Choleski Multivariate Stochastic Volatility

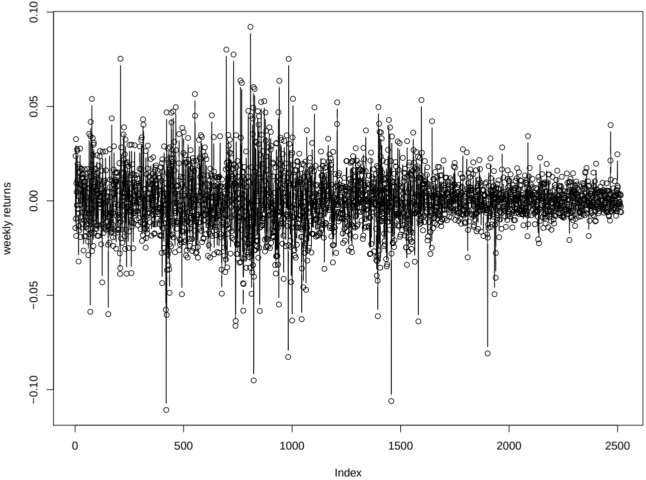
Robert McCulloch

McCombs School of Business
University of Texas

May 18, 2011

Joint with
Hedibert Lopes, Ruey Tsay
(University of Chicago, Booth School of Business)

Weekly returns on a stock in the S&P500.
$T = 2,516$.



There is variability in the variability !!!

Historically, a lot of statistics focused on “the mean”:

- anova
- regression

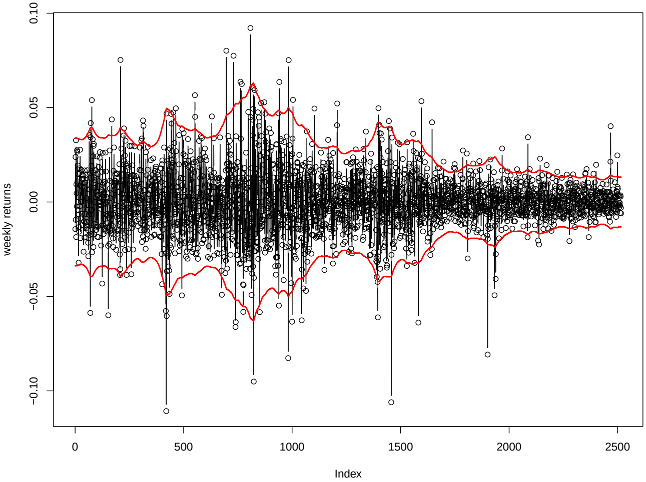
With financial data, models for the variability are of great importance.

volatility, risk , ...

We allow the variance to be *time-varying*: $Y_t \sim N(0, \sigma_t^2)$.

We assume throughout that our series has been “de-meaned” in some way.

Red is at $\pm\, 2\,\hat{\sigma}_t$.



Univariate Stochastic Volatility:

How do we get $\hat{\sigma}_t$?

$\sigma_t = e^{(d_t/2)}, \;\; d_t = log(\sigma_t^2).$

$Z_t, \epsilon_t \sim N(0,1), \text{iid}.$

$$\begin{aligned} Y_t &= e^{(d_t/2)}\, Z_t \\ d_t &= \alpha + \beta d_{t-1} + \tau\, \epsilon_t \end{aligned}$$

Observe $\{Y_t\}$, $\{d_t\}$ are the latent states.

This is a non-linear state-space model.

MCMC:

Let d_0 be the initial state, $d_0 \sim N(m_0, C_0)$.

$$\begin{array}{rcl} (d_0, \{d_t\}) & \mid & (\alpha, \beta, \tau), \{Y_t\} \\ (\alpha, \beta, \tau) & \mid & (d_0, \{d_t\}) \end{array}$$

For every draw of the $\{d_t\}$, compute the $\{\sigma_t\}$, then average to get the posterior mean.

Some important details omitted, but that is the jist of it.

