Quantitative Finance

ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: www.tandfonline.com/journals/rquf20

Bayesian nonparametric modelling of stochastic volatility

Efthimios Nikolakopoulos

To cite this article: Efthimios Nikolakopoulos (10 Jun 2025): Bayesian nonparametric modelling of stochastic volatility, Quantitative Finance, DOI: 10.1080/14697688.2025.2509561

To link to this article: https://doi.org/10.1080/14697688.2025.2509561

Published online: 10 Jun 2025.

Quantitative Finance, 2025
https://doi.org/10.1080/14697688.2025.2509561

Bayesian nonparametric modelling of stochastic volatility

EFTHIMIOS NIKOLAKOPOULOS *

Essex Business School, University of Essex, Colchester, UK

(*Received 30 December 2024; accepted 17 May 2025*)

This paper introduces a novel discrete-time stochastic volatility model that employs a countably infinite mixture of AR(1) processes, with a Dirichlet process prior, to nonparametrically model the latent volatility. Realized variance (RV) is incorporated as an ex post signal to enhance volatility estimation. The model effectively captures fat tails and asymmetry in both return and log(RV) conditional distributions. Empirical analysis of three major stock indices provides strong evidence supporting the nonparametric stochastic volatility. The results reveal that the volatility equation components exhibit significant variation over time, enabling the estimation of a more dynamic volatility process that better accommodates extreme returns and variance shocks. The new model delivers out-of-sample density forecasts with strong evidence of improvement, particularly for returns, log(RV), and the left region of the return distribution, including negative returns and extreme movements below -1% and -2%. The new approach provides improvements in forecasting the tail-risk measures of value-at-risk and expected shortfall.

Keywords: Stochastic volatility; Realized variance; Bayesian nonparametrics; Dirichlet process mixture; Density forecasting

JEL Classification: C11, C14, C58

1. Introduction

This paper introduces a new discrete-time stochastic volatility model, where the volatility process is modelled nonparametrically using a Dirichlet process mixture (DPM). Traditional approaches to stochastic volatility often rely on a Gaussian AR(1) process for the latent log-volatility. This paper proposes modelling log-volatility with a countably infinite mixture of AR(1) processes, enabling greater flexibility in capturing complex volatility dynamics. Realized variance (RV) is incorporated as an ex post signal to enhance volatility estimation. The proposed model effectively accounts for non-Gaussian features in both return and log(RV) conditional distributions. Empirical analysis of major stock indices demonstrates the model's ability to capture the evolving dynamics of volatility. The findings reveal that the components of the stochastic volatility equation exhibit significant time variation, enabling the estimation of a more dynamic process that accommodates negative returns and sudden variance shocks. The new model provides superior out-of-sample density forecasts and favourable Bayes factors compared to existing models. In a tail-risk application, the proposed framework improves the forecasts of value-at-risk and expected shortfall.

Discrete-time stochastic volatility (SV) models, first introduced by Taylor (1982), extend GARCH models (Engle 1982, Bollerslev 1986) by treating volatility as a stochastic and latent process. These models conceptualize volatility as the impact of an unobserved news flow process. In standard SV frameworks, return distribution is modelled as a mixture of normals, with the latent log-volatility following an AR(1) process driven by normal innovations.

Several extensions to the standard SV model address non-Gaussian return features. For instance, Mahieu and Schotman (1998), Liesenfeld and Jung (2000), Chib *et al.* (2002), Jacquier *et al.* (2004), Abanto-Valle *et al.* (2010) and, Nakajima and Omori (2012), introduce models with fat-tailed return distributions. Bayesian semiparametric approaches by Jensen and Maheu (2010), Delatola and Griffin (2011), Delatola and Griffin (2013), Jensen and Maheu (2014), Liu (2021) and Li *et al.* (2024), among others, capture in different ways asymmetry and fat tails in the return distribution. Realized variance has also been integrated into volatility estimation, as demonstrated by Takahashi *et al.* (2009), among

*Corresponding author. Email: e.nikolakopoulos@essex.ac.uk

others.† Another line of research explores regime-switching volatility models, as seen in So *et al.* (1998), Kalimipalli and Susmel (2004), Carvalho and Lopes (2007), Vo (2009), Virbickaitė and Lopes (2019), which challenge the restrictive Gaussian AR(1) assumption.

This paper adopts a Bayesian nonparametric framework for stochastic volatility modelling. The volatility equation components–namely the level, the AR(1) coefficient, and the variance–are modelled using a DPM (Ferguson 1973). This approach results in a countably infinite mixture of Gaussian AR(1) processes, enabling a highly dynamic volatility structure that captures variance shocks, jumps, and clustering. We incorporate RV in the proposed model to enhance volatility estimation.‡ We demonstrate that our new framework results in an infinite mixture of normals in log(RV) modelling, allowing for a non-Gaussian conditional distribution. We additionally include the return mean within the mixture framework to introduce asymmetry into the return conditional distribution, complementing the model's ability to handle fat tails through nonparametric variance modelling. The DPM framework learns from the data and adaptively uses the necessary number of clusters to approximate the underlying data distribution.

Our new model framework, to the best of our knowledge, is the first attempt to model the stochastic volatility nonparametrically. A related approach is found in the work of Dufays (2016), which applies a similar framework to GARCH models. Also, the proposed model extends multiple existing parametric and semiparametric approaches. Specifically, it directly builds on the SV models of Takahashi *et al.* (2009), Jensen and Maheu (2010), and Delatola and Griffin (2011), while also nesting the model of Liu (2021).

Empirical analysis of major stock indices, including the Dow Jones, Nasdaq 100, and S&P 500, demonstrates the proposed model's ability to capture the evolving dynamics of volatility. The volatility equation components form 4.5 – 5.8 clusters, on average. Results indicate that time-invariant parameters in SV modelling is a restrictive specification. All the SV equation components vary substantially through time. We find a trade-off between the AR(1) coefficient and the variance of log-volatility. The proposed model flexibly chooses between persistence and randomness to adequately capture the volatility dynamics. Out-of-sample predictive density plots, show clear non-Gaussian shapes, with asymmetry and fat tails, for the return, log(RV) and log-volatility densities.

Out-of-sample density forecasts provide strong evidence supporting the nonparametric stochastic volatility. Compared to widely used SV models, the proposed framework achieves superior predictive likelihood scores and Bayes factors for both return and log(RV) density forecasts. The flexible volatility modelling provides the best description of the underlying data generating process. Furthermore, the proposed framework delivers the best forecasts for specific regions of the return distribution, particularly for negative returns and losses exceeding 1% and 2%. In a tail-risk forecasting application, the value-at-risk and expected shortfall measures generated by the proposed framework deliver the best scores in the loss function of Taylor (2019), highlighting the framework's potential for risk measurement applications.

The paper is organized as follows. Section 2 is a brief discussion on RV and the benchmark model of Takahashi *et al.* (2009). Section 3 presents the proposed model, its estimation algorithm and the forecasting process. Section 4 is a brief simulation exercise. Section 5 is the empirical application. Section 6 concludes. An Appendix details the model estimation algorithm.

2. Realized stochastic volatility

We start this section with a brief discussion on ex post variance measures. The ex post variance of an asset's return is estimated with the nonparametric realized measures. In these, high-frequency (e.g. intraday) returns are used to estimate latent low-frequency (e.g. daily) variance. RV is the simplest realized measure. The initial assumption is that the logarithmic price of an asset during a period (day) t follows a continuous-time stochastic volatility process.§ In practice, for a trading day t, there are n intraday logarithmic returns $r_{t,i},\ i = 1,\dots,n$. Barndorff-Nielsen and Shephard (2002a) and Andersen *et al.* (2003) show that the quadratic returns variation at t, QV_t, can be approximated by the RV_t estimator which in its basic form is the summation of the intraday squared returns, $\mathrm{RV}_t = \sum_{i=1}^{n} r_{t,i}^2$. RV_t is a consistent ex post estimator of QV_t, under no market microstructure noise.¶

Takahashi *et al.* (2009) use the ex post variance as a second information source, alongside returns, in the stochastic volatility model. Their model here is referred to as RSV-N and conditional on the information set $\mathcal{I}_{t-1} = \{r_1, \mathrm{RV}_1, \dots, r_{t-1}, \mathrm{RV}_{t-1}\}$ is defined as

$$r_t = \mu + \exp(h_t/2)u_t, \quad u_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1), \tag{1}$$

$$\log(\mathrm{RV}_t) = \psi + h_t + \xi z_t, \quad z_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1), \tag{2}$$

$$h_t = \gamma + \delta h_{t-1} + \sigma e_t, \quad e_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1), \tag{3}$$

with model parameters $\theta = \{\mu, \psi, \xi^2, \gamma, \delta, \sigma^2\}$. r_t is the log return at time $t, t = 1,\dots,T$ and $\log(\mathrm{RV}_t)$ denotes the natural logarithm of realized variance at time t. h_t is the return's stochastic log-volatility at time t. Equations (1) and (2) are the measurement equations. The basic stochastic volatility model, referred to as SV-N, without RV, is defined with (1) and (3). In (2) $\log(\mathrm{RV}_t)$ is used as extra informational signal to assist the estimation of stochastic volatility. Parameter ψ captures the deviation between $\log(\mathrm{RV}_t)$ and h_t.

† Ex post variance measures have been used in various parametric frameworks, including GARCH (Hansen *et al.* 2012), asymmetric SV (Zhang and Zhao 2023), and multivariate SV models (Shirota *et al.* 2017, Yamauchi and Omori 2020).

‡ We only consider RV but any other ex post variance measure can be used as well.

§ For the theoretical foundation and applications of RV see Andersen *et al.* (2001a, 2001b), Barndorff-Nielsen and Shephard (2002a, 2002b), and Andersen *et al.* (2003).

¶ See Zhang *et al.* (2005) and Aït-Sahalia and Mancini (2008) for the use of *subsampling* for realized measures robust to noise and Barndorff-Nielsen *et al.* (2008) for the use of realized kernel measures robust to market frictions.

Following the standard approach in literature, the RSV-N model can be estimated using Bayesian analysis and with a hybrid of Gibbs and Metropolis-Hastings sampling steps.

3. Bayesian nonparametric stochastic volatility

The standard literature approach for stochastic volatility modelling is an AR(1) specification, as in (3). This is a restrictive assumption as shown by the the parametric extensions of So *et al.* (1998), Carvalho and Lopes (2007), Virbickaitė and Lopes (2019) and the semiparametric model of Jensen and Maheu (2014). We are building up on these extensions and we develop a SV model in which all the stochastic volatility equation components are time-varying. We nonparametrically model the latent volatility with the DPM. Thus we model stochastic volatility with a countably infinite mixture of Gaussian AR(1) processes instead of just one.

We extend the RSV-N model of Takahashi *et al.* (2009) with a DPM in the latent volatility equation. The ex post variance assists the latent volatility estimation under the nonparametric modelling. Conditional on the information set $\mathcal{I}_{t-1}$ the proposed model referred to as DPM-SV (Dirichlet process mixture in stochastic volatility) has the following hierarchical specification

$$r_t = m_t + \exp(h_t/2)u_t, \quad u_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1), \tag{4}$$

$$\log(\mathrm{RV}_t) = \psi + h_t + \xi z_t, \quad z_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1), \tag{5}$$

$$h_t = g_t + d_t h_{t-1} + v_t e_t, \quad e_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1), \tag{6}$$

$$m_t, g_t, d_t, v_t^2 | G \overset{\text{iid}}{\sim} G, \tag{7}$$

$$G \mid G_0, \alpha \sim \mathrm{DP}(\alpha, G_0), \tag{8}$$

$$G_0(m_t, g_t, d_t, v_t^2) \equiv \mathrm{N}(m_0, v_m^2) - \mathrm{N}(g_0, v_g^2) - \mathrm{N}(d_0, v_d^2)\mathbf{1}_{|d_t|<1} - \mathrm{IG}(v_0, s_0). \tag{9}$$

Model parameters are $\theta = \{\psi, \xi^2, \alpha\}$. Equations (7)–(9) place an infinite mixture of normals in returns and log-volatility. The mixing components are: the return mean m_t, the stochastic volatility level g_t, the AR(1) coefficient d_t and the volatility of volatility v_t. For identification we do not include parameters ψ and ξ^2 in the mixture.

The mixing components are distributed according to the latent G which is nonparametrically modelled with a Dirichlet process (DP) prior. A draw from a DP, $G \sim \mathrm{DP}(\alpha, G_0)$, is almost surely a discrete distribution and has two parameters, the base measure G_0 and the concentration parameter $\alpha > 0$. The DP is centred around G_0 since $\mathbb{E}[G] = G_0$ and α determines how close is G to G_0 since $\mathrm{Var}[G] = G_0[1 - G_0]/(\alpha + 1)$. In this case the base measure of the DP in (9) consists of three normal priors, N(.), for m_t, g_t and d_t, and an inverse gamma, IG(.), for v_t^2.

