Shrinkage Estimation for Multivariate Hidden Markov Mixture Models

Rainer von Sachs
ISBA, Université catholique de Louvain

Recent Advances in Changepoint Analysis
University of Warwick,
March 26-28, 2012

Joint work with M. Fiecas (UCSD), J. Franke and J. Tadjuidje (University of Kaiserslautern)

Outline

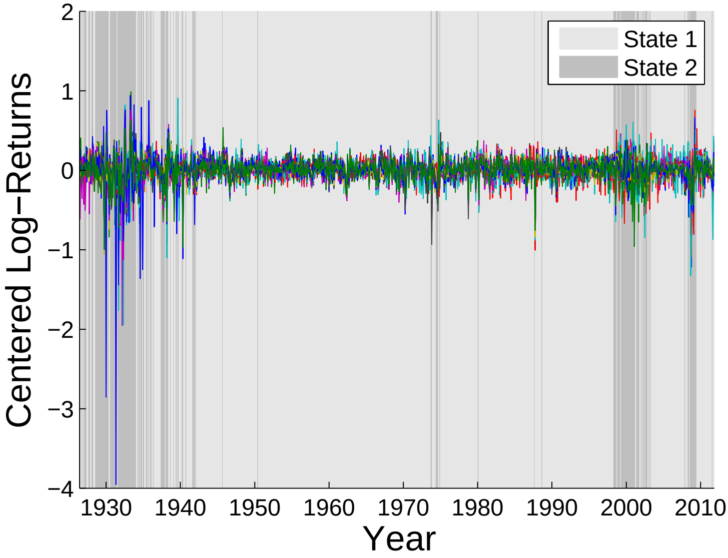
Motivating data analysis

Returns of US Industry Portfolio

- Monthly returns of $P = 30$ different industry sectors (NYSE, NASDAQ and AMEX)
- $T = 1026$ time points between July 1926 and December 2011
- Characteristics: changing market environment over time
- Model: Hidden (Markov) states based on pseudo-Gaussian likelihood

$$X_t = \sum_{k=1}^{K} S_{t,k} \, \boldsymbol{\Sigma}_k^{1/2} \varepsilon_t \ ,$$

US Industry Portfolio



Challenge

- Variance-covariance analysis of (mean-centered) log-returns
- Due to high dimensionality, sample covariance matrix (30×30) possibly not invertible and numerically unstable
- In this work: method to improve this estimator by **shrinkage**

Shrinkage for covariance estimation

- Developed by Ledoit et al (2004), Sancetta (2008)
- Can drastically improve the mean-squared error and the condition number
- (Theory: double asymptotics $P \to \infty$ and $T \to \infty$)
- Our contribution: Use for Hidden Markov mixtures

Multivariate Hidden Markov Mixture

Let S_t be a finite-state Markov chain with values in $\{e_1, \ldots, e_K\}$, where e_i is a unit vector in $\mathbb{R}^K$ and having the *ith* entry equal to 1.

$$P(S_t = e_j \mid S_{t-1} = e_i, S_{t-2}, \cdots) \quad = \quad P(S_t = e_j \mid S_{t-1} = e_i) \; = \; a_{ij}$$

Let assume

1. S_t aperiodic and irreducible, and
2. S_t is α-mixing with exponentially decreasing rate.

Then, S_t stationary with distribution given by $\pi_k = P(S_t = e_k)$.
We define,

$$X_t = \sum_{k=1}^{K} S_{t,k}(\mu_k + \mathbf{\Sigma}_k^{1/2}\varepsilon_t) \ , \tag{1}$$

where ε_t i.i.d. $(0, \mathbf{I}_p)$, independent of S_t and $S_{t,k} = 1$ iff $S_t = e_k$.
For ease of presentation: $\mu_k = 0 \; \forall \; k$.

Some literature: Francq and Roussignol (1997), Yang (2000): Switching Markov VAR; Francq and Zarkoïan (2001): Multivariate Markov switching ARMA; Franke et al (2010, 2011): Mixtures of nonparametric AR, Markov switching AR-ARCH.

Shrinkage Estimation of Covariance Matrices

Goal: Estimate $\mathbf{\Sigma}$ by more "regular" estimator than empirical covariance

$$\widehat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} X_t X_t'$$

Idea (Ledoit et al, 2004, Sancetta, 2008): Shrinkage of $\widehat{\Sigma}$:

$$\widehat{\Sigma}_s \quad = \quad (1 - W)\,\widehat{\Sigma} \; + \; W\;\alpha\mathbf{I}_p \text{ with } 0 \le W \le 1. \tag{2}$$

Shrink $\widehat{\Sigma}$ towards $\alpha\mathbf{I}_p$ such that $\mathrm{tr}(\alpha\mathbf{I}_p) = \mathbb{E}\,\mathrm{tr}\,\widehat{\Sigma} \approx \mathrm{tr}\,\mathbf{\Sigma}$.

