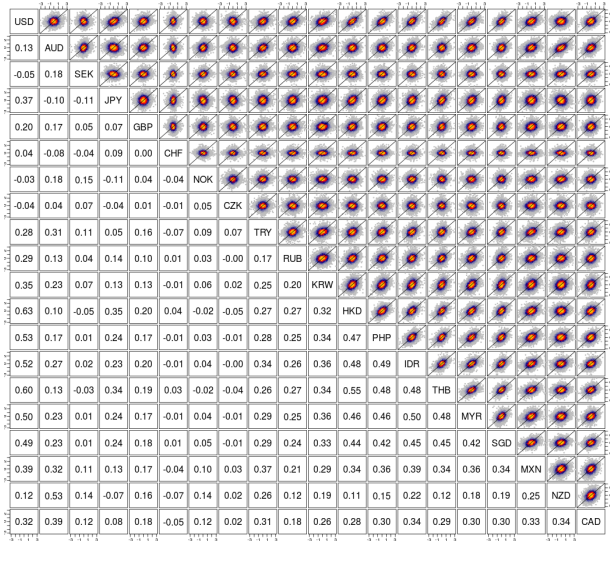
Sparse Bayesian Latent Factor Stochastic Volatility Models for Dynamic Covariance Estimation in High-Dimensional Financial Time Series



Gregor Kastner
WU Vienna | Austria

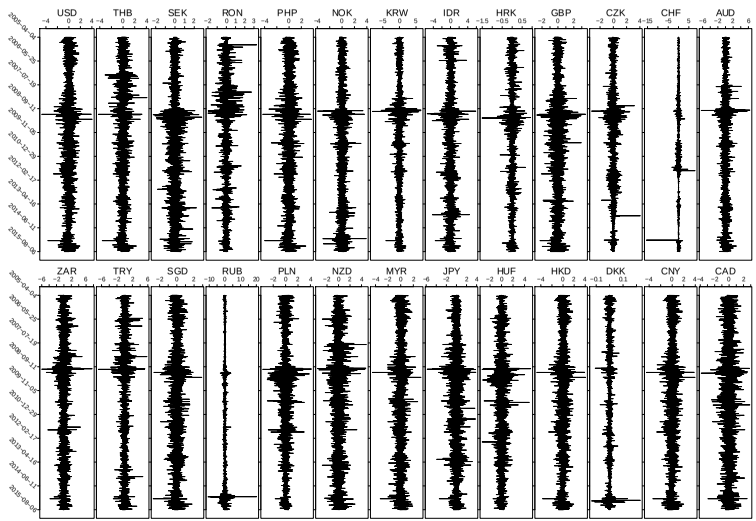
based on joint work with

Sylvia Frühwirth-Schnatter
WU Vienna | Austria

Hedibert Freitas Lopes
Insper São Paulo | Brazil

2015 NBER-NSF Time Series Conference | Vienna | September 26, 2015

1. Introduction
2. Factor Models
3. Application I
4. Sparse Factor Models
5. Application II
6. Estimation
7. Summary


USD
THB
SEK
RON
PHP
NOK
KRW
IDR
HRK
GBP
CZK
CHF
AUD
ZAR
TRY
SGD
RUB
PLN
NZD
MYR
JPY
HUF
HKD
DKK
CNY
CAD

The univariate SV model

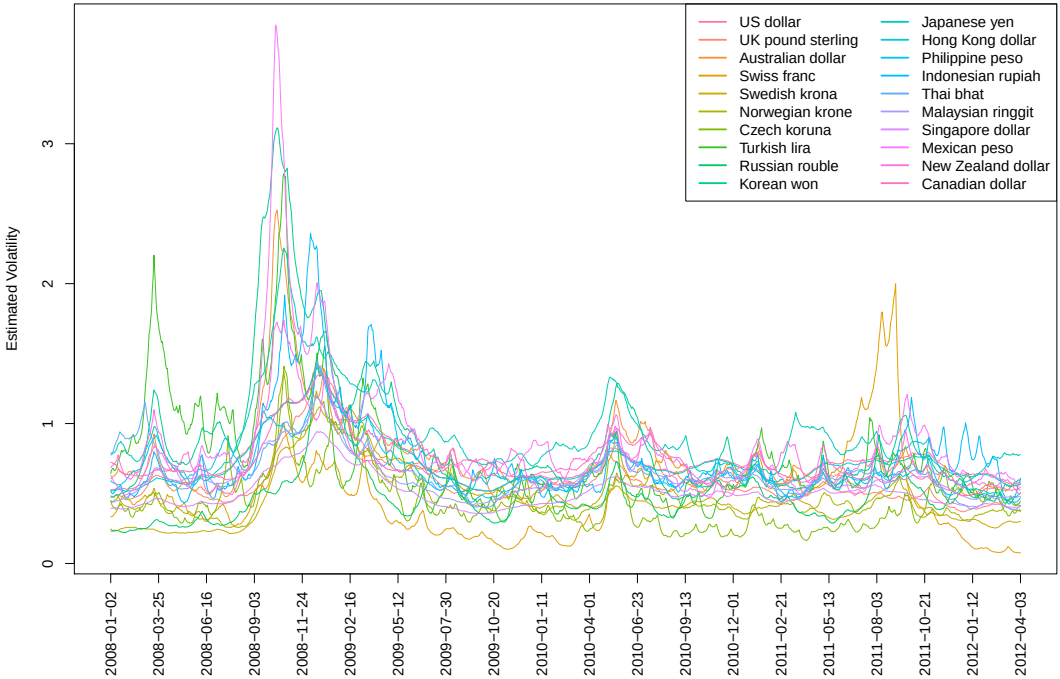
"Centered" version (e.g. Jacquier et al., 1994; Kim et al., 1998):

$$\begin{aligned} y_t &= e^{h_t/2}\varepsilon_t, & \varepsilon_t &\sim \mathcal{N}(0,1), \\ h_t &= \mu + \phi(h_{t-1} - \mu) + \sigma\eta_t, & \eta_t &\sim \mathcal{N}(0,1), \end{aligned}$$

with $\mathbf{h} = (h_0, \ldots, h_T)$, μ, ϕ and σ unknown. Features:

- have separate error terms for the mean and volatility equations (state space models).
- arise as discrete approximations to various diffusion processes in continuous-time asset pricing.
- shares many "nice" properties with (. . . -)GARCH . . .
- . . . while still showing important differences, e.g. a more realistic ACF of squared returns.
- shows less/no remaining nonlinear dependencies in residuals (Jacquier et al., 1994).

Individual SV models for daily EUR exchange rates (Kastner et al., 2014)



Factor analysis recap

Basic Static Factor Model: Common factors $\mathbf{f}_t = (f_{1t}, \ldots, f_{rt})'$ are related to $\mathbf{y}_t$ via

$$\mathbf{y}_t = \mathbf{\Lambda} \mathbf{f}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_m(\mathbf{0}, \mathbf{\Sigma}),$$

where $\mathbf{\Sigma} = \text{Diag}(\sigma_1^2, \ldots, \sigma_m^2)$ and $\mathbf{\Lambda}$ is the factor loading matrix.

Typically, orthogonal factors are used with $\mathbf{f}_t \sim \mathcal{N}_r(\mathbf{0}, \mathbf{1}_r)$. If furthermore $\mathbf{f}_t$ and $\boldsymbol{\varepsilon}_t$ are independent, we have

$$\mathbf{y}_t \sim \mathcal{N}_m(\mathbf{0}, \mathbf{\Omega}), \qquad \mathbf{\Omega} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Sigma}.$$

- Factor models yield good estimators of the covariance $\mathbf{\Omega}$ and (in particular) precision $\mathbf{\Omega}^{-1}$ (Fan et al., 2008).
- In econometrics, number of factors $r \ll m$ (dimension of $\mathbf{y}_t$) $\Rightarrow$ number of parameters grows linearly with m.
- For (financial) time series, static factor models need to be extended to settings where variances and covariances change over time $\Rightarrow$ combine factor models with SV models (Chib et al., 2006).

Factor SV

Use $m + r$ conditionally independent univariate SV models for h_{it}, $i = 1, \ldots, m + r$,

$$h_{it} = \mu_i + \phi_i(h_{i,t-1} - \mu_i) + \sigma_i \eta_{it}, \qquad \eta_{it} \sim \mathcal{N}(0, 1),$$

to define time-varying diagonal matrices

$$\mathbf{\Sigma}_t = \text{Diag}\left(e^{h_{1,t}}, \ldots, e^{h_{m,t}}\right), \quad \mathbf{V}_t = \text{Diag}\left(e^{h_{m+1,t}}, \ldots, e^{h_{m+r,t}}\right).$$

Model m-variate time series $\mathbf{y}_t$ through the factor model:

$$\begin{aligned} \mathbf{f}_t &\sim \mathcal{N}_r(\mathbf{0}, \mathbf{V}_t), \\ \mathbf{y}_t &= \mathbf{\Lambda} \mathbf{f}_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}_m(\mathbf{0}, \mathbf{\Sigma}_t). \end{aligned}$$

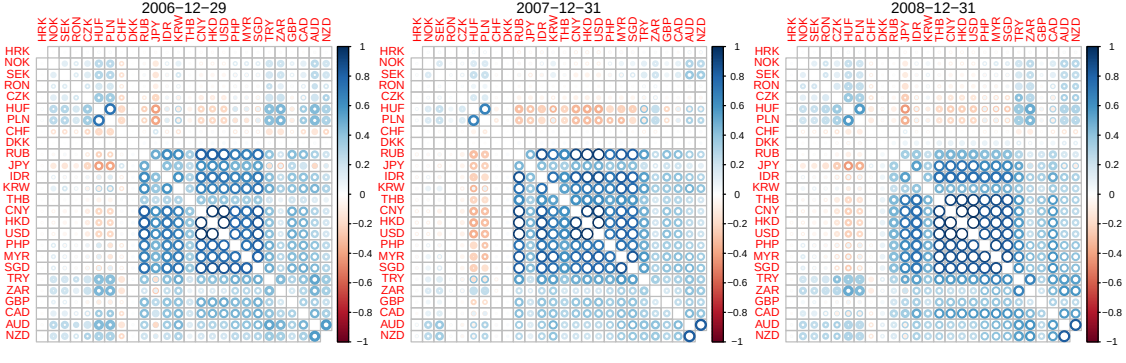
This yields a sparse representation of a time-varying covariance matrix $\mathbf{\Omega}_t$:

$$\mathbf{y}_t | \mathbf{h}_t \sim \mathcal{N}_m(\mathbf{0}, \mathbf{\Omega}_t(\mathbf{h}_t)), \quad \mathbf{\Omega}_t(\mathbf{h}_t) = \mathbf{\Lambda} \mathbf{V}_t(\mathbf{h}_t) \mathbf{\Lambda}' + \mathbf{\Sigma}_t(\mathbf{h}_t).$$

- Conditionally, $\mathbf{y}_t$ exhibits heteroskedasticity in many dimensions w/o their "curse".
- Marginally, $\mathbf{y}_t$ follows a process with a fat tailed stationary distribution.

