Bayesian Inference for Latent Factor Copulas and Application to Financial Risk Forecasting

Benedikt Schamberger †, Lutz F. Gruber † and Claudia Czado *

Department of Mathematics, Technical University of Munich, 85748 Garching, Germany;
schamberger.benedikt@gmail.com (B.S.); lutz.gruber@gmail.com (L.F.G.)
* Correspondence: cczado@ma.tum.de; Tel.: +49-89-289-17428
† These authors contributed equally to this work.

Abstract: Factor modeling is a popular strategy to induce sparsity in multivariate models as they scale to higher dimensions. We develop Bayesian inference for a recently proposed latent factor copula model, which utilizes a pair copula construction to couple the variables with the latent factor. We use adaptive rejection Metropolis sampling (ARMS) within Gibbs sampling for posterior simulation: Gibbs sampling enables application to Bayesian problems, while ARMS is an adaptive strategy that replaces traditional Metropolis-Hastings updates, which typically require careful tuning. Our simulation study shows favorable performance of our proposed approach both in terms of sampling efficiency and accuracy. We provide an extensive application example using historical data on European financial stocks that forecasts portfolio Value at Risk (VaR) and Expected Shortfall (ES).

1. Introduction

Copulas (Sklar 1959) are powerful models of multivariate dependence. Copula modeling has had profound impact on the field of financial econometrics by substantially improving the quality of key metrics of financial risk of portfolios such as Value at Risk (VaR) and Expected Shortfall (ES) (see Embrechts et al. (1999)).

Even in the simplest multivariate Gaussian model the number of correlation parameters increases quadratically in the number of variables d, and the quadratic growth in dependence parameters also extends to more general copula models. This illustrates that sparse modeling of multivariate dependence becomes critical for robust estimation and forecasting in increasingly high-dimensional problems. One approach to achieve parsimony is to use dynamic latent factors to capture dependence induced by common underlying economic activity components. Dynamic factor models, also known as index models, were first suggested by Geweke (1978) and Sargent and Sims (1977) based on ideas of Burns and Mitchell (1946), and later generalized by Forni et al. (2000, 2012), among others. Factor modeling has become popular in credit risk modeling (see for example Gordy (2000); Crouhy et al. (2000)), with the multivariate Gaussian and t copulas serving as the backbone of many credit risk models (Frey and McNeil 2003).

We present a Bayesian latent factor copula model that reduces the quadratic growth rate of the number of parameters to a linear one. Factor copulas have previously been successfully employed to scale the dimension of multivariate dependence models (Murray et al. (2013); Krupskii and Joe (2013, 2015); Oh and Patton (2017)). The pair copula construction-based factor model includes the Gaussian

factor model as a special case, and allows much more flexible fine-tuning to capture elaborate dependence characteristics than Murray et al. (2013) and Oh and Patton (2017)'s factor copula models.

Our proposed model builds on Krupskii and Joe (2013)'s, and shares their strategy to use bivariate pair copulas to couple the marginal variables with a latent factor. In contrast to their work, we present a Bayesian learning strategy that replaces frequentist uniform integration over the latent factor with Bayesian posterior simulation. A key advantage of Bayesian modeling over frequentist modeling is that credible intervals of the predictive distributions of derived quantities such as value at risk or expected shortfall can be easily obtained.

Copula models are typically estimated in two stages (see, for example, (Joe and Xu 1996) and (Joe 2005)): first, univariate marginal models are estimated to make the observed data independent and identically distributed on each of the margins by applying the cumulative distribution function of the marginal models to the observed data; second, a (multivariate) copula is estimated based on the transformed, marginally i.i.d. input data. The generally unfavorable view of two stage estimation procedures must be suspended in the context of copula estimation: the definition a copula requires that the marginal distributions of a copula be uniform on the unit interval; hence marginal models must be estimated to obtain said marginally i.i.d. uniform data before any copula modeling can be performed. Indeed, Czado et al. (2011)'s study confirms that joint Bayesian estimation of the marginal models and copula does not lead to improved estimates; the Markov chain Monte Carlo (MCMC) iterations until convergence of the marginal models produce non-uniform marginal input data for the copula, which leads to substantially increased MCMC runtime. An added benefit of estimating the marginal and dependence models separately is that inference methods for the copula can be applied universally as they do not depend on characteristics of the marginal data (which is marginally i.i.d. uniform, by definition).

We use adaptive rejection Metropolis sampling (ARMS) within Gibbs sampling (Gilks et al. 1995) for posterior simulation to eliminate the need for lengthy pilot runs to fine-tune proposal parameters: this strategy retains the Gibbs sampler's main conceptual idea to recursively simulate from the full conditional distributions, but uses ARMS to generate samples from full conditionals that cannot be drawn from directly. This way, there is no need for Metropolis-Hastings updates (Metropolis et al. (1953); Hastings (1970)), which typically require substantial effort from the user to carefully tune the proposals.

The proposed approach is then illustrated to forecast the VaR and ES of a portfolio of European bank stocks. The Bayesian factor copula model combines univariate daily one-day ahead from the marginal return time series, which are modeled by dynamic linear models (DLMs) with ARMA(1,1) structure (West and Harrison 1997).

The paper is organized as follows: we define the latent factor copula in Section 2, and present Bayesian theory as well as several MCMC-based strategies for posterior simulation in Section 3. Section 4 presents a simulation study that investigates the sampling performance of our simulation strategies, and Section 5 discusses an application to financial time series data. The paper concludes with summary comments in Section 6.

2. Latent Factor Copulas

A latent factor copula models the dependence among d variables $u_j \sim Unif(0,1)$, $j = 1{:}d$, by a linking copula $c_j(u_j, v; \theta_j)$ with copula parameter θ_j for each variable u_j and one common latent variable v. Each of the linking copulas c_j can be any bivariate copula; a detailed discussion of many bivariate copulas can be found in Joe (1997). Conditionally on the latent variable v, the density function of the latent factor copula is

$$c(u_1, \ldots, u_d | v) = \prod_{j=1{:}d} c_j(u_j, v; \theta_j). \tag{1}$$
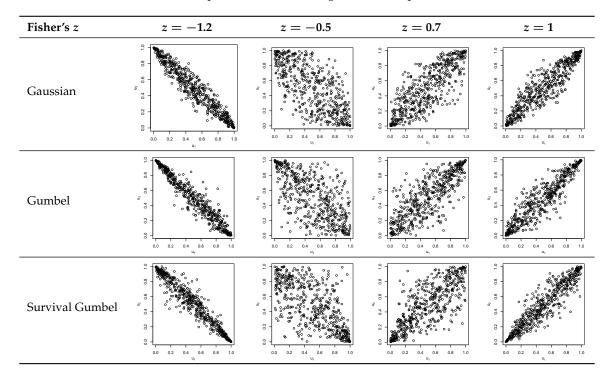
Linking Copulas

We will use the bivariate Gaussian and Gumbel copulas (Joe 1997) as candidate linking copulas; these are the same linking copulas used by Krupskii and Joe (2013). The Gaussian copula is tail-independent and symmetric, while the Gumbel copula is asymmetric and lower tail-dependent. Density functions and properties about these copulas are summarized in Table 1, while Table 2 shows pairs plots for several different parameter values. To further reduce model complexity, we assume a common linking copula family for all margins, but allow for a margin-specific copula parameter, that is, $c_j(u_j, v, \theta_j) = c(u_j, v, \theta_j)$.

Table 1. Density functions and Kendall's τ as a function of the parameter of the Gaussian and Gumbel pair copulas.

Copula	Density Function	$\tau = h(\theta)$
Gaussian	$c_{\mathcal{N}}(u_1, u_2; \theta) = \frac{1}{\sqrt{1-\delta^2}2\pi\phi(\Phi^{-1}(u_1))\phi(\Phi^{-1}(u_2))} \times$ $\exp\left(-\frac{\Phi^{-1}(u_1)+\Phi^{-1}(u_2)-2\delta\Phi^{-1}(u_1)\Phi^{-1}(u_2)}{2(1-\delta^2)}\right)$	$h_{\mathcal{N}}(\theta) = \frac{2}{\pi}\arcsin(\theta)$
Gumbel	$c_{\mathcal{G}}(u_1, u_2; \theta) = \frac{1}{u_1 u_2}(x_1 x_2)^{\theta-1} \times \exp\left(\frac{-x_1^\theta - x_2^\theta}{\theta}\right) \times$ $\left((1-\theta)\left(-x_1^\theta - x_2^\theta\right)^{\frac{1}{\theta}-2} + \left(-x_1^\theta - x_2^\theta\right)^{\frac{2}{\theta}-2}\right),$ where $x_1 = -\ln u_1$ and $x_2 = -\ln u_2$	$h_{\mathcal{G}}(\theta) = 1 - \frac{1}{\theta}$
Survival Gumbel	$c_{\mathcal{SG}}(u_1, u_2; \theta) = c_{\mathcal{G}}(1 - u_1, 1 - u_2; \theta)$	$h_{\mathcal{SG}}(\theta) = 1 - \frac{1}{\theta}$

Table 2. Pairs plots for the Gaussian, Gumbel and survival Gumbel copulas for different parameter values. As the Gumbel and survival Gumbel copulas only exhibit positive dependence, the 90° and 270° rotations of the Gumbel copulas are shown for negative Fisher z parameters.