Interesting: extreme eigenvalues are shrunken towards the "grand mean" $\dfrac{\mathrm{tr}\,\widehat{\Sigma}}{p}$.

Although bias is introduced, variance is highly reduced (for p large), and the MSE is reduced.

Optimal shrinkage weights

Choose optimal shrinkage weight W by minimizing MSE:

$$W^* = \arg \min_{W \in [0,1]} \mathbb{E}||\widehat{\Sigma}_s - \mathbf{\Sigma}||^2$$

where $||\mathbf{A}||^2 = \frac{1}{p}\,\mathrm{tr}(\mathbf{A}\mathbf{A}')$, the **scaled Frobenius norm**. The solution is

$$W^* = \frac{\mathbb{E}||\widehat{\Sigma} - \mathbf{\Sigma}||^2}{\mathbb{E}||\alpha I_p - \widehat{\Sigma}||^2} \wedge 1. \tag{3}$$

But: $\widehat{\Sigma}_s^*$ is not feasible as depending on unknowns, in particular $\alpha = \frac{1}{p}\,\mathrm{tr}(\mathbf{\Sigma})$.

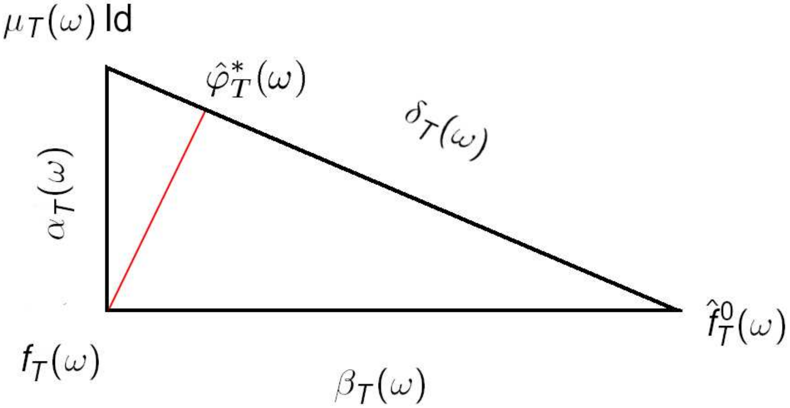
Interpretations:
W^* is also the PRIAL ("percentage relative improvement of average loss") of $\widehat{\Sigma}_s^*$ over $\widehat{\Sigma}$. It shows that (even under - correct - double asymptotics p and T to infinity) shrinkage is important.

Also, we have a Pythagorean for shrinkage (asymptotically, if biased $\widehat{\Sigma}$):

$$\alpha^2 + \beta^2 = \delta^2 ,$$

- $\alpha^2 =$ distance between "truth" and shrinkage target
- $\beta^2 =$ distance between "truth" and (unshrunken) estimator
- $\delta^2 =$ distance between (unshrunken) estimator and shrinkage target

A Pythagorean for Shrinkage



Shrinkage Estimation of Covariance Matrices (2)

How to estimate consistently the optimal shrinkage weight W^*?

$$W^* = \frac{\mathbb{E}||\widehat{\Sigma} - \mathbf{\Sigma}||^2}{\mathbb{E}||\alpha I_p - \widehat{\Sigma}||^2} \wedge 1. \tag{4}$$

Estimate

- α by $\widehat{\alpha} = \frac{1}{p}\,\mathrm{tr}(\widehat{\Sigma})$
- denominator by sample analogue $||\alpha I_p - \widehat{\Sigma}||^2$
- numerator by some less direct alternative approach which works even for correlated data X_t, suggested by Sancetta (2008):

Note that

$$\mathbb{E}||\widehat{\Sigma} - \mathbf{\Sigma}||^2 = \frac{1}{p}\sum_{i,j}^{p} \mathrm{var}(\widehat{\Sigma}_{ij}) = \frac{1}{p}\sum_{i,j}^{p} \frac{1}{T} f_{ij}(0) \ ,$$

where f_{ij} is the spectral density of (the time series) $Y_t^{ij} := X_{ti}\ X_{tj}$.
Estimate $f(\omega)(0)$ via some lag-window smoother over the empirical autocovariances of Y_t^{ij}.

