Bayesian Inference for a Semi-Parametric Copula-based Markov Chain

Mike Pitt & Kazim Azam

No 1051

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

THE UNIVERSITY OF WARWICK

Bayesian Inference for a Semi-Parametric Copula-based Markov Chain*

Kazim Azam [†]

Vrije Universiteit, Amsterdam

Michael Pitt [‡]

University of Warwick

July 9, 2014

Abstract

This paper presents a method to specify a strictly stationary univariate time series model with particular emphasis on the marginal characteristics (fat tailedness, skewness etc.). It is the first time in time series models with specified marginal distribution, a non-parametric specification is used. Through a Copula distribution, the marginal aspect are separated and the information contained within the order statistics allow to efficiently model a discretely-varied time series. The estimation is done through Bayesian method. The method is invariant to any copula family and for any level of heterogeneity in the random variable. Using count times series of weekly firearm homicides in Cape Town, South Africa, we show our method efficiently estimates the copula parameter representing the first-order Markov chain transition density.

JEL Classification: C11, C14, C20.
Keywords: Bayesian copula, discrete data, order statistics, semi-parametric, time series.

---

*We would like to thank Gianna Boero, Robert Kohn and Alessandro Palandri for the invaluable advices. Valuable comments from Charles Bos, Valentina Corradi, Peter Hoff, Lennart Hoogerheide, Siem Jan Koopman and Andre Lucas are much appreciated. All the participants from "Copulae In Mathematical And Quantitative Finance 2012" are also acknowledged. The comments from the participants at the Warwick Lunchtime Workshop are acknowledged. Responsibility for the remaining errors is solely ours.

[†]Address for correspondence: Department of Finance, VU University Amsterdam, De Boelelaan 1105, NL-1081HV Amsterdam, The Netherlands. E-mail: Kazim.Azam@vu.nl

[‡]E-mail: M.Pitt@warwick.ac.uk

1 Introduction

Limitations and rigidness of time series models is well documented. Their construction has always depended upon the exact nature of the data, and do not easily accommodate other types of data (non-normal random variables). By separating the characteristics specific to the data from the time-varying properties, we are able to specify a general method through copula to model any strictly stationary time series.

In Finance, copula models have recently gained popularity, as they provide an alternative to the assumption of normality in data (see Embrechts et al. (1999)). Copula applications range from Value at Risk (VaR) analysis of Cherubini and Luciano (2001) and Embrechts et al. (2003), to studies of financial contagion of Rodriguez (2007). Common to all these works is the emphasis on how copula models provide a solutions when the assumptions of normality and linear dependency fail. Bouyé et al. (2007) present a detailed coverage of copula methodology and other applications in finance. In Economics, instead, the copula based literature remains limited to few studies. In Economics and Econometrics, analysis involving discrete data is unavoidable and this represents a complication, as alike other joint analysis techniques, copula models do not cope well with the limitations posed by discrete data. Nevertheless, the advantage gained through non-normality and non-elliptical joint distribution, has lead researchers to address problems otherwise not possible. Smith (2003) uses copula framework to study self-selection problem. Demarta and McNeil (2005) among others, analyze categorical data from clinical trials. Zimmer and Trivedi (2006) employ a trivariate copula for dependency between health insurance status for couples and their demand for health care. Trivedi and Zimmer (2006) cover various aspects of copula estimation for discrete type data. Hoff (2007) proposes a Bayesian technique which addresses the discreteness of data in a semi-parametric copula setting, and employs it on cross-sectional data (U.S. labour force) of continuous and discrete type. Patton (2006) introduces copulas in modelling of economic time series (continuous data).

Most of the copula literature in general deals with modelling contemporaneous dependence among random variables, there has been quite a lot of work in specifying a Markov chain through a copula. Most of such work has been theoretical dealing with probability and weak dependence properties. Darswo and Olsen (1992) specify the necessary and sufficient conditions for constructing stationary first-order Markov models based on a copula. They show that the Chapman-Kolmogorov equations are satisfied for such models. Ibragimov (2009) extend the conditions presented by Darswo and Olsen (1992) for higher-order Markov models. Chen and Fan (2002), Chen et al. (2009) and Beare (2010a,b) present the persistence properties of stationary copula-based Markov chains. Chen and Fan (2002) state conditions under which copula-based Markov models are β-mixing with either exponential or polynomial decay rates, and the conditions are independent of the marginal distribution specification, but only dependent upon the copula specification, they also show EFGM and Gaussian copula are indeed geometric β-mixing. Beare (2010a) provides strong sufficient conditions for geometric β-mixing, which rules out copula families exhibiting asymmetric and tail dependence, for whom β-mixing with exponential decay rates is established. Lentzas and Ibragimov (2008) show a Clayton copula-based model behaves like a long memory time series with high persistence, but Chen et al. (2009) show in terms of the mixing properties, such a model is weakly dependent and short memory and models generated through Clayton, Gumbel and t-copula are indeed geometric β-mixing. Beare (2010b) shows for Archimedean copulas that the regular variation of the generator at zero and one implies geometric ergodicity. Joe (1997) shows in a fully-parametric copula setting that various Maximum likelihood (ML) based estimators to be consistent and asymptotically normal under some regularity conditions. Chen and Fan (2002) propose a semi-parametric copula estimation (empirically computed margins) using ML, and also prove consistency and asymptotic normality. Chen et al. (2009) state an efficient sieve ML estimation procedure for copula-based Markov chains.

We specify a univariate time series through a copula-based Markov chain, which is not only specific to discretely varying data (i.e. count data), but also to time series of continuous random variable type. The novelty of our paper is two fold, first, this is the first time in literature for time series models with specified marginal distribution (parametric), a non-parametric based specification is being introduced, which like the other models does acknowledge and identify marginal behaviour but is non-parametric. The transition density is given through a copula function. Secondly, this paper also introduces Bayesian methodology for copula-based Markov chain estimation. The advantages of such a technique is we are able to separate out using copula, the marginal behaviour (like fat tailedness and skewness) of a time series from the dependence structure (like asymmetric persistence in extreme values or other non-linear dependence structure) of any type of random variable. One such time series could be the number of patents a firm acquires over time, the outcomes could be of poisson type and having larger persistence over time for high number of patents obtained as compared to low number of patents. This is generally observed for growing firms.

The modelling is similar to an AR type process, where we specify the current value to be some function of its own lags. But unlike previous models, where the assumption on the marginal distribution dictates the conditional distribution (normality etc.), we model these distributions separately and they are not bounded to each other. Our method is general across various data types, as we make no assumption regarding the marginal distribution and treat it completely as unknown. The estimation of the copula parameters (conditional dependency) is based only on the order statistics of the observed time series which is similar to Hoff (2007), but he deals with cross-sectional data, and specifies a sampling scheme suitable only to a Gaussian copula. To keep the intuition clear, we model a first-order Markov chain, which can be adopted for high order processes. The marginal distribution is completely left unspecified, and we treat the uniform variables (generally obtained through the marginal distribution for copula modelling) as latent variables which along with the copula parameters are estimated in a Bayesian framework.

