Copulae for High-Frequency Data

Matthias R. Fengler
University St. Gallen, Fachbereich Financial Econometrics,
Bodanstrasse 6,
CH-9000 St. Gallen, Switzerland,
E-mail: matthias.fengler@unisg.ch

Ostap Okhrin
Humboldt-Universität zu Berlin, Ladislaus von Bortkiewicz Chair of Statistics,
Spandauer Straße 1,
10178 *Berlin, Germany*
E-mail: ostap.okhrin@wiwi.hu-berlin.de

Introduction

The modelling of the realized volatility and covariances has became a corner-stone in financial risk management and portfolio optimization. The first brick in intraday modelling has been laid by Andersen and Bollerslev (1998). Afterwards a series of papers providing estimators of the realized volatility appeared and a detailed investigation has been provided. More recently the attention shifted towards modelling the joint behavior of returns. A variety of time-series models with Gaussian innovations have thus been proposed. Apart of modelling volatility of returns separately, Andersen, Bollerslev, Diebold and Labys (1999) introduce the notion of realized correlation and realized covariance which helps to model high dimensional high frequency joint behaviour. Recent works dedicated to modelling the realized covariance/correlation matrix are by Hayashi and Yoshida (2005) and Audrino and Corsi (2009). Most crucial point in these papers, the residuals are assumed to be normal and the dependency is that of a normal structure. The Gaussian structure leads to imprecise modelling of the joint tails of the distribution as it has no upper nor lower tail dependency. Recent evidence shows, that a series of financial data are not following normal dependency, Lee and Long (2009). For static, non time dependent models a normal distribution can be replaced by the copula function, see Sklar (1959), Nelsen (2006), etc. Modelling with time varying copulae based innovations is increasing in popularity, see Patton (2004), Chen ad Fan (2006), Härdle, Giacomini, Spokoiny (2009), Härdle, Okhrin, Okhrin (2011), etc. To the authors' knowledge the only paper that is using intraday data to model the dependency is Breymann, Dias and Embrechts (2003). In this paper we propose a novel approach of time varying copulae using intraday data. Current work discusses different models of realized covariances and appropriate realized variances. The quality of this approach is compared by the out of sample forecast using an adaptive estimation of the copula model, and the simple rolling window applied to the residuals from some simple time varying model.

The paper is organized as follows. First we provide a discussion on different methods of estimation of the realized volatility and covariances. The second section deals with copula theory, mainly with the recent time varying models. The Section 3 states the main result of our model. A simulation study and an empirical part follows.

Realized volatility and correlation

In the paper we consider a d-dimensional process of log-prices $P = (P_1, \dots, P_d)^\top$ observed on the interval $[0;T]$, where T is considered to be an integer number measured in days. The observation times for the asset j at day t are denoted by $\mathfrak{t}_{j,t,1}, \mathfrak{t}_{j,t,2}, \dots$ and are assumed to be strictly increasing, from which follows that our observations are given by $P_{j,t,i} = P_{j,\mathfrak{t}_{j,t,i}},\ j = 1, \dots, d,\ i = 1, \dots, N_j,\ t = 1, \dots, T$. The observation points differ for all processes which leads to a different number of observations. The observed log-price process P is assumed to be driven by the efficient price process Y, which in our setup is a Brownian semimartingale with a jump component (BSMJ)

$$dY_t = \mu_t dt + \sigma_t dW_t + dJ_t$$

with μ_t being vector of predictable drifts, and σ_t cadlag volatility matrix process, and W_t is a vector of independent Brownian motions. The market microstructure effect is modeled through an additive component: $P_{j,t,i} = Y_{j,t,i} + U_{j,t,i}$, with $\mathsf{E}(U_{j,t,i}) = 0$, $\sum_h |h\Omega_h| < \infty$, where $\Omega_{j,h} = \mathrm{Cov}(U_{j,t,i}, U_{j,t,i-h})$. Later on we will estimate the quadratic variation of Y, i.e. $[Y] = \int_0^1 \Sigma_u du$, with $\Sigma = \sigma\sigma^\top$.

