SFB 649 Discussion Paper 2012-001

HMM in dynamic HAC models

Wolfgang Karl Härdle*
Ostap Okhrin*
Weining Wang*

* Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

HMM in dynamic HAC models[*†]

Wolfgang Karl Härdle[‡], Ostap Okhrin [§], Weining Wang[¶]

January 2, 2012

Abstract

Understanding the dynamics of high dimensional non-normal dependency structure is a challenging task. This research aims at attacking this problem by building up a hidden Markov model (HMM) for Hierarchical Archimedean Copulae (HAC), where the HAC represent a wide class of models for high dimensional dependency, and HMM is a statistical technique to describe time varying dynamics. HMM applied to HAC provide flexible modeling for high dimensional non Gaussian time series. Consistency results for both parameters and HAC structures are established in an HMM framework. The model is calibrated to exchange rate data with a VaR application, where the model's performance is compared with other dynamic models, and in the second application we simulate rainfall process.

Keywords: Hidden Markov model, Hierarchical Archimedean Copulae, Multivariate Distribution

JEL classification: C13, C14, G50

1 Introduction

Modelling high-dimensional time series is an often underestimated exercise of routine econometrical and statistical work. This slightly pejorative attitude towards day to day statistical analysis is unjustified since actually the calibration of time series models in high dimensions for standard data sizes is not only a difficulty on the numerical side but also poses a challenge on the mathematical side. Computationally speaking, integrated models for high dimensional time series become more evolved when the parameter space is too high. An example is the multivariate GARCH(1,1)

[*]The financial support from the Deutsche Forschungsgemeinschaft via SFB 649 Ökonomisches Risikö, Humboldt-Universität zu Berlin is gratefully acknowledged.

[†]We thank Prof. Cheng-Der Fuh for his comments

[‡]Professor at Humboldt-Universität zu Berlin and Director of CASE - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany. Email:haerdle@wiwi.hu-berlin.de

[§]Assistant Professor at the Ladislaus von Bortkiewicz Chair of Statistics of Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany. Email:ostap.okhrin@wiwi.hu-berlin.de

[¶]Research associate at the Ladislaus von Bortkiewicz Chair of Statistics of Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany. Email:wangwein@cms.hu-berlin.de

BEKK model that for even two dimensions has an associated parameter space of dimension 12. For moderate sample sizes, the parameter space dimension might well be in the range of the sample size or even bigger. This data situation has evoked a new strand of literature on dimension reduction via penalty methods.

In this paper we are taking a different route by calibrating an integrated dynamic model with unknown dependency structure among the d dimensional time series variables. More precisely, the unknown dependency structure may vary among a set of given dependencies. The specific dependence at each time t is unknown to the data analyst, but depend on the dependency patter at time $t-1$. Therefore, hidden Markov models (HMM) come naturally into play. This leaves us with the problem of specifying the set of dependencies.

An approach based assuming a multivariate Gaussian or mixed normal is limited in capturing important types of data features such as heavy tails, asymmetry and nonlinear dependence. Such a simplification might in practice be a too restrictive assumption and might lead to biased results. Copulae are one of possible approaches in solving these problems, see Joe (1996). Moreover, copulae allow us to separate the marginal distributions and the dependency model, see Sklar (1959). Over decades copula based models gained their popularity in various fields like finance, insurance, biology, hydrology, etc. Nevertheless, many basic multivariate copulae are still too restrictive and a simple extension by putting in more parameters would lead to the extreme of a total nonparametric approach that runs into a curse of dimensionality problem. A natural compromise is the class of hierarchical Archimedean copulae (HAC). A HAC allows a rich copula structure with a finite number of parameters. Recent works which have shown their flexibility are McNeil and Nešlehová (2009), Okhrin, Okhrin and Schmid (2009b), Whelan (2004).

Many attempts have been done to give insights into the dynamics of the copulae: Chen and Fan (2005) assumes the underlying sequence is Markovian; Patton (2004) considers an asset-allocation problem with a time-varying parameter of bivariate copulae; Rodriguez (2007) studies financial contagion using switching-parameter bivariate copulae. A likelihood based local adaptive method is an alternative approach to understand the time evolution, see Giacomini, Härdle and Spokoiny (2009), Härdle, Okhrin and Okhrin (2011). Figure 1 presents the LCP (local change point method) window analysis of HAC for exchange rate data. One observes that the structure (upper panel) stays very often the same for a long time, and the parameters (lower panel) are slowly varying over time. This indicates that the dynamics of HAC functions is likely driven by Markovian sequence connected with structures and parameter values. This suggests us a different path of modeling the dynamics: instead of taking a local point view, we adopt a global dynamic model (HMM) for the change of both tree structure and parameters of HAC along time horizon. Under HMM, a stochastic process Y with a not directly observable underlying Markov process X, it is needed to determine state of distributions of Y. It has been widely applied to speech recognition see Rabiner (1989), molecular biology, digital communications over unknown channels. Estimation and inference issues in HMM see Bickel, Ritov and Rydén (1998) and Fuh (2003), among others.

In this paper, we propose a new type of dynamic models, called HMMHAC, by incorporating HAC into an HMM framework. The theoretical problems like parameter consistency and structure consistency are solved. The expectation maximization (EM) algorithm is developed in this framework for parameter estimation. See section 2 for model description, section 3 for theorems for consistency. EM algorithm and computation issues are in section 4. Section 5 is for simulation study, and section 6 is for applications. The technical details are put into appendix.

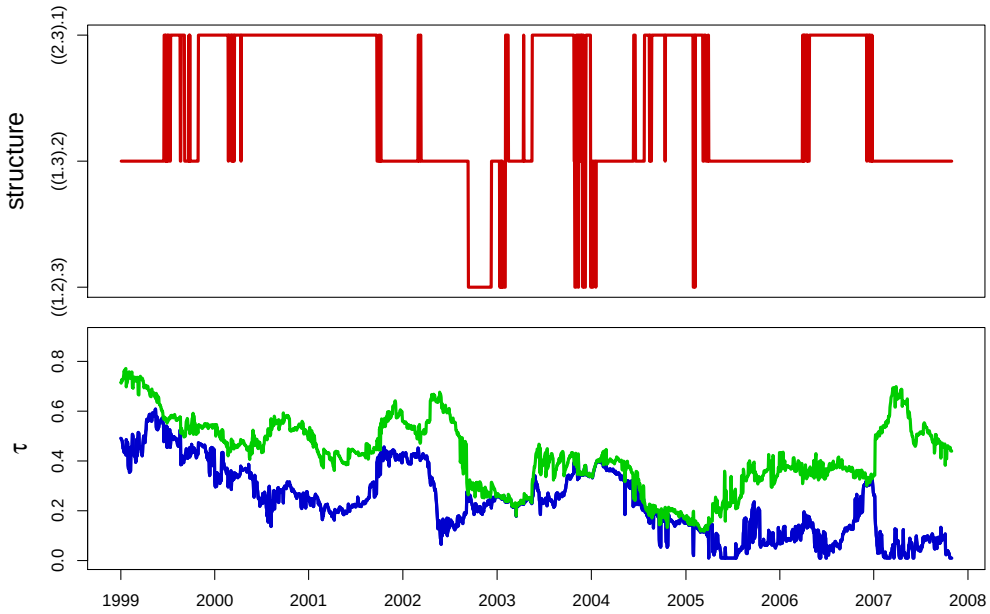


Figure 1: LCP for Exchange Rates: structure (upper) and parameters (lower, θ_1(green) and θ_2)(blue) for Gumbel HAC. $m_0 = 40$.

2 Model description

The hidden Markov model is regarded as a parameterized Markov random walk with the underlying Markov chain viewed as missing data, as in Leroux (1992), Bickel et al. (1998). Specifically, in our HMM HAC framework, let $\{X_t, t \geq 0\}$ be a stationary Markov chain on a finite state space $D = \{1, 2, \ldots, M\}$, with transition probability matrix $P^{v\times\omega} = [p_{ij}^{v\times\omega}]_{i,j=1,\ldots,M}$ and initial distribution $\pi^{v\times\omega} = \{\pi_i^{v\times\omega}\}_{i=1,\ldots,M}$, where $v \times \omega \in V \times \Omega \subseteq N^* \times \mathbb{R}^q$ denotes an element in the parameter space $V \times \Omega$ which parametrize this model, and q as the number of parameters (note that our parameter space is partially discrete (V), and partially continuous (Ω)). Suppose that a real-valued additive component $B_{t,j} = \sum_{k=0}^{t} Y_{k,j}, j \in 1, \ldots, d$ with $B_t = (B_{t,1}, B_{t,2}, \ldots, B_{t,d})^\top$ and $Y_k = (Y_{k,1}, Y_{k,2}, \ldots, Y_{k,d})^\top$ are r.v. taking values on $\mathbb{R}^d$, is adjoined to the chain such that $\{(X_n, B_t), t \geq 0\}$ is a Markov chain on $D \times \mathbb{R}^d$ and

$$\begin{aligned} & \mathrm{P}\{(X_t, B_t) \in A \times (B + b) | (X_{t-1}, B_{t-1}) = (i, b)\} \\ = \; & \mathrm{P}\{(X_1, B_1) \in A \times B | (X_0, B_0) = (i, 0)\} \\ = \; & \mathrm{P}(i, A \times B) = \sum_{j\in A} \int_{b\in B} p_{ij}^{v\times\omega} f_j\{b; s_j(v \times \omega), \boldsymbol{\theta}^{(j)}(v \times \omega)\}\mu(db), \end{aligned} \tag{1}$$

where $B, b \subseteq \mathbb{R}^d$, $A \subseteq D$, $f_j\{b; s^{(j)}(v \times \omega), \boldsymbol{\theta}^{(j)}(v \times \omega)\}$ is the conditional density of Y_t given X_{t-1}, X_t with respect to a σ-finite measure μ on $\mathbb{R}^d$, $\boldsymbol{\theta}(v \times \omega) \in \Theta, s(v \times \omega) \in S$, $j = 1, \ldots, M$ are the unknown parameters. That is, $\{X_t, t \geq 0\}$ is a Markov chain, given $X_0, X_1, \ldots, X_T$, with $Y_1, \ldots, Y_T$ being independent. We give a formal definition as follows. $\{B_t, t \geq 0\}$ is called a hidden Markov model if there is a Markov chain $\{X_t, t \geq 0\}$ such that the process $\{(X_t, B_t), t \geq 0\}$ satisfies (1). Note that in (1), the usual parameterization $\boldsymbol{\theta}^{(j)}(v \times \omega) = \boldsymbol{\theta}^{(j)}$, and $s^{(j)}(v \times \omega) = s^{(j)}$. Moreover, $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(M)}) \in R^{dM}$ are the unknown dependency parameters, $\mathbf{s} = (s^{(1)}, \ldots, s^{(M)})$ are the unknown structure parameters, and its true value is denoted by $\boldsymbol{\theta}^*$ and $\mathbf{s}^*$. For simplicity, we will use π_i for $\pi_i^{v\times\omega}$ and p_{ij} for $p_{ij}^{v\times\omega}$. See Figure 2 for a graphical illustration of HMM.

