Bayesian prediction of jumps in large panels of time series data

A. Alexopoulos * P. Dellaportas † O. Papaspiliopoulos ‡

April 11, 2019

Abstract

We take a new look at the problem of disentangling the volatility and jumps processes in a panel of stock daily returns. We first provide an efficient computational framework that deals with the stochastic volatility model with Poisson-driven jumps in a univariate scenario that offers a competitive inference alternative to the existing implementation tools. This methodology is then extended to a large set of stocks in which it is assumed that the unobserved jump intensities of each stock co-evolve in time through a dynamic factor model. A carefully designed sequential Monte Carlo algorithm provides out-of-sample empirical evidence that our suggested model outperforms, with respect to predictive Bayes factors, models that do not exploit the panel structure of stocks.

1 Introduction

It has been recognised in the financial literature that jumps in asset returns occur clustered in time and affect several stock markets within a few hours or days, see for example Aït-Sahalia et al. (2015). We work with a large panel of stocks from several European markets and we aim to identify and to predict joint tail risk expressed as probabilities of jumps in their daily returns. Our modelling assumptions are based on the well-known paradigms of stochastic volatility (SV) models (Taylor, 1982) combined with Poisson-driven jumps (Andersen et al., 2002). The resulting Bayesian hierarchical models require careful prior specifications and sophisticated, modern Markov chain Monte Carlo (MCMC) inference implementation strategies. Our purpose is to provide a general, robust modelling and algorithmic framework and test it in an out-of-sample forecasting scenario with real data.

Our first point of departure is the univariate SV model with jumps. We provide a new efficient MCMC algorithm combined with careful prior specification that improves upon existing MCMC strategies. Next, we extend this approach by utilising the panel structure of stock returns so that unobserved jump intensities are assumed to propagate over time through a dynamic factor model. The resulting Bayesian hierarchical model can

*Department of Statistical Science, University College London, UK. Email: a.alexopoulos@ucl.ac.uk.

†Department of Statistical Science, University College, London, UK. Email: p.dellaportas@ucl.ac.uk.

‡ICREA and Department of Economics, Universitat Pompeu Fabra, Barcelona, Spain. Email: omiros.papaspiliopoulos@upf.edu.

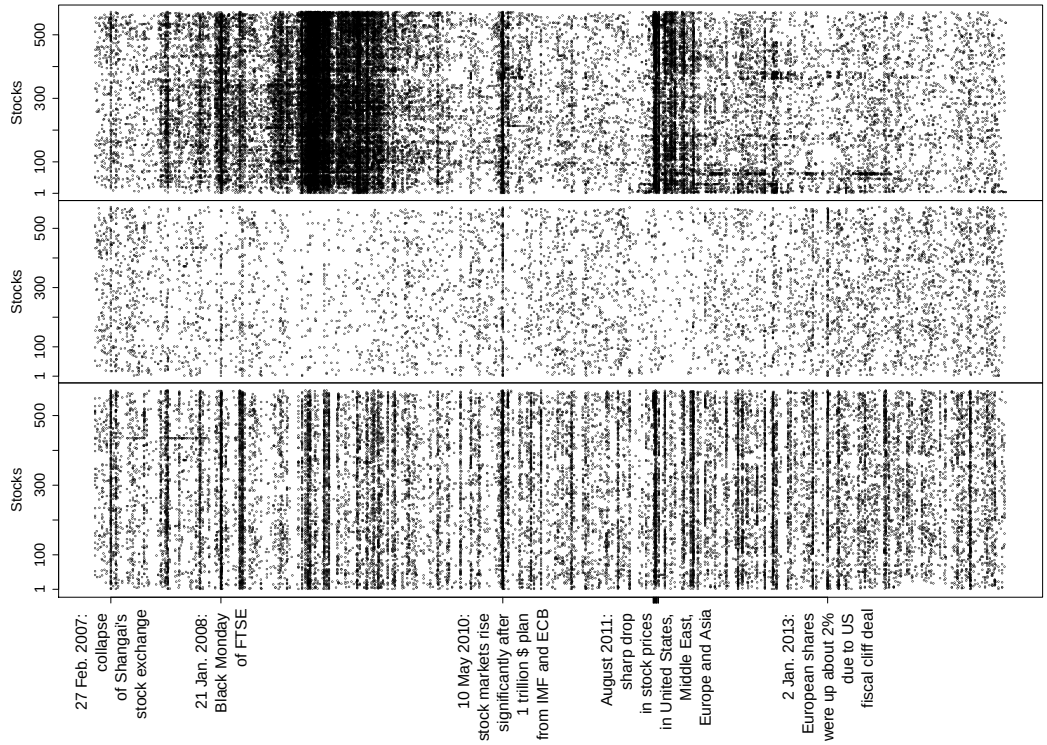

Figure 1: Each dot indicates a date between 10/1/2007 to 11/6/2014 (x-axis) in which at least one jump has been identified in observed daily log-returns of 571 stocks from the STOXX Europe 600 Index. The plot in the top panel is constructed by using an empirical method for detection of outliers in the data. The dots in the middle and in the bottom panels represent probabilities of jump greater than .5 for models with independent and jointly modelled across stocks intensities respectively.

be viewed as a generalisation of the Bates (1996), univariate SV model with jumps since it can be obtained as a special case by just assuming that jump intensities are independent across time and stocks. Furthermore, our computational techniques are relevant to other scientific areas such as, for example, neuroscience, whereas simultaneous recordings of neural spikes are often modelled through latent factor models (Buesing et al., 2014).

To illustrate our motivation, consider the daily returns from the stocks of the STOXX Europe 600 Index over a period 10/1/2007-11/6/2014. The top panel of Figure 1 has been created by empirically identifying a jump when the return exceeds three standard deviations, where the estimators of mean and variance of each series were robustly estimated as the median and the robust scale estimator of (Rousseeuw and Croux, 1993) respectively. The middle panel depicts a summary statistic of an SV model with jumps estimated separately for each stock return series: each dot denotes a stock return in which that particular day the probability of a jump in the series has been estimated to be greater than .5. Compared with the top panel, it provides a more sparse jump process indicating a successful separation of jumps from persistent moves modelled by the volatility process, a topic discussed in detail by, for example, Aït-Sahalia (2004). But the notable feature that inspired this work is that clustering and inter-dependence of jump intensities is evident in the middle panel and, interestingly, days in which jumps have been simultaneously identified in a large number of stocks coincide with days of events that affected the financial markets worldwide as pointed out at the x-axis of the plots. Thus, one may attempt joint, Markovian modelling of jump intensities across time which

might reveal some predictability of jumps. Indeed, the results of our proposed joint modelling formulation provides strong evidence for better out-of-sample predictability with estimated jump probabilities higher than .5 depicted in the lower panel of Figure 1.

We propose a modelling approach for the jump processes of SV models in which the time- and cross-sectional dependence of the jump intensities are driven by a latent, common across stocks, dynamic factor model. We carefully specify informative prior distributions to the parameters of the factor model so that the implied priors for the jump intensities in each stock are in concordance with the well-known (Eraker et al., 2003) a priori expectation for one jump every few months. Following Bhattacharya and Dunson (2011), we do not impose the usual (Aguilar and West, 2000) identifiability constraints on the factor loadings parameters. Our model is an alternative to the jump processes of SV models with jumps proposed by Aït-Sahalia et al. (2015). In their approach, the SV model is combined with a multivariate Hawkes process (Hawkes, 1971) which models the propagation of jumps over time and across assets by introducing a feedback from the jumps to their intensities and back. Their proposed model is estimated with a generalised version of the method of moments.

Our Bayesian inference implementation strategy is based on an MCMC algorithm that alternates sampling of the latent volatility process and its parameters with sampling of the jump process and its parameters. Both these steps require dawings of the high-dimensional paths of the latent volatilities and factors from their full conditional distributions. By noting that the target distributions are proportional to the product of intractable, non-linear likelihood functions with Gaussian priors we utilize the sampler proposed by Titsias and Papaspiliopoulos (2018) to simultaneously draw the whole path of each latent process at each MCMC iteration. This is an important feature of our methods for the following reasons. First updating each state of a volatility process separately results in Markov chain samplers with slow mixing (Shephard and Kim, 1994). Second, although Chib et al. (2002) and Nakajima and Omori (2009) use a mixture of Gaussian distributions approximation to also simultaneously update the whole volatility path, their methods rely on an importance sampling step which is quite problematic, see Johannes et al. (2009) for a detailed discussion and Section 3.4 of this paper for a simulation that illustrates this issue.

To compare the predictive performance of two different models by estimating predictive Bayes factors (Geweke and Amisano, 2010) we use sequential Monte Carlo methods. Since this turns out to be problematic in the case of jointly modelled jump intensities, we utilize the annealed importance sampling method developed by Neal (2001) to construct a sequential importance sampling algorithm in order to obtain the desired estimates.

The structure of the remaining of the paper is as follows. Section 2 presents our proposed modelling framework and Section 3 describes the methods that we develop to conduct Bayesian inference for the proposed model. In Section 4 we present the computational techniques that we use to assess the predictive performance of the proposed model. Section 5 presents results and insights from the application of our methods on the real dataset and Section 6 concludes with a small discussion.

2 Latent jump modelling

2.1 Notation

We denote the univariate Gaussian distribution with mean m and variance s^2 by $\mathcal{N}(m, s^2)$, and its density evaluated at x by $\mathcal{N}(x|m, s^2)$; $\mathcal{N}_d(X|M, S)$ denotes the density of the d-variate normal distribution with mean M and covariance matrix S evaluated at X; $Gam(\alpha, \beta)$ and $IGam(\alpha, \beta)$ denote the gamma, with mean α/β, and inverse gamma, defined as the reciprocal of gamma, distributions respectively; $U(\alpha, \beta)$ the uniform distribution on interval (α, β); $\Gamma(\cdot)$ denotes the gamma function. Index i is used for stocks and index t for times, e.g., r_{it} is the ith stock return at day t; when we introduce factor models index k denotes factor. Upper case letters denote vectors and matrices, and lower case letters denote scalars. For a collection of variables indexed by time or stock and denoted by a lower case character, the corresponding upper case character denotes their vector or matrix representation, e.g. R denotes the matrix whose elements are r_{it}; $R_i = (r_{i1}, \ldots, r_{iT})$, and $R_t = (r_{1t}, \ldots, r_{pt})$ are then generic rows and columns of R; all vectors are understood as column vectors. A subscript, of the form $(t+1) : (t+\ell)$, where ℓ is a positive integer, indicates the collection of vectors with subscripts $t+1, \ldots, t+\ell$; e.g. $R_{t+1:t+\ell}$ is the collection of vectors $R_{t+1}, \ldots, R_{t+\ell}$. Finally, the transpose of a vector or matrix is denoted by a prime; e.g. R' is the transpose of R.

2.2 Stochastic volatility with jumps

We model stock returns as

$$r_{it} = \exp(h_{it}/2)\epsilon_{it} + \sum_{\kappa=1}^{n_{it}} \xi_{it}^{\kappa}, \quad \epsilon_{it} \sim \mathcal{N}(0,1), \quad t = 1, \ldots, T \tag{1}$$

$$h_{it} = \mu_i + \phi_i(h_{i,t-1} - \mu_i) + \sigma_{i\eta}\eta_{it}, \quad \eta_{it} \sim \mathcal{N}(0,1), \quad t = 1, \ldots, T \tag{2}$$

with $h_{i0} \sim \mathcal{N}(\mu_i, \sigma_{i\eta}^2/(1-\phi_i^2))$ and ϵ_{it} and η_{it} are independent. The number of jumps of the ith stock at time t follow a Poisson distribution,

$$n_{it} \sim Poisson(\Delta_{it}\lambda_{it}), \tag{3}$$

where λ_{it} denotes the corresponding jump intensity, Δ_{it} is a time-increment associated to each return and $\xi_{it}^{\kappa} \sim \mathcal{N}(\mu_{i\xi}, \sigma_{i\xi}^2)$ is the κth jump size. We explicitly take into account that stock returns are computed over varying time increments Δ_{it}, such as in-between weekends and holidays, hence λ_{it} is the daily jump intensity.

We model the parameters μ_i, ϕ_i and $\sigma_{i\eta}^2$ of each log-volatility process and the parameters $\mu_{i\xi}$ and $\sigma_{i\xi}^2$ of the jump sizes as independent across stocks. For μ_i, ϕ_i and $\sigma_{i\eta}^2$ we follow the related literature (Kim et al. (1998), Omori et al. (2007), Kastner and Frühwirth-Schnatter (2014)) and take $\mu_i \sim \mathcal{N}(0, 10)$, $(\phi_i + 1)/2 \sim Beta(20, 1.5)$ and $\sigma_{i\eta}^2 \sim Gam(1/2, 1/2)$.

For $\mu_{i\xi}$ and $\sigma_{i\xi}^2$ we choose proper priors. In fact, improper priors will lead to improper posterior. This is due to the fact that according to the model there is positive probability that there are no jumps for the whole period of time, i.e., the event $n_{it} = 0$ for all t, has positive prior probability, and in such an event, $\mu_{i\xi}$ and $\sigma_{i\xi}^2$ are unidentifiable in the likelihood. Instead, we take $\sigma_{i\xi}^2 \sim IGam(3, \text{range}_i^2/18)$, where $\text{range}_i = \max_t\{r_{it}\} - \min_t\{r_{it}\}$, and $\mu_{i\xi} \sim \mathcal{N}(0, 5\text{range}_i^2)$. In this specification, $\text{E}[\sigma_{i\xi}^2] = \text{range}_i^2/36$, hence we match the variance of jump sizes with a quantity that relates to the sample variance of

returns. A same reasoning is followed for setting the prior variance of $\mu_{i\xi}$ and we have found that the choice of the multiplier, here taken 5, is not crucial in the analysis.

2.2.1 Modelling jumps independently across stocks and time

By modelling Λ_i as independent random vectors, we obtain p independent univariate SV with jumps models. Additionally, each vector can be taken as one of independent elements, $\lambda_{it} \sim Gam(\delta, c)$, in which case one obtains the models described, among others, by Bates (1996), Chib et al. (2002) and Eraker et al. (2003). It follows that n_{it} has, marginally with respect to λ_{it}, a negative binomial distribution with density

$$p(n_{it}) = \frac{\Gamma(\delta + n_{it})}{\Gamma(\delta) n_{it}!} \beta^{\delta} (1-\beta)^{n_{it}}, \tag{4}$$

where $\beta = c/(c + \Delta_{it})$. The mean of the distribution is $\delta(1-\beta)/\beta$ and the variance is $\delta(1-\beta)/\beta^2$. Taking $\delta \leq 1$ ensures that the density is monotonically decreasing. This is a feature we are interested in: jumps are meant to capture infrequently large price movements, and a priori we expect that with highest probability there is no jump and it is more likely to have less than more jumps, if any. We can choose c such that the probability of no jump at a given time increment Δ_{it}, is at least $\gamma \in (0,1)$, a user-specified threshold. Taking for example $\delta = 1$ and $c = 50$ results in $\gamma = 0.98$ with $\mathrm{E}(\lambda_{it}) = 0.02$. These choices are consistent with the prior expectation, in financial applications, for one jump every few months; see for example Eraker et al. (2003) for a more detailed discussion.

2.3 Joint modelling of jumps

To capture the dependence of the jumps over time and across stocks we model λ_{it} by using a dynamic factor model which is specified as follows. We first transform λ_{it} to y_{it} via

$$\lambda_{it} = \lambda^* (1 + e^{-y_{it}})^{-1}, \tag{5}$$

which implies that $\lambda_{it} < \lambda^*$. Then, we model the time-dependence of jumps via K latent factors F_t modelled as independent autoregressive processes. The cross-sectional dependence of jumps is captured by a $p \times K$ matrix of factor loadings W and a $p \times 1$ vector B. The joint model specification is

$$Y_t = B + W F_t, \tag{6}$$

$$F_t = A F_{t-1} + E_t, \quad t = 1, \ldots, T \tag{7}$$

$$F_0 \sim \mathcal{N}_K(0, \Sigma_F) \quad E_t \sim \mathcal{N}(0, I) \tag{8}$$

where A is a $K \times K$ diagonal matrix with elements $\alpha_1, \ldots, \alpha_K$ and Σ_F is a $K \times K$ diagonal matrix with elements $1/(1-\alpha_1^2), \ldots, 1/(1-\alpha_K^2)$.

