Estimating a Semiparametric Asymmetric Stochastic Volatility Model with a Dirichlet Process Mixture*

Mark J. Jensen
Federal Reserve Bank of Atlanta
Mark.Jensen@atl.frb.org

John M. Maheu
University of Toronto
& RCEA
jmaheu@chass.utoronto.ca

Abstract. In this paper we extend the parametric, asymmetric, stochastic volatility model (ASV), where returns are correlated with volatility, by flexibly modeling the bivariate distribution of the return and volatility innovations nonparametrically. Its novelty is in modeling the joint, conditional, return-volatility, distribution with a infinite mixture of bivariate, mean zero, Normal distributions having unknown mixture weights and covariance matrices. This semiparametric ASV model nests stochastic volatility models whose innovations are distributed as either Normal or Student-t distributions, plus the response in volatility to unexpected return shocks is more general than the fixed asymmetric response of the ASV model. The unknown mixture probabilities and parameters are modeled with a Dirichlet Process prior. This prior ensures a parsimonious, finite, posterior, mixture that bests represents the distribution of the innovations and a straightforward sampler of the conditional posteriors. We develop a Bayesian Markov chain Monte Carlo sampler to pfully characterize the parametric and distributional uncertainty. Nested model comparisons and out-of-sample predictions with the cumulative marginal-likelihoods, and one-day-ahead, predictive log-Bayes factors between semiparametric and parametric versions of the ASV model, shows our semiparametric model forecasting market returns more accurately. A major reason is volatility's expected response to unexpected market movements. When the market is tranquil, expected volatility's reaction to a negative (positive) price shock is to rise (decline, then rise as the size of the positive shock gets larger). However, during a volatile market the asymmetry and size of the change in expected volatility becomes muted. In other words, when times are good, no news is good news, but when times are bad, neither good nor bad news matters with regards to volatility.

Keywords: Bayesian nonparametrics, cumulative Bayes factor, Dirichlet process mixture, infinite mixture model, leverage effect, marginal likelihood, MCMC, non-normal, stochastic volatility, volatility-return relationship
JEL Classification: C11, C14, C53, C58

*We would like to thank the seminar participants at the 2010 Rimini Conference in Economics and Finance, the 10th Annual All-Georgia Financial Conference, the 2011 Seminar on Bayesian Inference in Econometrics and Statistics at Washington University in Saint Louis, the Department of Economics at Istanbul Bilgi University, the Econometrics and Statistics Colloquium at the Booth School Business at the University of Chicago, and Sim Kee Boon Institute for Financial Economics at Singapore Management University. Special thanks go to Jun Yu for providing us with the WinBugs code to his semiparametric ASV model, and to Mark Fisher, Hedibert Lopes, and Chayawat Ornthanalai for comments and suggestions. The views expressed here are ours and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

1 Introduction

In this paper we extend the parametric, asymmetric, stochastic volatility model (ASV), where returns are correlated with volatility, by flexibly modeling the bivariate distribution of the return and volatility innovations nonparametrically.[1] Log-volatility of the ASV belongs to the parametric, first-order autoregressive, class of stochastic volatility, which can accomodate stationarity as imposed in the literature (Jacquier et al. (1994) and Kim et al. (1998)) as well as nonstationary deviations from this assumption. The rest of the model is nonparametric in the sense that no assumptions are made about the underlying joint distribution of returns and volatility. Instead, the flexible Dirichlet process mixture (DPM) class of priors by Lo (1984) along with return data is used to estimate the joint distribution.

The version of the DPM used is an infinite mixture of bivariate, Normal distributions with mean zero, but unknown covariance matrices and mixture probabilities. This is used to fit the return and log-volatility distribution. As a mixture, each observations covariance matrix is distributed according to a Dirichlet process prior – a nonparametric prior over the value of the covariance matrix and the probability of its occurrence.

Others have nonparametrically modeled the return distribution of a stochastic volatility model, but the joint return, log-volatility distribution has not been. Jensen (2004), Griffin and Steel (2006), Jensen and Maheu (2010), Griffin and Steel (2011), and Delatola and Griffin (2011a), each apply a Dirichlet process mixture type prior to the return distribution. For the asymmetric stochastic volatility model, Delatola and Griffin (2011b) include a constant leverage effect and model the distribution of the log-squared return innovation nonparametrically with the infinite mixture, Constant Component Variance model of Griffin (2011). Yu (2011) directly models the leverage effect nonparametrically by modeling the correlation between returns and volatility with a fixed ordered linear spline, and Durham (2007) models the return distribution with a finite mixture model where its order is fixed a priori by the econometrician.

By relaxing the parametric distribution of the asymmetric stochastic volatility model, the approach taken in this paper is more flexible and better positioned to model non-Gaussian behavior. As pointed out by Das and Sundaram (1999), the parametric stochastic volatility model can only produce the level of skewness and kurtosis observed in the data when it takes on implausible parameter values. By design a nonparametric joint distribution allows the stochastic volatility model to capture these types of characteristics in the data without sacrificing the time dependent nature of the stochastic volatility model.

[1] See Harvey et al. (1994), Yu (2005) and Omori et al. (2007) or Eq. (20)-(21) for the parametric version of the asymmetric stochastic volatility model.

A Markov chain, Monte Carlo posterior algorithm for sampling the nonparametric and parametric portions of the model is presented. Our semiparametric sampler extends the univariate algorithm of the semiparametric stochastic volatility model by Jensen and Maheu (2010). A restricted version of the algorithm is also applied to a fully parametric, asymmetric, stochastic volatility model. Parameter, volatility, and distributional uncertainty are integrated out with the sampler of the posterior. These draws will be used to generate both the one-day-ahead predictive joint density for daily market returns and log-volatility and the marginal density for one-day-ahead returns.

Stochastic volatility models and econometric models in general are chosen based on their predictability (see Geweke (2001), and Geweke and Whiteman (2006)). This is understandable given the important role predictions play in valuing stocks and options, constructing portfolios, and creating hedging strategies. Strong evidence in favor of our semiparametric model relative to parametric models is provided by the sequential predictive likelihoods for returns. In particular, the DPM estimate of the unknown predictive distribution for returns is found to be robust over low and high volatility periods and to large return shocks.

Great emphasis is also placed on the ability of a model to forecast volatility.[2] Asymmetry, where an unexpected decline in price leads to higher volatility, whereas an increase in price causes volatility to decline, is common in volatility models (see Bekaert and Wu (2000), Chen and Ghysels (2011)). In the stochastic volatility model this asymmetry comes from returns and log-volatility being negatively correlated. Since the covariance matrix of the nonparametric return and log-volatility innovations distribution follows a random second-order effect, this paper's semiparametric volatility model exhibits asymmetry. However, the random second-order effect for the covariance also allows the correlation to change depending on market conditions. During a regular market the asymmetry is like that found by Chen and Ghysels (2011), moderate increases in stock prices reduce expected volatility, whereas any decline in stock prices or a large unexpected increase in prices leads to higher expected volatility. However, during volatile times, expected volatility barely increases following an unexpected shock to market returns. In other words, the asymmetric return and volatility relationship is not a high volatility phenomenon.

The paper is organized as follows. In the Section 2, the asymmetric, stochastic volatility model with a nonparametric Dirichlet process mixture prior for the unknown distribution is constructed. Section 3 spells out the nonparametric model's Markov chain, Monte Carlo sampler of all the unknown parameters and latent variables. In Section 5 we apply our semiparametric and an existing parametric asymmetric, stochastic volatility models to 28

[2]See Poon and Granger (2003) for an extensive overview of the volatility forecasting literature

years worth of daily market returns as measured by the value-weighted market portfolio from the Center of Research in Security Prices. In Section 6 and 7, we compare our empirical results by first using the Savage-Dickey density ratio to evaluate the Bayes factor in favor of a nested parametric versions of the nonparametric model. We then compare the forecasting performance of the models over the last two years of the data with their cumulative log predictive Bayes factors. In Section 8 we estimate volatility's response to unexpected changes in stock prices. A summary and conclusions are contained in Section 9.

2 Model

We model asset returns using the following semiparametric, asymmetric, stochastic, log-volatility model

$$y_t = \mu + \exp\{h_t/2\}\epsilon_t \tag{1}$$

$$h_{t+1} = \varphi h_t + \eta_t \tag{2}$$

where y_t is the continuously compounded daily return at time periods $t = 1, \ldots, n$, and h_{t+1} is the value of the latent, log-volatility, one-day-ahead. The absolute value of the autoregressive parameter, φ, is constrained to the unit interval.[3] Hence, h_{t+1} will be covariance stationary.

We relax all assumptions concerning the joint distribution of ϵ_t and η_t, and, instead, allow their distribution to be completely unknown and random as if the distribution were an additional unknown to the parameters and latent volatilities of the ASV model.

Being unknown and random, the joint distribution requires a prior, which can then be used to obtain the random distributions posterior once data has been collected. We choose the following Lo (1984) type Dirichlet process mixture prior (DPM)

$$\left(\begin{array}{c} \epsilon_t \\ \eta_t \end{array} \right) \Bigg| \Lambda_t \sim N(\mathbf{0}, \Lambda_t), \tag{3}$$

$$\Lambda_t \sim G, \tag{4}$$

$$G|G_0, \alpha \sim DP(\alpha, G_0), \tag{5}$$

to model the unknown distribution. In Eq. (3)-(5), the tth innovations are distributed as a bivariate Normal with a mean zero vector but with a random covariance matrix, $\Lambda_t = \left(\begin{array}{cc} \sigma^2_{y,t} & \sigma_{yh,t} \\ \sigma_{yh,t} & \sigma^2_{h,t} \end{array} \right)$, distributed as G. The distribution G is unknown with the Dirichlet

[3] Because the mean of h_{t+1} can be subsumed into the variance of ϵ_t, identification requires the mean of log-volatility to be zero; i.e., the intercept term of h_{t+1} must be set equal to zero.

process distribution prior, $DP(\alpha, G_0)$, of Ferguson (1973). The nonzero scalar α is the precision parameter and G_0 is the base distribution.

The DPM builds on the well known property that a flexible distribution can be found by mixing together a finite number of known distributions. It extends this concept by mixing together an infinite number of distributions. In its simplest and most basic form the Dirichlet process mixture models the innovation vector $(\epsilon_t, \eta_t)'$ as independent realizations from the same, unknown, distribution which we model as a mixture of distributions

$$\int F_N(0, \Lambda) G(d\Lambda), \tag{6}$$

where F_N is a Normal distribution function with mean zero and covariance matrix Λ, and G is a weighted mixture of the Λs.

Eq. (1)-(5) constitute the semiparametric, asymmetric, stochastic volatility model with DPM prior model (ASV-DPM). At first glance, the ASV-DPM model, with its mean zero, bivariate, Normal distribution function, might seem to lack the capacity to fit the non-Gaussian behavior of returns and log-volatility. This, however, is incorrect. Fixing the mean of F to zero only limits the DPM prior to the class of distributions having one mode. This is hardly a limitation since asset returns are not known to have distributions with more than one mode.

The DPM prior for the ASV model can be viewed in terms of a random, second-order, effects model, where Λ_t is the random effect, but with a slight twist. Unlike a random effects model where Λ_t is assumed to follow a parametric, Inverse-Wishart distribution, in the ASV-DPM model G is unknown and is modeled nonparametrically. As a unknown random distribution function, G enables the Λ_ts to be distributed with "multimodality", and more "skewness" and "kurtosis" than is possible with a parametric distribution. However, because G is nonparametric the second-order, random effects matrices, Λ_ts, do not have any financial or economic meaning. They are simply building blocks in fitting the unknown distribution of (ϵ_t, η_t).

Employing Sethuraman (1994) definition of $DP(\alpha, G_0)$, G will almost surely be equal to the discrete distribution

$$G(d\Lambda) = \sum_{j=1}^{\infty} \pi_j \delta_{\Sigma_j}(d\Lambda), \tag{7}$$

where $\delta_{\Sigma_j}(\cdot)$ is a degenerative distribution on the covariance matrix atom

$$\Sigma_j = \begin{pmatrix} \sigma^2_{y,j} & \sigma_{yh,j} \\ \sigma_{yh,j} & \sigma^2_{h,j} \end{pmatrix}, \tag{8}$$

with $\sigma_{yh,j} = \rho_j \sigma_{y,j} \sigma_{h,j}$. Each Σ_j is a unique covariance matrix randomly drawn from the DP prior's base distribution, G_0. To ensure conjugacy, G_0 is a Inverse-Wishart distribution with scale matrix S_0 and v_0 degrees of freedom, i.e.,

$$G_0 \equiv \text{Inv-Wish}(S_0, v_0). \tag{9}$$

The probability of Λ_t being equal to a particular Σ_j is π_j where $\pi_1 = V_1, \pi_j = V_j \prod_{j'<j}(1 - V_{j'})$, and $V_j \sim \text{Beta}(1, \alpha)$, for $\alpha > 0$.

In (7), G_0 is our "best" guess at the distribution of the Λ_ts. Because the V_js are distributed as $\text{Beta}(1, \alpha)$, their expected value is $E[V_j] = 1/(1+\alpha)$. It follows that for large values of α the π_js will be close to zero and, hence, G will be close to G_0. Formally,

$$\left.\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix}\right| \begin{aligned} \Lambda_t &\sim N(\mathbf{0}, \Lambda_t), \\ \Lambda_t &\sim G_0 \equiv \text{Inv-Wishart}(S_0, v_0), \end{aligned}$$

when $\alpha \to \infty$. This illustrates the flexibility of the ASV-DPM because, as $\alpha \to \infty$, the ASV-DPM converges to a parametric, ASV model where $(\epsilon_t, \eta_t)'$ are distributed as a bivariate, Student-t distribution with mean zero, covariance matrix, $v_0/(v_0 - 2)S_0$, and v_0 degrees of freedom.[4] In the Student-t case, $\exp\{h_t\}$ does not typically have finite unconditional moments, nor covariance stationary (see Nelson (1991)). Hence, the posterior moments of $\exp\{h_t\}$ depend on the data.

At the other end of the spectrum when α close to zero, the DP prior for G is a discrete distribution with support over a few unique covariance matrices, Σ_js. When taken to its extreme, $\Lambda_t = \Sigma$, for all t, and

$$\begin{aligned} \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} &\sim N(\mathbf{0}, \Sigma), \\ \Sigma &\sim \text{Inv-Wishart}(S_0, v_0) \end{aligned}$$

when $\alpha \to 0$. In other words, the ASV-DPM model converges to the parametric, normally distributed, ASV model of Harvey et al. (1994) with G_0 being the prior of Σ.