References in paper on my website.
(Chib, Shephard, Carter & Kohn, Fruwirth-Schnatter).

In particular we use the wonderful FFBS (forward-filter, backward-sample) algorithm.

Now let Y_t denote a vector of time-series.

Maybe *dependence structure and volatility* change over time.

We want to do *Multivariate* Stochastic Volatility.

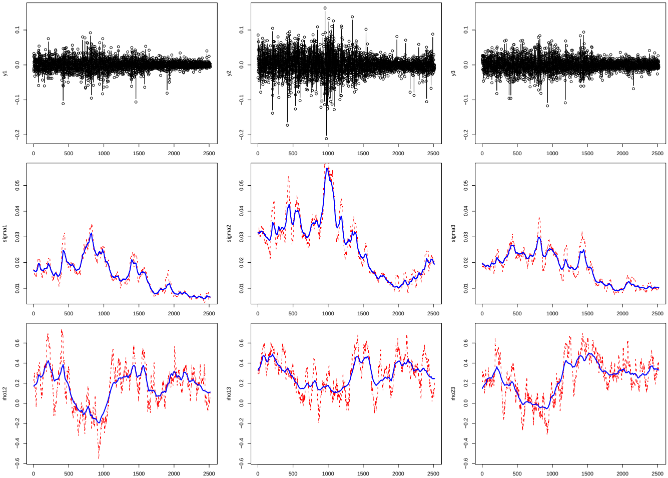
$$Y_t \sim N_p(0, \Sigma_t).$$

Problems:

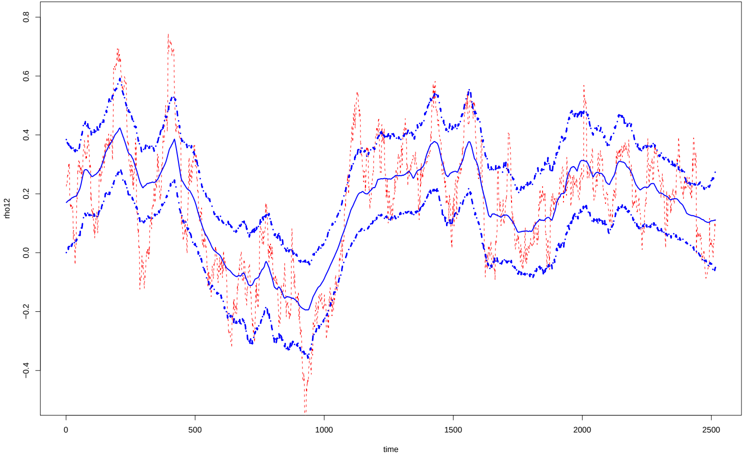
- Σ_t is a positive definite matrix.
- if p is large, $p(p+1)/2$ is very large.

$p = 50 \rightarrow p(p+1)/2 = 1275.$

p=3:
first row: data, second row: σ_{it}, third row: ρ_{ijt}.
blue are our stuff, red is sample covariance on a rolling window.



ρ_{12t}: pointsize posterior bands.



We rewrite Σ_t in terms of a series of time-varying regressions.

p=3:

$$\begin{array}{rcl r}
Y_{1t} & = & \exp(d_{t1}/2)\, Z_{1t} & p(Y_{1t}) \\
Y_{2t} & = & \phi_{21t}\, Y_{1t} + \exp(d_{2t}/2)\, Z_{2t} & p(Y_{2t} \mid Y_{1t}) \\
Y_{3t} & = & \phi_{31t}\, Y_{1t} + \phi_{32t}\, Y_{2t} + \exp(d_{3t}/2)\, Z_{t3} & p(Y_{3t} \mid Y_{2t}, Y_{1t})
\end{array}$$

With bigger p, you just keep going!

A high-dimensional non-linear state-space model.

Observe: $\{Y_{it}\}$, $i = 1, 2, \ldots, p$.

States: $\{d_{it}\}$, $i = 1, 2, \ldots, p$, $\{\phi_{ijt}\}$, $j = 1, 2, \ldots, (i-1)$.

