

# Mining local and tail dependence structures based on pointwise mutual information

Teruko Takada

Received: 27 January 2009 / Accepted: 22 March 2011 / Published online: 6 May 2011  
© The Author(s) 2011. This article is published with open access at Springerlink.com

**Abstract** The behavior of events that occur infrequently but have a large impact tends to differ from that of the central tendency, and identifying the tail dependence structure among key factors is critical for controlling risks. However, due to technical difficulties, conventional analyses of dependence have focused on the global average dependence. This article proposes a novel approach for analyzing the entire structure of nonlinear dependence between two data sets on the basis of accurate pointwise mutual information estimation. The emphasis is on fat-tailed distributions that tend to appear in events involving sudden large changes. The proposed pointwise mutual information estimator is sufficiently robust and efficient for exploring tail dependence, and its good performance was confirmed in an experimental study. The significance of the identified dependence structure was assessed using the proposed bootstrap procedure. New facts were discovered from its application to daily returns and volume on the New York stock Exchange (NYSE) Composite Index.

**Keywords** Nonparametric density estimation · Adaptive kernel · Risk control · Stock price changes · Volume · Dependence structure · Data visualization

## 1 Introduction

Many complex and dynamic systems are known to experience critical transitions, causing disastrous harm to society, such as in the form of financial bubbles, large earthquakes, or sudden climate changes. In financial markets, for example, sudden large changes in prices and volume appear periodically, and market behavior in such

---

Responsible editor: M.J. Zaki.

---

T. Takada (✉)  
Graduate School of Business, Osaka City University, Osaka, Japan  
e-mail: takada@bus.osaka-cu.ac.jp

extreme periods has been observed to be quite different from normal tendencies. For the purpose of risk control, it is crucial to understand the behavior of events that have large effects but occur only rarely. In traditional data mining, however, the usual assumption is an independence of the learning samples. Conventional statistics have focused on the analysis of central tendencies, and infrequent but large changes have been nuisance factors and are often removed from data sets. Consequently, it has been difficult to reveal complex spatio-temporal dependence structures in the areas of economics, finance, geophysics, and climate systems.

Capturing a *local* behavior is essentially a difficult task. [Mari and Kotz \(2001\)](#) stated that the number of works on local dependence remains quite low in contrast to the voluminous literature on global dependence measures. One reason for this is that the amount of information required for a sufficiently accurate estimation is larger for *local* dependence than for *global* dependence. This problem becomes even more difficult to resolve if the target phenomenon is nonlinear and occurs only infrequently: The structure to be identified is complicated, requiring a large amount of information, while the little amount of available data is short and noisy ([Barahona and Poon 1996](#)). Another difficulty is that events including sudden large changes tend to follow fat-tailed distributions. Estimation results can be significantly different depending on the selection of the estimator, as illustrated by Takada's Monte Carlo study ([2008](#)). In order to exploit the maximum amount of information from a limited amount of noisy data generated from complex dynamical systems, it is necessary to use a good dependence measure that can be estimated by a sufficiently efficient and robust method for exploring the tail dependence structure.

Mutual information (MI), a concept introduced by [Shannon \(1948\)](#), is an excellent measure for describing all types of nonlinear dependence without assuming specific functions or models. For further details on MI and some examples of its application, see [Cover and Thomas \(2006\)](#); [Pluim et al. \(2003\)](#) and [Papanas and Kugiumtzis \(2008\)](#), along with their included references. While accurate estimations are not easy even for a *global* average MI ([Kraskov et al. 2004](#)), the estimation of a *local* average MI suffers more from too few samples ([Pluim et al. 2003](#)). Therefore, the application of MI has been limited to the global average MI, that is, the expected value of pointwise MI. In a review of the literature on MI-based image processing by [Pluim et al. \(2003\)](#), few examples of local MI computations are listed. The only application areas of pointwise MI the authors can find are the computations of collocations in linguistics or text-mining related problems.

Considering the management of risks generated from complex dynamic systems, the objective of this article is to propose a statistical framework for comprehensive analysis of the dependence structure. This analysis is enabled through the use of the proposed pointwise MI estimator, whose efficiency and robustness is improved to a level sufficient for exploring a weak tail dependence structure. The major contributions of this article are as follows:

- To the best of the author's knowledge, this article reports the first visualization of a complete dependence structure expressed through pointwise MI.
- We proposed a pointwise MI estimator that provides a sufficiently efficient and robust estimation for investigating a tail dependence structure, without requiring

any prior knowledge or assumptions. The superior performance of the estimator relative to other approaches is demonstrated through an experimental study.

- The bootstrap procedure used for assessing the statistical significance of an identified dependence structure is proposed.
- New stylized facts obtained by the application of the proposed analysis to daily NYSE return and volume datasets are presented.

Pattern mining of the whole dependence structures proposed in this article has a large potential to reveal interesting nonlinear relationships. Moreover, this research can serve as the initial step toward inferring functional relationships among important variables.

The remainder of this article is organized as follows. Section 2 reviews related works. Section 3 illustrates how to interpret pointwise MI using an example of parametric bivariate densities. Section 4 describes the computational procedures used for an accurate and robust pointwise MI estimation, where a bootstrapping procedure for assessing the statistical significance of an identified local dependence structure is proposed. Section 5 compares the computational performance of the proposed methods with a major competing approach. In Sect. 6, the proposed analysis is applied toward an investigation of the dependence structure between daily NYSE index returns and volume. Section 7 provides some concluding remarks.

## 2 Related works

With special emphasis on capturing the tail dependence structure, previous works related to pointwise MI estimation are reviewed from the following three perspectives: (1) application of MI estimation, (2) MI estimation methods, and (3) other measures for local and tail dependence.

Global average MI is a nonlinear measure of dependence that is applied across multiple disciplines. Some recent examples of direct MI application are as follows: [Suzuki et al. \(2009\)](#) measured the association between conditions of yeast and biological processes through microarray datasets. [Afshin-Pour et al. \(2010\)](#) evaluated the performance of activation detection in real and experimental datasets.

In addition to a direct use, a large portion of the literature on this subject is motivated by either of the following two lines of research. The first line is the use of MI as a measure of testing independence for feature extraction and independent component analysis. For examples of this line of research, see [Battiti \(1994\)](#); [Kwaku and Choi \(2002\)](#) and [Hyvärinen and Oja \(2000\)](#).

The second line is the use of MI in the analysis of nonlinear complex dynamical systems, and consequently, causality detection. MI is considered to be a natural tool for analyzing complex systems. The driving characteristic of frequently occurring order-disorder transitions is entropy, which is deeply connected with MI. Reconstructing multi-dimensional phase portraits from a single or a few time series of observations is very important in nonlinear analysis, where an appropriate choice of time-delay is required. [Fraser and Swinney \(1986\)](#) proposed the use of a global average MI as a nonlinear version of an autocorrelation function for lag identification. [Wicks et al.](#)

(2007) show that a global average MI can be used to detect order-disorder transitions given the observations of the limited elements within a system.

MI can be used for inferring causality of spatio-temporally separated observed parameters: For example, solar wind conditions were studied by [March et al. \(2005\)](#), and information transmission within the brains of Alzheimer's patients was studied by [Jeong et al. \(2001\)](#). [Paluš et al. \(2001\)](#) used a conditional average MI for detecting the directionality of coupling from the phases of interacting oscillators. For more details on MI-based causality analysis, see the review by [Hlaváčková-Schindler et al. \(2007\)](#).

The results of comparative analyses of global average MI estimators have been mixed: While  $k$ -nearest neighbors (KNNs) and ordinary fixed-width kernel density estimators (FKDEs) have been shown to outperform classical histogram-based methods, the reported relative performance between KNN and FKDE is mixed depending on the conditions of the experiments. Results from [Khan et al. \(2007\)](#) in their estimation of global average MI based on very short data indicates that FKDE shows the best performance for highly noisy data, whereas the performance of KNN, is optimum for data with low noise levels.

Among the few measures used for capturing a local dependence structure, a function called a *copula* has gained considerable attention, particularly for analyzing multivariate dependence in financial risk management ([Cherubini et al. 2004](#); [McNeil et al. 2005](#)). A copula separates multivariate distributions into multivariate dependence structures and their marginal distributions. The main statistical advantage of a copula is that it is possible to replicate data sets through simulation using any type of marginal distribution. For further details on copulas, as well as a latest review on their use, and a description of their application in the area of finance, see [Nelsen \(2006\)](#); [Patton \(2009\)](#) and [McNeil et al. \(2005\)](#), respectively.

Recently, several extreme value copulas have been investigated: [Klüppelberg et al. \(2007\)](#) proposed the modelling of tail copulas via elliptical copulas, which is applied toward the calculation of spatio-temporal tail dependence for precipitation by [Kuhn et al. \(2007\)](#). General results on the tail dependence coefficients of Archimedean copulas have been given by [Charpentier and Segers \(2009\)](#). [Joe et al. \(2010\)](#) also studied the extreme dependence of vine copulas and developed various multivariate tail dependence functions. For an explanation and a list of references regarding extreme value copula, see a survey conducted by [Gudendorf and Segers \(2010\)](#). However, the application of copulas to certain phenomena including extreme events seems to require a degree of caution. Through example estimates of copulas used in foreign exchange rates and stock returns, [Malevergne and Sornette \(2006, Chapter 5\)](#) warned about the risks of model misspecification in the empirical determination of copulas.

Relative to previous works, this article greatly improves on the efficiency and robustness of the MI estimation. The pointwise MI estimator was first proposed by [Takada \(2001\)](#). This article emphasizes its efficiency by comparing it with other conventional method. The proposed pointwise MI estimator is based on the adaptive kernel density estimator, which combines the features of KNN and FKDE with additional devices for an improvement in accuracy. This enables an analysis of pointwise MI itself even within the tails, without requiring any prior knowledge or assumptions. To the author's best knowledge, this is the first attempt to visualize an entire dependence structure using pointwise MI for continuous variables.

