HMM for HAC

Cheng-Der Fuh, Wolfgang Karl Härdle

Ostap Okhrin, Weining Wang

Ladislaus von Bortkiewicz
Chair of Statistics
C.A.S.E. - Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin
Graduate Institute of Statistics National Central University, Taiwan
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de
http://www.stat.ncu.edu.tw/

Varying Dependency

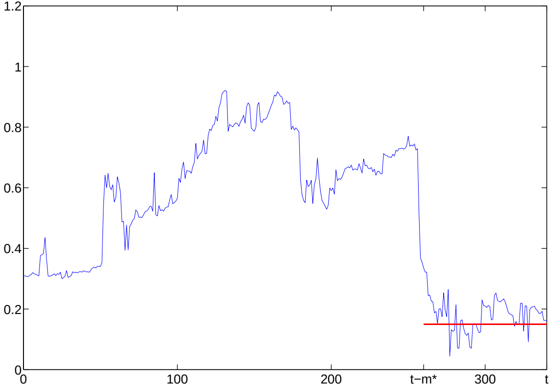


Figure 1: Dependence over time for DaimlerChrysler, Volkswagen, Bayer, BASF, Allianz and Münchener Rückversicherung, 20000101-20041231. Giacomini et. al (2009)

Copulae

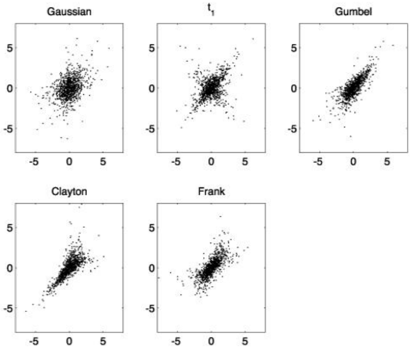


Figure 2: Copulae and Scatterplot

Copulae

A continuous function $C : [0,1]^d \to [0,1]$,

$$C(u_1, \dots, u_d) = F\{F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\}, \quad u_1, \dots, u_d \in [0,1],$$

where $F_1^{-1}(\cdot), \dots, F_d^{-1}(\cdot)$ the quantile functions.

- Separate dependency and marginal distributions
- Represent general dependency

Hierarchical Archimedean Copulae (HAC)

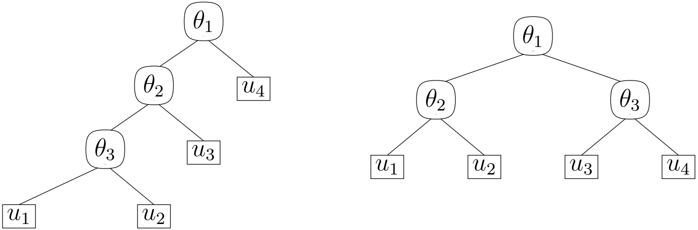


Figure 3: Fully and partially nested copulae of dimension $d = 4$ with structures $s = (((12)3)4)$ and $s = ((12)(34))$

Hierarchical Archimedean Copulae (HAC)

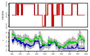
Compositions of simple Archimedean copulae, for example:

$$\begin{aligned} C(u_1,\dots,u_d) &= C_1\{C_2(u_1,\dots,u_{d-1}),u_d\} \\ &= \phi_1\{\phi_1^{-1}\circ\phi_2[\phi_2^{-1}\{C_3(u_1,\dots,u_{d-2})\}+\phi_2^{-1}(u_{d-1})]+\phi_1^{-1}(u_d)\}, \end{aligned}$$

where ϕ is completely monotone. $f(\cdot)$ corresponds to the density

$$f(\cdot) = c\{F_1(y_1),\dots,F_d(y_d),s,\theta\}f_1(y_1)\dots f_d(y_d),$$

where $c(\cdot)$ is the copulae density. Joe (1997) and Nelsen (2006).
$s \stackrel{\mathrm{def}}{=} \{(\dots(i_1\dots i_{j_1})\dots(\dots)\dots)\}$ denotes the structure of a HAC.



