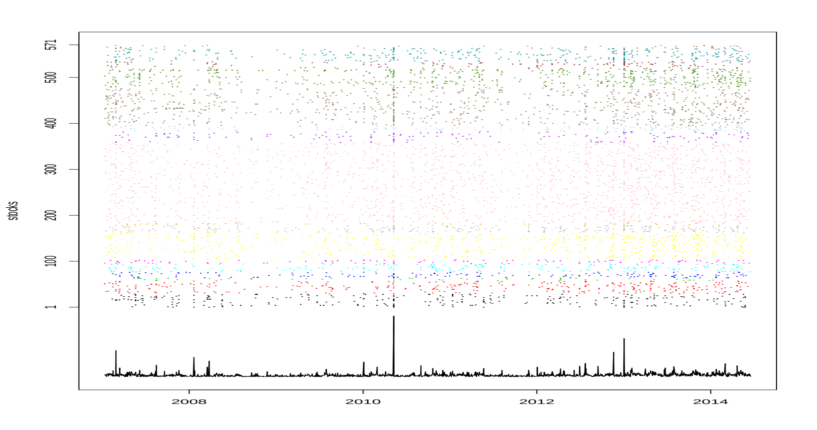
Identifying and predicting jumps in financial time series

Petros Dellaportas, UCL

with Alexopoulos (AUEB, Athens) and Papaspiliopoulos (UPF, Barcelona)

25/11/2016, Vienna

Jump identification from daily close prices for the 600 stocks of the EuroStoxx 600 index (January 8 of 2007 - November 5 of 2014).



Dots denote estimated probability of a jump > 0.5; Bottom graph depicts the total number of estimated jumps

Observe only returns y_t, everything else is unobserved:

$$y_t = exp(h_t/2)\epsilon_t + \sum_{j=1}^{N_t} \xi_j, \quad \epsilon_t \sim N(0,1)$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2)$$

- $N_t \sim Poisson(\Lambda_t \Delta_t)$, Δ_t is the distance between two successive observations
- $\Lambda_t \sim gamma(1, 50)$ (informative) so that the probability of no jump at time t is .98
- $\{\xi_j\}_{j=1}^{N_t} \stackrel{iid}{\sim} N(\mu_\xi, \sigma_\xi^2)$, $\mu_\xi \sim N(0, 5R^2)$, $\sigma_\xi^2 \sim IG(3, R^2/18)$ where R is the range of the data.
- $h_0 \sim N(\mu, \sigma_\eta^2/(1-\phi^2))$
- $\mu \sim N(0, 10)$
- $(\phi+1)/2 \sim Beta(20, 1.5)$ (informative)
- $\sigma_\eta^2 \sim \chi_1^2$, does not bound σ_η away from zero a priori, see Kastner and Frühwirth-Schnatter (2014) and Frühwirth-Schnatter and Wagner (2010).

Disentangling Volatility from Jumps

Bayesian inference with MCMC: how we sample the full conditional densities

- Sample simultaneously the vector of the log-volatility process
- Sample the number of jumps with rejection sampling
- Sample the parameters using interweaving; see Yu and Meng (2011) and Kastner and Frühwirth-Schnatter (2014).
- Contribution: we separate the volatility from the jump process without using any approximation of the model as it was proposed by Chib et al. (2002).

Sampling the volatility process

- Denote $Y = (y_1, y_2, \dots, y_T)$, $\theta = (\mu, \phi, \sigma_\eta^2, \mu_\xi, \sigma_\xi^2)$, $H = (h_1, \dots, h_T)$, $N = (N_1, \dots, N_T)$ and $\Xi = (\xi_1, \dots, \xi_T)$.
- Conditioning on number of jumps N and integrating out jump sizes Ξ, sample from

$$p(H|\theta, Y) \propto p(Y|H, \theta)p(H|\theta)$$

- Prior $p(H) = \mathcal{N}(H|M, Q^{-1})$
- There are many ways to do this -here we use an idea by Titsias that was first appeared in the discussion of the RSSB discussion paper by Girolami and Calderhead (2011).
- The advantage is that we sample the whole vector H as a block with one Metropolis move.

Sampling $p(H|\theta, Y) \propto p(Y|H,\theta)p(H|\theta)$

- Current state of H is H_n. Say we wish to use slice Gibbs:
- Introduce auxiliary variables U that live in the same space as H: $p(U|H_n) = \mathcal{N}(U|H_n + \frac{\delta}{2}\nabla \log p(Y|H_n), \frac{\delta}{2}I)$
- U injects Gaussian noise into U_n and shifts it by $(\delta/2)\nabla \log p(Y|H_n)$
- We cannot sample from $p(H|U)$ so we use a Metropolis step: Propose H^* from proposal q:

$$
\begin{aligned}
q(H^*|U) &= \frac{1}{\mathcal{Z}(U)}\mathcal{N}(H^*|U, \frac{\delta}{2}I)p(H^*) \\
&= \mathcal{N}(H^*|(I + \frac{\delta}{2}Q)^{-1}(U + \frac{\delta}{2}QM), \frac{\delta}{2}(I + \frac{\delta}{2}Q)^{-1}).
\end{aligned}
$$

where $\mathcal{Z}(U) = \int \mathcal{N}(H^*|U, \frac{\delta}{2}I)p(H^*)dH^*$.

