JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **27**: 269–295 (2012)
Published online 1 July 2010 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.1197

DYNAMIC STOCHASTIC COPULA MODELS: ESTIMATION, INFERENCE AND APPLICATIONS

CHRISTIAN M. HAFNER[a]* AND HANS MANNER[b]

[a] *Institut de Statistique and CORE, Université Catholique de Louvain, Louvain-la-Neuve, Belgium*
[b] *Chair of Statistics and Econometrics, University of Cologne, Cologne, Germany*

SUMMARY

We propose a new dynamic copula model in which 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. 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.

Received 15 July 2009; Accepted 18 April 2010

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, for example, Jacquier *et al*. (1994), Danielsson (1998), Kim *et al*. (1998), and Carnero *et al*. (2004) for comparisons between the two model classes. With the same number of parameters, or fewer, the SV models gain additional flexibility through the use of stochastic latent variables that drive volatilities. Until quite recently, SV models have suffered from a difficult estimation problem, due to the fact that 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, for example, 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 order 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 *et al*. (1994), multivariate SV models have become more popular in empirical finance to describe return volatilities and correlations, with direct applications, for example, to portfolio selection. Yu and Meyer (2006) are, to our knowledge, the first 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 (2007) suggest the

* Correspondence to: Christian M. Hafner, Institut de Statistique and CORE, Université Catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium. E-mail: christian.hafner@uclouvain.be

introduction of 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 equity returns, where the lower tail dependence is usually higher than upper tail dependence.

A natural, and in fact the most general way to model nonlinear dependencies is to use copulas; see, for example, the monograph of Joe (1997). A broad variety of copula models exists which allow the capture of, for example, asymmetries in tail dependencies. One advantage of using copulas is the decoupling of marginal distributions from the dependence. In many cases, the marginal distributions do not depend on the dependence parameter, such that one can first estimate the parameters of the marginal distributions, e.g. assuming univariate SV processes, and then in a second step the copula parameters. Furthermore, in situations where the marginal distribution cannot be assumed to be conditionally normal, the copula approach allows for the construction of new multivariate distributions by coupling the appropriate marginal distributions using a copula. A remarkable extension of copula theory has been made by Patton (2006), who introduced the notion of a conditional copula and proposed a model for the evolution of time-varying dependence parameters. As an alternative to Patton (2006), we propose a model in which one or more of the copula parameters follow 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 the better capture of asymmetric tail dependencies. In this sense our model can be viewed as a generalization of standard multivariate SV models. However, as we shall 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 Richard and Zhang (2007), and we extend it to the present model framework. We discuss specification tests such as parameter constancy and copula selection. In the 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 the model of Patton (2006) and a DCC model with GARCH(1,1) marginals. The latter is a natural competitor to our model and, due to its popularity, serves as a benchmark. This also holds if we replace the SV marginals with GARCH(1,1) to compare directly the dependence parts of the models, i.e. dynamic conditional correlation versus stochastic copula.

The remainder of the paper is organized as follows. The next section introduces the model, and 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

DOI: 10.1002/jae

C such that

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

Thus by the probability integral transform a copula is a multivariate[1] distribution function with uniform marginals and it fully captures the dependence between x and y irrespective of the marginal distributions. Examples and properties of the copulas we consider in this paper can be found in the Appendix. For a more detailed treatment, we refer to the book by Joe (1997).

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

$$(u_t, v_t) \sim C(u, v|\theta_t) \tag{2}$$

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 Ψ: $\mathbb{R} \to \Theta$ 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:

$$\lambda_t = \alpha + \beta\lambda_{t-1} + \nu\varepsilon_t \tag{3}$$

where ε_t is an i.i.d. $N(0,1)$ innovation. In principle, 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. Furthermore, the first-order autoregressive model allows for a simple state space representation, which we describe below.

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 of how to deal with the marginals below.

The above model has a nonlinear state space representation with observation equation $(u_t, \ v_t) \mid \lambda_t \sim C(u, v| \ \Psi(\lambda_t))$ and transition equation given by (3). A straightforward extension is possible when the marginals are unknown and included in the model, which is postponed to Section 2.3. Hence the estimation procedure used in this paper can be interpreted as a nonlinear filtering algorithm and is linked to the extensive literature on nonlinear filtering (see, for example, Doucet *et al.*, 2001). In particular, Durbin and Koopman (2000, 2001) discuss how importance sampling can be used to estimate time series models in state space form. In the following we describe how a particular importance sampling technique can be adapted to estimate our model.

2.2. Estimation

We are interested in estimating the parameter vector $\omega = (\alpha, \beta, \nu)$. Define $U = \{u_t\}_{t=1}^T$, $V = \{v_t\}_{t=1}^T$ 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 vector ω is

$$L(\omega; U, V) = \int f(U, V, \Lambda; \omega)\mathrm{d}\Lambda \tag{4}$$

[1] We only consider bivariate copulas in this paper, but most methods can in principle be extended to the multivariate case; see Section 2.8.

DOI: 10.1002/jae

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:

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

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:

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

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

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

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

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

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

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

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

DOI: 10.1002/jae

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

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

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:

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

for $t = T, \ldots, 1$ 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:

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

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

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

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 $t = T, \ldots, 1$ solve the back-recursive least squares regression problem

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

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

DOI: 10.1002/jae

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 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 versus 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 ith stock market returns r_{it} for $t = 1, \ldots, T$ 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 may be done using 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

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

As mentioned in the Introduction, it is easily possible to relax the assumption of conditional normality of the marginal distributions. For example ε_{it} may come from a t-distribution with different degrees of freedom for each series and the dependence may still be modeled using a Gaussian copula. This is not possible using generic multivariate distributions such as the multivariate Student t. Note that the joint model allows a nonlinear state space representation. Denote by $\Phi(\cdot)$ the standard normal cumulative distribution function (CDF). Then the observation equation is given by

$$(r_{it}, r_{jt}) \mid \lambda_t, h_{it}, h_{jt} \sim C(\Phi(\varepsilon_{it}), \Phi(\varepsilon_{jt}))$$

where in the SV model $\varepsilon_{it} = \exp(-h_{it}/2)r_{it}$ and the transition equations are given by (3) and (15), while in the GARCH model $\varepsilon_{it} = r_{it}/\sqrt{h_{it}}$ and the transition equations are given by (3) and (16). The state variables h_{1t} and h_{2t} of the marginal models do not depend on the states of the process λ_t driving the dependence, which justifies the sequential estimation of the marginals and the copula we describe below.

For either choice of the model for the marginals, 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

DOI: 10.1002/jae

$Y = \{y_t\}_{t=1}^T$ 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. 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 $\psi(X, Y) = (\partial\mathcal{L}_X/\partial\delta_1', \partial\mathcal{L}_Y/\partial\delta_2', \partial\mathcal{L}_C/\partial\omega')'$ denote the inference functions. Furthermore, let us define $D = E[\partial\psi(X, Y)/\partial\vartheta']$ and $M = E[\psi(X, Y)\psi(X, Y)']$. Then, as shown by Joe (2005), $\sqrt{T}(\hat{\vartheta} - \vartheta) \to_d N(0, V)$, where

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

We will use this result for inference in our applications.

Note that the possibility of two-step estimation is not exclusive to copula-based models, but has also been proposed for the DCC model by Engle (2002) and for the model by Asai and McAleer (2009). However, these procedures are in fact equivalent to using the copula decomposition of the log-likelihood function with the Gaussian copula.