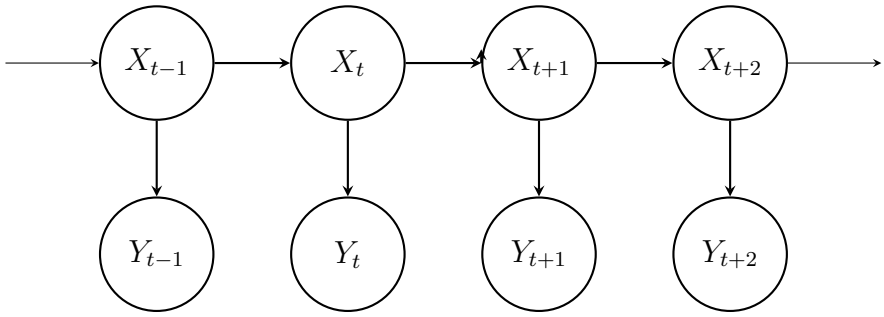


Figure 2: Graphical representation of the dependence structure of HMM, where X_t depends only on X_{t-1} and Y_t only on X_t.

For given d dimensional times series $y_1, \cdots, y_T, \in \mathbb{R}^d$ ($y_t = (y_{1t}, y_{2t}, y_{3t}, \ldots, y_{dt})^\top$) connected with unobservable (or missing) $x_1, \ldots, x_T$ from a hidden Markov model $\{B_t, t \geq 0\}$, define 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 full likelihood function given one realization of $\{x_t, y_t\}_{t=1}^T$ is:

$$p_T(y_1, \cdots, y_T; x_1, \ldots, x_T; v \times \omega) = \pi_{x_0} \prod_{t=1}^{T} p_{x_{t-1}x_t} f_{x_t}(y_t; \boldsymbol{\theta}^{(x_t)}, s^{(x_t)}), \tag{2}$$

and the likelihood for only the observations $\{y_t\}_{t=1}^T$ by marginalization :

$$p_T(y_1, \cdots, y_T; v \times \omega) = \sum_{x_0=1}^{M} \cdots \sum_{x_n=1}^{M} \pi_{x_0} \prod_{t=1}^{T} p_{x_{t-1}x_t} f_{x_t}(y_t; \boldsymbol{\theta}^{(x_t)}, s^{(x_t)}), \tag{3}$$

with the abbreviation of $p_T(y_1, \cdots, y_T; v \times \omega)$ as $p_T(y_{1:T}; v \times \omega)$

2.1 Parametrization of $f_{x_t}(y_t; \boldsymbol{\theta}^{(x_t)}, s^{(x_t)})(x_t = i)$ by HAC

The novelty of our approach lies in a special parametrization of $f_{x_t}(y_t; \boldsymbol{\theta}^{(x_t)}, s^{(x_t)})(x_t = i)$ (abbreviated as $f_i(.)$), which helps to properly understand the dynamics of a multivariate distribution. Up to now, typical parameterizations are mixtures of log-concave or elliptical symmetric densities, like those from Gamma or Poisson families, which are not flexible enough to model high dimensional time series. The advantage of the copula is that it splits the multivariate distribution into the margins and a pure dependency component. In other words, it captures the dependency between variables eliminating the impact of the marginal distributions.

Usually, copula comes into play, when one is interested in a simple but informative representation of the joint distribution of a d dimension r.v., say $Z_1, \ldots, Z_d$ with continuous cumulative distribution function (cdf)$F(\cdot)$. The theorem which guarantees the existence and uniqueness of copula functions states that there exists a unique function $C : [0,1]^d \to [0,1]$ satisfying

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

where $F_1^{-1,\mathfrak{m}}(u_1), \ldots, F_d^{-1,\mathfrak{m}}(u_d)$ are the quantile functions of the corresponding continuous marginal distributions $F_1^{\mathfrak{m}}(Z_1), \ldots, F_d^{\mathfrak{m}}(Z_d)$.

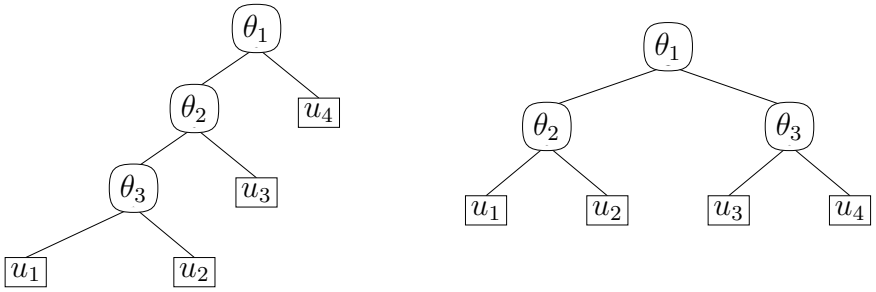


Figure 3: Fully and partially nested copulae of dimension $d = 4$ with structures $s = (((12)3)4)$ on the left and $s = ((12)(34))$ on the right

We need a further parametrization of the copula function, which are flexible enough to capture the tail dependency and have an explicit form and are simple in estimation and estimation. One candidate would be the family of Archimedean copulae, see Nelsen (2006).

$$C(u_1, \ldots, u_k) = \phi\{\phi^{-1}(u_1) + \cdots + \phi^{-1}(u_d)\}, \quad u_1, \ldots, u_d \in [0, 1], \tag{4}$$

where $\phi(.)$ is defined as the generator of the copula and most depends on the parameter θ. $\phi(.) \in \mathfrak{L} = \{\phi(.) : [0; \infty) \to [0, 1] \,|\, \phi(0) = 1, \phi(\infty) = 0; (-1)^j \phi^{(j)} \geq 0; j = 1, \ldots, \infty\}$, simplified assumptions on ϕ may be found in McNeil and Nešlehová (2009). As an example, the Gumbel generator is given by $\phi(.) = \exp(-x^{1/\theta})$ for $0 \leq x < \infty,\ 1 \leq \theta < \infty$.

However, multivariate Archimedean copulae are still restrictive, since the rendered dependency is symmetric with respect to the permutation of variables and the multivariate dependency structure depends on a single parameter of the generator function. For refined structure, we consider Hierarchical Archimedean Copulae (HAC) which are the compositions of simple Archimedean copulae. Conveniently, we denote the structure of a HAC as

$$s = \{(\ldots(i_1 \ldots i_{j_1}) \ldots (\ldots) \ldots)\},$$

where $i_\ell \in \{1, \ldots, d\}$ is a reordering of the indices of the variables. s_j denotes the structure of subcopulae with $s_d = s$. Further let the d-dimensional hierarchical Archimedean copula be denoted by $C(u_1, \ldots, u_d; s, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ the set of copula parameters. For example the fully nested HAC (see Figure 3, left) can be expressed as

$$\begin{aligned} & C(u_1, \ldots, u_d; s = s_d, \boldsymbol{\theta}) = C\{u_1, \ldots, u_d; ((s_{d-1})d), (\theta_1, \ldots, \theta_{d-1})^\top\} \\ = \ & \phi_{d-1,\theta_{d-1}}(\phi^{-1}_{d-1,\theta_{d-1}} \circ C\{u_1, \ldots, u_{d-1}; ((s_{d-2})(d-1)), (\theta_1, \ldots, \theta_{d-2})^\top\} + \phi^{-1}_{d-1,\theta_{d-1}}(u_d)), \end{aligned}$$

where $s = \{(\ldots(12)3)\ldots)d)\}$. Figure 3 presents the fully nested HAC with structure $s = (((12)3)4)$(left) and partially nested with $s = ((12)(34))$(right) in dimension $d = 4$. For more details of HAC, see Joe (1997), Whelan (2004), Savu and Trede (2006), Okhrin, Okhrin and Schmid (2009a).

The aforementioned tree structure with different generator functions would be too many to consider. To make the problem more concrete without loss of generality, we concentrate on one single generator family within one HAC, and the discussion is constrained to binary structures, i.e. at

each level of the hierarchy only two variables are joined together. This makes our model very flexible and simultaneously parsimonious.

Note for each HAC not only the parameters are unknown, but also the structure has to be determined. We adopt the computation procedure as in Okhrin et al. (2009b) to estimate the HAC structure and parameters, which leads to efficient and unbiased estimators. In this procedure, one estimates the marginal distributions either parametrically or nonparametrically. Then assuming that the marginal distributions are known, one selects the couple of variables with the strongest fit and denote the respective estimator of the parameter at the first level by $\hat{\theta}_1$ and the set of indices of the variables by I_1 . The selected couple is joined together to define the pseudo-variables $z_1 = C\{(I_1); \hat{\theta}_1, \phi_1\}$. Next, one proceeds in the same way by considering the remaining variables and the new pseudo-variable. At every level, the copula parameter is estimated by assuming that the margins as well as the copula parameters at lower levels are known. The considered procedure allows us to determine the estimated structure of the copula recursively.

Further, we incorporate the above mentioned procedure into the HMM framework. We denote the underlying Markov variable X_t as a dependency type variable. If $x_t = i$, the parameters $(s^{(i)}, \boldsymbol{\theta}^{(i)})$ determined by state $i = 1, \ldots, M$ take values on $S \times \Theta$, where S is discrete number of candidate states corresponding to different dependency structure of HAC, and Θ is a compact set in $\mathbb{R}^{d-1}$ where the HAC parameters take values. Therefore,

$$f_i(\cdot) = c\{F_1^{\mathfrak{m}}(y_1), F_2^{\mathfrak{m}}(y_2), \ldots, F_d^{\mathfrak{m}}(y_d), s^{(i)}, \boldsymbol{\theta}^{(i)}\} f_1^{\mathfrak{m}}(y_1) f_2^{\mathfrak{m}}(y_2) \cdots f_d^{\mathfrak{m}}(y_d), \tag{5}$$

with $f_i^{\mathfrak{m}}(y_i)$ are the marginal densities.