### 3 Visualizing local dependence structure using pointwise

#### MI: illustrative examples

The idea of plotting pointwise MI on the colored map of a dependence structure is quite novel even for known bivariate densities. Using examples of bivariate normal and Student- $t$  densities, this section illustrates how to interpret the dependence structures visualized through pointwise MI.

#### 3.1 Definition of the MI

Let  $X$  and  $Y$  be random variables of a probability space. The MI of  $\{X, Y\}$  at  $(x, y)$  was defined by [Shannon \(1948\)](#) as

$$\varphi_{XY}(x, y) = \log \left[ \frac{f_{XY}(x, y)}{f_X(x) f_Y(y)} \right], \quad (1)$$

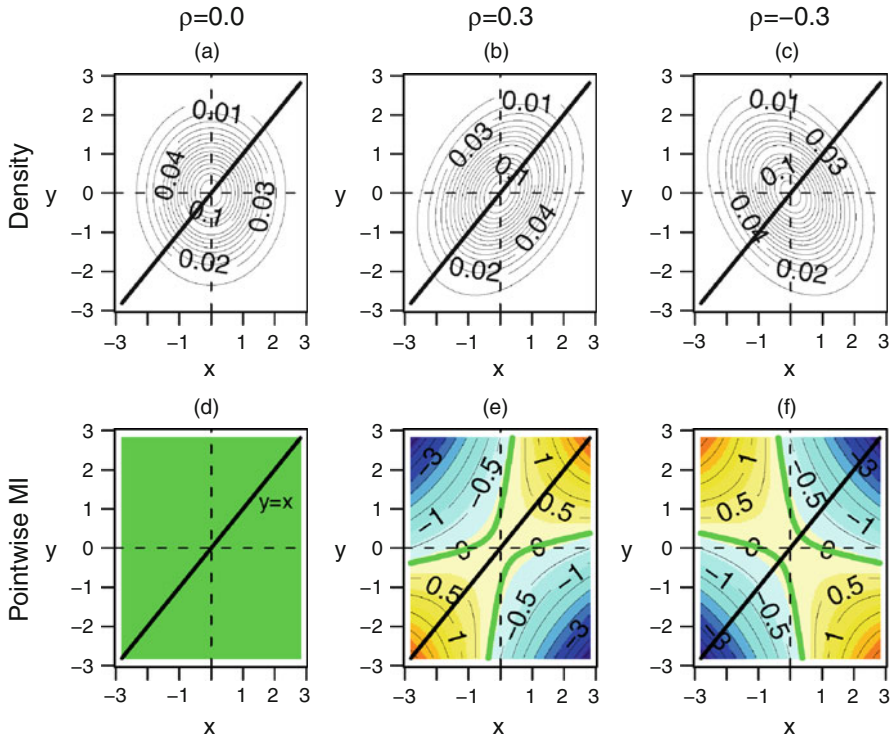
where  $f_{XY}(x, y)$  is the joint probability density of  $X$  and  $Y$  at the point  $(x, y)$ , and  $f_X(x)$  and  $f_Y(y)$  are the marginal densities of  $X$  at  $x$  and  $Y$  at  $y$ , respectively. The *global average* MI is then defined for a continuous case as

$$E [\varphi_{XY}(x, y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \log \left[ \frac{f_{XY}(x, y)}{f_X(x) f_Y(y)} \right] f_{XY}(x, y) dx dy. \quad (2)$$

The term, *mutual information* (MI) generally implies *global average* MI defined as in (2), probably because applications of MI defined as in (1) are rarely seen. To avoid confusion, we call representation (1) *pointwise* MI following the custom used in linguistics. The pointwise MI of  $X$  and  $Y$  measures the amount of information provided for event  $X = x$  by the occurrence of the event  $Y = y$ . On the other hand, global average MI is a measure of the central tendency of the local dependence structure captured by pointwise MI. See [Cover and Thomas \(2006\)](#) for further explanations on the uses of global average MI.

#### 3.2 Interpretation of pointwise MI: the case of bivariate normal density

Let  $X$  and  $Y$  follow bivariate normal distributions  $N[\boldsymbol{\mu}, \boldsymbol{\Omega}]$ , where  $\boldsymbol{\mu}$  is the mean vector of  $(X, Y)$ , and  $\boldsymbol{\Omega}$  is a covariance matrix with correlation coefficient  $\rho$ . Figure 1(a–c) show contour plots of bivariate standard normal density with different correlation levels at  $\rho = 0.3, -0.3$  and  $0.0$ . For all figures in this article, the plot range of  $x$  and  $y$  from known densities is set to include 99% quantiles of the target density. The joint density is spherical when it is independent, as in Fig. 1(a), and elliptical when there is dependence, as in Fig. 1(b) and (c). An event  $(x, y)$  located within the center occurs with high frequency, which may be regarded as a central tendency or normal-time event. In contrast, tail events occur only rarely, and are termed abnormal or non-central tendency events.



**Fig. 1** Theoretical bivariate normal density with different correlation coefficient,  $\rho$ , and the corresponding pointwise MI: The plot includes the 99% quantile

Figure 1(d–f) are the plots of the pointwise MI corresponding to the bivariate standard normal densities shown in Fig. 1(a–c). The contours of pointwise MI in the figures are colored in the following manner. The green line implies no dependence between  $X$  and  $Y$  ( $\varphi_{XY} = 0$ ). Positive dependence ( $\varphi_{XY} > 0$ ) is indicated by yellow to orange, and implies that  $Y$  tends to be large (small) when  $X$  is very large (small). Negative dependence ( $\varphi_{XY} < 0$ ) is indicated by light to dark blue, and implies that  $Y$  tends not to be large (small) when  $X$  is very large (small). Darker colors indicate larger absolute dependence. From the figures, it is clearly shown that pointwise MI is a function of the pair of values  $(x, y)$ , and provides a view of the entire dependence structure. For the non-correlation case with  $\rho = 0$  shown in Fig. 1(d), the corresponding pointwise MI becomes zero for all pairs of  $(x, y)$  in green only.

In the case of the positive correlation  $\rho = 0.3$  in Fig. 1(e), pointwise MI is close to zero around the center, while the degree of dependence is larger as it goes toward the tail of the distributions. In addition, the contour of pointwise MI is symmetric with regard to the line  $y = x$ , with positive dependence in the first and third quadrants, and negative dependence in the second and fourth quadrants. Observing the contour plot of the bivariate density estimates in Fig. 1(b), it is difficult to capture such complicated local relations as those illustrated in Fig. 1(e). In the case of the negative correlation,  $\rho = -0.3$  in Fig. 1(c), the region of positive dependence is observed in the second

and third quadrants and negative dependence in the first and fourth quadrants. Again, the plot is symmetric with regard to the line  $y = x$ . The pointwise MI is not rotation free. In rotating the  $x$  and  $y$  axes by 45 degrees, the coordinates are transformed into  $(x', y') = ((x - y)/\sqrt{2}, (x + y)/\sqrt{2})$ , and the generated pointwise MI no longer preserves the initial dependence.

### 3.3 Pointwise MI for assessing independence: the case of bivariate $t(3)$ density

First consider mutually independent bivariate  $t$  density. Let  $X$  and  $Y$  be mutually independent, and follow univariate Student- $t$  distributions with  $\nu$  degrees of freedom. The joint probability density function (pdf) of mutually independent bivariate  $t$  distributions with  $\nu$  degrees of freedom is defined as:

$$f_{IT}(x, y, \nu) = f_X(x, \nu) f_Y(y, \nu), \quad (3)$$

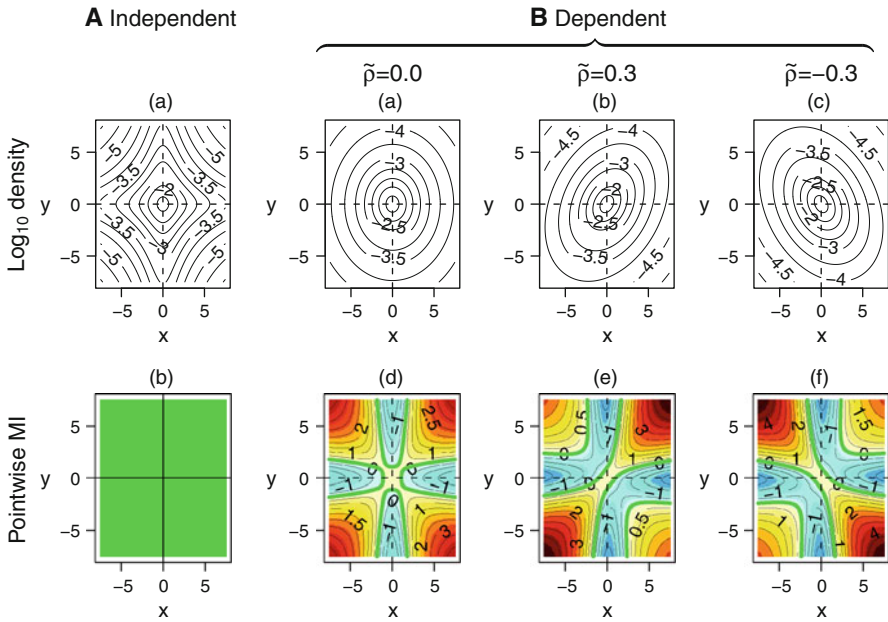
where  $f_X(x, \nu)$  and  $f_Y(y, \nu)$  are the pdf of the univariate Student- $t(\nu)$  density. The pointwise MI corresponding to the density defined in (3) is zero for all  $x$  and  $y$ .

