the univariate SVOL algorithm, with fat tails if desired.

Then the density $\pi(\mathrm{B}, \Omega|\mathbf{F}, \mathbf{y})$ follows from standard conjugate multivariate analysis.

When $\mathbf{F}$ is unobserved we need to simulate from the joint posterior $\pi(\mathbf{H}, \omega, \mathrm{B}, \Omega, \mathbf{F}|\mathbf{y})$. It is broken into into the conditionals $\pi(\mathbf{H}, \omega, \mathrm{B}, \Omega|\mathbf{F}, \mathbf{y})$, and $\pi(\mathbf{F}|\mathrm{B}, \Omega, \mathbf{H}, \omega, \mathbf{y})$. The first distribution

reduces to the case with known factors since it is conditional on $\mathbf{F}$. The second conditional can be decomposed as

$$\begin{aligned}\pi(\mathbf{F}|\mathrm{B},\Omega,\mathbf{H},\omega,\mathbf{y}) &\propto p(\mathbf{F}\mid\mathbf{H})\,p(\mathbf{y}\mid\mathbf{F},\mathrm{B},\Omega) = p(\mathbf{F}|\mathbf{H})\prod_{t=1}^{T} p(\mathbf{y}_t\mid F_t,\mathrm{B},\Omega)\\ &\propto |\Omega|^{-\frac{T}{2}}\exp\left(-\frac{1}{2}\mathrm{tr}\left(\Omega^{-1}(\mathbf{y}-\mathrm{B}\mathbf{F})'(\mathbf{y}-\mathrm{B}\mathbf{F})\right)\right)\exp\left(-\frac{1}{2}\mathrm{tr}\left(\mathbf{F}'\mathbf{H}\mathbf{F}\right)\right),\end{aligned}$$

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

3 A Simulation Study

3.1 Parameter Settings

We have shown that, if present in the data, the extension of the conditional distribution can have important practical consequences, for example on forecasts made with the model. We now incorporate parameter estimation to further our argument that these extensions are crucial. We do this by studying the sampling performance of the algorithms for various scenarios, or parameter settings. First, we show that, even if the data are generated by the basic model, estimating the extensions does not cause much performance deterioration. Then we show that if the extensions are present in the data, ignoring them and estimating the basic model instead does lead to significant performance deterioration. Finally, we show that the algorithms presented in section two above can reliably retrieve the extended models and result in better smoothing, especially for outliers.

The basic idea is to document performance when one estimates the right model, and also when one estimates the wrong model, i.e., the basic model when the data originate from an extended model or vice-versa. There are clearly many such combinations and we report on a subset of these for the sake of brevity. We will show results for both the parameters, table 1, and the smoothed volatilities, table 2.

The top row of table 1 gives the parameter values common to all the settings. They are $\delta=0.95, E_h=0.0009, V_h/E_h^2=1$, the values in a central cell in JPR (1994) used by others as well. E_h and V_h are the unconditional mean and variance of h_t. They follow from the assumption that h_t is lognormally distributed. These are realistic values for financial series. For the extensions, $\nu=10$ and $\rho=-0.6$ represent the presence of a strong deviation from the basic model. These deviations are stronger actually found in the financial series. In this section, we set the domain of ν at $[5,40]$ when we estimate it. The table shows 5 sets of simulations. We will refer to them as *setting* 1 to *setting* 5. Each setting includes 500 samples of either 500 or 1000 observations. For each sample the estimation is based upon the 15000 draws that follow a burn-in period of 5000 draws.

3.2 Results

Consider setting 1, the basic model with sample size $T=500$. The first two rows show the sample mean and RMSE of the posterior mean as an estimator of the true parameter. These numbers are from JPR (1994). The last three rows show the performance when we estimate ν even though it is

not needed since the data are generated by the basic model. The first row shows that this has no effect on the bias of σ_v and δ. The second row shows an increase in RMSE for δ, a decrease for σ_v. However the RMSE's remain small.

The third row shows a specification test based on the entire posterior not just the mean. We report the percentage of samples for which a 50% and a 90% posterior credibility interval covers the true value. The numbers in the third row for δ, σ_v, E_h prove the remarkable reliability of the posterior uncertainty delivered by the algorithm. The coverage intervals for V_h/E_h^2 shows the extent to which the algorithm allocates some variation in volatility to λ, slightly underestimating V_h/E_h^2.

We now look at the posterior mean of ν. Its sampling average is high; 26. For ν, the second row reports the sampling average of the 3 quartiles of the posterior distribution. So, the first quartile was 19 on average. The third row reports the percentage of the samples for which the quartile was larger than 20. This tells us how much information we can hope to get on ν for a given sample size. Here, the median was above 20 for 89% of the samples. The fourth column reports the sampling average of the posterior probability $p(\nu < 20)$, here 31%. The two numbers on the second row for $p(\nu < 20)$ show the fraction of samples which resulted in $p(\nu < 20) > 0.5$ and > 0.75 respectively. In this case, only 10% of the samples produced a $p(\nu < 20)$ above .5.

Overall, the results in setting 1 show that (1) the estimation of the fat-tail extension imposes little cost on the parameters of the basic model, and (2) the posteriors for ν unambiguously point towards values above 20. This is even for a sample as small as T=500.

Parameter setting 2 involves 500 samples of 500 observations with fat-tails ($\nu = 10$). The inspection of $\delta, \sigma_v, E_h, E_h^2/V_h$ shows the excellent performance of the algorithm. There are hardly any biases and the RMSE's are small. The 50% and 90% posterior credibility intervals are on the mark for these parameters. The posterior means for E_h and V_h/E_h^2 show that the algorithm allocates the right fraction of variability to h_t versus λ_t. With only 500 observations, the posterior mean of ν is now below 20 on average. The evidence against normality is strong though not overwhelming. The third row for ν show that only 4% of the samples produce a posterior first quartile above 20, which could be seen as misleading evidence in favor of normality. The posterior distribution of ν is indeed very different from that documented in setting 1 above. $p(\nu < 20)$ is now .58 on average. It is above .5 for 63% of the samples.

In setting 3, we consider $\nu = 10$ again, but with 1000 observations. The dramatic increase in precision for all the parameters is documented by their sampling RMSE's. The RMSE for δ is 0.02. The other RMSE's are now a third less than for T=500 observations. The identification of ν also improves, albeit less dramatically. The posterior median of ν is below 20 for only 71% of the samples. Now 42% of the samples result in $p(\nu < 20) > 0.75$. For these samples, we also estimated the basic model. This is the second set of numbers under setting 3 in table 1. Apart from δ, all the posterior distributions now have important biases. As we might expect, σ_v, E_h and V_h/E_h^2 are overestimated. This is confirmed by the posterior credibility intervals. For example, a 50% credibility interval for E_h covers the true value for only 21% of the samples. A 90% credibility interval for σ_v covers the true value for only 65% of the samples. So, not allowing for fat-tails when they are present results in misspecified posteriors for the parameters of the basic model.

In setting 4, we study the estimation of the correlated model when the true ρ is -0.6. This is arguably a very large value. The prior we use puts a relatively low weight on values of $|\rho \in [-0.6, -1]$. Yet, the sampling average of the posterior mean, -0.54, is close to this. The

posterior of ρ delivers coverage intervals close to their theoretical values of 0.5 and 0.9. The estimation of the other parameters does not suffer measurably from the introduction of ρ. Setting 5 shows the estimation of both $\rho = -0.6$ and $\nu = 10$ together in a full model. Again, the posteriors for the parameters other than ρ and ν are very well specified with little or no bias, low RMSE, and good coverage intervals.

We now see that the introduction of ρ does not affect the performance for ν. The average posterior mean of ν is 14. The average value of $p(\nu < 20)$ is 0.79, and it is above 0.5 for 93% of the samples. ρ however is affected by the introduction of ν. This average posterior mean for ρ is now -0.42, and the coverage intervals miss the value of -0.6. Nevertheless, the posterior mean is high and the correlation is unambiguously identified at no cost for the other parameters. Classic tests of convergence, such as the autocorrelation of the sequences of draws indicate that the algorithm has converged. We also experimented with different starting points and longer sequences of draws. Neither changed the results. However, a limited number of simulations, not included here, show that the use of larger samples improves the posterior of ρ even when ν is estimated. To an extent, ν and ρ compete for the interpretation of outliers, including the negative ones. Note that the sample size T=1000 is much smaller that the shortest sample used in the empirical section. We discuss this issue further in section 4.5.

Table 1 shows that (1) the costs of not estimating the extensions when they are present is high, and (2) the costs of unnecessarily estimating the extensions when they are not present are minor in comparison. The other stylized result apparent from these studies is that, with a short sample of T=1000 or less, one may want to do some sensitivity analysis to better understand the robustness of the posterior of ρ. We discuss this further in the empirical section. ρ and ν are probably revealed efficiently by only a subset of the observations. One may argue that the parameters are only a vehicle for the really interesting use of the model; the estimation of the time varying volatility. So we now document the performance of the algorithms with respect to the volatility estimates.

Table 2 reports the smoothing performance for setting 3 where $\nu = 10$, and $T = 1000$. For each observation we compute an error, equal to the difference between the posterior mean and the true value. We do this for the standard deviation $s_t \equiv \sqrt{h_t \lambda_t}$ and for $\sqrt{h_t}$. We then compute the Root Mean Squared Error and the Relative Mean Absolute Error by averaging over the 1000 observations of the 500 samples. The last three columns show the results broken down by the size of λ_t. The first (last) three rows show the performance when estimating the fat-tail (basic) model. The two models have very similar performance in fitting σ_t. This is actually not good for the basic model because the key here is to fit $\sqrt{h_t}$ properly, recall figure 1. If the basic model fits s_t nearly as well as the fat-tail model, then it must result in a misspecified $\sqrt{h_t}$. The second and third rows of results show the extent of the problem. In both absolute or relative terms, the fit of the basic model for $\sqrt{h_t}$ is significantly worse than that of the fat-tail model. It is worse by a factor of 21% on average. The worst fit occurs for the observations with the larger λ_t's. For example, the fit is 30% worse with λ_t in the top 10 percent. This confirms that the use of the fat-tail model greatly improves the quality of the volatility estimates.

4 Empirical Applications: Equities and Exchange Rates

4.1 The Data

We now report the results of the analysis of weekly and daily financial series. There is evidence in the literature that these series may have different volatility patterns. For example, higher frequency series have been argued to exhibit more volatility persistence. Lower frequency series have also been argued to exhibit less non-normality as the higher frequency jumps may average out over time. The weekly series are the equal and value weighted CRSP indices and three size sorted decile portfolios. We choose the first, tenth and fifth deciles, that is the 10% smallest and largest firms as well as a representative index of medium size firms. The sample period goes from 1962 to 1991. This constitute 1539 weekly returns. The returns are prefiltered for an AR(1) component as in JPR (1994).

The daily series include two stock index returns and three exchange rates. Whereas one can find good reasons for a leverage effect in equities, it is less clear with respect to exchange rates. The stock indexes are the S&P500 from 1980 to 1987 filtered to remove calendar effects see Gallant et al. (1992), and the CRSP value weighted index from 1962 to 1987. The latter is the dataset used by Nelson in his work on EGARCH. It contains 6409 observations. The S&P500 index contains only large firms while the CRSP index contains all the firms on the exchange. The exchange rates are the £, Deutsche Mark, and Canadian dollar to the US dollar. The sample periods go from January 1975 to December 1986 for the Canadian dollar, that is 3010 observations. It is the series used in Melino and Turnbull (1990) and JPR (1994). The £ and DM series go from January 1980 to June 1990, that is 2613 observations. We analyze the first difference of the logarithm of the exchange rate series. They display essentially no autocorrelation.

We first report on the model with fat-tails and no correlations, tables 3 and 4. We then report on the model with fat-tails and correlations in tables 5 and 6. This allows to include our sensitivity analysis of the results as we discuss the posteriors. Further sensitivity analysis is contained in table 7, including the estimation of the correlation model with no fat-tails.

4.2 Fat-tail Errors: Parameters

Let us first discuss the results of our analysis of the fat-tailed model of section 2.2 The degrees of freedom parameter ν will provide an inference on the severity of the departure from normality. Table 3 summarizes the posterior distribution of the parameters for the weekly series. Table 4 shows the results for the daily series.