We parameterize the linking copulas by the Fisher z-transformation of Kendall's τ, given by

$$z = z(\tau) = \frac{1}{2} \ln \left(\frac{1 + \tau}{1 - \tau} \right), \qquad \tau \in (-1, 1), \tag{2}$$

$$\tau = z^{-1}(z) = 1 - \frac{2}{e^{2z} + 1}, \qquad z \in \mathbb{R}. \tag{3}$$

This is a more convenient parameterization for analysis across different copula families than the copulas' natural parameters, because it makes the parameters comparable across different copula families. The domains of the Fisher z-transformation depend on the copula family; for example, the Fisher z parameter of a Gumbel copula takes values in $(0, \infty)$, while that of the Gaussian copulas takes values in $(-\infty, \infty)$.

Likelihood

Let $\boldsymbol{u} = (u_{tj})_{t=1:T, j=1:d}$ be T i.i.d. d-dimensional observations $\boldsymbol{u}_t = (u_{t1}, \ldots, u_{td})$, $t = 1{:}T$, in the copula domain $[0, 1]^d$; and denote the states of the latent factor by $\boldsymbol{v} = (v_t)_{t=1:T}$. The joint likelihood function $\ell(\boldsymbol{\theta}, \boldsymbol{v}; \boldsymbol{u})$ of the latent factor $\boldsymbol{v}$ and copula parameters $\boldsymbol{\theta} = (\theta_j)_{j=1:d}$ follows from (1) as

$$\ell(\boldsymbol{\theta}, \boldsymbol{v}; \boldsymbol{u}) = \prod_{t=1:T} \prod_{j=1:d} c_j(u_{tj}, v_t; \theta_j). \tag{4}$$

Substituting the Fisher z parameters $\boldsymbol{z} = (z_1, \ldots, z_d)$, where $z_i := z(h_j(\theta_j)$ (see (2) for $z(\cdot)$ and Table 1 for $h(\cdot)$), for the natural copula parameters $\boldsymbol{\theta}$, the joint likelihood of the latent factor $\boldsymbol{v}$ and Fisher z parameters $\boldsymbol{z} = (z_j)_{j=1:d}$ follows as

$$\ell(\boldsymbol{z}, \boldsymbol{v}; \boldsymbol{u}) = \prod_{t=1:T} \prod_{j=1:d} c_j \left(u_{tj}, v_t; h_j^{-1} \left(z^{-1}(z_j) \right) \right). \tag{5}$$

Frequentist Inference

Frequentist inference (Krupskii and Joe 2013) integrates over the latent variable v to obtain the unconditional density function of the latent factor copulas as

$$c(u_1, \ldots, u_d) = \int_0^1 \prod_{j=1:d} c_j(u_j, v; \theta_j) \, dv, \tag{6}$$

and then maximizes the unconditional likelihood function with respect to the copula parameters.

3. Bayesian Posterior Analysis

In Bayesian analyses, prior information about the unknown parameters is quantified in a "prior distribution." Once data is observed, the prior distribution and likelihood are combined to derive the posterior distribution—the distribution of the unknown parameters given the observed data. As the posterior distribution typically can not be derived in a closed form, Markov chain Monte Carlo (MCMC) methods are used to simulate from the posterior distribution.

3.1. Prior Distributions

Prior for Copula Parameters $\boldsymbol{z}$

Each linking copula's Fisher z parameter can take values in $D_{\mathcal{SG}} = D_{\mathcal{G}} = (0, \infty)$ or $D_{\mathcal{N}} = (-\infty, \infty)$, depending on whether it is a Gaussian or Gumbel copula. We propose normally-distributed priors on the Fisher z parameter, truncated to the allowable parameter range $D_{\mathcal{SG}} = D_{\mathcal{G}}$ and $D_{\mathcal{N}}$, respectively. Furthermore, we assume that the priors are independent over all linking copulas $j = 1{:}d$:

$$z_j \sim N_{D_j}(0, \sigma_j^2),$$
$$\pi(z) = \prod_{j=1:d} n_{D_j}(z_j; 0, \sigma_j^2), \tag{7}$$

where $N_D(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2 truncated to D with corresponding probability density function $n_D(\cdot; \mu, \sigma^2)$.

Prior for Latent Factor v

In line with the frequentist assumption that integrates out the latent factor v using a uniform density, we choose a uniform prior for v, independent across observations $t = 1{:}T$,

$$v_t \sim Unif(0, 1),$$
$$\pi(v) = \prod_{t=1:T} \pi(v_t) \equiv 1. \tag{8}$$

Given that the latent factor v_t feeds directly into a copula distribution, the uniform distribution on the unit interval is the only natural choice that does not violate the requirement of uniform copula margins.

Joint Prior for z and v

We assume no dependence between the latent factor v and copula parameters z, and choose the joint prior density $\pi(z, v)$ of z and v as the product of the individual prior densities,

$$\pi(z, v) = \pi(z)\pi(v). \tag{9}$$

Independence priors are a common choice in Bayesian modeling, and they do not prevent the posterior from exhibiting data-induced dependence; more comments below after the posterior density (10).

3.2. Posterior Distribution and Full Conditionals

Posterior Density

An analytical expression of the joint posterior density $p(z, v|u)$ of the copula parameters z and latent factor v, given observations u can only be obtained up to a normalizing constant,

$$p(z, v|u) \propto \ell(z, v; u)\pi(z, v). \tag{10}$$

Indeed, this posterior density cannot be factorized into independent marginal distributions for z and v, given that the likelihood function does not exhibit such a product form and thus induces multivariate dependencies in the posterior distribution.

Full Conditionals

The full conditionals of the latent factor $v = (v_t)_{t=1:T}$ and Fisher z parameters $z = (z_j)_{j=1:d}$ can be derived by straightforward calculation.

The joint posterior density $p(z, v|u)$ of z and v, given observations u can be written as

$$p(z, v \mid u) \propto \ell(z, v; u)\pi(z, v) = \ell(z, v; u)\pi(z)\pi(v) = \ell(z, v; u) \prod_{j=1:d} N_{D_j}(z_j)$$

$$= \left(\prod_{t=1:T} \prod_{j=1:d} c_j \left(u_{tj}, v_t; h_j^{-1}\left(z^{-1}(z_j)\right) \right) \right) \prod_{j=1:d} n_{D_j}(z_j; 0, \sigma_j^2); \tag{11}$$

here we used the likelihood from (5) and priors from (7) to (9). Then, for all $t = 1 : T$, the full conditional of v_t, given all other latent factor states $\boldsymbol{v}_{-t} := (v_1, \ldots, v_{t-1}, v_{t+1}, \ldots, v_T)$, observations $\boldsymbol{u}$ and Fisher z copula parameters $\boldsymbol{z}$ is

$$p(v_t \mid \boldsymbol{v}_{-t}, \boldsymbol{u}, \boldsymbol{z}) = \frac{p(\boldsymbol{z}, \boldsymbol{v}|\boldsymbol{u})}{p(\boldsymbol{z}, \boldsymbol{v}_{-t}|\boldsymbol{u})} \propto p(\boldsymbol{z}, \boldsymbol{v} \mid \boldsymbol{u}) \propto \prod_{j=1:d} c_j(u_{tj}, v_t; z_j), \tag{12}$$

where the prior densities of v_t and $\boldsymbol{z}$ are constant factors independent of v_t.

Similarly to (12), the full conditional of z_l, given all other Fisher z copula parameters $\boldsymbol{z}_{-j}$, observations $\boldsymbol{u}$ and latent factor $\boldsymbol{v}$, can be derived as

$$p(z_j \mid \boldsymbol{z}_{-j}, \boldsymbol{u}, \boldsymbol{v}) \propto \left(\prod_{t=1:T} c_j(u_{tj}, v_t; z_j) \right) n_{D_j}(z_j; 0, \sigma_j^2). \tag{13}$$

3.3. MCMC Methods for Posterior Simulation

The complex forms of the posterior and full conditional densities render direct sampling infeasible. We investigate the performance characteristics of several different sampling strategies:

- Metropolis-Hastings within Gibbs sampling (Metropolis et al. (1953); Hastings (1970); Gelfand and Smith (1990)):

  - MCM: Truncated normal proposals for $\boldsymbol{z}$ and Beta proposals for $\boldsymbol{v}$; proposal mode and curvature match those of the full conditionals;
  - EVM: Gamma proposals for $\boldsymbol{z}$ and Beta proposals for $\boldsymbol{v}$; proposal expectation and variance match those of the full conditionals;
  - IRW: Truncated normal random walk proposals for $\boldsymbol{z}$ and uniform independence proposals $\boldsymbol{v}$; and

- ARMGS: Adaptive rejection Metropolis sampling within Gibbs sampling (Gelfand and Smith (1990); Gilks et al. (1995)).

Details of the sampling schemes are given in Appendix A.1. We will investigate these strategies in a simulation study using Gumbel linking copulas (Section 4), and then adapt the best-performing method to also work for Gaussian linking copulas. Here the main difference will be the extension from only non-negative Fisher z values to positive and negative values.

4. Simulation Study

We test our four different MCMC sampling strategies in strong, weak, and mixed dependence scenarios. The goal is to verify that these samplers provide good estimates, and to determine if some perform better than others.