2.4. 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 state 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 predicted estimate of $\Psi(\lambda_t)$, which we denote $\Psi(\hat{\lambda}_{t|t-1})$ and is given by

$$E[\Psi(\lambda_t)|U_{t-1}, V_{t-1}] = \frac{\int \Psi(\lambda_t)\, p(\lambda_t|\Lambda_{t-1}, \omega) f(U_{t-1}, V_{t-1}, \Lambda_{t-1}; \omega) \mathrm{d}\Lambda_t}{\int f(U_{t-1}, V_{t-1}, \Lambda_{t-1}; \omega) \mathrm{d}\Lambda_{t-1}} \tag{19}$$

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

$$\frac{1}{N}\sum_{i=1}^{N}\left(\Psi[\tilde{\lambda}_t^{(i)}(\hat{\omega})]\prod_{t=1}^{t-1}\left[\frac{f(u_\tau, v_\tau, \tilde{\lambda}_t^{(i)}(\hat{a}_\tau)|U_{\tau-1}, V_{\tau-1}, \tilde{\Lambda}_{\tau-1}^{(i)}(\hat{a}_{\tau-1}), \hat{\omega})}{m(\tilde{\lambda}_{tau}^{(i)}(\hat{a}_\tau)|\tilde{\Lambda}_{\tau-1}^{(i)}(\hat{a}_{\tau-1}), \hat{a}_\tau)}\right]\right) \tag{20}$$

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, CRNs should be used to evaluate the numerator

DOI: 10.1002/jae

and the denominator. The integral in (19) must be evaluated for each $t = 1, \ldots, T$. In addition it needs to be mentioned that the number N of trajectories from the importance sampler must be much 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 predicted 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 calculating predicted states is to make use of the mechanics of EIS to obtain smoothed estimates of the latent process, i.e. to estimate each $\Psi(\lambda_t)$ using the complete information available at time T. 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 the smoothed estimate of $\{\Psi(\lambda_t)\}_{t=1}^T$ by

$$\Psi(\hat{\lambda}_{t|T}) = \frac{1}{N}\sum_{i=1}^{N} \Psi(\tilde{\lambda}_t^{(i)}(\hat{a}_t)) \quad \forall t = 1, \ldots, T \tag{21}$$

Using both simulated and real data, we obtained smoothed estimates of $\{\Psi(\lambda_t)\}_{t=1}^T$ 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 $\Psi(\hat{\lambda}_{t|T})$. Furthermore, it is computationally much cheaper and requires only a small N.

A third option is to calculate the smoothed estimate of $\Psi(\lambda_t)$ using only the information available at time t, i.e. U_t and V_t. This estimate, which is called the filtered estimate, will be denoted by $\Psi(\hat{\lambda}_{t|t})$. In order to understand how predicted estimates are obtained, one should be aware that these are in fact the one-step forecasts of the filtered state variable accounting for the nonlinearity of the transformation Ψ.

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) model[2] and the DCC-GARCH model of Engle (2002), where correlation is described by

$$\begin{aligned} \rho_{i,j,t} &= \frac{q_{i,j,t}}{\sqrt{q_{i,i,t}q_{j,j,t}}} \\ q_{i,j,t} &= \overline{\rho}_{i,j} + \alpha(\varepsilon_{i,t-1}\varepsilon_{j,t-1} - \overline{\rho}_{i,j}) + \beta(q_{i,j,t-1} - \overline{\rho}_{i,j}) \end{aligned} \tag{22}$$

with $\overline{\rho}$ 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.

[2] The correlation estimate is the smoothed correlation path.

DOI: 10.1002/jae

The conditional correlation in this model is given by

$$\rho_t = \Psi\left(\alpha + \beta\rho_{t-1} + \nu\frac{1}{10}\sum_{j=1}^{10}\Phi^{-1}(u_{t-j})\Phi^{-1}(v_{t-j})\right) \tag{23}$$

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

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: $\rho_t = \mathrm{mod}(t/200)$
5. DCC: correlation generated by (22) with $\overline{\rho} = 0.7$, $\alpha = 0.04$ 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 6–9 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 owing 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 I, 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, but it is remarkable that the SCAR model is capturing the dynamics of the latent process well under a large variety of data-generating processes that are quite different from the assumed stationary AR(1) structure. The DCC model 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). Note that in Asai and McAleer (2009) a similar simulation study is performed using DGPs very

Table I. Average MSEs for correlation estimates

Model/DGP	Constant	Jump	Sine	Ramp	DCC	AR(1)	Noise	RW	ARMA(2,2)
SCAR	0.0005	0.0066	0.0130	0.0261	0.0041	0.0087	0.0104	0.0081	0.0556
DCC	0.0010	0.0134	0.0318	0.0458	0.0012	0.0209	0.0104	0.0090	0.1096
Patton	0.0019	0.0416	0.0530	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	–	–	–	–	–	0.0350	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 rescaled by the inverse Fisher transform, whereas in the bottom panel the transformation $\rho_t = \lambda_t/\sup_t|\lambda_t|$ was chosen.

DOI: 10.1002/jae

close to 1–4. In terms of estimating ρ_t their dynamic correlation model is found to perform worse than the DCC, even though the volatilities in their DGP follow an SV model, which should strongly favor their SV-based model over the (misspecified) DCC-GARCH.

2.5. 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

$$H_0 : \theta_t = \overline{\theta}, \quad \forall \quad t = 1, \ldots, T \tag{24}$$

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

$$H_0 : \ \nu = 0 \tag{25}$$

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

$$\mathrm{LR}_C = -2(\mathrm{LL}_{\mathrm{res}} - \mathrm{LL}_{\mathrm{ur}}) \tag{26}$$

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 copula. The most simple, but still very reliable way of choosing the best-fitting copula, is to compare the values of the log-likelihood at the parameter estimates or the Akaike information criterion (AIC) when the competing models have 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, for example, 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.* (2009). 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

$$H_0 : C_i(u_t, v_t, \hat{\theta}_t) = C_0(u_t, v_t, \theta_t^0) \tag{27}$$

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 GoF test many of the tests proposed in the papers cited above such as 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

$$z_t = C_0(u_t|v_t, \theta_t^0) = \frac{\partial C_0(u_t, v_t, \theta_t^0)}{\partial v_t} \sim U(0, 1) \tag{28}$$

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)

DOI: 10.1002/jae

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

$$\hat{\mathbb{F}}(x) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{I}_{\{z_t < x\}} \tag{29}$$

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

$$T_{\mathrm{KS}} = \sup_x |F(x) - \hat{\mathbb{F}}(x)| \tag{30}$$

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

$$T_{\chi^2} = \sum_{i=1}^{k} \frac{(E(c_i) - c_i)^2}{E(c_i)} \overset{H_0}{\sim} \chi^2(k) \tag{31}$$

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 statistic is

$$T_{\mathrm{JB}} = \frac{T}{6}\left(S^2 + \frac{(K-3)^2}{4}\right) \overset{H_0}{\sim} \chi^2(2) \tag{32}$$

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

$$T_{\mathrm{AD}} = \sup_x \frac{\sqrt{T}|\hat{\mathbb{F}}(x) - F(x)|}{\sqrt{F(x)(1-F(x))}} \tag{33}$$

for which again tabulated critical values are used.

