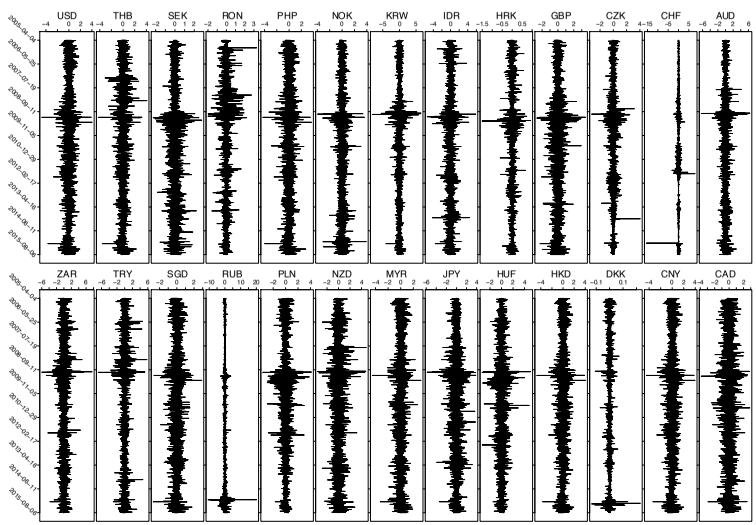
Outline

1. Introduction
2. Factor Models
3. Application I
4. Prediction
5. Sparse Factor Models
6. Application II
7. Efficient MCMC Sampling
8. 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

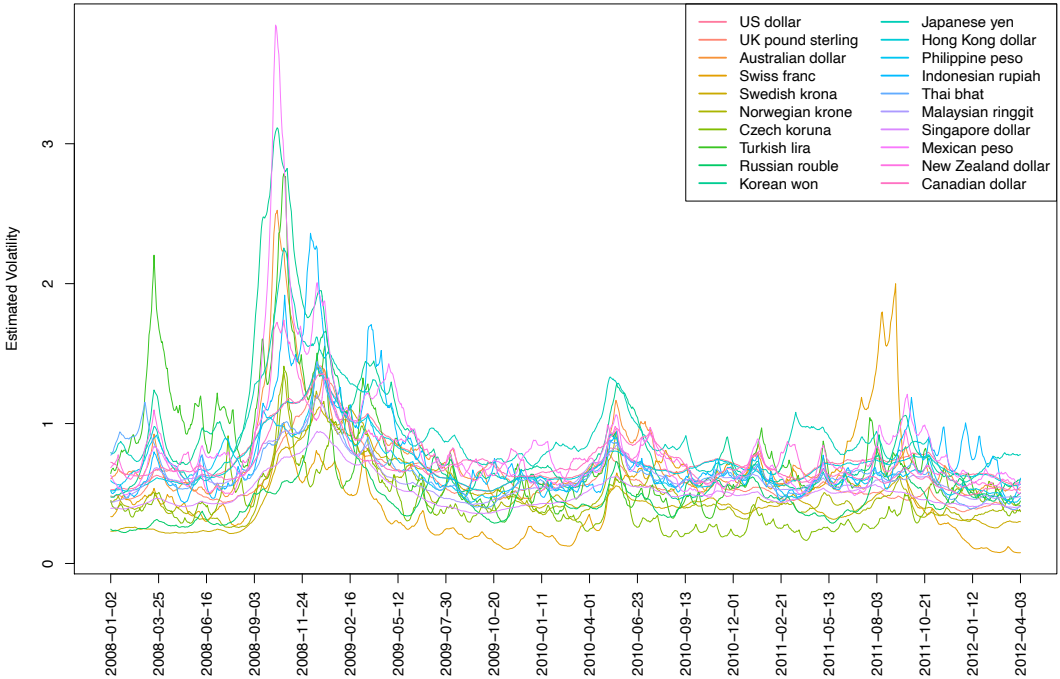
"Centered" version (e.g. Jacquier et al., 1994; Kim et al., 1998):

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

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 versions of 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



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 Priors

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.

Identification, applications, generalizations

For identification reasons, $\mathbf{\Lambda}$ is often assumed to be lower triangular with ones in the diagonal. Implies order dependence among the responses!

Alternatives:

- Let Λ_{jj} be unrestricted but set $\mu_{m+j} = 0$ for $j \in \{1, \ldots, r\}$,
- $\mathbf{\Lambda}$ lower triangular only if interpretation of factors is sought, otherwise unrestricted.

(Selected) applications in financial econometrics:

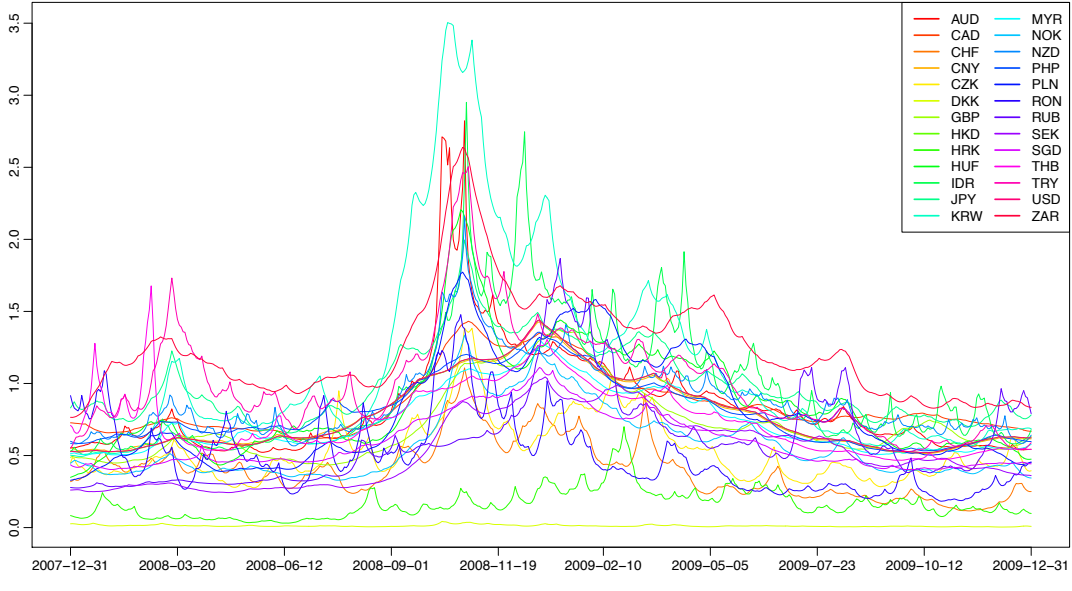
- Asset allocation (Aguilar and West, 2000),
- Asset pricing (Nardari and Scruggs, 2007),

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

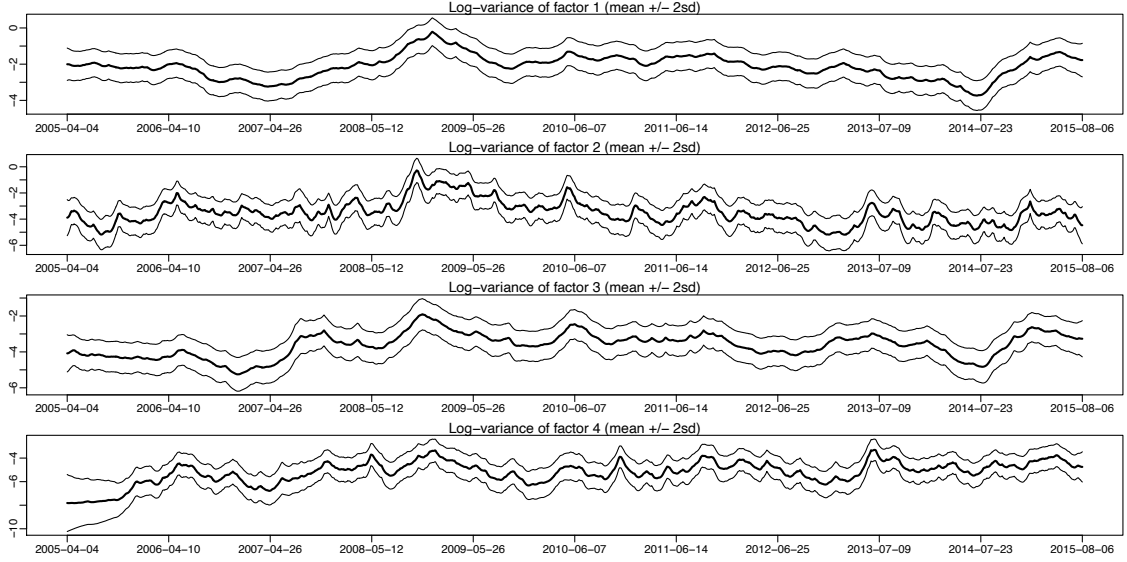
Generalizations:

