

Christian M. Hafner, Hans Manner

Dynamic stochastic copula models: Estimation,
inference and applications

RM/08/043

JEL code: C14, C22



Maastricht research school of **E**conomics
of **T**echnology and **O**rganizations

Universiteit Maastricht
Faculty of Economics and Business Administration
P.O. Box 616
NL - 6200 MD Maastricht

phone : ++31 43 388 3830
fax : ++31 43 388 4873

Dynamic stochastic copula models: Estimation, inference and applications

Christian M. Hafner¹

Hans Manner²

November 28, 2008

Abstract

We propose a new dynamic copula model where the parameter characterizing dependence follows an autoregressive process. As this model class includes the Gaussian copula with stochastic correlation process, it can be viewed as a generalization of multivariate stochastic volatility models. Despite the complexity of the model, the decoupling of marginals and dependence parameters facilitates estimation. We propose estimation in two steps, where first the parameters of the marginal distributions are estimated, and then those of the copula. Parameters of the latent processes (volatilities and dependence) are estimated using efficient importance sampling (EIS). We discuss goodness-of-fit tests and ways to forecast the dependence parameter. For two bivariate stock index series, we show that the proposed model outperforms standard competing models.

Keywords: Copula, stochastic volatility, stochastic dependence, efficient importance sampling

JEL Classification: C14, C22.

¹Institut de statistique and CORE, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium. Tel +32 10 47 43 06, fax +32 10 47 30 32 christian.hafner@uclouvain.be, corresponding author.

²Department of Quantitative Economics, Maastricht University, PoBOX 616, MD 6200, Maastricht, The Netherlands, h.manner@ke.unimaas.nl.

1 Introduction

Stochastic volatility (SV) models have attracted considerable interest in recent years, as they have been shown to offer a higher goodness-of-fit and flexibility than GARCH-type models when applied to financial time series, see e.g. Jacquier, Polson and Rossi (1994), Danielsson (1998), Kim, Shephard and Chib (1998), Carnero, Pena and Ruiz (2004) for comparisons between the two model classes. With the same number, or less, of parameters the SV models gain additional flexibility through the use of stochastic latent variables that drive volatilities. Until quite recently, SV models have suffered from the difficult estimation problem, due to the fact that the evaluation of the likelihood function amounts to solving an integral of dimension equal to the sample size. Several methods have been proposed to circumvent this problem, see e.g. Broto and Ruiz (2004) for a survey. The proposed methods still require in most cases some computational effort, but thanks to increased computing power one can now estimate univariate SV models for typical sample sizes in the realm of seconds or even fractions of seconds. This makes them attractive to the applied econometrician and attenuates the comparative advantage of GARCH models in terms of computational simplicity.

Starting with Harvey, Ruiz, and Shephard (1994), multivariate SV models have become more popular in empirical finance to describe return volatilities and correlations, with direct applications e.g. to portfolio selection. Yu and Meyer (2006) are, to our knowledge, the first ones to propose a multivariate SV model with stochastic correlations, where the Fisher transform of the correlation follows a Gaussian AR(1) process. Asai and McAleer (2009) propose a model that is similar in structure to the DCC model of Engle (2002) but where the correlations are driven by a stochastic VAR(1)-type process. Amisano and Casarin (2008) suggest to introduce a Markov-switching process to explain correlation dynamics in a multivariate SV framework. These model classes are very flexible, but they might fail to correctly describe the dependence between the series in situations where the dependence is nonlinear. If the multivariate return distribution conditional on the latent variables was elliptical, then correlations would be sufficient to describe dependence. However, one often observes non-elliptical distributions especially in finance, where the lower tail-dependence is usually much higher than upper tail-dependence.

A natural, and in fact the most general way to model non-linear dependencies is to use copulas, see e.g. the monograph of Joe (1997). A broad variety of copula models exists that allows to capture, for example, asymmetries in tail dependencies. One advantage of using copulas is the decoupling of marginal distributions from the dependence. This

allows to first estimate the marginal processes, e.g. assuming univariate SV processes, and then in a second step the copula parameters. We propose a model where one or more of the copula parameters follows a transformation of a stochastic process, e.g. a Gaussian autoregressive process. As this is again a stochastic latent variable, it is in the same spirit as standard multivariate SV models. As a special case, we obtain a multivariate SV model with stochastic correlation by using a Gaussian copula. The use of other, in particular asymmetric copulas allows to better capture asymmetric tail dependencies. In this sense our model can be viewed as a generalization of standard multivariate SV models. However, as we will argue, estimation is straightforward due to the two step estimation. In particular, for estimation we use the efficient importance sampling (EIS) algorithm of Liesenfeld and Richard (2003) and extend it to the present model framework. We discuss specification tests such as parameter constancy and copula selection. In case of nonlinear transformations of the latent process, forecasting becomes a nontrivial issue, but we show that in most cases simple solutions can be found.

In the application part of the paper we show that for two different bivariate stock index series, a stochastic copula model with SV marginals outperforms a DCC model with GARCH(1,1) marginals. This also holds if we replace the SV marginals by GARCH(1,1) to compare directly the dependence models, i.e. correlation versus stochastic copula.

The remainder of the paper is organized as follows. The next section introduces the model, discusses an estimation and inference method as well as the testing and forecasting problems. Section 3 discusses the application to bivariate stock index series, both at a daily and a weekly frequency and Section 4 concludes.

2 Specification, estimation and diagnostics

2.1 The stochastic copula model

Let x and y be two continuous random variables with joint distribution function $H(x, y)$ and marginal distribution F and G , respectively. Then by Sklar's theorem there exists a unique copula C such that

$$H(x, y) = C(F(x), G(y)). \tag{1}$$

Thus by the probability integral transform a copula is a multivariate¹ distribution function with uniform marginals and it fully captures the dependence between x and y irrespective of the marginal distributions. Examples and properties of copulas can be found e.g. in the book by Joe(1997).

Consider the bivariate time series process (u_t, v_t) for $t = 1, \dots, T$ with distribution function given by the following time-varying copula model

$$(u_t, v_t) \sim C(u, v|\theta_t), \quad (2)$$

where $\theta_t \in \Theta \subset \mathbb{R}^K$ is a random parameter vector of the copula function. In this paper we only consider copulas with one parameter, so that $K = 1$. We assume that θ_t is driven by an unobserved stochastic process λ_t such that $\theta_t = \Psi(\lambda_t)$, where $\Psi : \mathbb{R} \rightarrow \Theta$ is an appropriate transformation to ensure that the copula parameter remains in its domain and whose functional form depends on the choice of copula. The underlying dependence parameter λ_t , which is unobserved, is assumed to follow a Gaussian autoregressive process of order one,

$$\lambda_t = \alpha + \beta\lambda_{t-1} + \nu\varepsilon_t, \quad (3)$$

where ε_t is an i.i.d. $N(0, 1)$ innovation. Of course, higher order autoregressive models could be considered, but estimation would become very difficult and in all applications we found that AR(1) sufficiently describes persistence and the autocorrelation structure of the dependence parameter. We assume that the latent process is strictly stationary, i.e. $|\beta| < 1$. For identification reasons we also assume that the scale parameter of the innovations, ν , is positive. Note that we ignore specification of the marginals for the moment, so we assume that the observed variables have a $U(0, 1)$ distribution or, alternatively, we assume perfect knowledge of the marginals. We will discuss possibilities on how to deal with the marginals below.

2.2 Estimation

We are interested in estimating the parameter vector $\omega = (\alpha, \beta, \nu)$. Denote $U = \{u_t\}_{t=1}^T$, $V = \{v_t\}_{t=1}^T$ and $\Lambda = \{\lambda_t\}_{t=1}^T$ and let $f(U, V, \Lambda; \omega)$ be the joint density of the observable variables (U, V) and the latent process Λ . Then the likelihood function of the parameter

¹We only consider bivariate copulas in this paper, but most methods can in principle be extended to the multivariate case.

vector ω is

$$L(\omega; U, V) = \int f(U, V, \Lambda; \omega) d\Lambda. \quad (4)$$

If now $U_t = \{u_\tau\}_{\tau=1}^t$, and similarly for V_t and Λ_t , we can factorize the integrand of this likelihood function into a sequence of conditional densities as follows.

$$L(\omega; U, V) = \int \prod_{t=1}^T f(u_t, v_t, \lambda_t | U_{t-1}, V_{t-1}, \Lambda_{t-1}, \omega) d\Lambda. \quad (5)$$

Furthermore the joint density $f(u_t, v_t, \lambda_t | U_{t-1}, V_{t-1}, \Lambda_{t-1}, \omega)$ can be factorized into the copula density $c(u_t, v_t | \lambda_t, U_{t-1}, V_{t-1}, \omega)$ times the conditional density of λ_t given $(U_{t-1}, V_{t-1}, \Lambda_{t-1})$, which is $p(\lambda_t | U_{t-1}, V_{t-1}, \Lambda_{t-1}, \omega)$. Since p does not depend on the past observable variables (U_{t-1}, V_{t-1}) they can be omitted for the sake of notation. This gives the following likelihood function for ω

$$L(\omega; U, V) = \int \prod_{t=1}^T c(u_t, v_t | \lambda_t, U_{t-1}, V_{t-1}, \omega) p(\lambda_t | \Lambda_{t-1}, \omega) d\Lambda. \quad (6)$$

