

## Research Article

# Dynamically Measuring Statistical Dependencies in Multivariate Financial Time Series Using Independent Component Analysis

**Nauman Shah and Stephen J. Roberts**

*Machine Learning Research Group, Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK*

Correspondence should be addressed to Nauman Shah; [nauman.shah@eng.ox.ac.uk](mailto:nauman.shah@eng.ox.ac.uk)

Received 30 March 2013; Accepted 4 May 2013

Academic Editors: A. Krzyzak, C.-M. Kuo, S. Kwong, W. Liu, and F. Perez-Cruz

Copyright © 2013 N. Shah and S. J. Roberts. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present a computationally tractable approach to dynamically measure statistical dependencies in multivariate non-Gaussian signals. The approach makes use of extensions of independent component analysis to calculate information coupling, as a proxy measure for mutual information, between multiple signals and can be used to estimate uncertainty associated with the information coupling measure in a straightforward way. We empirically validate relative accuracy of the information coupling measure using a set of synthetic data examples and showcase practical utility of using the measure when analysing multivariate financial time series.

## 1. Introduction

The task of accurately inferring the statistical dependency structure (association) in multivariate systems has been an area of active research for many years, with a wide range of practical applications [1]. Many of these applications require real-time sequential analysis of dependencies in multivariate data streams with dynamically changing properties. However, most existing measures of dependence have some serious limitations; in terms of the type of data sets they are suitable for or in their computational complexities. If the data being analysed is generated using a known stable process, with known marginal and multivariate distributions, the degree of dependence can be estimated relatively easily. However, most real-world data sets have dynamically changing properties to which a single distribution cannot be assigned. Multivariate data generated in global financial markets is an example of such complex data sets. Financial data exhibits rapidly changing dynamics and is non-Gaussian in nature; this is especially true for financial data recorded at high frequencies [2]. In fact, as the scale over which financial returns are calculated decreases, their distribution becomes increasingly non-Gaussian, a feature referred to as aggregational Gaussianity. The recent explosive growth in availability and use of financial

data sampled at high frequencies therefore requires the use of computationally efficient algorithms which are suitable for dynamically analysing dependencies in non-Gaussian data streams.

The most commonly used measure of statistical dependence is linear correlation. However, practical use of the linear correlation measure has three main limitations; that is, it cannot accurately model dependencies between signals with non-Gaussian distributions [3]; it is restricted to measuring linear statistical dependencies and is very sensitive to outliers [4]. Rank correlation is another frequently used measure of association. However, it is only valid for monotonic functions and is not suitable for large data sets, as assigning ranks to a large number of observations is computationally demanding. Financial returns often have a large fraction of zero values, which result in tied ranks [5]. Rank correlation measures cannot accurately deal with the presence of tied ranks and hence the results obtained can be misleading [6]. Another widely used method for multivariate dependence analysis in the financial sector is the use of copulas [7]. However, copula-based methods also suffer from major limitations in practice; for example, computation of copula functions involves calculating multiple moments as well as integration of joint distributions, which require use of numerical

methods and hence become computationally complex [8]. Copula-based methods suffer from other major limitations as well, namely, the difficulties in accurate estimation of the copula functions, the empirical choice of the type of copulas, and problems in the design and use of time-dependent copulas [9]. Mutual information, the canonical measure of statistical dependence, is also used in practice. However, accurate computation of mutual information using finite data sets can be computationally complex (as we discuss later). In this paper, we present a computationally efficient independent component analysis (ICA) based approach to dynamically measure information coupling in multivariate non-Gaussian data streams as a proxy measure for mutual information.

The paper is organised as follows. We first discuss the need for developing an ICA-based information coupling measure and present the theoretical framework underlying the development of our approach. We then present a brief introduction to the principles of ICA and discuss our method of choice for accurately and efficiently inferring the ICA unmixing matrix in a dynamic environment. We then proceed to present the ICA-based information coupling metric and describe its properties. Finally, we present a set of synthetic and financial data examples which make use of the information coupling measure to estimate dependencies in non-Gaussian signals.