4.1. Simulation Setup

Scenarios

To explore the behavior of the Gibbs sampler under different dependence characteristics between the marginal data $\boldsymbol{u}$ and the latent factor $\boldsymbol{v}$, we consider three different scenarios: in the first scenario, all linking copulas show weak dependence, with Kendall's τ values of $0.1 \leq \tau \leq 0.2$; in the second scenario, we use high Kendall's τ values of $0.5 \leq \tau \leq 0.8$; and in the third scenario, we use a combination of strong and weak dependence between the marginal data and latent factor with Kendall's τ's of the linking copulas varying between $0.1 \leq \tau \leq 0.8$. Table 3 summarizes the Kendall's τ's, copula parameters θ, and Fisher z parameters $\boldsymbol{z}$ of all scenarios.

Table 3. Copula parameters used in the simulation to model the dependence between each marginal series $u_{:,j} := (u_{1j}, \ldots, u_{Tj})$ and the latent factor v.

	c_1	c_2	c_3	c_4	c_5
			Low τ		
τ	0.10	0.12	0.15	0.18	0.20
θ	1.11	1.14	1.18	1.21	1.25
z	0.10	0.13	0.15	0.18	0.20
			High τ		
τ	0.50	0.57	0.65	0.73	0.80
θ	2.00	2.35	2.86	3.64	5.00
z	0.55	0.65	0.78	0.92	1.10
			Mixed τ		
τ	0.10	0.28	0.45	0.62	0.80
θ	1.11	1.38	1.82	2.67	5.00
z	0.10	0.28	0.48	0.73	1.10

Simulation Data

We investigate $N = 100$ data sets from the latent factor copula model in three different scenarios. Each simulation data set contains $T = 200$ observations of dimension $d = 5$. We use Gumbel linking copulas for each pair $c_1, \ldots, c_5$. As an asymmetric extreme value copula with upper tail dependence and lower tail independence, the bivariate Gumbel copula can be a good fit to model financial data such as large negative stock returns: the probability of many stocks selling off simultaneously tends to be much higher than the probability of many stocks all making outsize gains on the same day. The density of the bivariate Gumbel copula quickly tends to infinity for $(u, u) \downarrow (0, 0)$ or $(u, u) \uparrow (1, 1)$, allowing to test the robustness of our proposed sampling routines for linking copulas that show difficult numerical behavior.

Posterior Simulation

For each of the $N = 100$ simulation data sets from the three scenarios described in Section 4.1, we generated $10,000$ MCMC samples with each sampling method from Section 3.3, using the starting values from our heuristic of Appendix A.2 and prior variances $\sigma_j^2 = 100^2$ for the Fisher z copula parameters.

The proposal variance for our IRW method were tuned in pilot runs with $1,000$ iterations each to achieve acceptance rates between 20% and 30% for each Fisher z parameter $z_1, \ldots, z_5$; we specified this range of desirous acceptance rates in line with Roberts et al. (1997)'s suggestion for tuning to 23%. Uniform $Unif(0, 1)$ proposals for the latent factor $v = (v_1, \ldots, v_T)$ lead to acceptance rates around 20%, which we accepted as good enough. These settings performed well across all three scenarios.

4.2. Results

Kendall's τ Pair Copula Parameters

For ease of interpretation, we analyze the results posterior samples in terms of Kendall's τ, which is a simple transformation from Fisher's z. Posterior trace plots show immediate convergence to an equilibrium state as well as good mixing behavior (see Figure 1); the latter is confirmed by posterior density plots. The runtimes of our different sampling strategies varied substantially: the slowest method (EVM) took 14 times as long as the fastest (IRW) to generate $10,000$ posterior samples (Table 4); the faster methods, IRW and ARMGS, also out-performed the slower ones in terms of effective sample size (ESS; Thiébaux and Zwiers (1984)) per minute. Table 4 shows that Metropolis-Hastings within Gibbs sampling with IRW updates was marginally faster than adaptive rejection Metropolis

sampling within Gibbs sampling (ARMGS), but the latter's coverage of credible intervals was better. Metropolis-Hastings within Gibbs Sampling with MCM updates provided the smallest point estimation errors (both mean absolute deviation and mean squared errors) of all analyzed Bayesian procedures, but it still trails frequentist maximum likelihood estimation (MLE). Overall, our ARMGS and IRW performed substantially better than MCM and EVM. As an adaptive procedure, ARMGS requires less effort of the user than IRW, making it the grand winner.
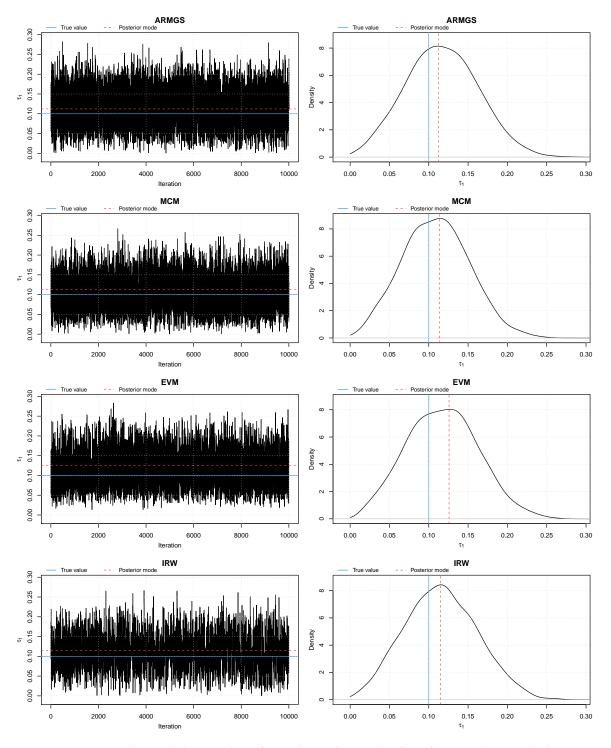

Figure 1. Trace plots and density plots of a single run from each of our four sampling methods on Kendall's τ scale. The output is based on $10,000$ posterior samples of z_1 in the mixed τ scenario.

Table 4. Mean absolute deviation (MAD), mean squared error (MSE), effective sample size (ESS) per minute, and realized coverage of 95% credible intervals (C.I.) averaged across all variables $\tau_1, \ldots, \tau_5$ and $N = 100$ replications. Runtime is for 10,000 MCMC iterations on a 16-core node.

	ARMGS	MCM	EVM	IRW	MLE
Low τ					
MAD	0.1088	**0.0900**	0.1105	0.1032	0.0739
MSE	0.0314	**0.0129**	0.0327	0.0262	0.0105
ESS/min	6.1	0.5	0.2	**9.2**	n/a
95% C.I.	**0.96**	0.61	**0.94**	0.93	n/a
High τ					
MAD	**0.0292**	**0.0292**	0.0293	0.0294	0.0212
MSE	**0.0014**	**0.0014**	**0.0014**	**0.0014**	0.0007
ESS/min	**23.7**	1.6	1.4	22.2	n/a
95% C.I.	**0.95**	0.91	**0.95**	**0.95**	n/a
Mixed τ					
MAD	0.0509	**0.0434**	0.0517	0.0506	0.0339
MSE	0.0043	**0.0030**	0.0045	0.0042	0.0017
ESS/min	21.4	1.9	1.1	**29.2**	n/a
95% C.I.	**0.93**	0.84	**0.93**	0.90	n/a
Average across All Scenarios					
Runtime	165s	664s	776s	**55s**	

Latent Factor States

Bayesian inference of our latent factor copula also provides posterior samples of the latent factor v, which can enable a more thorough analysis of dependence patterns in real-world applications. Again, adaptive rejection Metropolis sampling within Gibbs sampling (ARMGS) is the grand winner, showing the best combination of sampling efficiency (ESS/min), coverage of credible intervals and accuracy of point estimates (Table 5). The MCM and EVM methods produced very slightly better point estimates that ARMGS and IRW, but the latter two methods realized a more than 10 times higher effective sample size during the benchmark 10 min runtime than MCM and EVM.

Table 5. Mean absolute deviation (MAD), mean squared error (MSE), effective sample size (ESS) per minute, and realized coverage of 95% credible intervals (C.I.) averaged across all latent variables $v_1, \ldots, v_T$ and $N = 100$ replications.

	ARMGS	MCM	EVM	IRW
Low τ				
MAD	0.2808	0.2975	**0.2805**	0.2811
MSE	**0.1248**	0.1397	**0.1248**	0.1251
ESS/min	25.7	0.5	1.1	**63.5**
95% C.I.	**0.91**	0.71	0.89	0.88
High τ				
MAD	0.0709	**0.0678**	0.0710	0.0709
MSE	0.0095	**0.0087**	0.0095	0.0095
ESS/min	**43.7**	1.6	2.6	38.8
95% C.I.	**0.95**	0.80	**0.95**	0.94
Mixed τ				
MAD	0.0828	0.0898	**0.0824**	0.0826
MSE	0.0132	0.0154	**0.0131**	**0.0131**
ESS/min	25.8	1.5	2.0	**34.3**
95% C.I.	**0.88**	0.78	0.87	0.83

5. Application: Portfolio Value at Risk of European Financial Stocks

This section presents a case study to forecast value at risk (VaR) and expected shortfall (ES) using a dynamic multivariate model that consists of our latent factor copula and univariate time series DLMs. We analyze the daily log-returns of eight European bank stocks that are head-quartered in euro zone countries over a ten year period from January 2004 through December 2013. We use adjusted closing prices, which correct for dividends, (reverse) stock splits, and other measures to calculate the daily log-returns (data are downloaded from http://finance.yahoo.com). Table 6 lists details of the selected stocks, including their ticker symbols, which we will use to identify stocks in this section, as well as the exchange they are traded on.