This integral is T -dimensional and cannot be evaluated by analytical or numerical methods even for moderate sample sizes. However, it can be evaluated by simulation. In principle, one could simulate a large number N of trajectories $\{\tilde{\lambda}_t^{(i)}(\omega)\}_{t=1}^T$ from p , which we call the natural sampler, and evaluate the likelihood function by

$$\hat{L}_N(\omega; U, V) = \frac{1}{N} \sum_{i=1}^N \left[\prod_{t=1}^T c(u_t, v_t | \tilde{\lambda}_t^{(i)}(\omega), U_{t-1}, V_{t-1}, \omega) \right]. \quad (7)$$

However, as mentioned in Danielsson and Richard (1993) and Liesenfeld and Richard (2003) this estimator, labeled the “natural MC estimate”, is very inefficient for reasonably large sample sizes. This results from the fact that the trajectories $\{\tilde{\lambda}_t^{(i)}(\omega)\}_{t=1}^T$ are sampled independently of the observed variables U and V and thus do not make any use of the information available in the data.

A technique to handle such problems, which is proposed in Liesenfeld and Richard (2003) and Richard and Zhang (2007), is called efficient importance sampling (EIS). The main idea of EIS is to make use of the information on Λ contained in the observable variables U and V to construct a new sampler that exploits this information. Denote a sequence of this auxiliary sampler by $\{m(\lambda_t | \Lambda_{t-1}, a_t)\}_{t=1}^T$ indexed by the auxiliary parameters a_t , which need to be estimated. The likelihood function can be rewritten as

$$L(\omega; U, V) = \int \prod_{t=1}^T \left[\frac{f(u_t, v_t, \lambda_t | U_{t-1}, V_{t-1}, \Lambda_{t-1}, \omega)}{m(\lambda_t | \Lambda_{t-1}, a_t)} \right] \prod_{t=1}^T m(\lambda_t | \Lambda_{t-1}, a_t) d\Lambda, \quad (8)$$

which can be evaluated by using N trajectories $\{\tilde{\lambda}_t^{(i)}(a_t)\}_{t=1}^T$ drawn from the importance sampler m by

$$\tilde{L}_N(\omega; U, V) = \frac{1}{N} \sum_{i=1}^N \left(\prod_{t=1}^T \left[\frac{f(u_t, v_t, \tilde{\lambda}_t^{(i)}(a_t) | U_{t-1}, V_{t-1}, \tilde{\Lambda}_{t-1}^{(i)}(a_{t-1}, \omega)}{m(\tilde{\lambda}_t^{(i)}(a_t) | \tilde{\Lambda}_{t-1}^{(i)}(a_{t-1}), a_t)} \right] \right). \quad (9)$$

The challenge of EIS is to find a function m and a sequence of auxiliary parameters $\{a_t\}_{t=1}^T$ to provide a good match between the numerator and the denominator in (9) in order to reduce the MC sampling variance of \tilde{L}_N as much as possible. The auxiliary parameters need to be estimated, which must be done for each period t due to the high dimensionality of the problem. A good match between f and m is not possible period by period since the integral of f with respect to λ_t depends on Λ_{t-1} , while that of m is equal to one. Therefore we need a functional approximation $k(\Lambda_t; a_t)$ for f that is analytically integrable with respect to λ_t such that

$$m(\lambda_t | \Lambda_{t-1}, a_t) = \frac{k(\Lambda_t; a_t)}{\chi(\Lambda_{t-1}; a_t)}, \quad (10)$$

where χ denotes the integral of k with respect to λ_t . Note that we must account for the function χ when matching f and k , but since χ does not depend on λ_t it can be transferred back into the subproblem for period $t-1$. Then for each period t given a value for ω the following minimization problem must be solved

$$\hat{a}_t = \arg \min_{a_t} \sum_{i=1}^N \left(\log[f(u_t, v_t, \tilde{\lambda}_t^{(i)}(\omega) | U_{t-1}, V_{t-1}, \tilde{\Lambda}_{t-1}^{(i)}(\omega), \omega) \cdot \chi(\tilde{\Lambda}_t^{(i)}(\omega); \hat{a}_{t+1})] - c_t - \log k(\tilde{\Lambda}_t^{(i)}(\omega); a_t) \right)^2, \quad (11)$$

for $t = T, \dots, 1$ and $\chi(\Lambda_T; a_{T+1}) \equiv 1$. Since the trajectories of the underlying dependence process are drawn from the natural sampler, one should iterate this procedure and use draws $\{\tilde{\lambda}_t^{(i)}(\hat{a}_t)\}_{t=1}^T$ from the importance sampler in the next iteration until convergence of \hat{a}_t to fixed values. This requires the use of Common Random Numbers (CRNs) at each iteration. Furthermore, if $k(\Lambda; a_t)$ is chosen within the exponential family the least squares problem in (11) becomes linear. The likelihood function is then evaluated by substituting the estimated sequence $\{\hat{a}_t\}_{t=1}^T$ and N draws from the importance sampler m into (9).

Liesenfeld and Richard (2003) suggest the following decomposition of k

$$k(\Lambda_t; a_t) = p(\lambda | \lambda_{t-1}, \omega) \zeta(\lambda_t, a_t), \quad (12)$$

with $\zeta(\lambda_t, a_t)$ a Gaussian kernel. This decomposition further simplifies the least squares problem and the functional form of the likelihood as the natural sampler p now cancels out. The choice $\zeta(\lambda_t, a_t) = \exp(a_{1,t}\lambda_t + a_{2,t}\lambda_t^2)$ makes the least squares problem in (11) linear and thus greatly reduces the computational burden of the procedure. For non-Gaussian latent processes, however, the importance sampler should be chosen in a different way, which will most likely not result in a linear least squares problem. Given this choice the mean and variance of the importance sampler m , which depend on $a_t = (a_{1,t}, a_{2,t})$, are

$$\mu_t = \sigma_t^2 \left(\frac{\alpha + \beta\lambda_{t-1}}{\nu^2} + a_{1,t} \right), \quad \sigma_t^2 = \frac{\nu^2}{1 - 2\nu^2 a_{2,t}}. \quad (13)$$

The functional forms of p , k and χ are given by Liesenfeld and Richard (2003). For a given choice of the copula EIS can be implemented as follows.

1. Draw N trajectories $\{\tilde{\lambda}_t^{(i)}(\omega)\}_{t=1}^T$ from the natural sampler p .
2. For $t = T, \dots, 1$ solve the back-recursive least-squares regression problem

$$\begin{aligned} \log c(u_t, v_t | \theta_t(\omega)) + \log \chi(\tilde{\lambda}_t^{(i)}(\omega); \hat{a}_{t+1}) \\ = c_t + a_{1,t} \tilde{\lambda}_t^{(i)}(\omega) + a_{2,t} [\tilde{\lambda}_t^{(i)}(\omega)]^2 + \eta_t^{(i)}, \end{aligned} \quad (14)$$

with c_t and $\eta_t^{(i)}$ the regression constant and error term, respectively.

3. Draw N trajectories $\{\tilde{\lambda}_t^{(i)}(\hat{a}_t)\}_{t=1}^T$ from the importance sampler m and solve the least-squares problem in Step 2 again. Iterate Steps 2 and 3 this until convergence of $\{\hat{a}_t\}_{t=1}^T$.
4. Draw N trajectories $\{\tilde{\lambda}_t^{(i)}(\hat{a}_t)\}_{t=1}^T$ from the importance sampler m and evaluate the likelihood function given in (9).

The maximum likelihood estimator of the parameter vector ω is then obtained by maximizing the EIS likelihood function. In order to ensure its smoothness, the same CRNs are used for every evaluation of the likelihood function. Note that more than five iterations are not necessary in most cases. Concerning the number of trajectories N , a choice between 100 and 200 seems to be sufficient to keep the Monte Carlo variation small.

2.3 Including the marginals: One Step vs. Two Step Estimation

So far we have ignored that the marginals must also be modeled and estimated. Generally speaking this depends highly on the type of data one wants to model. Since in this paper we are mainly focusing on financial data observed with a daily or weekly frequency, in particular stock market returns, it is crucial to properly model the time-varying volatility and leptokurtosis. A natural model for the i -th stock market returns r_{it} for $t = 1, \dots, T$ is the stochastic volatility (SV) model proposed by Clark (1973) and Taylor (1986). In its simplest form it can be written as

$$\begin{aligned} r_{it} &= \exp(h_{it}/2)\varepsilon_{it} \\ h_{it} &= \alpha_i + \beta_i h_{it-1} + \nu_i \eta_{it}, \end{aligned} \tag{15}$$

where ε_{it} and η_{it} are mutually independent i.i.d. normal random variables with mean zero and variance one that are also uncorrelated with the innovations driving the dependence process. Estimation is most easily done by EIS as described in Liesenfeld and Richard (2003).

