Hidden Markov structures for dynamic copulae *

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

July 5, 2012

Abstract

Understanding the dynamics of a high dimensional non-normal dependency structure is a challenging task. A multivariate Gaussian or mixed normal time varying models are limited in capturing important types of data features such as heavy tails, asymmetry, and nonlinear dependencies. This research aims at tackling this problem by building up a hidden Markov model (HMM) for hierarchical Archimedean copulae (HAC). The HAC constitute a wide class of models for high dimensional dependencies, and HMM is a statistical technique for describing regime switching dynamics. HMM applied to HAC flexibly models high dimensional non-Gaussian time series.

In this paper we apply the expectation maximization (EM) algorithm for parameter estimation. Consistency results for both parameters and HAC structures are

[†]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

[‡]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

[§]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

established in an HMM framework. The model is calibrated to exchange rate data with a VaR application. This example is motivated by a local adaptive analysis that yields a time varying HAC model. We compare the forecasting performance with other classical dynamic models. In another, second, application we model a rainfall process. This task is of particular theoretical and practical interest because of the specific structure and required untypical treatment of precipitation data.

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 difficult on the numerical side but also on the mathematical side. Computationally speaking, integrated models for high dimensional time series become more involved when the parameter space is too large. An example is the multivariate GARCH(1,1) BEKK model that for even two dimensions has an associated parameter space of dimension 12. For moderate sample sizes, the parameter space dimension might 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 take 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 within a set of given dependencies. The specific dependence at each time t is unknown to the data analyst, but depends on the dependency pattern at time $t-1$. Therefore, hidden Markov models (HMM) naturally

2

come into play. This leaves us with the problem of specifying the set of dependencies.

An approach based on assuming a multivariate Gaussian or mixed normal is handicapped in capturing important types of data features such as heavy tails, asymmetry, and nonlinear dependencies. Such a simplification might in practice be too restrictive an assumption and might lead to biased results. Copulae are one possible approach to solving these problems. Moreover, copulae allow us to separate the marginal distributions and the dependency model, see Sklar (1959). In recent decades, copula-based models have gained 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 totally nonparametric approach that runs into the problem of the curse of dimensionality. A natural compromise is the class of hierarchical Archimedean copulae (HAC). An 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 (2009), Whelan (2004).

Many attempts have been made to obtain 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 for understanding the time evolution, and such a method is also named as local change point method (LCP), see Giacomini, Härdle and Spokoiny (2009), Härdle, Okhrin and Okhrin (2012). Figure 1 presents an analysis of HAC for exchange rate data using LCP on a moving window. One observes that the structure (upper panel) very often remains the same for a long time, and the parameters (lower panel) are only slowly varying over time. This indicates that the dynamics of HAC functions is likely to be driven by a Markovian sequence connected with the structures and parameter values. This suggests to us a different path of modeling the dynamics: instead of taking a local point of view, we adopt a global dynamic model HMM for the change of both the tree structure and the parameters of the HAC along the time horizon. In this situation, a stochastic process Y with a not directly observable
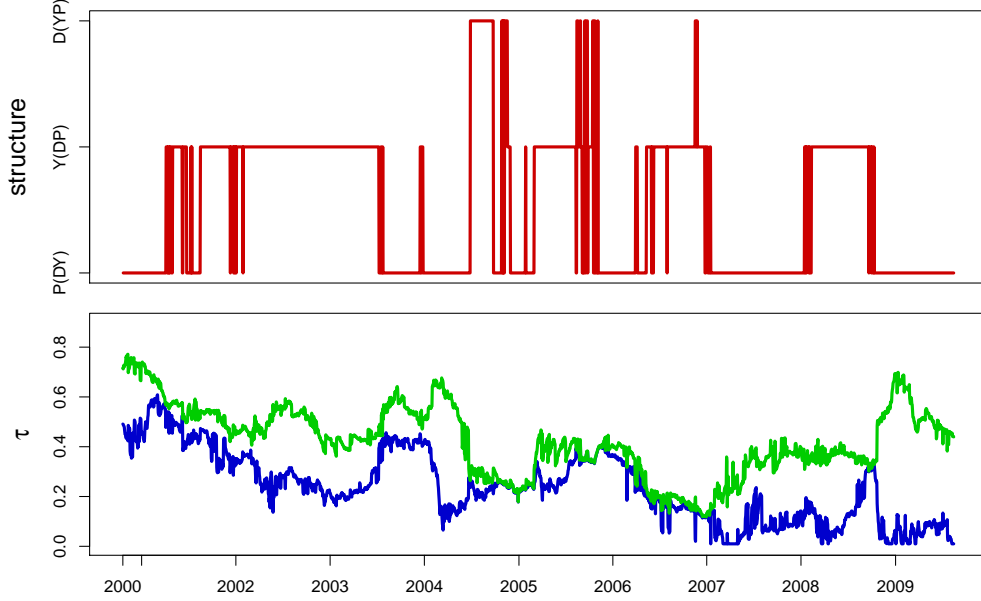
Figure 1: LCP for exchange rates: structure (upper) and parameters (lower, θ_1(green) and θ_2(blue)) for Gumbel HAC. $m_0 = 40$.

underlying Markov process X is needed to determine the state of distributions of Y. This has been widely applied to speech recognition, see Rabiner (1989), molecular biology, and digital communications over unknown channels. For 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 model, called HMM HAC, by incorporating HAC into an HMM framework. The theoretical problems such as parameter consistency and structure consistency are solved. The expectation maximization (EM) algorithm is developed in this framework for parameter estimation. See Section 2 for the model description, Section 3 for theorems about consistency. EM algorithm and computation issues are in Section 4. Section 5 is for the simulation study, and Section 6 is for applications. The technical details are put into the Appendix.

2 Model Description

2.1 Incorporating HAC into HMM

A hidden Markov model is a parameterized Markov random walk with an underlying Markov chain viewed as missing data, as in Leroux (1992), Bickel et al. (1998), and Gao and Song (2011). Specifically, in our HMM HAC framework, let $\{X_t, t \geq 0\}$ be a stationary Markov chain of order one on a finite state space $D = \{1, 2, \ldots, M\}$, with transition probability matrix $P = \{p_{ij}\}_{i,j=1,\ldots,M}$ and initial distribution $\pi = \{\pi_i\}_{i=1,\ldots,M}$.

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

Let $\{Y_t, t \geq 0\}$ be the associated observations, and they are adjoined with $\{X_t, t \geq 0\}$ in such a way that given $X_t = i, i = 1, \ldots, M$, the distribution of Y_t is fixed:

$$
\begin{aligned}
\mathrm{P}(X_t | X_{1:(t-1)}, Y_{1:(t-1)}) &= \mathrm{P}(X_t | X_{t-1}) & (3) \\
\mathrm{P}(Y_t | Y_{1:(t-1)}, X_{(1:t)}) &= \mathrm{P}(Y_t | X_t), & (4)
\end{aligned}
$$

where $Y_{1:(t-1)}$ stands for $\{Y_1, \ldots, Y_{t-1}\}$, $t < T$.