2.1 Parsimony with the DP

Parsimony, in other words, clustering or uniqueness in the covariances, Λ_t, of the ASV-DPM model is provided by the almost sure discreteness of Eq. (7). By modeling G as a DP prior

[4] Because the ASV-DPM is a mixture of Normals it only spans the set of Student-t distributions whose marginals distributions have the same degree of freedom. For the marginals to have different degrees of freedom one could replace the mixture of bivariate Normals with a mixture of copulas (see Burda and Prokhorov (2012) and Rey and Roth (2012)). This is a topic for future research.

there is guaranteed to be ties among the Λ_ts. To be explicit, the joint distribution of the covariances can be defined sequentially as $\pi(\Lambda_1, \Lambda_2, \ldots, \Lambda_n) \equiv \pi(\Lambda_1)\pi(\Lambda_2|\Lambda_1)...\pi(\Lambda_n|\Lambda_1, \ldots, \Lambda_{n-1})$ where $\pi(\Lambda_1) \equiv G_0$ and

$$\begin{aligned} \Lambda_t|G, \Lambda_1, \ldots, \Lambda_{t-1} &\sim G, & (10) \\ G|\Lambda_1, \ldots, \Lambda_{t-1} &\sim DP\left(\alpha, \frac{\alpha}{\alpha+t-1}G_0 + \sum_{t'=1}^{t-1}\frac{1}{\alpha+t-1}\delta_{\Lambda_{t'}}(dG)\right), & (11) \end{aligned}$$

for $t = 2, \ldots, n$ (see Blackwell and MacQueen (1973)). Integrating out G from each of the conditional distributions in Eq. (10), we obtain the conditional distribution for Λ_t

$$\Lambda_t|\Lambda_1, \ldots, \Lambda_{t-1} \sim \begin{cases} G_0 & \text{with probability } \frac{\alpha}{\alpha+t-1}, \\ \Sigma_j & \text{with probability } \frac{n_j}{\alpha+t-1}, \quad j = 1, \ldots, k_t, \end{cases} \tag{12}$$

where Σ_j, $j = 1, \ldots, k_t$ are the unique covariance matrices among the $\Lambda_{t'}$, $t' = 1, \ldots, t$, and k_t is the number of unique covariances.

Eq. (12) shows the self-reinforcing property of the DP where previously sampled values are more likely to be resampled in the future. The more $\Lambda_{t'}$s belonging to a cluster, the larger n_j will be and the greater the probability Λ_t being assigned Σ_j. On the other hand, if only a few $\Lambda_{t'}$ have been assigned Σ_j, both n_j and the likelihood of Λ_t being assigned Σ_j will be smaller. Notice also that there is a non-trivial chance, proportional to α, of a new covariance cluster being selected from G_0. If $\alpha \to \infty$, no clustering occurs, whereas there is extreme clustering when $\alpha \to 0$.

2.2 Orthogonal representation

The ASV-DPM model can be written in terms of orthogonal innovations by first defining the latent assignment variable $s_t = j$ when Λ_t equals the jth unique covariance Σ_j; i.e., when $\Lambda_t = \Sigma_j$, then $s_t = j$. Under the DP prior s_t is distributed

$$s_t \sim \sum_{j=1}^{\infty} \pi_j \delta_j,$$

where the probability weights, π_j, $j = 1, \ldots,$ are those defined in Eq. (7).

Incorporating s_t into the definition of the ASV-DPM model, we arrive at

$$\begin{aligned} y_t &= \mu + \sigma_{y,s_t} \exp\{h_t/2\}u_t, & (13) \\ h_{t+1} &= \varphi h_t + \sigma_{h,s_t} v_t, & (14) \\ \begin{pmatrix} u_t \\ v_t \end{pmatrix} \Big| s_t, \Upsilon_{s_t} &\sim N(\mathbf{0}, \Upsilon_{s_t}), & (15) \end{aligned}$$

$$s_t \sim \sum_{j=1}^{\infty} \pi_j \delta_j, \tag{16}$$

$$\Sigma_{s_t} \sim G_0, \tag{17}$$

where the correlation matrix $\Upsilon_j = \begin{pmatrix} 1 & \rho_j \\ \rho_j & 1 \end{pmatrix}$, with $\rho_j = \sigma_{hy,j}/(\sigma_{h,j}\sigma_{y,j})$.

Letting $\Upsilon_{s_t}^{1/2}$ represent the Cholesky decomposition, $\Upsilon_{s_t} \equiv \Upsilon_{s_t}^{1/2}\Upsilon_{s_t}^{1/2'}$, we pre-multiply $(u_t, v_t)'$ by the inverse of $\Upsilon_{s_t}^{1/2}$ to obtain the uncorrelated innovation vector

$$\begin{pmatrix} w_t \\ v_t \end{pmatrix} \equiv \left(\Upsilon_{s_t}^{1/2}\right)^{-1} \begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} (u_t - v_t\rho_{s_t})/\sqrt{1-\rho_{s_t}^2} \\ v_t \end{pmatrix}.$$

Solving for u_t in terms of w_t and substituting this into Eq. (13), the ASV-DPM model in terms of the orthogonal shocks $(w_t, v_t)'$ equals:

$$y_t = \mu + \sigma_{y,s_t} \exp\{h_t/2\}\rho_{s_t} v_t + \sigma_{y,s_t} \exp\{h_t/2\}\sqrt{1-\rho_{s_t}^2}\, w_t \tag{18}$$

$$h_{t+1} = \varphi h_t + \sigma_{h,s_t} v_t. \tag{19}$$

where $(w_t, v_t)' \overset{iid}{\sim} N(0, \mathbf{I_2})$. This form of the ASV-DPM model will be shown to be convenient for the posterior sampling of log-volatilities and φ.

3 Estimation

In this section we outline the likelihood-based approach we take to parameter inference, distributional uncertainty, and model comparison with the ASV-DPM model by sketching out the Markov chain Monte Carlo (MCMC) sampler. We leave the details to Appendix A. The MCMC sampler has a number of advantages. Along with providing parameter estimates, the MCMC sampler also estimates the latent volatilities and integrates out the uncertainty of the latent mixture variables from the DPM prior.

Let $y = (y_1, \ldots, y_n)'$ be the observed asset returns and $h = (h_0, h_1, \ldots, h_n, h_{n+1})'$ the vector of its unobserved log-volatilities. The ASV-DPM posterior distribution

$$\pi(\mu, \varphi, h, \{\Lambda_t\}, \alpha | y) \propto f(y|\mu, \varphi, h, \{\Lambda_t\}, \alpha)\pi(\mu)\pi(h|\varphi)\pi(\varphi)\pi(\{\Lambda_t\}|\alpha)\pi(\alpha)$$

does not have a closed form. As a result, we strategically group the unknown parameters, latent volatilities, and mixture order, identities and assignments into manageable blocks where the selected blocks conditional posterior distributions are either known or have a tractable form. A Markov chain is then constructed by iteratively sampling through each block's posterior distribution conditioning on the value of the other parameters and latent variables drawn earlier.

The blocks of conditional distributions are:

- $\pi(\{\Lambda_t\}|y, h, \mu, \varphi, \alpha)$
- $\pi(h|y, \mu, \varphi, \{\Lambda_t\})$
- $\pi(\varphi|y, h, \mu, \{\Lambda_t\})$
- $\pi(\mu|y, h, \{\Lambda_t\})$
- $\pi(\alpha|\{\Lambda_t\})$

Sampling from $\{\Lambda_t\}|y, h, \mu, \varphi, \alpha$ could be done by using the Polya urn, Gibbs sampler method of Escobar (1994). However, we choose the more efficient two-step, Poly urn based approach of West et al. (1994) and MacEachern and Müller (1998). This method makes draws from the equivalent distribution $\pi(\Sigma_1, \ldots, \Sigma_k, s|y, h, \mu, \varphi, \alpha)$, where the Σ_j, $j = 1, \ldots, k$, $k \leq n$, are the distinct Λ_t, $t = 1, \ldots, n$, and $s = (s_1, \ldots, s_n)'$ is the vector consisting of the assignment variables, s_t; i.e., $s_t = j$ when $\Lambda_t = \Sigma_j$. This sampler breaks up the draws of s_t, $t = 1, \ldots, n$ from the draws of Σ_j, $j = 1, \ldots, k$, resulting in less dependency than if Λ_t, $t = 1, \ldots, n$ were drawn directly. Hence, decreasing the number of sweeps needed for the sampler to span the support of the posterior distribution.

Since log-volatilities are highly correlated, step-by-step draws from $h_t|y, h_{t-1}, h_{t+1}$ will mix very slowly.[5] Hence, we propose a more efficient tailored, Metropolis-Hasting sampler of randomly drawn blocks of h. Draws from $\pi(h|y, \mu, \varphi, \{\Lambda_t\})$ are made by forming random partitions of h where the length of each subvector in the partition is equal to a random draw from a Poisson distribution. Each block will be different from sweep to sweep ensuring that the blocks are not drawn conditionally on the same adjacent volatilities. Each block volatility's conditional distribution is nonstandard so we use the Metropolis-Hasting sampler (see Chib and Greenberg (1995)) found in Appendix A.2.

Draws from $\pi(\varphi, |y, h, \{\Lambda_t\})$ are made with a Metropolis-Hastings sampler where the candidate distribution is the conditional distribution of φ in a SV model without leverage (see Appendix A.3 for details). Since $\pi(\mu) \sim N(m, \tau)$, $\pi(\mu|y, h, \{\Lambda_t\})$ is Normal with the mean and variance spelled out in Appendix A.4. Lastly, when the mixture order, k, identifying vector, s, and covariances $\{\Sigma_j\}$, are all known, the posterior of α depends only on k. So instead of directly drawing from $\pi(\alpha|\{\Lambda\})$ we instead use the two step algorithm of Escobar and West (1995) found in Appendix A.5 to sample from the equivalent distribution $\pi(\alpha|k)$.

[5] See Kim et al. (1998) and Chib et al. (2002) for evidence of this with the SV model.

4 Results from quasi-return data

To determine how well the ASV-DPM model approximates the unknown distribution, and how well the sampler spans the posterior distribution of the parameters, we apply our MCMC sampler to returns generated from three parametric, asymmetric, stochastic volatility models where

$$y_t = \mu + \exp\{h_t/2\}\epsilon_t, \tag{20}$$

$$h_{t+1} = \varphi h_t + \eta_t, \tag{21}$$

is the common structure, but the distributions of ϵ_t and η_t differ. They are:

1. Harvey et al. (1994) ASV model with $(\epsilon_t, \eta_t)' \sim N(\mathbf{0}, \Sigma)$, where $\Sigma = \begin{pmatrix} \sigma_y^2 & \rho\sigma_y\sigma_h \\ \rho\sigma_y\sigma_h & \sigma_h^2 \end{pmatrix}$ and $\rho \equiv \text{Corr}(\epsilon_t, \eta_t)$.

2. Jacquier et al. (2004) ASV-t model where $\epsilon_t \equiv \lambda_t^{1/2} z_t$ and $\lambda_t \overset{iid}{\sim}$ Inv-Gamma$(\nu/2, \nu/2)$ with $\nu = 10$; i.e., ϵ_t is distributed as a Student-t with ν degrees of freedom. The innovations $(z_t, \eta_t)' \sim N(0, \Sigma)$.

3. Durham (2007) ASV-MIX3 model where

$$\epsilon_t \sim 0.8386N(0.015, 1.029^2) + 0.0041N(-3.611, 2.651^2) + 0.1576N(0.015, 0.444^2),$$

   $\eta_t \sim N(0, \sigma_h^2)$ and $\text{Corr}(\epsilon_t, \eta_t) = \rho$. The mixture parameter values come from Durham (2007) and ensure a mean of zero and a variance of one for ϵ_t.

The value of the structual parameters in Eq. (20)-(21) and for Σ are set equal to the estimates reported in Table 2 of Section 5. These estimates come from estimating the ASV model with 7,319 daily returns (multiplied by a 100) of the Center of Research in Security Prices (CRSP) value-weighted portfolio index over the January 2, 1980 to December 31, 2008 time period. These parameter values are used to generate 1,000 quasi-returns from each of the three models.

We fit the ASV-DPM to returns simulated from the ASV, ASV-t and ASV-MIX3 models. The priors for the ASV-DPM model are $\pi(\mu) \equiv N(0, 0.1)$ and $\pi(\varphi) \equiv N(0, 100)I_{|\varphi|<1}$. For the DPM, we choose the base distribution $G_0 \equiv$ Inv-Wish(S_0, v_0) where $S_0 = \boldsymbol{I}_2$ and $v_0 = 10$. The prior for the DPM precision parameter is $\pi(\alpha) =$ Gamma$(2, 8)$ so that $E[\alpha] = 1/4$ and $\text{Var}[\alpha] = 1/32$.

Using the initial starting parameter values, we throw away the first 1,000 draws of log-volatility and then the following 10,000 draws of both the volatilities and parameters, before

keeping the last 30,000 parameter draws. Table 1 reports the posterior mean, standard deviation, and 95% probability interval of the parameters from the ASV-DPM model when applied to the quasi-returns. Also reported is the inefficiency measure of Geweke (1992) which is

$$1 + \frac{2R}{R-1} \sum_{\tau=1}^{L} K\left(\frac{\tau}{L}\right) \gamma(\tau),$$

where $K(\cdot)$ is Parzen's filter (see Percival and Walden (1993), p. 265), $\gamma(\cdot)$ is the sample autocorrelation function of the drawn parameter, R is the number of draws ($R = 30{,}000$), and L is the largest lag at which the autocorrelation function is computed ($L = 1000$). The inefficiency measure quantifies how well the sampler has converged to the target posterior density and how close it is to making uncorrelated draws by quantifying the losses from using correlated draws to compute the posterior properties like mean and standard deviation.

There is nothing out of the ordinary in the posterior results of Table 1. The ASV-DPM parameter estimates are reasonably close to their true value. The average number of mixture clusters k for the different quasi-return series reflect the increasing complexity of the data generating process's underlying distribution. Lastly, the inefficiency measures range from a high of 260.0 for the ASV-t model's φ to a low of 5.3 for the ASV-MIX3 model's μ. These levels of inefficiency are much smaller than those reported by Omori et al. (2007) with the single-move h sampler of Jacquier et al. (2004). This improvement in the inefficiency measures is likely due to our random block sampler of h.

Table 1: The posterior estimates from the ASV-DPM model as applied to 1,000 quasi-returns simulated from the ASV, ASV-t, and ASV-MIX3 models. Given the initial starting parameter values, the first 1,000 draws of log-volatility are discarded and then the next 10,000 draws of both the volatilities and parameters are thrown away, before keeping the last 30,000 draws.

	ASV				ASV-t				ASV-MIX3			
	mean	stdev	95% prob interval	ineff	mean	stdev	95% prob interval	ineff	mean	stdev	95% prob interval	ineff
α	0.1690	0.1414	(0.0207, 0.4727)	6.9	0.2254	0.1534	(0.0305, 0.6051)	26.1	0.2509	0.1646	(0.0347, 0.6545)	22.5
k	1.0	0.8454	(1, 4)	23.5	2.3884	1.3337	(1, 6)	76.0	2.7617	1.4616	(1, 6)	56.3
φ	0.9634	0.0118	(0.9386, 0.9849)	146.9	0.9611	0.0150	(0.9301, 0.9896)	260.0	0.9482	0.0169	(0.9108, 0.9774)	68.8
μ	0.0501	0.0237	(0.0032, 0.0962)	80.6	0.0552	0.0244	(0.0071, 0.1033)	50.2	0.0730	0.0195	(0.0345, 0.1111)	5.3

Returns were generated from the three parametric ASV models with the same structure, $y_t = \mu + \exp\{h_t/2\}\epsilon_t$ and $h_{t+1} = \varphi h_t + \eta_t$, where $\mu = 0.06$ and $\varphi = 0.97$, but the ASV innovations are distributed $(\epsilon_t, \eta_t)' \sim N(\mathbf{0}, \Sigma)$ with

$$\Sigma = \begin{pmatrix} 0.61 & -0.46 * \sqrt{0.61 * 0.04} \\ -0.46 * \sqrt{0.61 * 0.04} & 0.04 \end{pmatrix}.$$

The ASV-t return innovations are $\epsilon_t \equiv \lambda_t^{1/2} z_t$, with $\lambda_t \overset{iid}{\sim}$ Inv-Gamma$(10/2, 10/2)$, and $(z_t, \eta_t)' \sim N(\mathbf{0}, \Sigma)$. Lastly, the ASV-MIX3 innovations are distributed as $\epsilon_t \sim 0.8386 \cdot N(0.015, 1.029^2) + 0.0041 \cdot N(-3.611, 2.651^2) + 0.1576 \cdot N(0.015, 0.444^2)$ and $\eta_t \sim N(0, 0.04)$ with $\text{Corr}(\epsilon_t, \eta_t) = -0.46$.

