

Bayesian Prediction of Jumps in Large Panels of Time Series Data*

Angelos Alexopoulos^{†,‡}, Petros Dellaportas^{†,¶,§}, and Omiros Papaspiliopoulos^{||}

Abstract. We take a new look at the problem of disentangling the volatility and jumps processes of daily stock returns. We first provide a computational framework for the univariate stochastic volatility model with Poisson-driven jumps that offers a competitive inference alternative to the existing tools. This methodology is then extended to a large set of stocks for which we assume that their unobserved jump intensities co-evolve in time through a dynamic factor model. To evaluate the proposed modelling approach we conduct out-of-sample forecasts and we compare the posterior predictive distributions obtained from the different models. We provide evidence that joint modelling of jumps improves the predictive ability of the stochastic volatility models.

Keywords: dynamic factor model, forecasting stock returns, Markov chain Monte Carlo, stochastic volatility with jumps, sequential Monte Carlo.

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 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 estimation of SV models is preferably conducted by using likelihood-based approaches (Harvey et al., 1994) while Bayesian methods, in particular, are notably desirable since they deal efficiently with likelihood intractability problems; see for example Jarquier et al. (1994), Chib et al. (2002) and Kastner and Frühwirth-Schnatter (2014) for detailed discussions. This motivates us to utilize the Bayesian approach both for modelling and inference purposes. Our aim is to provide a general modelling approach for the time and cross-sectional dependence of jumps in multiple time series. The resulting Bayesian hierarchical models require careful prior specifications and sophisticated, modern Markov chain Monte

*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.

[†]Department of Statistical Science, University College London, United Kingdom

[‡]*Present address:* MRC Biostatistics Unit, University of Cambridge, United Kingdom, angelos@mrc-bsu.cam.ac.uk

[§]The Alan Turing Institute, London, United Kingdom, p.dellaportas@ucl.ac.uk

[¶]Department of Statistics, Athens University of Economics and Business, Greece

^{||}Department of Decision Sciences, Bocconi University, Milan, Italy, omiros@unibocconi.it

Carlo (MCMC) inference implementation strategies. To that end, we build the necessary algorithmic framework by developing robust computational methods. To evaluate the proposed model we focus on forecasting future stock returns. This is a problem of primary interest in financial statistics since it is closely related with risk management, portfolio allocation and asset pricing; see for example Aguilar and West (2000), Rapach and Zhou (2013) and Clements and Liao (2017) for more detailed discussions. We test the proposed methods 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 utilizing 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 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; see the supplementary material (Alexopoulos et al., 2021) for the details of the empirical method used for outlier detection. The middle panel depicts a summary statistic of a 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; by choosing .5 as threshold we avoid to over- or underestimate the number of jumps for each stock at each day, see the supplementary material for more details. 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). 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 a joint Markovian modelling of jump intensities across time which might reveal some predictability of jumps. Indeed, the results of our proposed joint modelling formulation provide 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

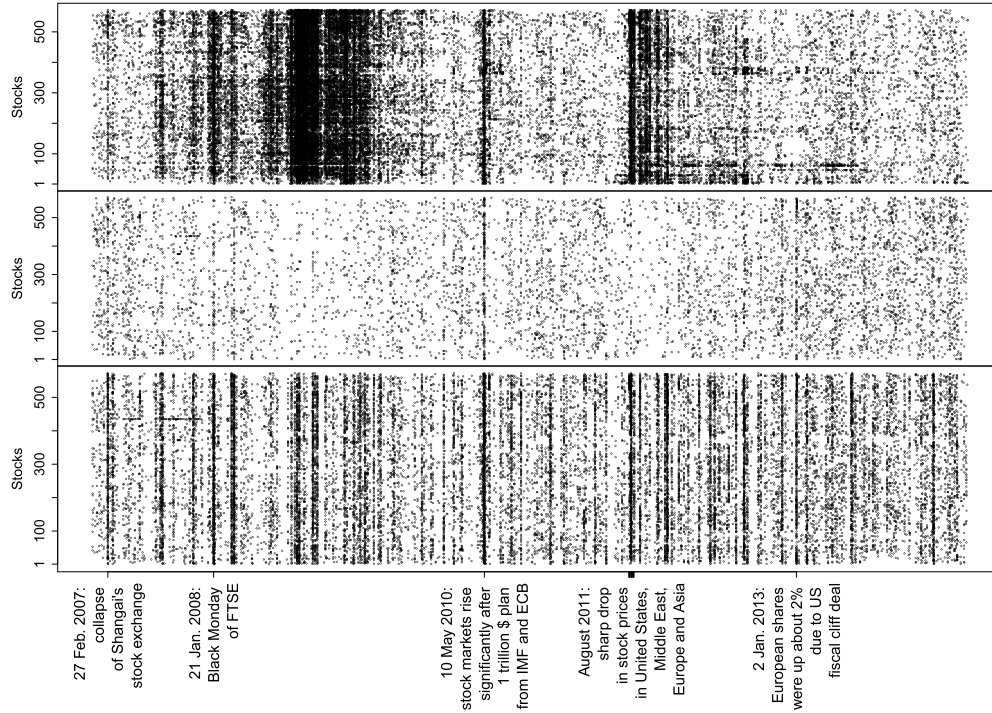


Figure 1: Each dot indicates a date between 10/1/2007 to 11/6/2014 (x-axis) in which at least one jump has been identified in observed daily log-returns of 571 stocks from the STOXX Europe 600 Index. The plot in the top panel is constructed by using an empirical method for detection of outliers in the 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.

intensities in each stock are in concordance with the well-known (Eraker et al., 2003) a priori expectation for one jump every few months. To avoid identifiability problems of the factor models extra assumptions for the form of the loadings matrix are needed. Nevertheless, it is recognised (Ghosh and Dunson, 2009) that MCMC algorithms developed under identifiability constraints exhibit slow convergence to the target distribution of interest. To improve the mixing properties of the proposed MCMC algorithm we follow Bhattacharya and Dunson (2011) and we do not impose the usual (see, e.g., Aguilar and West, 2000) identifiability constraints on the factor loadings parameters. We also note that there is a considerable amount of literature dedicated to the modelling of time series count data; see for example Fokianos et al. (2009), Jung et al. (2011), Pillow and Scott (2012) and Buesing et al. (2014). Our modelling perspective is applied on unobserved count data and, thus, the aim of the proposed latent factor model is twofold. First, since a few jumps are expected to occur during the observation period we borrow information across all the stocks in order to learn the unobserved processes that

drive the evolution of jumps across time and stocks. Second, we exploit the dimension reduction achieved by the latent factors to preserve the scalability of our method.

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 drawing 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 4.3 of this paper for a simulation that illustrates this issue.

To assess the predictive performance of the different models we compare the corresponding posterior predictive distributions by utilizing proper scoring rules (Gneiting and Raftery, 2007). We provide a full-fledged quantitative evaluation of the obtained forecasts by calculating logarithmic scores, such as predictive log Bayes factors (Geweke and Amisano, 2010), interval and continuous ranked probability scores (Gneiting and Katzfuss, 2014) as well as root mean squared errors. We estimate the posterior predictive distributions by employing sequential Monte Carlo methods such as the particle filters (Chopin and Papaspiliopoulos, 2020a) and the annealed importance sampling (Neal, 2001) algorithms.

The structure of the remaining of the paper is as follows. In Section 2 we provide a review of the literature related to univariate and multivariate SV models in econometrics as well as the modelling of jumps in financial applications. Section 3 presents our proposed modelling framework and Section 4 describes the methods that we develop to conduct Bayesian inference for the proposed model. In Section 5 we present the computational techniques that we use to assess the predictive performance of the proposed model. Section 6 presents results and insights from the application of our methods on the real dataset and Section 7 concludes with a small discussion.

2 Related work

There are two main classes of models that have been considered in the literature for the modelling of financial returns. The first class is based on observation-driven models where the time evolution of the variance of the returns is modelled by using past observations and variances. These models are known as autoregressive conditional heteroscedasticity (ARCH) models and have been developed by Engle (1982). A very popular parametrization, the generalized ARCH model, introduced later by Bollerslev (1986) while Bollerslev (1987) and Harvey and Chakravarty (2008) presented modifications that

account for the heavy tails of the distributions of the observed returns; see also Harvey (2013) and references therein. Creal et al. (2011) extended this modelling approach in the multivariate case; for the Bayesian analysis of these models see, for example, Vrontos et al. (2000) and Vrontos et al. (2003). The second class is known as parameter-driven models where the time evolution of the variance of the returns is modelled by employing latent stochastic processes. The models that belong to this class are the SV models which constitute the base of our proposed modelling approach.

The SV model proposed by Taylor (1982) in order to account for the time evolution of the volatilities in financial time series while Hull and White (1987) and Chesney and Scott (1989) have employed the SV model as discretization of continuous time SV diffusion processes. As mentioned earlier some early works on the Bayesian estimation of SV models include, but are not limited to, the papers presented by Jarquier et al. (1994), Kim et al. (1998) and Chib et al. (2002). Omori et al. (2007) dealt with efficient Bayesian estimation of the SV model by taking into account the correlation between the observation error and the error of the stochastic variance process; see also Silva et al. (2006) and Delatola et al. (2011) for other extensions of the basic SV model. A more recent contribution on the Bayesian analysis of univariate SV models has been offered by Kastner and Frühwirth-Schnatter (2014) while Zhang et al. (2020) developed a new class of SV models with autoregressive moving average errors. Considering a jump component to model the empirically observed fat tails of stock returns proposed, in his pioneering paper, by Merton (1976) while Bates (1996) and Andersen et al. (2002) combined jump diffusions with SV models. Bayesian estimation of SV models with jumps has been considered, among others, by Chib et al. (2002), Nakajima and Omori (2009) and Eraker et al. (2003) and Johannes et al. (2009) where the latter two papers account for jumps in the volatility process as well; see also Bandi and Renò (2016) for more recent research on the occurrence of contemporaneous jumps in the returns and volatilities.