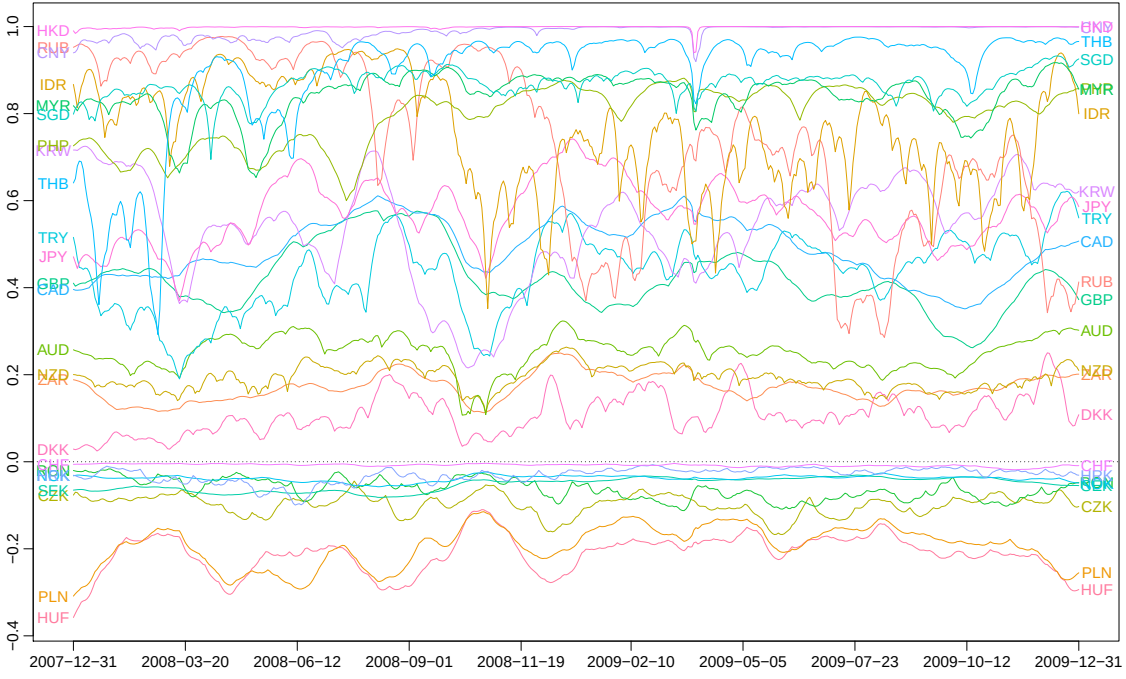
Application to exchange rate data

Some posterior correlation matrices (mean $\pm$ 2 $\times$ sd) Video

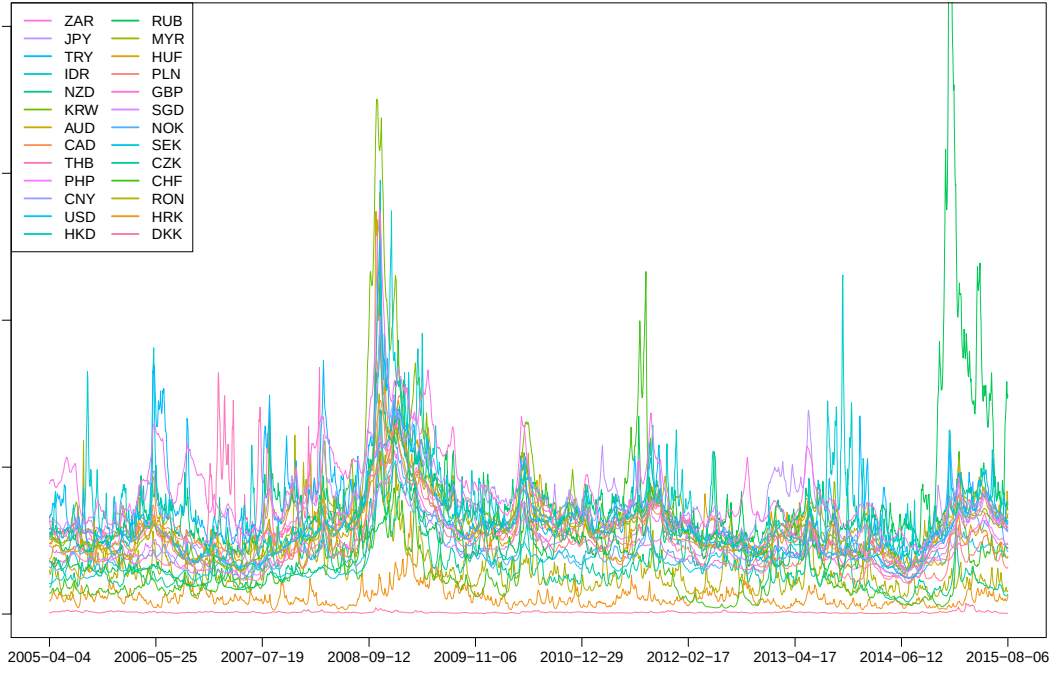


Quite some additional sparsity in terms of uncorrelated returns!

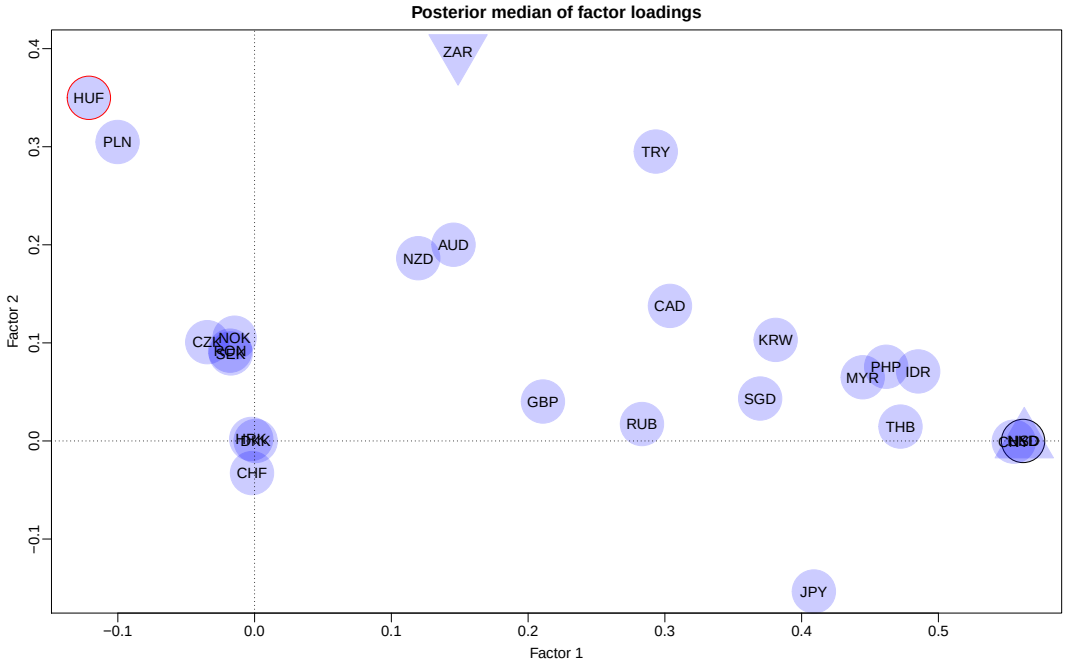
Mean posterior correlations with USD



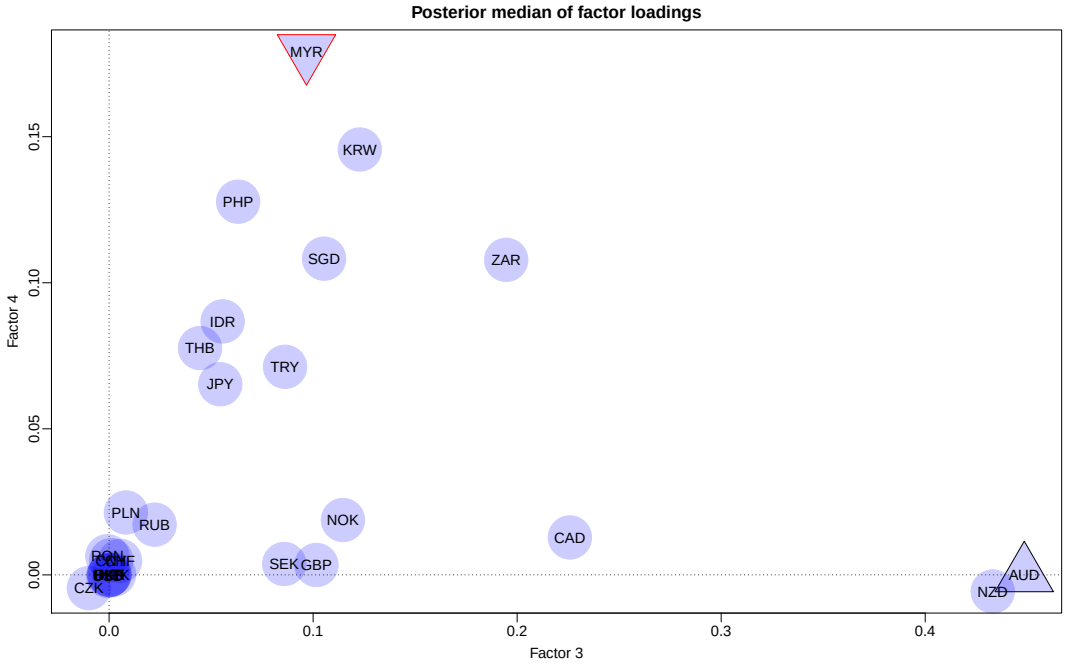
Mean posterior volatilities



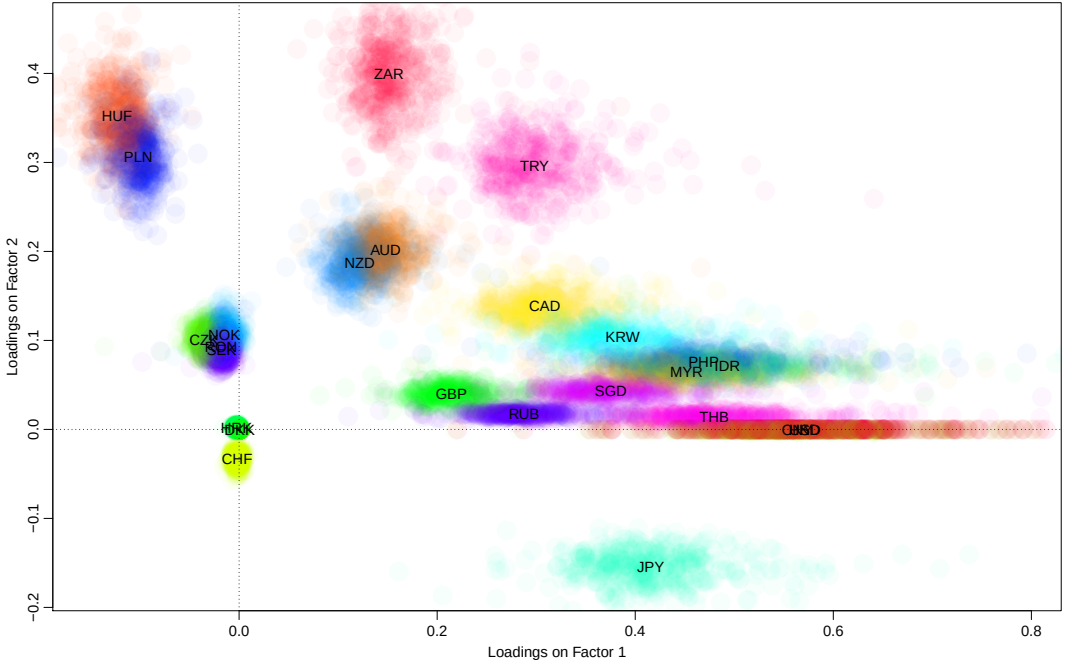
Median of loadings posterior $\mathbf{\Lambda}|\mathbf{y}$ (factors 1 and 2)



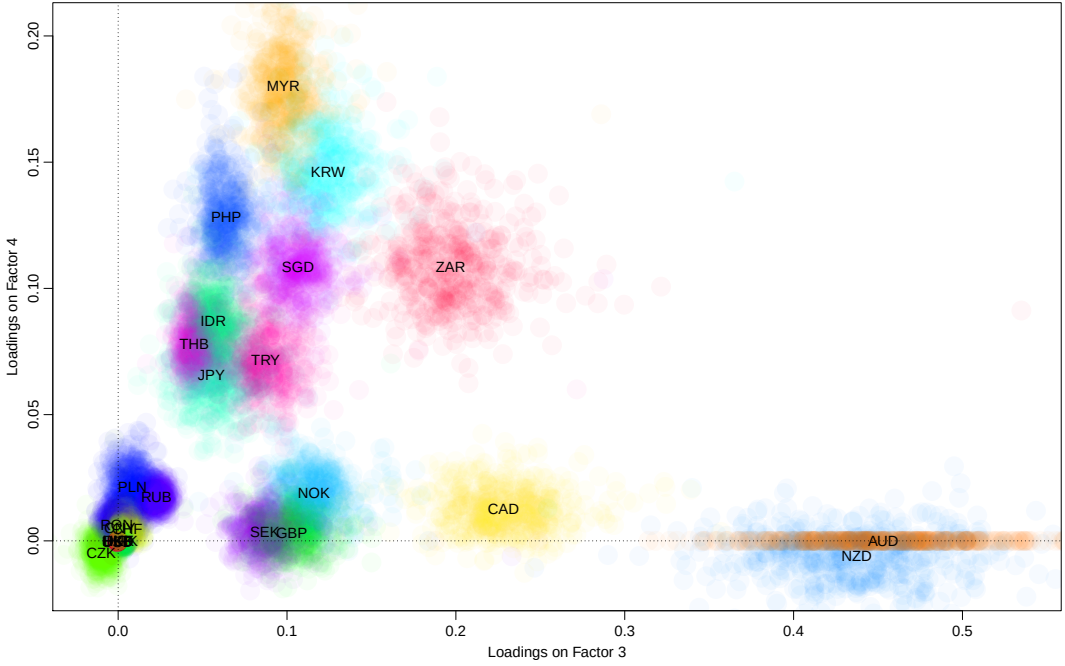
Median of loadings posterior $\mathbf{\Lambda}|\mathbf{y}$ (factors 3 and 4)



Posterior loadings distribution $\mathbf{\Lambda}|\mathbf{y}$ (factors 1 and 2)



Posterior loadings distribution $\mathbf{\Lambda}|\mathbf{y}$ (factors 3 and 4)



Sparse factor models

Factor models are a sparse representation of $\mathbf{\Omega}$ and $\mathbf{\Omega}^{-1}$. To achieve additional sparsity, use shrinkage priors a.k.a. penalized likelihood.