Consider first the weekly data. The decile results show that the volatility of the smaller firms is less persistent, see the posterior means δ for d1, d5, d10. Also, the smaller firms exhibit a higher volatility of volatility. This is shown by the posterior means of σ_v and V_h/E_h^2 for d1, d5, and d10. Table 3 also provides posterior means for δ, σ_v and V_h/E_h^2 from the basic model. It confirms that the introduction of ν has lowered the posterior estimates of the volatility of volatility. This is revealed by the comparison of σ_v and V_h/E_h^2 for EW, VW, d5, d10. A fraction of what was construed in the basic model as volatility of volatility is now reallocated to the λ_t's.

We now turn to the posterior distribution of ν. They are all centered at or below 21. The smaller firms, represented in EW and d1, exhibit the thickest tails. The odds ratios of $\nu < 20/\nu > 20$

confirm this. Recall that our prior distribution on ν implies a prior ratio of 1.5 to 1. These results also confirm the simulations. One should not expect extreme precision in the inference on ν. The posterior distributions are relatively spread out. This 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 a valid representation of the posterior uncertainty.

One might argue that the results are hard to interpret because of the choice of prior domain [5, 30]. How strong is the evidence against normality? We conduct a sensitivity analysis by estimating the models with $\nu \in [5, 60]$. The results are reported in the last row of table 3 (and 4). The difference between small and large firm portfolios is now clear. The upper bound at 30 did not significantly affect inference for EW and d1, which posterior location have not changed. However, the posterior means for VW, d5, and d10 have moved up to 25, with 95^{th} percentile around 57. Also, the prior specification on ν did not affect the other parameters. So we do not report them. Note finally that even for VW, d5, and d10, that have relatively high values of ν, the mere *freeing* of ν hence of λ_t produces estimates of σ_v and V_h/E_h^2 lower than with the basic model.

Table 4 reports the daily results. The persistence of daily series does seem higher, as shown by δ. This is however by no means an overwhelming pattern. The evidence of low ν is very strong for the exchange rates and the daily S&P500. The Canadian \$ results depart from this pattern. The posterior mean of ν is 26. There appears to be mass close to the upper bound. This is reflected in the posterior odds ratio of 0.13 to 1 in favor of $\nu > 20$. The last row contains the estimation with $\nu \in [5, 60]$. The posterior mean for ν is then 47. For the other series, the posteriors are not affected by the choice of domain for ν.

The comparison of the S&P500 and the CRSP index is also interesting. The S&P500 exhibits more fat-tailness than the CRSP. This can be due to the fact that the S&P estimates are for 1980-1987 only. A day such as in October 19, 87 may have more weight than in the CRSP estimation which includes 35 years of data.

The data provide strong evidence against normality. However, they do not support a very low degree of freedom for the Student-t distribution. This is in contrast with the findings of the ARCH literature. As we might expect, after the extra mixing induced by the SVOL model results in somewhat less fat-tailed conditional distributions than the ARCH models. Still it does in no ways obviate the need to model the fat-tailness of this conditional distribution. Finally, even though the estimated persistence of the fat-tailed models is often higher than that of the basic model, there is still no hint of non stationarity in volatility even for daily data.

We now show one diagnostic which confirms that the data do not require very low degrees of freedoms. Consider the S&P 500 daily returns. We generate the marginal distribution of the fat-tail SVOL model by simulating 500 samples from the model, with parameters equal to the posterior means in table 4. We sort each sample. Then we average the sorted observations over the samples. This produces an empirical estimate of the quantiles of the marginal distribution of the fat-tail SVOL. We use it in a qq-plot against the actual data. We omit the -20% return of Oct 19, 1987 in order to make the Q-Q plot more informative. Granted that no model fits that day very well, a plot including it prevents us from seeing how well the rest of the distribution is fitted.

The top plot of figure 6 shows the Q-Q plots of the returns versus the marginal distribution of

the fat-tail model with the parameters as in table 4. We then build another marginal distribution using $\nu = 5$ instead. The bottom plot shows a Q-Q plot of that distribution against the stock S&P500 returns. The straight line is a 45 degree line along which the plot should be in case of perfect distributional equality. The dotted line is a least square fit lines. It allows to ignore differences in unconditional mean and variance and concentrate on the fit of higher moments. The top panel shows the close fit between the marginal distribution of the data and the fat-tail SVOL model estimated in table 4. The bottom plot shows that a smaller value of ν leads to a very poor fit even in the bulk of the distribution. First, this confirms the difference between SVOL and ARCH models for which very low degrees of freedoms were required in the conditional distribution. Second, it confirms that the algorithm converges on appropriate values for ν. Very small values of ν would be unreasonable and most likely the result of a convergence failure.

4.3 Scale Parameters λ_t

A unique feature of our hierarchical framework is the ability to infer about the draws of the scale-mixing parameter λ_t, for each observation. This is a convenient method of outlier analysis which we document. If a particular observation is outlying, the model centers the posterior $p(\lambda_t \mid \mathbf{y})$ at a large value. This re-weighs the outlier and brings $p(h_t \mid \mathbf{y})$ more in line with the rest of the observations. We can then identify outliers by looking at each marginal posterior $p(\lambda_t|\mathbf{y})$.

The top plot of Figure 7 shows the daily S&P 500 index returns. Obviously, October 19, 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 bottom plot show the daily posterior means $\mathrm{E}(\lambda_t \mid \mathbf{y})$. This estimation was done with ν fixed. For robustness of the result we estimated ρ as well, see the next section and table 7 for the parameter estimates. We note that the outlier diagnostic does not require the estimation of ν. The implied prior on $\sqrt{\lambda_t} \mid \nu = 15$ can be seen in figure 2. It is quite spread out. This prior mean is the horizontal line on the bottom plot of figure 7. Higher quantiles could be plotted and this would allow for the flagging of outlying observation.

4.4 Correlated and Fat-tailed Errors

We now augment the fat-tail model and allow for correlated errors. Tables 5 and 6 summarize the posterior distributions for the correlated case for the weekly and daily series. In table 6, we add estimates for the CRSP return for the period 80-87. This is to match the S&P500 estimation period. Recall that the daily CRSP value-weighted is the series used in Nelson's seminal (1991) paper where he introduces the EGARCH model with asymmetric conditional variance.

Nelson's estimates of the EGARCH parameters are on the edge of non-stationarity. This is the case for most GARCH estimates as well. Compare the persistences in tables 5 and 6, with those in tables 3 and 4. Despite differences in point estimates, there is still no evidence of unit roots in the volatility equation. The posteriors damp down near 1 and no appreciable mass is put on the region above .99. This occurs in spite of a prior which is locally uniform around 1.0. So the introduction of the correlation has not appreciably affected the persistence. The difference between the persistences reported by ARCH and SVOL models is robust to the class of SVOL model used. Our results are consistent with the model free conclusions of Christoffersen and Diebold (1997).

We now look at the correlation ρ. It is large, in the vicinity of -0.4 for all weekly stock indexes, albeit a bit smaller for the small firm index. The correlation for the daily CRSP return over the 38 year period is also high, -0.43. It is lower for the daily S&P500, only -0.18. Even then, the evidence of leverage effect is overwhelming. The posterior probability of $\rho < 0$ is equal to 0.96. The posterior mean of ρ for the CRSP from 1980 to 1987 is -0.23. It is consistent with the S&P500 for the same period, and lower than the estimate for the entire 38 year period. This casts some doubt over the assumption of constant parameters over 38 years for the daily data. The posterior mean of δ for the 80-87 period is 0.978, lower than the value of 0.989 of the 38 year period. This is also consistent with some parameter shift. However the two posterior stil have large overlaps as shown by their quantiles.

The leverage effect is significantly lower for exchange rates, -0.02 for the Deutsche Mark, -0.13 for the British Pound. The Canadian dollar again stands as an oddity with $\rho = -0.29$, more like an equity than an exchange rate. One could conceive of a leverage effect for an exchange rate. This would require some economic rationale to attributes a consistent meaning to the direction of the rate over the sample period. This may be the reason why the Canadian rate exhibits a behavior different from, say, the Deutsche Mark. Note that ρ is precisely estimated. Recall that we use the refinement described in section 2.3.2 for the blanket $q(h_t|.)$. Without it, the standard deviation of ρ for the CRSP return would not be around 0.04 as in table 6, but 0.16.

We now discuss the estimates for ν. They differ sometimes from those in tables 3 and 4, but not in a significant manner. For the weekly returns, the posterior means in table 5 (ρ estimated) are close to those reported in table 3. The evidence for $\nu < 20$ which was already strong for EW is now stronger. For VW and d10, the evidence was weak, it is now weaker. These effects are not dramatic however. Table 6 shows strong evidence in favor of $\nu < 20$ for both the CRSP 80-87 and the daily S&P500, and for the exchange rates. The exception was the Canadian dollar. It still is. The introduction of ρ does not affect the posterior distribution of ν which is nearly identical to that in table 4. So, the introduction of ρ does not remove the need for fat-tailness where it was documented in the pure fat-tail models. The odds ratios for normality are qualitatively similar to those of tables 3 and 4.

We also note that the variances of the various posterior distributions in tables 5 and 6 are of similar to those in tables 3 and 4. This shows that the introduction of ρ has not decreased the precision of estimation of the other parameters. This confirms the simulation result discussed earlier, that the introduction of the additional parameters does not adversely affect the estimation of the basic model parameters.

We have taken a practical view that $\nu > 20$ resulted in error nearly normal error distributions. We now allow for the different view that normality is only when $\nu = \infty$. This is on the basis that even for $\nu > 20$, the model estimated does not behave as the basic model owing to the freedom left to the λ_t's. Under this strict view, one may argue that the odds ratios shown in tables 3 to 6, $p(\nu < 20)/p(\nu > 20)$ are not very interesting. We address this issue as follows. We generate the posterior distribution of ν resulting from the estimation of a fat-tail model when the data are in fact generated by the basic model. This posterior distribution is a function of the sample size, and the prior chosen for ν, which we both control for. To achieve stability we generate multiple samples of the basic model and 20000 draws of the posteriors for each sample. We can then compare this normal-errors based posterior to the data-based posterior.

In figure 8, we plot the posterior distributions for ν and ρ. We overlay ν with the diffuse

prior which we used and the normal-errors based posterior discussed above. Consider the left plots showing the posteriors and prior for ν. The dots represent the normal-errors based posterior. The evidence easy to interpret. The weekly EW errors, top plot, are unambiguously fat-tailed. The data have in fact moved us from the flat prior, i.e. the horizontal line, *away* from the normal-based posterior, not toward it. For the weekly VW returns, the data seem to have actually moved us toward the normal-based posterior. The errors of the daily VW index, bottom plot, on the other hand have evidence some fat-tailness, clearly centered around 20. The bottom plot of figure 8c represents the Canadian Dollar. This is an extreme, and only, example of the data pointing toward strict normality. The posterior is undistinguishable from what would obtain, had the data been generated by the basic SVOL model.

4.5 Posterior Sensitivity and Convergence

We use figure 8 to highlight the difference between the prior and posterior distributions for ρ as well. The right plots show the posterior and prior for ρ. This serve to reassure us that our prior did not prevent the data from producing larger values of ρ. In fact, the prior which we chose is nearly locally uniform for each of the 8 series plotted.

We conducted diagnostics on the convergence and information content of the chain for all the estimations presented here. The autocorrelations of the new parameters were not high. The autocorrelation of ρ in general starts below 0.6 and decays very fast. Further, the distributions obtained are always stable. We check this by computing the quantiles for various subsamples of the whole chain, and verifying their stability. We use an over-cautious burn-in period of 5000 draws. In nearly all of the cases a few hundred draws would have been more than enough to dissipate even ludicrous initial conditions.

The autocorrelation of the λ_t's were close to 0 even at lag 1. The time series plots of ν shows that it navigated the entire space for the results which we present here. One must warn that in some instances, the draws of ν can lock into the lower bound of the domain and never escape. This is obviously easily detected, see Eraker et al. (1998). Here we used a lower bound of 5, as discussed in section 2. However, tests based on a single run may fail to uncover convergence problems. The evidence from the sampling behavior of the posterior distribution is a much more robust test of convergence. Note that even there, an unfavorable result may be hard to interpret. This is because nothing guarantees that a posterior mean should be unbiased in small sample. So a biased sampling behavior may result from proper convergence to a biased posterior or failure of convergence to an unbiased posterior.