Let $\boldsymbol{\theta}^{(i)} = (\theta_{i1}, \ldots, \theta_{id-1})^\top$ be the dependency parameters of copulae starting with the lowest up to the highest level connected a fixed state $x_t = i$ and the $f_i(.)$. The multistage maximum likelihood estimator $\hat{s}^{(i)}, \hat{\boldsymbol{\theta}}^{(i)}$ solves the system

$$\left(\frac{\partial \mathcal{L}_1}{\partial \theta_{i1}}, \ldots, \frac{\partial \mathcal{L}_{d-1}}{\partial \theta_{id-1}} \right)^\top = \mathbf{0}, \tag{6}$$

$$\begin{aligned} \text{where} \quad \mathcal{L}_j &= \sum_{t=1}^{T} w_{it} l_{ij}(Y_t), \text{ for } j = 1, \ldots, d-1, \\ l_{ij}(Y_t) &= \log \Big(c\big[\{\hat{F}_m^{\mathfrak{m}}(y_{tm}, \boldsymbol{\alpha}_m)\}_{m \in \{1,\ldots,d\}};\, s^{(j)}, \{\theta_{i\ell}\}_{\ell=1,\ldots,d-1}\big] \prod_{m \in \{1,\ldots,d\}} \hat{f}_m^{\mathfrak{m}}(y_{tm}, \boldsymbol{\alpha}_m) \Big) \\ &\quad \text{for } j = 1, \ldots, d-1,\ t = 1, \ldots, T. \end{aligned}$$

where $\hat{F}_m^{\mathfrak{m}}(\cdot)$ is an estimator (either nonparametric or parametric, depending on the data) of the marginal cdf $F_m^{\mathfrak{m}}(\cdot)$ and if estimated margins are parametrical then $\hat{F}_m^{\mathfrak{m}}(\cdot) = F_m^{\mathfrak{m}}(\cdot, \hat{\boldsymbol{\alpha}}_m)$. Marginal densities $\hat{f}_m^{\mathfrak{m}}(\cdot)$ are estimated accordingly to the c.d.f.s, and w_{it} is the weight associated with state i and time t, see (17). Chen and Fan (2006) and Okhrin et al. (2009b) provide asymptotic behavior of the estimates.

2.2 Likelihood estimation

For the estimation of the HMM HAC model, we adopt the EM algorithm, Dempster, Laird and Rubin (1997). In the context of HMM, the EM algorithm is also known as the Baum-Welch algorithm. Let us recall the description in the setting of HMM on HAC.

Recall the definition of a Markov chain:

$$\begin{aligned} \mathrm{P}(X_0 = i) &= \pi_i, && (7)\\ \mathrm{P}(X_t = j | X_{t-1} = i) &= p_{ij} && (8)\\ &= \mathrm{P}(X_t = j | X_{t-1} = i, X_{t-2} = x_{t-2}, \ldots, X_1 = x_1, X_0 = x_0), \\ &\quad i, j = 1, \ldots, M \end{aligned}$$

In addition, at time t, given $X_t = i$, the distribution of Y_t is fixed. Namely, the following holds:

$$\begin{aligned} \mathrm{P}(X_t | X_{1:(t-1)}, Y_{1:(t-1)}) &= \mathrm{P}(X_t | X_{t-1}) && (9)\\ \mathrm{P}(Y_t | Y_{1:(t-1)}, X_{(1:t)}) &= \mathrm{P}(Y_t | X_t), && (10) \end{aligned}$$

where $Y_{1:(t-1)}$ stands for $\{Y_1, \ldots, Y_{t-1}\}$, $t < T$.
Recall the full likelihood $p_T(y_{1:T}; x_{1:T}; \upsilon \times \omega)$ in (2) and the partial likelihood $p_T(y_1, \ldots, y_T; \upsilon \times \omega)$ in (3), and the log likelihood :

$$\log\{p_T(y_1, \ldots, y_T; \upsilon \times \omega)\} = \log\{\sum_{x_0=1}^{M} \cdots \sum_{x_n=1}^{M} \pi_{x_0} \prod_{t=1}^{T} p_{x_{t-1}x_t} f_{x_t}(y_t; s^{(x_t)}, \boldsymbol{\theta}^{(x_t)}, s^{(x_t)})\} \qquad (11)$$

The EM algorithm suggests to estimate a sequence of parameters $\mathfrak{g}_{(i)} \stackrel{\text{def}}{=} (P_{(i)}, \mathbf{s}_{(i)}, \boldsymbol{\theta}_{(i)})$ (for the ith iteration) by iterative maximization of $\mathcal{Q}(\mathfrak{g}; \mathfrak{g}_{(i)})$ with $\mathcal{Q}(\mathfrak{g}; \mathfrak{g}_{(i)}) \stackrel{\text{def}}{=} \mathsf{E}_{\mathfrak{g}_{(i)}}(\log p_T(Y_{1:T}; X_{1:T}; \upsilon \times \omega) | Y)$, ($Y$ stands for $Y_{1:T}$), namely, one conducts the following two steps:

- (a) E-step : compute $\mathcal{Q}(\mathfrak{g}; \mathfrak{g}_{(i)})$,
- (b) M-step : choose the update parameters $\mathfrak{g}_{(i+1)} = \arg\max_{\mathfrak{g}} \mathcal{Q}(\mathfrak{g}; \mathfrak{g}_{(i)})$.

The essence of the EM algorithm is that $\mathcal{Q}(\mathfrak{g}; \mathfrak{g}_{(i)})$ can be used as a surrogate for $\log p_T(y_1, \ldots, y_T;_1, \ldots, x_T; \theta)$, see Cappé, Moulines and Rydén (2005).

In our setting, we may write $\mathcal{Q}(\mathfrak{g};\mathfrak{g}_{(i)})$ as :

$$\begin{aligned}
\mathcal{Q}(\mathfrak{g};\mathfrak{g}_{(i)}) &= \sum_{i=1}^{M} \mathsf{E}_{\mathfrak{g}_{(i)}}[\log\{\mathbf{1}\{X_0=i\}\pi_i f_i(y_0)\}|Y] && (12)\\
&\quad + \sum_{t=1}^{T}\sum_{i=1}^{M}\sum_{j=1}^{M} \mathsf{E}_{\mathfrak{g}_{(i)}}[\log\{\mathbf{1}\{X_t=j\}\mathbf{1}\{X_{t-1}=i\}p_{ij}f_j(y_t)\}|Y]\\
&= \sum_{i=1}^{M} \mathsf{E}_{\mathfrak{g}_{(i)}}[\mathbf{1}\{X_0=i\}\log\{\pi_i f_i(y_0)\}|Y] && (13)\\
&\quad + \sum_{t=1}^{T}\sum_{i=1}^{M}\sum_{j=1}^{M} \mathsf{E}_{\mathfrak{g}_{(i)}}[\mathbf{1}\{X_t=j\}\mathbf{1}\{X_{t-1}=i\}\log\{p_{ij}\}|Y]\\
&\quad + \sum_{t=1}^{T}\sum_{i=1}^{M} \mathsf{E}_{\mathfrak{g}_{(i)}}[\mathbf{1}\{X_t=i\}\log f_i(y_t)|Y] && (14)\\
&= \sum_{i=1}^{M} \mathrm{P}_{\mathfrak{g}_{(i)}}(X_0=i|Y)\log\{\pi_i f_i(y_0)\} + \sum_{t=1}^{T}\sum_{i=1}^{M}\sum_{j=1}^{M} \mathrm{P}_{\mathfrak{g}_{(i)}}(X_{t-1}=i,X_t=j|Y)\log\{p_{ij}\}\\
&\quad + \sum_{t=1}^{T}\sum_{i=1}^{M} \mathrm{P}_{\mathfrak{g}_{(i)}}(X_t=i|Y)\log f_i(y_t), && (15)
\end{aligned}$$

where $f_i(\cdot)$ is as in (5) and margins may be estimated nonparametrically as $\hat{F}_d^{\mathfrak{m}}(x) = (T+1)^{-1}\sum_{i=1}^{T}\mathbf{1}(X_i \le x)$. The E-step, in which $\mathrm{P}_{\mathfrak{g}_{(i)}}(X_t=i|Y), \mathrm{P}_{\mathfrak{g}_{(i)}}(X_{t-1}=i,X_t=j|Y)$ are evaluated, is carried out by forward-backward algorithm shown in the appendix, and the M-step is explicit in p_{ij}s and π_is. Recall $f_i(\cdot)$ is defined from last section as $c\{F_1^{\mathfrak{m}}(y_1), F_2^{\mathfrak{m}}(y_2),\ldots,F_d^{\mathfrak{m}}(y_d), s^{(i)}, \boldsymbol{\theta}^{(i)}\}f_1^{\mathfrak{m}}(y_1)f_2^{\mathfrak{m}}(y_2)\cdots f_d^{\mathfrak{m}}(y_d)$. Adding constraints to (15) yields:

$$\mathfrak{L}(\mathfrak{g},\lambda;\mathfrak{g}') = \mathcal{Q}(\mathfrak{g};\mathfrak{g}') + \sum_{i=1}^{M}\lambda_i(1-\sum_{j=1}^{M}p_{ij}) \qquad (16)$$

For the M-step, we need to take the first order partial derivative, and plug into the (16). So dependency parameters $\boldsymbol{\theta}$ and structure parameters $\mathbf{s}$ needs to be estimated iteratively, for $\boldsymbol{\theta}^{(i)}$:

$$\frac{\partial\mathfrak{L}(\mathfrak{g},\lambda;\mathfrak{g}')}{\partial\theta_{ij}} = \sum_{t=1}^{T}\mathrm{P}(X_t=i|Y)\partial\log f_i(y_t)/\partial\theta_{ij}, \qquad (17)$$

where, $j=1,\ldots,d-1$. To simplify the procedure, we adopt the HAC estimation method (6) with weights in terms of $w_{it} \stackrel{\text{def}}{=} \mathrm{P}(X_t=i|Y)$. To reduce the number of parameters to be estimated, we may fix $\pi_i, i=1,\ldots,M$ as it influences only the first observation X_o which may be consider also as given and fixed. The estimation of the transition probabilities p_{ij} follows:

$$\begin{aligned}
\frac{\partial\mathfrak{L}(\mathfrak{g},\lambda;\mathfrak{g}')}{\partial p_{ij}} &= \sum_{t=1}^{T}\frac{\mathrm{P}(X_{t-1}=i,X_t=j|Y)}{p_{ij}} - \lambda_i && (18)\\
\frac{\partial\mathfrak{L}(\mathfrak{g},\lambda;\mathfrak{g}')}{\lambda_i} &= 1-\sum_{j=1}^{M}p_{ij}. && (19)
\end{aligned}$$

Equating expression in (18) and (19) yields:

$$\hat{p}_{i,j} = \frac{\sum_{t=1}^{n} \mathrm{P}(X_{t-1} = i, X_t = j|Y)}{\sum_{t=1}^{n}\sum_{j=1}^{M} \mathrm{P}(X_{t-1} = i, X_t = j|Y)} \tag{20}$$

3 Theoretical Results

Assumptions

A.1 $\{X_t\}$ is stationary and irreducible

A.2 The family of mixture of at most M elements $\{f(y, \boldsymbol{\theta}_j, s_j) : \boldsymbol{\theta}_j \in \Theta, s_j \in S\}$ is identifiable w.r.t. the parameters and structures:

$$\sum_{j=1}^{M} \alpha_j f(y, \boldsymbol{\theta}_j, s_j) = \sum_{j=1}^{M} \alpha'_j f(y, \boldsymbol{\theta}'_j, s'_j) a.e. \Longrightarrow \sum_{j=1}^{M} \alpha_j \delta_{\boldsymbol{\theta}_j, s^{(j)}} = \sum_{j=1}^{M} \alpha'_j \delta_{\boldsymbol{\theta}'_j, s'_j}, \tag{21}$$

Define $\delta_{\boldsymbol{\theta}_j, s_j}$ as the distribution function for a point mass in $\Theta \times S$, and it only make sense to say $\boldsymbol{\theta}_j = \boldsymbol{\theta}'_j$ when $s_j = s'_j$. The property of identifiability is nothing else as the construction of the finite mixture model. For more details on mixture models we refer to McLanchlan and Peel (2000). As copula is a special form of the multivariate distribution, similar techniques may be applied to get identifiability also in the case of copulae. The family of copula mixtures has been thoroughly investigated in Caia, Chen, Fan and Wang (2006) during developing the estimation techniques. In that general case one should be careful as the general copula class is very wide and its mixture identification may cause some problems because of different densities forms. Construction of the HAC itself narrows the class. Imposing same generator functions on all levels of the HAC we restrict the family to the vector of parameters and the tree structure. Some discussion on this can been found in Okhrin et al. (2009b). Our numerical preliminary analysis shows that HAC fulfills identifiability property for all used in the study structures and parameters. To be more sure we assume throughout the paper that the copula model is identifiable.