- Accept H^* with Metropolis-Hastings probability $\min(1, r)$:

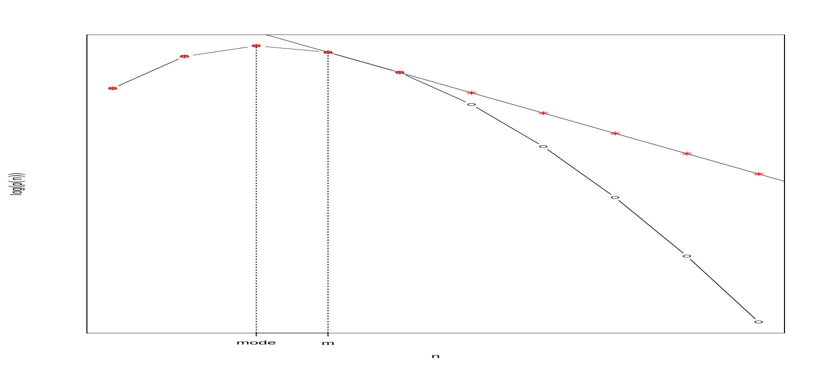
$$
\begin{aligned}
r &= \frac{p(Y|H^*)p(U|H^*)p(H^*)}{p(Y|H_n)p(U|H_n)p(H_n)} \frac{q(H_n|U)}{q(H^*|U)} \\
&= \frac{p(Y|H^*)p(U|H^*)p(H^*)}{p(Y|H_n)p(U|H_n)p(H_n)} \frac{\frac{1}{\mathcal{Z}(U)}\mathcal{N}(H_n|U, \frac{\delta}{2}I)p(H_n)}{\frac{1}{\mathcal{Z}(U)}\mathcal{N}(H^*|U, \frac{\delta}{2}I)p(H^*)} \\
&= \frac{p(Y|H^*)\mathcal{N}(U|H^* + \frac{\delta}{2}G_y, \frac{\delta}{2}I)}{p(Y|H_n)\mathcal{N}(U|H_n + \frac{\delta}{2}G_t, \frac{\delta}{2}I)} \frac{\mathcal{N}(H_n|U, \frac{\delta}{2}I)}{\mathcal{N}(H^*|U, \frac{\delta}{2}I)} \\
&= \frac{p(Y|H^*)}{p(Y|H_n)} \exp\left\{-(U - H_n)^T G_t + (U - H^*)^T G_y - \frac{\delta}{4}(||G_y||^2 - ||G_t||^2)\right\}
\end{aligned}
$$

  where $G_t = \nabla \log p(Y|H_n)$, $G_y = \nabla \log p(Y|H^*)$ and $||Z||$ denotes the Euclidean norm of a vector Z.

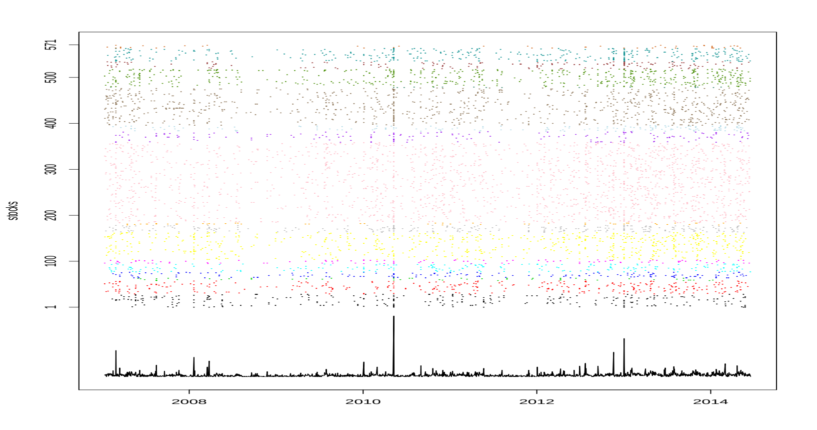
- The Gaussian prior terms $p(H_n)$ and $p(H^*)$ have been cancelled out from the acceptance probability, so their evaluation is not required: the resulting $q(H^*|U)$ is invariant under the Gaussian prior.
- Tune δ to achieve an acceptance rate of around $50 - 60\%$.

Rejection sampling for the number of jumps

- Target: the discrete log-concave distribution with density $p(N_t|h_t, \Lambda_t, y_t, \theta)$.
- Proposal: Choose as m any point after the mode, use as a proposal the red-dotted discrete density: it is a geometric after m.



Jump identification from daily close prices for the 600 stocks of the EuroStoxx 600 index (January 8 of 2007 - November 5 of 2014).