It is known that parameters of latent factor models are only identifiable up to certain transformations, such as orthogonal rotations and sign changes (Aguilar and West, 2000). However, latent factor models are also used as low-rank predictive models in which case the out-of-sample prediction is of prior importance; see Bhattacharya and Dunson (2011). This is the perspective we adopt here, where we try to borrow strength from the information in large panels of stocks in order to predict with higher accuracy individual jumps by using a parsimonious factor model. Therefore, our prior specification imposes no identifiability constraints on the parameters of the factor model and are taken as $w_{ik} \overset{iid}{\sim} \mathcal{N}(0, \sigma_w^2)$, $b_i \overset{iid}{\sim} \mathcal{N}(\mu_b, \sigma_b^2)$, and $\alpha_k \overset{iid}{\sim} U(-1, 1)$.

The specification of the hyperparameters of the priors assigned on the parameters A, B and W require extra care because they affect, through (5), the imposed prior on λ_{it}. The fact that our modelling formulation of Section 2.2.1 was based on informative priors $Gam(1, 50)$ provides a way to elicit informative prior specification of the hyperparameters σ_w^2, σ_b^2 and μ_b. We first set $\lambda^* = 0.15$ which is the 0.0005 upper quantile of the $Gam(1, 50)$ density. Then, we choose the hyperparameters σ_w^2, μ_b and σ_b^2 to be such that the mean, the variance and the mode for the resulting prior on each λ_{it} in (5) are comparable with the corresponding quantities of the $Gam(1, 50)$ distribution. Based on calculations presented in the supplementary material we set $\sigma_w^2 = 0.5$, $\sigma_b^2 = 1$ and $\mu_b = -5$. In Table 1 we display the quantities of interest in the two prior distributions. To evaluate the differences of the two priors we also examine the impact of different choices for their hyperparameters in the forecasts of future observations. See in the supplementary material for the related results.

Table 1: Mean, variance and mode of the prior distribution assumed for the jump intensities λ_{it} in the independent model and for the prior induced by the dynamic factor model.

Prior for λ_{it}	Mean	Variance	Mode
$Gam(1, 50)$	0.020	0.00040	0
Factor model	0.003	0.00004	0.0001

3 MCMC sampling from the posterior of interest

Let $\Theta_i = (\mu_i, \phi_i, \sigma_{i\eta}^2, \mu_{i\xi}, \sigma_{i\xi}^2, b_i)$. Our interest lies on the joint posterior of parameters and latent states $p(H, N, \Xi, W, \Theta, F, A|R)$. To draw samples from the posterior of interest we design an MCMC algorithm which proceeds as follows. At each MCMC iteration we perform stock-specific updates of H_i, N_i, Ξ_i, W_i, Θ_i and then we update the latent factors F and their parameters A.

An important characteristic of the designed MCMC algorithm is that before updating the log-volatilities H_i and the number of jumps N_i we integrate out the jump sizes Ξ_i. This results to an efficient simultaneous update of the log-volatilities vector and a direct sampling of the full conditional of the number of jumps N_i. Algorithm 1 summarizes the steps of the MCMC sampler. The algorithm presents how the full conditional densities of vectors of parameters are sampled by exploiting the conditional independence structure of the hierarchical model defined by the equations (1)-(3) and (5)-(8). In Sections 3.1, 3.2 and 3.3 we will describe in detail the more elaborate steps $4, 6, 7$ and 12 of Algorithm 1 whereas the details of the other sampling steps are given in the supplementary material.

Algorithm 1 MCMC that targets $p(H, N, \Xi, W, \Theta, F, A|R)$.

1: Set the number of iterations S.
2: **for** $s = 1, \ldots, S$ **do**
3: **for** $i = 1, \ldots, p$ **do**
4: Sample from $p(H_i, \phi_i, \sigma^2_{i\eta}|\mu_i, \mu_{i\xi}, \sigma^2_{i\xi}, N_i, R_i)$ by using Metropolis-Hastings.
5: Sample from $p(\mu_i|H_i, \phi_i, \sigma^2_{i\eta})$ which is a Gaussian density.
6: Sample from $p(N_i|H_i, W_i, F, R_i, b_i, \mu_{i\xi}, \sigma^2_{i\xi})$ by using rejection sampling.
7: Sample from $p(\Xi_i|H_i, N_i, R_i, \mu_{i\xi}, \sigma^2_{i\xi})$ which is a Gaussian density.
8: Sample from $p(\mu_{i\xi}|N_i, \Xi_i, \sigma^2_{i\xi})$ which is a Gaussian density.
9: Sample from $p(\sigma^2_{i\xi}|N_i, \Xi_i, \mu_{i\xi})$ which is a inverse Gamma density.
10: Sample from $p(b_i, W_i|N_i, F)$ by using Metropolis-Hastings.
11: **end for**
12: Sample from $p(F, A|N, B, W)$ by using Metropolis-Hastings.
13: **end for**

We also emphasize that by switching off certain steps of Algorithm 1 we obtain a novel MCMC sampler for existing models. More precisely, Bayesian inference for univariate SV with jumps models can be conducted if the step in the 10th line is skipped and the step in the 12th line replaced with the step of drawing the independent jump intensities of the model. Switching off the steps in lines $6-10$ and 12 results in an MCMC algorithm for univariate SV without jumps models. By removing the steps in lines $4-9$ of Algorithm 1 we obtain a sampler for a Cox model (Cox, 1955) with intensities driven by latent factors.

3.1 Sampling the number of jumps and jump sizes

To sample each n_{it} we first integrate out the jump sizes ξ_{it} and then we construct a rejection sampling algorithm to draw from $p(n_{it}|h_{it}, \mu_{i\xi}, \sigma^2_{i\xi}, b_i, W_i, F_t, r_{it})$. This algorithm is based on the following result.

Proposition 1. For each $i = 1, \ldots, p$ and $t = 1, \ldots, T$ the discrete distribution with probability mass function $p(n_{it}|h_{it}, \mu_{i\xi}, \sigma^2_{i\xi}, b_i, W_i, F_t, r_{it})$ is log-concave for $n_{it} \geq 1$.

Proof. See the supplementary material. □

For the remaining of this subsection let $p(n)$ denote $p(n_{it}|h_{it}, \mu_{i\xi}, \sigma^2_{i\xi}, b_i, W_i, F_t, r_{it})$. Due to Proposition 1, the target distribution is a unimodal distribution. Let m be an integer at the right of the mode of the distribution. We define the probability mass function (pmf)

$$q_m(n) = \begin{cases} \frac{p(n)}{c}, & \text{if } n < m \\ \frac{p(m)}{c}\left(\frac{p(m+1)}{p(m)}\right)^{n-m}, & \text{if } n \geq m \end{cases}$$

with $c = \sum_{n=0}^{m-1} p(n) + p(m)^2/(p(m) - p(m+1))$. The pmf $q_m(n)$ is proportional to $p(n)$ for $n < m$ and proportional to the geometric pmf with parameter, $(p(m) - p(m+1))/p(m)$, for $n \geq m$. By noting that if a random variable X follows the exponential distribution with parameter, $\log(p(m)) - \log(p(m+1))$, then $\lfloor X \rfloor$ follows the geometric distribution with parameter $(p(m) - p(m+1))/p(m)$, we observe that $q_m(n)$, for $n \geq m$, corresponds to an exponential density that goes through the points $(m, p(m))$ and $(m+1, p(m+1))$. Log-concavity of the distribution implies that $cq_m(n) \geq p(n)$, for $n \geq m$, since any straight line that goes through the points $(m, \log(p(m)))$ and $(m+1, \log(p(m+1)))$ bounds from

above the logarithm of $p(n)$. See Figure 2 for an illustration. A rejection algorithm to sample from the distribution with pmf $p(n)$ proceeds as summarized by Algorithm 2.

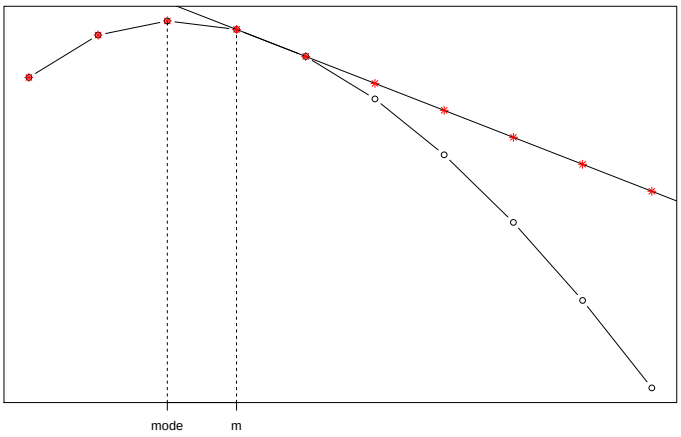

Figure 2: The pmf $p(n)$ and its bounding line in the log-scale. Circles: $\log(p(n))$, stars: $\log(q_m(n))$.

Algorithm 2 Rejection sampler for the number of jumps.

Generate $U \sim U(0,1)$
if $U \leq \frac{\sum_{n=0}^{m-1} p(n)}{c}$ **then**
generate n from $q_m(n)$ truncated from the right at the point $m-1$, by using the inversion method.
else
repeat:
generate $V \sim U(0,1)$
generate n from the geometric pmf with parameter $(p(m)-p(m+1))/p(m)$, truncated from the left at the point m.
until $Vcq_m(n) \leq p(n)$
end if
return n

3.1.1 Jump sizes

Drawing the jump sizes ξ_{it} given the number of jumps n_{it} is based on the following proposition. Let I_n and $\mathbf{1}_n$ be the $n \times n$ identity matrix and an n-dimensional vector with ones respectively.

Proposition 2. According to the model defined in (1) and (2), $p(\xi_{it}|n_{it}, h_{it}, \mu_{i\xi}, \sigma^2_{i\xi}, r_{it})$ is equal to the Gaussian density

$$\mathcal{N}_{n_{it}}\left(\xi_{it} \,\middle|\, (\mu_{i\xi}\frac{e^{h_{it}}}{e^{h_{it}}+n_{it}\sigma^2_{i\xi}} + \frac{r_{it}}{e^{h_{it}}+n_{it}\sigma^2_{i\xi}}n_{it}\sigma^2_{i\xi})\mathbf{1}_{n_{it}}, \sigma^2_{i\xi}I_{n_{it}} - \frac{\sigma^4_{i\xi}}{e^{h_{it}}+n_{it}\sigma^2_{i\xi}}\mathbf{1}_{n_{it}}\mathbf{1}'_{n_{it}}\right),$$

for each $i = 1, \ldots, p$.

Proof. See the supplementary material. □

3.2 Sampling latent volatilities and their parameters

In the 4th line of Algorithm 1 we sample the whole vector H_i jointly with the parameters ϕ_i and $\sigma_{i\eta}^2$. We make this choice because the prior covariance matrix of H_i depends on the values of these parameters and this dependence results in high correlation between H_i and $\phi_i, \sigma_{i\eta}^2$ (Titsias and Papaspiliopoulos, 2018). To sample the latent log-volatility paths H_i and the parameters $\phi_i, \sigma_{i\eta}^2$ we first integrate out the jump sizes Ξ_i from the target density and we obtain

$$p(H_i, \phi_i, \sigma_{i\eta}^2 | \mu_i, \mu_{i\xi}, \sigma_{i\xi}^2, N_i, R_i) \propto \pi(H_i|\phi_i, \sigma_{i\eta}^2, \mu_i)\pi(\phi_i)\pi(\sigma_{i\eta}^2) \prod_{t=1}^{T} \mathcal{N}(r_{it}|n_{it}\mu_{i\xi}, n_{it}\sigma_{i\xi}^2 + e^{h_{it}}) \tag{9}$$

where $\pi(H_i|\phi_i, \sigma_{i\eta}^2, \mu_i)$ denotes the prior of the latent autoregressive log-volatility process defined in (2) and $\pi(\phi_i)$ and $\pi(\sigma_{i\eta}^2)$ denote the densities of the priors for the parameters ϕ_i and $\sigma_{i\eta}^2$ as described in Section 2.2.

To draw samples from (9) we utilize the auxiliary gradient-based sampler proposed by Titsias and Papaspiliopoulos (2018). The sampler is a Metropolis-Hastings step that combines auxiliary variables, the Gaussian prior of the latent path H_i and a Taylor expansion of the intractable likelihood of the observations. After the update of H_i, ϕ_i and $\sigma_{i\eta}^2$ we condition on their new values in order to draw μ_i from its full conditional, Gaussian, distribution. See the supplementary material for details.

To further improve the sampling of the parameters $\phi_i, \sigma_{i\eta}^2, \mu_i$ we follow Kastner and Frühwirth-Schnatter (2014) and we use the, so-called, ancillarity-sufficiency interweaving strategy proposed by Yu and Meng (2011). After drawing H_i, $\phi_i, \sigma_{i\eta}^2$ and μ_i as described above we transform the vector H_i to a new vector $\tilde{H}_i$ via $\tilde{h}_{it} = (h_{it} - \mu_i)/\sigma_{i\eta}$. Then, we draw again from the conditional densities of $\phi_i, \sigma_{i\eta}^2$ and μ_i which are now conditioned on $\tilde{H}_i$ rather than on H_i. These last updates are performed through univariate Metropolis-Hastings steps. Before moving to the next step of the MCMC algorithm we also update deterministically the path H_i by setting $h_{it} = \sigma_{i\eta}\tilde{h}_{it} + \mu_i$; see in the supplementary material for illustrations of the MCMC performance achieved by using the interweaving strategy.

3.3 Sampling latent jump intensities

Following the modelling approach presented in Section 2.2.1 it is easy to see that the full conditional density of λ_{it} is an independent $Gam(n_{it} + \delta, c + \Delta_{it})$.

In the case of modelling the jump intensities by using the dynamic factor model defined in (5)-(8) we need to draw samples from the distribution with density

$$p(F, A|N, B, W) \propto \pi(A)\mathcal{N}_{K(T+1)}(F|0, C_A) \prod_{t=1}^{T}\prod_{i=1}^{n} e^{-\lambda_{it}}\lambda_{it}^{n_{it}}, \tag{10}$$

where $\lambda_{it} = \lambda^*/\big(1 + \exp(-b_i - W_i' F_t)\big)$, $\pi(A)$ denotes the prior density of A and C_A is the covariance matrix, which depends on the values of the parameters A, of the autoregressive process defined in (7) and (8). We utilize again the auxiliary gradient-based sampler of Titsias and Papaspiliopoulos (2018) to obtain the desired samples. We also employ the ancillarity-sufficiency interweaving strategy developed by Yu and Meng (2011) to reduce the autocorrelation of the samples of the parameters in A. After the end of the auxiliary gradient-based sampler, we set $\Gamma_0 = DF_0$, where D is a $K \times K$ diagonal matrix with

elements $(1-\alpha_k^2)^{1/2}$ and $\Gamma_t = F_t - AF_{t-1}$. Then we use a random walk Metropolis-Hastings step that targets the density $p(A|\Gamma, N, B, W)$. See the supplementary material for more details and illustration for the improvements of MCMC mixing.

3.4 Approximation by using a mixture of Gaussian distributions

For a given stock, sampling the whole path of the latent log-volatilities H_i at each MCMC iteration could be performed by utilizing the methodology developed by Chib et al. (2002) and Nakajima and Omori (2009). Their methods are based on the idea of Kim et al. (1998) to approximate a SV without jumps model by using a mixture of Gaussian distributions.

