Centre for Computational Finance and Economic Agents

WP031-08

Working Paper Series

Wing Lon Ng and Mark Trede

High-frequency Index Returns: The Stylized Facts Revised

September 2008

University of Essex

CCFEA

www.ccfea.net

High-frequency Index Returns: The Stylized Facts Revised*

Wing Lon Ng† Mark Trede‡
University of Essex University of Münster

September 2008

*The authors are very grateful to conference participants at Singapore Econometrics Group Meeting 2006, at ESEM 2007 and at CEF 2008 Conference for their helpful discussion. In particular, they are indebted to Olivier Scaillet, Ike Mathur, Timothy Crack and Robert Hudson for their suggestions. This research was supported by the Deutsche Forschungsgemeinschaft through research grant "NG65/3-1" and through the SFB 649 "Economic Risk" at the Humbolt University Berlin.

†Centre for Computational Finance and Economic Agents (CCFEA), University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom, E-Mail: wlng@essex.ac.uk, Tel.: +44 (0) 1206 874684

‡**Corresponding author**: Institute of Econometrics, University of Münster, Am Stadtgraben 9, 48143 Münster, Germany, E-mail: mark.trede@uni-muenster.de, Tel.: +49 (0) 251 8325006

High-frequency Index Returns: The Stylized Facts Revised

Abstract

This paper applies a nonparametric copula-based approach to analyze the first-order autocorrelation of returns in high frequency financial time series and revise their stylized facts. Using EUREX tick data from the German DAX index, it can be shown that the short-term temporal dependence structure of price movements is neither negatively nor positively autocorrelated as often claimed as a stylized fact in the literature. The dependence structure between consecutive returns as well as the sign of their correlation is highly dependent on the sampling interval.

For a long time, the literature on financial econometrics and quantitative finance dealt with daily asset prices or returns and their volatility. The increasing availability of high-frequency and ultra high-frequency data shifted the research focus towards the micro structure of asset markets (Andersen, Bollerslev, and Diebold 2005, Russell and Engle 2005). A well-established stylized fact of index (or portfolio) returns is their positive first-order autocorrelation (Mech 1993, Boudoukh, Richardson, and Whitelaw 1994, Koutmos 1997, Ogden 1997). The positive index autocorrelation is in contrast to the negative autocorrelation generally found in individual stocks (Dacorogna, Gençay, Müller, Olsen, and Pictet 2001, Goodhart and Figliuoli 1991, Tsay 2002). According to the classical transaction model of Roll (1984) for stock markets, negative autocorrelation of returns are caused by the bid-ask-bounce. Bollerslev and Domowitz (1993), for example, describe this phenomenon as an outcome of market makers skewing the spread into particular directions when they have order imbalances. A well-known microstructure model reconciling negative individual autocorrelation with positive index autocorrelation is based on nonsynchronous trading (Lo and MacKinlay 1990). Other prominent explanations of the positive index autocorrelation are time-varying expected returns or risk premia Conrad and Kaul (1989), and nonsynchronous information flows Badrinath, Kale, and Noe (1995).

The empirical studies on the temporal dependence structure of returns apply the Pearson-type autocorrelation coefficient, which by construction only measures the "aggregated" or "overall" linear dependence. Nonlinear effects cannot be revealed and are therefore neglected in most studies. However, a negative linear correlation is by no means incompatible with a nonlinear positive dependence. Attempts to discover nonlinearities in financial data are for instance phase portraits, often used in the physical sciences to detect chaotic phenomena in dynamic systems (Szpiro 1998). In its simplest version, it represents a scatterplot, in which a time series is plotted against its lagged values, see Wöhrmann (2005) for an overview. Another drawback of many previous empirical studies is that they do not take into account the effect of the sampling frequency. However, the intertemporal dependence structure might, of course, vary with the sampling interval: There could be positive dependence of returns when sampling in one-second intervals while the dependence could be negative for sampling intervals of one minute (Cai, Hudson, and Keasey 2003).

In fact, the sign and the magnitude of the first order autocorrelation has become an important issue in the real world trading practice as many submissions in electronic markets are not executed by human traders but algo-traders, i.e., algorithmic trading programs that initiate certain strategies depending on high-frequency price movements. In most cases, the software is trained to forecast the next transaction prices and the pre-specified trading rules are kept rather simple. As the trading decision has to be made quickly in order to benefit from short-term statistical arbitrage, an accurate dependence measure is required.

The objective of this paper is to re-investigate the empirical properties of high-frequency returns and bring out new stylized facts of the first order intertemporal dependence. We avoid both the restriction to linear dependence and the limitation to a single sampling frequency. A flexible copula-based approach is developed to describe the temporal dependence of high-frequency EUREX returns. Since time series can be regarded as drawings from a multivariate distribution, one may split this distribution into two components: (a) the marginal distributions and (b) the dependence structure determined by its copula. This paper focuses on univariate stationary return processes, in which the copula controls the temporal dependence. The unconditional distributions are left unspecified, allowing all kinds of marginal distributions. A salient feature of very high frequency return data is a singularity at zero return. If there is no trading during an interval the price does not change and the return is zero. This happens with positive (and at very short intervals with rather high) probability. We carefully describe statistical methods capable to adequately take into account the singularity.

The copulas are estimated nonparametrically to detect the first order temporal dependence of the data exploratively. We find that the stylized fact of positive first order autocorrelation is too crude and should be broken down into the following more sophisticated stylized facts: (a) There are always patterns of *both* (not necessarily linear) positive and negative dependence, regardless of the sampling frequency. (b) The overall direction of the dependence structure depends on the sampling frequency. When sampling at high frequencies (i.e. short sampling intervals), a globally positive dependence with local negative patterns is visible. When sampling at lower frequencies (i.e. longer sampling intervals) the picture reverses and we observe a globally negative dependence with local positive patterns. These complex dependence structures show a high degree of nonlinearity that cannot be revealed by the linear correlation coefficient. Traders (either human or algorithmic) relying on a linear measure will systematically cut out the existence of nonlinearly dependent price movements and, thus, may suffer from losses due to an under-estimation of price jumps in the high-frequency transaction process.

The outline of this paper is as follows. Section 1 introduces the copula approach with special emphasis on a rigorous treatment of singularities in the margins and singular components of the copula. Section 2 describes the estimation procedure. In Section 3, the data and our results are presented. Section 4 concludes.