Table 6. Ticker symbol, company name and exchange of the selected bank stocks.

Ticker	Company Name	Exchange
ACA.PA	Credit Agricole S.A.	Euronext - Paris
BBVA.MC	Banco Bilbao Vizcaya Argentaria	Madrid Stock Exchange
BNP.PA	BNP Paribas SA	Euronext - Paris
CBK.DE	Commerzbank AG	XETRA
DBK.DE	Deutsche Bank AG	XETRA
GLE.PA	Societe Generale Group	Euronext - Paris
ISP.MI	Intesa Sanpaolo S.p.A.	Borsa Italiana
SAN.MC	Banco Santander	Madrid Stock Exchange

Since the data set includes stocks from several European countries, trading days may differ due to varying bank holidays for each exchange. Therefore, we use the union of all trading days, where for stocks that are not traded at a given day, the asset value of the previous day is reused. This method introduces some additional zero returns into the data, but avoids altering actual historical returns which would happen if we intersected the trading days of different exchanges. The so expanded data set consists of 2,606 daily 8-dimensional log-returns.

5.1. Marginal Analysis

Figure 2 shows the daily log-returns of Credit Agricole (ACA); the historical returns of the other stocks look very similar. The initial period from 2004 through 2007 was relatively calm, and was followed by a more turbulent period beginning in 2008. Market nervousness intensified during the second half of 2008, when, at the height of the sub-prime mortgage crisis, Lehman Brothers collapsed. Volatility receded in the second half of 2009 and remained subdued until the second half of 2011, when the European debt crisis hit the markets. Volatility remained at elevated levels through 2012, and then showed a gradual decline in 2013.
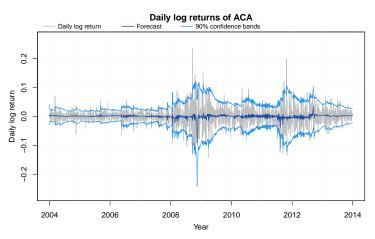

Figure 2. Daily log-returns (grey) of ACA from 2004 through 2013 with DLM one-day ahead 90% confidence bands (blue) and forecast value (dark blue).

Time Series DLMs

Since we are interested in online learning and forecasting of the return time series, we use the dynamic linear model (DLM), which is also a model of dynamic stochastic volatility. While the economics literature tends to favor GARCH models to model financial time series data, we view the DLM as a better choice as it allows for fast and efficient fully Bayesian online learning and forecasting based on analytical expressions of the posterior and predictive distributions, which can be evaluated in fractions of a second, and can also model stochastic, time-varying volatilities. If we were using GARCH models, we would have to perform lengthy MCMC-based posterior simulation for each trading day t in our dataset to generate sequential out-of-sample predictions; this computational disadvantage disqualifies GARCH models from further consideration in our study.

We use an ARMA(1,1)-structure in our marginal time series DLM with beta discounting for on-line learning of stochastic observational variances (West and Harrison 1997). Denote the daily log-returns of the j-th series by $(y_{jt})_{t=1:T}$; the model is defined by observation Equation (14)

$$y_{jt} = \mu_{jt} + \phi_{jt}y_{j,t-1} + \psi_{jt}\epsilon_{j,t-1} + \epsilon_{jt} =: \boldsymbol{F}_{jt}'\boldsymbol{\theta}_{jt} + \epsilon_{jt}, \tag{14}$$

which relates the observations y_{jt} to a time-varying local level μ_{jt}, ARMA(1,1)-structure with dynamic regression coefficients ϕ_{jt} and ψ_{jt}, and conditionally normally-distributed observation error $\epsilon_{jt} \sim N(0, 1/\lambda_{jt})$; the predictors $\boldsymbol{F}_{jt} = (1, y_{j,t-1}, \epsilon_{j,t-1})^\top$ are just those of a regular local-level ARMA(1,1) model.

Appendix C summarizes the analytic filtering equations to sequentially learn the model, and elaborates on the system equations for the prior evolutions of the states $\boldsymbol{\theta}_{jt} = (\mu_{jt}, \phi_{jt}, \psi_{jt})^\top$ and observational precision $\lambda_{jt} > 0$ (see Equations (A8) and (A10)).

Results

We started the analysis with initial normal-gamma priors as follows:

$$(\boldsymbol{\theta}_{j1}, \lambda_{j1})|\mathcal{D}_0 \sim NG(\boldsymbol{a}_{j1}, \boldsymbol{R}_{j1}, r_{j1}, c_{j1}), \tag{15}$$

where $\boldsymbol{a}_{j1} = (0,0,0)^\top$, $\boldsymbol{R}_{j1} = 10^{-3} \cdot \boldsymbol{I}_3$, $r_{j1} = 5$ and $c_{j1} = 10^{-3}$, and the *NG* notation is as explained in Appendix C. For forward filtering, we used the discount factors $\beta = 0.97$ and $\delta = 0.99$; these values provide a desirable balance between robustness of forecasts and rate of adaption to market movements, and are similar to those used in existing literature (e.g., Gruber and West 2016).

Figure 2 shows the DLM's sequential out-of-sample trend forecasts and confidence bands for ACA. It can be seen that the one-day ahead forecasts quickly adapt to changes in the volatility of the underlying stock. Table 7 shows that with our choice of discount factors, the coverage of the obtained forecast confidence intervals are very close to their theoretical values.

Table 7. Percentage of observed log-returns above and below sequentially out-of-sample 90% forecast interval.

	ACA	BBVA	BNP	CBK	DBK	GLE	ISP	SAN
Above 95% bound	4.90%	4.90%	4.43%	4.86%	4.22%	4.73%	4.73%	4.61%
Below 5% bound	4.35%	5.07%	4.65%	4.43%	5.29%	4.78%	5.33%	5.16%

Copula data is obtained by transforming the observed log-returns y_{jt} using the cumulative distribution function of the one-day ahead forecast t-distribution of y_{jt} given information set $\mathcal{D}_{t-1}$ (see Equation (A4) of Appendix C). The forecast t-distribution already accounts for uncertainty of the model parameters; it is obtained by integrating out the parameters of the normal innovation/residual distribution weighted by their respective step ahead prior distributions. The so transformed data lie in the unit interval $(0,1)$, and we will denote it by u_{jt}. We use the observations from 2004 as an initial

learning data set and ignore them in further analyses; the 2, 345 remaining 8-dimensional transformed observations (u_{jt}) define our copula data set.

5.2. Bayesian Latent Factor Copula Analysis

We used the adaptive rejection Metropolis Gibbs sampler introduced in Section 3.3 for Bayesian posterior simulation of the parameters and latent factors of our copula model. We partitioned the resulting copula data (u_{jt}) by calendar year (each year consists of roughly 260 daily observations), and select a separate copula model for each year. These annually updated models allow proper out-of-sample analysis, which would not be possible if we had estimated the model using the entire data set.

For each of the nine years, we considered three different linking copulas: the Gaussian, Gumbel and survival Gumbel copulas. We simulated 11, 000 posterior samples for each year and choice of bivariate linking copulas. We chose the starting values as discussed in Appendix A.2, and discarded the first 1, 000 of MCMC iterations as burn-in, resulting in a posterior sample size of 10, 000.

The Kendall's τ dependence parameters between each series $j = 1{:}8$ and the latent factor v are solidly positive, taking values between 0.3 and 0.8. Furthermore, there is substantial variation of the dependence parameters over the years (Figure 3). The posterior mode estimates for the three linking copulas are generally similar, with slight differences noticeably only in the details. We stress again that the credible intervals obtained through Bayesian posterior simulation allow for quantification of estimation uncertainty, which is not available in frequentist analysis.

Next, we verify our implicit assumption that the bivariate distribution of every series' copula data and the latent factor v is the one of the selected linking copula. Figure 4 shows normalized contour plots of the filtered Credit Agricole (ACA) and latent factor pairs ($u_{\mathrm{ACA},t}, \hat{v}_t$) for all three linking copulas; here we estimated each latent factor state v_t by its sample posterior mean across all 10, 000 MCMC iterations. The theoretical contours of the densities of the linking copulas (with posterior mean parameter) are provided for context. A visual inspection of Figure 4 reveals that the data are best modeled by a Gaussian linking copula, given that the data lack asymmetries so typical for the (survival) Gumbel copula.
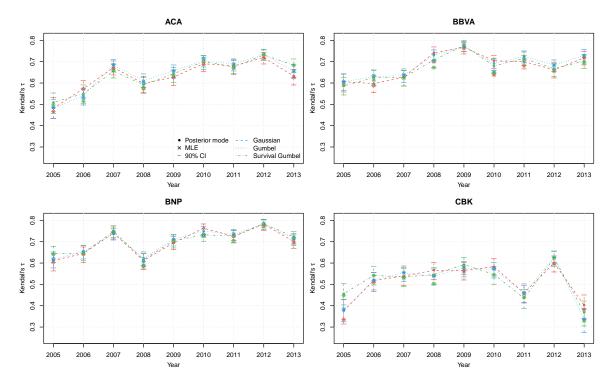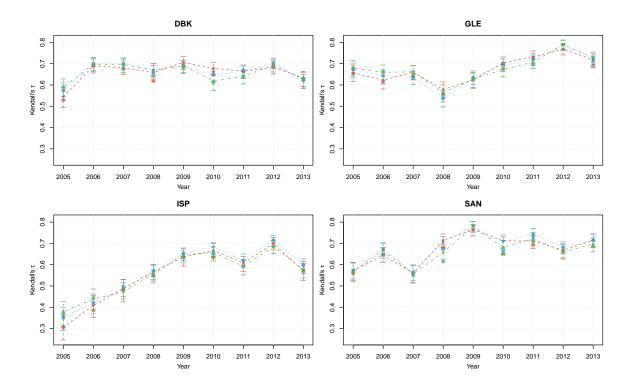