The advantage of their approach is that the update of H_i can be conducted by drawing samples directly from its full conditional distribution which is a $(T+1)$-variate Gaussian with tridiagonal inverse covariance matrix. The jumps sizes and their parameters should be updated after integrating out the mixture component indicators from the likelihood of the approximated model, otherwise convergence of the corresponding MCMC algorithm will be slow (Nakajima and Omori, 2009). Integrating out the mixture component indicators results in a partially collapsed Metropolis-Hastings within Gibbs sampler which has to be implemented with care since permuting the order of the updates can change the stationary distribution of the chain (Van Dyk and Jiao, 2015). If, for example, the update of the mixture indicators is conducted in the step which is intermediate between the update of the latent log-volatilities and the update of their parameters, then the Markov chain will not converge to the desired stationary distribution.

More importantly, by using the methods developed in Chib et al. (2002) and Nakajima and Omori (2009), samples from the posterior of the parameters of the exact model are obtained with an importance sampling step. In this step, weights that are equal to the ratio between the likelihood of the exact and the approximated model, are assigned to the MCMC samples obtained from the approximated model. Although in the case of SV models without jumps the described importance sampling step works without problems, it is known (Johannes et al., 2009) that in the case of models with jumps such weights may have large variance since a few or only one of them could be much larger than the others. Thus, the output of the importance sampling step will not be suitable for any Monte Carlo calculation. For an illustration of the problem we conducted a simulation experiment in which we calculated the required importance sampling weights for SV models with and without jumps. In particular, we simulated $p = 1$ time series with $T = 1,500$ log-returns from each model by utilizing equations (1)-(2) and omitting the jump component in (1) to simulate from the model without jumps. In the case of the SV model with jumps we assumed independent jump intensities over time following the modelling approach presented in Section 2.2.1. We chose values for the parameters of the log-volatility process which are common in the financial literature (Kim et al., 1998); $\mu_1 = -0.85, \phi_1 = 0.98$ and $\sigma_{1\eta} = 0.15$ and we simulated the sizes of possible jumps from the Gaussian distribution with $\mu_{1\xi} = 0$ and $\sigma_{1\xi} = 3.5$ (Eraker et al., 2003). Then, we drew $3,000$ (thinned) posterior samples of the parameters and latent states of the approximated models by using the corresponding MCMC algorithms.

Figure 3 displays the log-normalised importance sampling weights plus the logarithm of their number. The variance of the normalized importance sampling weights should become smaller as the importance sampling approximation improves. This implies that the histograms in Figure 3 should be centered at zero in the case of an accurate enough approximation; see the corresponding plots in Kim et al. (1998) and Omori et al. (2007). The left panel of Figure 3 indicates that there is little effect of the mixture approximation

in the case of the model without jumps, but the right panel points out that this is not true in the case of the SV with jumps model. The bad performance of the approximation strategy in models with jumps is due to the fact that the mixture model has been proposed by Kim et al. (1998) and improved by Omori et al. (2007) for SV models without jumps. By adding a jump component in the model the mixture approximation has to take into account the uncertainty in the estimation of the jump process and this is not the case in the approximation used by Chib et al. (2002) and Nakajima and Omori (2009).

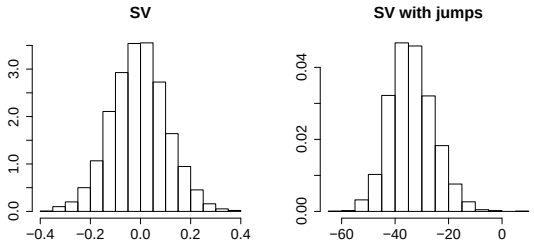

Figure 3: Histograms of normalized weights multiplied by their number, in the log-scale, calculated as the ratio between the likelihood of the exact and the approximated, by a mixture of Gaussian distributions, model in the case of stochastic volatility (SV) models with and without jumps.

4 Out-of-sample forecasting

To compare the predictive performance of different SV models we develop a computational framework for out-of-sample forecasting. Based on T in-sample and ℓ out-of-sample observations we estimate the predictive Bayes factors (Geweke and Amisano, 2010) by utilizing sequential Monte Carlo methods (Doucet et al., 2001). For the remaining of this section we assume that the parameters Θ, A, W are fixed to their in-sample estimated posterior means and we omit them from our notation.

4.1 Prediction with independent SV models

Let $\mathbb{M}_{\text{svj}}$ and $\mathbb{M}_{\text{sv}}$ be the model that consists of p independent, univariate SV models with and without jumps respectively. Let also $R_{1:T}$ and $R_{T+1:T+\ell}$ denote in-sample and out-of-sample observations. We compare the predictive performance of $\mathbb{M}_{\text{sv}}$ and $\mathbb{M}_{\text{svj}}$ by considering predictive Bayes factors defined as

$$\frac{p(R_{T+1:T+j}|R_{1:T},\mathbb{M}_{\text{svj}})}{p(R_{T+1:T+j}|R_{1:T},\mathbb{M}_{\text{sv}})},\quad j=1,\ldots,\ell. \tag{11}$$

For both $\mathbb{M}_{\text{svj}}$ and $\mathbb{M}_{\text{sv}}$ we have that

$$p(R_{T+1:T+j}|R_{1:T})=\prod_{i=1}^{p}\prod_{t=T+1}^{T+j}p(r_{it}|r_{i,1:t-1}),$$

so to estimate the Bayes factors in (11) we need to estimate the quantities $p(r_{it}|r_{i,1:t-1})$. Based on the formula

$$p(h_{i,0:t}|r_{i,1:t})=\frac{p(h_{i,0:t-1}|r_{i,1:t-1})p(r_{it}|h_{it})p(h_{it}|h_{i,t-1})}{p(r_{it}|r_{i,1:t-1})}, \tag{12}$$

we use sequential Monte Carlo methods, also known as particle filters, to estimate the desired quantities. It is known, see for example in Chopin et al. (2013), that from the output of a particle filter algorithm that targets the posterior in (12) we also obtain an unbiased estimator for its normalizing constant. See in the supplementary material for the pseudocode of the particle filter algorithm that we employ to obtain unbiased estimators of the normalizing constants in (12).

4.2 Prediction by using joint modelled jump intensities

Let $\mathbb{M}_{\text{mvj}}$ be our proposed SV with jumps model, defined by equations (1)-(3) and (5)-(8). To compare the predictive performance of $\mathbb{M}_{\text{mvj}}$ with the performance of competing models by utilizing predictive Bayes factors, we need to estimate the marginal likelihood $p(R_{T+1:T+j}|R_{1:T}, \mathbb{M}_{\text{mvj}})$ for each $j = 1, \ldots, \ell$. This could be achieved with a particle filter algorithm, constructed similarly to the case of independent SV models, as described in section 4.1. However, in the case of $\mathbb{M}_{\text{mvj}}$, particle filter algorithms deliver poor estimations of the marginal likelihoods since the importance weights, calculated at each of their steps, have large variance and the estimation of $p(R_{T+1:T+j}|R_{1:T}, \mathbb{M}_{\text{mvj}})$ is based on a few samples with large weights.

We develop an alternative sequential importance sampling algorithm to obtain the desired estimates which is based on the annealed importance sampling (AIS) proposed by Neal (2001). AIS utilizes the method of simulated annealing (Kirkpatrick et al., 1983) to construct an importance sampling algorithm for drawing weighted samples from an unnormalized target distribution. The proposal distribution of the algorithm is based on a sequence of auxiliary distributions and Markov transition kernels that leave them invariant. As in any importance sampling algorithm, the sample mean of the importance weights is an estimation of the ratio between the normalizing constants of the target and proposal distributions. Moreover, the output of AIS can be used for the estimation of expectations with respect to any of the auxiliary distributions used for the construction of the algorithm. Here we exploit these features of AIS to estimate the marginal likelihoods. In the rest of the section we provide a detailed description of how we apply AIS, $\mathbb{M}_{\text{mvj}}$ is suppressed throughout except when it is necessary.

Let $g_0, \ldots, g_\ell$ be the required sequence of auxiliary distributions. By noting that each g_j must be proportional to a computable function and satisfy $g_j \neq 0$ wherever $g_{j-1} \neq 0$ we set $g_\ell(H_{0:T+\ell}, F_{0:T+\ell}) = p(H_{0:T+\ell}, F_{0:T+\ell}|R_{1:T+\ell})$ and

$$\begin{aligned} g_j(H_{0:T+\ell}, F_{0:T+\ell}) &= p(H_{0:T+\ell}, F_{0:T+\ell}|R_{1:T+j}) \\ &= p(H_{0:T+j}, F_{0:T+j}|R_{1:T+j}) \prod_{t=T+j+1}^{T+\ell} p(H_t|H_{t-1})p(F_t|F_{t-1}), \end{aligned}$$

for $j = 0, \ldots, \ell - 1$. Samples from g_j, $j \geq 1$, are obtained by drawing samples from g_{j-1} using a few iterations of Algorithm 1 and the Markov transition densities $p(H_{T+j}|H_{T+j-1})$ and $p(F_{T+j}|F_{T+j-1})$. Note that g_0 is easily sampled by utilizing samples already drawn from $p(H_{0:T}, F_{0:T}|R_{1:T})$. To compute the importance weights of the obtained samples we first note that g_j are proportional to the computable functions

$$g_j^*(H_{0:T+\ell}, F_{0:T+\ell}) = p(H_{0:T+j}, F_{0:T+j}, R_{1:T+j}) \prod_{t=T+j+1}^{T+\ell} p(H_t|H_{t-1})p(F_t|F_{t-1}),\ j < \ell,$$

and $g_\ell^*(H_{0:T+\ell}, F_{0:T+\ell}) = p(H_{0:T+\ell}, F_{0:T+\ell}, R_{1:T+\ell})$. Therefore, the importance weight for every sampled point is

$$\Omega_{T+j}^s = \Omega_{T+j-1}^s \frac{g_j^*(H_{0:T+\ell}^s, F_{0:T+\ell}^s)}{g_{j-1}^*(H_{0:T+\ell}^s, F_{0:T+\ell}^s)} = \Omega_{T+j-1}^s p(R_{T+j}|H_{T+j}^s, F_{T+j}^s),\ j \geq 2,$$

and $\Omega_{T+1}^s = p(R_{T+1}|H_{T+1}^s, F_{T+1}^s)$, $s = 1, \ldots, S$, where S is the number of samples obtained by AIS and

$$p(R_t|H_t^s, F_t^s) = \prod_{i=1}^p p(r_{it}|h_{it}^s, F_t^s) = \prod_{i=1}^p \sum_{n_{it}=0}^\infty \mathcal{N}(r_{it}|n_{it}\mu_{i\xi}, \exp(h_{it}^s) + n_{it}\sigma_{i\xi}^2) p(n_{it}|\Delta_{it}\lambda_{it}^s). \tag{13}$$

We also note that the infinite sum in (13) can be truncated without affecting the results of the algorithm (Johannes et al., 2009). See in the supplementary material for the pseudocode that we used to construct the described AIS algorithm.

The sample means of the weights calculated at each step of AIS provide estimators of the marginal likelihoods $p(R_{T+1:T+j}|R_{1:T})$. However, the resulting weights have large variance and the corresponding estimates will be inaccurate. Nevertheless, and in contrast with a particle filter that targets $p(H_{0:T+\ell}, F_{0:T+\ell}|R_{1:T+\ell})$, we can use the samples $\{H_{0:T+j}^s, F_{0:T+j}^s\}_{s=1}^S$ to estimate the quantities $p(r_{i,T+1:T+j}|R_{1:T})$ for each stock separately. We propose the following sampling procedure. First, note that

$$p(r_{it}|R_{1:t-1}) = \int p(r_{it}|h_{it}, F_t) p(h_{it}|h_{i,t-1}) p(F_t|F_{t-1}) p(h_{i,0:t-1}, F_{0:t-1}|R_{1:t-1})\, \mathrm{d}h_{i,0:t}\, \mathrm{d}F_{0:t}\,.$$

This implies that by using only the ith element of the product in (13) we obtain an estimator of the ith marginal of $p(R_{T+1:T+j}|R_{1:T})$ as follows. We set $\hat{p}(r_{i,T+1:T+j}|R_{1:T}) = \sum_{s=1}^S \omega_{i,T+j}^s / S$, where

$$\omega_{it}^s = \omega_{i,t-1}^s p(r_{it}|h_{it}^s, F_t^s),$$

and $\omega_{i,T+1}^s = 1$. Then, $\hat{p}(r_{i,T+1:T+j}|R_{1:T})$ is an accurate estimator of $p(r_{i,T+1:T+j}|R_{1:T}, \mathbb{M}_{\text{mvj}})$ since it is based on importance sampling weights with reduced variance and can be used for the estimation of Bayes factors of the form

$$\frac{p(r_{i,T+1:T+j}|R_{1:T}, \mathbb{M}_{\text{mvj}})}{p(r_{i,T+1:T+j}|R_{1:T}, \mathbb{M}_{\text{svj}})}, \tag{14}$$

where the denominator can be estimated with the methods presented in Section 4.1. Finally, by assuming that the joint predictive density of all returns is estimated by the product of their marginal predictive densities, as in the independent across stocks SV with jumps models, we estimate the Bayes factors

$$\frac{\prod_{i=1}^p p(r_{i,T+1:T+j}|R_{1:T}, \mathbb{M}_{\text{mvj}})}{\prod_{i=1}^p p(r_{i,T+1:T+j}|R_{1:T}, \mathbb{M}_{\text{svj}})}. \tag{15}$$

5 Real data application

Here we present results from the application of the developed methodology on real data. In the supplementary material we present results from simulation studies used to test our methodology. Further specifics of the data analysis, such as parameter estimates for the log-volatility processes, can be found in Alexopoulos (2017).

5.1 Data and MCMC details

We applied the developed methods on time series consisted of daily log-returns from stocks of the STOXX Europe 600 Index downloaded from Bloomberg between 10/1/2007 to 11/6/2014. We removed 29 stocks by requiring each stock to have at least 1000 traded days and no more than 10 consecutive days with unchanged price. The final dataset consisted of $p = 571$ stocks and 1947 traded days. We used daily log-returns observed in the first $T = 1917$ days as in-sample observations to estimate all models described through our article and the last $\ell = 30$ observations to test their predictive ability. In our dynamic factor model we used $K = 2$ factors. We ran all the MCMC algorithms for $100,000$ iterations using a burn-in period of $30,000$ sampled points and we collected $1,000$ (thinned) posterior samples.

5.2 Results

5.2.1 Bayesian inference for parameters and latent states

The middle and bottom panels of Figure 1 present a summary visual outcome of models $\mathbb{M}_{\text{mvj}}$ and $\mathbb{M}_{\text{svj}}$. Figure 4 depicts the posterior means of the factor paths along with the posterior distributions of their persistent parameters α_1 and α_2. The size of α_1 and α_2 indicates strong evidence of autocorrelation in the factor processes. Figure 5 demonstrates the ability of our Bayesian methods to estimate several quantities of interest. It depicts the posterior probabilities of the number of jumps for each weekday. This is clearly related with the parameter Δ_{it} in (3) which takes into account the time distance of two successive observations.