## 2. Measuring Dependencies Using ICA: A Conceptual Overview

Mutual information is the canonical measure of statistical dependence in multivariate systems [10]. It is a quantitative measurement of how much information the observation of one variable gives us regarding another variable. Whilst the computation of mutual information is conceptually straightforward when the full probability density functions (pdf) of the variables under consideration are available, it is often difficult to accurately estimate mutual information directly using finite data sets. This is especially true in high-dimensional spaces, in which computation of mutual information requires the estimation of multivariate joint distributions, a process which is unreliable (being exquisitely sensitive to the joint pdf over the variables of interest) as well as computationally expensive [11]. Existing approaches to compute mutual information include methods that are based on ranking of variables, kernel density estimation, k-nearest neighbours, and the Edgeworth approximation of differential entropy [12, 13]. However, for most finite data sets, none of these techniques, all of which impose a trade-off between computational complexity and accuracy, outperforms the other methods and all these approaches can be extremely sensitive to the presence of noise [12]. The accuracy of these approaches is also highly sensitive to the choice of the model parameters, such as the number of kernels or neighbours. Therefore, in most practical cases, the direct use of mutual information is not feasible. However, as we discuss below, it is possible to make use of information encoded in the ICA

unmixing matrix to calculate information coupling as a proxy measure for mutual information.

Let us first take a look at the conceptual basis on which we can use ICA as a tool for measuring statistical dependencies. According to its classical definition, ICA estimates an unmixing matrix such that the mutual information between the independent source signals is minimised [14]. Hence, we can consider the unmixing matrix to contain information about the degree of mutual information between the observed signals. Although the direct computation of mutual information can be very expensive, alternative efficient approaches to ICA, which do not involve direct computation of mutual information, exist. Hence, it is possible to indirectly obtain an estimate for mutual information by using the ICA-based information coupling measure as a proxy. Now let us consider some properties of financial returns which make them well-suited to be analysed using ICA (as this paper is focused on financial applications, therefore we consider the case of measuring dependencies in financial data; however, similar ideas can be applied to most real-world systems which give rise to non-Gaussian data). Financial markets are influenced by many independent factors, all of which have some finite effect on any specific financial time series. These factors can include, among others, news releases, price trends, macroeconomic indicators, and order flows. We hypothesise that the observed multivariate financial data may hence be generated as a result of linear combination of some hidden (latent) variables [15, 16]. This process can be quantitatively described by using a linear generative model, such as principal component analysis (PCA), factor analysis (FA), or ICA. As financial returns have non-Gaussian distributions with heavy tails, therefore PCA and FA are not suitable for modelling multivariate financial data, as both these second-order approaches are based on the assumption of Gaussianity [17]. ICA, in contrast, takes into account non-Gaussian nature of the data being analysed by making use of higher-order statistics. ICA has proven applicability for multivariate financial data analysis; some interesting applications are presented in [15, 16, 18]. These, and other similar studies, make use of ICA primarily to extract the underlying latent source signals. However, all relevant information about the source mixing process is contained in the ICA unmixing matrix, which hence encodes dependencies. Therefore, in our analysis we only make use of the ICA unmixing matrix (without extracting the independent components) to measure information coupling. The ICA-based information coupling model we present in this paper can be used to directly measure statistical dependencies in high-dimensional spaces. This makes it particularly attractive for a range of practical applications in which relying solely on pair-wise analysis of dependencies is not feasible (there is surprisingly little work done towards addressing the important issue of estimating the dependency structure in high-dimensional multivariate systems, although there has been interest in this field for a long time [19]. High-dimensional analysis of information coupling has various important applications in the financial sector, including active portfolio management, multivariate financial risk analysis, statistical arbitrage, and pricing and hedging of various instruments [9]).

### 3. Independent Components, Unmixing and Non-Gaussianity

Mixing two or more unique signals, to a set of mixed observations, results in an increase in the dependency of the pdfs of the mixed signals. The marginal pdfs of the observed mixed signals become more Gaussian due to the central limit theorem [20]. The mixing process also results in a reduction in the independence of the mixed signal distribution and hence increase in mutual information associated with it. Moreover, there is a rise in the stationarity of the mixed signals, which have flatter spectra as compared to the original sources [21]. Consider a set of  $N$  observed signals  $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_N(t)]^T$  at the time instant  $t$ , which are a mixture of  $M$  source signals  $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_M(t)]^T$ , mixed linearly using a mixing matrix  $\mathbf{A}$ , with observation noise  $\mathbf{n}(t)$ , as per