Figure 3. *Cont.*

Figure 3. Kendall's τ posterior modes with corresponding 90% credible intervals of copula parameters of the latent factor copula model with Gumbel (red), Gaussian (blue) and survival Gumbel (green) linking copulas.
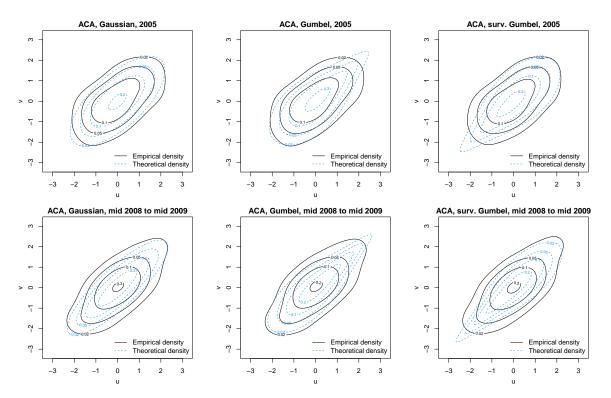

Figure 4. Contour plots with standard normal margins for ACA and latent factor pairs $(u_{\mathrm{ACA},t}, \hat{v}_t)$ for 2005 (**Top**) and 1 July 2008 to 30 June 2009 (**Bottom**). Empirical densities are indicated by solid black lines, while dotted blue lines show the theoretical densities.

5.3. Value at Risk and Expected Shortfall Forecasts

Value at risk (VaR) is one of the most widely used metrics to measure risk of financial assets. Furthermore, many current industry regulations such as Basel II/III (Basel Committee on Banking Supervision 2006, 2011) and Solvency II (European Parliament and Council 2009) use VaR to measure the risk exposure of banks and insurance companies. In recognition of the importance of VaR in real-world applications, we provide a blueprint for implementation as well as rigorous statistical verification of our model's forecasting performance.

Expected shortfall (ES) is a risk measure that is based on VaR, and is slated to replace VaR as the prime risk measure imposed by regulators (see, for example, Basel Committee on Banking Supervision (2013)). In contrast to VaR, ES is a coherent risk measure, and a worst-case value can be derived by using the comonotonic copula (McNeil et al. 2005, Remark 6.17.). However, ES is only defined for random variables with a finite mean, which might be problematic when modeling operational losses or the losses resulting from natural catastrophes.

Forecasting Method

As in Section 5.2, we estimated the copula dependence model using data from the previous year so that the forecasts for the current year are obtained out-of-sample; for example, we learned the copula using data from 2005, and used this copula without any further adjustments to generate forecasts for each trading day in 2006. We combine the yearly changing copula with the daily updated DLMs to obtain up-to-date forecasts for each trading day. Our DLM-filtered copula data cover the years 2005 to 2013, so the resulting forecasts were for 2006 to 2014. For each day, we drew $1,000$ samples from that year's copula, and transformed the $(0,1)$-copula data to the observation scale using the sequential step-ahead forecast t-distributions of the marginal time series DLMs.

We used a family-restricted regular vine copula (RV_R) as a benchmark model to test our latent factor copula against. We allowed the pair copula families to be Gaussian, Student's t, (survival) Gumbel, Frank or BB1; these copulas can describe tail-dependent and tail-independent, and symmetric and asymmetric dependence characteristics. The tree structure and pair copula families were selected using Dißmann et al. (2013)'s selection strategy as implemented in Schepsmeier et al. (2014)'s R package *VineCopula*. We then generated Bayesian parameter estimates using Gruber and Czado (2015)'s MCMC-based C++ software (we only used the within-model moves, not the between-model moves for model selection). In a last step, we simulated $(0,1)$ copula observations from the posterior sample, and transformed them to the observation level as above.

Portfolio Composition

We use a constant mix strategy with equal weights in all stocks and daily re-balancing; this is without loss of generality. Figure 5 shows the portfolio value process of this portfolio. After an initial gain of more than 40%, the portfolio value plunged to less 50% of the initial value, and never fully recovered to pre-crisis highs; it should be noted that the dismal performance of this portfolio is the result of investing in financials exclusively, and not the result of an overly poor portfolio weights strategy.
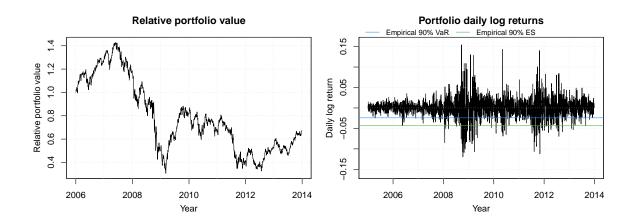
Figure 5. Historical relative portfolio value of a constant mix strategy with equal weights in the 8 bank stocks ACA, BBVA, BNP, CBK, DBK, GLE, ISP and SAN for years 2006 to 2013. Portfolio weights are readjusted daily. The 90% empirical VaR is shown in blue and the 90% empirical ES is in green.

Value at Risk Forecasts

Figure 6 shows the dynamic VaR forecasts provided by our Gaussian latent factor copula (in conjunction with marginal time series DLMs); our VaR forecasts are the 10% quantiles of each day's $1,000$ simulated portfolio returns. This figure shows that our dynamic combined multivariate time series model adapts quickly to changing market volatilities and eliminates the clustered occurrence of VaR violations found in an empirical constant volatility model (see right sub-figure of Figure 5). The realized frequency of VaR violations at the predicted levels was closest to the theoretical value of 10% for forecasts from the Bayesian Gumbel latent factor copula (10.17%), followed by the frequentist Gumbel latent factor copula (10.60%), and the regular vine copula (9.64%); furthermore, the Bayesian version of the latent factor copula always out-performed the frequentist version with the same choice of linking copula (Table 8).
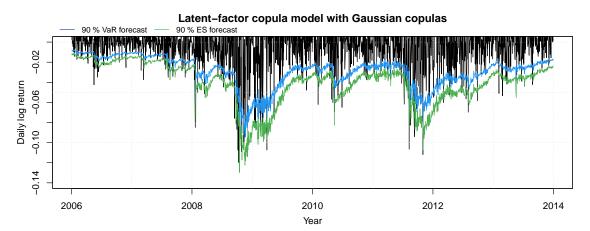

Figure 6. Daily log-returns of the equally weighted constant mix portfolio with (negative) 90% VaR (blue line) and ES (green line) forecasts.

In further analyses, we utilized the conditional coverage test suggested by (Christoffersen 2011, Chapter 13) to evaluate the performance of our models' VaR forecasts. This test jointly analyzes the rate of VaR violations as well as whether they occur independently. Specifically, the null hypothesis of this test is that the hit sequence of $(1 - \alpha\%)$ VaR violations is a first-order Markov chain, where violations happen with probability α and are independent of the current state of the chain. Again, the

Bayesian latent factor copula with Gumbel linking copulas fares best, showing the highest *p*-value of 0.44 (Table 8). Most noteworthy is that all frequentist latent factor copula models can be rejected at the 10% level, while all Bayesian latent factor copulas and the Bayesian regular vine copula pass.

Table 8. Frequency of 90% VaR violations and 90% ES of an equals weights portfolio of all eight selected financial stocks, and *p*-values of conditional coverage test of VaR violations. RV_R denotes the multivariate regular vine copula model; the other columns are for the dynamic factor copula model with the respective linking copula families. The best values are emphasized in bold.

	RV_R	Gumbel		Gaussian		Survival Gumbel	
		ARMGS	**MLE**	**ARMGS**	**MLE**	**ARMGS**	**MLE**
90% VaR viol.	9.64%	**10.17%**	10.60%	9.59%	9.36%	9.35%	9.17%
90% ES	4.07%	4.00%	3.88%	4.10%	4.08%	**4.13%**	4.08%
p-value, cond. coverage test	0.24	**0.44**	0.07	0.30	0.01	0.10	0.03

Although regulators often require higher confidence levels such as 99%, 99.5% or even 99.9%, producing robust estimates at these levels is not statistically feasible unless unrealistically big training data is available (see, e.g., McNeil et al. 2005, Example 7.15). In our study, no reliable results could be obtained about such extreme quantiles.

Expected Shortfall Forecasts

The empirical 90% ES of our equal weights portfolio from 2006 through 2014 was 4.58%, and the forecasts range between 4.00% and 4.13% for Bayesian latent factor copulas, 3.88% and 3.08% for frequentist latent factor copulas, and 4.07% for the regular vine copula (Table 8).

There are no universally accepted backtests for ES, as academic and regulatory discussions are ongoing (Embrechts et al. 2014). Part of the problem is that ES is not elicitable, which means that there is no obvious way to compare different ES forecasts. In this paper, we compare ES forecasts with the observed ES at the 90% level instead: using this decision criterion, our Bayesian latent factor model with survival Gumbel linking copulas yields performed best. Again, all frequentist latent factor copulas are out-performed by their Bayesian peers (Table 8).