G has a support of infinite distributions. Let $\mu = \{\mu_1, \mu_2, \ldots\}$, $\gamma = \{\gamma_1, \gamma_2, \ldots\}$, $\delta = \{\delta_1, \delta_2, \ldots\}$ and $\sigma^2 = \{\sigma_1^2, \sigma_2^2, \ldots\}$ denote the unique points of support in G with base measure prior's

$$\mu_j \overset{\text{iid}}{\sim} \mathrm{N}(m_0, v_m^2), \quad \gamma_j \overset{\text{iid}}{\sim} \mathrm{N}(g_0, v_g^2), \quad \delta_j \overset{\text{iid}}{\sim} \mathrm{N}(d_0, v_d^2)\mathbf{1}_{|\delta_j|<1},$$
$$\sigma_j^2 \overset{\text{iid}}{\sim} \mathrm{IG}(v_0, s_0), \tag{10}$$

and $j = 1, 2, \ldots$. In practice, a finite set $\{(r_1, \mathrm{RV}_1, h_1), \ldots, (r_T, \mathrm{RV}_T, h_T)\}$ will be associated with a finite set $\{(m_1, g_1, d_1, v_1^2), \ldots, (m_T, g_T, d_T, v_T^2)\}$ of draws from G in (7). The DPM allows data clustering in identical sets of (m_t, g_t, d_t, v_t^2). This truncates the theoretically infinite mixture into a practical finite mixture with k unique clusters, $\{\mu_j, \gamma_j, \delta_j, \sigma_j^2\}_{j=1}^k$, $k < T$. The DPM concentration parameter α controls the number of mixture clusters. Estimating α in a Bayesian fashion makes the DPM cluster consistent (Ascolani *et al.* 2023). The DPM-SV model has the flexibility to capture non-Gaussian behaviour in both returns and $\log(\mathrm{RV}_t)$. It also nests all parametric specifications. For instance, when $\alpha \to 0$ then all data are assigned to one cluster $\{\mu_1, \gamma_1, \delta_1, \sigma_1^2\}$ and stochastic volatility follows a Gaussian AR(1). In this case the DPM-SV is equivalent to RSV-N. If $\alpha \to \infty$ then we have as many mixing clusters as data points. If on top of that the mixing clusters have an identical $\mu_j = \mu$, then returns follow a Student's t-distribution. The model learns from the data and adaptively uses the necessary number of clusters to approximate the underlying data distribution.

Typically, finite mixture models face a label-switching issue, which necessitates parameter constraints to address. This issue, however, does not arise in DPM models. In these, the mixture components are primarily a tool for approximating the underlying data distribution, and the number of clusters often lacks intrinsic interpretability beyond representing a flexible mixture of normals. In our case, we impose the mixture mainly on a latent variable. Consequently, the number of clusters primarily provides flexibility in modelling volatility, allowing the model to better capture latent dynamics and accommodate shocks.

The DPM-SV model can also be written with the Sethuraman (1994) stick-breaking specification as

$$r_t = \mu_{s_t} + \exp(h_t/2)u_t, \quad u_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1), \tag{11}$$

$$\log(\mathrm{RV}_t) = \psi + h_t + \xi z_t, \quad z_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1), \tag{12}$$

$$h_t = \gamma_{s_t} + \delta_{s_t} h_{t-1} + \sigma_{s_t} e_t, \quad e_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1), \tag{13}$$

$$s_t \sim \mathrm{Multinominal}(w), \quad w = \{w_1, w_2, \ldots\}, \tag{14}$$

$$w_j = v_j \prod_{l=1}^{j-1} (1 - v_l), \quad v_j \overset{\text{iid}}{\sim} \mathrm{Beta}(1, \alpha), \tag{15}$$

and base measure (10). w is the infinite set of weights associated with the mixing clusters, with the stick-breaking prior as in (15), $w_1 = v_1$ and $\sum_{j=1}^{\infty} w_j = 1$. $s_t = 1, 2, \ldots,$ is a cluster (or state) indicator auxiliary variable which maps each observation set (r_t, h_t) to a cluster j. The above representation shows clearly the model's assumption that there is a latent state variable governing the return and volatility dynamics.

The mixture conditional distribution of stochastic volatility is,

$$p(h_t \mid \mathcal{I}_{t-1}, w, \gamma, \delta, \sigma^2) = \sum_{j=1}^{\infty} w_j \mathrm{N}\left(h_t \mid \gamma_j + \delta_j h_{t-1}, \sigma_j^2\right). \tag{16}$$

Conditional on h_t the return conditional distribution has the following stick-breaking representation

$$p(r_t \mid \mathcal{I}_{t-1}, w, h_t, \mu) = \sum_{j=1}^{\infty} w_j \mathrm{N}\left(r_t \mid \mu_j, \exp(h_t)\right). \tag{17}$$

The DPM-SV model has the ability, through h_t, to capture non-Gaussian effects in the conditional distribution of $\log(\mathrm{RV})$. Conditional on h_t, the $\log(\mathrm{RV})$ conditional distribution has the following stick-breaking representation

$$\begin{aligned} &p(\log(\mathrm{RV}_t) \mid \mathcal{I}_{t-1}, w, h_t, \gamma, \delta, \sigma^2) \\ &= \sum_{j=1}^{\infty} w_j \mathrm{N}\left(\log(\mathrm{RV}_t) \mid \psi + \gamma_j + \delta_j h_{t-1}, \xi^2 + \sigma_j^2\right). \end{aligned} \tag{18}$$

We also estimate a restricted version of the DPM-SV model with parametric return mean, $\mu_j = \mu, \forall j$. In this restricted version the latent state variable governs only the stochastic volatility dynamics. This enables a straightforward comparison with RSV-N to examine the impact of nonparametric modelling of stochastic volatility. We refer to this model as DPM-SV-h and it is specified with equations (1), (2), (13)–(15) and base measure $G_0(\gamma_j) \overset{\text{iid}}{\sim} \mathrm{N}(g_0, v_g^2), G_0(\delta_j) \overset{\text{iid}}{\sim} \mathrm{N}(d_0, v_d^2)\mathbf{1}_{|\delta_j|<1}, G_0(\sigma_j^2) \overset{\text{iid}}{\sim} \mathrm{IG}(v_0, s_0), \forall j$. The DPM-SV-h model captures fat tails in return distribution but lacks the ability to account for asymmetry. However, it effectively models non-Gaussian effects, namely asymmetry and fat tails, in $\log(\mathrm{RV})$.

3.1. DPM-SV model estimation

In this section, we discuss the estimation algorithm for the DPM-SV model, which is newly proposed. This algorithm can be easily modified to estimate the DPM-SV-h model.

For the DPM model we use the stick-breaking formulation and the slice sampler by Walker (2007) and Kalli *et al.* (2011), which truncates the infinite mixture into a finite number κ, $\kappa \le k < T$, of unique clusters, $\{\mu_j, \gamma_j, \delta_j, \sigma_j^2\}_{j=1}^{\kappa}$, with at least one data observation assigned in each cluster. To do so, the parameter space is expanded by introducing two latent vectors. The first one, already discussed, is the cluster indicator $s_{1:T}$. The second auxiliary vector is $u_{1:T} = \{u_1, \dots u_T\}$, with $u_t \in (0, 1)$, which helps to convert the infinite sum in (16), (17) and (18) into a finite mixture.

The joint posterior of the DPM-SV model $p(\{\mu_j, \gamma_j, \delta_j, \sigma_j^2\}_{j=1}^{\infty}, s_{1:T}, u_{1:T}, h_{1:T}, \theta \mid \mathcal{I}_T)$ is proportional to

$$\begin{aligned} &p(\theta)p(w_{1:k}) \prod_{j=1}^{k} p(\mu_j, \gamma_j, \delta_j, \sigma_j^2) \\ &\times \prod_{t=1}^{T} \mathbf{1}_{u_t < w_{s_t}} \mathrm{N}\left(r_t | \mu_{s_t}, \exp(h_t)\right) \mathrm{N}\left(\log(\mathrm{RV}_t) | \psi + h_t, \xi^2\right) \\ &\mathrm{N}\left(h_t | \gamma_{s_t} + \delta_{s_t} h_{t-1}, \sigma_{s_t}^2\right). \end{aligned} \tag{19}$$

with the first line being the prior and the second one the likelihood. $\mathbf{1}$ is the indicator function and k is the smallest positive integer that satisfies the condition $\sum_{j=1}^{k} w_j > 1 - \min(u_{1:T})$. The above posterior does not have a known form. We follow standard Markov chain Monte Carlo (MCMC) techniques and a hybrid algorithm of Gibbs and Metropilis-Hastings steps to sample from a series of conditional distributions.

After giving starting values to $\theta, k, w_{1:k}, s_{1:T}, \mu_{1:k}, \gamma_{1:k}, \delta_{1:k}, \sigma_{1:k}^2, \alpha$ and $h_{1:T}$, we collect a large number of posterior draws $\{\theta^{(i)}, k^{(i)}, s_{1:T}^{(i)}, \mu_{1:k}^{(i)}, \gamma_{1:k}^{(i)}, \delta_{1:k}^{(i)}, \sigma_{1:k}^{2(i)}, \alpha^{(i)}, h_{1:T}^{(i)}\}_{i=1}^{R}$ by iterating through the following MCMC steps:

(1) Sample h_t from $p(h_t | h_{-t}, \mathcal{I}_t, \theta, \mu_{1:k}, \gamma_{1:k}, \delta_{1:k}, \sigma_{1:k}^2, s_{1:T})$, $t = 1, \dots, T$.†
(2) Sample (a) ψ from $p(\psi \mid \mathrm{RV}_{1:T}, h_{1:T}, \xi^2)$ and (b) ξ^2 from $p(\xi^2 \mid \mathrm{RV}_{1:T}, h_{1:T}, \psi)$.
(3) Sample (a) $\mu_{1:k}$ from $p(\mu_{1:k} | r_{1:T}, h_{1:T}, s_{1:T})$, (b) $\gamma_{1:k}$ from $p(\gamma_{1:k} | r_{1:T}, h_{1:T}, \delta_{1:k}, \sigma_{1:k}^2, s_{1:T})$ (c) $\delta_{1:k}$ from $p(\delta_{1:k} | r_{1:T}, h_{1:T}, \gamma_{1:k}, \sigma_{1:k}^2, s_{1:T})$ and (d) $\sigma_{1:k}^2$ from $p(\sigma_{1:k}^2 | r_{1:T}, h_{1:T}, \gamma_{1:k}, \delta_{1:k}, s_{1:T})$.
(4) Update $w_{1:k}, u_{1:T}, k | s_{1:T}, \alpha$.
(5) Sample s_t from $p(s_t | \mathcal{I}_t, h_{1:T}, \theta, \mu_{1:k}, \gamma_{1:k}, \delta_{1:k}, \sigma_{1:k}^2, w_{1:k}, u_{1:T}, k)$, $t = 1, \dots, T$.
(6) Sample α from $p(\alpha | \kappa, T)$.

See the Appendix for details.

3.2. Forecasting

The interest of Bayesian nonparametrics forecasting is on the predictive density. This is approximated in the DPM framework by first integrating out the uncertainty about the future cluster of the mixture parameters. Conditional on $\mathcal{I}_t = \{r_{1:t}, \mathrm{RV}_{1:t}\}$, the predictive densities of return, $\log(\mathrm{RV})$ and stochastic volatility, for the DPM-SV model, can be approximated with the use of R posterior draws as

$$p(r_{t+1} \mid \mathcal{I}_t, s_{t+1}, h_{t+1}) \approx \frac{1}{R} \sum_{i=1}^{R} \mathrm{N}\left(r_{t+1} \,\middle|\, \mu_{s_{t+1}^{(i)}}^{(i)}, \exp\left(h_{t+1}^{(i)}\right)\right), \tag{20}$$

$$\begin{aligned} &p(\log(\mathrm{RV}_{t+1}) \mid \mathcal{I}_t, s_{t+1}, h_{t+1}) \\ &\approx \frac{1}{R} \sum_{i=1}^{R} \mathrm{N}\left(\log(\mathrm{RV}_{t+1}) \,\middle|\, \psi^{(i)} + h_{t+1}^{(i)}, \xi^{2(i)}\right), \end{aligned} \tag{21}$$

† A popular approach in the standard SV models is sampling $h_{1:T}$ in random blocks (Fleming and Kirby 2003). This would be challenging to use since we do not have fixed SV equation parameters. Instead we use an extension of the single-move SV sampler from Kim *et al.* (1998). Also, the path dependence issue observed in regime-switching GARCH models (e.g. Bauwens *et al.* 2010) is not found in the SV model literature (e.g. So *et al.* 1998, Carvalho and Lopes 2007) since the latent volatility is not a deterministic function of its previous values.

$$p(h_{t+1}|\mathcal{I}_t, s_{t+1}) \approx \frac{1}{R}\sum_{i=1}^{R} \mathrm{N}\left(h_{t+1}\left|\gamma^{(i)}_{s^{(i)}_{t+1}} + \delta^{(i)}_{s^{(i)}_{t+1}} h^{(i)}_t, \sigma^{2(i)}_{s^{(i)}_{t+1}}\right.\right), \tag{22}$$

$$\text{where} \quad s^{(i)}_{t+1} = \begin{cases} j, & \text{if } \sum_{l=0}^{j-1} w^{(i)}_l < \phi < \sum_{l=0}^{j} w^{(i)}_l, \\ k^{(i)}+1, & \text{if } \phi \geq \sum_{l=0}^{k^{(i)}} w^{(i)}_l, \end{cases} \tag{23}$$

with $w^{(i)}_o = 0, j \leq k^{(i)}$, $\phi \sim \mathrm{U}(0,1)$ and future volatility conditional on $s^{(i)}_{t+1}$ is simulated as

$$h^{(i)}_{t+1} \mid s^{(i)}_{t+1}, \mathcal{I}_t \sim \mathrm{N}\left(\gamma^{(i)}_{s^{(i)}_{t+1}} + \delta^{(i)}_{s^{(i)}_{t+1}} h^{(i)}_t, \sigma^{2(i)}_{s^{(i)}_{t+1}}\right). \tag{24}$$

The future value of $s^{(i)}_{t+1}$ in (23) is one of the existing clusters with probability equal to the associated weights and there is a nonzero probability of introducing a new cluster $(\mu^{(i)}_{k^{(i)}+1}, \gamma^{(i)}_{k^{(i)}+1}, \delta^{(i)}_{k^{(i)}+1}, \sigma^{2(i)}_{k^{(i)}+1})$ from the base measure G_0.