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

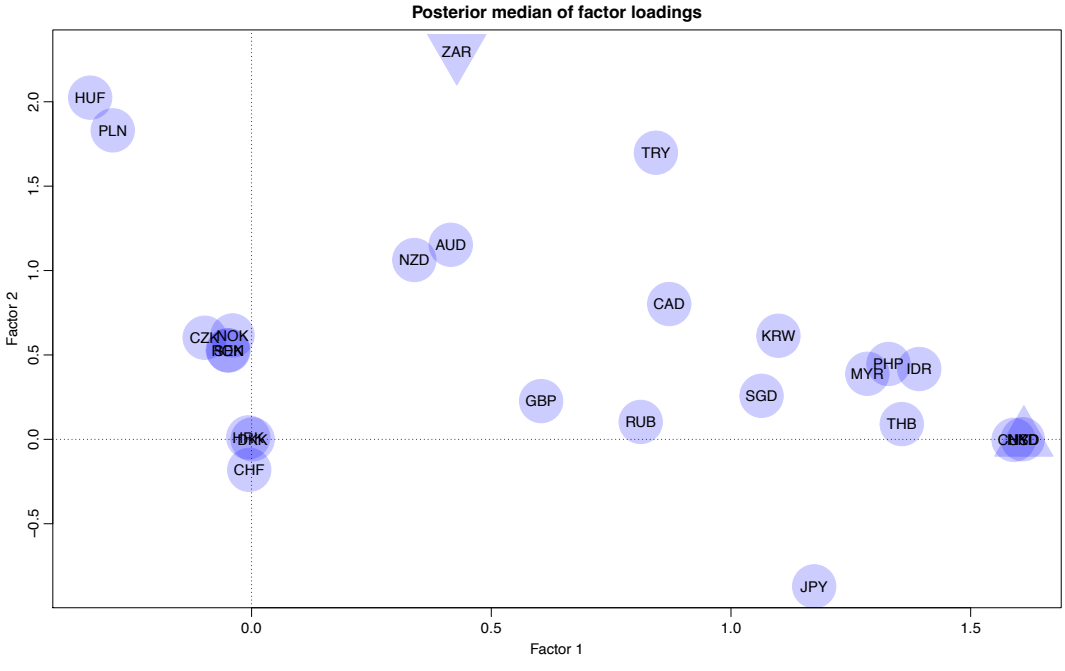
Exchange rate data: Mean of marginal posterior volatilities



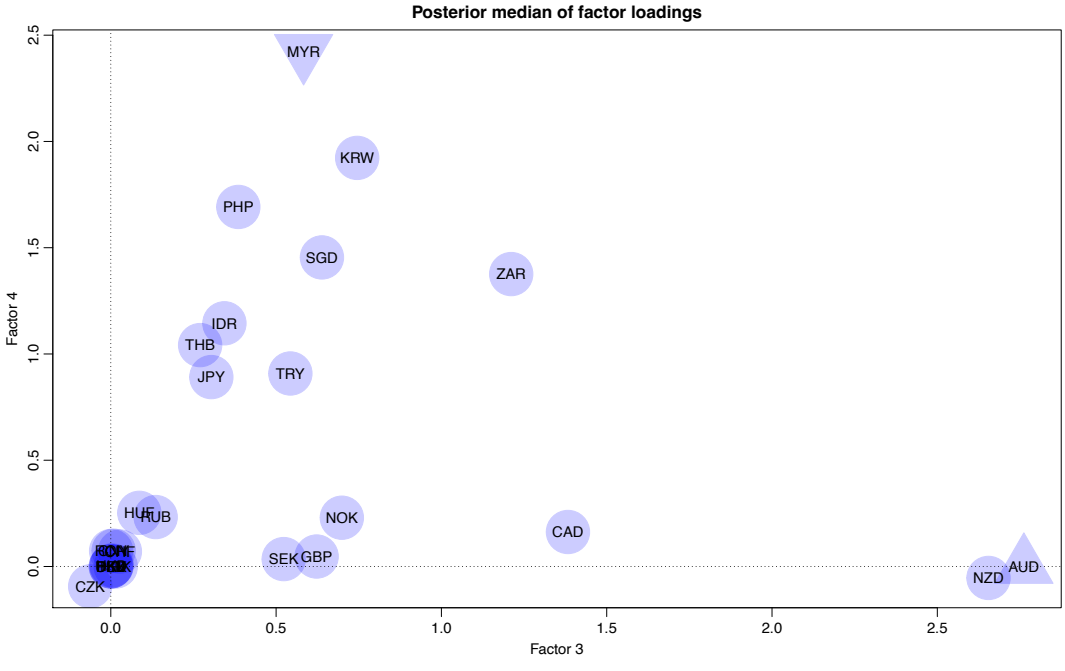
Marginal posterior distributions of the factor log-variances (mean $\pm$ 2 $\times$ sd)



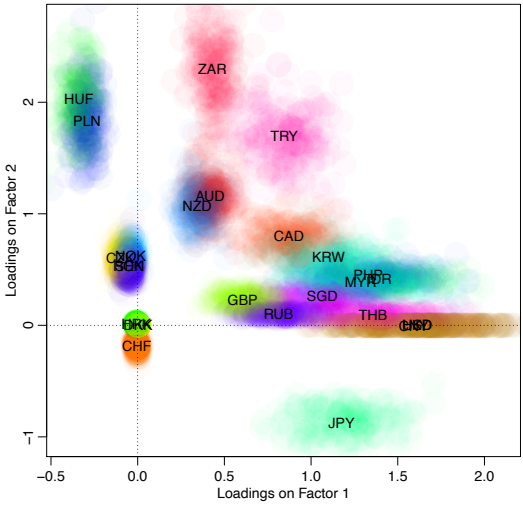
Median of loadings posterior $\mathbf{\Lambda}|\mathbf{y}$ (factors 1 and 2)

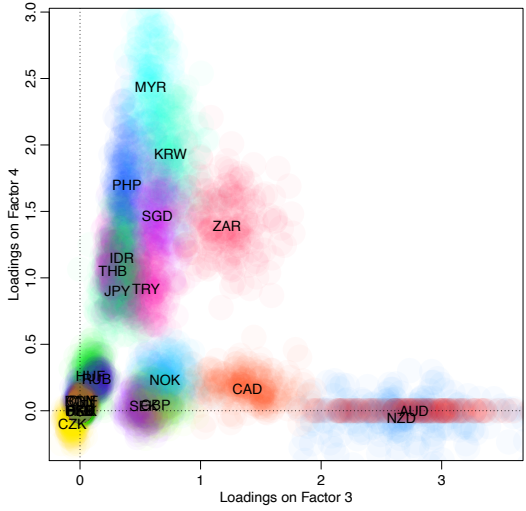


Median of loadings posterior $\mathbf{\Lambda}|\mathbf{y}$ (factors 3 and 4)



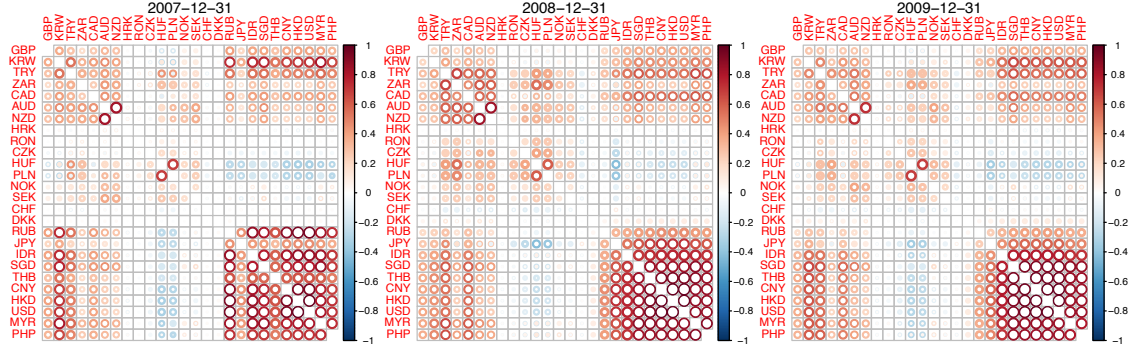
Posterior loadings distribution $\mathbf{\Lambda}|\mathbf{y}$