At time t, $\Sigma_t \Leftrightarrow (d_{it}, \ldots, d_{pt}, \phi_{12t}, \ldots, \phi_{p(p-1)t})$.

State Equations:

$$d_{it} = \alpha_i + \beta_i \, d_{i(t-1)} + \tau_i \, \epsilon_t.$$

$$\phi_{ijt} = \alpha_{ij} + \beta_{ij} \, \phi_{ij(t-1)} + \tau_{ij} \, \epsilon_t.$$

Priors:

$$p(d_{i0}), \;\; p(\phi_{ij0}), \;\; p(\alpha_i, \beta_i, \tau_i), \;\; p(\alpha_{ij}, \beta_{ij}, \tau_{ij}).$$

MCMC:

Let $\circ$ denote "everything else"
(all other parameters and all the data).

$$\begin{aligned}(d_{i0}, \{d_{it}\}_{t=1}^{T}) &\mid \circ \\ (\phi_{ij0}, \{\phi_{ijt}\}_{t=1}^{T}) &\mid \circ \\ (\alpha_i, \beta_i, \tau_i) &\mid \circ \\ (\alpha_{ij}, \beta_{ij}, \tau_{ij}) &\mid \circ\end{aligned}$$

You just draw each state sequence and the associated AR1 parameters one at a time.

$$\begin{array}{rcl} Y_{1t} & = & \exp(d_{t1}/2)\, Z_{1t} \\ Y_{2t} & = & \phi_{21t}\, Y_{1t} + \exp(d_{2t}/2)\, Z_{2t} \\ Y_{3t} & = & \phi_{31t}\, Y_{1t} + \phi_{32t}\, Y_{2t} + \exp(d_{3t}/2)\, Z_{t3} \end{array}$$

$(d_{i0}, \{d_{it}\}_{t=1}^{T}) \mid \circ :$

$i = p = 3$:

$$\tilde{Y}_t = Y_{3t} - \phi_{31t}\, Y_{1t} - \phi_{32t}\, Y_{2t} = \exp(d_{3t}/2)\, Z_{t3}$$

All $(d_{i0}, \{d_{it}\})$ can be drawn as a univariate stochastic volatility.

$$\begin{array}{rcl} Y_{1t} & = & \exp(d_{t1}/2)\, Z_{1t} \\ Y_{2t} & = & \phi_{21t}\, Y_{1t} + \exp(d_{2t}/2)\, Z_{2t} \\ Y_{3t} & = & \phi_{31t}\, Y_{1t} + \phi_{32t}\, Y_{2t} + \exp(d_{3t}/2)\, Z_{t3} \end{array}$$

$(\phi_{ij0}, \{\phi_{ijt}\}_{t=1}^{T}) \mid \circ :$

$i = p = 3, j = 2$:

$$\tilde{Y}_t = Y_{3t} - \phi_{31t}\, Y_{1t} = \phi_{32t}\, Y_{2t} + \exp(d_{3t}/2)\, Z_{t3}$$

All $(\phi_{ij0}, \{\phi_{ijt}\})$ can be drawn as in a simple dynamic linear model (DLM).

This draw is done very simply using FFBS.

AR coefficients:

$$(\alpha_i, \beta_i, \tau_i) \,|\, \circ = (\alpha_i, \beta_i, \tau_i) \,|\, (d_{i0}, \{d_{it}\})$$

$$(\alpha_{ij}, \beta_{ij}, \tau_{ij}) \,|\, \circ = (\alpha_{ij}, \beta_{ij}, \tau_{ij}) \,|\, (\phi_{ij0}, \{\phi_{ijt}\})$$

Now I can tell you what the paper is about!!

Want to do this for "large" p.

$p = 50$: there are $p(p-1)/2 = 1,225$ ϕ sequences to draw.

