



Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: [www.elsevier.com/locate/jeconom](http://www.elsevier.com/locate/jeconom)

# High dimensional dynamic stochastic copula models<sup>☆</sup>

Drew D. Creal<sup>\*</sup>, Ruey S. Tsay

University of Chicago, Booth School of Business, USA



## ARTICLE INFO

### Article history:

Available online 25 March 2015

### JEL classification:

C32

G32

### Keywords:

State space models  
Dynamic copulas  
Bayesian estimation  
Particle filters  
Credit default swaps

## ABSTRACT

We build a class of copula models that captures time-varying dependence across large panels of financial assets. Our models nest Gaussian, Student's  $t$ , grouped Student's  $t$ , and generalized hyperbolic copulas with time-varying correlations matrices, as special cases. We introduce time-variation into the densities by writing them as factor models with stochastic loadings. The proposed copula models have flexible dynamics and heavy tails yet remain tractable in high dimensions due to their factor structure. Our Bayesian estimation approach leverages a recent advance in sequential Monte Carlo methods known as particle Gibbs sampling which can draw large blocks of latent variables efficiently and in parallel. We use this framework to model an unbalanced, 200-dimensional panel consisting of credit default swaps and equities for 100 US corporations. Our analysis shows that the grouped Student's  $t$  stochastic copula is preferred over seven competing models.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

Copulas provide a general approach to measuring dependence among groups of random variables. They are an important tool in econometrics for pricing portfolios of assets and measuring their risk. A major issue in the recent development of copulas for financial applications is the ability of the parameters that measure dependence to change through time. This is especially relevant in the aftermath of the 2007–2008 financial crisis, when existing copula models were heavily criticized for failing to capture the actual risk in portfolios of credit default swaps and mortgages.

The main contribution of this paper is a class of copula models with time-varying dependence parameters that provide flexibility in low dimensions yet remain tractable in high dimensions. The class of conditional copula densities we consider includes Gaussian, Student's  $t$ , grouped Student's  $t$ , and generalized hyperbolic copulas, among others. We introduce time-variation into their dependence parameters by representing them as factor models with stochastic loadings. The proposed copula models have flexible

dynamics and heavy tails. Their factor structure introduces parsimony and provides simple expressions for the inverse and determinant of large matrices, which makes implementation feasible as the cross-section dimension increases. In our application, we model daily returns on credit default swaps and equities jointly for 100 major US corporations with a cross-sectional dimension of 200.

Sklar's theorem states that the joint distribution of a group of random variables can be decomposed into their univariate marginals and the copula function that couples the marginals together; see Sklar (1959), and for surveys of copulas McNeil et al. (2005) in risk management and Patton (2009) in econometrics. Copulas are distribution functions over the unit hypercube with uniform marginals. A realization from a copula is an  $n \times 1$  vector of observed variables  $u_t = (u_{1t}, \dots, u_{nt})$  with  $0 \leq u_{it} \leq 1$ . We are interested in the class of copulas that can be represented as a non-linear, non-Gaussian state space model

$$u_t \sim p(u_t | \Lambda_t, X_t, \theta), \quad t = 1, \dots, T \quad (1)$$

$$\Lambda_{t+1} = \mu + \Phi_\lambda (\Lambda_t - \mu) + \eta_t, \quad \eta_t \sim N(0, \Sigma) \quad (2)$$

where  $\Lambda_t$  is an unobserved vector of state variables,  $X_t = (X_{1t}, \dots, X_{nt})$  are observable covariates, and  $\theta$  contains all the unknown parameters of the model including any parameters governing the dynamics (2). In our models, the dependence structure of the conditional copula changes through time as a function of the state variables  $\Lambda_t$ . In the terminology of state space models (see Cappé et al. (2005), Durbin and Koopman (2012)), the conditional copula density (1) is the observation density and the transition density  $p(\Lambda_{t+1} | \Lambda_t, \theta)$  of the state variables  $\Lambda_t$  in (2) is linear

<sup>☆</sup> The authors thank Denis Pelletier, Jing Cynthia Wu, and seminar participants at CIREQ, Chicago Booth marketing for helpful comments. Both authors gratefully acknowledge financial support from the Fama Miller Center for Research in Finance and Chicago Booth. Drew Creal gratefully acknowledges additional financial support from the William Ladany Faculty Scholar Fund at the University of Chicago Booth School of Business.

<sup>\*</sup> Correspondence to: The University of Chicago Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA.

E-mail address: [dcreal@chicagobooth.edu](mailto:dcreal@chicagobooth.edu) (D.D. Creal).

and Gaussian. We discuss further details of the model including identifying restrictions below.

Inference in this class of models is challenging for two reasons. First, for many empirically relevant observation densities  $p(u_t | \Lambda_t, X_t, \theta)$ , the computations become infeasible as the cross-section dimension  $n$  becomes large because estimation requires repeatedly inverting a large matrix to evaluate  $p(u_t | \Lambda_t, X_t, \theta)$ . Second, the likelihood of the model is a high-dimensional integral over the path of the latent state variables

$$p(u_{1:T} | X_{1:T}, \theta) = \int p(u_{1:T} | \Lambda_{1:T}, X_{1:T}, \theta) p(\Lambda_{1:T} | \theta) d\Lambda_{1:T} \quad (3)$$

where  $u_{1:T} = (u_1, \dots, u_T)$ ,  $\Lambda_{1:T} = (\Lambda_1, \dots, \Lambda_T)$ , and  $X_{1:T} = (X_1, \dots, X_T)$ . For stochastic copula models, this integral has no closed-form solution.

Our solution to these problems has two parts. First, we focus on a class of densities  $p(u_t | \Lambda_t, X_t, \theta)$  induced from a factor copula model. The factor structure reduces the number of parameters to estimate and provides simple expressions for the inverse and determinant of relevant matrices. This makes computations feasible in high-dimensions. Second, we develop Bayesian procedures that leverage recent developments in the literature on Monte Carlo methods called particle Markov chain Monte Carlo (PMCMC); see Andrieu et al. (2010). Specifically, we use a particle Gibbs sampler to draw from the joint posterior distribution  $p(\theta, \Lambda_{1:T} | u_{1:T}, X_{1:T})$  by iterating between the full conditional distributions  $p(\theta | u_{1:T}, X_{1:T}, \Lambda_{1:T})$  and  $p(\Lambda_{1:T} | u_{1:T}, X_{1:T}, \theta)$ . The particle Gibbs sampler draws the latent state variables in large blocks improving the mixing of the algorithm. Due to the factor structure, the draws can be performed in parallel either in the cross-section, time dimension, or both. To the best of our knowledge, this is the first application of the particle Gibbs sampler in econometrics.

As is typical in financial applications using copulas, we use a two-step procedure to estimate a full joint distribution from the observed data  $y_t = (y_{1t}, \dots, y_{nt})'$  for  $t = 1, \dots, T$ . First, we specify models for the marginal distributions  $F(y_{it} | y_{i,1}, \dots, y_{i,t-1}, \psi_i)$  for  $i = 1, \dots, n$ , where  $\psi_i$  denotes the parameters of the  $i$ th marginal. From the marginals, we calculate the probability integral transforms  $u_{it} = F(Y_{it} \leq y_{it} | y_{i,1}, \dots, y_{i,t-1}, \psi_i)$  for  $i = 1, \dots, n$  and  $t = 1, \dots, T$  and, in the second step, take these  $u_t = (u_{1t}, \dots, u_{nt})$  as data to estimate the copula. This two step procedure is admittedly not fully Bayesian but it simplifies the estimation problem considerably once the cross-section dimension  $n$  is moderately large.<sup>1</sup> Importantly, the marginal distributions of the copula may be different than the marginal distributions  $F(y_{it} | y_{i,1}, \dots, y_{i,t-1}, \psi_i)$  of the data  $y_{it}$ . In our work, we model the marginal distributions with univariate stochastic volatility models that have leverage, heavy tails, and asymmetry. Although the CDF's of stochastic volatility models are not known in closed-form, we show how to calculate the probability integral transforms using the particle filter.

This paper is related to three different parts of the literature on copulas: the building of copulas with time-varying dependence parameters, the estimation of factor copulas, and Bayesian inference for copulas. Our paper makes contributions to each of these literatures as well as a growing literature on the modeling of credit default swaps (CDS).

The econometric literature on copulas with time-varying dependence parameters has grown since they were first proposed by Patton (2006). Time-varying copulas have been built from Markov-switching models (Pelletier, 2006), observation-driven models (Creal et al., 2011, 2013), and parameter-driven models (Hafner and Manner, 2012); see Manner and Reznikova (2012) for a survey. The papers closest to our paper are Hafner and Manner (2012) and Oh and Patton (2013). Hafner and Manner (2012) is the first and only paper to estimate a state space copula model like (1) and (2). Our work extends theirs beyond the setting of univariate state variables to significantly higher dimensions. Oh and Patton (2013) use the generalized autoregressive score (GAS) framework of Creal et al. (2013) to drive the factor loadings of conditional factor copulas through time. Their paper therefore provides an observation-driven alternative to the models discussed here. As GAS models are observation-driven, their copulas do not require integrating over the path of the latent variables  $p(\Lambda_{1:T} | \theta)$  as in (3) in order to calculate the likelihood.

Factor models are popular in econometrics and statistics for modeling high dimensional data based on a lower-dimensional, parsimonious structure. The single factor, time-invariant Gaussian copula of Li (2000) was the industry standard for modeling credit defaults prior to the sub-prime financial crisis. Recently, Oh and Patton (2012) and Krupskii and Joe (2013) have proposed more general copulas built from factor models with Oh and Patton (2013) estimating time-varying versions of this model. Murray et al. (2013) provide Bayesian methods for Gaussian factor copulas with constant factor loadings.

Bayesian estimation of copulas has drawn considerable interest with work on Gaussian copula regression models (Pitt et al., 2006), skewed multivariate distributions (Smith et al., 2012), and vine copulas (Min and Czado, 2010), among others. Smith (2011) provides a survey of the Bayesian literature on copulas. Our paper extends the Bayesian literature to dynamic copulas using state space models such as (1) and (2). We note that our estimation procedures based on the Particle Gibbs sampler can be applied to other popular copulas such as Clayton and Gumbel copulas.