We use a real data application, which is based on the count of weekly firearm homicides observed in Cape Town, South Africa. Crime in general is quite persistent, and our period of analysis consists of a time when there was urbanisation in and around Cape Town. Hence we could model such persistent through copulas. We successfully capture temporal dependence first through a Gumbel copula, and then to show our method is invariant to different copula families (also applying a Gaussian copula separately). In terms of the Bayesian methods applied, standard diagnostics are used to confirm the Markov Chain Monte Carlo (MCMC) performs well.

The paper sets out with a brief discussion on copula models, most importantly on copula-based Markov chain specification. In Section 3, we set out the modelling framework based on the order statistics. The two-stage Bayesian sampling scheme for the latent copula arguments and the copula parameters in Section 3. We then in Section 4 apply our technique on to a real data application and concluding in Section 5.

2 Copula-based Markov chain

2.1 Copula Definition

A copula is multivariate distribution able to separate out the marginal distribution of the random variables, and the dependence among them. This enables modelling dependency more rigorously and entails features not available within other commonly employed multivariate distributions (like a Multivariate Normal distribution). Let us consider a bivariate case for simplicity, then the theorem of Sklar (1959) states that a bivariate distribution H can be decomposed into a copula C capturing the dependence and two marginal distribution F and G of random variables X and Y respectively

$$H(x, y) = C(F(x), G(y)), \tag{2.1}$$

which could be reformulated as

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)).$$

$C : [0, 1] \times [0, 1] \mapsto [0, 1]$ is the joint Cumulative Distribution Function (CDF) with the uniform margins as input arguments. The marginals F and G transform the observed x and y to uniformly distributed variables, u and v respectively. C is a uniquely defined copula if H is continuous, otherwise it is defined over the range of both F and G. The attraction of copula models is, that it allows to disentangle characteristics specific to the marginals from the dependence strcture, allowing flexibility and their independent selection process. There exists a wide choice of copulas, differing in the type of dependence they capture (i.e. asymmetric and tail dependence). Joe (1997) and Nelsen (2007) provide a detailed survey of various copula families and their properties.

2.2 Markov chain through Copula

We can specify a strictly stationary process through a copula, assuming $\{Y_t\}_{t=1}^{T}$ is Markov chain of order one, and let F be its true invariant continuous distribution along with H being the joint distribution of Y_{t-1} and Y_t. Then there exists a unique copula capturing the transition distribution, given as,

$$H(y_t, y_{t-1}) = C(F(y_t), F(y_{t-1})). \tag{2.2}$$

The above equation would be true for al the $T-1$ pairs. Given that F is continuous, then we can take the partial derivatives w.r.t both y_t and y_{t-1} of (2.2) which yields

$$\frac{\partial H^2(y_t, y_{t-1})}{\partial y_t \partial y_{t-1}} = h(y_t, y_{t-1}) = c(F(y_t), F(y_{t-1})).f(y_t).f(y_{t-1}), \tag{2.3}$$

where h is the joint density and c denotes the copula density. From (2.3) we can now state the conditional density of y_t given y_{t-1} as

$$h_{t|t-1}(y_t|y_{t-1}) = c(F(y_{t-1}), F(y_t)).f(y_t), \tag{2.4}$$

where $h_{t|t-1}$ denotes the conditional density. If F was assumed to be a normal distribution and C a Gaussian copula, then (2.4) would correspond to normal conditional density, hence an AR(1) process with normally distributed errors.

Suppose now F is discrete distribution, hence Y_t takes non-negative integers. Then (2.2) would be given as,

$$\begin{aligned} H(y_t, y_{t-1}) &= C(F(y_t), F(y_{t-1})) - C(F(y_t - 1), F(y_{t-1})) \\ &- C(F(y_t), F(y_{t-1} - 1)) + C(F(y_t - 1), F(y_{t-1} - 1)). \end{aligned} \tag{2.5}$$

Then we can also derive the conditional density equation.

There are quite a few problems associated with employing discrete copula. First, the copula in the joint distribution H is not identified, which implies there could be several functions A in (2.2), such that,

$$H(y_t, y_{t-1}) = A(F(y_t), F(y_{t-1})). \tag{2.6}$$

We can solve the above equation for A, but it is not necessarily a copula function or even a distribution function. This is referred to as the identifiability issue, as we require some set of copulas for which we can replace A by, such copulas of course require their own lower and upper bounds. We refer interested reader to Genest and Nešlehová (2007) for more details.

Within continuous copula setup, the copula dependence parameter is usually converted to a meaningful measure such as Kendall's tau or Spearman's rho. Both of these measures are defined on $[0, 1]$ and are independent upon the marginal specifications. However, in case of discrete data both measures are not invariant to the choice of the marginal distributions. Also rank-based estimators are not useful due to the ties observed in the ranks of discrete data, and dealing with them by either splitting or ignoring creates a bias in the estimate of the copula parameters.

Trivedi and Zimmer (2006) mention maximization of likelihood with discrete margins poses computational difficulties and proposes to perform continuation transformation, where each discrete margin is made continuous by adding some noise (Uniform $[0, 1]$ draw), then proceed with copula estimation, with continuous margins, but such a process of course creates a bias through misspecification.

Hoff (2007) calls marginal parameters as nuisance parameters, especially for discrete data, and derives a likelihood which treats the copula arguments as latent variables and relies on the fact that they have the same order statistics, as the observed data. Other similar Bayesian methods are specified in Pitt et al. (2006) and Smith and Khaled (2012) to deal with discrete margins. Generally the interpretation of θ (copula parameter) does not have the same meaning in case of discrete margins as it has for continuous margins.

We now provide a method similar to Hoff (2007) for specifying a copula-based Markov chain for all types of data, which is invariant to the choice of the copula family.

3 Framework

Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_T)$ be a strictly stationary time series originating from an unknown marginal distribution F, and $\mathbf{U} = (U_1, U_2, \ldots, U_T)$ be the series of uniforms, with each instance being in $[0, 1]$. Let the Markov chain be generated through copula C, then we can specify the

Data Generating Process (DGP) for a first-order Markov chain as

$$c(u_t|u_{t-1};\Theta),\ \text{for } t = 2,\dots,T,$$
$$y_t = j,\ \text{if } \max\Big\{u_s;\ F : j-1 \mapsto u_s\Big\} < u_t < \min\Big\{u_s;\ F : j+1 \mapsto u_s\Big\},$$
$$\text{where } j \in J \text{ (discrete outcomes).}$$

Where Θ is the parameter vector associated to C, and j is a discrete observation belonging to set of possible values in J. Each $u_t \in [0,1]$ is generated through the conditional copula density c, and the corresponding y_t determined through the maximum and minimum of the uniforms corresponding to the neighbouring order statistics of j, as seen in Figure 1. The DGP described above is set out for a discretely-varied time series, in case we are dealing with a time series of continuous type random variables, the correspondence of u_t to y_t is one-to-one.