Dots denote estimated probability of a jump > 0.5; Bottom graph depicts the total number of estimated jumps

Jumps prediction

- Key idea: Use a Bayesian hierarchical model for all $i = 1, , \ldots, p = 600$ stocks to borrow strength

$$N_{it} \sim Poisson(\Lambda_{it}\Delta_{it})$$
$$\Lambda_{it} = \lambda \left(1 + \exp(-b_i - W_i F_t)\right)^{-1}$$

$\lambda = 0.15$ is the maximum intensity of each stock, W is a $p \times K$ matrix of factor loadings with rows W_i and each F_t are K-dimensional time-varying independent factors

$$F_t = AF_{t-1} + e_t, \quad t = 2, \ldots, T,$$

$e_t \sim N_K(0, I_K)$, $A = diag(\alpha_1, \ldots, \alpha_K)$.

The full model

$$y_{it} = exp(h_{it}/2)\epsilon_{it} + \sum_{j=1}^{N_{it}} \xi_{ij}, \quad \epsilon_{it} \sim N(0,1)$$

$$h_{it} = \mu_i + \phi_i(h_{i,t-1} - \mu_i) + \eta_{it}, \quad \eta_{it} \sim N(0, \sigma^2_{i\eta})$$

$$\{\xi_{ij}\}_{j=1}^{N_{it}} \overset{iid}{\sim} N(\mu_{i\xi}, \sigma^2_{i\xi}), \mu_{i\xi} \sim N(0, 5R_i^2), \sigma^2_{i\xi} \sim IG(3, R_i^2/18)$$

$$N_{it} \sim Poisson(\Lambda_{it}\Delta_{it})$$

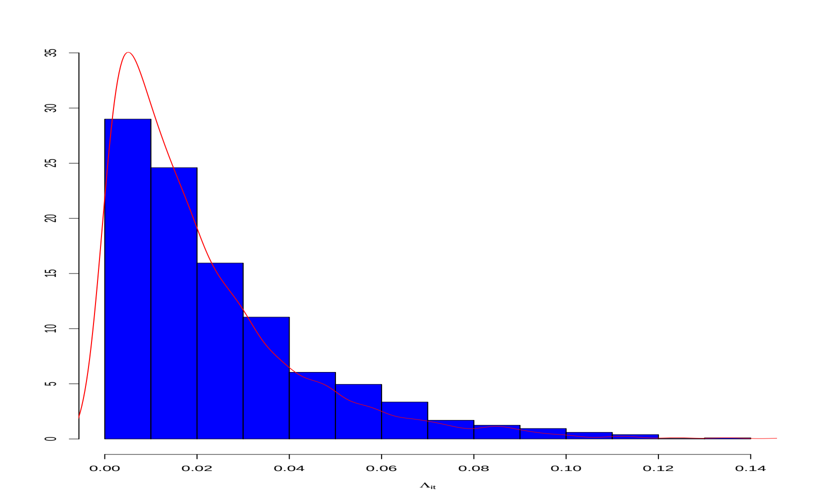
$$\Lambda_{it} = \lambda\left(1 + \exp(-b_i - W_i F_t)\right)^{-1}$$

$$F_t = AF_{t-1} + e_t$$

MCMC

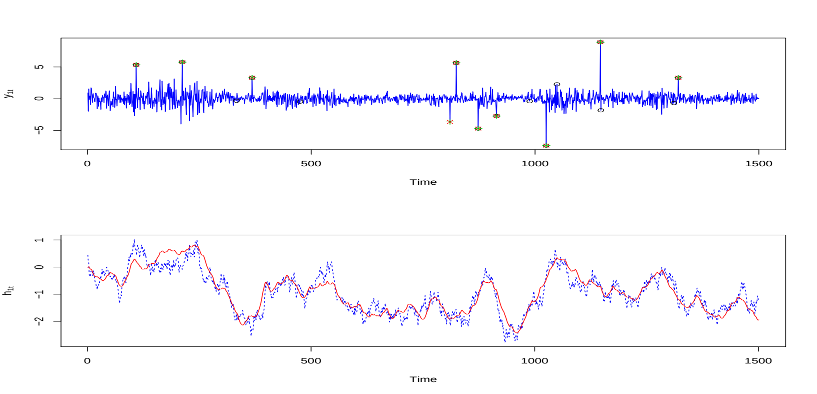
- We integrate out the jump sizes Ξ and we target the posterior distribution
$$p(\theta, h, N, F|Y)$$
where the parameters and the data correspond to all stocks.
- $p(h_i|N, F, \theta)$: Metropolis as in the 1-dim
- $p(N_i|H, F, \theta)$: rejection sampling as in the 1-dim
- $p(F|H, N, \theta)$: Label and sign switching are not taken care of during MCMC, see also Aßmann et al. (2016); we choose informative prior distributions for the parameters of the factor process such that 2 jumps are expected on average every 100 days. We sample all factors simultaneously based on the auxiliary Metropolis algorithm described for the 1-d case.
- $p(\theta|H, N, F)$: we use interweaving

We choose prior distributions for the parameters b, the factor loadings in the matrix W and for the persistent parameters of the matrix A such that the induced prior (histogram) for the intensity Λ_{it} of the ith stock at time t is comparable with the $Gamma(1, 50)$ prior (red line) used in the univariate model.