6. Summary Comments

Latent factor copulas offer a simple and elegant way to parsimoniously model multivariate dependence and achieve scalability to high dimensions. Allowing for different choices of linking copulas, they provide sufficient flexibility to be used in a wide range of multivariate applications.

Our proposed Bayesian inference strategy utilizes adaptive rejection Metropolis sampling within Gibbs sampling for posterior simulation (Section 3). A main benefit to the end-user is that no tuning of MCMC proposals is required, given that ARMS is an adaptive method. At the same time, our analyses showed that adaptive rejection Metropolis sampling within Gibbs sampling ("ARMGS" in earlier sections) is a potent strategy that out-performed regular Metropolis-Hastings samplers with ease (see Section 4).

Our case study on forecasting the portfolio Value at Risk (VaR) and Expected Shortfall (ES) demonstrated that Bayesian latent factor copulas generated substantially improved risk forecasts than frequentist ones, and also out-performed the benchmark regular vine copula in terms of forecast accuracy (Section 5). We expect that these findings will impact industry best practices as well as regulatory requirements to potentially increase resilience of world financial markets.

Methods for model selection of linking copulas as well as time-varying extensions of the latent factor copula model are the subject of ongoing research. For example, we could potentially expect that the dependence parameters of the linking copulas change over time. Furthermore, Bayesian extensions allowing for several latent factors and/or structured factor copulas as proposed in Krupskii and Joe (2015) are future research directions.

Author Contributions: All authors contributed equally.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. MCMC Sampling Methods

Appendix A.1. Details on MCMC Sampling Schemes

Mode and Curvature Matching (MCM)

This strategy uses $Beta(\alpha_t, \beta_t)$-distributed ($\alpha_t, \beta_t > 0$) proposal distributions for each latent factor $v_t \in (0,1)$, and a truncated normal distribution $N_{[a,b]}(\mu_j, \sigma_j^2)$ for the Fisher z parameters z_j of each linking copula.

The mode and curvature of the $Beta(\alpha_t, \beta_t)$ distribution with density $f(v_t; \alpha_t, \beta_t)$ are

$$\text{mode}_{v_t}(\alpha_t, \beta_t) = \frac{\alpha_t - 1}{\alpha_t + \beta_t - 2} \text{ for } \alpha_t, \beta_t > 1; \text{ and}$$

$$\text{curv}_{v_t}(v_t, \alpha_t, \beta_t) := \frac{\partial^2 f(v_t; \alpha_t, \beta_t)}{\partial v_t^2}$$

$$= \frac{1}{B(\alpha_t, \beta_t)} \left(v_t^{\alpha_t - 3}(1 - v_t)^{\beta_t - 3} \left(\alpha_t^2 + v_t^2(\alpha_t + \beta_t - 3)(\alpha_t + \beta_t - 2) \right. \right.$$

$$\left. \left. -2(\alpha_t - 1)v_t(\alpha_t + \beta_t - 3) - 3\alpha_t + 2) \right) \right).$$

The mode and curvature of the truncated normal $N_{[a,b]}(\mu_j, \sigma_j^2)$ proposals for z_j are

$$\text{mode}_{z_j}(\mu_j) = \mu_j \text{ for } a \leq \mu_j \leq b; \text{ and}$$

$$\text{curv}_{z_j}(z_j, \mu_j, \sigma_j) = \frac{e^{-\frac{(z_j - \mu_j)^2}{2\sigma_j^2}} (\mu_j^2 - \sigma_j^2 + z_j^2 - 2\mu_j z_j)}{\sqrt{2\pi}\,\sigma_j^5 \left(\Phi\left(\frac{b - z_j}{\sigma_j}\right) - \Phi\left(\frac{a - z_j}{\sigma_j}\right) \right)} \text{ for } z_j \in [a, b].$$

Numerically unstable behavior of the copula density functions at perfect dependence requires truncation of the allowable parameter range of the linking copulas. We limit the range of allowable Kendall's τ to $[0, 0.99]$, which translates to allowable Fisher z parameters $z_j \in [z(0), z(0.99)] =: [a, b]$.

The proposal parameters α_t, β_t, μ_j and σ_j^2 are chosen such that the mode and curvature of the proposal distributions agree with the ones of the respective full conditional distributions. The calculations are performed using numerical methods.

Expectation and Variance Matching (EVM)

This strategy is basically the same as the previous one, but it matches the proposals' and full conditionals' expectation and variances instead of their modes and curvatures. In another deviation from the previous approach, we use *Gamma*-distributed proposals for the Fisher z parameters here instead of truncated normal proposals used above.

The expectation and variance of the $Beta(\alpha_t, \beta_t)$ proposals for the latent factor v_t are

$$\mathbb{E}_{v_t}(\alpha_t, \beta_t) = \frac{\alpha_t}{\alpha_t + \beta_t}; \text{ and}$$

$$\mathbb{V}_{v_t}(\alpha_t, \beta_t) = \frac{\alpha_t \beta_t}{(\alpha_t + \beta_t)^2(\alpha_t + \beta_t + 1)}.$$

Given expectation $0 < \mathbb{E}_{v_t} < 1$ and variance $0 < \mathbb{V}_{v_t} < \infty$ of the full conditional distribution of v_t, this simple functional form enables analytic solution for the parameters α_t and β_t of the proposal distribution:

$$\alpha_t = \frac{-\mathbb{E}_{v_t}(\mathbb{E}_{v_t}^2 - \mathbb{E}_{v_t} + \mathbb{V}_{v_t})}{\mathbb{V}_{v_t}}; \text{ and}$$

$$\beta_t = \alpha_t \left(\frac{1}{\mathbb{E}_{v_t}} - 1 \right).$$

For non-negative Fisher z parameters z_j, $j = 1 : d$, of the linking copulas, we utilize $Gamma(s_j, r_j)$-distributed proposals, $s_j, r_j > 0$, with expectation and variance

$$\mathbb{E}_{z_j}(s_j, r_j) = \frac{s_j}{r_j}; \text{ and}$$

$$\mathbb{V}_{z_j}(s_j, r_j) = \frac{s_j}{r_j^2}.$$

Given expectation $0 < \mathbb{E}_{z_j} < \infty$ and variance $0 < \mathbb{V}_{z_j} < \infty$ of the full conditional distribution of z_j, the proposal parameters s_j and r_j follow as

$$s_j = \frac{\mathbb{E}_{z_j}^2}{\mathbb{V}_{z_j}}; \text{ and}$$

$$r_j = \frac{s_j}{\mathbb{E}_{z_j}}.$$

Again, the expectation and variance of the full conditional distributions are evaluated using a numerical scheme.

Independence and Random Walk Samplers (IRW)

A numerically less complex alternative to the previous methods, this strategy employs an independence sampler for the latent factors v_t, $t = 1 : T$, and a random walk sampler for the Fisher z parameters z_j, $j = 1 : d$.

Proposals for each latent factor v_t are drawn from a $Unif(0, 1)$ distribution, and are independent of the current state of the sampling chain.

The proposal for the $(r + 1)$-st iteration z_j^{r+1} of parameter z_j is drawn from a truncated normal distribution $N_{[a,b]}(\mu_j^{r+1}, \sigma_j^2)$ with random walk means $\mu_j^{r+1} = z_j^r$. The proposal variances σ_j^2 are chosen such that the acceptance rate is roughly 23%, which is the optimal acceptance rate for multivariate normal random walk proposals (Roberts et al. 1997).

Adaptive Rejection Metropolis within Gibbs Sampling (ARMGS)

The last sampling method is adaptive rejection Metropolis sampling (ARMS) from the full conditionals of each variable v_t, $t = 1 : T$, and z_j, $j = 1 : d$; this strategy is called adaptive rejection Metropolis sampling within Gibbs sampling (ARMGS; Gilks et al. (1995)). ARMS extends acceptance rejection sampling by a adaptive exponential envelope and a Metropolis-Hastings acceptance/rejection step. The acceptance/rejection step ensures proper sampling in case of a density that is not fully covered by the exponential envelope.

A major advantage of ARMS is that it adapts automatically to characteristics of the target distribution. For univariate log-concave target densities, ARMS creates an envelope for rejection sampling, while for other univariate target densities, ARMS uses acceptance/rejection sampling. In spite of being applicable to sample from general univariate target distributions, ARMS demonstrates

good performance in many situations and is particularly appealing in the context of Gibbs sampling with complex full conditional distributions.

Block ARMS is a multivariate extension to ARMS. Instead of sampling each v_t, $t = 1 : T$, and z_j, $j = 1 : d$, individually, some or all of them can be sampled jointly from their multivariate block full conditional distribution. As this strategy did not perform well in our initial analyses (Schamberger 2015, Chapter 8), we will not discuss it here.

Appendix A.2. Strategy to Find Starting Values for MCMC

All methods discussed above require starting values for all parameters v and z. Good starting values can improve convergence speed of these MCMC methods and can also be used as an initial values for maximum likelihood estimation.

We propose a two-step heuristic that first chooses starting values for the Fisher z parameters z of the linking copulas, and then chooses starting values for the latent factor v.