Let $f_j\{\cdot; \boldsymbol{\theta}^{(j)}, s^{(j)}\}$ be the conditional density of Y_t given $X_t = j$ with $\boldsymbol{\theta} \in \Theta, \mathbf{s} \in S$, $j = 1, \ldots, M$ being the unknown parameters. That is, $\{X_t, t \geq 0\}$ is a Markov chain, given $X_0, X_1, \ldots, X_T$, with $Y_0, Y_1, \ldots, Y_T$ being independent. Note that $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(M)}) \in \mathbb{R}^{(d-1)M}$ are the unknown dependency parameters, $\mathbf{s} = (s^{(1)}, \ldots, s^{(M)})$ are the unknown HAC structure parameters, and its true value is denoted by $\boldsymbol{\theta}^*$ and $\mathbf{s}^*$. See Figure 2 for a graphical illustration, and in Appendix 7.2 we have a more strict formulation of the definition of a HMM.
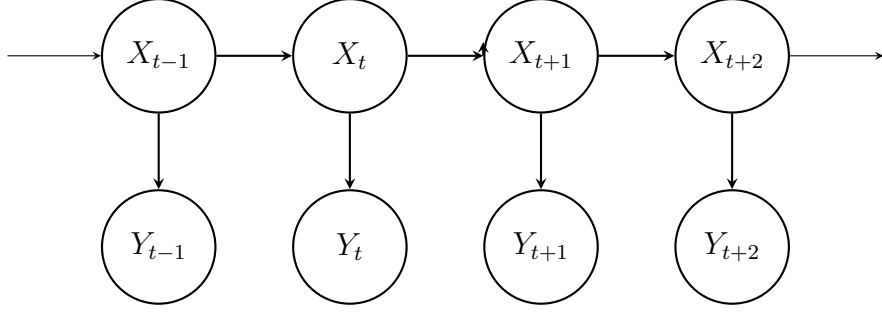
Figure 2: Graphical representation of the dependence structure of HMM, where X_t depends only on X_{t-1} and Y_t only on X_t.

For given d dimensional time series $y_1, \ldots, y_T \in \mathbb{R}^d$ $(y_t = (y_{1t}, y_{2t}, y_{3t}, \ldots, y_{dt})^\top)$ connected with unobservable (or missing) $x_1, \ldots, x_T$ from the given hidden Markov model, 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

$$p_T(y_{1:T}; x_{1:T}) = \pi_{x_0} \prod_{t=1}^T p_{x_{t-1}x_t} f_{x_t}(y_t; \boldsymbol{\theta}^{(x_t)}, s^{(x_t)}), \tag{5}$$

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

$$p_T(y_{1:T}) = \sum_{x_0=1}^M \cdots \sum_{x_n=1}^M \pi_{x_0} \prod_{t=1}^T p_{x_{t-1}x_t} f_{x_t}(y_t; \boldsymbol{\theta}^{(x_t)}, s^{(x_t)}). \tag{6}$$

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 $f_i(\cdot)$), which helps to properly understand the dynamics of a multivariate distribution. Up to now, typical parameterizations have been mixtures of log-concave or elliptical symmetric densities, such as 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 its margins and a pure dependency component. In other words, it captures the dependency between variables eliminating the impact of the marginal distributions. Technical details and properties about copulae are to be found in the Appendix 7.1.

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

$$f_i(\cdot) = c\{F_1^{\mathfrak{m}}(y_1), F_2^{\mathfrak{m}}(y_2), \ldots, F_d^{\mathfrak{m}}(y_d), \boldsymbol{\theta}^{(i)}, s^{(i)}\} f_1^{\mathfrak{m}}(y_1) f_2^{\mathfrak{m}}(y_2) \cdots f_d^{\mathfrak{m}}(y_d), \qquad (7)$$

with $f_i^{\mathfrak{m}}(y_i)$ the marginal densities, $F_i^{\mathfrak{m}}(y_i)$ the marginal cdf, $c(\cdot)$ the copula density, and see more details in Appendix 7.1.

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

$$\left(\frac{\partial \mathcal{L}_1}{\partial \theta_{i1}}, \ldots, \frac{\partial \mathcal{L}_{d-1}}{\partial \theta_{id-1}}\right)^\top = \mathbf{0}, \qquad (8)$$

$$\text{where} \quad \mathcal{L}_j = \sum_{t=1}^{T} w_{it} l_{ij}(Y_t), \text{ for } j = 1, \ldots, d-1,$$

$$l_{ij}(Y_t) = \log\left(c\left[\{\hat{F}_m^{\mathfrak{m}}(y_{tm})\}_{m \in \{1,\ldots,j\}}; \{\theta_{i\ell}\}_{\ell=1,\ldots,j-1}, s_m^{(i)}\right] \prod_{m \in \{1,\ldots,j\}} \hat{f}_m^{\mathfrak{m}}(y_{tm})\right)$$

$$\text{for } t = 1, \ldots, T.$$

where j denote the layers of a tree structure, and $\hat{F}_m^{\mathfrak{m}}(\cdot)$ is an estimator (either nonparametric with $\hat{F}_m^{\mathfrak{m}}(x) = (T+1)^{-1} \sum_{t=1}^{T} \mathbf{1}(Y_{tm} \leq x)$ or parametric $\hat{F}_m^{\mathfrak{m}}(\cdot) = F_m^{\mathfrak{m}}(\cdot, \hat{\boldsymbol{\alpha}}_m)$) of the marginal cdf $F_m^{\mathfrak{m}}(\cdot)$. The marginal densities $\hat{f}_m^{\mathfrak{m}}(\cdot)$ are estimated according to the cdfs, and w_{it} is the weight associated with state i and time t, see (12). Chen and Fan (2006) and Okhrin et al. (2009) provide the 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 full likelihood $p_T(y_{1:T}; x_{1:T})$ in (5) and the partial likelihood $p_T(y_{1:T})$ in (6), and the log likelihood:

$$\log\{p_T(y_{1:T})\} = \log\{\sum_{x_0=1}^{M} \cdots \sum_{x_n=1}^{M} \pi_{x_0} \prod_{t=1}^{T} p_{x_{t-1}x_t} f_{x_t}(y_t; \boldsymbol{\theta}^{(x_t)}, s^{(x_t)})\}. \tag{9}$$

The EM algorithm suggests estimating a sequence of parameters $\mathfrak{g}_{(r)} \overset{\text{def}}{=} (P_{(r)}, \boldsymbol{\theta}_{(r)}, \mathbf{s}_{(r)})$ (for the rth iteration) by iterative maximization of $\mathcal{Q}(\mathfrak{g}; \mathfrak{g}_{(r)})$ with

$$\mathcal{Q}(\mathfrak{g}; \mathfrak{g}_{(r)}) \overset{\text{def}}{=} \mathsf{E}_{\mathfrak{g}_{(r)}}\{\log p_T(Y_{1:T}; X_{1:T})|Y_{1:T}\}.$$

Namely, one carries out the following two steps:

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

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

The essence of the EM algorithm is that $\mathcal{Q}(\mathfrak{g}; \mathfrak{g}_{(r)})$ can be used as a surrogate for $\log p_T(y_{1:T}; x_{1:T}; \theta)$, see Cappé, Moulines and Rydén (2005).

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

$$
\begin{aligned}
\mathcal{Q}(\mathfrak{g}; \mathfrak{g}_{(r)}) &= \sum_{i=1}^{M} \mathsf{E}_{\mathfrak{g}_{(r)}}[\mathbf{1}\{X_0 = i\} \log\{\pi_i f_i(y_0)\}|Y_{1:T}] + \sum_{t=1}^{T}\sum_{i=1}^{M} \mathsf{E}_{\mathfrak{g}_{(r)}}[\mathbf{1}\{X_t = i\} \log f_i(y_t)|Y_{1:T}] \\
&\quad + \sum_{t=1}^{T}\sum_{i=1}^{M}\sum_{j=1}^{M} \mathsf{E}_{\mathfrak{g}_{(r)}}[\mathbf{1}\{X_t = j\}\mathbf{1}\{X_{t-1} = i\} \log\{p_{ij}\}|Y_{1:T}] \\
&= \sum_{i=1}^{M} \mathsf{P}_{\mathfrak{g}_{(r)}}(X_0 = i|Y_{1:T}) \log\{\pi_i f_i(y_0)\} + \sum_{t=1}^{T}\sum_{i=1}^{M} \mathsf{P}_{\mathfrak{g}_{(r)}}(X_t = i|Y_{1:T}) \log f_i(y_t) \\
&\quad + \sum_{t=1}^{T}\sum_{i=1}^{M}\sum_{j=1}^{M} \mathsf{P}_{\mathfrak{g}_{(r)}}(X_{t-1} = i, X_t = j|Y_{1:T}) \log\{p_{ij}\}, \tag{10}
\end{aligned}
$$

where $f_i(\cdot)$ is as in (7). The E-step, in which $\mathsf{P}_{\mathfrak{g}_{(r)}}(X_t = i|Y_{1:T}), \mathsf{P}_{\mathfrak{g}_{(r)}}(X_{t-1} = i, X_t = j|Y_{1:T})$ are evaluated, is carried out by the forward-backward algorithm and the M-step

is explicit in the p_{ij} and the π_i. Adding constraints to (10) yields

$$\mathcal{L}(\mathfrak{g}, \lambda; \mathfrak{g}') = \mathcal{Q}(\mathfrak{g}; \mathfrak{g}') + \sum_{i=1}^{M} \lambda_i (1 - \sum_{j=1}^{M} p_{ij}). \tag{11}$$