There also exists a growing literature on the modeling of portfolios of CDS in high dimensions. Creal et al. (2014) build a rating system for thousands of firms based on databases of CDS, Lucas et al. (2014) measure European sovereign default risk, Oh and Patton (2013) provide systemic risk measures for the US firms, and Christoffersen et al. (2013) estimate dynamic conditional copula models for panels of CDS, intensities, and equities.

This paper continues as follows. In Section 2, we discuss classes of observation densities  $p(u_t | \Lambda_t, X_t, \theta)$  that lead to flexible yet tractable copula models. In Section 3, we design MCMC algorithms for estimation of factor copulas with and without time-varying factor loadings. Section 4 includes an empirical application to an unbalanced, 200 dimensional panel of credit default swaps and equities. Section 5 concludes.

## 2. Copula-based state space models

### 2.1. Stochastic factor copula models

A flexible class of factor copula models can be expressed as

$$u_{it} = P(x_{it} | \theta), \quad i = 1, \dots, n, \quad t = 1, \dots, T \quad (4)$$

$$x_{it} = \tilde{\lambda}'_{it} z_{1t} + (\beta_i \odot \tilde{\lambda}_{it})' z_{2t} + \sigma_{it} \varepsilon_{it},$$

$$z_t \sim p(z_t | \theta), \quad \varepsilon_{it} \sim p(\varepsilon_{it} | \theta), \quad (5)$$

where  $\tilde{\lambda}_{it}$  is a  $p \times 1$  vector of (scaled) factor loadings,  $\tilde{\lambda}_{it}$  is a  $k \times 1$  vector of (scaled) observable covariates,  $z_t = (z'_{1t}, z'_{2t})'$  is a  $(p+k) \times 1$  vector of common latent factors with mean zero

<sup>1</sup> Most papers in econometrics estimating copulas use a two step procedure. A fully Bayesian procedure requires estimating the copula and marginals simultaneously in one global MCMC algorithm because the parameters of the marginals  $\psi_i$  enter the copula. Implementing a global MCMC algorithm only impacts how the parameters  $\psi_i$  of the marginals are drawn and not how the parameters and state variables in the copula (1) and (2) are drawn, which is our primary interest.

and  $\text{Cov}(z_t) = I_{p+k}$ ,  $\varepsilon_{it}$  are i.i.d. shocks with mean zero and variance one, and  $P(x_{it}|\theta)$  is the marginal distribution function of  $x_{it}$ . We use  $\odot$  to denote the Hadamard product. The common factor and idiosyncratic shocks are independent of one another for  $t = 1, \dots, T$  and  $i = 1, \dots, n$ . The model (4)–(5) is a special case of the observation density in (1).

The state vector  $\Lambda_t$  in (2) determines the evolution of the factor loadings  $\tilde{\lambda}_{it}$  and scale parameter  $\sigma_{it}^2$ . These quantities are related to  $\Lambda_t$  as

$$\Lambda_t = \text{vec}(\lambda_t) \quad \lambda'_t = (\lambda'_{1t}, \dots, \lambda'_{nt})'$$

$$\tilde{\lambda}_{it} = \frac{\lambda_{it}}{\sqrt{1 + \lambda'_{it}\lambda_{it} + (\beta_i \odot X_{it})'(\beta_i \odot X_{it})}}, \quad (6)$$

$$\sigma_{it}^2 = \frac{1}{1 + \lambda'_{it}\lambda_{it} + (\beta_i \odot X_{it})'(\beta_i \odot X_{it})}, \quad (7)$$

where we define the rescaled covariate in (5) as  $\tilde{X}_{it} = X_{it}/\sqrt{1 + \lambda'_{it}\lambda_{it} + (\beta_i \odot X_{it})'(\beta_i \odot X_{it})}$ . The matrix  $\lambda_t$  has dimension  $n \times p$ . The rescaling of  $\lambda_{it}$  and  $X_{it}$  ensures that the marginal transformation  $P(x_{it}|\theta)$  does not depend on the parameters  $\beta = (\beta'_1, \dots, \beta'_n)'$  or the parameters governing the transition density  $p(\Lambda_{t+1}|\Lambda_t; \theta)$ . This will be important for inference in higher dimensions.

For identification, we assume that the top  $p$  rows of  $\lambda'_t$  are lower triangular with positive entries on the diagonal. To satisfy the positivity restriction, we specify the dynamics of these  $p$  diagonal elements in logarithms, while all other remaining elements in  $\Lambda_t$  have no sign restrictions. Identification also requires either no constant terms in  $X_{it}$  or setting  $\mu = 0$  in (2). In Section 2.2.5, we discuss extensions of the model that allow the number of shocks driving  $\Lambda_t$  to be smaller than  $np$ .

The densities of the idiosyncratic shocks  $p(\varepsilon_{it}|\theta)$  and the common factor  $p(z_t|\theta)$  completely determine the conditional copula density  $p(u_t|\Lambda_t, X_t, \theta)$ .<sup>2</sup> The conditional observation density of  $u_t$  is directly related to the density of  $x_t = (x_{1t}, \dots, x_{nt})$  by the change-of-variables

$$p(u_t|\Lambda_t, X_t, \theta) = \frac{1}{\prod_{i=1}^n p(x_{it}|\theta)} p(x_t|\Lambda_t, X_t, \theta),$$

$$x_{it} = P^{-1}(u_{it}|\theta). \quad (8)$$

The first term is the Jacobian of the transformation from  $x_t$  to  $u_t$ ,  $P^{-1}(u_{it}|\theta)$  are the inverse of the marginal distributions, and  $p(x_{it}|\theta)$  are the marginal densities. From (5), the joint density of  $x_t$  is found by integrating out the common factor

$$p(x_t|\Lambda_t, X_t, \theta) = \int p(x_t|\Lambda_t, X_t, z_t, \theta) p(z_t|\theta) dz_t. \quad (9)$$

The marginal density and distribution function of  $x_{it}$  are

$$p(x_{it}|\theta) = \int p(x_{it}|z_t, \lambda_{it}, X_{it}, \theta) p(z_t|\theta) dz_t, \quad (10)$$

$$P(x_{it}|\theta) = \int P(x_{it}|z_t, \lambda_{it}, X_{it}, \theta) p(z_t|\theta) dz_t. \quad (11)$$

These are not functions of either  $\lambda_{it}$  or  $X_{it}$  due to the rescaling in (7). Combining (9)–(11) gives the conditional copula density in (8).

In this paper, we choose distributions for  $z_t$  and  $\varepsilon_{it}$  that lead to models where the marginal distributions of the copula  $P(x_{it}|\theta)$  are

known in closed form and there exist routines to compute their inverse. Knowledge of the marginals is important for practical reasons. The inverse  $x_{it} = P^{-1}(u_{it}|\theta)$  must be recalculated during the MCMC algorithm every time a parameter within the marginals changes. In some special cases (e.g. the Gaussian copula), the transformation does not depend on any parameters and we can calculate  $x_{it} = P^{-1}(u_{it}|\theta) = \Phi^{-1}(u_{it})$  once and take  $x_{it}$  as our data. In other cases (e.g. the Student's  $t$  copula), the marginal distributions depend on  $\theta$  and the values of  $x_{it}$  change as  $\theta$  changes. This is why rescaling of  $\lambda_{it}$  in (7) is important. Changes in the parameters  $\beta = (\beta'_1, \dots, \beta'_n)'$  or those in the transition density  $p(\Lambda_{t+1}|\Lambda_t, \theta)$  impact the copula but they do not enter the marginal distributions.

## 2.2. Special cases within the proposed family

In this section, we describe a family of conditional copulas  $p(u_t|\Lambda_t, X_t; \theta)$  where the marginal densities remain known. The family of models considered includes, among others, the Gaussian, Student's  $t$ , and skewed Student's  $t$  copulas. These copulas are characterized by a conditional correlation matrix  $R_t$  given by

$$R_t = \tilde{C}'_t \tilde{C}_t + D_t, \quad (12)$$

where  $\tilde{C}'_t = \left[ \tilde{\lambda}'_t (\beta \odot \tilde{X}_t)' \right]$  is a  $n \times (p+k)$  matrix composed of the  $n \times p$  matrix of scaled loadings  $\tilde{\lambda}'_t = (\tilde{\lambda}'_{1t}, \dots, \tilde{\lambda}'_{nt})$  and the  $n \times k$  matrix of covariates  $(\beta \odot \tilde{X}_t)'$ .  $D_t$  is a  $n \times n$  diagonal matrix with entries  $\sigma_{it}^2$ . Many of these copulas and their properties are described by McNeil et al. (2005) for the time-invariant case  $R_t = R$  and when there is no factor structure. We follow McNeil et al. (2005) and define these copulas using  $R_t$  as the correlation matrix.

This class of copula models has two major advantages in high dimensions. First, from a statistical perspective, the factor structure dramatically reduces the number of parameters. The conditional correlation matrix  $R_t$  has  $n(n-1)/2$  elements while the factor model can have significantly fewer depending on  $p$  and  $k$ . The second major advantage for high dimensional inference is computational. The inverse and determinant of  $R_t$  are simple expressions

$$R_t^{-1} = D_t^{-1} - D_t^{-1} \tilde{C}_t \left( I_{p+k} + \tilde{C}'_t D_t^{-1} \tilde{C}_t \right)^{-1} \tilde{C}'_t D_t^{-1},$$

$$|R_t| = \left| I_{p+k} + \tilde{C}'_t D_t^{-1} \tilde{C}_t \right| |D_t|.$$

Quadratic forms  $x'_t R_t^{-1} x_t$  that enter the densities  $p(u_t|\Lambda_t, X_t, \theta)$  can be calculated efficiently, which is critical as the cross-section dimension  $n$  gets large.