The shrinkage estimator $\hat{\mathbf{\Sigma}}_k^s$ (based on *estimated* optimal shrinkage weights) asymptotically has the same propoerties as the optimal shrinkage estimator $\mathbf{\Sigma}^{s,*}$, improving the PRIAL over $\widehat{\Sigma}$.

Also: both estimators improve the *condition number* of the matrix $\widehat{\Sigma}$ defined to be the ratio of its largest to its smallest eigenvalue.

Covariance estimators built via oracle states

Paradigm: Ignore that state variables S_t are unknown, develop covariance shrinkage estimators. Consider

$$\widetilde{\boldsymbol{\Sigma}}_k^{(0)} = \frac{1}{\sum_{t=1}^T S_{t,k}} \sum_{t=1}^T S_{t,k} X_t X_t' = \frac{1}{T_k} \sum_{t=1}^T S_{t,k} X_t X_t' \quad \text{Oracle State Covariance estimator}$$

Problem of numerical stability if the effective sample size $\sum_{t=1}^T S_{t,k} = T_k$ is not large enough compared to the dimension of the process p.
Way out: work with

$$\boldsymbol{\Sigma}_k^{(0)} = \frac{1}{T} \sum_{t=1}^T S_{t,k} X_t X_t' = \pi_k^{(0)} \widetilde{\boldsymbol{\Sigma}}_k^{(0)},$$

with

$$\pi_k^{(0)} = \frac{1}{T} \sum_{t=1}^T S_{t,k},$$

Observe that $\boldsymbol{\Sigma}_k^{(0)}$ estimates

$$\mathbb{E}\, \boldsymbol{\Sigma}_k^{(0)} = \pi_k \boldsymbol{\Sigma}_k$$

but it is still potentially ill-conditioned: **Shrinkage**

Shrinkage based Oracle State Covariance Estimator

Goal: Shrink $\mathbf{\Sigma}_k^{(0)}$ towards $\alpha_k \mathbf{I}_p$, with $\alpha_k = \frac{1}{p}\operatorname{tr}(\pi_k \mathbf{\Sigma}_k)$ which can be unbiasedly and $\sqrt{T}-$consistently estimated by

$$\alpha_k^{(0)} = \frac{1}{p} \operatorname{tr} \mathbf{\Sigma}_k^{(0)} .$$

Let

$$\mathbf{\Sigma}_k^s = (1 - W_k)\, \mathbf{\Sigma}_k^{(0)} + W_k\, \alpha_k^{(0)} \mathbf{I}_p \quad \text{with } 0 \leq W_k \leq 1.$$

and optimize the $\{W_k\}$ by

$$\tilde{W}_k^{(0)} = \arg \min_{W_k \in [0,1]} \mathbb{E}||(1 - W_k)\, \mathbf{\Sigma}_k^{(0)} + W_k\, \alpha_k^{(0)} \mathbf{I}_p - \pi_k \mathbf{\Sigma}_k||^2.$$

We get

$$\begin{aligned} \tilde{W}_k^{(0)} &= \frac{\mathbb{E}||\mathbf{\Sigma}_k^{(0)} - \pi_k \mathbf{\Sigma}_k||^2 - \sum_{i=1}^{p} \operatorname{cov}(\operatorname{tr} \mathbf{\Sigma}_{ii,k}^{(0)}, \alpha_k^{(0)})}{\mathbb{E}||\alpha_k^{(0)} \mathbf{I}_p - \mathbf{\Sigma}_k^{(0)}||^2} \wedge 1 \\ &\approx \frac{\mathbb{E}||\mathbf{\Sigma}_k^{(0)} - \pi_k \mathbf{\Sigma}_k||^2}{\mathbb{E}||\alpha_k^{(0)} \mathbf{I}_p - \mathbf{\Sigma}_k^{(0)}||^2} \wedge 1 \end{aligned}$$

using that

$$\sum_{i=1}^{p} \operatorname{cov}(\operatorname{tr} \mathbf{\Sigma}_{ii,k}^{(0)}, \alpha_k^{(0)}) = \frac{1}{p} \operatorname{var}(\operatorname{tr} \mathbf{\Sigma}_k^{(0)}) = p\, \mathbb{E}(\alpha_k^{(0)} - \alpha_k)^2 = O\left(\frac{p}{T}\right) .$$