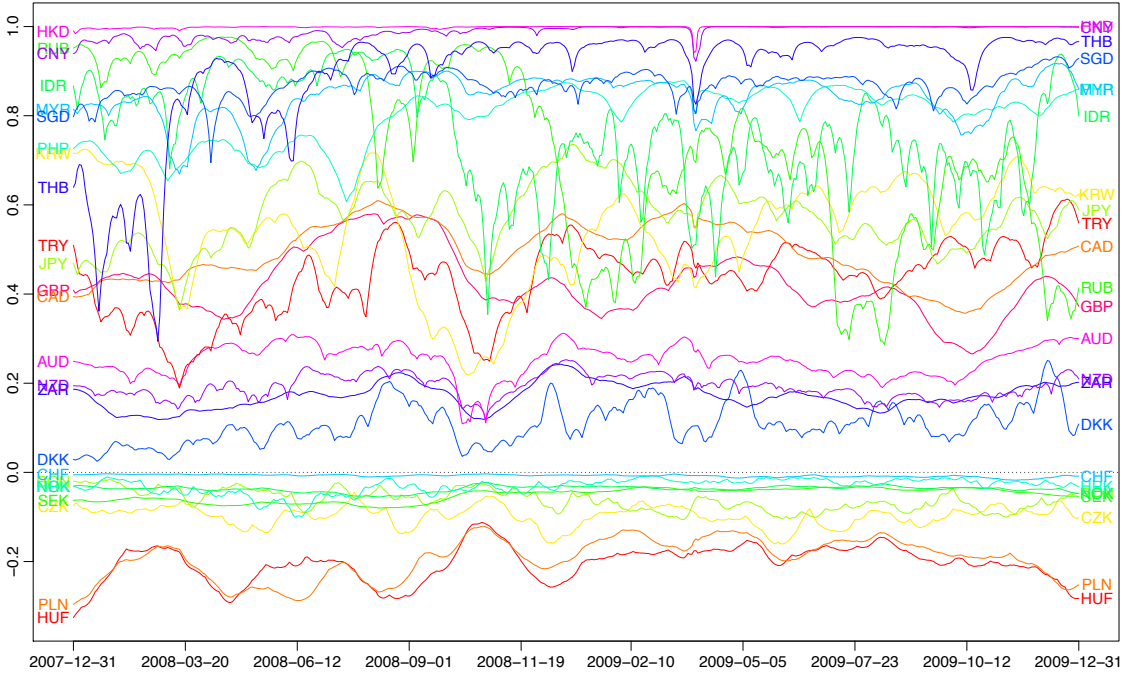


Exchange rate data

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



Mean posterior correlations with USD



A Bayesian predictive framework

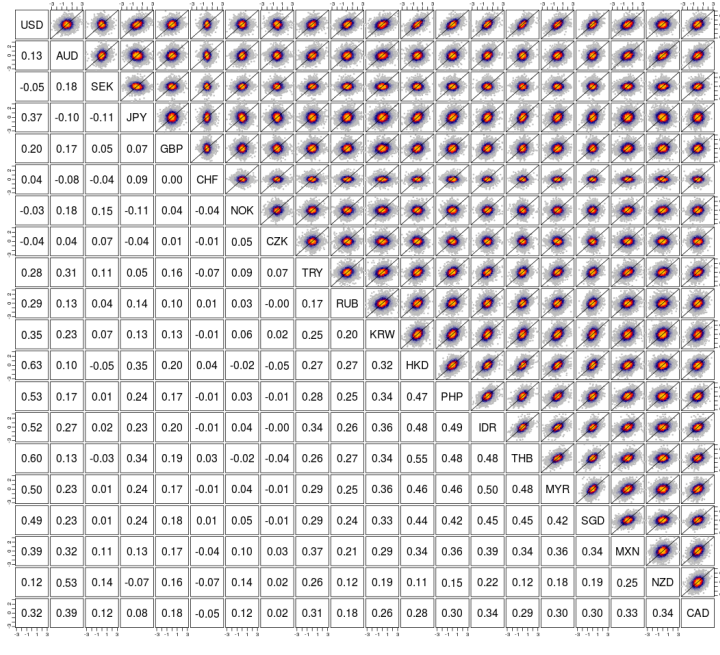
Within a Bayesian framework, a natural way of assessing the predictive performance of a given model is through its predictive density (sometimes also referred to as posterior predictive distribution). It is given through

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

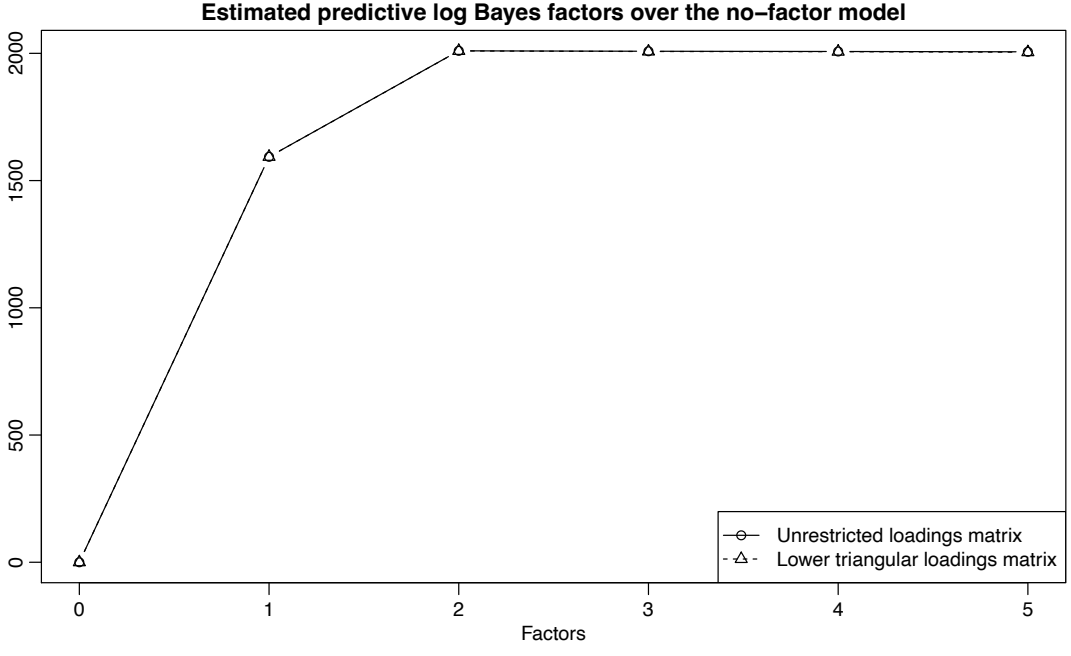
where $\boldsymbol{\kappa}$ denotes the vector of all unobservables, i.e., parameters and possible latent variables.

Video

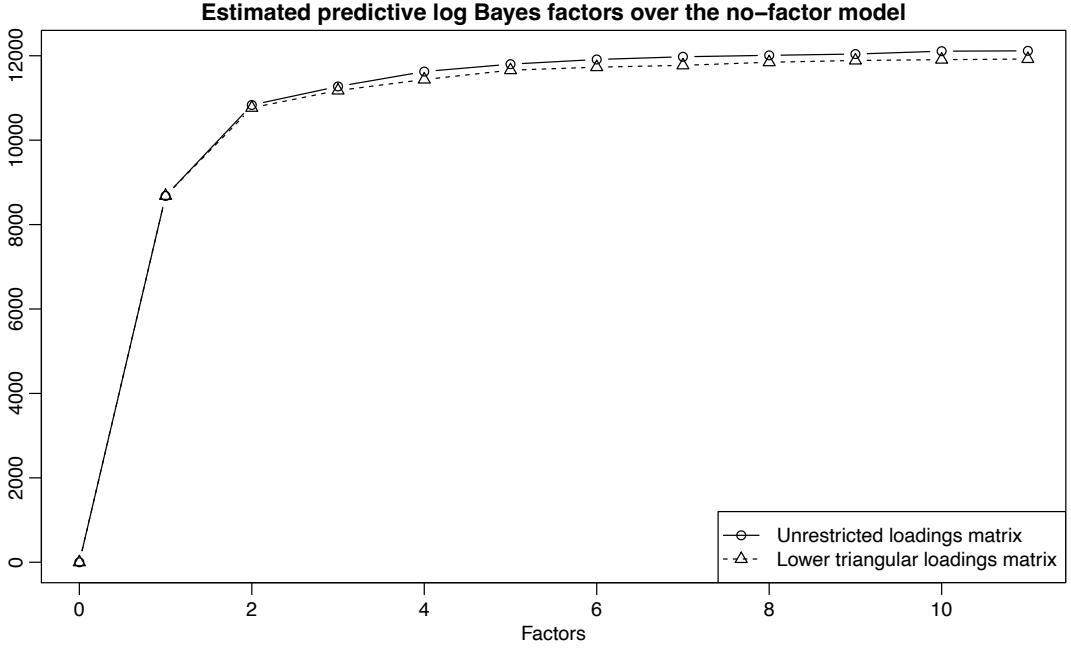
Bivariate predictive marginals (exemplary day) Video



Simulated data: $r_{\text{true}} = 2$



Exchange rate data



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, 2018)
  - 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)
- ...