4.1 Predictive density

We also compare the joint predictive densities, $f((y_{n+1}, h_{n+2})'|y, M)$, $M =$ ASV-DPM, ASV, by numerically calculating and graphing each of the two model's density. Let $\theta = (\mu, \varphi, h, \{\Sigma_j\}, s, \alpha)$. Then the predictive density for the ASV-DPM is

$$\begin{aligned} f\left(\left(\begin{array}{c} y_{n+1} \\ h_{n+2} \end{array}\right)\middle| y, \text{ASV-DPM}\right) &\equiv \int f\left(\left(\begin{array}{c} y_{n+1} \\ h_{n+2} \end{array}\right)\middle|\theta\right)\pi(\theta|y)\,d\theta, \\ &\approx R^{-1}\sum_{l=1}^{R} f\left(\left(\begin{array}{c} y_{n+1} \\ h_{n+2} \end{array}\right)\middle|\theta^{(l)}\right), \end{aligned} \tag{22}$$

where $R = 30{,}000$, $\theta^{(l)} = \left(\mu^{(l)}, \varphi^{(l)}, h^{(l)}, \left\{\Sigma_j^{(l)}\right\}, s^{(l)}, \alpha^{(l)}\right)$ is the lth draw from the posterior $\pi(\theta|y)$ and $G(d\Lambda)$ has already been integrated out of the unknown joint distribution of ϵ_t and η_t (see Gelfand and Mukhopadhyay (1995)) to arrive at

$$\begin{aligned} f\left(\left(\begin{array}{c} y_{n+1} \\ h_{n+2} \end{array}\right)\middle|\theta\right) &= \frac{\alpha}{\alpha+n} f_{MSt}\left(\left(\begin{array}{c} y_{n+1} \\ h_{n+2} \end{array}\right)\middle|\left(\begin{array}{c} \mu \\ \varphi h_{n+1} \end{array}\right), \left(\frac{H_{n+1}S_0H_{n+1}}{v_0-1}\right)^{-1}, v_0-1\right) \\ &\quad + \sum_{j=1}^{k}\frac{n_j}{\alpha+n} f_N\left(\left(\begin{array}{c} y_{n+1} \\ h_{n+2} \end{array}\right)\middle|\left(\begin{array}{c} \mu \\ \varphi h_{n+1} \end{array}\right), H_{n+1}\Sigma_j H_{n+1}\right), \end{aligned} \tag{23}$$

with $H_{n+1} = \left(\begin{array}{cc} e^{h_{n+1}/2} & 0 \\ 0 & 1 \end{array}\right)$. By averaging over the weighted draws of the parameters and the unknown volatilities, the predictive density integrates out both parameter and log-volatility uncertainty leaving a distribution dependent on only the return series, y.

In the ASV model, the predictive density integrates out μ, φ, h, Σ, from the sampling distribution,

$$f_N\left(\left(\begin{array}{c} y_{n+1} \\ h_{n+2} \end{array}\right)\middle|\left(\begin{array}{c} \mu \\ \varphi h_{n+1} \end{array}\right), H_{n+1}\Sigma H_{n+1}\right). \tag{24}$$

Similarly, the marginal posterior predictive density of y_{n+1} can be approximated by averaging over the randomly sampled draws of θ from the posterior $\pi(\theta|y)$. For the ASV-DPM model the marginal posterior predictive density for returns equals

$$\begin{aligned} f\left(y_{n+1}|y, \text{ASV-DPM}\right) &= \int f(y_{n+1}|\theta)\pi(\theta|y)d\theta, \\ &\approx R^{-1}\sum_{l=1}^{R} f\left(y_{n+1}\middle|\theta^{(l)}\right), \end{aligned} \tag{25}$$

where

$$f\left(y_{n+1}\middle|\theta^{(l)}\right) = \frac{\alpha^{(l)}}{\alpha^{(l)}+n} f_{St}\left(y_{n+1}\middle|\mu^{(l)}, \left(\frac{s_{11}e^{h_{n+1}^{(l)}}}{v_0-1}\right)^{-1}, v_0-1\right)$$

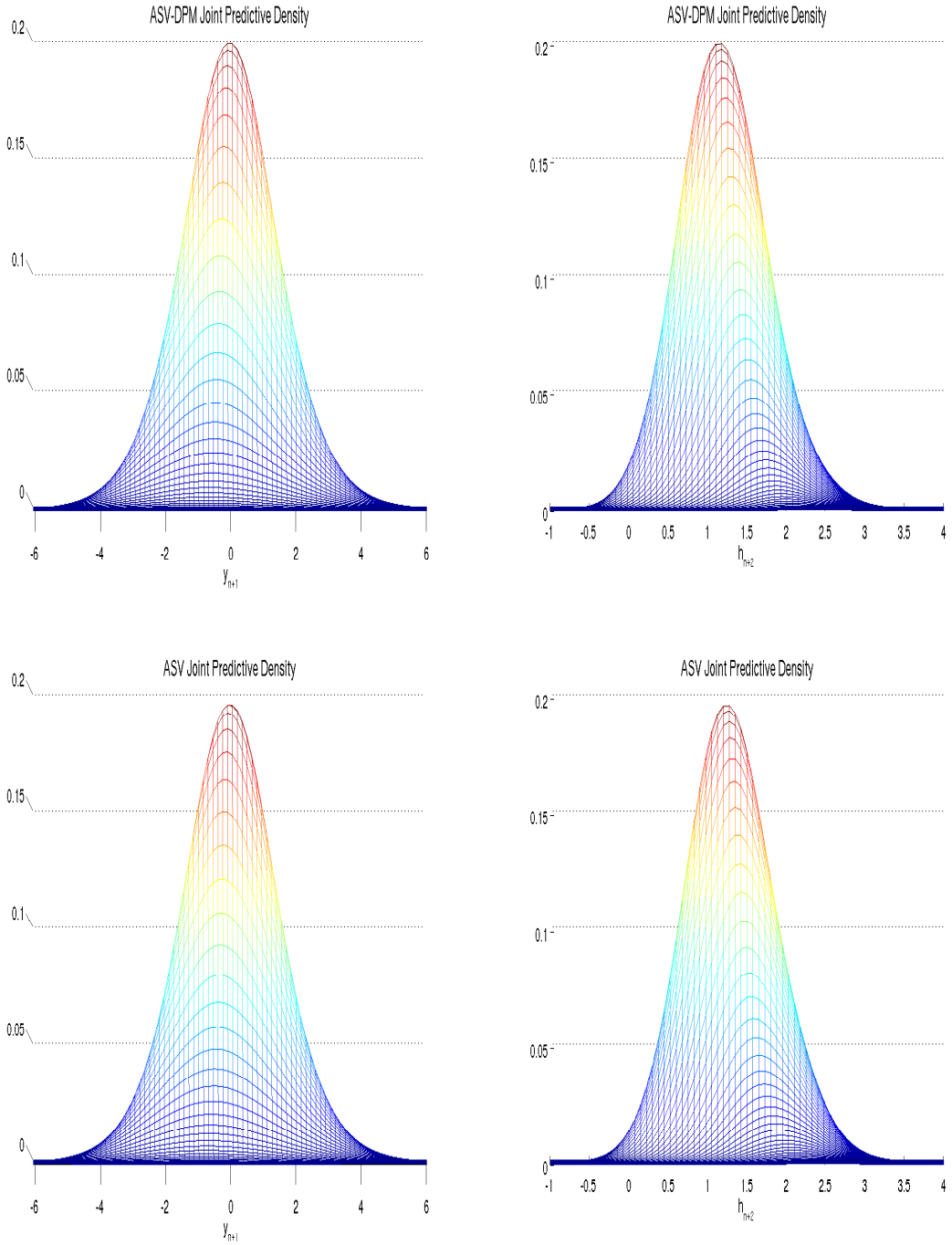
$$+\frac{1}{\alpha^{(l)}+n}\sum_{j=1}^{k^{(l)}} n_j^{(l)} f_N\left(y_{n+1}\left|\mu^{(l)}, e^{h_{n+1}^{(l)}}\sigma_{y,j}^{2(l)}\right.\right). \tag{26}$$

Given a realization of y_{n+1}, Eq. (26) is the predictive likelihood for the ASV-DPM model.

In addition to estimating the ASV-DPM model we now estimate the ASV model for the quasi-returns generated from the ASV. The priors for the ASV are those for the ASV-DPM, meaning the prior for Σ is G_0. The sampler of the ASV model is also the same as the ASV-DPM model's except with k and s_t, $t = 1, \ldots, n$, fixed and set equal to 1.

In Figure 1 we plot the joint predictive densities of both models. Since the densities are three-dimensional, we plot each model's density from two vantage points – the lefthand side figures plot the joint densities from the y_{n+1}-axis vantage point, and the righthand figures plot the densities, but from the h_{n+2}-axis perspective. The two model's densities are nearly identical in their shape and location. Both densities are centered at $y_{n+1} = 0$ and $h_{n+2} = 1.25$, and both show a slight upward skewness in the h_{n+2} dimension. If there is a difference to be found it is in their height. The ASV-DPM model's predictive density reaches a maximum density value of 0.2 that is slightly larger than the ASV, indicating the predictive distribution of the ASV-DPM model is leptokurtotic relative to the predictive distribution of the ASV model. This does not come as a surprise. From the results in Table 1, there are sweeps where the DPM sampler drew a mixture representation with two or more clusters. Furthermore, in Eq. (23) we see that the ASV-DPM predictive density includes the Student-t density function with $v_0 - 1$ degrees of freedom – a known leptokurtotic distribution.

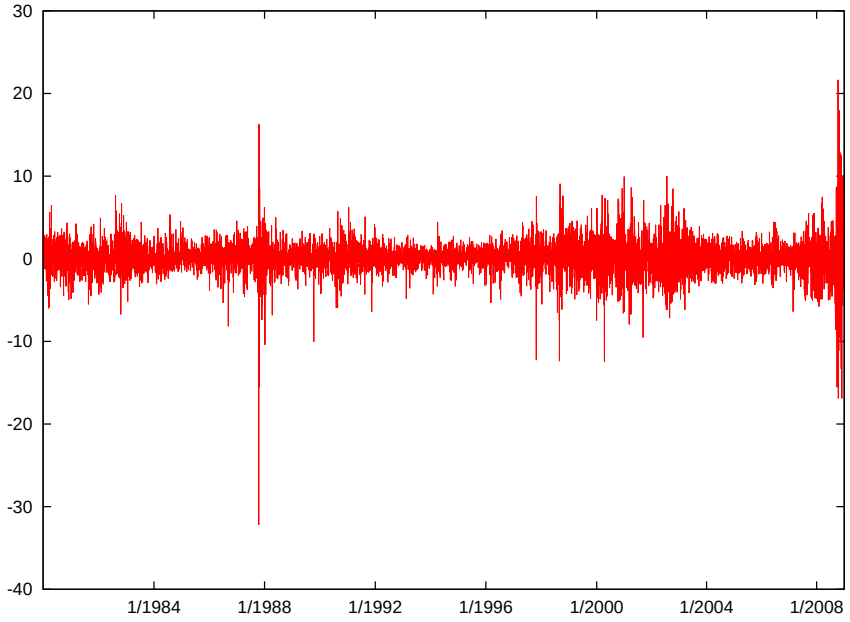
Figure 1: Joint predictive densities of the ASV-DPM and ASV model from the vantage points of the y_{n+1}-axis (lefthand side plots) and h_{n+2}-axis (righthand side plots) as applied to return data simulated from the ASV model.



5 Empirical Application

We analyze the ASV-DPM and ASV models by applying them to 7,319 daily returns (multiplied by a 100) over the period of January 2, 1980 to December 31, 2008 from the Center of Research in Security Prices (CRSP) value-weighted portfolio index. In Figure 2, we plot the value-weighted portfolio returns. The chosen time period is ideal since market returns exhibit a number of different dynamics. For example, the pre- and post-1987 market crash periods, the tech bubble of the late 90s, and the financial crisis of 2008. Over the entire sample, returns average 0.045 and have a variance of 1.12. Daily market returns appear to be asymmetrically distributed with fat-tails as is evident in a negative skewness of -0.757 and an excess kurtosis measure of 19.296.

Figure 2: CRSP value-weighted portfolio daily compounded returns from January 2, 1980 to December 31, 2008 (in percentages).



The priors applied to the ASV-DPM and ASV models are the same as were used in Section 4. They are $\pi(\mu) \equiv N(0, 0.1)$, $\pi(\varphi) \equiv N(0, 100)I_{|\varphi|<1}$, $G_0 \equiv$ Inv-Wish$(\boldsymbol{I}_2, 10)$, and $\pi(\alpha) =$ Gamma$(2, 8)$. For the ASV model, the prior for Σ is the Inv-Wish$(\boldsymbol{I}_2, 10)$ distribution and k and s_t, $t = 1, \ldots, n$, are set equal to 1. To reduce the influence of the starting values, we first perform 1000 sweeps over the log-volatilities using the step-by-step volatility sampler of Kim et al. (1998) while holding the other parameters constant. We then let the entire sampler of Section 3 iterate 40,000 times, keeping only the last 30,000

Table 2: Posterior estimates of the ASV-DPM and ASV models for daily compounded CRSP value-weighted portfolio returns from January 2, 1980 to December 29, 2008 (7319 observations, 41,000 draws with the first 1000 draws of log-volatility followed by the next 10,000 draws of all the unknowns being discarded).

	ASV-DPM				ASV			
	mean	stdev	95% prob interval	ineff	mean	stdev	95% prob interval	ineff
f	0.0423	0.0191	(0.0057, 0.0727)	39.38				
α	0.3188	0.1609	(0.0862, 0.7007)	16.53				
k	4.4220	1.3120	(3, 8)	56.52				
φ	0.9799	0.0030	(0.9739, 0.9855)	136.94	0.9740	0.0037	(0.9662, 0.9809)	73.38
μ	0.0665	0.0082	(0.0506, 0.0826)	163.02	0.0611	0.0083	(0.0448, 0.0774)	56.50
σ_y^2					0.6100	0.0514	(0.5112, 0.7169)	337.98
σ_h^2					0.0391	0.0046	(0.0308, 0.0490)	131.62
ρ					-0.4682	0.0371	(-0.5402,-0.3944)	40.78

f is the predictive likelihood, $f(y_{n+1}|y)$, where $y_{n+1} = 3.0911$ is the daily return on January 2, 2009.

draws from the two models for inference purposes.