Estimation of the shrinkage weights

As in Sancetta (2008), estimate the numerator

$$\mathbb{E}||\mathbf{\Sigma}_k^{(0)} - \pi_k \mathbf{\Sigma}_k||^2$$

by an estimator of the spectral density $f_{ij}(\omega)$ of $Y_t^{ij,k} = S_{tk} X_{ti} X_{tj}$ at frequency $\omega = 0$.
Reminder: Use lag-window smoothed empirical autocovariances.
Indeed, for $K(u) \geq 0$, $K(u) = K(-u)$ and $K(0) = 1$ and some $b > 0$

$$\hat{f}_{ij,k}^{b}(0) = \sum_{s=-T+1}^{T-1} K\left(\frac{s}{b}\right) \Gamma_{ij,k}^{(0)}(s).$$

Then, we get an estimator of the optimal shrinkage weights $W_k^{(0)}$ as follows:

$$\hat{W}_k^{(0)} = \frac{\frac{1}{p}\frac{1}{T}\sum_{i,j=1}^{p} \hat{f}_{ij,k}^{b}(0)}{\left\|\mathbf{\Sigma}_k^{(0)} - \alpha_k^{(0)} \mathbf{I}_p\right\|^2} \wedge 1.$$

Some asymptotic theory

Theorem

Under the above assumptions on S_t, with $\mathbb{E}\|\varepsilon_t\|^8 < \infty$, a kernel $K(u)$ with "usual" poperties and a bandwidth $b = b_T \to \infty$ such that $\frac{b_T}{\sqrt{T}} \to 0$. Moreover, assume, with p fixed,

A1) $\alpha_k \mathbf{I}_p \neq \pi_k \boldsymbol{\Sigma}_k$

Then

a) $W_k^* = \dfrac{\mathbb{E}||\boldsymbol{\Sigma}_k^{(0)} - \pi_k \boldsymbol{\Sigma}_k||^2}{\mathbb{E}||\alpha_k \mathbf{I}_p - \boldsymbol{\Sigma}_k^{(0)}||^2} \wedge 1 \asymp \dfrac{1}{T}$

b) $\left(\hat{W}_k^{(0)} - W_k^*\right) = o_p(T^{-1})$

c) $\left\|\hat{\boldsymbol{\Sigma}}_k^s - \pi_k \boldsymbol{\Sigma}_k\right\| = \|\boldsymbol{\Sigma}_k^{s,*} - \pi_k \boldsymbol{\Sigma}_k\| \, (1 + o_p(\frac{1}{\sqrt{T}}))$

That is, the shrinkage estimator $\hat{\boldsymbol{\Sigma}}_k^s$ based on estimated optimal weights $\hat{W}_k^{(0)}$ is asymptotically as performant as $\boldsymbol{\Sigma}_k^{s,*}$ based on true optimal weights W_k^* (and both reduce the risk of the unshrunken estimator $\boldsymbol{\Sigma}_k^{(0)}$) .

Variations of this Theorem

Asymptotically growing dimension $p = p(T) \to \infty$:

A2) $\frac{1}{p^\gamma} \|\alpha_k \mathbf{I}_p - \pi_k \boldsymbol{\Sigma}_k\|^2 \to c > 0$ for some $2 > \gamma > 0$ such that $\frac{p^{2-\gamma}}{T} \to 0$.

Then, with $a_T = \frac{T}{p^{2-\gamma}}$

a) $W_k^* = \frac{\mathbb{E}||\boldsymbol{\Sigma}_k^{(0)} - \pi_k \boldsymbol{\Sigma}_k||^2}{\mathbb{E}||\alpha_k \mathbf{I}_p - \boldsymbol{\Sigma}_k^{(0)}||^2} \wedge 1 \asymp \frac{p^{2-\gamma}}{T}$

b) $a_T \left(\hat{W}_k^{(0)} - W_k^*\right) = o_p(1)$

c) $\left\|\hat{\boldsymbol{\Sigma}}_k^s - \pi_k \boldsymbol{\Sigma}_k\right\| = \|\boldsymbol{\Sigma}_k^{s,*} - \pi_k \boldsymbol{\Sigma}_k\| \left(1 + o_p\left(\frac{1}{\sqrt{a_T}}\right)\right)$

Interpretation: Everything is scaled with the amount of cross-covariance converging to zero with $p \to \infty$.

If $\mu_k \neq 0$: Estimator $\widehat{\mu}_k = \frac{\sum_t S_{tk} X_{tk}}{\sum_t S_{tk}}$ leads to additional but asymptotically vanishing bias.