Two key ideas:

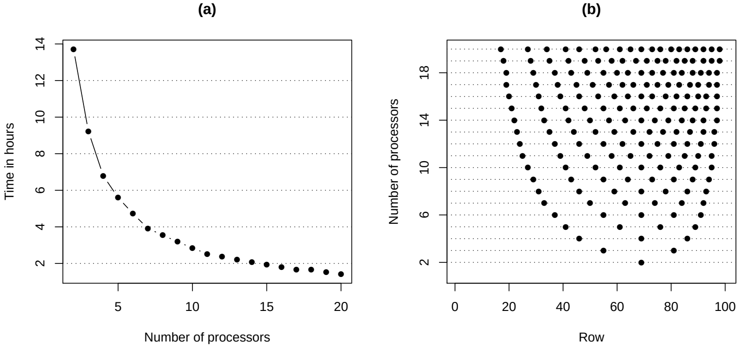
- Draws are done equation by equation.
  Can do blocks of equations separately using parallel computing.
- Need very non-standard, highly informative prior on the (α, β, τ).

Parallel Computing:

(a): time to compute as a function of the number of processors.
(b): how to allocate equations across processors.

Note: takes 10 times longer to run a univariate SV than a DLM (a d than a ϕ).

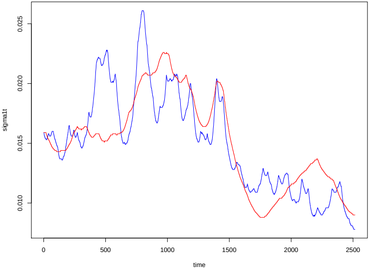
$p = 100$.



Prior:

With so many things going on, the prior is inevitably influential. It is more like part of the model.

Here is the posterior mean of $\{\sigma_{1t}\}$ from two different priors.



$$d_{it} = \alpha_i + \beta_i \, d_{i(t-1)} + \tau_i \, \epsilon_t .$$

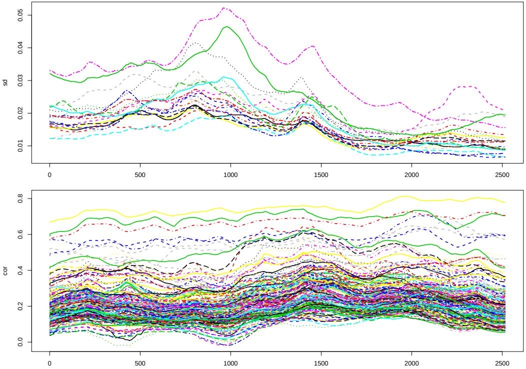
Clearly, prior on τ is a key.

The prior on the (α, β, τ) for the ϕ series is even more important than for the *d*.

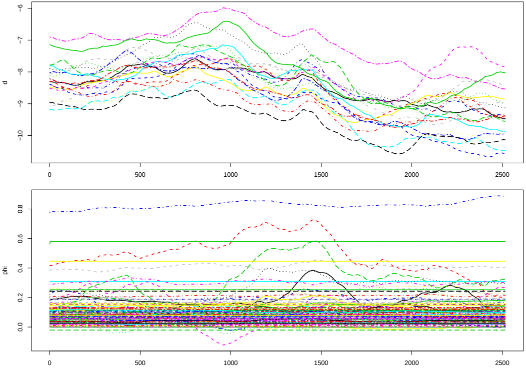
There are many more of them *and* the data is less informative about them.

Example: p=20:

top: all the time varying standard deviations.
bot: all the time varying correlations.



top: all the time varying d.
bot: all the time varying ϕ.



many of the ϕ, flat - line !!!

Key Idea in Bayesian statistics:

> *Prior guides search for parsimony in high dimensional systems !!!!*

Here, “parsimony” means ϕ that don’t change much and ϕ that are close to 0.

Prior Specification:

Start by rescaling each series thinking $\Sigma_t \approx I$.

This can be as simple as subtracting off the sample means and dividing by the sample standard deviations.

Have to be able to specify just about any kind of prior we want and draw (α, β, τ) jointly.

But (β, τ) on a bi-variate grid (grid size 100 works).