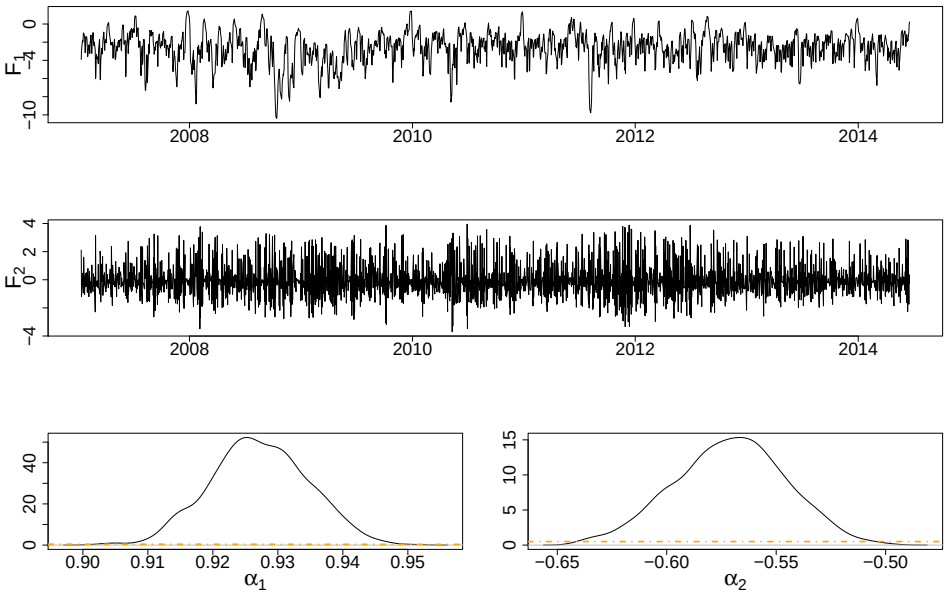

Figure 4: First and second row: posterior mean of the paths of the two autoregressive factors F_1 and F_2 that we used to model the evolution of the jump intensities across stocks and across time. Third row: posterior (solid lines) and prior (dotted lines) distributions of the persistent parameters α_1 and α_2.

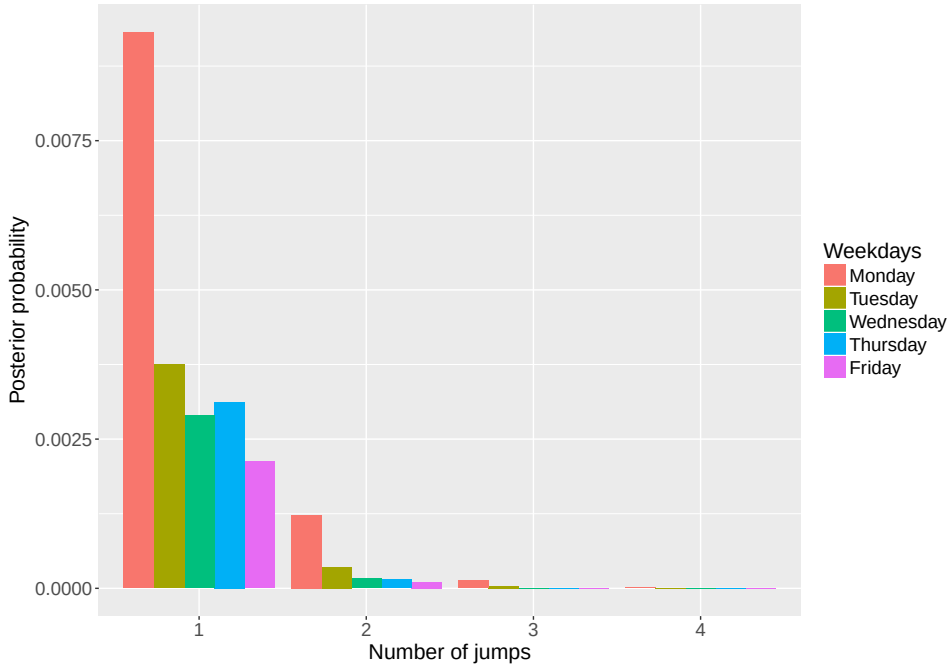

Figure 5: Posterior distributions for the number of jumps in 571 stocks from the STOXX 600 Europe Index conditional on the weekday. The order in which the bars are appearing from left to right is with respect to the order of the weekdays. The probabilities of zero jumps are not displayed.

5.2.2 Predictive performance

Figure 6 compares the predictive performance of independent, univariate SV models with and without jumps. The Figure is constructed by considering the logarithms of the Bayes factors defined in (11) for the $\ell = 30$ out-of-sample observations of our dataset. Thus, positive log-Bayes factor indicates that the predictions obtained with univariate SV models with jumps are more accurate than those of SV models without jumps. According to Kass and Raftery (1995) a value of the log-Bayes factor greater than 5 provides strong evidence in favour of the model with jumps. The increasing nature of the log-Bayes factor reassures that as more data are collected and as more jumps are identified in stock returns, the evidence that the model with jumps outperforms the model without jumps increases. In Figure 7 we compare the predictive performance of our proposed SV with jumps model in which the jump intensities are modelled by using a dynamic factor model with SV models in which the jump intensities are independent over time and across stocks. The Figure compare the two models by presenting the logarithms of the Bayes factors in (15). Clearly, there is strong evidence that our proposed model outperforms the independent modelling of stock returns with a SV with jumps formulation.

6 Discussion

We have developed a general modelling framework together with carefully designed MCMC algorithms to perform Bayesian inference for SV models with Poisson-driven jumps. We have shown that for the data we applied our models there is evidence that, with respect to predictive Bayes factors, (i) univariate SV models with jumps outperform univariate SV models without jumps and (ii) models that jointly model jump intensities with a dynamic

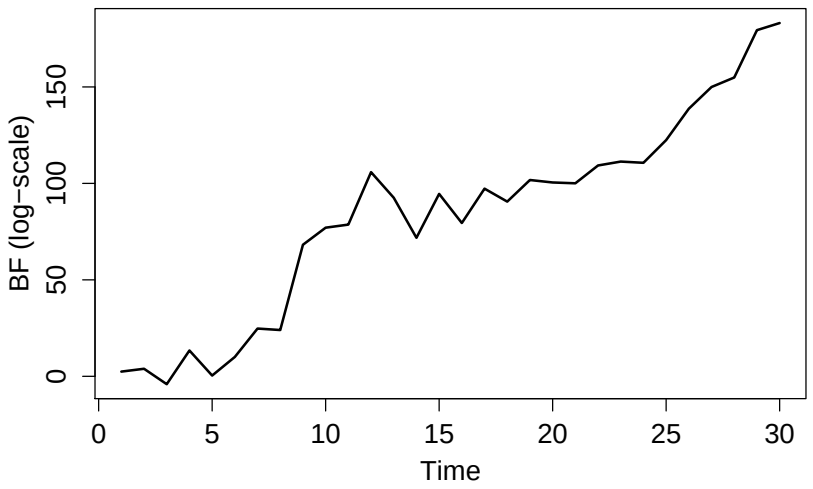

Figure 6: Logarithm of Bayes factors (BF) in favour of univariate SV models with jump intensities modelled independently across time and stocks against univariate SV without jumps models for $\ell = 30$ out-of-sample observations. The displayed Bayes factors are defined in (11).

factor model outperform SV models with jumps that are applied independently across stocks. We feel that (ii) is an interesting result that adds considerable insight to the growing literature of modelling financial returns with jumps, adding to the observation by Aït-Sahalia et al. (2015) that there is indeed predictability in jump intensities.

There are various issues that have not been addressed in this paper. The interesting problem of choosing the number of factors has not been dealt with and left for future research. Other modelling aspects include the relaxing of independence of E_t in (8) or the assumption that A is lower diagonal as in the models proposed in Dellaportas and Pourahmadi (2012).

A series of interesting financial questions can be addressed with our models by exploiting the fact that panel of stock returns carry additional, possibly important, information. For example, in our dataset one can explore the effects of country and sector effects or could investigate whether the joint tail risk dependence does or does not change before, during and after the financial crisis in Europe. We feel that our proposed models together with our algorithmic guidelines will serve as useful tools for such future research pursuits in financial literature.

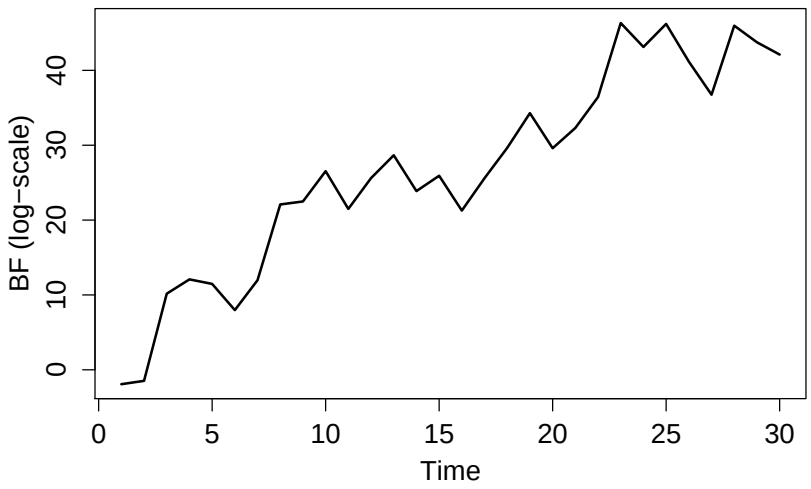

Figure 7: Bayes factors (BF) in the log-scale for SV models with jointly (numerator) and independently (denominator) modelled jumps intensities. The x-axis in both plots refers to the $\ell = 30$ out-of-sample observations. The displayed Bayes factors are defined in (15).

Acknowledgements

The authors gratefully acknowledge the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: ARISTEIA-LIKEJUMPS-436, and the Alan Turing Institute for EPSRC grant EP/N510129/1.

Supplementary material

Choosing hyperparameters for the factor model

We have assumed that $b_i \overset{iid}{\sim} \mathcal{N}(\mu_b, \sigma_b^2)$, $w_{ik} \overset{iid}{\sim} \mathcal{N}(0, \sigma_w^2)$ and $\alpha_k \overset{iid}{\sim} U(-1, 1)$, $i = 1, \ldots, p$ and $k = 1, \ldots, K$. Our aim is to specify the hyperparameters μ_b, σ_b^2 and σ_w^2 to be such that the resulting prior on λ_{it} has mode, mean and variance close to the corresponding values of the $Gam(1, 50)$ distribution which was used as prior in the case of modelling the jump intensities independently over time and across stocks.

From (5) and (6) we have that

$$y_{it} = \log\big(\lambda_{it}/(\lambda^* - \lambda_{it})\big) \sim \mathcal{N}(\mu_y, \sigma_y^2), \;\; \lambda_{it} \in (0, \lambda^*), \tag{16}$$

where we have set, see Section 2.3, $\lambda^* = 0.15$. To specify μ_y and σ_y^2 we note that equations $(6) - (8)$ imply

$$\mathrm{E}(y_{it}|\mu_b) = \mu_b \;\text{ and }\; \mathrm{Var}(y_{it}|A, \sigma_w^2, \sigma_b^2) = \sigma_w^2 \sum_{k=1}^{K} (1 - \alpha_k^2)^{-1} + \sigma_b^2.$$

Thus, by assuming that $\alpha_k = \mathrm{E}(\alpha_k) = 0$ we have

$$\mu_y = \mu_b \;\text{ and }\; \sigma_y^2 = K\sigma_w^2 + \sigma_b^2. \tag{17}$$

From (16) we have that the density of λ_{it} is

$$\frac{\lambda^* \exp\left\{-\frac{1}{2\sigma_y^2}\left(\log\left(\frac{\lambda_{it}}{\lambda^* - \lambda_{it}}\right) - \mu_y\right)^2\right\}}{\sigma_y \lambda_{it}(\lambda^* - \lambda_{it})\sqrt{2\pi}}, \;\; \lambda_{it} \in (0, \lambda^*). \tag{18}$$

Therefore, the location of its mode satisfies the equation

$$\log\left(\frac{\lambda_{it}}{\lambda^* - \lambda_{it}}\right) = \frac{\lambda^* \mu_y + \sigma_y^2(2\lambda_{it} - \lambda^*)}{\lambda^*}, \tag{19}$$

which can be solved, for given values of λ^*, μ_y and σ_y^2, with numerical methods. Here we use the R-package `rootSolve` (Soetaert and Herman, 2009) to find all the roots of the equation in (19).

Since our aim is to match μ_y, σ_y^2 and the mode of the density in (18) with the corresponding quantities implied from the $Gam(1, 50)$ distribution we first note that if $\lambda_{it} \sim Gam(1, 50)$, right truncated at λ^*, then for $\lambda^* = 0.15$ we have that

$$\mathrm{E}\big[\log\big(\lambda_{it}/(\lambda^* - \lambda_{it})\big)\big] \approx -2.4 \text{ and } \mathrm{Var}\big[\log\big(\lambda_{it}/(\lambda^* - \lambda_{it})\big)\big] \approx 2, \tag{20}$$

and that the mode of the $Gam(1, 50)$ distribution is located at zero. By noting that in our applications we have used $K = 2$ latent factors and taking into account (17) we set $\sigma_w^2 = 0.5$ and $\sigma_b^2 = 1$. Then, we have that $\sigma_y^2 = 2$ which is close to the variance that we obtain for y_{it} in the univariate case; see equation (20). To specify μ_b except from matching the means in (17) and (20) we are also trying to set the mode of the density in (18) close to zero. Table 2 displays the mean, the variance and the mode of (18) for three different values of μ_b. Note that by choosing $\mu_b = -2.4$ we have that the mean and

the variance of the density in (18) are quite close to the corresponding moments of the $Gam(1, 50)$ distribution while by taking smaller values for μ_b the mode in (18) becomes closer to zero.

Table 2: Mean, variance and mode of the density in (18) for different values of μ_b.

	Mean	Variance	Mode
$\mu_b = -2.4$	0.021	0.00054	0.002
$\mu_b = -5$	0.003	0.00004	0.0001
$\mu_b = -10$	2×10^{-5}	4×10^{-8}	10^{-6}

Proof of Proposition 1

Proof. A discrete distribution with pmf $p(n)$ is called log-concave when the inequality $p(n)^2 \geq p(n-1)p(n+1)$ is satisfied for all n. Let $\sigma_{it}^2 = \exp(h_{it})$, in the case of joint modelled jump intensities we have that $\lambda_{it} = \lambda^*/\big(1 + \exp(-b_i - W_i' F_t)\big)$ and in the case of independent modelling λ_{it} is a constant. We denote with $p(n)$ the pmf $p(n_{it} = n | h_{it}, \mu_{i\xi}, \sigma_{i\xi}^2, b_i, W_i, F_t, r_{it})$. By integrating out the jump sizes Ξ_i from the likelihood we have that

$$
\begin{aligned}
p(n) &\propto p(r_{it}|n, h_{it}, \mu_{i\xi}, \sigma_{i\xi}^2) p(n|\lambda_{it}\Delta_{it}) \\
&= (2\pi(\sigma_{it}^2 + n\sigma_{i\xi}^2))^{-1/2} \exp\left(-\frac{1}{2} \frac{(r_{it} - n\mu_{i\xi})^2}{\sigma_{it}^2 + n\sigma_{i\xi}^2} - \lambda_{it}\Delta_{it} \right) \frac{(\lambda_{it}\Delta_{it})^n}{n!}, \quad n = 0, 1, 2, ...
\end{aligned}
$$