An alternative to the SV model is the large class of GARCH models, introduced by Engle (1982) and Bollerslev (1986). The standard GARCH(1,1) model is given by

$$\begin{aligned} r_{it} &= \sqrt{h_{it}}\varepsilon_{it} \\ h_{it} &= \alpha_i + \gamma_i r_{it-1}^2 + \beta_i h_{it-1}. \end{aligned} \tag{16}$$

For either choice of the marginal model denote the parameter vector for component i by δ_i for $i = 1, 2$ and ω for the copula. Assume we observe the processes $X = \{x_t\}_{t=1}^T$ and $Y = \{y_t\}_{t=1}^T$ with marginal distributions $F(X; \delta_1)$ and $G(Y; \delta_2)$. Then, as a consequence of Sklar's theorem, the joint log-likelihood function can be decomposed into the marginal likelihood and the copula likelihood,

$$\mathcal{L}(\delta_1, \delta_2, \omega; X, Y) = \mathcal{L}_X(\delta_1; X) + \mathcal{L}_Y(\delta_2; Y) + \mathcal{L}_C(\omega; F(X; \delta_1), G(Y; \delta_2)). \tag{17}$$

In principle, (17) can be estimated w.r.t. all parameters to give a fully efficient maximum likelihood estimate whose covariance matrix is given by the inverse of the Fisher information matrix. However, joint estimation of all parameters is computationally very expensive in our situation, in particular when SV models are chosen for the margins, because in that case all three components of (17) would need to be evaluated jointly by an algorithm such as EIS. In some cases it may not even be possible to reach convergence at all. To solve this

problem, one may maximize the marginal log-likelihood functions in a first step, and then maximize the copula log-likelihood conditional on the estimated marginals. This method has been labeled the inference function for margins (IFM) estimator by Joe (1997, 2005). Under weak regularity conditions, this estimator is consistent, although not fully efficient for the copula parameters. Standard errors of the two step estimator $\hat{\omega}$ can be obtained as follows. Let us denote the full parameter vector as $\vartheta = (\delta'_1, \delta'_2, \omega')'$ and let $\psi(X, Y) = (\partial\mathcal{L}_X/\partial\delta'_1, \partial\mathcal{L}_Y/\partial\delta'_2, \partial\mathcal{L}/\partial\omega')'$ denote the inference functions. Furthermore, let us denote $D = E[\partial\psi(X, Y)/\partial\vartheta']$ and $M = E[\psi(X, Y)\psi(X, Y)']$. Then, as shown by Joe (2005), $\sqrt{T}(\hat{\vartheta} - \vartheta) \rightarrow_d N(0, V)$, where

$$V = D^{-1}M(D^{-1})'. \quad (18)$$

We will use this result for inference in our applications.

2.4 Examples of copulas

Below we give some possibilities for the choice of the (one parameter) copula density c along with appropriate transformations Ψ .

2.4.1 Frank Copula

The density of the Frank copula is given by

$$c_{Frank}(u, v; \theta) = \frac{\exp((1 + u + v)\theta)(\exp(\theta) - 1)\theta}{\{\exp(\theta) + \exp((u + v)\theta) - \exp(\theta + u\theta) - \exp(\theta + v\theta)\}^2}. \quad (19)$$

The fact that the parameter of the Frank copula lies in $(-\infty, \infty) \setminus 0$ makes it particularly attractive in our case, since we can choose $\Psi(x) = x$, which implies that the time-varying parameter θ_t itself follows a Gaussian AR(1) process. The Frank copula belongs to the family of Archimedean copulas and it implies a rotationally symmetric dependence structure with no tail dependence. It allows for both positive and negative dependence.

2.4.2 Clayton Copula

The density of the Clayton copula is

$$c_{Clayton}(u, v; \theta) = u^{(-1-\theta)}v^{(-1-\theta)}(u^{-\theta} + v^{-\theta} - 1)^{(-2-1/\theta)}(1 + \theta), \quad (20)$$

with $\theta \in (0, \infty)$. This suggests the transformation $\Psi(x) = \exp(x)$ and implies that the dependence parameter has a log-normal distribution. The Clayton copula is also

Archimedean, but it is not rotationally symmetric, has lower tail dependence and only allows for positive dependence.

2.4.3 Gumbel Copula

The density of the Gumbel copula is

$$c_{Gumbel}(u, v; \theta) = \frac{\{\log(u) \log(v)\}^{(\theta-1)} \{ [(-\log(u))^\theta + (-\log(v))^\theta]^{1/\theta} + \theta - 1 \}}{[(-\log(u))^\theta + (-\log(v))^\theta]^{(2-1/\theta)} uv} \times \exp \{ [(-\log(u))^\theta + (-\log(v))^\theta]^{1/\theta} \}, \quad (21)$$

with $\theta \in [1, \infty)$. An obvious choice for the transformation is $\Psi(x) = \exp(x) + 1$, which again implies log-normality of the dependence parameter. The Gumbel copula also belongs to the Archimedean class, it is asymmetric, has upper tail dependence and only allows for positive dependence.

2.4.4 Gaussian Copula

Denoting $x = \Phi^{-1}(u)$ and $y = \Phi^{-1}(v)$, where $\Phi(\cdot)$ is the cdf of a standard normal r.v., the density of the Gaussian copula is given by

$$c_{Gaussian}(u, v; \theta) = \frac{1}{\sqrt{1-\theta^2}} \exp \left(\frac{2\theta xy - x^2 - y^2}{2(1-\theta^2)} + \frac{x^2 + y^2}{2} \right), \quad (22)$$

with $\theta \in (-1, 1)$. For the transformation, we will use the inverse Fisher transform $\Psi(x) = (\exp(2x) - 1)/(\exp(2x) + 1)$. Note that this choice is natural as the Fisher transform is the variance stabilizing transformation for the correlation coefficient (see van der Vaart, 1998).

2.4.5 Survival (rotated) Copulas

Instead of considering the distribution of u and v one can also consider the copula of $1 - u$ and $1 - v$, which is known as the survival or rotated copula. Its density is the original density rotated by 180 degrees and thus the idea only makes sense for asymmetric copulas, which in our selection of candidate models are the Gumbel and the Clayton family. For a parametric copula C_θ the distribution function of the survival copula is given by $C_\theta(1 - u, 1 - v) + u + v - 1$, whereas its density is $c_\theta(1 - u, 1 - v)$.

2.5 Estimating the underlying process

Even though the parameters of the underlying process are of interest themselves, it is of crucial importance to obtain an estimate of the sequence of the underlying latent variable λ_t and of the function $\Psi(\lambda_t)$. To this end we need to compute the conditional expectation of $\Psi(\lambda_t)$ given the information on the past observable variables U_{t-1} and V_{t-1} . This is known as the filtered estimates of $\Psi(\lambda_t)$, which we denote $\Psi(\hat{\lambda}_{t|t-1})$ and is given by

$$E[\Psi(\lambda_t)|U_{t-1}, V_{t-1}] = \frac{\int \Psi(\lambda_t) p(\lambda_t | \Lambda_{t-1}, \omega) f(U_{t-1}, V_{t-1}, \Lambda_{t-1}; \omega) d\Lambda_t}{\int f(U_{t-1}, V_{t-1}, \Lambda_{t-1}; \omega) d\Lambda_{t-1}}. \quad (23)$$

Using the estimated parameters $\hat{\omega}$ the denominator is simply the likelihood function using the first $t-1$ observations, which is $L(\omega; U_{t-1}, V_{t-1})$. The numerator can be evaluated by

$$\frac{1}{N} \sum_{i=1}^N \left(\Psi[\tilde{\lambda}_t^{(i)}(\hat{\omega})] \prod_{t=1}^{t-1} \left[\frac{f(u_\tau, v_\tau, \tilde{\lambda}_\tau^{(i)}(\hat{a}_\tau) | U_{\tau-1}, V_{\tau-1}, \tilde{\Lambda}_{\tau-1}^{(i)}(\hat{a}_{\tau-1}, \hat{\omega}))}{m(\tilde{\lambda}_\tau^{(i)}(\hat{a}_\tau) | \tilde{\Lambda}_{\tau-1}^{(i)}(\hat{a}_{\tau-1}), \hat{a}_\tau)} \right] \right), \quad (24)$$

with $\{\tilde{\lambda}_\tau^{(i)}(\hat{a}_\tau)\}_{\tau=1}^{t-1}$ a trajectory from the EIS sampler for $L(\omega; U_{t-1}, V_{t-1})$ and $\tilde{\lambda}_t^{(i)}(\hat{\omega})$ a draw from the natural sampler $p(\lambda_t | \tilde{\Lambda}_{t-1}^{(i)}, \hat{\omega})$. As before CRN's should be used to evaluate the numerator and the denominator. The integral in (23) must be evaluated for each $t = 1, \dots, T$. In addition it needs to be mentioned that the number N of trajectories from the importance sampler must be a lot higher than for estimation purposes in order to ensure numerical stability at each evaluation of the likelihood function. We recommend a number of about 500, which means that evaluation of a filtered path for 2500 observations requires about 2-3 hours of computation on an Intel dual core processor.