1 Copulas and Singularities

This section briefly introduces the most important definitions and properties of copulas. In finance, many studies using copulas focus on the contemporaneous dependence between two or more random variables, see Fermanian and Scaillet (2005) for an overview. In contrast, in this paper we will model the temporal dependence structure of a univariate time series $\{X_t\}_{t=1}^{N}$ via copulas (see Chen and Fan (2006a), Chen and Fan (2006b) and Patton (2006)). For a comprehensive survey of the basic theory of copulas, the reader is referred to Joe (1997), Mari and Kotz (2001), Cherubini, Luciano, and Vecchiato (2004) and, in particular, Nelsen (1990), chap. 2, to which we refer in the following. Since high frequency data usually have a significant proportion of observations with zero price changes, any realistic model of the theoretical return distribution needs a singular component. Strictly speaking, the existence of price ticks implies that return distributions are to be modelled as discrete random variables with countably infinite possible outcomes. We do, however, stick to the usual simplification and model return distributions as continuous random variables – apart of course from the important singular component at zero.

Consider two random variables X and Y with marginal distribution functions $F_X(x) = P(X \leq x)$ and $F_Y(y) = P(Y \leq y)$ and joint distribution function $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$. The random vector (X, Y) may have singular and absolutely continuous components. Denote the generalized inverse distribution function (generalized quantile function) of X as

$$F_X^{-1}(p) = \inf\{x : F_X(x) \geq p\}$$

for $0 < p < 1$, and similarly for F_Y^{-1}. While the distribution functions are right-continuous, the generalized inverse distribution functions are left-continuous. In particular, they are continuous if the distribution function is strictly increasing.

A copula is a bivariate distribution function with standard uniform margins. Sklar (1959) proves that there always exists a copula C such that for all $x, y \in \mathbb{R}$

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y)). \tag{1}$$

The copula C is unique on $\text{Ran}F_X \times \text{Ran}F_Y$. Restricting the domain to $\text{Ran}F_X \times \text{Ran}F_Y$, the copula is called subcopula and denoted C'. Subcopulas can be extended to copulas, but only in a non-unique way. If F_X and F_Y are continuous then C is unique on the unit square and $C' = C$. Using the generalized inverses the following holds for all $(u, v) \in \text{Ran}F_X \times \text{Ran}F_Y$,

$$C'(u, v) = F_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v)). \tag{2}$$

Note that the equality $C(u, v) = F_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v))$ does not hold on the entire unit square if there are singularities in X or Y. Of course, the subcopula C' defined by (2) can be extended to a copula C.

Any copula C can be decomposed into an absolutely continuous component C_A and a singular component C_S

$$C(u, v) = C_A(u, v) + C_S(u, v)$$

where

$$\begin{aligned} C_A(u, v) &= \int_0^u \int_0^v \frac{\partial^2 C(s, t)}{\partial s \partial t} dt ds \\ C_S(u, v) &= C(u, v) - C_A(u, v). \end{aligned}$$

By construction, the second derivative of the copula with respect to both arguments is defined almost everywhere on the unit square. Since both margins of a copula are standard uniformly distributed, the singular component C_S does not have atoms. Note that neither C_A nor C_S needs to be a copula. In particular, it is possible that $C_A = 0$ (in which case C is a singular copula) or $C_S = 0$ (then C is an absolutely continuous copula). If the copula is absolutely continuous the second derivative $\partial^2 C(u,v)/\partial u \partial v$ is called copula density and denoted $c(u,v)$. For simplicity we call $\partial^2 C(u,v)/\partial u \partial v$ copula density even if $C_A(1,1) < 1$. The simplest way to extend a subcopula to a copula is to assume that the copula density is constant on the complement of $\mathrm{Ran}F_X \times \mathrm{Ran}F_Y$ (i.e. bilinear interpolation, see Nelsen (1990)).

Singular components can be thought of as curves in the unit square on which the random variable $(F_X(X), F_Y(Y))$ lies with positive probability; a well known example is the Marshall-Olkin copula (Marshall and Olkin 1988, Mari and Kotz 2001). However, each single point of the curves has zero probability, i.e. no atoms. Note that vertical and horizontal curves are not possible as they would imply a singular component of the marginal distributions. Hence, the copula of a return distribution (to be discussed in more detail in Section 4) has no singular components even though there is a positive probability of zero returns, i.e. a singularity in the marginal distributions.

If the random variables X and Y are interpreted in a time series context, the copula can be used to define a first-order Markov process (Darsow, Nguyen, and Olsen 1992). In general, time series models are constructed as $X_t = g\left(X_{t-1}, X_{t-2}, \ldots, \varepsilon_t\right)$, where the current variable is explained as a function of past observations and a random innovation ε_t. Applying the copula concept has the advantage that the temporal dependence structure of the stochastic process can be modeled in a flexible way without restrictive assumptions such as linearity (Savu and Ng (2005)).

Let $\{X_t\}_{t=0,1,\ldots}$ denote a stochastic process in discrete time. We assume that $\{X_t\}_{t=0,1,\ldots}$ is strictly stationary, implying $F_{X_t}(x) = F_{X_{t-1}}(x)$ for all $x \in \mathbb{R}$ and $t = 0, 1, \ldots$, and write F_X for the marginal distribution function. The joint distribution function of consecutive observations X_{t-1} and X_t is simply denoted as F

$$\begin{aligned} F\left(x_{t-1}, x_t\right) &= P\left(X_{t-1} \leq x_{t-1}, X_t \leq x_t\right) \\ &= C\left(F_X\left(x_{t-1}\right), F_X\left(x_t\right)\right). \end{aligned}$$

The copula C can also be regarded as the joint distribution function of the transformed random variables $U = F_X\left(X_{t-1}\right)$ and $V = F_X\left(X_t\right)$. We will now turn to the question of how to estimate the copula and the marginal distributions.

2 Estimation

As we are mainly interested in the intertemporal dependence structure of the returns, no particular parametric form for the marginals is assumed. Any problems concerning misspecification or overfitting are ruled out. Since recent studies have shown that temporal aggregation and the sampling frequency have an essential impact on the resulting stochastic process (Lee, Gleason, and Mathur 1999, Cai, Hudson, and Keasey 2003, Aït-Sahlia and Mykland 2003), one must take these effects into account. Therefore, the estimation is not only performed on the original data observed at the 1 second interval, but also on various thinned return processes with increasing observation intervals from 2 seconds up to 30 minutes. Let P_t be the price of an asset at time t, observed at a certain sampling frequency, then $\{R_t\}_{t=1}^{N}$ with

$$R_t = (P_t - P_{t-1})/P_{t-1} \times 100\% \tag{3}$$

represents the return process. When estimating copulas it is immaterial whether the returns are defined in a discrete or continuous fashion (i.e., $\ln P_t - \ln P_{t-1}$). Since discrete returns are a strictly increasing transformation of continuous returns their copulas are identical. The differences only show up in the marginal distribution.