We now provide further evidence of the reliability of the posterior distributions of ρ for the series studied. This evidence is independent from convergence tests based upon the sequence of draws. Recall that our simulation experiments revealed some bias in the posterior of ρ when both ρ and ν are estimated, while there was no such bias for the correlated normal errors model. We might be concerned that the treatment of ν affects the inference on ρ.

To address this concern, we conducted a sensitivity analysis of the posterior of ρ (and other parameters) to the treatment of ν. Table 7 reports the posterior 5^{th}, 50^{th}, and 95^{th} quantiles of $\rho, E_h, V_h/E_h^2$ for three models for each series. The first model is with correlated normal errors. The second fixes $\nu = 15$ and estimates the λ_t's. The value 15 was chosen to represent a fairly low value of ν while leaving it fixed. The third model recalls the full model estimates of tables 5

and 6 for easy comparison. The last column of table 7 shows that the posterior distributions of ρ are not affected by the treatment of ν. That is, whether ν is large or small, and whether there is uncertainty about it or not, does not change the inference on ρ for these series. This further confirms that the posteriors of ρ for the full model have converged.

There may be two reasons for this difference with the simulation results. Recall that in the simulation we had a high *true* ρ of -0.6 and a sample size $T = 1000$. First, the combination of the large ρ and the small sample size (relative to the financial series), may have caused the prior to shrink ρ away from -0.6. This phenomenom may not be present with longer series better described with smaller values of ρ. Second, the smaller sample size used in the simulations may allow for an interaction (uncertainty) between the effect of ν and ρ when both are free. The larger sample sizes of the financial series provide enough information to resolve this interaction. This indicates merely that it is difficult to identify ρ from ν in small samples.

Finally table 7 shows that the basic model parameters react to the treatment of ν, exactly in the expected manner. The central columns report the distributions of δ, E_h and V_h/E_h^2 for the three models. For each series, the first row is the normal errors model ($\nu = \infty$). The second row then shows the results for the full model if the posterior mean of ν is larger than 15. Otherwise the second row shows the result of the model with a fixed $\nu = 15$. So for each series, going from row 1 to row 3 represents decreasing values of ν. Of course one must remember that for the full model, ν navigates around its posterior mean with draws above and below 15. The inspection of the column V_h/E_h^2 reveals the pervasive fact that the lower ν is, the lower the volatility of volatility is. Similarly, a lower ν is often associated with a higher δ and a lower average variance E_h.

5 Conclusion

A Bayesian hierarchical approach to the modeling of stochastic volatility is presented. MCMC algorithms coupled with data augmentation are developed. This allows for efficient parameter and volatility estimation for a large class of stochastic volatility models. This approach also provides an outlier diagnostic and a diagnostic for assessing conditional normality. Namely, we implement this framework for two major extensions of the basic stochastic volatility model, fat-tailed conditional errors and a correlation between conditional observable and variance errors, i.e., the leverage effect. We describe an algorithm for multivariate stochastic volatility models, see Aguilar and West (1998) for an implementation.

We show that these extensions can produce very different volatility patterns from the basic SVOL model, and the failure to incorporate them can lead to dangerously biased forecasts. We document the performance of the algorithm. First, we show that the incorporation of the extensions does not have an adverse effect on the estimation of the basic parameters even if the extensions are unwarranted. Second, the algorithm retrieves the extension parameters and the volatilities efficiently, effectively leading to predictions significantly different from the basic SVOL model.

The empirical analysis performed for weekly and daily financial series shows that the distributional extensions modeled are pervasive in the data. These extensions do modify the prediction of variance, most significantly during the periods of high variances or around isolated outliers. Economic models based on variance forecasts, especially asset allocation, risk management, and short term option pricing will most benefit from the incorporation of these extensions. They in fact give

flexibility to the first four moments of the conditional distribution of returns or growth rates, while keeping the model parsimomious.

The algorithm which we implemented is reliable and fast. The extensions have been modelled so has to involve direct uncorrelated draws of the additional parameters. For example, on a 360 Mhz workstation, for a sample size of 1000 observations, 15000 draws of the full model were generated in 7 minutes. A fruitful avenue for future research would be the parsimonious incorporation of these features in multivariate models of stochastic volatility.

References

Andersen, T. (1994). 'Comment on Bayesian Analysis of Stochastic Volatility" *Journal of Business and Economics Statistics*, 12, 4, 371-417.

Aguilar, O., and West, M. (1998). "Bayesian Dynamic Factor Models and Variance Matrix Discounting for Portfolio Allocation", ISDS Discussion paper 98-03.

Black, F. (1976), "Studies of Stock Market Volatility Changes", Proceedings of the American Statistical Association, Business and Economic Statistics Section, pp. 177-181.

Bollerslev, T., Chou, R. and Kroner, K. (1992), "ARCH Modeling 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 Non-conjugate Bayesian Models Using the Gibbs Sampler", *Canadian Journal of Statistics*, 19, 399-405.

Chib, S., Kim, S. and Shephard, N. (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models", *Review of Economic Studies*, 65, 361-393.

Chib, S., and Greenberg, (1995), "Understanding the Metropolis-Hastings Algorithm", *The American Statistician*, 49, 327-335.

Christoffersen, P., and Diebold, F. (1997), "How relevant is volatility forecasting for risk management", Working paper, Penn. Economics department.

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

Eraker, B., Jacquier, E. and Polson, N. (1998), "Pifalls in MCMC Algorithms", Department of Econometrics and Statistics, Graduate School of Business, University of Chicago Technical Report.

Fridman, M., and Harris, L. (1998), 'A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models", *J. Business and Economics Statistics*, 16, 3, 284-291.

Gallant, A.R., Hsieh, D., and Tauchen, G. (1994), "Estimation of Stochastic Volatility Models with Diagnostics", *Duke Economics Working Paper #95-36*,

Geweke, J. (1989). "Exact Predictive Densities for Linear Models with ARCH disturbances". *Journal of Econometrics*, 40, 63-86.

Geweke, J. (1993). "Bayesian Treatment of the Independent Student-t Linear Model". *Journal of Applied Econometrics*, 8.

Geweke, J. (1994). "Priors for Macroeconomic Time Series and Their Application". *Econometric Theory*, v10, p 609-632.

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

Geweke, J. (1994c). 'Comment on Bayesian Analysis of Stochastic Volatility" *J. Business and Economics Statistics*, 12, 4, 371-417.

Geweke, J. and Zhou, G. (1996). 'Measuring the Pricing Error of the Arbitrage Pricing Theory" *Review of Financial Studies*, 9, 2, 556-587.

Glosten, L., Jagannathan, R., and Runkle, D. (1993), "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks" *Journal of Finance*, 48, 1779-1801.

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.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods using Markov Chains and their Applications" *Biometrika*, 57, 97-109.

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, 371-417.

Jacquier, E., Polson, N. and Rossi, P. (1995), "Priors and Models of Stochastic Volatility Models", Rodney White Center working paper, Wharton.

Kim, S. and Shephard, N. (1994), "Comment of Bayesian Analysis of Stochastic Volatility by Jacquier, Polson, and Rossi, " *Journal of Business and Economics Statistics*, 12, 4, 371-417.

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

Mc Culloch, R., Polson, N. and Rossi, P. (1998), "Bayesian Analysis of the Multinomial Probit with Fully Identified Parameters", University of Chicago working paper.

Melino, A. and Turnbull, S.M. (1990), "Pricing Foreign Currency Options with Stochastic Volatility", *Journal of Econometrics*, 45, 239-265.

Mengersen, K., and Robert, C. (1998), "MCMC Convergence Diagnostics: A Review (with discussion)", in Bayesian Statistics 6, Proceedings of the Sixth Valencia International Meeting, Oxford Univ. Press, Edited by Bernardo, Berger, Dawid and Smith.

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

Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. Annals of Statistics, 22, 1701-1762.

Uhlig, H. (1997). "Bayesian Vector Autoregression with Stochastic Volatility". *Econometrica*,

Wild, P. and Gilks, W. (1993), "Adaptive Rejection Sampling from Log-Concave Densities," *Journal of the Royal Statistical Society*, Ser. C, 42, 701-708.

Table 1
Sampling Performance of the Algorithms: Parameters

δ: 0.95	σ_v: 0.26	ν: 10	$Prob(\nu < 20)$	ρ: -0.6	E_h: 9 $\times 10^{-4}$	$\frac{V_h}{E_h^2}$: 1
Setting 1: Basic Model, T=500						
Estimate basic model						
0.92	0.28					
0.046	0.065					
Estimate ν						
0.92	0.24	26	0.31		0.86	0.93
0.075	0.06	19 26 33			0.24	0.74
50 90	49 91	40	10 1		47 88	40 81
Setting 2: Fat tails only, $\nu = 10$, $\rho = 0$, T = 500						
Estimate ν						
0.92	0.25	19	0.58		1.00	1.03
0.09	0.06	12 17 25			0.32	0.67
49 91	49 90	4	63 26		46 87	44 85
Setting 3: Fat tails only, $\nu = 10$, $\rho = 0$, T = 1000						
Estimate ν						
0.94	0.26	18	0.65		0.98	1.02
0.02	0.04	12 16 22			0.21	0.40
48 91	52 90	5	71 42		45 86	43 84
Estimate basic model						
0.93	0.31				1.16	1.19
0.03	0.064				0.34	0.46
31 73	26 65				21 60	45 88
Setting 4: Correlation only, ρ = -0.6, T = 1000						
0.94	0.27			-0.54	0.92	1.00
0.023	0.034			0.09	0.18	0.33
40 81	48 87			41 81	45 87	47 89
Setting 5: Fat tails and correlation, $\nu = 10$, ρ= -0.6, T = 1000						
0.94	0.27	14	0.79	-0.42	0.93	1.02
0.025	0.039	10 14 18		0.19	0.19	0.34
41 83	49 91	0	93 61	2 10	52 87	49 91

The sampling distributions are based on 500 samples simulated for **each** parameter setting numbered 1 to 5. For each sample, the posterior quantities are based on 15000 draws of the samplers, after discarding up to 5000 draws. The header row indicates the parameter values used. For each model, the first row reports the sampling average of the posterior means. The second row is the sampling RMSE of the posterior mean. The two numbers in the third row are the percentage of samples for which a 50% and 90% credibility intervals covered the true value. For ν, the second row is the average of the $25^{th}, 50^{th}$, and 75^{th} percentiles, and the third row is the percentage of the samples for which the 25^{th} percentile was larger than 20. The column after ν reports the average of the posterior probabilities $Prob(\nu < 20)$, and below it, the percentage of samples for which this probability was larger than 0.5, and 0.75.

Table 2
Sampling Performance of the Algorithms: Smoothing

Criterion	Parameter	All	Top 50% λ_t's	Top 10% λ_t's	Top 5% λ_t's
ν estimated					
RMSE	s_t	0.01043	0.01207	0.01978	0.02372
RMSE	$\sqrt{h_t}$	0.00713	0.00723	0.00762	0.00786
MAE %	$\sqrt{h_t}$	21.4	22.1	23.7	24.6
Basic model estimated					
RMSE	s_t	0.01065	0.01196	0.01981	0.02394
RMSE	$\sqrt{h_t}$	0.00823	0.00860	0.00975	0.01037
MAE %	$\sqrt{h_t}$	25.9	27.3	30.6	32.3
	Nobs	500000	250785	50441	25159

The sampling distributions are based on the 500 samples of dataset 3 of table 1. That is, $\delta = 0.95, V_h/E_h^2 = 1, \nu = 10, \rho = 0, T = 1000$. We report the Root Mean Squared Error and the Relative Mean Absolute Error for the posterior means as estimators of $\sqrt{h_t}$ and $s = \sqrt{h_t \times \lambda_t}$. This is done for all the observations, and for those with λ_t larger than the 50^{th}, 10^{th}, and 5^{th} percentile of the prior $p(\lambda \mid \nu = 10$.