To conduct joint modelling of the observed financial time series, multivariate SV models have been considered more than twenty years ago; see for example Harvey et al. (1994) and Asai et al. (2006) for a review. The existing multivariate SV models assume either constant correlations over time or some form of dynamic correlation modelling through factor models with factors being independent univariate stochastic volatility models. The latter approach has been considered, for example, by Chib et al. (2006), Kastner et al. (2017) and Kastner (2019). In a similar framework Chan and Eisenstat (2018) and Kastner and Huber (2020) deal with Bayesian estimation of time varying parameter vector autoregressive models with stochastic volatility. An alternative multivariate SV modelling approach proposed recently by Dellaportas et al. (2015) where all the elements of the covariance matrix of the observed time series evolve over time. More closely related to the present paper, Bollerslev et al. (2008), Jacod et al. (2009) and Clements and Liao (2017) study the presence of common or not common jumps in multiple time series. Aït-Sahalia et al. (2015) model the dependencies of jumps by combining the SV model with a multivariate Hawkes process (Hawkes, 1971) that models the propagation of jumps over time and across assets by introducing a feedback from the jumps to their intensities and back. They employ the generalised version method of moments to estimate the parameters of the model. We close this literature review

section by noting that tools for the identification of non-observed events that affect the evolution of time series have been also developed by Hamilton (1989) and Billio et al. (2012) that deal with the problem of identifying structural changes in time series of economic indices.

3 Latent jump modelling

3.1 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 ; $\text{Gam}(\alpha, \beta)$ and $\text{IGam}(\alpha, \beta)$ 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 i th 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} ; $R_i = (r_{i1}, \dots, r_{iT})$, and $R_t = (r_{1t}, \dots, r_{pt})$ 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 $t+1, \dots, t+\ell$; e.g. $R_{t+1:t+\ell}$ is the collection of vectors $R_{t+1}, \dots, R_{t+\ell}$. The transpose of a vector or matrix is denoted by the superscript \top ; e.g. R^\top is the transpose of R ; $\mathbb{1}\{\cdot\}$ denotes an indicator function that takes value 1 if the event in brackets is true and 0 otherwise; for a function $f(x)$ we set $\nabla =: (\partial f/\partial x_1, \dots, \partial f/\partial x_d)$.

3.2 Stochastic volatility with jumps

We model stock returns as

$$r_{it} = \exp(h_{it}/2)\epsilon_{it} + \sum_{\kappa=1}^{n_{it}} \xi_{it}^{\kappa}, \quad \epsilon_{it} \sim \mathcal{N}(0, 1), \quad t = 1, \dots, T, \quad (3.1)$$

$$h_{it} = \mu_i + \phi_i(h_{i,t-1} - \mu_i) + \sigma_{i\eta}\eta_{it}, \quad \eta_{it} \sim \mathcal{N}(0, 1), \quad t = 1, \dots, T \quad (3.2)$$

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 i th stock at time t follow a Poisson distribution,

$$n_{it} \sim \text{Poisson}(\Delta_{it}\lambda_{it}), \quad (3.3)$$

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 κ th 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 $\mu_i \sim \mathcal{N}(0, 10)$, $(\phi_i + 1)/2 \sim \text{Beta}(20, 1.5)$ and $\sigma_{i\eta}^2 \sim \text{Gam}(1/2, 1/2)$.

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 $\sigma_{i\xi}^2 \sim \text{IGam}(3, \text{range}_i^2/18)$, 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, $E[\sigma_{i\xi}^2] = \text{range}_i^2/36$, 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.

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, $\lambda_{it} \sim \text{Gam}(\delta, c)$, 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

$$p(n_{it}) = \frac{\Gamma(\delta + n_{it})}{\Gamma(\delta)n_{it}!} \beta^\delta (1 - \beta)^{n_{it}}, \quad (3.4)$$

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 $E(\lambda_{it}) = 0.02$. 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.

3.3 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

$$\lambda_{it} = \lambda^* (1 + e^{-y_{it}})^{-1}, \quad (3.5)$$

which implies that $\lambda_{it} < \lambda^*$. The parameter λ^* is, thus, the upper bound for the jump intensities of all the stocks at any time point. This is an important feature of

our modelling approach since it allows to incorporate the necessary (Chib et al., 2002; Eraker et al., 2003) prior information about the expected number of jumps which are typically described as large but infrequent moves. Furthermore, as discussed below, by carefully choosing this upper bound we are able to construct a prior for the jump intensities which is similar to the $Gam(1, 50)$ distributions that we used in the case of univariate SV models with jumps. We also note that this type of information is not necessary only in Bayesian applications; see for example Honore (1998) for a similar discussion related to maximum likelihood estimation of models with jumps. 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

$$Y_t = B + WF_t, \quad (3.6)$$

$$F_t = AF_{t-1} + E_t, \quad t = 1, \dots, T, \quad (3.7)$$

$$F_0 \sim \mathcal{N}_K(0, \Sigma_F) \quad E_t \sim \mathcal{N}(0, I), \quad (3.8)$$

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

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 $w_{ik} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_w^2)$, $b_i \stackrel{iid}{\sim} \mathcal{N}(\mu_b, \sigma_b^2)$, and $\alpha_k \stackrel{iid}{\sim} U(-1, 1)$.

The specification of the hyperparameters of the priors assigned on the parameters A, B and W require extra care because they affect, through (3.5), the imposed prior on λ_{it} . The fact that our modelling formulation of Section 3.2 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. As discussed earlier, this an important characteristic of our modelling approach since it penalizes the estimation of a large number of jumps which is not consistent with the prior information for one or two jumps every 100 days. A small sensitivity analysis, not reported here, confirms that choosing a value considerably greater than 0.15 does not allow the efficient separation of the jumps from the underlying volatility process of the stock daily returns. Moreover, by using values notably smaller than 0.15 we estimate quite low jump intensities and this can result in less accurate prediction of future stock returns; in the supplementary material we compare predictions obtained under different prior assumptions for the expected number of jumps. We have concluded that a value for λ^* between 0.1 and 0.2 does not have

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

Table 1: Mean, variance and mode of the $Gam(1, 50)$ prior distribution assumed for the jump intensities λ_{it} in the independent model and for the prior induced by the dynamic factor model.

a substantial effect on the results presented in the rest of the paper. 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 (3.5) 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 between the two priors we also examine the impact of different choices for their hyperparameters in the forecasts of future observations. See the supplementary material for the related results.

4 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 in an efficient simultaneous update of the log-volatilities vector and 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 equations (3.1)–(3.3) and (3.5)–(3.8). In Sections 4.1 and 4.2 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.

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.

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

- 1: Set the number of iterations S .
 - 2: **for** $s = 1, \dots, S$ **do**
 - 3: **for** $i = 1, \dots, p$ **do**
 - 4: Sample from $p(H_i, \phi_i, \sigma_{i\eta}^2 | \mu_i, \mu_{i\xi}, \sigma_{i\xi}^2, N_i, R_i)$ by using Metropolis-Hastings.
 - 5: Sample from $p(\mu_i | H_i, \phi_i, \sigma_{i\eta}^2)$ which is a Gaussian density.
 - 6: Sample from $p(N_i | H_i, W_i, F, R_i, b_i, \mu_{i\xi}, \sigma_{i\xi}^2)$ by using rejection sampling.
 - 7: Sample from $p(\Xi_i | H_i, N_i, R_i, \mu_{i\xi}, \sigma_{i\xi}^2)$ which is a Gaussian density.
 - 8: Sample from $p(\mu_{i\xi} | N_i, \Xi_i, \sigma_{i\xi}^2)$ which is a Gaussian density.
 - 9: Sample from $p(\sigma_{i\xi}^2 | N_i, \Xi_i, \mu_{i\xi})$ which is an 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**
-

4.1 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 $p(n_{it} | h_{it}, \mu_{i\xi}, \sigma_{i\xi}^2, b_i, W_i, F_t, r_{it})$. This algorithm is based on Proposition 1. Notice also that a discrete distribution with pmf $p(n)$ is called log-concave if $p(n)^2 \geq p(n-1)p(n+1)$ is satisfied for all n ; see for example Devroye (1986) for more details. For the remaining of this subsection the unnormalized $p(n_{it} | h_{it}, \mu_{i\xi}, \sigma_{i\xi}^2, b_i, W_i, F_t, r_{it})$ is denoted by $\tilde{p}(n)$.

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

Proof. See the supplementary material. □

Due to Proposition 1, the target distribution is a unimodal distribution (Devroye, 1986). Let m be an integer at the right of the mode of the distribution. We define the probability mass function (pmf)

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

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

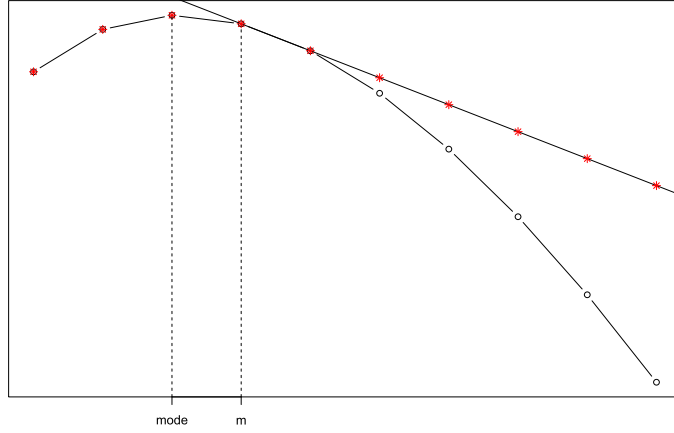


Figure 2: $\tilde{p}(n)$ and its bounding line in the log-scale. Circles: $\log(\tilde{p}(n))$, stars: $\log(q_m(n))$.