For the M-step, we need to take the first order partial derivative, and plug into (11). So, the dependency parameters $\boldsymbol{\theta}$ and the structure parameters $\mathbf{s}$ need to be estimated iteratively, for $\boldsymbol{\theta}^{(i)}$ ($\boldsymbol{\theta}^{(i)} = \{\theta_{i1}, \dots, \theta_{i(d-1)}\}$):

$$\frac{\partial \mathcal{L}(\mathfrak{g}, \lambda; \mathfrak{g}')}{\partial \theta_{ij}} = \sum_{t=1}^{T} \mathrm{P}_{\mathfrak{g}'}(X_t = i | Y_{1:T}) \partial \log f_i(y_t) / \partial \theta_{ij}. \tag{12}$$

To simplify the procedure, we adopt the HAC estimation method (8) with weights in terms of $w_{it} \stackrel{\text{def}}{=} \mathrm{P}_{\mathfrak{g}'}(X_t = i | Y_{1:T})$. We also fix $\pi_i, i = 1, \dots, M$ as it influences only the first observation X_0 which may be considered also as given and fixed. The estimation of the transition probabilities p_{ij} follows:

$$\frac{\partial \mathcal{L}(\mathfrak{g}, \lambda; \mathfrak{g}')}{\partial p_{ij}} = \sum_{t=1}^{T} \frac{\mathrm{P}_{\mathfrak{g}'}(X_{t-1} = i, X_t = j | Y_{1:T})}{p_{ij}} - \lambda_i, \tag{13}$$

$$\frac{\partial \mathcal{L}(\mathfrak{g}, \lambda; \mathfrak{g}')}{\partial \lambda_i} = 1 - \sum_{j=1}^{M} p_{ij}. \tag{14}$$

Equating (13) and (14) yields:

$$\hat{p}_{ij} = \frac{\sum_{t=1}^{T} \mathrm{P}_{\mathfrak{g}'}(X_{t-1} = i, X_t = j | Y_{1:T})}{\sum_{t=1}^{T} \sum_{j=1}^{M} \mathrm{P}_{\mathfrak{g}'}(X_{t-1} = i, X_t = j | Y_{1:T})} \tag{15}$$

3   Theoretical Results

Assumptions

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

A.2 The family of mixtures of at most M elements $\{f(y; \boldsymbol{\theta}^{(i)}, s^{(i)}) : \boldsymbol{\theta}^{(i)} \in \Theta, s^{(i)} \in S\}$ is

identifiable w.r.t. the parameters and structures:

$$\sum_{i=1}^{M} \alpha_i f(y; \boldsymbol{\theta}^{(i)}, s^{(i)}) = \sum_{i=1}^{M} \alpha_i' f(y; \boldsymbol{\theta}'^{(i)}, s'^{(i)}) \quad a.e. \tag{16}$$

$$\Longrightarrow \sum_{i=1}^{M} \alpha_j \delta_{\boldsymbol{\theta}^{(i)}, s^{(i)}} = \sum_{i=1}^{M} \alpha_i' \delta_{\boldsymbol{\theta}'^{(i)}, s'^{(i)}}, \tag{17}$$

defining $\delta_{\boldsymbol{\theta}^{(i)}, s^{(i)}}$ as the distribution function for a point mass in Θ associated with the structure $s^{(i)}$, noting that $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}'^{(i)}$ is only meaningful when $s^{(i)} = s'^{(i)}$. The property of identifiability is nothing else than the construction of a finite mixture model, McLanchlan and Peel (2000). As a copula is a special form of a 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) while developing 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 the different forms of the densities. The very construction of the HAC narrows this class. Imposing the same generator functions on all levels of the HAC, we restrict the family to the vector of parameters and the tree structure, see also Okhrin et al. (2009). Moreover, we restrict the classes to only binary trees with distinct parameters to avoid identifiability issues induced by the case of the same parameter values on each layers of a tree. Our preliminary numerical analysis shows that the HAC fulfills the identifiability property for all the structures and parameters used in this study.

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 $i = 1, \dots, M$, $\forall s^{(i)} \in S$.

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 $\delta > 0$.

Denote as $p_T(y_{1:T}; v, \omega)$ the density in (6) with parameters $\{v, \omega\} \in \{V, \Omega\}$ as described in the Appendix 7.2. Define $\hat{\boldsymbol{\theta}}^{(i)}, \hat{s}^{(i)}$ as $\hat{\boldsymbol{\theta}}^{(i)}(\hat{v}, \hat{\omega})$ and $\hat{s}^{(i)}(\hat{v}, \hat{\omega})$ with $(\hat{v}, \hat{\omega})$ as the point where $p_T(y_{1:T}; v, \omega)$ achieve its maximum value over the parameter space $\{V, \Omega\}$.

It is known that HMM is not itself identifiable as the permutation of states would yield the same value for $p_T(y_{1:T}; v, \omega)$. We assume therefore $\boldsymbol{\theta}^{*(j)}$s and $s^{*(j)}$s to be 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. *Under A.1–A.5, we find the corresponding structure:*

$$\lim_{T \to \infty} \max_{i \in 1, \dots, M} \mathrm{P}(\hat{s}^{(i)} = s^{*(i)}) = 1, \forall i. \tag{18}$$

Moreover,

Theorem 3.2. *Assume A.1–A.5, and $\{Y_t\}_{t=1}^T$ are i.i.d and generated from an HAC HMM model with parameters $\{s^{*(i)}, \theta^{*(i)}, \pi^*, \{p_{ij}^*\}_{i,j}\}$. The parameter $\hat{\boldsymbol{\theta}}^{(i)}$ satisfies, $\forall \varepsilon > 0$:*

$$\lim_{T \to \infty} \min_{i \in 1, \dots, M} \mathrm{P}(|\hat{\boldsymbol{\theta}}^{(i)} - \boldsymbol{\theta}^{*(i)}| > \varepsilon | \hat{s}^{(i)} = s^{*(i)}) = 0. \tag{19}$$

For the proof, we refer to the Appendix.

4 Simulation

The estimation performance of HMM HAC is evaluated in this section: subsection I considers four states with very disjoint copulae parameters, while subsection II considers three realistic states. We show that our algorithm converges after a few iterations with moderate estimation errors. Throughout the simulation study, we keep the marginal distributions 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 corresponding to four states ($M = 4$):

$$C\{u_3, C(u_1, u_2; \theta_1 = 4.00); \theta_2 = 1.5\},$$

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

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

$$C\{u_1, C(u_2, u_3; \theta_1 = 40.0); \theta_2 = 20.0\}.$$

The quite different state parameters help to easily visualize the dependency states. The transition probability matrix is

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

Sample size $T = 2000$ and $\pi = (0.25, 0.25, 0.25, 0.25)^\top$. Note that we set the diagonal elements of P close to 1, since it is realistic to assume that the states remain the same with a high probability. Figure 3 represents the underlying states and a marginal plot of the generated three dimensional time series. No state switching pattern is evident from the marginal plots. Figure 4, however, clearly displays the switching of dependency patterns. The black, red, green, and blue dots correspond to the observations from different states. The green points represent the highest correlation state, whereas the red has smaller correlation. The remaining colors blue and black represent states 1 and 2 as described above. One clearly sees that how the HMM changes the dependency structures.