For the estimation of $[Y]$ from discrete, non-synchronous and noisy price observations, we will use the realized kernel method introduced by Barndorff-Nielsen, Hansen, Lunde, Shephard (BNHLS) (2008) and is considered to be consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. The multivariate realized kernel is defined as

$$K(P) = \sum_{h=-H}^{H} k\left(\frac{|h|}{H+1}\right)\Gamma_h,$$

with Γ_h being a matrix of autocovariances given by

$$\Gamma_h = \begin{cases} \sum_{j=|h|+1}^{n} p_j p_{j-h}^\top, & h \geq 0 \\ \sum_{j=|h|+1}^{n} p_{j-h} p_j^\top, & h < 0 \end{cases},$$

and $k(x)$ being the weight function of the Parzen kernel. For a discussion on the bandwidth selection we refer to the web appendix of BNHLS (2008).

Time Varying Copula

The advantage of the copula is that it allows to split the multivariate distribution into the margins and a pure dependency component it captures the dependency between variables eliminating the impact of the marginal distributions. Formally copulae where introduced in Sklar (1959). The main result states that if F is an arbitrary d-dimensional continuous distribution function of the random variables $X_1, \dots, X_d$, then the associated copula is unique and defined as the continuous function $C : [0,1]^d \to [0,1]$ which satisfies the equality

$$C(u_1, \dots, u_d) = F\{F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\}, \quad u_1, \dots, u_d \in [0,1],$$

where $F_1^{-1}(\cdot), \dots, F_d^{-1}(\cdot)$ are the quantile functions of the corresponding marginal distributions. One of the classes that overcomes the drawback of elliptical copulae of having no explicit form of the cdf is the class of Archimedean copulae

$$(1) \quad C(u_1, \dots, u_k) = \phi\{\phi^{-1}(u_1) + \cdots + \phi^{-1}(u_d)\}, \quad u_1, \dots, u_d \in [0,1],$$

where $\phi \in \mathfrak{L} = \{\phi : [0;\infty) \to [0,1] \,|\, \phi(0) = 1,\ \phi(\infty) = 0;\ (-1)^j \phi^{(j)} \geq 0;\ j = 1, \dots, \infty\}$. The function ϕ is called the generator of the copula and commonly depends on a single parameter θ. A detailed review of the properties of Archimedean copulae can be found in McNeil and Neslehova (2008).

In this paper we consider a classical rolling window approach (with the window width w) and the approach based on a local constant copula approximation. Corresponding theory and applications may be found in Giacomini, Härdle, Spokoiny (2009) and Cizek, Härdle, Spokoiny (2009).

Let θ_t denote the time varying but unknown copula parameter. The idea is to select for each time point t_0 an interval I at which θ_t can be well approximated by a constant θ^*. The discrepancy between two copulae $C(\cdot;\theta)$ and $C(\cdot;\theta')$ is measured by the Kullback-Leibler divergence $\mathcal{K}\{C(\cdot;\theta'),C(\cdot;\theta)\} = \mathsf{E}_{\theta'} \log \frac{c(\cdot;\theta')}{c(\cdot;\theta)}$, where c is the copula density. The aim is to select I as close as possible to the so-called "oracle" choice interval I_{k^*}, defined as the largest interval $I = [t_0 - m_{k^*}; t_0]$, for which the small modelling bias condition (SMB) is fulfilled, i.e. $\triangle_I(\theta) = \sum_{t\in I} \mathcal{K}\{C(\cdot;\theta_t),C(\cdot;\theta)\} \leq \triangle$, for some $\triangle \geq 0,\ \theta$. The LCP is based on sequentially testing the hypotheses of homogeneity on intervals I_k. We select I_k with $k = -1, 0, 1, \ldots$ as the sequence of intervals $I_k \subset I_{k+1}$, starting with $k = 1$. If there are no change points in $\mathcal{T}_k \subset I_k \setminus I_{k-1}$ we accept I_k as an interval with constant copula parameter and structure. At the next step we take $\mathcal{T}_{k+1}$ and test it for homogeneity. We repeat the steps until rejection or the largest possible interval I_K is accepted, leading to an interval $I_{\hat{k}}$.