### 2.2.1. Conditionally Gaussian factor copula

The conditionally Gaussian factor copula with correlation matrix  $R_t$  can be written as

$$u_{it} = \Phi(x_{it}), \quad i = 1, \dots, n, \quad t = 1, \dots, T$$

$$x_{it} = \tilde{\lambda}'_{it} z_{1t} + (\beta_i \odot \tilde{X}_{it})' z_{2t} + \sigma_{it} \varepsilon_{it},$$

$$z_t \sim N(0, I_{p+k}), \quad \varepsilon_{it} \sim N(0, 1).$$

The inverses  $x_{it} = \Phi^{-1}(u_{it})$  only need to be evaluated once and  $x_{it}$  can be taken as the observed data. Murray et al. (2013) consider Bayesian inference for this model when the factor loadings are constant  $\lambda_{it} = \lambda_i$  and  $\beta_i = 0$ .

<sup>2</sup> Although the copula is defined by continuous r.v.'s  $z_t$  and  $\varepsilon_{it}$ , the marginal distributions of the data  $y_{it} = F^{-1}(u_{it}|y_{i,1:t-1}, \psi_i)$  may be discrete.

### 2.2.2. Conditionally grouped Student's $t$ factor copula

Conditional on  $\Lambda_t$ , a factor copula having a grouped Student's  $t$  density can be written as

$$u_{it} = T(x_{it} | v_j), \quad i = 1, \dots, n, \quad t = 1, \dots, T$$

$$x_{it} = \sqrt{\zeta_{t,j}^{-1}} \left[ \tilde{\lambda}_{it}' \tilde{z}_{1t} + (\beta_i \odot \tilde{X}_{it})' \tilde{z}_{2t} + \sigma_{it} \tilde{\varepsilon}_{it} \right],$$

$$\tilde{z}_t \sim N(0, I_{p+k}), \quad \tilde{\varepsilon}_{it} \sim N(0, 1),$$

$$\zeta_{t,j} \sim \text{Inv-Gamma}\left(\frac{v_j}{2}, \frac{v_j}{2}\right), \quad j = 1, \dots, G.$$

Each series  $i$  belongs to one of  $G$  groups for  $j = 1, \dots, G$  and for  $t = 1, \dots, T$ .<sup>3</sup> Observations within the same group  $j$  share the same mixing variable  $\zeta_{t,j}$  and degrees of freedom parameter  $v_j$ . We use the notation  $\zeta_{t,j}^i$  to indicate that series  $i$  belongs to group  $j$ . For a given value of  $R_t$ , series in the same group exhibit joint extreme tail dependence because they share a common mixing variable. When there is only one group, the model reduces to the standard Student's  $t$  copula. Despite different degrees of freedom parameters, the marginal distributions of  $x_{it}$  remain Student's  $t$  distributions meaning that the CDF and its inverse are standard. However, any change in the degrees of freedom  $v_j$  requires inverting the CDF's to recalculate  $x_{it} = T^{-1}(x_{it} | v_j)$ .

A slightly different definition of the grouped  $t$  copula with constant correlation matrix  $R_t = R$  was proposed by Daul et al. (2003) and its properties were studied by Demarta and McNeil (2005). In their model, the inverse gamma mixing variables are generated from a common uniform random variable at each date.<sup>4</sup> This slightly alters the properties of their copula, as the common uniform variable makes the copula co-monotonic.

### 2.2.3. Conditionally skewed Student's $t$ factor copula

Skewness can be introduced into the marginal distributions of the copula through the skewed Student's  $t$  factor model. The model has the structure

$$u_{it} = \text{skew-T}(x_{it} | v, \gamma_i), \quad i = 1, \dots, n, \quad t = 1, \dots, T$$

$$x_{it} = \gamma_i \zeta_t + \sqrt{\zeta_t} \left[ \tilde{\lambda}_{it}' \tilde{z}_{1t} + (\beta_i \odot \tilde{X}_{it})' \tilde{z}_{2t} + \sigma_{it} \tilde{\varepsilon}_{it} \right],$$

$$\tilde{z}_t \sim N(0, I_{p+k}), \quad \tilde{\varepsilon}_{it} \sim N(0, 1),$$

$$\zeta_t \sim \text{Inv-Gamma}\left(\frac{v}{2}, \frac{v}{2}\right),$$

where  $\gamma_i$  are skewness parameters and all  $n$  series share a common mixing variable  $\zeta_t$ . A positive (negative) value of  $\gamma_i$  determines whether the marginal is positively (negatively) skewed. The marginal distributions of the skewed Student's  $t$  copula are all univariate-skewed Student's  $t$  distributions. A “grouped” skewed Student's  $t$  copula can be developed to add flexibility in modeling the tails of the copula. Although the marginal densities are known in closed-form, the distribution function is not and will have to be calculated numerically. This limits the model (at present) to lower dimensions. Lucas et al. (2014) use an observation-driven version of this model to estimate co-dependence between CDS spreads during the recent Euro sovereign debt crisis.

### 2.2.4. Additional copulas

The skewed- $t$  model can be extended to the larger family of conditionally generalized hyperbolic (GH) copulas by choosing the distribution of the mixing variable  $\zeta_t$  to be a generalized inverse Gaussian (GIG) distribution, see Chapter 3 of McNeil et al. (2005). The marginal distributions of  $x_{it}$  are GH distributions. The GH model adds additional parameters but it does not exhibit joint extreme tail dependence (other than the skewed  $t$  sub-class).

### 2.2.5. Other dynamics for the state variables

In some applications, the cross-sectional dimension  $n$  may be large making it valuable to build models where the  $np \times 1$  state vector  $\Lambda_t$  is driven by only  $r$  shocks with  $r \ll np$ . The simplest approach decomposes the covariance matrix in (2) as  $\Sigma = Q\Upsilon Q'$  with  $Q$  an  $np \times r$  matrix and  $\Upsilon$  an  $r \times r$  matrix. This effectively makes the shock a singular, Gaussian random variable. Alternatively, we could define  $\Lambda_t$  as a  $r \times 1$  vector and set  $\text{vec}(\lambda_t) = A + B\Lambda_t$  where  $A$  is a  $np \times 1$  vector and  $B$  is a  $np \times r$  matrix of parameters. Restrictions may be placed on  $\mu$ ,  $\Phi_\lambda$ , and  $\Sigma$  to reduce the number of parameters.

The factor loadings  $\lambda_t$  may follow other types of stochastic processes and the basic ideas discussed above do not change. For example, the factor loadings  $\lambda_t$  may exhibit one-time structural breaks, regime-switching (Pelletier, 2006), or dynamic equicorrelations (Engle and Kelly (2012)). Parameter-driven versions of the dynamic equicorrelation models provide a good benchmark model for comparison to factor copulas. They are particularly easy to estimate because they have only a few parameters and state variables. As an additional contribution, we describe dynamic models for one and two-block equicorrelation matrices in the online appendix that, to the best of our knowledge, are new to the literature.

### 2.2.6. Discussion

Oh and Patton (2012), Oh and Patton (2013) and Krupskii and Joe (2013) consider factor copula models where the distributions of  $z_t$  and  $\varepsilon_{it}$  in (5) may be non-Gaussian. When the distribution of the common factor  $z_t$  is heavy-tailed (e.g. a Student's  $t$  distribution), these models have joint, positive extreme tail dependence and are slightly more general than the models that we estimate here. However, for arbitrary choices of  $z_t$  and  $\varepsilon_{it}$ , the marginal distributions  $P(x_{it} | \theta)$  are not known in closed-form and consequently neither are the observation densities  $p(u_t | \Lambda_t, X_t, \theta)$  even if the factor loadings are constant  $\Lambda_t = \Lambda$ .

When the factor loadings are stochastic and the observation density  $p(u_t | \Lambda_t, X_t, \theta)$  is not known in closed form because either or both of the marginals in (10) and (11) contain an unknown integral, the likelihood function (3) has an integral inside another integral. Estimation becomes computationally more demanding. One option is to calculate the integrals for the marginals (10) and (11) by numerical integration assuming that any error introduced at this stage is negligible, especially relative to the Monte Carlo error of integrating over the path of the state variables.

A second option is to calculate each of the marginal distributions by Monte Carlo inside the MCMC algorithm. Algorithms that use Monte Carlo methods (e.g. importance sampling) inside the acceptance ratio of a Metropolis–Hastings algorithm have recently drawn considerable attention in the theoretical literature on MCMC; see, Andrieu and Roberts (2009), Andrieu et al. (2010), and Flury and Shephard (2011). The main point of this literature is that the resulting Markov chain targets the correct stationary distribution as long as the estimates of any unknown integrals within the MH acceptance ratio are unbiased. This approach to estimating factor copulas when the marginal distributions are unknown is currently feasible only when the cross-section dimension  $n$  is small. We expect that future research will make this feasible through “massively” parallel computing on graphics processing units (GPUs) because each value of  $x_{it} = P^{-1}(u_{it} | \theta)$  can typically be calculated independently for all  $i, t$ .

<sup>3</sup> It is possible for each series  $i$  to change its group membership over time. We leave this extension for future research.

<sup>4</sup> To generate the mixing variables in their framework, draw  $U_t \sim U(0, 1)$  and invert the inverse gamma CDF as  $\zeta_{t,j} = \text{Inv-Gamma}^{-1}(U_t | \frac{v_j}{2}, \frac{v_j}{2})$  for  $j = 1, \dots, G$  and  $t = 1, \dots, T$ . The posterior distribution is then defined over the uniform variables  $U_t$ . In the MCMC algorithm of Section 3.1, draws are taken of  $U_t$  instead of  $\zeta_{t,j}$ .