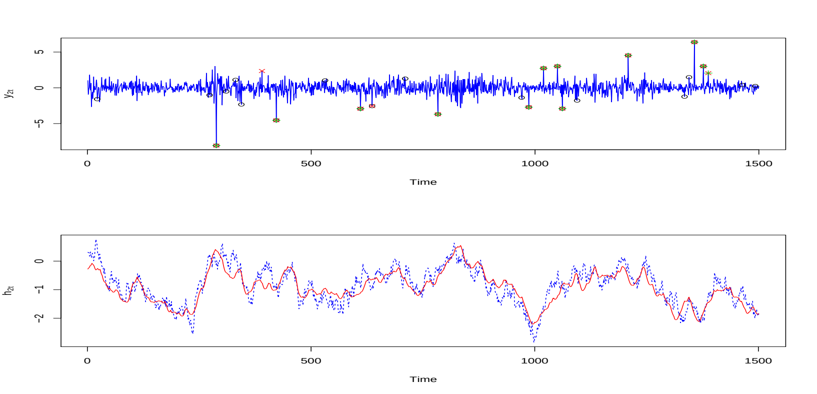
Separating volatility and jumps: simulation results

Simulated log returns (top) and their volatility path (bottom -blue) for 3 of the $p = 300$ time series of length $T = 1500$ with $K = 2$ factors. Bottom, Red: posterior mean of the volatility path. Black circles: simulated jumps. Red and green crosses: estimated probability of jump greater than 50% and 70%.



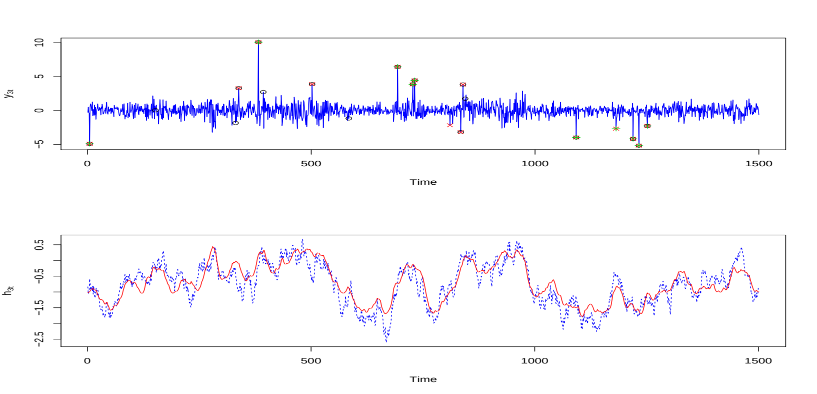
Separating volatility and jumps

Simulated log returns (top) and their volatility path (bottom -blue) for 3 of the $p = 300$ time series of length $T = 1500$ with $K = 2$ factors. Bottom, Red: posterior mean of the volatility path. Black circles: simulated jumps. Red and green crosses: estimated probability of jump greater than 50% and 70%.



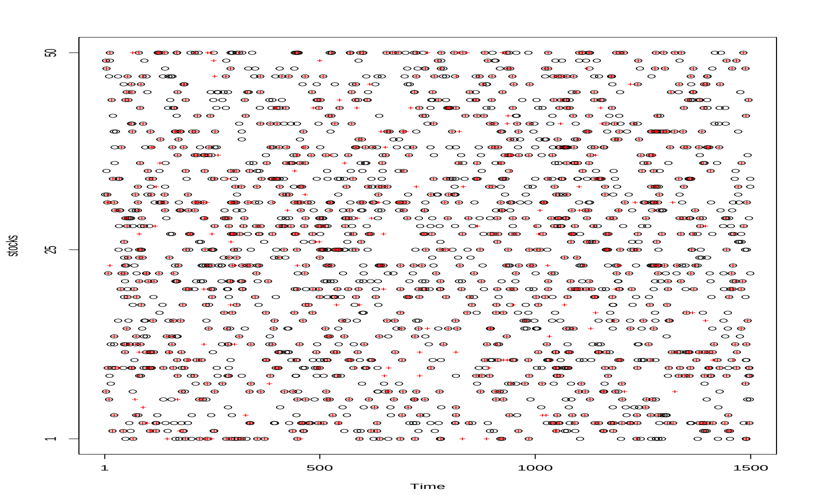
Separating volatility and jumps

Simulated log returns (top) and their volatility path (bottom -blue) for 3 of the $p = 300$ time series of length $T = 1500$ with $K = 2$ factors. Bottom, Red: posterior mean of the volatility path. Black circles: simulated jumps. Red and green crosses: estimated probability of jump greater than 50% and 70%.