An alternative to the computationally expensive method of filtering is to make use of the mechanics of EIS. As mentioned above, EIS exploits all the information available in the data to produce efficient samples of the underlying process $\{\lambda_t\}_{t=1}^T$. Thus as a byproduct of EIS when evaluating the likelihood function we obtain a smoothed estimate of $\{\lambda_t\}_{t=1}^T$ by

$$\hat{\lambda}_{t|T} = \frac{1}{N} \sum_{i=1}^N \tilde{\lambda}_t^{(i)}(\hat{a}_t) \quad \forall t = 1, \dots, T. \quad (25)$$

Using both simulated and real data, we obtained smoothed estimates of $\{\lambda_t\}_{t=1}^T$ that were very close to the filtered and true (in case of simulated data) paths of the underlying process. Furthermore, the smoothed estimates are much less noisy, due to the fact that they are calculated as an average, but also since they make efficient use of the complete

information contained in the sample at each $\hat{\lambda}_{t|T}$. Furthermore, it is computationally much cheaper and requires only a small N .

A third option is to calculate the smoothed estimate of λ_t using only the information available at time t , i.e. U_t and V_t . This estimate, which may be called the ‘updated’ estimate shall be denoted by $\hat{\lambda}_{t|t}$.

Since the main objective of time-varying correlation models is to estimate the correlation path over time, we conduct a small simulation study to see how competing methods for estimating time-varying correlation compare in the sense of being closer to the true correlation path. We draw a sample of size 1000 from a Gaussian copula model with a variety of underlying correlation dynamics. Four competing models are fit, the estimates for time-varying correlations are retrieved and mean square errors (MSE) are computed. The models are a constant copula, the stochastic copula autoregressive (SCAR) model² and the DCC GARCH model of Engle (2002), where correlation is described by

$$\begin{aligned}\rho_{i,j,t} &= \frac{q_{i,j,t}}{\sqrt{q_{i,i,t}q_{j,j,t}}} \\ q_{i,j,t} &= \bar{\rho}_{i,j} + \alpha(\varepsilon_{i,t-1}\varepsilon_{j,t-1} - \bar{\rho}_{i,j}) + \beta(q_{i,j,t-1} - \bar{\rho}_{i,j}),\end{aligned}\tag{26}$$

with $\bar{\rho}$ the unconditional sample correlation and $\varepsilon_{i,t}$ the standardized GARCH residual for variable i at time t . Note that in our simulation study we do not consider volatility dynamics, but only the model for correlation. The fourth model we consider is the conditional copula specification from Patton (2006), which is similar to the DCC model, but does not assume any marginal distribution. The conditional correlation in this model is given by

$$\rho_t = \Psi\left(\alpha + \beta_1\rho_{t-1} + \beta_2\frac{1}{10}\sum_{j=1}^{10}\Phi^{-1}(u_{t-j})\Phi^{-1}(v_{t-j})\right),\tag{27}$$

where Ψ is chosen to be the inverse Fisher transform. The correlations follow several processes, both deterministic and stochastic, which are:

1. Constant: $\rho_t = 0.5$
2. Jump: $\rho_t = 0.2 + 0.5I_{t>500}$
3. Sine: $\rho_t = 0.5 + 0.4\cos(2\pi t/200)$
4. Ramp: $\rho_t = \text{mod}(t/200)$

²The correlation estimate is the smoothed correlation path.

5. DCC: Correlation generated by (26) with $\bar{\rho} = 0.7$, $\alpha = 0.04$ and $\beta = 0.95$
6. AR(1): $\lambda_t = 0.03 + 0.97\lambda_{t-1} + 0.1\varepsilon_t$
7. Noise: $\lambda_t = 0.3 + 0.1\varepsilon_t$
8. Random Walk (RW): $\lambda_t = \lambda_{t-1} + 0.01\varepsilon_t$
9. ARMA(2,2): $\lambda_t = 0.01 + 0.65\lambda_{t-1} + 0.3\lambda_{t-2} + 0.1\varepsilon_t + 0.05\varepsilon_{t-1} + 0.03\varepsilon_{t-2}$

where ε_t is a $N(0, 1)$ random variable. For the DCC model the parameters were those estimated for the Dow Jones-NASDAQ data in the empirical application. For processes 6 to 9 we consider two transformations to keep correlation between -1 and 1 at all times: The first one is the inverse Fisher transform, whereas the second one is $\rho_t = \lambda_t / \sup_t |\lambda_t|$. The number of trajectories N in the EIS sampler is chosen to be 200. We repeat the simulation 1000 times for the constant copula and the DCC model and only 100 times for the SCAR and Patton (2006) model due to computational complexity. We report the average of the mean squared distance between the true and the estimated correlation path. Results, which are reported in Table 1, show that the SCAR model clearly outperforms its competitors both under deterministic and stochastic correlation dynamics and regardless of the transformation. Not surprisingly, all models do worse when they are misspecified. The DCC only performs better when it is also the data generating process, which is not surprising as in that case the innovations driving the correlation are known. Still, the SCAR model does well and outperforms the specification of Patton (2006).

Table 1: Average MSEs for correlation estimates

Model/DGP	constant	jump	Sine	ramp	DCC	AR(1)	Noise	RW	ARMA(2,2)
SCAR	0.0005	0.0066	0.013	0.0261	0.0041	0.0087	0.0104	0.0081	0.0556
DCC	0.001	0.0134	0.0318	0.0458	0.0012	0.0209	0.0104	0.009	0.1096
Patton	0.0019	0.0416	0.053	0.0589	0.0062	0.0215	0.0134	0.0119	0.1193
Constant	0.0004	0.0631	0.0806	0.0838	0.0111	0.0356	0.0093	0.0154	0.1909
Alternative transformation									
SCAR						0.0173	0.0267	0.0128	0.0402
DCC						0.0282	0.0268	0.0247	0.0714
Patton						0.0315	0.0277	0.0437	0.0701
Constant						0.035	0.0259	0.0972	0.1025

Note: Average mean square errors of time-varying correlation estimates. In the top panel the underlying process λ_t is re-scaled by the inverse Fisher transform, whereas in the bottom panel the transformation $\rho_t = \lambda_t / \sup_t |\lambda_t|$ was chosen.

2.6 Testing

We now consider two hypotheses that may be of interest in empirical modeling using the stochastic copula model. The first is whether the dependence parameter is actually time-varying. Formally, the null hypothesis can be written as

$$H_0 : \theta_t = \bar{\theta}, \quad \forall \quad t = 1, \dots, T, \quad (28)$$

where $\bar{\theta}$ is the time-constant copula parameter. In terms of our model parameters in (3) this null hypothesis can also be stated as

$$H_0 : \nu = 0. \quad (29)$$

We test this hypothesis with a simple likelihood ratio test. Let LL_{res} be the log-likelihood of the model under H_0 and LL_{ur} the log-likelihood of the unrestricted model. The test statistic is

$$LR_C = -2(LL_{res} - LL_{ur}). \quad (30)$$

Since β is unidentified under the null and furthermore ν is on the boundary of the parameter space the asymptotic distribution of LR_C is non-standard. However, we can obtain approximate critical values by simulation. These critical values, which are approximately 4.65, 6.44 and 9.99, at 10%, 5% and 1% level of significance, respectively, are close to those of a χ^2 distribution with 2 degrees of freedom.