A.3 $\{X_t\}_{t=1}^{T}$ is a time homogeneous Markov chain that is ergodic

A.4 $\mathsf{E}\{|\log f_i(y, \boldsymbol{\theta}^{(i)}, s^{(i)})|\} < \infty$, for $i = 1, \ldots, M$, $\forall s \in S$.

A.5 For every $\boldsymbol{\theta} \in \Theta$, and any particular structure considered $s \in S$,

$$\mathsf{E}[\sup_{\|\boldsymbol{\theta}' - \boldsymbol{\theta}\| < \delta} \{f_i(Y_1, \boldsymbol{\theta}', s)\}^{+}] < \infty,$$

for some $\delta > 0$.

Define $\hat{\boldsymbol{\theta}}^{(i)}, \hat{s}^{(i)}$ as $\hat{\boldsymbol{\theta}}^{(i)}(\hat{v} \times \hat{\omega})$ and $\hat{s}^{(i)}(\hat{v} \times \hat{\omega})$ with $(\hat{v} \times \hat{\omega})$ as the point over the whole parameter space $V \times \Omega$ where $p(y_{1:T}; v \times \omega)$ achieve the maximum value.

It is known that HMM is not itself identifiable as with the permutation of states $p_T(y_{1:T;v\times\omega})$ would take the same value. We assume $\boldsymbol{\theta}^{*(j)}$s and $s^{*(j)}$s are distinct in the sense that: for any $s^{*(i)} = s^{*(j)}, i \neq j$ we have $\boldsymbol{\theta}^{*(i)} \neq \boldsymbol{\theta}^{*(j)}$.

Theorem 3.1 *Assume A.1- A.5, and $\{Y_t\}_{t=1}^{T}$ are i.i.d and generated from HAC HMM model with parameters $\{s_i^*, \theta_i^*, \pi^*, [p_{ij}^*]_{i,j}\}$. The parameter $\hat{\boldsymbol{\theta}}^{(i)}$ satisfies:*

$$\lim_{n\to\infty} \mathrm{P}(\hat{\boldsymbol{\theta}}^{(i)} = \boldsymbol{\theta}^{*(i)}) = 1, \forall i, 1, \ldots, M \tag{22}$$

given the selected structure $\hat{s}^{(1)}, \hat{s}^{(2)}, \ldots, \hat{s}^{(M)}$.

Moreover,

Theorem 3.2 *Under A.1- A.5, we find the corresponding structure:*

$$\lim_{n\to\infty} \mathrm{P}(\hat{s}^{(i)} = s^{*(i)}) = 1, \forall i. \tag{23}$$

For the proof we refer to the appendix.

4 Simulation

The estimation performance of HMMHAC is evaluated in this section, Subsection I considers four states with very disjoined copulae parameters, while subsection II considers three states realistically calibrated from exchange rates data. We show that our algorithm converges after a few iterations with moderate estimation errors. Throughout the simulation study, we keep the marginal distribution fixed.

4.1 Simulation I

In this setup, a three dimensional generating process has fixed marginal distributions: $Y_{t1} \sim \mathsf{N}(0,1)$, $Y_{t2} \sim t(3)$, $Y_{t3} \sim \mathsf{N}(0,3)$. The dependence structure is modeled through HAC with Gumbel generators, and four different dependency parameters and structures correspond to four states ($M = 4$).

$$\begin{aligned}
&C\{u_3, C(u_1, u_2; \theta_1 = 4.00); \theta_2 = 1.5\}\\
&C\{u_1, C(u_2, u_3; \theta_1 = 10.0); \theta_2 = 4.0\}\\
&C\{u_2, C(u_1, u_3; \theta_1 = 30.0); \theta_2 = 10.0\}\\
&C\{u_1, C(u_2, u_3; \theta_1 = 40.0); \theta_2 = 20.0\}
\end{aligned}$$

As can be seen, we consider quite different state parameters, which helps to easily visualize dependency states. The transition probability matrix is given by:

$$P = \{p_{ij}\}_{i,j} = \begin{pmatrix} 0.985 & 0.001 & 0.003 & 0.006 \\ 0.005 & 0.990 & 0.003 & 0.003 \\ 0.005 & 0.005 & 0.991 & 0.001 \\ 0.005 & 0.004 & 0.003 & 0.990 \end{pmatrix}$$

of sample size $T = 2000$. With $\pi = (0.25, 0.25, 0.25, 0.25)^\top$. Note that we set the diagonal elements of P close 1, since it is realistic to assume the states stay the same with a high probability. Figure 4 represents underlying states, and marginal plot of the generated three dimensional time series. A state switching pattern is not evident from the marginal plots. Figure 5 however clearly displays the switching of dependency patterns. The black, red, green, blue dots corresponding to the observations from different states.

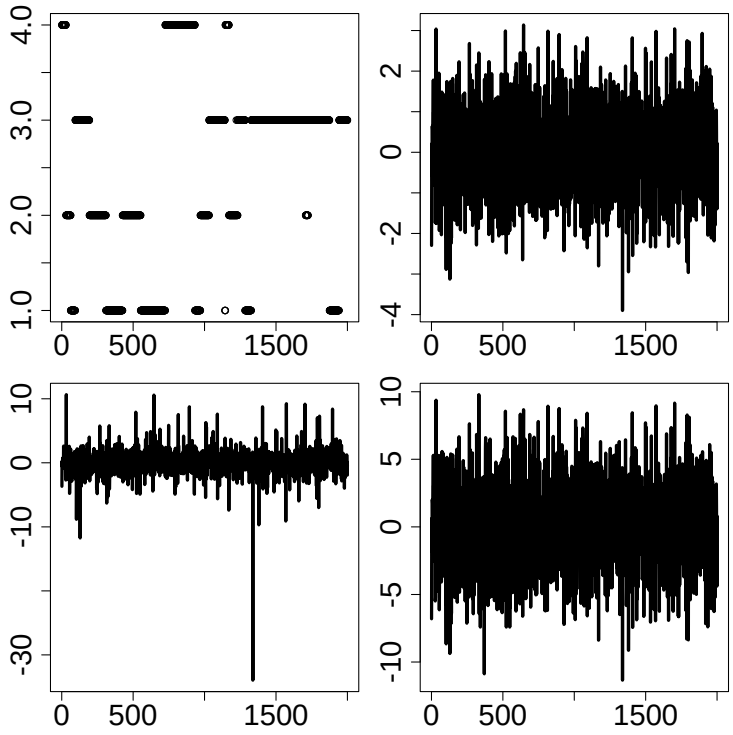


Figure 4: The underlying sequence x_t (upper left panel), marginal plots of (y_{t1}, y_{t2}, y_{t3}).

Figure 6 displays the first 7 iterations a (The parameters stay constant after). Since starting values play an important role, a moving window estimation is proposed to decide the initial parameters. The blue and the red dotted line show respectively how the estimators behave with the initial values close to the true (red) and initial values obtained from our algorithm (blue). The upper panel of Figure 6 shows the number of wrongly estimated states at each iteration; the middle panel represents the (L_1) difference of the true transition matrix from the estimated ones; the lower panel is the sum of estimated parameter errors of the four states with the correctly estimated states. One can see that our choice of initial values can perform as good as the true one.

4.2 Simulation II

Let us consider now a Monte Carlo setup where the setting employs parameters calibrated from data, see application I. The three states with $M = 3$ are taken as follows:

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

the transition matrix is chosen as:

$$P = \begin{pmatrix} 0.72 & 0.15 & 0.13 \\ 0.23 & 0.64 & 0.13 \\ 0.03 & 0.02 & 0.95 \end{pmatrix}$$

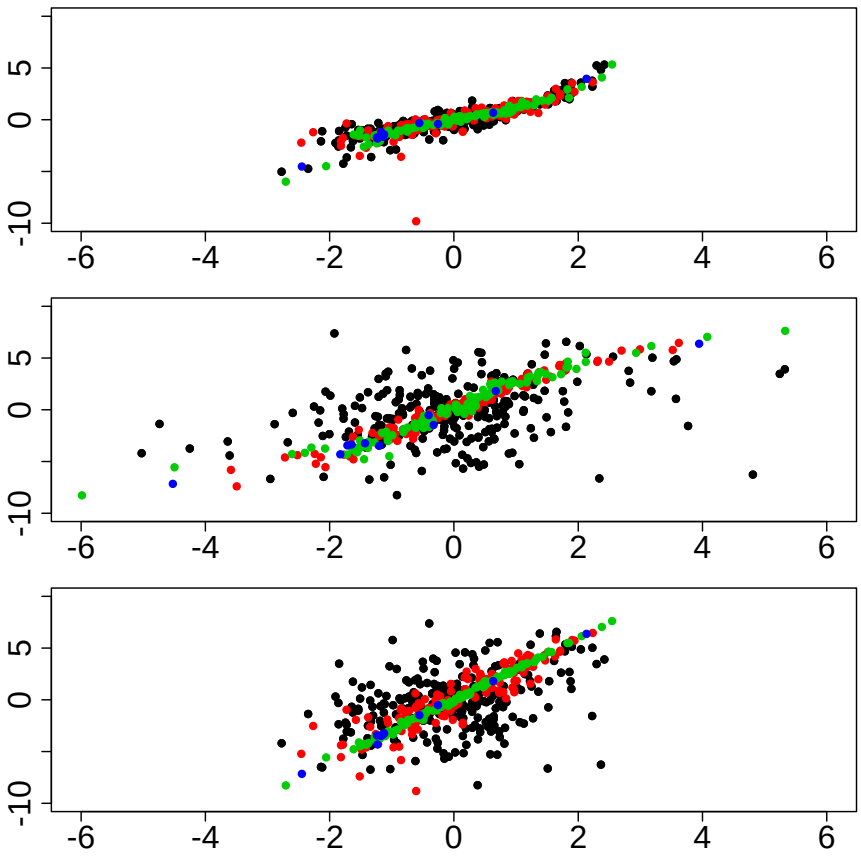


Figure 5: Snapshots of pairwise scatter plots of dependency structures ($t = 500, \ldots, 1000$), the 1st against 2nd (upper), the 2nd against 3rd (middle), and the 1st against 3rd (lower).