We can also approximate the joint return and log(RV) predictive density. Conditional on s_{t+1} and h_{t+1}, returns and log(RV) have independent conditional distributions and their joint predictive density can be approximated as

$$\begin{aligned} &p(r_{t+1}, \log(\mathrm{RV}_{t+1}) \mid \mathcal{I}_t, s_{t+1}, h_{t+1}) \\ &\approx \frac{1}{R}\sum_{i=1}^{R} \mathrm{N}\left(r_{t+1}\left|\mu^{(i)}_{s^{(i)}_{t+1}}, \exp\left(h^{(i)}_{t+1}\right)\right.\right) \\ &\quad \mathrm{N}\left(\log(\mathrm{RV}_{t+1})\left|\psi^{(i)} + h^{(i)}_{t+1}, \xi^{2(i)}\right.\right). \end{aligned} \tag{25}$$

The predictive density serves as the building block for the predictive likelihood (PL) (Geweke 1994). This measure provides an *out-of-sample* density forecast record that facilitates straightforward model comparisons. The log-predictive likelihood (log-PL) for a τ-length vector of returns, $r_{T-\tau+1:T}$ (with $1 < \tau < T$), is the summation of individual log predictive densities evaluated at the r_{t+1} data as

$$\log-\mathrm{PL}(r_{T-\tau+1:T} \mid \mathcal{I}_T) = \sum_{t=T-\tau}^{T-1} \log\left(p(r_{t+1} \mid \mathcal{I}_t, s_{t+1}, h_{t+1})\right). \tag{26}$$

Similarly, the log predictive likelihood for a τ-length vector of realized variances, $\log(\mathrm{RV})_{T-\tau+1:T}$ is calculated as

$$\begin{aligned} &\log-\mathrm{PL}(\log(\mathrm{RV})_{T-\tau+1:T} \mid \mathcal{I}_T) \\ &= \sum_{t=T-\tau}^{T-1} \log\left(p(\log(\mathrm{RV}_{t+1}) \mid \mathcal{I}_t, s_{t+1}, h_{t+1})\right), \end{aligned} \tag{27}$$

and the joint return, log(RV) predictive likelihood is calculated as

$$\begin{aligned} &\log-\mathrm{PL}(r_{T-\tau+1:T}, \log(\mathrm{RV})_{T-\tau+1:T} \mid \mathcal{I}_T) \\ &= \sum_{t=T-\tau}^{T-1} \log\left(p(r_{t+1}, \log(\mathrm{RV}_{t+1}) \mid \mathcal{I}_t, s_{t+1}, h_{t+1})\right). \end{aligned} \tag{28}$$

The *out-of-sample* density forecasting exercise is done with recursive posterior model estimations and likelihood evaluations.

In a two-model comparison, the log-Bayes factor (log-BF) is the difference between the log-predictive likelihoods of the two models. As noted by Kass and Raftery (1995), a higher log-BF indicates preference for one model over the other, with values greater than 3 indicating strong preference and very strong preference for values greater than 5. The Bayes factor measure favours complicated model specifications only when they provide better data density explanation.

In many finance applications, the objective is to forecast specific regions of a distribution. For example, in risk measurement, the focus is often on the left tail of the return distribution. We evaluate the model tail forecasting with the measure of tail predictive density (Diks *et al.* 2011). Following Jensen and Maheu (2013), the region predictive density of $r_{t+1} < \eta$, $\eta \in \mathbb{R}$, is defined as

$$\begin{aligned} &p(r_{t+1} \mid r_{t+1} < \eta, \mathcal{I}_t, s_{t+1}, h_{t+1}) \\ &= \frac{p(r_{t+1} \mid \mathcal{I}_t, s_{t+1}, h_{t+1}) \mathbf{1}_{r_{t+1}<\eta}}{\int_{-\infty}^{\eta} p(y_{t+1} \mid \mathcal{I}_t, s_{t+1}, h_{t+1})\, \mathrm{d}y_{t+1}} \\ &\approx \frac{\frac{1}{R}\sum_{i=1}^{R} \mathrm{N}\left(r_{t+1}\left|\mu^{(i)}_{s^{(i)}_{t+1}}, \exp\left(h^{(i)}_{t+1}\right)\right.\right) \mathbf{1}_{r_{t+1}<\eta}}{\frac{1}{R}\sum_{i=1}^{R} \Phi\left(\left(\eta - \mu^{(i)}_{s^{(i)}_{t+1}}\right) / \exp\left(h^{(i)}_{t+1}/2\right)\right)}, \end{aligned} \tag{29}$$

where $\Phi(.)$ denotes the standard Gaussian c.d.f. The denominator in (29) is an integrating constant ensuring that the predictive density integrates to one. The region log-predictive likelihood can be calculated as the log-PL in (26).

3.3. Benchmark models

We compare our proposed models DPM-SV and DPM-SV-h with popular SV models. These are:

(1) The standard Gaussian SV-N model (Kim *et al.* 1998), defined with equations (1) and (3).

(2) The SV model with fat tails of Jacquier *et al.* (2004) which is referred to as SV-t and defined as

$$r_t = \mu + \exp(h_t/2)\lambda_t^{1/2} u_t, \quad u_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1) \tag{30}$$

$$\lambda_t \sim \mathrm{IG}(\nu/2, \nu/2), \tag{31}$$

and volatility equation (3). Conditional on λ_t, returns follow a Student's *t*-distribution with ν degrees of freedom.

(3) The SV-DPM of Jensen and Maheu (2010) which uses a DPM in the return distribution while having a Gaussian AR(1) stochastic volatility. This model has the following stick-breaking form

$$r_t = \mu_{s_t} + \lambda_{s_t} \exp(h_t/2) u_t, \quad u_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1), \tag{32}$$

$$h_t = \delta h_{t-1} + \sigma_v e_t, \quad e_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1), \tag{33}$$

with s_t following a Multinominal distribution as in (14), stick-breaking mixture weights as in (15) and base measure $G_0(\mu_j, \lambda_j^2) \equiv \mathrm{N}(m_0, v_m^2) - \mathrm{IG}(v_l, s_l), \forall j$.

(4) The RSV-N model discussed in Section 2.
(5) The final benchmark model is the RSV-DPM of Liu (2021) which has the following stick-breaking representation

$$r_t = \mu_{s_t} + \lambda_{s_t} \exp(h_t/2) u_t, \quad u_t \overset{\text{iid}}{\sim} \mathrm{N}(0, 1), \tag{34}$$

$$\log(\mathrm{RV}_t) = \psi_{s_t} + h_t + \xi_{s_t} z_t, \quad z_t \overset{\text{iid}}{\sim} \mathrm{N}(0, 1), \tag{35}$$

$$h_t = \gamma + \delta h_{t-1} + \sigma_v e_t, \quad e_t \overset{\text{iid}}{\sim} \mathrm{N}(0, 1), \tag{36}$$

with cluster variable s_t as in (14), mixture weights as in (15) and base measure $G_0(\mu_j, \lambda_j^2, \psi_j, \xi_j^2) \equiv \mathrm{N}(m_0, v_m^2) - \mathrm{IG}(v_l, s_l) - \mathrm{N}(c_0, v_c^2) - \mathrm{IG}(v_0, s_0), \forall j$. The difference of this model compared to our proposed DPM-SV(-h) lies in the stochastic volatility equation. Here it is fully parametric with fixed intercept, persistence and volatility of volatility parameters. The DPM-SV framework flexibly models these latent volatility dynamics with time-varying components.

4. Simulation exercise

In this section we perform a simulation exercise to illustrate the models estimation. We generate $T = 2000$ return and $\log(\mathrm{RV})$ values from

$$r_t \sim \mathrm{N}(0, \exp(h_t)),$$
$$\log(\mathrm{RV}_t) \sim \mathrm{N}(0.1 + h_t, 0.1).$$

We test two different data generating processes for the log-volatility. The first one is a standard 1-state Gaussian volatility, generated as

$$h_t \sim \mathrm{N}(-0.1 + 0.95 h_{t-1}, 0.05).$$

The second volatility generating process is the following 2-states mixture of normals

$$h_t \sim \begin{cases} \mathrm{N}(0.95 h_{t-1}, 0.05), & \text{w.p. } 0.9 \\ \mathrm{N}(-0.1 + 0.80 h_{t-1}, 0.2), & \text{w.p. } 0.1. \end{cases}$$

In DPM-SV we set the priors as: $\mu_j \sim \mathrm{N}(0, 0.1)$, $\gamma_j \sim \mathrm{N}(0, 0.1)$, $\delta_j \sim \mathrm{N}(0.9, 0.1)\mathbf{1}_{|\delta_j|<1}$ and $\sigma_j^2 \sim \mathrm{IG}(10/2, 1/2)$, with $\mathbb{E}(\sigma_j^2) = 1/8$, $j = 1, 2, \ldots$. The precision parameter of DPM has a Gamma prior, $\alpha \sim \Gamma(2, 10)$, with $\mathbb{E}(\alpha) = 0.2$. For the $\log(\mathrm{RV})$ measurement equation we set $\psi \sim \mathrm{N}(0, 1)$ and $\xi^2 \sim \mathrm{IG}(10/2, 1/2)$. For the rest of the models we use the above priors in their associated parameters.

Table 1 displays the estimation results from models RSV-N, RSV-DPM, DPM-SV and DPM-SV-h for the two data generating processes. Overall, the posterior mean and median values are very close to the true ones. The 95% density intervals for most of the parameters contain the true value. The DPM-SV(-h) models use an average number of clusters close to the true number. RSV-DPM uses roughly the same number of clusters for both cases. In the 2-states case, the latent volatility dynamics cannot be detected by the RSV-DPM specification. The RSV-N and RSV-DPM stochastic volatility equation parameters are close to a weighted average of the 2-states true values. A similar trend is also observed in the models empirical application that follows.

5. Empirical application

5.1. Data

We consider three datasets consisting of daily open-to-close† (log) returns and their ex post measures of realized variance from Oxford-Man Institute's Realized Library‡ (Heber *et al.* 2009). The datasets are: Dow Jones Industrial Average Index (DJI), Nasdaq 100 (IXIC) and S&P 500 Index (SPX). Summary statistics are in table 2. Returns have been converted to percentages while RV has been scaled by 100^2.

5.2. Posterior estimation

Table 3 reports the posterior estimation results for the three datasets. The priors used are the same as in Section 4. Results are from 30,000 posterior draws, after discarding 20,000 ones. We report the posterior mean and 95% posterior density interval.

The proposed DPM-SV and DPM-SV-h models use a mixture of normals in stochastic volatility modelling. On average the SV equation mixing components form 4.5–5.8 clusters. The DPM precision parameter α is estimated in the range 0.3405–0.4277 and its trace plots in figure 1 show good convergence. The benchmark RSV-DPM model uses less than 3 clusters on average.

In the $\log(\mathrm{RV})$ measurement equations, the parameters from the proposed DPM-SV-h model are close to the benchmark RSV-N. In the more flexible DPM-SV, for the cases of IXIC and SPX, variance parameter ξ^2 is lower in value compared to the other two models and parameter ψ is closer to zero. This indicates less noise between $\log(\mathrm{RV})$ and the estimated volatility.