Dependency Dynamics

- Multivariate GARCH: DCC, CCC, BEKK, Silvennoinen and Teräsvirta (2009)
- Patton (2004): asset allocation, time varying copulae
- Rodriguez (2007): switching-parameter bivariate copulae.
- Giacomini, Härdle and Spokoiny (2009), Härdle, Okhrin and Okhrin (2011): local adaptive estimation

Local Change Point Detection

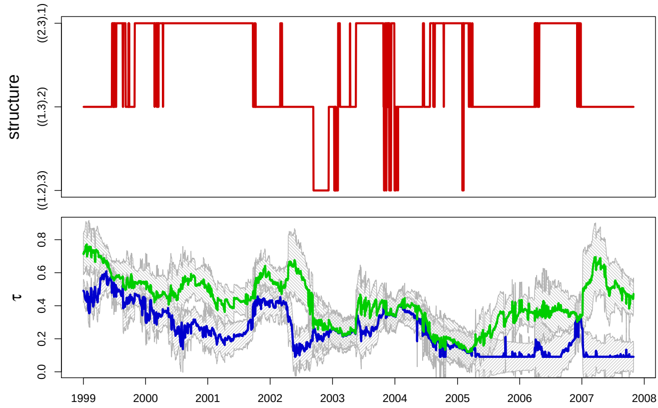


Figure 4: Dependence over time for JPY/USD, GBP/USD and EUR/USD, [19990104-20090814]. Härdle et. al (2011)

Hidden Markov Models

Stochastic process driven by an underlying Markov process, Bickel, Ritov and Ryden (1998), Fuh (2003):

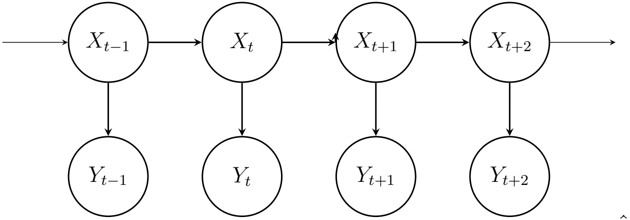
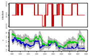


Figure 5: Graphical representation of the dependence structure of HMM



Outline

1. Motivation ✓
2. Model Set-up
3. Simulations
4. Applications
5. Further work

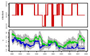
Hidden Markov Models

Observe i.i.d. $Y = (Y_1, Y_2, \ldots, Y_T)^\top \in \mathbb{R}^d$, where $\mathcal{L}(Y_t)$ is driven by a Markov Chain X_t , $t = 1, \ldots, T$, X_t takes value on $1, \ldots, M$. States $X_t = i$ denote structures (s_i^*, θ_i^*).

$$
\begin{aligned}
\mathrm{P}(X_t | X_{1:(t-1)}, Y_{1:(t-1)}) &= \mathrm{P}(X_t | X_{t-1}) \qquad (1)\\
\mathrm{P}(Y_t | Y_{1:(t-1)}, X_{(1:t)}) &= \mathrm{P}(Y_t | X_t), \qquad (2)
\end{aligned}
$$

$\{X_t, Y_t\}$ follows an HMM.

Andrei Markov on BBI: 



Likelihood

Define $p_{ij} = \mathrm{P}(X_t = j | X_{t-1} = i)$ the transition probability, π_i the initial probability, $f_i(\cdot) = f_i(\cdot; s_i, \theta_i)$ the HAC-based density and $\mathfrak{g} \stackrel{\text{def}}{=} (\{\mathbf{s}, \theta\}, p_{ij})$ $(i = 1, \ldots, M,\ j = 1, \ldots, M)$.

$$Z_{i,t} \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } X_t = i \\ 0, & \text{if } X_t \neq i, \end{cases} \quad i = 1, \ldots, M,$$

From Figure 5

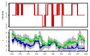
$$\begin{pmatrix} Z_{1t} = 0 \\ Z_{2t} = 1 \\ Z_{3t} = 0 \end{pmatrix}_{t \leq 990705} \quad \begin{pmatrix} Z_{1t} = 0 \\ Z_{2t} = 0 \\ Z_{3t} = 1 \end{pmatrix}_{990705 < t \leq 990710} \quad \ldots$$