To prove log-concavity for $n = 1, 2, ...$, we need to show that for $n = 2, 3, ...$

$$
2\log(p(n)) - \log(p(n-1)) - \log(p(n+1)) \geq 0
$$

or equivalently that,

$$
\begin{aligned}
\log\left((\sigma_{it}^2 + n\sigma_{i\xi}^2)^2 - \sigma_{i\xi}^4\right) - \log\big(\sigma_{it}^2 + n\sigma_{i\xi}^2\big)^2 + \frac{2\left(\sigma_{i\xi}^2 (r_{it} - n\mu_{i\xi})^2 + \mu_{i\xi}(\sigma_{it}^2 + n\sigma_{i\xi}^2)\right)^2}{(\sigma_{it}^2 + n\sigma_{i\xi}^2)\left((\sigma_{it}^2 + n\sigma_{i\xi}^2)^2 - \sigma_{i\xi}^4\right)} \\
+ \log(n+1)^2 - \log(n)^2 \geq 0.
\end{aligned}
$$

or that

$$
\frac{((\sigma_{it}^2 + n\sigma_{i\xi}^2)^2 - \sigma_{i\xi}^4)(n+1)^2}{n^2(\sigma_{it}^2 + n\sigma_{i\xi}^2)^2} e^L \geq 1
$$

where $L = \frac{2\left(\sigma_{i\xi}^2 (r_{it} - n\mu_{i\xi})^2 + \mu_{i\xi}(\sigma_{it}^2 + n\sigma_{i\xi}^2)\right)^2}{(\sigma_{it}^2 + n\sigma_{i\xi}^2)\left((\sigma_{it}^2 + n\sigma_{i\xi}^2)^2 - \sigma_{i\xi}^4\right)}$. The last inequality holds; $L > 0$ and it is readily shown that the nominator is greater than the denominator for all $n = 2, 3....$ □

Proof of Proposition 2

Proof. Let $\mathbf{1}_n'$ be an n-dimensional column vector with ones and I_n the $n \times n$ identity matrix. For each $i = 1, \ldots, p$ and $t = 1, \ldots, T$ we have that $\xi_{it}^{\kappa} \overset{iid}{\sim} N(\mu_{i\xi}, \sigma_{i\xi}^2)$, $\kappa =$

$1, \ldots, n_{it}$. Let $\sigma_{it}^2 = \exp(h_{it})$ then from the canonical form of the Gaussian distribution we have that

$$p(\xi_{it}|n_{it}, r_{it}, h_{it}, \mu_{i\xi}, \sigma_{i\xi}^2) \propto \mathcal{N}\left(r_{it}\Big|\sum_{\kappa=1}^{n_{it}}\xi_{it}^{\kappa}, \sigma_{it}^2\right)\prod_{\kappa=1}^{n_{it}}\mathcal{N}\big(\xi_{it}^{\kappa}|\mu_{i\xi}, \sigma_{i\xi}^2\big)$$
$$\propto \mathcal{N}_{n_{it}}\big(\xi_{it}|Q_{i\xi}^{-1}c, Q_{i\xi}^{-1}\big)$$

where $c = (\frac{\mu_{i\xi}}{\sigma_{i\xi}^2} + \frac{r_{it}}{\sigma_{it}^2})\mathbf{1}_{n_{it}}$ and $Q_{i\xi} = Q_1 + Q_2$, with $Q_1 = \sigma_{i\xi}^{-2}I_{n_{it}}$ and $Q_2 = \sigma_{it}^{-2}\mathbf{1}_{n_{it}}\mathbf{1}_{n_{it}}'$. From Miller (1981) we have that

$$Q_{i\xi}^{-1} = (Q_1 + Q_2)^{-1} = Q_1^{-1} - \frac{1}{1+\nu}Q_1^{-1}Q_2Q_1^{-1}$$

where $\nu = tr(Q_2Q_1^{-1})$ or $(1+\nu)^{-1} = \dfrac{\sigma_{it}^2}{\sigma_{it}^2 + n_{it}\sigma_{i\xi}^2}$. Therefore

$$Q_{i\xi}^{-1} = \sigma_{i\xi}^2 I_{n_{it}} - \frac{\sigma_{i\xi}^4}{\sigma_{it}^2 + n_{it}\sigma_{i\xi}^2}\mathbf{1}_{n_{it}}\mathbf{1}_{n_{it}}' \text{ and } Q_{i\xi}^{-1}c = (\frac{\mu_{i\xi}}{\sigma_{i\xi}^2} + \frac{r_{it}}{\sigma_{it}^2})\frac{\sigma_{i\xi}^2\sigma_{it}^2}{\sigma_{it}^2 + n_{it}\sigma_{i\xi}^2}\mathbf{1}_{n_{it}}.$$

□

Auxiliary gradient-based sampler

Here we describe how we apply the auxiliary gradient-based sampler (Titsias and Papaspiliopoulos, 2018) to draw, at each MCMC iteration, the whole latent paths of the log-volatilities H_i jointly with the parameters ϕ_i and $\sigma_{i\eta}^2$ and the whole path of the latent factors F jointly with the parameters A.

Sampling latent log-volatilities and parameters

To sample, for a fixed i, H_i jointly with ϕ_i and $\sigma_{i\eta}^2$, we need to draw samples from the distribution with density given by (9). Following the method presented in Section 4 in Titsias and Papaspiliopoulos (2018) we augment the target with auxiliary variables $Z_i \sim \mathcal{N}_{T+1}\big(Z_i|H_i + (\gamma_i/2)\nabla g(H_i|R_i, N_i, \mu_{i\xi}, \sigma_{i\xi}^2), (\gamma_i/2)I_{T+1}\big)$, where $g(H_i|R_i, N_i, \mu_{i\xi}, \sigma_{i\xi}^2)$ denotes the logarithm of the product of the Gaussian densities in (9). Then we consider the expanded target

$$p(Z_i, H_i, \phi_i, \sigma_{i\eta}^2|\mu_i, \mu_{i\xi}, \sigma_{i\xi}^2, N_i, R_i) \propto \exp\{g(H_i|R_i, N_i, \mu_{i\xi}, \sigma_{i\xi}^2)\}\mathcal{N}_{T+1}\big(H_i|\mu_i, C_{H_i}\big)$$
$$\times \mathcal{N}_{T+1}\big(Z_i|H_i + (\gamma_i/2)\nabla g(H_i|R_i, N_i, \mu_{i\xi}, \sigma_{i\xi}^2), (\gamma_i/2)I_{T+1}\big)\pi(\phi_i)\pi(\sigma_{i\eta}^2),$$

where $\pi(\phi_i)$, $\pi(\sigma_{i\eta}^2)$ denote the prior densities of the parameters ϕ_i and $\sigma_{i\eta}^2$ and C_{H_i} denotes the covariance matrix of the autoregressive log-volatility process in (2). We sample from the expanded target by iterating between the steps:

(i) $Z_i \sim \mathcal{N}_{T+1}\big(Z_i|H_i + (\gamma_i/2)\nabla g(H_i|R_i, N_i, \mu_{i\xi}, \sigma_{i\xi}^2), (\gamma_i/2)I_{T+1}\big)$

(ii) Sample from $p(H_i, \phi_i, \sigma_{i\eta}^2|Z_i, \mu_i, \mu_{i\xi}, \sigma_{i\xi}^2, N_i, R_i)$ by using a Metropolis-Hastings in which we transform $\hat{\phi}_i = \log\big((\phi_i+1)/(1-\phi_i)\big)$ and $\hat{\sigma}_{i\eta}^2 = 2\log\sigma_{i\eta}$ and we propose new values $(H_i^*, \hat{\phi}_i^*, \hat{\sigma}_{i\eta}^{2*})$ by first drawing $(\hat{\phi}_i^*, \hat{\sigma}_{i\eta}^{2*})$ from the bivariate Gaussian

density centred at $(\hat{\phi}_i, \hat{\sigma}^2_{i\eta})$ with covariance matrix $\kappa_i I_2$ and then draw H_i^* from

$$q\big(H_i^*\big|\phi_i^*, \sigma^{2*}_{i\eta}, Z_i\big) = \mathcal{N}\bigg(H_i^*\Big|\frac{2}{\gamma_i}Q(Z_i + \frac{\gamma_i}{2}M_i^*), Q\bigg) = \frac{\mathcal{N}(H_i^*\big|\mu_i, C^*_{H_i})\mathcal{N}(Z_i\big|H_i^*, (\gamma_i/2)I_{T+1})}{\mathcal{Z}(Z_i, \mu_i, \phi_i^*, \sigma^{2*}_{i\eta})}, \tag{21}$$

with $Q = (C^{*-1}_{H_i} + \frac{2}{\gamma_i}I_{T+1})^{-1}$, $M_i^* = \dfrac{\mu_i(1-\phi_i^*)}{\sigma^{2*}_{i\eta}}\big(1,\ 1-\phi_i^*, \ldots, 1-\phi_i^*,\ 1\big)$ and

$$\mathcal{Z}(Z_i, \mu_i, \phi_i^*, \sigma^{2*}_{i\eta}) = \mathcal{N}_{T+1}(Z_i|\mu_i, C^*_{H_i} + (\gamma_i/2)I_{T+1}),$$

where $\phi_i^* = (e^{\hat{\phi}_i^*} - 1)/(e^{\hat{\phi}_i^*} + 1)$ and $\sigma^{2*}_{i\eta} = e^{\hat{\sigma}^{2*}_{i\eta}}$. By using the notation $C^*_{H_i}$ we emphasize that C_{H_i} depends on the values of the parameters ϕ_i^* and $\sigma^{2*}_{i\eta}$. The values $(H_i^*, \hat{\phi}_i^*, \hat{\sigma}^{2*}_{i\eta})$ are accepted or rejected according to the ratio

$$\exp\big\{g\big(H_i^*|N_i, \mu_{i\xi}, \sigma^2_{i\xi}\big) - g\big(H_i|R_i, N_i, \mu_{i\xi}, \sigma^2_{i\xi}\big) + u\big(Z_i, H_i^*\big) - u\big(Z_i, H_i\big)\big\} \times \frac{\mathcal{Z}(Z_i, \mu_i, \phi_i^*, \sigma^{2*}_{i\eta})\pi\big(\phi_i^*\big)\pi\big(\sigma^{2*}_{i\eta}\big)J(\hat{\phi}_i^*, \hat{\sigma}^{2*}_{i\eta})}{\mathcal{Z}\big(Z_i, \mu_i, \phi_i, \sigma^2_{i\eta}\big)\pi\big(\phi_i\big)\pi\big(\sigma^2_{i\eta}\big)J(\hat{\phi}_i, \hat{\sigma}^2_{i\eta})} \tag{22}$$

where $u\big(Z_i, H_i\big) = \big(Z_i - H_i - \frac{\gamma_i}{4}\nabla g\big(H_i\big|R_i, N_i, \mu_{i\xi}, \sigma^2_{i\xi}\big)\big)'\nabla g\big(H_i\big|N_i, \mu_{i\xi}, \sigma^2_{i\xi}\big)$ and $J(\cdot,\cdot)$ denotes the Jacobian of the inverse of the transformation $(\phi_i, \sigma^2_{i\eta}) \to (\hat{\phi}_i, \hat{\sigma}^2_{i\eta})$.

The values of the parameters γ_i and κ_i are tuned to achieve acceptance rate in the range of $50\% - 60\%$ and $20\% - 30\%$ respectively (Titsias and Papaspiliopoulos, 2018). Moreover, we note that a single iteration of the auxiliary gradient-based sampler consists of two steps: the first for sampling H_i alone and the second for joint sampling $(H_i, \phi_i, \sigma^2_{i\eta})$. These two steps are necessary at least for the burn-in phase since the acceptance histories of the first step allow to tune γ_i while the ones from the second step allow to tune κ_i. Finally, we note that $C^{-1}_{H_i}$ has a tridiagonal form, thus sampling from the proposal in (21) and the calculation of the acceptance ratio are conducted, efficiently, based on the Cholesky decomposition of tridiagonal matrices. After drawing $(H_i, \phi_i, \sigma^2_{i\eta})$ we sample μ_i from its full conditional $\mathcal{N}(m_i, s_i^2)$, where $m_i = \frac{s_i^2\left[(1-\phi_i^2)h_{i0} + (1-\phi_i)\sum_{t=1}^T (h_{it} - \phi_i h_{i,t-1})\right]}{\sigma^2_{i\eta}}$ and $s_i^2 = \frac{\sigma^2_{i\eta}}{\sigma^2_{i\eta}/10 + (1-\phi_i^2) + T(1-\phi_i)^2}$.

Ancillarity-sufficiency interweaving strategy for log-volatility parameters

In order to implement the ancillarity-sufficiency interweaving strategy (Yu and Meng, 2011) for the parameters μ_i, ϕ_i and $\sigma^2_{i\eta}$ we first consider the model defined by equations (1) and (2) with the non-centred parametrization of the log-volatilities. This is specified, for each $i = 1, \ldots, p$ and $t = 1, \ldots, T$, by

$$r_{it} = \exp\left(\frac{\mu_i + \sigma_{i\eta}\tilde{h}_{it}}{2}\right)\epsilon_{it} + \sum_{\kappa=1}^{n_{it}} \xi_{it}^{\kappa}, \quad \epsilon_{it} \sim N(0,1)$$

$$\tilde{h}_{it} = \phi_i \tilde{h}_{i,t-1} + \eta_{it}, \quad \eta_{it} \sim N(0,1), \quad \tilde{h}_{i0} \sim N(0, 1/(1-\phi_i^2)),$$

where ϵ_{it} and η_{it} are independent random variables and $\tilde{h}_{it} = (h_{it} - \mu_i)/\sigma_{i\eta}$. Then, we utilize random walk Metropolis-Hastings steps to draw μ_i and $\sigma^2_{i\eta}$ from their full conditionals obtained from

$$p(\mu_i, \sigma^2_{i\eta}|\mu_{i\xi}, \sigma^2_{i\xi}, \tilde{H}_i, R_i) \propto \pi(\mu_i)\pi(\sigma^2_{i\eta}) \prod_{t=1}^{T} \mathcal{N}\big(r_{it}\big|n_{it}\mu_{i\xi}, n_{it}\sigma^2_{i\xi} + \exp\{\mu_i + \sigma_{i\eta}\tilde{h}_{it}\}\big),$$

and we sample ϕ_i from

$$p(\phi_i|\tilde{H}_i) \propto \pi(\phi_i)\mathcal{N}\big(\tilde{h}_{i0}|0, 1/(1-\phi_i^2)\big)\prod_{t=2}^{T}\mathcal{N}\big(\tilde{h}_{it}|\phi_i\tilde{h}_{i,t-1}, 1\big),$$

by using a Metropolis-Hastings step with proposal distribution

$$\mathcal{N}\left(\frac{\sum_{t=1}^{T}\tilde{h}_{it}\tilde{h}_{i,t-1}}{\sum_{t=1}^{T}\tilde{h}_{i,t-1}}, \frac{1}{\sum_{t=1}^{T}\tilde{h}_{i,t-1}}\right).$$

Simulation experiments for samplers with and without interweaving

Figures 8 and 9 evaluate the efficiency of the auxiliary gradient-based sampler in drawing the parameters of the latent log-volatility path of univariate SV models. We compare three sampling schemes; the auxiliary gradient-based sampler with interweaving, without interweaving, and a scheme in which instead of interweaving we iterate twice the auxiliary sampler. To evaluate the different sampling schemes we compare the effective sample size (ESS) of the samples drawn from the posterior distributions of interest. The ESS of S samples drawn by using an MCMC algorithm can be estimated as v^2S/ν_0 where v^2 is the sample variance of the posterior samples and ν_0 is an estimation of the spectral density of the Markov chain at zero. We used the R-package `coda` (Plummer et al., 2006) to estimate the ESS of the posterior samples obtained from the samplers under comparison.