2.6. 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, for example, in 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

$$\hat{\lambda}_{T+r} = \mu + \beta^r(\hat{\lambda}_T - \mu) \tag{34}$$

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

$$\sigma^2_{T+r} = \nu^2(1-\beta^{2r})/(1-\beta^2) \tag{35}$$

DOI: 10.1002/jae

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 one-step-ahead forecast of $\theta_t = \exp(\lambda_t)$ is

$$E(\theta_t|\mathcal{F}_{t-1}) = \exp\left(\mu_t + \frac{\sigma_t^2}{2}\right) \tag{36}$$

For μ_t, the conditional expectation of λ_t given $\mathcal{F}_{t-1}$, we insert its linear forecast given in (34). From this it follows by straightforward calculations that the r-step-ahead forecast of θ is

$$\hat{\theta}_{T+r} = \exp\left(\hat{\lambda}_{T+r} + \frac{\sigma_{T+r}^2}{2}\right) \tag{37}$$

The confidence bands for these forecasts can be calculated 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 owing to the nonlinearity of the inverse Fisher transform. We therefore use a second-order Taylor approximation of $\Psi(\lambda_t)$ around μ_t:

$$\Psi(\lambda_t) \approx \Psi(\mu_t) + \Psi'(\mu_t)(\lambda_t - \mu_t) + \frac{1}{2}\Psi''(\mu_t)(\lambda_t - \mu_t)^2 \tag{38}$$

Taking the conditional expectation we have

$$\begin{aligned} E[\Psi(\lambda_t)|\mathcal{F}_{t-1}] &\approx E\left\{\Psi(\mu_t) + \Psi'(\mu_t)(\lambda_t - \mu_t) + \frac{1}{2}\Psi''(\mu_t)(\lambda_t - \mu_t)^2|\mathcal{F}_{t-1}\right\} \\ &= \Psi(\mu_t) + \Psi'(\mu_t)E((\lambda_t - \mu_t)|\mathcal{F}_{t-1}) + \frac{\Psi''(\mu_t)}{2}E((\lambda_t - \mu_t)^2|\mathcal{F}_{t-1}) \\ &= \Psi(\mu_t) + \frac{\Psi''(\mu_t)}{2}\nu^2 \end{aligned} \tag{39}$$

Then the r-step-ahead forecast is

$$\hat{\theta}_{T+r} = \Psi(\hat{\lambda}_{T+r}) + \frac{-4(\exp(2\hat{\lambda}_{T+r}) - 1)\exp(2\hat{\lambda}_{T+r})}{(\exp(2\hat{\lambda}_{T+r}) + 1)^3}\sigma_{T+r}^2 \tag{40}$$

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.

2.7. Tail Properties of the Gaussian SCAR Model

Manner and Segers (2009) study the dependence in the tails of mixtures of elliptical copulas. For the special case of the Gaussian copula they show that when the random correlation parameter ρ can get close to one with positive probability the tails of the mixture copula are much heavier than those of the static Gaussian copula. This situation applies to the Gaussian SCAR model. In particular, it is shown that the mixture copula falls into the newly defined category of *near asymptotic dependence* in the tails, in contrast to the property of *near asymptotic independent tails* as studied

DOI: 10.1002/jae

in Ledford and Tawn (1996). *Near asymptotic dependence* means that for every $\varepsilon > 0$ it holds that $u^{1+\varepsilon} \ll C(u, u) \ll u$ as $u \downarrow 0$, where $a(u) \ll b(u)$ means that $a(u) = o(b(u))$, i.e. $a(u)/b(u) \to 0$. This implies that, although $P[V < u|U < u] \to 0$ as $u \to 0$, the speed of convergence is extremely slow and at any practically relevant quantile the tail probability is substantially different from zero and hence at finite samples one gets the impression of asymptotically dependent tails. Consequently, one can expect the model to fit financial data well and in most situations the popular t-copula loses its advantage of being able to capture tail dependence. Economically, this property also has a compelling interpretation. Just like GARCH or SV models imply fat tails in the univariate case, stochastic correlations create fat tails (in the sense of joint extremes) in the unconditional multivariate distribution. This might be seen as an extension of the 'mixtures of distributions hypothesis' for stock prices introduced by Clark (1973).

2.8. Possible Extensions

So far we have only considered stochastic copula models for bivariate copulas with a single dependence parameter. Here we discuss potential extension for bivariate models and some ideas that look promising for extending the model to larger dimensions.

2.8.1 Extensions of the Bivariate Model

Consider the conditional copula mixture density with time-varying mixing parameter

$$c_{\text{mix}}(u, v; \theta_1, \theta_2, \kappa_t) = \kappa_t c_1(u, v; \theta_1) + (1 - \kappa_t) c_2(u, v; \theta_2) \tag{41}$$

where the mixing parameter $\kappa_t = \Psi(\lambda_t)$ with λ_t following (3) and $\Psi(x) = 1/(1 + \exp(x))$ to keep the mixing parameter in (0,1) at all times. For example, c_1 and c_2 could be the Gaussian and the rotated Gumbel copulas and hence a decrease in κ_t would cause an increase in lower tail dependence. Such a model could describe tranquil periods, when a Gaussian dependence structure is appropriate, and crisis periods with larger overall dependence and in particular greater dependence for losses.

The symmetrized Joe–Clayton copula introduced by Patton (2006) is a flexible two-parameter copula that is parameterized in terms of λ_L and λ_U, the coefficients of lower and upper tail dependence. A time-varying version of this model, as considered in Patton (2006), allows for changing degrees of asymmetry, as well as a time-varying overall level of dependence. Again $\Psi(x) = 1/(1 + \exp(x))$ could be chosen. For estimation, the two-component EIS procedure of Liesenfeld and Richard (2003) could be used. However, the estimation is likely to be computationally very burdensome. Furthermore, in Patton (2006) the time-varying version of this model hardly has a better fit than a dynamic Gaussian copula, but three additional parameters need to be estimated.

An exogenous variable x may be integrated into the models by replacing (3) by

$$\lambda_t = \alpha + \beta\lambda_{t-1} + \delta x_t + \nu\varepsilon_t \tag{42}$$

The variable x may be deterministic, such as a trend or a dummy representing a structural break, but it may also be an economic variable that is expected to explain correlation. A possibility would be to include trade volume, which could serve as a proxy for the impact of volatility on correlation.

2.8.2 Multivariate Modeling

Extending the SCAR model to more than two dimensions is far from trivial. Archimedean copulas such as the Gumbel or Clayton copula have straightforward multivariate analogues, with the

DOI: 10.1002/jae

drawback that a single dependence parameter describes the complete dependence between all variables, which may be too restrictive for most applications. Similarly, for Gaussian copulas one could restrict all correlations to be equal, leading to a model similar to the dynamic equicorrelation model of Engle and Kelly (2008).

Nevertheless, the recent advances in multivariate copula modeling using so-called vine copulas introduced by Bedford and Cooke (2002) and applied to financial data by Chollete *et al*. (2009), for example, seem to offer a promising approach. We do not discuss the construction of vine copulas (also called pair copula constructions) here, but we note that the log-likelihood function decomposes conveniently into a sum of the marginal log-likelihood and bivariate (conditional) copula log-likelihoods. We refer to Aas *et al*. (2009) for a accessible exposition of the details on pair copula constructions and their estimation. Thus the estimation of a rather complex model can be done in a number of relatively simple steps and the time-varying dependence processes can be estimated individually, but conditional on the estimated dependence process in earlier steps.

For Gaussian copulas an alternative to vine copulas could be to construct a time-varying correlation matrix R_t by simply joining correlations obtained from bivariate Gaussian SCAR models. The obvious disadvantage of this approach is that R_t may not be positive definite at all points in time, which is likely to be a serious problem for larger dimensions. This may be overcome by using shrinkage methods such as those proposed in Ledoit and Wolf (2004). However, optimal shrinkage rates and appropriate targets need to be developed. Moreover, it is not clear whether such an estimate for R_t has good properties given the inefficient estimation and the bias due to shrinkage.

3. APPLICATION

In this section we present an empirical application of our model and compare it with competing models using two different datasets on stock indices: one measured at the daily frequency and during mostly tranquil market conditions; the second one at the weekly frequency and including a financial crisis at the end of the sample. The analysis is split into estimation and in-sample validation, and out-of-sample comparison.

3.1. Daily Data: Dow Jones and Nasdaq

The first dataset we consider to illustrate our model and compare with competing models consists of daily observations of the Dow Jones Industrial Average and the Nasdaq composite ranging from 26 March 1990 until 23 March 2000. The same dataset has been considered in Engle (2002). Returns are calculated as the first difference of the natural logarithm multiplied by 100.