Likelihood

The likelihood and log likelihood of Y and X can be expressed as:

$$L(Y,X,\{\theta,s\}) = \{\sum_{i=1}^{M} Z_{i,0}\pi_i f_i(y_0)\} \prod_{t=1}^{T}\{\sum_{i=1}^{M}\sum_{j=1}^{M} Z_{i,t-1}Z_{j,t}p_{ij}f_j(y_t)\}$$

$$\log L(Y,X,\{\theta,s\}) = \sum_{i=1}^{M} Z_{i,0}\log\{\pi_i f_i(y_0)\} + \sum_{t=1}^{T}\sum_{i=1}^{M}\sum_{j=1}^{M} Z_{j,t}Z_{i,t-1}\log\{p_{ij}f_j(y_t)\}$$



EM algorithm

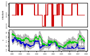
Following Dempster, Laird and Rubin (1997)

(a) E-step : compute $\mathcal{Q}(\mathfrak{g}; \mathfrak{g}^{(\nu)})$,

(b) M-step : choose the update parameters

$$\mathfrak{g}^{(\nu+1)} = \arg\max_{\mathfrak{g}} \mathcal{Q}(\mathfrak{g}; \mathfrak{g}^{(\nu)}),$$

where $\mathcal{Q}(\mathfrak{g}; \mathfrak{g}^{(\nu)}) \stackrel{\text{def}}{=} \mathsf{E}_{\mathfrak{g}^{(\nu)}}\{\log L(Y, X, \theta, \mathsf{s}) | Y\}$.

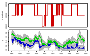


EM algorithm – E-step

$$\begin{aligned}\mathcal{Q}(\mathfrak{g};\mathfrak{g}') &= \sum_{i=1}^{M} \mathrm{P}(X_0 = i|Y)\log\{\pi_i f_i(Y_0)\} \\ &+ \sum_{t=1}^{T}\sum_{i=1}^{M}\sum_{j=1}^{M} \mathrm{P}(X_{t-1} = i, X_t = j|Y)\log\{p_{ij}\} \\ &+ \sum_{t=1}^{T}\sum_{i=1}^{M} \mathrm{P}(X_t = i|Y)\log f_i(Y_t)\end{aligned}$$

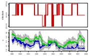
Likelihood with constraints:

$$\mathfrak{L}(\mathfrak{g},\lambda;\mathfrak{g}') = \mathcal{Q}(\mathfrak{g};\mathfrak{g}') + \sum_{i=1}^{M}\lambda_i(1-\sum_{j=1}^{M}p_{ij}). \qquad (3)$$



EM algorithm – M-step

$$
\begin{aligned}
\{\hat{\theta}_i^{(\nu)}, \hat{s}_i^{(\nu)}\} &= \text{arg max}_{s_i,\theta_i} \sum_{t=1}^{T} \mathrm{P}(X_t = i|Y)\mathfrak{L}(\mathfrak{g}_i, \lambda; \mathfrak{g}') \\
\{\hat{\theta}_i^{(j)}\} &= \text{arg zero}_{\theta_i} \sum_{t=1}^{T} \mathrm{P}(X_t = i|Y)\partial \log f_i(y_t)/\partial\theta_i, \\
& \quad i \in 1, \cdots, M
\end{aligned}
$$



Theoretical Results

Theorem

Under certain conditions, we can consistently find the corresponding structure:

$$\lim_{n\to\infty} \mathrm{P}(\hat{s}_i = s_i^*) = 1, \forall i \in 1, \cdots, M. \tag{4}$$

Theorem

Given the selected structure $\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_M$ *and* $\hat{s}_i = s_i^*$*, the estimator* $\hat{\theta}_i$ *satisfies:*

$$\lim_{n\to\infty} \mathrm{P}(\hat{\theta}_i = \theta_i^*) = 1, \forall i, \tag{5}$$

Simulation, I

Transition matrix: $\begin{pmatrix} 0.985 & 0.005 & 0.005 & 0.005 \\ 0.001 & 0.990 & 0.005 & 0.004 \\ 0.003 & 0.003 & 0.991 & 0.003 \\ 0.006 & 0.003 & 0.001 & 0.990 \end{pmatrix}$, $n = 1000$, $d = 3$, $M = 4$,homogeneous marginal distribution: $\mathrm{N}(0,1)$, $t(3)$, $\mathrm{N}(0,3)$

$$C\{u_3, C(u_1, u_2; \theta_1 = 4.0); \theta_2 = 1.5\}$$
$$C\{u_1, C(u_2, u_3; \theta_1 = 15.0); \theta_2 = 4.0\}$$
$$C\{u_2, C(u_1, u_3; \theta_1 = 30.0); \theta_2 = 10.0\}$$
$$C\{u_1, C(u_2, u_3; \theta_1 = 55.0); \theta_2 = 30.0\}$$