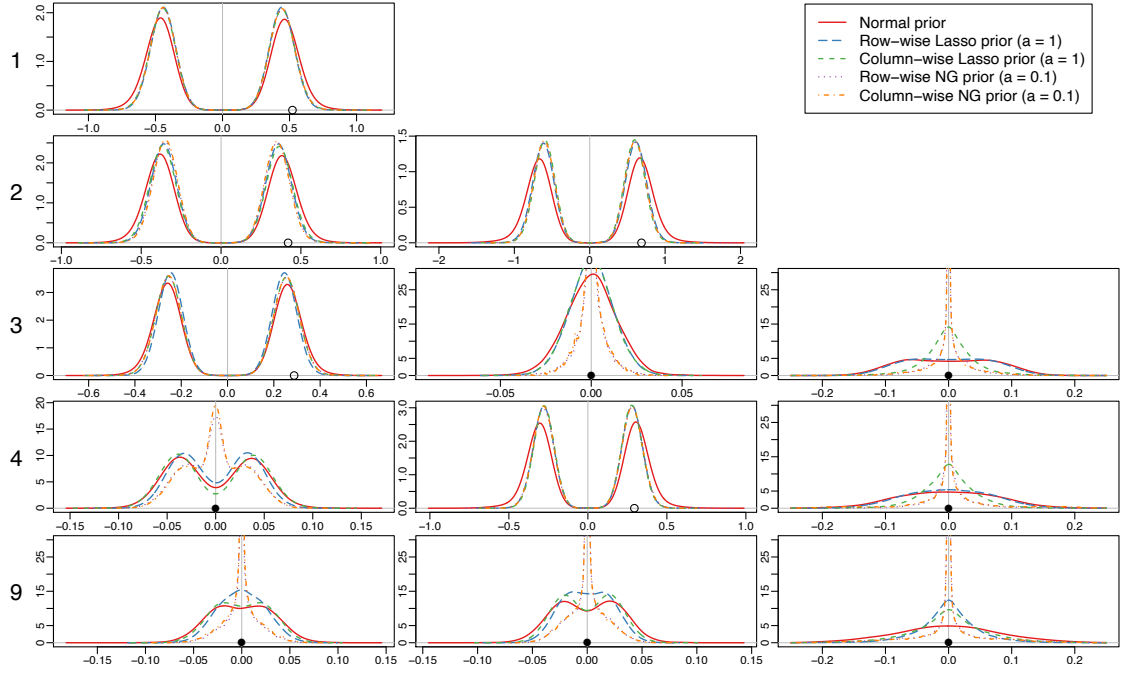
Sparse SV factor models

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

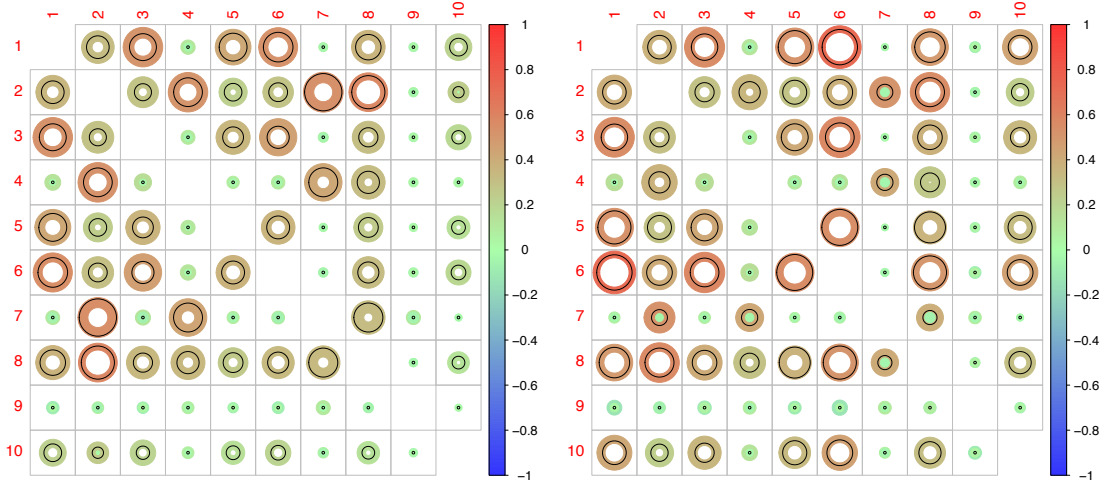
$$\Lambda_{ij}|\lambda_j, \tau_{ij} \sim \mathcal{N}\left(0, \tau_{ij}^2/\lambda_j^2\right), \quad \lambda_j^2 \sim \mathcal{G}\left(c, d\right), \quad \tau_{ij}^2 \sim \mathcal{G}\left(a, a\right).$$

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

Marginal factor loadings posteriors, $r_{\text{true}} = 2$, $r = 3$, $m = 10$, $T = 1000$

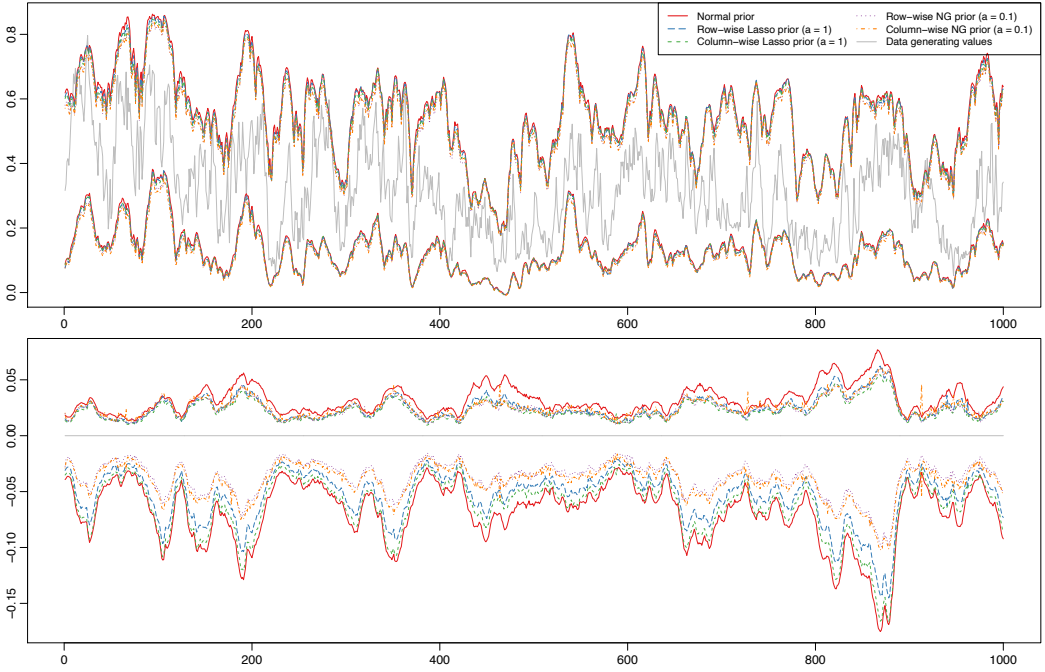


Posterior correlations at two points in time, $r_{\text{true}} = 2$, $r = 3$, $m = 10$, $T = 1000$



Left: $t = 1$, Right: $t = 1000$. Upper parts: NG-prior ($a = 0.1$), lower parts: Normal prior.

Two posterior correlations over time, $r_{\text{true}} = 2$, $r = 3$, $m = 10$, $T = 1000$



Bayesian Lasso vs. NG prior

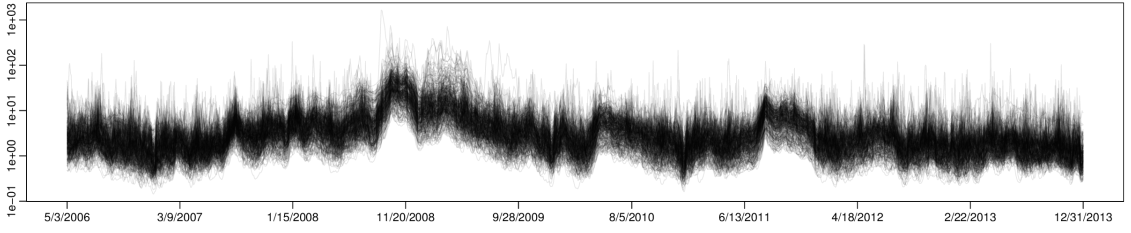
	1	2	3	4	5	6	7	8	9	10
Average 1	1.05	1.03	1.05	1.07	1.04	1.05	1.08	1.03	1.32	1.05
1		1.00	1.00	1.08	1.00	1.00	1.10	1.00	1.29	1.00
2	1.00		1.00	1.00	1.00	1.00	1.00	1.00	1.26	1.00
3	1.00	1.01		1.07	1.00	1.00	1.09	1.00	1.31	0.99
4	3.04	1.00	2.85		1.06	1.07	1.00	1.00	1.30	1.10
5	1.00	1.01	1.00	3.01		1.00	1.06	1.00	1.31	1.00
6	1.00	1.01	1.00	3.00	1.00		1.07	1.00	1.32	0.99
7	3.02	1.00	2.81	1.00	2.87	2.77		1.00	1.28	1.09
8	1.00	1.00	1.00	1.01	1.01	1.01	1.01		1.27	1.00
9	2.85	2.61	2.70	2.69	2.73	2.71	2.77	2.54		1.31
10	1.00	1.01	1.00	2.84	1.00	1.00	2.89	1.01	2.72	
Average 2	1.45	1.12	1.40	2.01	1.40	1.41	2.04	1.12	2.79	1.40

Table: Relative RMSEs of pairwise correlations. Above the diagonal: Row-wise Lasso ($a_i = 1$) vs. benchmark standard Gaussian prior with geometric means (first row). Below the diagonal: Row-wise Normal-Gamma ($a_i = 0.1$) vs. row-wise Lasso prior ($a_i = 1$) with geometric means (last row). Numbers greater than one mean that the former prior performs better than the latter. All values reported are medians of 100 repetitions where $c_i = d_i = 0.001$.

Application to S&P 500 members

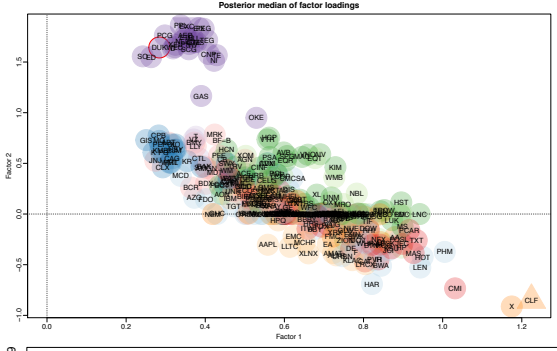
- Four-factor (more on that later) SV model to stock prices listed in the S&P 500 index.
- Only firms which have been listed from November 1994 onwards, resulting in $m = 300$ stock prices on 5001 days, ranging from 11/1/1994 to 12/31/2013.
- Data was obtained from Bloomberg Terminal in January 2014.
- Investigate $T = 5000$ demeaned percentage log-returns.
- Time-varying covariance matrix with 45150 nontrivial elements on 5000 days can be well explained by 4 factors with many factor loadings shrunken to 0.