Algorithm 2 Rejection sampler for the number of jumps.

Generate $u \sim U(0, 1)$
if $v \leq \frac{\sum_{n=0}^{m-1} \tilde{p}(n)}{c}$ **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 $(\tilde{p}(m) - \tilde{p}(m + 1))/\tilde{p}(m)$, truncated from the left at the point m .
 until $Vcq_m(n) \leq \tilde{p}(n)$
end if
return n

$(m, \log(\tilde{p}(m)))$ and $(m + 1, \log(\tilde{p}(m + 1)))$ bounds from above the logarithm of $\tilde{p}(n)$. See Figure 2 for an illustration. A rejection algorithm to sample from the distribution with pmf $p(n_{it}|h_{it}, \mu_{i\xi}, \sigma_{i\xi}^2, b_i, W_i, F_t, r_{it})$ proceeds as summarized by Algorithm 2. Note that Algorithm 2 implies the evaluation of $\tilde{p}(0), \dots, \tilde{p}(m + 1)$ at each iteration of the proposed MCMC algorithm.

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 by equations (3.1) and (3.2) we have*

that $p(\xi_{it}|n_{it}, h_{it}, \mu_{i\xi}, \sigma_{i\xi}^2, r_{it})$ is equal to the Gaussian density

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

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

Proof. See the supplementary material. \square

4.2 Log-volatilities and factors: MCMC for latent Gaussian models

In the 4th and 12th lines of Algorithm 1 we sample the paths of the log-volatilities H_i and factors F respectively. Moreover, each path is updated jointly with the parameters of the corresponding autoregressive models defined by equations (3.2) and (3.8). These steps correspond to the joint sampling of latent paths and parameters of latent Gaussian models. A latent Gaussian model is defined in terms of a non-Gaussian likelihood and a latent Gaussian field controlled by some parameters. By denoting with \mathbf{X} a d -dimensional random vector that corresponds either to a latent log-volatility or to a factor path, with \mathbf{x} its realization and with θ any parameters we target the density

$$p(\mathbf{x}, \theta) \propto \exp\{g(\mathbf{x})\} \mathcal{N}(\mathbf{x}|0, C_\theta) \pi(\theta), \quad (4.1)$$

where $g(\mathbf{x})$ is the log-likelihood function, C_θ denotes the covariance matrix of the latent Gaussian field, parametrized in terms of θ and $\pi(\theta)$ denotes the density of the prior for θ . We have assumed, without loss of the generality, that the latent Gaussian field has zero mean. We also note that only throughout the present Section we violate our notation and we use lower-case letters to denote the realizations of a random vector; the dependence on any data has been suppressed.

To draw samples from the target in (4.1) two approaches have dominated in the literature. One of them is to alternate sampling from the conditionals $p(\mathbf{x}|\theta)$ and $p(\theta|\mathbf{x})$. Sampling from $p(\mathbf{x}|\theta)$ is quite challenging since usually corresponds to a high dimensional distribution. To draw the desired samples several well-established methods, based on the Metropolis-Hastings algorithm, have been used, see e.g. Roberts and Stramer (2002), Murray et al. (2010), Girolami and Calderhead (2011) and Cotter et al. (2013); in the latter approach the proposal distribution utilizes information both from the likelihood and the Gaussian prior of the latent states and this is an important feature of the proposed sampler (Titsias and Papaspiliopoulos, 2018). Sampling from $p(\theta|\mathbf{x})$ is also challenging since the parameters θ are highly correlated with the Gaussian latent states of the model. See, for example, Papaspiliopoulos et al. (2007), Murray and Adams (2010) and Yu and Meng (2011) for efficient methods that have been developed for the update of θ , conditional on the latent Gaussian field, in a more general class of Bayesian hierarchical models. In the second approach, joint sampling of the latent states \mathbf{x} and the parameters θ is conducted. This perspective is considered, among others, by Knorr-Held and Rue (2002) where the latent states and the parameters are updated in, carefully selected, blocks and by Titsias and Papaspiliopoulos (2018) where a sampler

which utilizes auxiliary variables to perform updates that are both prior and likelihood informed has been constructed.

In the present paper we follow the latter approach and we update jointly the latent states and the parameters of a given model by combining the methods proposed by Yu and Meng (2011) and Titsias and Papaspiliopoulos (2018). Following Yu and Meng (2011) we interweave the so-called (Papaspiliopoulos et al., 2007) centred and non-centred parametrization of the latent states and by utilizing the algorithm developed by Titsias and Papaspiliopoulos (2018) we sample jointly \mathbf{x} and θ under each one of the two parametrizations. We show that by utilizing the proposed Metropolis-Hastings step we find efficiency gains relative to alternative samplers that can be used to update the latent log-volatilities, the factor paths and their parameters. In the rest of the Section we present the proposed methodology in the case of a general latent Gaussian model and in the supplementary material we provide the details for the application of the method in order to sample the latent factors F as well as each log-volatility path H_i and their parameters.

Metropolis-Hastings for the joint update of latent paths and parameters

To construct a Metropolis-Hastings step in order to sample jointly \mathbf{x} and θ we work as in Titsias and Papaspiliopoulos (2018) and we introduce auxiliary variables $\mathbf{u} \sim \mathcal{N}_d(\mathbf{x}, (\delta/2)I_d)$. We consider, thus, the expanded target

$$p(\mathbf{x}, \theta, \mathbf{u}) \propto \exp\{g(\mathbf{x})\} \mathcal{N}_d(\mathbf{x}|0, C_\theta) \mathcal{N}_d(\mathbf{u}|\mathbf{x}, (\delta/2)I_d) \pi(\theta). \quad (4.2)$$

Samples from the expanded target can be drawn by first drawing \mathbf{u} from $p(\mathbf{u}|\mathbf{x}) = \mathcal{N}_d(\mathbf{u}|\mathbf{x}, (\delta/2)I_d)$ and then employing the Metropolis-Hastings algorithm to draw from the intractable $p(\mathbf{x}, \theta|\mathbf{u})$. To perform the latter update we utilize a proposal distribution of the form

$$q(\mathbf{x}', \theta'|\mathbf{x}, \theta, \mathbf{u}) = q(\mathbf{x}'|\mathbf{x}, \theta', \mathbf{u}) q(\theta'|\theta). \quad (4.3)$$

To construct $q(\mathbf{x}'|\mathbf{x}, \theta', \mathbf{u})$ we approximate the intractable log-likelihood $g(\cdot)$ by employing a first order Taylor expansion of $g(\cdot)$ around the current value \mathbf{x} . We obtain, then, the proposal density

$$\begin{aligned} q(\mathbf{x}'|\mathbf{x}, \theta', \mathbf{u}) &\propto \exp\{g(\mathbf{x}) + \nabla g(\mathbf{x})^\top (\mathbf{x}' - \mathbf{x})\} \mathcal{N}_d(\mathbf{x}'|0, C_{\theta'}) \mathcal{N}_d(\mathbf{u}|\mathbf{x}', (\delta/2)I_d) \\ &\propto \mathcal{N}_d\left(\mathbf{x}' \mid \frac{2}{\delta} Q_{\theta'} (\mathbf{u} + \frac{\delta}{2} \nabla g(\mathbf{x})), Q_{\theta'}\right), \end{aligned} \quad (4.4)$$

where $Q_{\theta'} = (C_{\theta'}^{-1} + \frac{2}{\delta} I_d)^{-1}$. To reduce the costs of generating from the proposal in (4.4) as well as evaluating the resulting acceptance ratio we follow Titsias and Papaspiliopoulos (2018) and we define new auxiliary variables \mathbf{z} by using the reparametrization

$$\mathbf{z} \equiv \mathbf{u} + (\delta/2) \nabla g(\mathbf{x}) \sim \mathcal{N}_d(\mathbf{x} + (\delta/2) \nabla g(\mathbf{x}), (\delta/2) I_d).$$

The expanded target in (4.2) becomes now

$$p(\mathbf{x}, \mathbf{z}, \theta) \propto \exp\{g(\mathbf{x})\} \mathcal{N}_d(\mathbf{x}|0, C_\theta) \mathcal{N}_d(\mathbf{z}|\mathbf{x} + (\delta/2) \nabla g(\mathbf{x}), (\delta/2) I_d) \pi(\theta),$$

Algorithm 3 Auxiliary gradient-based sampler to draw samples from (4.1).

1. Sample $z \sim \mathcal{N}_d(x + (\delta/2)\nabla g(x), (\delta/2)I_d)$.
 2. Propose $\theta' \sim q(\theta'|\theta)$ and $x' \sim q(x'|\theta', z)$ and accept them according to the Metropolis-Hastings ratio in (4.6).
-

while the proposal in (4.4) can be written as

$$q(x'|\theta', z) = \mathcal{N}_d(x' | \frac{2}{\delta} Q_{\theta'} z, Q_{\theta'}) = \frac{1}{\mathcal{Z}(z, \theta')} \mathcal{N}_d(x'|0, C_{\theta'}) \mathcal{N}_d(z|x', C_{\theta'}), \quad (4.5)$$

where $\mathcal{Z}(z, \theta') = \mathcal{N}_d(z|0, C_{\theta'} + (\delta/2)I_d)$. Notice that, importantly, the proposed x' becomes conditionally independent from the current x given the auxiliary variable z . In particular, the proposal $q(x', \theta'|x, \theta, u)$ in (4.3) becomes

$$q(x', \theta'|x, \theta, z) = q(x'|\theta', z)q(\theta'|\theta),$$

and the acceptance ratio of the move is

$$\exp \{g(x') - g(x) + \nu(z, x') - \nu(z, x)\} \times \frac{\mathcal{Z}(z, \theta')\pi(\theta')q(\theta'|\theta)}{\mathcal{Z}(z, \theta)\pi(\theta)q(\theta'|\theta)}, \quad (4.6)$$

where $\nu(z, x) = (z - x - (\delta/4)\nabla g(x))^\top \nabla g(x)$. In our application we choose $q(\theta'|\theta)$ to be a simple random walk proposal distribution with step-size κ . Algorithm 3 summarizes the proposed Metropolis-Hastings step for the joint update of x and θ . Finally, we note that following Titsias and Papaspiliopoulos (2018) we learn κ and δ during the burn-in period by decomposing step 2 in Algorithm 3 in two sub-steps. We first update x alone and tune δ in order to achieve an acceptance ratio between 50% and 60%. Then, we update (x, θ) jointly and we tune κ to obtain acceptance ratio in the range of 20% to 30%.