Simulation, I

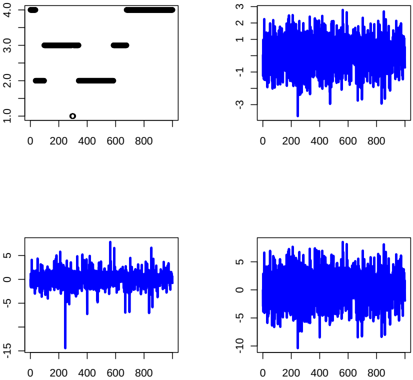


Figure 6: The underlying sequence X_t (upper left panel), marginal plots of (y_1, y_2, y_3).

Simulation, I

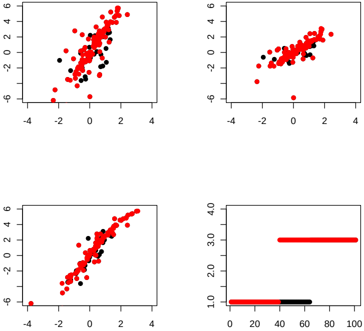


Figure 7: Snapshots of pairwise scatter plots of dependency structures, windows of width 100 with 25 observations lag.

Simulation, I

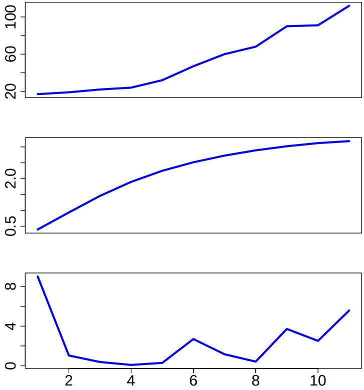


Figure 8: The convergence of states (upper panel), transition matrix (middle panel), parameters (lower panel).

Simulation, II

Realistic setting from data, three states,

$$C\{u_1, C(u_2, u_3; \theta_1 = 1.3); \theta_2 = 1.05\}$$
$$C\{u_2, C(u_3, u_1; \theta_1 = 2.0); \theta_2 = 1.35\}$$
$$C\{u_3, C(u_1, u_2; \theta_1 = 4.5); \theta_2 = 2.85\}$$

Transition matrix: $\begin{pmatrix} 0.72 & 0.15 & 0.13 \\ 0.23 & 0.64 & 0.13 \\ 0.03 & 0.02 & 0.95 \end{pmatrix}$, $n = 2000$, $d = 3$

Simulation, II

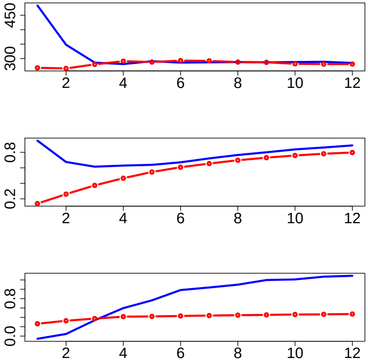


Figure 9: The convergence of states (upper panel), transition matrix (middle panel), parameters (lower panel).

Simulation, II

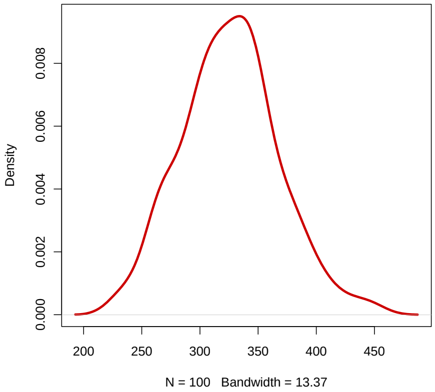


Figure 10: The error of misidentification of states by 100 samples

Application

JPN/EUR, GBP/EUR and USD/EUR, from `DataStream`, [4.1.1999; 14.8.2009], 2771 obs.
Fit to each marginal time series of log-returns a univariate GARCH(1,1) process:

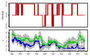
$$X_{j,t} = \mu_{j,t} + \sigma_{j,t}\varepsilon_{j,t} \text{ with } \sigma^2_{j,t} = \omega_j + \alpha_j \sigma^2_{j,t-1} + \beta_j (X_{j,t-1} - \mu_{j,t-1})^2,$$

and $\omega > 0$, $\alpha_j \geq 0$, $\beta_j \geq 0$, $\alpha_j + \beta_j < 1$.

Application

	$\hat{\mu}_j$	$\hat{\omega}_j$	$\hat{\alpha}_j$	$\hat{\beta}_j$	BL	KS
JPY	4.85e-05	2.99e-07	0.06	0.94	0.73	1.70e-05
	(1.15e-04)	(1.02e-07)	(7.49e-03)	(7.64e-03)		
GBP	6.34e-05	1.44e-07	0.06	0.93	0.01	2.10e-04
	(7.39e-05)	(5.11e-08)	(8.75e-03)	(9.12e-03)		
USD	1.76e-04	1.19e-07	0.03	0.97	0.87	1.65e-03
	(1.10e-04)	(5.92e-08)	(4.14e-03)	(4.28e-03)		

Table 1: Results of the fitting of univariate GARCH(1,1) to exchange rates. The last two columns provide the *p*-values of the BL test for autocorrelations with 12 lags and KS test for normality applied to the residuals.



Application

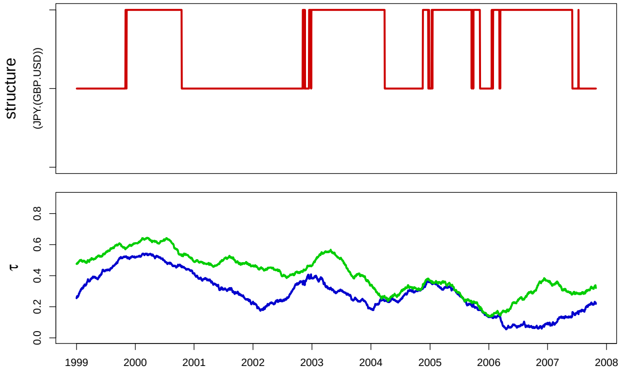


Figure 11: Rolling window for Exchange Rates: structure (upper) and parameters (lower, θ_1 and θ_2) for Gumbel HAC. $w = 250$.

Application

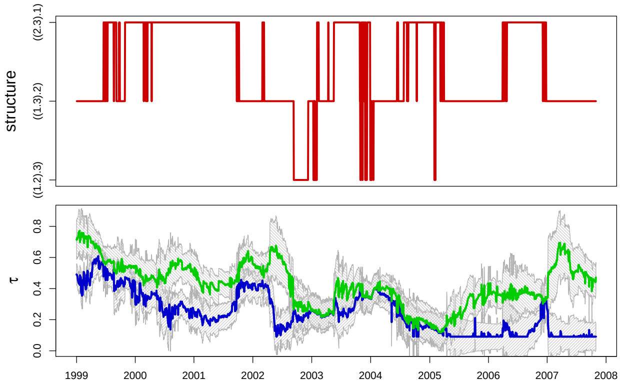


Figure 12: LCP for Exchange Rates: structure (upper) and parameters (lower, θ_1 and θ_2) for Gumbel HAC. $m_0 = 40$.

Application

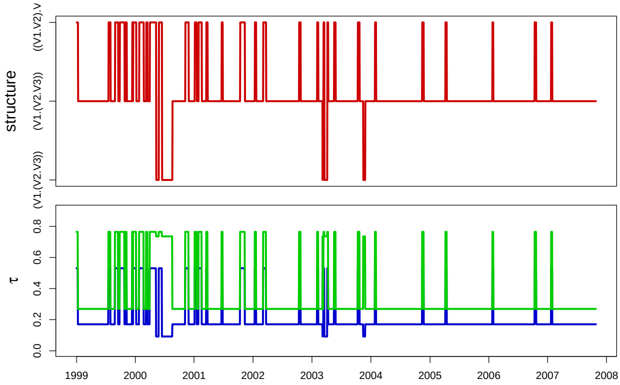


Figure 13: HMM for Exchange Rates: structure (upper) and parameter (lower, θ_1 and θ_2) for Gumbel HAC.