Due to stationarity, the marginal distribution function $F_R(r) = P(R_t \leq r)$ is time-invariant. It is estimated nonparametrically using the empirical distribution function

$$\hat{F}_R(r) = \frac{1}{N+1} \sum_{t=1}^{N} \mathbf{1}(R_t \leq r);$$

the empirical distribution function is re-scaled by the asymptotically negligible factor $N/(N+1)$ in order to avoid computational problems at the boundaries of the copula. Singularities in the margins manifest themselves in $\hat{F}_R$ as steps larger than $1/(N+1)$. Usually, the copula estimation with nonparametrically estimated margins proceeds by using the empirical distribution function to map the observations into the unit interval,

$$U_t = \hat{F}_R(R_t). \tag{4}$$

This approach needs some modifications if there are ties (i.e., singularities). Obviously, $(N+1)\hat{F}_R(R_t)$ is the rank of R_t in $R_1, \ldots, R_N$ if the value of R_t appears only once. If the value of R_t is a singularity, $(N+1)\hat{F}_R(R_t)$ is the maximal rank of all observations at (or below) the singularity. Hence, U_t is no longer evenly spaced on the unit interval. This has to be taken into account for the statistical inference.

Statistical inference for copulas is by now well developed (Charpentier, Fermanian, and Scaillet 2007, Chen and Huang 2007). However, most studies assume that the random variables under consideration are absolutely continuous which is not the case for high-frequency returns. We present two modified nonparametric estimation procedures that can cope with singularities in the marginal distributions.

First, only the subcopula (2) is estimated by explicitly restricting its domain to $\mathrm{Ran}F_X \times \mathrm{Ran}F_Y$. The density in each part of the domain is estimated as if it were a separate copula, observing its bounded support. Weights are attached to each part according to its proportion of observations. The subcopula is then extended to a copula by bilinear interpolation. Note that we assume that there are no singularities in the subcopula C' but only in the marginal distribution F_R.

Second, the entire copula is estimated at once using a randomized bilinear interpolation. The mapping (4) does not spread the observations evenly over the unit interval if there are ties. A straightforward remedy is to assign random ranks to ties. Contrary to the common approach, the observation R_t is not mapped by $\hat{F}_R(R_t)$ but by

$$U_t = \frac{1}{N+1}\mathrm{Rank}_{\mathrm{rand}}(R_t), \tag{5}$$

where $\mathrm{Rank}_{\mathrm{rand}}(R_t)$ is the randomized rank in case of ties. In this way, the singularities are extended over the unit square in a randomized way akin to bilinear interpolation. In a different context Brockwell (2007) has suggested a generalized Rosenblatt transformation that is similar to our approach. Note that the randomization only affects those areas of the copula that are not uniquely defined. Other mappings are, of course, allowed as long as $U_1, \ldots, U_N$ are evenly distributed along the unit interval. Equation

(1) will hold for any such mapping.

Having transformed the observations R_t into U_t, $t = 1, \ldots, N$, one can estimate the copula density by means of an ordinary product kernel (Fermanian and Scaillet (2003)). However, since the copula is only defined on the unit square, one has to take the boundary bias into account that occurs when using fixed symmetric kernel functions. To resolve this problem, one can use the mirroring technique suggested by Gijbels and Mielniczuk (1990), or apply non-fixed beta kernels as proposed by Chen (1999). We opted for the latter technique because of its superior computation performance. Let

$$\begin{aligned} k(p, q, u) &= \frac{u^{p-1}(1-u)^{q-1}}{B(p, q)} \\ &= u^{p-1}(1-u)^{q-1} \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \end{aligned}$$

be the density of a random variable U with Beta(p, q)-distribution. The copula density can be estimated as (Härdle, Müller, Sperlich, and Werwatz 2003)

$$\hat{c}(u, v) = \frac{1}{N-1} \sum_{t=2}^{N} \left(K(u, h, U_{t-1}) \cdot K(v, h, U_t) \right)$$

where

$$K(u, h, r) = \begin{cases} k\left(\varphi(u), \frac{1-u}{h}, r\right) & \text{if} \quad u \in [0, 2h) \\ k\left(\frac{u}{h}, \frac{1-u}{h}, r\right) & \text{if} \quad u \in [2h, 1-2h] \\ k\left(\frac{u}{h}, \varphi(1-u), r\right) & \text{if} \quad u \in (1-2h, 1] \end{cases}$$

and

$$\varphi(u) = 2h^2 + 2.5 - \sqrt{4h^4 + 6h^2 + 2.25 - u^2 - \frac{u}{h}}.$$

In addition to its good computation performance, the beta kernel estimation has several other advantages (Charpentier, Fermanian, and Scaillet 2007, chap. 2): Beta kernels naturally match the support of the density to be estimated. The kernel's form changes smoothly depending on which part of the density is estimated. The beta kernel density estimator is unbiased even at the boundaries of the support. Further, using the nonparametrically estimated marginal distributions usually reduces the variance of the density estimator.

3 Data and empirical results

The high frequency data set D3047 containing the DAX performance index is extracted from the EUREX database. The sample includes observations at one second sampling frequency from 2^{nd} January until 28^{th} April 2006, observed for 90 trading days over 18 weeks. Daily trading starts at 9 a.m. and ends at 5.45 p.m. For our analysis, the time series is sampled several times, using successively increasing time intervals (1, 2, 3,..., 9, 10, 15, 20,..., 1800 seconds). For each sampling frequency, returns are computed using (3); overnight returns are deleted. Descriptive statistics of the return processes at various sampling frequencies are given in table 1.

Initially, the singularity at 0 has a large probability mass of about 0.5. It is quickly declining with decreasing sampling frequency; there are hardly any zero returns at a sampling interval of 5 min or more. Both skewness and kurtosis tend to be the smaller, the lower the sampling frequency. Note that the minimum (maximum) does not uniformly decrease (increase).

Figure 1 depicts the (univariate) distribution function of the returns at the one second sampling frequency (left) and its generalized inverse (right). The range of the empirical distribution function is $\mathrm{Ran}F_R = [0, 0.236] \cup [0.761, 1]$.

	Sampling interval (in sec)						
	1	10	30	60	300	600	1800
#obs.	2612106	261136	86990	43454	8625	4271	1435
% zeros	52.48	4.46	3.50	3.04	0.97	0.14	0.00
mean	0.0000	0.0000	0.0001	0.0002	0.0007	0.0009	0.0068
median	0.0000	0.0000	0.0000	0.0000	0.0010	0.0017	0.0100
sd	0.0032	0.0111	0.0206	0.0300	0.0669	0.0934	0.1889
skewness	1.72	0.16	−0.11	0.06	−0.19	−0.44	−0.54
kurtosis	421.85	46.35	20.78	14.68	9.53	7.92	13.41
min	−0.23	−0.26	−0.37	−0.34	−0.51	−0.67	−1.62
max	0.46	0.45	0.42	0.43	0.62	0.64	1.25

Table 1: Descriptive statistics of the returns

Hence, the supcopula is only defined in the four corners of the unit square defined by $\mathrm{Ran}F_R \times \mathrm{Ran}F_R$.