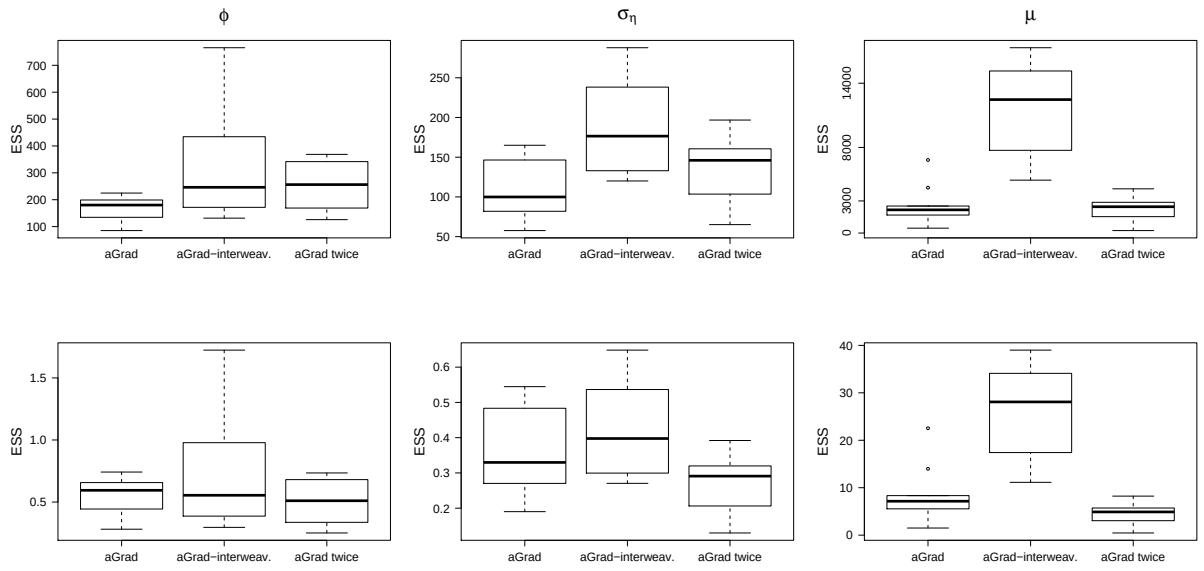

Figure 8: Top: the effective sample size (ESS) of the posterior samples of the parameters μ, σ_η and ϕ obtained by applying the auxiliary gradient-based (aGrad) sampler with and without interweaving in 10 replicates of a dataset consisted of $1,500$ log-returns simulated from the univariate SV model with $\mu = -0.85$, $\sigma_\eta = 0.15$ and $\phi = 0.98$. Bottom: the corresponding ESS divided by the number of seconds required for its sampling scheme.

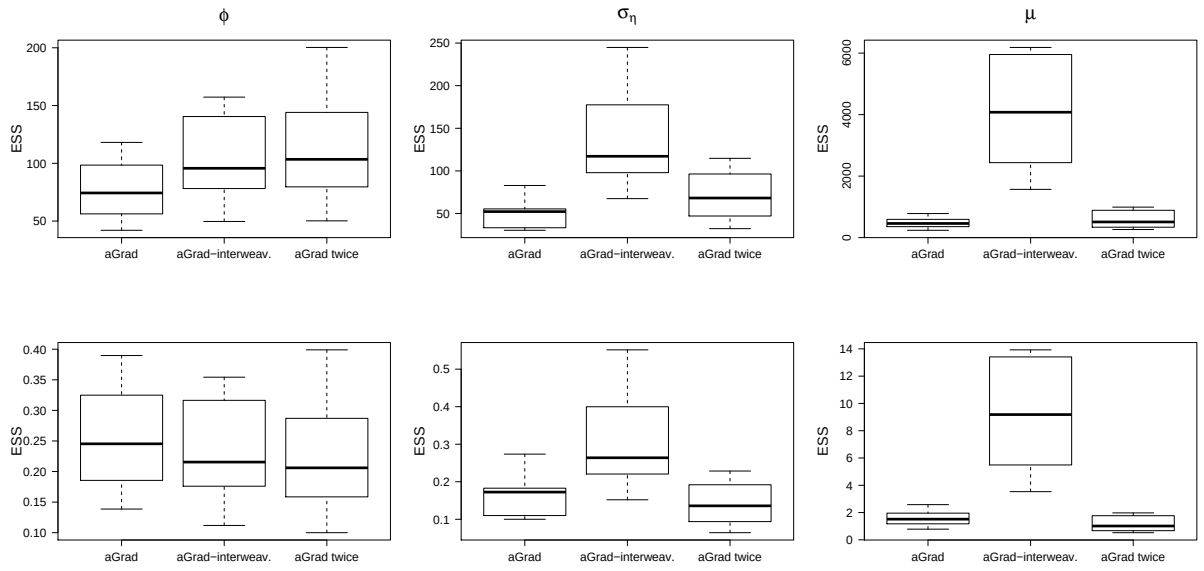

Figure 9: Top: the effective sample size (ESS) of the posterior samples of the parameters μ, σ_η and ϕ obtained by applying the auxiliary gradient-based (aGrad) sampler with and without interweaving in 10 replicates of a dataset consisted of $1,500$ log-returns simulated from the univariate SV model with $\mu = -0.85$, $\sigma_\eta = 0.15$ and $\phi = 0.90$. Bottom: the corresponding ESS divided by the number of seconds required for its sampling scheme.

The comparisons of the three sampling schemes are based on 10 replicates of two datasets consisted of $T = 1,500$ log-returns. These were simulated from the univariate SV model without jumps which is defined by (1) and (2) with $p = 1$ and by omitting the jump component in (1). For the remaining of this subsection we omit the subscript i from our notation; e.g. $\mu \equiv \mu_i$ and $H \equiv H_i$. We set $\mu = -0.85$, $\sigma_\eta = 0.15$ to generate both datasets and $\phi = 0.98$ in the first and $\phi = 0.90$ in the second. In each replicate of the datasets we applied the sampling schemes under comparison in order to draw samples for the latent log-volatility path H and its parameters μ, σ_η and ϕ. As described in Section 3.2 we update H jointly with ϕ and σ_η^2 by using the auxiliary gradient-based sampler to draw from (9) (with $n_t = 0$) and then we draw μ from its full conditional. In the case of combining the sampler with the ancillarity-sufficiency interweaving strategy we follow the procedure described in the previous section in order to update the parameters μ, σ_η and ϕ under the non-centered parametrization of the log-volatility process. We ran all the MCMC algorithms for $30,000$ iterations and we discarded the first $10,000$ as burn-in period obtaining, thus, $20,000$ posterior samples for the log-volatility path and its parameters for each replicate of the two simulated datasets. In the Figures 8 and 9 we present, for the datasets with $\phi = 0.98$ and $\phi = 0.90$ respectively, the boxplots with the ESS of the posterior samples of the parameters μ, σ_η and ϕ obtained from the 10 replicates. The two figures indicate that by combining the auxiliary gradient-based sampler with the ancillarity-sufficiency interweaving strategy the ESS of σ_η and μ is clearly improved, while the three schemes have almost the same efficiency in drawing ϕ.

Sampling the latent factors and their parameters

In the case of sampling the path of the latent factors F jointly with the parameters A the auxiliary gradient-based sampler is described as follows. The target distribution is

defined by (10). We first consider the expanded target

$$p(Z_F, F, A|N, B, W) \propto \exp\{g_F(F|N, B, W)\}\mathcal{N}_{K(T+1)}\big(F|0, C_A\big)$$
$$\mathcal{N}_{K(T+1)}\bigg(Z_F|F + (\gamma_F/2)\nabla g_F(F|N, B, W), (\gamma_F/2)I_{K(T+1)}\bigg)\pi(A),$$

where $g_F(F|N, B, W)$ denotes the logarithm of the unnormalized Poisson likelihood functions in (10) and $\pi(A)$ the prior density of A.

We sample from the expanded target by iterating between the steps:

(i) $Z_F \sim \mathcal{N}_{K(T+1)}\bigg(Z_F|F + (\gamma_F/2)\nabla g_F(F|N, B, W), (\gamma_F/2)I_{K(T+1)}\bigg)$

(ii) Sample from $p(F, A|Z_F, N, B, W)$ by using a Metropolis-Hastings in which we first transform the elements α_k, $k = 1, \ldots, K$, to $\hat{\alpha}_k = \log\big((\alpha_k + 1)/(1 - \alpha_k)\big)$ and we propose jointly new values F^* and $\hat{A}^*$ by first drawing $\hat{\alpha}_1^*, \ldots, \hat{\alpha}_K^*$ from a K-variate Gaussian distribution with mean $(\hat{\alpha}_1, \ldots, \hat{\alpha}_K)$ and covariance matrix κI_K and then

$$F^* \sim \mathcal{N}_{K(T+1)}\big(F^*\big|\frac{2}{\gamma_F}Q_F Z_F, Q_F\big),$$

where $Q_F = \big(C_{A^*}^{-1} + \frac{2}{\gamma_F}I_{K(T+1)}\big)^{-1}$ and $C_{A^*}^{-1}$ is constructed by using the proposed values $\alpha_k^* = (e^{\hat{\alpha}_k^*} - 1)/(e^{\hat{\alpha}_k^*} + 1)$. The acceptance ratio of the step is given by

$$\exp\big\{g_F(F^*|N, B, W) - g_F(F|N, B, W) + u_F\big(Z_F, F^*\big) - u_F\big(Z_F, F\big)\big\}\frac{\mathcal{Z}\big(Z_F, A^*\big)J(\hat{A}^*)}{\mathcal{Z}\big(Z_F, A\big)J(\hat{A})},$$

where we have that $u_F(Z_F, F) = \big(Z_F - F - \frac{\gamma_F}{4}\nabla g_F(F|N, B, W)\big)'\nabla g_F(F|N, B, W)$, $\mathcal{Z}(Z_F, A) = \mathcal{N}_{K(T+1)}\big(Z_F\big|0, C_A + (\gamma_F/2)I_{K(T+1)}\big)$ and $J(\cdot)$ denotes the Jacobian of the inverse of the transformation $(\alpha_1, \ldots, \alpha_k) \to (\hat{\alpha}_1, \ldots, \hat{\alpha}_k)$.

We tune the parameters γ_F and κ to achieve acceptance rate in the range of $50\% - 60\%$ and $20\% - 30\%$ respectively while a single iteration of the auxiliary gradient-based sampler consists of two steps; sampling F alone and joint sampling of (F, A). These two steps are necessary at least for the burn-in phase since the acceptance histories of the first step allow to tune γ_F while the ones from the second step allow to tune κ. The inverse covariance C_A^{-1} of the latent path F has a block-tridiagonal form, thus sampling from the corresponding proposal and the calculation of the acceptance ratio are conducted, efficiently, based on Cholesky decomposition.

Using the interweaving strategy to improve sampling of A

In Section 3.3 we combine the auxiliary gradient-based sampler with the ancillarity-sufficiency interweaving strategy in order to reduce the autocorrelation of the parameters in A. More precisely, after sampling jointly the paths F of the latent factors and the parameters A we set $\Gamma_0 = DF_0$, where D is a $K \times K$ diagonal matrix with elements $(1 - \alpha_k^2)^{1/2}$, $k = 1, \ldots, K$, and $\Gamma_t = F_t - AF_{t-1}$, for $t = 1, \ldots, T$. Then we use a random walk Metropolis-Hastings step that targets the density

$$p(A|\Gamma, N, B, W) \propto \pi(A)\prod_{t=1}^{T}\prod_{i=1}^{n} e^{-\lambda_{it}}\lambda_{it}^{n_{it}},$$

with $\lambda_{it} = \lambda^*/(1+\exp(-b_i - W_i' F_t))$. We note that F_t and Γ_t are related by $F_0 = D^{-1}\Gamma_0$ and $F_t = AF_{t-1} + \Gamma_t$, $t \geq 1$.

Here, we compare the efficiency of three different MCMC schemes for the update of the parameters A. These are the auxiliary gradient-based sampler with interweaving, without interweaving, and a scheme in which instead of interweaving we iterate twice the auxiliary sampler. To simplify our experiment we omit equations (1) and (2) and we simulate observations from the model defined by equations (3) and (5) – (8). This a Cox model with intensities driven by latent dynamic factors and Bayesian inference for this model is conducted by switching off certain steps (see Section 3) of Algorithm 1. To compare the three sampling schemes we simulated $p = 100$ time series consisted of $T = 1,500$ observations n_{it}, $t = 1, \ldots, T$. We used $K = 2$ autoregressive latent factors for the jump intensities, we set $\alpha_1 = 0.8$ and $\alpha_2 = 0.4$ and we generated $w_{ik} \overset{iid}{\sim} N(0,1)$. We ran the MCMC algorithms, that correspond to the three different schemes for the update of F and A, for $30,000$ iterations and we discarded the first $10,000$ as burn-in period. Figure 10 presents the estimated autocorrelation functions of the parameters in A for each one of the three sampling schemes.

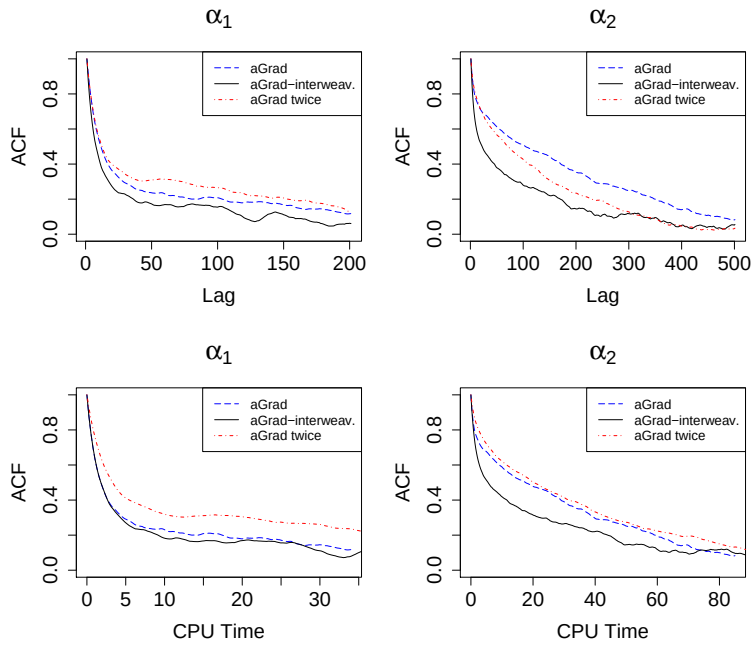

Figure 10: Estimated autocorrelation functions (ACF) of the parameters α_1 and α_2 of the latent autoregressive factors used for the intensities of the simulated time series from the model defined by equations (3) and (5) – (8). The three sampling schemes under comparison are based on the auxiliary gradient-based (aGrad) sampler. Top: ACF against time-lag, bottom: ACF against CPU time.

Sampling $\mu_{i\xi}$ and $\sigma^2_{i\xi}$

The full conditional distribution of $\mu_{i\xi}$ is

$$N\left(\frac{5\text{range}_i^2 \sum_{t=1}^{T}\sum_{\kappa=1}^{n_{it}} \xi_{it}^{\kappa}}{5\text{range}_i^2 \sum_{t=1}^{T} n_{it} + \sigma^2_{i\xi}}, \frac{5\text{range}_i^2 \sigma^2_{i\xi}}{5\text{range}_i^2 \sum_{t=1}^{T} n_{it} + \sigma^2_{i\xi}}\right), \quad i = 1, \ldots, p$$

and the full conditional distribution of $\sigma_{i\xi}^2$ is

$$IGam\left(3+0.5\sum_{t=1}^{T}n_{it},\text{range}_i^2/18+0.5\sum_{t=1}^{T}\sum_{\kappa=1}^{n_{it}}(\xi_{it}^{\kappa}-\mu_{i\xi})^2\right),\ \ i=1,\ldots,p,$$

Particle filter and annealed importance sampling

Here we present the pseudocode for the implementation of the particle filter and annealed importance sampling algorithms that we constructed in order to estimate predictive Bayes factors. A superscript s in Algorithm 3 means that the operation is performed for all $s=1,...,S$. The infinite sum in (23) is truncated at 10 without affecting the performance of the algorithm, see also Johannes et al. (2009) for a discussion. The step in the 9th line of Algorithm 3 is conducted by using multinomial resampling, see Doucet and Johansen (2009) for a detailed presentation of resampling methods. The `for` loop that begins in the 2nd line of Algorithm 4 is performed in parallel. In the 11th line of Algorithm 4 we use a few iterations of Algorithm (1) initialized with the samples obtained from the $(j-1)$th step.