Then use

$$p(\alpha, \beta, \tau) = p(\beta, \tau)\, p(\alpha \mid \beta, \tau)$$

with $p(\alpha \mid \beta, \tau)$ normal.

$$\phi_t = \alpha + \beta\,\phi_{t-1} + \tau\,\epsilon_t.$$

$$\alpha \,|\, \beta \sim N(0, \sigma_\alpha^2(1 - \beta^2)).$$

When $\beta \approx 0$, the state flat-lines, and α is the level.

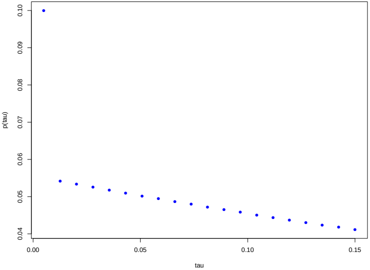
When $\beta \approx 1$ we are close to a random-walk, want $\alpha \approx 0$.

Because we can integrate out α, the draw from the posterior can be done using

$$p(\alpha, \beta, \tau \mid \circ) = p(\beta, \tau \mid \circ)\, p(\alpha \mid \beta, \tau, \circ)$$

where the first draw is on a grid, and the second draw is just a normal.

τ prior:



Pick the probability of the smallest value.

For the other values:

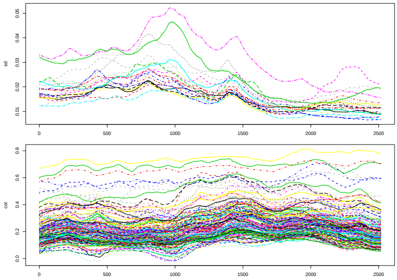
$$p(\tau) \propto \exp(-c|\tau - \tau_{min}|)$$

Push towards small value of τ = smooth state.

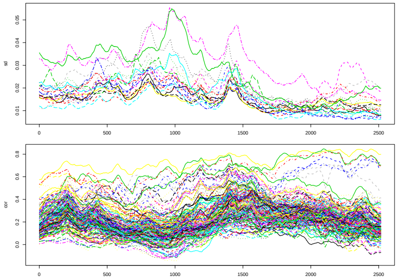
Here, $c = 2$, in practice, $c = 100$.

Prior Parameters: $\tau_{min}, \tau_{max}, p(\tau = \tau_{min}), c$.

$c = 200$.


sd
0.01
0.02
0.03
0.04
0.05
cor
0.0
0.2
0.4
0.6
0.8
0
500
1000
1500
2000
2500

$c = 100$.


sd
0.01
0.02
0.03
0.04
0.05
cor
0.0
0.2
0.4
0.6
0.8
0
500
1000
1500
2000
2500

$$\phi_t = \alpha + \beta\,\phi_{t-1} + \tau\,\epsilon_t.$$

The full prior on (α, β, τ) mixes over configurations of interest.

$$p(\alpha,\beta,\tau) = p_{01}\,\delta_{\{\alpha=0,\beta=1\}}\,p(\tau \mid \beta = 1) + p_{00}\,\delta_{\{\alpha=0,\beta=0\}}\,p(\tau \mid \beta = 0) + \\ p_{u0}\,\delta_{\{\beta=0\}}\,p(\tau \mid \beta = 0)\,p(\alpha \mid \beta = 0) + p_{uu}\,p(\beta)p(\tau \mid \beta \neq 0)\,p(\alpha \mid \beta).$$

- p_{01}: prob of random walk.
- p_{00}: prob of flat-line at 0.
- p_{u0}: prob of flat-line, not at 0.
- p_{uu}: prob $\beta \in (0, 1)$.
- when $\beta = 0$, may want an even smaller τ.