The first-order autocorrelation of the price process as a function of the sampling frequency is shown in Figure 2 (upper left panel). The correlation coefficient is positive and increasing at high frequencies, having a peak of 0.091 at 7 sec. The autocorrelation coefficient decreases rather quickly and crosses the abscissa at about 150 sec. The minimum is reached at about 7 min where the correlation coefficient is roughly -0.05. For larger intervals the autocorrelation is increasing again, at the 15 min sampling interval and beyond the first-order autocorrelation coefficient oscillates around 0. Removing all zero returns one can calculate the non-null return autocorrelation shown in Figure 2 (upper rigth panel). Apart from the one second sampling interval (where the autocorrelation coefficient is 3% when all observations are included but 4.5% when null returns are removed) the difference between

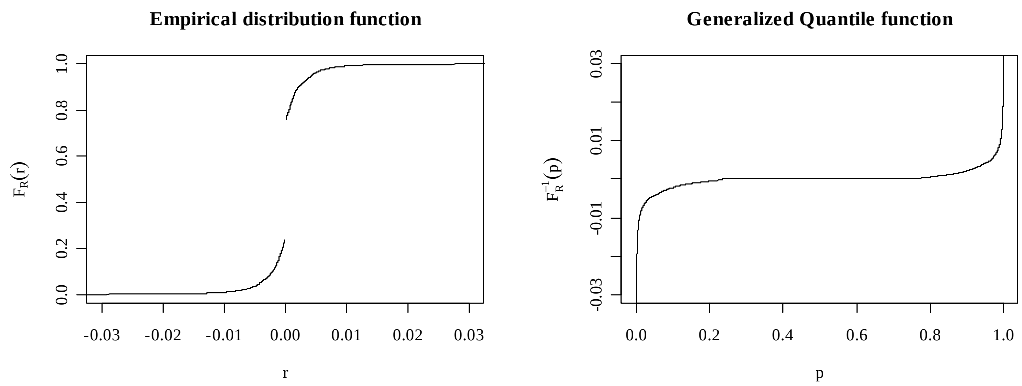

Figure 1: Empirical distribution function and quantile function of the returns at the one second sampling interval

the two autocorrelation functions is hardly discernible.

In contrast to the first-order autocorrelation of the returns, the first-order autocorrelation of their ranks does not eventually revert to the abscissa (see Figure 2, lower panels). Even for sampling intervals as large as half an hour there is a substantial negative rank autocorrelation of about -6%. It does not really matter whether the ranks are computed by (4) or (5), the only exception being the rank correlation at the one second interval which is positive for random ranks but negative when assigning the maximal rank to ties.

Although the information contained in the "global" autocorrelation coefficient and rank correlation coefficient is already revealing that there are substantial nonlinearities in the return dynamics, it does neither show what is happening locally nor where the nonlinearities are located. To gain a deeper insight we take a look at the nonparametrically estimated subcopulas and copula densities[1] for four specific sampling intervals (see figures 3 to 8). As expected, the entire dependence structure cannot be accurately captured by a single measure.

Figure 3 depicts the copula when analyzing the returns at the 1 second interval. As more than half of the observations are zero-returns, the middle of the copula appears flat, similar to the product copula. This is, however, a statistical artefact of our method of extending the subcopula by randomized bilinear interpolation. Concentrating on the subcopula density $\hat{c}_A$, Figure 4 indicates both positive and negative dependencies in the four corners. Extremely small returns tend to be followed by extremely small or (less often) extremely large returns. The reverse holds for extremely large returns. Due to the singularity, the subcopula shows a "cross" in the center, similar to the well-known compass rose that displays several "rays" radiating from the origin with the thickest streams pointing towards the "north", "east","south" and "west" of the compass. This pattern was first documented by Huang and Stoll (1994) and later reinvestigated by Crack and Ledoit (1996). As shown by Krämer and Runde (1997) and Szpiro (1998), this phenomenon is caused by the discreteness of price changes in financial markets, resulting in a finite number of possible (often clustered) returns. As long as the jumps take discrete ticks, this phenomenon also holds for portfolios and indices, due to rounding.

Adopting the allegories of the literature, Figure 5 presents the copula of 5 sec returns and seems to resemble a rose with one blossom in the middle. The petals in the four corners of the unit-square reveal again that extreme returns are associated with subsequent both positive and negative extreme

[1]The CML-procedure of the Aptech software GAUSS 5.0 was used for estimating the copula densities.

returns (where positive dependence patterns are still stronger than the negative ones). The subcopula in Figure 6 again shows a cross in the middle. Of course, the width of the cross is smaller, as the number of zero-returns at this sampling frequency has decreased. The singular component vanishes at all higher frequencies, see Figure 10 below, therefore, the subcopula equals the copula.

Increasing the observation interval to 30 seconds (Figure 7), the petals in the south-western and the north-eastern corner dominate the petals in the other corners, indicating that negative returns tend to be followed by negative ones, and positive returns by positive ones. This pattern clearly shows an overall positive dependence of consecutive index returns being in line with the conventional stylized fact of positive autocorrelation. Taking a closer look at the centre of the copula in Figure 7, one can recognize a slightly negative dependence within the "interquartile-square". While the global rank correlation is positive there is a negative rank correlation for non-extreme returns. This result shows similarities to the so-called overreaction phenomena that has been widely studied in behavioral finance and financial psychology (see, for example, Bikhchandani, Hirshleifer, and Welch (1992) and Caginalp, Porter, and Smith (2000)), but not at the high-frequency level. It is to be emphasized that these nonlinear dependence structures have a high degree of complexity and cannot be detected by a simple scalar measure, such as the linear correlation coefficient. Traders (either human or algorithmic) relying on a linear measure will misleadingly conclude that less extreme negative (positive) returns tend to be followed by negative (positive) ones. Ignoring the existence of nonlinearly correlated price movements, the trader will execute wrong order submissions and, thus, might suffer from losses cumulated over a short period of time due to the underestimation of certain price jumps in the high-frequency transaction process.

Figure 9 shows the local correlations, i.e., the correlation coefficients calculated from only one quadrant of the copula or – in case of a singularity at zero – from one part of the subcopula. The balance of the two counter-acting effects switches again when the observation interval is further increased (see Figure 8). The density in the north-western and the south-eastern corners is increasing so strongly that the overall dependence becomes negative.