Figure 2 displays the DPM-SV posterior mixture components and smooth volatility against the associated RSV-N parameters and volatility. The estimated log-volatilities from both models have similar patters but, the DPM-SV one shows more flexibility in the extreme low and high volatility days.

All the SV equation components vary substantially through time. Time-invariant log-volatility parameters is a restrictive specification. Among the series the patterns are similar. γ_{s_t} fluctuates around the parametric γ and around zero with values ranging from -0.7 to 1. δ_{s_t} is mostly close to 0.95 and δ but, often adjusts to lower values, as low as 0.65. $\sigma_{s_t}^2$ is mostly below 0.05 and σ^2 from SV-N and often spikes to values larger than σ^2 from RSV-N.

There is an observed trade-off between δ_{s_t} and $\sigma_{s_t}^2$ in figure 2. Table 4 reports the correlations of the posterior estimated volatility mixing components. The correlation between δ_{s_t}

† Following a big part of the relevant literature we only consider trading hours returns but overnight returns can be included as well.
‡ The Oxford-Man Realized Library was discontinued in 2022.

Table 1. Model estimation results on simulated data.

	True	RSV-N			RSV-DPM			DPM-SV-h			DPM-SV		
		Mean	Median	95% D.I.	Mean	Median	95% D.I.	Mean	Median	95% D.I.	Mean	Median	95% D.I.
Panel A: 1 state													
μ	0.00	0.006	0.006	[− 0.008, − 0.003]				0.017	0.017	[0.014, 0.019]			
ψ	0.10	0.110	0.110	[0.044, 0.174]				0.142	0.142	[0.077, 0.206]	0.142	0.142	[0.080, 0.207]
ξ^2	0.10	0.110	0.110	[0.100, 0.121]				0.097	0.097	[0.088, 0.107]	0.097	0.097	[0.088, 0.107]
γ	− 0.10	− 0.101	− 0.100	[− 0.131, − 0.099]	− 0.094	− 0.094	[− 0.126, − 0.065]						
δ	0.95	0.950	0.950	[0.934, 0.965]	0.952	0.952	[0.936, 0.966]						
σ^2	0.05	0.052	0.052	[0.044, 0.062]	0.050	0.050	[0.042, 0.059]						
α					0.144	0.118	[0.018, 0.414]	0.145	0.121	[0.018, 0.407]	0.138	0.114	[0.016, 0.398]
κ	1.00				1.393	1.000	[1.000, 3.000]	1.404	1.000	[1.000, 3.000]	1.305	1.000	[1.000, 3.000]
Panel B: 2 states													
μ	0.00	0.006	0.006	[− 0.013, 0.001]				− 0.026	− 0.026	[− 0.033, − 0.019]			
ψ	0.10	0.145	0.145	[0.085, 0.206]				0.122	0.123	[0.050, 0.183]	0.127	0.127	[0.064, 0.191]
ξ^2	0.10	0.105	0.105	[0.093, 0.117]				0.091	0.091	[0.081, 0.102]	0.092	0.091	[0.081, 0.103]
γ		− 0.020	− 0.020	[− 0.034, − 0.007]	− 0.005	− 0.004	[− 0.007, 0.016]						
δ		0.910	0.910	[0.886, 0.932]	0.921	0.921	[0.900, 0.941]						
σ^2		0.072	0.072	[0.060, 0.086]	0.070	0.070	[0.059, 0.083]						
α					0.133	0.110	[0.015, 0.379]	0.172	0.144	[0.021, 0.481]	0.194	0.158	[0.022, 0.560]
κ	2.00				1.479	1.000	[1.000, 3.000]	2.137	2.000	[1.000, 5.000]	2.113	2.000	[1.000, 5.000]

Notes: Results are from 20,000 MCMC posterior draws after 10,000 burnin sweeps. The simulated data are 2000 return and log(RV) values generated as $r_t \sim \mathrm{N}(0, \exp(h_t))$ and $\log(\mathrm{RV}_t) \sim \mathrm{N}(0.1 + h_t, 0.1)$. The log-volatility in panel A is generated as $h_t \sim \mathrm{N}(-0.1 + 0.95h_{t-1}, 0.05)$, while in panel B follows a 2-states mixture generated as $h_t \sim \mathrm{N}(0.95h_{t-1}, 0.05)$, with probability 0.9 and $h_t \sim \mathrm{N}(-0.1 + 0.8h_{t-1}, 0.2)$, with probability 0.1.

Table 2. Summary statistics.

		Mean	St. Dev.	Skewness	Kurtosis	Min	Max
DJI	returns	0.020	1.097	− 0.033	11.790	− 8.405	10.754
	RV	1.082	2.609	14.003	332.380	0.019	86.240
	log(RV)	− 0.653	1.109	0.380	3.419	− 3.952	4.457
IXIC	returns	− 0.012	1.308	0.020	10.540	− 8.046	14.908
	RV	1.350	2.813	9.956	170.158	0.017	73.045
	log(RV)	− 0.431	1.122	0.433	3.136	− 4.071	4.291
SPX	returns	0.007	1.119	− 0.192	11.594	− 9.351	10.220
	RV	1.067	2.422	11.940	258.703	0.012	77.477
	log(RV)	− 0.678	1.132	0.340	3.306	− 4.408	4.350

Note: Data period: January 3rd, 2000 – December 31st, 2021. DJI: 5512 days. IXIC: 5516 days. SPX: 5515 days.

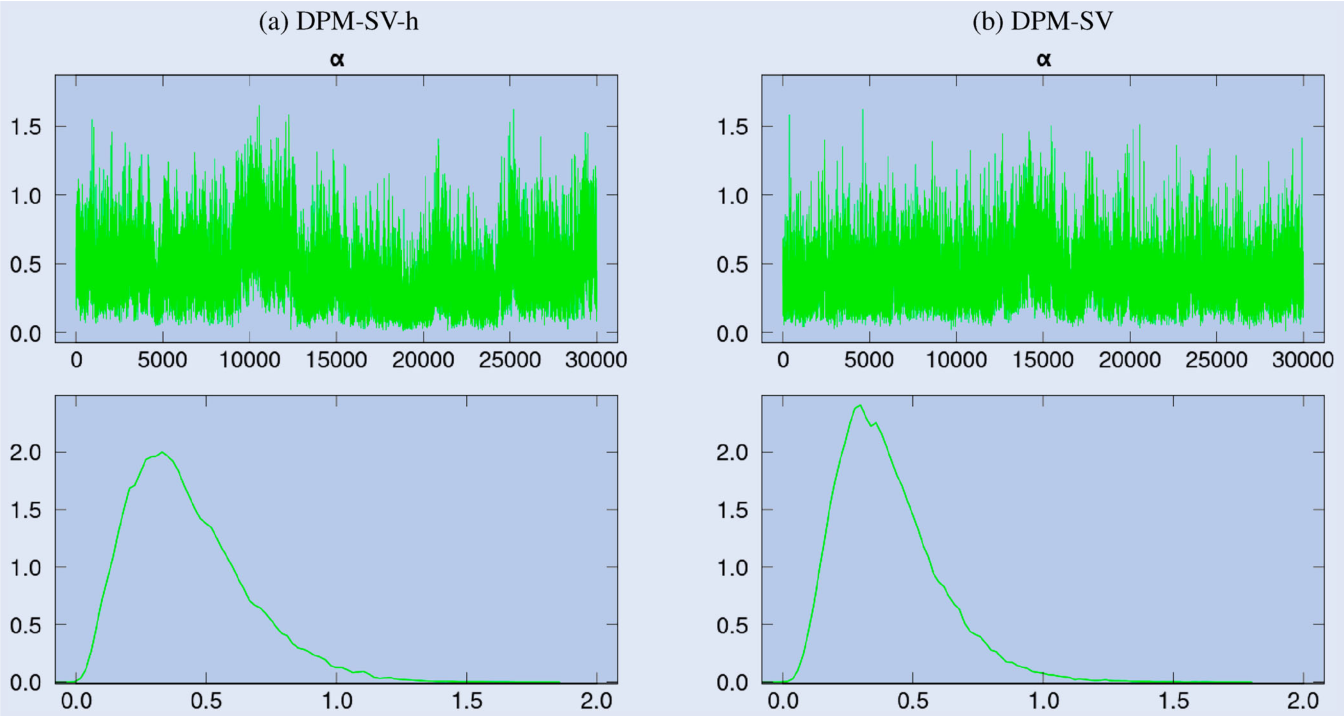

Figure 1. Trace plots of 30,000 MCMC posterior draws of DPM precision parameter, α, and their smoothed histogram, for DJI dataset. (a) DPM-SV-h (b) DPM-SV.

and $\sigma^2_{s_t}$ is consistently below -0.72. The DPM-SV(-h) models flexibly choose between persistence and randomness to adequately capture the volatility dynamics. The correlation between the level γ_{s_t} and the other two components varies substantially depending on the data. Including the return mean in the mixture mostly reduces the absolute value of the correlations.

5.3. *Out-of-sample forecasting*

Panel A of table 5 reports the cumulative log-predictive likelihood for returns, from (26), and the log-Bayes factor of each model against SV-N. These are calculated from $\tau = 2500$ recursive posterior estimations and likelihood evaluations at the out-of-sample returns. The forecasting period covers almost 10 years, from early 2012 until the end of 2021. Across all datasets, DPM-SV emerges as the best model, with the highest log-predictive likelihood scores and log-Bayes factors. The log-Bayes factor between DPM-SV and the restrictive DPM-SV-h ranges from 9.05 to 19.4, showing decisive evidence in favour of DPM-SV.

Overall, among the models, DPM-SV provides the best description of the underlying return-generating process. This performance can be attributed to the fact that DPM-SV effectively nests all the benchmarks. For instance, parametrically restricting δ_{s_t} and $\sigma^2_{s_t}$ in DPM-SV results in a conditional return distribution equivalent to RSV-DPM. The time-varying δ_{s_t} and $\sigma^2_{s_t}$ provide DPM-SV with the flexibility to estimate a more dynamic latent volatility. The difference between DPM-SV-h and RSV-N provides strong evidence supporting the nonparametric SV modelling.

Panel B of table 5 reports the cumulative log-predictive likelihood for log(RV), from (27), and the log-Bayes factor of each model against RSV-N. For all data series, there is decisive evidence favouring nonparametric SV modelling over the benchmarks. This is explained by the fact that both DPM-SV and DPM-SV-h nest the benchmarks in modelling log(RV). Compared to the semiparametric log(RV) modelling

Table 3. Posterior estimation results.

	SV-N		SV-t		RSV-N		RSV-DPM		DPM-SV-h		DPM-SV	
	Mean	95% D.I.	Mean	95% D.I.	Mean	95% D.I.	Mean	95% D.I.	Mean	95% D.I.	Mean	95% D.I.
Panel A: DJI												
μ	0.0631	[0.057, 0.069]	0.0569	[0.050, 0.064]	0.0727	[0.069, 0.076]			0.0725	[0.069, 0.076]		
ν			15.3146	[9.983, 26.297]								
ψ					−0.0780	[−0.117, −0.038]			−0.0750	[−0.116, −0.036]	0.0871	[0.039, 0.139]
ξ^2					0.1948	[0.181, 0.208]			0.1899	[0.177, 0.203]	0.1895	[0.178, 0.201]
γ	−0.0121	[−0.020, −0.005]	−0.0126	[−0.020, −0.006]	−0.0313	[−0.043, −0.020]	−0.0314	[−0.042, −0.021]				
δ	0.9759	[0.968, 0.983]	0.9798	[0.973, 0.986]	0.9469	[0.936, 0.957]	0.9543	[0.944, 0.964]				
σ^2	0.0579	[0.046, 0.073]	0.0479	[0.037, 0.060]	0.1115	[0.097, 0.127]	0.0951	[0.082, 0.109]				
α							0.2163	[0.042, 0.550]	0.4277	[0.100, 0.960]	0.4048	[0.123, 0.875]
κ							2.6802	[2.000, 5.000]	5.8244	[3.000, 11.000]	5.5063	[4.000, 9.000]
Panel B: IXIC												
μ	0.0538	[0.047, 0.061]	0.0545	[0.047, 0.063]	0.0829	[0.078, 0.088]			0.0827	[0.078, 0.088]		
ν			24.8326	[14.846, 37.147]								
ψ					−0.2321	[−0.272, −0.193]			−0.2282	[−0.269, −0.189]	−0.0069	[−0.052, 0.040]
ξ^2					0.1641	[0.152, 0.177]			0.1620	[0.150, 0.174]	0.1502	[0.139, 0.162]
γ	−0.0021	[−0.007, 0.003]	−0.0031	[−0.008, 0.002]	−0.0109	[−0.020, −0.002]	−0.0212	[−0.031, −0.012]				
δ	0.9829	[0.976, 0.989]	0.9844	[0.978, 0.990]	0.9483	[0.938, 0.958]	0.9528	[0.943, 0.962]				
σ^2	0.0389	[0.030, 0.049]	0.0354	[0.027, 0.045]	0.1114	[0.098, 0.127]	0.1020	[0.089, 0.116]				
α							0.2258	[0.045, 0.566]	0.3405	[0.069, 0.804]	0.3770	[0.116, 0.810]
κ							2.8096	[2.000, 5.000]	4.5393	[2.000, 9.000]	5.1011	[4.000, 8.000]
Panel C: SPX												
μ	0.0580	[0.052, 0.064]	0.0540	[0.047, 0.061]	0.0750	[0.071, 0.079]			0.0741	[0.071, 0.078]		
ν			18.0444	[10.928, 30.714]								
ψ					−0.1198	[−0.160, −0.081]			−0.1170	[−0.158, −0.078]	0.0819	[0.034, 0.131]
ξ^2					0.1644	[0.152, 0.177]			0.1645	[0.153, 0.177]	0.1555	[0.145, 0.167]
γ	−0.0111	[−0.019, −0.004]	−0.0118	[−0.019, −0.005]	−0.0322	[−0.044, −0.021]	−0.0361	[−0.048, −0.025]				
δ	0.9775	[0.970, 0.985]	0.9802	[0.973, 0.987]	0.9450	[0.934, 0.955]	0.9496	[0.939, 0.960]				
σ^2	0.0553	[0.043, 0.070]	0.0480	[0.037, 0.061]	0.1253	[0.110, 0.141]	0.1142	[0.100, 0.130]				
α							0.2185	[0.043, 0.557]	0.3545	[0.076, 0.847]	0.3954	[0.121, 0.849]
κ							2.7012	[2.000, 5.000]	4.7598	[2.000, 9.000]	5.3851	[4.000, 9.000]

Notes: Results are from 30,000 MCMC posterior draws after 20,000 burnin sweeps. The benchmark model specifications are in Section 3.3. The priors used are discussed in Section 4. In the SV-t model, parameter ν is the Student's-t degrees of freedom, with a uniform prior, $p(\nu) \sim \mathrm{U}(3, 50)$.