- Point mass priors
  - Basic factor model (West, 2003; Carvalho et al., 2008; Frühwirth-Schnatter and Lopes, 2015)
  - Bayesian dedicated factor analysis (Conti et al., 2014)
  - Sparse dynamic factor models (Kaufmann and Schumacher, 2013)
- Continuous shrinkage priors, e.g. in sparse Bayesian infinite factor models (Bhattacharya and Dunson, 2011)
- Latent thresholding approaches (Nakajima and West, 2013; Zhou et al., 2014)
- ...

Sparse SV factor models

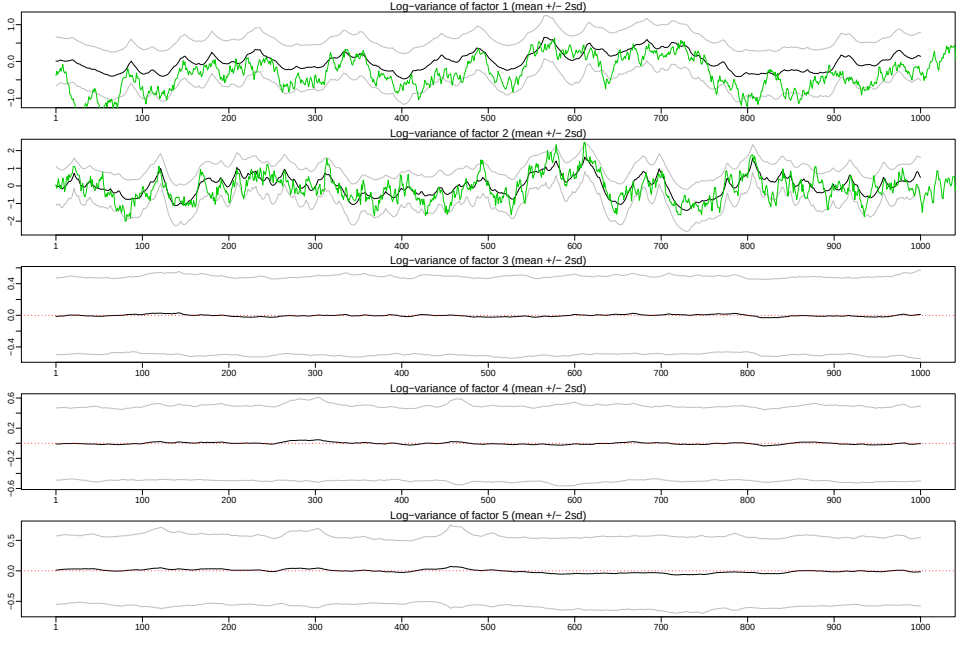
Use continuous shrinkage priors such as the Normal-Gamma prior (Griffin and Brown, 2010):

$$\Lambda_{ij}|\lambda_j, \tau_{ij} \sim \mathcal{N}\left(0, \tau_{ij}^2/\lambda_j^2\right), \quad \lambda_j^2 \sim \mathcal{G}(c, d), \quad \tau_{ij}^2 \sim \mathcal{G}(a, a).$$

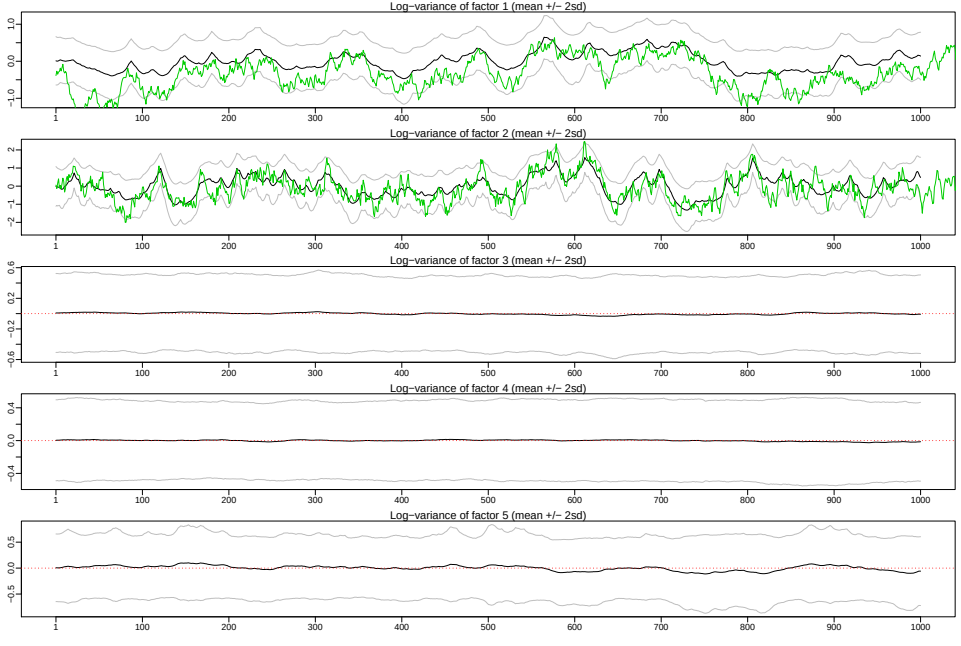
Visualization

- $\mathbb{V}(\Lambda_{ij}|\lambda_j^2) = 1/\lambda_j^2$, $\mathbb{V}(\Lambda_{ij}) = d/(c-1)$ if it exists
- Excess kurtosis of Λ_{ij} is $3/a$ if it exists
- Shrink globally (column-wise) through λ_j^2
- Adjust locally (element-wise) through τ_{ij}: $\tau_{ij} < 1$ more, $\tau_{ij} > 1$ less shrinkage
- Bayesian Lasso (Park and Casella, 2008) arises for $a = 1$

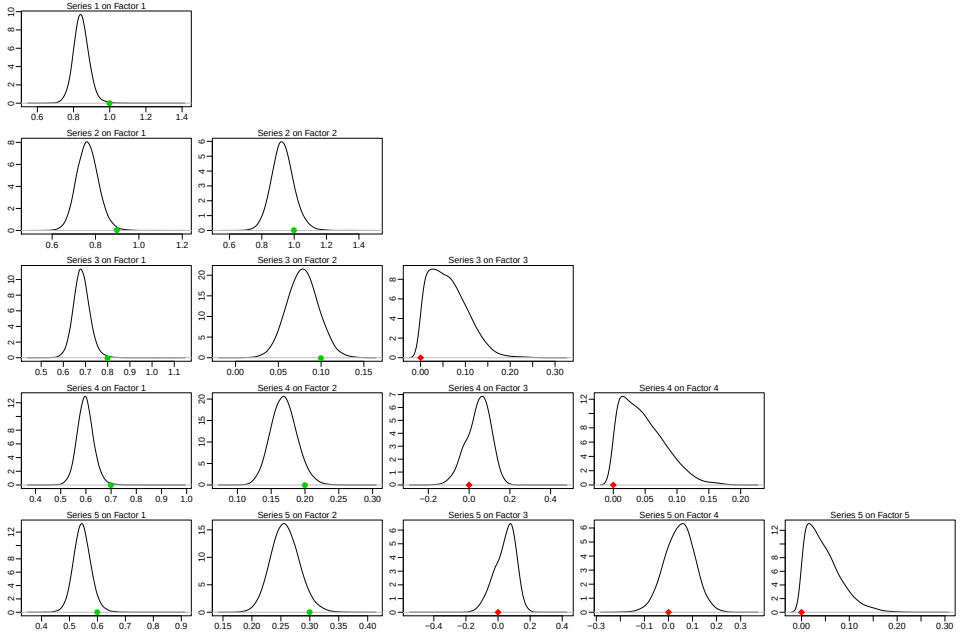
Simulation: $r_{\text{true}} = 2$, Normal prior ($m = 10,\ T = 1000$)



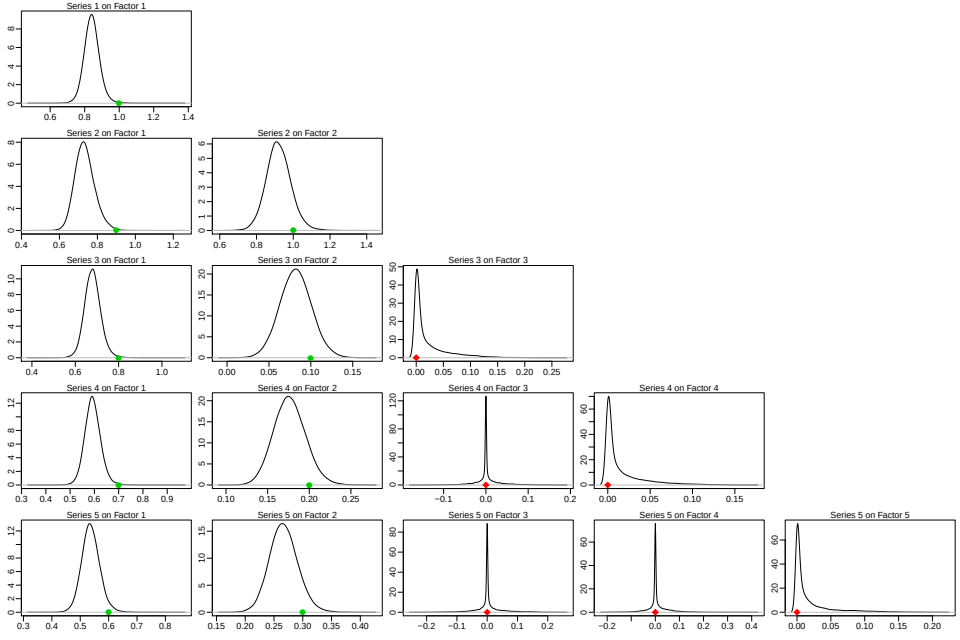
Simulation: $r_{\text{true}} = 2$, NG prior, $(m = 10,\ T = 1000)$



Simulation: $r_{\text{true}} = 2$, Normal prior ($m = 10,\ T = 1000$)



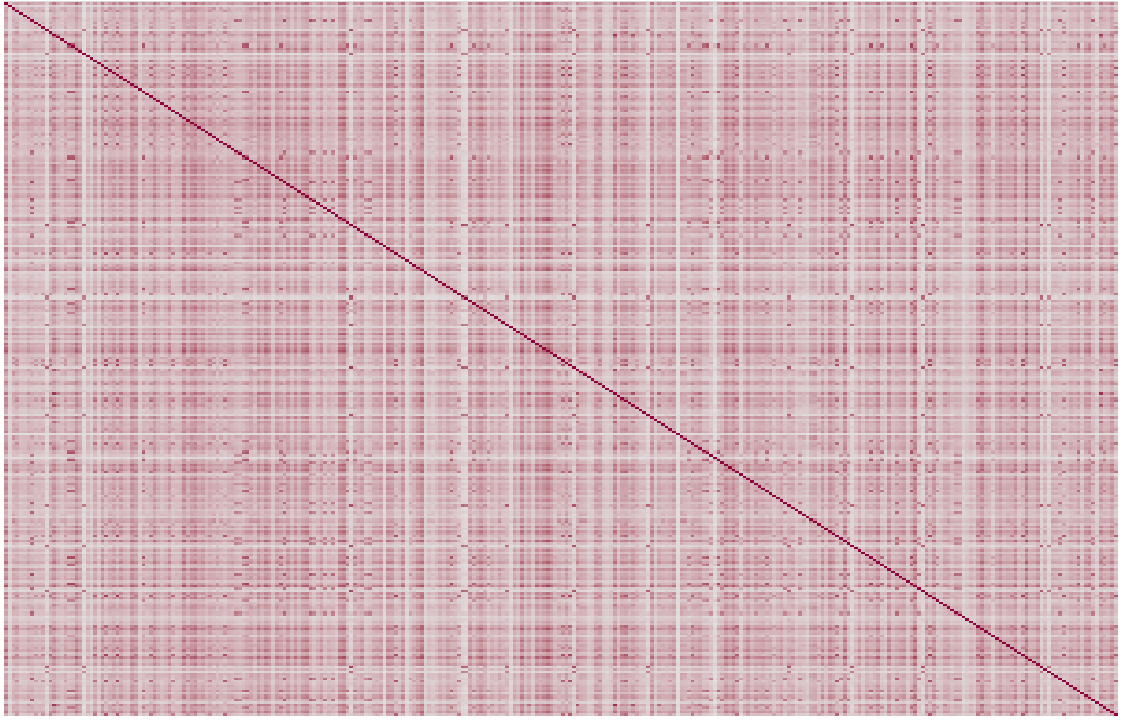
Simulation: $r_{\text{true}} = 2$, NG prior, $(m = 10,\ T = 1000)$