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t). \quad (1)$$

Independent component analysis (ICA) attempts to find an unmixing matrix  $\mathbf{W}$ , such that the  $M$  recovered source signals  $\mathbf{b}(t) = [b_1(t), b_2(t), \dots, b_M(t)]^T$  are given by

$$\mathbf{b}(t) = \mathbf{W}(\mathbf{x}(t) - \mathbf{n}(t)). \quad (2)$$

For the case where observation noise  $\mathbf{n}(t)$  is assumed to be normally distributed with a mean of zero, the least squares expected value of the recovered source signals is given by:

$$\hat{\mathbf{b}}(t) = \mathbf{W}\mathbf{x}(t), \quad (3)$$

where  $\mathbf{W}$  is the pseudo-inverse of  $\mathbf{A}$ ; that is,

$$\mathbf{W} = \mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T. \quad (4)$$

In the case of square mixing,  $\mathbf{W} = \mathbf{A}^{-1}$ .

Most ICA approaches make implicit or explicit assumptions regarding the parametric model of the pdfs of the independent sources [21]; for example, Gaussian mixture distributions are used as source models in [22, 23], while [24] makes use of a flexible source density model given by the generalised exponential distribution. In our analysis, we use a reciprocal cosh source model as a canonical heavy-tailed distribution; namely [21],

$$p(s_i) = \frac{1}{Z_0 \cosh(s_i)}, \quad (5)$$

where  $s_i$  is the  $i$ th source and  $Z_0$  is a normalising constant. This analytical fixed source model has no adjustable parameters; therefore, it has considerable computational advantages over alternative source models. Also, as this source model is heavier in the tails, it is able to accurately model heavy-tailed unimodal non-Gaussian distributions, such as financial returns.

**3.1. Inference.** For our analysis, we make use of the *icadec* algorithm [21, 24] to infer the unmixing matrix. This algorithm constrains the unmixing matrix to the manifold of

decorrelating matrices, thus offering rapid computation. The algorithm also gives accurate results compared to other related ICA approaches [24] and allows us to obtain a confidence measure for the unmixing matrix. Here we present a brief overview of this algorithm; an in-depth description is presented in [24].

The independent source signals obtained using ICA,  $\mathbf{b}(t)$ , must be at least linearly decorrelated for them to be classed as independent. The *icadec* algorithm makes use of this property of the independent components to efficiently infer the unmixing matrix. For a set of observed signals,  $\mathbf{X}$ , where  $\mathbf{X} = [\mathbf{x}(t)]_{t=1}^{t=T}$ ; the set of recovered independent components,  $\mathbf{B} = [\mathbf{b}(t)]_{t=1}^{t=T}$ , is given by  $\mathbf{B} = \mathbf{W}\mathbf{X}$ . The independent components are linearly decorrelated if

$$\mathbf{B}\mathbf{B}^T = \mathbf{W}\mathbf{X}\mathbf{X}^T\mathbf{W}^T = \mathbf{D}^2, \quad (6)$$

where  $\mathbf{D}$  is a diagonal matrix of scaling factors. The singular value decomposition of the set of observed signals is given by

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \quad (7)$$

where  $\mathbf{U}$  and  $\mathbf{V}$  are orthogonal matrices with the columns of  $\mathbf{U}$  being the principal components of  $\mathbf{X}$ , and  $\mathbf{\Sigma}$  is a diagonal matrix of the singular values of  $\mathbf{X}$ . It can be shown that the decorrelating matrix,  $\mathbf{W}$ , can then be written as [24]