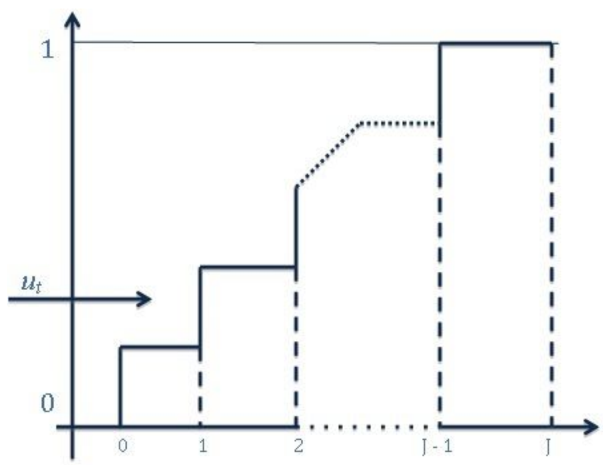


Figure 1: Mapping of generated u_t

If F is known, and belongs to either a parametric family or non-parametric (empirical distribution), then one of the estimators stated in Section 2.4 could be employed for Θ. For a continuous margin such estimators would yield constant and asymptotically normal estimates of Θ, but in case F is a discrete distribution, ML methods can fail with convergence of the likelihood and results will be biased from using continuous transformation.

We treat F as completely unknown, and hence $\mathbf{U}$ is unobtainable and considered as a series of latent variables. The only available information available related to F is that it is a non-decreasing monotonic function, and could either be a continuous or a discrete distribution. In case the margin is continuous, F^{-1} will be a one-to-one mapping function, and for a discrete margin a many-to-one function. It is the first time a completely non-parametric specification has been assumed for the marginal distribution using for Markov chain type time series framework, both for continuous and discrete type outcomes. This overcomes any form of marginal misspecification, and allows to combine any type of marginal behaviour with non-gaussian type of temporal dependence.

Given that F is non-decreasing, we know the order statistics of the uniforms generated through the unknown F will be dictated by the order statistics of the observed $\mathbf{Y}$, and this is the only information known with certainty. But there is still the uncertainty of the actual value of $\mathbf{U}$, and the degree of uncertainty depends upon the discrete data (low count implying more uncertainty). In case we have a time series of binary outcomes, there is really only two ranks, and hence we have more uncertainty. We can provide a formal definition for the order statistics of the time series.

Definition 3.1 *Let the rank of the observation at time t, y_t, be denoted as k_t. Hence $y_t = y^{(k_t)}$, and for each t*

$$y^{(k_t-1)} < y_t < y^{(k_t+1)}, \text{ and,}$$
$$u^{(k_t-1)} < u_t < u^{(k_t+1)}.$$

$y^{(k_t)}$ is the order statistic of y_t. u_t has the same rank k_t, as y_t.

Definition 3.1 simply states that given that F is non-decreasing and monotonic, the unobserved $\mathbf{U}$ have to obey the same order statistics as that of $\mathbf{Y}$. We keep the time indexing on the ranks, as unlike cross-sectional analysis, the time stamp on an observation is important. Note, we have strict-inequality for the ranks of the observed data, implying any ties are left unresolved.

As we are capturing the temporal dependence of the series $\mathbf{U}$, each instance u_t is related to its neighbor in time and the corresponding order statistic through $\mathbf{Y}$. This is perhaps best understood by employing a Directed Acyclic Graph (DAG) in Figure 2, where we see that $\mathbf{U}$ is a Markov chain, and the observation y_t are independent of each other conditional upon u_t. Starting from u_1 the chain moves forward till the last value u_T. We see how each u_t is connected to its neighbors in time (u_{t-1} and u_{t+1}) and the corresponding y_t, from which the

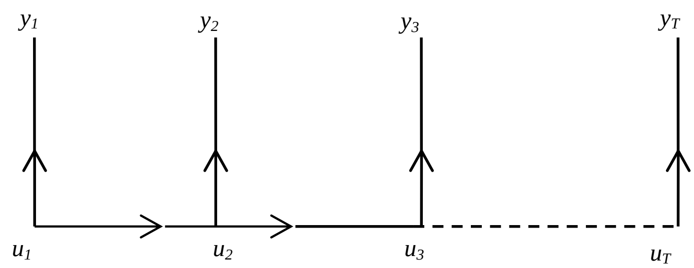


Figure 2: DAG of Latent Variable

only information retrieved is k_t, the rank.

Using the above framework of the DAG, we can now model the structure of the Markov chain in a Bayesian setup.

4 Bayesian Sampling Scheme

We specify a general sampling scheme to estimate the copula parameters, which will capture the temporal dependence within a stationary Markov chain. The estimation technique is general to any copula family and makes no assumption regarding the marginal distribution, and hence can accommodate both continuous and discrete type margins. The sampling scheme can be separated into two different stages, the first stage involves sampling $\mathbf{U}$ conditional upon the copula parameter Θ, and in the second stage we draw Θ conditional upon $\mathbf{U}$. This implies all the uniforms are considered as auxiliary variables and have to be sampled.

Before proceeding with the sampling scheme, let us define the necessary notation. Through Bayes theorem, let the posterior of Θ be given as

$$p(\Theta|\mathbf{U}) \propto p(\Theta) \times p(\mathbf{U}|\mathbf{Y};\Theta), \tag{4.1}$$

where π denotes the posterior of Θ. To make our scheme general across various copula families, we need to re-parametrize the copula parameter vector Θ. Copula families support different ranges, for a Gaussian and Student-t copula the dependence parameter lies in $[-1, 1]$. Whereas for most Archimedean copulas the upper or lower bounds are defined up to infinity, like the range of Clayton copula parameter is $(0, \infty)$. We can transform them all to be defined over the real line $\mathbb{R}$. Let the mapping be $\mathbf{Z}(\Theta) = \Psi$, where $\Psi \in \mathbb{R}$ and $\mathbf{Z}$ (See Appendix A for various transformations) represents a vector of functions, which has the same dimension as Θ. This re-parametrization will change the posterior defined in (4.1). The prior distribution $p(\theta)$ has to be transformed over to prior of Ψ, and we have to consider the Jacobian matrix associated to such a reformulation. The prior $p(\Psi)$ for Ψ will then be defined as

$$p(\Psi) \propto p(\Theta)\left|\frac{\partial \Theta}{\partial \Psi}\right|,$$

where $\left|\frac{\partial \Theta}{\partial \Psi}\right|$ is the determinant of the Jacobian matrix. Regardless of which copula is now chosen, we have a support over the real line for the parameters associated. The posterior defined in (4.1) now becomes

$$p(\Psi|\mathbf{U}) \propto p(\Psi) \times p(\mathbf{U}|\mathbf{Y};\Psi).$$

Finally, we can now proceed with specifying the two stage sampling scheme, where first we sample from $p(\mathbf{U}|\mathbf{Y};\Psi)$, followed by sampling from $p(\Psi|\mathbf{U})$.