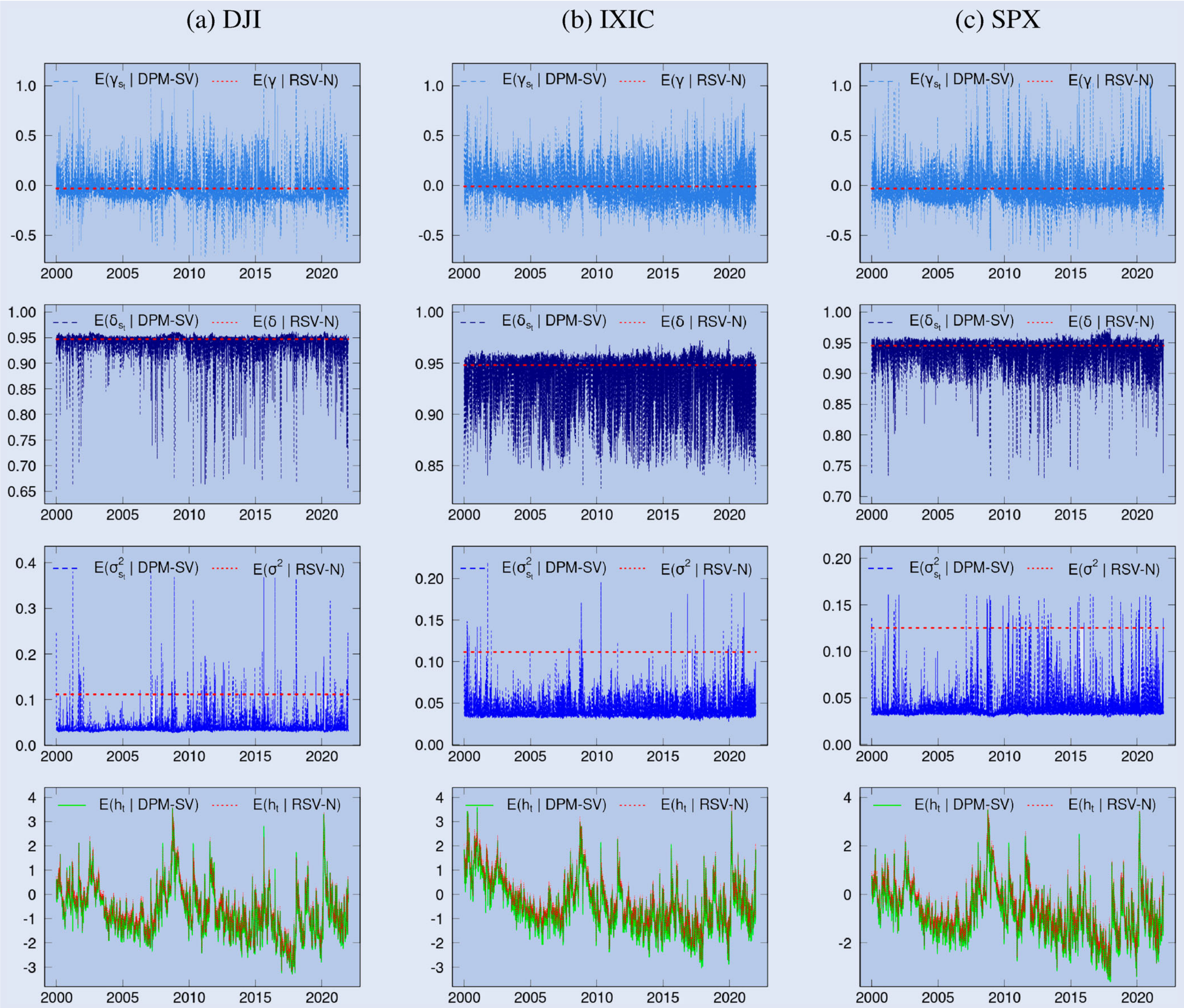

Figure 2. Full sample posterior estimations of the time-varying stochastic volatility equation mixing components from the DPM-SV model against the parametric RSV-N specification. From top to bottom: intercept $\mathbb{E}(\gamma_{s_t}|\text{Model})$, AR(1) coefficient $\mathbb{E}(\delta_{s_t}|\text{Model})$, variance of volatility $\mathbb{E}(\sigma^2_{s_t}|\text{Model})$ and smooth estimation of stochastic volatility $\mathbb{E}(h_t|\text{Model})$. (a) DJI (b) IXIC (c) SPX.

Table 4. Posterior correlations of the volatility mixing components.

	DJI		IXIC		SPX	
	DPM-SV-h	DPM-SV	DPM-SV-h	DPM-SV	DPM-SV-h	DPM-SV
$\text{Corr}[\mathbb{E}(\gamma_{s_t}|\mathcal{I}_T), \mathbb{E}(\delta_{s_t}|\mathcal{I}_T)]$	0.373	0.147	− 0.894	− 0.476	− 0.331	− 0.301
$\text{Corr}[\mathbb{E}(\gamma_{s_t}|\mathcal{I}_T), \mathbb{E}(\sigma^2_{s_t}|\mathcal{I}_T)]$	0.239	0.139	0.963	0.553	0.587	0.108
$\text{Corr}[\mathbb{E}(\delta_{s_t}|\mathcal{I}_T), \mathbb{E}(\sigma^2_{s_t}|\mathcal{I}_T)]$	− 0.796	− 0.809	− 0.968	− 0.824	− 0.937	− 0.724

Notes: This table reports the correlations of the posterior estimated volatility mixture components, $\mathbb{E}(\gamma_{s_t}|\mathcal{I}_T)$, $\mathbb{E}(\delta_{s_t}|\mathcal{I}_T)$ and $\mathbb{E}(\sigma^2_{s_t}|\mathcal{I}_T)$, $t = 1, \ldots, T$.

of RSV-DPM, the DPM-SV(-h) models, as seen in (18), provide more flexibility through h_t. Both RSV-DPM and DPM-SV(-h), model log(RV) semiparametrically, combining in its conditional mean and variance a parametric component with a nonparametric DPM component. The difference is in the volatility persistence adjustment from δ_{s_t}, which is absent in RSV-DPM. This adjustment enhances the flexibility of DPM-SV(-h) models, improving their volatility forecasts.

Panel C of table 5 reports the joint return and log(RV) cumulative log-predictive likelihood, from (27), and the log-Bayes factor of each model against RSV-N. The results provide decisive evidence in favour of the proposed nonparametric SV modelling. The DPM-SV(-h) models, against the RSV benchmarks, have log-Bayes factors values higher that 16. Between the two model specifications the general DPM-SV is very strongly preferred against

Table 5. Density forecasts.

	DJI		IXIC		SPX	
	log-PL	log-BF	log-PL	log-BF	log-PL	log-BF
Panel A: Returns						
SV-N	– 2569.25		– 2932.50		– 2537.95	
SV-t	– 2563.17	6.08	– 2928.24	4.26	– 2537.14	0.81
SV-DPM	– 2564.02	5.23	– 2931.18	1.32	– 2537.31	0.64
RSV-N	– 2466.57	102.68	– 2835.87	96.63	– 2418.74	119.21
RSV-DPM	– 2459.28	109.97	– 2835.14	97.36	– 2411.74	126.21
DPM-SV-h	– 2461.99	107.26	– 2832.35	100.15	– 2417.50	120.45
DPM-SV	**– 2452.94**	**116.31**	**– 2812.95**	**119.55**	**– 2402.69**	**135.26**
Panel B: Realized variance						
RSV-N	– 2500.86		– 2534.57		– 2520.26	
RSV-DPM	– 2502.27	– 1.41	– 2529.76	4.81	– 2522.16	– 1.9
DPM-SV-h	– 2498.32	2.54	– 2525.09	9.48	– 2510.21	10.05
DPM-SV	**– 2491.64**	**9.22**	**– 2520.93**	**13.64**	**– 2500.46**	**19.80**
Panel C: Joint return and log(RV) density						
RSV-N	– 4770.70		– 5175.80		– 4736.10	
RSV-DPM	– 4766.95	3.75	– 5179.70	– 3.9	– 4740.50	– 4.4
DPM-SV-h	– 4750.63	20.07	– 5151.57	24.23	– 4713.12	22.98
DPM-SV	**– 4725.48**	**45.22**	**– 5091.91**	**83.89**	**– 4656.25**	**79.85**
Forecasting period	9/1/2012 – 31/12/2021		17/1/2012 – 31/12/2021		11/1/2012 – 31/12/2021	

Notes: The table reports the log-predictive likelihood (log-PL) and log-Bayes Factors (log-BF) from 2500 out-of-sample model forecasts. The log-BFs are computed against the basic benchmark model on top, SV-N for panel A and RSV-N for panels B and C. Bold indicates the best value among the models in each column.

Table 6. DPM-SV sensitivity analysis to the choice of $G_0(\sigma^2_{s_t})$.

$G_0(\sigma^2_{s_t})$		log-PL		
Base measure	Mean	Returns	log(RV)	Joint
IG(22/2, 1/2)	0.050	– 1050.58	– 991.42	– 1931.91
IG(10/2, 1/2)	0.125	– 1048.91	– 991.66	– 1927.58
IG(6/2, 1/2)	0.250	– 1048.91	– 992.96	– 1930.49
IG(4/2, 1/2)	0.500	– 1049.72	– 995.39	– 1930.71

Notes: This table reports the DPM-SV model log-predictive likelihoods (log-PL) for the SPX dataset. These are calculated from 1000 out-of-sample forecasts (27/12/2017–31/12/2021) for different choices of the base measure $G_0(\sigma^2_{s_t})$, from very informative to less informative.

the restrictive DPM-SV-h with log-Bayes factors values exceeding 25.

SV models are known for their sensitivity to volatility of volatility prior choice. Table 6 presents a sensitivity analysis in the DPM-SV model, based on log-PL, for different volatility of volatility base measure choices, from very informative, $\mathbb{E}(\sigma^2_{s_t}) = 0.05$, to less informative, $\mathbb{E}(\sigma^2_{s_t}) = 0.5$. The log-Bayes Factors among the models range from 0 to 4.33 Specifically, for the return density, log-Bayes Factors are in values $0 - 1.67$, indicating indifference among the different cases. Overall, the analysis implies robustness of DPM-SV to the volatility of volatility base measure choice.

Figure 3 offers insights into the models' forecasting performance. It illustrates the cumulative log-Bayes factors over time of the proposed DPM-SV against the restricted DPM-SV-h and the benchmarks RSV-DPM and RSV-N. For all datasets, DPM-SV exhibits consistent ongoing gains after 2015. DPM-SV-h and RSV-N are competitive during the period 2012–2015, while RSV-DPM is very competitive during the first two forecasting years. The plots highlight the DPM-SV(-h) models ability to capture variance shocks, with significant predictive likelihood gains during notable market events such as the flash crash on August 24, 2015 and the market drop on February 8, 2018.