The second important hypothesis is that of the correct choice of the copula. The most simple, but still extremely reliable way of choosing the best fitting copula is to compare the value of the log-likelihood at the parameter estimates or the Akaike Information Criterion (AIC) when competing models have the different numbers of parameters. Although this works quite well, we have no guarantee that the model chosen in such a way fits the data well. There are various approaches to the problem of goodness-of-fit (GoF) testing in copulas, see e.g. Genest and Rivest (1993), Genest et al. (2006), Chen and Fan (2006), Junker et al. (2006), Patton (2006), or for a comparison of some tests Genest et al. (2008). Let $C_i(u_t, v_t, \hat{\theta}_t)$ be our candidate copula with estimated parameter $\hat{\theta}_t$ at time t and let $C_0(u_t, v_t, \theta_t^0)$ be the true copula where θ_t^0 denotes the true parameter at time t . Our hypothesis is

$$H_0 : C_i(u_t, v_t, \hat{\theta}_t) = C_0(u_t, v_t, \theta_t^0). \quad (31)$$

As an estimator of θ_t we consider the smoothed paths $\hat{\theta}_{t|T}$ resulting from the importance sampler. Concerning the choice of the goodness-of-fit test many of the tests proposed in the papers cited above such as e.g. the bivariate χ^2 test, are not suitable for time-varying copula models. A class of tests that is easily adaptable for the time-varying case is based on the fact that the conditional copula, i.e. the copula of u given v (or of v given u), is uniformly distributed, which is an application of the Rosenblatt transformation. In our case this means

$$z_t = C_0(u_t|v_t, \theta_t^0) = \frac{\partial C_0(u_t, v_t, \theta_t^0)}{\partial v_t} \sim U(0, 1). \quad (32)$$

Testing the copula specification therefore means testing whether $\hat{z}_t = C(\hat{u}_t|\hat{v}_t, \hat{\theta}_{t|T})$ has a $U(0, 1)$ distribution. For static copulas a closely related test has been considered in Breymann et al. (2003) and studied via Monte Carlo simulations in Dobric and Schmid (2007). For testing the uniformity of \hat{z}_t , various tests are available and we will consider the Kolmogorov-Smirnov (KS) test, the χ^2 test, the Anderson-Darling (AD) test and the Jarque-Bera (JB) test for normality, for which we need to apply the transformation $\Phi^{-1}(\hat{z}_t)$. Let the empirical distribution function of \hat{z}_t be

$$\hat{\mathbb{F}}(x) = \frac{1}{T} \sum_{t=1}^T \mathbf{I}_{\{z_t < x\}} \quad (33)$$

and let the CDF under the null hypothesis be $F(x)$. Then the KS test statistic is defined as

$$T_{KS} = \sup_x |F(x) - \hat{\mathbb{F}}(x)|, \quad (34)$$

for which critical values have been tabulated. In practice, the supremum is replaced by the maximum over the observations. For the χ^2 test, consider splitting the domain $(0, 1)$ in k bins and let c_i be the number of observations in bin i . Then the statistic of interest is

$$T_{\chi^2} = \sum_{i=1}^k \frac{(E(c_i) - c_i)^2}{E(c_i)} \sim \chi^2(k), \quad (35)$$

where the expectations of c_i are taken under the null model. Next, let S be sample skewness and K be the sample kurtosis of $\Phi^{-1}(\hat{z}_t)$. Then the JB test is

$$T_{JB} = \frac{T}{6} \left(S^2 + \frac{(K - 3)^2}{4} \right) \sim \chi^2(2). \quad (36)$$

Finally, the AD, which is a refinement of the KS test that is suitable to test deviations in the tails of the distribution, is given by

$$T_{AD} = \sup_x \frac{\sqrt{T} |\hat{F}(x) - F(x)|}{\sqrt{F(x)(1 - F(x))}}, \quad (37)$$

for which again tabulated critical values are used.

2.7 Forecasting

A big advantage of specifying the underlying dependence process by an AR(1) structure is that it allows for easy forecasting. In contrast to the DCC model we can compute r-step ahead forecasts without making any approximations as outlined in Engle and Sheppard (2001). The techniques for forecasting AR(1) processes are standard and can be found in e.g. Hamilton (1994). With an estimate $\hat{\lambda}_T$, for which the smoothed estimate $\hat{\lambda}_{T|T}$ is suitable, the r-step ahead forecast of λ is given by

$$\hat{\lambda}_{T+r} = \mu + \beta^r (\hat{\lambda}_T - \mu), \quad (38)$$

where $\mu = \alpha/(1 - \beta)$. The mean squared r-period-ahead forecast error for λ is

$$\sigma_{T+r}^2 = \nu^2(1 - \beta^{2r})/(1 - \beta^2). \quad (39)$$

Unfortunately, only in the case of the Frank copula are we interested in forecasting λ_t itself, but generally we want to forecast a nonlinear transformation thereof. In the case of the Clayton and Gumbel copulas we can use the following results. For $\lambda_t | \mathcal{F}_{t-1} \sim N(\mu_t, \sigma_t)$ the 1-step ahead forecast of $\theta_t = \exp(\lambda_t)$ is

$$E(\theta_t | \mathcal{F}_{t-1}) = \exp\left(\mu_t + \frac{\sigma_t^2}{2}\right). \quad (40)$$

For μ_t , the conditional expectation of λ_t given \mathcal{F}_{t-1} , we insert its linear forecast given in (38). From this it follows by straightforward calculations that the r-step ahead forecast of θ is

$$\hat{\theta}_{T+r} = \exp\left(\hat{\lambda}_{T+r} + \frac{\sigma_{T+r}^2}{2}\right). \quad (41)$$

The confidence bands for these forecasts can be used by using the corresponding quantiles of the log-normal distribution with parameters $\hat{\lambda}_{T+r}$ and σ_{T+r}^2 .

Forecasting the correlation coefficient of the normal copula is not as straightforward due the nonlinearity of the inverse Fisher transform. We therefore use a Taylor approximation of $\Psi(\lambda_t)$ around μ_t ,

$$\Psi(\lambda_t) = \Psi(\mu_t) + \Psi'(\mu_t)(\lambda_t - \mu_t) + \frac{1}{2}\Psi''(\mu_t)(\lambda_t - \mu_t)^2. \quad (42)$$

Taking the conditional expectation we have

$$\begin{aligned} E[\Psi(\lambda_t)|\mathcal{F}_{t-1}] &= E\left\{\Psi(\mu_t) + \Psi'(\mu_t)(\lambda_t - \mu_t) + \frac{1}{2}\Psi''(\mu_t)(\lambda_t - \mu_t)^2|\mathcal{F}_{t-1}\right\} \\ &= \Psi(\mu_t) + \Psi'(\mu_t)E((\lambda_t - \mu_t)|\mathcal{F}_{t-1}) + \frac{\Psi''(\mu_t)}{2}E((\lambda_t - \mu_t)^2|\mathcal{F}_{t-1}) \\ &= \Psi(\mu_t) + \frac{\Psi''(\mu_t)}{2}\nu^2. \end{aligned} \quad (43)$$

Then the r-step ahead forecast is

$$\hat{\theta}_{T+r} = \Psi(\hat{\lambda}_{T+r}) + \frac{-4(\exp(2\hat{\lambda}_{T+r}) - 1)\exp(2\hat{\lambda}_{T+r})}{(\exp(2\hat{\lambda}_{T+r}) + 1)^3}\sigma_{T+r}^2. \quad (44)$$

Of course, higher order approximations could be used, but we did not find any substantial differences in our applications. Confidence bands are obtained by applying the inverse Fisher transform to the corresponding quantiles of the normal distribution.

3 Application

3.1 Daily data: Dow Jones and NASDAQ

The first data set we consider to illustrate our model and compare it to competing models are daily observations of the Dow Jones Industrial Average and the NASDAQ composite ranging from March 26, 1990 until March 23, 2000. The same data set has been considered in Engle (2002). Returns are calculated as the first difference of the natural logarithm and are multiplied by 100. In a first estimation step, a stochastic volatility (SV) model is fit to the demeaned returns.³ The SV model is estimated by EIS as described in Liesenfeld and Richard (2003). As a comparison we also estimate a standard GARCH(1,1) model. Parameter estimates and the values of the maximized log-likelihood function are given in Table 2, and the GARCH and smoothed SV volatilities can be found in Figure 1.

³An AR(1) model for the conditional mean was also considered, but estimates for the volatility and dependence models were almost identical.

Table 2: Estimates of GARCH and SV models: Dow Jones and Nasdaq

GARCH	DJ	NQ	SV	DJ	NQ
α	0.006 (0.0044)	0.0308 (0.0173)	α	-0.0112 (0.0051)	-0.0054 (0.0042)
γ	0.048 (0.0184)	0.115 (0.0435)	β	0.9786 (0.0076)	0.9733 (0.0078)
β	0.945 (0.0214)	0.863 (0.0529)	ν	0.1573 (0.0254)	0.2072 (0.0266)
logl	-3195.38	-3693.89	logl	-3137.37	-3638.03

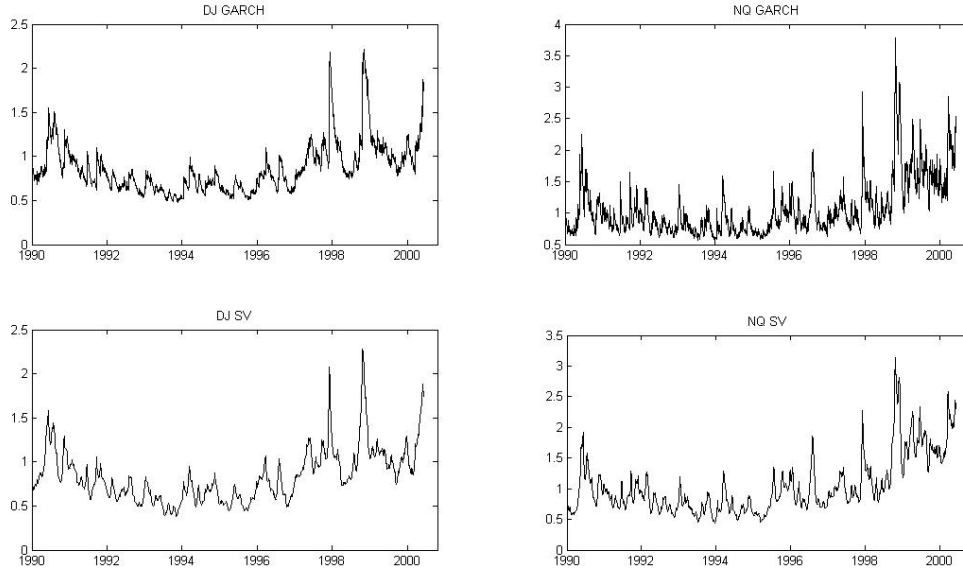
Not surprisingly, the SV model provides a better fit than the standard GARCH model due to its higher flexibility, and the GARCH model estimates a slightly higher degree of persistence. However, the estimated volatility series look very similar.