Realized Copula

From the Hoeffdings lemma it is known that the covariance σ_{ij} for random variables X_i and X_j with marginal distributions F_i and F_j, with finite first and second moments and the copula C_θ is given by:

$$\sigma_{ij}(\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C_\theta\{F_i(x_i), F_j(x_j)\} - F_i(x_i)F_j(x_j)dx_i dx_j$$

Following theory of the intraday data we suppose the daily returns $r_t \sim \mathsf{N}(0; I\hat{V})$

Usually this integral has no explicit form, except as for the normal distribution, where one gets $\sigma = \theta$. For $d = 2$ one has to invert $\sigma_{i,j}(\theta)$ to get the estimator of the copula parameter $\theta_{i,j} = \sigma_{i,j}^{-1}(\theta)$. This can be problematic, when one deals with a high dimensional problem, simple Archimedean copulae depend only on the one parameter, thus, there should be a method, allowing some approximation of the copula parameter. Let us first define the function which measures losses from the estimator of the volatility, without any model assumption (in the copula sense) and based on the copula: $g_{i,j}(\theta) = \sigma_{i,j} - \sigma_{i,j}(\theta)$. Here we propose several estimators of the copula parameter based on the realized covariance

1. An average of the all possible copula parameters based on the pairs

$$\hat{\theta}_{av} = \frac{2}{d(d-1)} \sum_{i\leq j} \hat{\theta}_{i,j}$$

2. For the second method we propose one-stage method of moments estimator with the identity weight matrix $\mathbf{W} = \mathbf{I}_{d\times d}$. By denoting $\mathbf{g}(\theta) = \{g_{i,j}(\theta)\}_{i\leq j}^\top$ the estimator is given by

$$\theta_{GMMI} = \arg\min_\theta \mathbf{g}^\top(\theta)\mathbf{g}(\theta) \tag{2}$$

This method is equivalent to the first one in the two dimensional case.

3. The two-stage GMM estimator with $\mathbf{W} = \{E[\mathbf{g}(\theta_{GMMI})\mathbf{g}^\top(\theta_{GMMI})]\}^{-1}$, given by

$$\hat{\theta}_{GMM} = \arg\min_\theta \mathbf{g}^\top(\theta)\mathbf{W}\mathbf{g}(\theta) \tag{3}$$

	HL		HLgmm		LinCor	
	mean	R.mse	mean	R.mse	mean	R.mse
$\xi = 0,\ \lambda = (30, 45, 60)$						
RC_{390}	1.349	0.651	1.340	0.662	1.356	0.645
RC_{26}	1.929	0.300	1.980	0.364	1.938	0.286
K	1.843	0.199	1.856	0.214	1.836	0.201
$\xi = 0.001,\ \lambda = (30, 45, 60)$						
RC_{390}	1.176	0.824	1.170	0.830	1.183	0.817
RC_{26}	1.842	0.320	1.907	0.382	1.857	0.306
K	1.667	0.348	1.706	0.321	1.668	0.346

Table 1: Simulation results for the 3-dimensional realized copula using the factor SV model with nonsynchronous observations and measurements noise. Bias and root mean square error are reported.

4. An estimator based on the transformation of the linear correlation coefficient of the normal distribution to the Kendall's τ, and the consequent transformation of the Kendall's τ to the copula parameter. Same as for the first method, this works in two dimensions. For $d > 2$ we average all bivariate estimates

$$\tau_{Gauss,i,j} = 2\arcsin(\sigma_{i,j})/\pi, \quad \hat{\theta}_{Gumbel,i,j} = \frac{1}{1-\hat{\tau}}, \quad \hat{\theta}_K = \frac{2}{d(d-1)} \sum_{i \leq j} \hat{\theta}_{Gumbel,i,j}$$

Empirical findings show, that this is almost equivalent to the first method.

Simulation Study

In this simulation study we considered almost the same model as in BNHLS(2011) with the only one difference of the copula based dependence:

$$\begin{aligned} X &= Y + U,\ dY = \mu\, dt + \sigma dW_c \\ \sigma^{(i)} &= \exp\{\beta_0 + \beta_1 \varrho^{(i)}\} \text{ with } d\varrho^{(i)} = \alpha\varrho^{(i)}dt + dB^{(i)} \text{ and } \varrho^{(i)}(0) = \mathsf{N}(0, (-2\alpha)^{-1}) \\ U_j|\sigma, Y &\sim \mathsf{N}(0, \omega) \text{ with } \omega^2 = \xi^2 \sqrt{N^{-1}\sum_{j=1}^{N} \sigma^{(i)4}(j/N)} \\ (\mu, \beta_0, \beta_1, \alpha, \varrho, \theta, \xi^2) &= (0.03, -5/16, 1/8, -1/40, -0.3, 2.0, \{0, 0.001\}) \end{aligned}$$

where dW_c is the copula dependent vector of Brownian motions, what means, that increments are copula dependent with parameter θ. U is an additive noise, and ξ^2 is the signal-to-noise ratio. For the simulation we apply Euler scheme with the sample size $N = 23400$ which corresponds to trading every second in 6.5 hours. To introduce nonsynchronous trading we implement d independent Poisson processes with intensity λ, for $d = 2$ we use $\lambda = (30, 45, 60)$. Bias and root mean square error for all models in two and three dimensional case are provided in the Table 1.