### 3. Bayesian estimation

#### 3.1. MCMC algorithm

##### 3.1.1. Broad principles

Given a prior  $p(\theta)$ , an MCMC algorithm with data augmentation draws from the posterior

$$p(\theta, z_{1:T}, \Lambda_{1:T}, \zeta_{1:T} | u_{1:T}, X_{1:T})$$

$$\propto p(u_{1:T} | X_{1:T}, z_{1:T}, \zeta_{1:T}, \Lambda_{1:T}, \theta) p(z_{1:T}, \Lambda_{1:T}, \zeta_{1:T} | \theta) p(\theta).$$

A standard Gibbs sampler that draws from this distribution iterates for  $j = 1, \dots, N$  as

1. Draw  $z_{1:T}^{(j)} \sim p(z_{1:T} | u_{1:T}, X_{1:T}, \Lambda_{1:T}^{(j-1)}, \zeta_{1:T}^{(j-1)}, \theta^{(j-1)})$ .
2. Draw  $\Lambda_{1:T}^{(j)} \sim p(\Lambda_{1:T} | u_{1:T}, X_{1:T}, z_{1:T}^{(j)}, \zeta_{1:T}^{(j-1)}, \theta^{(j-1)})$ .
3. Draw  $\zeta_{1:T}^{(j)} \sim p(\zeta_{1:T} | \Lambda_{1:T}, X_{1:T}, z_{1:T}^{(j)}, \Lambda_{1:T}^{(j)}, \theta^{(j-1)})$ .
4. Draw  $\theta^{(j)} \sim p(\theta | u_{1:T}, X_{1:T}, z_{1:T}^{(j)}, \zeta_{1:T}^{(j)}, \Lambda_{1:T}^{(j)})$ .

In our implementation, parts of the Gibbs sampler are by-now standard. We focus here on the non-standard part of the algorithm in Step 2 that is new to the paper. Details for the remaining steps are in Appendix A.1 Priors for the model are discussed in Section 4.2.

Two principles that improve the performance of MCMC algorithms are: (i) to condition on as few parameters as possible when drawing from the full conditional distributions; (ii) to draw the parameters or state variables in large blocks. These principles improve the mixing of the Markov chain leading to faster convergence. We describe how to apply these principles for factor copula models and we highlight some computational trade offs that exist.

In the models of Section 2.2, the common factor  $z_t$  and the mixing variables  $\zeta_t$  can typically be integrated out of all the full conditionals. This helps satisfy condition (i) above because we can draw parameters without conditioning on  $z_t$  or  $\zeta_t$ . Nevertheless, conditioning on  $z_t$  can have benefits in high dimensions. The  $x_{it}$  are independent conditional on  $z_t$  whereas they are not if  $z_t$  is marginalized out. By conditioning on  $z_t$ , most steps of the MCMC algorithm can be performed in parallel either in the cross-section, time dimension, or both. For example, under an assumption that  $\Phi_\lambda$  and  $\Sigma$  are diagonal, the factor loadings  $\lambda_{i,1:T} = (\lambda_{i,1}, \dots, \lambda_{i,T})$  can be sampled in blocks independently of one another. Although this requires conditioning on  $z_t$ , it has little impact on the mixing of the MCMC algorithm when the cross-section dimension  $n$  is large and  $p$  is small. This is because uncertainty over  $z_t$  decreases rapidly as  $n$  becomes large for a fixed value of  $p$ . Intuitively, for each date  $t$ , a central limit theorem applies in the cross-section as  $n \rightarrow \infty$ . We recommend conditioning on  $z_{1:T}$  in high dimensions and marginalizing over it in lower dimensions.

##### 3.1.2. Drawing the factor loadings in blocks

An important ingredient of our MCMC algorithm is how we draw the factor loadings  $\Lambda_{1:T}$  from their full conditional distribution  $p(\Lambda_{1:T} | u_{1:T}, X_{1:T}, z_{1:T}, \zeta_{1:T}, \theta)$ . We use a recent advance in sequential Monte Carlo methods known as the particle Gibbs (PG) sampler, see Andrieu et al. (2010). The algorithm allows us to draw paths of the state variables  $\lambda_{i,1:T} = (\lambda_{i,1}, \dots, \lambda_{i,T})$  for  $i = 1, \dots, np$  in large blocks. This satisfies condition (ii) above. Particle filters are simulation based algorithms that sequentially approximate continuous, marginal distributions by discrete distributions through a set of stochastic support points called “particles” and probability masses; see Creal (2012) for a survey.

In our work, we draw from the full conditional distribution  $p(\Lambda_{1:T} | u_{1:T}, X_{1:T}, z_{1:T}, \zeta_{1:T}, \theta)$  by drawing each path  $\lambda_{i,1:T} =$

$(\lambda_{i,1}, \dots, \lambda_{i,T})$  separately for  $i = 1, \dots, np$ . Under the assumption that  $\Phi_\lambda$  and  $\Sigma$  are diagonal, path  $\lambda_{i,1:T}$  does not depend on path  $\lambda_{k,1:T}$  once we condition on  $z_{1:T}$ . This means that we can draw paths at the same time using parallel computing. At each date  $t$ , the marginal filtering distribution of the state variable  $\lambda_{i,t}$  is  $p(\lambda_{i,t} | u_{1:t}, X_{1:t}, z_{1:t}, \zeta_{1:t}, \theta)$ . The particle filter approximates

each marginal through a collection of  $M$  particles  $\{\lambda_{i,t}^{(m)}, w_t^{(m)}\}_{m=1}^M$  where  $\lambda_{i,t}^{(m)}$  is a point on the support of the distribution and  $w_t^{(m)}$  is the probability mass at that point. Collecting the particles for all dates  $t = 1, \dots, T$ , the particle filter approximates the joint distribution  $p(\lambda_{i,1:T} | u_{1:T}, X_{1:T}, z_{1:T}, \zeta_{1:T}, \theta)$ . The PG sampler draws one path of the state variables from this discrete approximation. As the number of particles  $M$  goes to infinity, the PG sampler draws from the exact full conditional distribution.

We describe the algorithm under the assumption that  $\Phi_\lambda$  and  $\Sigma$  are diagonal in (2). The PG sampler starts with a set of existing particles  $\lambda_{i,1:T}^{(1)}$  that were drawn from the previous iteration.

For  $t = 1, \dots, T$ , run:

- For  $m = 2, \dots, M$ , draw from a proposal:  $\lambda_{i,t}^{(m)} \sim q(\lambda_{i,t} | \lambda_{i,t-1}^{(m)}, u_t, X_t, z_t, \zeta_t, \theta)$ .

- For  $m = 1, \dots, M$ , calculate the importance weight:

$$w_t^{(m)} \propto \frac{p(u_t | z_t, X_t, \lambda_{i,t}^{(m)}, \zeta_t, \theta) p(\lambda_{i,t}^{(m)} | \lambda_{i,t-1}^{(m)}, \theta)}{q(\lambda_{i,t}^{(m)} | \lambda_{i,t-1}^{(m)}, u_t, X_t, z_t, \zeta_t, \theta)}.$$

- For  $m = 1, \dots, M$ , normalize the weights:  $\hat{w}_t^{(m)} = \frac{w_t^{(m)}}{\sum_{m=1}^M w_t^{(m)}}$ .
- Conditionally resample the particles  $\{\lambda_{i,t}^{(m)}\}_{m=1}^M$  with probabilities  $\{\hat{w}_t^{(m)}\}_{m=1}^M$ . In this step, the first particle  $\lambda_{i,t}^{(1)}$  always gets resampled and may be randomly duplicated.

In the particle filter, the proposal distribution  $q(\lambda_{i,t} | \lambda_{i,t-1}, u_t, X_t, z_t, \zeta_t, \theta)$  is chosen by the researcher to approximate the target distribution  $p(u_t | z_t, X_t, \lambda_{i,t}, \zeta_t, \theta) p(\lambda_{i,t} | \lambda_{i,t-1}, \theta)$  as closely as possible, while having heavier tails. At time  $t = 1$ , the proposal distribution simplifies to  $q(\lambda_{i,1} | u_1, X_1, z_1, \zeta_1, \theta)$  as it does not depend on any earlier particles.

The PG sampler is a standard Gibbs sampler but defined on an extended probability space that includes all the random variables that are generated by a particle filter. Implementation of the PG sampler is different than a standard particle filter due to the “conditional” resampling algorithm used in the last step. Specifically, in order for draws from the particle filter to be a valid Markov transition kernel on the extended probability space, Andrieu et al. (2010) note that there must be positive probability of sampling the existing path of the state variables that were drawn at the previous iteration. The pre-existing path must survive the resampling steps of the particle filter. The conditional resampling step within the algorithm forces this path to be resampled at least once. We use the conditional multinomial resampling algorithm from Andrieu et al. (2010), although other resampling algorithms exist, see Chopin and Singh (2013).

In the original PG sampler, the particles  $\{\lambda_{i,t}^{(m)}\}_{m=1}^M$  are stored for  $t = 1, \dots, T$  and a single trajectory is sampled using the probabilities from the last iteration  $\{\hat{w}_T^{(m)}\}_{m=1}^M$ . An important improvement upon the original PG sampler was introduced by Whiteley (2010), who suggested drawing the path of the state variables from the discrete particle approximation using the backwards sampling algorithm of Godsill et al. (2004). On the

forwards pass, we store the normalized weights and particles  $\{\hat{w}_t^{(m)}, \lambda_{i,t}^{(m)}\}_{m=1}^M$  for  $t = 1, \dots, T$ . We draw a path of the state variables  $(\lambda_{i,1}^*, \dots, \lambda_{i,T}^*)$  from this discrete distribution.

At  $t = T$ , draw a particle  $\lambda_{i,T}^* = \lambda_{i,T}^{(m)}$  with probability  $\hat{w}_T^{(m)}$ .

For  $t = T - 1, \dots, 1$ , run:

- For  $m = 1, \dots, M$ , calculate the backwards weights:  $w_{t|T}^{(m)} \propto \hat{w}_t^{(m)} p(\lambda_{i,t+1}^* | \lambda_{i,t}^{(m)}, \theta)$
- For  $m = 1, \dots, M$ , normalize the weights:  $\hat{w}_{t|T}^{(m)} = \frac{w_{t|T}^{(m)}}{\sum_{m=1}^M w_{t|T}^{(m)}}$ .
- Draw a particle  $\lambda_{i,t}^* = \lambda_{i,t}^{(m)}$  with probability  $\hat{w}_{t|T}^{(m)}$ .