Oracle estimators: Example with two hidden states

Let

$$S_t \in \{0,1\}, \varepsilon_t \sim \mathcal{N}(0, \mathbf{I}_p)$$

and define

$$X_t = (S_t \mathbf{\Sigma}_1^{1/2} + (1 - S_t)\mathbf{\Sigma}_2^{1/2})\varepsilon_t$$

Additionally,

$$\begin{aligned} \mathbf{A} &= (a_{ij})_{1 \le i,j \le 2} \\ &= \begin{pmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{pmatrix} \end{aligned}$$

Furthermore

$$p = 20 \qquad T = 256$$

Let

$$\mathbf{R}_1 = \begin{pmatrix} 1 & .2 & .2 & .2 \\ .2 & 1 & .2 & .2 \\ .2 & .2 & 1 & .2 \\ .2 & .2 & .2 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{R}_2 = \begin{pmatrix} 1 & .5 & .5 & .5 \\ .5 & 1 & .5 & .5 \\ .5 & .5 & 1 & .5 \\ .5 & .5 & .5 & 1 \end{pmatrix}.$$

and construct the block-diagonal correlation matrix

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 & \mathbf{0}_4 & \cdots & \cdots & \mathbf{0}_4 \\ \mathbf{0}_4 & \mathbf{R}_2 & \ddots & & \vdots \\ \vdots & \ddots & \mathbf{R}_1 & \ddots & \vdots \\ \vdots & & \ddots & \mathbf{R}_2 & \mathbf{0}_4 \\ \mathbf{0}_4 & \cdots & \cdots & \mathbf{0}_4 & \mathbf{R}_1 \end{pmatrix}$$

We then consider

$$\mathbf{B} = \sqrt{5}\mathbf{R}, \quad \mathbf{D} = \sqrt{10}\mathbf{I}_{20} \quad \text{and } \mathbf{H} = \begin{pmatrix} \sqrt{1}\mathbf{I}_5 & \mathbf{0}_5 & \cdots & \mathbf{0}_5 \\ \mathbf{0}_5 & \sqrt{2}\mathbf{I}_5 & \ddots & \vdots \\ \vdots & \ddots & \sqrt{3}\mathbf{I}_5 & \mathbf{0}_5 \\ \mathbf{0}_5 & \cdots & \mathbf{0}_5 & \sqrt{5}\mathbf{I}_5 \end{pmatrix}$$

1. Distribution of estimated shrinkage weights:

Simulation 1		Simulation 2		Simulation 3	
$\mathbf{\Sigma}_1 = \mathbf{H}$	$\mathbf{\Sigma}_2 = \mathbf{D}$	$\mathbf{\Sigma}_1 = \mathbf{B}$	$\mathbf{\Sigma}_2 = \mathbf{D}$	$\mathbf{\Sigma}_1 = \mathbf{B}$	$\mathbf{\Sigma}_2 = \mathbf{H}$

Mode of estimated shrinkage weight

in State 1/ State 2	in State 1/ State 2	in State 1/ State 2
0.65 / 1.0	0.3 / 1.0	0.3 / 0.65

2. Improvement for Average Loss:

$$\text{PRIAL}(\hat{\mathbf{\Sigma}}_k^s) = 100 \times \frac{\mathbb{E}||\mathbf{\Sigma}_k^{(0)} - \pi_k \mathbf{\Sigma}_k||^2 - \mathbb{E}||\hat{\mathbf{\Sigma}}_k^s - \pi_k \mathbf{\Sigma}_k||^2}{\mathbb{E}||\mathbf{\Sigma}_k^{(0)} - \pi_k \mathbf{\Sigma}_k||^2}. \tag{5}$$

which yields

	State	Covariance Matrix	Precision Matrix
Simulation 1	1	49.293	95.626
	2	70.876	89.640
Simulation 2	1	22.132	85.653
	2	71.837	89.544
Simulation 3	1	19.054	84.965
	2	45.727	86.194

Table: PRIALs per state for the covariance matrix and the precision (=inverse covariance) matrix when the true state is known.

Condition numbers

Let $\lambda_1 \geq \ldots \geq \lambda_p$ be the ordered eigenvalues of a matrix $\mathbf{\Sigma}$.

$$\text{cond}(\mathbf{\Sigma}) = \frac{\lambda_1}{\lambda_p} \equiv \quad \text{condition number of } \mathbf{\Sigma}. \tag{6}$$

The larger cond($\mathbf{\Sigma}$), the more numerically unstable the inversion of the matrix.

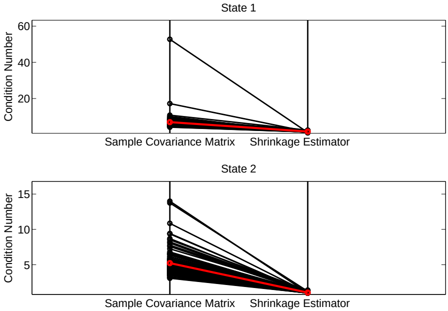


Figure: Comparison of condition numbers for each Monte Carlo iteration in Simulation 1 for (top) state 1 and (bottom) state 2. The mean decreasing trend is shown in red.