3.1.1 Estimation and In-Sample Validation

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 II, and the GARCH and smoothed SV volatilities can be found in Figure 1. 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.

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

DOI: 10.1002/jae

Table II. Estimates of GARCH and SV models: Dow Jones and Nasdaq

GARCH	DJ	NQ	SV	DJ	NQ
α	0.0060	0.0308	α	−0.0112	−0.0054
	(0.0044)	(0.0173)		(0.0051)	(0.0042)
γ	0.0480	0.1150	β	0.9786	0.9733
	(0.0184)	(0.0435)		(0.0076)	(0.0078)
β	0.9450	0.8630	ν	0.1573	0.2072
	(0.0214)	(0.0529)		(0.0254)	(0.0266)
logl	−3195.38	−3693.89	logl	−3137.37	−3638.03

Note: Estimation results for the models of the marginal distributions. Standard errors are reported in parentheses, and logl is the value of the log-likelihood function.

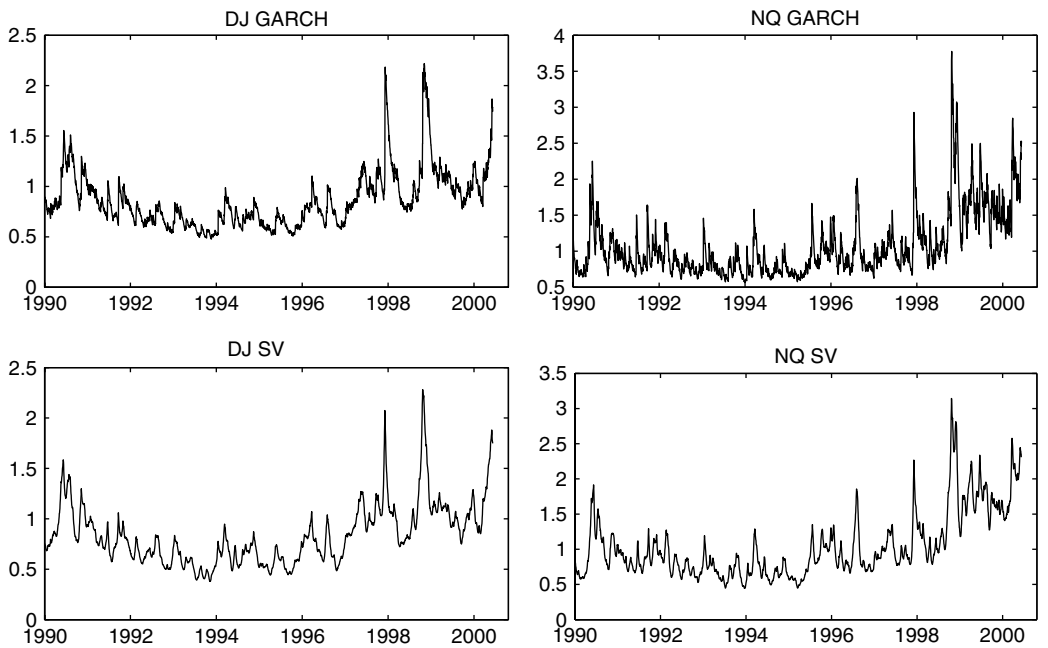

Figure 1. Volatility estimates: Dow Jones and Nasdaq

Using the estimated standard deviations, the data are 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 and the results of Joe (2005, p. 405). Table III presents the results of the estimation.

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. In terms of the log-likelihood value, the Gaussian copula is the best-fitting model, followed by the survival Gumbel copula. Comparing one-step

[4] We 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.

DOI: 10.1002/jae

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

Copula/logl	Two-step			One-step		
	α	β	ν	α	β	ν
Normal	0.0302	0.9679	0.0824	0.0261	0.9720	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	0.9750	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 model as marginal distribution. Left: two-step estimation using the inference function for margins method of Joe (2005); right: one-step estimation using full maximum likelihood. Standard errors are reported in parentheses, and logl is the value of the log-likelihood function.

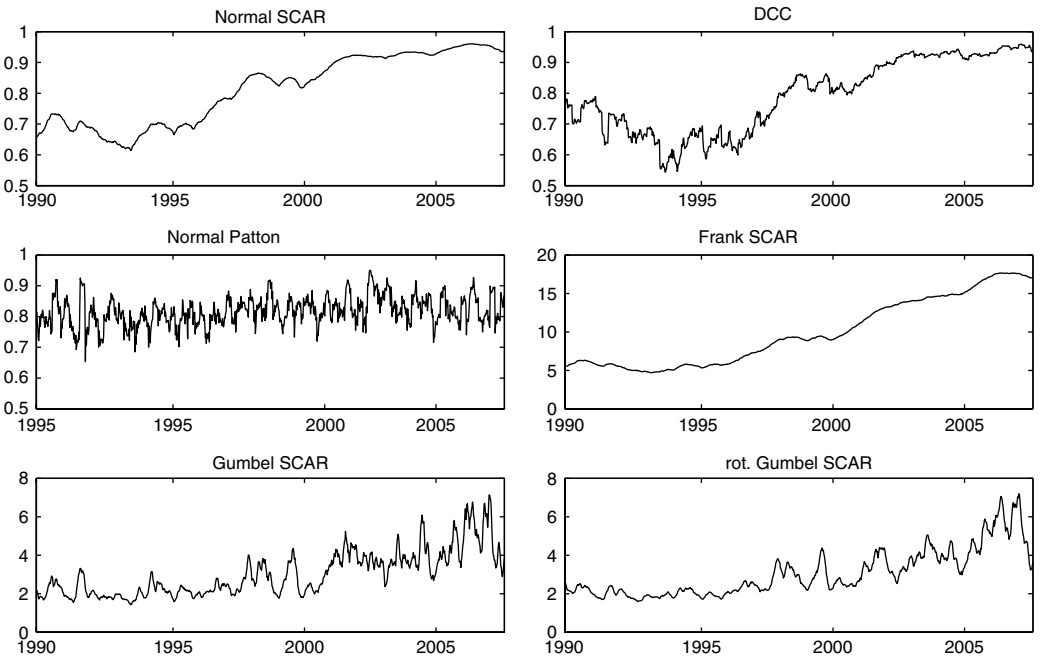

Figure 2. Time path of the dependence parameters: Dow Jones and Nasdaq

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 are they 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 paths from the DCC and Patton models for comparison. The path of the dependence parameters looks quite similar for all stochastic copula models, although the scale is different. The DCC correlation is slightly noisier, which is likely due to the fact that it is based on one-step-ahead forecasts. Economically, the decrease in dependence in 2000 was explained by Engle (2002) as the 'sector rotation between

DOI: 10.1002/jae

Table IV. Goodness-of-fit tests: Dow Jones and Nasdaq

	Normal	Rot. Gumbel	Frank	Clayton	DCC	Patton
Two-step						
KS	0.2130	0.0046	0.0777	0.0001	0.1006	0.0392
χ^2	0.0012	0.0000	0.0001	0.0000	0.0001	0.0000
JB	0.0348	0.0000	0.0000	0.0000	0.0223	0.0014
AD	0.0303	0.0000	0.0000	0.0000	0.0155	0.0009
One-step						
KS	0.2105	0.0043	0.0996	0.0001	0.1854	0.0404
χ^2	0.0527	0.0000	0.0000	0.0000	0.0010	0.0000
JB	0.0913	0.0000	0.0000	0.0000	0.0049	0.0010
AD	0.0584	0.0000	0.0000	0.0000	0.0039	0.0003

Note: p-values for the null hypothesis of correct specification of the copula function for the two-step (top) and one-step estimation (bottom) of the Dow Jones and Nasdaq data. KS, χ^2, JB and AD are defined in equations (30)–(33).

new economy stock and "brick and mortar" stocks'. Note that the Patton model fails to capture this drop in dependence.

