HMM for HAC

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
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de

Varying Dependency

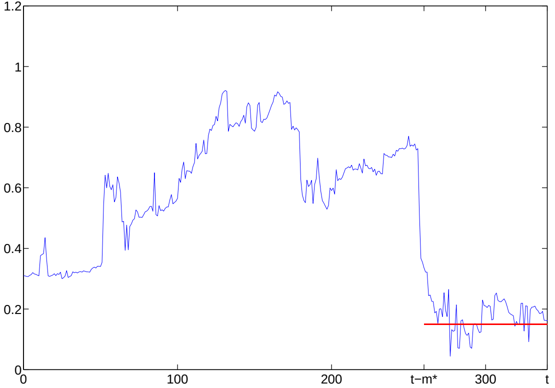
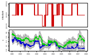


Figure 1: Dependence over time for DaimlerChrysler, Volkswagen, Bayer, BASF, Allianz and Münchener Rückversicherung, 20000101-20041231. Giacomini et. al (2009)



Copulae

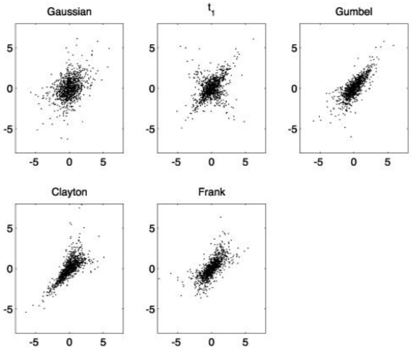


Figure 2: Copulae and Scatterplot

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)

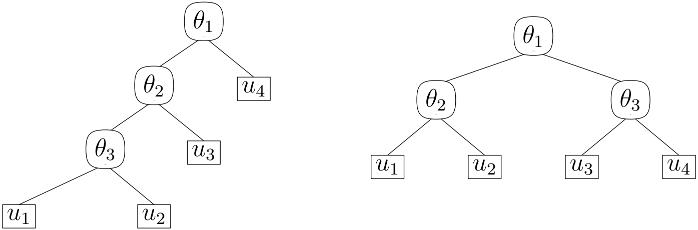


Figure 3: Fully and partially nested copulae of dimension $d = 4$ with structures $s = (((12)3)4)$ and $s = ((12)(34))$

Hierarchical Archimedean Copulae (HAC)

Compositions of simple Archimedean copulae, for example:

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

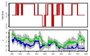
where ϕ is completely monotone. $f(\cdot)$ corresponds to the density

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

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

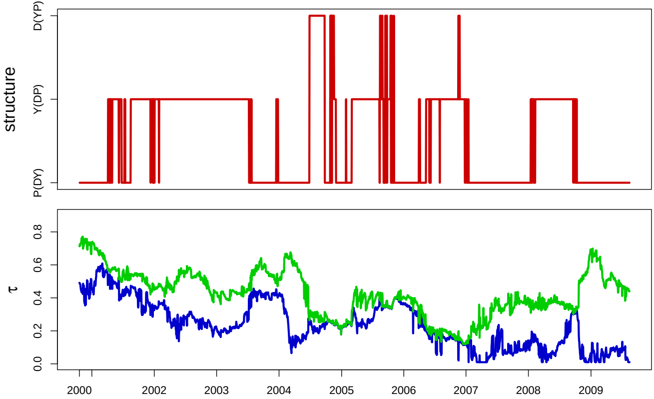


Figure 4: Dependence over time for JPY/USD, GBP/USD and EUR/USD, [19990104-20090814]. Härdle et. al (2011)

Hidden Markov Models

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

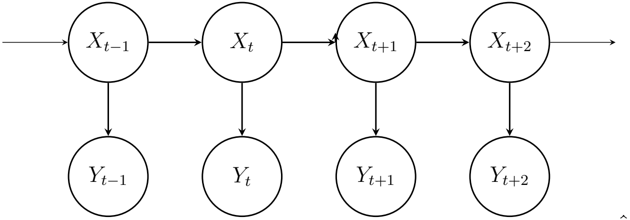


Figure 5: Graphical representation of the dependence structure of HMM

Outline

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

Hidden Markov Models

Observe i.i.d. $Y = (Y_1, Y_2, \dots, Y_T)^\top \in \mathbb{R}^d$, where $\{Y_t\}_{t \geq 0}$ is connected with an underlying Markov Chain $\{X_t\}_{t \geq 0}$, $t = 1, \dots, T$, X_t takes value on $1, \dots, M$.