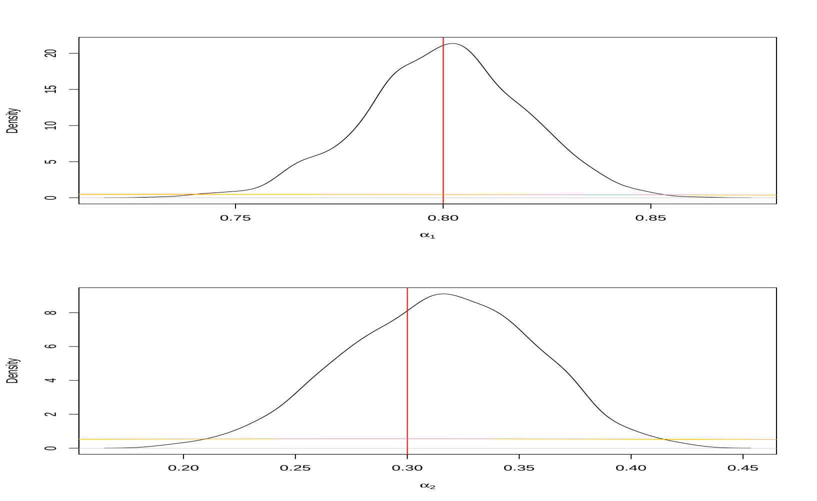
Jump identification

Black circles: simulated jumps for 50 of the $p = 300$ simulated times series. Red crosses: estimated probability of at least one jump greater than 50%.



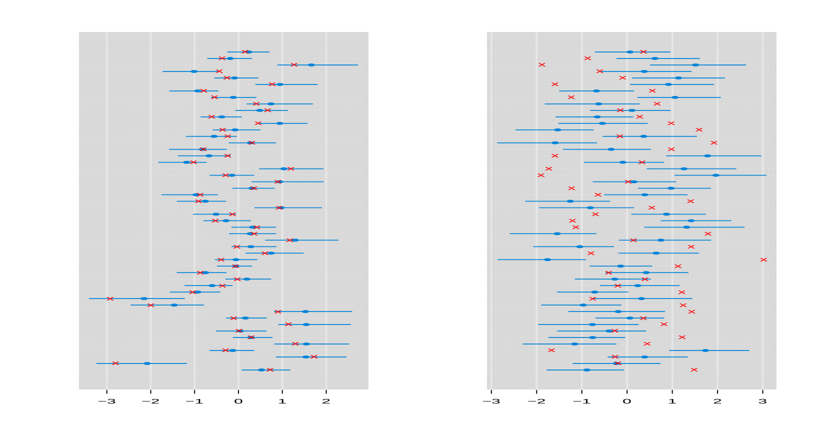
Posterior distributions I

Posterior (black lines) and prior (orange lines) distributions of the persistent parameters of the simulated latent factors. The vertical lines represent the simulated values.



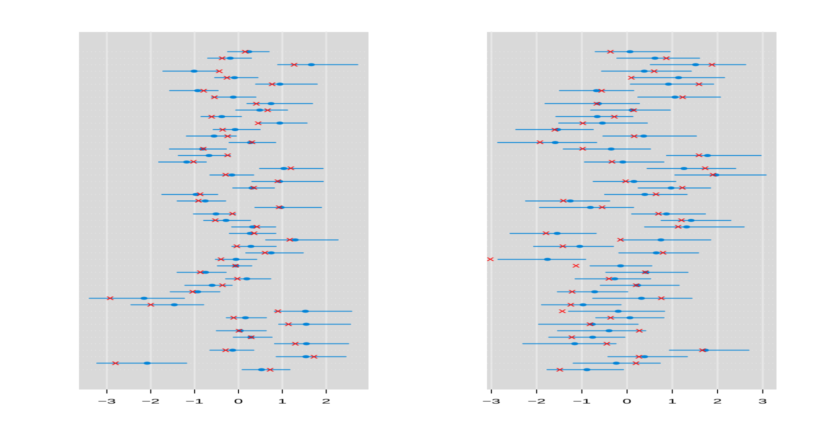
Posterior distributions for factor loadings (W)

95% credible intervals; red crosses indicate the simulated values.



Posterior distributions for factor loadings (W) after sign switching

95% credible intervals; red crosses indicate the simulated values.



Centered, Non Centered and Interweaving

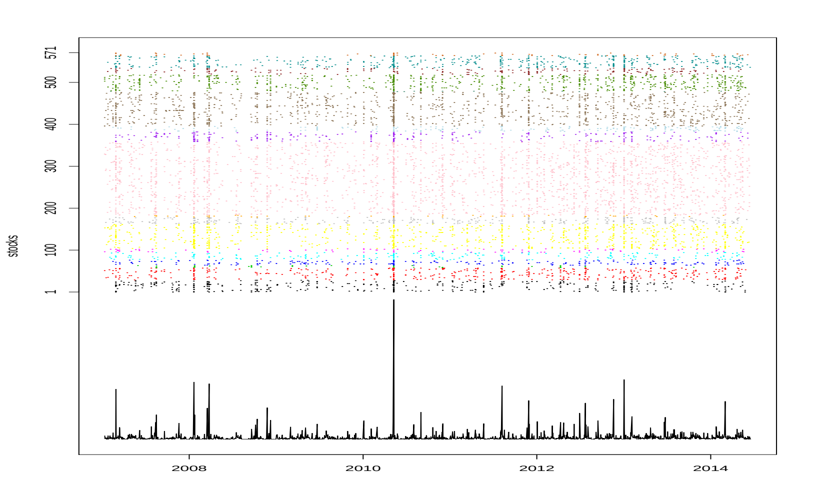
- Simulation of $p = 150$ series and $K = 2$ factors ($\lambda = 0.15$, $T = 1500$). The table belows *ESS* per unit of time for α_1,α_2 based on $M = 3000$ iterations (thinned by 10).
- $ESS = M/IF$ where $IF = \gamma_0/s^2$.
  - γ_0 estimated spectral density of the Markov chain at zero.
  - s^2 is the sample variance of the MCMC draws.