Maximum Likelihood Estimation

Recall
$S_t \in \{0,1\}, \varepsilon_t \sim \mathcal{N}(0, \mathbf{I}_p)$ and define

$$X_t = (S_t \mathbf{\Sigma}_1^{1/2} + (1 - S_t)\mathbf{\Sigma}_2^{1/2})\varepsilon_t \tag{7}$$

Additionally,

$$\mathbf{A} = \begin{pmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{pmatrix}$$

Given $S_{t,k} = 1$ and

$$\begin{aligned} f(X_t, \mu_k, \mathbf{\Sigma}_k) &= \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}_k|^{1/2}} \exp\left(-\frac{1}{2}(X_t - \mu_k)'\mathbf{\Sigma}_k^{-1}(X_t - \mu_k)\right) \\ &= f\left(X_t \mid S_t = e_k, \lambda\right) \end{aligned}$$

where λ is the vector of parameters for the K different probability density functions.

More generally, if we were given a hidden sample path $S = (S_1, \ldots, S_T)$, one could have defined

$$L(X_1, \ldots, X_T \mid \lambda, S) = \prod_{t=1}^{T} f(X_t \mid S_t, \lambda) = \prod_{t=1}^{T} \left(\sum_{k=1}^{K} S_{t,k} f(X_t, \mu_k, \mathbf{\Sigma}_k) \right)$$

and therefore the extended likelihood could have been written as

$$P(X, S \mid \lambda) = L(X_1, \ldots, X_T \mid \lambda, S) L(S \mid \lambda). \tag{8}$$

Unfortunately the hidden process is unknown. Therefore,

$$L(X_1, \ldots, X_T \mid \theta) = \sum_{\text{all possible } S} P(X, S \mid \theta). \tag{9}$$

where θ is the vector of parameters of the K different probability distribution functions and the transition probabilities.

A direct optimization of the likelihood function could be numerically cumbersome - we prefer to use the *EM-Algorithm*.

Idea of the EM-Algorithm

Algorithm 1 EM Algorithm

1. Initialize a good starting value of the parameter $\theta^{(0)}$
2. E-Step: Assume the parameter are known and compute the estimated state variables $\hat{S}_{t,k}$ (making use of the forward and backward procedure see e.g. Rabiner (1989))
3. M-Step: Assume the hidden state variables are known and update the parameter estimates by optimizing the cost function (quasi-likelihood function) with respect to the unknown parameters
4. Iterate the E-step and M-Step until a stopping criterion is satisfied.

The M-Step

1. Transition probabilities

$$\hat{a}_{i,j} = \frac{\text{Expected number of transitions from state i to state j}}{\text{Expected number of transitions from i to anywhere}}, \ i,j = 1, \ldots, K \tag{10}$$

2. Initial Distribution

$$\hat{\pi}_k = \frac{\sum_t \hat{S}_{t,k}}{T} \tag{11}$$

3. State means

$$\hat{\mu}_k = \frac{\sum_t \hat{S}_{t,k} X_t}{\sum_t \hat{S}_{t,k}} \tag{12}$$

4. State covariances

$$\hat{\mathbf{\Sigma}}_k^s = (1 - \hat{W}_k^{(0)})\mathbf{\Sigma}_k^{(0)} + \hat{W}_k^{(0)} \alpha_k^{(0)} I_p,$$

where

$$\mathbf{\Sigma}_k^{(0)} = \frac{\sum_t \hat{S}_{t,k}(X_t - \hat{\mu}_k)(X_t - \hat{\mu}_k)'}{T}, \ k = 1, \ldots, K,$$

and then set

$$\hat{\mathbf{\Sigma}}_k = \frac{1}{\hat{\pi}_k} \hat{\mathbf{\Sigma}}_k^s.$$

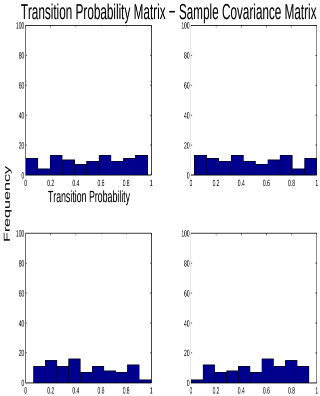
where $\hat{\mathbf{\Sigma}}_k^s$ is the shrinkage estimator.