Comparing all nonparametric copulas, one can see that the rose is the clearlier visible the shorter the observation interval. This result is in line with the analysis of Wang, Hudson, and Keasey (2000), who found "that the compass rose becomes more apparent as the frequency of observations increases". In contrast to phase portraits, where the pattern is sometimes not discernible due to the huge number of rays, the copula is always able to reveal the underlying dependence structure of the data. Figures 9 and 10 show the evolution of the local rank correlation coefficient and its weights (i.e. estimated probability mass) within each quadrant of the (sub-)copula as a function of the sampling frequency. There are always patterns of both positive and negative dependence, regardless of the sampling frequency. Hence, the overall direction of the rank correlation coefficient is not unique and strongly depends on the sampling frequency. The overall aggregated dependence within the unit square is either negative or positive, but there are always "local" dependence patterns in the opposite direction. These two antagonistic effects cannot be discovered by common linear regression or correlation coefficients.

4 Conclusion

This paper re-investigates the return process of high-frequent EUREX tick data; the copula-based statistical analysis shows that the conventional stylized fact of positive first-order autocorrelation of index returns has to be modified in two respects: First, allowing for nonlinearities, there are patterns of both positive and negative dependence. Second, the relative strength of the positive and negative dependence is a matter of the sampling frequency. When sampling at high frequencies, the positive relationship within the unit-square is stronger than the negative one, whereas the dependence of the less frequently observed data is negative. The copula rose shows that the global rank correlation coefficient cannot capture some important nonlinear local dependencies.

The advantage of the copula approach is its capability to separate the temporal dependence from the marginal distribution of the stationary times series, enabling more flexibility. The bivariate distribution of consecutive returns R_t and R_{t-1} is split into two components: the marginal distribution without any parametric assumptions, and the serial dependence of the return process captured by the copula. The pronounced singularity at zero return visible in the marginal distribution of high-frequency returns is carefully taken into account.

The stylized facts established in this paper cannot be explained by current theoretical models of the stock market mechanism. Bringing in line theory with the new empirical evidence is a major challenge for future research.

References

Aït-Sahlia, Y., and P. Mykland (2003): "The Effects of Random and Discrete Sampling when Estimating Continuous-Time Diffusions," *Econometrica*, 71(2), 483–549.

Andersen, T., T. Bollerslev, and F. Diebold (2005): "Some Like it Smooth, and Some Like it Rough: Untangling Continous and Jump Components in Measuring, Modeling and Forecasting Asset Return Volatility," Working paper, Duke University.

Badrinath, S., J. R. Kale, and H. Noe, Thomas (1995): "Of Shepherds, Sheep, and the Cross-Autocorrelations in Equity Returns," *Review of Financial Studies*, 8(2), 401–430.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100(5), 992–1026.

Bollerslev, T., and I. Domowitz (1993): "Trading Patterns and Prices in the Interbank Foreign Exchange Market," *Journal of Finance*, 48(4), 1421–1443.

Boudoukh, J., M. P. Richardson, and R. F. Whitelaw (1994): "A Tale of Three Schools: Insights on Autocorrelations of Short-Horizon Stock Returns," *Review of Financial Studies*, 7(3), 539–573.

Brockwell, A. E. (2007): "Universal Residuals: A Multivariate Transformation," *Statistics and Probability Letters 77*, pp. 1473–1478.

Caginalp, G., D. Porter, and V. L. Smith (2000): "Overreactions, Momentum, Liquidity, and Price Bubbles in Laboratory and Field Stock Markets," *Journal of Psychology and Financial Markets*, 1(1), 24–48.

Cai, C., R. Hudson, and K. Keasey (2003): "Trading frequency and the compass rose," *Applied Economics Letters*, 10(8), 511–517.

Charpentier, A., J.-D. Fermanian, and O. Scaillet (2007): "The Estimation of Copulas: Theory and Practice," in *Copulas: From Theory to Application in Finance*, ed. by J. Rank, chap. 2. Riskbooks, London.

Chen, S. X. (1999): "Beta Kernel Estimators for Density Functions," *Computational Statistics & Data Analysis*, 31(2), 131–145.

Chen, S. X., and T.-M. Huang (2007): "Nonparametric estimation of copula functions for dependence modelling," *Canadian Journal of Statistics*, 35(2).

Chen, X., and Y. Fan (2006a): "Estimation and Model Selection of Semiparametric Copula-Based Multivariate Dynamic Models Under Copula Misspecification," *Journal of Econometrics*, 135(1-2), 125–154.

——— (2006b): "Estimation of Copula-Based Semiparametric Time Series Models," *Journal of Econometrics*, 130(2), 307–335.

Cherubini, U., E. Luciano, and W. Vecchiato (2004): *Copula Methods in Finance*. John Wiley & Sons, New York etc.

Conrad, J., and G. Kaul (1989): "Mean Reversion in Short-Horizon Expected Returns," *Review of Financial Studies*, 2(2), 225–240.

Crack, T., and O. Ledoit (1996): "Robust Structure without Predictability: The "Compass Rose" Pattern of the Stock Market," *The Journal of Finance*, 51, 751–762.

Dacorogna, M., R. Gençay, U. Müller, R. Olsen, and O. Pictet (2001): *An Introduction to High-Frequency Finance*. Academic Press, New York.

Darsow, W., B. Nguyen, and E. Olsen (1992): "Copulas and Markov Processes," *Illinois Journal of Mathematics*, 36(4), 600–642.

Fermanian, J.-D., and O. Scaillet (2003): "Nonparametric Estimation of Copulas for Time Series," *Journal of Risk*, 5(4), 25–54.

——— (2005): "Some Statistical Pitfalls in Copula Modeling for Financial Applications," in *Capital Formation, Governance and Banking*, ed. by E. Klein, pp. 59–74. Nova Science Publ.

Gijbels, I., and J. Mielniczuk (1990): "Estimating the Density of a Copula Function," *Communications in Statistics: Theory and Methods*, 19(2), 445–464.

Goodhart, C., and L. Figliuoli (1991): "Every minutes counts in financial markets," *Journal of International Money and Finance*, 10(1), 23–52.

Härdle, W., M. Müller, S. Sperlich, and A. Werwatz (2003): *Nonparametric and Semiparametric Models*. Springer, Heidelberg.

Huang, R. D., and H. R. Stoll (1994): "Market Microstructure and Stock Return Predictions," *Review of Financial Studies*, 7, 179–213.

JOE, H. (1997): *Multivariate Models and Dependence Concepts*, Monographs on Statistics and Applied Probability, 73. Chapman & Hall, London.

KOUTMOS, G. (1997): “Feedback Trading and the Autocorrelation Pattern of Stock Returns: Further Empirical Evidence,” *Journal of International Money and Finance*, 16(4), 625–636.

KRÄMER, W., AND R. RUNDE (1997): “Chaos and the compass rose,” *Economics Letters*, 54(2), 113–118.

LEE, C., K. GLEASON, AND I. MATHUR (1999): “A Comprehensive Examination of the Compass Rose Pattern in Future Markets,” *Journal of Future Markets*, 19(5), 541–564.

LO, A., AND A. C. MACKINLAY (1990): “An Econometric Analysis of Nonsynchronous Trading,” *Journal of Econometrics*, 45(1–2), 181–211.