Interweaving strategy for the parameters

Although drawing the latent path x jointly with the parameters θ aims to overcome convergence problems that occurring due to their high dependence, it is well-known that the parametrization of latent Gaussian models as well as of more general Bayesian hierarchical models plays a key role in the efficiency of MCMC algorithms (Papaspiliopoulos et al., 2007). To address convergence issues that arise from the parametrizations of the latent log-volatility models and from the latent factor model for the jump intensities we employ the, so-called, ancillarity–sufficiency interweaving strategy (ASIS) developed by Yu and Meng (2011); see also Kastner and Frühwirth-Schnatter (2014) for the benefits of the ASIS in the case of univariate SV models and Simpson et al. (2017) where the ASIS has been applied in more general dynamic linear models. In the case of a latent Gaussian model with latent states x and parameters θ the application of ASIS is briefly described as follows. After the joint update of x and θ with the Metropolis-Hastings step described by Algorithm 3 we transform the latent path x to its non-centred parametrization \tilde{x} by using an one-to-one transformation. We re-draw then the parameters θ from $p(\theta|\tilde{x})$ before transforming \tilde{x} back to x by using the inverse transformation.

To employ ASIS in our application we first note that equations (3.1)–(3.2) and (3.5)–(3.8) define the centred parametrizations (Papaspiliopoulos et al., 2007) of the corresponding models. The non-centred parametrization for the latent states of a latent Gaussian model can be found by following the reasoning developed by Yu and Meng (2011): under a non-centred parametrization x has to be an ancillary statistic for θ , i.e., the distribution of x needs to be free of θ . We note that there are cases where a many-to-one transformation is needed (Papaspiliopoulos et al., 2003) but this is not required in our application. It is easy to see that in the case of the latent log-volatilities h_{it} defined in (3.2) the variables $\tilde{h}_{it} = (h_{it} - \mu_i)/\sigma_{i\eta}$ are in a non-centred parametrization while in the case of the latent factors defined by (3.7) the transformation $\Gamma_t = F_t - AF_{t-1}$ results in a non-centred factor model for the jump intensities. ASIS is completed by re-drawing the log-volatility parameters μ_i, ϕ_i and $\sigma_{i\eta}^2$ and the factor parameters A under the non-centred parametrizations and by using the inverse transformation to return back to h_{it} and F_t .

In the supplementary material we provide all the details for our implementation of ASIS along with simulation studies which emphasize the gains obtained by using the ASIS. By describing briefly the results we note that combining ASIS with the auxiliary gradient-based sampler (Algorithm 3) improves substantially the efficiency of the proposed MCMC algorithm (Algorithm 1) in drawing samples from the parameters μ_i and $\sigma_{i\eta}^2$ of the log-volatility processes (see Figures S.2 and S.3 in the supplementary material) as well as from the parameters in the matrix A of the latent factor model for the jump intensities; see Figure S.4 in the supplementary material.

4.3 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 (3.1)–(3.2) and omitting the jump component in (3.1) 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 3.2. 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).

5 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 assess the forecasts obtained from the various models by calculating proper scoring rules (Gneiting and Raftery, 2007). A scoring rule assigns a numerical score to a probabilistic forecast by taking into account the posterior predictive distribution and the realized observation and it is proper if the expected score is maximized when the estimated predictive distribution coincides with the true distribution.

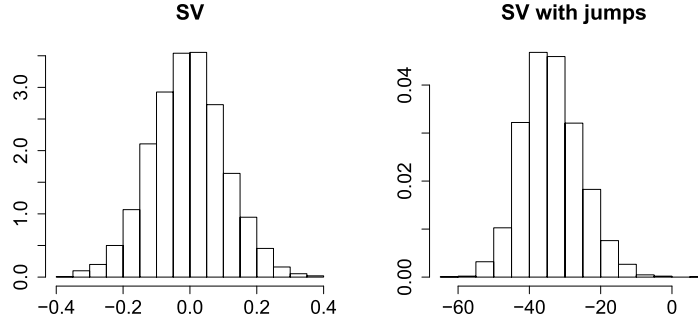


Figure 3: Histograms of normalized weights multiplied by their number, in the log-scale, calculated as the ratio between the likelihood of the exact and the approximated, by a mixture of Gaussian distributions, model in the case of stochastic volatility (SV) models with and without jumps.

Proper scoring rules allow for the joint evaluation of the calibration (compatibility of forecasts and observations) and the sharpness (concentration of the predictive distributions); see, for example, Geweke and Amisano (2010) and Krüger et al. (2020) for more detailed discussions. We focus on three proper scoring rules: (i) logarithmic scores, and particularly, the predictive log Bayes factors which compare the models with respect to their marginal likelihoods, (ii) continuous ranked probability scores which are employed for the assessment of the predictive cumulative distribution functions (CDFs) and (iii) interval scores which are utilized for the evaluation of the prediction credible intervals. We complete a full-fledged assessment of the predictive performance of the different models by calculating root mean squared errors of the obtained forecasts.

To calculate these summary measures we draw samples from the posterior predictive distributions of interest and we estimate their densities by utilizing sequential Monte Carlo methods (Chopin and Papaspiliopoulos, 2020b). In the rest of the Section we present the computational framework that we built in order to conduct out-of-sample forecasts with univariate SV models with and without jumps as well as by using the proposed joint modelling approach for the jump intensities. To evaluate the accuracy of the obtained estimators we calculate the effective sample size of the samples drawn from the posterior predictive distributions. The effective sample size of a given sample is commonly employed (Neal, 2001; Chopin et al., 2013) to assess the variance of estimators obtained from weighted rather than independent samples; see in the supplementary material for more details. For the remaining of this section we assume that the parameters Θ , A and W are fixed to their in-sample estimated posterior means and we omit them from our notation.

5.1 Prediction with independent SV models

Let \mathbb{M}_{svj} and \mathbb{M}_{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. To calculate predictive Bayes factors and the continuous ranked probability and interval scores we consider the posterior predictive distributions with densities $p(r_{it}|r_{i,1:t-1})$, $t = T + 1, \dots, T + \ell$. The predictive log Bayes factors are defined as

$$\log(p(R_{T+1:T+j}|R_{1:T}, \mathbb{M}_{\text{svj}})) - \log(p(R_{T+1:T+j}|R_{1:T}, \mathbb{M}_{\text{sv}})), \quad j = 1, \dots, \ell, \quad (5.1)$$

where for both \mathbb{M}_{svj} and \mathbb{M}_{sv} we have that

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

Moreover, if l_{it} and u_{it} denote the $\alpha/2$ and $1 - \alpha/2$ quantiles of a $(1 - \alpha) \times 100\%$ prediction interval for the value of r_{it} then its interval score is given by the formula

$$(u_{it} - l_{it}) + \frac{2}{\alpha}(l_{it} - r_{it})\mathbb{1}\{r_{it} < l_{it}\} + \frac{2}{\alpha}(r_{it} - u_{it})\mathbb{1}\{r_{it} > u_{it}\}. \quad (5.2)$$

We note, finally, that the calculation of the continuous ranked probability scores that we employ for the evaluation of predictive CDFs is also based on samples from $p(r_{it}|r_{i,1:t-1})$; see the supplementary material for details.

We draw the desired samples and we estimate the densities $p(r_{it}|r_{i,1:t-1})$ based on the formula

$$p(h_{i,0:t}|r_{i,1:t}) = \frac{p(h_{i,0:t-1}|r_{i,1:t-1})p(r_{it}|h_{it})p(h_{it}|h_{i,t-1})}{p(r_{it}|r_{i,1:t-1})}, \quad (5.3)$$

and by utilizing sequential Monte Carlo methods also known as particle filters (Chopin and Papaspiliopoulos, 2020a). From the output of a particle filter algorithm that targets the posterior in (5.3) we are able to draw samples from the posterior predictive distribution of r_{it} and to estimate the normalizing constants $p(r_{it}|r_{i,1:t-1})$. See in the supplementary material for more details and for the pseudocode of the particle filter algorithm.

5.2 Prediction by using joint modelled jump intensities

Let \mathbb{M}_{mvj} be our proposed SV with jumps model, defined by equations (3.1)–(3.3) and (3.5)–(3.8). To compare the predictive performance of \mathbb{M}_{mvj} with the performance of competing models we need to estimate the predictive densities $p(R_{T+j}|R_{1:T+j-1}, \mathbb{M}_{\text{mvj}})$, for each $j = 1, \dots, \ell$. This could be achieved with a particle filter algorithm, constructed similarly to the case of independent SV models, as described in Section 5.1. However, in the case of \mathbb{M}_{mvj} , particle filter algorithms deliver poor estimations of the marginal likelihoods since the importance weights, calculated at each of their steps, have large variance.