Table 3
Posterior Analysis for Weekly Series - Fat Tails Model

	EW	VW	d1	d5	d10
δ	0.960	0.959	0.914	0.935	0.959
basic	0.91	0.95	0.93	0.91	0.93
	(0.016)	(0.014)	(0.025)	(0.016)	(0.014)
	0.931, 0.982	0.935, 0.979	0.872, 0.956	0.906, 0.959	0.89, 0.97
σ_v	0.21	0.20	0.34	0.26	0.20
basic	0.39	0.23	0.32	0.32	0.29
	(0.044)	(0.035)	(0.06)	(0.04)	(0.04)
	0.14, 0.29	0.15, 0.26	0.24, 0.43	0.20, 0.33	0.15, 0.27
$\frac{V_h}{E_h^2}$	0.82	0.71	1.11	0.76	0.72
basic	1.61	0.80	1.10	0.92	0.93
	(0.23)	(0.19)	(0.25)	(0.17)	(0.19)
	0.47, 1.36	0.43, 1.17	0.71, 1.66	0.50, 1.15	0.44, 1.19
$E_h \times 10^3$	0.43	0.42	0.58	0.46	0.44
	(0.07)	(0.06)	(0.07)	(0.05)	(0.06)
	0.32, 0.60	0.32, 0.55	0.46, 0.73	0.37, 0.58	0.34, 0.57
$\nu < 31$	15	19	18	19	21
	(6.1)	(13.7)	(6.4)	(6.0)	(5.4)
	8, 14, 27	10, 19, 29	8, 17, 29	10, 18, 29	12, 21, 29
$\frac{P(\nu<20)}{P(\nu>20)}$	2.1	1.42	2.02	1.67	0.91
$\nu < 61$	15	25	19	25	25
	8, 13, 37	12, 20, 57	9, 18, 29	10, 20, 57	10, 21, 57

The first number is the posterior mean. For δ, σ_v, and V_h/E_h^2, the second row is the posterior mean from the basic model, see JPR (1994). Then the number between parentheses is the posterior standard deviation. The two numbers below are the posterior 5^{th} and 95^{th} quantiles. For ν we also give the posterior median, and the results based upon prior bounds at 30 and 60. ρ is fixed at 0. EW $\equiv$ Equal-weighted NYSE; VW $\equiv$ Value-weighted NYSE; d1 $\equiv$ small, d5 $\equiv$ medium, d10 $\equiv$ large firms; weekly returns, 7/62-12/91. T = 1539. Returns are prefiltered to remove AR(1) and monthly seasonals from the mean equation.

Table 4
Posterior Analysis for Daily Series - Fat Tails Model

	CRSP	SP500	U.K.£	DM	CD$
δ	0.989	0.954	0.98	0.967	0.951
basic		0.97	0.96	0.95	0.95
	(0.003)	(0.013)	(0.006)	(0.009)	(0.010)
	0.984, 0.993	0.931, 0.973	0.970, 0.990	0.949, 0.979	0.934, 0.962
σ_v	0.127	0.218	0.11	0.167	0.255
basic		0.17	0.17	0.13	0.26
	(0.01)	(0.025)	(0.017)	(0.023)	(0.027)
	0.108, 0.145	0.17, 0.26	0.086, 0.141	0.134, 0.21	0.22, 0.30
$\frac{V_h}{E_h^2}$	1.12	0.73	0.43	0.53	1.02
basic		0.54	0.52	0.69	1.10
	(0.27)	(0.15)	(0.12)	(0.11)	(0.18)
	0.74, 1.78	0.50, 1.07	0.27, 0.72	0.35, 0.79	0.73, 1.41
E_h	$\times 10^4$	$\times 1$	$\times 10^4$	$\times 10^4$	$\times 10^5$
	0.61	1.01	0.38	0.42	0.50
	(0.09)	(0.12)	(0.05)	(0.04)	(0.05)
	0.47, 0.82	0.81, 1.26	0.30, 0.49	0.34, 0.51	0.42, 0.61
$\nu < 31$	19	13	9	11	26
	(4.2)	(4.7)	(1.8)	(3.2)	(3.7)
	13, 18, 27	8, 12, 21	7, 9, 14	8, 11, 17	18, 26, 30
$\frac{P(\nu<20)}{P(\nu>20)}$	1.45	17.5	331	51.6	0.13
$\nu < 61$	21	11	10	11	47
	13, 20, 35	7, 10, 17	7, 10, 15	8, 12, 19	28, 49, 60
Nobs	6409	2023	2613	2613	3010

The results shown are as in table 3. CRSP: daily VW returns from 1962 to 1987. S&P500-daily change in log of the index, filtered to remove calendar effects, see Gallant, Rossi, and Tauchen (1992); 1/2/80-12/30/87. 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. CD$ daily noon interbank market spot rates from Bank of Canada supplied by Melino and Turnbull (1990); 1/2/75-12/10/86.

Table 5
Posterior Analysis for Weekly Series - Full Model

	EW	VW	d1	d5	d10
δ	0.965	0.962	0.90	0.931	0.950
	(0.009)	(0.01)	(0.027)	(0.017)	(0.014)
	0.949, 0.979	0.945, 0.977	0.851, 0.940	0.901, 0.956	0.925, 0.972
σ_v	0.21	0.208	0.39	0.28	0.23
	(0.02)	(0.019)	(0.06)	(0.03)	(0.03)
	0.176, 0.245	0.177, 0.240	0.29, 0.48	0.23, 0.34	0.18, 0.29
$\frac{V_h}{E_h^2}$	0.95	0.84	1.22	0.85	0.79
	(0.26)	(0.22)	(0.25)	(0.18)	(0.19)
	0.58, 1.62	0.53, 1.38	0.81, 1.77	0.58, 1.26	0.50, 1.25
$E_h \times 10^3$	0.49	0.48	0.59	0.46	0.44
	(0.08)	(0.07)	(0.07)	(0.05)	(0.06)
	0.36, 0.67	0.32, 0.55	0.47, 0.74	0.37, 0.58	0.34, 0.56
$\nu < 31$	16	23	20	18	21
	(4.9)	(4.8)	(6.3)	(6.0)	(5)
	9, 15, 25	14, 23, 30	9, 20, 29	9, 17, 29	12, 21, 29
$\frac{P(\nu<20)}{P(\nu>20)}$	4.5	0.5	1	1.85	0.8
ρ	-0.41	-0.44	-0.16	-0.44	-0.40
	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)
	-0.52, -0.30	-0.54, -0.33	-0.26, -0.06	-0.54, -0.34	-0.52, -0.26
$P(\rho < 0)$	1	1	0.994	1	1

The first number is the posterior mean. The number between parentheses is the posterior standard deviation. The two numbers between brackets are the posterior 5^{th} and 95^{th} quantiles. For ν we also give the posterior median. The series analyzed are as in Table 3.

Table 6
Posterior Analysis for Daily Series - Full Model

	CRSP	CRSP 80-87	SP500	U.K.£	DM	CD$
δ	0.989	0.978	0.971	0.970	0.962	0.957
	(0.002)	(0.01)	(0.008)	(0.008)	(0.009)	(0.009)
	0.984, 0.992	0.959, 0.992	0.957, 0.982	0.957, 0.981	0.948, 0.974	0.941, 0.972
σ_v	0.131	0.124	0.139	0.14	0.174	0.243
	(0.01)	(0.028)	(0.020)	(0.021)	(0.021)	(0.025)
	0.117, 0.147	0.09, 0.18	0.11, 0.174	0.111, 0.175	0.14, 0.212	0.21, 0.29
$\frac{V_h}{E_h^2}$	1.20	0.49	0.42	0.43	0.51	1.05
	(0.28)	(0.15)	(0.10)	(0.12)	(0.10)	(0.20)
	0.81, 1.89	0.29, 0.88	0.27, 0.65	0.27, 0.72	0.35, 0.74	0.75, 1.50
E_h	$\times 10^4$	$\times 10^4$	$\times 1$	$\times 10^4$	$\times 10^4$	$\times 10^5$
	0.62	0.70	0.97	0.39	0.41	0.51
	(0.094)	(0.11)	(0.11)	(0.05)	(0.04)	(0.05)
	0.46, 0.84	0.53, 0.95	0.78, 1.20	0.32, 0.47	0.34, 0.50	0.42, 0.63
$\nu < 31$	20	10	16	11	12	24
	(4.2)	(2.6)	(5.2)	(2.8)	(3.7)	(2.4)
	14, 20, 28	7, 10, 15	9, 15, 26	7, 10, 16	8, 11, 19	18, 26, 30
$\frac{P(\nu<20)}{P(\nu>20)}$	1.4	25	4.1	149.9	26	0.15
ρ	-0.48	-0.23	-0.18	-0.13	-0.018	-0.29
	(0.04)	(0.08)	(0.09)	(0.08)	(0.07)	(0.06)
	-0.54, -0.42	-0.37, -0.08	-0.33, -0.02	-0.26, -0.00	-0.14, 0.097	-0.38, -0.20
$P(\rho < 0)$	1	0.99	0.96	0.953	0.59	1
Nobs	6409	2023	2023	2613	2613	3010

The numbers are as in table 5. The series analyzed are as in table 4. The second series is the CRSP index returns from 1980 to 1987.

Table 7
Sensitivity of posteriors to the treatment of ν

	ν	δ	$\frac{V_h}{E_h^2}$	E_h	ρ
EW	∞	.910 .937 .958	.72 1.03 1.6	.40 .495 .64	-.55 -.45 -.35
	15	.917 .944 .965	.65 0.95 1.5	.35 .434 .57	-.55 -.46 -.36
	free (15)	.949 .965 .979	.58 0.89 1.6	.36 .476 .67	-.52 -.42 -.30
VW	∞	.916 .944 .965	.61 0.88 1.4	.38 .470 .61	-.58 -.47 -.36
	free (23)	.945 .962 .977	.53 0.79 1.4	.32 .470 .55	-.54 -.44 -.33
	15	.922 .948 .967	.53 0.80 1.3	.33 .409 .53	-.56 -.47 -.35
D1	∞	.855 .893 .925	.92 1.27 1.8	.55 .655 .81	-.23 -.14 -.04
	free (20)	.851 .898 .940	.81 1.19 1.8	.47 .589 .74	-.26 -.16 -.06
	15	.874 .912 .944	.76 1.09 1.6	.47 .570 .72	-.26 -.16 -.05
D10	∞	.918 .944 .964	.57 0.84 1.3	.39 .478 .62	-.51 -.39 -.27
	free (21)	.925 .951 .972	.50 0.76 1.3	.34 .430 .56	-.52 -.40 -.26
	15	.928 .954 .974	.49 0.74 1.2	.34 .418 .55	-.52 -.41 -.29
SP500	∞	.947 .968 .982	.35 0.52 .84	.92 1.12 1.4	-.35 -.22 -.07
	free (16)	.957 .972 .982	.27 0.41 .65	.78 .959 1.2	-.33 -.18 -.02
	15	.968 .981 .991	.29 0.46 .88	.77 .977 1.3	-.35 -.21 -.05
CRSP	∞	.984 .988 .992	.85 1.19 1.9	.54 .680 .93	-.56 -.51 -.45
	free (20)	.984 .989 .992	.81 1.14 1.9	.46 .611 .84	-.54 -.48 -.42
	15	.984 .989 .992	.80 1.13 1.9	.46 .592 .81	-.53 -.47 -.41
UK £	∞	.932 .960 .978	.39 0.56 .82	.41 .485 .59	-.25 -.13 -.01
	15	.965 .979 .990	.28 0.44 .76	.34 .415 .53	-.29 -.16 -.03
	free (10)	.957 .971 .981	.27 0.42 .72	.32 .385 .47	-.26 -.13 -.00
DM	∞	.932 .953 .968	.45 0.61 .88	.43 .503 .60	-.11 -.04 +.10
	15	.944 .962 .976	.39 0.55 .82	.37 .437 .53	-.13 -.02 +.09
	free (11)	.948 .962 .974	.35 0.50 .74	.34 .408 .50	-.14 -.02 +.10

Three models are estimated. The first model fixes all λ_t's to 1. ν is mentioned as ∞. The second fixes $\nu = 15$ but estimates the λ_t's. The third model is the full model. It is the row where ν is mentioned as *free*. The number in parentheses is the posterior mean of ν. For each series, we show first the model with $\nu = \infty$. Then we show the full model if the posterior mean of ν is larger than 15. Otherwise we show the model with $\nu = 15$ fixed. For the other parameters, the three numbers are the 5^{th}, 50^{th}, and 95^{th} percentiles.