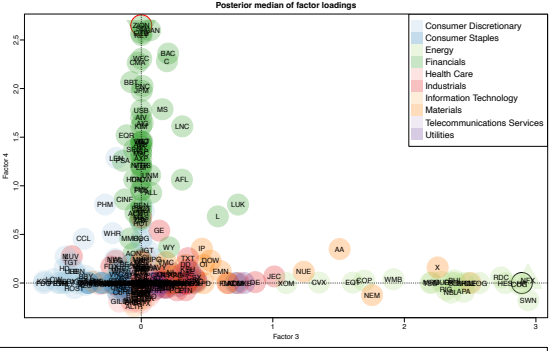
Marginal conditional variances (posterior mean, log scale)

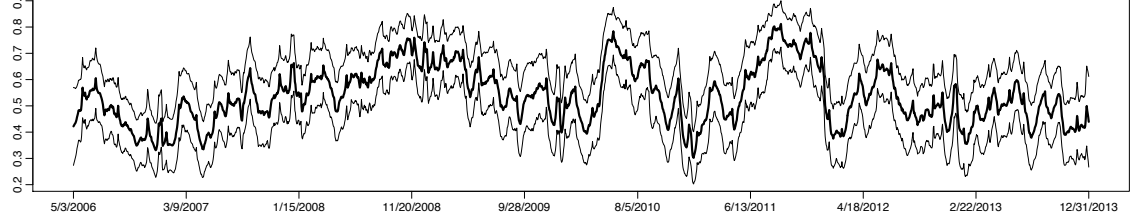


- Substantial co-movement (generally)
- Idiosyncratic deviations (at certain stretches in time)

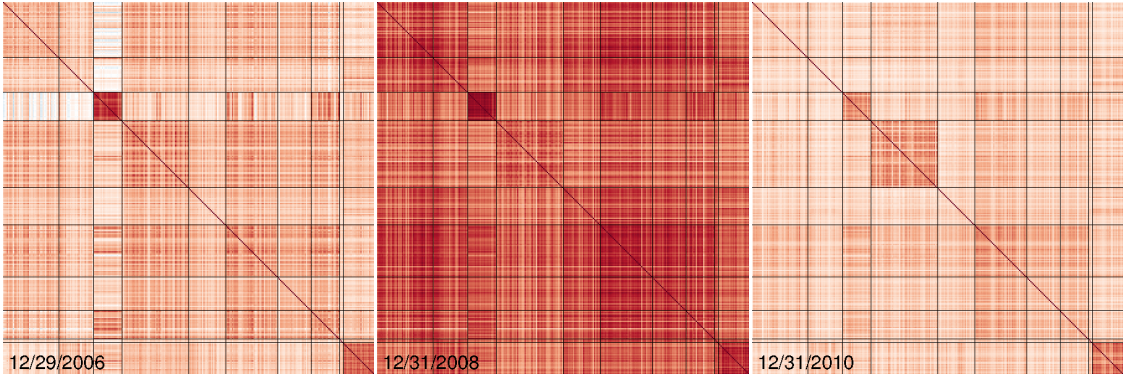
Median factor loadings and joint communalities (mean $\pm$ 2 sd)





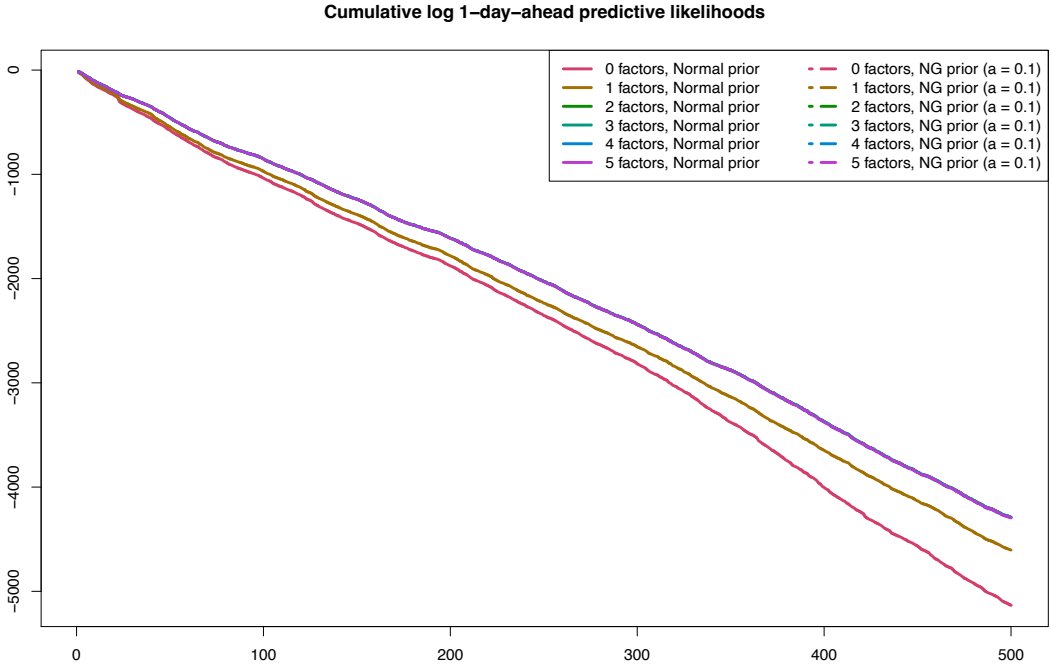


Mean posterior correlations

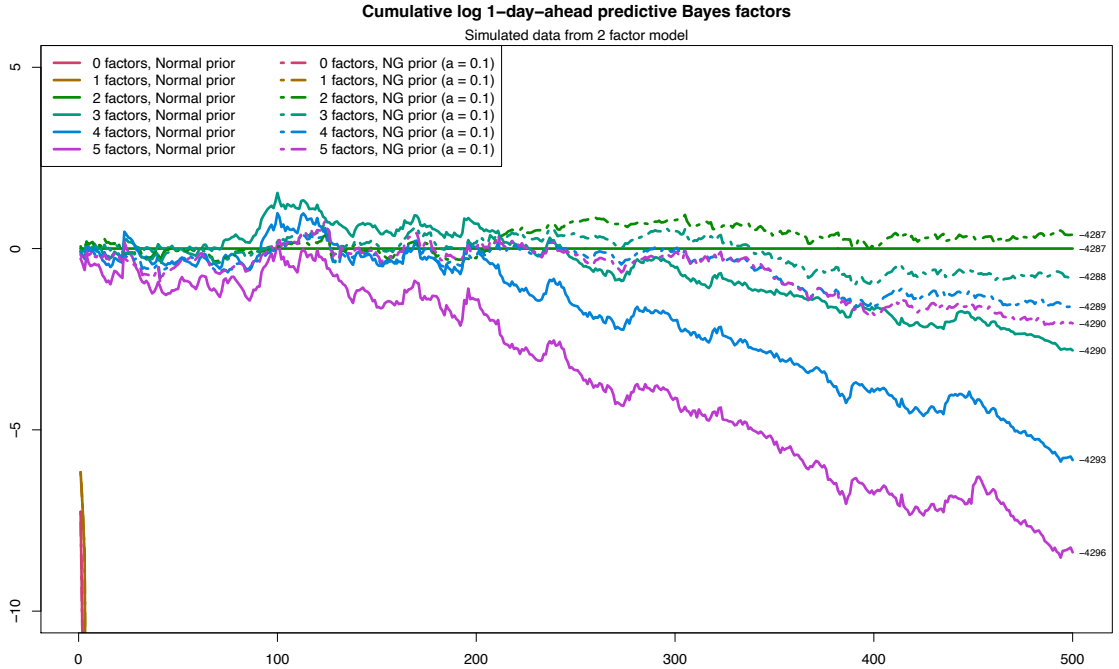


Video

A note about prediction (simulated data, $m = 10$, $r = 2$)



A note about prediction (simulated data, $m = 10$, $r = 2$)

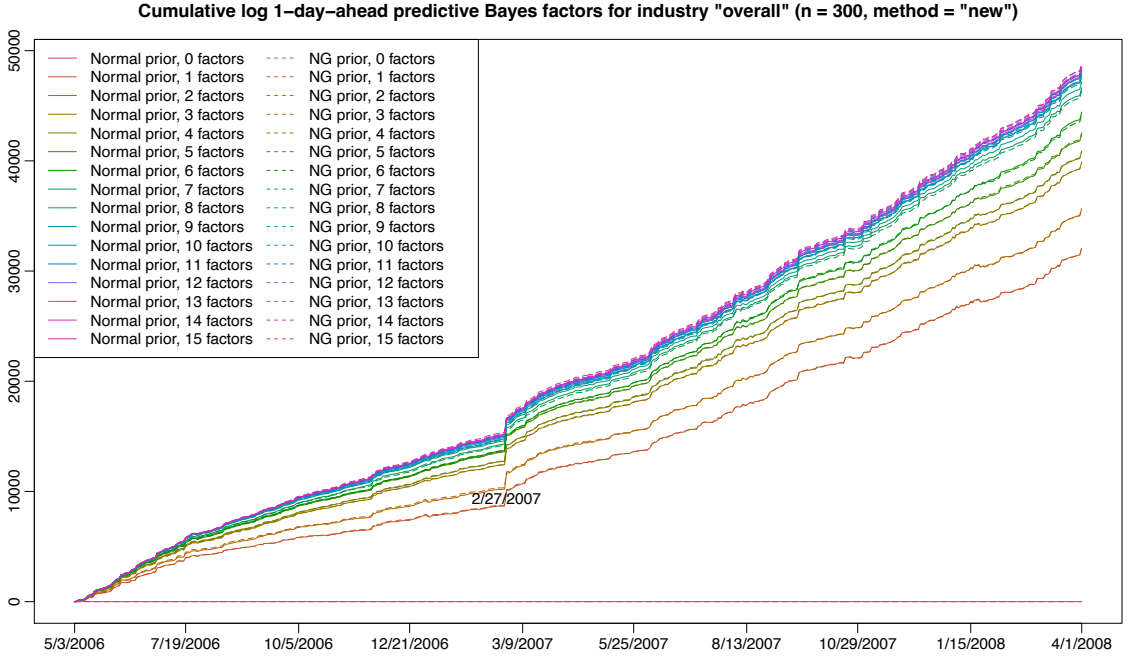


A note about prediction (simulated data, $m = 10$, $r = 2$)