The draw  $\lambda_{i,1:T} = (\lambda_{i,1}^*, \dots, \lambda_{i,T}^*)$  is a draw from the full-conditional distribution.

The additional backwards sampling pass dramatically improves the mixing of the Markov chain and allows the PG sampler to run with very few particles. In our work, we use  $M = 100$  particles. Recently, [Chopin and Singh \(2013\)](#) have analyzed the theoretical properties of the PG sampler, proving that it is uniformly ergodic. They also prove that the PG sampler with backwards sampling strictly dominates the original PG sampler in terms of asymptotic efficiency.

When  $\Phi_\lambda$  and  $\Sigma$  in (2) are not diagonal, the paths of the state variable  $\lambda_{i,1:T}$  and  $\lambda_{k,1:T}$  are not independent for  $i \neq k$  even if we condition on the common factor  $z_t$ . There are two ways to proceed. The original PG sampler suggests that we can draw the entire path of the state vector  $\Lambda_{1:T}$  jointly in one block. This works well when the dimension of the state vector is small either because the cross-section  $n$  is small or because the state variables exhibit a factor structure. For example, we estimate the one block and two block equicorrelation models (see the online appendix) by drawing the entire state vector  $\Lambda_{1:T}$  all at once. Alternatively, when the dimension of the state vector is large, we can draw  $\lambda_{i,1:T}$  conditional on all other paths  $\lambda_{i^-,1:T}$  that are not path  $i$ . In other words, we can draw from the full conditional distribution  $p(\lambda_{i,1:T} | u_{1:T}, X_{1:T}, Z_{1:T}, \lambda_{i^-,1:T}, \theta)$  for  $i = 1, \dots, n$ .

## 4. Application to equities and credit default swaps

### 4.1. Data

We have collected daily equity returns and log differences in credit default swap (CDS) rates for 100 US corporations from January 2, 2008 to February 28, 2013. All 100 firms are components of the S&P 500 index. The CDS rate is the 5 year contract with the XR clause. The data on equity prices are from the Center for Research in Securities Prices (CRSP) and the data on the 5 year CDS are from the Markit Corporation. We restrict attention to days when equity markets are open. This makes for a cross-section of  $n = 200$  series with  $T = 1299$  observations. From these  $nT$  total days, there are 2487 observations that are missing at random.

### 4.2. Prior distributions

To reduce the number of parameters in the model, we take  $\Phi_\lambda$  and  $\Sigma$  to be diagonal in (2) and set  $\beta_i = 0$ . We place a normal prior on the long-run mean of the factors  $\mu_i \sim N(0.4, 2)$  for  $i = 1, \dots, n$ , which expresses a belief that the long-run correlations are positive. We use a normal prior on the diagonal elements of  $\Phi_{\lambda,ii} \sim N(0.985, 0.001)$  truncated to the stationarity region and an inverse gamma prior on the diagonal elements of  $\Sigma$ , i.e.  $\Sigma_{ii} \sim \text{Inv Gamma}(20, 0.25)$ . The prior mean of 0.985 for  $\Phi_{\lambda,ii}$  coupled with a small mean for the diagonal elements of  $\Sigma$  indicates a prior

belief in persistent factors with a smooth evolution through time. For the initial conditions of the state variable, we inflate the scale on the initial distribution  $\lambda_{i,1} \sim N(\mu_i, \Sigma_{ii} * 100)$ . Collectively, our priors on  $(\mu, \Phi_\lambda, \Sigma)$  allow us to have conditionally conjugate updates during the Gibbs sampler. The degrees of freedom parameters  $\nu_j$  for  $j = 1, \dots, G$  have a shifted-gamma distribution, i.e.  $\nu_j = 2 + \tilde{\nu}$  where  $\tilde{\nu} \sim \text{Gamma}(2.5, 2)$ . This guarantees that the degrees of freedom is greater than two.

For comparison purposes, we also estimate several factor copulas with constant factor loadings as well as two equicorrelation models with Student's  $t$  errors, see the online appendix for details. In the former case, we place a normal prior on each  $p \times 1$  vector of factor loadings  $\lambda_i \sim N(0.2 \cdot \iota_p, 2 \cdot \mathbf{I}_p)$ . For the one and two block equicorrelation models, the state vectors have dimension one and three respectively. For the equicorrelation models, the priors on the degrees of freedom  $\nu$  and the parameters of the transition density  $(\mu, \Phi_\lambda, \Sigma)$  are the same as above.

### 4.3. Marginal distributions

We model the marginal distribution for each of the  $n = 200$  series using univariate stochastic volatility models with leverage and skewed Student's  $t$  errors for the conditional distribution.<sup>5</sup> Let  $y_{it}$  denote the log-return for the  $i$ th series. The model is specified as

$$y_{it} = W_{it}\beta_{y,i} + \gamma_{y,i}\delta_{it} + \sqrt{\delta_{it}} \exp(h_{it}/2) \varepsilon_{y,it},$$

$$\varepsilon_{y,it} \sim N(0, 1), \quad i = 1, \dots, n$$

$$h_{i,t+1} = \mu_{h,i} + \phi_{h,i}(h_t - \mu_{h,i}) + \sigma_{h,i}\varepsilon_{h,it},$$

$$\varepsilon_{h,it} \sim N(0, 1),$$

$$\delta_{i,t} \sim \text{Inv-Gamma}\left(\frac{\nu_{y,i}}{2}, \frac{\nu_{y,i}}{2}\right), \quad \text{corr}(\varepsilon_{y,it}, \varepsilon_{h,it}) = \rho_i$$

where  $\beta_{y,i}$  are regression parameters,  $W_{it}$  are exogenous covariates,  $\nu_{y,i}$  is the degrees of freedom,  $\gamma_{y,i}$  determines the skewness, and  $\rho_i$  is the leverage parameter. Integrating the mixing variable  $\delta_{it}$  from the model results in a conditional likelihood  $f(y_{it} | W_{it}, h_{it}, \psi_i)$  that is a skewed Student's  $t$  distribution where  $\exp(h_{it}/2)$  is a time-varying scale parameter. The parameters of the model are  $\psi_i = (\beta_{y,i}, \gamma_{y,i}, \phi_{h,i}, \mu_{h,i}, \sigma_{h,i}^2, \rho_i, \nu_{y,i})$  for each series  $i = 1, \dots, n$ . For the covariates  $W_{it}$ , we include an intercept and five (three) lags of  $y_{it}$  for cds spreads (equities). Priors for the parameters of the model are available in the online appendix.

We estimate the SV models by extending the MCMC algorithm in [Omori et al. \(2007\)](#) to include the skewness parameter  $\gamma_{y,i}$ , see the online appendix for details. In each MCMC algorithm, we take 20,000 draws and throw away the first 2000 draws as a burn-in. Unlike GARCH or GAS models that are commonly used for marginal distributions when working with copulas, the CDFs for SV models are not known in closed form. The CDFs can be evaluated by simulation using the particle filter. After estimating each SV model, we calculate the posterior mean  $\bar{\psi}_i$  and the probability integral transforms  $\bar{u}_{it} = F(Y_{it} \leq y_{it} | y_{i,1:t-1}, \bar{\psi}_i)$  using the particle filter with  $M = 100,000$  particles, see the online appendix. In the next step, we take these values  $\bar{u}_t = (\bar{u}_{1t}, \dots, \bar{u}_{nt})$  for  $t = 1, \dots, T$  as our data to estimate the copula.<sup>6</sup> In practice, we reverse the sign of the

<sup>5</sup> Our prior for the degrees of freedom only assigns mass greater than three. However, for some series (particularly CDS), the tails of the distribution are heavy enough that it made sense to drop the skewness parameter and estimate a symmetric Student's  $t$  model.

<sup>6</sup> In a fully Bayesian approach when the marginals and copula are estimated simultaneously, the probability integral transforms  $u_{it} = F(Y_{it} \leq y_{it} | y_{i,1:t-1}, \psi_i)$  need to be re-calculated by the particle filter for every draw of  $\psi_i$  in the MCMC algorithm. This is feasible when the cross-section is small using the particle Metropolis–Hastings sampler, see [Andrieu et al. \(2010\)](#).

**Table 1**

Summary statistics from the univariate marginals across the 100 US corporations.

		$v_y$	$\rho$	$\gamma_y$	$\mu_h$	$\phi_h$	$\sigma_h^2$
CDS	mean	5.585	−0.039	−0.044	1.007	0.941	0.206
	95% hpdi-width	4.052	0.272	0.349	0.825	0.059	0.215
	min	2.718	−0.112	−0.347	−0.348	0.746	0.061
	max	10.880	0.174	0.232	2.398	0.980	0.762
	lnef	12.95	8.80	12.51	7.41	7.88	14.04
Equity	mean	12.853	−0.358	0.0536	0.822	0.974	0.057
	95% hpdi-width	10.243	0.344	0.630	1.184	0.029	0.046
	min	4.041	−0.567	−0.378	−0.652	0.946	0.036
	max	21.741	0.119	0.535	2.229	0.993	0.110
	lnef	9.52	7.21	10.48	6.02	7.17	8.97

The table reports for each parameter the average (across 100 firms) posterior mean, the average width of the 95% highest posterior density intervals, and the minimum and maximum posterior mean. Also reported are the inefficiency factors (averaged across 100 firms) for each of the parameters.

**Table 2**

Estimation results for alternative copula models.

	Gauss const. $\lambda$	Stud- $t$ const. $\lambda$	gr-Stud- $t$ const. $\lambda$	Stud- $t$ equi	Stud- $t$ block equi	Gauss t.v. $\lambda_{it}$	Stud- $t$ t.v. $\lambda_{it}$	gr-Stud- $t$ t.v. $\lambda_{it}$
LPS	−41.76	−45.75	−51.89	−42.28	−47.05	−53.53	−55.72	−57.23
AIC	−108092.4	−118482.5	−134382.2	−109835.4	−122215.9	−137868.9	−143584.5	−147455.5
BIC	−107058.6	−117443.4	−133275.9	−109814.7	−122164.2	−134762.2	−140477.7	−144281.5
#parms	200	201	214	4	10	600	601	614
$\nu$	–	31.95	[10.65, 19.57]	54.95	41.47	–	32.35	[12.46, 25.06]

The table reports the log-predictive score (LPS), AIC, and BIC for a total of eight models. Three models have constant loadings and three have time-varying loadings. The distributions are Gaussian, Student's  $t$ , and grouped Student's  $t$ . We also report an equicorrelation model and a two-block equicorrelation model with Student's  $t$  errors. The table also reports the posterior mean for  $\nu$  for the Student's  $t$  models and the range in the posterior mean of  $\nu$  for the grouped Student's  $t$  models.