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 draw samples from the posterior predictive distributions of interest and to estimate their densities. In the rest of the section we provide a detailed description of how we apply AIS, \mathbb{M}_{mvj} is suppressed throughout except when it is necessary.

Let g_0, \dots, g_ℓ be the required sequence of auxiliary distributions. By noting that each g_j must be proportional to a computable function and satisfies $g_j \neq 0$ 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 $j = 0, \dots, \ell - 1$. 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 transition densities $p(H_{T+j} | H_{T+j-1})$ and $p(F_{T+j} | F_{T+j-1})$. To sample from g_0 we utilize samples from $p(H_{0:T}, F_{0:T} | R_{1:T})$. We compute the importance weights of the obtained samples by first noting that g_j are proportional to the computable functions

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

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

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

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

$$\begin{aligned} p(R_t | H_t^s, F_t^s) &= \prod_{i=1}^p p(r_{it} | h_{it}^s, F_t^s) \\ &= \prod_{i=1}^p \sum_{n_{it}=0}^{\infty} \mathcal{N}(r_{it} | n_{it} \mu_{i\xi}, \exp(h_{it}^s) + n_{it} \sigma_{i\xi}^2) p(n_{it} | \Delta_{it} \lambda_{it}^s). \end{aligned} \quad (5.4)$$

We also note that the infinite sum in (5.4) can be truncated without affecting the results of the algorithm (Johannes et al., 2009). We refer to 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

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

This implies that by using only the i th element of the product in (5.4) we obtain an estimator of the i th 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 $\omega_{i,T+1}^s = 1$. Thus, $p(r_{i,T+1:T+j}|R_{1:T}, \mathbb{M}_{\text{mvj}})$ is estimated by $\hat{p}(r_{i,T+1:T+j}|R_{1:T})$ accurately since the latter is based on importance sampling weights with reduced variance and can be used for the estimation of log Bayes factors of the form

$$\log(p(r_{i,T+1:T+j}|R_{1:T}, \mathbb{M}_{\text{mvj}})) - \log(p(r_{i,T+1:T+j}|R_{1:T}, \mathbb{M}_{\text{svj}})), \quad (5.5)$$

where $p(r_{i,T+1:T+j}|R_{1:T}, \mathbb{M}_{\text{svj}})$ can be estimated with the methods presented in Section 5.1. We also note that 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 approximate log Bayes factors

$$\sum_{i=1}^p \log(p(r_{i,T+1:T+j}|R_{1:T}, \mathbb{M}_{\text{mvj}})) - \sum_{i=1}^p \log(p(r_{i,T+1:T+j}|R_{1:T}, \mathbb{M}_{\text{svj}})). \quad (5.6)$$

Finally, $\{h_{i,0:T+j}^s, F_{0:T+j}^s\}_{s=1}^S$ can be utilized in order to draw samples from the posterior predictive distributions needed to calculate the continuous ranked probability and the interval scores; see the supplementary material for details.

6 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 also found in the supplementary material as well as in Alexopoulos (2017).

6.1 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

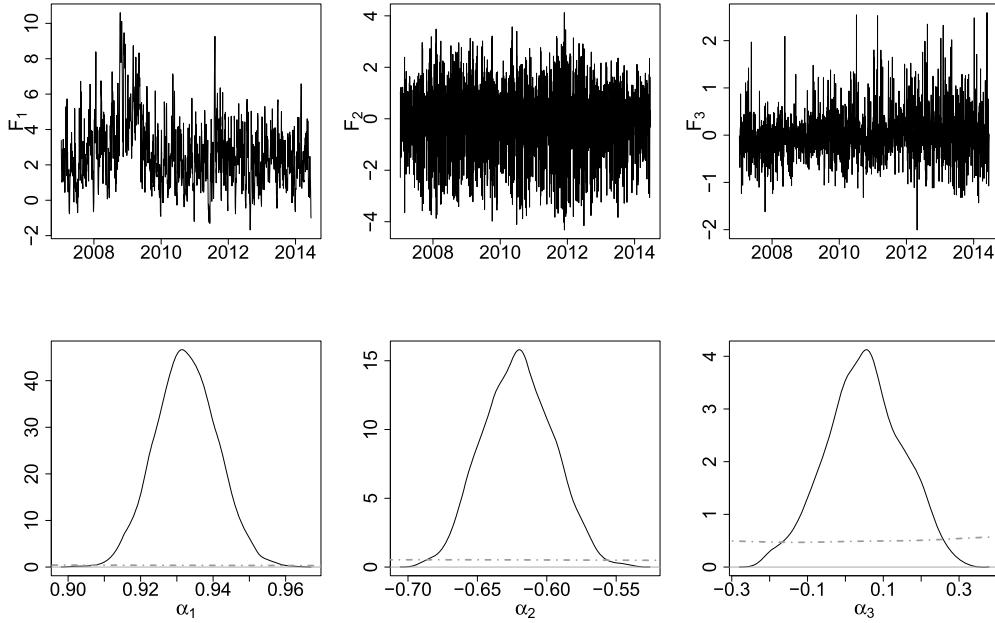


Figure 4: First row: posterior mean of the paths of the three autoregressive factors F_1 , F_2 and F_3 that we used to model the evolution of the jump intensities across stocks and across time. Second row: posterior (solid lines) and prior (dotted lines) distributions of the persistent parameters α_1 , α_2 and α_3 .

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 $K = 3$ factors. We ran all the MCMC algorithms for 250,000 iterations using a burn-in period of 100,000 sampled points and we collected 1,000 (thinned) posterior samples. In the supplementary material we give the computing times for the real and simulated data analysis that we have implemented and we evaluate the accuracy of the sequential Monte Carlo methods that we used to conduct the assessment of the predictive performance of the different models.

6.2 Results

Bayesian inference for parameters and latent states

Figure 4 depicts the posterior means of the factor paths along with the posterior distributions of their persistent parameters α_1 , α_2 and α_3 . The posterior densities of α_1 and α_2 indicate strong evidence of autocorrelation in the first two factor processes while that

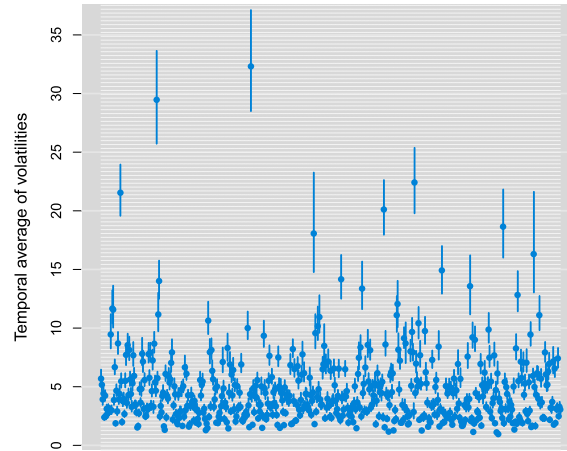


Figure 5: 95% credible intervals of the posterior distributions of the temporal averages of the volatility processes. Each dotted blue line corresponds to a different stock.

of α_3 suggests that a model with only $K = 2$ factors adequately captures the dependence of the jumps over times and across stocks. Therefore, all results in the following are based on $K = 2$ factors.

By noting that we have applied our methods on a very high-dimensional dataset we present posterior distributions for summary statistics related to the latent volatility and jump processes. Figure 5 shows, for each stock, the 95% credible interval of the temporal average of the volatility process. From the visual inspection of the Figure we note that the stocks can be separated in those with low, medium and high volatilities. More precisely, the posterior temporal average of the volatilities for the majority of the stocks is estimated lower than 5, for a few number of stocks the temporal average ranges between 5 and 15 while for less than 10 stocks the posterior temporal average has 95% credible interval with lower limit greater than 15. See also the supplementary material for additional results regarding Bayesian inference about the log-volatility processes and their parameters. Figure 6 presents the 95% credible intervals for the posterior distribution of the temporal sum of jumps for each stock with respect to the economic sector that it belongs to. The Figure indicates that for each stock have been identified 100 jumps on average between 10/1/2007 and 11/6/2014. We also note that the number of stocks for which almost no jumps have been estimated is very small and such stocks mainly belong to the financial, the industrial and the consumer- cyclical and non-cyclical, the energy and the communications sectors. In the supplementary material we present more results related to the Bayesian analysis of the jump processes of the SV models while Figure 1 offers a visual illustration of the estimated jumps underlying the time evolution of the stock returns.