Empirical Study

The empirical part of this work is based on data taken from NYSE's Trades and Quotes (TAQ) database, for Citigroup (C), IBM and Alcoa (AA) for the period from 01.01.2007 till 31.07.2009 which

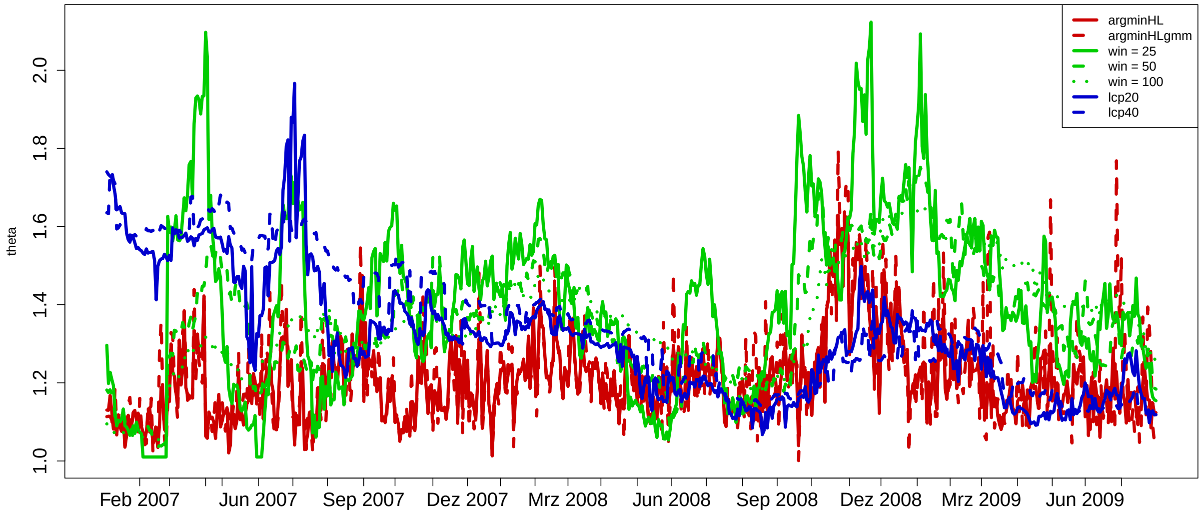

Figure 1: Estimation of the copula parameter based on all methods discussed above.

covers the very turbulent periods of the financial crisis in the mid of 2008. This period covers 638 days in total and daily transactions from 9:30 till 16:00. We apply the cleaning procedure proposed by BNHLS (2008).

Let us denote by w the portfolio, which is represented by the number of assets for a specified stock, $w = \{w_1, \dots, w_d\}$, $w_i \in \mathbb{Z}$. The value V_t of the portfolio w is given non-recursively by $V_t = \sum_{j=1}^{d} w_j S_{j,t}$ and the profit and loss function by $L_{t+1} = \sum_{j=1}^{d} w_j S_{j,t} \{\exp(X_{j,t+1}) - 1\}$. For every model we simulate returns assuming no autocorrelation, by taking the respective dependency parameter (realized copula, copula from LCP or rolling window) and appropriate volatility. If we take the rolling window or LCP method, volatility is just the variance of the returns taken from the respective interval. For the realized copula volatility is a realized volatility. The distribution function of L, dropping the time index, is given by $F_L(x) = P(L \leq x)$. As usual the VaR at level α from a portfolio w is defined as the α-quantile from F_L i.e. $\text{VaR}(\alpha) = F_L^{-1}(\alpha)$. F_L depends on the d-dimensional distribution of log-returns F_X. Thus, modelling their distribution is essential to obtain the quantiles from F_L. Afterwards backtesting is used to evaluate the performance of the specified copula family $\mathcal{C}$. The estimated values for the VaR are compared with the true realisations $\{l_t\}$ of the P&L function, an *exceedance* occuring for each l_t smaller than $\widehat{VaR}_t(\alpha)$. The ratio of the number of exceedances to the number of observations gives the *exceedances ratio* $\hat{\alpha}$: $\hat{\alpha} = \frac{1}{T-r} \sum_{t=r}^{T} \mathbf{I}\{l_t < \widehat{VaR}_t(\alpha)\}$.