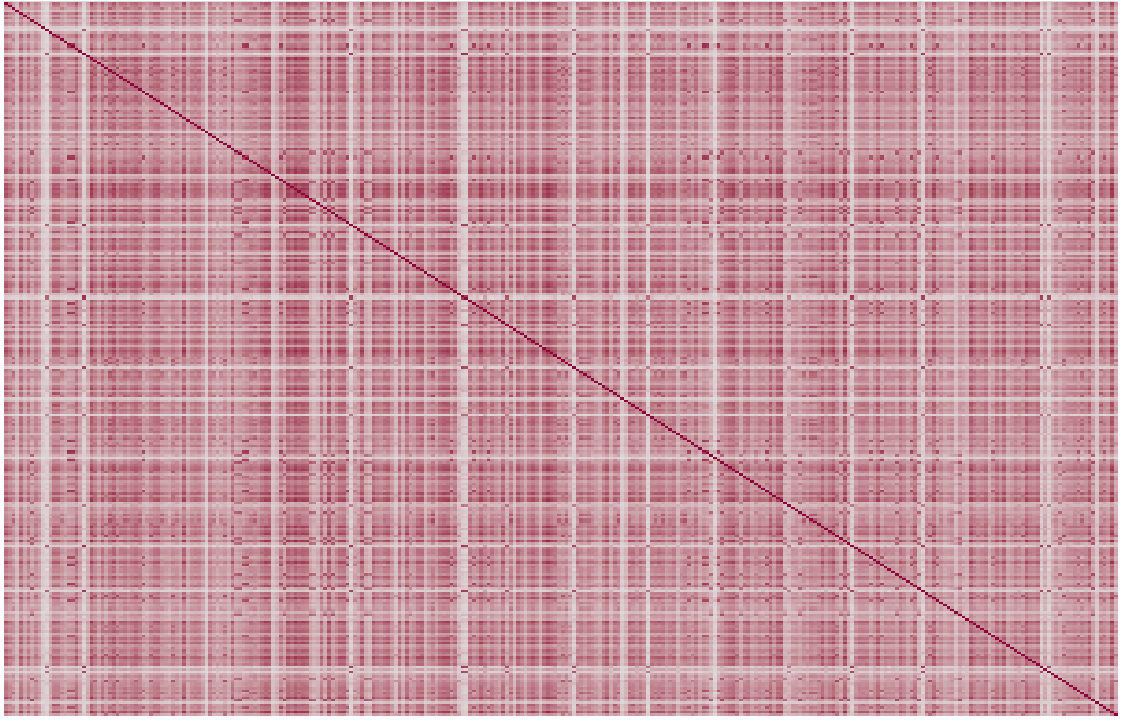
Application to S&P 500 members

- Seven-factor SV model to stock prices listed in the S&P 500 index.
- Only firms which have been listed from November 1994 onwards, resulting in $m = 300$ stock prices on 5001 days, ranging from 11/1/1994 to 12/31/2013.
- Data was obtained from Bloomberg Terminal in January 2014.
- Investigate $T = 5000$ demeaned percentage log-returns.
- Time-varying covariance matrix with 45150 elements on 5000 days can be well explained by 7 factors with many factor loadings shrunken to 0.

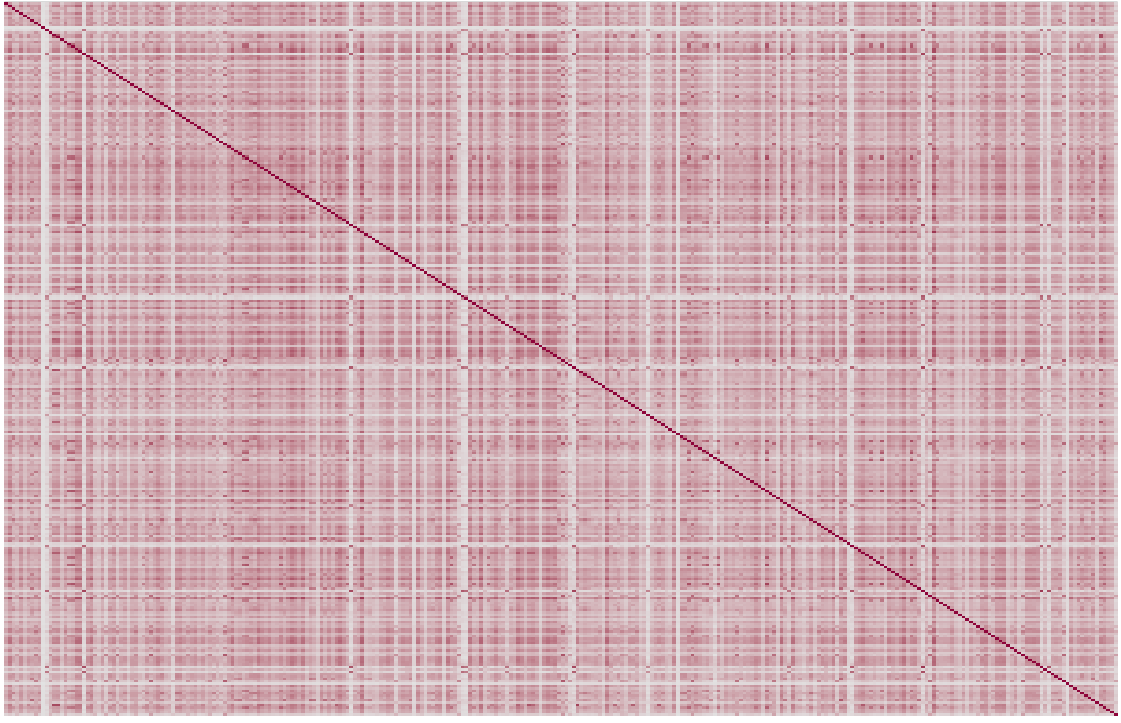
Median correlation matrix on 11/14/2007



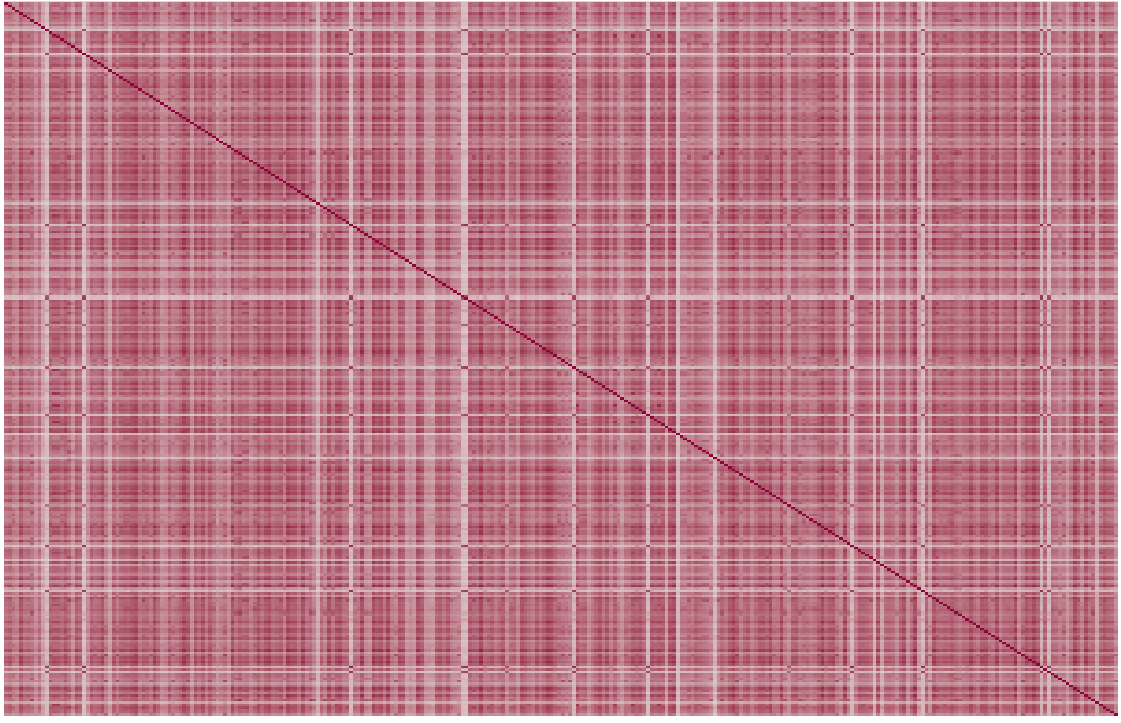
Median correlation matrix on 10/29/2008



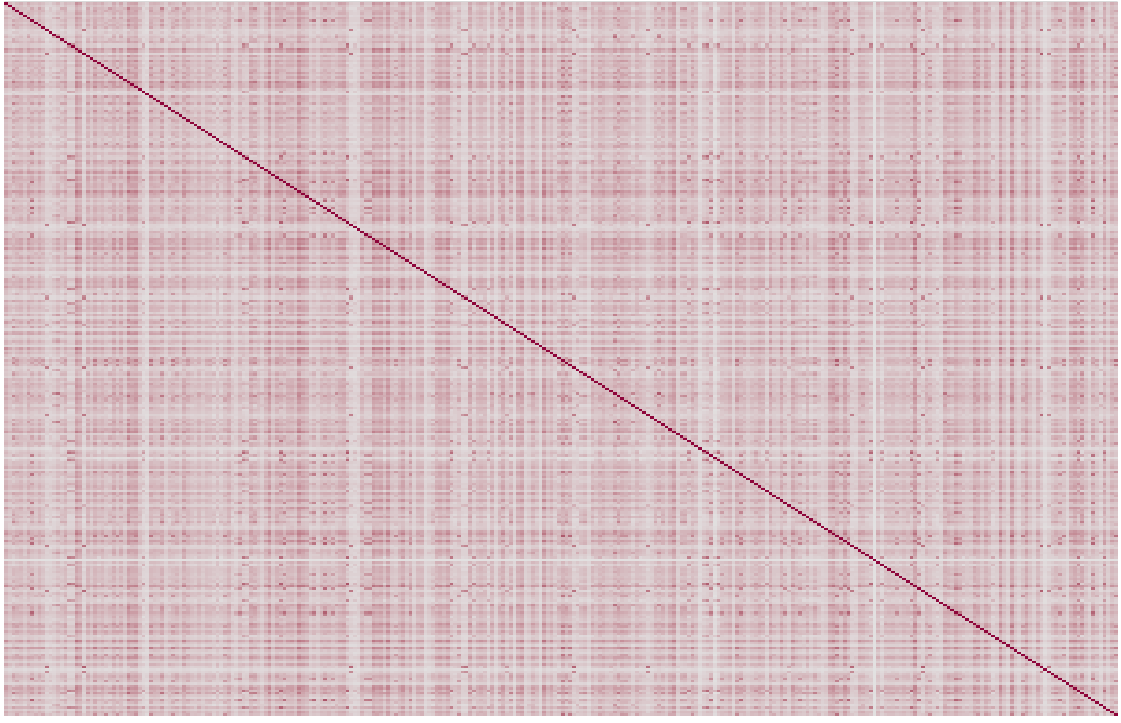
Median correlation matrix on 8/5/2009



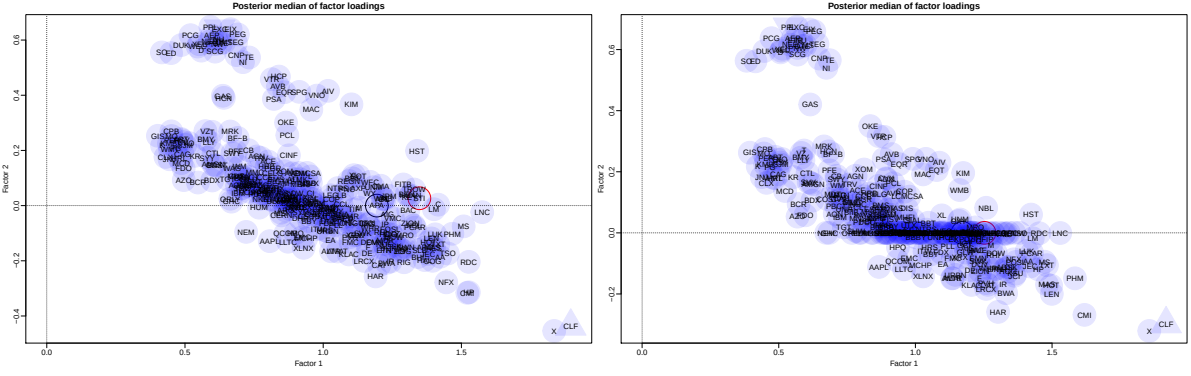
Median correlation matrix on 5/12/2010



Median correlation matrix on 4/27/2011



S&P 500: Loadings on factors 1/2 (of 7), N prior (left) and NG prior (right)



Practical implementation

Full conditional Gibbs sampling with efficient solutions for expensive or non-standard steps:

- Sampling the prior scalings τ_{ij}^2 of the shrinkage prior: efficient adaptive rejection sampling (Hörmann and Leydold, 2013), implemented in R-Package **GIGrvg** (Leydold and Hörmann, 2014).
- Sampling the univariate SVs: efficient auxiliary mixture sampling (Kastner and Frühwirth-Schnatter, 2014), implemented in R-Package **stochvol** (Kastner, forthcoming).
- Everything in C/C++ using **Armadillo** (Sanderson, 2010) linked to R via **RcppArmadillo** (Eddelbuettel and Sanderson, 2014).