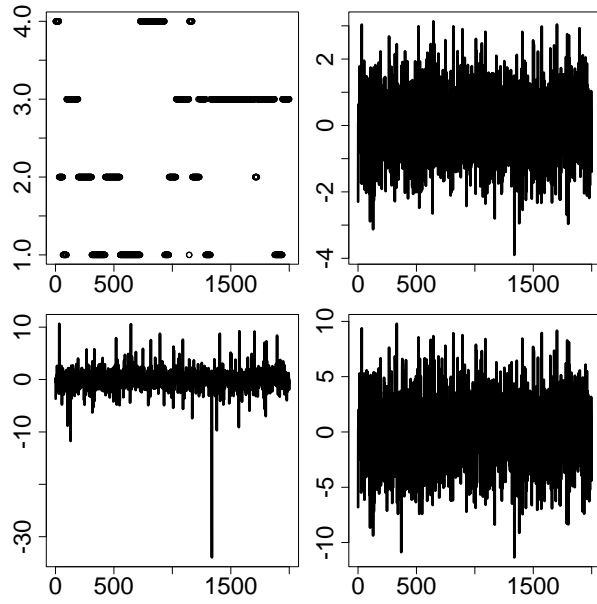Figure 5 displays the first seven iterations. (The parameters remain constant after that).

Figure 3: The underlying sequence x_t (upper left panel), marginal plots of (y_{t1}, y_{t2}, y_{t3}).
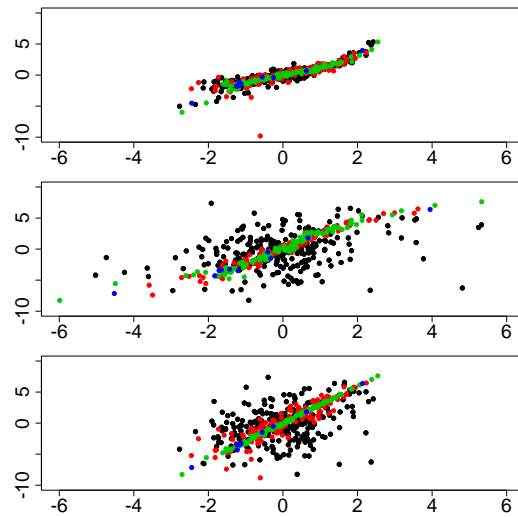

Figure 4: Snapshots of pairwise scatter plots of dependency structures ($t = 500, \ldots, 1000$), the (y_{t1}) vs. (y_{t2}) (upper), the (y_{t2}) vs. (y_{t3}) (middle), and the (y_{t1}) vs. (y_{t3})(lower).
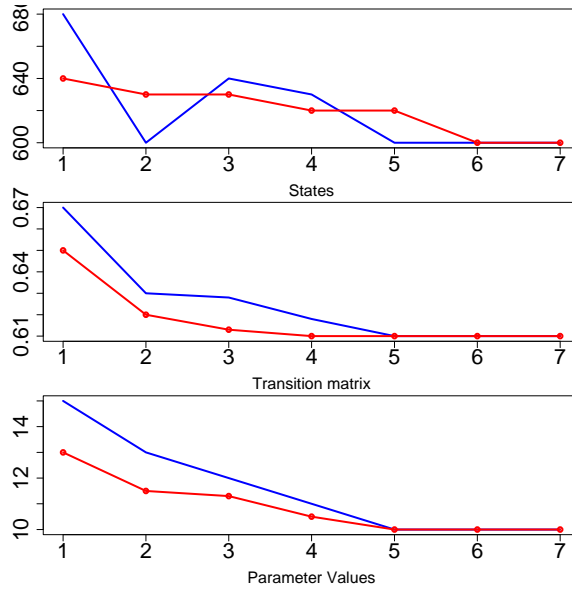
Figure 5: The convergence of states (upper panel), transition matrix (middle panel), and parameters (lower panel). Estimation starts from a near true value (red); starts from values provided by our proposal (blue).

Since the starting values may influence the result, 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 (blue) obtained from the proposed algorithm . The upper panel of Figure 5 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 the estimated parameter errors of the four states with the correctly estimated states. One can see that our choice of initial values can perform as well as the true one.

4.2 Simulation II

Let us consider now a Monte Carlo setup where the setting employs more realistic models. 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:

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

and sample size $T = 2000$. The iteration procedure stops after eleven steps. Figure 6 presents the deviations of the estimated states, the transition matrix, and the parameters from their true values. The estimation error is presented in the same fashion as in Figure 5. To judge the estimation quality, a histogram of the estimation error from 400 samples is presented in Figure 7. The proportion of the misspecified states is centered around roughly $15\% - 17\%$.

5 Applications

To see how HMM HAC performs on a real data set, applications to financial and rainfall data are offered. A good model for the dynamics of exchange rates gives insights into exogenous economic conditions, such as the business cycle. It is also helpful for portfolio risk management and decisions on asset allocation. We demonstrate the performance of our proposed technique by applying it to forecasting the VaR of a portfolio and compare it with multivariate GARCH models (DCC, BEKK, etc.) The backtesting results show
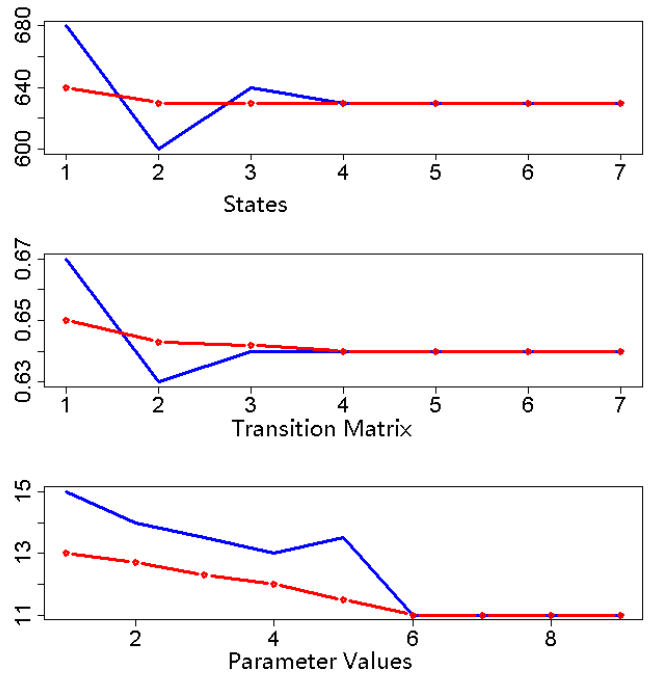
Figure 6: The convergence of states (upper panel), transition matrix (middle panel), parameters (lower panel). Estimation starts from near true value (red); starts from values attained by our proposal (blue)
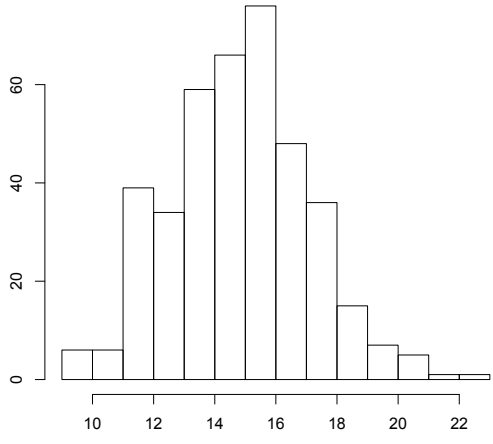

Figure 7: The percentage of misidentification of states from 400 samples

that the VaR calculated from HMM HAC performs significantly better.

The second application is on modeling a rainfall process. HMM is a conventional model for rainfall data, however, bringing HMM and HAC together for modeling the multivariate rainfall process is an innovative modeling path.

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.

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

$$Y_{j,t} = \mu_{j,t} + \sigma_{j,t}\varepsilon_{j,t} \text{ with } \sigma_{j,t}^2 = \omega_j + \alpha_j \sigma_{j,t-1}^2 + \beta_j (Y_{j,t-1} - \mu_{j,t-1})^2 \tag{20}$$

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 the GBP/EUR rate. After this intertemporal correction, we work only with the residuals.