Figure 1: Fat Tail and Basic SVOL: Smoothing and Prediction

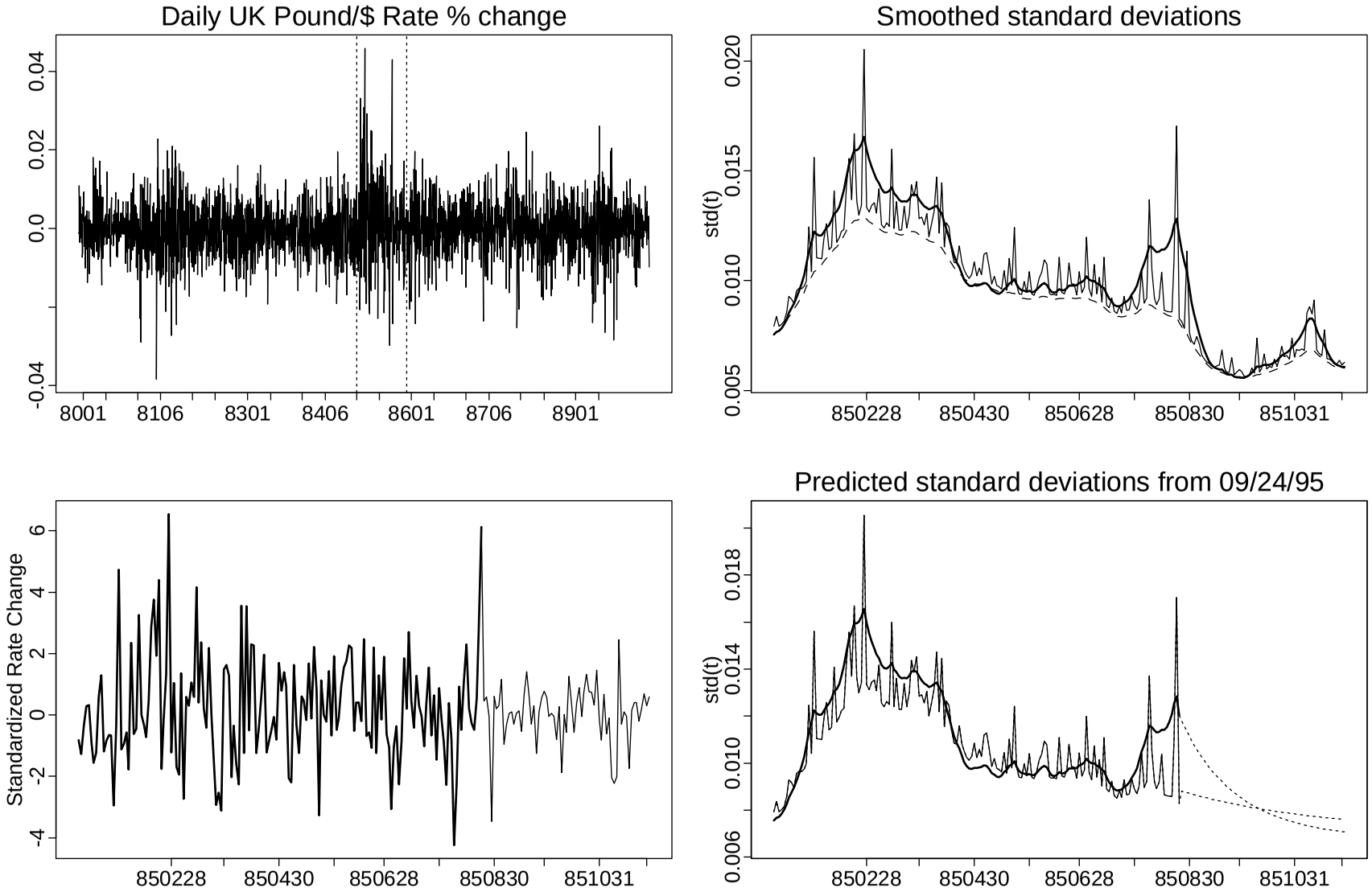


Figure 2: Fat Tail - nu=10 vs Basic SVOL Model

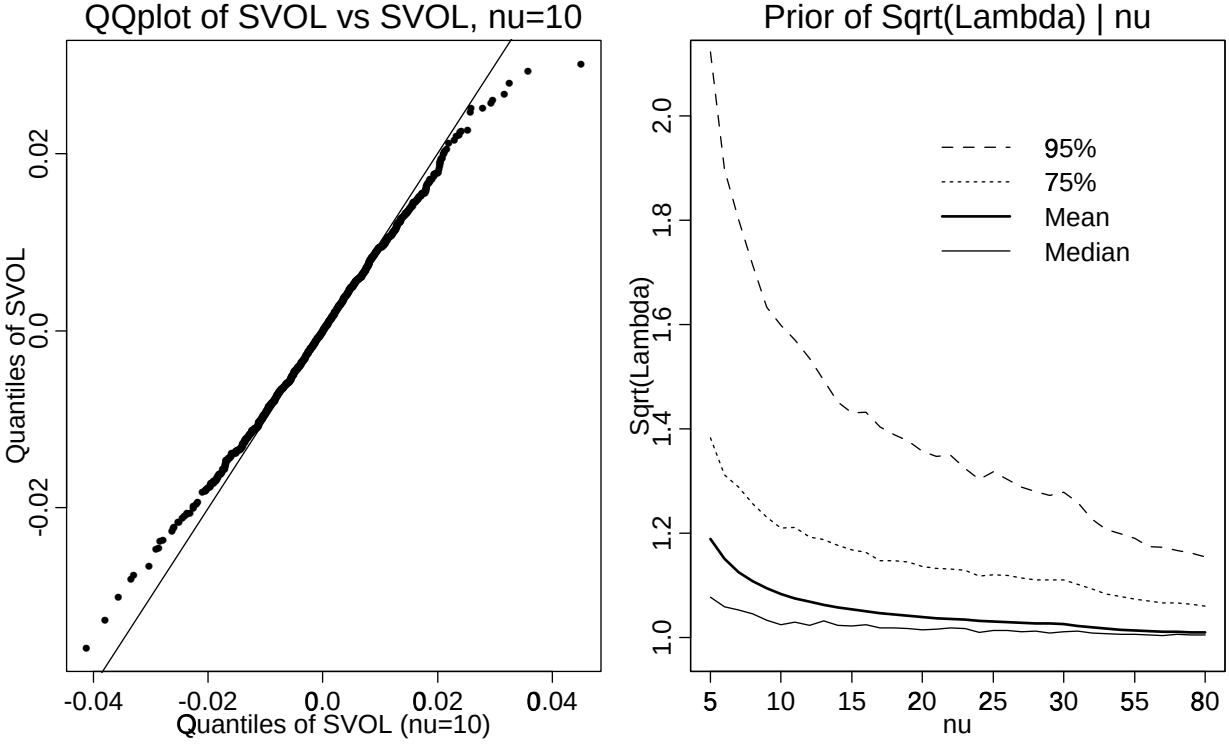
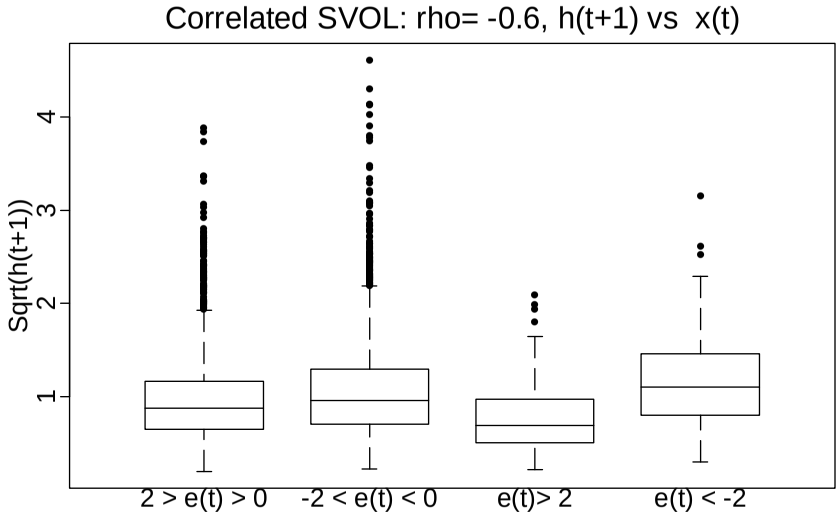
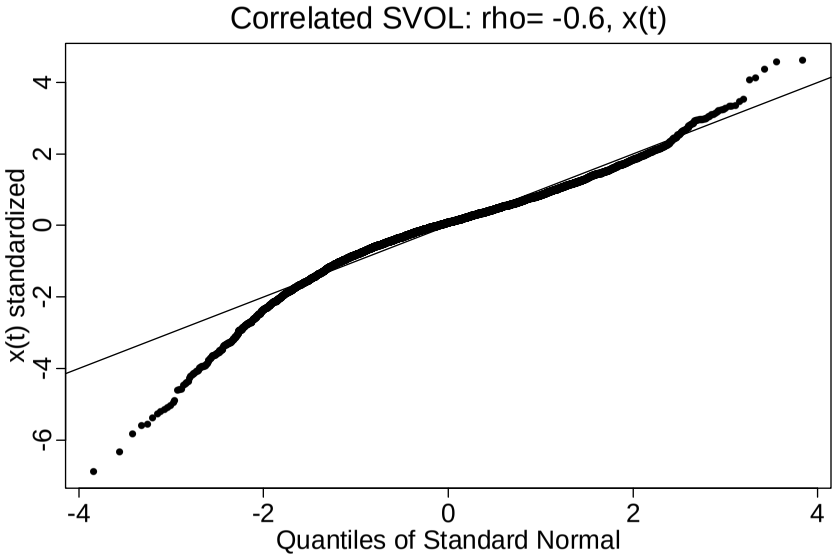
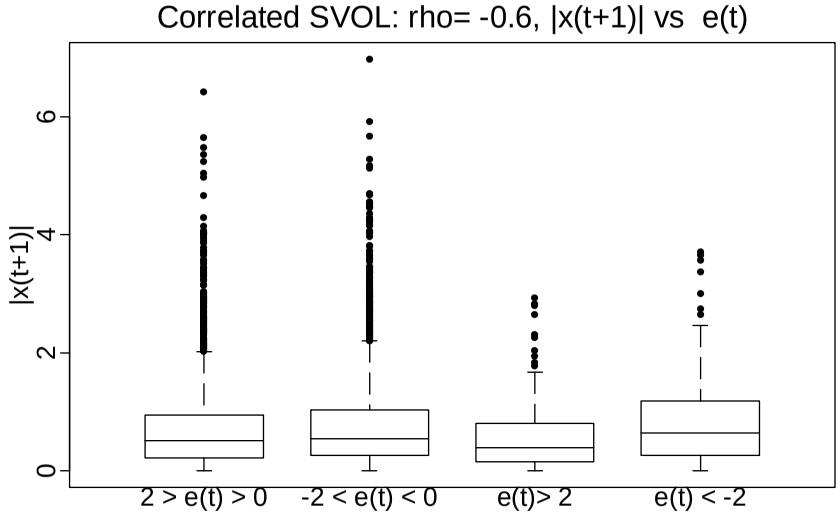