Using the estimated standard deviation, the data is transformed into $U(0, 1)$ random variables using the probability integral transform, and the stochastic copula model is estimated.⁴ In order to assess the stability of our estimation procedure, we also estimate the model in one-step to obtain fully efficient estimates using the estimates of the two-step estimation as starting values. For the one-step estimator, standard errors are obtained by the inverse of the numerically evaluated Hessian. Standard errors for the two step estimates are obtained by evaluating (18) using numerical derivatives. Table 3 presents the results of the estimation. In terms of the value of log-likelihood, the Gaussian copula is the best fitting model, followed by the survival Gumbel copula. This is confirmed by the outcomes of the GoF-tests reported in Table 4.

The Gaussian copula estimated in one step is the only model that passes all the four tests. Note that the likelihood-ratio tests for constancy were also performed and led to a rejection of the null of constant dependence with p-values being essentially zero for all models. The first thing to observe is that for all models the dependence process is highly persistent, which is in line with findings in earlier studies. Next, it is surprising that the Gaussian copula outperforms the two asymmetric models for which losses have a higher degree of dependence than large returns, even though asymmetric models are

⁴We choose N , the number of simulated trajectories in the EIS sampler, equal to 500 in order to eliminate the Monte Carlo variation. A much smaller number such as 200 gives almost exactly the same estimates.

Figure 1: Volatility estimates: Dow Jones and Nasdaq



preferred when considering a static copula model⁵. It seems that to some extent, this asymmetry is accounted for by the time-varying dependence parameter. Furthermore, the asymmetric models, in particular the Clayton copula, may simply underestimate the dependence for larger observations and this may outweigh the advantage of allowing for lower tail dependence. Also note that our bivariate Gaussian copula model does not imply that returns are normally distributed, since time-varying volatilities imply fat tails in the marginal distributions. Moreover, the Gaussian copula with dynamic stochastic correlations implies a non-Gaussian copula of the bivariate stationary distribution that is likely to have near tail dependence⁶. Comparing one-step and two-step estimates, one can see that they do not differ substantially. The standard errors for the two-step estimates are slightly higher, as expected. Only for the rotated Gumbel copula they are about equal, which could be attributed to numerical imprecisions. In Figure 2, the smoothed estimate of the copula parameter is shown. It also includes the correlation path from a DCC model as a comparison. The path of the dependence parameters looks quite similar for all stochastic copula models, although the scale is different. The DCC correlation is a

⁵Estimation results for static copulas are available upon request

⁶Unreported simulation results show that a Gaussian copula with time-varying correlation has more similarities with a t-copula than with a static Gaussian copula.

Table 3: Estimates of the stochastic copula model: Dow Jones and Nasdaq.

copula/logl	α	β	ν	α	β	ν
Normal	0.0302	0.9679	0.0824	0.0261	0.972	0.0754
902.169	(0.0099)	(0.0103)	(0.0147)	(0.0089)	(0.0093)	(0.0135)
rot. Gumbel	-0.0015	0.9795	0.1078	-0.0013	0.9786	0.1111
866.7934	(0.0023)	(0.0084)	(0.0236)	(0.0024)	(0.0085)	(0.0235)
Frank	0.1527	0.975	0.5457	0.1422	0.9768	0.5229
790.9929	(0.0613)	(0.0098)	(0.1228)	(0.0537)	(0.0086)	(0.1103)
Clayton	0.0136	0.9611	0.1549	0.0154	0.9636	0.1445
752.0481	(0.0071)	(0.0175)	(0.0407)	(0.0067)	(0.0137)	(0.0319)

Note: Estimation results for the stochastic copula models with a Gaussian SV models as marginal distribution. Two-step estimation results in columns 2 to 4 and one-step estimation results in columns 7 to 9.

bit noisier, which is due to the fact that it is based on one-step ahead forecasts.

Next we compare the in-sample-fit of our stochastic copula autoregressive (SCAR) model to two competing models, the DCC model of Engle (2002) and the conditional Gaussian copula model of Patton (2006). The models are compared both in terms of the log-likelihood of the copula, where the margins were chosen to be normal GARCH(1,1) models, and of the full model, where we chose SV margins for the Patton (2006) and the SCAR model. Both the DCC and the Patton (2006) likelihoods are evaluated using the estimated correlation paths, whereas the correlation process is integrated out of the SCAR likelihood. However, the likelihoods of the competing models should also be compared using all the information available at time t . To this end we also evaluate the likelihood functions of the SCAR and the SV models using $\hat{\lambda}_{t|t}$, the smoothed estimate of λ_t using only the information available until time t . In terms of in-sample-fit, the SCAR model outperforms its competitors. Using the updated estimates of the latent process the difference in the likelihoods increases.

Finally, we use the techniques described in Section 2.7 to obtain out-of-sample forecasts of the copula dependence process. We use the last observation of the smoothed dependence

Table 4: Goodness-of-Fit tests: Dow Jones and Nasdaq.

Two step	Normal	rot Gumbel	Frank	Clayton
KS	0.213	0.0046	0.0777	0.0001
χ^2	0.0012	0	0.0001	0
JB	0.0348	0	0	0
AD	0.0303	0	0	0
One step	Normal	rot Gumbel	Frank	Clayton
KS	0.2105	0.0043	0.0996	0.0001
χ^2	0.0527	0	0	0
JB	0.0913	0	0	0
AD	0.0584	0	0	0

Note: P-values for the null hypothesis of correct specification of the copula function for the two-step (top panel) and one-step estimation (bottom panel) of the Dow Jones and Nasdaq data.

Figure 2: Time path of the dependence parameters: Dow Jones and Nasdaq

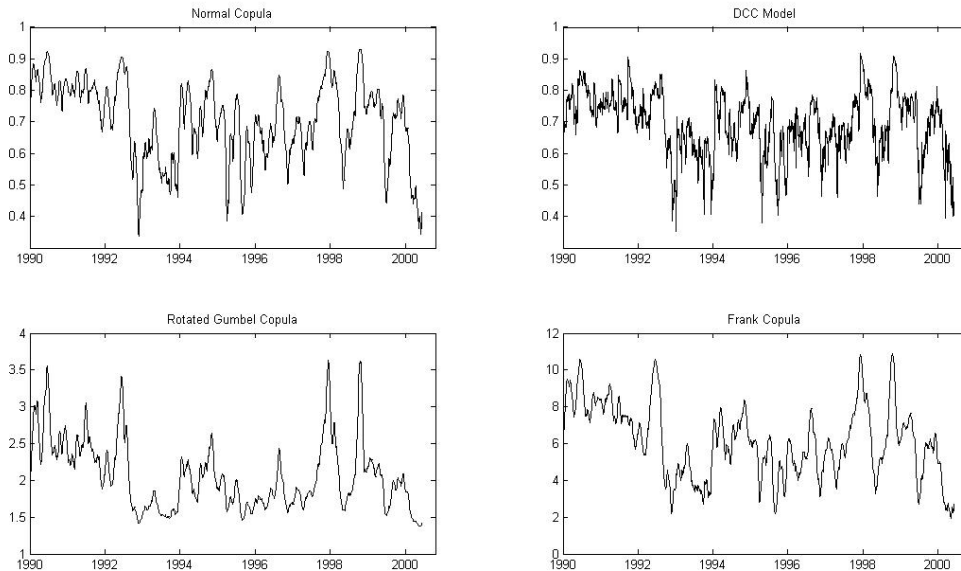


Table 5: Comparison of in-sample-fit: Dow Jones and Nasdaq.