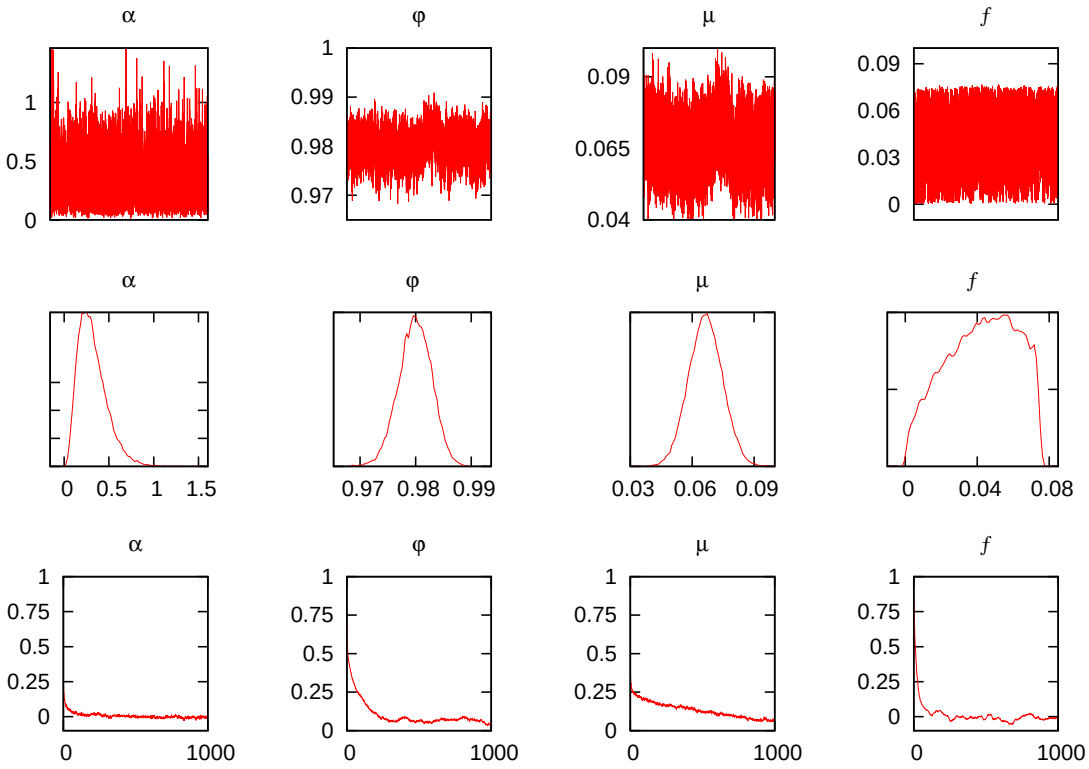
Table 2 reports the posterior mean, standard deviation, and 95% Bayesian probability interval for the parameters of the ASV-DPM and ASV models, along with the posterior estimate of the predictive likelihood, $f(y_{n+1}|y)$, for the ASV-DPM model. Posterior draws of the ASV-DPM parameters along with their smoothed histogram and autocorrelation function are plotted in Figure 3. Although some of the posterior draws are correlated, our sampler overall mixes well and produces realizations from over the entire posterior distribution.

The posterior mean for the ASV-DPM model's unconditional mean of return, μ, at 0.067 is slightly larger than the ASV model's estimate of 0.061. But their posterior standard deviations have the same value of 0.008. Hence, the model's posterior distributions, $\pi(\mu|y)$, are similar in shape but the ASV-DPM is shifted slightly to the right. Although the difference in the estimates of μ are small, it can still have an effect on one's expected median wealth say 20 years into the future. For example the median expected wealth per unit of investment over twenty years ($e^{\mu/100*365*20}$) is 128.32 for the ASV-DPM model as opposed to 86.51 for the ASV.

Persistence in volatility, as captured by the posterior distribution of the autoregressive parameter, φ, is close to being the same for the two models. In the ASV-DPM model, the posterior mean of φ is 0.98, whereas, in the ASV model it is 0.97. The standard deviations are also similar with the ASV-DPM model's equaling 0.003 and the ASV slightly larger at

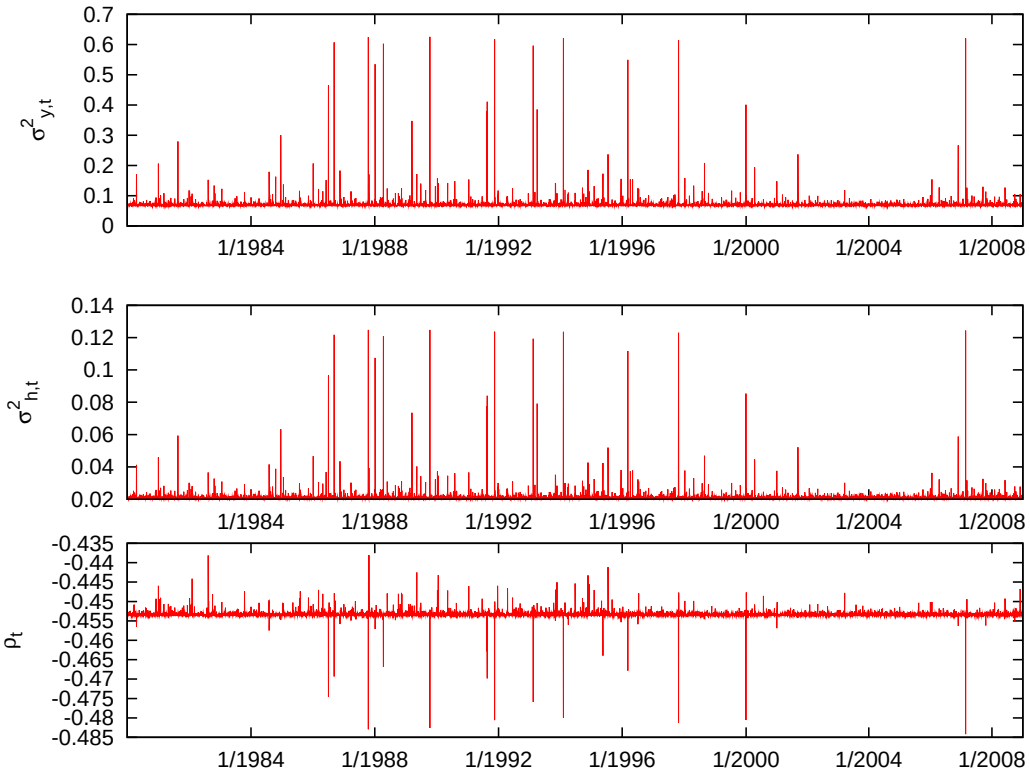
0.004. Values of φ so close to one is evidence of a strongly persistent volatility process where a shock to volatility impacts its future values and lives on for a very long time in either model. Even under a nonparametric distribution, the φ for ASV-DPM finds clustering in volatility where large and small fluctuations follow similar type of behavior.

Figure 3: 30,000 MCMC draws, their smoothed histogram and autocorrelation function, from the ASV-DPM model of the daily compounded CRSP value-weighted portfolio returns from January 2, 1980 to December 29, 2008.



The average number of posterior clusters is $k = 4.42$. In other words, the DPM uses on average 4.42 bivariate Normal densities to model the unknown return, log-volatility distribution. We calculate the average posterior value for each covariance matrix, Λ_t, $t = 1, \ldots, 7319$ and plot them in Figure 4. The first graph is of the average posterior draw of $\sigma^2_{y,t}$, the second graph $\sigma^2_{h,t}$, and the third is of ρ_t. The behavior over time in $\sigma^2_{y,t}$ and $\sigma^2_{h,t}$ are indistinguishable. Both move at the same time and in the same direction. This dynamic suggests that a market where $\sigma^2_{y,t}$ is large also has a large $\sigma^2_{h,t}$. Instances where the $\sigma^2_{y,t}$ and $\sigma^2_{h,t}$ are both large also helps identify mixture clusters with relatively low probability, but, which only occur when the market declines. For example, those instances where $\sigma^2_{y,t} >$

Figure 4: Time plots of the posterior average of $\sigma^2_{y,t}$, $\sigma^2_{h,t}$, ρ_t, t=1,...,7319



0.5, $\sigma^2_{h,t} > 0.1$, $\rho_t < -0.457$ only occur eleven times out of the 7319 observations, but each of these epsiodes corresponds to a market decline ranging from the 2% market drop on February 4, 1994 to the 17% decline during the October '87 crash.[6] This cluster contrasts with the one where $\sigma^2_{y,t} \approx 0.07$, $\sigma^2_{h,t} \approx 0.02$, $\rho_t \approx -0.455$. Because most of the sample consists of this second cluster, it has the largest mixture probability and, hence, represents the market's 'typical' covariance.

Figure 4 also identifies a third mixture cluster. This third cluster consists of Λ_ts whose $\sigma^2_{y,t}$ are between 0.15 and 0.4, and whose $\sigma^2_{h,t}$ are between 0.04 and 0.08. The cluster's correlation is no different from the -0.45 found for the 'typical' mixture cluster. In total there are 29 occurences of the third cluster with twenty-two occuring when returns were negative, and, except for four of these days, declining 2%. On the days were returns were positive the market increased betweeen 2 to 4%. Hence, unlike the first cluster where returns

[6]The exact dates of these episodes, with the market return in parenthesis, are: 9/11/1986 (-4.4%), 10/19/1987 (-17.1%), 1/8/1988 (-5.5%), 4/14/1988 (-3.6%), 10/13/1989 (-5.3%), 11/14/1991 (-3.4%), 2/16/1993 (-2.6%), 2/4/1994 (-2.3%), 3/8/1996 (-2.8%), 10/27/1997 (-6.5%), 2/27/2007 (-3.4%).

Table 3: Posterior estimates of the ASV-DPM model with the precision parameter prior Gamma(0.1, 0.1) for daily compounded CRSP value-weighted portfolio returns from January 2, 1980 to December 29, 2008 (7319 observations, 41,000 draws with the first 1000 draws of log-volatility followed by the next 10,000 draws of all the unknowns being discarded).

	mean	stdev	95% prob interval	ineff
f	0.0443	0.0188	(0.0067, 0.07307)	33.17
α	0.5085	0.3767	(0.0613, 1.4901)	54.19
k	5.3737	2.3133	(3, 11)	96.01
φ	0.9799	0.0030	(0.9736, 0.9853)	140.39
μ	0.6684	0.0081	(0.0510, 0.0825)	150.24

f is the posterior predictive likelihood of y_{n+1}.

were always negative, this third cluster occurs during a market gain or loss. Its leverage effect is also smaller.

5.1 Robustness to $\pi(\alpha)$

To gauge the sensativity of the ASV-DPM results to $\pi(\alpha)$, we re-estimate the ASV-DPM model using the more diffuse prior $\pi(\alpha) \equiv \text{Gamma}(0.1, 0.1)$. Under this prior $E[\alpha] = 1$ and $\text{Var}[\alpha] = 10$. Table 3 contains the results. Similar to Jensen and Maheu (2010) findings for the semiparametric stochastic volatility model, the structurual parameters, μ, φ, are unchanged relative to Table 2 results. On the other hand, the DPM parameters, α and k, both go up. The precision parameter goes from 0.32 to 0.51 and k increases from a posterior median of 4 to 5 clusters. In addition, $\pi(\alpha)$ effect on the predictive likelihood, f, is negligble. Neither the predictive likelihood's posterior mean, standard deviation, probability interval, or inefficiency level are impacted by the change to $\pi(\alpha)$.

5.2 Fit

A point of possible contention with the ASV-DPM model is that it fits an unknown joint distribution where one of the random variables, h, is not observed. To address this concern we compare the draws of h from the ASV-DPM model with those from the ASV model. If the ASV-DPM model is unable to fit a nonparametric distribution to the latent volatility process, we would expect to find the draws from the smooth distribution $\pi(h|y)$ to be less precise and have a very different posterior mean from the parametric model. We would also expect the location and spread of the ASV-DPM model's joint predictive density to not match up with the ASV model's predictive distribution.

In Figure 5 we plot two graphs. In Figure 5(a) we plot the difference between the ASV and ASV-DPM sampler's standard deviations of $\pi(h|y)$ over the trading days, January 2, 1980 to December 29, 2008, and in Figure 5(b) the sample means of the two model's draws of h. Except for a few trading days, the standard deviation of the ASV draws are slightly larger than the ASV-DPM models. Because of the large difference between the ASV-DPM and ASV standard deviation on those days where the ASV-DPM standard deviations is larger, on average, the ASV-DPM standard deviations are 0.0026 larger than the ASV. On these days market volatility was generally low.

In Figure 5(b), the two model's smoothed volatilities, $E[h|y]$, are similar in their level and pattern. However, there are differences, such as when volatility reaches a local peak. In these instances, the ASV model's smooth volatility is larger than the ASV-DPM model. Because the instantaneous variance of returns, σ_y^2, in the ASV model is constant, large changes in the return process correspond to large changes in volatility. This contrasts with the ASV-DPM model where σ_{y,s_t}^2 is flexible and, as we saw in Figure 4, changes value when it needs to adapt to a market decline. As a result the ASV-DPM model's volatilities are less volatile and do not increase by as much as the ASV model.

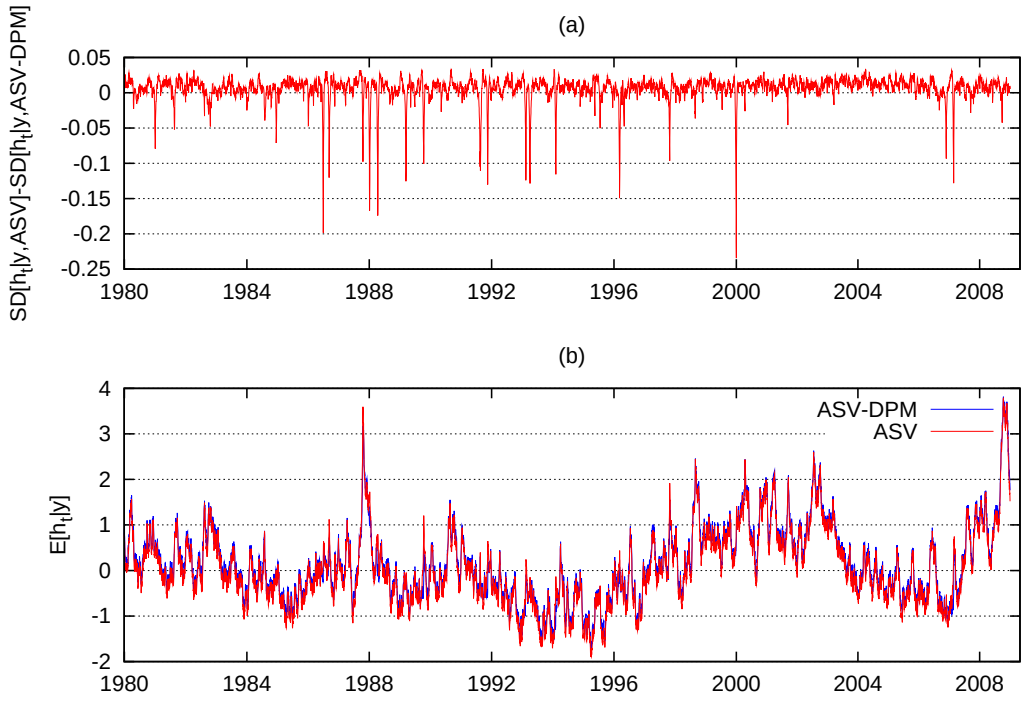
6 Nested Model Comparison

The DP precision parameter α can be understood as being a tuning parameter to the number of unique Λ_ts; i.e., the degree of clustering imposed by the DP prior. Under the $DP(\alpha, G_0)$ prior, a data set of length n is expected to have $E[k_n] = \sum_{i=1}^{n} \alpha/(\alpha + i - 1)$ clusters. By increasing or decreasing the value of α, the ASV-DPM model is more or less likely to add new clusters. As α approaches zero, the probability of a second unique covariance matrix, $\alpha/(1+\alpha)$, approaches zero, as does the probability of there being a second, third, fourth, etc. cluster. It follows that the ASV-DPM model is equivalent to the parametric ASV model when α is equal to zero.