MARI, D. D., AND S. KOTZ (2001): *Correlation and Dependence.* Imperial College Press, London.

MARSHALL, A. W., AND I. OLKIN (1988): “Families of Multivariate Distributions,” *Journal of the American Statistical Association*, 83(403), 834–841.

MECH, T. S. (1993): “Portfolio Return Autocorrelation,” *Journal of Financial Economics*, 34, 307–344.

NELSEN, R. B. (1990): *An Introduction to Copulas*, Lectures Notes in Statistics, 139. Springer, New York etc.

OGDEN, J. P. (1997): “Empirical Analyses of Three Explanations for the Positive Autocorrelation of Short-Horizon Stock Index Returns,” *Review of Quantitative Finance and Accounting*, 9, 203–217.

PATTON, A. J. (2006): "Estimation of Multivariate Models for Time Series of Possibly Different Lengths," *Journal of Applied Econometrics*, 21(2), 147–173.

ROLL, R. (1984): "A Simple Explicit Measure of the Effective Bid-Ask-Spread in an Efficient Market," *The Journal of Finance*, 39(4), 1127–1139.

RUSSELL, J. R., AND R. F. ENGLE (2005): "A Discrete-State Continuous-Time Model of Financial Transactions Prices and Times: The Autoregressive Conditional Multinomial-Autoregressive Conditional Duration Model," *Journal of Business & Economic Statistics*, 23(2), 166–180.

SAVU, C., AND W. L. NG (2005): "The SCoD Model - Analyzing Durations with a Semiparametric Copula Approach," *International Review of Finance*, 5(1-2), 55–74.

SKLAR, A. (1959): "Fonctions de répartition à n dimensions et leurs marges," *Publications de l'Institut de Statistique de Paris*, 8, 229–231.

SZPIRO, G. G. (1998): "Tick size, the compass rose and market nanostructure," *Journal of Banking and Finance*, 22(12), 1559–1569.

TSAY, R. (2002): *Analysis of Financial Time Series*. John Wiley & Sons, New York.

WANG, E., R. HUDSON, AND K. KEASEY (2000): "Tick size and the compass rose: further insights," *Economics Letters*, 68(2), 119–125.

WÖHRMANN, P. (2005): "Predictability in Discrete Stock Prices," Working Paper 245, Institute for Empirical Research in Economics, University of Zürich.