REFERENCES (RÉFERENCES)

Andersen, T., Bollerslev, T., Diebold, F.X. and Labys, P. (1999), (Understanding, Optimizing, Using and Forecasting) Realized Volatility and Correlation, Manuscript, Northwestern University, Duke University and University of Pennsylvania. Published in revised form as "Great Realizations," Risk, March 2000, 105-108.

Andersen, T., Bollerslev, T., Diebold, F.X. and Labys, P. (2003), Modeling and Forecasting Realized Volatility, *Econometrica*, **71**, 529-626.

Audrino F., Corsi F. (2010). Modelling Tick-by-Tick Realized Correlations. *Computational Statistics & Data Analysis* **54**11, 2372-2382.

Barndorff-Nielsen, O., Hansen, P., Lunde, A., and Shephard, N. (2008), Realised Kernels in Practice: Trades and Quotes, *Econometrics Journal*, **4**, 132.

	1%	5%	10%
lcp ($m_0 = 20$)	0.0470 (46.536)	0.095 (22.308)	0.134 (7.827)
lcp ($m_0 = 40$)	0.0564 (66.756)	0.100 (26.654)	0.145 (13.204)
rol ($w = 25$)	0.0297 (16.482)	0.087 (15.798)	**0.123** (3.769)
rol ($w = 50$)	0.0376 (28.850)	0.086 (14.611)	0.125 (4.265)
rol ($w = 100$)	0.0454 (43.399)	0.089 (17.024)	0.134 (7.827)
HLgmm	0.0266 (12.261)	0.086 (14.611)	0.139 (9.975)
HL	0.0297 (16.482)	0.089 (17.024)	0.144 (12.358)
K	0.0282 (14.314)	0.087 (15.798)	0.142 (11.538)
Gauss	**0.0031** (4.150)	**0.023** (11.630)	0.059 (13.361)

Table 2: $\hat{\alpha}$ for different models, for given $\alpha \in \{0.01, 0.05, 0.1\}$. With bold we represent closest $\hat{\alpha}$ for given α. The study has been done for 10^6 simulations. In the brackets we represent the Kupiec test statistic, which should χ^2_1 distributed.

Barndorff-Nielsen, O., Hansen, P., Lunde, A., and Shephard, N. (2011), Multivariate Realized Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-synchronous Trading, *Journal of Econometrics*, **162** (2), 149-169.

Breymann, W., Dias, A., Embrechts, P.: Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance* **3**(1), 1-16

Chen, X. and Fan, Y. (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspesification, *Journal of Econometrics* **135**: 125154.

Cizek, P., Härdle, W. and Spokoiny, V. (2009). Adaptive pointwise estimation in time-inhomogeneous conditional heteroscedasticity models, *Econometrics Journal* **12**(2): 248-271.

Harris, F., McInish, T., Shoesmith, G., and Wood, R. (1995), Cointegration, Error Correction and Price Discovery on Informationally-linked Security Markets, *Journal of Financial and Quantitative Analysis*, **30**, 563-581.

Härdle, W. K., Okhrin, O., Okhrin, Y., (2011). Dynamic Structured Copula Models, submitted

Hayashi, T., Yoshida, N., (2005), "On covariance estimation of non-synchronously observed diffusion processes, *Bernoulli* **11(2)**, 359379

Giacomini, E., Härdle, W. K. and Spokoiny, V. (2009). Inhomogeneous dependence modeling with time-varying copulae, *Journal of Business and Economic Statistics* **27**(2): 224-234.

McNeil, A. J. and Neslehova, J. (2009). Multivariate Archimedean copulas, d-monotone functions and l1 norm symmetric distributions, *Annals of Statistics* **37**(5b): 30593097.

Nelsen, R. B. (2006). An Introduction to Copulas, Springer Verlag, New York.

Patton, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation, *Journal of Financial Econometrics* **2**: 130168.

Sklar, A. (1959). Fonctions dé repartition á n dimension et leurs marges, *Publ. Inst. Stat. Univ. Paris* **8**: 299-231.

Spokoiny, V. (2009). Multiscale local change point detection with applications to value-at-risk, *The Annals of Statistics* **37**(3): 14051436.