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

$\{X_t, Y_t\}$ 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
- $f_j\{b; s^{(j)}, \theta^{(j)}\}$ (abbreviated as $f_i(.)$) the HAC-based density
- $\mathfrak{g} \stackrel{\text{def}}{=} (\{\mathsf{s}, \theta\}, p_{ij})$ $(i = 1, \ldots, M,\ j = 1, \ldots, M)$.

Likelihood

For given d dimensional time series $y_1, \cdots, y_T, \in \mathbb{R}^d$
$(y_t = (y_{1t}, y_{2t}, y_{3t}, \ldots, y_{dt})^\top)$ π_{x_t} as the π_i for
$x_0 = i, i = 1, \ldots, M$, and $p_{x_{t-1}x_t} = p_{ji}$ for $x_{t-1} = j$ and $x_t = i$.
The likelihood of Y and X can be expressed as:

$$p_T(y_1, \cdots, y_T; x_1, \ldots, x_T) = \pi_{x_0} \prod_{t=1}^{T} p_{x_{t-1}x_t} f_{x_t}(y_t; \theta^{(x_t)}, s^{(x_t)})$$

EM algorithm

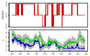
Following Dempster, Laird and Rubin (1997)

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

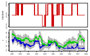


EM algorithm – E-step

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

Likelihood with constraints:

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

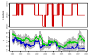


EM algorithm – M-step

$$\{\hat{\theta}^{(i)}_{(\nu)}, \hat{s}^{(i)}_{(\nu)}\} = \text{arg max}_{s^{(i)},\theta^{(i)}} \sum_{t=1}^{T} \mathrm{P}(X_t = i|Y)\mathfrak{L}(\mathfrak{g}_i, \lambda; \mathfrak{g}')$$

$$\{\hat{\theta}_{ij}\} = \text{arg zero}_{\theta_{ij}} \sum_{t=1}^{T} \mathrm{P}(X_t = i|Y)\partial \log f_i(y_t)/\partial\theta_{ij},$$

$$i \in 1, \cdots, M$$



Theoretical Results

Theorem

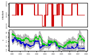
Under certain conditions, we can consistently find the corresponding structure:

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

Theorem

Given the selected structures $\{\hat{s}^{(i)}\}s$*, 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

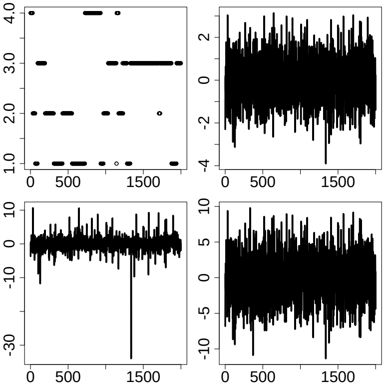


Figure 6: The underlying sequence X_t (upper left panel), marginal plots of (y_1, y_2, y_3).

Simulation, I

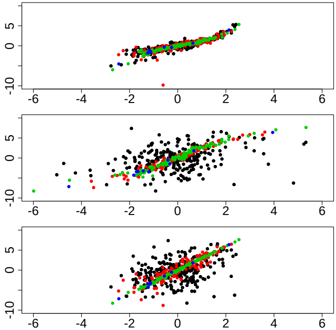


Figure 7: Snapshots of pairwise scatter plots of dependency structures ($t = 500, \ldots, 1000$), the 1st against 2nd (upper), the 2nd against 3rd (middle), and the 1st against 3rd (lower).