Figure 3: Correlation and Asymmetry, SVOL and the Daily CRSP return







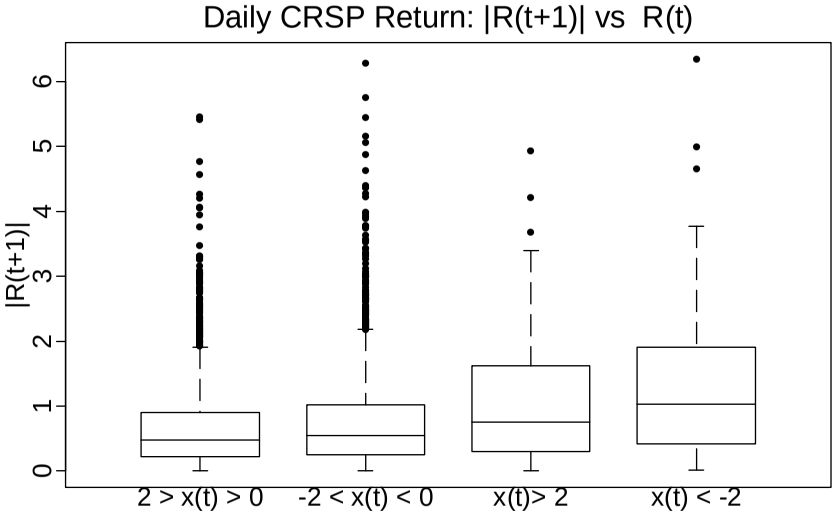


Figure 4: Correlated SVOL, Correction for Additional Term in p(h)

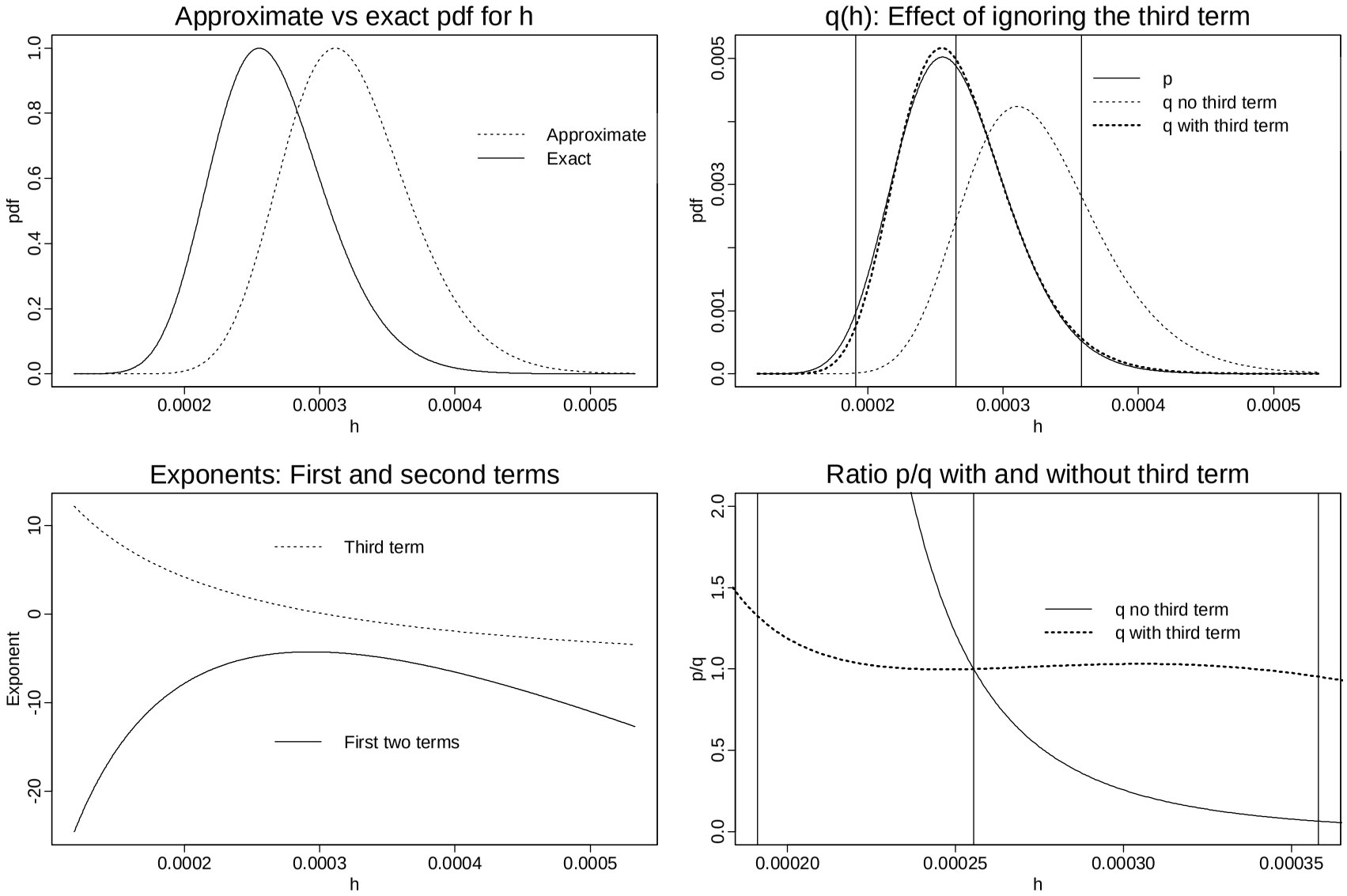
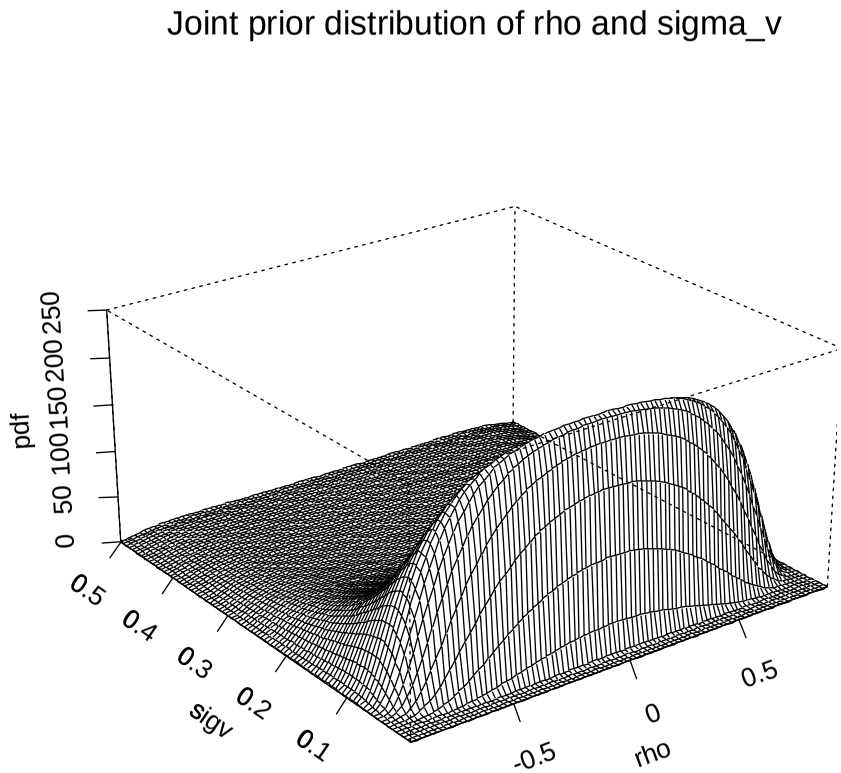
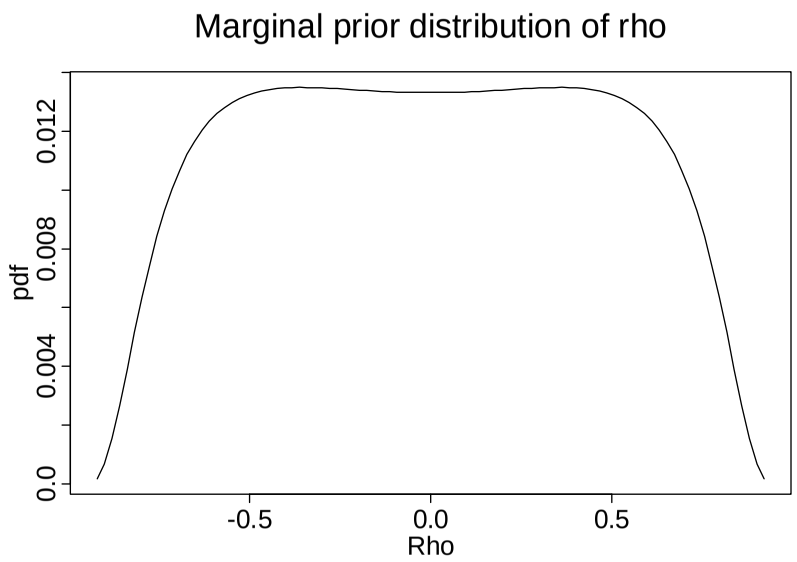


Figure 5: Priors for the Correlated SVOL Model





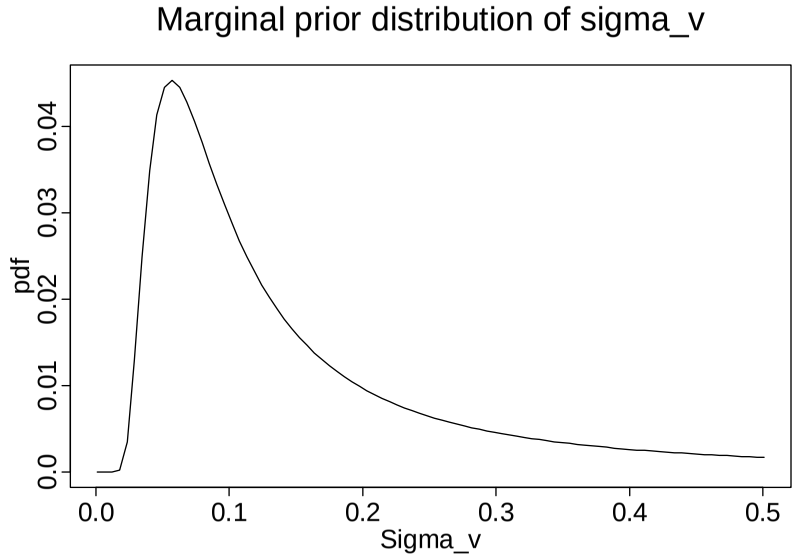
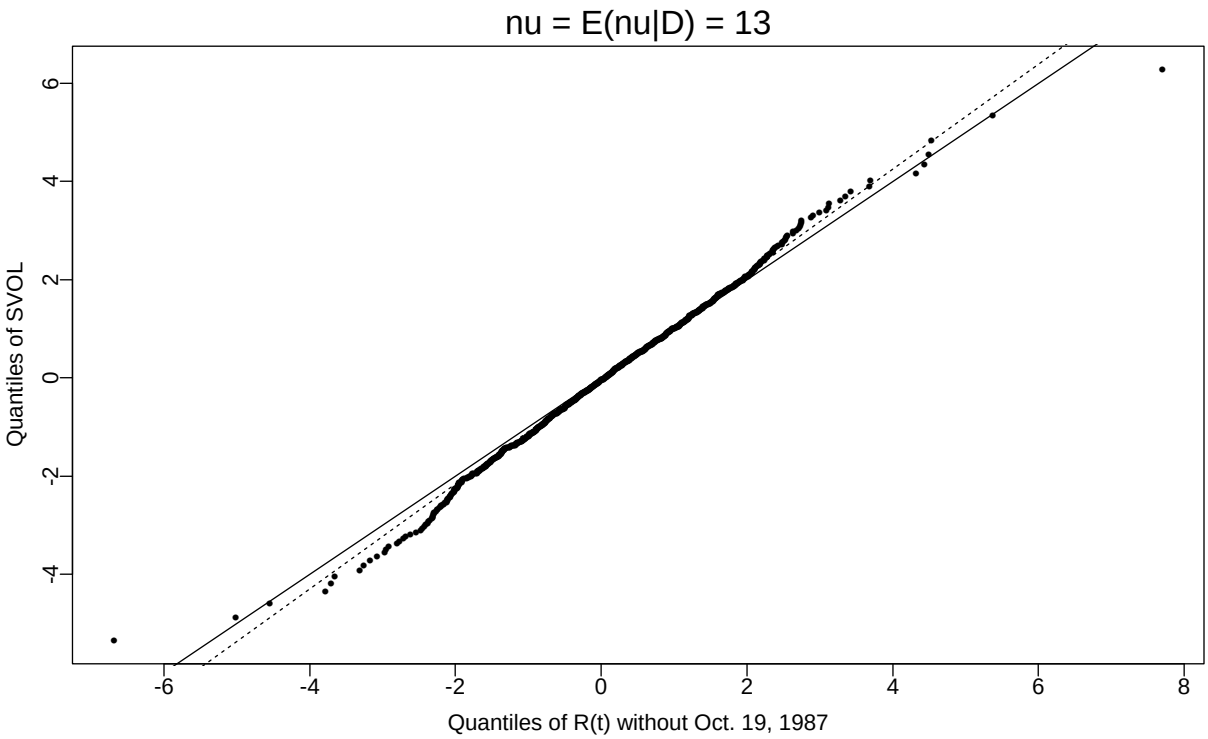