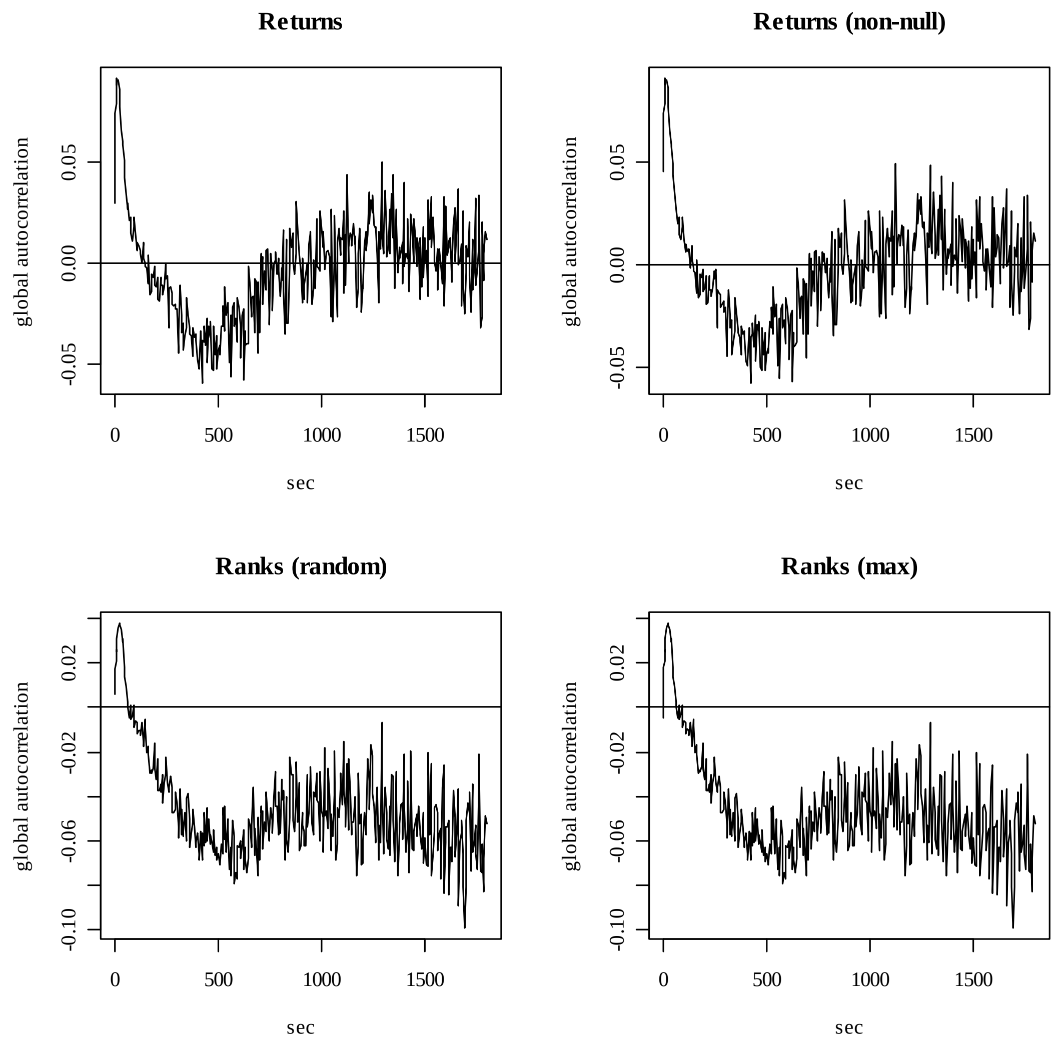

Figure 2: Correlations and rank correlations

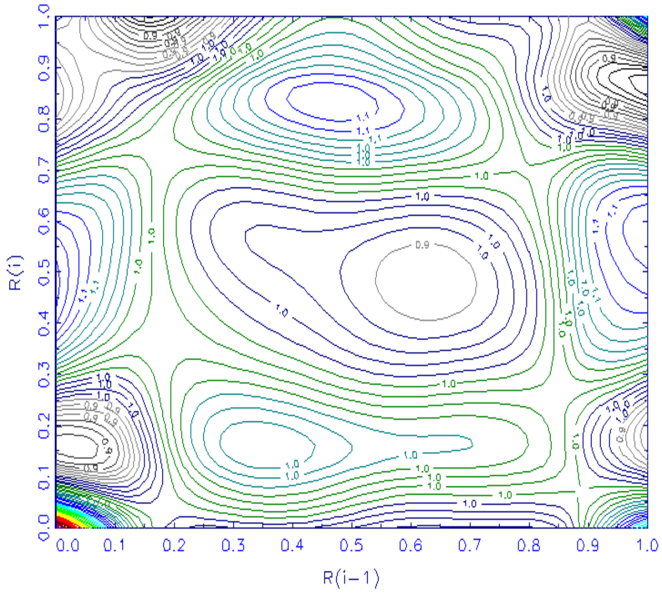

Figure 3: Contour plot of the copula at 1 sec sampling interval

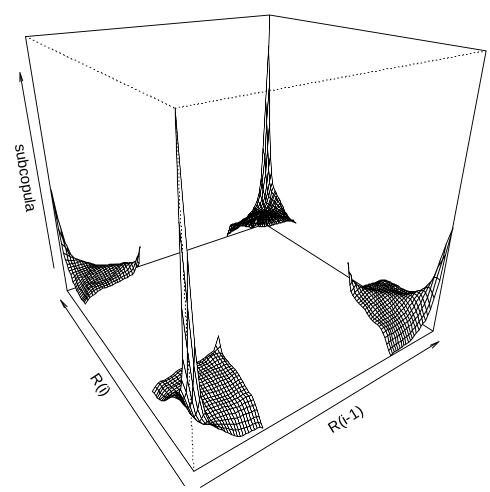

Figure 4: Density of subcopula at 1 sec sampling interval

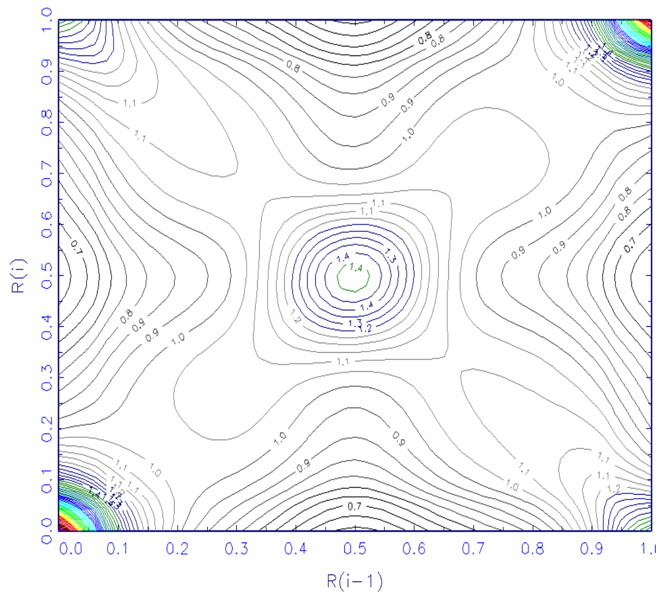

Figure 5: Contourplot of the copula at 5 sec sampling interval

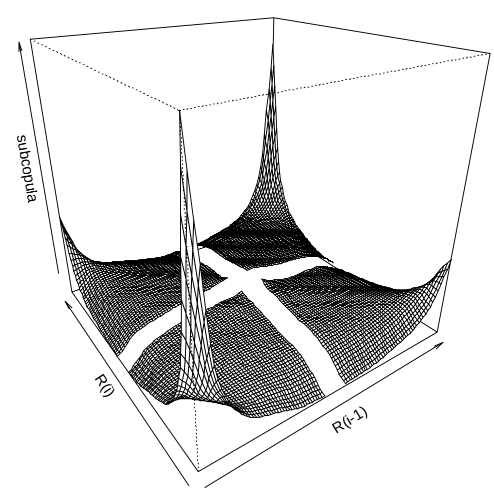