Next, consider a general bivariate  $t$  density where  $X$  and  $Y$  are mutually dependent. The joint pdf of a dependent bivariate central  $t$  distribution with  $\nu$  degrees of freedom is defined as

$$f_T(x, y, \nu, \tilde{\rho}) = \frac{\Gamma((\nu + 2)/2)}{\pi \nu \sqrt{1 - \tilde{\rho}^2} \Gamma(\nu/2)} \left[ 1 + \frac{x^2 + y^2 - 2\tilde{\rho}xy}{\nu(1 - \tilde{\rho}^2)} \right]^{-(\nu+2)/2}, \quad (4)$$

where  $\tilde{\rho}$  is the correlation coefficient between  $X$  and  $Y$ . This is the bivariate version of the most common form of multivariate  $t$  distributions: it directly generalizes the univariate Student- $t$  distribution in the same manner in which the multivariate normal distribution generalizes the univariate normal distribution. See [Kotz and Nadarajah \(2004\)](#) for further explanation of multivariate  $t$  distributions. Note that the pointwise MI does not reduce to zero when  $\tilde{\rho} = 0$ . Thus,  $X$  and  $Y$  following the joint density defined as (4) with  $\tilde{\rho} = 0$  are not mutually independent.

Figure 2 illustrates the contour plots of joint  $t(3)$  density and the corresponding pointwise MI. The contours of the densities are presented in  $\log_{10}$  scale to emphasize the tail behavior. Figure 2A refers to the independent case defined in (3), and Fig. 2B refers to the correlated case defined in (4). The independent joint  $t(3)$  density in Fig. 2A(a) is plotted as a rhomboid contour, while the dependent joint  $t(3)$  density in Fig. 2B(a) is plotted as a spherical contour. Differing from the case of the correlated standard normal density, a weak positive dependence is observed at the tails. Given these density contours, identifying an independent case is difficult. On the other hand, the pointwise MI plot provides a very clear information regarding the dependence. The pointwise MI for the independent joint  $t(3)$  density in Fig. 2A(b) is shown in green only, indicating the local independence for all  $x$  and  $y$  values. The pointwise MI for dependent joint  $t(3)$  density with  $\tilde{\rho} = 0$  in Fig. 2B(d) exhibits a significant



**Fig. 2** Theoretical bivariate  $t(3)$  densities and the corresponding pointwise MI: (Panels **A**) Cases of independent  $t(3)$  density defined in (3); (Panels **B**) Cases of dependent densities defined in (4) where  $\tilde{\rho}$  denotes the degree of dependence. The plot includes the 99% quantile

positive dependence along the positive and negative 45-degree lines, and the significance becomes stronger as it goes toward the tails. In this way, the pointwise MI comprehensively uncovers a hidden dependence structure, which is impossible for the case of global average statistics.

#### 4 Pointwise mutual information estimation method

For the purpose of risk control, it is important to analyze target events that include sudden large periodic changes. Such events tend to follow fat-tailed distributions. Moreover, the data tend to include noise and outliers. Estimation of such distributions is difficult using conventional approaches. An accurate tail estimation becomes even more difficult. This section proposes a robust estimation method that enables capturing the tail dependence structure of target events with a fat-tailed distribution.

##### 4.1 Bivariate adaptive kernel density estimation

The estimation of the pointwise MI proposed herein is based on the bivariate adaptive kernel density estimation method of [Breiman et al. \(1977\)](#), and refined by [Abramson \(1982\)](#), which is known to be good at estimating fat-tailed and multi-modal densities. Different from conventional kernel methods with a fixed bandwidth, the bandwidth of the adaptive kernel method varies locally. Because of this locality in fitting, this method has the best performance among the major approaches in estimating densities



with fat tails and multi-modes. See [Takada \(2008\)](#) and [Hwang et al. \(1994\)](#) for a comparative analysis of competing estimators in univariate cases and multivariate cases, respectively. Another appealing point is that a good performance can be achieved using a relatively simple algorithm.

Let  $d$ -dimensional data set with sample size  $n$  be  $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ ,  $i = 1, \dots, n$ , and a  $d$ -dimensional point be  $\mathbf{x} = (x_1, \dots, x_d)$ . The conventional kernel density estimator with a Gaussian kernel has the following form:

$$\hat{f}_{FK}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^n \phi\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right), \quad (5)$$

where  $\phi$  denotes the standard Gaussian density, and fixed bandwidth  $h$  controls the degree of smoothing. An optimal choice of  $h$  for the Gaussian kernel suggested by [Silverman \(1986\)](#) is given as  $h = \{4/(d+2)\}^{1/(d+4)} n^{-1/(d+4)}$ .

The adaptive kernel density estimator of [Breiman et al. \(1977\)](#), and refined by [Abramson \(1982\)](#), is defined as

$$\hat{f}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{(h\lambda_i)^d} \phi\left(\frac{\mathbf{x} - \mathbf{X}_i}{h\lambda_i}\right), \quad (6)$$

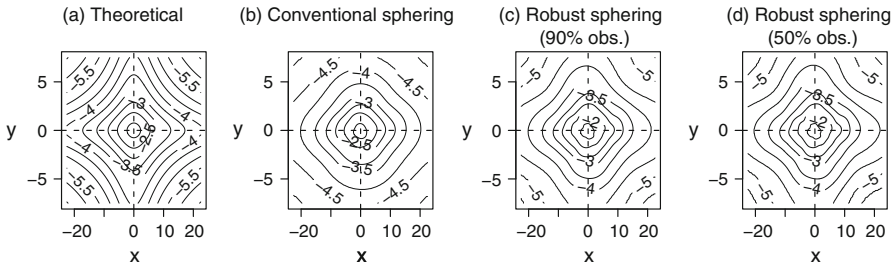
where  $\lambda_i$  is a local bandwidth factor, which narrows the bandwidth  $h\lambda_i$  near the modes and widens it at the tails. The determination of  $\lambda_i$  is based on the pilot estimate  $\tilde{f}(x)$  from the fixed kernel method, which is defined in (5) as  $\lambda_i = \{\tilde{f}(X_i)/g\}^{-1/2}$ , where  $\log g = n^{-1} \sum \log \tilde{f}(X_i)$ .

## 4.2 Robust sphering

In order to avoid extreme differences of spread in the various coordinate directions, pre-scaling or data sphering is generally desirable in multivariate settings. This has the effect of skewing the shape of kernels to better fit to the data. To assure the accuracy of a sufficient tail estimation for exploring tail behavior, we employ the robust sphering of [Rousseeuw and van Zomeren \(1990\)](#).

The conventional approach of [Fukunaga \(1972\)](#) first spheres  $d$ -dimensional observed data  $\mathbf{M}$  into  $\mathbf{Z}$  using  $\mathbf{Z}_i = \boldsymbol{\Omega}^{-1/2}(\mathbf{X}_i - \bar{\mathbf{X}})$ , where  $\bar{\mathbf{X}}$  and  $\boldsymbol{\Omega}$  are the mean vector and covariance matrix of  $\mathbf{X}$ , respectively. After estimating the density of sphered data,  $\mathbf{Z}$ , the estimates are transformed back as  $\hat{f}(\mathbf{m}) = (\det \boldsymbol{\Omega})^{-1/2} \hat{f}(\mathbf{z})$ . In the robust sphering of [Rousseeuw and van Zomeren \(1990\)](#), non-robust  $\bar{\mathbf{X}}$  and  $\boldsymbol{\Omega}$  in Fukunaga sphering are replaced with robust alternatives: The center of the minimum volume ellipsoid covers  $q\%$  of the observations, and the covariance matrix is determined based on the same ellipsoid multiplied by a correction factor to obtain consistency in multivariate normal distributions. The authors recommend  $q = 50\%$  from a robustness perspective.

If the target data are known to follow fat-tailed distributions such as the case of financial asset returns, robust sphering is advantageous to the [Fukunaga \(1972\)](#) method,



**Fig. 3** Effects of robust sphering in estimating scaled independent bivariate  $t(3)$  densities: Scale of  $x$  is set as three-times of the scale of  $y$

which we demonstrate by using the example of bivariate  $t(3)$  density estimation. The choice of density comes from the fact that the value of the algebraic tail index  $\alpha$  for most financial data has been reported to be about three to five. For example of this, see Huisman et al. (2001) and Lin and Kao (2008). Figure 3 compares the effect of employing a different sphering method in estimating the independent bivariate  $t(3)$  density defined in (3). The range of  $x$  is modified to three-times that of  $y$  in order to observe the sphering effects. The corresponding theoretical density contours at the  $\log_{10}$  scale are plotted in Fig. 3(a).

Figure 3(b) shows the joint density estimates using Fukunaga (1972) sphering with the sample mean and covariance matrix. Figure 3(c) and (d) are the results of robust sphering developed by Rousseeuw and van Zomeren (1990) using a minimum volume ellipsoid covering  $m = 90$  and 50% of the observations, respectively. The estimates based on robust sphering are significantly better than those based on Fukunaga sphering, which are much too smoothed. In addition, the robust sphering based on a 50% observation is slightly better than that based on a 90% observation, which is in agreement with the claim of Rousseeuw and van Zomeren (1990) that  $q = 50\%$  is a desirable choice from a robustness perspective. To investigate weak local (tail) dependence of a fat-tailed data series, which tends to have a more serious effect from the outliers than the other cases, it is desirable to employ a robust sphering procedure.

### 4.3 Problems of artifacts in the tail estimates

A sufficiently accurate estimation for exploiting the tail relationship requires special attention in estimating the denominator of the pointwise MI,  $f_x(x)f_y(y)$ . Herein, it is shown that the naive estimators for  $f_x(x)f_y(y)$  as a product of univariate density estimates are problematic, and therefore new estimators are proposed. Table 1 is a list of different pointwise MI estimators, the performance of which are compared in this and the following sections.