Figure 6: SP500 Daily Returns vs Marginal SVOL Distributions



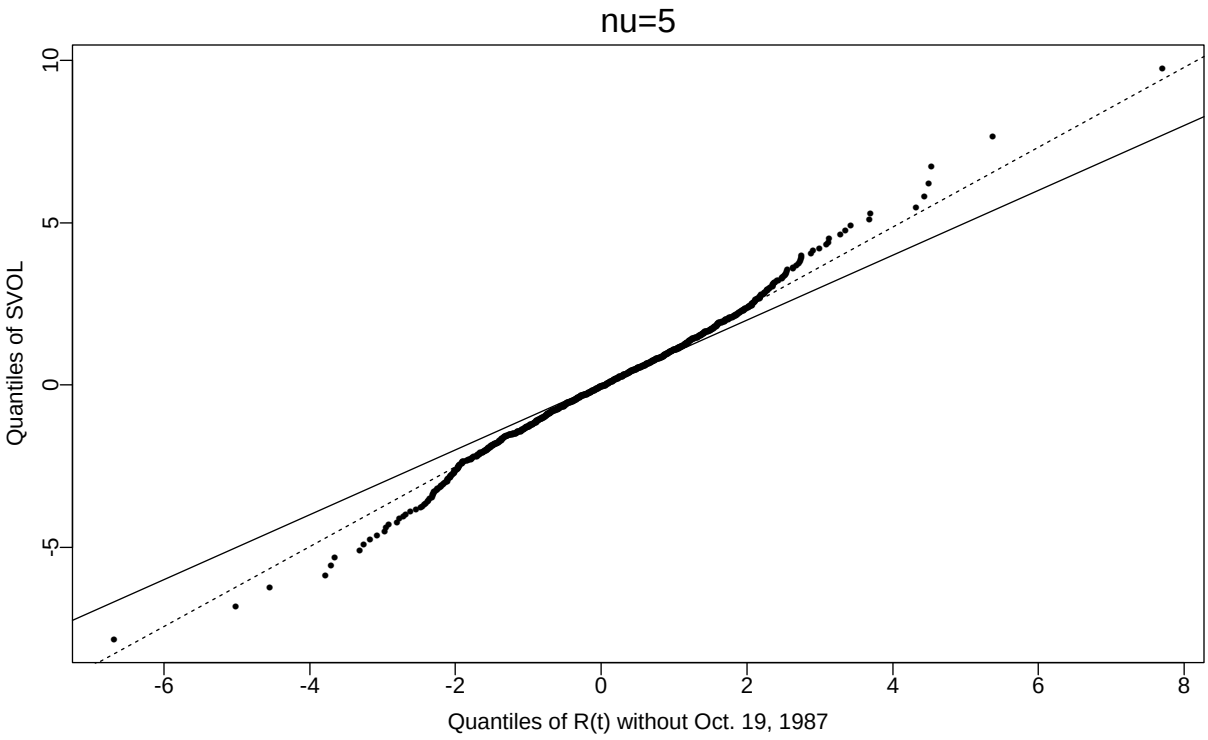
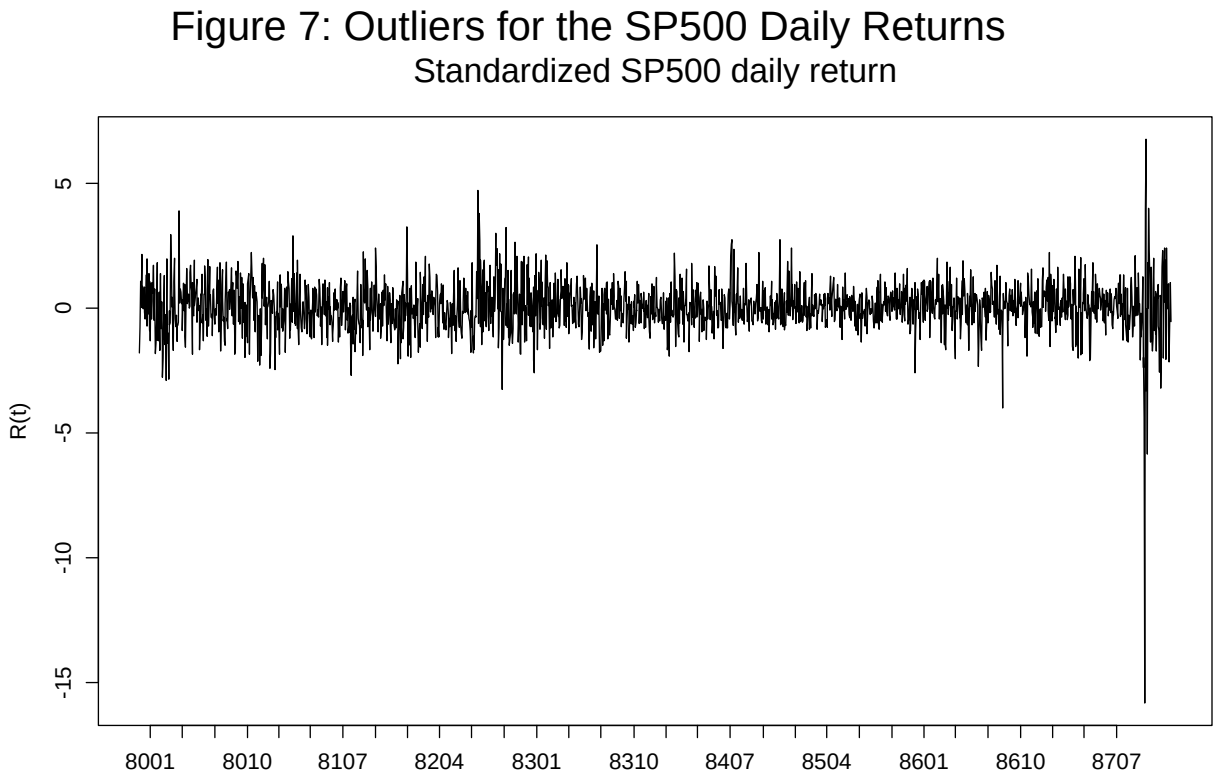