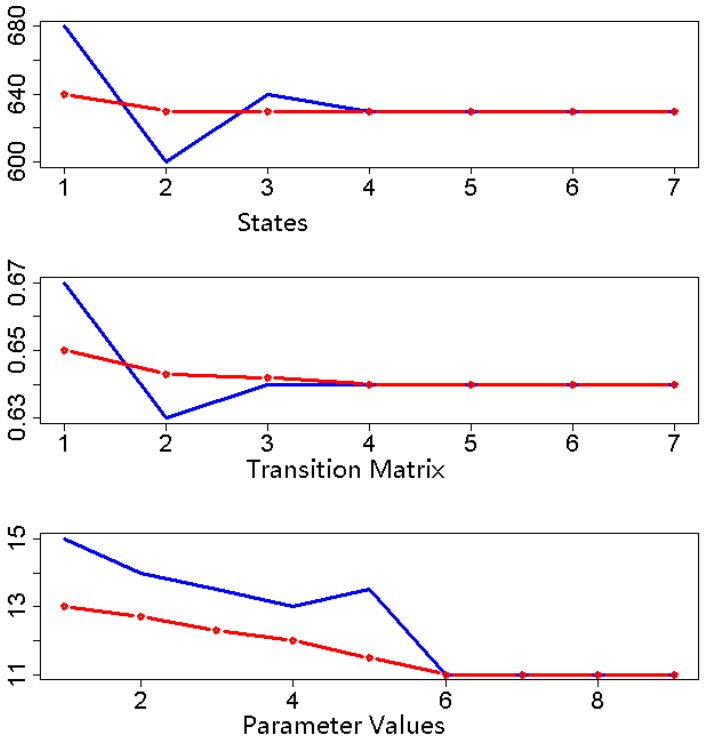


Figure 6: The convergence of states (upper panel), transition matrix (middle panel), parameters (lower panel). Estimation starts from near true value (red); starts from values attained by our proposal (blue)

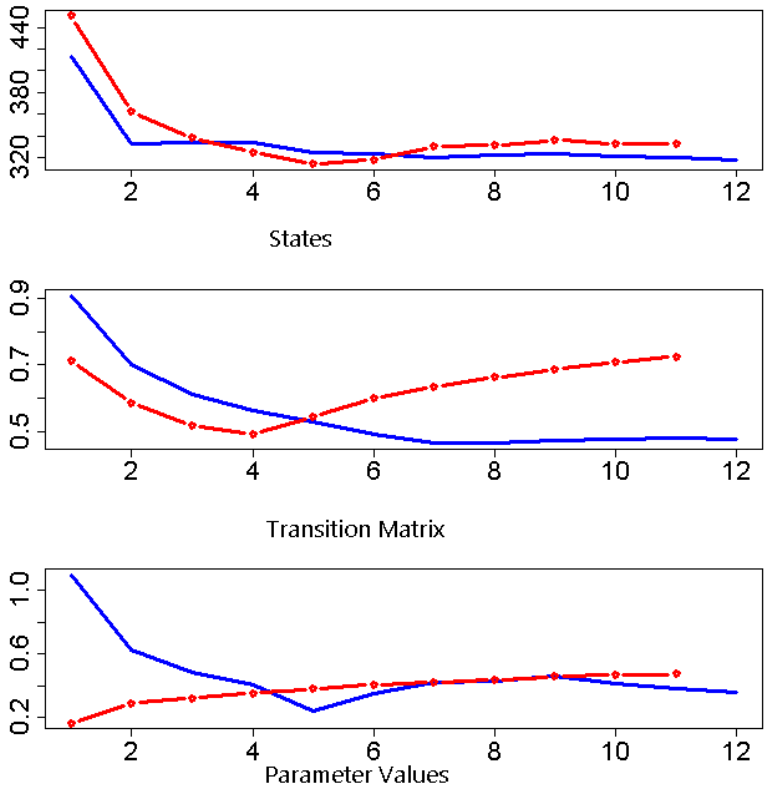


Figure 7: The convergence of states (upper panel), transition matrix (middle panel), parameters (lower panel). Estimation starts from near true value (red); starts from values attained by our proposal (blue)

sample size $T = 2000$. The iteration procedure stops after 12 steps. Figure 7 presents respectively the deviation of estimated states, transition matrix, and parameters from their true value . The estimation error is presented in the same fashion as Figure 6. To judge the estimation quality, a histogram of the estimation error from 100 samples is presented in Figure 8. It can be seen that on average only %15 of the states can not be correctly estimated.

5 Applications

To see how HMM HAC performs on a real data set, applications on financial and rainfall data are offered. A good model for the dynamics of exchange rates give insights into exogenous economic conditions, like the business cycle. It is also helpful for portfolio risk management and decisions on asset allocation. We demonstrate the forecast performance of the proposed technique by estimating VaR of the portfolio and compare it with multivariate Garch models like DCC, BEKK, etc. The backtesting results show that VaR calculated from HMMHAC performs significantly better.
The second application is on modeling rainfall process. HMM is a conventional model for rainfall data, however, bringing HMM and HAC together for modeling the multivariate rainfall process is

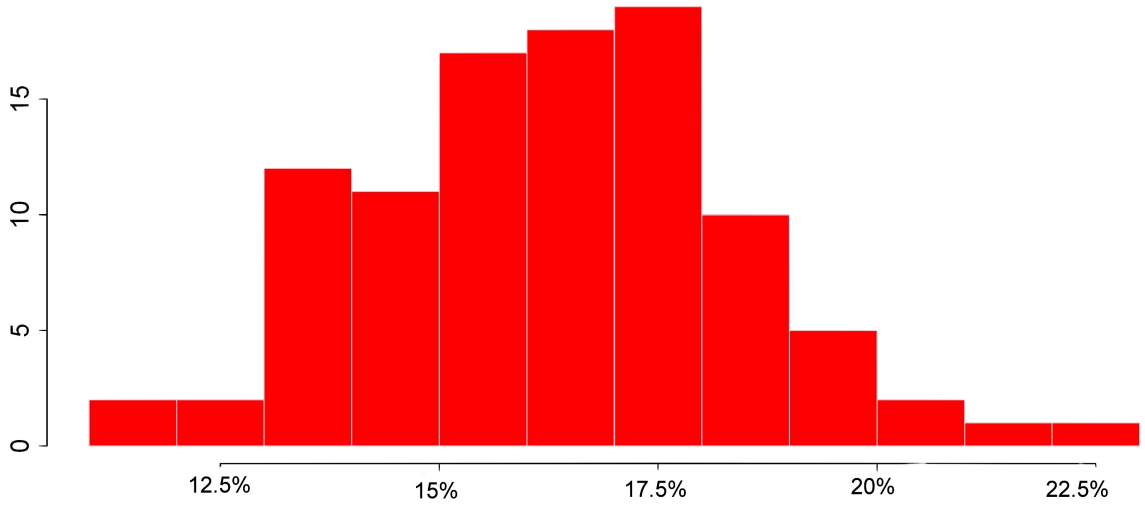


Figure 8: The error of misidentification of states by 100 samples

totally new in our work. We illustrate the estimation procedure and evaluate its performance by checking how far the model is from the reality.

5.1 Application I

5.1.1 Data

The data set consists of the daily values for the exchange rates JPY/EUR, GBP/EUR and USD/EUR. The covered period is [4.1.1999; 14.8.2009], resulting in 2771 observations, Härdle et al. (2011).

To eliminate intertemporal conditional heteroscedasticity we fit to each marginal time series of log-returns a univariate GARCH(1,1) process

$$Y_{j,t} = \mu_{j,t} + \sigma_{j,t}\varepsilon_{j,t} \text{ with } \sigma^2_{j,t} = \omega_j + \alpha_j \sigma^2_{j,t-1} + \beta_j (Y_{j,t-1} - \mu_{j,t-1})^2 \tag{24}$$

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

The residuals exhibit the typical behavior: they are not normally distributed, which motivates nonparametric estimation of the margins. From the results of the Box-Ljung test, whose p-values are $0.73, 0.01$ and 0.87 for JPY/EUR, GBP/EUR and USD/EUR, we conclude that the autocorrelation of the residuals is strongly significant only for GBP/EUR rate. After this intertemporal correction we work only with the residuals.

The dependency variation is measured by Kendall and Pearson's correlation coefficients: Figure 9 shows the variation of both coefficients calculated in a rolling window of width $r = 250$. Their dynamic behavior is similar, but not identical. This motivates once more a time varying copula based model.

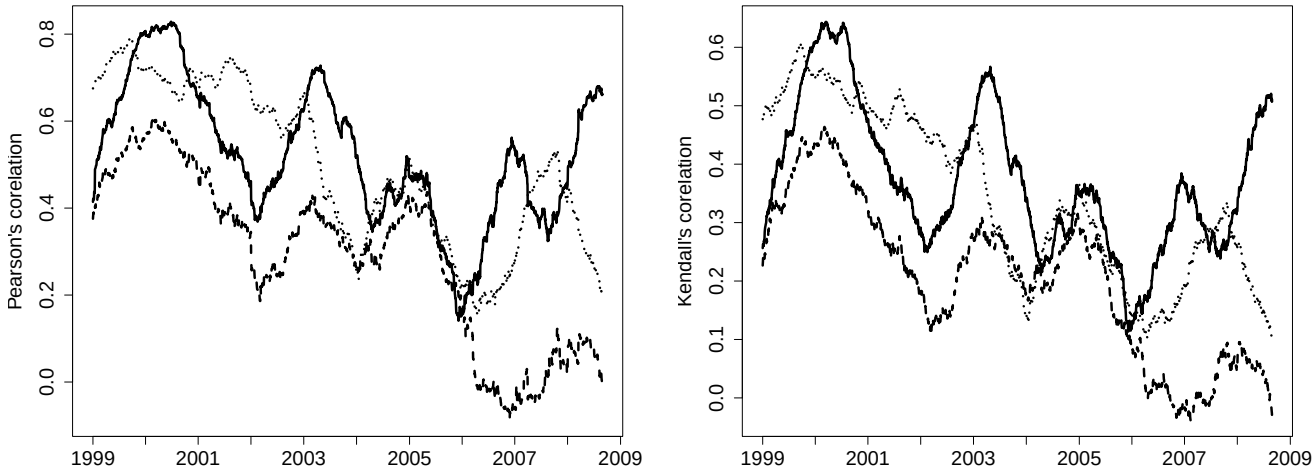


Figure 9: Rolling window estimators of Pearson's (left) and Kendall's (right) correlation coefficients between the GARCH(1,1) residuals of exchange rates: JPY and USD (solid line), JPY and GBP (dashed line), GBP and USD (dotted line). The width of the rolling window is set to 250 observations.