To better understand the models forecasting differences we turn to plots of log-predictive distributions. Figure 4 illustrates the one day ahead out-of-sample predictive density plots from the DPM-SV(-h) models compared to the RSV benchmarks, for the crash of Monday February 5th, 2018. The densities are computed over a grid of values with data up to and including Friday February 2nd, 2018.

The DPM-SV model captures non-Gaussian features in returns, log(RV), and stochastic volatility. Both DPM-SV and DPM-SV-h models yield a posterior distribution for the stochastic volatility that is clearly a mixture–exhibiting fat tails, asymmetry, and substantial deviation from the Gaussian shape.

Simulating h_{t+1} from the mixture distribution in (24), produces fat-tailed return densities† , as well as fat-tailed and asymmetric log(RV) densities. For returns, both DPM-SV and DPM-SV-h predictive density plots display thicker tails compared to RSV-DPM. The DPM-SV model, in particular, provides stronger evidence of tail asymmetry, more evident than in RSV-DPM. The proposed DPM-SV model allocates more probability mass to the left tail of the return distribution, resulting in higher predictive likelihood scores for the realized return values. In log(RV), all models incorporating DPM exhibit density shapes distinct from RSV-N. The DPM-SV(-h) models assign more probability mass to upward shocks than to downward RV movements.

† See also Jin and Maheu (2016).

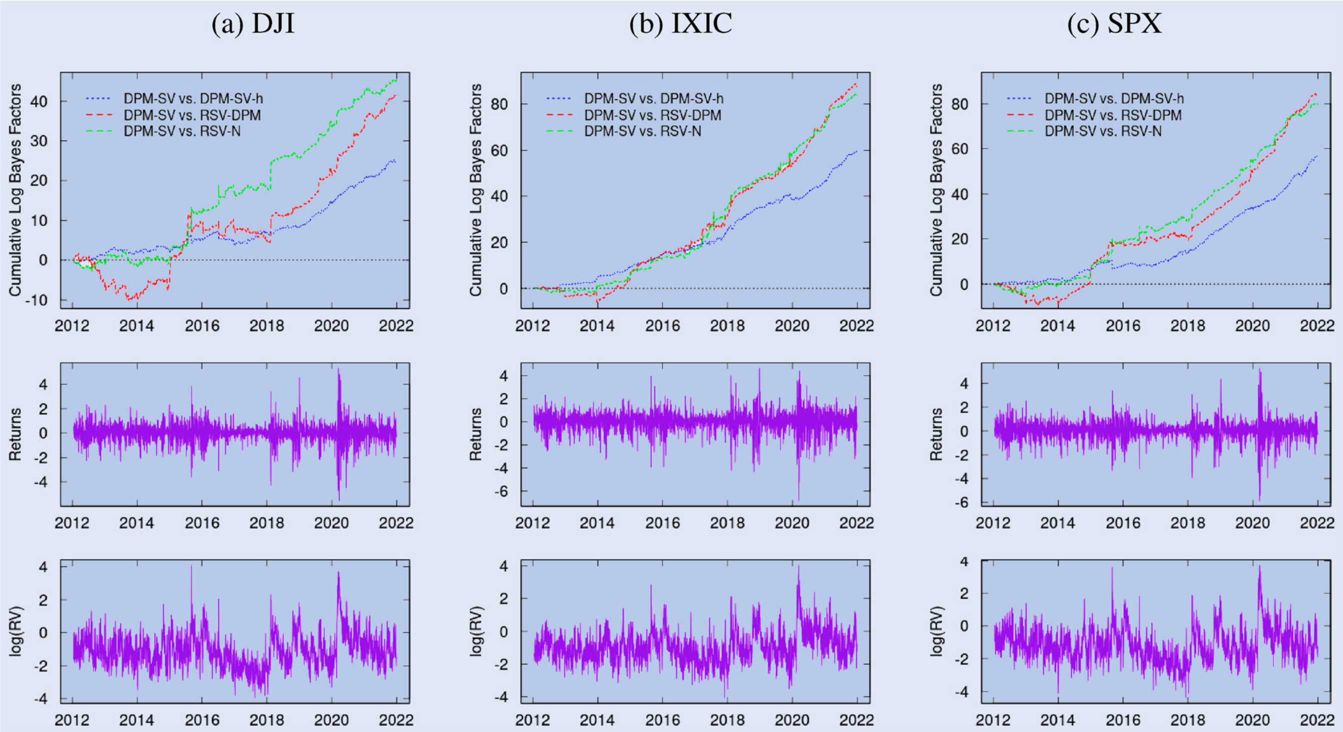

Figure 3. Joint return and log(RV) density forecasting. Top: cumulative log-Bayes factors for the pairs DPM-SV vs. DPM-SV-h (blue), DPM-SV vs. RSV-DPM (red) and DPM-SV vs. RSV-N (green). Middle: returns. Bottom: log(RV). (a) DJI (b) IXIC (c) SPX.

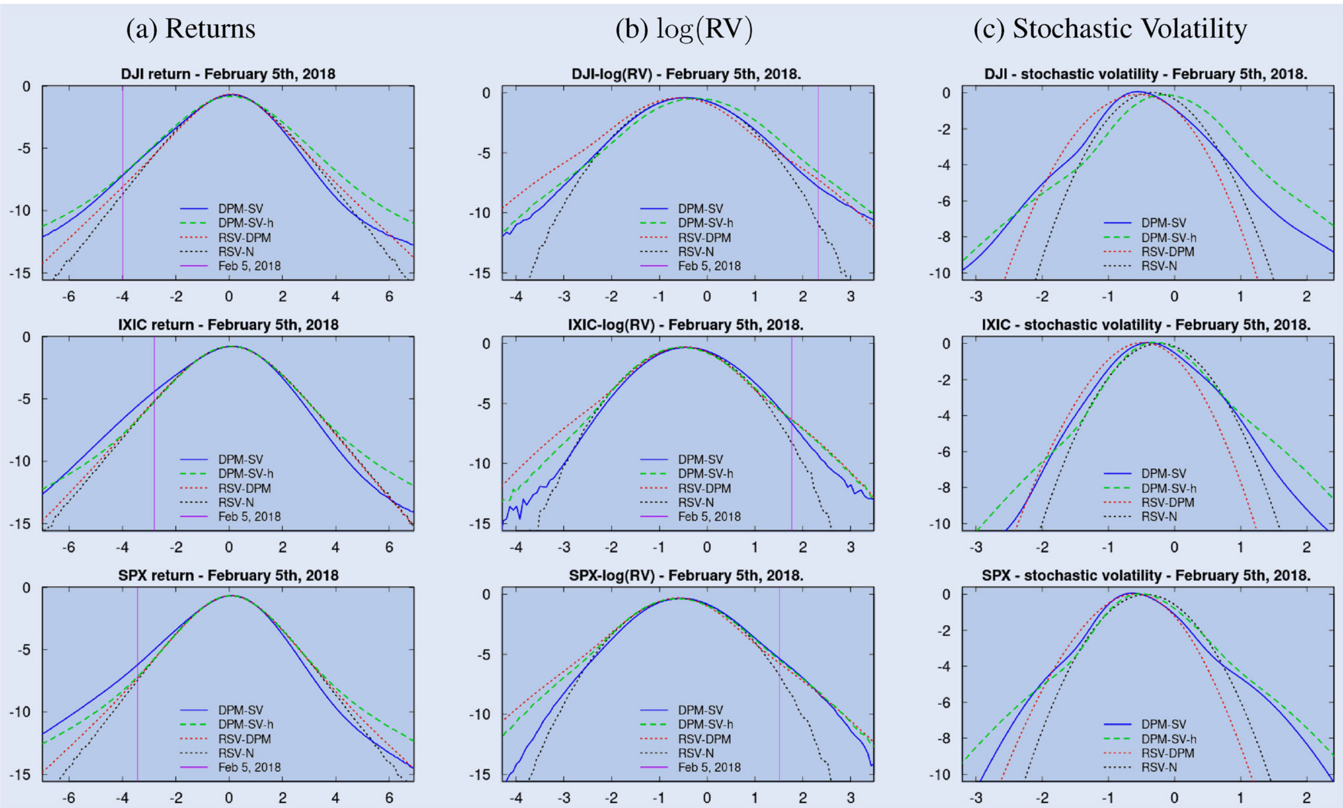

Figure 4. One day ahead out-of-sample logarithmic predictive density plots for Monday February 5th, 2018. The densities are computed over a grid of values from model posterior estimations with data up to and including Friday February 2nd, 2018. (a) Returns (b) log(RV) (c) Stochastic volatility.

Table 7. Region density forecasts.

	DJI			IXIC			SPX		
$\eta(\%)$:	-2%	-1%	0%	-2%	-1%	0%	-2%	-1%	0%
SV-N	– 35.54	– 143.02	– 465.58	– 46.71	– 181.28	– 668.85	– 36.49	– 137.20	– 459.27
SV-t	– 34.36	– 138.91	– 455.71	– 44.00	– 177.33	– 664.15	– 35.32	– 132.16	– 452.25
SV-DPM	– 34.98	– 141.31	– 455.55	– 44.88	– 175.11	– 666.28	– 34.07	– 132.35	– 450.59
RSV-N	– 33.34	– 134.68	– 412.81	– 37.98	– 164.91	– 621.98	– 33.41	– 130.27	– 398.31
RSV-DPM	– 32.95	**– 131.95**	– 409.12	– 38.42	– 165.68	– 622.86	– 33.01	– 127.56	– 395.36
DPM-SV-h	**– 32.10**	– 133.56	– 407.98	– 38.62	– 160.97	– 616.77	– 32.87	– 128.12	– 394.45
DPM-SV	– 32.24	– 132.14	**– 404.97**	**– 37.69**	**– 157.71**	**– 605.08**	**– 31.80**	**– 125.71**	**– 389.77**
Forecasts	45	206	1146	69	252	1113	47	201	1138
Forecasting period	9/1/2012 – 31/12/2021			17/1/2012 – 31/12/2021			11/1/2012 – 31/12/2021		

Notes: The table displays the region log-predictive likelihood (log-PL), constructed from: $p(r_{t+1}|r_{t+1} < \eta, \mathcal{I}_t)$, for $\boldsymbol{\eta} = \{\mathbf{-2, -1, 0}\}$. The chosen regions are negative returns ($r_{t+1} < 0$) and greater than 1% ($r_{t+1} < -1$) and 2% ($r_{t+1} < -2$) losses. Bold indicates the best value among the models.

Table 8. Tail-risk forecasting results.

	DJI		IXIC		SPX	
$\epsilon(\%)$:	1%	5%	1%	5%	1%	5%
SV-N	2.0847	1.6208	2.2532	1.8118	2.1350	1.6439
SV-t	2.0689	1.6284	2.2436	1.8122	2.0995	1.6385
SV-DPM	2.1574	1.6412	2.3091	1.8431	2.1844	1.6660
RSV-N	1.9937	1.5464	2.2204	1.7474	2.0438	1.5580
RSV-DPM	1.9551	1.5439	2.1871	1.7420	2.0244	1.5515
DPM-SV-h	1.9241	1.5411	2.1792	1.7411	2.0064	1.5480
DPM-SV	**1.9051**	**1.5256**	**2.0819**	**1.7183**	**1.9325**	**1.5338**
Forecasting period	9/1/2012 – 31/12/2021		17/1/2012 – 31/12/2021		11/1/2012 – 31/12/2021	

Notes: This table displays the average out-of-sample loss function of Taylor (2019) $\mathcal{L}^{\epsilon}(r_{t+1}, \mathrm{VaR}^{\epsilon}_{t+1}, \mathrm{ES}^{\epsilon}_{t+1})$ from (37). Bold indicates the best value among the models.

The ability of the proposed models to accommodate return plummets is further demonstrated in their forecasts for specific regions of the return density. We examine three cases relevant to risk measurement: negative returns, and losses greater than 1% and 2%. The region density forecasting results are presented in table 7. DPM-SV(-h) models achieve the highest region log-PL score in all cases except one, where RSV-DPM is marginally the best, with a log-Bayes factor of 0.19. The DPM-SV model is the best in 7 out of 9 occasions. This result is attributed to the tail asymmetry of the DPM-SV model, as shown in figure 4. For negative returns, the DPM-SV outperforms the best benchmark in each case, with log-Bayes factors exceeding 3. For losses greater than 1%, the log-Bayes factor of DPM-SV against the best benchmark ranges from -0.19 to 7.2, while for losses greater than 2%, the log-Bayes factor of DPM-SV(-h) against the best benchmark ranges from 0.29 to 1.21. We further explore the tail benefits of the proposed models with the following tail-risk forecasting application.