Simulation, I

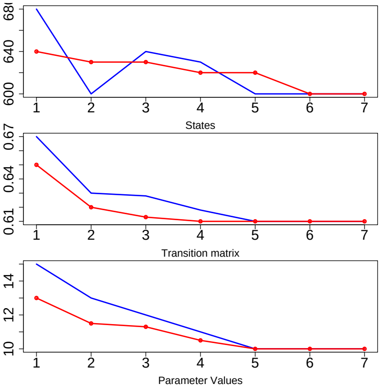


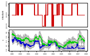
Figure 8: The convergence of states (upper panel), transition matrix (middle panel), parameters (lower panel).

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

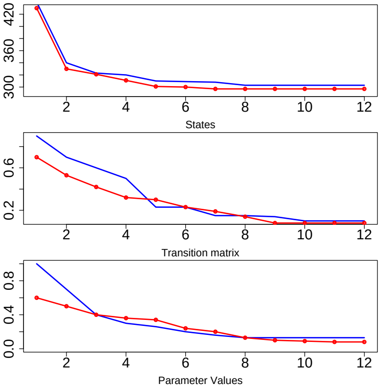


Figure 9: The convergence of states (upper panel), transition matrix (middle panel), parameters (lower panel).

Simulation, II

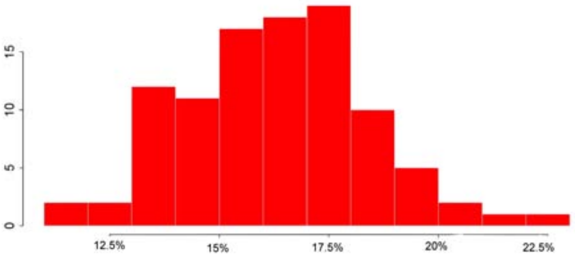
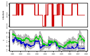


Figure 10: The error of misidentification of states by 100 samples



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:

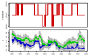
$$X_{j,t} = \mu_{j,t} + \sigma_{j,t}\varepsilon_{j,t} \text{ with } \sigma^2_{j,t} = \omega_j + \alpha_j \sigma^2_{j,t-1} + \beta_j (X_{j,t-1} - \mu_{j,t-1})^2,$$

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

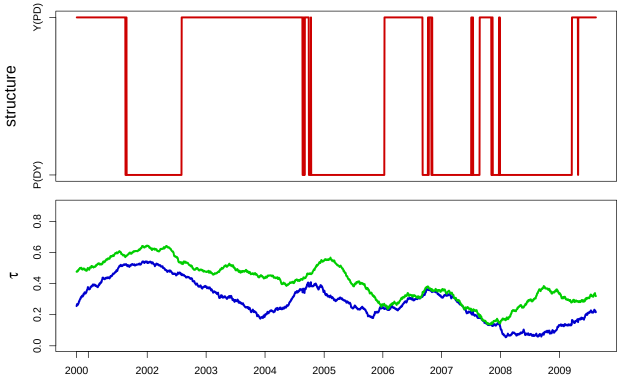


Figure 11: Rolling window for Exchange Rates: structure (upper) and parameters (lower, θ_1 and θ_2) for Gumbel HAC. $w = 250$.

Application

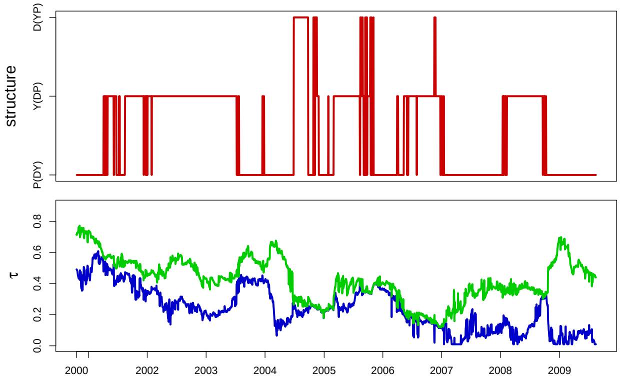
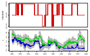


Figure 12: LCP for Exchange Rates: structure (upper) and parameters (lower, θ_1 and θ_2) for Gumbel HAC. $m_0 = 40$.