4.1 Sampling from $p(\mathbf{U}|\mathbf{Y};\Psi)$

We see from Figure 2 how each instance of $\mathbf{U}$ is linked by its neighbours in time and the corresponding order statistic from $\mathbf{Y}$. Assuming a first-order Markov chain, we can write the conditional probability for each u_t, where $1 < t < T$ as

$$p(u_t|\mathbf{U}_{\backslash \mathbf{t}}, \mathbf{Y};\Psi) = p(u_t|u_{t-1}, u_{t+1}, y_t;\Psi), \tag{4.2}$$

where $\mathbf{U}_{\backslash \mathbf{t}}$ is the complete series $\mathbf{U}$ without u_t. Given that we have a Markov chain of order one and using the information from the DAG, the conditioning of $\mathbf{U}_{\backslash \mathbf{t}}$ can be reduced to u_t's connected neighbours in time (i.e. u_{t-1} and u_{t+1}). As we mentioned previously, the only information available from conditioning u_t on y_t is, if $y^{(k_t)}$ is the order statistic of y_t then k_t is also the rank of u_t. This implies u_t has to lie between $u^{(k_t-1)} < u_t < u^{(k_t+1)}$ to maintain the order statistics, and the size of the interval depends upon the degree on discreteness. So we can simplify (4.2) as

$$p(u_t|u_{t-1}, u_{t+1}, u^{(k_t-1)}, u^{(k_t+1)};\Psi) = p(u_t|u_{t-1}, u_{t+1};\Psi)\mathcal{I}(u^{(k_t-1)} < u_t < u^{(k_t+1)}). \tag{4.3}$$

We cannot directly sample from (4.3) through a bivariate copula density (a first-order Markov chain corresponding to bivariate copula), but applying Bayes theorem further, we can write $p(u_t|u_{t-1}, u_{t+1}; \Psi)$ as

$$p(u_t|u_{t-1}, u_{t+1}; \Psi) \propto p(u_t|u_{t-1}; \Psi) \times p(u_{t+1}|u_t; \Psi).$$

Now we have two conditional distributions and we introduce the conditional copula, let C be the copula distribution, and let $c_{t|t-1}$ be the bivariate conditional copula density of u_t given u_{t-1}, corresponding to C (see Appendix A.1 for various copulas density formulation). We model the conditional probability of u_t through the conditional copula density as,

$$p(u_t|u_{t-1}; \Psi) = c_{t|t-1}(u_t|u_{t-1}; \Psi).$$

Hence the copula represents the transition density of the Markov chain. The same holds for conditioning of u_{t+1} to u_t. Now we can easily sample from the conditional distribution $C_{t|t-1}(u_t|u_{t-1}; \Psi)$, and evaluate the draw through Metropolis-Hasting (M-H) algorithm using $c_{t+1|t}(u_{t+1}|u_t; \Psi)$. The sampling scheme is given as

for each $u_t, (t = 1, \ldots, T)$,

compute $u^{(k_t-1)}$ and $u^{(k_t+1)}$, given $(y^{(k_t-1)} < y_t < y^{(k_t+1)})$,

sample $u_t^*$from$C_{t|t-1}(u_t|u_{t-1}; \Psi)\mathcal{I}(u^{(k_t-1)} < u_t < u^{(k_t+1)})$,

compute $\alpha_u = \min \left\{1, \dfrac{c_{t+1|t}(u_{t+1}|u_t^*; \Psi)}{c_{t+1|t}(u_{t+1}|u_t; \Psi)}\right\}$.

Sampling from $C_{t|t-1}$ is easier than $c_{t|t-1}$, as most copulas have a closed inverse form for the conditional distribution. The above scheme is repeated for all u_t. The truncated intervals are updated if the drawn u_t^* is accepted. The sampling is performed in the order of the $\mathbf{U}$ dictated by the time order, but the intervals have to be maintained regardless of time. Through this scheme we obtain an updated sample of $\mathbf{U}$. For large enough T (for continuous data) and high count data (discrete data), the interval $(u^{(k_t-1)}, u^{(k_t+1)})$ becomes smaller and the acceptance probability α_u gets close to one. In fact anything uniformly sampled through the interval can be accepted, and we would not need to pass them through the M-H step to evaluate the conditional copula of u_{t+1}. Missing values can also be generated through this scheme. We could also consider higher-order Markov chain, for example for a second-order Markov chain, we will require a trivariate copula and proceed with similar sampling procedure.

4.2 Sampling from the Posterior $p(\Psi|\mathbf{U})$

Now we can proceed with sampling from the posterior of Ψ. Unlike the Gaussian copula (see Hoff (2007)), most copula families do not have the full conditional available to sample from, and a Markov Chain Monte Carlo (MCMC) based on M-H algorithm has to be adopted.

Within the M-H framework, we have to choose an adequate proposal distribution $g(\Psi)$. A multivariate t-distribution with the mean equal to the mode of the posterior, and the variance equal to the negative inverse of the information matrix computed at the mode, can be used as a proposal distribution. Such a Laplace-type proposal has been used in the literature (see Chib and Greenberg (1998), Chib and Winkelmann (2001) and Pitt et al. (2006)). A multivariate t-distribution with ν degrees-of-freedom as opposed to a Normal distribution is preferred, as it would dominate in the tails of the true density. The advantage of such a proposal density is we do not need to consider tuning of parameters, to attain some acceptance probability.