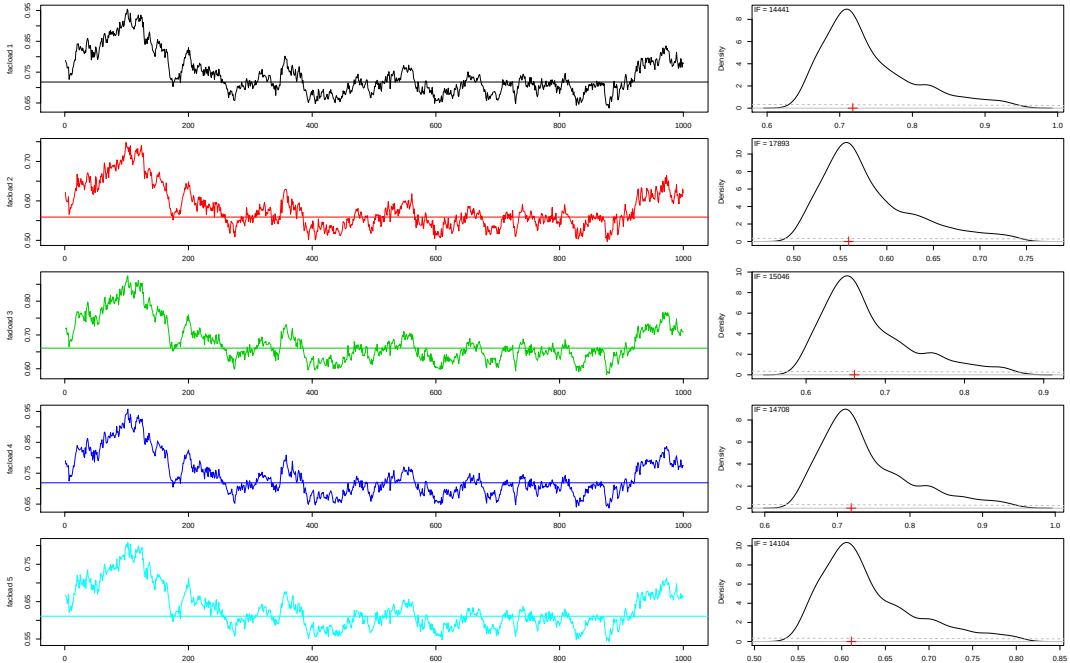
Practical implementation

Full conditional Gibbs sampling with efficient solutions for expensive or non-standard steps:

- Sampling the prior scalings τ_{ij}^2 of the shrinkage prior: efficient adaptive rejection sampling (Hörmann and Leydold, 2013), implemented in R-Package **GIGrvg** (Leydold and Hörmann, 2014).
- Sampling the univariate SVs: efficient auxiliary mixture sampling (Kastner and Frühwirth-Schnatter, 2014), implemented in R-Package **stochvol** (Kastner, forthcoming).
- Everything in C/C++ using **Armadillo** (Sanderson, 2010) linked to R via **RcppArmadillo** (Eddelbuettel and Sanderson, 2014).

However...

Extreme mixing problems for simulated data (thinning of 100!)



Boosting MCMC for state space models

Chib et al. (2006): no Gibbs sampling, MH-algorithm with sophisticated proposals to sample $\mathbf{\Lambda}$ without conditioning on $\mathbf{f}$.

More general strategies:

- Parameterizations matter for MCMC (Gelfand et al., 1995; Papaspiliopoulos et al., 2007), in particular for state space models – centered versus non-centered parameterizations (Frühwirth-Schnatter, 2004; Strickland et al., 2008).
- Marginal data augmentation (van Dyk and Meng, 2001): move to a non-identified state space model through a working parameter (Frühwirth-Schnatter and Lopes, 2015; Conti et al., 2014).
- Ancillarity-sufficiency interweaving strategy (ASIS) (Yu and Meng, 2011): move between parameterizations during sampling.

Interweaving the factor loadings

Exploit non-identifiability of factor models to move between the three levels of the hierarchical latent state representation.

Baseline representation:

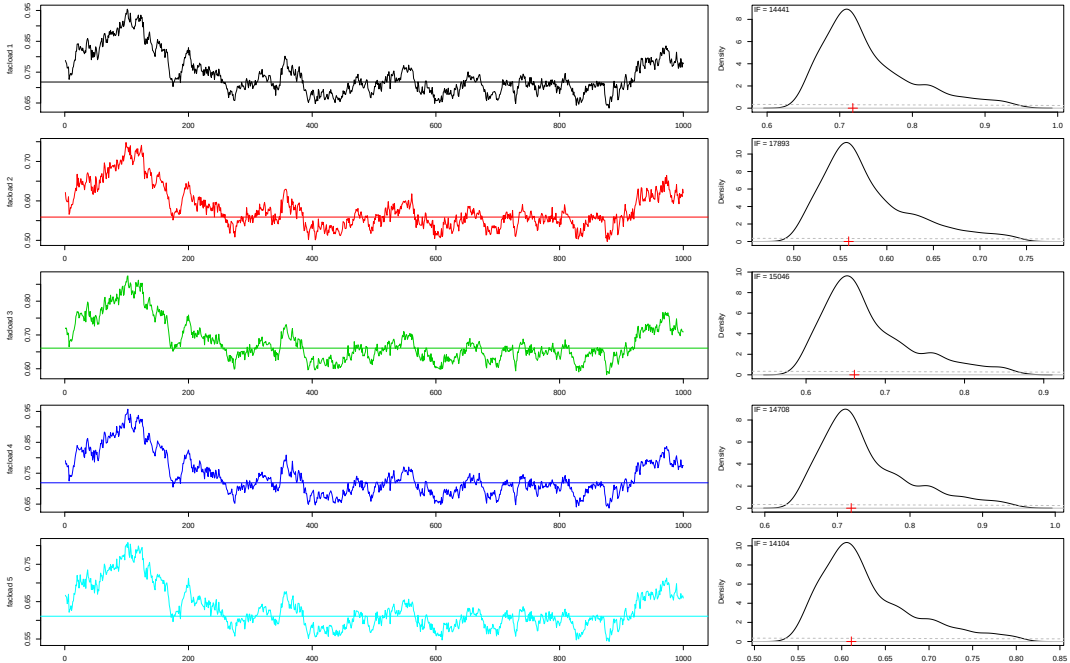
$$\mathbf{y}_t = \mathbf{\Lambda} \mathbf{f}_t + \boldsymbol{\varepsilon}_t,$$

with $\Lambda_{ii} \neq 0, i = 1, \ldots r$, and

$$\begin{aligned} f_{it} &\sim \mathcal{N}\left(0, 1 \cdot e^{h_{m+i,t}}\right), \quad i = 1, \ldots, r, \\ h_{m+i,t} &= \mu_{m+i} + \phi_{m+i}(h_{m+i,t} - \mu_{m+i}) + \eta_{m+i,t}, \end{aligned}$$

with $\mu_{m+i} = 0$.

Extreme mixing problems for simulated data (thinning of 100!)



Interweaving the factor loadings

Exploit non-identifiability of factor models to move between the three levels of the hierarchical latent state representation.

Shallow interweaving representation:

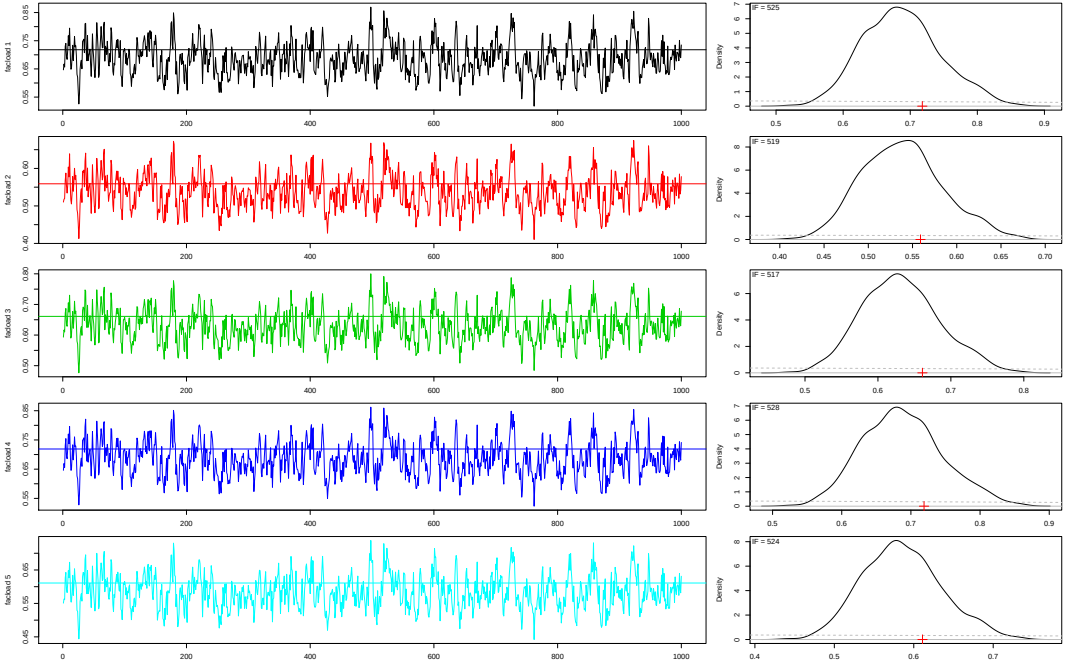
$$\mathbf{y}_t = \mathbf{\Lambda}^* \mathbf{f}_t^* + \boldsymbol{\varepsilon}_t,$$

with $\Lambda_{ii}^* = 1, i = 1, \ldots r$, and

$$\begin{aligned} f_{it}^* &\sim \mathcal{N}\left(0, \Lambda_{ii}^2 \cdot e^{h_{m+i,t}}\right), \quad i = 1, \ldots, r, \\ h_{m+i,t} &= \mu_{m+i} + \phi_{m+i}(h_{m+i,t} - \mu_{m+i}) + \eta_{m+i,t}, \end{aligned}$$

with $\mu_{m+i} = 0$.

Shallow interweaving



Interweaving the factor loadings

Exploit non-identifiability of factor models to move between the three levels of the hierarchical latent state representation.

Deep interweaving representation:

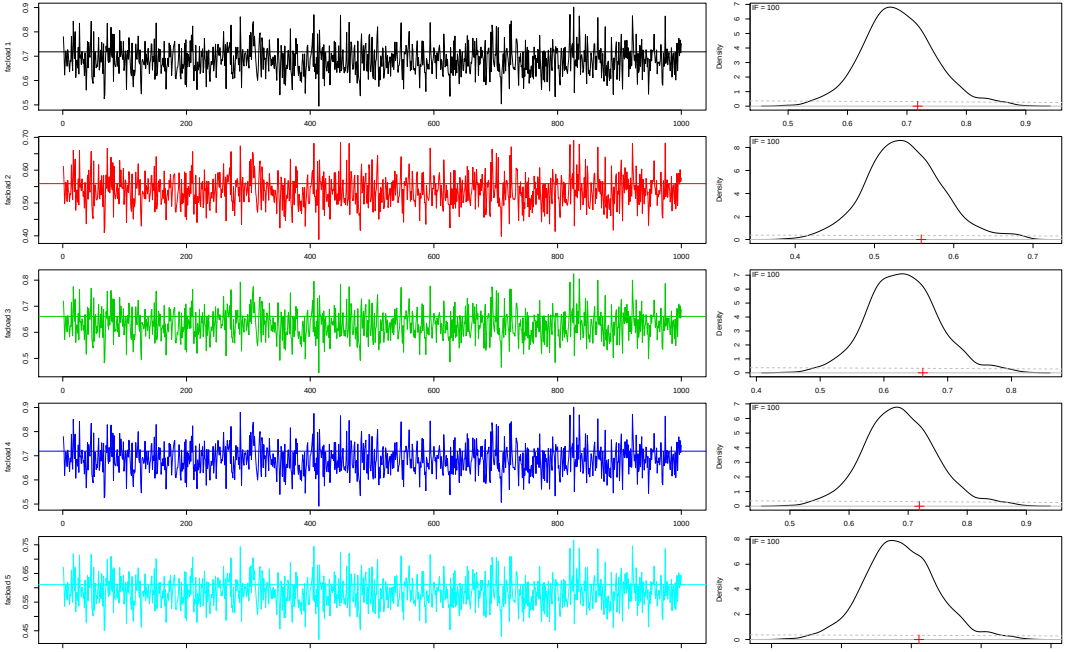
$$\mathbf{y}_t = \mathbf{\Lambda}^* \mathbf{f}_t^* + \varepsilon_t,$$

with $\Lambda_{ii}^* = 1, i = 1, \ldots r$, and

$$\begin{aligned} f_{it}^* &\sim \mathcal{N}\left(0, 1 \cdot e^{h_{m+i,t}}\right), \quad i = 1, \ldots, r, \\ h_{m+i,t} &= \mu_{m+i} + \phi_{m+i}(h_{m+i,t} - \mu_{m+i}) + \eta_{m+i,t}, \end{aligned}$$

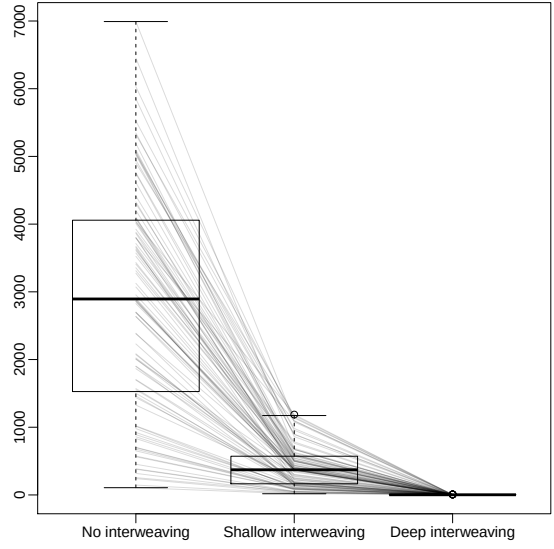
with $\mu_{m+i} = \log \Lambda_{ii}^2$.