In the acronyms listed in Table 1, the FKUD pointwise MI estimator in Table 1 is that used by Moon et al. (1995) and Khan et al. (2007), where FK denotes the conventional fixed bandwidth density estimator defined in (5). UD denotes a naive estimator for  $f_x(x)f_y(y)$ , which is a product of univariate density estimates. AK in Table 1 denotes the adaptive kernel density estimator defined in (6).

**Table 1** Acronyms used for different pointwise MI estimators

Acronyms	Pointwise MI estimator
FKUD	$\log \left[ \frac{f_{FK}(x,y)}{f_{FK}(x)f_{FK}(y)} \right]$
AKUD	$\log \left[ \frac{f_{AK}(x,y)}{f_{AK}(x)f_{AK}(y)} \right]$
AKMD	$\log \left[ \frac{f_{AK}(x,y)}{\int_x f_{AK}(x,y)dx \int_y f_{AK}(x,y)dy} \right]$
AKSD	$\log \left[ \frac{f_{AK}(x,y)}{f_{AK}(x,y_{shuffled})} \right]$

$f_{FK}$ : conventional fixed bandwidth kernel density estimator

$f_{AK}$ : adaptive kernel density estimator

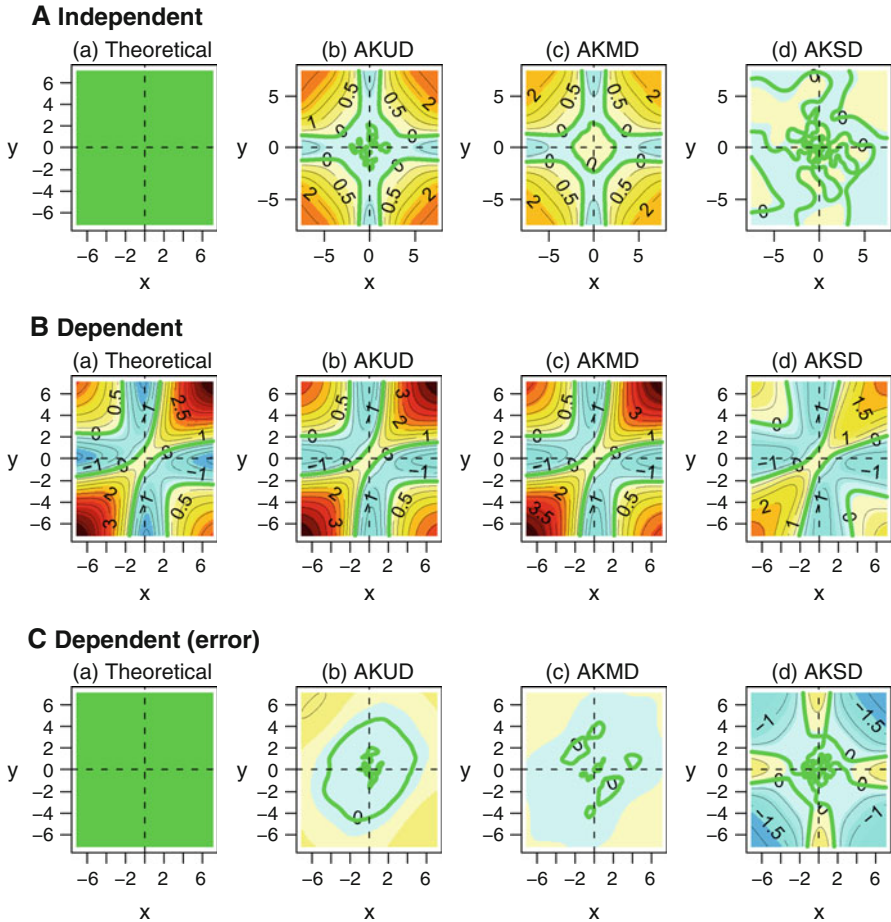
$(x, y_{shuffled})$ : Bivariate  $(x,y)$  sequence where the order of  $y$  is shuffled to eliminate dependence between  $x$  and  $y$

Figure 4A shows pointwise MI estimates of the independent bivariate  $t(3)$  density defined in (3) with a sample size of 10000, where the estimates for  $f_x(x)f_y(y)$  are obtained in different ways. Figure 4A(b) shows pointwise MI estimates from the AKUD estimator, which is a natural estimator based on the adaptive kernel density estimator defined in Table 1. In Fig. 4A(b), artificial dependence is observed in the tails. This is because the precisions of the bivariate density estimation and univariate density estimation are different, and their difference becomes more significant as they move toward the tails where the amount of available data is limited. Taking the logarithm of the ratio in the computation of pointwise MI enlarges the difference, particularly when the dependence is weak. The noise in the tails is more significant for the  $t(3)$  density than for the Gaussian density, probably because fitting fat tails is generally more difficult than fitting Gaussian tails.

A tail artifact occurs due to data scarcity at the tails, which is inevitable. In order to eliminate noises coming from the procedural differences, we devised an AKMD estimator whose denominator is the product of marginal densities of the joint density,  $f_{x|y}(x, y)$ . MD implies a marginal density estimation. The AKMD estimates shown in Fig. 4A(c), show a reduced artificial dependence, but the level is still harmful.

#### 4.4 AKSD estimation

For a practical tail analysis, the pointwise MI estimator is required to identify the existence of dependence for all data types, including faint or no dependence cases. We propose computing  $f_x(x)f_y(y)$  using the bivariate density estimation of shuffled data series  $\{X, Y_{shuffled}\}$ , where the dependence between  $X$  and  $Y$  is completely eliminated by randomizing the order. We call the proposed estimator as AKSD, where SD denotes bivariate density estimation in the shuffled sequences. In the AKSD estimation, to eliminate the noises coming from procedural differences, we can apply an identical bivariate density estimation procedure to both the nominator and denominator of pointwise MI. In a way, the pointwise MI measures the degree of independence between the outcome  $X$  and outcome  $Y$ , and the denominator,  $f_x(x)f_y(y)$ , may be



**Fig. 4** The effects of employing different procedures to obtain an estimate of  $f(x)f(y)$ : (Panels **A**) Pointwise MI estimates of independent bivariate  $t(3)$  density; (Panels **B**) Pointwise MI estimates of dependent bivariate  $t(3)$  density; (Panels **C**) Pointwise MI estimation errors of dependent bivariate  $t(3)$  density; Sample size  $n = 10000$ . See Table 1 for an explanation of the acronyms used

regarded as a test statistics for pointwise independence. Thus, a comparison of the densities of the given  $\{X, Y\}$  and shuffled no-dependence  $\{X, Y_{\text{shuffled}}\}$  is reasonable for the purpose of identifying the existence of pointwise dependence.

Let  $Y = \{Y_1, \dots, Y_{i-1}, Y_i, Y_{i+1}, \dots, Y_{j-1}, Y_j, Y_{j+1} \dots, Y_n\}$ . The algorithm for shuffling the order of  $Y$  is as follows:

1. Choose two components of  $Y$ ,  $Y_i$  and  $Y_j$ , where  $i \neq j$ , based on the uniform deviates, and swap them to obtain  $\tilde{Y} = \{Y_1, \dots, Y_{i-1}, Y_j, Y_{i+1}, \dots, Y_{j-1}, Y_i, Y_{j+1}, \dots, Y_n\}$ .
2. Repeat the first step  $10n$  times to obtain two independent series,  $\{X, \tilde{Y}_{\text{shuffled}}\}$ .

Figure 4A(d) shows the estimation results of AKSD, which illustrates the characteristics of white noise as theoretically expected. No artifacts can be observed at the tails. Therefore, the proposed AKSD estimator seems better than AKUD and AKMD for the purpose of exploring subtle pointwise dependence.

Applying an identical bivariate density estimation procedure to compute  $f_{XY}(x, y)$  and  $f_X(x)f_Y(y)$  has the effect of unifying the accuracy or resolution to the level of bivariate density estimation. Using AKUD and AKMD estimation, we can obtain a more accurate estimate of  $f_X(x)f_Y(y)$  than through AKSD estimation. However, more information does not always bring about better estimation results. From the view point of “scale space theory” in the literature on computer vision, applying an identical procedure to both estimates and calculating the ratio, plays the same role as that played by scale resolution in a visual system. A similar example is the minimum distance estimations of Basu and Lindsay (1994) and Takada (2009): Minimizing the distance between a nonparametrically smoothed density and a *smoothed* analytical model density has been proven to provide higher efficiency than cases in which analytical model density itself is used. For further discussion of this view, see Chaudhuri and Marron (2000) and Lindeberg (1994).

However, some limitations are observed in AKSD estimation when a certain level of dependence exists. Figure 4B shows the pointwise MI estimates of dependent bivariate  $t(3)$  density defined in (4) with  $\tilde{\rho} = 0.3$  and a sample size of 10000, and Fig. 4C shows the corresponding error plot. While AKSD estimates correctly capture the dependence structure, the image is fainter than the theoretical level. From Fig. 4C, we can observe that AKMD estimates have the smallest amount of error in this case. While some negative patterns exist in the AKSD estimation of the error plot in Fig. 4C(d), the level of error is slightly lower than in the tail artifacts observed in Fig. 4A(c).

#### 4.5 Assessment of statistical significance

In pointwise MI estimation, errors in a data-scarce region are unavoidable. In practice, therefore, it is desirable to be able to distinguish whether the estimated dependence structure is significant or mere noise when the densities of the target data are unknown. This subsection proposes an assessment procedure for the significance of pointwise MI dependence: The statistical significance of the estimated dependent structure is compared with the 95% quantile range of pointwise MI estimates of replicated non-dependent data sets, which are generated from the original data sets to eliminate dependence.