We choose a flat prior for Θ, $p(\Theta) = 1$. Hence the re-parametrized Ψ's prior will be $p(\Psi) \propto \left|\frac{\partial\Theta}{\partial\Psi}\right|$. To use a Laplace approximation, we need to employ a Maximum a Posterior Probability (MAP).

$$\widehat{\Psi}_{MAP}(\mathbf{U}) = \arg\max_{\Psi} p(\mathbf{U}|\mathbf{Y};\Psi)p(\Psi),$$

where $\widehat{\Psi}$ denotes the estimated mode of Ψ. The log of the posterior can then be written as

$$\log p(\Psi|\mathbf{U}) \approx \log g(\Psi|\widehat{\Psi};V).$$

$V = -I^{-1}$ and $I = \left[\frac{\partial^2 p(\mathbf{U}|\mathbf{Y};\Psi)p(\Psi)}{\partial\Psi\partial\Psi'}\right]_{\Psi=\widehat{\Psi}}$ is the information matrix evaluated at the mode. We can finally draw from

$$\Psi \simeq t_\nu(\widehat{\Psi}, V),$$

let the drawn value be denoted as Ψ^* and the current value be Ψ, then Ψ^* can be evaluated using M-H with the following acceptance probability

$$\alpha_\Psi = \min\left\{1, \frac{\pi(\Psi^*)t_\nu(\Psi|\widehat{\Psi}, V)}{\pi(\Psi)t_\nu(\Psi^*|\widehat{\Psi}, V)}\right\},$$

where $t_\nu(\Psi|\widehat{\Psi}, V)$ denotes the density of $t_\nu(\widehat{\Psi}, V)$ evaluated at Ψ. The acceptance ratio is very high, as the proposal density is close to the true density. We arbitrarily choose $\nu = 6$ to dominate in the tails. A new mode $\widehat{\Psi}$ is found at each iteration, conditional on the new updated $\mathbf{U}$ sample. To avoid computational burden of finding a new mode at each iteration, we could update the mode every hundredth time or so, but we cannot not change it as the $\mathbf{U}$ get sampled at each iteration. Numerical techniques such as Newton-Raphson are required to locate the mode, and they are found within few steps of the algorithm search. Re-estimating the mode at each iteration could become computationally intensive, and could be avoided by updating it at regular intervals.

5 Data Example (Firearm Homicides)

We now present a real data application, where there could be persistence within a discretely-varied time series. To be specific, we are looking at the weekly firearm homicides in Cape Town, South Africa from the period of 1st January 1986 up to 31st February 1991 (275 observations)[1], as given in Figure 3.

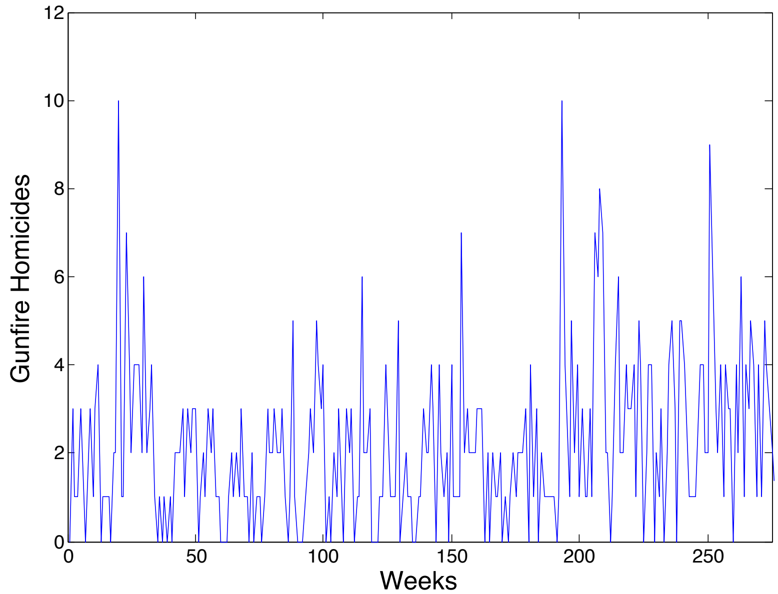
Generally the period of the data sample corresponds to a time when areas in and around Cape Town experienced rapid urbanization. Through such development local gangs form in these areas, and clustering of high count of homicides could be associated with it. Cape Town also attracted people from different regions of South Africa to settle in, and differences in social norms could be another contributing reason.

We could employ various copula models to capture some specific form of temporal dependence present in such a series. Two copula families will be employed here, first a Gumbel copula which is a one-parameter copula exhibiting greater dependence in the right tail. Secondly, to emphasis that our method is general across copula families a standard Gaussian copula with no tail dependence parameter will be used, for the transition density.

The MCMC scheme is performed for 50,000 iterations. To deal with posterior correlation,

[1]Data obtained from MacDonald and Zucchini (1997).

Figure 3: Firearm Homicides South Africa



we perform thinning by saving every 30th iterate for posterior analysis, also the first 5000 iterations are discarded for burn-in. Figure 4 presents the trace plot for both the Gumbel and the Gaussian copula, where we see the chain mixes well. We also present the autocorrelations plot of the final posterior sample for both of the used copulas in figure 5. The autocorrelation is lower than 0.02 after the 4th and 3rd lag for the Gumbel and the Gaussian copula parameter respectively. Further to ensure convergence to the stationary distribution, we performed multiple runs from different initial values. The reason for finding such high correlation, is due to the sampling of each u_t at a time, and then sampling Ψ based on the whole sample of $\mathbf{U}$. In a time series framework it is difficult to overcome such posterior correlation.

Apart from obtaining the posterior mean of Θ, we also compute the Kendall's Tau rank correlation coefficient for the copulas, which provides a meaningful interpretation of the dependence as it is defined over $[-1, 1]$. Computed as

$$\tau = 4 \iint_{[0,1]^2} C(u, v) dC(u, v) - 1. \tag{5.1}$$

For most copula families, a simple analytical solution for (5.1) exists, else it has to be numerically computed. From using both of the copulas, we see there is evidently some temporal dependence present within this time series, suggesting persistence. The Kendall's tau measure