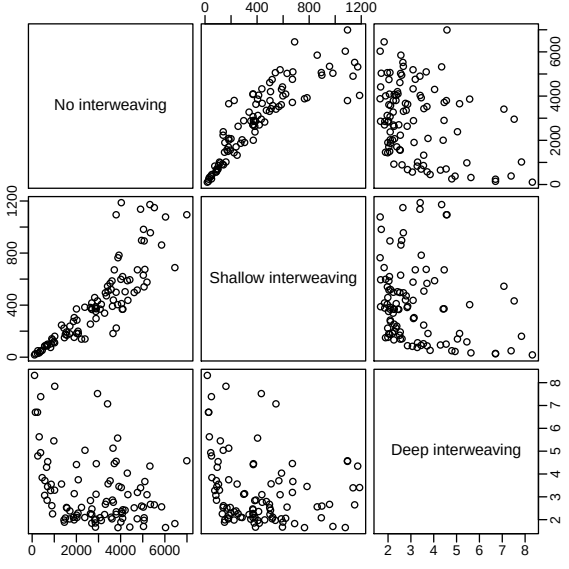
Deep interweaving



What about sample variation?

Estimated inefficiency factors for posterior draws from $p(\Lambda_{11}|\mathbf{y}_i)$ where $\mathbf{y}_i, i \in \{1, \ldots, 100\}$, denote artificially generated data with identical parameters.





Average IFs for factor loadings matrix $\mathbf{\Lambda}$

No interweaving		
1	2902.86	
2	2614.82	991.35
3	2932.20	214.87
4	2952.25	648.48
5	2926.91	865.40
6	2772.75	924.57
7	2264.23	948.54
8	1398.57	961.15
9	563.30	963.82
10	107.00	965.31

Shallow interweaving		
1	412.96	
2	394.06	148.37
3	410.95	62.70
4	408.42	110.23
5	405.60	130.05
6	394.07	138.14
7	371.41	141.90
8	318.16	143.91
9	215.74	145.16
10	74.11	146.12

Deep interweaving		
1	3.23	
2	6.48	3.83
3	3.26	9.04
4	3.45	5.59
5	3.90	4.30
6	4.78	3.77
7	6.36	3.52
8	9.11	3.41
9	13.83	3.33
10	20.91	3.32

Some final remarks

Sparse factor SV models are...

- a "hybrid approach" to dynamic covariance modeling, i.e. we get parsimony through
  - factor structure, plus additional
  - shrinkage
- estimable within a fully Bayesian framework, in particular via Gibbs sampling in combination with ASIS
- computationally doable (but by no means cheap!) in "vast dimensional" ($T = 10000+$, $m = 500+$, $r = 50+$) applications
- available through **R** package **factorstochvol** – almost! :)

References I

Aguilar, O. and M. West (2000). "Bayesian dynamic factor models and portfolio allocation". Journal of Business & Economic Statistics 18, pp. 338–357. DOI: 10.1080/07350015.2000.10524875 (cit. on p. 56).

Bhattacharya, A. and D. B. Dunson (2011). "Sparse Bayesian infinite factor models". Biometrika 98. DOI: 10.1093/biomet/asr013 (cit. on p. 15).

Carvalho, C. M. et al. (2008). "High-dimensional sparse factor modeling: Applications in gene expression genomics". Journal of the American Statistical Association 103, pp. 1438–1456. DOI: 10.1198/016214508000000869 (cit. on p. 15).

Chib, S., F. Nardari, and N. Shephard (2006). "Analysis of high dimensional multivariate stochastic volatility models". Journal of Econometrics 134, pp. 341–371. DOI: 10.1016/j.jeconom.2005.06.026 (cit. on pp. 6, 31).

Conti, G. et al. (2014). "Bayesian exploratory factor analysis". Journal of Econometrics 183, pp. 31–57. DOI: doi:10.1016/j.jeconom.2014.06.008 (cit. on pp. 15, 31).

Eddelbuettel, D. and C. Sanderson (2014). "RcppArmadillo: Accelerating R with high-performance C++ linear algebra". Computational Statistics and Data Analysis 71, pp. 1054–1063. DOI: 10.1016/j.csda.2013.02.005 (cit. on pp. 28, 29).

Fan, J., Y. Fan, and J. Lv (2008). "High dimensional covariance matrix estimation using a factor model". Journal of Econometrics 147, pp. 186–197 (cit. on p. 6).

Frühwirth-Schnatter, S. (2004). "Efficient Bayesian parameter estimation". In: State Space and Unobserved Component Models: Theory and Applications. Ed. by A. Harvey, S. J. Koopman, and N. Shephard. Cambridge: Cambridge University Press, pp. 123–151 (cit. on p. 31).

Frühwirth-Schnatter, S. and H. F. Lopes (2015). Parsimonious Bayesian Factor Analysis when the Number of Factors is Unknown. Research Report (cit. on pp. 15, 31).

References II

Gelfand, A., S. Sahu, and B. Carlin (1995). "Efficient parametrisations for normal linear mixed models". Biometrika 82, pp. 479–488. DOI: 10.1093/biomet/82.3.479 (cit. on p. 31).

Geweke, J. and G. Amisano (2010). "Comparing and evaluating Bayesian predictive distributions of asset returns". International Journal of Forecasting 26, pp. 216–230. DOI: 10.1016/j.ijforecast.2009.10.007 (cit. on p. 48).

Griffin, J. E. and P. J. Brown (2010). "Inference with Normal-Gamma prior distributions in regression problems". Bayesian Analysis 5, pp. 171–188. DOI: 10.1214/10-BA507 (cit. on p. 16).

Hörmann, W. and J. Leydold (2013). "Generating generalized inverse Gaussian random variates". Statistics and Computing, pp. 1–11. DOI: 10.1007/s11222-013-9387-3 (cit. on pp. 28, 29).

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). "Bayesian analysis of stochastic volatility models". Journal of Business & Economic Statistics 12, pp. 371–389. DOI: 10.1080/07350015.1994.10524553 (cit. on p. 4).

Kastner, G. (2015). **stochvol**: Efficient Bayesian inference for stochastic volatility (SV) models. R package version 1.2.0. URL: http://CRAN.R-project.org/package=stochvol.

Kastner, G. (forthcoming). "Dealing with stochastic volatility in time series using the R package **stochvol**". Journal of Statistical Software. URL: http://cran.r-project.org/web/packages/stochvol/vignettes/article.pdf (cit. on pp. 28, 29, 48).

Kastner, G. and S. Frühwirth-Schnatter (2014). "Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models". Computational Statistics and Data Analysis 76, pp. 408–423. DOI: 10.1016/j.csda.2013.01.002 (cit. on pp. 28, 29, 54).

Kastner, G., S. Frühwirth-Schnatter, and H. F. Lopes (2014). "Analysis of exchange rates via multivariate Bayesian factor stochastic volatility models". In: The Contribution of Young Researchers to Bayesian Statistics – Proceedings of BAYSM2013. Ed. by E. Lanzarone and F. Ieva. Vol. 63. Springer Proceedings in Mathematics & Statistics. Springer, pp. 181–186. DOI: 10.1007/978-3-319-02084-6_35 (cit. on p. 5).

References III

Kaufmann, S. and C. Schumacher (2013). Bayesian estimation of sparse dynamic factor models with order-independent identification. Working Paper 13.04. Swiss National Bank, Study Center Gerzensee. URL: http://ideas.repec.org/p/szg/worpap/1304.html (cit. on p. 15).

Kim, S., N. Shephard, and S. Chib (1998). "Stochastic volatility: Likelihood inference and comparison with ARCH models". Review of Economic Studies 65, pp. 361–393. DOI: 10.1111/1467-937X.00050 (cit. on pp. 4, 54).

Leydold, J. and W. Hörmann (2014). **GIGrvg**: Random variate generator for the GIG distribution. R package version 0.2. URL: http://CRAN.R-project.org/package=GIGrvg (cit. on pp. 28, 29).

Lopes, H. F. and C. M. Carvalho (2007). "Factor stochastic volatility with time varying loadings and Markov switching regimes". Journal of Statistical Planning and Inference 137, pp. 3082–3091. DOI: 10.1016/j.jspi.2006.06.047 (cit. on p. 56).

Nakajima, J. and M. West (2013). "Dynamic Factor Volatility Modeling: A Bayesian Latent Threshold Approach". Journal of Financial Econometrics 11, pp. 116–153. DOI: 10.1093/jjfinec/nbs013 (cit. on pp. 15, 56).

Nardari, F. and J. T. Scruggs (2007). "Bayesian analysis of linear factor models with latent factors, multivariate stochastic volatility, and APT pricing restrictions". Journal of Financial and Quantitative Analysis 42, pp. 857–891. DOI: 10.1017/S0022109000003422 (cit. on p. 56).

Omori, Y. et al. (2007). "Stochastic volatility with leverage: Fast and efficient likelihood inference". Journal of Econometrics 140, pp. 425–449. DOI: 10.1016/j.jeconom.2006.07.008 (cit. on p. 61).

Papaspiliopoulos, O., G. Roberts, and M. Sköld (2007). "A general framework for the parameterization of hierarchical models". Statistical Science 22, pp. 59–73 (cit. on p. 31).

Park, T. and G. Casella (2008). "The Bayesian Lasso". Journal of the American Statistical Association 103, pp. 681–686. DOI: 10.1198/016214508000000337 (cit. on p. 16).

References IV

Sanderson, C. (2010). Armadillo: An open source C++ linear algebra library for fast prototyping and computationally intensive experiments. Technical Report, NICTA. URL: http://arma.sourceforge.net/armadillo_nicta_2010.pdf (cit. on pp. 28, 29).

Strickland, C. M. et al. (2008). Efficient Bayesian Estimation of Multivariate State Space Models. Tech. rep. Queensland, Australia: Queensland University of Technology (cit. on p. 31).

van Dyk, D. and X.-L. Meng (2001). "The art of data augmentation". Journal of Computational and Graphical Statistics 10, pp. 1–50 (cit. on p. 31).

West, M. (2003). "Bayesian factor regression in the large p, small n paradigm". In: ed. by J. M. Bernardo et al. CHECK. Oxford: Oxford University Press, pp. 307–326 (cit. on p. 15).

Yu, Y. and X.-L. Meng (2011). "To center or not to center: that is not the question—An ancillarity-suffiency interweaving strategy (ASIS) for boosting MCMC efficiency". Journal of Computational and Graphical Statistics 20, pp. 531–570. DOI: 10.1198/jcgs.2011.203main (cit. on p. 31).

Zhou, X., J. Nakajima, and M. West (2014). "Bayesian forecasting and portfolio decisions using dynamic dependent sparse factor models". International Journal of Forecasting 30, pp. 963–980. DOI: 10.1016/j.ijforecast.2014.03.017 (cit. on pp. 15, 56).

A Bayesian predictive framework Video

Within a Bayesian framework, a natural way of assessing the predictive performance of a given model is through its predictive density (sometimes also referred to as posterior predictive distribution). It is given through

$$p(\mathbf{y}_{t+1}|\mathbf{y}^{o}_{[1:t]}) = \int_{\boldsymbol{\kappa}} p(\mathbf{y}_{t+1}|\mathbf{y}^{o}_{[1:t]}, \boldsymbol{\kappa}) \times p(\boldsymbol{\kappa}|\mathbf{y}^{o}_{[1:t]}) \, \mathrm{d}\boldsymbol{\kappa}, \tag{1}$$

where $\boldsymbol{\kappa}$ denotes the vector of all unobservables, i.e., parameters and possible latent variables.