Construction of the quantile range, or confidence band, is conducted through bootstrapping. See Efron and Tibshirani (1993) for details of the bootstrapping method. The computational procedure is as follows:

1. Given observed data set  $\{X, Y\}$ , remove the dependence by following the shuffling algorithm described in Sect. 4.4, to obtain  $N$  sets of non-dependent pairs of  $\{X, Y_{\text{shuffled}}\}$ .
2. Compute pointwise MI of the  $N$  sets of  $\{X, Y_{\text{shuffled}}\}$  in an identical manner used for estimating the given data set  $\{X, Y\}$ .

3. Compute 2.5% and 97.5% of the sample quantile points of the pointwise MI estimates of  $N$  sets of  $\{X, Y_{\text{shuffled}}\}$  to obtain a 95% confidence band for non-dependence.

The confidence band with non-dependent shuffled data sets describes the noise level of the estimation procedure: If the estimated pointwise MI is within the 95% confidence band, it implies that the detected dependence is not statistically significant. Thus, it helps in discriminating noises or artifacts with significant dependence. This assessment can be applied to any type of estimator. An example use of the confidence band is demonstrated in Sect. 6.

## 5 Experimental analysis

The performance of pointwise MI estimation using the AKSD and AKMD estimators is thoroughly examined by changing the sample size and tail fatness. Their performance is compared in the same setting with those of the widely used FKDE-based pointwise MI estimator first used in Moon et al. (1995), whose superior performance in short and noisy data was demonstrated in the comparative studies by Khan et al. (2007). The inclusion of other approaches is avoided, since our purpose is not a comparison of the performance of several density estimators.

### 5.1 Implementations and data

The definitions of the estimators are given in Sects. 4.3 and 4.4, and are summarized in Table 1. Considering the applications used to control the risks arising from infrequent but sudden large changes, this study focuses on fat-tailed distributions. We choose bivariate Student- $t$  distributions, whose degree of freedom coincides with the algebraic tail parameter,  $\alpha$ . In addition to mutually dependent cases, mutually independent cases are compared for analyzing the problem of artifacts occurring within the tails.

Mutually independent bivariate  $t(\nu)$  deviates are generated as  $\{X, Y\}$ , where  $X$  and  $Y$  are separately generated as univariate Student- $t(\nu)$  deviates. Mutually dependent bivariate  $t(\nu)$  deviates,  $\{X, Y\}$ , following the pdf defined in (4) are generated as follows:  $\{X, Y\} = \{Z_x(S/\sqrt{\nu})^{-1}, Z_y(S/\sqrt{\nu})^{-1}\}$ , where  $Z_x$  and  $Z_y$  jointly follow a standardized bivariate normal distribution with variance-covariance matrix  $\Sigma$ , and being independent of  $Z_x$  and  $Z_y$ ,  $S$  is distributed as  $\chi_\nu$ . This is the same procedure as that employed in Mathematica. Since the FKUD estimator used by Moon et al. (1995) and Khan et al. (2007) does not employ a sphering procedure, the scales of  $X$  and  $Y$  of the simulated data are kept equal to avoid the possibility that the sphering effect dominates the differences in the performance.

### 5.2 Results: size effect

The effect of changing sample sizes  $n$  for estimation of the pointwise MI performance is compared with respect to the FKUD, AKMD, and AKSD estimators. Random samples with sizes  $n = 100, 1000, 10000$ , and  $100000$  are simulated from independent

bivariate  $t$  density defined in (3) and dependent bivariate  $t$  density defined in (4) with  $\tilde{\rho} = 0.3$ . As in the example cases of financial asset returns, the degree of freedom is set at  $\nu = 3$ , which is a widely observed level of tail fatness.

Using the independent bivariate  $t(3)$  data sets, we study the effects of changing the sample sizes  $n$  to those in the tail artifact problems. The pointwise MI estimation performance of three methods are compared. Figure 5 shows the results. Figure 5A plots the theoretical value, which indicates the zero pointwise MI for all  $(x, y)$ . The FKUD estimator shows a positive dependence in the region where the data are observed. Judging from the FKDE joint density estimates, the black area corresponds to a data scarce area. We prepared 23 color levels, and the pointwise MI contour levels are plotted in 0.5 steps. The region in black is considered to be either positive or negative estimates overflowed from the preset color range. The AKMD estimator also shows artificial positive dependence in all tails, and the error level increases as the sample size increases. On the other hand, the AKSD estimator exhibits a white noise image for all sample sizes. For identifying subtle dependence, the AKSD estimator is therefore considered to be the best method.

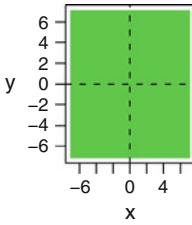
The effect of sample size on the change in performance in estimating the pointwise MI of dependent bivariate  $t(3)$  deviates is illustrated in Fig. 6. The corresponding errors from the true value are plotted in Fig. 7. The performance of the FKUD estimator is significantly poor. In Fig. 7(a–d), positive artificial errors are observed. The performance improves as the sample size increases, but a sample size of even 100000 seems insufficient for the FKUD estimator to provide an accurate pointwise MI estimate. Relative to the FKUD estimator, the AKMD estimator is very close to the theoretical image. Its error plot shows the smallest amount of errors among the three estimators, and the amount of errors is reduced as the sample size increases.

Relative to the AKMD estimator, the AKSD estimator seems advantageous for small sample sizes. In the case of larger sample sizes, however, the estimated image becomes fainter than the theoretical image. A good aspect is that the estimated pattern of the dependence structure is correct. The error patterns shown in Fig. 7 are similar to an independent bivariate  $t(3)$  density contour, suggesting that major errors come from estimating  $f_x(x)f_y(y)$ . Since the error pattern indicates mostly negative pointwise MI, the risk of misunderstanding is lower than in the opposite case: While a negative pointwise MI indicates that the events  $(x, y)$  are not likely to occur simultaneously, our major interest is in events that are likely to simultaneously occur.

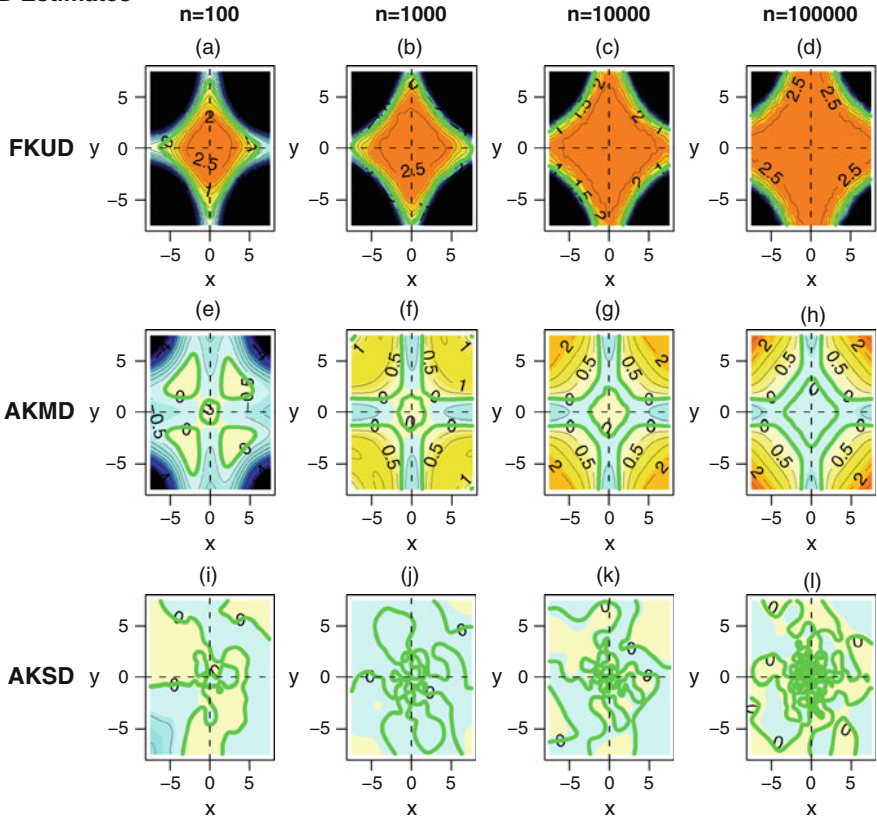
### 5.3 Results: effect from tail fatness

For the purpose of examining the tail effect on the pointwise MI estimation performance, the selected densities of the experimental data sets are dependent bivariate  $t$  densities with degrees of freedom 2, 5, 10 and 30. The sample size is set as  $n = 10000$ . The pointwise MI estimation results are illustrated in Fig. 8, and corresponding errors from the true value are plotted in Fig. 9. The performance of the FKUD estimator improves as the tail fatness decreases, but significant positive errors are still observed even for the case of  $t(30)$ . On the other hand, the AKMD estimator provides the closest image to the theoretical one. However, the amount of positive errors in the data-scarce

**A Theoretical**



**B Estimates**

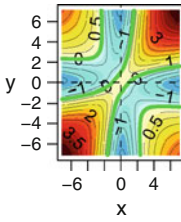


**Fig. 5** Size effect on the pointwise MI estimation of independent bivariate  $t(3)$  density: (Panels **A**) Theoretical pointwise MI; (Panels **B**) Pointwise MI estimates. See Table 1 for an explanation of the acronyms used