To illustrate the mixture prior, we simulate from the simple DLM.

$$\begin{aligned} Y_t &= x_t\, s_t + Z_t, \\ s_t &= \alpha + \beta\, s_{t-1} + \tau\, \epsilon_t. \end{aligned}$$

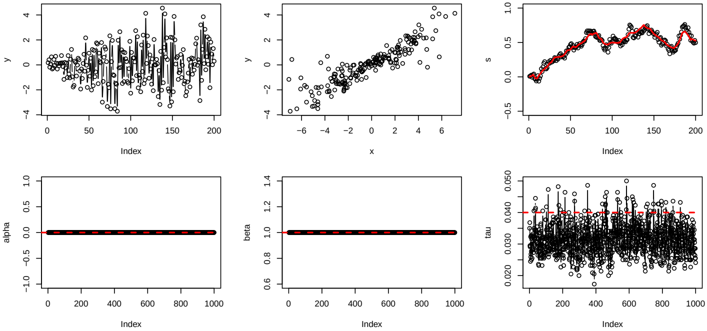
Gibbs:

$$(s_0, \{s_t\}) \mid (\alpha, \beta, \tau), (y, x), \quad (\alpha, \beta, \tau) \mid (s_0, \{s_t\}), (y, x)$$

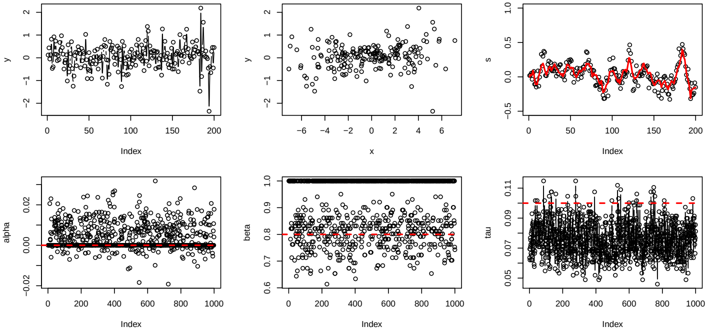
prior 0:
$p_{00} = .15, p_{u0} = .15, p_{01} = .5, \tau_{max} = .15, c_\tau = 100.$

prior 1:
$p_{00} = .05, p_{u0} = .05, p_{01} = .85, \tau_{max} = .05, c_\tau = 200.$

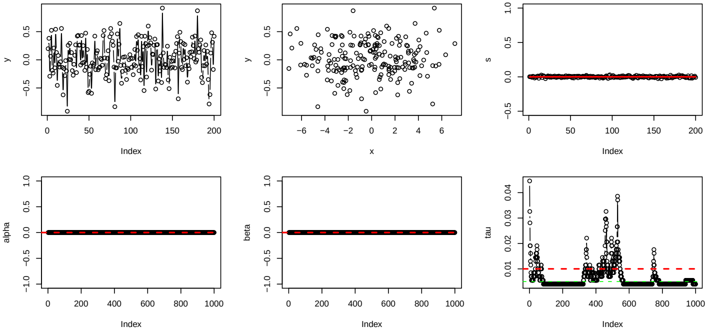
Using prior 1.



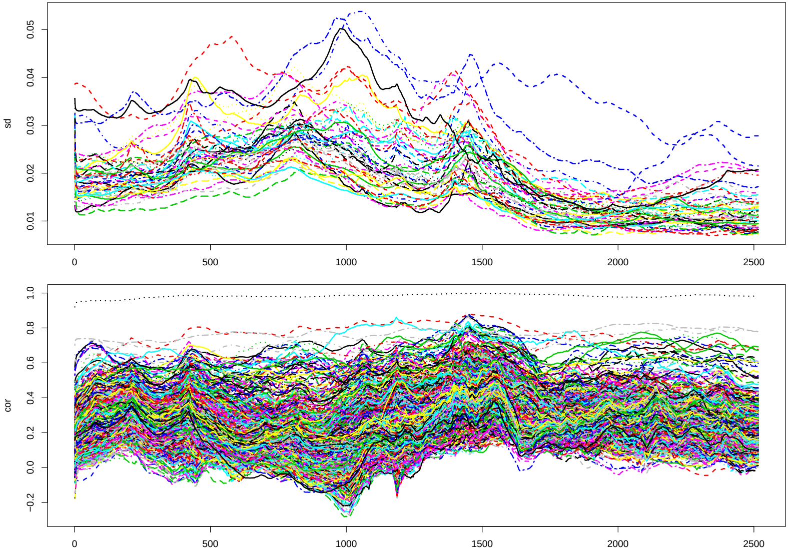
Using prior 0.