	no fac	1 fac	2 fac	3 fac	4 fac	5 fac
Standard Gaussian	−845.80	−316.41		−2.81	−5.83	−8.38
Row-wise NG ($c_i = d_i = 1$)	−845.80	−317.75	0.27	−0.44	−1.74	−2.34
Col-wise NG ($c_j = d_j = 1$)	−845.80	−317.45	0.38	−0.77	−1.61	−2.06
Row-wise NG ($c_i = d_i = 0.01$)	−845.80	−317.71	0.40	−0.29	−1.36	−1.97
Col-wise NG ($c_j = d_j = 0.01$)	−845.80	−317.52	1.04	−0.12	−0.88	−2.07
Standard Gaussian	−787.41	−255.92		−4.98	−7.75	−12.67
Row-wise NG ($c_i = d_i = 1$)	−787.41	−256.85	2.90	1.73	0.87	−0.05
Col-wise NG ($c_j = d_j = 1$)	−787.41	−256.44	3.14	1.17	0.90	0.09
Row-wise NG ($c_i = d_i = 0.01$)	−787.41	−256.81	3.18	1.98	1.28	0.53
Col-wise NG ($c_j = d_j = 0.01$)	−787.41	−255.90	2.85	1.61	0.84	0.16

Table: Estimated log Bayes factors at $t = 1500$ against the 2-factor model using a standard Gaussian prior, where data up to $t = 1000$ is treated as the training sample. Lines 1 to 5 correspond to cumulative 1-day-ahead Bayes factors, lines 6 to 10 correspond to 10-days-ahead predictive Bayes factors.

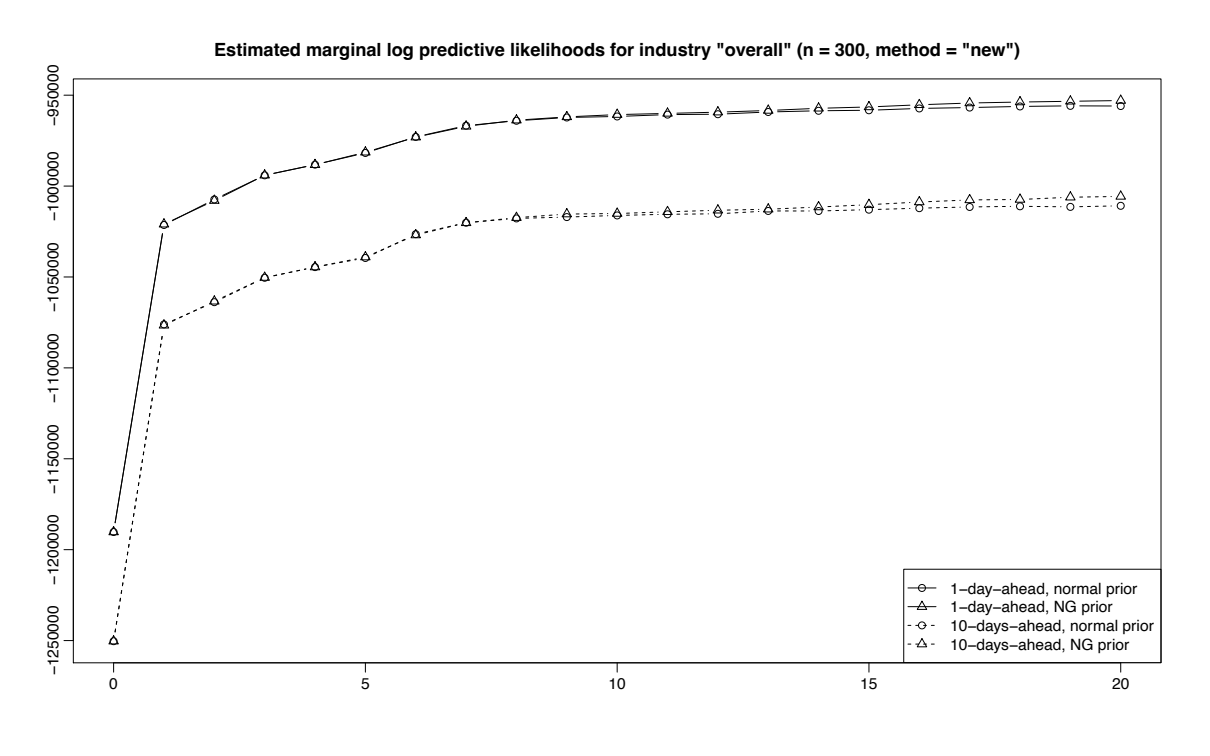
A note about prediction (S&P 500 data, $m = 300$, excerpt)



A note about prediction (S&P 500 data, $m = 300$)

	Mean	2/27/2007	8/4/2011	8/8/2011
1 factors, Gaussian	84.83	1415.13	1189.78	1446.03
2 factors, Gaussian	91.88	1455.39	1261.38	1462.23
3 factors, Gaussian	98.55	1449.31	1237.97	1479.84
4 factors, Gaussian	101.50	1449.21	1254.36	1502.17
10 factors, Gaussian	114.81	1482.36	1271.59	1557.63
20 factors, Gaussian	117.69	1481.15	1293.98	1568.60
1 factors, NG prior	85.03	1421.45	1194.08	1445.51
2 factors, NG prior	91.56	1450.88	1239.42	1461.73
3 factors, NG prior	98.57	1448.90	1231.94	1471.58
4 factors, NG prior	101.47	1443.86	1240.68	1536.77
10 factors, NG prior	115.32	1482.32	1283.90	1546.43
20 factors, NG prior	119.21	1499.08	1293.74	1578.54

Table: Average and top 3 daily 1-day-ahead log predictive gains over the no-factor SV model.


Estimated marginal log predictive likelihoods for industry "overall" (n = 300, method = "new")
−950000
−1000000
−1050000
−1100000
−1150000
−1200000
−1250000
0
5
10
15
20
1−day−ahead, normal prior
1−day−ahead, NG prior
10−days−ahead, normal prior
10−days−ahead, NG prior

Implied minimum variance portfolios and psuedo log predictive scores

- Consider the global minimum variance portfolio implied by $\hat{\boldsymbol{\Sigma}}_{t+1}$ (the point estimate or posterior mean estimate, respectively) which uniquely defines the optimal portfolio weights

$$\boldsymbol{\omega}_{t+1} = \frac{\hat{\boldsymbol{\Sigma}}_{t+1}^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'\hat{\boldsymbol{\Sigma}}_{t+1}^{-1}\boldsymbol{\iota}}.$$

- Consider pseudo log predictive scores (PLPSs), i.e. use data up to time t and determine a point estimate $\hat{\boldsymbol{\Sigma}}_{t+1}$ for $\boldsymbol{\Sigma}_{t+1}$ and evaluate a multivariate normal density $\mathcal{N}_m\left(\mathbf{0}, \hat{\boldsymbol{\Sigma}}_{t+1}\right)$ at the actually observed value $\boldsymbol{y}_{t+1}^o$ to obtain the one-day-ahead PLPS for time $t+1$.

Mimimum variance portfolio performance and pseudo LPS (1000 days)

	SD	Avg	Sharpe	PLPS
Equal weight portfolio	29.02	0.00	0.00	
MA (500 days)	16.56	−2.39	−0.14	−320.17
EWMA ($\alpha = 0.94$)	35.90	−4.43	−0.12	-3×10^9
EWMA ($\alpha = 0.99$)	18.42	−4.18	−0.23	−1409.97
Ledoit-Wolf	12.53	6.14	0.49	25.68
FF+MOM	16.82	12.68	0.75	74.56
no-factor SV	22.53	1.02	0.05	0.00
N-FSV 1	19.03	9.45	0.50	66.29
N-FSV 2	18.92	9.32	0.49	75.25
N-FSV 3	15.50	7.36	0.48	81.47
N-FSV 4	14.58	8.26	0.57	87.07
N-FSV 10	12.94	12.45	0.96	97.74
N-FSV 20	12.57	7.81	0.62	101.82
N-FSV 50	12.41	5.03	0.40	106.57
NG-FSV 1	18.98	9.44	0.50	66.29
NG-FSV 2	18.84	9.37	0.50	75.21
NG-FSV 3	15.52	7.20	0.46	81.39
NG-FSV 4	14.85	8.36	0.56	86.80
NG-FSV 10	12.94	12.00	0.93	97.74
NG-FSV 20	12.02	13.33	1.11	102.30
NG-FSV 50	12.42	10.78	0.87	107.23

MCMC 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, 2017).
- Sampling the univariate SVs: efficient auxiliary mixture sampling (Kastner and Frühwirth-Schnatter, 2014), implemented in R-Package **stochvol** (Kastner, 2016).
- Everything in C/C++ using **Armadillo** (Sanderson, 2010) linked to R via **RcppArmadillo** (Eddelbuettel and Sanderson, 2014).