$$\mathbf{W} = \mathbf{D}\mathbf{Q}\mathbf{\Sigma}^{-1}\mathbf{U}^T, \quad (8)$$

where  $\mathbf{Q}$  is a real orthogonal matrix. To obtain an estimate for the ICA unmixing matrix, we need to optimise a given contrast function (we use log-likelihood of the data, as described later). There are a variety of optimisation approaches which can be used; our approach of choice is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method, which gives the best estimate of the minimum of the negative log-likelihood in a computationally efficient manner and also provides us with an estimate for the Hessian matrix. However, parameterising the optimisation problem directly by the elements of  $\mathbf{Q}$  makes it a constrained minimisation problem for which BFGS is not applicable. Therefore, to convert it into an unconstrained minimisation problem, we constrain  $\mathbf{Q}$  to be orthonormal by parameterising its elements as the matrix exponential of a skew-symmetric matrix  $\mathbf{J}$  (nonzero elements of this matrix are known as the Cayley coordinates), whose above diagonal elements parameterise  $\mathbf{Q}$  [21] (for  $M$  sources and  $N$  observed signals, the ICA unmixing matrix may be optimised in the  $(1/2)M(M+1)$  dimensional space of decorrelating matrices rather than in the full  $MN$  dimensional space, as  $(1/2)M(M-1)$  and  $M$  parameters are required to specify  $\mathbf{Q}$  and  $\mathbf{D}$  respectively. This feature offers considerable computational benefits (especially in high-dimensional spaces) and the resulting matrix hence obtained is guaranteed to be decorrelating [24]):

$$\mathbf{Q} = \exp(\mathbf{J}). \quad (9)$$

Using this parameterisation makes it possible to apply BFGS to any contrast function; the contrast function used as part of

the *icadec* algorithm is an expression for the log-likelihood of the data, as described below.

Using (1) and assuming that the observation noise ( $\mathbf{n}$ ) is normally distributed with a mean of zero and having an isotropic covariance matrix with precision  $\beta$ , the distribution of the observations (as a preprocessing step, we normalise each observed signal to have a mean of zero and unit variance) conditioned on  $\mathbf{A}$  and  $\mathbf{s}$  (where we drop the time index  $t$  for ease of presentation) is given by

$$p(\mathbf{x} | \mathbf{A}, \mathbf{s}) = \mathcal{N}(\mathbf{x}; \mathbf{A}\mathbf{s}, \beta^{-1}\mathbf{I}), \quad (10)$$

where  $\mathbf{A}\mathbf{s}$  is the mean of the normal distribution and  $\beta^{-1}\mathbf{I}$  is its covariance. The likelihood of an observation occurring is given by

$$p(\mathbf{x} | \mathbf{A}) = \int p(\mathbf{x} | \mathbf{A}, \mathbf{s}) p(\mathbf{s}) d\mathbf{s}. \quad (11)$$

Assuming that the distribution over sources has a single dominant peak, in this case given by the maximum likelihood source estimates  $\hat{\mathbf{s}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{x}$ , the integral in (11) can be analysed by using a simplified (computationally efficient) variant of Laplace's method, as shown in [21, 25]:

$$\begin{aligned} p(\mathbf{x} | \mathbf{A}) &= \int p(\mathbf{x} | \mathbf{A}, \mathbf{s}) p(\mathbf{s}) d\mathbf{s} \\ &\approx p(\mathbf{x} | \mathbf{A}, \hat{\mathbf{s}}) p(\hat{\mathbf{s}}) (2\pi)^{M/2} \det(\mathbf{G})^{-1/2}, \end{aligned} \quad (12)$$

where  $\mathbf{G}$  is the Hessian matrix:

$$\mathbf{G} = - \left[ \frac{\partial^2 \log p(\mathbf{x} | \mathbf{A}, \mathbf{s})}{\partial \mathbf{s}_i \partial \mathbf{s}_j} \right]_{\mathbf{s}=\hat{\mathbf{s}}}. \quad (13)$$

Taking log of the expanded form of (10) gives

$$\log p(\mathbf{x} | \mathbf{A}, \mathbf{s}) = \frac{N}{2} \log \left( \frac{\beta}{2\pi} \right) - \frac{\beta}{2} (\mathbf{x} - \mathbf{A}\mathbf{s})^T (\mathbf{x} - \mathbf{A}\mathbf{s}) \quad (14)$$

which, via (13), results in  $\mathbf{G} = \beta \mathbf{A}^T \mathbf{A}$ . The log-likelihood,  $\ell \equiv \log p(\mathbf{x} | \mathbf{A})$ , is therefore

$$\begin{aligned} \ell &= \frac{N-M}{2} \log \left( \frac{\beta}{2\pi} \right) - \frac{\beta}{2} (\mathbf{x} - \mathbf{A}\hat{\mathbf{s}})^T (\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}) \\ &\quad + \log p(\hat{\mathbf{s}}) - \frac{1}{2} \log \det(\mathbf{A}^T \mathbf{A}). \end{aligned} \quad (15)$$