scheme	$\alpha_1 = 0.8$	$\alpha_2 = 0.8$
A : centered	50	70
B : Non centered	20	50
C : Interw. (Cent.-Non Cent.)	400	200
D : Interweaving (Non Cent.- Cent.)	150	60

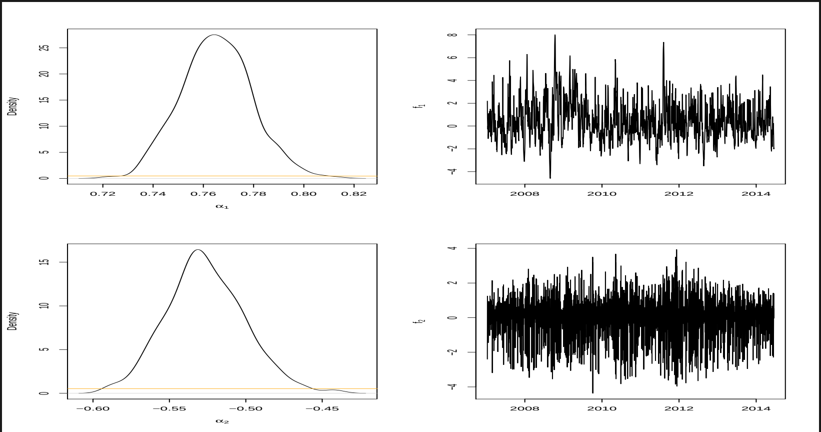
Results on real data

- Our dataset contains daily close prices for the 600 stocks of the EuroStoxx 600 index from the January 8 of 2007 until the November 5 of 2014.
- After removing stocks with less than 1500 prices and stocks with more than 10 consecutive zero returns we applied our MCMC algorithm on the log-returns of each one of the rest 571 stocks.
- We only present here results based on subset of this dataset -running in progress! We have used $K = 2$ factors in the intensity process.

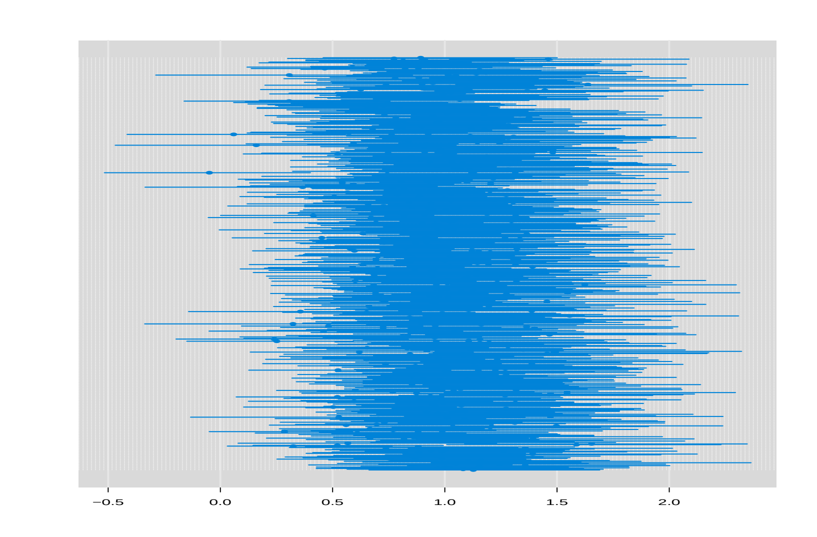
Identified jumps



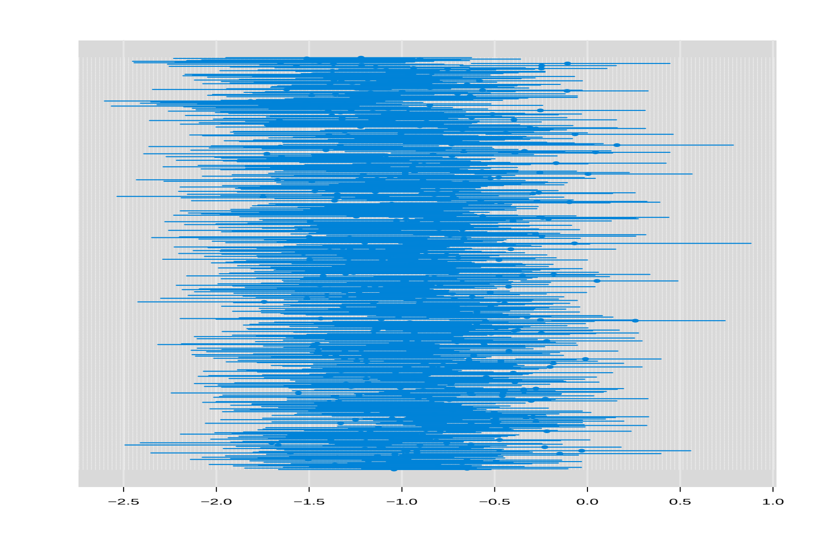
Posterior (black) and prior (orange) distributions for the persistent parameters of the $K = 2$ factors and posterior mean of the path of the two latent factors (right).