Figure 7: Outliers for the SP500 Daily Returns



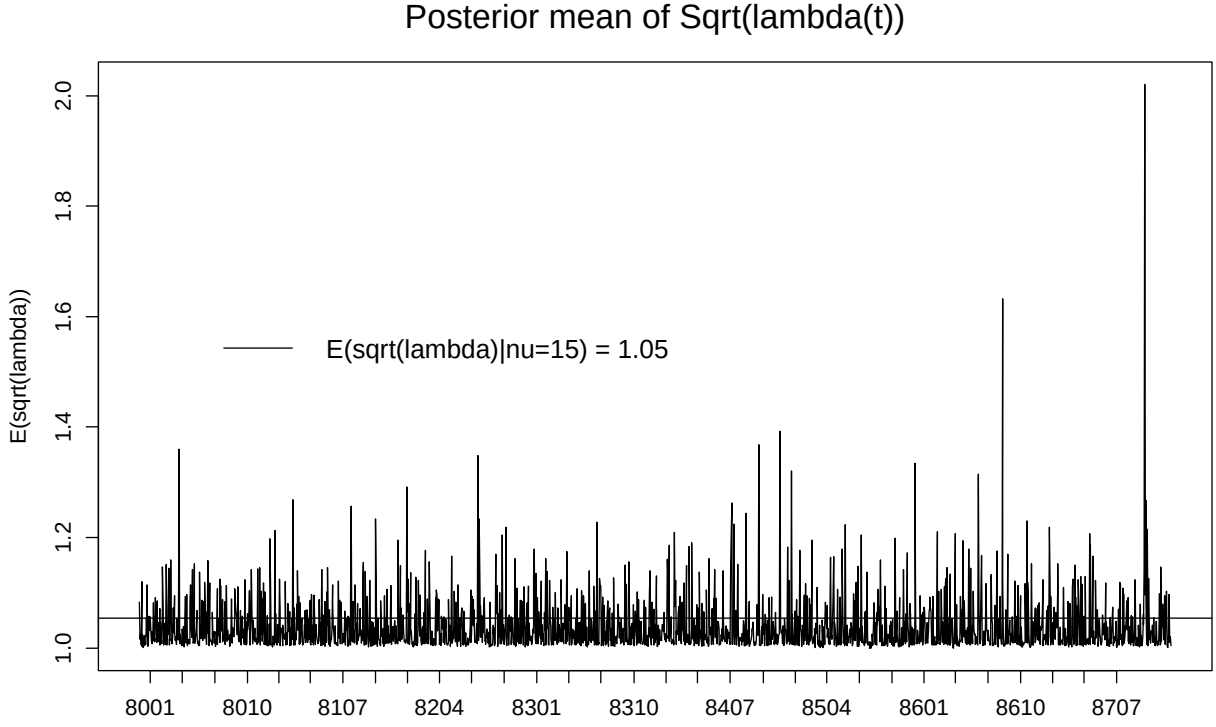


Figure 8a: Stock Market Indexes, Posterior Distributions of Nu and Rho

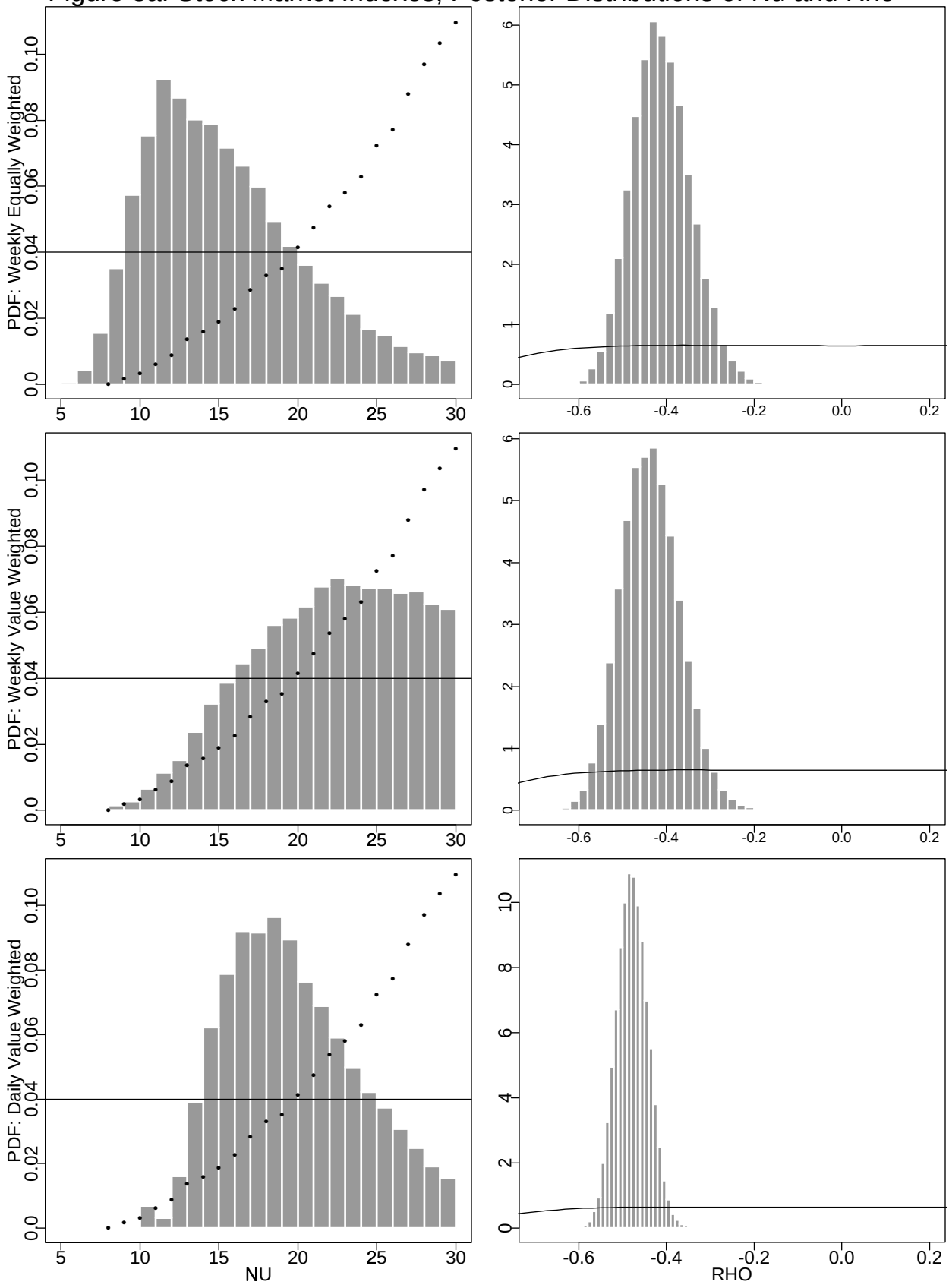


Figure 8b: Stock Markets 1980-1987, Posterior Distributions of Nu and Rho

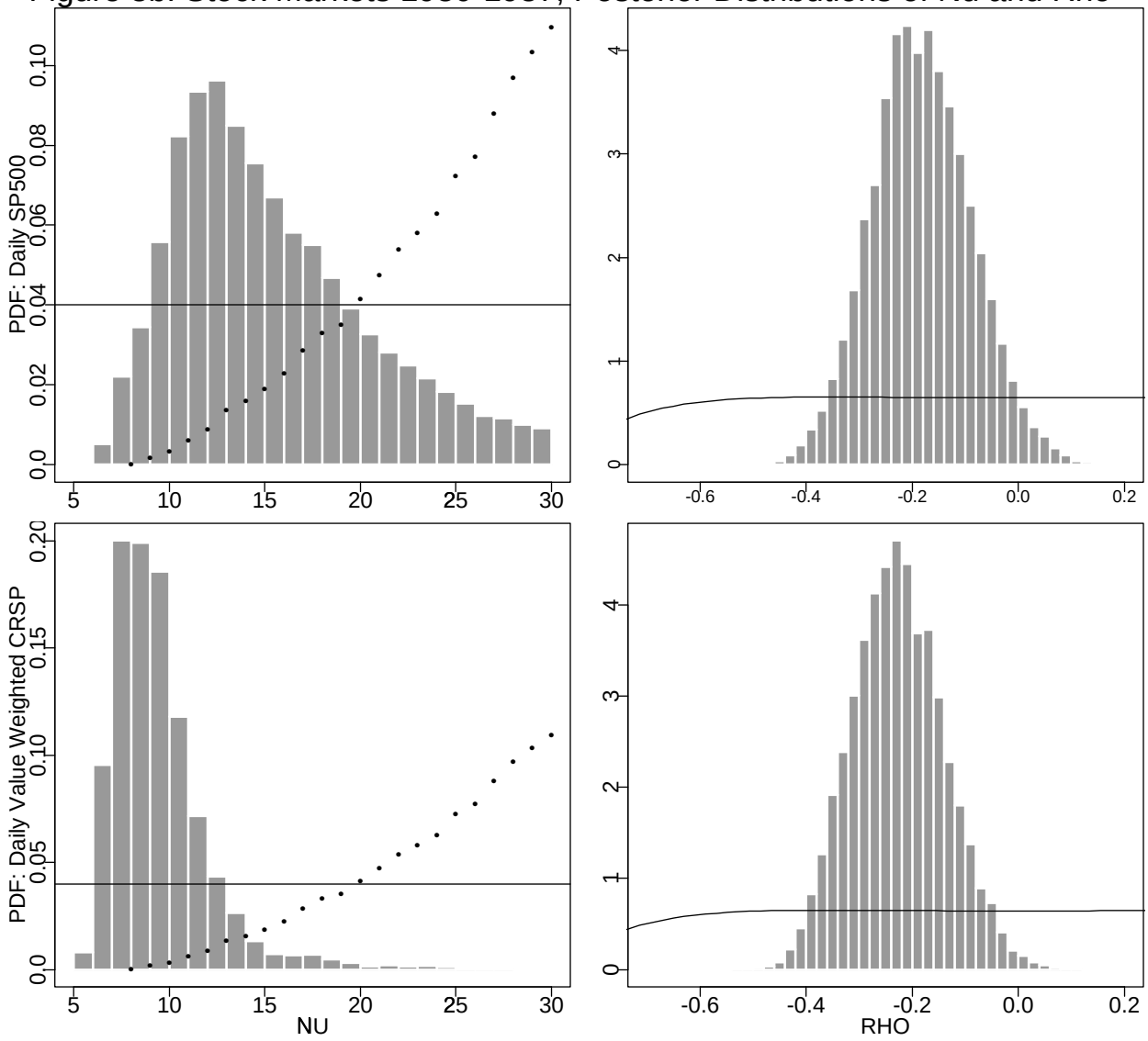
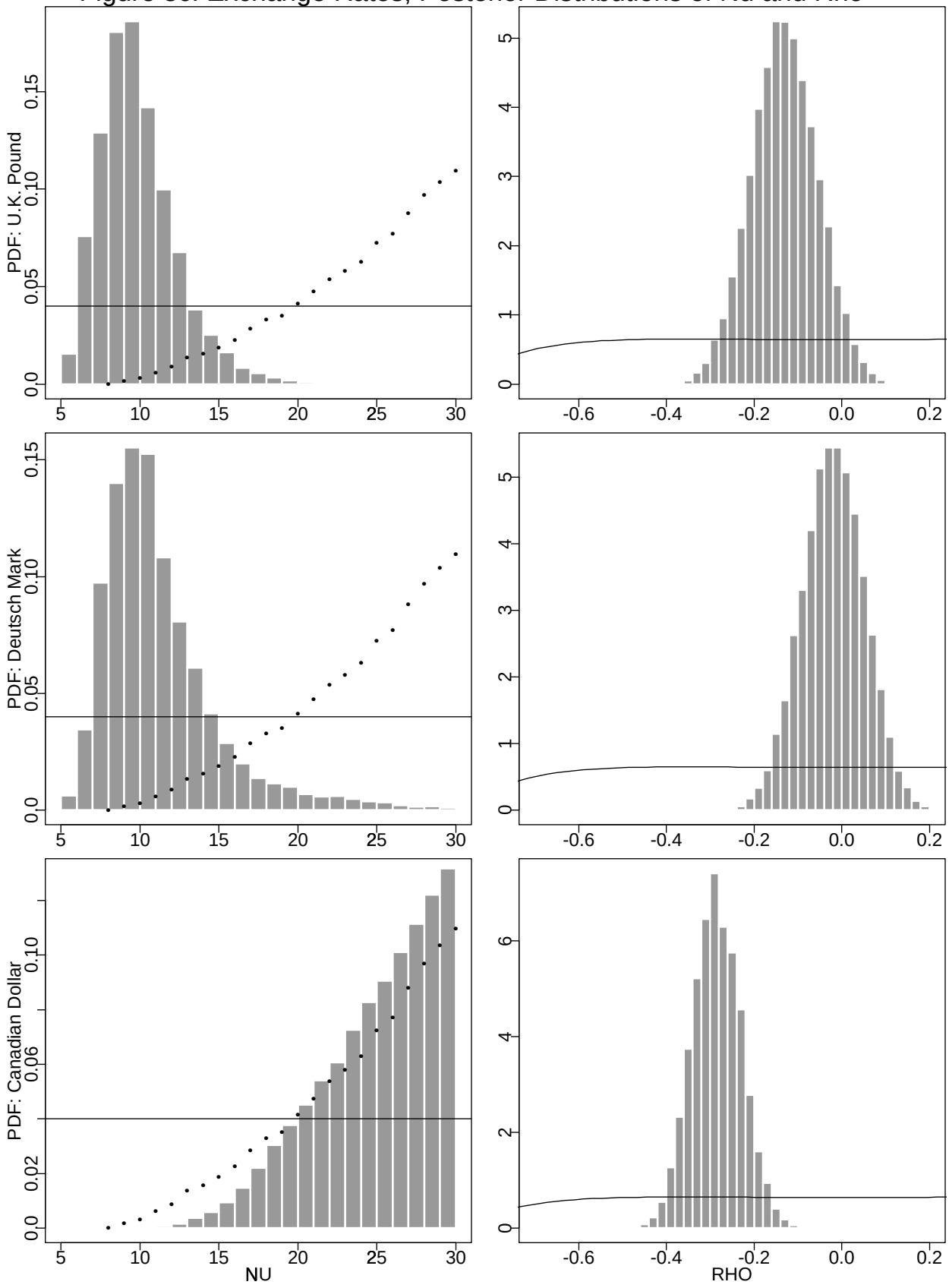


Figure 8c: Exchange Rates, Posterior Distributions of Nu and Rho



Liste des publications au CIRANO *

Cahiers CIRANO / *CIRANO Papers* (ISSN 1198-8169)

- 99c-1 Les Expos, l'OSM, les universités, les hôpitaux : Le coût d'un déficit de 400 000 emplois au Québec — Expos, Montréal Symphony Orchestra, Universities, Hospitals: The Cost of a 400,000-Job Shortfall in Québec / Marcel Boyer
- 96c-1 Peut-on créer des emplois en réglementant le temps de travail ? / Robert Lacroix
- 95c-2 Anomalies de marché et sélection des titres au Canada / Richard Guay, Jean-François L'Her et Jean-Marc Suret
- 95c-1 La réglementation incitative / Marcel Boyer
- 94c-3 L'importance relative des gouvernements : causes, conséquences et organisations alternative / Claude Montmarquette
- 94c-2 Commercial Bankruptcy and Financial Reorganization in Canada / Jocelyn Martel
- 94c-1 Faire ou faire faire : La perspective de l'économie des organisations / Michel Patry

Série Scientifique / *Scientific Series* (ISSN 1198-8177)

- 99s-25 Inference for the Generalization Error / Claude Nadeau et Yoshua Bengio
- 99s-24 Mobility and Cooperation: On the Run / Karl-Martin Ehrhart et Claudia Keser
- 99s-23 Input Price Discrimination, Access Pricing, and Bypass / Ngo Van Long et Antoine Soubeyran
- 99s-22 Existence and Uniqueness of Cournot Equilibrium: A Contraction Mapping Approach / Ngo Van Long et Antoine Soubeyran
- 99s-21 Sources of Productivity Growth: Technology, Terms of Trade, and Preference Shifts / Thijs ten Raa et Pierre Mohnen
- 99s-20 Remarks on Environmental Regulation, Firm Behavior and Innovation / Bernard Sinclair-Desgagné
- 99s-19 Subvention gouvernementale et partage du travail : Une analyse économique – II / Paul Lanoie et François Raymond
- 99s-18 Subvention gouvernementale et partage du travail : Une analyse économique – I / Paul Lanoie et Ali Béjaoui
- 99s-17 Content Horizons for Forecasts of Economic Time Series / John W. Galbraith
- 99s-16 Modelling the Role of Organizational Justice: Effects on Satisfaction and Unionization Propensity of Canadian Managers / Michel Tremblay et Patrice Roussel
- 99s-15 Pricing Discretely Monitored Barrier Options by a Markov Chain / Jin-Chuan Duan, Evan Dudley, Geneviève Gauthier et Jean-Guy Simonato
- 99s-14 Shame and Guilt in Lancashire: Enforcing Piece-Rate Contracts / Michael Huberman

---

* Vous pouvez consulter la liste complète des publications du CIRANO et les publications elles-mêmes sur notre site World Wide Web à l'adresse suivante :
http://www.cirano.umontreal.ca/publication/documents.html