Application

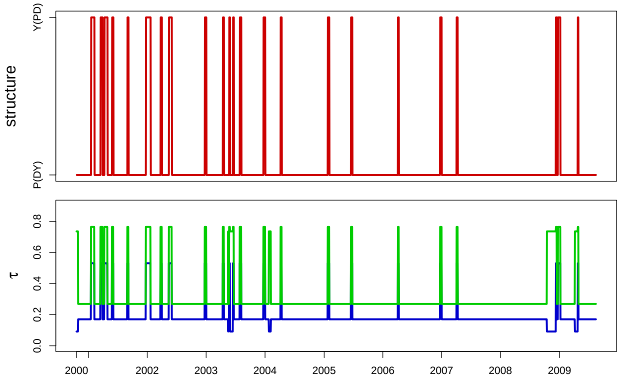
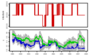


Figure 13: HMM for Exchange Rates: structure (upper) and parameter (lower, θ_1 and θ_2) for Gumbel HAC.



Movie

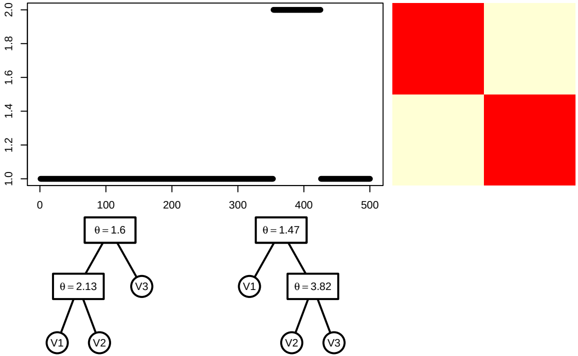
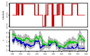


Figure 14: States(top left and bottom), Transition matrix(top right)



Application

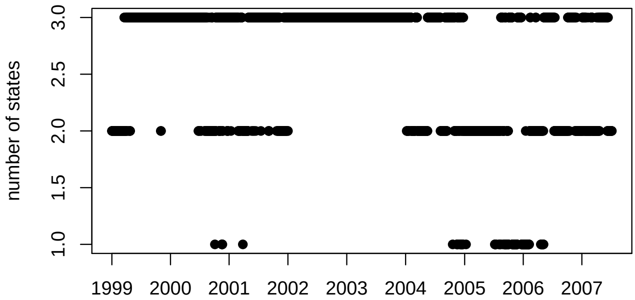
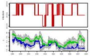


Figure 15: Plot of estimated number of states



VaR

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

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

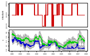
$$\hat{\alpha}_{\mathbf{w}} = \frac{1}{T}\sum_{t=1}^{T} \mathrm{I}\{L_t < \widehat{VaR}_t(\alpha)\}.$$

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

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

The performance of models is measured through

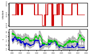
$$A_W = \frac{1}{|W|}\sum_{\mathbf{w}\in W} e_{\mathbf{w}}, \qquad D_W = \left\{\frac{1}{|W|}\sum_{\mathbf{w}\in W}(e_{\mathbf{w}} - A_W)^2\right\}^{1/2}.$$



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 Data

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

$$f_k^{\mathrm{m}}\{y_{tk}\} = \begin{cases} 1 - p_k^{x_t} & y_{tk} = 0 \\ p_k^{x_t}\varphi\{(y_{tk} - \mu^{x_t}(k))/(\sigma^{x_t}(k))\}/\sigma^{x_t}(k) & y_{tk} > 0 \end{cases}$$

gamma distribution:

$$f_k^{\mathrm{m}}\{y_{tk}\} = \begin{cases} 1 - p_k^{x_t} & y_{tk} = 0 \\ p_k^{x_t}\gamma\{y_{tk};\alpha(k)^{x_t},\beta(k)^{x_t}\} & y_{tk} > 0, \end{cases}$$

where k in $1,\ldots,d$.