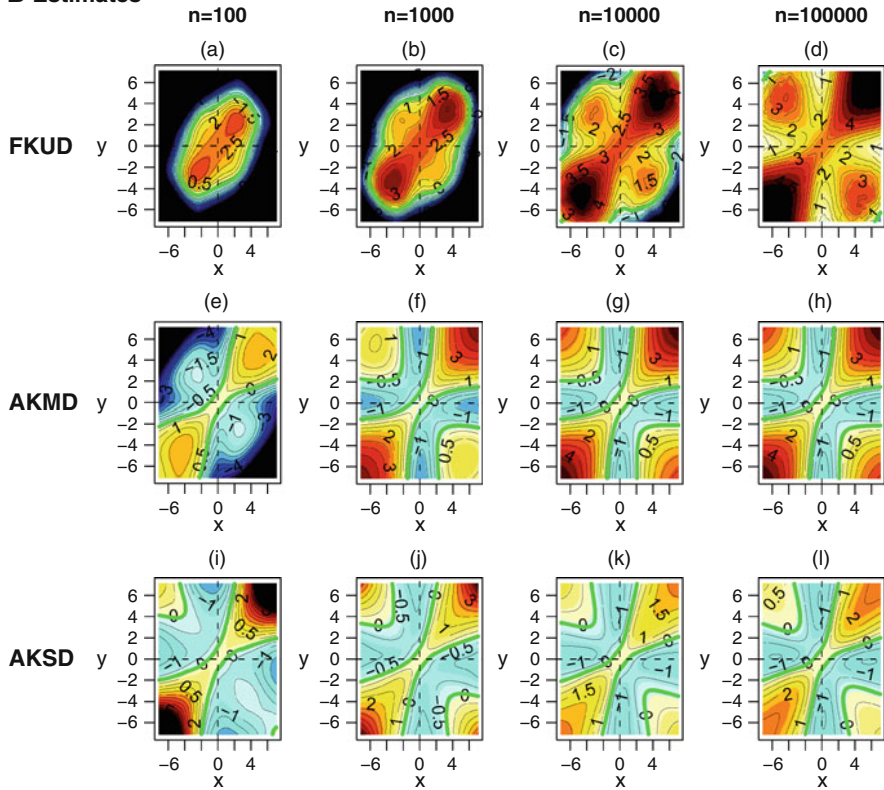
area slightly increase as the tail fatness decreases. Small positive errors for the AKSD estimates are found in Fig. 9(i) and (l) with  $\nu = 2$  and 30. The errors in Figs. 4C(d) and 9(j), (k) are mostly slightly negative. A tail fatness range of  $\nu$  from 3 to 10 is frequently observed in fat-tailed phenomena. Therefore, in practice, the possibility of misunderstanding the dependence structure from the AKSD estimates seems relatively low.



**A Theoretical**

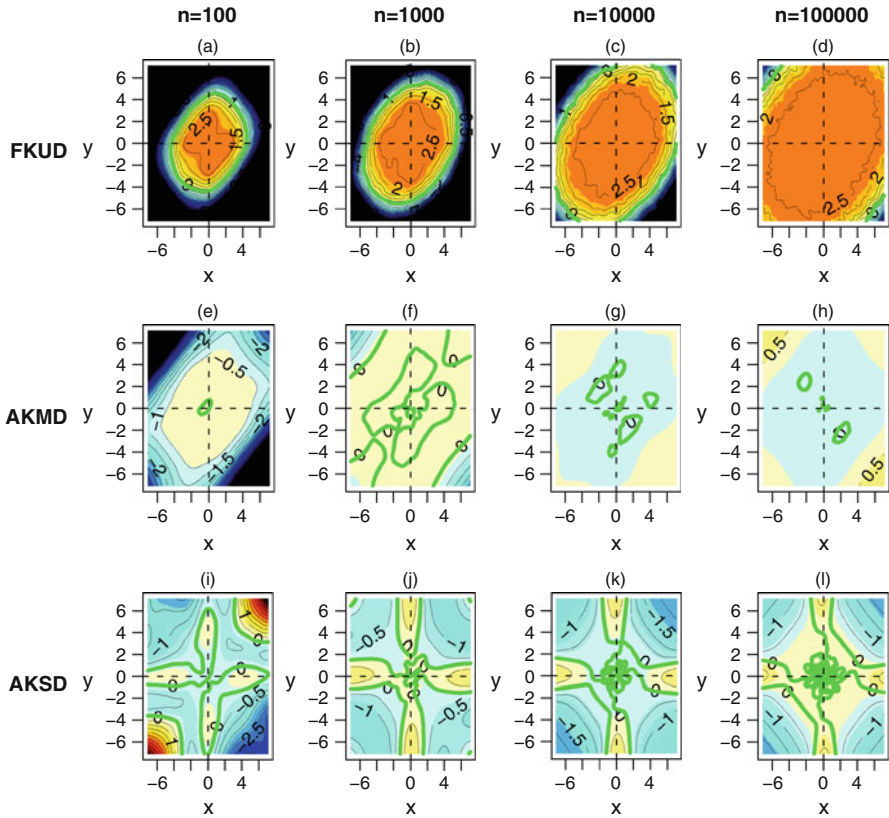


**B Estimates**



**Fig. 6** Size effect on the pointwise MI estimation of dependent bivariate  $t(3)$  density: (Panels A) Theoretical pointwise MI; (Panels B) Pointwise MI estimates. See Table 1 for an explanation of the acronyms used

The results from the experimental study are summarized as follows: The accuracy of the FKSD estimator for pointwise MI estimation is not sufficient for practical use. Whereas the AKMD estimator exhibits excellent performance in estimating dependent bivariate  $t$  densities, it has a problem of artifacts at the tails in estimating independent bivariate densities. The AKSD estimator does not have such a tail artifact problem, and can correctly capture the dependence structure.



**Fig. 7** Size effect on the pointwise MI estimation error of dependent bivariate  $t(3)$  density: See Table 1 for an explanation of the acronyms used

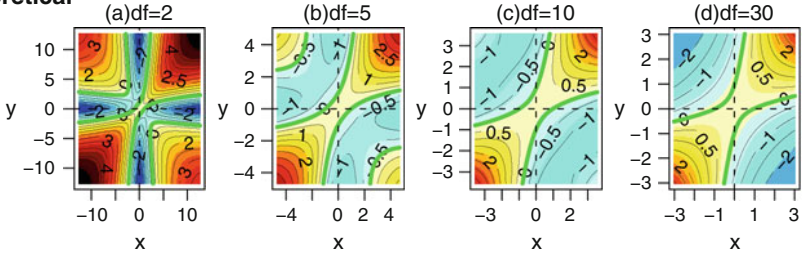
## 6 Empirical application: stock price changes and volume

The relationship between price and quantity is the fundamental building block of economic theory. This section illustrates an application example of the proposed local dependence analysis by investigating the contemporaneous relationship between stock price changes and volume.

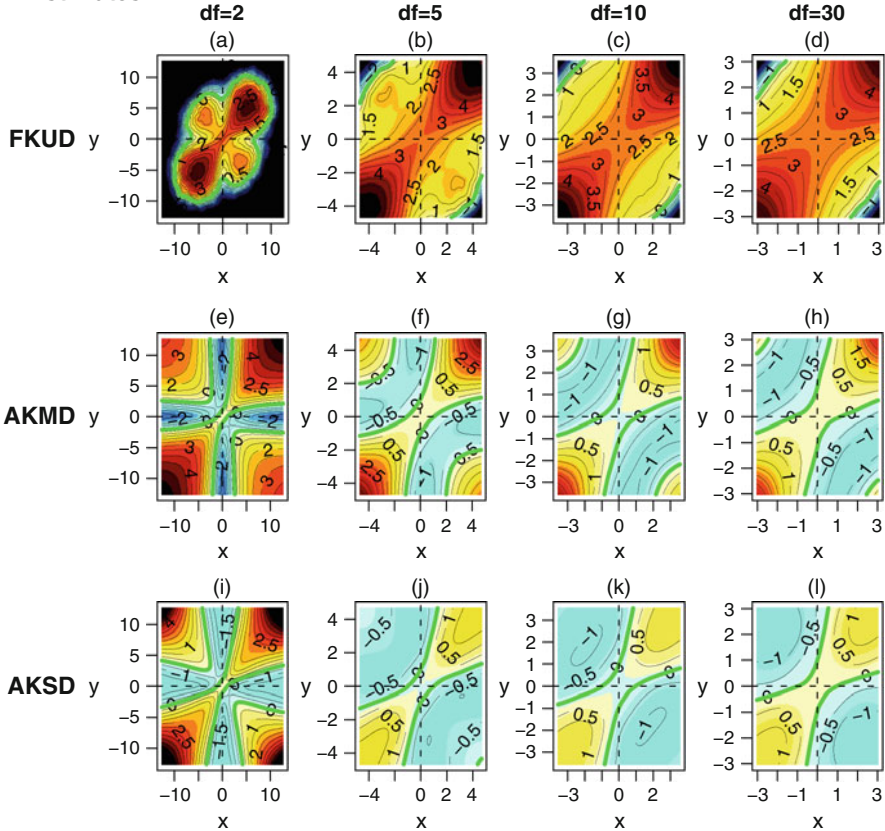
### 6.1 Related studies and background discussion

The estimation methods used for investigating return-volume relations have been mostly based on linear regression, and a *global* relationship has been the major concern of many studies. Only few studies have investigated a conditional relationship, where linear regression is used by dividing the data range into two to three sections. However, these estimates are too partial to capture an entire dependence structure. The study by Gallant et al. (1992) applying the joint SNP density estimation of the daily Standard and Poor’s composite stock price index and the total volume of shares