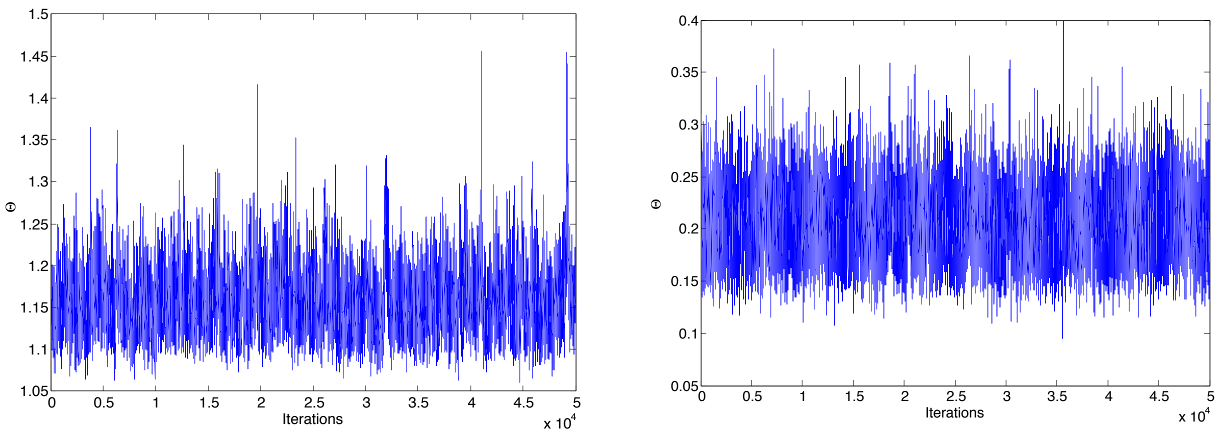


Figure 4: Trace Plots: Gumbel & Gaussian Copula

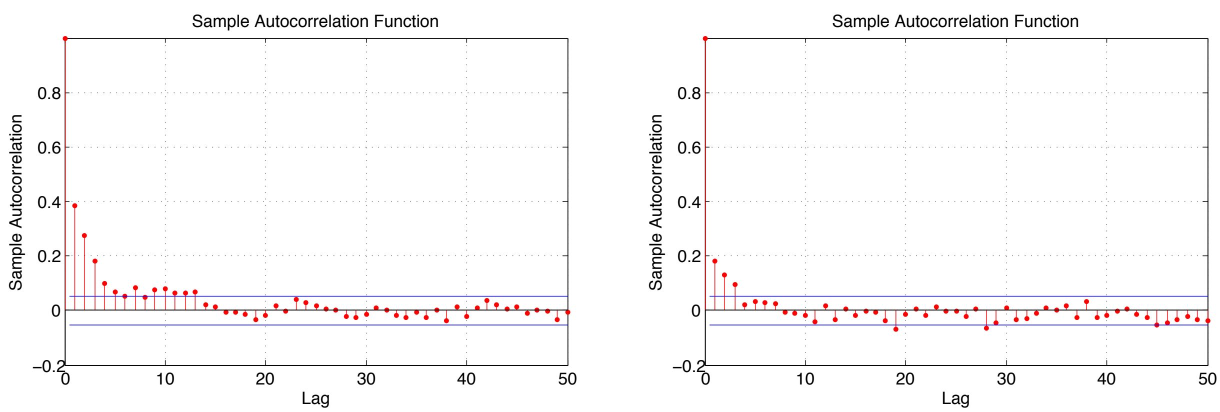


Figure 5: Autocorrelation Plots: Gumbel & Gaussian Copula

Table 1: Posterior Distribution Inference (Firearm Homicides)

Copula C	$E(\pi(\Theta))$ (Posterior mean)	τ (Kendall's tau)	λ_U (Upper tail dep.)
Gumbel	1.278	0.138	0.182
	(0.069)	(0.026)	(0.033)
Gaussian	0.209	0.134	-
	(0.031)	(0.020)	

through both copulas is the same. For the Gaussian copula the tail dependence (lower and upper) parameter is 0, but for the Gumbel copula we can compute the upper tail dependence parameter through,

$$\lambda_U = \lim_{1^-} \frac{1 - 2u + C(u,u)}{u},$$

where $\lim_{1^-}$ is the limit approaching from the left. From Table 1 we can see there is significant upper tail dependence, which implies, through Gumbel copula there seems to be higher probability of observing high homicides counts this week if in the previous week there were high number of homicides.

Apart from the persistence observed within the time series, there could be other explanatory variables to determine the current period firearm homicides, like number of police deployment in local boroughs of Cape town over time, policy implementation, number of people moving to Cape Town from rural areas around and others. Such covariates can be considered either within copula framework where we could choose separate copula families to present the contemporaneous dependence or if our interest is not on these covariates but solely on the temporal dependence, we can filter out the unexplained variation in homicides through a regression where the covariates are considered.

6 Conclusion

Time series models fail to be flexible enough to capture complex temporal dependence for data of different types (binary, count, ordered etc.). The entanglement of the marginal distribution with the conditional (transition) distribution, confine most models to specific problems. Models constructed for data of continuous type cannot normally be applied to discrete data type. Copula models are generally used to capture contemporaneous dependence, but recently they have been considered to describe a time series process (see **?**, Chen and Fan (2002) and Chen et al. (2009) among others). Lentzas and Ibragimov (2008) show a copula-based Markov chain can act like long memory process. **?** proposes a parametric copula-based Markov chain for continuous data, where both the marginal distribution and the copula are parametrically specified. Chen and Fan (2002) present a semi-parametric copula-based Markov chain, where the marginal distribution is computed through an empirical distribution and employing Maximum Likelihood for the copula parameters, yields consistent and asymptotically normal results.

However, for discrete data types such methods can create computational problems (algorithm failing to converge) and induce bias.

Hoff (2007) proposes a technique for cross-sectional data of mixed type (continuous-discrete), where the marginals are left completely unspecified and based only on the information contained in the order statistics a multivariate Gaussian copula is estimated. We extended Hoff's technique in a time series framework and make it general across various copula families. Such a technique can accommodate time series of both discretely and continuously varying random variables. A Bayesian sampling scheme is proposed, where we first sample the latent uniform variables conditional upon the copula parameters, and then the copula parameter conditional upon the sampled uniform series. To make the methodology general across copula families, a Laplace-type proposal for the posterior is presented, where each draw is evaluated through a Metropolis-Hasting algorithm. The novelty of our work is in the class of time series models with specified margins, it is the first time a non-parametric specification has been proposed for the marginal density, avoiding misspecification for discrete data which is quite easy and not having to estimate nuisance parameters associated to the marginal distribution.

We use a real data application based on the weekly count of firearm homicides in Cape Town, South Africa. Employing both a Gumbel and a Gaussian copula separately, we capture the significant persistence present within such a time series. Various quantities of interest, like tail dependence can also be computed through copulas. The MCMC technique works well. Appropriate thinning and discarding of the initial posterior draws is performed to ensure no autocorrelation in the final posterior sample.

A Copula Families and Conditional Distribution

A.1 Copula Transformations

We can obtain analytical formulas for important copula properties through the general distribution function C. For the copula density c given the distribution function is twice differentiable

$$c(u,v|\Theta) = \frac{\partial^2 C(u,v|\Theta)}{\partial u \partial v}.$$

And the conditional distribution of a copula given as

$$C_{u|v}(u|v;\Theta) = \frac{\partial C(u,v|\Theta)}{\partial v}.$$

The conditional copula density $c_{u|v}(u|v;\Theta) = c(u,v|\Theta)$.

A.2 Clayton Copula

The Clayton copula is an Archimedean copula exhibiting strong joint left tail dependency. It is a one-parameter based copula. The various transformations for Clayton copula are given as

$C(u,v|\alpha) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}),$

$c(u,v|\alpha) = (1+\alpha)[uv]^{-\alpha-1}(u^{-\alpha} + v^{-\alpha} - 1)^{-2-1/\alpha},$

$c_{u|v}(u|v;\alpha) = v^{-\alpha-1}(u^{-\alpha} + v^{-\alpha} - 1)^{-1-1/\alpha},$

$C^{-1}_{u|v}(u|v;\theta) = \left[(uv^{\theta+1})^{-\theta/(1+\theta)} + 1 - v^{-\theta} \right]^{-1/\theta},$

where $u, v \in [0,1]$, and the copula parameter $\alpha \in [0, \infty]$.