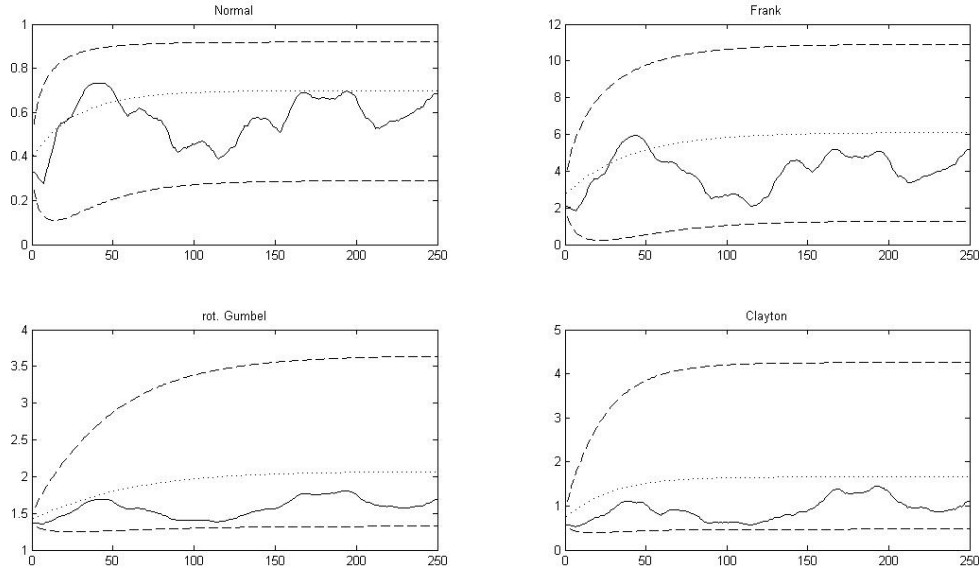
	Copula	Full model
SCAR	945.48	-5867.05
SCAR($\hat{\lambda}_{t t}$)	1058.70	-5501.20
DCC	938.2	-5951.62
Patton	890.2069	-5937.43
Patton($\hat{\lambda}_{t t}$)	-	-5769.50

Note: Log-likelihood of the Gaussian copula using GARCH(1,1) margins (left column) and of the full bivariate model (right columns). Margins in the full models are normal SV models, except for the DCC, where they are GARCH(1,1). $\hat{\lambda}_{t|t}$ indicates that the log-likelihood is evaluated using the smoothed estimate of the latent process with information until time t .

process as our initial observation and forecast over a horizon ranging from 1 to 250 trading days. As the dependence process is not observable, it needs to be estimated to check the performance of the forecasts. This is done by re-estimating the model using 250 additional observations and computing the smoothed path of the dependence parameter. Figure 3 presents the forecasts together with 95% confidence bands and the smoothed path. It is noteworthy that the confidence band for the Normal, rotated Gumbel and Clayton models are asymmetric, thus taking the distributional assumptions on the dependence process into account. Although they are on different scales, their width is comparable when measured in terms of Kendall's tau.⁷ After about 100 days, the forecast distribution corresponds to the stationary distribution. Thus, the width of the forecast bands at long horizons corresponds to the large variation of the dependence paths in Figure 2. Also note that the 1-day ahead forecasts are surprisingly far off from the realization of the path. This is due to the fact that the complete sample has been re-estimated using additional data, which changed the estimates of the dependence process, in particular near the end of the original sample. This could be avoided comparing r -step ahead forecasts only with realizations of the process using $T + r$ observations, where T is the size of the original sample. The forecasts seem to be reasonably precise considering the difficult task

⁷Kendall's tau measures the probability that two pairs of observations are concordant minus the probability that they are discordant and is a popular dependence measure for copulas.

Figure 3: Out-of-sample forecasts of the dependence parameter: Dow Jones and Nasdaq



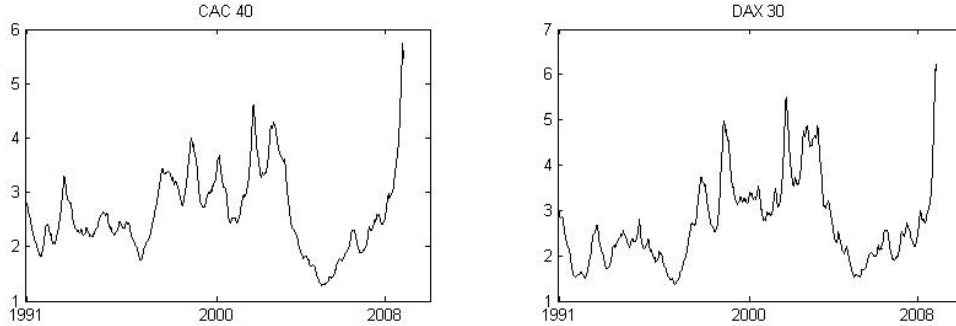
Note: The dotted lines denote the out-of-sample forecasts $\hat{\theta}_{T+r}$, the dashed lines are the 95% confidence intervals and the solid lines are the smoothed estimates $\hat{\theta}_{t|T}$.

of forecasting unobserved dependence parameters and the realizations stay within the confidence bands at all times.

3.2 Weekly data: CAC and DAX

For the second application, we consider stock index returns observed at a weekly frequency. The series are the French CAC 40 and the German DAX 30 from November 26, 1990 until November 3, 2008 and thus also cover the recent financial crisis. As above, a SV model is chosen for the margins and the SCAR model with different choices of copulas is considered for the dependence model. Only the two-step estimator is used as the estimates are quite close to those using one-step estimation, but numerical problems made it impossible to obtain standard errors. Volatility estimates can be found in Figure 4. The high levels of volatility during recent financial turmoils, but also around the 9/11 terrorist attacks are notable. Estimates for the SCAR model are reported in Table 6 and Figure 5. While

Figure 4: Volatility estimates: CAC and DAX



in a static copula framework, a model with excess upper tail dependence dominates⁸, in the time-varying case the Gaussian copula is again clearly the best fitting model, both when looking at the log-likelihood and the GoF tests. Similar to the example above, the SCAR model fits slightly better than the DCC model and clearly outperforms the conditional copula model by Patton (2006). The estimates for the persistence parameter differ quite significantly across different copulas. The symmetric models, namely the Gaussian and Frank copulas, show very high persistence. In fact, the Frank copula attains the upper bound in the constrained optimization, so its dependence process is likely to be integrated. Differences can also be seen from the time paths of the dependence parameters in Figure 5. This shows that choosing different copulas does not necessarily just result in similar shapes of the dependence process with differently scaled dependence parameters, but that an asymmetric model may in fact imply different dynamics over time. It is also encouraging that the SCAR model is able to capture movements of the dependence parameter that resemble trends or regime shifts without explicitly including such features in the model. Finally, forecasts of a 50 week horizon differ across the different models, although comparison is rather difficult due to the differences in the scales and underlying models of the parameters.

⁸This is rather surprising, as excess lower tail dependence for stock market returns is almost a stylized fact and one of the main motivations to use copulas. Results for the static copula estimation are available from the authors upon request.