The dependency variation is measured by Kendall's and Pearson's correlation coefficients: Figure 8 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.
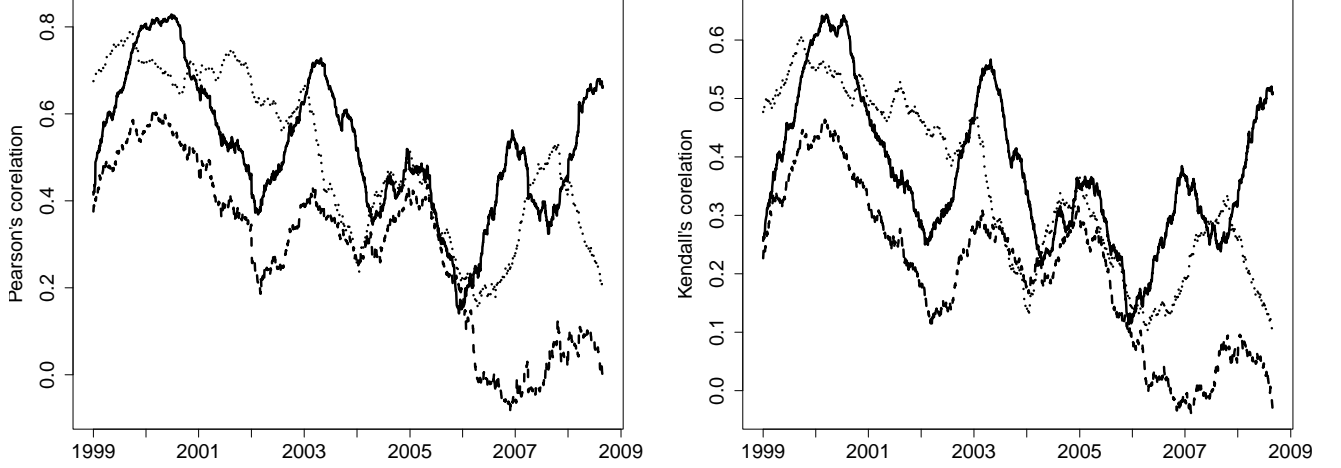
Figure 8: Rolling window estimators of Pearson's (left) and Kendall's (right) correlation coefficients between the GARCH(1,1) residuals of exchange rates: JPY and USD (solid line), JPY and GBP (dashed line), GBP and USD (dotted line). The width of the rolling window is set to 250 observations.

5.1.2 Fitting an HMM model

Figures 1, 9, and 10 summarize the analysis using three methods: moving window, LCP, and HMM HAC. LCP uses moving windows, with varying sizes. To be more specific, LCP is a scaling technique which determines a local homogeneous window at each time point Härdle et al. (2012). In contrast to LCP, HMM HAC is based on a global modeling concept rather than a local one. One observes relatively smooth changes of the parameters, see Figures 1 and 9. HMM HAC is as flexible as LCP, as can be seen from Figures 1, 9, and 10, since the structure estimated also takes three values and is confirmed by the variations of structures estimated from LCP. Moreover, the moving window analysis or LCP can serve as a guideline for choosing the initial values for our HMM HAC. Figure 11 displays the number of states for HMM HAC for rolling windows with a length of 500 observations.

A VaR estimation example is to show the good performance of HMM HAC. We generate $N = 10^4$ paths with $T = 2219$ observations, and $|W| = 1000$ combinations of different portfolios, where $W = \{(1/3, 1/3, 1/3)^\top \bigcup [\mathbf{w} = (w_1, w_2, w_3)^\top]\}$, 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
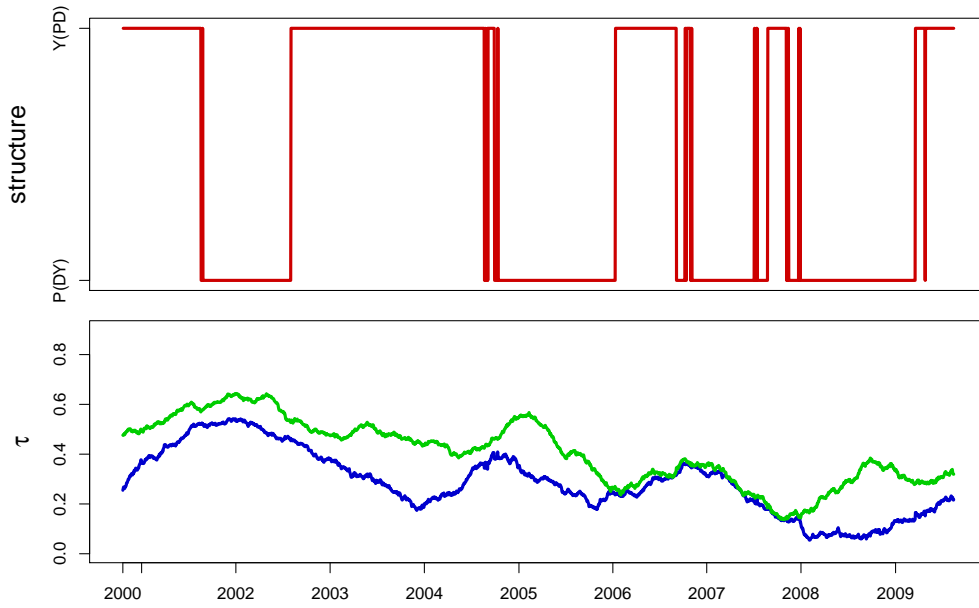
18

Figure 9: Rolling window for exchange rates: structure (upper) and dependency parameters (lower, θ_1 and θ_2) for Gumbel HAC. Rolling window size $win = 250$.
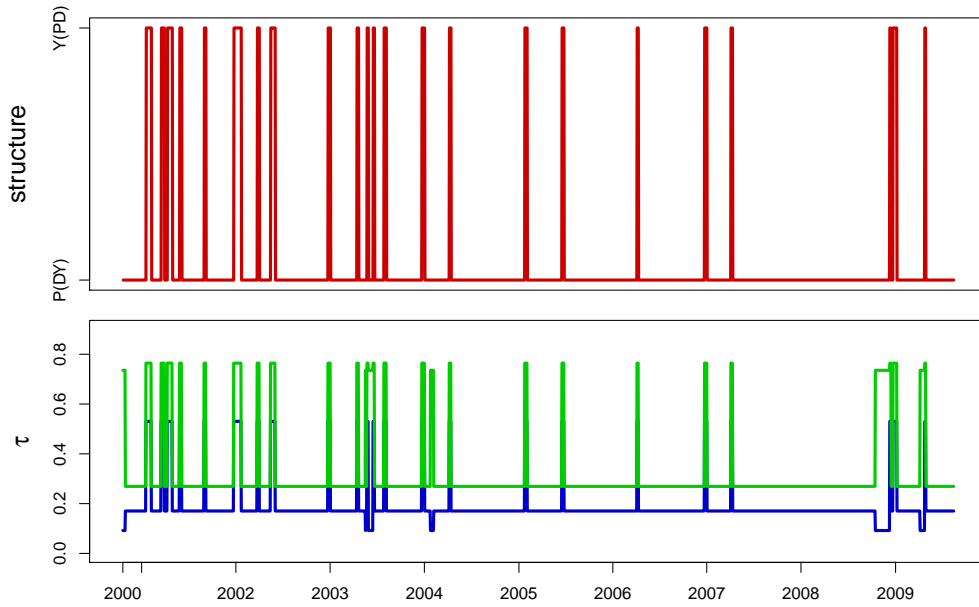

Figure 10: HMM for exchange rates: structure (upper) and dependency parameters (lower, θ_1 and θ_2) for Gumbel HAC.
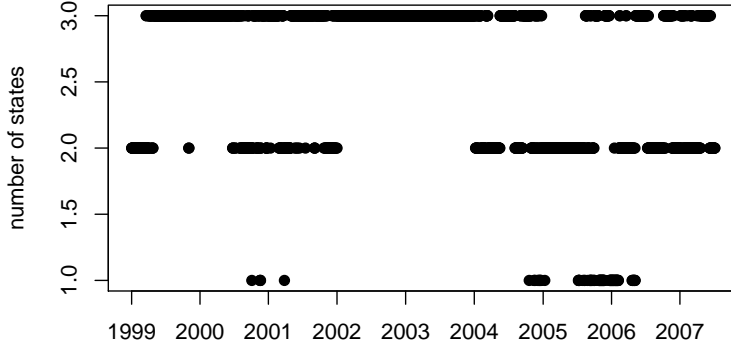
Figure 11: Plot of estimated number of states for each window

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

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

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.