A.3 Gumbel Copula

Gumbel copula is an Archimedean copula exhibiting strong joint right tail dependency. It is a one-parameter based copula. The various transformations for Gumbel copula are given as

$C(u,v|\alpha) = \exp\left[-(\widetilde{u}^{\alpha} + \widetilde{v}^{\alpha})^{1/\alpha} \right],$

$$c(u,v|\alpha) = C(u,v|\alpha)(uv)^{-1}\frac{(\widetilde{u}.\widetilde{v})^{\alpha-1}}{(\widetilde{u}^{\alpha}+\widetilde{v}^{\alpha})^{2-(1/\alpha)}}\Bigg[(\widetilde{u}^{\alpha}+\widetilde{v}^{\alpha})^{1/\alpha}+\alpha-1\Bigg],$$

$$C_{u|v}(u|v;\alpha) = v^{-1}\exp\Bigg[-(\widetilde{u}^{\alpha}+\widetilde{v}^{\alpha})^{1/\alpha}\Bigg]\Bigg[1+(\frac{\widetilde{u}}{\widetilde{v}})^{\alpha}\Bigg]^{-1+(1/\alpha)},$$

where $\widetilde{u} = -\text{Log}\, u$ and $\widetilde{v} = -\text{Log}\, v$, $u, v \in [0,1]$, and the copula paramater $\alpha \in [0,\infty]$.
The inverse of the conditional gumbel distribution can not be solved analytically and hence we have to rely on a numerical method.

A.4 Gaussian Copula

Gaussian copula is an Elliptical copula, which is completely symmetric and has zero probability for any left/right extreme dependency. It is a one-parameter based copula. The various transformations for Gaussian copula are given as

$$C(u,v|\rho) = C(u,v|\rho) = \int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)}\frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}}\Bigg\{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}\Bigg\},$$

$$c(u,v|\rho) = \frac{1}{(1-\rho^2)^{1/2}}\exp\Bigg\{\frac{[\rho^2(\Phi^{-1}(u))^2-2\rho\Phi^{-1}(u)\Phi^{-1}(v)+\rho^2(\Phi^{-1}(v))^2]}{2(1-\rho^2)}\Bigg\},$$

$$C_{u|v}(u|v;\rho) = \Phi\Bigg(\frac{\Phi^{-1}(u)-\rho\Phi^{-1}(v)}{\sqrt{1-\rho^2}}\Bigg),$$

$$C^{-1}_{u|v}(u|v;\rho) = \Phi\Bigg(\Phi^{-1}(u)\sqrt{1-\rho^2}+\rho\Phi^{-1}(v)\Bigg).$$

$\Phi^{-1}(.)$ is the standard normal quantile function. $u, v \in (0,1)$, and the correlation paramater $\rho \in (-1,1)$.

A.5 Student-t Copula

Student-t copula is an Elliptical copula and is symmetric. It has tail dependency dictated by the degrees of freedom. It is two-parameter based copula. The various transformations for Student-t copula are given as

$$C(u,v|\Theta) = C(u,v|\rho,\nu) = \int_{-\infty}^{t_\nu^{-1}(u)}\int_{-\infty}^{t_\nu^{-1}(v)}\frac{\Gamma(\frac{(\nu+2)}{2})}{\Gamma(\frac{\nu}{2})\pi\nu(1-\rho^2)^{\frac{1}{2}}}\exp\Bigg\{1+\frac{(x^2-2\rho xy+y^2)}{\nu(1-\rho^2)}\Bigg\}^{-(\nu+2)/2},$$

$$c(u,v|\rho,\nu) = \frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})}{\sqrt{1-\rho^2}[\Gamma(\frac{\nu+1}{2}]^2]},$$

$$\times \frac{([1+\frac{(t_\nu^{-1}(u))^2}{\nu}][1+\frac{(t_\nu^{-1}(v))^2}{\nu}])^{\frac{\nu+1}{2}}}{[1+\frac{(t_\nu^{-1}(u))^2+(t_\nu^{-1}(v))^2-2\rho t_\nu^{-1}(u)t_\nu^{-1}(v)}{\nu(1-\rho^2)}]^{\frac{\nu+2}{2}}},$$

$$C_{u|v}(u|v;\rho,\nu) = t_{\nu+1}\left\{\frac{t_\nu^{-1}(u)-\rho t_\nu^{-1}(v)}{\sqrt{\frac{\nu+(t_\nu^{-1}(v))^2(1-\rho^2)}{\nu+1}}}\right\},$$

$$C^{-1}_{u|v}(u|v;\rho,\nu) = t_\nu\left\{t^{-1}_{\nu+1}(u)\sqrt{\frac{\nu+(t_\nu^{-1}(v))^2(1-\rho^2)}{\nu+1}}+\rho t_\nu^{-1}(v)\right\},$$

where $\Gamma(a)$ is the gamma function equal to

$$\Gamma(a) = \int_0^\infty x^(a-1)e^{-x}dx.$$

$t^{-1}(.)$ denotes the standardized student-t quantile function. $u, v \in [0,1]$ and the correlation paramater $\rho \in (-1,1)$, and the degrees-of-freedom parameter $\nu > 0$.

B Re-Parametrization

Copula families generally vary across in terms of their respective parameters ranges. We propose a re-parameterization for a copula. Such a method unifies the range and makes the proposal's distribution support over the real line $\mathbb{R}$.

If the transformation is $\mathbf{Z} : \Theta \mapsto \Psi$, then for various copula families,

Clayton Copula

$\alpha \in (0,\infty)$, $\mathbf{Z}(\alpha) = \exp(\psi)$.

Gumbel Copula

$\alpha \in [1,\infty)$, $\mathbf{Z}(\alpha) = \exp(\psi) + 1$.

Gaussian Copula

$\rho \in [-1, 1], \qquad \mathbf{Z}(\rho) = \dfrac{1 - e^{-\psi}}{1 + e^{-\psi}}.$

Student-t Copula

$\rho \in [-1, 1], \qquad \mathbf{Z}_1(\rho) = \dfrac{1 - e^{-\psi_1}}{1 + e^{-\psi_1}}.$

$\nu \in [1, \infty), \qquad \mathbf{Z}_2(\nu) = \exp(\psi_2) + 1.$

C Sampling from a Truncated Copula

Each u_t is sampled from the conditional copula distribution conditional on u_{t-1} and truncated through the order statistics. As we showed before the neighbouring instances are simply to ensure truncation.