As already noted in the introduction of the paper the top panel of Figure 1, which has been created by using an empirical method for jump identification, visualizes a dense estimation of the underlying jump process. The middle panel which corresponds to

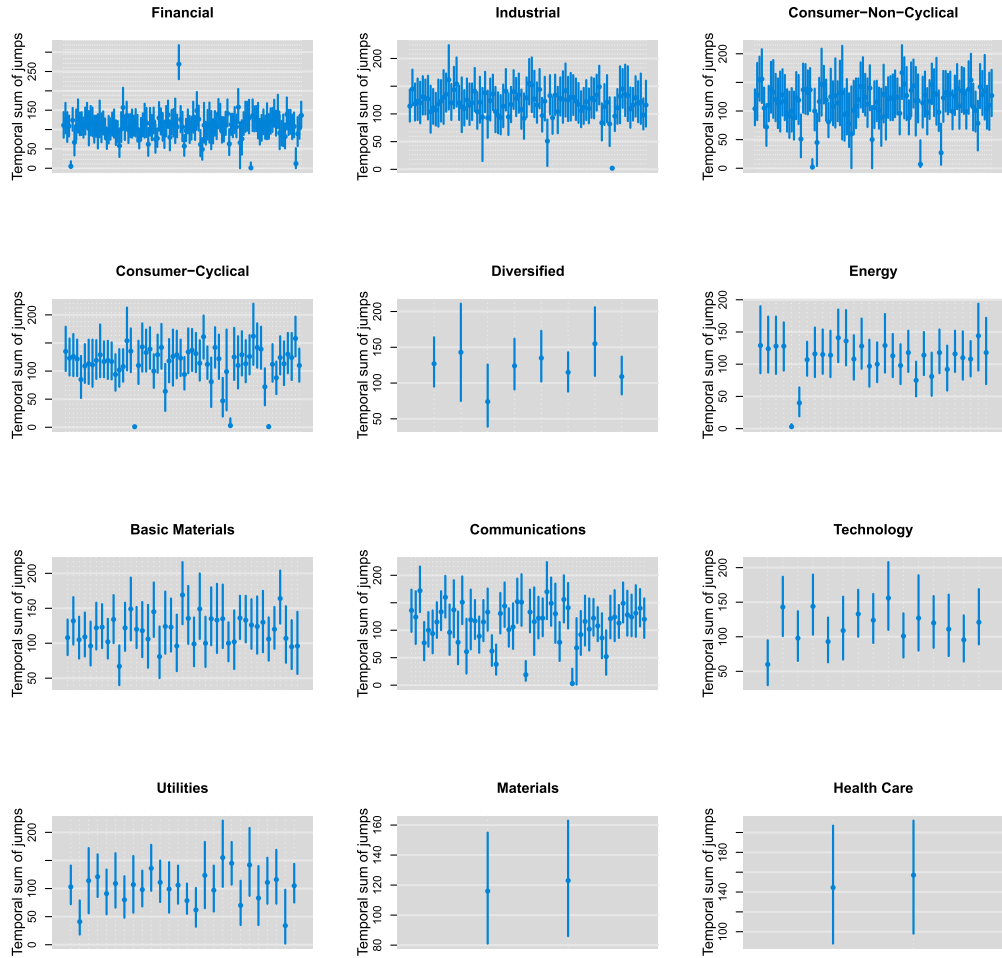


Figure 6: 95% credible intervals of the posterior distributions of the temporal sum of jumps. Each dotted blue line corresponds to a different stock and each plot is consisted of credible intervals for the stocks that belong to the economic sector indicated by its title.

independent modelling of jumps depicts a very sparse jump process whereas the bottom panel indicates less number of jumps than the empirical method but more than the independent SV models. To evaluate the estimation of the jump processes in the three panels we first emphasize that the true jumps in the stock prices are unobserved events. Therefore, drawing conclusions for the accuracy of the jumps estimation by only visually inspecting the Figure is difficult. We note, however, that the dense jump structure obtained from the empirical method can be explained by the fact that the time evolution of the volatility of the stock returns is not taken into account; resulting in a possible overestimation of the number of jumps. The latter combined with the more narrow

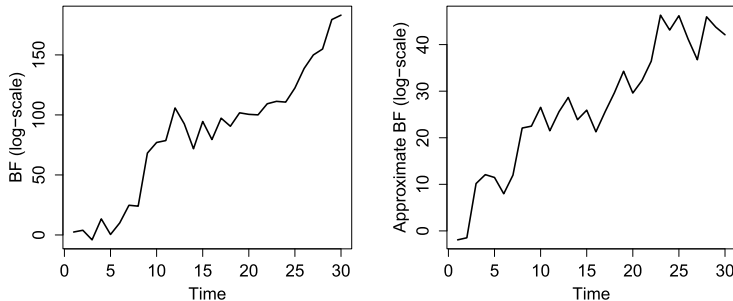


Figure 7: Left: Log Bayes factors (BF) in favour of univariate SV models with jump intensities modelled independently across time and stocks against univariate SV without jumps models; see (5.1) for the corresponding formula. Right: Logarithm of approximate BF for SV models with jointly (numerator) and independently (denominator) modelled jumps intensities; see (5.6) for the corresponding formula. The x-axis in both plots refers to the $\ell = 30$ out-of-sample time points.

credible intervals for the log-volatilities delivered by SV models with jumps (see Figure S.12 in the supplementary material) indicates that an efficient jump identification results in accurate estimation of the log-volatility process without increasing unnecessarily its level due to the presence of large price movements and, thus, more precise forecasts of future returns are expected. See also the following Section where we compare the predictive performance of SV models with and without jumps. To assess the jump structures estimated by employing independent and joint models for the jump intensities of SV models we rely on their forecasting performance. In particular, as also described below, the proposed joint modelling approach for the jump intensities results in a more accurate prediction of future returns. We conclude, therefore, that the estimation of the underlying jumps process depicted by the third panel of Figure 1 is closer to the real jump activity of the stock prices allowing further improvement in the estimation of the log-volatilities and more accurate prediction of future stock returns.

Predictive performance

Here we evaluate the predictive performance of the proposed SV model with jointly modelled jump intensities as well as the corresponding performance of univariate SV models with and without jumps which are also specified by equations (3.1)–(3.3) by making suitable assumptions. In the supplementary material we compare the forecasting performance of the proposed model with alternative SV models which account for the heavy tails of the distribution of the observed returns.

The left panel of Figure 7 compares the predictive performance of independent, univariate SV models with and without jumps by presenting predictive log Bayes factors. The Figure is constructed by considering the log Bayes factors defined in (5.1) 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

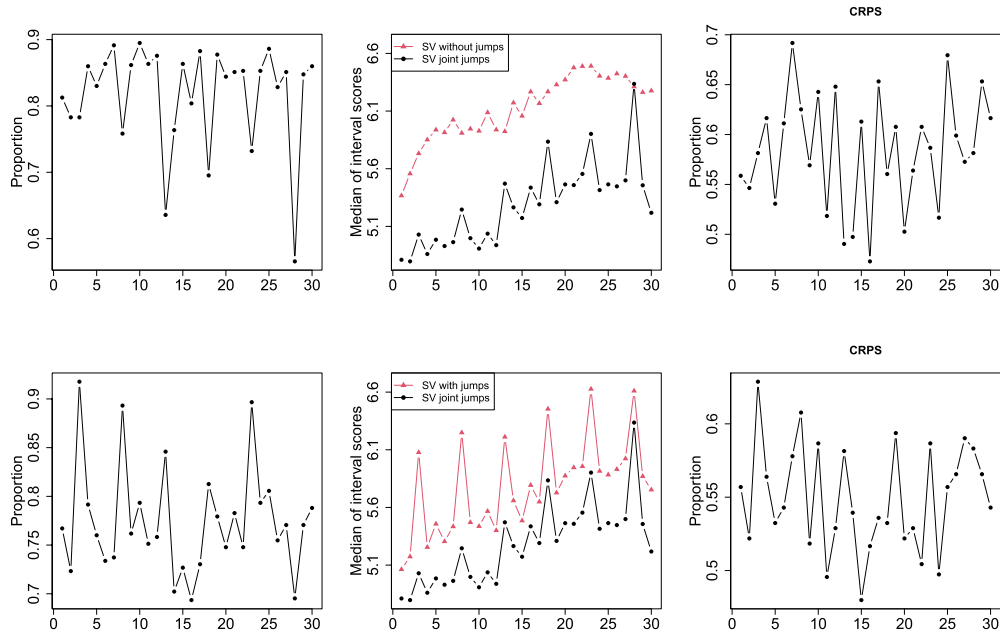


Figure 8: First column: proportion of stocks for which the interval scores for the 95% prediction intervals of the SV model with jointly modelled jumps is lower than the corresponding score for the SV models without (top) and with independent (bottom) jumps. Second column: medians of interval scores across the 571 stocks. Third column: proportion of stocks for which the continuous ranked probability score (CRPS) for the SV model with jointly modelled jumps is lower than the corresponding score for the SV models without (top) and with independent (bottom) jumps. The x-axis in all plots refers to the $\ell = 30$ out-of-sample time points.

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 the right panel of 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 compares the two models by presenting the logarithms of the approximate log Bayes factors in (5.6). Clearly, there is strong evidence that our proposed model outperforms the independent modelling of stock returns with a SV with jumps formulation.

Figure 8 compares the forecasts of the proposed model with those obtained from SV models without jumps and with independent jumps respectively by employing their probabilistic evaluation (Gneiting and Raftery, 2007). Joint modelling of the intensities

leads to 95% prediction credible intervals with the lowest interval scores at each one of the out-of-sample points for the vast majority of the stocks. The definition of the interval score in (5.2) rewards the forecaster that delivers narrow prediction intervals and when the observation is outside the predicted interval a penalty proportional to the significance level of the interval is incurred. It is thus indicated that the proposed SV model delivers the most accurate prediction intervals compared to those obtained from univariate SV models with and without jumps. The fact that Figure 8 displays the median instead of the mean of the 571 interval scores emphasizes an important characteristic of our modelling approach. The efficient identification of the in-sample jumps has resulted in estimating lower volatilities for the stock returns; see Figures S.11 and S.12 in the supplementary material. This is depicted in the conducted out-of-sample exercise by more narrow prediction intervals for a given coverage probability. Moreover, since the prediction of future jumps is based on common across the stocks, autoregressive factors, we expect that jumps that occur suddenly in a small number of stocks at a given day are not easily predictable. On the other hand, the large volatility intervals delivered from SV models without jumps or models with less accurate jump identification can accidentally predict more efficiently these types of movements. The median is chosen to avoid taking into account such predictions. The superiority of the proposed approach is also emphasized by Figure 8 by the lower continuous ranked probability scores that correspond to the proposed SV model than those of the univariate SV models. This implies that the posterior predictive distributions of the proposed model are more concentrated than those obtained from univariate SV models with and without jumps while deliver predictions that are not easily distinguishable from the materialized observations. This is a desired property of probabilistic forecasts described as the main goal of probabilistic forecasting; see for example Gneiting and Katzfuss (2014). Finally, Figure S.14 in the supplementary material which presents root mean squared errors for the predictions obtained from the different models indicates that joint modelling of the jump intensities improves slightly the point forecasts. We conclude that the proposed modelling approach offers a more accurate identification of jumps and consequently a more precise estimation of the returns volatility compared to the one conducted by using SV models without or with independent jumps. As a result, the predictive performance of the proposed model outperforms the other models. This makes the suggested model a competitive alternative for practitioners who are interested in obtaining accurate predictions for the stocks of large portfolios.