At the other end of the spectrum, the prior probability of drawing a new cluster for Λ_t goes to one as $\alpha \to \infty$. Because the DP prior for G is no longer discrete, but is instead equal to the distribution G_0, when $\alpha \to \infty$, the prior for Λ_t, $t = 1, \ldots, n$, is G_0. Hence, as $\alpha \to \infty$, the ASV-DPM model is the ASV-t model – a parametric ASV model whose innovations are distributed as a bivariate, Student-t with mean-vector zero, covariance matrix, $S_0/(v_0 - 3)$ and $v_0 - 1$ degrees of freedom (see Eq. (43)).

Because the ASV and ASV-t models are nested versions of the ASV-DPM that depend on the value of α, Bayes factors in favor of the nested models can be computed using the

Figure 5: (a) difference between the ASV and ASV-DPM standard deviation from the MCMC draws of $\pi(h_t|y)$, and (b) the average draw of $h_t|y$ from the ASV and ASV-DPM. Both figures are for the period of January 2, 1980 to December 31, 2008.



Savage-Dickey density ratio test (see Dickey (1971)). In general, the Savage-Dickey density ratio favors the nested model $M' : \alpha = \alpha_0$, where, in our case, $\alpha_0 = \{0, \infty\}$, over the general model $M : \alpha$, where $\alpha \geq 0$, when Bayes-factor

$$BF(\alpha = \alpha_0) \equiv \frac{m(y|M')}{m(y|M)}, \tag{27}$$

$$= \frac{\pi(\alpha = \alpha_0|y, M)}{\pi(\alpha = \alpha_0)}, \tag{28}$$

where m is the marginal likelihood, is large. In our case, M = ASV-DPM and M' = ASV, ASV-t, respectively.

The limit $\alpha_0 \to \infty$ does not lend itself easily to the Savage-Dickey density ratio, so we transform α into the random variable $u \equiv \alpha/(\alpha + 1)$ and assume u is distributed as the maximum entropy prior

$$\pi(u|\lambda) = \frac{e^{\lambda u}\lambda}{e^{\lambda} - 1}, \qquad \lambda \in \mathbb{R}, \tag{29}$$

over the unit interval, $u \in [0, 1]$. The corresponding maximum entropy prior for α is

$$\pi(\alpha|\lambda) = \frac{\lambda \exp\{\alpha\lambda/(1 + \alpha)\}}{(e^{\lambda} - 1)(1 + \alpha)^2}. \tag{30}$$

Under this transformation, $u \to 0$, as $\alpha \to 0$, and $u \to 1$, as $\alpha \to \infty$. The random variable u is thus, the prior probability of there being a second mixture cluster. If $u = 1$, then Λ_2 will be drawn from G_0. Whereas, if $u = 0$, the probability of drawing a second cluster is zero and $\Lambda_2 = \Lambda_1$.

Let the nested ASV and ASV-t models $M' : u = u_0$, be, respectively, $u_0 = \{0, 1\}$ and the unrestricted model $M : u \in [0, 1]$. In terms of u, the Savage-Dickey density ratio in favor of the nested model M' is

$$BF(u = u_0) \equiv \frac{\pi(u = u_0|y, M)}{\pi(u = u_0)}.$$

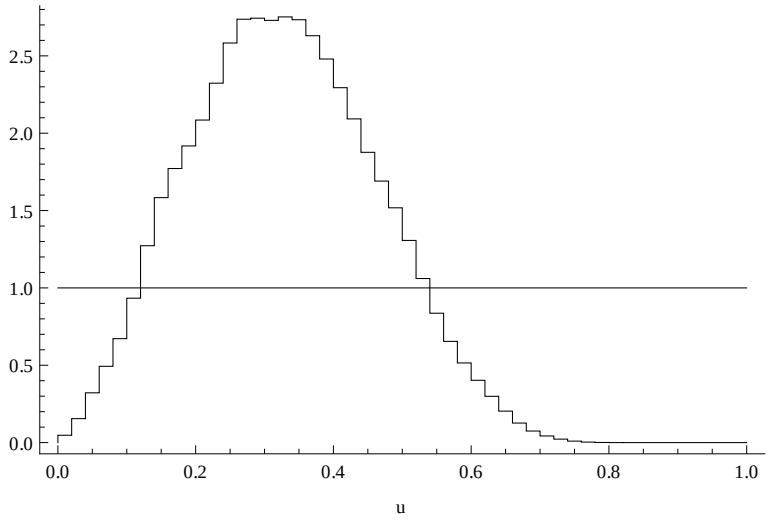
Unlike the Gamma prior we used earlier for α, the maximum entropy prior does not lend itself to a standard distribution for the conditional posterior of α, so, we modify the Escobar and West (1995) sampler of α. Since drawing either α or u requires a Metropolis sampler, we could choose to sample either one. However, because α is defined on the positive real line, whereas u is constrained to the unit interval, we choose to draw α and use a random walk proposal with unit variance to generate the candidate draws.

Denote the candidate draw by α'. It will be accepted as a draw from $\pi(\alpha|y)$ with probability

$$\frac{\pi(k|\alpha', n)\pi(\alpha'|\lambda)}{\pi(k|\alpha, n)\pi(\alpha|\lambda)} = \frac{\alpha'^k \Gamma(\alpha')/\Gamma(\alpha' + n)\pi(\alpha'|\lambda)}{\alpha^k \Gamma(\alpha)/\Gamma(\alpha + n)\pi(\alpha|\lambda)}, \tag{31}$$

where α is the draw from the previous sweep and k is the number of clusters from the current sweep (see Escobar and West (1995) for the formula of the likelihood function, $\pi(k|\alpha, n)$). For each draw of α we compute the corresponding draw of u and evaluate u's empirical posterior distribution $\pi(u|y)$ at zero and one. If the Savage-Dickey ratio at these points is greater than one then there is evidence in favor of the nested model, M'.

Figure 6: The prior, $\pi(u|\lambda = 0) \equiv \text{Unif}(0, 1)$, and the empirical posterior density $\pi(u|y, \text{ASV-DPM})$.



We choose the maximum entropy prior for u where $\lambda = 0$ when testing the nested ASV and ASV-t models against the ASV-DPM. When $\lambda = 0$, the maximum entropy prior is a uniform prior over the unit interval. Except for the prior of u, the MCMC sampler is the same as Section 5.

In Figure 6, we plot the uniform prior and empirical posterior density of u. The Savage-Dickey density ratio in favor of the ASV model ($u_0 = 0$) has a 1-in-20 chance of occurring and indicates that there is some likelihood for the parametric ASV model but not much. Since $\pi(u = 1|y) = 0$, there is virtually no evidence supporting the ASV-t model ($u_0 = 1$).

Though there is little empirical evidence supporting the nested ASV and ASV-t models, in Figure 6, there is still a range of values for u_0 where a sharp hypothesis is supported. If we restrict u_0 to values where $\pi(u = u_0|y, \text{ASV-DPM})$ is greater than $\pi(u_0)$, we find a number of Bayes-factors in favor of the restricted ASV-DPM model. From Figure 6, the data supports a ASV-DPM model where u is between 0.125 and 0.525, with a mode of $u \approx 0.3$. Given the relationships, $E[k_n] = \sum_{i=1}^{n} \alpha/(\alpha + i - 1)$ and $\alpha \equiv u/(1 - u)$, these

posterior values of u support ASV-DPM models with approximately 2 to 10 clusters.

6.1 Prior sensitivity

Because M' is a sharp hypothesis, the prior for u under M' is the Dirac delta function, $\pi(u|M') = \delta_{u_o}(u)$. For the unrestricted model, M, the prior is the maximum entropy distribution of Eq. (29) with hyperparameter $\lambda \in \mathbb{R}$. As $\lambda \to -\infty$, the prior $\pi(u|\lambda, M) \to \delta_0(u)$; i.e., the prior of the unrestricted model converges to that of the sharp ASV model hypothesis. As $\lambda \to \infty$, $\pi(u|\lambda, M) \to \delta_1(u)$ and the prior is that of the ASV-t model.

The marginal likelihood for M is

$$m(y|\lambda, M) = \int l(y|u, \lambda, M)\pi(u|\lambda, M)\, du, \tag{32}$$

and the marginal likelihood for the nested model M' is

$$\begin{aligned} m(y|M') &= \int l(y|u, \lambda, M)\delta_{u_0}(u)\, du & (33)\\ &= l(y|u = u_0, M). & (34) \end{aligned}$$

The Bayes factor in favor of the restricted model M' written in terms of these marginal likelihoods is

$$BF(u = u_0|y, \lambda, M) = \frac{l(y|u = u_0, M)}{m(y|\lambda, M)}.$$

Under this prior for u, there exists limits where the unrestricted model equals the restricted. The two restricted models are $u_0 = 0$ (ASV) and $u_0 = 1$ (ASV-t). To obtain the ASV model with the unrestricted model, the prior

$$\lim_{\lambda \to -\infty} \pi(u|\lambda, M) = \delta_0(u),$$

and for the ASV-t model

$$\lim_{\lambda \to \infty} \pi(u|\lambda, M) = \delta_1(u).$$

Thus, the Bayes factor in favor of the ASV model can be written in terms of M's marginal likelihood function as

$$BF(u = 0|y, \lambda, M) = \frac{m(y|\lambda \to -\infty, M)}{m(y|\lambda, M)} \tag{35}$$

and for the ASV-t the Bayes factor can be expressed as

$$BF(u = 1|y, \lambda, M) = \frac{m(y|\lambda \to \infty, M)}{m(y|\lambda, M)}. \tag{36}$$

Eq. (35) illustrates how the Bayes factor of a sharp hypothesis is influenced by the prior even when the posterior is robust to the prior. The situation occurs when the marginal likelihood of the unrestricted model is sensitive to the prior. For example, the maximum entropy prior causes the Bayes factor in favor of the ASV model to get closer and closer to one for more and more negative values of λ.

We use the prior $\pi(u|\lambda = -10, M)$ to compute the Bayes factor favoring the ASV model and the prior $\pi(u|\lambda = 10, M)$ for the Bayes factor favoring the ASV-t. These priors give the benefit of doubt to the restricted ASV-DPM model. In Figure 7 we graph in subplot (a) the prior $\pi(u|\lambda = 10, M)$ and in subplot (b) the prior $\pi(u|\lambda = -10, M)$. Each plot also contains the empirical distribution of u using the respective prior.

Using these two maximum entropy priors, the Savage-Dickey density ratio for the two sharp hypothesis reinforces the findings of Section 6. In the case where the prior lends support to a ASV model ($\lambda = -10$), the density ratio at $u_0 = 0$ in Figure 7(b) shows there to be slightly less than a 1-in-10 chance of a ASV model. The Savage-Dickey density ratio goes to zero in Figure 7(a) when the prior heavily weights ASV-DPM models having a large number of clusters. Neither prior changes the likelihood of the data coming from a ASV-t model. In each case $\pi(u = 1|y, M)$ is zero.

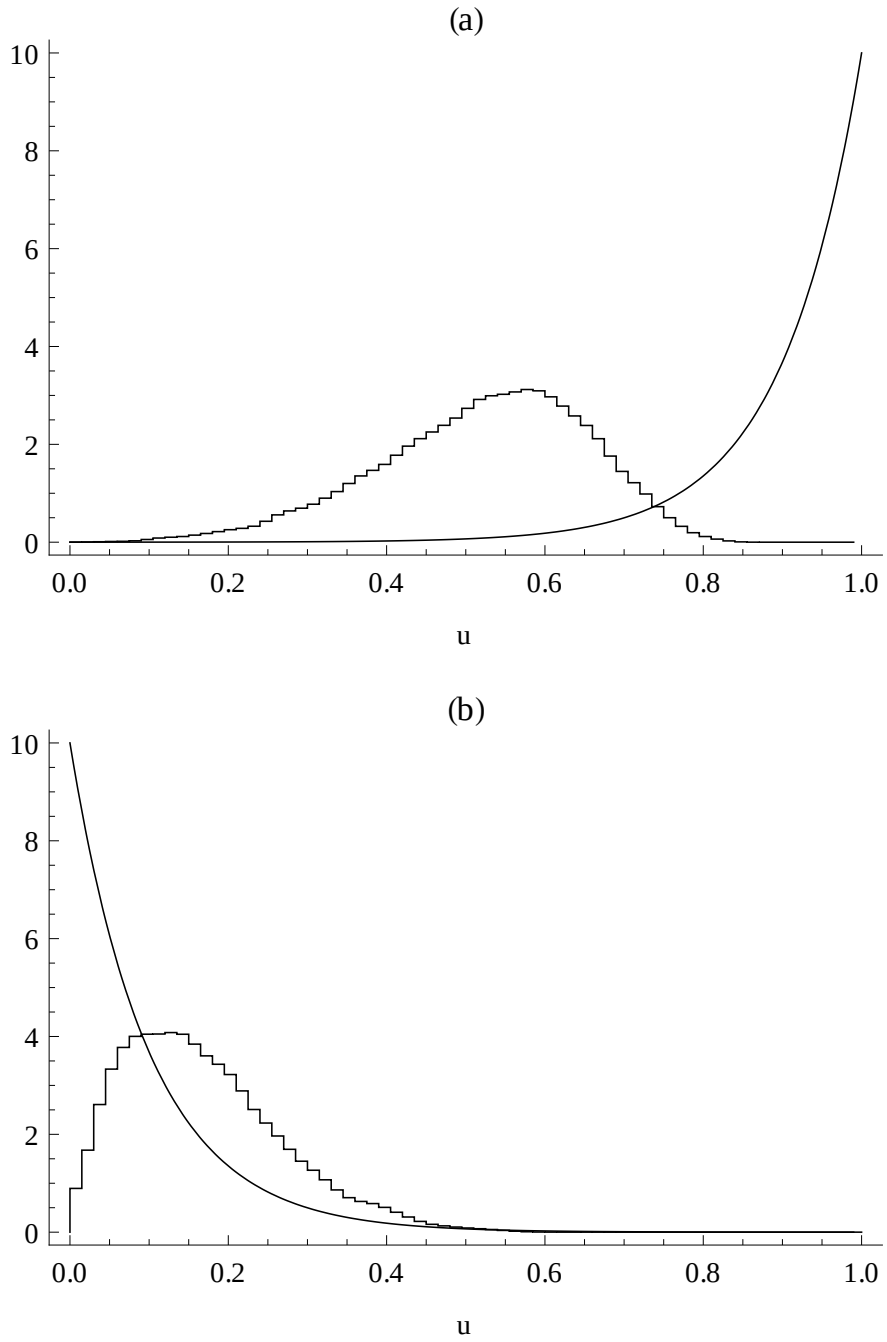
Figure 7 also shows how the expected number of posterior clusters is robust to the prior. As was the case with the uniform prior, the data supports a u between 0.125 and 0.5 (0.75) when $\lambda = -10$ ($\lambda = 10$). Again these are sharp hypothesis that the data can support.