**A Theoretical**

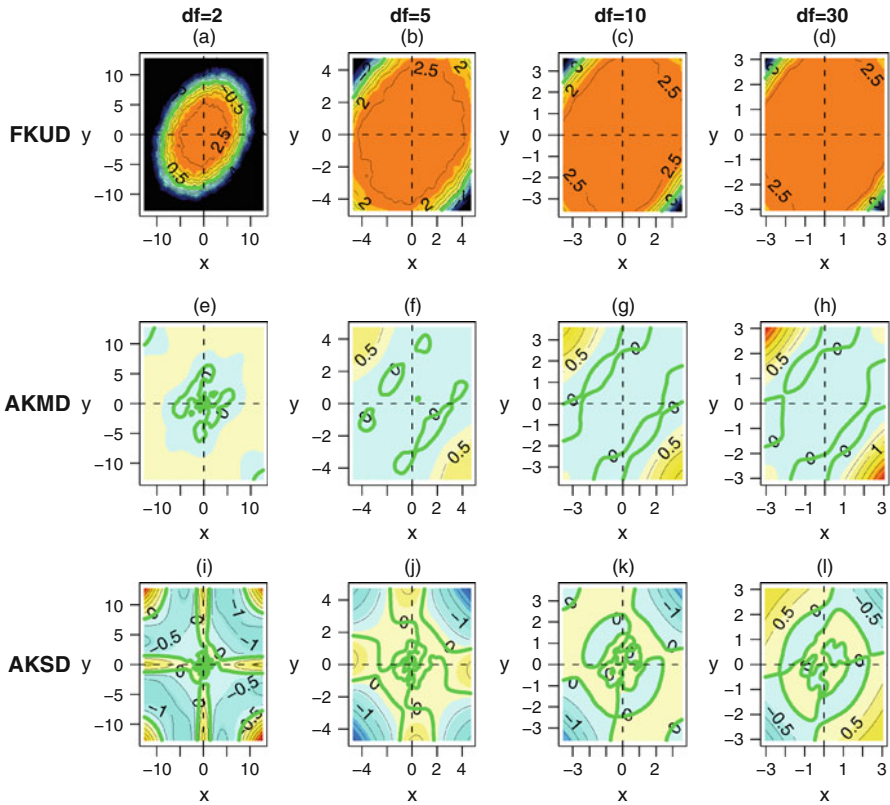


**B Estimates**



**Fig. 8** Tail effect on the pointwise MI estimation of dependent bivariate  $t$  density: (Panels A) Theoretical pointwise MI; (Panels B) Pointwise MI estimates. Sample size  $n = 10000$ . See Table 1 for an explanation of the acronyms used

traded at the NYSE, is the first trial to capture a whole dependence structure between stock returns and volume. However, their approach, which is based on the SNP estimator, has serious defects in estimating fat-tailed densities, as shown in a comparative study by Takada (2008). In particular, the credibility of their findings regarding tail behavior is suspicious due to artificial bumps frequently created in the SNP



**Fig. 9** Tail effect on the pointwise MI estimation error of dependent bivariate  $t$  density: Sample size  $n = 10000$ . See Table 1 for an explanation of the acronyms used

estimation. Accordingly, our proposed local dependence analysis based on pointwise MI is expected to reveal unknown features.

Many empirical studies have documented significantly positive global correlations between returns and volume. The empirical findings in the first review of [Karpoff \(1987\)](#) on the return-volume relationship can be summarized as follows: The positive dependence between price changes and volume is larger in cases of positive price changes than in cases of negative price changes. However, more detailed findings have been difficult to achieve thus far due to technical constraints.

### 6.2 Data and adjustments

The raw data consist of the daily closing prices of the NYSE composite index and the total number of shares traded on the NYSE. The sample period is from December 31, 1965 to February 29, 2008 and comprises of 10609 raw observations. The adjustment of the raw data series is conducted in two steps. In the first step, the raw price series,  $P_t$ , is first differenced in the logs as  $\log P_t - \log P_{t-1}$ . The log of raw volume series,

$\log V_t$ , is nonlinearly detrended by  $\log V_t$ —the centered one-year moving average of  $\log V_t$ . Due to this detrending, the size of the adjusted data is shortened to 10358 for the period from June 30, 1966, to August 30, 2007.

In the next step, systematic calendar effects and seasonalities are removed from  $\log P_t - \log P_{t-1}$  and detrended  $\log V_t$ , in order to obtain the adjusted returns and detrended log volume series,  $\Delta p_t$  and  $v_t$ . The selection of dummy variables used in the seasonal adjustment regression follows those of Gallant et al. (1992): Day of the week, gap in trading days, month variables from March through November, and January effects comprise of each week of December and January. The total number of dummies is 24, which has sufficient degrees of freedom for the 10358 adjusted observations. Let  $Y$  be a vector of pre-adjusted data with length  $n$ , and  $\mathbf{X}$  be a matrix consisting of adjustment dummy vectors. The seasonal adjustment regression is  $Y = \mathbf{X}'\beta + \epsilon$ , where  $\beta$  denotes the vector of least squares coefficients, and the vector of residuals,  $\epsilon$ . The adjusted series,  $Y_{\text{adj}}$ , is obtained as  $Y_{\text{adj}} = Y - \mathbf{X}'\beta$ . The seasonalities of the two series, which are found to be significant, are in agreement with those detected by Gallant et al. (1992). The basic statistics of the adjusted series,  $\Delta p_t$  and  $v_t$ , are summarized in Table 2.

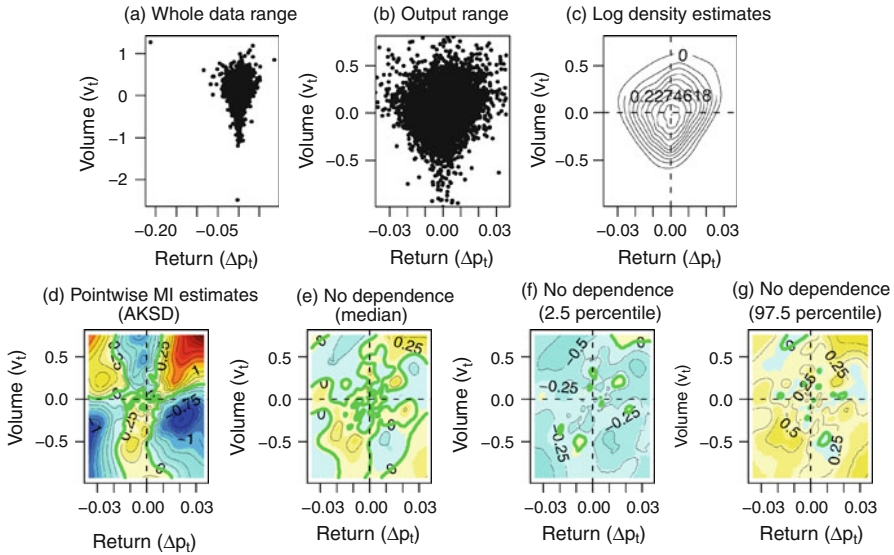
Sample observations tend to include outliers, which are more likely for fat-tailed distributions such as those of financial asset returns. Figure 10(a) plots the contemporaneous adjusted returns and volume,  $\{\Delta p_t, v_t\}$ , and thus all data are included. The existence of outliers compress the mass of the data into a small region, worsening the resolution level in investigating subtle patterns. To avoid this problem, we set the bounds for an output range used in generating joint density estimates. The scatter plot of Fig. 10(b) illustrates the output range used for the density estimation. Let  $n$  be the number of observations,  $\{\Delta p_t, v_t\}$ . The bound is then set at the 0.2% extreme data point (or 0.002  $n$ -th extremes) in each coordinate direction. If grid of either  $\Delta p_t$  or  $v_t$  of a data point exceeds the bound, it is eliminated from the output range. Since the same data can be removed more than once in this algorithm, the total amount of data removed from the output region is usually about 0.4% of the sample size. Note that the information obtained from all data, including those outside the output region, are used in estimating the densities.

### 6.3 Results

We analyze the contemporaneous dependence between the adjusted daily NYSE composite index returns,  $\Delta p_t$ , and the adjusted detrended log volume,  $v_t$ . Figure 10(b) shows that the variability of price changes tend to be larger when volume is larger. Figure 10(c) shows the bivariate adaptive kernel density estimates of the NYSE data

**Table 2** Basic statistics for the full sample period of adjusted returns and volume

	Mean	Median	SD	Skewness	Kurtosis
$\Delta p_t$	0.0003	0.0004	0.0090	-1.42	37.93
$v_t$	-0.0006	-0.0091	0.1954	-0.10	8.56



**Fig. 10** Joint density and pointwise MI estimates of contemporaneous NYSE index returns and volume: (Panels (a–c)) Scatter plot and adaptive kernel log density estimates; (Panel (d)) AKSD estimates of pointwise MI; (Panels (e–g)) The median and the 95% confidence bands of the null profile for assessing the significance of Panel (e). See Table 1 for an explanation of AKSD. In Panels (d–g), the *green line* implies no dependence. *Yellow to orange* imply positive pointwise MI, indicating that the volume tends to be high (low) when the returns are very high (low). *Blue* implies negative pointwise MI indicating that the volume tends not to be high (low) when the returns are very high (low)

sets based on all the adjusted data shown in Fig. 10(a). The joint density estimate in Fig. 10(c) is almost symmetric, and interesting feature cannot be observed. The ordinary Pearson correlation between the corresponding returns and volume is 0.10.

From the pointwise MI plot of Fig. 10(d), we can explore rich information on the local dependence between the stock returns and volume of the adjusted NYSE data. Considering the fact that an image estimated by the AKSD method is fainter than the true image, the contours are set doubly finer than in the cases of experimental studies in Sect. 5. In Fig. 10(d), the significant positive pointwise MI indicated in orange is found in the first quadrant, which implies that positive large price changes tend to accompany a high volume. The blue area in the fourth quadrant implies that a low volume is not likely to occur when price changes are positive and large. It is suggested that investors tend to trade actively when prices increase. Clear patterns are revealed that cannot be found from conventional approaches using a scatter plot, joint density estimates, or global measures of dependence.