Algorithm 3 Paticle filter algorithm that targets (12) for $t=T+1,\ldots,T+\ell$.

1: **Set** the number of desired samples S and give as **inputs**; in sample and out of sample observations $r_{i,1:T}$ and $r_{i,T+1:T+\ell}$ from the ith stock and S samples from $p(h_{i,1:T}|r_{i,1:T})$.
2: **for** $t=T+1,\ldots,T+\ell$ **do**
3: **if** $t=T+1$ **then**
4: **Sample** $h_{it}^s\sim N(\mu_i+\phi_i(h_{i,t-1}^s-\mu_i),\sigma_{i\eta}^2)$
5: **Set**

$$\omega_t^s=p(r_{it}|h_{it})=\left\{\begin{array}{ll}\sum_{n_{it}=0}^{\infty}\mathcal{N}(r_{it}|n_{it}\mu_{i\xi},e^{h_{it}}+n_{it}\sigma_{i\xi}^2)p(n_{it}), & \text{if SV with jumps}\\ \mathcal{N}(r_{it}|0,e^{h_{it}}), & \text{if SV without jumps}\end{array}\right\}\quad(23)$$

6: **Set** $\hat{p}(r_{it}|r_{i,1:t-1})=\sum_{s=1}^{S}\omega_t^s/S$
7: **else**
8: **Compute** $ESS_{t-1}=\dfrac{(\sum_{d=1}^{s}\omega_{t-1}^s)^2}{\sum_{s=1}^{S}(\omega_{t-1}^s)^2}$.
9: **If** $ESS_{t-1}<S/2$ obtain ancestor variables o_{t-1}^s and set $\omega_{t-1}^s=1$ **else** set $o_{t-1}^s=s$.
10: **Sample** $h_{it}^s\sim N(\mu_i+\phi_i(h_{i,t-1}^{o_{t-1}^s}-\mu_i),\sigma_{i\eta}^2)$ and **set** $\omega_t^s=\omega_{t-1}^s p(r_{it}|h_{it}^s)$
11: **If** $ESS_{t-1}<S/2$ set $\hat{p}(r_{it}|r_{i,1:t-1})=\sum_{s=1}^{S}\omega_t^s/S$ **otherwise** set $\hat{p}(r_{it}|r_{i,1:t-1})=\sum_{s=1}^{S}\omega_t^s/\sum_{s=1}^{S}\omega_{t-1}^s$
12: **end if**
13: **end for**
14: **return** $\hat{p}(r_{i,T+1}|r_{i,1:T}),\ldots,\hat{p}(r_{i,T+\ell}|r_{i,1:T+\ell-1})$

Algorithm 4 Annealed importance sampling algorithm that targets $p(H_{0:T+\ell}, F_{0:T+\ell}|R_{1:T+\ell})$.

Set the number of desired samples S and give as **inputs**; S samples from $p(H_{0:T}, F_{0:T}|R_{1:T})$ and the out of sample observations $R_{T+1:T+\ell}$.
for $s = 1, \ldots, S$ **do**
  for $j = 1, \ldots, \ell$ **do**
    if $j = 1$ **then**
      Sample $F^s_{T+1} \sim N(AF^s_T, I_K)$
      for $i = 1, \ldots, p$ **do**
        Sample $h^s_{i,T+1} \sim N(\mu_i + \phi_i(h^s_{iT} - \mu_i), \sigma^2_{i\eta})$
        Set $\omega^s_{i,T+1} = p(r_{i,T+1}|h^s_{i,T+1}, F^s_{T+1})$; the ith element of the product in (13)
      end for
    else
      Sample $(F^s_{T+j-1}, H^s_{T+j-1}) \sim p(F_{0:T+j-1}, H_{0:T+j-1}|R_{1:T+j-1})$
      Sample $F^s_{T+j} \sim N(AF^s_{T+j-1}, I_K)$
      for $i = 1, \ldots, p$ **do**
        Sample $h^s_{iT+j} \sim N(\mu_i + \phi_i(h^s_{i,T+j-1} - \mu_i), \sigma^2_{i\eta})$
        Set $\omega^s_{i,T+j} = \omega^s_{i,T+j-1} p(r_{i,T+j}|h^s_{i,T+j}, F^s_{T+j})$
      end for
    end if
  end for
end for
return $\{\omega^s_{it}\}_{s=1}^S$ for each $i = 1, \ldots, p$ and $t = T+1, \ldots, T+\ell$

Approximation with a mixture of Gaussian distributions

The approximation of the model defined by equations $(1)-(3)$ with a mixture of Gaussian distributions relies on the transformation

$$r^*_{it} = \log\left(r_{it} - \sum_{\kappa=1}^{n_{it}} \xi^\kappa_{it}\right)^2 = h_{it} + \log\left(\epsilon^2_{it}\right), \quad t = 1, \ldots, T,$$

and on the assumption that $\log(\epsilon^2_{it})|u_{it} \sim \mathcal{N}(m_{u_{it}}, s^2_{u_{it}})$ where $u_{it} \in \{1, \ldots, 10\}$ denotes a mixture component indicator. The means and the variances of the mixture components as well as the probabilities $p(u_{it} = j)$, $j = 1, \ldots, 10$, can be chosen as suggested by Omori et al. (2007). See Chib et al. (2002) and Nakajima and Omori (2009) for details.

Results on simulated data

Results for independent SV with jumps models

By using the approach described in Section 2.2.1 to model the jump intensities in (3) we obtain p independent univariate SV with jumps models. Bayesian inference for their parameters and latent states can be conducted with the MCMC algorithm derived from Algorithm 1 by omitting the 10th step and by substituting the 12th step 9 (sampling

latent factors) with the step of sampling the independent jump intensities from their inverse Gamma full conditionals.

To test our methods we simulated $T = 1,500$ log-returns from $p = 4$ independent univariate models as defined by equations $(1) - (3)$ with $Gam(1, c)$ priors for the jump intensities. We used $c = 50$ to simulate the first and the second and $c = 100$ to simulate the third and the fourth time series in order to represent observed log-returns with more and less jumps respectively. Following the related literature (see for example Chib et al. (2002) and Eraker et al. (2003)) we set $\mu_i = -0.85$, $\phi_i = 0.98$ and $\sigma_{i\eta} = 0.12$ for each $i = 1, 2, 3, 4$ and $\mu_{1\xi} = \mu_{3\xi} = -3$ and $\mu_{2\xi} = \mu_{4\xi} = 0$. We ran the MCMC algorithm for $80,000$ iterations from which we discarded the first $20,000$ iterations as burn-in period and we were storing the sampling outcome of every 60th iteration. Figure 11 presents the simulated log-returns along with the times in which jumps have been generated and times with estimated posterior probability of at least one jump greater than 50%. Figure 12 displays the simulated paths for the corresponding log-volatilities and the 95% credible intervals of their posterior distributions. By the visual inspection of the figures it is clear that with the proposed methods we have separated the jumps from the log-volatility process in all the simulated frameworks.

In order to check the ability of the proposed methods to identify jumps in time series with varying volatility of the log-volatility process we conducted the following experiment. We simulated $p = 4$ time series, from the model defined by equations $(1) - (3)$, with $T = 1,500$ observations in each one by using different values for $\sigma_{i\eta}$, $i = 1, \ldots, 4$; $\sigma_{1\eta} = 0.1$, $\sigma_{2\eta} = 0.15$, $\sigma_{3\eta} = 0.25$ and $\sigma_{4\eta} = 0.35$. We used $Gam(1, 50)$ priors for the jump intensities, we set $\mu_{1\xi} = \mu_{2\xi} = 0$ and the same values as in the previous experiment for all the other parameters. The same MCMC algorithm as in the first experiment was used to obtain posterior samples for the parameters and the latent states of the model. Figure 13 presents the simulated log-returns for each stock and indicates the times in which we estimated posterior probabilities of at least one jump greater than 50%, while Figure 14 displays the posterior probabilities of at least one jump for each one of the $1,500$ time points of each stock. The figures demonstrate that using the proposed methods we are able to disentangle the volatility from the jump process in all the different scenarios for the log-volatility process, since in the majority of the simulated jumps have been identified and only jumps with size that is in the range of the volatility process are not detected.

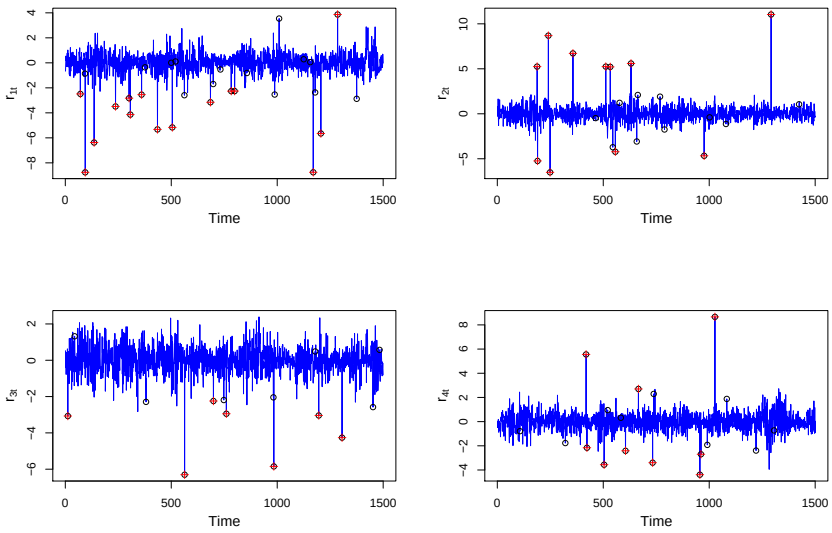

Figure 11: Simulated log-returns from 4 independent univariate SV with jumps models defined by equations $(1)-(3)$ with $Gam(1,50)$ (top) and $Gam(1,100)$ (bottom) prior for the jump intensities with $\mu_{1\xi}=\mu_{3\xi}=-3$ (left column) and $\mu_{2\xi}=\mu_{4\xi}=0$ (right column). We set $T=1,500$, $\mu_i=-0.85$, $\phi_i=0.98$, $\sigma_{i\eta}=0.12$, $\sigma_{i\xi}=3.5$, $i=1,2,3,4$. Black circles indicate times in which a jump has been simulated. Red crosses indicate times with estimated posterior probabilities of at least one jump greater than 50%.

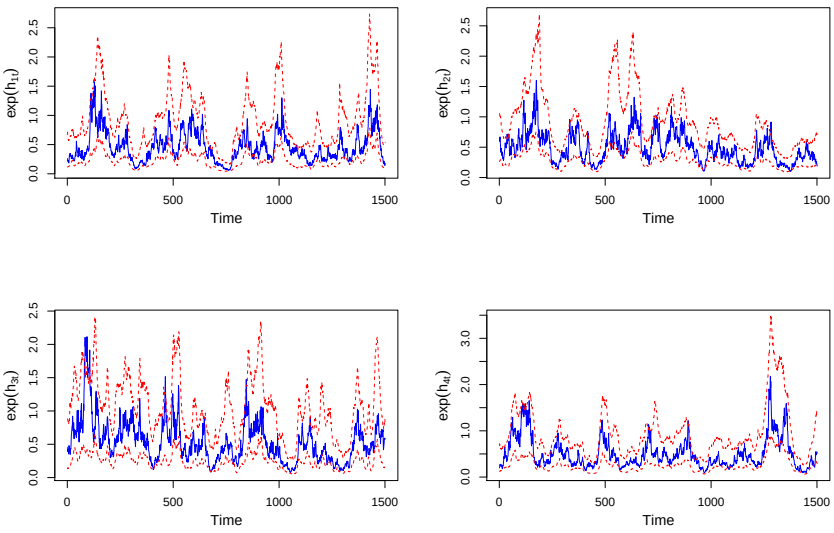

Figure 12: Simulated paths of the volatility process (solid lines) and 95% posterior credible intervals (dashed lines). The log-returns are simulated from 4 independent univariate SV with jumps models defined by equations $(1)-(3)$ with $Gam(1,50)$ (top) and $Gam(1,100)$ (bottom) prior for the jump intensities and with $\mu_{1\xi}=\mu_{3\xi}=-3$ (left column) and $\mu_{2\xi}=\mu_{4\xi}=0$ (right column). We set $T=1,500$, $\mu_i=-0.85$, $\phi_i=0.98$, $\sigma_{i\eta}=0.12$, $\sigma_{i\xi}=3.5$, $i=1,2,3,4$.

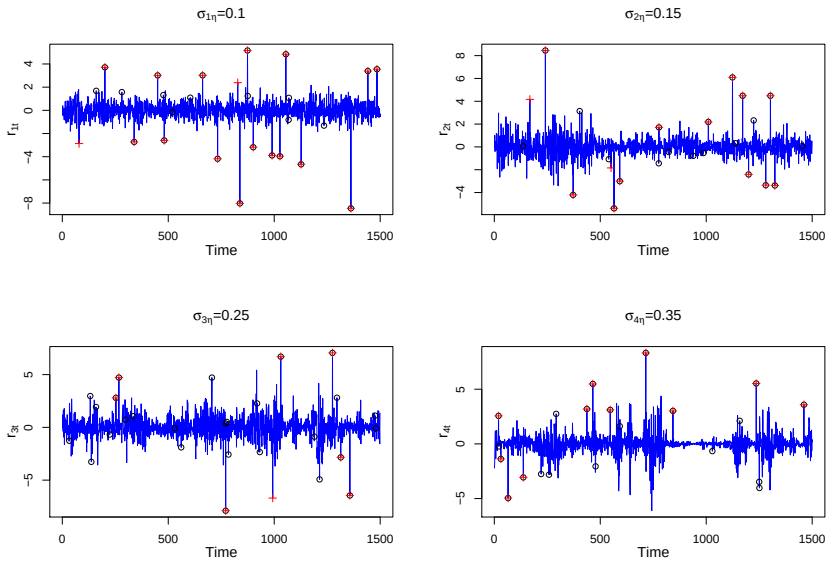

Figure 13: Simulated log-returns from $p = 4$ independent univariate SV with jumps models defined by equations $(1) - (3)$ with $Gam(1, 50)$ priors for the jump intensities. We set $T = 1,500$, $\mu_i = -0.85$, $\phi_i = 0.98$, $\mu_{i\xi} = 0$ and $\sigma_{i\xi} = 3.5$ and we used different values for $\sigma_{i\eta}$, $i = 1, 2, 3, 4$. Black circles: times in which a jump has been simulated. Red crosses: times with estimated posterior probabilities of at least one jump greater than 50%.

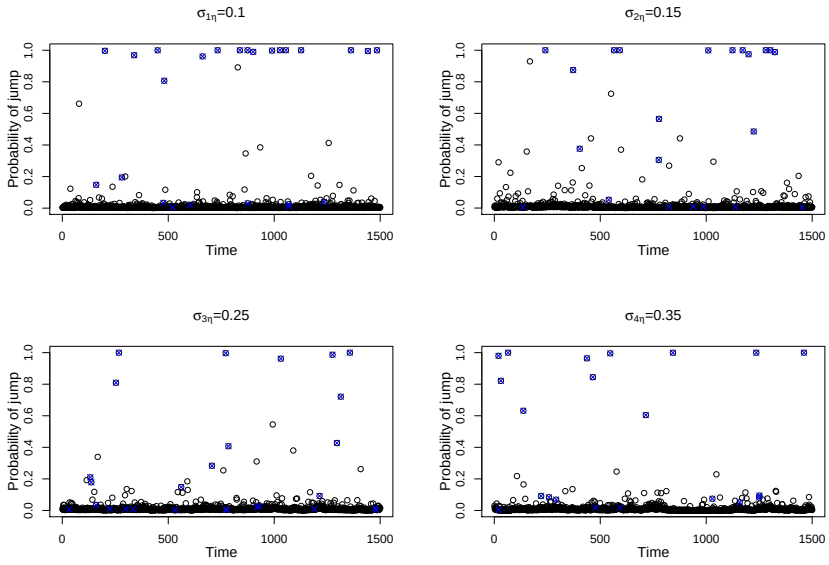

Figure 14: Black circles: Posterior probabilities for at least one jump in simulated log-returns from independent univariate SV with jumps models defined by equations $(1) - (3)$ with $Gam(1, 50)$ priors for the jump intensities. We set $T = 1,500$, $\mu_i = -0.85$, $\phi_i = 0.98$, $\mu_{i\xi} = 0$ and $\sigma_{i\xi} = 3.5$ and we used different values for $\sigma_{i\eta}$, $i = 1, 2, 3, 4$. Blue crosses: times of simulated jumps.