Movie

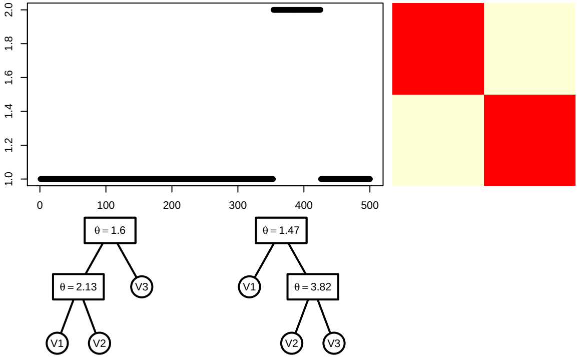


Figure 14: States(top left and bottom), Transition matrix(top right)

Application

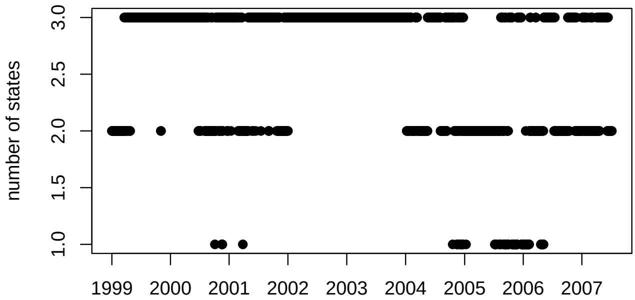


Figure 15: Plot of estimated number of states

VaR

$T = 2219$, $N = 10^4$ is the sample size, $\omega = 1000$ portfolios.

The P&L function is $L_{t+1} = \sum_{i=1}^{3} w_i(y_{i,t+1} - y_{i,t})$,, $w_i = 1/3$ The VaR of at level α is $VaR(\alpha) = F_L^{-1}(\alpha)$

$$\hat{\alpha}_{\mathbf{w}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{I}\{L_t < \widehat{VaR}_t(\alpha)\}.$$

The distance between $\hat{\alpha}$ and α

$$e_{\mathbf{w}} = (\hat{\alpha}_{\mathbf{w}} - \alpha)/\alpha.$$

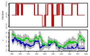
The performance of models is measured through

$$A_W = \frac{1}{|W|}\sum_{\mathbf{w}\in W} e_{\mathbf{w}}, \qquad D_W = \left\{\frac{1}{|W|}\sum_{\mathbf{w}\in W}(e_{\mathbf{w}} - A_W)^2\right\}^{1/2}.$$

Backtesting

	w	0.1	0.05	0.01
HMM, RGum	500	0.0980	0.0507	0.0128
HMM, Gum	500	0.0981	0.0512	0.0135
Rolwin, RGum	250	0.1037	0.0529	0.0151
Rolwin, Gum	250	0.1043	0.0539	0.0162
LCP, $m_0 = 40$	468	0.0973	0.0520	0.0146
LCP, $m_0 = 20$	235	0.1034	0.0537	0.0169
DCC	500	0.0743	0.0393	0.0163

Table 2: VaR backtesting results, $\bar{\hat{\alpha}}$, where "Gum" denotes the Gumbel copula and "RGum" the rotated Gumbel one.



Backtesting

	w	0.1	0.05	0.01
HMM, RGum	500	-0.0204 (0.013)	0.0147 (0.012)	0.2827 (0.064)
HMM, Gum	500	-0.0191 (0.008)	0.0233 (0.018)	0.3521 (0.029)
Rolwin, RGum	250	0.0375 (0.009)	0.0576 (0.012)	0.5076 (0.074)
Rolwin, Gum	250	0.0426 (0.009)	0.0772 (0.030)	0.6210 (0.043)
LCP, $m_0 = 40$	468	-0.0270 (0.010)	0.0391 (0.018)	0.4553 (0.037)
LCP, $m_0 = 20$	235	0.0344 (0.009)	0.0735 (0.026)	0.6888 (0.050)
DCC	500	-0.2573 (0.015)	-0.2140 (0.015)	0.6346 (0.091)