Panels (e–g) of Fig. 10 provide the noise level of the estimation results of Fig. 10(d), based on 500 replicated non-dependent profiles obtained by randomizing the order of volume sequences using the procedure provided in Sect. 4.5. Figure 10(e) shows a median image of a sample of no dependence. Figure 10(f) and (g) represent 2.5 and 97.5 sample percentiles, giving a 95% confidence band for the no dependence case. For darker colors than those shown in the confidence band in Fig. 10(f) and (g), the corresponding dependence is considered to be significant.

Comparing the pointwise MI estimate in Fig. 10(d) with the noise level indicated in Fig. 10(e–g), we can infer that a subtle positive dependence might exist in the second and third quadrants. In other words, the accompanied volume size is different depending on the size of the price decrease: Trading is generally inactive when price decreases are small. Whereas many investors do not have strong incentives for buying, stock holders hesitate to realize losses by selling in the hope of a price increase. During the phase of a large price decrease, however, trading becomes active: While many stock holders will want to get out of the market, some will want to go bargain hunting.

A similar graphical pattern can be observed in Fig. 2.11 of Takada (2001), which illustrates the pointwise MI estimates of the NYSE and NASDAQ total market returns and volume data from January 1966 to May 2000, along with six individual stock data sets of different market capitalization by digit in US dollar for the sample period of July 1962 to December 1998. The individual stock data consist of General Electric, IBM, Ford Motor, Amerada Hess Corp., Universal Corp., and Edo Corp. Figure 10(d) is consistent with the stylized facts provided in the previous literature: A positive dependence between price changes and volume is higher for a positive price change than for a negative price change. Moreover, our pointwise MI estimates reveal a complete dependence structure, which enables the interpretation of investor behavior.

## 7 Discussion and conclusion

The proposed local and tail dependence analysis enables the capture of an entire dependence structure, providing insight into the functional relationship between two variables. By examining the statistical significance, even a weak tail dependence can be identified given data sets of relatively small sample sizes. The limitation of the proposed AKSD method is that the estimated pointwise MI pattern is fainter in the color plot, but the dependence structure is correctly captured. A comprehensive analysis of the error patterns of the AKMD and AKSD estimators might give us some clues on how to improve the pointwise MI estimation performance given the observations following unknown distributions.

As a direction for future research, there is a good possibility of extending into three to four dimensions or more. One reason for this possibility is the good pointwise MI estimation results achieved for sample sizes 100 to 1000 in two dimensions. Another reason is based on Monte Carlo study by Hwang et al. (1994). They reported that the performance of an adaptive kernel density estimator in fitting fat-tailed and multi-modal multivariate densities is the best for up to three or four dimensions. A complete dependence structure at higher dimensions is very useful, especially for capturing functional relationships among key variables. However, depending on the available sample size, the combination of pairwise results may be more informative.

**Acknowledgements** The author is deeply grateful to Roger Koenker and Shoji Takada for their many helpful suggestions and constructive comments. The author also wishes to thank the anonymous reviewers for their fruitful comments for improving this article. Financial support to the author from Ishii Memorial Securities Research Promotion Foundation, the Japan Society for the Promotion of Science (Grant-in-

Aid for Scientific Research) and the Japan Science and Technology Agency (PRESTO) are gratefully acknowledged.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Abramson IS (1982) On bandwidth variation in kernel estimates—a square root law. *Ann Stat* 10:1217–1223
- Afshin-Pour B, Soltanian-Zadeh H, Hossein-Zadeh GA, Grady CL, Strother SC (2010) A mutual information-based metric for evaluation of fMRI data-processing approaches. *Hum Brain Mapp*. doi:[10.1002/hbm.21057](https://doi.org/10.1002/hbm.21057)
- Basu A, Lindsay BG (1994) Minimum disparity estimation for continuous models: efficiency, distributions and robustness. *Ann Inst Stat Math* 46:683–705
- Barahona M, Poon CS (1996) Detection of nonlinear dynamics in short, noisy time series. *Nature* 381:215–217
- Battiti R (1994) Using mutual information for selecting features in supervised neural net learning. *IEEE Trans Neural Netw* 5:537–550
- Breiman L, Meisel W, Purcell E (1977) Variable kernel estimates of multivariate densities. *Technometrics* 19:135–144
- Charpentier A, Segers J (2009) Tails of multivariate Archimedean copulas. *J Multivar Anal* 100:1521–1537
- Chaudhuri P, Marron JS (2000) Scale space view of curve estimation. *Ann Stat* 28:408–428
- Cherubini U, Luciano E, Vecchiato A (2004) Copulas methods in finance. Wiley, New York
- Cover TM, Thomas JA (2006) Elements of information theory. Wiley, New York
- Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman and Hall, New York
- Fraser AM, Swinney HL (1986) Independent coordinates for strange attractors from mutual information. *Phys Rev A* 33:1134–1140
- Fukunaga K (1972) Introduction to statistical pattern recognition. Academic Press, New York
- Gallant AR, Rossi E, Tauchen G (1992) Stock prices and volume. *Rev Financ Stud* 5:199–242
- Gudendorf G, Segers J (2010) Extreme-value copulas. Copula theory and its applications. Lecture notes in statistics, vol 198. Springer, Heidelberg, pp 127–145
- Hlaváčková-Schindler K, Paluš M, Vejmelka M, Bhattacharya J (2007) Causality detection based on information-theoretic approaches in time series analysis. *Phys Rep* 441:1–46
- Huisman R, Koedijk KG, Kool CJM, Palm F (2001) Tail-index estimates in small samples. *J Bus Econ Stat* 19:208–216
- Hwang JN, Lay SR, Lippman A (1994) Nonparametric multivariate density estimation: a comparative study. *IEEE Trans Signal Process* 42:2795–2810
- Hyvärinen A, Oja E (2000) Independent component analysis: algorithms and applications. *Neural Netw* 13:411–430
- Joe H, Li H, Nikoloulopoulos AK (2010) Tail dependence functions and vine copulas. *J Multivar Anal* 101:252–270
- Jeong J, Gore JC, Peterson BS (2001) Mutual information analysis of the EEG in patients with Alzheimer's disease. *Clin Neurophysiol* 112:827–835
- Karpoff J (1987) The relation between price changes and trading volume: a survey. *J Financ Quant Anal* 22:109–126
- Khan S, Bandyopadhyay S, Ganguly AR, Saigal S, Erickson DJIII, Protopoulos V, Ostrouchov G (2007) Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data. *Phys Rev E* 76:026209
- Klüppelberg C, Kuhn G, Peng L (2007) Estimating the tail dependence function of an elliptical distribution. *Bernoulli* 13(1):229–251
- Kotz S, Nadarajah S (2004) Multivariate t distributions and their applications. Cambridge University Press, Cambridge
- Kraskov A, Stögbauer H, Grassberger P (2004) Estimating mutual information. *Phys Rev E* 69:066138



- Kuhn G, Khan S, Ganguly AR, Branstetter ML (2007) Geospatial-temporal dependence among weekly precipitation extremes with applications to observations and climate model simulations in South America. *Adv Water Resour* 30:2401–2423
- Kwaku N, Choi C-H (2002) Input feature selection by mutual information based on Parzen window. *IEEE Trans Pattern Anal Mach Intell* 24:1667–1671
- Lin CH, Kao TC (2008) Multiple structural changes in the tail behavior: evidence from stock index futures returns. *Nonlinear Anal: Real World Appl* 9:1702–1713
- Lindeberg T (1994) Scale space theory in computer vision. Kluwer, Boston
- March TK, Chapman SC, Dendy RO (2005) Mutual information between geomagnetic indices and the solar wind as seen by WIND: implications for propagation time estimates. *Geophys Res Lett* 32:L04101
- Mari DD, Kotz S (2001) Correlation and dependence. Imperial College Press, London
- Malevergne Y, Sornette D (2006) Extreme financial risks. Springer, Berlin
- McNeil AJ, Frey R, Embrechts P (2005) Quantitative risk management. Princeton, NJ
- Moon YI, Rajagopalan B, Lall U (1995) Estimation of mutual information using kernel density estimators. *Phys Rev E* 52:2318
- Nelsen RB (2006) An introduction to copulas. Springer, Berlin
- Paluš M, Komárek V, Hrnčíř Z, Štěrbová K (2001) Synchronization as adjustment of information rates: detection from bivariate time series. *Phys Rev E* 63:046211
- Papana A, Kugiumtzis D (2008) Evaluation of mutual information estimators on nonlinear dynamic systems. *Nonlinear Phenom Complex Syst* 11:225–232
- Patton A (2009) Copula-based models for financial time series. In: Andersen TG et al (ed) Handbook of financial time series. Springer Verlag, Berlin pp 767–785
- Pluim JPW, Maints JBA, Viergever MA (2003) Mutual-information-based registration of medical images: a survey. *IEEE Trans Med Imag* 22:986–1004
- Rousseeuw PJ, van Zomeren BC (1990) Unmasking multivariate outliers and leverage points. *J Am Stat Assoc* 85:633–639
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423 623653
- Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London
- Suzuki T, Sugiyama M, Kanamori T, Sese J (2009) Mutual information estimation reveals global associations between stimuli and biological processes. *BMC Bioinform* 10(Suppl 1). doi:[10.1186/1471-2105-10-S1-S52](https://doi.org/10.1186/1471-2105-10-S1-S52)
- Takada T (2001) Density estimation for robust financial econometrics. PhD thesis, University of Illinois at Urbana-Champaign
- Takada T (2008) Asymptotic and qualitative performance of nonparametric density estimators: a comparative study. *Econom J* 11:573–592
- Takada T (2009) Simulated minimum Hellinger distance estimation of stochastic volatility models. *Comput Stat Data Anal* 53:2390–2403
- Wicks RT, Chapman SC, Dendy RO (2007) Mutual information as a tool for identifying phase transitions in dynamical complex systems with limited data. *Phys Rev E* 75:051125