We considered four main models: HMM HAC 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 (the LCP finds the optimal length of window in the past by a sequence of tests on windows of increasing sizes, m_0 is a starting window size); and DCC, see Engle (2002), 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:

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

20

	Window$\backslash\alpha$	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, $\bar{\bar{\alpha}}$, where "Gum" denotes the Gumbel copula and "RGum" the rotated survival Gumbel one.

	Window$\backslash\alpha$	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 $A_W(D_W)$

Tables 1 and 2 show the backtesting performance for the described models. One concludes that HMM HAC performs better than the concurring moving window, LCP, or DCC, as A_w and D_w are typically smaller in absolute value.

5.2   Application II

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

modeling the marginal distributions: one is to take y_{tk} to be censored normal distributions, with the following equation:

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

with $k = 1, \ldots, d$ 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 $\mu^{x_t}(k), \sigma^{x_t}(k)$ the mean and standard deviation parameters at time t for location k.

A second proposal for the marginal distributions are the gamma distributions:

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

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

$$c_d(\mu, \theta) = \begin{cases} c_c(\mu, \theta), & y_{tk} > 0, \forall k \\ \partial C_c(\mu, \theta)/\partial \mu_{k_1} \ldots \partial \mu_{k_B}, & k_i \in \{y_{tk_i} > 0\}, i \in 1, \ldots, E \end{cases} \tag{21}$$

where E denotes the number of wet places among the d locations, the C_c are the continuous copula functions, and c_c are the continuous copula densities. Our formulation is simpler than that of Ailliot et al. (2009) since the copulae have closed-form cdfs, so we do not need additional effort to calculate an integral. The representation in (21) is, however, more general, as we consider copulae for capturing the dependencies.

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

Figure 12: Map of Guangxi, Guangdong, Fujian in China

different weather types. As an example, we take every June's daily rainfall.

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

with B is the set of days which are the first day of June for each year. We use here 50 years of rainfall data from three locations in China: Guangxi, Guangdong, and Fujian (Figure 12). The graphical correlation can naturally be captured by the fitting of different copulae state parameters.

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:

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

In our data situation, gamma distributions fit better as marginals. The states filtered out represent different weather types. The third states are the most humid states, with a high

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	$\widehat{\mathrm{Corr}}(Y_{t,1}, Y_{t,2})$
$1 - -2$	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.

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 samples of artificial time series of 1500 observations were generated from the fitted model and compared with the original data. Table 4 presents the true Pearson 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 13 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 $F_k^{\mathrm{m}}(y_{tk})$:

$$\log\{1 - F_k^{\mathrm{m}}(y_{tk})\}. \tag{22}$$

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

Figure 14 contains the autocorrelations and cross-correlations of the real data and the generated time series. Unfortunately, our generated time series do not show a similar autocorrelation or cross-correlation. Since there is usually more than one significant lag of autocorrelation or cross-correlation, the simulated time series mostly only have one lag.
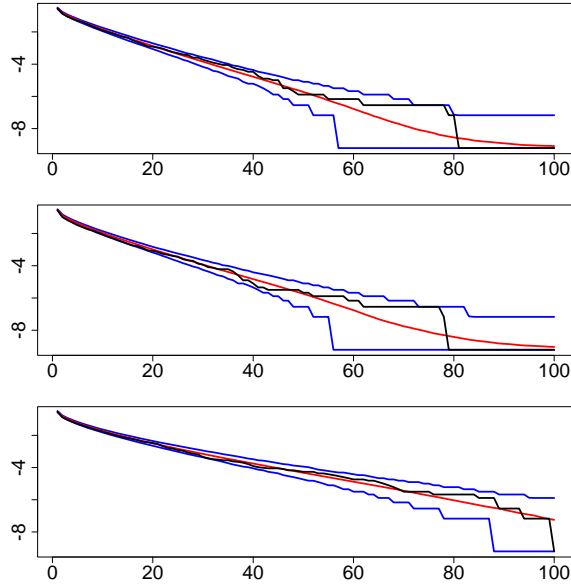
Figure 13: Log-survivor-function (red) and 95% prediction intervals (blue) of the simulated distribution for the fitted model with sample log-survivor-function superimposed (black)

This is the issue also observed in Ailliot et al. (2009). The precipitation can be modelled first by a vector autoregressive (VAR) type model, adjusted for zero observations. An alternative could be a Markov switching model, Cappé et al. (2005).

6   Conclusion

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 leads to a rich field for further work on dynamic models with dependency structures. This method is helpful in studying financial contagion at an extreme level over time, and naturally it can help in deriving conditional risk measures, such as CoVaR, Adrian and Brunnermeier (2011). As we have shown, dynamic copula models are good enough to mimic financial markets as well as nature.

7 Appendix

7.1 Copulae

Let $Z_1, \ldots, Z_d$ be r.v. with continuous cumulative distribution function (cdf) $F(\cdot)$. The Sklar theorem guarantees the existence and uniqueness of copula functions by stating 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 $F_1^{\mathfrak{m}}(Z_1), \ldots, F_d^{\mathfrak{m}}(Z_d)$.

One of the families, which are flexible enough to capture a tail dependency, have an explicit form, and are simple to estimate is the family of Archimedean copulae, see Nelsen (2006),

$$C(u_1, \ldots, u_k) = \phi\{\phi^{-1}(u_1) + \cdots + \phi^{-1}(u_d)\}, \quad u_1, \ldots, u_d \in [0,1], \tag{23}$$

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

In this work we consider less restrictive compositions of simple Archimedean copulae leading to a Hierarchical Archimedean Copula (HAC) $C(u_1, \ldots, u_d; \boldsymbol{\theta}, s)$, where $s = \{(\ldots (i_1 \ldots i_{j_1}) \ldots (\ldots) \ldots)\}$ denotes the structure of HAC, with $i_\ell \in \{1, \ldots, d\}$ being a reordering of the indices of the variables and s_j the structure of the subcopulae with $s_d = s$, and $\boldsymbol{\theta}$ is the set of copula parameters. For example, the fully nested HAC (see

Figure 15, left) can be expressed by

$$C(u_1, \ldots, u_d; \boldsymbol{\theta}, s = s_d) = C\{u_1, \ldots, u_d; (\theta_1, \ldots, \theta_{d-1})^\top, ((s_{d-1})d)\}$$

$$= \phi_{d-1,\theta_{d-1}}(\phi_{d-1,\theta_{d-1}}^{-1} \circ C\{u_1, \ldots, u_{d-1}; (\theta_1, \ldots, \theta_{d-2})^\top, ((s_{d-2})(d-1))\} + \phi_{d-1,\theta_{d-1}}^{-1}(u_d)),$$

where $s = \{(\ldots(12)3)\ldots)d)\}$. On the RHS of Figure 15 we have the partially nested HAC with $s = ((12)(34))$ in dimension $d = 4$. For more details of HAC, see Joe (1997), Whelan (2004), Savu and Trede (2010), Okhrin et al. (2009).

Not all generator functions can be mixed within one HAC. To make the problem more concrete, 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 yet also parsimonious.

Note that for each HAC not only are the parameters unknown, but also the structure has to be determined. We adopt the computation procedure in Okhrin et al. (2009) 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 denotes the corresponding 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. This procedure allows us to determine the estimated structure of the copula recursively.