7 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 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 proposed MCMC algorithms can be easily extended in order to conduct inference for the more realistic (Harvey and Shephard, 1996) class of SV models which allow for correlation between the error terms ϵ_{it} and η_{it} in (3.1) and (3.2), a phenomenon which is commonly known as leverage effect. In this case samples from their full conditional distributions can be obtained by using the auxiliary gradient-based sampler (Titsias and Papaspiliopoulos, 2018) to draw each correlation parameter jointly with ϕ_i , $\sigma_{i\eta}^2$ and H_i . Other modelling aspects include the relaxing of independence of E_t in (3.8) or the assumption that A is lower diagonal as in the models proposed in Dellaportas and Pourahmadi (2012).

Finally, a series of interesting financial questions can be addressed with our models by exploiting the fact that panel of stock returns carries 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.

Supplementary Material

Web-based supporting material for “Bayesian prediction of jumps in large panels of time series data” by A. Alexopoulos, P. Dellaportas and O. Papaspiliopoulos (DOI: [10.1214/21-BA1268SUPP](https://doi.org/10.1214/21-BA1268SUPP); .pdf). The supplementary material contains proof of results provided in the main text, simulation experiments and additional results from the real data analysis.

References

- Aguilar, O. and West, M. (2000). “Bayesian dynamic factor models and portfolio allocation.” *Journal of Business & Economic Statistics*, 18(3): 338–357. [2](#), [3](#), [8](#)
- Aït-Sahalia, Y. (2004). “Disentangling diffusion from jumps.” *Journal of Financial Economics*, 74(3): 487–528. [2](#)
- Aït-Sahalia, Y., Cacho-Diaz, J., and Laeven, R. J. (2015). “Modeling financial contagion using mutually exciting jump processes.” *Journal of Financial Economics*, 117(3): 585–606. [1](#), [5](#), [27](#)
- 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. [20](#)
- Alexopoulos, A., Dellaportas, P., and Papaspiliopoulos, O. (2021). “Supplementary Ma-

- terial of “Bayesian Prediction of Jumps in Large Panels of Time Series Data”.
Bayesian Analysis. doi: <https://doi.org/10.1214/21-BA1268SUPP>. 2
- Andersen, T. G., Benzoni, L., and Lund, J. (2002). “An empirical investigation of continuous-time equity return models.” *The Journal of Finance*, 57(3): 1239–1284. 1, 5
- Asai, M., McAleer, M., and Yu, J. (2006). “Multivariate stochastic volatility: a review.” *Econometric Reviews*, 25(2-3): 145–175. MR2256285. doi: <https://doi.org/10.1080/07474930600713564>. 5
- Bandi, F. M. and Renò, R. (2016). “Price and volatility co-jumps.” *Journal of Financial Economics*, 119(1): 107–146. 5
- 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. 2, 5, 7
- Bhattacharya, A. and Dunson, D. B. (2011). “Sparse Bayesian infinite factor models.” *Biometrika*, 98(2): 291–306. MR2806429. doi: <https://doi.org/10.1093/biomet/asr013>. 3, 8
- Billio, M., Casarin, R., Ravazzolo, F., and van Dijk, H. K. (2012). “Combination schemes for turning point predictions.” *The Quarterly Review of Economics and Finance*, 52(4): 402–412. 6
- Bollerslev, T. (1986). “Generalized autoregressive conditional heteroskedasticity.” *Journal of Econometrics*, 31(3): 307–327. MR0853051. doi: [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1). 4
- Bollerslev, T. (1987). “A conditionally heteroskedastic time series model for speculative prices and rates of return.” *The Review of Economics and Statistics*, 542–547. 4
- Bollerslev, T., Law, T. H., and Tauchen, G. (2008). “Risk, jumps, and diversification.” *Journal of Econometrics*, 144(1): 234–256. MR2439931. doi: <https://doi.org/10.1016/j.jeconom.2008.01.006>. 5
- Buesing, L., Machado, T. A., Cunningham, J. P., and Paninski, L. (2014). “Clustered factor analysis of multineuronal spike data.” In *Advances in Neural Information Processing Systems*, 3500–3508. 2, 3
- Chan, J. C. and Eisenstat, E. (2018). “Bayesian model comparison for time-varying parameter VARs with stochastic volatility.” *Journal of Applied Econometrics*, 33(4): 509–532. MR3824524. doi: <https://doi.org/10.1002/jae.2617>. 5
- Chesney, M. and Scott, L. (1989). “Pricing European currency options: A comparison of the modified Black-Scholes model and a random variance model.” *Journal of Financial and Quantitative Analysis*, 267–284. 5
- Chib, S., Nardari, F., and Shephard, N. (2002). “Markov chain Monte Carlo methods for stochastic volatility models.” *Journal of Econometrics*, 108(2): 281–316. MR1894758. doi: [https://doi.org/10.1016/S0304-4076\(01\)00137-3](https://doi.org/10.1016/S0304-4076(01)00137-3). 1, 4, 5, 7, 8, 15, 16
- Chib, S., Nardari, F., and Shephard, N. (2006). “Analysis of high dimensional mul-

- tivariate stochastic volatility models.” *Journal of Econometrics*, 134(2): 341–371. MR2328413. doi: <https://doi.org/10.1016/j.jeconom.2005.06.026>. 5
- Chopin, N., Jacob, P. E., and Papaspiliopoulos, O. (2013). “SMC²: an efficient algorithm for sequential analysis of state space models.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(3): 397–426. MR3065473. doi: <https://doi.org/10.1111/j.1467-9868.2012.01046.x>. 17
- Chopin, N. and Papaspiliopoulos, O. (2020a). *An introduction to sequential Monte Carlo*, chapter 10, 219–165. Springer. MR4215639. 4, 18
- Chopin, N. and Papaspiliopoulos, O. (2020b). *An introduction to sequential Monte Carlo*. Springer. MR4215639. 17
- Clements, A. and Liao, Y. (2017). “Forecasting the variance of stock index returns using jumps and cojumps.” *International Journal of Forecasting*, 33(3): 729–742. 2, 5
- Cotter, S. L., Roberts, G. O., Stuart, A. M., White, D., et al. (2013). “MCMC methods for functions: modifying old algorithms to make them faster.” *Statistical Science*, 28(3): 424–446. MR3135540. doi: <https://doi.org/10.1214/13-STS421>. 12
- 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. MR0092301. 9
- Creal, D., Koopman, S. J., and Lucas, A. (2011). “A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations.” *Journal of Business & Economic Statistics*, 29(4): 552–563. MR2879242. doi: <https://doi.org/10.1198/jbes.2011.10070>. 5
- Delatola, E.-I., Griffin, J. E., et al. (2011). “Bayesian nonparametric modelling of the return distribution with stochastic volatility.” *Bayesian Analysis*, 6(4): 901–926. MR2869968. doi: <https://doi.org/10.2139/ssrn.1664423>. 5
- Dellaportas, P., Plataniotis, A., and Titsias, M. (2015). “Scalable inference for a full multivariate stochastic volatility model.” *arXiv preprint arXiv:1510.05257*. 5
- Dellaportas, P. and Pourahmadi, M. (2012). “Cholesky-GARCH models with applications to finance.” *Statistics and Computing*, 22(4): 849–855. MR2913787. doi: <https://doi.org/10.1007/s11222-011-9251-2>. 27
- Devroye, L. (1986). *Non-uniform random variate generation*. Springer-Verlag. URL http://books.google.gr/books?id=mEw_AQAAIAAJ MR0836973. doi: <https://doi.org/10.1007/978-1-4613-8643-8>. 10
- Engle, R. F. (1982). “Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation.” *Econometrica: Journal of the Econometric Society*, 987–1007. MR0666121. doi: <https://doi.org/10.2307/1912773>. 4
- Eraker, B., Johannes, M., and Polson, N. (2003). “The impact of jumps in volatility and returns.” *The Journal of Finance*, 58(3): 1269–1300. 3, 5, 7, 8, 16
- Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009). “Poisson autoregression.” *Jour-*