Factor 1: 95% credible intervals for each stock loading

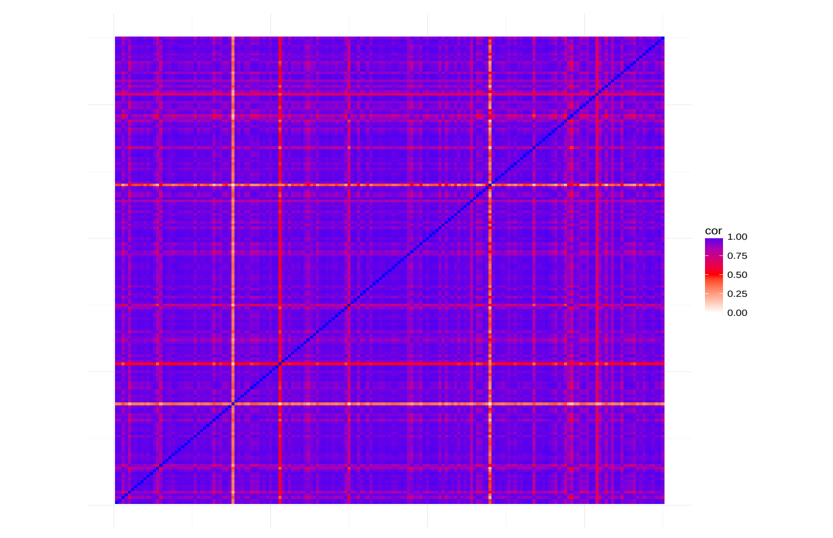


Factor 2: 95% credible intervals for each stock loading



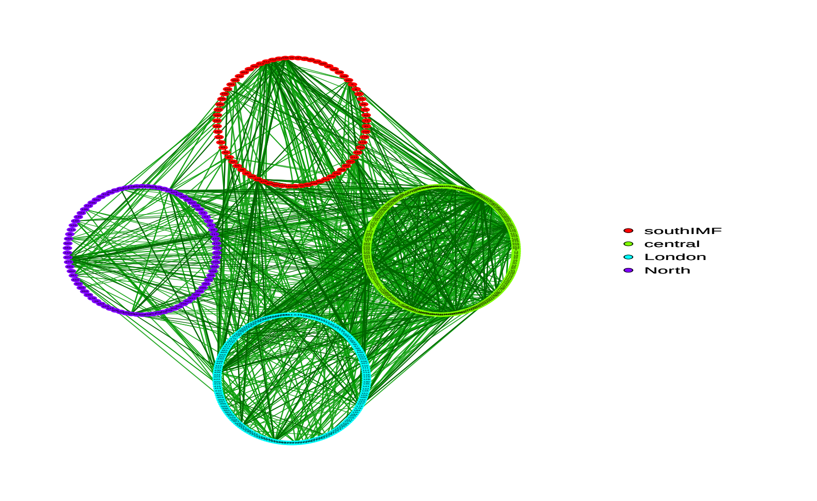
Posterior correlations

Posterior mean correlations of jump intensities: 175 stocks of the London stock exchange.



Network based on jump intensities

Nodes represent stocks, edges are present when the posterior mean correlation of jump intensities is > 0.9.



Model comparison

For two competing models $\mathbb{M}_1$ and $\mathbb{M}_2$ and out of sample observations $y_{T+1}, \ldots, y_{T+n}$ we compute the sequence of Bayes factors

$$\frac{p(y_{T+1}|y_{1:T}, \mathbb{M}_1)}{p(y_{T+1}|y_{1:T}, \mathbb{M}_2)}, \ldots, \frac{p(y_{T+1}, y_{T+2}, \ldots, y_{T+n}|y_{1:T}, \mathbb{M}_1)}{p(y_{T+1}, y_{T+2}, \ldots, y_{T+n}|y_{1:T}, \mathbb{M}_2)},$$

where for every model $\mathbb{M}$ and $j = 1, \ldots, n$

$$p(y_{T+1}, y_{T+2}, \ldots, y_{T+j}|y_{1:T}, \mathbb{M}) = \prod_{t=T+1}^{T+j} p(y_t|y_{1:t-1}, \mathbb{M}).$$

Problem

Assume θ is known and equal with the posterior mean. For each $t = T+1, \dots, T+n$, we need to compute the likelihood increments

$$p(y_t | y_{1:t-1}, \mathbb{M}) = \int p(h_{0:t-1}, F_{1:t-1} | y_{1:t-1}) p(y_t | F_t, h_t) p(h_t | h_{t-1}) p(F_t | F_{t-1}) dh_{0:t} dF_{1:t}$$

Typically the above integral is computed using SMC methods but since in our case the likelihood function is the product of the terms $p(y_{it} | F_t, h_{it})$ it contains a lot of information for the latent state F_t resulting in poor MC estimation of the integral (Beskos et al., 2014).