The outcomes of the GoF tests are reported in Table IV. 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. We also include the results for the Patton model for the Gaussian copula and the DCC model. Note that the DCC was estimated on the transformed variables $\Phi^{-1}(u)$ and $\Phi^{-1}(v)$, where Φ denotes the CDF of the standard normal distribution. This was done to prevent differences in the copula fit due to different marginal distributions. The Gaussian SCAR model estimated in one step is the only model that passes all four tests. The Gaussian copula using the DCC and Patton models, producing copula log-likelihoods of 876.51 and 837.98, respectively, clearly have a worse in-sample fit. 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 preferred when considering a static copula model.[5] 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 recall from Section 2.7 that the Gaussian SCAR model implies near-asymptotic tail dependence and hence can capture dependencies in the extremes at finite samples.

3.1.2 Out-of-Sample Comparison

As a start we use the techniques described in Section 2.6 to obtain out-of-sample forecasts of the copula dependence process. We use the last observation of the smoothed dependence 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 the 250 out-of-sample observations and computing the smoothed path of the dependence parameter. Figure 3 presents the forecasts together with 95% confidence bands and the smoothed path. Note that the confidence band for the normal, rotated Gumbel and Clayton models are asymmetric, 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. After about 100 days, the forecast

[5] Estimation results for static copulas are available upon request.

DOI: 10.1002/jae

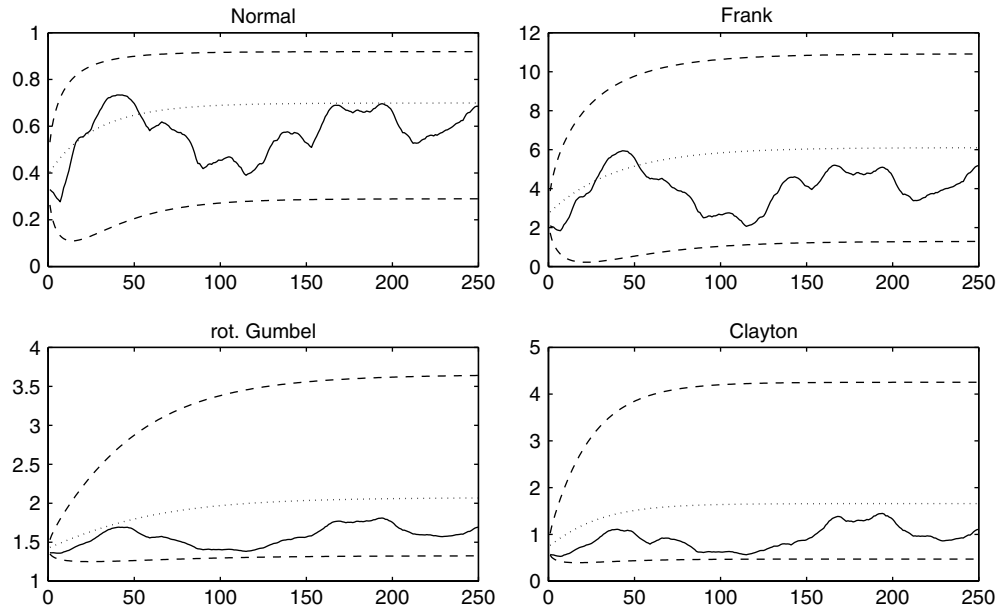

Figure 3. Out-of-sample forecasts of the dependence parameter: Dow Jones and Nasdaq. Note: The dotted lines denote the out-of-sample forecasts $\hat{\theta}_{T+r}$, the dashed lines are the 95% confidence intervals and the solid lines are the smoothed estimates $\hat{\theta}_{t|T}$

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

So far our results suggest that the stochastic copula model based on the Gaussian copula fits that data well in-sample and that the forecasts of the dependence path look reasonable when compared with the dependence implied by the model using the out-of-sample data. However, to show the usefulness of the forecasts of the copula parameter, it is preferable to compare the forecasts to some measure that does not depend on a model. Given that the Gaussian copula had the best in-sample fit, we exclusively focus on evaluating correlation and covariance forecasts in this section, which makes comparison with competing models and the use of established criteria possible. As in Pelletier (2006) we measure the closeness of an r-step forecast of the covariance matrix, which we denote by $\hat{H}_{t+r}$, to the true covariance matrix by the following two criteria:

$$\mathrm{RMSE}_r = \left(\frac{1}{K^2} \sum_{i=1}^{K} \sum_{j=1}^{K} (\hat{H}_{i,j,t+r} - r_{i,t+r} r_{j,t+r})^2 \right)^{1/2} \tag{43}$$

$$\mathrm{MAD}_r = \frac{1}{K^2} \sum_{i=1}^{K} \sum_{j=1}^{K} |\hat{H}_{i,j,t+r} - r_{i,t+r} r_{j,t+r}| \tag{44}$$

DOI: 10.1002/jae

where K is the number of assets. The second criterion may be preferable because it is more robust to outliers. A third criterion on which we base our out-of-sample comparison could be linked to the economic value of the forecasts, namely the variance of the global minimum variance portfolio (MVP) constructed using the forecast covariance matrix, see Fleming *et al*. (2001). The portfolio weights are given by

$$w_{t+r} = \frac{\hat{H}_{t+r}^{-1}\iota}{\iota'\hat{H}_{t+r}^{-1}\iota} \tag{45}$$

where ι is a $(K \times 1)$ vector of ones.

We compare 1-, 5-, 10- and 20-step forecasts of the constant conditional correlation (CCC) GARCH, DCC-GARCH, SCAR-SV and the SCAR model with GARCH margins (SCAR-G). The SCAR model with GARCH margins was included to study the difference in forecasts that can be attributed to the correlation model. The parameters of the models are only estimated once using the in-sample period. We restrict the in-sample period to the last 1000 observations from the original sample to avoid effects of potentially unstable parameters, and since we need to run the computationally heavy importance sampler every time we want to forecast to evaluate the current volatility and correlation. For the out-of-sample period we consider the following 250 trading days.

Table V reports the results of the analysis. The MAD and RMSE suggest that the SCAR-SV model has the best out-of-sample fit. Looking at σ^2_{MVP} the simple CCC-GARCH model gives the best results, except for $r = 1$, where the SCAR-SV slightly outperforms the DCC. This surprising finding is in line with the out-of-sample analysis of Pelletier (2006), who also found a good performance of a constant correlation model. This finding may be attributed to this specific dataset. It is remarkable that, comparing only the dynamic correlation models, the SCAR does better than the DCC in all cases, except for σ^2_{MVP} and $r = 1$.

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 1 January 1990 until 24 August 2009 and thus also cover the recent financial crisis. The last 100 observations are put aside for

Table V. Out-of-sample fit: Dow Jones and Nasdaq

		CCC	DCC	SCAR-SV	SCAR-G
MAD	$r = 1$	4.69	4.63	**4.35**	4.62
	$r = 5$	4.76	4.72	**4.32**	4.70
	$r = 10$	4.74	4.70	**4.27**	4.69
	$r = 20$	4.27	4.25	**3.85**	4.23
RMSE	$r = 1$	10.29	10.32	**10.25**	10.31
	$r = 5$	10.50	10.53	**10.50**	10.52
	$r = 10$	10.60	10.63	**10.59**	10.62
	$r = 20$	**9.67**	9.69	9.69	9.68
σ^2_{MVP}	$r = 1$	1.45	1.43	**1.43**	1.44
	$r = 5$	**1.37**	1.46	1.45	1.43
	$r = 10$	**1.36**	1.49	1.50	1.47
	$r = 20$	**1.25**	1.31	1.31	1.29

Note: This table presents the mean absolute deviation (MAD) and root mean square error (RMSE) between the forecast covariance and the cross-product of the out-of-sample data. The bottom panel shows the variance of the minimum variance portfolio (MVP) with weights given in (45).