7 Predictability

To compare the ASV-DPM model with the ASV and SV-DPM-P models, and more generally any model of returns, we compute the marginal likelihood of each model using the product of its one-step-ahead predictions. As we have mentioned, in addition to integrating out parameter uncertainty, the marginal likelihood of stochastic volatility models also integrates out the uncertainty associated with the latent volatilities. In the past, particle filters have been applied to stochastic volatility models to integrate out volatility (see Chib et al. (2002)). However, the marginal likelihood for the ASV-DPM model also requires integrating out the latent DP covariance matrices. Basu and Chib (2003) have a way of doing this but only for a DPM type model, not a DPM model with stochastic volatility. The ASV-DPM model requires a particle filter for integrating out the latent volatilities and DP parameters, making the Basu and Chib approach infeasible.[7]

[7] An appealing alternative to the approach we take here is the sequential Monte Carlo method to estimating and filtering DPM type model of Carvalho et al. (2010).

Figure 7: The empirical posterior density of u under the maximum entropy priors (a) $\pi(u|\lambda = 10)$ and (b) $\pi(u|\lambda = -10)$.



Because of the additional computation costs involved in integrating out the volatilities and DPM covariances, and also because of the increased availability of parallel computing, Beowulf clusters, and quad-core processors, we choose to compute the ASV model's marginal likelihood sequentially with one-step-ahead predictive likelihoods. Given the low cost of multi-thread computing and availability of multiple processors, a large number of individual and independent MCMC draws can be conducted on the histories of the return series. For the ASV-DPM, ASV and SV-DPM-P models we simulataneously carry out 50 separate and unique MCMCs on 5 servers where each server has two quad-core processors. As a result a model's marginal likelihood is computed well within the length of a day.

Our approach is as follows. Let the vectors $y^{t-1} = (y_1, \ldots, y_{t-1})'$, where $t = 2, \ldots, n$, denote the histories of returns up to time period $t-1$. By the law of conditional probability, the marginal likelihood can be expressed in terms of the one-step-ahead predictive likelihoods

$$m(y|M) = \prod_{t=1}^{n} f(y_t|y^{t-1}, M), \tag{37}$$

where M = ASV-DPM, ASV, SV-DPM-P. It is helpful to write the posterior predictive densities as

$$f(y_t|y^{t-1}, M) \equiv \int f(y_t|y^{t-1}, \Theta_M, M)\pi(\Theta_M|y^{t-1}, M)\, d\Theta_M, \tag{38}$$

where Θ_M is a vector of the unobservable parameters and volatilities. The posterior density

$$\pi(\Theta_M|y^{t-1}, M) \propto \pi(\Theta_M|M) \prod_{\tau=1}^{t-1} f(y_\tau|y^{\tau-1}, \Theta_M, M),$$

where $\pi(\Theta_M|M)$ is the Mth model's prior density for Θ_M. For the ASV-DPM model the one-step-ahead posterior predictive likelihoods, $f(y_t|y^{t-1}, \text{ASV-DPM})$, is found in Eq. (26) except it will be evaluated at y_t and y^{t-1} replaces y.

When t is small, $\pi(\Theta_M|M)$ will influence the one-step-ahead predictive likelihood. So, in practice, the product in Eq. (37) generally does not begin at $t = 1$, but will instead start further into the data set. However with DPM models the densities $f(y_\tau|y^{\tau-1}, \Theta_M, M)$, M = ASV-DPM, SV-DPM-P are unknown and must be learned by observing data. Whereas the ASV model's predictive density function is already known and similar to the prior it will influence $f(y_t|y^{t-1}, \text{ASV})$ and cause it to have a relative advantage over the DPM models. Hence, it will be informative to see how the ASV-DPM learns relative to the ASV by computing $f(y_t|y^{t-1}, M)$ for $t = 2, \ldots, n$. Our marginal likelihoods will then be $m_2(y|M) = \prod_{t=2}^{n} f(y_t|y^{t-1}, M)$.

Each of the $n-2$ MCMC posterior conditional samplers is independent from the others. Given this independence we only need to supply a particular sampler with one of the $n-2$ histories y^{t-1} before letting it run. First, we farm out as many histories as there are processors available. On a quad-core, multi-threaded, computer this generally equals 10 potential MCMC samplers. When a processor's task of sampling R draws from the posterior distribution conditioned on that particular history is completed, the processor computes the one-step-ahead predictive likelihood in Eq. (26) and returns it for later use. If predictive likelihoods for other histories still need to be computed, the processor will request another history and sample from its posterior. Once all $n-2$ predictive likelihoods have been computed the marginal likelihood is calculated.

7.1 Cumulative Bayes-factor

From the posterior predictive likelihoods, $f(y_t|y^{t-1}, M)$, the cumulative log-Bayes factor (CLBF) of the ASV-DPM relative to the other two models equals

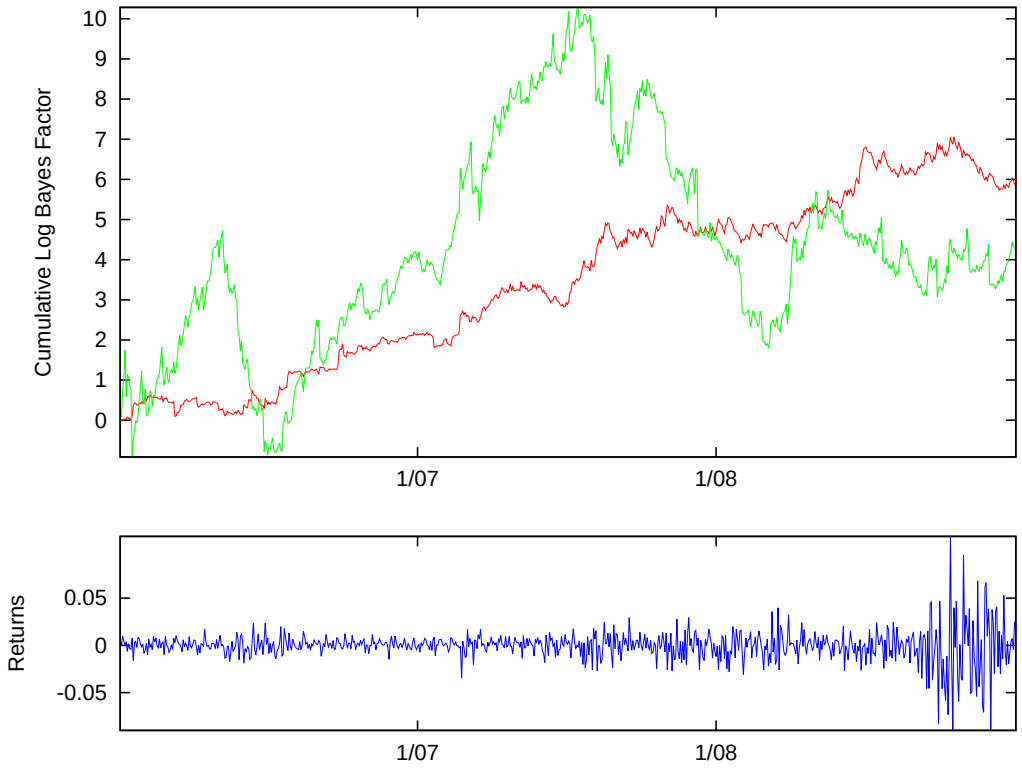
$$\log\left(\frac{m_L(y^\tau|\text{ASV-DPM})}{m_L(y^\tau|M)}\right) = \sum_{t=2}^{\tau} \log\left(\frac{f(y_t|y^{t-1}, \text{ASV-DPM})}{f(y_t|y^{t-1}, M)}\right), \quad \tau = 2, \ldots, n, \quad (39)$$

where M = ASV, SV-DPM-P. Each point on the CLBF represents the log-Bayes factor between the relative models for the history y^τ. By plotting the CLBF over τ we are able to identify those instances where one model out preforms the others. Day-to-day change in the CBLF depict how well the models perform relative to forecasting the next day's return as illustrated by the quantities $\log f(y_t|y^{t-1}, \text{ASV-DPM}) - \log f(y_t|y^{t-1}, M)$.

Because of the large computational costs involved in computing the predictive likelihoods over the 7,319 daily returns of Section 5, we investigate the cumulative log-Bayes factors using the value-weighted CRSP portfolio returns from January 3, 2006 to December 31, 2008 (755 trading days). For the three models we compute the posterior predictive likelihoods, $f(y_t|y^{t-1}, M)$, from $t = 2$ (Jan. 4, 2006) to 755 (Dec 31, 2008), by sampling the model's unknowns 11,000 times and discarding the first 1,000 draws. The cumulative log-Bayes factors are plotted in the top panel of Figure 8. In the bottom panel, we plot the daily return for the CRSP value-weighted portfolio over the corresponding period.

Unlike Geweke and Amisano (2010), who find the CBLF for a handful of symmetrical stochastic volatility models jumping by as much as 50 points on a particular days, the ASV-DPM model's CLBF do not move more than a point on a single day, but, instead exhibit general behavior and trends. For example, the ASV model's CLBF initially muddles around with values less than one until the middle of 2006. This muddling is a time of learning

Figure 8: Cumulative log-Bayes factor, $\log\left(\frac{m_L(y^\tau|\text{ASV-DPM})}{m_L(y^\tau|M)}\right)$, of the ASV-DPM model relative to the ASV (red line) and SV-DPM-P (green line) models from January 4, 2006 to December 31, 2008 using CRSP value-weighted portfolio returns (blue line) back to January 3, 2006.



for the ASV-DPM as it identifies the necessary number of mixture covariances and their locations. On the other hand, the ASV's normality assumption allows it to start reducing the uncertainty around its unknown covariance right from the beginning. After this initial period of learning, the CLBF begins to climb steadily, experiencing a few short lived drops along the way, in favor of the ASV-DPM model. During the summer of 2008 the ASV-DPM briefly enjoys a period where the CLBF exceeds seven before declining over the financial crisis to 6.0. This general upward trend in the ASV model's CLBF is evidence of the ASV-DPM model producing on average better day-ahead predictions of market returns than the ASV model. According to Jefferies (1961), a final CLBF value of six is "very strong" evidence in favor of the ASV-DPM model over the ASV model.

Leverage appears to play less of a role when markets are volatile than it does when market volatility is low. In Figure 8, the CLBF between the model with leverage, ASV-DPM, and the one without, SV-DPM-P, rises when the market is tranquil from mid-2006 to mid-2007 (as seen in the smoothness of CRSP returns over this period). During this time the CLBF rises to over 10, "decisive" evidence in favor of the ASV-DPM model for the first half of the return series. Returns then become more volatile and the SV-DPM-P model begins to produce on average better forecasts until volatility recedes after the first quarter of 2008. By the end of 2008 the CLBF is 4.3. This is "strong" evidence in favor of the ASV-DPM model relative to the SV-DPM-P for the entire series.

8 Volatility Response

Ever since Black (1976) proposed the leverage hypothesis and French et al. (1987) the volatility feedback effect, many have studied how volatility reacts to changes in market returns.[8] Yu (2005) and Asai and McAleer (2009) both establish the theoretical relationship for stochastic volatility and returns when leverage is present. Yu (2005) derives the volatility-return relationship for the asymmetric stochastic volatility model and Asai and McAleer (2009) for the multivariate stochastic volatility model, but neither empirically investigates the relationship.

Volatility's response in the ASV-DPM to a change in market returns is derived from Eq. (22) – the joint posterior predictive density, $f((y_{n+1}, h_{n+2})'|y^n)$. This predictive density dispenses with parameter and latent volatility uncertainty by integrating out the unknowns over their posterior.[9] For a 200×200 array of equally spaced values of y_{n+1} and h_{n+2}, we

[8] See Bekaert and Wu (2000) for a review of the research prior to 2000 and Chen and Ghysels (2011) for more recent work on the subject.

[9] The volatility-return relationship of Yu (2005) and Asai and McAleer (2009) do not integrate out this

compute $f((y_{n+1}, h_{n+2})'|y^n)$. In other words, we evaluate the predictive density at four-thousand points, $\{y_{n+1,j}, h_{n+2,j'}\}_{j,j'}$, where $j, j' = 1, \ldots, 200$. Tomorrow's 200 different returns are centered around the next day's actual return and the log-volatilities are centered around the average draw of h_{n+1} from the MCMC burnin.

The ASV-DPM's predictive joint density contains a healthy amount of information concerning the return-volatility relationship. However, since the predictive posterior density is bivariate and depends on today's return, y_n, it is difficult to visualize all this information in a density plot. We try to summarize some of the more interesting features of the return-volatility relationship by computing and plotting the conditional expectation of next period's log-volatility, $E[h_{n+2}|y^n, y_{n+1}]$, over a range of values for y_{n+1}. If tomorrow's return is y'_{n+1}, next period's expected log-volatility will equal

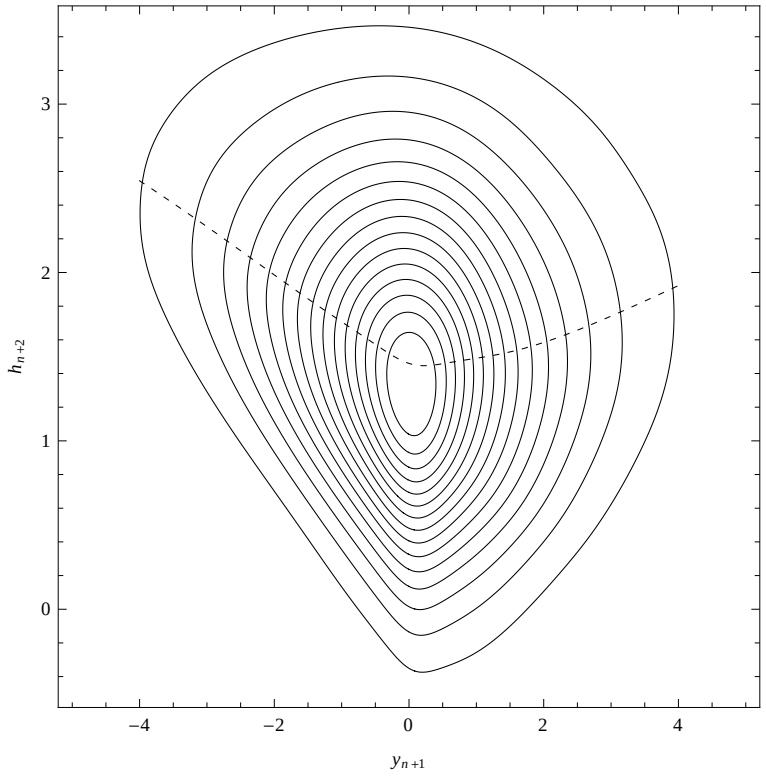
$$\begin{aligned} E[h_{n+2}|y^n, y'_{n+1}] &= \int h_{n+2}\, f(h_{n+2}|y^n, y'_{n+1})\, dh_{n+2} \\ &\approx \sum_{j=1}^{200} h_{n+2,j} \left[\frac{f(y'_{n+1}, h_{n+2,j}|y^n)}{\sum_{j'=1}^{200} f(y'_{n+1}, h_{n+2,j'}|y^n)} \right]. \end{aligned} \tag{40}$$

We compute the ASV-DPM model's joint predictive density using the MCMC draws from Section 7 and the CRSP portfolio returns from January 3, 2006 to December 31, 2008. In Figure 9, we plot fifteen of the contour lines from the predictive density (solid lines) along with $E[h_{n+2}|y^n, y_{n+1}]$ (dashed line). In the figure, the contour lines bow up and out when tomorrow's return is negative, while the contours are nearly linear over small values of h_{n+2}; i.e., conditional on returns being negative, the predictive distribution for h_{n+1} is not symmetrical but is skewed upward. Skewness is also present when $y_{n+1} > 0$. However, these contour lines are less (more) bowed out for large (small) values of h_{n+1}. When market returns are positive, the conditional distribution of tomorrow's log-volatility is skewed upward, but less so relative to when y_{n+1} is negative. Thus, tomorrow's volatility is likely to be higher following either a market decline or increase, but because of the difference in the degree of skewness, volatility's expected response is asymmetric.