MORE

MCMC implementation

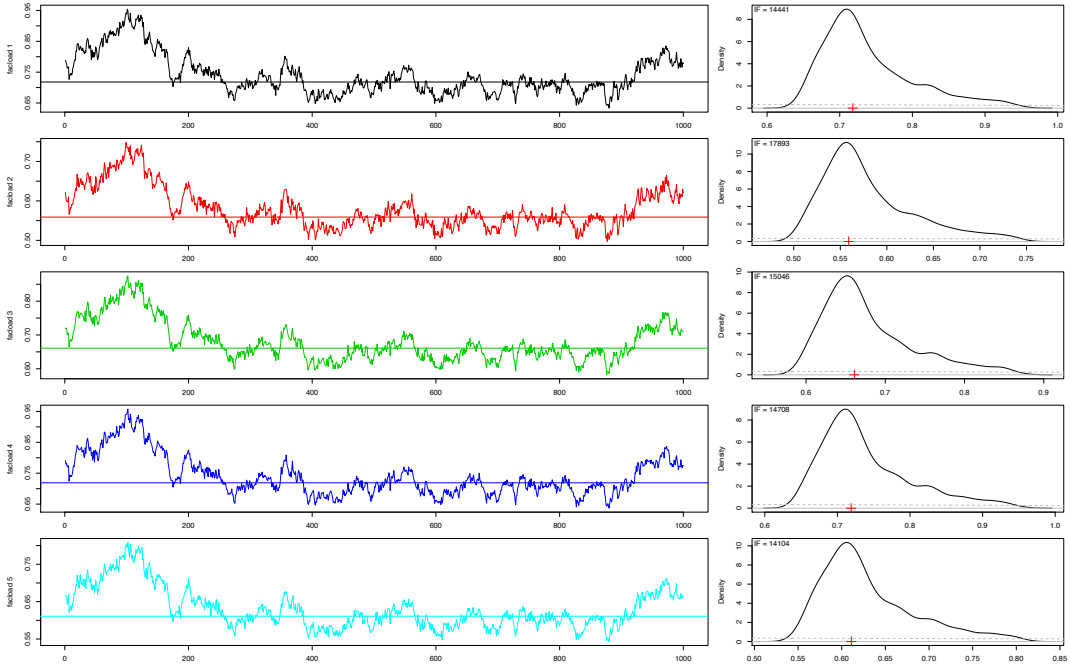
Full conditional Gibbs sampling with efficient solutions for expensive or non-standard steps:

- 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, 2017).
- Sampling the univariate SVs: efficient auxiliary mixture sampling (Kastner and Frühwirth-Schnatter, 2014), implemented in R-Package **stochvol** (Kastner, 2016).
- Everything in C/C++ using **Armadillo** (Sanderson, 2010) linked to R via **RcppArmadillo** (Eddelbuettel and Sanderson, 2014).

MORE

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, 2018; 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:

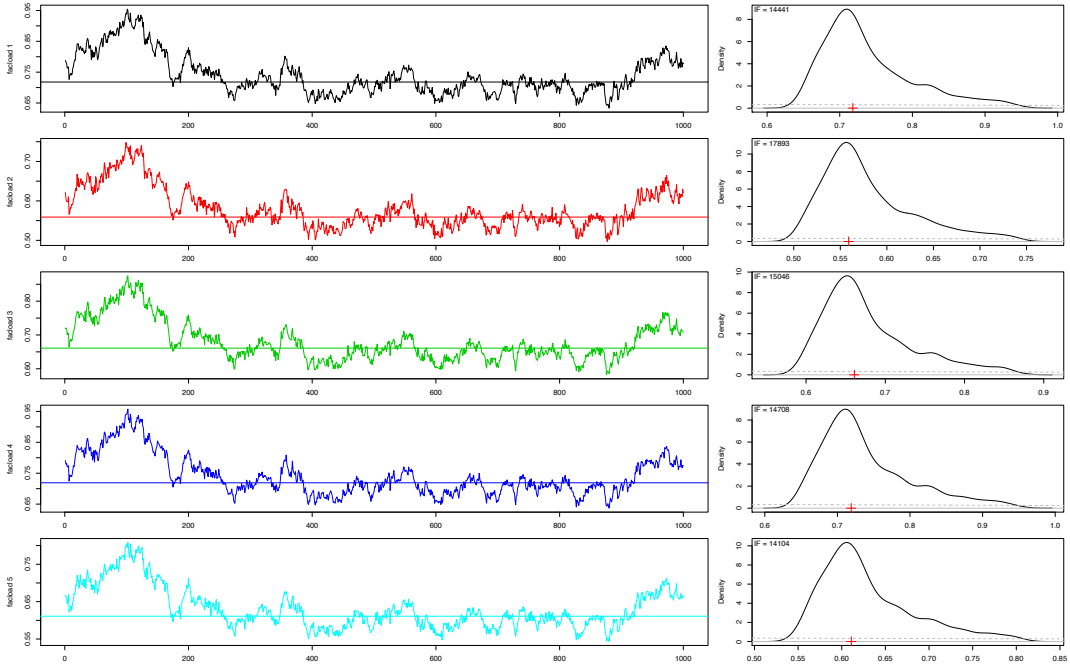
$$\mathbf{y}_t = \mathbf{\Lambda} \mathbf{f}_t + \varepsilon_t,$$

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:

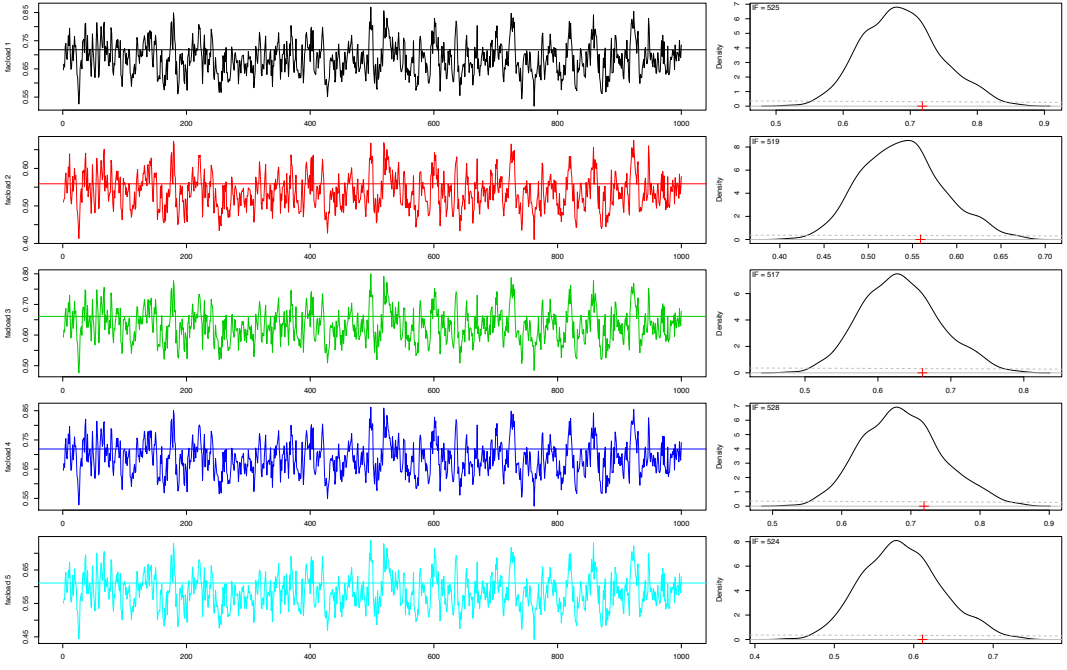
$$\mathbf{y}_t = \mathbf{\Lambda}^* \mathbf{f}_t^* + \boldsymbol{\varepsilon}_t,$$

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:

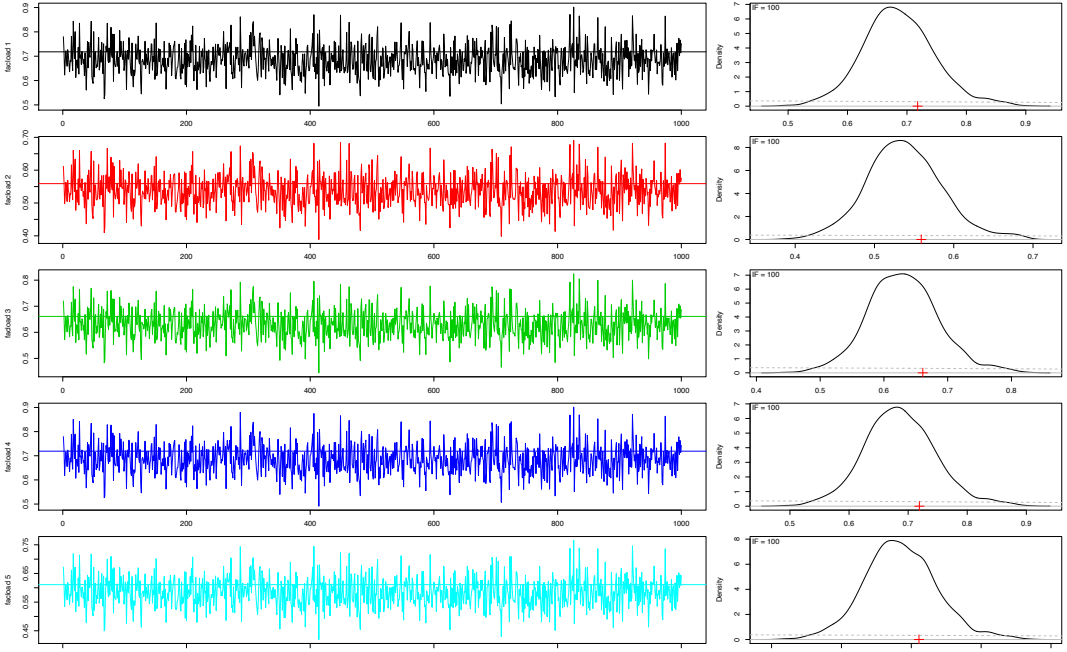
$$\mathbf{y}_t = \mathbf{\Lambda}^* \mathbf{f}_t^* + \boldsymbol{\varepsilon}_t,$$

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

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

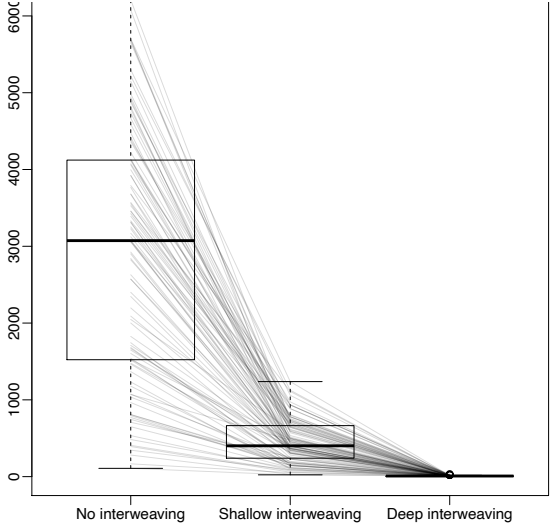
with $\mu_{m+i} = \log \Lambda_{ii}^2$.