5.1.2 Fitting an HMM model

Figure 1, 10, and 11 summarize the analysis using three methods: moving window, LCP, HMMHAC. LCP uses moving windows, with varying sizes. To be more specific, LCP is a multiple testing scaling technique which determines a local homogeneous window at each time point Härdle et al. (2011). In contrast to LCP, HMMHAC is based on a global modeling concept rather than a local one. One observes, a relative smooth changes of parameters, see Figure 1 and 10 . HMMHAC is as flexible as LCP as can be seen from Figure 1, 10, and 11, since the structure estimated taken also three values and confirms with the variations of structures estimated from LCP. Moreover, the moving window analysis or LCP can serve as guidelines for choosing the initial values for our HMMHAC. Figure 12 displays the number of states for HMMHAC for rolling windows with a length of 500 observations.

A VaR estimation example is to show the good performance of HMMHAC. We generate $N = 10^4$ pathes with $T = 2219$ observations, and $|W| = 1000$ combinations of different portfolios, where $W = \{(1/3, 1/3, 1/3) \bigcup [\mathbf{w} = (w_1, w_2, w_3)]\}$, with $w_i = w_i' / \sum_{i=1}^{3} w_i'$, $w_i' \in U(0, 1)$. The Profit Loss (P&L) function of a weighted portfolio based on assets y_{td} is $L_{t+1} \stackrel{\text{def}}{=} \sum_{d=1}^{3} w_i (y_{t+1d} - y_{td})$, with weights $\mathbf{w} = (w_1, w_2, w_3) \in W$. The VaR of a particular portfolio at level $0 < \alpha < 1$ is defined as $VaR(\alpha) \stackrel{\text{def}}{=} F_L^{-1}(\alpha)$, where the $\hat{\alpha}_{\mathbf{w}}$ is estimated as a relative fraction of violations, see Table 1:

$$\hat{\alpha}_{\mathbf{w}} \stackrel{\text{def}}{=} T^{-1} \sum_{t=1}^{T} \mathbf{I}\{L_t < \widehat{VaR}_t(\alpha)\},$$

and the distance between $\hat{\alpha}_{\mathbf{w}}$ and α is as:

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

If the portfolio distribution is i.i.d., and a well calibrated model is properly mimicking the true underlying asset process, $\hat{\alpha}_{\mathbf{w}}$ is close to its nominal level α. The performance is measured through an average of $\alpha_{\mathbf{w}}$ over all $|W|$ portfolios see Table 1.

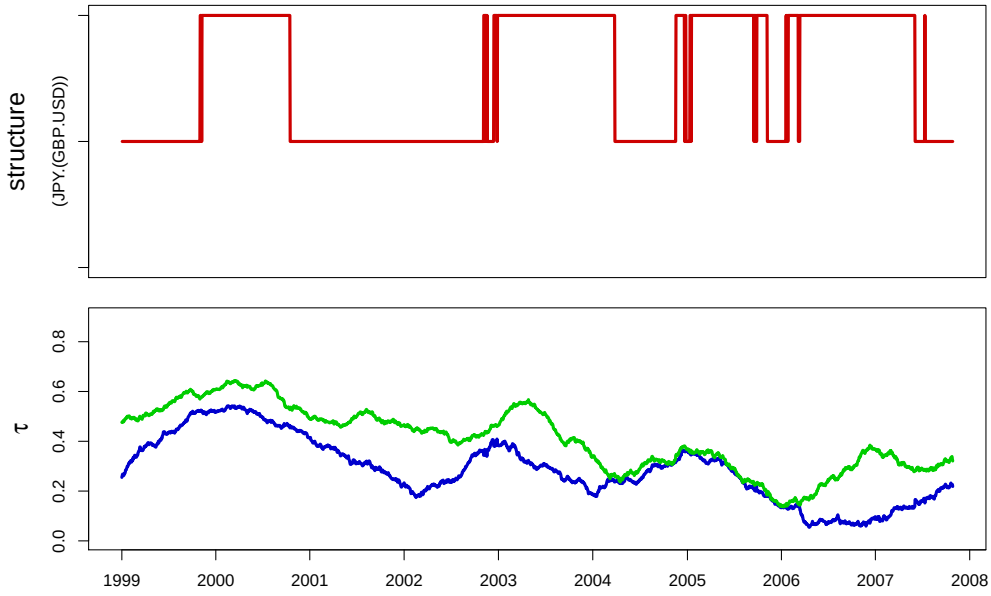


Figure 10: Rolling window for Exchange Rates: structure (upper) and dependency parameters (lower, θ_1 and θ_2) for Gumbel HAC. $w = 250$.

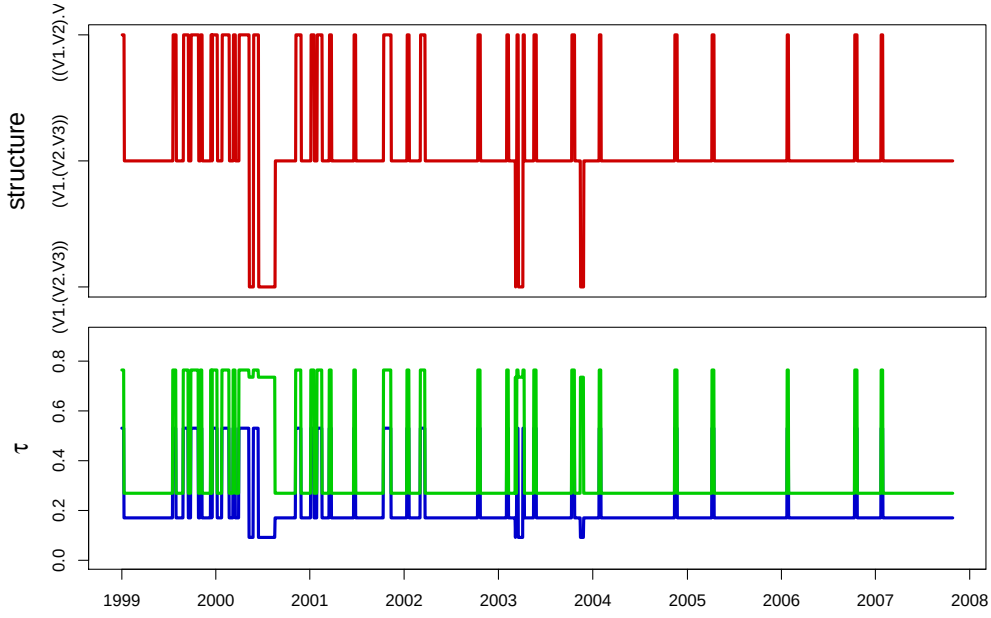


Figure 11: HMM for Exchange Rates: structure (upper) and dependency parameters (lower, θ_1 and θ_2) for Gumbel HAC.

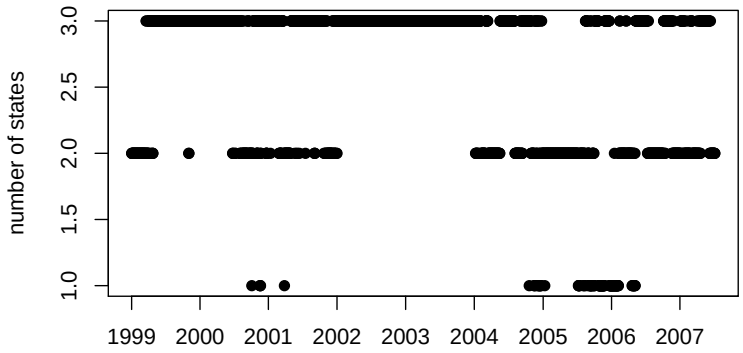


Figure 12: Plot of estimated number of states

	Window\α	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 1: VaR backtesting results, $\bar{\hat{\alpha}}$, where "Gum" denotes the Gumbel copula and "RGum" the rotated Gumbel one.

We considered four main models: HMMHAC for 500 observation windows for Gumbel and rotated Gumbel ; multiple rolling window with 250 observations windows; LCP with $m_0 = 20$ and $m_0 = 40$ with Gumbel Copulae; DCC based on 500 observation windows. For all the models we made an out of sample forecast. To better evaluate the performance we calculated the average and SD of e_W as.

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

Table 1 and 2 show backtesting performance for the described models. One concludes that HMMHAC performs better than the concurring moving window, LCP, DCC, as A_w and D_w are typically smaller.

5.2 Application II

Our goal is to propose a realistic model for rainfall, which can be used to forecast or simulate rainfall. The difficulty for modeling precipitation data is the nonzero point mass at zero of the rainfall distribution. Another difficulties arises when one incorporates spatial relationships, see

Window\α		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 2: Robustness relative to $A_W(D_W)$

Ailliot, Thompson and Thomson (2009) for an HMM application. However, Ailliot et al. (2009) only consider Gaussian dependency among locations, and the method is computationally expensive.

We extend Ailliot et al. (2009) to a copula framework. Different from application I, the marginal distribution here will be varying over states. We propose two methods in modeling marginal distributions, one is considering y_{tk} to be censored normal distributions, with the following equation:

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

with $k = 1, \cdots, d$ as the location, $\varphi(\cdot)$ as the standard normal density, $p_k^{x_t}$ as the rainfall occurrence probability for the location k and state x_t, and $\mu^{x_t}(k), \sigma^{x_t}(k)$ being mean and standard deviation parameters at time t, for location k.

A second proposal for the marginal distributions are gamma distributions:

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

where again the $\alpha(k)^{x_t}, \beta(k)^{x_t}$ are the shape and scale parameter for state x_t and location k. We take the joint distribution function to be a truncated version of a continuous copulae function, with the copulae density $c_d(\cdot)$ denoted through:

$$c_d(\mu, \theta) = \begin{cases} c_c(\mu, \theta) & , y_{tk} > 0, \forall k \\ \partial C_c(\mu, \theta)/\partial\mu_{k_1} \ldots \partial\mu_{k_B} & , k_i \in \{y_{tk_i} > 0\}, i \in 1, \ldots, E \end{cases} \tag{25}$$

where E is denoted as the number of wet places among the d locations, the $C_c(\cdot)$ as the continuous copulae function, and $c_c(.)$ as the continuous copulae density. Our formulation is simpler than Ailliot et al. (2009) since the copulae has closed form c.d.f., so we do not need additional effort to calculate integration. Representation in (25) is however more general, as we consider copulae to capture the dependency.