By using (8), we obtain  $\mathbf{A} = \mathbf{U}\Sigma\mathbf{Q}^T\mathbf{D}^{-1}$ . Hence, the log-likelihood becomes [21]

$$\ell = \frac{N-M}{2} \log \left( \frac{\beta}{2\pi e} \right) + \log p(\hat{\mathbf{s}}) + \log \det(\Sigma^{-1}\mathbf{D}), \quad (16)$$

where  $e$  is the average reconstruction error. Noting that we use a reciprocal cosh source model (as given by (5)), it can be shown that taking the derivative of this log-likelihood expression with respect to  $\mathbf{D}$  and  $\mathbf{J}$  (which parameterises  $\mathbf{Q}$ ), and by following the resulting likelihood gradient using a BFGS optimiser, makes it possible to efficiently compute an optimum value for the ICA unmixing matrix; details of this procedure are presented in [24].

**3.2. Dynamic Mixing.** The standard (offline) ICA model uses all available data samples at times  $t = 1, 2, \dots, T$  of the observed signals,  $\mathbf{x}(t)$ , to estimate a single static unmixing matrix,  $\mathbf{W}$ . The unmixing matrix obtained provides a good estimate of the mixing process for the complete time series and is well suited for offline data analysis. However, many time series, such as financial data streams, are highly dynamic in nature with rapidly changing properties and therefore require a source separation method that can be used in a sequential manner. This issue is addressed here by using a sliding-window ICA model [26]. This model makes use of a sliding-window approach to sequentially update the current unmixing matrix using information contained in the previous window and can easily handle nonstationary data. The unmixing matrix for the current window,  $\mathbf{W}(t)$ , is used as a prior for computing the unmixing matrix for the next window,  $\mathbf{W}(t+1)$ . This results in significant computational efficiency as fewer iterations are required to obtain an optimum value for  $\mathbf{W}(t+1)$ . The algorithm also results in an improvement in the source separation results obtained when the mixing process is drifting and addresses the ICA permutation and sign ambiguity issues [21], by maintaining a fixed (but of course arbitrary) ordering of recovered sources through time.

## 4. Information Coupling

We now proceed to derive the ICA-based information coupling metric. Later in this section we discuss the practical advantages this metric offers when used to analyse multivariate financial time series.

**4.1. Coupling Metric.** Let  $\mathbf{W}$  be any arbitrary square ICA unmixing matrix (for the purpose of brevity and clarity, we only consider the case of square mixing while deriving the metric here. However, the metric derived in this section is valid for nonsquare mixing as well, and the corresponding derivation can be undertaken using a similar approach as presented here but converting the nonsquare ICA unmixing matrices in each instance into square matrices by padding them with zeros):

$$\mathbf{W} \in \mathbb{R}^{N \times N}. \quad (17)$$

Since multiplication of  $\mathbf{W}$  by a diagonal matrix does not affect the mutual information of the recovered sources; therefore, we row normalise the unmixing matrix in order to address the ICA scale indeterminacy problem (most ICA algorithms, including *icadec*, suffer from the scale indeterminacy problem; that is, the variances of the independent components cannot be determined; this is because both the unmixing matrix and the source signals are unknown and any scalar multiplication on either will be lost in the mixing process). Row normalisation implies that the elements  $w_{ij}$  of the unmixing matrix  $\mathbf{W}$  are constrained, such that each row of the matrix is of unit length; that is,

$$\sum_{j=1}^N w_{ij}^2 = 1 \quad (18)$$



for all rows  $i$ . For a set of observed signals to be completely decoupled, their latent independent components must be the same as the observed signals; therefore, the row-normalised unmixing matrix for decoupled signals ( $\mathbf{W}_0$ ) must be a permutation of the identity matrix ( $\mathbf{I}$ ):

$$\mathbf{W}_0 = \mathbf{P}\mathbf{I} \in \mathbb{R}^{N \times N}, \quad (19)$$

where  $\mathbf{P}$  is a permutation matrix. For the case where the observed signals are completely coupled, all the latent independent components must be the same; therefore, the row-normalised unmixing matrix for completely coupled signals ( $\mathbf{W}_1$ ) is given by

$$\mathbf{W}_1 = \frac{1}{\sqrt{N}}\mathbf{K} \in \mathbb{R}^{N \times N}, \quad (20)$$

where  $\mathbf{K}$  is the unit matrix (a matrix of ones).