This asymmetric response in volatility is seen in the dashed line of $E[h_{n+2}|y^n, y_{n+1}]$. When the market does not move, tomorrow's expected log-volatility is 1.46. With each percentage point decline in the daily market return, expected log-volatility increases at approximately the rate of 0.3. This contrasts with a market increase, where expected log-volatility responds in a nonlinear manner. Expected log-volatility ever so slightly declines for market gains smaller than 0.2%. increasing at a rate of 0.1. But as returns get even larger expected log-volatility rises at a faster rate of 0.2. Even though this rate of increase

uncertainty and depend on the value of the estimated parameters.

Figure 9: Contour lines (solid lines) of the ASV-DPM model's joint predictive density, $f((y_{n+1}, h_{n+2})'|y^n)$, and the conditional expected value of log-volatility given tomorrow's return, $E[h_{n+2}|y^n, y_{n+1}]$, (dashed line) where y^n contains 755 daily CRSP value-weighted portfolio returns from January 3, 2006, to December 31, 2008.

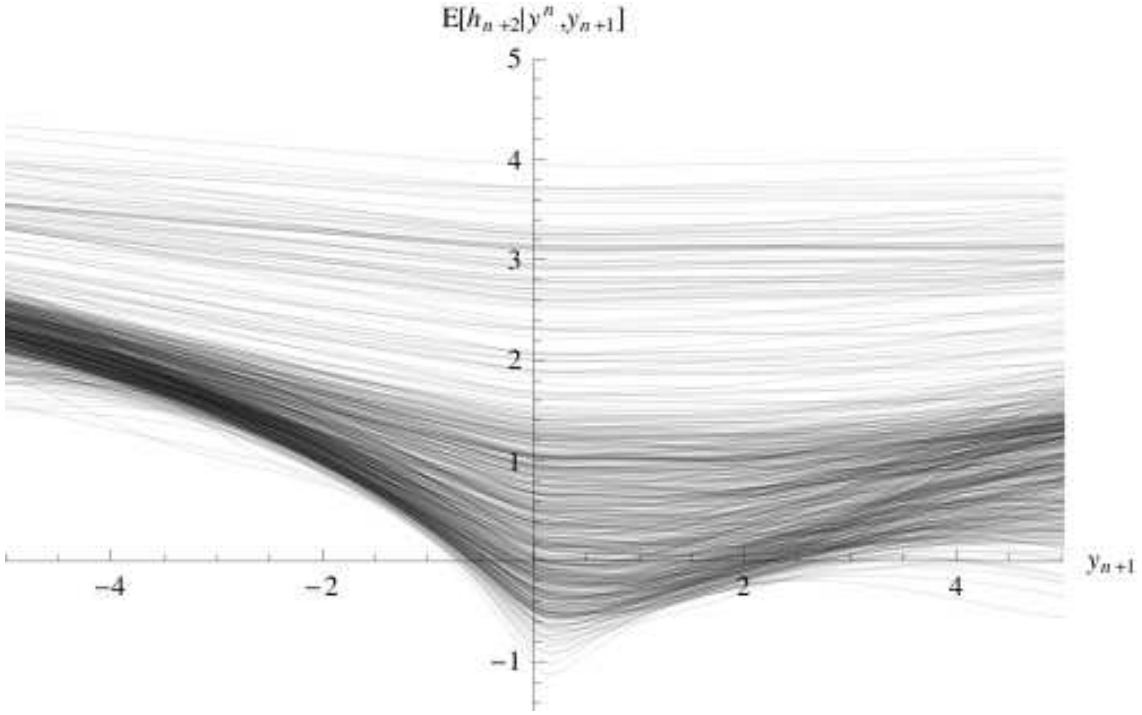


in expected log-volatility is smaller than had the market dropped, tomorrow's expected volatility will be larger after any sizable move in the market, be it negative or positive. Thus, in the words of Campbell and Hentschel (1992) and Chen and Ghysels (2011),"No news is good news," when talking about market returns and volatility.

8.1 Time-varying volatility response

Like volatility, the volatility-return relationship can vary. This begs the question, does the dashed line in Figure 9 represent the typical joint predictive distribution of tomorrow's market return and log-volatility, or is it an abnormality? To answer this question we delete the last 540 returns from the return history and sequentially estimate the volatility-return relationship beginning with the return history ending on November 8, 2006 and adding one return at a time until we reach December 31, 2008.

Figure 10: The ASV-DPM conditional expected log-volatility, $E[h_{n+2}|y^n, y_{n+1}]$, plotted against future return, y_{n+1}, for the return series, $y^n = (y_1, \ldots, y_n)$, $n = 216, \ldots, 755$.



In Figure 10, we plot $E[h_{n+2}|y^n, y_{n+1}]$ for each of the 540 return histories, y^n, $n = 216, \ldots, 755$. This figure illustrates how the volatility-return relationship depends on the return history. The value of $E[h_{n+2}|y^n, y_{n+1} = 0]$ ranges from -1 to 4, with the highest concen-

tration falling between -0.5 to 1.5. Histories associated with the largest $E[h_{n+2}|y^n, y_{n+1} = 0]$ end during the most volatile period of September, 2008. The $E[h_{n+2}|y^n, y_{n+1} = 0]$ near the origin have y^n that end during calmer more normal market periods. Current volatility explains 89 percent of the variation in $E[h_{n+2}|y^n, y_{n+1} = 0]$ when regressing it on $E[h_{n+1}|y^n]$. Hence, as one would expect tomorrow's expected volatility is strongly correlated with today's volatility.

Many studies of the leverage effect assume that debt is unaffected by changes in the stock market, and is hence, riskless; e.g. see Schwert (1989) and Figlewski and Wang (2000). Christie (1982) shows that by allowing the value of debt to change in the same direction as market capital, the leverage effect is weakened. Under risky debt, a negative market return leads to an increase in volatility, but the size of the impact declines as leverage grows. By most practitioners, leverage was viewed as being high during 2008. The return histories ending during this period have the highest values for $E[h_{n+2}|y^n, y_{n+1}]$, but also the flattest.

The degree of asymmetry in $E[h_{n+2}|y^n, y_{n+1}]$ is also affected. Asymmetry is greatest when $E[h_{n+2}|y^n, y_{n+1} = 0]$ is negative and today's volatility is low. At the other end of the spectrum, volatility's response to the market is nearly symmetrical when $E[h_{n+2}|y^n, y_{n+1} = 0]$ is greater than 2 and the return history ends during volatile bear markets. Volatility response is not only symmetrical for these histories, there is almost no volatility response at all.

The degree of shading in Figure 10 provides a rough density measure of the different volatility-return responses. Dark bands over the negative returns show that for many of the 540 volatility-return responses a drop in the market today corresponds to an increase in expected log-volatility. A similar but not quite as dark a band can also be seen for positive daily returns.[10] The lighter lines are the volatility-return relationships whose histories end during rare, but very volatile, times. As mentioned above, these responses are symmetrical and nearly flat. Thus, we see that during "normal" times, no news is good news for volatility, which is even more true during tranquil times. During the most turbulent of markets, news of any sort just does not matter for expected volatility. This is a notable feature of the semiparametric ASV model. When markets are highly volatile, a return shock must be large in order to affect the market's expectations about tomorrow's volatility. Whereas, on a typical day only a little bit of news is required to cause expected volatility to increase.

[10] The dark band in Figure 10 has a shape very similar to Chen and Ghysels (2011) nonparametric news impact curve on p. 49.

9 Conclusion

In this paper we extended the asymmetric, stochastic, log-volatility model whose innovations are correlated and normally distributed by modeling the uncertainty in their joint distribution with a nonparametric, bivariate Dirichlet process mixture prior. We provide a sampling algorithm to integrate out the parameter, volatility and distributional uncertainty of our semiparametric, asymmetric, stochastic volatility model. Our algorithm is also used to compute the log Bayes predictive forecast of the semiparametric model relative to the parametric version. These log Bayes predictions are used to evaluate and compare the forecasting abilities of the two models.

The nonparametric prior increases the flexibility of the asymmetric stochastic volatility model by allowing the correlation between its innovations to take on infinite number of values, while being manageable and parsimonious by taking on a finite number of correlations for a finite length data set. This flexibility is important when forecasting market returns, especially when the market transitions from a low to high volatility state or the market suddenly declines. In the empirical case study, forecasts from our semiparametric asymmetric stochastic volatility model capture these types of episodes, whereas the parametric model does not. This leads to the daily predictive Bayes factors favoring the nonparametric asymmetric stochastic volatility model more often than the parametric version.

The flexibility of having more than one value to model the correlation between volatility and return is also important for modeling the volatility-return relationship and the response in expected volatility to a unexpected change in market returns. In particular, the size and and degree of asymmetry in the response of volatility to an unexpected change in market prices can vary. Using the semiparametric model of the asymmetric stochastic volatility, the response in the expected value of volatility to a decline in stock prices versus an increase is highly asymmetric when volatility is currently low and the market is calm. However, if volatility is high and the market irregular, the asymmetry nearly disappears and the response in expected volatility becomes muted. In other words, during normal times, just a little bit of news affects volatility, whereas, during a turbulent market, it takes a significant amount of news to impact volatility.

A Details of MCMC sampler

A.1 Sampler of Λ

Define $\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)'$, where $\boldsymbol{z}_t = ((y_t - \mu)\exp\{-h_t/2\}, h_{t+1} - \varphi h_t)'$, then Λ_t conditional posterior distribution is :

$$\begin{aligned} \Lambda_t|\{\Lambda_{t'} : t' \neq t\}, \boldsymbol{z}_t, \alpha &\sim \frac{\alpha}{\alpha + n - 1} g(\boldsymbol{z}_t) G(d\Lambda|\boldsymbol{z}_t) \\ &\quad + \frac{1}{\alpha + n - 1} \sum_{t' \neq t} f_N(\boldsymbol{z}_t|\Lambda_{t'}) \delta_{\Lambda_{t'}}(d\Lambda), \end{aligned} \tag{41}$$

where $g(\boldsymbol{z}_t) \equiv \int f_N(\boldsymbol{z}_t|\boldsymbol{0}, \Lambda)\ G_0(\Lambda) d\Lambda$, and, by the law of conditional probability, $G(d\Lambda|\boldsymbol{z}_t) \propto f_N(\boldsymbol{z}_t|\boldsymbol{0}, \Lambda) G_0(d\Lambda)$. Applying the prior information of Eq. (3) and (9) it follows that the density for the distribution $G(d\Lambda|\boldsymbol{z}_t)$ is:

$$\begin{aligned} g(\Lambda|\boldsymbol{z}_t) &\propto |\Lambda|^{-1/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\boldsymbol{z}_t\boldsymbol{z}_t'\Lambda^{-1}\right\} \frac{|S_0|^{v_0/2}}{|\Lambda|^{(v_0+3)/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\Lambda^{-1}S_0\right\}, \\ &= \frac{|S_0|^{v_0/2}}{|\Lambda|^{(v_0+4)/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}(S_0 + \boldsymbol{z}_t\boldsymbol{z}_t')\Lambda^{-1}\right\}, \end{aligned} \tag{42}$$

This is the kernel to the Inverse-Wishart distribution, $G(d\Lambda|\boldsymbol{z}_t) \equiv \text{Inv-Wish}(S_0 + \boldsymbol{z}_t\boldsymbol{z}_t', v_0 + 1)$ (see Zellner (1971), p.395). Integrating out Λ from $G(d\Lambda|\boldsymbol{z}_t)$ results in the marginal likelihood function:

$$g(\boldsymbol{z}_t) = f_{MSt}(\boldsymbol{z}_t|\boldsymbol{0}, (S_0/(v_0 - 1))^{-1}, v_0 - 1), \tag{43}$$

In other words, the bivariate Student-t density function with $v_0 - 1$ degrees of freedom, mean-zero vector, and covariance, $S_0/(v_0 - 3)$ (see Zellner (1971), Eq. (B.20), p. 383 for the exact formula of f_{MSt}).

Let n_j be the number of observations where $s_t = j$, $k^{(t)}$ to be the distinct number of Σ_j in $\{\Lambda_{t'} : t' \neq t\}$, and $n_j^{(t)}$ the number of observations where $s_{t'} = j$, $t' \neq t$. For a given h, μ, φ and α, draws from the posterior $\pi(\{\Sigma_j\}, s|\boldsymbol{z}, \alpha) \equiv \pi(\{\Sigma_j\}, s|y, h, \mu, \varphi, \alpha)$ are made in 2-steps:

1. s and k are drawn by sampling s_t, $t = 1, \ldots, n$, from:

$$s_t|\{\Lambda_{t'} : t' \neq t\}, \boldsymbol{z}_t, \alpha \sim \begin{cases} \frac{\alpha}{\alpha+n-1} g(\boldsymbol{z}_t)\ \delta_0(ds_t) \\ \frac{1}{\alpha+n-1} \sum_{j=1}^{k^{(t)}} n_j^{(t)} f_N(\boldsymbol{z}_t|\boldsymbol{0}, \Sigma_j)\ \delta_j(ds_t). \end{cases} \tag{44}$$

   If zero is drawn for s_t, k increases by one, s_t is set equal to the new value of k, and a new Σ_k is drawn from the distribution whose density is Eq. (42). Otherwise, s_t equals the randomly drawn j.