Starting Values for z:

In order to estimate z, one series $j \in \{1, \ldots, d\}$ is chosen as a proxy for latent factor v. We do this by selecting the 1-truncated c-vine copula that maximizes the sum of absolute Kendall's τ's of all linking copulas and choosing the root node variable j^* of this c-vine as the proxy for v. The starting values of the Fisher z parameters $z_{-j^*} = (z_j)_{j \in \{1:d\} \setminus \{j^*\}}$ are set to the Fisher z parameter of the pair copula linking the corresponding series $u_{:,j}$, $j \in \{1:d\} \setminus \{j^*\}$, with the root node variable $u_{:,j^*}$. For the Fisher z parameter z_{j^*} of the linking copula that connects the root node variable $u_{:,j^*}$ to the latent factor v, we use the highest absolute value of Fisher z parameters appearing in the c-vine as the starting value.

Starting Values for v:

Given starting values for the Fisher z parameters z, we set the latent factors $v = (v_1, \ldots, v_T)$ to the modes of their full conditional densities.

Formal Procedure:

We generate starting values for posterior sampling of z and v, given T observations of d-dimensional data $u = (u_{t,:})_{t=1:T}$ with uniform marginals as follows.

1. Calculate the empirical $d \times d$ Kendall's τ matrix $T = (\tau_{ij})_{i=1:d, j=1:d}$, where τ_{ij} denotes the empirical Kendall's τ between series $u_{:,i}$ and $u_{:,j}$.
2. Find the column j^* of T that maximizes the sum of absolute Kendall's τ's,

$$j^* = \arg\max_{j=1:d} \left\{ \sum_{i=1:d} |\tau_{ij}| \right\}$$

3. Set the starting values $z^0 = (z_1^0, \ldots, z_d^0)$ for the Fisher z parameters z of the linking copulas to

$$z_i^0 = z\left(\tau_{ij^*}\right) \text{ for } i \in \{1:d\} \setminus \{j^*\}$$

$$z_{j^*}^0 = z\left(\max_{i=\{1:d\}\setminus\{j^*\}} \tau_{ij^*}\right)$$

4. Set the starting values $v^0 = (v_t^0)_{t=1:T}$ of the latent factor v to

$$v_t^0 = \arg\max_{v \in (0,1)} \left\{ p(v \mid u_t, z^0) \right\},$$

where p is the full conditional density of v_t given by (12).

Appendix B. Detailed Simulation Results

Table A1. Mean absolute deviation (MAD), mean squared errors (MSE), effective sample size (ESS) per minute, and coverage of 90% and 95% credible intervals for Kendall's τ posterior sample. Results are averages over 100 independent replications of the analysis.

		Low τ					High τ					Mixed τ				
		ARMGS	**MCM**	**EVM**	**IRW**	**MLE**	**ARMGS**	**MCM**	**EVM**	**IRW**	**MLE**	**ARMGS**	**MCM**	**EVM**	**IRW**	**MLE**
τ_1	MAD	0.0829	0.0762	0.0898	0.0829	0.0656	0.0359	0.0352	0.0359	0.0357	0.0260	0.0431	0.0427	0.0432	0.0431	0.0321
	MSE	0.0203	0.0104	0.0269	0.0203	0.0099	0.0020	0.0020	0.0020	0.0020	0.0011	0.0029	0.0028	0.0029	0.0029	0.0016
	ESS/min	9.5	0.5	0.3	13.7	n/a	37.2	2.5	2.1	35.5	n/a	46.8	3.1	1.6	60.2	n/a
	90% C.I.	0.92	0.65	0.93	0.92	n/a	0.89	0.88	0.89	0.88	n/a	0.97	0.96	0.97	0.97	n/a
	95% C.I.	0.97	0.76	0.96	0.95	n/a	0.94	0.94	0.94	0.94	n/a	0.98	0.98	0.97	0.98	n/a
τ_2	MAD	0.0936	0.0832	0.0831	0.0916	0.0652	0.0334	0.0338	0.0335	0.0336	0.0259	0.0487	0.0486	0.0483	0.0492	0.0388
	MSE	0.0246	0.0116	0.0156	0.0229	0.0076	0.0017	0.0018	0.0017	0.0018	0.0010	0.0037	0.0037	0.0036	0.0038	0.0022
	ESS/min	7.9	0.6	0.2	11.6	n/a	33.2	2.2	1.8	31.5	n/a	37.6	2.9	1.8	46.7	n/a
	90% C.I.	0.91	0.58	0.92	0.89	n/a	0.83	0.81	0.84	0.81	n/a	0.85	0.85	0.87	0.82	n/a
	95% C.I.	0.96	0.67	0.96	0.93	n/a	0.92	0.88	0.92	0.92	n/a	0.92	0.91	0.92	0.92	n/a
τ_3	MAD	0.1070	0.0913	0.1144	0.0930	0.0811	0.0276	0.0279	0.0276	0.0277	0.0198	0.0406	0.0399	0.0407	0.0406	0.0290
	MSE	0.0298	0.0131	0.0354	0.0163	0.0137	0.0012	0.0012	0.0012	0.0012	0.0006	0.0026	0.0025	0.0026	0.0026	0.0013
	ESS/min	5.1	0.5	0.2	8.1	n/a	25.9	1.8	1.5	23.9	n/a	19.1	2.2	1.5	31.1	n/a
	90% C.I.	0.91	0.50	0.86	0.88	n/a	0.90	0.87	0.90	0.90	n/a	0.87	0.87	0.86	0.84	n/a
	95% C.I.	0.96	0.61	0.95	0.93	n/a	0.95	0.93	0.96	0.95	n/a	0.93	0.93	0.93	0.92	n/a
τ_4	MAD	0.1142	0.0983	0.1144	0.1231	0.0787	0.0252	0.0239	0.0254	0.0253	0.0172	0.0404	0.0401	0.0407	0.0401	0.0288
	MSE	0.0302	0.0145	0.0301	0.0369	0.0120	0.0010	0.0009	0.0010	0.0010	0.0005	0.0026	0.0025	0.0026	0.0026	0.0013
	ESS/min	4.7	0.6	0.2	6.8	n/a	15.7	1.2	1.0	14.4	n/a	2.8	0.8	0.4	6.8	n/a
	90% C.I.	0.88	0.42	0.89	0.83	n/a	0.86	0.86	0.86	0.87	n/a	0.84	0.74	0.86	0.82	n/a
	95% C.I.	0.94	0.49	0.91	0.89	n/a	0.93	0.90	0.92	0.93	n/a	0.89	0.79	0.89	0.86	n/a
τ_5	MAD	0.1463	0.1008	0.1507	0.1251	0.0787	0.0241	0.0249	0.0244	0.0245	0.0170	0.0817	0.0459	0.0855	0.0801	0.0406
	MSE	0.0524	0.0150	0.0557	0.0346	0.0094	0.0009	0.0010	0.0010	0.0010	0.0004	0.0099	0.0034	0.0106	0.0094	0.0023
	ESS/min	3.3	0.5	0.1	5.6	n/a	6.7	0.4	0.4	5.8	n/a	0.8	0.4	0.1	1.0	n/a
	90% C.I.	0.94	0.49	0.91	0.92	n/a	0.97	0.80	0.98	0.97	n/a	0.81	0.48	0.82	0.71	n/a
	95% C.I.	0.98	0.51	0.93	0.94	n/a	1.00	0.92	1.00	1.00	n/a	0.93	0.57	0.92	0.81	n/a

Appendix C. Sequential Learning of DLMs

Priors at time t:

The prior for the states $(\boldsymbol{\theta}_{jt}, \lambda_{jt})$ given time $t-1$ information set $\mathcal{D}_{t-1}$ is normal-gamma

$$(\boldsymbol{\theta}_{jt}, \lambda_{jt})|\mathcal{D}_{t-1} \sim NG(\boldsymbol{a}_{jt}, \boldsymbol{R}_{jt}, r_{jt}, c_{jt}) \tag{A1}$$

with parameters $\boldsymbol{a}_{jt} \in \mathbb{R}^3$, $\boldsymbol{R}_{jt} \in \mathbb{R}^{3 \times 3}$, $r_{jt}, c_{jt} > 0$. The normal-gamma distribution is

$$\boldsymbol{\theta}_{jt}|\lambda_{jt}, \mathcal{D}_{t-1} \sim N(\boldsymbol{a}_{jt}, \boldsymbol{R}_{jt}/(c_{jt}\lambda_{jt})) \tag{A2}$$

$$\lambda_{jt}|\mathcal{D}_{t-1} \sim G(r_{jt}/2, r_{jt}c_{jt}/2). \tag{A3}$$

Forecasts at time t:

The unconditional forecast distribution of y_{jt} given time $t-1$ information set $\mathcal{D}_{t-1}$ is obtained by integrating observation Equation (14) over the prior distribution of $\boldsymbol{\theta}_{jt}$ and λ_{jt}:

$$y_{jt}|\mathcal{D}_{t-1} \sim T_{r_{jt}}(f_{jt}, q_{jt}), \tag{A4}$$

with degrees of freedom r_{jt}, mean $f_{jt} = \boldsymbol{F}'_{jt}\boldsymbol{a}_{jt}$ and variance factor $q_{jt} = \boldsymbol{F}'_{jt}\boldsymbol{R}_{jt}\boldsymbol{F}_{jt} + c_{jt}$. Here $T_{r_{jt}}(f_{jt}, q_{jt})$ denotes a non-standard t distribution, which is a location-scale transformation of a t distribution with r_{jt} degrees of freedom, $T_{r_{jt}}$:

$$T_{r_{jt}}(f_{jt}, q_{jt}) = f_{jt} + \sqrt{q_{jt}}T_{r_{jt}}. \tag{A5}$$

Posteriors at time t:

Upon observation of y_{jt}, the posterior distribution of $(\boldsymbol{\theta}_{jt}, \lambda_{jt})$ given information set $\mathcal{D}_t := \{\mathcal{D}_{t-1}, y_{jt}\}$ is normal-gamma

$$(\boldsymbol{\theta}_{jt}, \lambda_{jt})|\mathcal{D}_t \sim NG(\boldsymbol{m}_{jt}, \boldsymbol{C}_{jt}, n_{jt}, s_{jt}) \tag{A6}$$

with parameters $\boldsymbol{m}_{jt} = \boldsymbol{a}_{jt} + \boldsymbol{A}_{jt}e_{jt}$, $\boldsymbol{C}_{jt} = (\boldsymbol{R}_{jt} - \boldsymbol{A}_{jt}\boldsymbol{A}'_{jt}q_{jt})z_{jt}$, $n_{jt} = r_{jt} + 1$, $s_{jt} = z_{jt}c_{jt}$; these can be calculated using the adaptive coefficient vector $\boldsymbol{A}_{jt} = \boldsymbol{R}_{jt}\boldsymbol{F}_{jt}/q_{jt}$, forecast error $e_{jt} = y_{jt} - f_{jt}$, and volatility update factor $z_{jt} = (r_{jt} + e_{jt}^2/q_{jt})/n_{jt}$.