PITs for CDS spreads by studying  $\bar{u}_{it} = 1 - F(Y_{it} \leq y_{it} | y_{i,1:t-1}, \bar{\psi}_i)$ . This makes all the correlations across equities and CDS positive on average.

To evaluate the fit of the marginal distributions, we transform the probability integral transforms  $\bar{u}_{it}$  into Gaussian variables  $\bar{x}_{it} = \Phi^{-1}(\bar{u}_{it})$  for  $i = 1, \dots, n$  and  $t = 1, \dots, T$ . We then perform a Kolmogorov–Smirnov test for equality with the normal distribution on the value of the  $\bar{x}_{it}$ 's for each firm's CDS and equity series separately. Across the 100 firms, only one model failed the null hypothesis for the equity series while 11 models failed for the CDS series. We also tested for serial correlation in the Gaussian variables  $\bar{x}_{it}$ , and found no evidence of serial correlation for equities and mild violations for the CDS series. Upon inspection of the latter series, we found that the CDS quotes for these series became stale for short periods of time resulting in periods with zero or close-to-zero log-returns.

In Table 1, we report summary statistics for the posterior distribution of the univariate SV models applied to CDS and equity returns across the 100 firms. The table includes the average posterior mean, the average width in the 95% highest posterior density intervals (HPDI), the minimum posterior mean, and the maximum posterior mean. The first two are averages across the 100 firms. On average, CDS returns have extremely heavy tails with an average degrees of freedom of 5.6, which is significantly heavier than equities. The volatility of CDS returns is less persistent with an average AR(1) parameter of 0.941 versus a value of 0.974 for equities. The leverage effect does not appear to be important on average for CDS while it is a major feature of equities. The 95% HPDI's for  $\rho$  covered zero for 94% of firms' CDS and only 5% out of 100 firm's equity. Similarly, skewness does not seem to play a major role on average, as the 95% HPDI covered zero for 87% of the firm's CDS and 83% of the firm's equity.

In Table 1, we also report for each parameter the inefficiency factor (averaged across the 100 firms) of the MCMC draws. The inefficiency factor (or integrated autocorrelation time) is defined as  $\text{lnef} = 1 + 2 \sum_{k=1}^{\infty} \rho(k)$  where  $\rho(k)$  is the autocorrelation at lag  $k$  of the MCMC chain. It measures the degree of autocorrelation with smaller values indicating better mixing due to better MCMC algorithms. The inefficiency factors are comparable to estimates

reported for other univariate SV models, see Chib et al. (2002) and Omori et al. (2007). These provide evidence that the MCMC algorithm mixes reasonably well for equities and slightly less well for CDS.

#### 4.4. Copula estimation

Given the probability integral transforms from the univariate SV models, we estimate a total of eight different copula models with no exogenous covariates. These include three copulas with one factor and constant factor loadings (Gaussian, Student's  $t$ , grouped Student's  $t$ ), three factor copulas with one time-varying loading (Gaussian, Student's  $t$ , and grouped Student's  $t$ ), and a single block and a two-block equicorrelation model (Student's  $t$ ). In the two-block equicorrelation model, we separate all the CDS into one block and all the equities into the other block. For the factor copulas, there are a total of  $n(n-1)/2 = 19,900$  correlations, which will be time-varying for models with random factor loadings. We run the MCMC algorithm on each copula for 30,000 draws after a burn-in of 5000 draws. We thin the Markov chain by taking each 10th draw.

For the grouped Student's  $t$  copulas, we created a total of  $G = 14$  industry groups of unequal size based on each firm's SIC code. The industries are Oil, Food & Beverage, Pharmaceuticals, Plastics & Chemicals, Paper products, Steel & Refining, Home Appliances, Electronics, Transportation, Water & Natural Gas, Retail, Insurance, Finance (less insurance), Services (advertising). On average each group has 7 firms making for 14 series, as the debt and equity of each firm are always in the same group.

Table 2 contains the AIC, BIC, and log-predictive score (LPS) for each of the copula models.<sup>7</sup> These are evaluated at the posterior mean  $\bar{\theta}$  of the draws. The likelihood for each of the models is calculated by the particle filter with the exception of the Gaussian

<sup>7</sup> The log-predictive score of Good (1952) ranks models according to their predictive ability with smaller values being preferred. The LPS =  $-\frac{1}{T} \sum_{t=1}^T \log p(u_t | u_{1:t-1}, X_{1:t}, \bar{\theta})$  where  $p(u_t | u_{1:t-1}, X_{1:t}, \bar{\theta})$  is the contribution to the likelihood and  $\bar{\theta}$  is the posterior mean.

**Table 3**  
Results for the grouped Student's- $t$  copula with time-varying factor loadings.

	Oil	Food & Bev.	Pharma.	Plastics	Paper	Steel	Home App.
$\nu$	14.41 (1.381)	17.67 (3.042)	15.09 (2.968)	20.06 (4.508)	23.01 (4.005)	21.94 (2.862)	17.18 (6.328)
Inef	3.45	5.36	6.06	6.21	5.09	4.01	10.15
$\mu$	[0.632, 1.428] (0.098)	[0.606, 0.764] (0.067)	[0.540, 0.818] (0.068)	[0.706, 0.931] (0.074)	[0.234, 1.078] (0.088)	[0.813, 1.102] (0.080)	[−0.060, 0.976] (0.070)
Inef	1.01	1.06	1.05	1.10	1.07	1.01	1.5
$\Phi_\lambda$	[0.938, 0.962] (0.013)	[0.921, 0.954] (0.015)	[0.916, 0.956] (0.015)	[0.924, 0.956] (0.014)	[0.930, 0.958] (0.014)	[0.914, 0.959] (0.014)	[0.928, 0.958] (0.016)
Inef	1.46	1.49	1.59	1.47	1.81	1.63	1.45
$\Sigma \times 10$	[0.108, 0.153] (0.023)	[0.107, 0.174] (0.028)	[0.099, 0.166] (0.027)	[0.110, 0.180] (0.030)	[0.108, 0.160] (0.029)	[0.100, 0.164] (0.028)	[0.098, 0.162] (0.028)
Inef	4.20	4.27	4.79	4.77	4.61	4.58	4.69
#firms	10	8	5	7	8	6	4
	Electronics	Trans.	Wat. & N. Gas	Retail	Insur.	Finance	Services
$\nu$	18.73 (2.406)	21.70 (3.157)	12.60 (1.452)	19.23 (2.350)	17.76 (1.805)	12.46 (1.222)	25.06 (5.828)
Inef	4.16	4.65	4.28	4.35	3.83	3.84	5.97
$\mu$	[0.598, 1.033] (0.082)	[−1.455, 1.261] (0.090)	[0.627, 0.800] (0.073)	[0.656, 0.814] (0.066)	[−1.422, 1.627] (0.106)	[0.776, 1.338] (0.086)	[0.669, 0.866] (0.070)
Inef	1.08	1.01	1.10	1.01	1.50	1.09	1.20
$\Phi_\lambda$	[0.927, 0.957] (0.015)	[0.928, 0.958] (0.015)	[0.927, 0.955] (0.014)	[0.919, 0.953] (0.016)	[0.940, 0.968] (0.012)	[0.934, 0.963] (0.014)	[0.917, 0.958] (0.015)
Inef	1.59	1.63	1.48	1.67	1.54	1.60	1.74
$\Sigma \times 10$	[0.106, 0.161] (0.028)	[0.110, 0.167] (0.029)	[0.110, 0.157] (0.028)	[0.101, 0.157] (0.026)	[0.104, 0.158] (0.027)	[0.103, 0.150] (0.027)	[0.103, 0.181] (0.031)
Inef	4.43	4.62	4.62	4.43	4.36	4.73	4.77
#firms	8	8	9	7	7	8	5

The table reports estimates from the posterior distribution of the grouped Student's  $t$  copula with  $G = 14$  industry groups. This includes the posterior mean and standard deviation for  $\nu_j$  for  $j = 1, \dots, G$ . The parameters  $(\mu, \Phi_\lambda, \Sigma)$  are the range in posterior means across all series  $i$  that are in group  $j$ . For each of these values, we report the average posterior standard deviation of  $\mu_i, \Phi_{\lambda,ii}, \Sigma_{\lambda,ii}$  across all series  $i$  that are in group  $j$ . The row labeled 'Inef' is the inefficiency factor of the MCMC draws.

and Student's  $t$  copulas with constant factor loadings, as these are known in closed-form, see the online appendix. Calculating the likelihood function in high dimensions for a factor copula with time-varying loadings is challenging due to the dimension of the state vector. Theoretical results for particle filtering indicate that the variance of the estimator of the likelihood is directly related to the dimension of the state vector. In practice, to estimate the likelihood for the factor models, we run the particle filter multiple times (100) and take the average log-likelihood across the runs. The results in Table 2 indicate that the grouped Student's  $t$  copula with time-varying factor loadings is the preferred model followed by the Student's  $t$  model with time-varying loadings. The grouped Student's  $t$  copula with constant loadings is also preferred to other models with time-invariant loadings by a marked margin.

The degrees of freedom parameter is estimated to be significantly higher for the Student's  $t$  distributions than for the grouped Student's  $t$  models. This is understandable as the number of firms in a group increases (i.e. the cross-section per group becomes large) a central limit theorem takes hold. Consequently, grouped Student's  $t$  models are more flexible than Student's  $t$  copulas in handling diversified heavy tails.