Figure 5: Time path of the dependence parameters: CAC and DAX

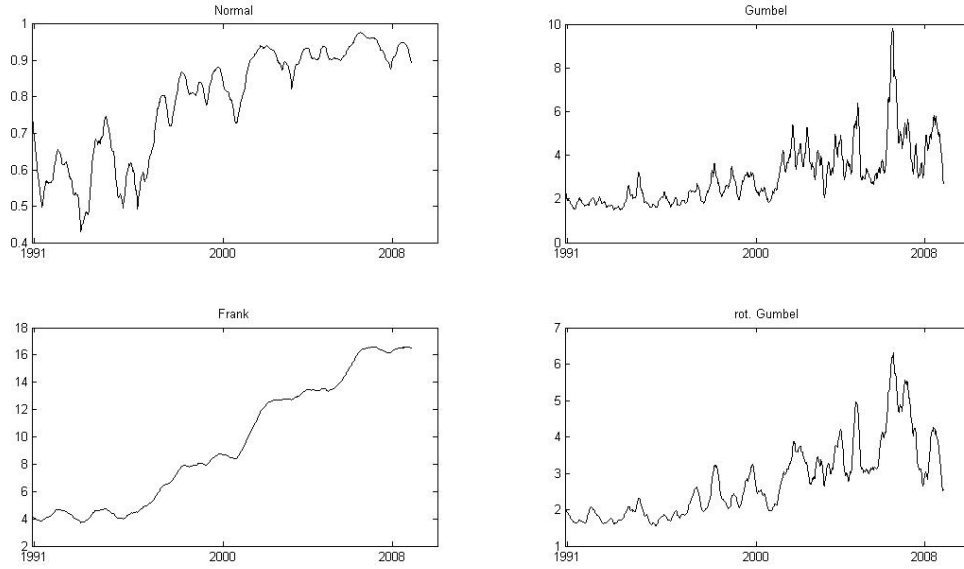


Figure 6: Out-of-sample forecasts of the dependence parameter: CAC and DAX

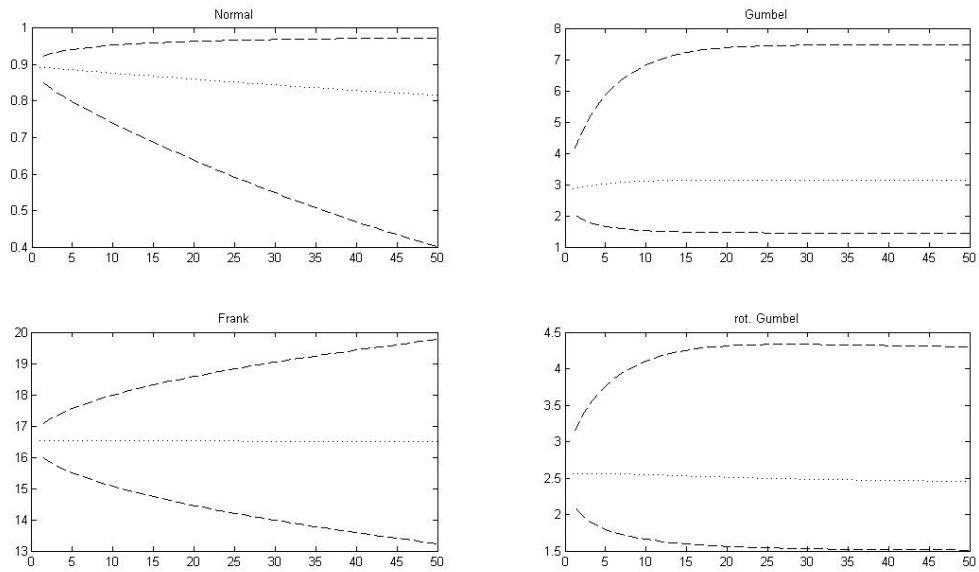


Table 6: Estimation and GoF for CAC and DAX returns.

	Normal	Gumbel	Frank	rot. Gumbel
logl	555.36	526.11	514.69	495.32
α	0.0095	0.0425	0.0002	0.0143
β	0.9908	0.9194	0.9999	0.9432
ν	0.0772	0.2681	0.2359	0.1581
GoF				
KS	0.8076	0.3352	0.1134	0.0233
χ^2	0.0583	0.0408	0.4775	0.0444
JB	0.4209	0.0126	0.0286	0.001
AD	0.086	0.0014	0.0307	0

Note: Two-step estimation results of the SCAR models and p-values for the null hypothesis of correct specification of the copula function for the weekly CAC and DAX returns.

4 Conclusions

We have proposed a stochastic copula model with a latent stochastic process driving the copula parameter. The model is discussed in various respects concerning specification, estimation, testing and forecasting. A simulation study compares the performance of the stochastic copula autoregressive (SCAR) model with that of the DCC model of Engle (2002) and of the Patton (2006) model for alternative scenarios. In cases where all considered models are misspecified, the SCAR model clearly outperforms its competitors. In an empirical application we considered two index series, one on a daily, the other on a weekly frequency. In most cases, the SCAR model based on the Gaussian copula fits the data well and again outperforms the DCC and Patton (2006) models. Out-of-sample forecasts of the dependence parameter confirm the usefulness of the model.

One extension of the model would be to allow for several copula parameters to depend on the latent variable. This would further complicate the estimation problem, but should still be feasible. An interesting approach would be to define the copula as a mixture of two basic copulas such as Clayton and Gaussian with fixed parameters, and then letting the mixture parameter depend on the latent variable. This seems promising as it combines

the flexibility of copula mixtures with a time-varying dependence framework, and we leave it for future research. Another extension would be to consider higher dimensional copulas. One could either consider one separate process for each pair of variables or achieve some dimension reduction by letting the time-varying correlation be driven by a limited number of common factors that are modeled as stochastic processes. Finally, one could extend the model to allow for exogenous variables driving the correlation dynamics. A possibility would be to include trade volume, which could serve as a proxy for the impact of volatility on correlation. Another option would be to include a function of time so as to capture trends or regime shifts in the dependence parameters.

Acknowledgements

We would like to thank Franz Palm, Johan Segers, Jean-Pierre Urbain and the participants of the Maastricht University econometrics seminar for comments and suggestions. The second author gratefully acknowledges financial support from METEOR.

References

- [1] Amisano, G. and R. Casarin (2007), "Particle filters for Markov Switching Stochastic Correlation Models," Proceedings of the 2007 Intermediate Conference "Risk and Prediction", Cleup, Padua, pp. 305-316.
- [2] Asai, M. and M. McAleer (2009), "The structure of dynamic correlations in multivariate stochastic volatility models," *Journal of Econometrics*, forthcoming.
- [3] Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics* 31, 307-327.
- [4] Breymann, W, A. Dias and P. Embrechts (2003), "Dependence structures for multivariate high-frequency data in finance," *Quantitative Finance*, 3(1), 1-14.
- [5] Broto, C. and E. Ruiz (2004), "Estimation methods for stochastic volatility models: A survey," *Journal of Economic Surveys*, 18(5), 613-649.
- [6] Carnero, M.A., D. Peña and E. Ruiz (2004), "Persistence and Kurtosis in GARCH and Stochastic Volatility Models," *Journal of Financial Econometrics*, 2, 319-342.

- [7] Chen, X. and Y. Fan (2006), "Estimation and Model Selection of Semiparametric Copula-based Multivariate Dynamic Models under Copula Misspecification," *Journal of Econometrics*, 135, 125-154.
- [8] Clark, P.K. (1973), "A subordinate stochastic process model with finite variance for speculative prices", *Econometrica*, 41, 135-155.
- [9] Danielsson, J. (1994), "Stochastic volatility in asset prices estimation with simulated maximum likelihood," *Journal of Econometrics*, 64(1-2), 375-400.
- [10] Danielsson, J. (1998), "Multivariate stochastic volatility models: Estimation and comparison with VGARCH models," *Journal of Empirical Finance*, 5, 155-173.
- [11] Dobric, J. and F. Schmid (2007), "A goodness of fit test for copulas based on Rosenblatt's transformation," *Computational Statistics and Data Analysis*, 51, 4633-4642.
- [12] Engle, R. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation", *Econometrica*, 50, 987-1008.
- [13] Engle, R. (2002), "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models," *Journal of Business and Economic Statistics*, 20(3), 339-350.
- [14] Engle, R. and K. Sheppard (2001), "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH," NBER Working Paper 8554, National Bureau of Economic Research.
- [15] Genest, C. and L.-P. Rivest (1993), "Statistical Inference Procedures for Bivariate Archimedean Copulas," *Journal of the American Statistical Association*, 88(423), 1034-1043.
- [16] Genest, C., J.-F. Quessy and B. Rémillard (2006), "Goodness-of-fit procedures for copula models based on the probability integral transformation," *Scandinavian Journal of Statistics*, 33, 337-366.
- [17] Genest, C., B. Rémillard and D. Beaudoin (2008), "Goodness-of-fit tests for copulas: A review and a power study," *Insurance: Mathematics and Economics*, forthcoming.
- [18] Hamilton, J. (1994), *Time Series Analysis*. Princeton: Princeton University Press.

- [19] Harvey, A. E. Ruiz and N. Shephard (1994), "Multivariate stochastic variance models," *Review of Economic Studies*, 61, 247-264.
- [20] Jacquier, E., N.G. Polson and P.E. Rossi (1994), "Bayesian analysis of stochastic volatility models (with discussion)," *Journal of Business and Economic Statistics*, 12, 371-389.
- [21] Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Chapman & Hall/CRC.
- [22] Joe, H. (2005), "Asymptotic efficiency of the two-stage estimation method for copula-based models," *Journal of Multivariate Analysis*, 94(2), 401-419.
- [23] Junker, M., A. Szimayer and N. Wagner (2006), "Nonlinear term structure dependence: Copula functions, empirics, and risk implications," *Journal of Banking and Finance*, 30, 1171-1199.
- [24] Kim, S., N. Shephard and S. Chib (1998), "Stochastic volatility: Likelihood inference and comparison with ARCH models," *Review of Economic Studies*, 65, 361-393.
- [25] Liesenfeld, R. and J.F. Richard (2003), "Univariate and multivariate volatility models: Estimation and diagnostics," *Journal of Empirical Finance*, 10, 505-531.
- [26] Liesenfeld, R. and J.F. Richard (2006), "Classical and Bayesian Analysis of Univariate and Multivariate Stochastic Volatility Models," *Econometric Reviews*, 25(2-3), 335-360.
- [27] Patton, A. (2006), "Modelling Asymmetric Exchange Rate Dependence," *International Economic Review*, 47(2), 527-556.
- [28] Richard, J.F. and W. Zhang (2007), "Efficient high-dimensional importance sampling," *Journal of Econometrics*, 141(2), 1385-1411.
- [29] Sandmann, G. and S.J. Koopman (1998), "Estimation of stochastic volatility models via Monte Carlo maximum likelihood," *Journal of Econometrics*, 87(2), 271-301.
- [30] Taylor, S.J. (1986), *Modelling Financial Time Series*, JohnWiley and Sons, Chichester.
- [31] van der Vaart, A.W. (1998), *Asymptotic Statistics*, Cambridge University Press.

- [32] Yu, J. and R. Meyer (2006), "Multivariate stochastic volatility models: Bayesian estimation and model comparison," *Econometric Reviews*, 25(2-3), 361-384.