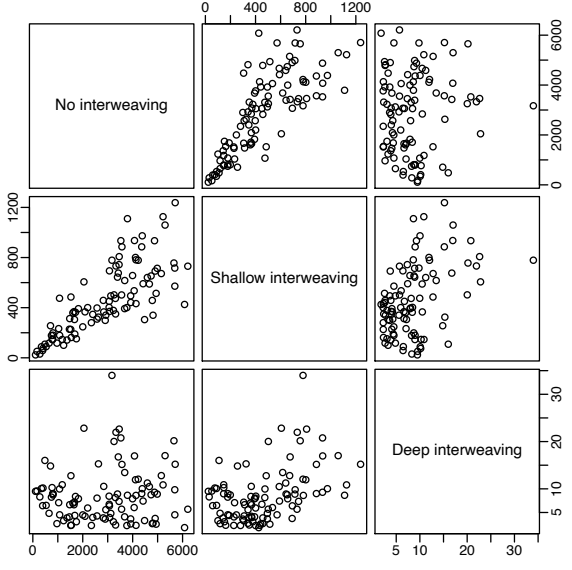
Deep interweaving



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
- due to the modular nature of MCMC, can be used as a "plug-and-play" tool for, e.g., fancy mean models
- available through preliminary **R** package **factorstochvol** (Kastner, 2017a)

More details in two papers (Kastner et al., 2017; Kastner, forthcoming)

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. 8).

Bhattacharya, A. and D. B. Dunson (2011). "Sparse Bayesian Infinite Factor Models". Biometrika 98, pp. 291–309. DOI: 10.1093/biomet/asr013 (cit. on p. 20).

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. 20).

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, 41).

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. 20, 41).

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. 38, 39).

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. 41).

Frühwirth-Schnatter, S. and H. F. Lopes (2018). Sparse Bayesian Factor Analysis when the Number of Factors is Unknown. arXiv pre-print 1804.04231. URL: https://arxiv.org/abs/1804.04231 (cit. on pp. 20, 41).

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. 41).

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. 21).

Hörmann, W. and J. Leydold (2013). "Generating Generalized Inverse Gaussian Random Variates". Statistics and Computing 24, pp. 1–11. DOI: 10.1007/s11222-013-9387-3 (cit. on pp. 38, 39).

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. (2016). "Dealing with Stochastic Volatility in Time Series Using the R Package stochvol". Journal of Statistical Software 69, pp. 1–30. DOI: 10.18637/jss.v069.i05 (cit. on pp. 38, 39).

Kastner, G. (2017a). factorstochvol: Bayesian Estimation of (Sparse) Latent Factor Stochastic Volatility Models. R package version 0.8.4. URL: https://CRAN.R-project.org/package=factorstochvol (cit. on p. 50).

Kastner, G. (2017b). stochvol: Efficient Bayesian inference for stochastic volatility (SV) models. R package version 1.3.3. URL: https://CRAN.R-project.org/package=stochvol.

Kastner, G. (forthcoming). "Sparse Bayesian Time-Varying Covariance Estimation in Many Dimensions". Journal of Econometrics. URL: https://arxiv.org/abs/1608.08468 (cit. on p. 50).

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. 38, 39, 55).

References III

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.

Kastner, G., S. Frühwirth-Schnatter, and H. F. Lopes (2017). "Efficient Bayesian Inference for Multivariate Factor Stochastic Volatility Models". Journal of Computational and Graphical Statistics 26, pp. 905–917. DOI: 10.1080/10618600.2017.1322091 (cit. on p. 50).

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. 20).

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, 55).

Leydold, J. and W. Hörmann (2017). GIGrvg: Random Variate Generator for the GIG Distribution. R package version 0.5. URL: http://CRAN.R-project.org/package=GIGrvg (cit. on pp. 38, 39).

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. 8).

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. 8, 20).

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. 8).

References IV

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. 59).

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. 41).

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. 21).

Sanderson, C. (2010). Armadillo: An Open Source C++ Linear Algebra Library for Fast Prototyping and Computationally Intensive Experiments. Tech. rep. URL: http://arma.sourceforge.net/armadillo_nicta_2010.pdf (cit. on pp. 38, 39).

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. 41).

van Dyk, D. and X.-L. Meng (2001). "The art of data augmentation". Journal of Computational and Graphical Statistics 10, pp. 1–50. DOI: 10.1198/10618600152418584 (cit. on p. 41).

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. 20).

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. 41).

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. 8, 20).

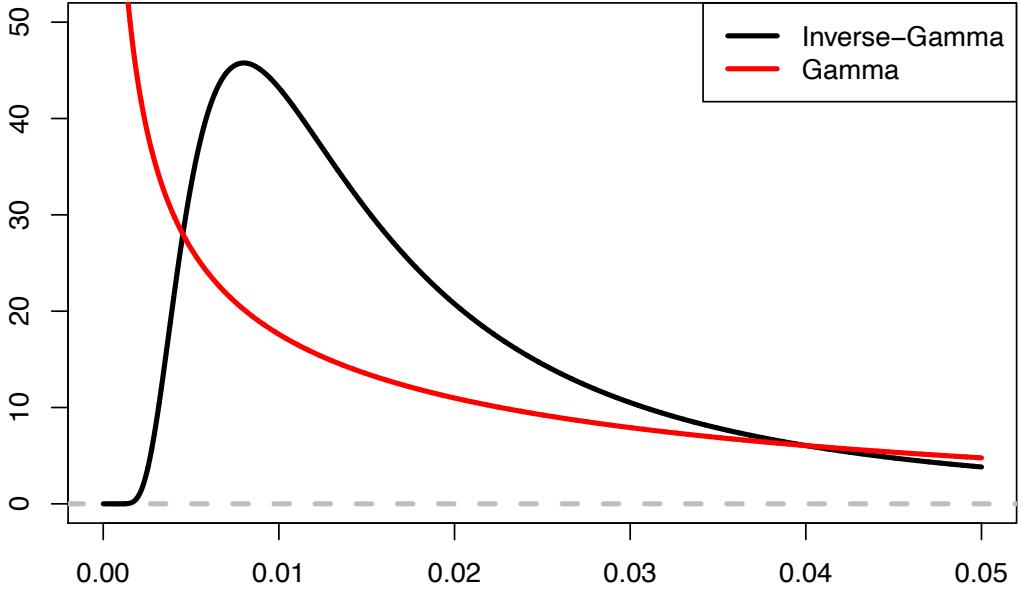
Priors

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

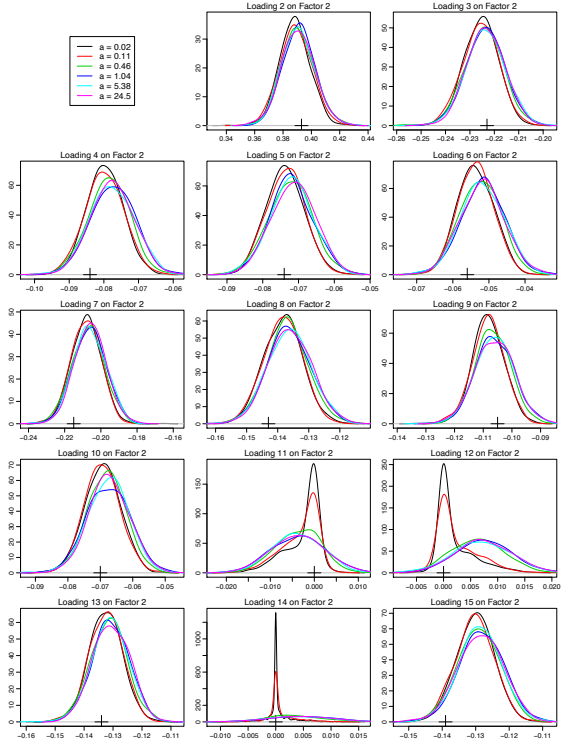
$$p(\phi) = \frac{1}{2B(a_0, b_0)} \left(\frac{1+\phi}{2}\right)^{a_0 - 1} \left(\frac{1-\phi}{2}\right)^{b_0 - 1}$$

- $\sigma^2 \sim B_\sigma \times \chi^2_1 = \mathcal{G}\left(\frac{1}{2}, \frac{1}{2B_\sigma}\right)$ (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?



BACK


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

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