- nal of the American Statistical Association*, 104(488): 1430–1439. MR2596998. doi: <https://doi.org/10.1198/jasa.2009.tm08270>. 3
- Geweke, J. and Amisano, G. (2010). “Comparing and evaluating Bayesian predictive distributions of asset returns.” *International Journal of Forecasting*, 26(2): 216–230. 4, 17
- Ghosh, J. and Dunson, D. B. (2009). “Default prior distributions and efficient posterior computation in Bayesian factor analysis.” *Journal of Computational and Graphical Statistics*, 18(2): 306–320. MR2749834. doi: <https://doi.org/10.1198/jcgs.2009.07145>. 3
- Girolami, M. and Calderhead, B. (2011). “Riemann manifold Langevin and hamiltonian Monte Carlo methods.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2): 123–214. MR2814492. doi: <https://doi.org/10.1111/j.1467-9868.2010.00765.x>. 12
- Gneiting, T. and Katzfuss, M. (2014). “Probabilistic forecasting.” *Annual Review of Statistics and Its Application*, 1: 125–151. 4, 26
- Gneiting, T. and Raftery, A. E. (2007). “Strictly proper scoring rules, prediction, and estimation.” *Journal of the American statistical Association*, 102(477): 359–378. MR2345548. doi: <https://doi.org/10.1198/016214506000001437>. 4, 16, 25
- Hamilton, J. D. (1989). “A new approach to the economic analysis of nonstationary time series and the business cycle.” *Econometrica: Journal of the econometric society*, 57(3): 357–384. MR0996941. doi: <https://doi.org/10.2307/1912559>. 6
- Harvey, A., Ruiz, E., and Shephard, N. (1994). “Multivariate stochastic variance models.” *The Review of Economic Studies*, 61(2): 247–264. 1, 5
- Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, volume 52. Cambridge University Press. MR3135260. doi: <https://doi.org/10.1017/CB09781139540933>. 5
- Harvey, A. C. and Chakravarty, T. (2008). “Beta-t(e) garch.” 4
- Harvey, A. C. and Shephard, N. (1996). “Estimation of an asymmetric stochastic volatility model for asset returns.” *Journal of Business & Economic Statistics*, 14(4): 429–434. 27
- Hawkes, A. G. (1971). “Spectra of some self-exciting and mutually exciting point processes.” *Biometrika*, 58(1): 83–90. MR0278410. doi: <https://doi.org/10.1093/biomet/58.1.83>. 5
- Honore, P. (1998). *Pitfalls in estimating jump-diffusion models*. Department of Finance, Faculty of Business Administration, Aarhus School of Business. 8
- Hull, J. and White, A. (1987). “The pricing of options on assets with stochastic volatilities.” *The journal of finance*, 42(2): 281–300. 5
- Jacod, J., Todorov, V., et al. (2009). “Testing for common arrivals of jumps for discretely

- observed multidimensional processes.” *The Annals of Statistics*, 37(4): 1792–1838. MR2533472. doi: <https://doi.org/10.1214/08-AOS624>. 5
- Jarquier, E., Polson, N., and Rossi, P. (1994). “Bayesian analysis of stochastic volatility models (with discussion).” *Journal Business of Econometrics Statistics*, 12 (4), 371–417. MR1940631. doi: <https://doi.org/10.1198/073500102753410408>. 1, 5
- Johannes, M. S., Polson, N. G., and Stroud, J. R. (2009). “Optimal filtering of jump diffusions: Extracting latent states from asset prices.” *Review of Financial Studies*, 22(7): 2759–2799. 4, 5, 16, 19
- Jung, R. C., Liesenfeld, R., and Richard, J.-F. (2011). “Dynamic factor models for multivariate count data: An application to stock-market trading activity.” *Journal of Business & Economic Statistics*, 29(1): 73–85. MR2789392. doi: <https://doi.org/10.1198/jbes.2009.08212>. 3
- Kass, R. E. and Raftery, A. E. (1995). “Bayes factors.” *Journal of the American Statistical Association*, 90(430): 773–795. MR3363402. doi: <https://doi.org/10.1080/01621459.1995.10476572>. 25
- Kastner, G. (2019). “Sparse Bayesian time-varying covariance estimation in many dimensions.” *Journal of Econometrics*, 210(1): 98–115. MR3944765. doi: <https://doi.org/10.1016/j.jeconom.2018.11.007>. 5
- Kastner, G. and Frühwirth-Schnatter, S. (2014). “Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models.” *Computational Statistics & Data Analysis*, 76: 408–423. MR3209449. doi: <https://doi.org/10.1016/j.csda.2013.01.002>. 1, 5, 7, 14
- Kastner, G., Frühwirth-Schnatter, S., and Lopes, H. F. (2017). “Efficient Bayesian inference for multivariate factor stochastic volatility models.” *Journal of Computational and Graphical Statistics*, 26(4): 905–917. MR3765354. doi: <https://doi.org/10.1080/10618600.2017.1322091>. 5
- Kastner, G. and Huber, F. (2020). “Sparse Bayesian vector autoregressions in huge dimensions.” *Journal of Forecasting*, 39(7): 1142–1165. MR4161021. doi: <https://doi.org/10.1002/for.2680>. 5
- Kim, S., Shephard, N., and Chib, S. (1998). “Stochastic volatility: likelihood inference and comparison with ARCH models.” *The Review of Economic Studies*, 65(3): 361–393. 5, 7, 15, 16
- Kirkpatrick, S., Gelatt, C. D., Vecchi, M. P., et al. (1983). “Optimization by simulated annealing.” *Science*, 220(4598): 671–680. MR0702485. doi: <https://doi.org/10.1126/science.220.4598.671>. 19
- Knorr-Held, L. and Rue, H. (2002). “On block updating in Markov random field models for disease mapping.” *Scandinavian Journal of Statistics*, 29(4): 597–614. MR1988414. doi: <https://doi.org/10.1111/1467-9469.00308>. 12
- Krüger, F., Lerch, S., Thorarinsdottir, T., and Gneiting, T. (2020). “Predictive inference based on Markov chain Monte Carlo output.” *International Statistical Review*. 17

- Merton, R. C. (1976). “Option pricing when underlying stock returns are discontinuous.” *Journal of Financial Economics*, 3(1-2): 125–144. 5
- Murray, I., Adams, R., and MacKay, D. (2010). “Elliptical slice sampling.” In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 541–548. 12
- Murray, I. and Adams, R. P. (2010). “Slice sampling covariance hyperparameters of latent Gaussian models.” In *Advances in neural information processing systems*, 1732–1740. 12
- Nakajima, J. and Omori, Y. (2009). “Leverage, heavy-tails and correlated jumps in stochastic volatility models.” *Computational Statistics & Data Analysis*, 53(6): 2335–2353. MR2665832. doi: <https://doi.org/10.1016/j.csda.2008.03.015>. 4, 5, 15, 16
- Neal, R. M. (2001). “Annealed importance sampling.” *Statistics and Computing*, 11(2): 125–139. MR1837132. doi: <https://doi.org/10.1023/A:1008923215028>. 4, 17, 19
- Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). “Stochastic volatility with leverage: Fast and efficient likelihood inference.” *Journal of Econometrics*, 140(2): 425–449. MR2408914. doi: <https://doi.org/10.1016/j.jeconom.2006.07.008>. 5, 7, 16
- Papaspiliopoulos, O., Roberts, G. O., and Sköld, M. (2003). “Non-centered parameterisations for hierarchical models and data augmentation.” *Bayesian Statistics*, 7: 307–326. MR2003180. 15
- Papaspiliopoulos, O., Roberts, G. O., and Sköld, M. (2007). “A general framework for the parametrization of hierarchical models.” *Statistical Science*, 22(1): 59–73. MR2408661. doi: <https://doi.org/10.1214/088342307000000014>. 12, 13, 14, 15
- Pillow, J. W. and Scott, J. G. (2012). “Fully Bayesian inference for neural models with negative-binomial spiking.” In *NIPS*, 1907–1915. 3
- Rapach, D. and Zhou, G. (2013). “Forecasting stock returns.” In *Handbook of economic forecasting*, volume 2, 328–383. Elsevier. 2
- Roberts, G. O. and Stramer, O. (2002). “Langevin diffusions and Metropolis-Hastings algorithms.” *Methodology and Computing in Applied Probability*, 4(4): 337–357. MR2002247. doi: <https://doi.org/10.1023/A:1023562417138>. 12
- Rousseeuw, P. J. and Croux, C. (1993). “Alternatives to the median absolute deviation.” *Journal of the American Statistical Association*, 88(424): 1273–1283. MR1245360. 2
- Shephard, N. and Kim, S. (1994). “Comment.” *Journal of Business & Economic Statistics*, 12(4): 406–410. 4
- Silva, R. S., Lopes, H. F., and Migon, H. S. (2006). “The extended generalized inverse Gaussian distribution for log-linear and stochastic volatility models.” *Brazilian Journal of Probability and Statistics*, 67–91. MR2340185. 5
- Simpson, M., Niemi, J., and Roy, V. (2017). “Interweaving Markov chain Monte Carlo

- strategies for efficient estimation of dynamic linear models.” *Journal of Computational and Graphical Statistics*, 26(1): 152–159. MR3610416. doi: <https://doi.org/10.1080/10618600.2015.1105748>. 14
- 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. 1, 5
- Titsias, M. K. and Papaspiliopoulos, O. (2018). “Auxiliary gradient-based sampling algorithms.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4): 749–767. MR3849342. doi: <https://doi.org/10.1111/rssb.12269>. 4, 12, 13, 14, 27
- Van Dyk, D. A. and Jiao, X. (2015). “Metropolis-Hastings within partially collapsed Gibbs samplers.” *Journal of Computational and Graphical Statistics*, 24(2): 301–327. MR3357383. doi: <https://doi.org/10.1080/10618600.2014.930041>. 15
- Vrontos, I. D., Dellaportas, P., and Politis, D. N. (2000). “Full Bayesian inference for GARCH and EGARCH models.” *Journal of Business & Economic Statistics*, 18(2): 187–198. 5
- Vrontos, I. D., Dellaportas, P., and Politis, D. N. (2003). “A full-factor multivariate GARCH model.” *The Econometrics Journal*, 6(2): 312–334. MR2028238. doi: <https://doi.org/10.1111/1368-423X.t01-1-00111>. 5
- Yu, Y. and Meng, X.-L. (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. MR2878987. doi: <https://doi.org/10.1198/jcgs.2011.203main>. 12, 13, 14, 15
- Zhang, B., Chan, J. C., and Cross, J. L. (2020). “Stochastic volatility models with ARMA innovations: An application to G7 inflation forecasts.” *International Journal of Forecasting*, 36(4): 1318–1328. 5

Acknowledgments

The authors are grateful for the comments and feedback from the anonymous associate editor and referees, which significantly helped to improve this work.