Results for joint modelled jump intensities

We present results from the application of our methods on $p = 100$ time series simulated from the model defined by equations (1)-(3) and (5)-(8). We used $K = 2$ latent factors

with $\alpha_1 = 0.85$ and $\alpha_2 = -0.45$ to simulate the paths of the latent factors. We note that although we have chosen α_1 and α_2 that represent the values which we have found in the real data analysis presented in Section 5, we have also tested that our methods work well in several cases. The factor loadings in W were simulated from standard normal distributions and we simulated $b_i \sim \mathcal{N}(-2.45, 1)$, $i = 1, \ldots, p$. For the parameters of the latent log-volatility processes we used $\phi_i = 0.98$, $\sigma_{i\eta} = 0.12$ and $\mu_i = -0.85$ and we set $\mu_{i\xi} = 0$ and $\sigma_{i\xi} = 3.5$ for each $i = 1, \ldots, p$. Then we ran the MCMC algorithm summarized by Algorithm 1 for $100,000$ iterations from which we discarded the first $30,000$ as burn-in period and we were saving the outcome of every 70th iteration.

Each line with black circles in Figure 15 refers to a stock while the circles indicate times in which a jumps has been simulated in the price of the stock. The red crosses indicate the times where the posterior probability of at least one jump has estimated to be greater than 50%. Figure 16 presents the posterior distributions for the parameters α_1 and α_2.

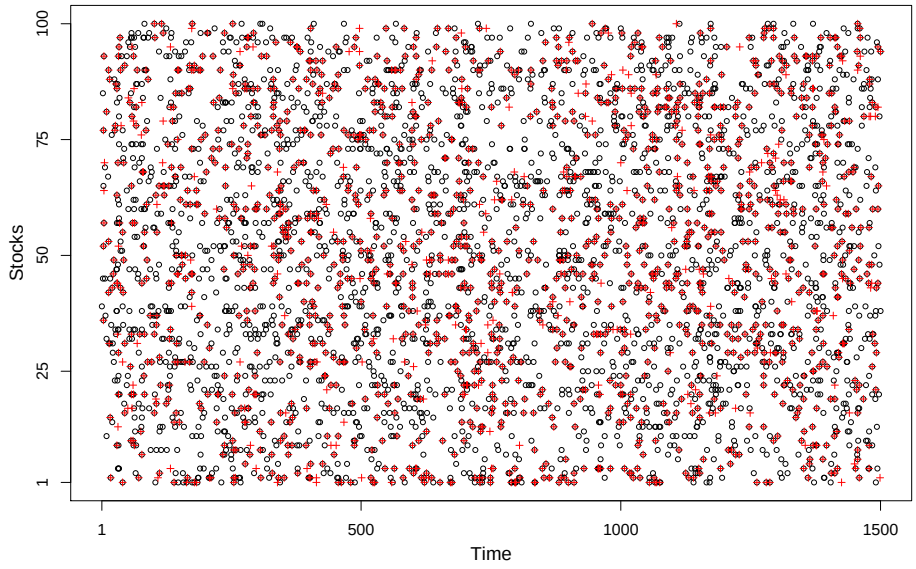

Figure 15: Simulated jump times (black circles) for each of the $p = 100$ simulated time series with $T = 1,500$ log-returns simulated from the model defined by the equations (1)-(3) and (5)-(8). The red crosses indicate times with posterior probability of having at least one jump greater than 50%.

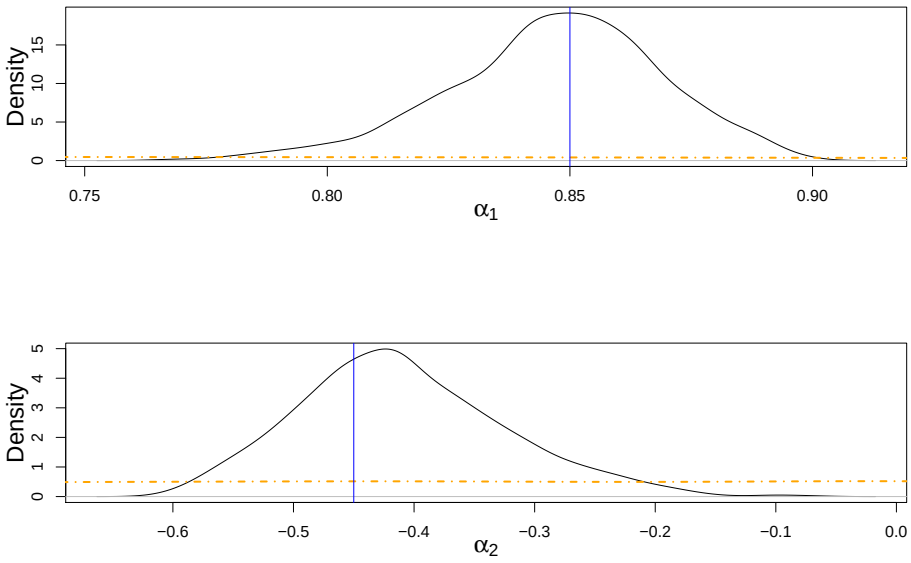

Figure 16: Posterior (solid lines) and prior (dashed lines) distributions for the parameters α_1 and α_2 of the dynamic factor model used for the joint modelling of the jump intensities. The vertical lines indicate the true values.

Predictions by using different prior-hyperparameters for the jump intensities

In the case of modelling independently, over time and across stocks, the jump intensities we have assumed that $\lambda_{it} \sim Gam(\delta, c)$ and by following the related literature, see for example Chib et al. (2002), we set $\delta = 1$ and $c = 50$ in order the prior expectation of λ_{it} to be 0.02 with prior variance 0.0004. To model jointly the jump intensities we used the dynamic factor model specified by equations (5)-(8). We specified hyperparameters σ_w^2, σ_b^2 and μ_b such that the induced prior on λ_{it} to be similar to the gamma distributions used in the case of independent modelling. Table 2 displays the mean, the variance and the mode of the prior induced, by the dynamic factor model, on each λ_{it}, for three different choices of the hyperparameters μ_b, σ_b^2 and σ_w^2. Since our aim with joint modelling the jump intensities was to improve the predictions of future observations we perform a sensitivity analysis of the predictions obtained from independent and joint modelled jump intensities with different choices for the prior-hypeparameters of the jump intensities.

Figure 17 compares the predictive performance of univariate SV models without jumps with the performance of SV models with jumps in which the jump intensities are modelled independently over time and across stocks by using the gamma prior distributions presented in Table 3. The figure is constructed by considering the logarithms of the Bayes factors defined by (11) for the $\ell = 30$ out-of-sample observations of our real dataset presented in Section 5. Figure 18 presents the logarithms of the Bayes factors in (14) and the logarithm of the Bayes factors in (15) for the 30 out-of-sample observations by using the priors on the jump intensities that described in Table 2.

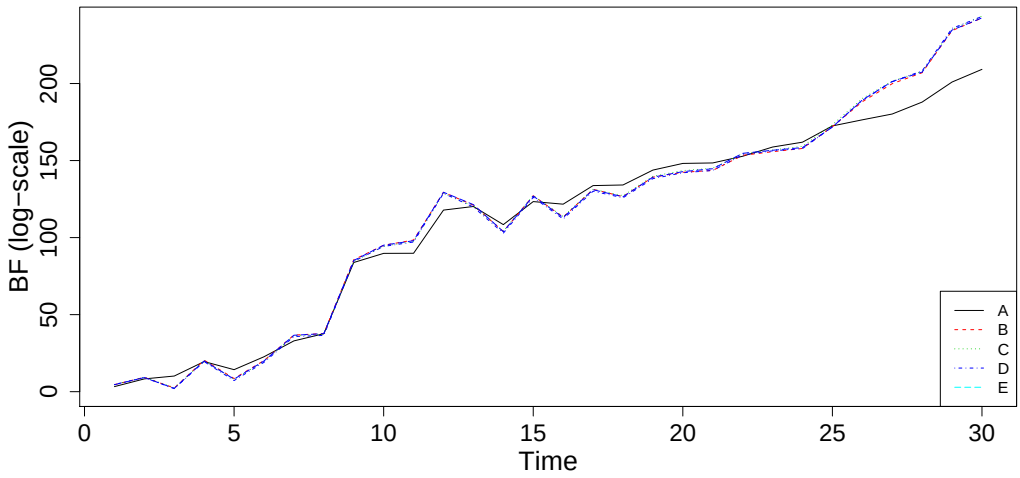

Figure 17: Logarithm of Bayes factors (BF), defined in (11), in favour of univariate SV models with jumps against univariate SV without jumps models for each one of the $\ell = 30$ out-of-sample observations. The letters A, B, C, D and E in the legend correspond to the prior distributions displayed in Table 3.

Table 3: Mean, variance and mode of different gamma priors assumed for the jump intensities λ_{it} in the case of independent models.

Prior for λ_{it}	Mean	Variance	Mode
A: $Gam(1, 250)$	0.004	0.00016	0
B: $Gam(1, 50)$	0.02	0.0004	0
C: $Gam(1.05, 52.63)$	0.02	0.00038	0.001
D: $Gam(1.33, 66.67)$	0.02	0.00029	0.005
E: $Gam(2, 100)$	0.02	0.0002	0.01

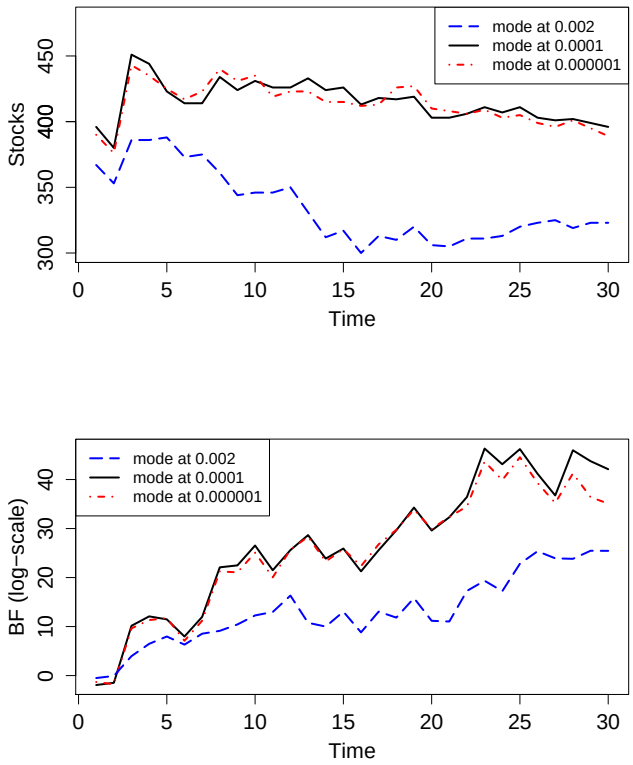

Figure 18: Top: number of stocks with positive log-Bayes factors, defined in (14), in favour of the proposed SV model with jointly modelled jump intensities against SV models with independent intensities. Bottom: logarithms of the Bayes factors, defined in (15), in favour of the proposed SV model with jointly modelled jump intensities against SV models with independent intensities. The quantities have been calculated for each one of the $\ell = 30$ out-of-sample observations. The different lines in each plot correspond to induced, by the dynamic factor model, priors on the jump intensities with different modes.

References

Aguilar, O. and M. West (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics 18*(3), 338–357.

Aït-Sahalia, Y. (2004). Disentangling diffusion from jumps. *Journal of Financial Economics 74*(3), 487–528.

Aït-Sahalia, Y., J. Cacho-Diaz, and R. J. Laeven (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics 117*(3), 585–606.

Alexopoulos, A. (2017). *Bayesian modelling of high dimensional financial data using latent Gaussian models*. Ph. D. thesis, Athens University of Economics and Business. Available online at `www.pyxida.aueb.gr/index.php?op=view_object&object_id=5343`.

Andersen, T. G., L. Benzoni, and J. Lund (2002). An empirical investigation of continuous-time equity return models. *The Journal of Finance 57*(3), 1239–1284.

Bates, D. S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *The Review of Financial Studies 9*(1), 69–107.

Bhattacharya, A. and D. B. Dunson (2011). Sparse Bayesian infinite factor models. *Biometrika 98*(2), 291–306.

Buesing, L., T. A. Machado, J. P. Cunningham, and L. Paninski (2014). Clustered factor analysis of multineuronal spike data. In *Advances in Neural Information Processing Systems*, pp. 3500–3508.

Chib, S., F. Nardari, and N. Shephard (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics 108*(2), 281–316.

Chopin, N., P. E. Jacob, and O. Papaspiliopoulos (2013). SMC^2: an efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(3), 397–426.

Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological) 17*(2), 129–164.

Dellaportas, P. and M. Pourahmadi (2012). Cholesky-GARCH models with applications to finance. *Statistics and Computing 22*(4), 849–855.

Doucet, A., N. De Freitas, and N. Gordon (2001). An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo methods in practice*, pp. 3–14. Springer.

Doucet, A. and A. M. Johansen (2009). A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering 12*(656-704), 3.

Eraker, B., M. Johannes, and N. Polson (2003). The impact of jumps in volatility and returns. *The Journal of Finance 58*(3), 1269–1300.

Geweke, J. and G. Amisano (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting 26*(2), 216–230.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika 58*(1), 83–90.

Johannes, M. S., N. G. Polson, and J. R. Stroud (2009). Optimal filtering of jump diffusions: Extracting latent states from asset prices. *Review of Financial Studies 22*(7), 2759–2799.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Kastner, G. and S. Frühwirth-Schnatter (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis 76*, 408–423.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies 65*(3), 361–393.

Miller, K. S. (1981). On the inverse of the sum of matrices. *Mathematics Magazine 54*(2), 67–72.

Nakajima, J. and Y. Omori (2009). Leverage, heavy-tails and correlated jumps in stochastic volatility models. *Computational Statistics & Data Analysis 53*(6), 2335–2353.

Neal, R. M. (2001). Annealed importance sampling. *Statistics and Computing 11*(2), 125–139.

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

Plummer, M., N. Best, K. Cowles, and K. Vines (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R News 6*(1), 7–11.

Rousseeuw, P. J. and C. Croux (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association 88*(424), 1273–1283.

Shephard, N. and S. Kim (1994). Comment. *Journal of Business & Economic Statistics 12*(4), 406–410.

Soetaert, K. and P. M. Herman (2009). *A Practical Guide to Ecological Modelling. Using R as a Simulation Platform*. Springer. ISBN 978-1-4020-8623-6.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75. *Time series analysis: theory and practice 1*, 203–226.

Titsias, M. K. and O. Papaspiliopoulos (2018). Auxiliary gradient-based sampling algorithms. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80*(4), 749–767.

Van Dyk, D. A. and X. Jiao (2015). Metropolis-hastings within partially collapsed gibbs samplers. *Journal of Computational and Graphical Statistics 24*(2), 301–327.

Yu, Y. and X.-L. Meng (2011). To center or not to center: That is not the question-an ancillarity–sufficiency interweaving strategy (ASIS) for boosting MCMC efficiency. *Journal of Computational and Graphical Statistics 20*(3), 531–570.