In Table 3, we report the parameter estimates for the grouped Student's  $t$  copula with time-varying factor loadings separated into each of the  $G = 14$  industry groups. The estimates of  $\nu$  are the posterior mean and standard deviation while the reported intervals for  $\mu, \Phi_\lambda$ , and  $\Sigma$  are the range in posterior means across the series in each industry group. Below these values in parentheses are the average value of the posterior standard deviation, which have been adjusted for the serial correlation in the MCMC draws. The estimates of  $\nu$  and the associated uncertainty vary considerably from one industry group to another. The smallest estimated value is 12.46 for the Finance industry and the largest is 25.06 for the Services industry. The latent factor loadings are reasonably highly autocorrelated but not as high as traditionally found for daily equity returns of univariate GARCH and SV models.

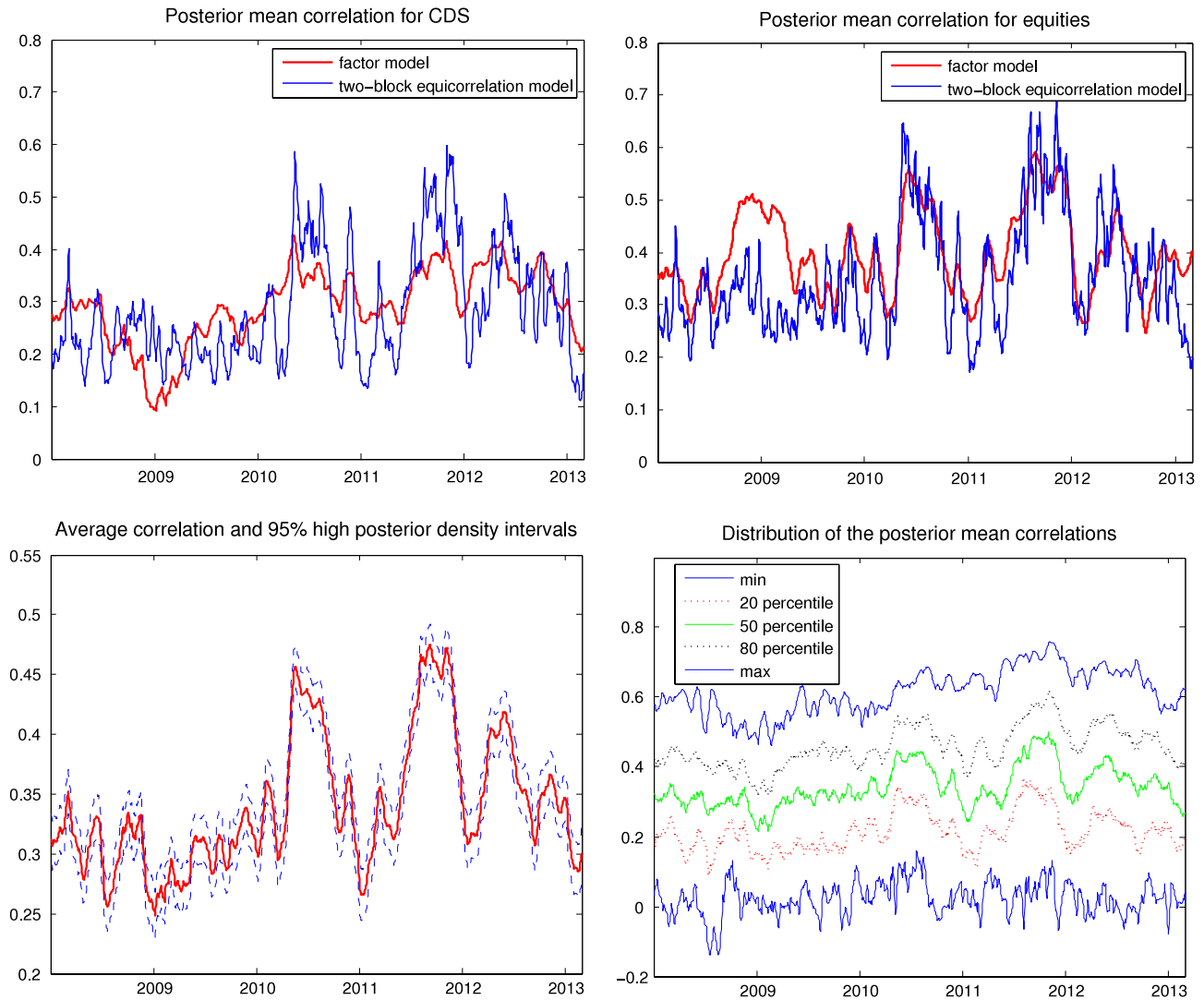
Included in Table 3 are the inefficiency factors for each parameter. For  $(\mu, \Phi_\lambda, \Sigma)$ , these are the average values for these parameters across firms in a group. In general, these indicate that the MCMC algorithm is mixing reasonably well after thinning it by the 10th draw. The autocorrelations for the degrees of freedom  $\nu$  and volatility  $\Sigma$  are higher, while  $\mu$  and  $\Phi_\lambda$  mix more rapidly.

In the top left of Fig. 1, we plot the posterior mean (smoothed) estimate of the conditional correlation between all CDS series and, in the top right, we plot the same estimates for equities. In these graphs, estimates from the grouped Student's  $t$  factor copula are compared to the two-block equicorrelation model. For the factor models, the posterior mean correlation for CDS and equities are calculated at each date  $t$  by averaging all values of the conditional correlation across series  $i$  and  $k$  that are both CDS, i.e.  $\rho_{t, \text{CDS}} = \frac{1}{100(100-1)/2} \sum_{i,k} R_{t,ik}$  if both  $i$  and  $k$  are CDS. A similar procedure is used for equities. The estimates from both models share the same broad trend for both CDS and equities. In addition, the average path of the correlation across equities and CDS have similar dynamics. The conditional correlations are greater after the financial crisis with peaks in both the middle of 2010 and the end of 2011. This is confirmed in the bottom left panel, which plots the average posterior mean across all assets in the entire portfolio along with the 95% highest posterior density intervals.

In the lower right panel of Fig. 1, we provide a summary of the variation in the posterior mean correlations across the 100 time series paths for equity and CDS of each firm. At each date, we calculate the posterior mean for each conditional correlation and then report the minimum, maximum, and (20, 50, 80)th percentiles. This gives an indication of the wide range of correlation patterns found across series.

In Fig. 2, we plot summary statistics from the posterior distribution for several individual firms including Coca Cola, AIG, Boeing, and Goldman Sachs. Each row represents a different firm. The first column includes the estimated posterior mean of the conditional volatility of each firm's CDS, the second column is the





**Fig. 1.** Time plots of average conditional correlations of two copula models. Top left: average conditional correlations across all firm's CDS returns compared to the estimated conditional correlation between CDS from the two-block equicorrelation model. Top right: average correlations across all firm's equity returns compared to the estimated conditional correlation between equities from the two-block equicorrelation model. Bottom left: the average conditional correlations for all assets and the 95% highest posterior density intervals. Bottom right: the minimum, maximum, 20%, 50%, 80% values of the posterior mean conditional correlation between equity and CDS across firms.

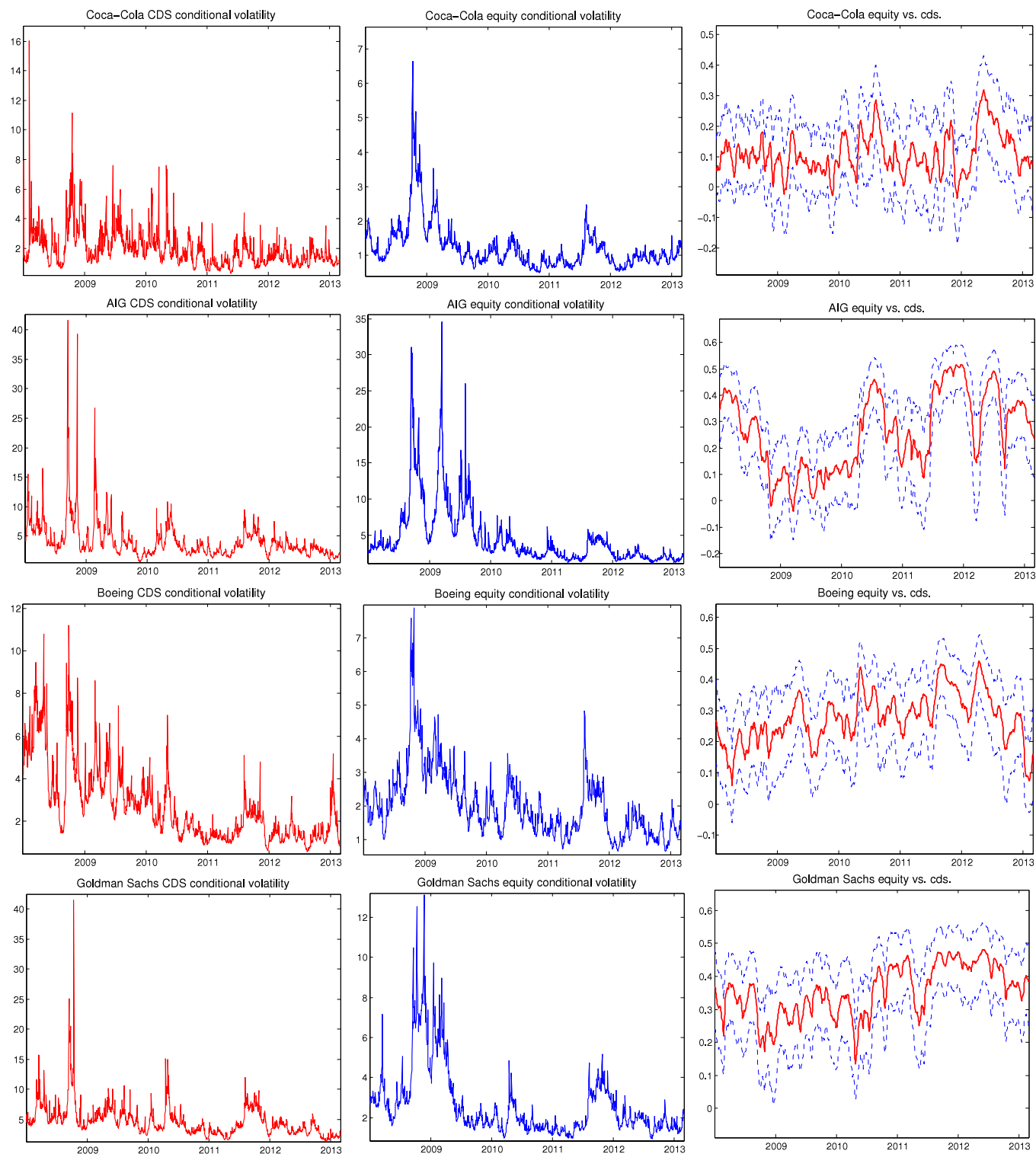
conditional volatility of each firm's equity, and the final column is Kendall's (conditional) rank correlation between a firm's CDS and equity. From the plots, we make the following observations. First, the CDS spread is more volatile than the equity return for each individual firm. In particular, the range in volatility for the CDS spread of Goldman Sachs is about three times greater than its equity return. Second, as expected, the volatility of CDS is higher during the recent credit crisis and declines after the crisis. Third, the volatility of equity returns is also higher at the end of 2008, and increases at the latter part of 2011. Finally, Kendall's tau between the CDS spread and equity return of individual firm varies markedly over time. In general, Kendall's tau appears to be higher after the recent financial crisis.