Details

A Bayesian predictive framework

In the factor SV case, Equation 1 is a (very) high-dimensional integral which cannot be solved analytically. Nevertheless, it may be evaluated at an arbitrary point $\mathbf{x}$ through Monte Carlo integration,

$$p(\mathbf{x}|\mathbf{y}^o_{[1:t]}) \approx \frac{1}{M} \sum_{m=1}^{M} p(\mathbf{x}|\mathbf{y}^o_{[1:t]}, \boldsymbol{\kappa}^{(m)}_{[1:t]}), \tag{2}$$

where $\boldsymbol{\kappa}^{(m)}_{[1:t]}$ stands for the m^{th} draw from the respective posterior distribution up to time t. If Equation 2 is evaluated at $\mathbf{x} = \mathbf{y}^o_{t+1}$, we refer to it as the (one-step-ahead) predictive likelihood at time $t+1$, denoted PL_{t+1}.

Moreover, draws from the posterior predictive distribution can be obtained by simulating values $\mathbf{y}^{(m)}_{t+1}$ from the distribution given through the density $p(\mathbf{y}_{t+1}|\mathbf{y}^o_{[1:t]}, \boldsymbol{\kappa}^{(m)}_{[1:t]})$, the summands of Equation 2.

A Bayesian predictive framework

It is worth pointing out that log predictive likelihoods also carry an intrinsic connection to the log marginal likelihood, defined through

$$\log ML = \log p(\mathbf{y}^o) = \log \int_{\boldsymbol{\kappa}} p(\mathbf{y}^o|\boldsymbol{\kappa}) \times p(\boldsymbol{\kappa}) \, \mathrm{d}\boldsymbol{\kappa}.$$

This real number corresponds to the logarithm of the normalizing constant which appears in the denominator of Bayes' law and is often used for evaluating model evidence. It can straightforwardly be decomposed into the sum of the logarithms of the one-step-ahead predictive likelihoods:

$$\log ML = \log p(\mathbf{y}^o) = \log \prod_{t=1}^{n} p(\mathbf{y}_t^o|\mathbf{y}_{[1:t-1]}^o) = \sum_{t=1}^{n} \log PL_t.$$

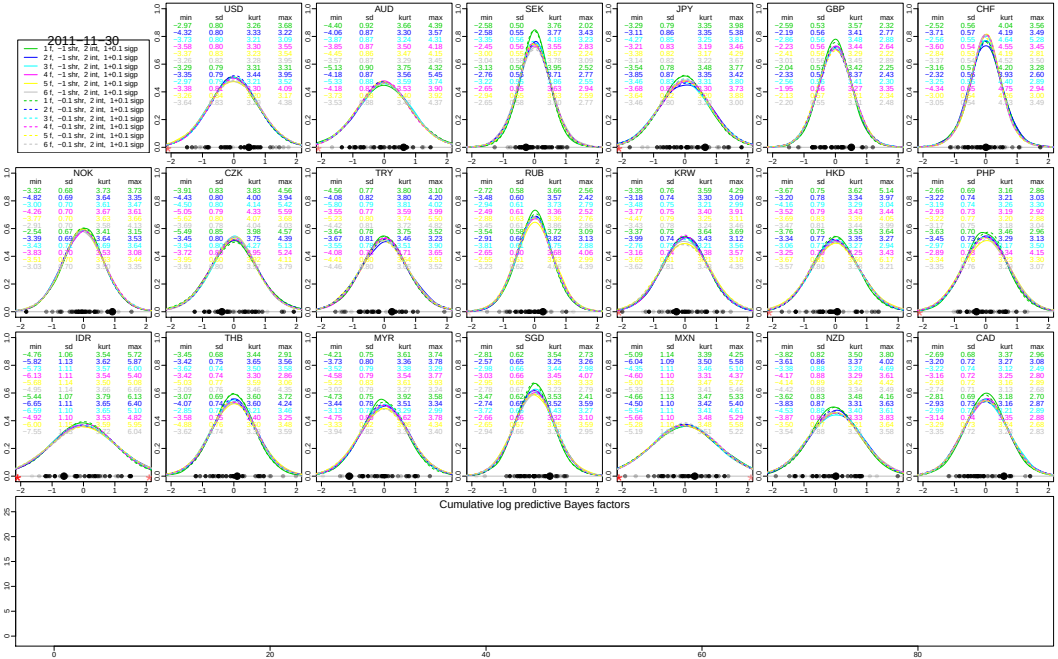
A Bayesian predictive framework

Cumulative sums of log PL_t also allow for model comparison through cumulative log predictive Bayes factors. Letting $PL_t(A)$ denote the predictive likelihood of model A at time t, and $PL_t(B)$ the corresponding value of model B, the cumulative log predictive Bayes factor at time u (and starting point s) in favor of model A over model B is straightforwardly given through

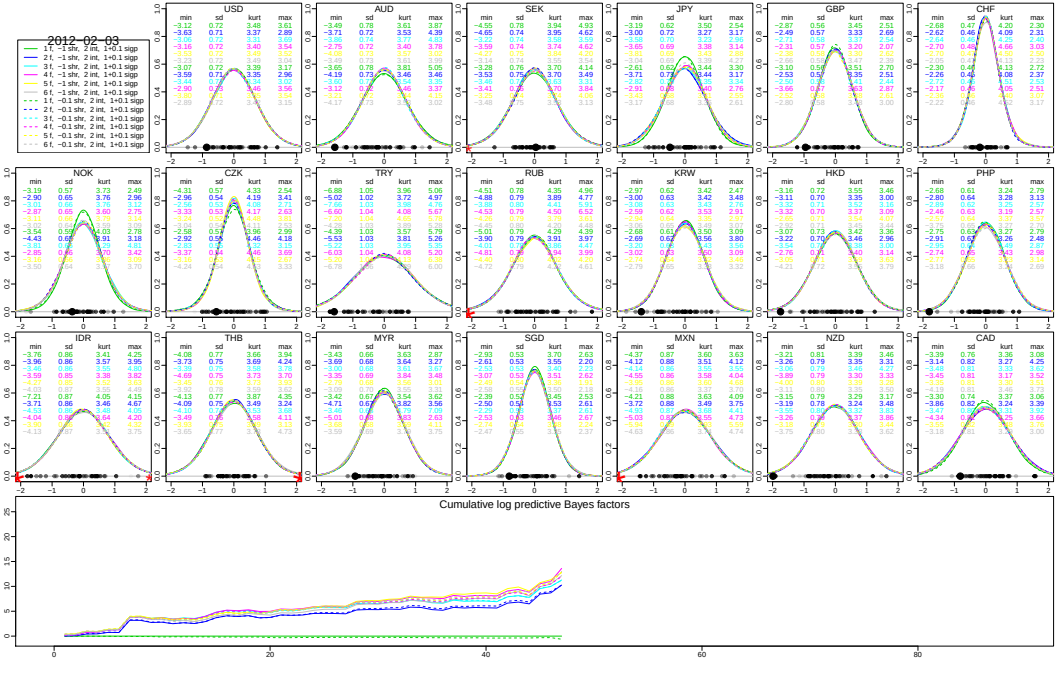
$$\log \left[\frac{p_A(\mathbf{y}^o_{[s+1:u]} | \mathbf{y}^o_{[1:s]})}{p_B(\mathbf{y}^o_{[s+1:u]} | \mathbf{y}^o_{[1:s]})} \right] = \sum_{t=s+1}^{u} \log \left[\frac{PL_t(A)}{PL_t(B)} \right]$$

When the cumulative log predictive Bayes factor is positive at a given point in time, there is evidence in favor of model A, and vice versa. In this context, information up to time s is regarded as prior information, while out-of-sample predictive evaluation starts at time $s+1$. Note that the usual (overall) log Bayes factor is a special case of this equation for $s = 0$ and $u = T$. More details in Geweke and Amisano (2010), Kastner (forthcoming). BACK

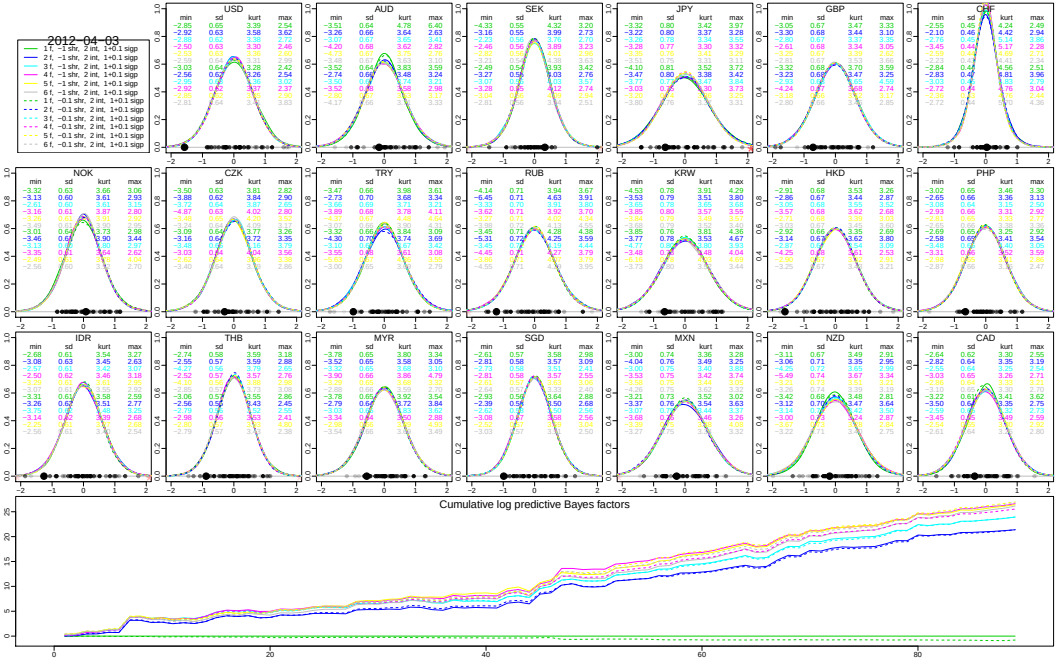
Exchange rates: Predictive marginals for Nov 30, 2011



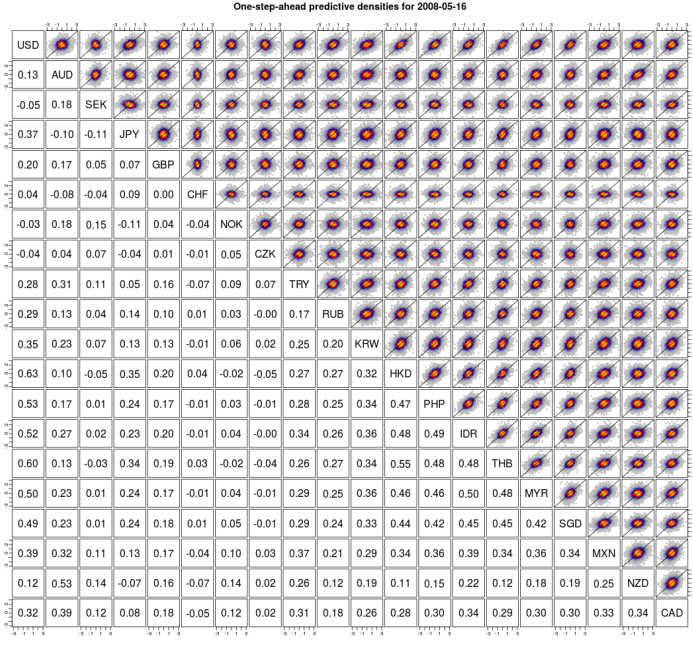
Exchange rates: Predictive marginals for Feb 3, 2012



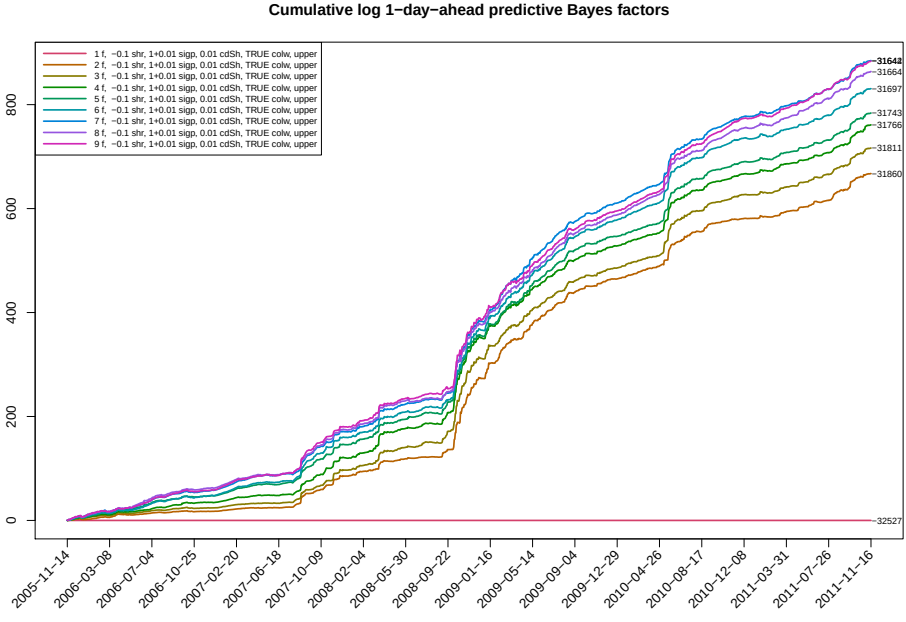
Exchange rates: Predictive marginals for Apr 3, 2012



Bivariate predictive marginals (exemplary day) Video



An open question. . .