Assume that the daily rainfall observations from the same month are yearly independent realizations of a common underlying hidden Markov model, whose states represents different weather

Figure 13: Map of Guangxi, Guangdong, Fujian in China

types. So the likelihood is different. As an example, we take every June's daily rainfall.

$$\begin{aligned}
& \log p_T(y_{1:T}, x_{1:T}; v \times \omega) \\
= \; & \log\{\sum_{i=1}^{M} \mathbf{1}\{x_0 = i\}\pi_i f_i(y_0)\} + \sum_{t=1}^{T} \log\{\sum_{i=1}^{M}\sum_{j=1}^{M} \mathbf{1}\{x_t = j\}\mathbf{1}\{x_{t-1} = i\}p_{ij} f_j(y_t)\} \\
= \; & \sum_{i=1}^{M} \mathbf{1}\{x_0 = i\} \log\{\pi_i f_i(y_0)\} + \sum_{t=1}^{T}\sum_{i=1}^{M}\sum_{j=1}^{M} \mathbf{1}\{x_t = j\}\mathbf{1}\{x_{t-1} = i\} \log\{p_{ij} f_j(y_t)\} \\
& + \sum_{t\in B}\sum_{i=1}^{M}\{\mathbf{1}\{x_t = i\}\{\log(\pi_i)\} - \sum_{j=1}^{M} \mathbf{1}\{x_t = j\}\mathbf{1}\{x_{t-1} = i\} \log(p_{ij})\}.
\end{aligned}$$

where B is the set of days as the first day in June for each year. We use here 50 years of rainfall data from three locations in China, Guangxi, Guangdong, Fujian (Figure 13). The graphical correlation can naturally be captured by the fitting of different copulae state parameter.

Table 3 presents with a truncated Gumbel the estimated three states, the corresponding different marginal distributions and copula parameters, with estimated initial probability: $\hat{\pi}_{X_t} = (0.298, 0.660, 0.042)$ and estimated transition probability matrix:

$$\begin{pmatrix} 0.590 & 0.321 & 0.298 \\ 0.188 & 0.742 & 0.660 \\ 0.329 & 0.271 & 0.042 \end{pmatrix}.$$

In our data situation, gamma distributions fit better as marginal. The states filtered out represents different weather types. The third states is the most humid state with a high rainfall occurrence probabilities, while the second states are drier, and the first are the driest. From the parameters of the gamma distributions, one sees the variance increases from the first to the third states, which indicates a higher chance for heavy rainfall for the humid states.

To validate our model, 1000 sample of artificial time series of 1500 observations are generated from the fitted model and compared with the original data. Table 4 presents the true Pearson

X_t	Shape	Scale	Occur Prob
1	(0.442,0.429,0.552)	(139.33,116.70,169.66)	(0.252,0.256,0.439)
2	(0.671,0.618,0.561)	(273.83,253.25,427.46)	(0.806,0.786,0.683)
3	(0.636,1.125,0.774)	(381.09,264.83,514.08)	(0.667,1.000,0.944)

Table 3: Rainfall occurrence probability and shape, scale parameters estimated from HMM (data $1957-2006$) .

Location	True	$\widehat{\text{Corr}}(Y_{t,1}, Y_{t,2})$
$1-2$	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 4: True correlations, simulated averaged correlations from 1000 samples their 5%confidence intervals. 1 Fujian, 2 Guangdong, 3 Guangxi

correlation compared with the estimated ones from the generated time series. The 5% confidence intervals of the estimators cover the true correlation, which implies that the simulated rainfall can describe the real correlation of the data quite well. Figure 14 shows a marginal plot of the log survival function derived from the empirical cdf of the real data and generated data. The log survival function is a transformation of the marginal cdf $F^{\mathfrak{m}}(y_{tk})$:

$$\log\{1 - F^{\mathfrak{m}}(y_{tk})\}. \tag{26}$$

Again we show that the 95% confidence interval can cover the true curve fairly well.

Figure 15 are the autocorrelation and cross autocorrelation of the real data and the generated time series. Unfortunately, our generated time series do not show the similar auto correlation and cross auto correlation. Since there is usually more than one significant lag of auto correlation or cross correlation, but the simulated time series mostly only have one lag.

6 Conclusion

In this project, we propose a dynamic model for multivariate time series with non Gaussian dependency. The idea has an easy extension to HMM for general Copula models, and implies a rich class further work on dynamic models for dependency structures. This method is helpful in studying the financial contagion at extreme level over time, and naturally it can help to derive the conditional risk measures, such as CoVaR. As we have shown, dynamics copula models are good enough for mimicking financial markets as well as nature.

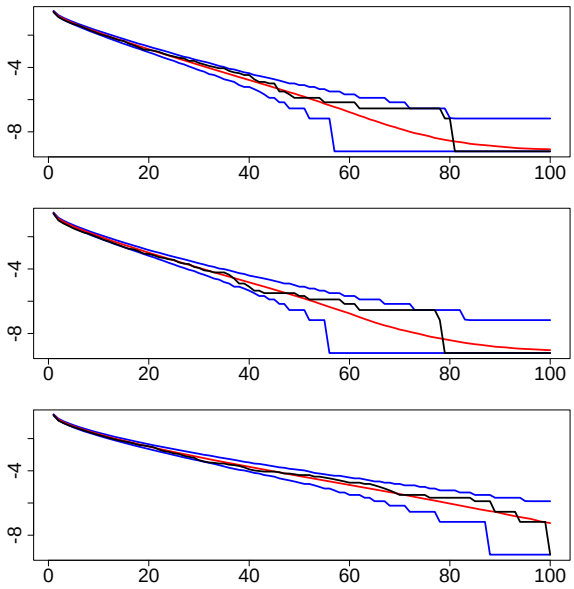


Figure 14: Log-survivor-function (red) and 95% prediction intervals (blue) of the simulated distribution for the fitted model with sample log-survivor-function superimposed (black)

7 Appendix

7.1 Proof of Theorem 3.1, 3.2

Recall the associated parameter space being $V \times \Omega$, where V consists of a set of discrete finite elements and Ω is associated with parameters $\boldsymbol{\theta}, [p_{ij}]_{i,j}$. Defined $\mathbf{s}^*$ and $\boldsymbol{\theta}^*$ is associated with the point $v^0 \times \omega^0$ in the parameter space, consider the following definitions:

$$q_T(Y_{1:T}; v^0 \times \omega^0) \stackrel{\text{def}}{=} \max_{j \in 1,\ldots,M} p_T(y_{1:T} | x_1 = j, ; v^0 \times \omega^0) \tag{27}$$

$$H(v^0 \times \omega^0) \stackrel{\text{def}}{=} \mathsf{E}_{v^0 \times \omega^0}\{-\log p(Y_0 | Y_{-1}, Y_{-2}, \ldots ; v_0 \times \omega_0)\},$$

where $Y_{-1}, \ldots, Y_{-T}$ are finite number of past values of the process.

$$H(v^0 \times \omega^0, v \times \omega) \stackrel{\text{def}}{=} \mathsf{E}_{v^0 \times \omega^0}\{\log p_T(Y_{1:T}; v \times \omega)\}$$

Theorem 7.1 (Leroux (1992)) *Under A.1- A.5*

$$\begin{aligned} \lim_{T \to \infty} T^{-1} \mathsf{E}_{v^0 \times \omega^0}\{\log p_T(Y_{1:T}; v^0 \times \omega^0)\} &= -H(v^0 \times \omega^0)) \\ \lim_{T \to \infty} T^{-1} \log p_T(Y_{1:T}; v^0 \times \omega^0)) &= -H(v^0 \times \omega^0)), \end{aligned}$$

with probability 1*, under* $v^0 \times \omega^0$*, and*

$$\begin{aligned} \lim_{T \to \infty} T^{-1} \mathsf{E}_{v^0 \times \omega^0}\{\log p_T(Y_{1:T}; v \times \omega)\} &= H(v^0 \times \omega^0, v \times \omega) \\ \lim_{T \to \infty} T^{-1} p_T(Y_{1:T}; v \times \omega) &= H(v^0 \times \omega^0, v \times \omega), \end{aligned}$$

with probability 1*, under* $v_0 \times \omega_0$.

To prove the consistency of our estimated parameter, we try to restate the theorems of consistency in Leroux (1992) for our parameter space. One needs to show that for $V^c \times \Omega^c$ which does not contain any point of the equivalent class of $v^0 \times \omega^0$ (equivalent class of $v^0 \times \omega^0$ is defined in Leroux (1992)), we have with probability 1

$$\lim_{T\to\infty} \max_{v\in V^c} \log \sup_{\omega\in\Omega^c} p_T(Y_{1:T}; v \times \omega) - \log p_T(Y_{1:T}; v^0 \times \omega^0) \to -\infty, \tag{28}$$

which is implied from, for any closed subset C of Ω^c, exists a sequence of open subsets of $\mathcal{O}_{\omega_h}$ with $h = 1, \ldots, H$ with $C \subseteq \cup_{h=1}^{H} \mathcal{O}_{\omega_h}$, such that

$$\lim_{T\to\infty} \max_{v\in V^c} \max_{h} \log \sup_{\omega\in\mathcal{O}_{\omega_h}} p_T(Y_{1:T}; v \times \omega) - \log p_T(Y_{1:T}; v^0 \times \omega^0) \to -\infty. \tag{29}$$

To prove (29), we have the modified definition:

$$H(v^0 \times \omega^0, v \times \omega; \mathcal{O}_{\omega_h}) \stackrel{\text{def}}{=} \lim_{T} \log \max_{v\in V^c} \sup_{\omega'\in\omega^0} q_T(Y_{1:T}, v \times \omega')/T. \tag{30}$$

It can be derived that

$$H(v^0 \times \omega^0, v \times \omega) < H(v^0 \times \omega^0, v^0 \times \omega^0), \tag{31}$$

for $v \times \omega$ and $v^0 \times \omega^0$ does not lie in the same equivalent class. (31) is a consequence of the identifiability condition A.2, and it would lead to: $\exists \varepsilon > 0, T_\varepsilon$ and $\mathcal{O}_\omega$, such that,

$$\mathsf{E} \log \sup_{\omega'\in\mathcal{O}_\omega} q_{T_\varepsilon}(v_\omega \times \omega')/T_\varepsilon < \mathsf{E} \log q_{T_\varepsilon}(v_\omega \times \omega)/T_\varepsilon + \varepsilon < H(v^0 \times \omega^0, v^0 \times \omega^0) - \varepsilon,$$

with

$$v_\omega \stackrel{\text{def}}{=} \text{argmax}_{v\in V^c} H(v^0 \times \omega^0, v \times \omega', \mathcal{O}_\omega).$$

Also because $\max_{v\in V^c} \log \sup_{\omega'\in\mathcal{O}_\omega} p_T(Y_{1:T}, v \times \omega')/T$ and $\max_{v\in V^c} \log \sup_{\omega'\in\mathcal{O}_\omega} q_T(Y_{1:T}, v \times \omega')/T$ have the same limit value, there exists a constant $\varepsilon > 0$,

$$\lim_{T\to\infty} \max_{v\in V^c} \log \sup_{\omega'\in\mathcal{O}_{\omega_h}} p_T(y_{1:T}, v \times \omega')/T = H(v^0 \times \omega^0, v \times \omega; \mathcal{O}_{\omega_h}) \le H(v^0 \times \omega^0, v^0 \times \omega^0) - \varepsilon.$$

Then (29) follows.