Figure 6: Density of subcopula at 5 sec sampling interval

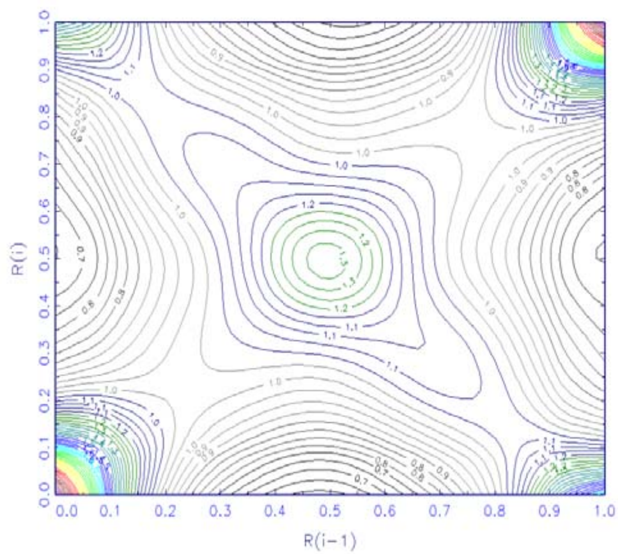

Figure 7: Contourplot of the copula at 30 sec sampling interval

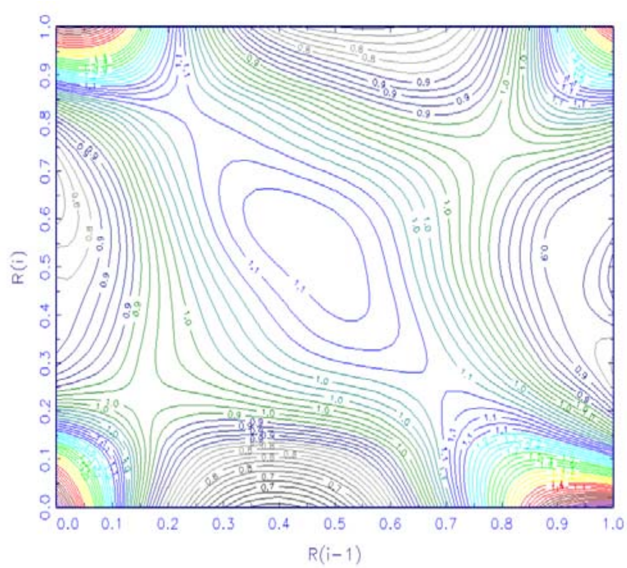

Figure 8: Contourplot of the copula at 300 sec sampling interval

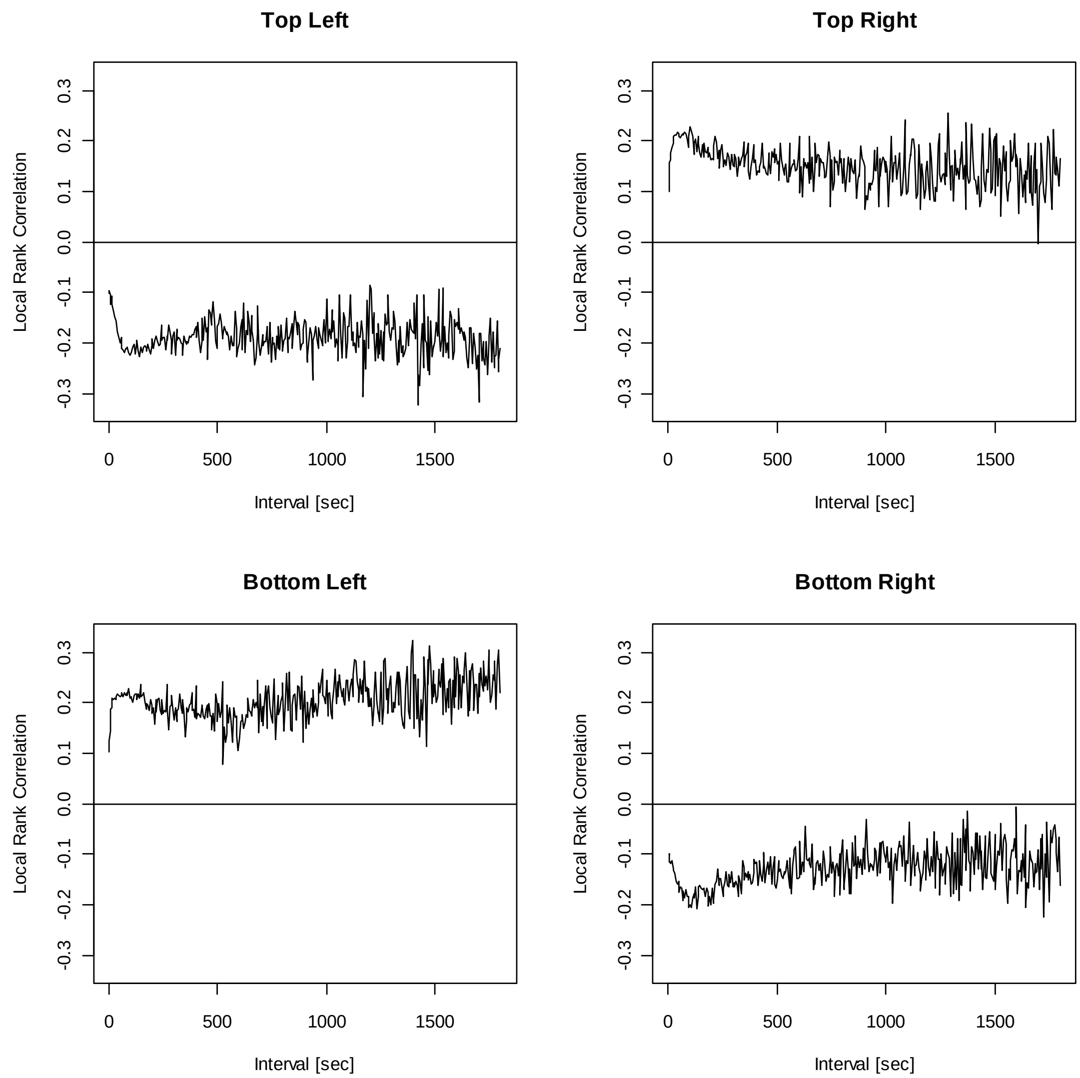

Figure 9: Local rank correlations within each quadrant of the subcopula

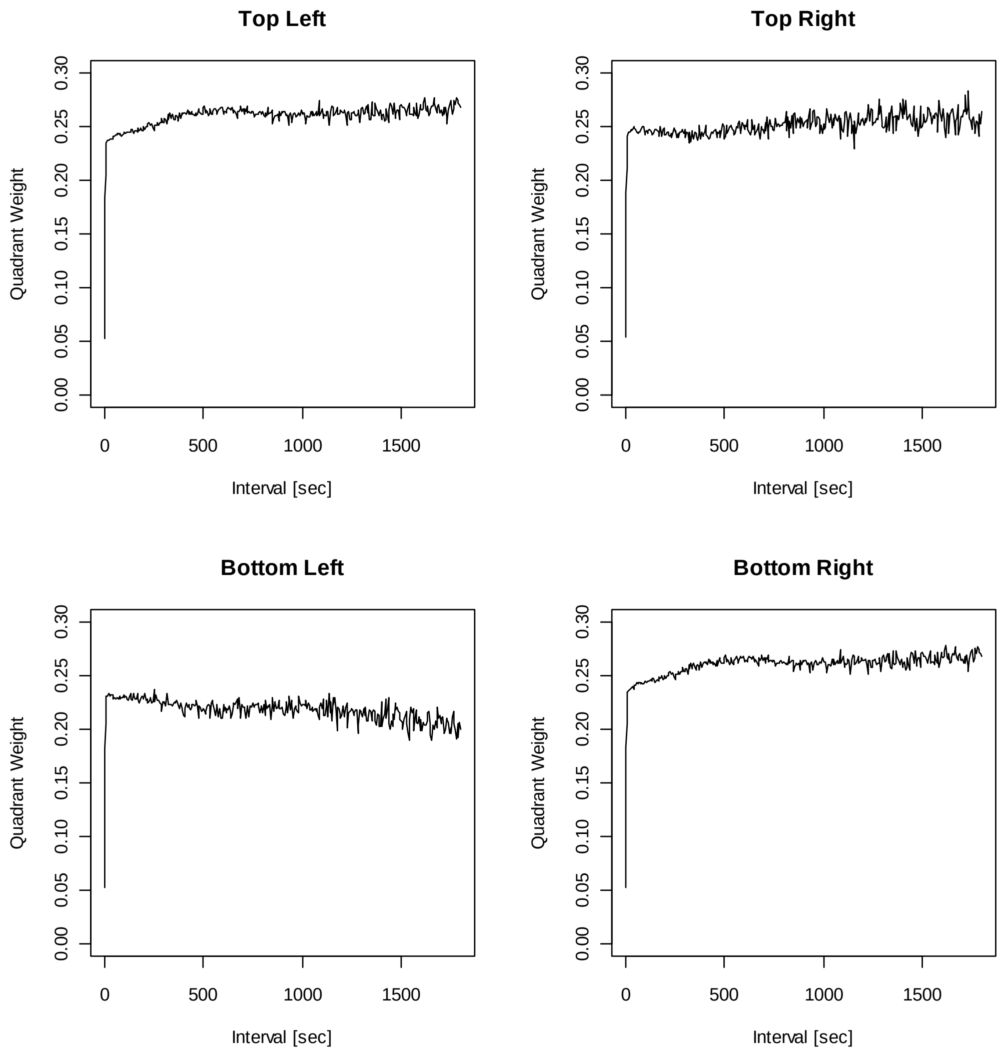

Figure 10: Weights of the quadrants of the subcopula