5.3.1. Tail-risk forecasting. A popular application of volatility models is in tail-risk forecasting for the measures of value-at-risk (VaR) and expected shortfall (ES). For r_{t+1}, the $\mathrm{VaR}^{\epsilon}_{t+1}$ at level ϵ is the conditional tail quantile defined as $P[r_{t+1} \leq \mathrm{VaR}^{\epsilon}_{t+1} \,|\, \mathcal{I}_t] = \epsilon$. This denotes the least potential loss for return r_{t+1} with probability ϵ. VaR does not consider the loss magnitude. This is better evaluated with ES which is the average expected loss, conditional on exceeding the VaR, and is defined as $\mathrm{ES}^{\epsilon}_{t+1} = \mathbb{E}[r_{t+1} \,|\, r_{t+1} \leq \mathrm{VaR}^{\epsilon}_{t+1}, \mathcal{I}_t]$.

To estimate the out-of-sample VaR and ES, we simulate a large number of return draws from each model's predictive density $\{r^{(i)}_{t+1}\}_{i=1}^{R}$. For the DPM-SV the predictive density is in (20). The empirical $\epsilon\%$ quantile of the simulated return values is the $\mathrm{VaR}^{\epsilon}_{t+1}$. The expected shortfall is calculated as

$$\mathrm{ES}^{\epsilon}_{t+1} = \frac{\sum_{i=1}^{R} r^{(i)}_{t+1} \mathbf{1}_{r^{(i)}_{t+1} \leq \mathrm{VaR}^{\epsilon}_{t+1}}}{\sum_{i=1}^{R} \mathbf{1}_{r^{(i)}_{t+1} \leq \mathrm{VaR}^{\epsilon}_{t+1}}}.$$

The joint VaR and ES model forecasts are compared with the loss function of Taylor (2019) defined as

$$\begin{aligned} &\mathcal{L}^{\epsilon}(r_{t+1}, \mathrm{VaR}^{\epsilon}_{t+1}, \mathrm{ES}^{\epsilon}_{t+1}) \\ &= -\log\left(\frac{\epsilon - 1}{\mathrm{ES}^{\epsilon}_{t+1}}\right) \\ &\quad - \frac{\left(r_{t+1} - \mathrm{VaR}^{\epsilon}_{t+1}\right)\left(\epsilon - \mathbf{1}_{r_{t+1} \leq \mathrm{VaR}^{\epsilon}_{t+1}}\right)}{\epsilon \mathrm{ES}^{\epsilon}_{t+1}} + \frac{r_{t+1}}{\mathrm{ES}^{\epsilon}_{t+1}}. \end{aligned} \tag{37}$$

Smaller average $\mathcal{L}^{\epsilon}(r_{t+1}, \mathrm{VaR}^{\epsilon}_{t+1}, \mathrm{ES}^{\epsilon}_{t+1})$ values over the out-of-sample data indicate more accurate tail measures.

Table 8 displays the average out-of-sample loss from (37) for the 1% and 5% VaR and ES measures. Across all the three indices considered, the proposed DPM-SV model produces the best VaR and ES values, with the least average loss score. The second best model is consistently the restricted DPM-SV-h. These results further highlight the forecasting benefits of the proposed flexible volatility framework for extreme tail events.

Overall, the proposed DPM-SV(-h) models offer enhanced density forecasts along with an improved ability to capture tail risks and volatility shocks, making them highly appealing for financial risk measurement applications.

6. Concluding remarks

This paper introduces a novel discrete-time stochastic volatility model, where the latent volatility is modelled nonparametrically using a Dirichlet process mixture. Our findings reveal that the stochastic volatility equation components exhibit significant variation over time. The model dynamically balances volatility persistence and randomness, effectively capturing variance shocks and extreme returns. Out-of-sample density forecasts demonstrate substantial improvement, particularly for returns, log(RV), and the left tail of the return distribution. In addition, the proposed model offers economic gains by enhancing joint forecasts of the tail-risk measures value-at-risk and expected shortfall.

The proposed framework offers two potential extensions. First, it can be adapted to the asymmetric stochastic volatility model. Second, the independent infinite mixture can be replaced with a Markovian structure. These will be pursued in future research.

Acknowledgments

The author thanks John M. Maheu, Ronald J. Balvers, Narat Charupat, Arnaud Dufays, the participants of the 13th and the 15th RCEA Bayesian Econometrics workshops, and two anonymous referees for comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Efthimios Nikolakopoulos http://orcid.org/0000-0002-5579-0834

References

Abanto-Valle, C.A., Bandyopadhyay, D., Lachos, V.H. and Enriquez, I., Robust bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. *Comput. Stat. Data. Anal.*, 2010, **54**(12), 2883–2898.

Aït-Sahalia, Y. and Mancini, L., Out of sample forecasts of quadratic variation. *J. Econom.*, 2008, **147**(1), 17–33.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Ebens, H., The distribution of realized stock return volatility. *J. Financ. Econ.*, 2001a, **61**(1), 43–76.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P., The distribution of realized exchange rate volatility. *J. Am. Stat. Assoc.*, 2001b, **96**(453), 42–55.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P., Modeling and forecasting realized volatility. *Econometrica*, 2003, **71**(2), 579–625.

Ascolani, F., Lijoi, A., Rebaudo, G. and Zanella, G., Clustering consistency with dirichlet process mixtures. *Biometrika*, 2023, **110**(2), 551–558.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A. and Shephard, N., Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 2008, **76**(6), 1481–1536.

Barndorff-Nielsen, O.E. and Shephard, N., Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc.: Ser. B (Statistical Methodology)*, 2002a, **64**(2), 253–280.

Barndorff-Nielsen, O.E. and Shephard, N., Estimating quadratic variation using realized variance. *J. Appl. Econom.*, 2002b, **17**(5), 457–477.

Bauwens, L., Preminger, A. and Rombouts, J.V., Theory and inference for a markov switching garch model. *Econom. J.*, 2010, **13**(2), 218–244.

Bollerslev, T., Generalized autoregressive conditional heteroskedasticity. *J. Econom.*, 1986, **31**(3), 307–327.

Carvalho, C.M. and Lopes, H.F., Simulation-based sequential analysis of markov switching stochastic volatility models. *Comput. Stat. Data. Anal.*, 2007, **51**(9), 4526–4542.

Chib, S., Nardari, F. and Shephard, N., Markov chain monte carlo methods for stochastic volatility models. *J. Econom.*, 2002, **108**(2), 281–316.

Delatola, E.-I. and Griffin, J.E., Bayesian nonparametric modelling of the return distribution with stochastic volatility. *Bayesian Anal.*, 2011, **6**(4), 901–926.

Delatola, E.-I. and Griffin, J.E., A bayesian semiparametric model for volatility with a leverage effect. *Comput. Stat. Data. Anal.*, 2013, **60**, 97–110.

Diks, C., Panchenko, V. and Van Dijk, D., Likelihood-based scoring rules for comparing density forecasts in tails. *J. Econom.*, 2011, **163**(2), 215–230.

Dufays, A., Infinite-state markov-switching for dynamic volatility. *J. Financial Econom.*, 2016, **14**(2), 418–460.

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

Escobar, M.D. and West, M., Bayesian density estimation and inference using mixtures. *J. Am. Stat. Assoc.*, 1995, **90**(430), 577–588.

Ferguson, T.S., A bayesian analysis of some nonparametric problems. *Ann. Stat.*, 1973, **1**, 209–230.

Fleming, J. and Kirby, C., A closer look at the relation between garch and stochastic autoregressive volatility. *J. Financial Econom.*, 2003, **1**(3), 365–419.

Geweke, J., Bayesian comparison of econometric models. Technical report, Citeseer, 1994.

Hansen, P.R., Huang, Z. and Shek, H.H., Realized garch: A joint model for returns and realized measures of volatility. *J. Appl. Econom.*, 2012, **27**(6), 877–906.

Heber, G., Lunde, A., Shephard, N. and Sheppard, K., Oxford-Man Institute's Realized Library, version 0.1, 2009.

Jacquier, E., Polson, N.G. and Rossi, P.E., Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *J. Econom.*, 2004, **122**(1), 185–212.

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

Jensen, M.J. and Maheu, J.M., Bayesian semiparametric multivariate GARCH modeling. *J. Econom.*, 2013, **176**(1), 3–17.

Jensen, M.J. and Maheu, J.M., Estimating a semiparametric asymmetric stochastic volatility model with a dirichlet process mixture. *J. Econom.*, 2014, **178**, 523–538.
Jin, X. and Maheu, J.M., Bayesian semiparametric modeling of realized covariance matrices. *J. Econom.*, 2016, **192**(1), 19–39.
Kalimipalli, M. and Susmel, R., Regime-switching stochastic volatility and short-term interest rates. *J. Empirical Finance*, 2004, **11**(3), 309–329.
Kalli, M., Griffin, J.E. and Walker, S.G., Slice sampling mixture models. *Stat. Comput.*, 2011, **21**(1), 93–105.
Kass, R.E. and Raftery, A.E., Bayes factors. *J. Am. Stat. Assoc.*, 1995, **90**(430), 773–795.
Kim, S., Shephard, N. and Chib, S., Stochastic volatility: Likelihood inference and comparison with arch models. *Rev. Econ. Stud.*, 1998, **65**(3), 361–393.
Li, C., Maheu, J.M. and Yang, Q., An infinite hidden markov model with stochastic volatility. *J. Forecast.*, 2024, **43**(6), 2187–2211.
Liesenfeld, R. and Jung, R.C., Stochastic volatility models: Conditional normality versus heavy-tailed distributions. *J. Appl. Econom.*, 2000, **15**(2), 137–160.
Liu, J., A Bayesian semiparametric realized stochastic volatility model. *J. Risk Financial Manage.*, 2021, **14**(12), 617.
Mahieu, R.J. and Schotman, P.C., An empirical application of stochastic volatility models. *J. Appl. Econom.*, 1998, **13**(4), 333–360.
Nakajima, J. and Omori, Y., Stochastic volatility model with leverage and asymmetrically heavy-tailed error using gh skew student's t-distribution. *Comput. Stat. Data. Anal.*, 2012, **56**(11), 3690–3704.
Sethuraman, J., A constructive definition of dirichlet priors. *Stat. Sin.*, 1994, **4**, 639–650.
Shirota, S., Omori, Y., Lopes, H.F. and Piao, H., Cholesky realized stochastic volatility model. *Econom. Stat.*, 2017, **3**, 34–59.
So, M.E.P., Lam, K. and Li, W.K., A stochastic volatility model with markov switching. *J. Bus. Econ. Stat.*, 1998, **16**(2), 244–253.
Takahashi, M., Omori, Y. and Watanabe, T., Estimating stochastic volatility models using daily returns and realized volatility simultaneously. *Comput. Stat. Data. Anal.*, 2009, **53**(6), 2404–2426.
Taylor, J.W., Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric laplace distribution. *J. Bus. Econ. Stat.*, 2019, **37**(1), 121–133.
Taylor, S.J., Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961–75. *Time Series Anal.: Theory Practice*, 1982, **1**, 203–226.
Virbickaitė, A. and Lopes, H.F., Bayesian semiparametric markov switching stochastic volatility model. *Appl. Stoch. Models. Bus. Ind.*, 2019, **35**(4), 978–997.
Vo, M.T., Regime-switching stochastic volatility: Evidence from the crude oil market. *Energy Econ.*, 2009, **31**(5), 779–788.
Walker, S.G., Sampling the dirichlet mixture model with slices. *Commun. Stat. – Simul. Comput.*®, 2007, **36**(1), 45–54.
Yamauchi, Y. and Omori, Y., Multivariate stochastic volatility model with realized volatilities and pairwise realized correlations. *J. Bus. Econ. Stat.*, 2020, **38**(4), 839–855.
Zhang, L., Mykland, P.A. and A-Sahalia, Y., A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Am. Stat. Assoc.*, 2005, **100**(472), 1394–1411.
Zhang, Z. and Zhao, R., Improving the asymmetric stochastic volatility model with ex-post volatility: The identification of the asymmetry. *Quant. Finance*, 2023, **23**(1), 35–51.

Appendix

DPM-SV sampling steps

Following are details for the DPM-SV model estimation steps discussed in Section 3.1.

(1) Sample h_t from $p(h_t \mid h_{-t}, \mathcal{I}_t, \theta, \mu_{1:k}, \gamma_{1:k}, \delta_{1:k}, \sigma^2_{1:k}, s_{1:T})$, $t = 1, \ldots, T$.