Let a and b be the lower and upper limit, for a draw u to lie in. Then as seen, the conditional copula distribution is given as

$$C_{u|v}(u|v; \Theta) = \frac{\partial C(u, v|\Theta)}{\partial v}.$$

Let $C^{Tr}_{u|v}(u|v; \Theta, a, b)$ be the truncated conditional copula, such that

$$\begin{aligned} C^{Tr}_{u|v}(u|v; \Theta, a, b) &= \frac{\int_a^u c_{t|v}(t|v; \Theta) dt}{C_{b|v}(b|v; \Theta) - C_{a|v}(a|v; \Theta)}, \\ &= \frac{C_{u|v}(u|v; \Theta) - C_{a|v}(a|v; \Theta)}{C_{b|v}(b|v; \Theta) - C_{a|v}(a|v; \Theta)}. \end{aligned}$$

Let $C^{Tr}_{u|v}(u|v; \Theta, a, b) = w$, then we can re-arrange equation above to

$$w \cdot [C_{b|v}(b|v; \Theta) - C_{a|v}(a|v; \Theta)] + C_{a|v}(a|v; \Theta) = C_{u|v}(u|v; \Theta).$$

Let the L.H.S in above equation be equal x, so

$x = C_{u|v}(u|v;\Theta)$, where now by simply inverting the conditional distribution we get

$u = C^{-1}_{x|v}(x|v;\Theta)$.

$u \in (a,b)$. Hence we simply need the inverse of the conditional distribution to successfully draw truncated instances.

References

Al-Osh, M. A. and Alzaid, A. A. First-order integer-valued autoregressive (inar(1)) process. *Journal of Time Series Analysis*, 8(3):261–275, 1987.

Baillie, Richard T.; Bollerslev, Tim, and Mikkelsen, Hans Ole. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3–30, September 1996.

Beare, Brendan K. Copulas and temporal dependence. *Econometrica*, 78(1):395–410, 2010a.

Beare, Brendan K. Archimedean copulas and temporal dependence. University of california at san diego, economics working paper series, Department of Economics, UC San Diego, Sep 2010b.

Bollerslev, Tim. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3): 307–327, April 1986.

Bouyé, Eric; Durrleman, Valdo; Nikeghbali, Ashkan; Riboulet, Gaël, and Roncalli, Thierry. Copulas for finance - a reading guide and some applications. *Social Science Research Network Working Paper Series*, November 2007.

Chen, Xiaohong and Fan, Yanqin. Estimation of copula-based semiparametric time series models. Working Papers 0226, Department of Economics, Vanderbilt University, October 2002.

Chen, Xiaohong; Wu, Wei Biao, and Yi, Yanping. Efficient estimation of copula-based semiparametric markov models. Cowles Foundation Discussion Papers 1691, Cowles Foundation for Research in Economics, Yale University, February 2009.

Cherubini, U. and Luciano, E. Value-at-risk trade-off and capital allocation with copulas. *Economic Notes*, 30 (2), 2001.

Chib, Siddhartha and Greenberg, Edward. Analysis of multivariate probit models. *Biometrika*, pages 347–361, 1998.

Chib, Siddhartha and Winkelmann, Rainer. Markov chain monte carlo analysis of correlated count data. *Journal of Business & Economic Statistics*, 19(4):428–35, 2001.

W.Darswo, B. Nguyen and Olsen, E. Copulas and markov processes. *Illinois Journal of Mathematics*, 36: 600–642, 1992.

Demarta, Stefano and McNeil, Alexander J. The t copula and related copulas. *International Statistical Review*, 73:111–129, 2005.

Embrechts, P.; Lindskog, F., and Mcneil, A. Modelling Dependence with Copulas and Applications to Risk Management. In *Handbook of Heavy Tailed Distributions in Finance*, chapter 8, pages 329–384. 2003.

Embrechts, Paul; McNeil, Alexander, and Straumann, Daniel. Correlation and dependence in risk management: Properties and pitfalls. In *Risk Management: Value At Risk And Beyond*, pages 176–223. Cambridge University Press, 1999.

Engle, Robert F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, July 1982.

Freeland, R.K. *Statistical analysis of discrete time series with applications to the anaylsis of worker's compensation claims data.* PhD thesis, The University of British Columbia, Canada, 1998.

Freeland, R.K. and McCabe, B.P.M. Forecasting discrete valued low count time series. *International Journal of Forecasting*, 20(3):427 – 434, 2004.

Genest, C. and Nešlehová, J. A primer on copulas for count data. *Astin Bulletin*, 37(2):475, 2007.

Hoff, Peter D. Extending the rank likelihood for semiparametric copula estimation. *Ann. Appl. Stat.*, 1(1): 265–283, 2007.

Ibragimov, Rustam. Copula-based characterizations for higher order markov processes. *Econometric Theory*, 25(03):819–846, June 2009.

Joe, H. *Multivariate Models and Dependence Concepts.* Chapman & Hall/CRC, 1997.

Lentzas, G and Ibragimov, Rustam. Copula and long memory. 2008.

MacDonald, I.L. and Zucchini, W. *Hidden Markov and Other Models for Discrete- valued Time Series: A Practical Introduction using R.* Monographs on Statistics and Applied Probability. Taylor & Francis, 1997. ISBN 9780412558504.

Mikosch, T. and Stărică, C. *Limit Theory for the Sample Autocorrelations and Extremes of a Garch (1,1) Process*. Preprint // Department of Mathematics, Chalmers University of Technology, Göteburg University. Department, Univ, 1998.

Nelsen, R. B. *An Introduction to Copulas*. Springer, 2007.

Patton, A. J. Modelling Asymmetric Exchange Rate Dependence. *International Economic Review*, 47(2): 527–556, 2006.

Pitt, Michael; Chan, David, and Kohn, Robert. Efficient bayesian inference for gaussian copula regression models. *Biometrika*, 93(3):537–554, September 2006.

Rodriguez, Juan Carlos. Measuring financial contagion: A copula approach. *Journal of Empirical Finance*, 14 (3):401–423, June 2007.

Sklar, A. Fonctions de répartition à n dimensions et leurs marges. *Publications de l Institut Statistique de l'Univwesitè de Paris*, 8:229–31, 1959.

Smith, Michael S. and Khaled, Mohamad A. Estimation of copula models with discrete margins via Bayesian data augmentation. 107(497):290–303, 2012.

Smith, Murray D. Modelling sample selection using archimedean copulas. *Econometrics Journal*, 6(1):99–123, 2003.

Trivedi, Pravin K. and Zimmer, David M. Copula Modeling: An Introduction for Practitioners. *Foundations and Trends in Econometrics*, 1(1):1–111, 2006. ISSN 1551-3076.

Zimmer, David M. and Trivedi, Pravin K. Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand. *Journal of Business & Economic Statistics*, 24:63–76, 2006.