Solution

By observing that for each $t = T+1, \ldots, T+n$, the marginal likelihood increment $p(y_{it}|y_{1:t-1})$ of the ith stock is given by

$$\int p(h_{i(0:t-1)}, F_{1:t-1}|y_{1:t-1})p(y_{it}|F_t, h_{it})p(h_{it}|h_{i(t-1)})p(F_t|F_{t-1})dh_{i(0:t)}dF_{1:t}$$

we use annealed importance sampling (Neal, 2001) to obtain estimates of the marginal likelihoods

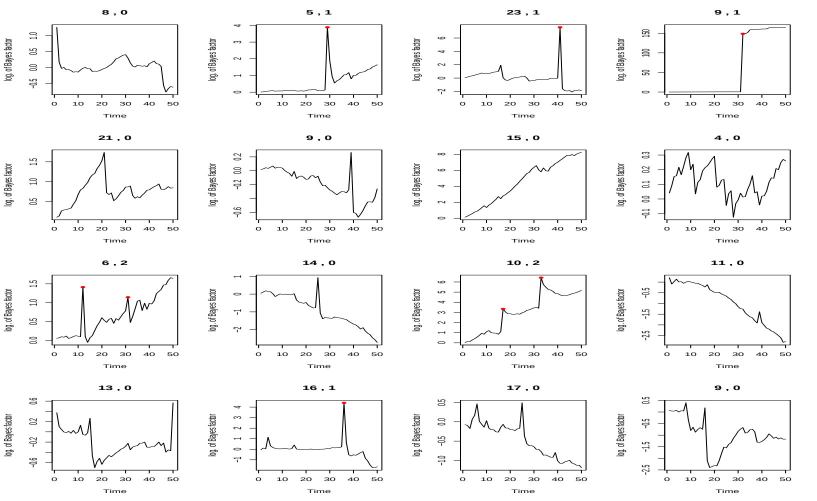
$$p(y_{i(T+1)}|y_{1:T}, \mathbb{M}), \ldots, p(y_{i(T+1)}, y_{i(T+2)}, \cdots, y_{i(T+n)}|y_{1:T}, \mathbb{M})$$

and we compute the predictive Bayes factors

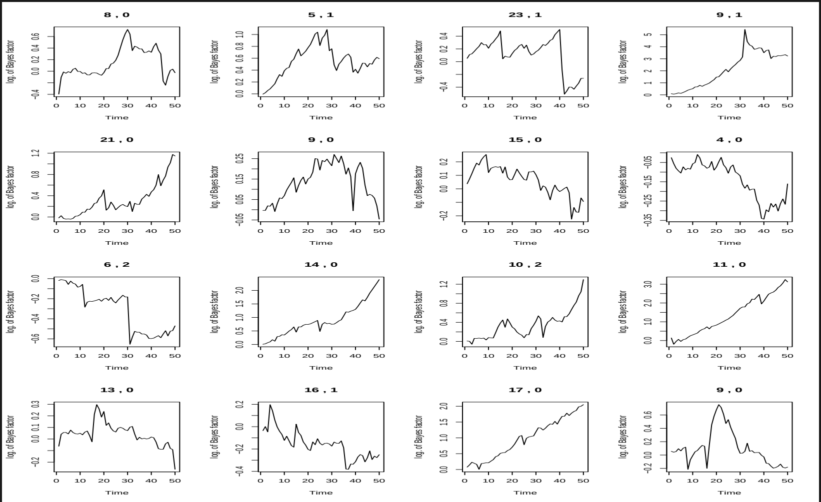
$$\frac{p(y_{i(T+1)}|y_{1:T}, \mathbb{M}_1)}{p(y_{i(T+1)}|y_{1:T}, \mathbb{M}_2)}, \ldots, \frac{p(y_{i(T+1)}, y_{i(T+2)}, \cdots, y_{i(T+n)}|y_{1:T}, \mathbb{M}_1)}{p(y_{i(T+1)}, y_{i(T+2)}, \cdots, y_{i(T+n)}|y_{1:T}, \mathbb{M}_2)},$$

to compare the competing models $\mathbb{M}_1$ and $\mathbb{M}_2$.

Log-Bayes factor of the univariate SV model with jumps against a univariate SV model without jumps for 16 stocks of London. Here $T = 1967$ and out of sample size is $n = 50$. For each stock we report estimated in-sample and out-of-sample (red circles) jumps.



Log-Bayes factor of the factor SV model with jumps against a SV model with jumps for 16 stocks of London. Here $T = 1967$ and out of sample size is $n = 50$. For each stock we report estimated in-sample and out-of-sample jumps.



Discussion

we have presented:

- A new algorithm for univariate SV with jumps
- A way to forecasting of jump intensities through joint modelling of many stocks
- Evidence for better predictive performance