DOI: 10.1002/jae

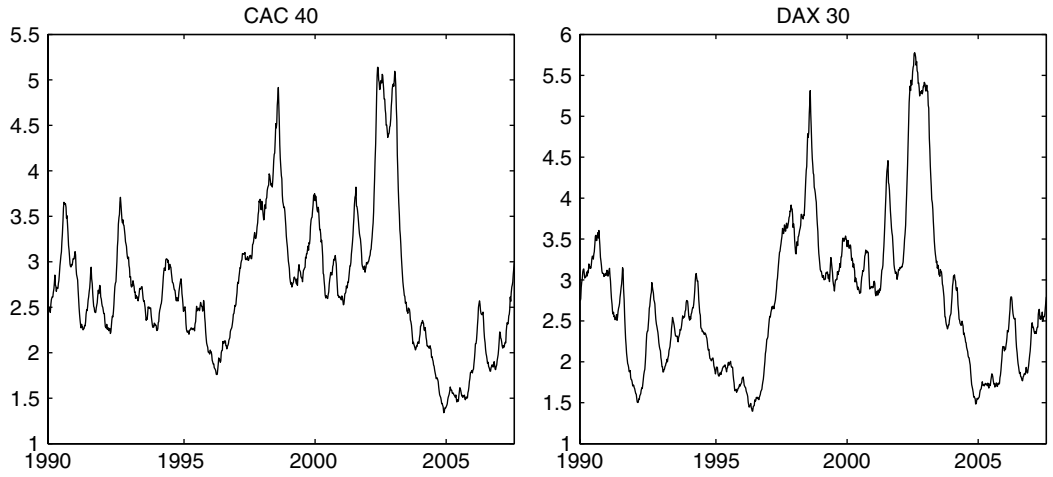

Figure 4. Volatility estimates: CAC and DAX

the out-of-sample evaluation. As this sample covers a rather large span of time it is likely that not only the conditional, but also the unconditional dependence has changed. If this is indeed the case, the assumption of stationary AR(1) dynamics for the latent dependence process is no longer valid, meaning that estimation and inference become questionable. Nevertheless, we saw in Section 2.4 that the latent process can still be estimated quite well in such situations.

3.2.1 Estimation and In-Sample Validation

As above, an SV model is chosen for the margins and the SCAR model with different choices of copulas is considered for the dependence model. Volatility estimates are depicted in Figure 4. Note the high levels of volatility at the end of the 1990s and around the 9/11/2001 terrorist attacks. At the end of the in-sample period volatility seems to increase again, but the effect of the credit crisis mainly falls into the out-of-sample period. Estimates for the SCAR model and its competitors are reported in Table VI and Figure 5. Only the two-step estimator is used as the estimates are quite close to those using one-step estimation. Since for two of the models the estimate for β achieves the upper bound of 0.9999, the numerical derivatives could not be evaluated and hence standard errors of the estimates could not be obtained. This is an obvious problem and it is an indication that at the upper bound of the permitted parameter space $\beta = 0.9999$ is not the optimum.

Table VI. Estimation and GoF for CAC and DAX returns

	Normal	Gumbel	Frank	Rot. Gumbel	DCC	Patton
logl	575.20	531.15	532.03	543.44	565.11	498.80
α	0.0001	0.0284	0.0001	0.0219	0.0209	3.1171
β	0.9999	0.9219	0.9999	0.9617	0.9792	−1.9257
ν	0.0294	0.2749	0.2330	0.1764	—	1.0604
Goodness of fit						
KS	0.2052	0.0321	0.5434	0.1546	0.0326	0.0226
χ^2	0.3174	0.0692	0.5964	0.0345	0.1438	0.0000
JB	0.0044	0.0030	0.4691	0.0062	0.0010	0.0010
AD	0.0185	0.0000	0.6784	0.0006	0.0002	0.0000

Note: Two-step estimation results of the SCAR models and p-values for the null hypothesis of correct specification of the copula function using the tests in (30)–(33) for the weekly CAC and DAX returns.

DOI: 10.1002/jae

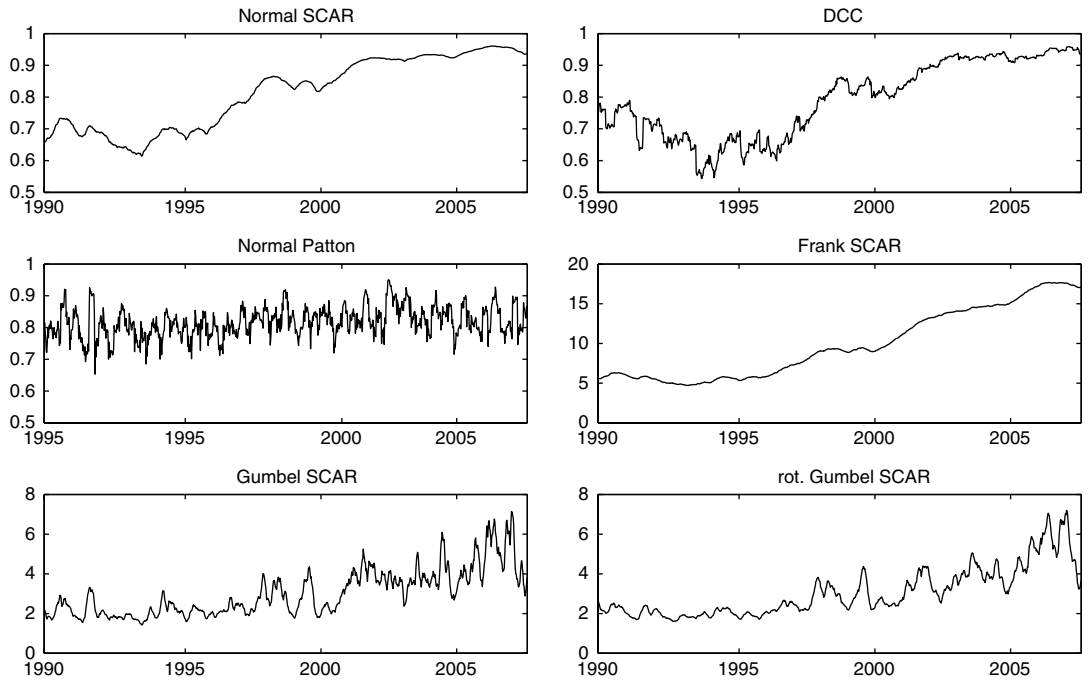

Figure 5. Time path of the dependence parameters: CAC and DAX

As mentioned above, this is probably due to the fact that the true dependence process does not follow a stationary AR(1) process and therefore assuming such a structure can only be seen as an approximation to the true data-generating process.