Calculation of either Kendall's or Spearman's rank correlation (or tail dependence measures) are particularly easy for time-varying Gaussian and Student's  $t$  copulas. Bivariate marginal distributions remain within the same family of distributions and these measures are simple closed-form transformations of the traditional linear correlation  $\tau_{t,ik} = \frac{2}{\pi} \arcsin(R_{t,ik})$  for series  $i$  and  $k$ . This expression for Kendall's  $\tau$  does not hold for all bivariate series  $i$  and  $k$  in the grouped Student's  $t$  copula. However, for

bivariate series  $i$  and  $k$  that are in the same group, the closed-form formula for Kendall's tau continues to apply for the grouped Student's  $t$  copula employed in the paper. This is because we use the same mixing variable and the same degrees of freedom in each group. If we condition on all the mixing variables at each date, the  $n \times 1$  vector  $x_t$  is conditionally Gaussian. Bivariate marginals within the same group are Gaussian, i.e.  $(x_{it}, x_{kt}) \sim N(0, \zeta_{t,j} R_{t,ik})$ . Integrating out  $\zeta_{t,j}$ , we obtain a Student's  $t$  distribution for the pair  $(x_{it}, x_{kt})$ .

## 5. Conclusion

We built a class of copula models with time-varying dependence parameters by representing the copulas as factor models with stochastic loadings. The class of conditional copula densities included Gaussian, Student's  $t$ , grouped Student's  $t$ , and generalized hyperbolic copulas, with time-varying conditional correlation matrices. The factor structure of the copula models simplifies computation because calculation of the inverse, determinant, and quadratic forms involving the conditional correlation matrix are simple, low dimensional operations. Consequently, the models can



**Fig. 2.** Posterior estimates of conditional volatility and rank correlations for four firms. First column: estimated conditional volatility of CDS; Second column: estimated conditional volatility of equity; Third column: estimated conditional (Kendall) rank correlations with 95% highest posterior density intervals between equity returns and CDS returns from the grouped Student's  $t$  factor copula. First row: Coca-Cola; Second row: AIG; Third row: Boeing Corp.; Bottom row: Goldman Sachs.

easily be used to estimate portfolios with hundreds of assets. We applied the methods to an unbalanced, 200 dimensional panel of CDS and equities for 100 US corporations. Our analysis shows that the grouped Student's  $t$  copula with time-varying factor loadings fits the data better than seven other competing models, and the model is capable of describing the time-varying correlations and tail dependence between assets in the same group. The model also

provides time-varying dependence between returns of CDS and equity of a given company.

#### Appendix A. Supplementary data

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.jeconom.2015.03.027>.

## References

- Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle Markov chain Monte Carlo methods (with discussion). *J. R. Stat. Soc. Ser. B* 72 (2), 1–33.
- Andrieu, C., Roberts, G.O., 2009. The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* 37 (2), 697–725.
- Cappé, O., Moulines, E., Rydén, T., 2005. *Inference in Hidden Markov Models*. Springer Press, New York.
- Chib, S., Nardari, F., Shephard, N., 2002. Markov chain Monte Carlo methods for stochastic volatility models. *J. Econometrics* 108 (2), 281–316.
- Chopin, N., Singh, S.S., 2013. On the particle Gibbs sampler. Working paper, ENSAE. <http://arxiv.org/abs/1304.1887>.
- Christoffersen, P.F., Jacobs, K., Jin, X., Langlois, H., 2013. Dynamic dependence and diversification in corporate credit. Working Paper, C.T Bauer College of Business, University of Houston.
- Creal, D.D., 2012. A survey of sequential Monte Carlo methods for economics and finance. *Econometric Rev.* 31 (3), 245–296.
- Creal, D.D., Gramacy, R.B., Tsay, R., 2014. Market-based credit ratings. *J. Bus. Econom. Statist.* 32 (3), 430–444.
- Creal, D.D., Koopman, S.J., Lucas, A., 2011. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *J. Bus. Econom. Statist.* 29 (4), 552–563.
- Creal, D.D., Koopman, S.J., Lucas, A., 2013. Generalized autoregressive score models with applications. *J. Appl. Econometrics* 28 (5), 777–795.
- Daul, S., Giorgi, E., Lindskog, F., McNeil, A.J., 2003. The grouped- $t$  copula with an application to credit risk. *RISK* 16, 73–76.
- Demarta, S., McNeil, A.J., 2005. The  $t$  copula and related copulas. *Internat. Statist. Rev.* 73 (1), 111–129.
- Durbin, J., Koopman, S.J., 2012. *Time Series Analysis by State Space Methods*, second ed. Oxford University Press, Oxford, UK.
- Engle, R.F., Kelly, B., 2012. Dynamic equicorrelation. *J. Bus. Econom. Statist.* 30 (2), 212–228.
- Flury, T., Shephard, N., 2011. Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models. *Econometric Theory* 27 (5), 933–956.
- Godsill, S.J., Doucet, A., West, M., 2004. Monte Carlo smoothing for nonlinear time series. *J. Amer. Statist. Assoc.* 99 (465), 156–168.
- Good, I.J., 1952. Rational decisions. *J. R. Stat. Soc. Ser. B* 14, 107–114.
- Hafner, C.M., Manner, H., 2012. Dynamic stochastic copula models: estimation, inference and applications. *J. Appl. Econometrics* 27 (2), 269–295.
- Krupskii, P., Joe, H., 2013. Factor copula models for multivariate data. *J. Multivariate Anal.* 120, 85–101.
- Li, D.X., 2000. On default correlation: a copula function approach. *J. Fixed Income* 9 (4), 43–54.
- Lucas, A., Schwaab, B., Zhang, X., 2014. Conditional Euro area sovereign default risk. *J. Bus. Econom. Statist.* 32 (2), 271–284.
- Manner, H., Reznikova, O., 2012. A survey of time-varying copulas: specification, simulations, and application. *Econometric Rev.* 31 (6), 654–687.
- McNeil, A.J., Frey, R., Embrechts, P., 2005. *Quantitative Risk Management*. Princeton University Press, Princeton, NJ.
- Min, A., Czado, C., 2010. Bayesian inference for multivariate copulas using pair-copula constructions. *J. Financ. Econometrics* 8 (4), 511–546.
- Murray, J.S., Dunson, D.B., Carin, L., Lucas, J.E., 2013. Bayesian Gaussian copula factor models for mixed data. *J. Amer. Statist. Assoc.* 108 (502), 656–665.
- Oh, D.H., Patton, A.J., 2012. Modelling dependence in high dimensions with factor copulas. Unpublished manuscript, Department of Economics, Duke University.
- Oh, D.H., Patton, A.J., 2013. Time-varying systemic risk: evidence from a dynamic copula model of CDS spreads. Unpublished manuscript, Department of Economics, Duke University.
- Omori, Y., Chib, S., Shephard, N., Nakajima, J., 2007. Stochastic volatility with leverage: fast and efficient likelihood inference. *J. Econometrics* 140 (2), 425–449.
- Patton, A.J., 2006. Modelling asymmetric exchange rate dependence. *Internat. Econom. Rev.* 47 (2), 527–556.
- Patton, A.J., 2009. Copula-based models for financial time series. In: Anderson, T.G., Davis, R.A., Kreiss, J.P., Mikosch, T. (Eds.), *Handbook of Financial Time Series*. Springer Verlag, New York, pp. 767–786.
- Pelletier, D., 2006. Regime switching for dynamic correlations. *J. Econometrics* 131 (1–2), 445–473.
- Pitt, M.K., Chan, D., Kohn, R., 2006. Efficient Bayesian inference for Gaussian copula regression models. *Biometrika* 93 (3), 537–554.
- Sklar, A., 1959. Fonctions de répartition à  $n$  dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* 8, 229–231.
- Smith, M.S., 2011. Bayesian approaches to copula modelling. In: P. Damien, P. Dellaportas, N. Polson, D. Stephens (Eds.), *Hierarchical models and MCMC: A Tribute to Adrian Smith*.
- Smith, M.S., Gan, Q., Kohn, R., 2012. Modeling dependence using skew  $t$  copulas: Bayesian inference and applications. *J. Appl. Econometrics* 27 (3), 500–522.
- Whiteley, N., 2010. Discussion on particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B* 72 (3), 306–307.

## Further reading

- Chib, S., Greenberg, E., 1994. Bayes inference in regression models with ARMA(p,q) errors. *J. Econometrics* 64 (1–2), 183–206.
- Roberts, G.O., Rosenthal, J.S., 2009. Examples of adaptive MCMC. *J. Comput. Graph. Stat.* 18, 349–367.