2. Discard the Σ_js from Step 1 and use the sampled s and k to iteratively draw new Σ_j, $j = 1, \ldots, k$, from:

$$\begin{aligned}
\pi(\Sigma_j|\boldsymbol{z}, s, k) &\propto \prod_{t:s_t=j} f_N(\boldsymbol{z}_t|\mathbf{0}, \Sigma_j) G_0(d\Sigma) && (45)\\
&\propto \prod_{t:s_t=j} |\Sigma_j|^{-1/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\boldsymbol{z}_t\boldsymbol{z}_t'\Sigma_j^{-1}\right\} \\
&\quad \times \frac{|S_0|^{v_0/2}}{|\Sigma_j|^{(v_0+3)/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}S_0\Sigma_j^{-1}\right\} && (46)\\
&= \frac{|S_0|^{v_0/2}}{|\Sigma_j|^{(v_0+n_j+3)/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\sum_{t:s_t=j} \boldsymbol{z}_t\boldsymbol{z}_t' + S_0\right)\Sigma_j^{-1}\right\} && (47)\\
&\sim \text{Inv-Wish}\left(S_0 + \sum_{t:s_t=j} \boldsymbol{z}_t\boldsymbol{z}_t', v_0 + n_j\right). && (48)
\end{aligned}$$

A.2 Sampler of h

Given a particular partition of h sequentially draw each volatility block conditional on the value of the other volatilities. The conditional distribution of the volatility block $h_{(t',\tau)} = (h_{t'}, h_{t'+1}, \ldots, h_\tau)'$, with $1 \leq t' \leq \tau < n$, and $l_{t'} \equiv \tau - t' + 1$ being randomly distributed as a Poisson distribution, Pois(λ_h), is:

$$\begin{aligned}
\pi(h_{(t',\tau)}|y_{(t',\tau)}, h_{-(t',\tau)}) &\propto \pi(h_{(t',\tau)}|h_{\tau+1}, h_{t'-1}) f(y_{(t',\tau)}|h_{(t',\tau)}, h_{\tau+1}) \\
&= \exp\left\{-\frac{1}{2}\frac{(h_{t'} - \varphi h_{t'-1})^2}{\sigma^2_{h,s_{t'}}}\right\} \prod_{t=t'}^{\tau} \exp\left\{-\frac{1}{2}\left[\frac{(h_{t+1} - \varphi h_t)^2}{\sigma^2_{h,s_t}} + h_t\right]\right\} \\
&\quad \times \prod_{t=t'}^{\tau} \exp\left\{-\frac{1}{2}\frac{\left(y_t - \mu - e^{h_t/2}\sigma_{y,s_t}\rho_{s_t}(h_{t+1} - \varphi h_t)/\sigma_{h,s_t}\right)^2}{(1-\rho^2_{s_t})\exp\{h_t\}\sigma^2_{y,s_t}}\right\}. && (49)
\end{aligned}$$

For the first block, $\pi(h_{(1,\tau)}|h_{\tau+1}, h_0)$ depends on h_0. The unknown h_0 is modeled with the prior $\pi(h_0) \sim N(0, \sigma^2_{h,0}/(1-\varphi^2))$, where $\sigma^2_{h,0} \equiv E[G_0(d\Lambda)]_{2,2}$ is the expected variance of log-volatility from the the DPM base distribution. By drawing h_0 from $\pi(h_0|h_1) \equiv N(\varphi h_1, \sigma^2_{h,0})$, it is integrated out.

If the draw of $l_{t'}$ causes τ to be greater than or equal to n, the conditional distribution of the volatility block is the same as above except $\tau = n$. For the last block of h we integrate out the one period ahead, out of sample, volatility, h_{n+1}, by replacing it with a random draw from:

$$\pi(h_{n+1}|y, h_n) \sim N(\bar{h}_{n+1}, \bar{\sigma}^2_{h_{n+1}}), \tag{50}$$

where

$$\bar{h}_{n+1} = \varphi h_n + \frac{(y_n - \mu)\rho_{s_n}\sigma_{h,s_n}e^{-h_n/2}}{\sigma_{y,s_n}}, \qquad \bar{\sigma}^2_{h_{n+1}} = \left(\frac{\rho^2_{s_n}}{(1-\rho^2_{s_n})\sigma^2_{h,s_n}} + \frac{1}{\sigma^2_{h,s_n}}\right)^{-1},$$

and h_n is from the previous sweep of the sampler.

The conditional distributions $\pi(h_{(t,\tau)}|y_{(t,\tau)}, h_{-(t,\tau)})$ are nonstandard, so a Metropolis-Hasting sampler is used. Candidate draws of $h_{(t,\tau)}$ are made from a l_t-variate Student-t distribution having mean vector $\boldsymbol{m}$, covariance matrix S, and ξ degrees of freedom. The candidate mean vector, $\boldsymbol{m}$, is set equal to the argument maximizing $\pi(h_{(t',\tau)}|h_{\tau+1}, h_{t'-1})f(y_{(t',\tau)}|h_{(t',\tau)}, h_{\tau+1})$, and the candidate covariance, S, equals the negative inverted Hessian of the unnormalized conditional distribution evaluated at $\boldsymbol{m}$.

A.3 Sampler of φ

Draws from $\pi(\varphi|y, h, \{\Lambda_t\})$ are made with a Metropolis-Hasting sampler whose candidate distribution is $N(\widehat{\varphi}, \widehat{\sigma}^2_\varphi)$ where:

$$\widehat{\varphi} = \widehat{\sigma}^2_\varphi \left(\frac{\mu_\varphi}{\sigma^2_\varphi} + \sum_{t=1}^{n-1}\frac{h_{t+1}h_t}{\sigma^2_{h,s_t}}\right), \qquad \widehat{\sigma}^2_\varphi = \left(\frac{1}{\sigma^2_\varphi} + \sum_{t=1}^{n-1}\frac{h_t^2}{\sigma^2_{h,s_t}}\right)^{-1}.$$

A draw, φ', from this candidate distribution will be accepted with probability $\alpha(\varphi', \varphi) = \min\left\{\frac{g(\varphi')}{g(\varphi)}\frac{f_N(\varphi|\widehat{\varphi},\widehat{\sigma}^2_\varphi)}{f_N(\varphi'|\widehat{\varphi},\widehat{\sigma}^2_\varphi)}, 1\right\}$ where

$$\begin{aligned} g(\varphi) &\propto f(y|\varphi, h, \{\Sigma_j\}, s, h_{n+1})\pi(h|\varphi, \{\Sigma_j\}, s)\pi(h_{n+1}|h_n, \varphi)\pi(h_0|\varphi)\pi(\varphi), \\ &= \prod_{t=1}^{n}\exp\left\{-\frac{1}{2}\left[\frac{\left(y_t - \mu - \rho_t e^{h_t/2}\sigma_{y,s_t}(h_{t+1} - \varphi h_t)/\sigma_{h,s_t}\right)^2}{(1-\rho^2_{s_t})\sigma^2_{y,s_t}e^{h_t}} + \frac{(h_{t+1} - \varphi h_t)^2}{\sigma_{h,s_t}}\right]\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}\left[\frac{(h_1 - \varphi h_0)^2}{\sigma^2_{h,0}} + \frac{h_0^2}{\sigma^2_{h,0}/(1-\varphi^2)}\right]\right\} f_N(\varphi|\mu_\varphi, \sigma^2_\varphi)I_{|\varphi|<1}. \end{aligned}$$

If the candidate φ' is rejected, the previous sweep's value of φ is kept as the current sweep's draw from $\pi(\varphi|y, h, \{\Lambda_t\})$.

A.4 Sampler of μ

To draw from $\pi(\mu|y, h, \{\Lambda_t\})$, let $\pi(\mu) \sim N(m, \tau)$. Since $\pi(\mu)$ is Normal, draws of μ are made from the conjugate posterior $N(\hat{\mu}, \hat{\tau})$ where:

$$\hat{\mu} = \hat{\tau}\left(\frac{m}{\tau} + \sum_{t=1}^{n}\frac{\tilde{y}_t}{\tilde{\sigma}^2_t}\right), \qquad \hat{\tau} = \left(\frac{1}{\tau} + \sum_{t=1}^{n}\frac{1}{\tilde{\sigma}^2_t}\right)^{-1},$$

and $\tilde{y} = y_t - \rho_t e^{h_t/2}\sigma_{y,s_t}(h_{t+1} - \varphi h_t)/\sigma_{h,s_t}$ and $\tilde{\sigma}^2_t = (1-\rho^2_{s_t})\sigma^2_{y,s_t}\exp\{h_t\}$.

A.5 Sampler of α

The two step algorithm of Escobar and West (1995) is used to sample the ASV-DPM model's precision parameter from $\pi(\alpha|\{\Lambda_t\})$ equivalent distribution $\pi(\alpha|k)$. More precisely, assume $\pi(\alpha) \sim \Gamma(a, b)$, where $a > 0$ and $b > 0$. Draws from $\pi(\alpha|k)$ are made by

1. Sampling the random variable ξ from $\pi(\xi|\alpha, k) \sim Beta(\alpha + 1, n)$
2. Sampling α from the mixture $\pi(\alpha|\xi, k) \sim \pi_\xi \Gamma(a+k, b-\ln \xi)+(1-\pi_\xi)\Gamma(a+k-1, b-\ln \xi)$, where $\pi_\xi/(1 - \pi_\xi) = (a + k - 1)/[n(b - \ln \xi)]$.

References

Asai, M. and McAleer, M.: 2009, Multivariate stochastic volatility, leverage and news impact surfaces, *Econometric Journal* **12**, 292–309.

Basu, S. and Chib, S.: 2003, Marginal likelihood and Bayes factors for Dirichlet process mixture models, *Journal of the American Statistical Association* **98**(461), 224–235.

Bekaert, G. and Wu, G.: 2000, Asymmetric volatility and risk in equity markets, *Review of Financial Studies* **13**, 1–42.

Black, F.: 1976, Studies in stock price volatility changes, *Proceedings of the 1976 Meetings of the Business and Economics Statistics Section of the American Statistical Association* pp. 177–181.

Blackwell, D. and MacQueen, J.: 1973, Ferguson distributions via polya urn schemes, *The Annals of Statistics* **1**, 353–355.

Burda, M. and Prokhorov, A.: 2012, Copula based factorization in Bayesian multivariate mixture models, *Technical report*, University of Toronto, Department of Economics.

Campbell, J. Y. and Hentschel, L.: 1992, No news is good news: An asymmetric model of changing volatility in stock returns, *Journal of Financial Economics* **31**, 281–318.

Carvalho, C. M., Lopes, H. F., Polson, N. G. and Taddy, M. A.: 2010, Particle learning for general mixtures, *Bayesian Analysis* **5**(4), 709–740.

Chen, X. and Ghysels: 2011, News - good or bad - and its impact on volatility predictions over multiple horizons, *Review of Financial Studies* **24**, 46–81.

Chib, S. and Greenberg, E.: 1995, Understanding the Metropolis-Hastings algorithm, *American Statistician* **49**, 327–335.

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

Christie, A. A.: 1982, The stochastic behavior of common stock variances: Value, leverage and interest rate effects, *Journal of Financial Economics* **10**, 407–432.

Das, S. R. and Sundaram, R. K.: 1999, Of smiles and smirks: A term structure perspective, *Journal of Financial and Quantitave Analysis* **34**(2), 211–239.

Delatola, E.-I. and Griffin, J. E.: 2011a, Bayesian nonparametric modelling of the return distribution with stochastic volatilty, *Bayesian Analysis* **6**, 1–26.

Delatola, E.-I. and Griffin, J. E.: 2011b, A Bayesian semiparametric model for volatility with a leverage effect, *Technical report*, University of Kent.

Dickey, J. D.: 1971, The weighted likelihood ratio, linear hypotheses on normal location parameters, *Annals of Mathematical Statistics* **42**, 204–223.

Durham, G. B.: 2007, SV mixture models with application to S&P 500 index return, *Journal of Financial Economics* **85**, 822–856.

Escobar, M. D.: 1994, Estimating normal means with a Dirichlet process prior, *Journal of the American Statistical Association* **89**(425), 268–277.

Escobar, M. D. and West, M.: 1995, Bayesian density estimation and inference using mixtures, *Journal of the American Statistical Association* **90**(430), 577–588.

Ferguson, T.: 1973, A Bayesian analysis of some nonparametric problems, *The Annals of Statistics* **1**(2), 209–230.

Figlewski, S. and Wang, X.: 2000, Is the "leverage effect" a leverage effect?, *Technical report*, NYU Stern School of Business.

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

Gelfand, A. E. and Mukhopadhyay, S.: 1995, On nonparametric Bayesian inference for the distribution of a random sample, *The Canadian Journal of Statistics* **23**(4), 411–420.

Geweke, J.: 1992, Evaluating the accuracy of sampling based approaches to the calculation of posterior moments, *in* Bernardo, Berger, Dawid and Smith (eds), *Bayesian Statistics*, Vol. 4, Oxford: Clarendon Press.

Geweke, J.: 2001, Bayesian econometric and forecasting, *Journal of Econometrics* **100**, 11–15.

Geweke, J. and Amisano, G.: 2010, Comparing and evaluating bayesian predictive distributions of asset returns, *International Journal of Forecasting* **26**, 216–230.

Geweke, J. and Whiteman, C.: 2006, Bayesian forecasting, *in* G. Elliot, C. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*, Elsevier, Amsterdam.

Griffin, J. E.: 2011, Default priors for density estimation with mixture models, *Bayesian Analysis* **5**, 45–64.

Griffin, J. E. and Steel, M. F. J.: 2006, Ordered-based dependent Dirichlet processes, *Journal of the American Statistical Association* **101**, 179–194.

Griffin, J. E. and Steel, M. F. J.: 2011, Stick-breaking autoregressive processes, *Journal of Econometrics* **162**, 383–396.

Harvey, A., Ruiz, E. and Shephard, N.: 1994, Multivariate stochastic variance models, *The Review of Economic Studies* **61**(2), 247–264.

Jacquier, E., Polson, N. G. and Rossi, P. E.: 1994, Bayesian analysis of stochastic volatility models, *Journal of Business & Economic Statistics* **12**, 371–417.

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

Jefferies, H.: 1961, *Theory of Probability*, Claredon Press.

Jensen, M. J.: 2004, Semiparametric Bayesian inference of long-memory stochastic volatility models, *Journal of Time Series Analysis* **25**(6), 895–922.

Jensen, M. J. and Maheu, J. M.: 2010, Bayesian semiparametric stochastic volatility modeling, *Journal of Econometrics* **157**(2), 306 – 316.

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

Lo, A. Y.: 1984, On a class of Bayesian nonparametric estimates. i. density estimates, *The Annals of Statistics* **12**, 351–357.

MacEachern, S. N. and Müller, P.: 1998, Estimating mixture of Dirichlet process models, *Journal of Computational and Graphical Statistics* **7**(2), 223–238.

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

Omori, Y., Chib, S., Shephard, N. and Nakajima, J.: 2007, Stochastic volatility with leverage: Fast and efficient likelihood inference, *Journal of Econometrics* **140**(2), 425–449.

Percival, D. B. and Walden, A. T.: 1993, *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*, Cambridge University Press.

Poon, S. and Granger, C. W.: 2003, Forecasting volatility in financial markets: A review, *Journal of Economics Literature* **41**, 478–539.

Rey, M. and Roth, V.: 2012, Copula mixture model for dependency-seeking clustering, *Technical report*, Proceeding of the 29th International Conference on Machine Learning.

Schwert, G.: 1989, Why does stock market volatility change over time?, *Journal of Finance* **44**, 1115–1154.

Sethuraman, J.: 1994, A constructive definition of Dirichlet priors, *Statistica Sinica* **4**, 639–650.

West, M., Muller, P. and Escobar, M.: 1994, Hierarchical priors and mixture models with applications in regression and density estimation, *in* P. R. Freeman and A. F. Smith (eds), *Aspects of Uncertainty*, John Wiley.

Yu, J.: 2005, On leverage in a stochastic volatility model, *Journal of Econometrics* **127**(2), 165–178.

Yu, J.: 2011, A semiparametric stochastic volatility model, *Journal of Econometrics* **176**, 473–482.

Zellner, A.: 1971, *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York, New York.