Table 3: Robustness relative to $A_W(D_W)$

Rainfall

- nonzero point mass of the rainfall distribution on a certain location
- marginal distributions: censored normal

Rainfall Data

- Non-zero point mass at 0, need censored distributions
- Marginals, censored normal:

$$f_{X_t,k}\{y_t(k)\} = \begin{cases} 1 - p^{X_t,k} & y_t(k) < 0 \\ p^{X_t,k}\varphi\{(y - \mu^{X_t,k})/(\sigma^{X_t,k})\}/\sigma^{X_t,k} & y_t(k) \geq 0 \end{cases}$$

where $\varphi(.)$ is the standard normal pdf.

Rainfall

Joint distribution function:

$$c_d(\mu,\theta) = \begin{cases} c_c(\mu,\theta) & Y_t(k) > 0, \forall k \\ \partial C_c(\mu,\theta)/\partial\mu_{j_1}\dots\partial\mu_{j_B} & , j_i \in \{y_t(j_i) > 0\} \end{cases}$$

Likelihood:

$$\begin{aligned} &\log L(Y, X, \theta, \mathrm{s}) \\ &= \sum_{i=1}^{M} Z_{i,0} \log\{\pi_i f_i(y_0)\} + \sum_{t=1}^{T}\sum_{i=1}^{M}\sum_{j=1}^{M} Z_{j,t} Z_{i,t-1} \log\{p_{i,j} f_j(y_t)\} \\ &+ \sum_{t\in B}\sum_{i=1}^{M}\{Z_{i,t}\{\log(\pi_i)\} - \sum_{j=1}^{M} Z_{j,t-1} Z_{i,t} \log(p_{i,j})\}, \end{aligned}$$

where B is the starting date of each month.

Rainfall

- Data: Daily rainfall from Fujian, Guangdong, Guangxi, every June 1950 – 2006.
- Treat each month as independent realization of a HMM.

state	Occur Prob			Mean		
1	0.174	0.112	0.106	2.813	3.040	3.297
2	0.808	0.777	0.742	-5.908	-4.393	-3.527
3	0.173	0.715	0.715	2.482	-3.322	-3.519
	Variance					
1	2.992	2.498	2.642			
2	6.787	5.776	5.442			
3	2.639	5.839	6.186			

States and Transition Matrix

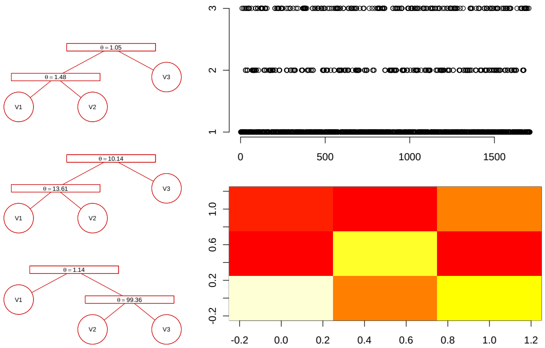


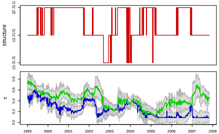
Figure 16: Tree structure for Copulae parameter (left panel),estimated underlying states and transition matrix

HMM for HAC

Cheng-Der Fuh, Wolfgang Karl Härdle

Ostap Okhrin, Weining Wang

Ladislaus von Bortkiewicz
Chair of Statistics
C.A.S.E. - Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin
Graduate Institute of Statistics National Central University, Taiwan
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de
http://www.stat.ncu.edu.tw/



Algorithm

Following Rabiner (1989),

estimate $x_1, x_2, \ldots, x_T$ (the underlying Markov chain) which maximizes $\mathrm{P}(Y|\lambda)$.