7.2   Proof of Theorems 3.1 and 3.2

In the HMM HAC framework, let $\{X_t, t \geq 0\}$ with transition probability matrix $P^{v,\omega} = [p_{ij}^{v,\omega}]_{i,j=1,\ldots,M}$ and initial distribution $\pi^{v,\omega} = \{\pi_i^{v,\omega}\}_{i=1,\ldots,M}$, where $\{v, \omega\} \in \{V, \Omega\}$ de-

notes an element in the parameter space $\{V, \Omega\}$ which parametrizes this model, and q is the number of continuous parameters (note that our parameter space is partly discrete (V), and partly 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 with $Y_k = (Y_{k,1}, Y_{k,2}, \ldots, Y_{k,d})^\top$ a r.v. taking values on $\mathbb{R}^d$, is adjoined to the chain such that $\{(X_t, B_t), t \geq 0\}$ is a Markov chain on $D \times \mathbb{R}^d$ and

$$
\begin{aligned}
& \mathrm{P}\{(X_t, B_t) \in A \times (B + b) | (X_{t-1}, B_{t-1}) = (i, b)\} \quad \quad (24) \\
= \ & \mathrm{P}\{(X_1, B_1) \in A \times B | (X_0, B_0) = (i, 0)\} \\
= \ & \mathrm{P}(i, A \times B) = \sum_{j \in A} \int_{b \in B} p_{ij}^{v \times \omega} f_j\{b; \boldsymbol{\theta}^{(j)}(v \times \omega), s^{(j)}(v \times \omega)\} \mu(db),
\end{aligned}
$$

where $B, b \subseteq \mathbb{R}^d$, $A \subseteq D$, $f_j\{b; \boldsymbol{\theta}^{(j)}(v, \omega), s^{(j)}(v, \omega)\}$ is the conditional density of Y_t given X_{t-1}, X_t with respect to a σ-finite measure μ on $\mathbb{R}^d$, and $\boldsymbol{\theta}(v, \omega) \in \Theta, s(v, \omega) \in S$, $j = 1, \ldots, M$ 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. $\{B_t, t \geq 0\}$ 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 (24). Note that in (24), the usual parameterization $\boldsymbol{\theta}^{(j)}(v, \omega) = \boldsymbol{\theta}^{(j)}$, and $s^{(j)}(v, \omega) = s^{(j)}$. Recall the associated parameter space $\{V, \Omega\}$, where V consists of a set of discrete finite elements and Ω is associated with the parameters $\boldsymbol{\theta}, [p_{ij}]_{i,j}$. Define $\mathbf{s}^*$ and $\boldsymbol{\theta}^*$ associated with the point $\{v^0, \omega^0\}$ in the parameter space, as in the following definitions:

$$
q_T(Y_{1:T}; v^0, \omega^0) \overset{\text{def}}{=} \max_{j \in 1, \ldots, M} p_T(Y_{1:T} | x_1 = j; v^0, \omega^0) \quad \quad (25)
$$

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

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

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

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

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

$$\lim_{T\to\infty} T^{-1} \log p_T(Y_{1:T}; v^0, \omega^0) \;=\; -H(v^0, \omega^0),$$

with probability 1, under (v^0, ω^0), and

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

$$\lim_{T\to\infty} T^{-1} \log p_T(Y_{1:T}; v, \omega) \;=\; H(v^0, \omega^0, v, \omega),$$

with probability 1, under (v^0, ω^0).

Lemma 7.2. $\forall v_i, u_j, i, j \in 1, \ldots, M$ *as weights, the difference between M linear combination of states would lead to*

$$\sum_{i=1}^{M} v_i f(y, \boldsymbol{\theta}_{s^{(i)}}, s^{(i)}) \neq \sum_{j=1}^{M} \mu_j f(y, \boldsymbol{\theta}_{s'^{(j)}}, s'^{(j)}). \tag{26}$$

Proof. For each $s^{(i)}, i \in 1, \ldots, M$ associated with dependency parameter $\boldsymbol{\theta}_{s^{(i)}} \in \mathbb{R}_+^d$.
So

$$\sum_{i=1}^{M} v_i \delta_{s^{(i)}} \neq \sum_{j=1}^{M} \mu_j \delta_{s'^{(j)}}, a.e. \tag{27}$$

implies

$$\sum_{i=1}^{M} v_i \delta_{s^{(i)}} \delta_{\boldsymbol{\theta}_{s^{(i)}}} \neq \sum_{j=1}^{M} \mu_j \delta_{s'^{(j)}} \delta_{\boldsymbol{\theta}_{s'^{(j)}}}, a.e.. \tag{28}$$

$\square$

Also if (27), then the corresponding point in the parameter space (v, ω) would lead to $\mathcal{K}(v^0, \omega^0; v, \omega)$, and (v, ω) would not be in the equivalent class of (v^0, ω^0) as long as the point v and v^0 are different as (27) , (the equivalence class of v^0 is defined in Leroux (1992)), and the divergence between (v, ω) and (v^0, ω^0) is defined as $\mathcal{K}(v^0, \omega^0; v, \omega) \stackrel{\text{def}}{=} H(v^0, \omega^0, v^0, \omega^0) - H(v^0, \omega^0, v, \omega)$. It is connected to the log likelihood ratio process, and one can prove that if either (27) or (28) holds (A.2), (26) will hold, and it will

lead to $\mathcal{K}(v^0, \omega^0; v, \omega) > 0$. Namely, the divergence can distinguish points from different equivalent classes.

Next, we study whether plugging in nonparametric estimated margins would affect the consistency results by analyzing the uniform convergence of $\hat{f}(y, \boldsymbol{\theta}_j, s_j)$.

As $\hat{f}(y, \boldsymbol{\theta}^{(i)}, s^{(i)}) = c\{\hat{F}_1^{\mathfrak{m}}(y_1), \hat{F}_2^{\mathfrak{m}}(y_2), \ldots, \hat{F}_d^{\mathfrak{m}}(y_d), \hat{\boldsymbol{\theta}}^{(i)}, \hat{s}^{(i)}\} \hat{f}_1^{\mathfrak{m}}(y_1) \hat{f}_2^{\mathfrak{m}}(y_2) \cdots \hat{f}_d^{\mathfrak{m}}(y_d)$, we have according to the uniform consistency of copulae density,

$$\sup_{u_1,\ldots,u_d \in [0,1]^d} |\hat{c}(u_1, u_2, \ldots, u_d) - c(u_1, u_2, \ldots, u_d)| = o_p(1) \tag{29}$$

and according to Bickel and Rosenblatt (1973),

$$\sup_{x \in B} |\hat{f}_i^{\mathfrak{m}}(x) - f_i^{\mathfrak{m}}(x)| = o_p(1) \tag{30}$$

Therefore,

$$\sup_{y \in B^d} |\hat{f}(y, \hat{\boldsymbol{\theta}}_j, \hat{s}_j) - f(y, \boldsymbol{\theta}_j, s_j)| = o_p(1),$$

where B, B^d are compact sets. So the plug in estimation would not contaminate the consistency results.

To prove the consistency of our estimation of this parameter, we restate the theorems of consistency in Leroux (1992) for our parameter space. One needs to show that first for the discrete subspace V^c which does not contain any point of the equivalence class of v^0, for $v \in V^c$ and any arbitrary value of $\omega \in \Omega$, it holds, with probability 1,

$$\lim_{T \to \infty} \left[\max_{v \in V^c} \log \sup_{\omega \in \Omega} p_T(Y_{1:T}; v, \omega) - \log p_T(Y_{1:T}; v^0, \omega^0) \right] \to -\infty. \tag{31}$$

The fact follows directly from lemma 7.2 (the identifiability of the states parameters), and its consequence $\mathcal{K}(v^0, \omega^0; v, \omega) > 0$. Theorem 3.1 is proved.

To prove Theorem 3.2, note that $\lim_{T \to \infty} \min_{i \in 1, \ldots, M} \mathrm{P}(|\hat{\boldsymbol{\theta}}^{(i)} - \boldsymbol{\theta}^{*(i)}| > \varepsilon | \hat{s}^{(i)} = s^{*(i)})$ is conditioning on the event $\{\hat{s}^{(i)} = s^{*(i)}\}$ which asymptotically holds with probability 1.