To estimate the stochastic volatility vector h_t, $t = 1, \ldots, T$ we use an extension of the single-move sampler from Kim *et al.* (1998). Following the Bayes rule twice, the conditional posterior above is proportional to

$$p(r_t \mid \mu_{s_t}, h_t)\, p(\log(\mathrm{RV}_t) \mid h_t, \psi, \xi^2)\, p(h_t \mid h_{-t}, \gamma_{s_t}, \delta_{s_t}, \sigma^2_{s_t}), \tag{A1}$$

$$p(r_t \mid \mu_{s_t}, h_t)\, p(\log(\mathrm{RV}_t) \mid h_t, \psi, \xi^2)\, p(h_t \mid h_{t-1}, \gamma_{s_t}, \delta_{s_t}, \sigma^2_{s_t}))$$

$$p(h_{t+1} \mid h_t, \gamma_{s_{t+1}}, \delta_{s_{t+1}}, \sigma^2_{s_{t+1}}), \tag{A2}$$

$$p(r_t \mid \mu_{s_t}, h_t)\, p(\log(\mathrm{RV}_t) \mid h_t, \psi, \xi^2)\, f_{\mathrm{N}}(h_t \mid \bar{h}, v_h^2), \tag{A3}$$

$$p(r_t \mid \mu_{s_t}, h_t)\, f_{\mathrm{N}}(h_t \mid M_h, V_h), \tag{A4}$$

with $f_{\mathrm{N}}(.)$ being the normal density. The posterior in (A2) has an unknown form. By combining the last two densities we get that $p(h_t \mid h_{-t}, \gamma_{s_t}, \delta_{s_t}, \sigma^2_{s_t}) \propto f_{\mathrm{N}}(h_t \mid \bar{h}, v_h^2)$ with

$$\bar{h} = \frac{\delta_{s_{t+1}} \sigma^2_{s_t} (h_{t+1} - \gamma_{s_{t+1}}) + \sigma^2_{s_{t+1}} (\delta_{s_t} h_{t-1} + \gamma_{s_t})}{\sigma^2_{s_{t+1}} + \delta^2_{s_{t+1}} \sigma^2_{s_t}},$$

$$v_h^2 = \frac{\sigma^2_{s_t} \sigma^2_{s_{t+1}}}{\sigma^2_{s_{t+1}} + \delta^2_{s_{t+1}} \sigma^2_{s_t}}.$$

Using the $\log(\mathrm{RV}_t)$ data in (A3) we get that $h_t | h_{-t}, \log(\mathrm{RV}_t), \theta, \gamma_{1:k}, \delta_{1:k}, \sigma^2_{1:k}, s_{1:T} \sim \mathrm{N}(M_h, V_h)$ with

$$M_h = \frac{(\log(\mathrm{RV}_t) - \psi) v_h^2 + \bar{h} \xi^2}{v_h^2 + \xi^2},$$

$$V_h = \frac{\xi^2 v_h^2}{v_h^2 + \xi^2}.$$

The posterior in (A4) does not have a known form so a Metropolis-Hastings (MH) algorithm is used to sample from it. The proposal distribution is found following the results of Kim *et al.* (1998). They show that $\exp(-h_t)$ is bounded and

$$\begin{aligned} p(r_t \mid \mu_{s_t}, h_t) &\propto f(r_t, h_t, \mu_{s_t}) \\ &= \exp\{-.5 h_t - .5 \exp(-h_t)(r_t - \mu_{s_t})^2\} \\ &\le \exp\{-.5 h_t - .5 \exp(-M_h)(r_t - \mu_{s_t})^2 (1 + M_h - h_t)\} \\ &= g(r_t, h_t, \mu_{s_t}, M_h). \end{aligned}$$

Combining this with (A4) we get the proposal

$$\begin{aligned} &p(r_t \mid \mu_{s_t}, h_t)\, f_{\mathrm{N}}(h_t \mid M_h, V_h) \\ &\quad \le g(r_t, h_t, \mu_{s_t}, M_h) f_{\mathrm{N}}(h_t \mid M_h, V_h) \propto f_{\mathrm{N}}(h_t \mid M, V_h), \end{aligned}$$

with $M = M_h + .5 V_h((r_t - \mu_{s_t})^2 \exp(M_h) - 1)$. The candidate $h_t' \sim \mathrm{N}(M, V_h)$ is accepted as a draw of h_t with probability

$$\min \left\{ \frac{p(h_t' | h_{-t}, \mathcal{I}_t, \theta, \mu_{1:k}, \gamma_{1:k}, \delta_{1:k}, \sigma^2_{1:k}, s_{1:T}) / \mathrm{N}(h_t' \mid M, V_h)}{p(h_t | h_{-t}, \mathcal{I}_t, \theta, \mu_{1:k}, \gamma_{1:k}, \delta_{1:k}, \sigma^2_{1:k}, s_{1:T}) / \mathrm{N}(h_t \mid M, V_h)}, 1 \right\}.$$

(2) Sample (a) ψ from $p(\psi \mid \mathrm{RV}_{1:T}, h_{1:T}, \xi^2)$ and (b) ξ^2 from $p(\xi^2 \mid \mathrm{RV}_{1:T}, h_{1:T}, \psi)$.

Let $x_t = \log(\mathrm{RV}_t) - h_t$, $t = 1, \ldots, T$, from the measurement equation (2)

$$x_t = \psi + \xi z_t, \quad z_t \overset{\text{iid}}{\sim} \mathrm{N}(0, 1).$$

Parameters ψ and ξ^2 are sampled with Gibbs draws from linear model conjugate priors as

(a) With prior $p(\psi) \sim \mathrm{N}(c_0, c^2)$ parameter ψ is sampled with a Gibbs draw from

$$p(\psi \mid \mathrm{RV}_{1:T}, h_{1:T}, \xi^2) \sim \mathrm{N}(M_\psi, V_\psi)$$

$$\text{with} \quad M_\psi = V_\psi \left(\xi^{-2} \sum_{t=1}^{T} x_t + c^{-2} c_0 \right) \quad \text{and}$$

$$V_\psi = \left(T\xi^{-2} + c^{-2} \right)^{-1}.$$

(b) With prior $p(\xi^2) \sim \mathrm{IG}(v_\xi, s_\xi)$ parameter ξ^2 is sampled with a Gibbs draw from

$$p(\xi^2 \mid \mathrm{RV}_{1:T}, h_{1:T}, \psi)$$
$$\sim \mathrm{IG}\left(\frac{T + v_\xi}{2}, \frac{\sum_{t=1}^{T} (x_t - \psi)^2 + s_\xi}{2} \right).$$

(3) Sample (a) $\mu_{1:k}$ from $p(\mu_{1:k} \mid r_{1:T}, h_{1:T}, s_{1:T})$, (b) $\gamma_{1:k}$ from $p(\gamma_{1:k} \mid r_{1:T}, h_{1:T}, \delta_{1:k}, \sigma^2_{1:k}, s_{1:T})$ (c) $\delta_{1:k}$ from $p(\delta_{1:k} \mid r_{1:T}, h_{1:T}, \gamma_{1:k}, \sigma^2_{1:k}, s_{1:T})$ and (d) $\sigma^2_{1:k}$ from $p(\sigma^2_{1:k} \mid r_{1:T}, h_{1:T}, \gamma_{1:k}, \delta_{1:k}, s_{1:T})$.

The mixing parameters are drawn with Gibbs sampling from linear model conjugate priors:

(a) $p(\mu_j) \sim \mathrm{N}(m_0, v_m^2)$, $j = 1, \ldots, k$. Its conditional posterior is proportional to

$$p(\mu_j \mid r_{1:T}, h_{1:T}, s_t = j)$$
$$\propto p(\mu_j) \prod_{t:s_t=j} p(r_t|\mu_j, h_t)$$
$$\sim \mathrm{N}\left(m_\mu, v_\mu^{-1} \right)$$

$$\text{with} \quad v_\mu = \sum_{t:s_t=j} \exp(-h_t) + u_0^{-2} \quad \text{and}$$

$$m_\mu = v_\mu^{-1} \left(\sum_{t:s_t=j} r_t \exp(-h_t) + m_0 v_m^{-2} \right).$$

For the parametric mean in DPM-SV-h, the above sampling step is used with $t = 1, \ldots, T$.

(b) $p(\gamma_j) \sim \mathrm{N}(g_0, v_g^2)$, $j = 1, \ldots, k$. Its conditional posterior is proportional to

$$p(\gamma_j \mid h_{1:T}, \delta_j, \sigma_j^2, s_t = j)$$
$$\propto p(\gamma_j) \prod_{t:s_t=j} p(h_t \mid \gamma_j, \delta_j, \sigma_j^2)$$
$$\sim \mathrm{N}\left(m_\gamma, v_\gamma^2 \right)$$

$$\text{with} \quad v_\gamma^2 = \frac{\sigma_j^2 v_g^2}{n_j v_g^2 + \sigma_j^2} \quad \text{and}$$

$$m_\gamma = v_\gamma^2 \left(\sigma_j^{-2} \sum_{t:s_t=j} (h_t - \delta_j h_{t-1}) + g_0 v_g^{-2} \right),$$

where $n_j = \sum_{t:s_t=j} \mathbf{1}\{s_t = j\}$, is the number of observations in the cluster j.

(c) $p(\delta_j) \sim \mathrm{N}(d_0, v_d^2)\mathbf{1}_{|\delta_j|<1}$, $j = 1, \ldots, k$. Its conditional posterior is proportional to

$$p(\delta_j \mid h_{1:T}, \gamma_j, \sigma_j^2, s_t = j)$$
$$\propto p(\delta_j) \prod_{t:s_t=j} p(h_t \mid \gamma_j, \delta_j, \sigma_j^2)$$
$$\sim \mathrm{N}\left(m_\delta, v_\delta^2 \right)$$

$$\text{with} \quad v_\delta^2 = \frac{v_d^2}{v_d^2 \sum_{t:s_t=j} h_{t-1}^2 \sigma_j^{-2} + 1} \quad \text{and}$$

$$m_\delta = v_\delta^2 \left(\sum_{t:s_t=j} h_{t-1}(h_t - \gamma_j) \sigma_j^{-2} + \frac{d_0}{v_d^2} \right).$$

(d) $p(\sigma_j^2) \sim \mathrm{IG}(v_0/2, s_0/2)$, $j = 1, \ldots, k$. Its conditional posterior is proportional to

$$p(\sigma_j^2 \mid h_{1:T}, \delta_j, \gamma_j, s_t = j)$$
$$\propto p(\sigma_j^2) \prod_{t:s_t=j} p(h_t \mid \gamma_j, \delta, \sigma_j^2)$$
$$\sim \mathrm{IG}\left(\frac{n_j + v_0}{2}, \frac{\sum_{t:s_t=j} (h_t - \gamma_j - \delta_j h_{t-1})^2 + s_0}{2} \right).$$

(4) Update $w_{1:k}, u_{1:T}, k \mid s_{1:T}, \alpha$.

(a) Update the mixture weights in $w_{1:k} \mid s_{1:T}, \alpha$ with a stick-breaking process as

$$v_j \mid s_{1:T}, \alpha \sim \mathrm{B}\left(1 + \sum_{t=1}^{T} \mathbf{1}_{s_t=j}, \alpha + \sum_{t=1}^{T} \mathbf{1}_{s_t>j} \right),$$

$$w_1 = v_1, \quad w_j = v_j \prod_{l=1}^{j-1} (1 - v_l), \quad j = 2, \ldots, k.$$

(b) Update the slice vector $u_{1:T} \mid w_{1:k}, s_{1:T}$ from a uniform draw as: $u_t \mid w_{1:k}, s_{1:T} \sim \mathrm{U}(0, w_{s_t})$.

(c) Update the number of mixture clusters k to the smallest positive integer that satisfies: $\sum_{j=1}^{k} w_j > 1 - \min(u_{1:T})$. If new clusters are needed to satisfy the inequality, their mixing components are drawn from the base measure (9).

(5) Sample s_t from $p(s_t \mid \mathcal{I}_t, h_{1:T}, \theta, \mu_{1:k}, \gamma_{1:k}, \delta_{1:k}, \sigma^2_{1:k}, w_{1:k}, u_{1:T}, k)$, $t = 1, \ldots, T$.

Each element s_t of the vector $s_{1:T}$ takes an integer value j which is drawn from a multinomial distribution with probabilities

$$p(s_t = j \mid r_{1:T}, h_{1:T}, \mu_{1:k}, \gamma_{1:k}, \delta_{1:k}, \sigma^2_{1:k}, w_{1:k}, u_{1:T}, \alpha)$$
$$\propto \mathbf{1}_{u_t<w_j} \mathrm{N}\left(r_t \mid \mu_j, \exp(h_t) \right) \mathrm{N}\left(h_t \mid \gamma_j + \delta_j h_{t-1}, \sigma_j^2 \right),$$

for $j = 1, \ldots, k$. The number of active clusters κ, can be calculated as the ones with at least one assigned data observation.

(6) Sample α from $p(\alpha \mid \kappa, T)$.

The DPM precision parameter α with a gamma prior $\alpha \sim \Gamma(a_0, b_0)$ is drawn following the two steps algorithm of Escobar and West (1995):

(i) draw the random variable $\xi \mid \alpha, k \sim \mathrm{B}(\alpha + 1, T)$.

(ii) sample α from

$$\alpha \mid \xi \sim \pi_\xi \Gamma(a_0 + \kappa, b_0 - \log(\xi))$$
$$+ (1 - \pi_\xi) \Gamma(a_0 + \kappa - 1, b_0 - \log(\xi)),$$

with $\frac{\pi_\xi}{1-\pi_\xi} = \frac{a_0+k-1}{T(b_0-\log(\xi))}$.