Histogram Transition Probabilities: Simulation 1

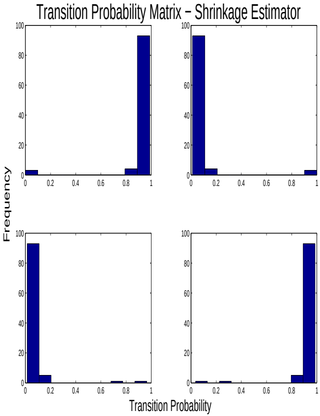
Recall Transition matrix

$$\mathbf{A} = \begin{pmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{pmatrix}$$

Recall Simulation 1: $\mathbf{\Sigma}_1 = \mathbf{H} \quad \mathbf{\Sigma}_2 = \mathbf{D}$ state separation small



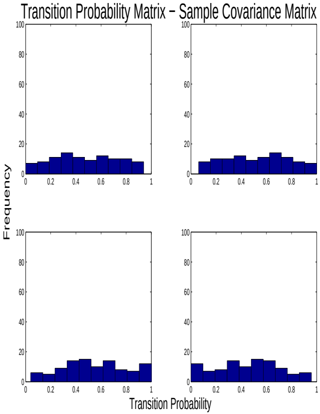
(a) Simulation 1 - Sample Covariance Matrix



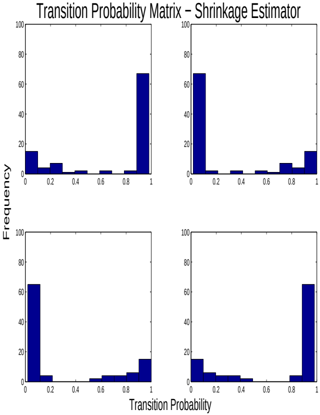
(b) Simulation 1 - Shrinkage Estimator

Histogram Transition Probabilities: Simulation 2

Recall Simulation 2: $\mathbf{\Sigma}_1 = \mathbf{B} \quad \mathbf{\Sigma}_2 = \mathbf{D}$ state separation moderate



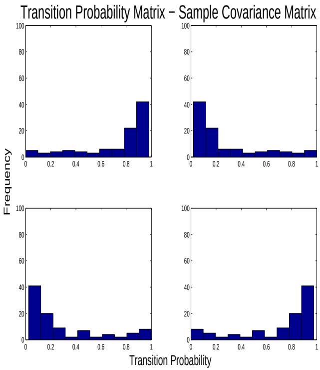
(c) Simulation 2 - Sample Covariance Matrix



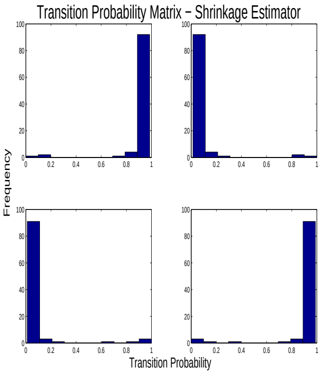
(d) Simulation 2 - Shrinkage Estimator

Histogram Transition Probabilities: Simulation 3

Recall Simulation 3: $\boldsymbol{\Sigma}_1 = \mathbf{B} \quad \boldsymbol{\Sigma}_2 = \mathbf{H}$ state separation larger

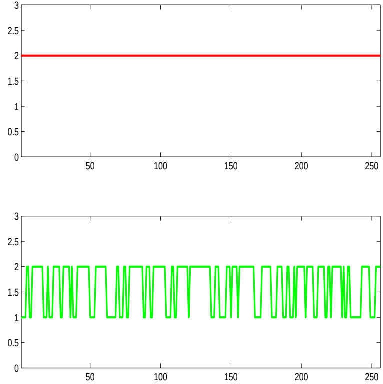


(e) Simulation 3 - Sample Covariance Matrix

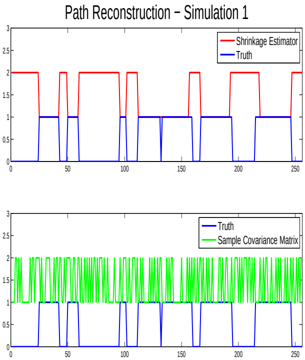


(f) Simulation 3 - Shrinkage Estimator

Sample reconstructed paths - using the Viterbi Algorithm

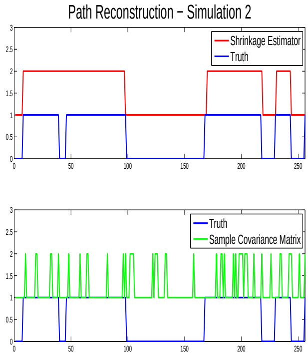


(g) Simulation with $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{D}$

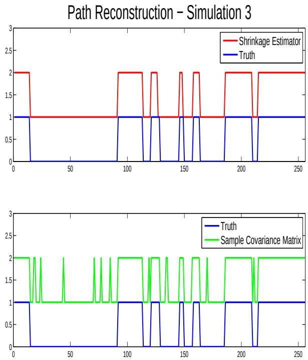


(h) Simulation 1

Sample reconstructed paths (cont'd)