While in a static copula framework asymmetric models provide a better fit, the Gaussian copula is again clearly the best-fitting model in the time-varying case in terms of the log-likelihood value. The GoF tests, on the other hand, suggest that the Frank copula is the appropriate model. Similar to the example above, the SCAR model fits slightly better than the DCC model and clearly outperforms the Gaussian conditional copula model of Patton (2006). Note that the estimate of β for the Patton model, which is not restricted to be smaller than one in absolute value, does not have the expected sign. However, different starting values lead to the same parameter estimates and this is in fact in line with Patton (2006), who estimates a (large) negative persistence parameter for his time-varying Joe–Clayton copula. For the SCAR model, 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, attaining the upper bound in the constrained optimization. Hence their dependence processes are likely to be integrated and the stationary AR(1) model can only be seen as an approximation. 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 both the SCAR and the DCC model are able to capture movements of the dependence parameter that resemble trends or regime shifts without explicitly including such features in the model. The price to pay for this flexibility is that it is not possible to obtain standard errors in such cases. The strong rise in stock market dependence

DOI: 10.1002/jae

Table VII. Out-of-sample fit: CAC and DAX

		CCC	DCC	SCAR-SV	SCAR-G
MAD	$r = 1$	17.27	17.55	**16.98**	17.51
	$r = 5$	17.68	18.00	**17.49**	17.94
	$r = 10$	18.26	18.50	**18.20**	18.46
	$r = 20$	18.13	18.21	**17.80**	18.19
RMSE	$r = 1$	33.28	**33.18**	33.21	33.19
	$r = 5$	35.14	35.07	**34.89**	35.06
	$r = 10$	36.13	36.00	**35.97**	36.01
	$r = 20$	34.24	34.07	**33.75**	34.09
σ^2_{MVP}	r = 1	19.07	19.01	**18.67**	18.91
	$r = 5$	19.81	19.72	**19.60**	19.67
	$r = 10$	20.79	20.88	**20.66**	20.84
	$r = 20$	**20.93**	21.85	21.31	21.42

Note: This table presents the mean absolute deviation (MAD) and root mean square error (RMSE) between the forecast covariance and the cross-product of the out-of-sample data. The bottom panel shows the variance of the minimum variance portfolio (MVP) with weights given in (45).

is likely to be a consequence of the European integration process and the introduction of the euro in 1999.

3.2.2 Out-of-Sample Comparison

The out-of-sample analysis is conducted in the same way as for the Dow Jones Nasdaq data using the last 100 weeks of the sample. The results are reported in Table VII. Given that the credit crisis in 2008 is covered, it is not surprising that now the CCC model does not beat the models that allow for correlation dynamics. The SCAR-SV model outperforms its competitors in the majority of cases. Comparing the DCC and SCAR with GARCH margins, the latter does better in terms of MAD and σ^2_{MVP}, indicating better forecasting performance of the SCAR model, in particular when measured in economic terms. It is remarkable that the model works well in a situation in which it is important to make good forecasts and when larger changes in dependence are likely to occur.

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 pairs of stock 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. The out-of-sample analysis shows good performance of the model compared with the DCC and CCC-GARCH models. In particular, for the data including the credit crisis, the model performs very well compared to its competitors.

We have discussed a number of possible extensions of the model. In particular, the modeling of more than two assets appears to be an interesting, but challenging, venue for future research. Furthermore, an interesting question is whether there are specific applications where asymmetric

DOI: 10.1002/jae

dynamic copulas outperform Gaussian ones. In such cases, the flexible and nonlinear dependence structures of these copulas can be exploited.

ACKNOWLEDGEMENTS

We would like to thank Franz Palm, Johan Segers, Jean-Pierre Urbain, the participants of the Maastricht University econometrics seminar, the Humboldt–Copenhagen 2009 conference, the 2009 NESG meeting, the Oberseminar Finanz- und Versicherungsmathematik in Munich, two anonymous referees, and the editor Herman van Dijk for comments and suggestions. The first author gratefully acknowledges financial support from the contract *Projet d'Actions de Recherche Concertées* no. 07/12/002 of the Communauté française de Belgique, granted by the Académie universitaire Louvain. The second author gratefully acknowledges financial support from METEOR for the international travel grant.

REFERENCES

Aas K, Czado C, Frigessi A, Bakken H. 2009. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* **44**(2): 182–198.

Amisano G, Casarin R. 2007. Particle filters for Markov switching stochastic correlation models. In *Proceedings of the 2007 Intermediate Conference 'Risk and Prediction'*, Cleup, Padua; 305–316.

Asai M, McAleer M. 2009. The structure of dynamic correlations in multivariate stochastic volatility models. *Journal of Econometrics* **150**(2): 182–192.

Bedford T, Cooke RM. 2002. Vines: a new graphical model for dependent random variables. *Annals of Statistics* **30**(4): 1031–1068.

Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**: 307–327.

Breymann W, Dias A, Embrechts P. 2003. Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance* **3**(1): 1–14.

Broto C, Ruiz E. 2004. Estimation methods for stochastic volatility models: a survey. *Journal of Economic Surveys* **18**(5): 613–649.

Carnero MA, Peña D, Ruiz E. 2004. Persistence and Kurtosis in GARCH and stochastic volatility models. *Journal of Financial Econometrics* **2**: 319–342.

Chen X, Fan Y. 2006. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics* **135**: 125–154.

Chollete L, Heinen A, Valdesogo A. 2009. Modeling international financial returns with a multivariate regime switching copula. *Journal of Financial Econometrics* **7**(4): 437–480.

Clark PK. 1973. A subordinate stochastic process model with finite variance for speculative prices. *Econometrica* **41**: 135–155.

Danielsson J. 1998. Multivariate stochastic volatility models: estimation and comparison with VGARCH models. *Journal of Empirical Finance* **5**: 155–173.

Danielsson J, Richard JF. 1993. Quadratic acceleration for simulated maximum likelihood evaluation. *Journal of Applied Econometrics* **8**: 153–173.

Dobric J, Schmid F. 2007. A goodness of fit test for copulas based on Rosenblatt's transformation. *Computational Statistics and Data Analysis* **51**: 4633–4642.

Doucet A, de Freitas N, Gordon N (eds). 2001. *Sequential Monte Carlo Methods in Practice*. Springer: New York.

Durbin J, Koopman SJ. 2000. Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). *Journal of the Royal Statistical Society, Series B* **62**: 3–56.

Durbin J, Koopman SJ. 2001. *Time Series Analysis by State Space Models*. Oxford University Press: New York.

Engle R. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50**: 987–1008.

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.

DOI: 10.1002/jae

Engle RF, Kelly B. 2008. Dynamic equicorrelation. Technical report, Stern Business School.
Engle RF, Sheppard K. 2001. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. NBER Working Paper 8554, National Bureau of Economic Research.
Fleming J, Kirby C, Ostdiek B. 2001. The economic value of volatility timing. *Journal of Finance* **34**: 157–170.
Genest C, Rivest L-P. 1993. Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association* **88**(423): 1034–1043.
Genest C, Quessy J-F, Rémillard B. 2006. Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics* **33**: 337–366.
Genest C, Rémillard B, Beaudoin D. 2009. Goodness-of-fit tests for copulas: a review and a power study. *Insurance: Mathematics and Economics* **44**: 199–213.
Hamilton J. 1994. *Time Series Analysis*. Princeton University Press: Princeton, NJ.
Harvey A, Ruiz E, Shephard N. 1994. Multivariate stochastic variance models. *Review of Economic Studies* **61**: 247–264.
Jacquier E, Polson NG, Rossi PE. 1994. Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics* **12**: 371–389.
Joe H. 1997. *Multivariate Models and Dependence Concepts*. Chapman & Hall/CRC: Boca Raton, FL.
Joe H. 2005. Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis* **94**(2): 401–419.
Junker M, Szimayer A, Wagner N. 2006. Nonlinear term structure dependence: Copula functions, empirics, and risk implications. *Journal of Banking and Finance* **30**: 1171–1199.
Kim S, Shephard N, Chib S. 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* **65**: 361–393.
Ledford AW, Tawn JA. 1996. Statistics for near independence in multivariate extreme values. *Biometrika* **86**(1): 169–187.
Ledoit O, Wolf M. 2004. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* **88**: 365–411.
Liesenfeld R, Richard JF. 2003. Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance* **10**: 505–531.
Manner H, Segers J. 2009. Tails of correlation mixtures of elliptical copulas. DP0938, Institut de statistique, Université catholique de Louvain.
Patton A. 2006. Modelling asymmetric exchange rate dependence. *International Economic Review* **47**(2): 527–556.
Pelletier D. 2006. Regimes switching for dynamic correlations. *Journal of Econometrics* **131**: 445–473.
Richard JF, Zhang W. 2007. Efficient high-dimensional importance sampling. *Journal of Econometrics* **141**(2): 1385–1411.
Taylor SJ. 1986. *Modelling Financial Time Series*. Wiley: Chichester.
van der Vaart AW. 1998. *Asymptotic Statistics*. Cambridge University Press: Cambridge, UK.
Yu J, Meyer R. 2006. Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews* **25**(2–3): 361–384.