Therefore it is suffice to prove, for any $\hat{s}^{(i)} = s^{(i)}$

$$\lim_{T \to \infty} \min_{i \in 1,\dots,M} P(|\hat{\boldsymbol{\theta}}^{(i)} - \boldsymbol{\theta}^{*(i)}| > \varepsilon) = 0. \tag{32}$$

To show (32), one needs to show that for (V^c, Ω^c) which does not contain any point of the equivalence class of (v^0, ω^0), we have, with probability 1,

$$\lim_{T \to \infty} \{\log \sup_{\omega \in \Omega^c} p_T(Y_{1:T}; v^0, \omega) - \log p_T(Y_{1:T}; v^0, \omega^0)\} \to -\infty, \tag{33}$$

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

$$\lim_{T \to \infty} \{\max_h \log \sup_{\omega \in \mathcal{O}_{\omega_h}} p_T(Y_{1:T}; v^0, \omega) - \log p_T(Y_{1:T}; v^0, \omega^0)\} \to -\infty. \tag{34}$$

To prove (34), we have the modified definition:

$$H(v^0, \omega^0, v^0, \omega; \mathcal{O}_{\omega_h}) \overset{\text{def}}{=} \lim_T \log \sup_{\omega' \in \omega^0} q_T(Y_{1:T}, v^0, \omega')/T. \tag{35}$$

It can be derived that

$$H(v^0, \omega^0, v^0, \omega) < H(v^0, \omega^0, v^0, \omega^0), \tag{36}$$

for (v^0, ω) and (v^0, ω^0) does not lie in the same equivalence class. Then (36) is a consequence of the identifiability condition A.2, and this leads to: $\exists \varepsilon > 0, T_\varepsilon$ and $\mathcal{O}_\omega$ such that

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

Also because $\log \sup_{\omega' \in \mathcal{O}_\omega} p_T(Y_{1:T}, v^0, \omega')/T$ and $\log \sup_{\omega' \in \mathcal{O}_\omega} q_T(Y_{1:T}, v^0, \omega')/T$ have the

same limit value, there exists a constant $\varepsilon > 0$,

$$\lim_{T\to\infty} \log \sup_{\omega' \in \mathcal{O}_{\omega_h}} p_T(y_{1:T}, v^0, \omega')/T = H(v^0, \omega^0, v^0, \omega; \mathcal{O}_{\omega_h}) \le H(v^0, \omega^0, v^0, \omega^0) - \varepsilon.$$
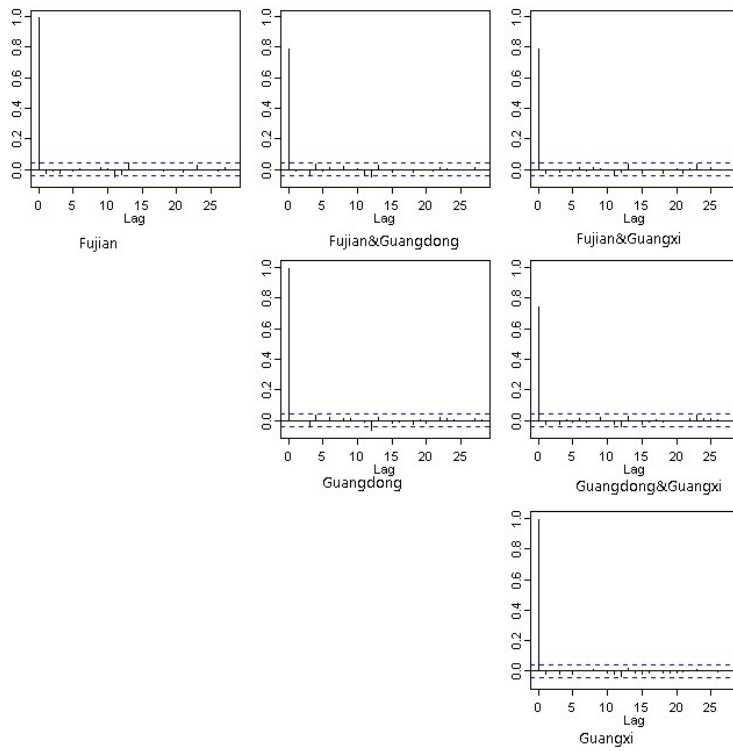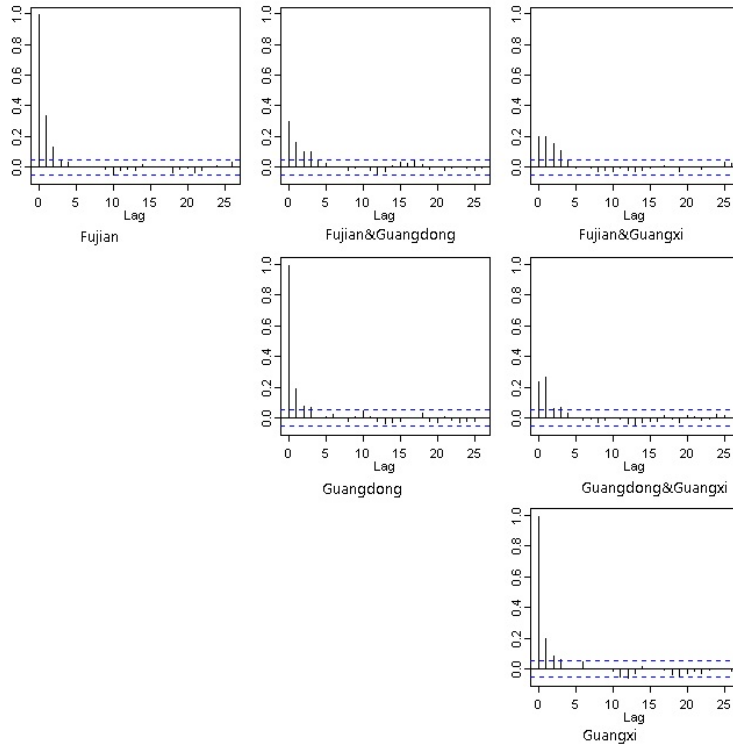
Now (34) follows.

References

Adrian, T. and Brunnermeier, M. K. (2011). CoVaR, *Staff Reports 348*, Federal Reserve Bank of New York.

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.

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates, *The Annals of Statistics* **1**: 1071–1095.

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.

Engle, R. (2002). Dynamic conditional correlation, *Journal of Business and Economic Statistics* **20**(3): 339–350.

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

Gao, X. and Song, P. X.-K. (2011). Composite likelihood em algorithm with applications to multivariate hidden markov model, *Statistica Sinica* **21**: 165–185.

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. (2012). Dynamic structured copula models, *Submitted for publication 13.02.2012* .

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. (2009). On the structure and estimation of hierarchical archimedean copulas, *Under Revision of Journal of Econometrics* .

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. (2010). Hierarchical Archimedean copulas, *Quantitative Finance* **10**: 295–304.

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 14: Autocorrelations and cross-correlations of the simulated rainfall and original time series
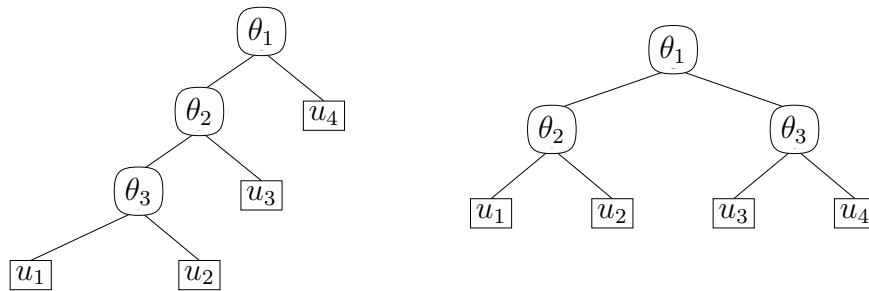
Figure 15: Fully and partially nested copulae of dimension $d = 4$ with structures $s = (((12)3)4)$ on the left and $s = ((12)(34))$ on the right