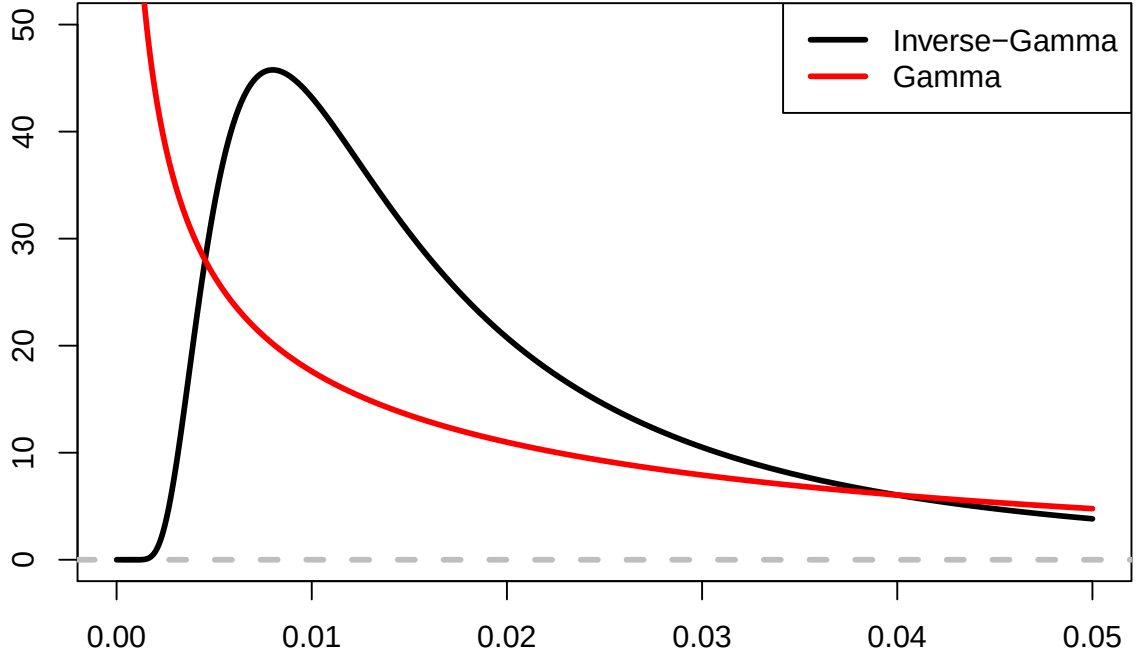
Priors

- $\Lambda_{ij} \sim \mathcal{N}(0, B_\Lambda)$
- $\mu \sim \mathcal{N}(b_\mu, B_\mu)$
- $(\phi + 1)/2 \sim \mathcal{B}(a_0, b_0)$ as in Kim et al. (1998), implying

$$p(\phi) = \frac{1}{2B(a_0, b_0)} \left(\frac{1+\phi}{2}\right)^{a_0 - 1} \left(\frac{1-\phi}{2}\right)^{b_0 - 1}$$

- $\sigma^2 \sim B_\sigma \times \chi^2_1 = \mathcal{G}\left(\frac{1}{2}, \frac{1}{2B_\sigma}\right)$ (cf. Kastner and Frühwirth-Schnatter, 2014)
- $h_0 | \mu, \phi, \sigma \sim \mathcal{N}(\mu, \sigma^2/(1-\phi^2))$ (stationary distribution)

Inverse Gamma prior for σ^2?



Identification, applications, generalizations

For identification reasons, we (currently) set

- $\mu_i = 0$ for $i \in \{m+1, \ldots, m+r\}$,
- $\boldsymbol{\Lambda}$ lower triangular.

$\Longrightarrow$ (certain) order dependence among the responses!

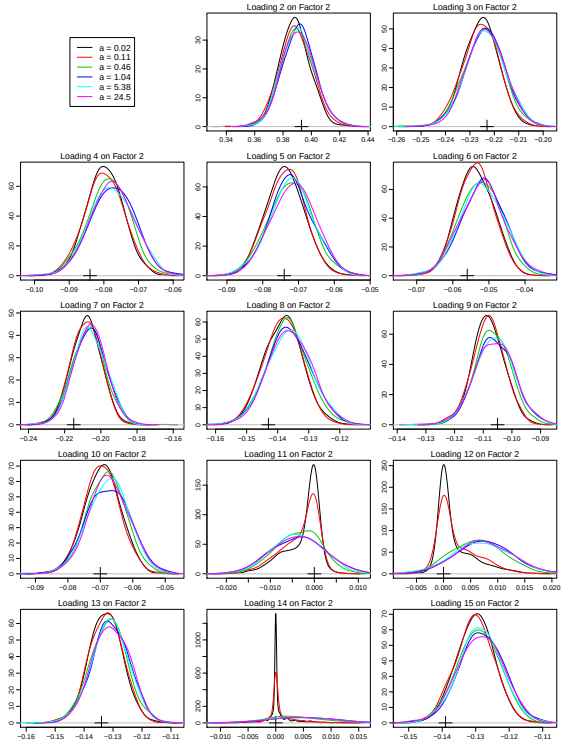
(Selected) applications in financial econometrics:

- Asset allocation (Aguilar and West, 2000),
- Asset pricing (Nardari and Scruggs, 2007),

extending arbitrage pricing theory (APT) and capital asset pricing models (CAPM) for time-varying covariance matrices.

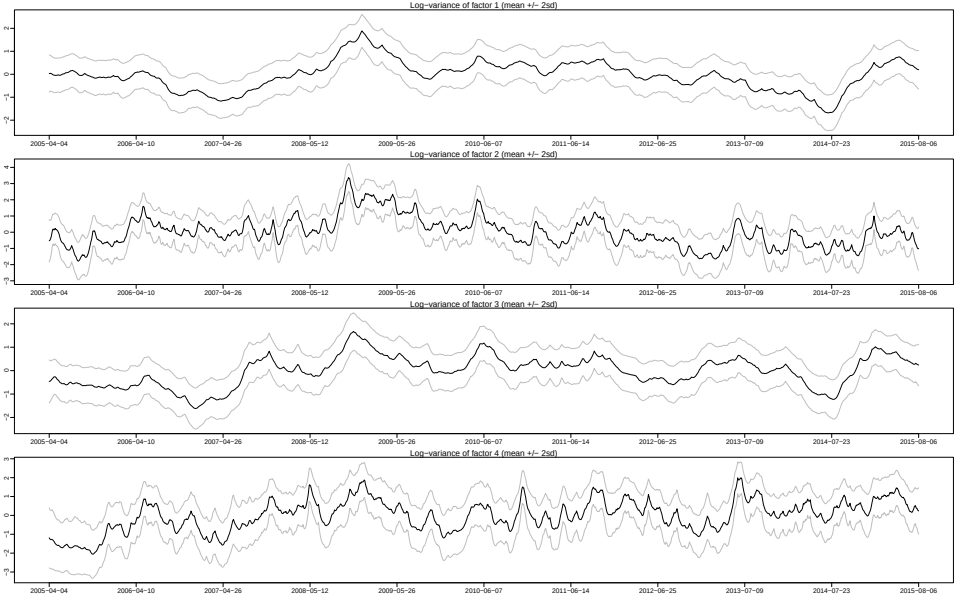
Generalizations:

- Time-varying loadings (Lopes and Carvalho, 2007),
- Time-varying patterns of sparsity (Nakajima and West, 2013; Zhou et al., 2014).

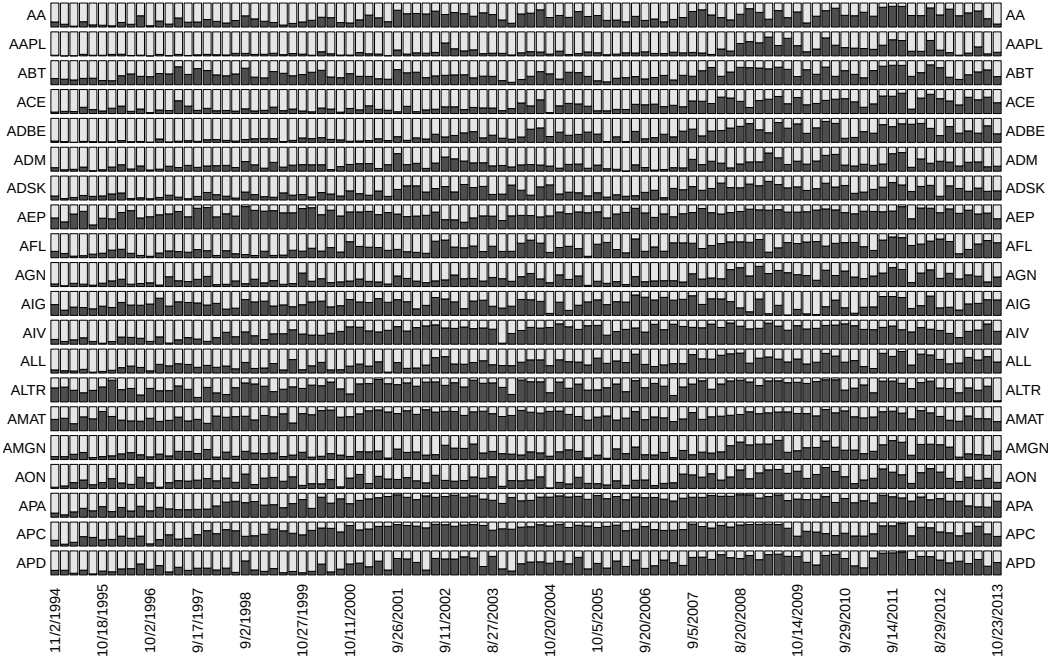

a = 0.02
a = 0.11
a = 0.46
a = 1.04
a = 5.38
a = 24.5
Loading 2 on Factor 2
Loading 3 on Factor 2
Loading 4 on Factor 2
Loading 5 on Factor 2
Loading 6 on Factor 2
Loading 7 on Factor 2
Loading 8 on Factor 2
Loading 9 on Factor 2
Loading 10 on Factor 2
Loading 11 on Factor 2
Loading 12 on Factor 2
Loading 13 on Factor 2
Loading 14 on Factor 2
Loading 15 on Factor 2

BACK

Marginal posterior distributions of the factor log-variances $h_{i,t}$ for $i \in \{27, \ldots, 30\}$ (mean $\pm$ 2 $\times$ sd)



Median posterior communalities (series 1–20)



MCMC sampling scheme

Choose $T(m+2r)+mr+4m+3r$ starting values for $\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, \mathbf{h}, \mathbf{f}, \boldsymbol{\Lambda}$, in our applications that's around around 100k or 2m, respectively. Then repeat:

a) Perform $m+r$ univariate SV updates for $\mathbf{h}_i, (\mu_i), \phi_i, \sigma_i$ by using the sampler in R-package **stochvol**.

b) Sample the factor loadings, constituting m independent r-variate regression problems with T observations.

c) Sample the latent factors, constituting T independent r-variate regression problems with m observations.

Note that in principle these steps can be straightforwardly parallelized.

Univariate SV Estimation

$\log(\varepsilon_t^2)|r_t \sim \mathcal{N}(m_{r_t}, s_{r_t}^2)$ for $r_t \in \{1, \ldots, 10\}$ (Omori et al., 2007):

1. Sample $\boldsymbol{h} = (h_0, \ldots, h_T)$ AWOL:
   - $\boldsymbol{h} = (\boldsymbol{L}')^{-1}(\boldsymbol{L}^{-1}\boldsymbol{c} + \varepsilon)$ with $\boldsymbol{\Omega} = \boldsymbol{L}\boldsymbol{L}'$ and $\varepsilon \sim \mathcal{N}_T(\boldsymbol{0}, \boldsymbol{I}_T)$
2. Sample μ, ϕ, σ (C):
   - MH-step with independence proposal for $(\mu, \phi)|\mathbf{h}, \sigma^2$
   - MH-step with independence proposal for $\sigma^2|\mathbf{h}, \mu, \phi$

2* Calculate $\tilde{\boldsymbol{h}} = \frac{\boldsymbol{h}-\mu}{\sigma}$

2** Sample μ, ϕ, σ (NC):
   - MH-step with independence proposal for $\phi|\tilde{\boldsymbol{h}}$
   - Gibbs-step for $(\mu, \sigma)|\boldsymbol{y}, \tilde{\boldsymbol{h}}, \boldsymbol{r}$ with possible sign switching for σ

2*** Calculate $\boldsymbol{h} = \mu + \sigma\tilde{\boldsymbol{h}}$ ("bookkeeping")

3. Draw indicators r_t from easily available posterior

BACK