DOI: 10.1002/jae

APPENDIX: COPULAS

This Appendix reviews some copula-based dependence measures and shortly discusses the copulas that are used in this paper. For a complete review of copulas and their properties the interested reader is referred to Joe (1997).

Copula-Based Dependence Measures

Since the parameters of different copulas are usually hard to compare, the degree of dependence is often measured in terms of Kendall's τ. This is a rank correlation measure that lies in $[-1,1]$ and does not depend on the scale of variables of interest (unlike the Pearson correlation coefficient). Two pairs of observations (x_i, y_i) and (x_j, y_j) are called concordant if $(x_i - x_j)(y_i - y_j) > 0$ and discordant otherwise. Kendall's τ is defined as the difference between the probability of concordance minus the probability of discordance. Given a copula C the relation

$$\tau = 4E(C(U, V)) - 1 \tag{46}$$

holds. For most one-parameter copulas there exists a one-to-one relationship between its parameter θ and Kendall's τ.

Whereas Kendall's τ measures the overall dependence, a further popular dependence measure is the coefficient of tail dependence, which captures the dependence in the extremes of the distribution. The coefficient of lower and upper tail dependence are given by

$$\lambda_L = \lim_{u \downarrow 0} P[U < u | V < u] = \lim_{u \downarrow 0} \frac{C(u, u)}{u} \tag{47}$$

$$\lambda_U = \lim_{u \uparrow 1} P[U > u | V > u] = \lim_{u \uparrow 1} \frac{1 - 2u + C(u, u)}{1 - u} \tag{48}$$

The upper and lower tail dependence coefficients can often be expressed in terms of the copula parameters.

Examples of Copulas

Below we give some possibilities for the choice of the (one-parameter) copula density c, along with appropriate transformations Ψ we use for the SCAR specification. Contour plots of these copulas with standard normal margins and an overall dependence level corresponding to $\tau = 0.5$ can be found in Figure 6.

Frank Copula
The density of the Frank copula is given by

$$c_{\text{Frank}}(u, v; \theta) = \frac{\exp((1 + u + v)\theta)(\exp(\theta) - 1)\theta}{\{\exp(\theta) + \exp((u + v)\theta) - \exp(\theta + u\theta) - \exp(\theta + v\theta)\}^2} \tag{49}$$

Kendall's τ related to θ through $\tau = 1 - \frac{4(1 - D_1)(\theta)}{\theta}$, where D is the Debye function $D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} \mathrm{d}t$. It does not exhibit tail dependence. The fact that the parameter of the Frank copula lies in $(-\infty, \infty) \backslash 0$ 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

DOI: 10.1002/jae

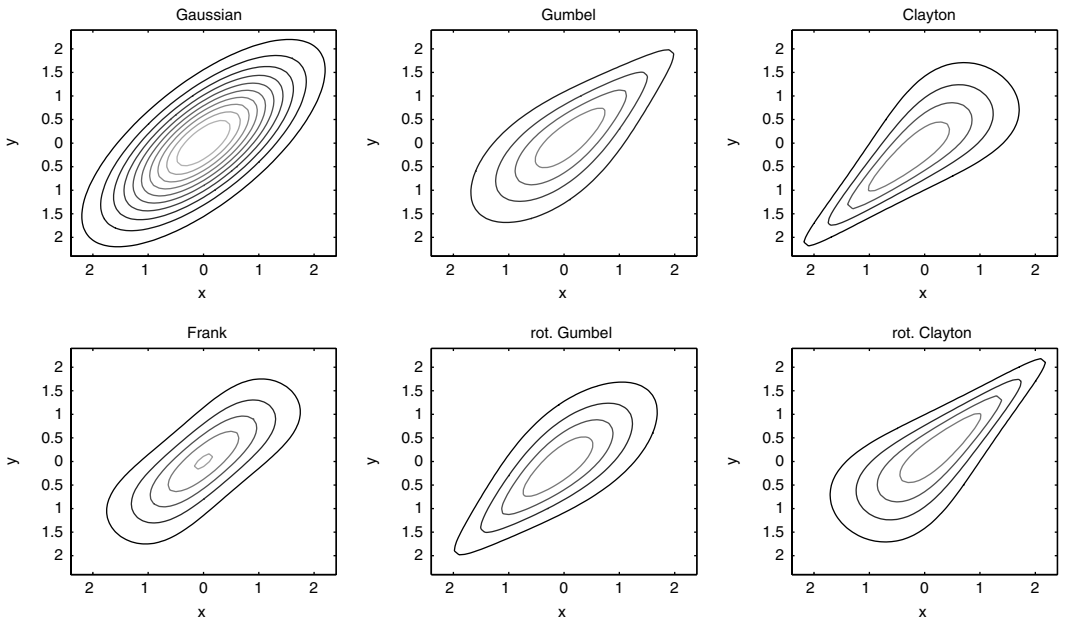

Figure 6. Contour plots. Note: contour plots of the copulas we consider with standard normal margins and $\tau = 0.5$

AR(1) process. The Frank copula belongs to the family of Archimedean copulas and it implies a rotationally symmetric dependence structure. It allows for both positive and negative dependence.

Clayton Copula
The density of the Clayton copula is

$$c_{\text{Clayton}}(u, v; \theta) = u^{(-1-\theta)} v^{(-1-\theta)} (u^{-\theta} + v^{-\theta} - 1)^{(-2-1/\theta)} (1 + \theta) \tag{50}$$

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 relations $\tau = \frac{\theta}{\theta + 2}$ and $\lambda_L = 2^{-1/\theta}$ hold, whereas $\lambda_U = 0$ for all θ. The Clayton copula is also Archimedean, but it is not rotationally symmetric and it only allows for positive dependence.

Gumbel Copula
The density of the Gumbel copula is

$$c_{\text{Gumbel}}(u, v; \theta) = \frac{\{\log(u)\log(v)\}^{(\theta-1)} \{[(-\log(u))^{\theta} + (-\log(v))^{\theta}]^{1/\theta} + \theta - 1\}}{[(-\log(u))^{\theta} + (-\log(v))^{\theta}]^{(2-1/\theta)} uv} \times \exp\{[(-\log(u))^{\theta} + (-\log(v))^{\theta}]^{1/\theta}\} \tag{51}$$

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. For the Gumbel copula $\tau = 1 - \frac{1}{\theta}$, $\lambda_U = 2 - 2^{1/\theta}$ and $\lambda_L = 0$. It also belongs to the Archimedean class and only allows for positive dependence.

DOI: 10.1002/jae

Gaussian Copula

Defining $x = \Phi^{-1}(u)$ and $y = \Phi^{-1}(v)$, where $\Phi(\cdot)$ is the CDF of a standard normal random variable, the density of the Gaussian copula is given by

$$c_{\text{Gaussian}}(u, v; \theta) = \frac{1}{\sqrt{1-\theta^2}} \exp\left(\frac{2\theta xy - x^2 - y^2}{2(1-\theta^2)} + \frac{x^2 + y^2}{2}\right) \tag{52}$$

with $\theta \in (-1, 1)$. For the transformation we 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). The expression for Kendall's tau is $\tau = \frac{2}{\pi}\arcsin(\theta)$ and, like the Frank copula, the Gaussian copula has no tail dependence.

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° 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 $c_\theta(1 - u, 1 - v)$.

DOI: 10.1002/jae