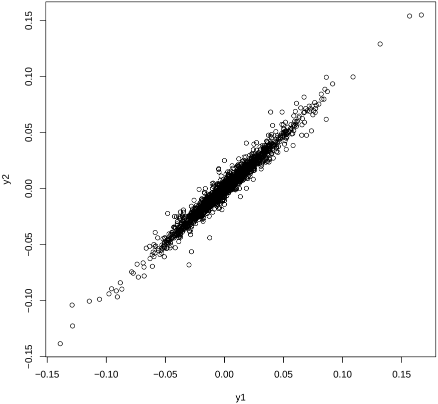
Using prior 0.



$p = 50$:

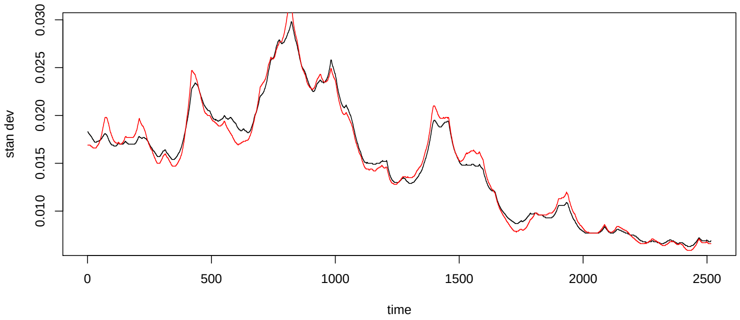

sd
0.01
0.02
0.03
0.04
0.05
0
500
1000
1500
2000
2500
cor
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000
2500

This is the scatter-plot of the pair that have the high and near constant correlation.



Run $p = 20$ with one order, then reverse the order.

This is the standard deviation of the first (last) series.



Portfolio Weights:

For larger p it is quite difficult to display the fit of the model.

In order to illustrate the fit and use of the model, consider a simple application in the case where all of our series are asset returns.

Let w_t denote the portfolio weights of the *global minimum variance portfolio*. That is, w minimizes $w' \, \Sigma \, w$ subject to the constraint that $\sum w_i = 1$.

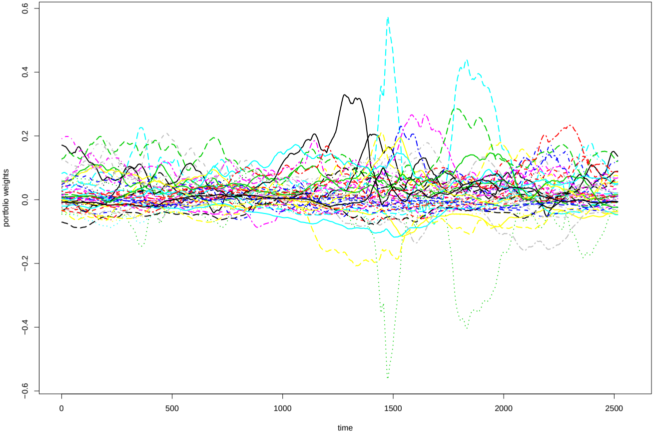
$$w(\Sigma) = \frac{\Sigma^{-1} \, \iota}{\iota' \, \Sigma^{-1} \, \iota}$$

where ι is a vector of ones.

$p = 3$:


portfolio weight
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000
2500
time

$p = 50$:



portfolio weights
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
0
500
1000
1500
2000
2500
time

Concluding Remarks:

This is a hard problem !!!!

Mixture prior gives you a lot of flexibility.

Can get a constant state with $\beta = 1$ or $\beta = 0$, so the prior identifies which one you get.

Could order the series so that the first so many span the space of factors and then adjust the priors so that it is easier to load on these factors.

Bayesian setup allows us to adjust to outliers, jumps, ...