Evolution to time $t+1$:

The step-ahead prior for $(\boldsymbol{\theta}_{j,t+1}, \lambda_{j,t+1})$ given time t information set $\mathcal{D}_t$ is normal-gamma

$$(\boldsymbol{\theta}_{j,t+1}, \lambda_{j,t+1})|\mathcal{D}_t \sim NG(\boldsymbol{a}_{j,t+1}, \boldsymbol{R}_{j,t+1}, r_{j,t+1}, c_{j,t+1}), \tag{A7}$$

where $\boldsymbol{a}_{j,t+1} = \boldsymbol{m}_{jt}$, $\boldsymbol{R}_{j,t+1} = \boldsymbol{C}_{jt}/\delta$, $r_{j,t+1} = \beta n_{jt}$, $c_{j,t+1} = s_{jt}$, and $\beta, \delta \in (0,1)$ are discount factors. This normal-gamma distribution is obtained via normal evolution of $\boldsymbol{\theta}_{jt}$ in system Equation (A8) and beta discount evolution of λ_{jt} in system Equation (A10).

The system equation for state vector $\boldsymbol{\theta}_{jt}$ is

$$\boldsymbol{\theta}_{j,t+1} = \boldsymbol{\theta}_{jt} + \boldsymbol{\omega}_{j,t+1}, \quad \boldsymbol{\omega}_{j,t+1} \sim N(\boldsymbol{0}, \boldsymbol{W}_{j,t+1}), \tag{A8}$$

where the innovation variances $\boldsymbol{W}_{j,t+1} = (1 - \delta)/\delta \boldsymbol{C}_{jt}$ are set to inflate the posterior variance $\boldsymbol{C}_{jt}$ by the reciprocal of discount factor δ:

$$\boldsymbol{R}_{j,t+1} = \boldsymbol{C}_{jt} + \boldsymbol{W}_{j,t+1} = \boldsymbol{C}_{jt} + \frac{1-\delta}{\delta}\boldsymbol{C}_{jt} = \frac{1}{\delta}\boldsymbol{C}_{jt}. \tag{A9}$$

The precisions λ_{jt} evolve through the application of Beta-shocks with discount factor β,

$$\lambda_{j,t+1} = \frac{\gamma_{j,t+1}}{\beta}\lambda_{jt}, \quad \gamma_{j,t+1} \sim Beta\left(\frac{\beta n_{jt}}{2}, \frac{(1-\beta)n_{jt}}{2}\right). \tag{A10}$$

References

Basel Committee on Banking Supervision. 2006. International Convergence of Capital Measurement and Capital Standards: A Revised Framework Comprehensive Version. Basel: Bank for International Settlements.

Basel Committee on Banking Supervision. 2011. Basel III: A global regulatory framework for more resilient banks and banking systems. Basel: Bank for International Settlements.

Basel Committee on Banking Supervision. 2013. Fundamental Review of the Trading Book: A Revised Market Risk Framework. Basel: Bank for International Settlements.

Burns, Arthur F., and Wesley C. Mitchell. 1946. *Measuring Business Cycles*. Cambridge: The National Bureau of Economic Research.

Christoffersen, Peter F. 2011. *Elements of Financial Risk Management*. New York: Academic Press.

Crouhy, Michel, Dan Galai, and Robert Mark. 2000. A comparative analysis of current credit risk models. *Journal of Banking & Finance* 24: 59–117.

Czado, Claudia, Florian Gärtner, and Aleksey Min. 2011. Analysis of Australian electricity loads using joint Bayesian inference of D-Vines with autoregressive margins. In *Dependence Modeling: Vine Copula Handbook*. Singapore: World Scientific Publishing, pp. 265–80.

Dißmann, Jeffrey, Eike Christian Brechmann, Claudia Czado, and Dorota Kurowicka. 2013. Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis* 59: 52–69.

Embrechts, Paul, Er Mcneil, and Daniel Straumann. 1999. Correlation: Pitfalls and Alternatives. Available online: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.638.3329 (accessed on 24 June 2014.)

Embrechts, Paul, Giovanni Puccetti, Ludger Rüschendorf, Ruodu Wang, and Antonela Beleraj. 2014. An academic response to Basel 3.5. *Risks* 2: 25–48.

European Parliament and Council. 2009. On the Taking-Up and Pursuit of the Business of Insurance and Reinsurance (Solvency II). *Official Journal of the European Union* 58: 10.

Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. 2000. The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics 82*: 540–54.

Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. 2012. The generalized dynamic factor model. *Journal of the American Statistical Association* 100: 830–40.

Frey, Rüdiger, and Alexander J. McNeil. 2003. Dependent defaults in models of portfolio credit risk. *Journal of Risk* 6: 59–92.

Geweke, John. 1978. *The Dynamic Factor Analysis of Economic Time Series Models*. Madison: University of Wisconsin.

Gelfand, Alan E., and Adrian F. M. Smith. 1990. Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association* 85: 398–409.

Gilks, Wally R., Nicola G. Best, and Keith K. C. Tan. 1995. Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics* 44: 455–73.

Gordy, Michael B. 2000. A comparative anatomy of credit risk models. *Journal of Banking & Finance* 24: 119–49.

Gruber, Lutz, and Claudia Czado. 2015. Sequential Bayesian Model Selection of Regular Vine Copulas. *Bayesian Analysis* 10: 937–63.

Gruber, Lutz, and Mike West. 2016. GPU-Accelerated Bayesian Learning and Forecasting in Simultaneous Graphical Dynamic Linear Models. *Bayesian Analysis* 11: 125–49.

Hastings, W. Keith. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57: 97–109.

Joe, Harry. 1997. *Multivariate Models and Multivariate Dependence Concepts*. New York: CRC Press.

Joe, Harry, and James Jianmeng Xu. 1996. *The Estimation Method of Inference Functions for Margins for Multivariate Models*. Technical Report No. 166. Vancouver: Department of Statistics, University of British Columbia.

Joe, Harry. 2005. Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis* 94: 401–19.

Krupskii, Pavel , and Harry Joe. 2013. Factor copula models for multivariate data. *Journal of Multivariate Analysis* 120: 85–101.

Krupskii, Pavel , and Harry Joe. 2015. Structured factor copula models: Theory, inference and computation. *Journal of Multivariate Analysis* 138: 53–73.

McNeil, Alexander J., Rüdiger Frey, and Paul Embrechts. 2005. *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton: Princeton University Press.

Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. 1953. Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics* 21: 1087–92.

Murray, Jared S., David B. Dunson, Lawrence Carin, and Joseph E. Lucas. 2013. Bayesian Gaussian Copula Factor Models for Mixed Data. *Journal of the American Statistical Association* 108: 656–65.

Oh, Dong Hwan, and Andrew J. Patton. 2017. Modelling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics* 35: 139–54.

Roberts, Gareth O., Andrew Gelman, and Walter R. Gilks. 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability* 7: 110–20.

Sargent, Thomas J., and Christopher A. Sims. 1977. Business cycle modeling without pretending to have too much a priori economic theory. *New Methods in Business Cycle Research* 1: 145–68.

Schepsmeier, Ulf, Jakob Stoeber, Eike Christian Brechmann, and Benedikt Graeler. 2014. *VineCopula: Statistical Inference of Vine Copulas*. R Package Version 1.3.

Schamberger, Benedikt. 2015. Bayesian Analysis of the One-Factor Copula Model with Applications to Finance. Master's dissertation, Technische Universität München, München, Germany. Available online: http://mediatum.ub.tum.de/doc/1253486/1253486.pdf (accessed on 4 January 2016).

Sklar, Abe. 1959. *Fonctions de Répartition À N Dimensions et Leurs Marges*. Paris: Publications de L'Institut de Statistique de l'Université de Paris, vol. 8, pp. 229–31.

Thiébaux, H. Jean, and Francis W. Zwiers. 1984. The interpretation and estimation of effective sample size. *Journal of Climate and Applied Meteorology* 23: 800–11.

West, Mike, and Jeff Harrison. 1997. *Bayesian Forecasting & Dynamic Models*, 2nd ed. New York: Springer Verlag.