Viterbi Algorithm:

- Initialization : $\delta_1(i) = \pi_i f_i(y_1)$, $1 \le i \le M$, $\psi_1(i) = 0$.
- Recursion :

$$\begin{aligned} \delta_t(i) &= \max_{1 \le i \le M} \{\delta_{t-1}(i) p_{ij}\} f_j(y_t), 2 \le t \le T, 1 \le j \le M, \\ \psi_t(j) &= \arg\max_{1 \le i \le M} \psi_{t-1}(i) p_{ij} \end{aligned} \tag{6}$$

Algorithm

- Termination :

$$p^* = \max_{1 \le i \le M} \{\delta_T(i)\}$$
$$q_T^* = \arg\max_{1 \le i \le M} \{\delta_T(i)\}$$

- Path (State Sequence) back tracking : $q_t^* = \psi_{t+1}(q_{t+1}^*)$, $t = T-1, T-2, \dots, 1$

Algorithm

Update parameters, let

$$\begin{aligned} \alpha_t(i) &= \mathrm{P}(y_1, y_2, \dots, y_t, x_t = i | \lambda^{(0)}) \\ \beta_t(i) &= \mathrm{P}(y_{t+1}, y_{t+2}, \dots, T | x_t = i, \lambda^{(0)}) \end{aligned}$$

They can be estimated efficiently by the follow algorithm:

- $\alpha_1(i) = \pi_i f_i(y), 1 \leq i \leq M$
- Induction : $\alpha_{t+1}(j) = \sum_{i=1}^{M} \alpha_t(i) p_{ij} f_j(y_{t+1})$
- Termination: $\mathrm{P}(Y|\lambda) = \sum_{i=1}^{M} \alpha_t(i)$

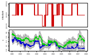
Algorithm

- $\beta_T(i) = 1,\ 1 \leq i \leq M$.
- $\beta_t(i) = \sum_{j=1}^{N} p_{ij} f_j(y_{t+1}) \beta_{t+1}(j),\ t = T-1, T-2, \ldots, 1,$ $1 \leq i \leq M$

$$\begin{aligned} \xi_t(i,j) &\stackrel{\text{def}}{=} \mathrm{P}(x_t = i, x_{t+1} = j | Y, \lambda) \\ r_t(i) &\stackrel{\text{def}}{=} \mathrm{P}(x_t = i | Y, \lambda) \end{aligned}$$

So they can be estimated by:

$$\begin{aligned} \xi_t(i,j) &= \frac{\alpha_t(i) p_{ij} f_j(y_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) p_{ij} f_j(y_{t+1}) \beta_{t+1}(j)} \\ r_t(j) &= \sum_{j=1}^{N} \xi_t(i,j) \end{aligned}$$

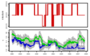


Algorithm

Therefore, update equations are:

$$\pi_i^{(k)} = r_i^{(k-1)}(i)$$

$$p_{i,j}^{(k)} = \frac{\sum_{t=1}^{T-1} \xi_t^{(k-1)}(i,j)}{\sum_{t=1}^{T-1} r_t^{(k-1)}(i)}$$



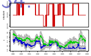
References

Bickel, P. J., Ritov, Y. and Ryden, T.
Asymptotic normality of the maximum-likelihood estimator for general hidden markov models
Annals of Statistics (1998), 26(4): 1614-1635.

Chen, X. and Fan, Y.
Estimation of copula-based semiparametric time series models,
Journal of Econometrics, (2005), 130(2): 307-335.

Dempster, A., Laird, N. and Rubin, D.
Maximum likelihood from incomplete data via the EM algorithm (with discussion),
J. R. Statistics Soc. (1997), 39: 1-38.

Giacomini, E., Härdle, W. and Spokoiny, V.
Inhomogeneous dependence modeling with time-varying copulae,
Journal of Business and Economic Statistics (2009), 27(2): 224-234.



References

Härdle, W. K., Okhrin, O. and Okhrin, Y.
Time varying hierarchical Archimedean copulae,
SFB 649 Discussion Paper (18),(2010)

McNeil, A. J. and Neślehova
Multivariate Archimedean copulas, d-monotone functions and l1 norm symmetric distributions,
Annals of Statistics (2009) 37(5b): 3059-3097.

Joe, H.
Multivariate Models and Dependence Concepts
(1997) Chapman and Hall, London.

Patton A. J.
On the Out-of-sample Importance of Skewness and Asymmetric Dependence for Asset Allocation,
Journal of Financial Econometrics 2 (2004) 130-168.