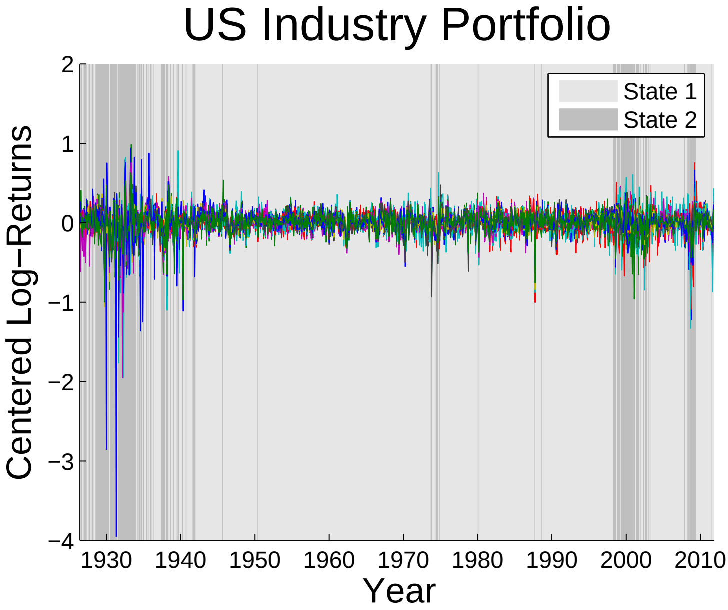


(i) Simulation 2



(j) Simulation 3

Data analysis example - revisited



Analysis of portfolio data - results

- BIC indicated 2 states with estimated transition probability matrix (see paper)

$$\widehat{\mathbf{A}} = \begin{pmatrix} 0.9418 & 0.0582 \\ 0.2565 & 0.7435 \end{pmatrix}.$$

- Industry portfolios prefer less volatile state 1 over state 2 (Great Depression 1930s, dot-com bubble early 2000s, and recent financial crisis late 2007)
- Inspection of the correlation matrix: stronger correlations in state 2
  1. "games and recreation" industry, highly correlated with many of the other industries;
  2. "chemicals, textiles, construction, steel, machinery, electrical equipment, automobiles, transportation equipment, and metal mining" correlated with one another;
  3. and with "business equipment, supplies, transportation, wholesale, retail, restaurants and hotels, banking and trading",...

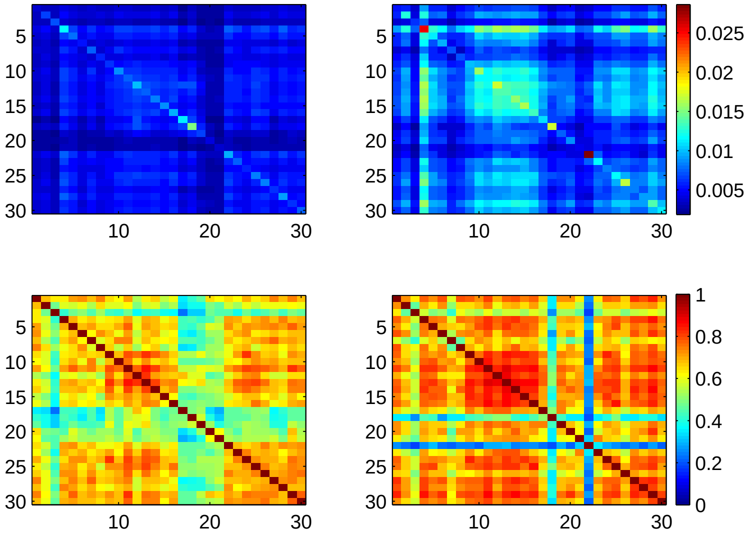


Figure: Estimated variance-covariance (top) and correlation (bottom) matrix state 1 (left) and 2 (right). The brighter the color, the greater the value.

Conclusion

Shrinkage for Covariance estimation in Hidden Markov Models

- improves MSE and
- reduces condition number,
- in particular when effective sample sizes per state are small (in the order of dimension p):
- allows for numerically more stable and invertible estimators
- stabilizes EM and Path reconstruction