Rainfall

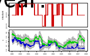
$$c_d(\mu,\theta) = \begin{cases} c_c(\mu,\theta) & , y_{tk} > 0, \forall k \\ \partial C_c(\mu,\theta)/\partial\mu_{k_1}\dots\partial\mu_{k_B} & , k_i \in \{y_{tk_i} > 0\}, i \in 1,\dots,E \end{cases}$$

$$\log p_T(y_{1:T}, x_{1:T}; v \times \omega) = \sum_{i=1}^{M} \mathbf{1}\{x_0 = i\} \log\{\pi_i f_i(y_0)\} +$$

$$\sum_{t=1}^{T}\sum_{i=1}^{M}\sum_{j=1}^{M} \mathbf{1}\{x_t = j\}\mathbf{1}\{x_{t-1} = i\} \log\{p_{ij} f_j(y_t)\}$$

$$+\sum_{t\in B}\sum_{i=1}^{M}\{\mathbf{1}\{x_t = i\}\{\log(\pi_i)\} - \sum_{j=1}^{M}\mathbf{1}\{x_t = j\}\mathbf{1}\{x_{t-1} = i\}\log(p_{ij})\}.$$

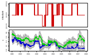
where B is the set of days as the first day in June for each year.



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

X_t	trans. prob			$\hat{\pi}_{X_t}$	α	β	p_k
1	0.590	0.321	0.089	0.298	0.442	139.33	0.748
					0.429	116.70	0.744
					0.552	169.66	0.561
2	0.188	0.742	0.070	0.660	0.671	273.83	0.194
					0.618	253.25	0.213
					0.561	427.46	0.317
3	0.329	0.271	0.400	0.042	0.636	381.09	0.333
					1.125	264.83	0.000
					0.774	514.08	0.056

Table 4: Transition probs of the model together with the initial probs. (data 1957 – 2006), rainfall occurrence prob. and mean, variance estimated from HMM.



Correlations

State	True	$\widehat{\text{Corr}}(X_{t_1}, X_{t_2})$
$1-2$	0.308	0.300 (0.235, 0.373)
$2-3$	0.261	0.411 (0.256, 0.586)
$1-3$	0.203	0.130 (0.058, 0.215)

Table 5: True correlations, simulated averaged correlations from 1000 samples their confidence intervals 1%.

Log-survivor-function

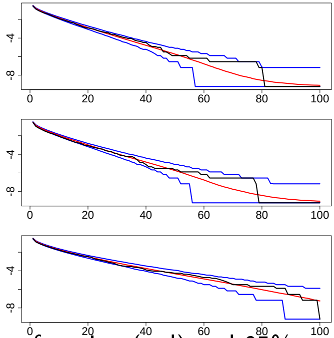


Figure 16: Log-survivor-function (red) and 95% prediction intervals (blue) of the simulated distribution for the fitted model with sample log-survivor-function superimposed (black)

States and Transition Matrix

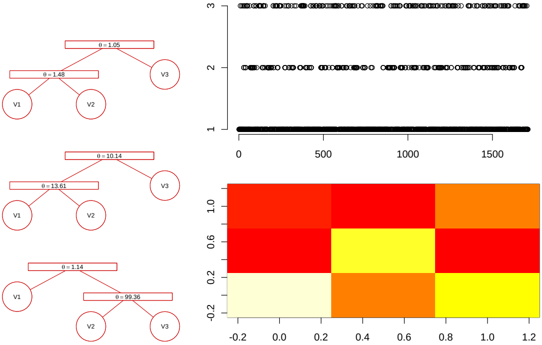
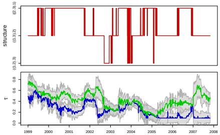


Figure 17: Tree structure for Copulae parameter (left panel),estimated underlying states and transition matrix

HMM for HAC

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
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de



Algorithm

Following Rabiner (1989),

estimate $x_1, x_2, \dots, x_T$ (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 :

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

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

Algorithm

Update parameters, let

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

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

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

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

$$r_t(i) \stackrel{\text{def}}{=} \mathrm{P}(x_t = i | Y, \lambda)$$

So they can be estimated by:

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

$$r_t(j) = \sum_{j=1}^{N} \xi_t(i,j)$$

Algorithm

Therefore, update equations are:

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

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

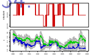
References

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.