7.2 Estimation and Algorithm

If original data has GARCH dependence structure, we deGARCH by taking the residuals of the GARCH model.

The estimation procedure start with initializing the formal unknown parameters, specifically, we fix M and initialize parameters $\lambda^{(0)} = \{s_{(0)}^{(i)}, \theta_{(0)}^{(i)}, \pi_{(0)}, P_{(0)}\}$ by some preliminary analysis, typically a moving window analysis.

1) We estimate $x_1, x_2, \ldots, x_T$ (the realization of the underlying Markov chain) which maximize $\mathrm{P}(Y|\lambda)$.

To achieve this goal, we use the Viterbi Algorithm, see Rabiner (1989):

- Initialization : $\delta_1(i) = \pi_i f_i(y_1)$, $1 \le i \le M$, $\psi_1(i) = 0$. However in the practice it is better to consider $\log \delta_1(i)$ as in this case convergence is faster.
- Recursion :

$$\begin{aligned} \delta_t(i) &= \max_{1\le j\le M}\{\delta_{t-1}(i)p_{ij}\}f_i(y_t), 2 \le t \le T, 1 \le j \le M, \\ \psi_t(j) &= \arg\max_{1\le i\le M} \psi_{t-1}(i)p_{ij} \end{aligned} \tag{32}$$

- Termination :

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

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

We propose to estimate the marginal distribution function of the realizations of Y_t nonparametrically, namely,

$$\hat{F}_i(s) = (T+1)^{-1}\sum_{t=1}^{T} \mathbf{1}\{y_{ti} \le s\}$$

2) After estimating the optimal sequence x^* from 1), the next step would be to update parameters $\lambda^{(0)}$. Define :

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

They can be estimated efficiently by the following algorithm:

- $\alpha_1(i) = \pi_i f_i(y), 1 \le i \le M$
- Induction : $\alpha_{t+1}(j) = \sum_{i=1}^{M} \alpha_t(i)p_{ij}f_j(y_{t+1})$. Following Rabiner (1989) we use a computationally more efficient by setting $\alpha_{t+1}(j) = \sum_{i=1}^{M} \alpha_t(i)p_{ij}f_j(y_{t+1}) / \sum_{i=1}^{M} \alpha_t(i)$
- Termination: $\mathrm{P}(Y|\lambda) = \sum_{i=1}^{M} \alpha_t(i)$
- $\beta_T(i) = 1$, $1 \le i \le M$.
- $\beta_t(i) = \sum_{j=1}^{M} p_{ij}f_j(y_{t+1})\beta_{t+1}(j)$, $t = T-1, T-2, \ldots, 1$, $1 \le i \le M$, similarly to the case with α, we define $\beta_t(i) = \sum_{j=1}^{N} p_{ij}f_j(y_{t+1})\beta_{t+1}(j) / \sum_{j=1}^{M} \beta_{t+1}(j)$

Updates of $\pi_{(0)}$ and $P_{(0)}$, define:

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

which can be estimated by:

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

Therefore, update equations for the step k are:

$$\begin{aligned} \pi_{i,(k)} &= r_i^{(k-1)}(i) \\ p_{ij,(k)} &= \frac{\sum_{t=1}^{T-1}\xi_t^{(k-1)}(i,j)}{\sum_{t=1}^{T-1}r_t^{(k-1)}(i)} \end{aligned}$$

Given the updates of $\pi_{i,(k)}$, and $p_{ij,(k)}$, the coefficients of copulae θ can be reestimated by (20) with the weights r_t for each observation. In this case structure s and parameters $\boldsymbol{\theta}$ are estimated jointly.

References

Ailliot, P., Thompson, C. and Thomson, P. (2009). Space-time modeling of precipitation by using a hidden markov model and censored gaussian distributions, *Journal of the Royal Statistical Society* **58**: 405–426.

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

Caia, Z., Chen, X., Fan, Y. and Wang, X. (2006). Selection of copulas with applications in finance, *Working Paper* . available at http://www.economics.smu.edu.sg/femes/2008/papers/219.pdf.

Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in Hidden Markov Models*, Springer Verlag.

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

Chen, X. and Fan, Y. (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspesification, *Journal of Econometrics* **135**: 125–154.

Dempster, A., Laird, N. and Rubin, D. (1997). Maximum likelihood from incomplete data via the em algorithm (with discussion), *J. Roy. Statistical Society B* **39**: 1–38.

Fuh, C.-D. (2003). SPRT and CUSUM in hidden Markov Models, *Ann. Statist.* **31**(3): 942–977.

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

Härdle, W. K., Okhrin, O. and Okhrin, Y. (2011). Time varying hierarchical archimedean copulae, *Submitted for publication* .

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters, *in* L. Rüschendorf, B. Schweizer and M. Taylor (eds), *Distribution with fixed marginals and related topics*, IMS Lecture Notes – Monograph Series, Institute of Mathematical Statistics.

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

Leroux, B. G. (1992). Maximum-likelihood estimation for hidden markov models, *Stochastic Processes and their Applications* **40**: 127–143.

McLanchlan, G. and Peel, D. (2000). *Finite Mixture Models*, Wiley.

McNeil, A. J. and Nešlehová, J. (2009). Multivariate Archimedean copulas, d-monotone functions and l_1 norm symmetric distributions, *Annals of Statistics* **37**(5b): 3059–3097.

Nelsen, R. B. (2006). *An Introduction to Copulas*, Springer Verlag, New York.

Okhrin, O., Okhrin, Y. and Schmid, W. (2009a). On the structure and estimation of hierarchical archimedean copulas, *Under Revision of Journal of Econometrics* .

Okhrin, O., Okhrin, Y. and Schmid, W. (2009b). Properties of Hierarchical Archimedean Copulas, *SFB 649 Discussion Paper 2009-014*, Sonderforschungsbereich 649, Humboldt-Universität zu Berlin, Germany. available at http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2009-014.pdf.

Patton, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation, *Journal of Financial Econometrics* **2**: 130–168.

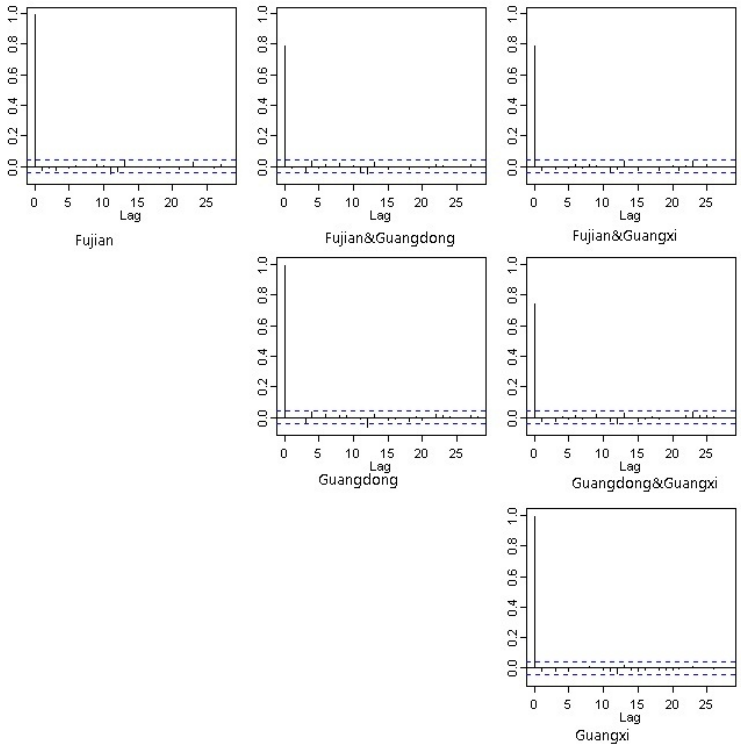
Rabiner, L. R. (1989). A tutorial on Hidden Markov Models and selected applications in speech recognition, *Proceedings of IEEE* **77**(2).

Rodriguez, J. C. (2007). Measuring financial contagion: a copula approach, *Journal of Empirical Finance* **14**: 401–423.

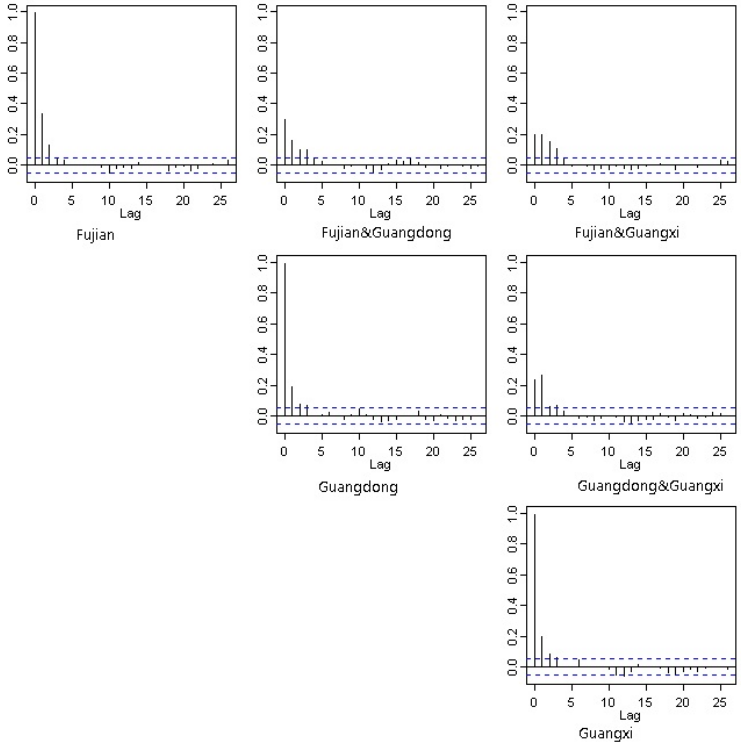
Savu, C. and Trede, M. (2006). Hierarchical Archimedean copulas, *Discussion paper*, University of Muenster.

Sklar, A. (1959). Fonctions dé repartition á n dimension et leurs marges, *Publ. Inst. Stat. Univ. Paris* **8**: 299–231.

Whelan, N. (2004). Sampling from Archimedean copulas, *Quantitative Finance* **4**: 339–352.



(a) the simulated rainfall time series.



(b) the original rainfall time series.

Figure 15: Autocorrelations and cross correlations of the simulated rainfall and original time series

SFB 649 Discussion Paper Series 2012

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "HMM in dynamic HAC models" by Wolfgang Karl Härdle, Ostap Okhrin and Weining Wang, January 2012.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".