To calculate coupling, we need to consider the *distance* between any arbitrary unmixing matrix ( $\mathbf{W}$ ) and the zero coupling matrix ( $\mathbf{W}_0$ ). The distance measure we use is the generalised 2-norm, also called the spectral norm, of the difference between the two matrices [27], although we can use some other norms as well to get similar results. The spectral norm of a matrix corresponds to its largest singular value and is the matrix equivalent of the vector Euclidean norm. Hence, the distance,  $d(\mathbf{W}, \mathbf{W}_0)$ , between the two matrices can be written as

$$d(\mathbf{W}, \mathbf{W}_0) = \|\mathbf{W} - \mathbf{W}_0\|_2, \quad (21)$$

where  $\|\cdot\|_2$  is the spectral norm of the matrix. As  $\mathbf{W}_0$  is a permutation of the identity matrix, therefore,

$$d(\mathbf{W}, \mathbf{W}_0) = \|\mathbf{W} - \mathbf{P}\mathbf{I}\|_2. \quad (22)$$

As the spectral norm of a matrix is independent of its permutations, therefore, we may define another permutation matrix ( $\hat{\mathbf{P}}$ ) such that

$$d(\mathbf{W}, \mathbf{W}_0) = \|\hat{\mathbf{P}}\mathbf{W} - \mathbf{I}\|_2. \quad (23)$$

For this equation, the following equality holds:

$$d(\mathbf{W}, \mathbf{W}_0) = \|\hat{\mathbf{P}}\mathbf{W}\|_2 - 1. \quad (24)$$

Again, noting that  $\|\hat{\mathbf{P}}\mathbf{W}\|_2 = \|\mathbf{W}\|_2$ , we have

$$d(\mathbf{W}, \mathbf{W}_0) = \|\mathbf{W}\|_2 - 1. \quad (25)$$

We normalise this measure with respect to the range over which the distance measure can vary, that is, the distance between matrices representing completely coupled ( $\mathbf{W}_1$ ) and decoupled ( $\mathbf{W}_0$ ) signals. From (20) we have that  $\mathbf{W}_1 = (1/\sqrt{N})\mathbf{K}$ ; therefore,

$$d(\mathbf{W}_1, \mathbf{W}_0) = \|\mathbf{W}_1 - \mathbf{W}_0\|_2 = \left\| \frac{1}{\sqrt{N}}\mathbf{K} - \mathbf{P}\mathbf{I} \right\|_2. \quad (26)$$

Using the same analysis as presented previously, this equation can be simplified to

$$d(\mathbf{W}_1, \mathbf{W}_0) = \frac{1}{\sqrt{N}}\|\mathbf{K}\|_2 - 1. \quad (27)$$

For a  $N$ -dimensional square unit matrix, the spectral norm is given by  $\|\mathbf{K}\|_2 = N$ . Therefore, for a row-normalised unit matrix, the spectral norm is  $(1/\sqrt{N})\|\mathbf{K}\|_2 = \sqrt{N}$ . Hence, if  $\mathbf{W}$  is row-normalised, (27) can be written as

$$d(\mathbf{W}_1, \mathbf{W}_0) = \sqrt{N} - 1. \quad (28)$$

The normalised information coupling metric ( $\eta$ ) is then defined as

$$\eta = \frac{d(\mathbf{W}, \mathbf{W}_0)}{d(\mathbf{W}_1, \mathbf{W}_0)}. \quad (29)$$

Substituting (25) and (28) into (29), the normalised information coupling between  $N$  observed signals is given by

$$\eta = \frac{\|\mathbf{W}\|_2 - 1}{\sqrt{N} - 1}. \quad (30)$$

We can consider the bounds of  $\eta$  as described below. Suppose  $\mathbf{M}$  is an arbitrary real matrix. We can look upon the spectral norm of this matrix ( $\|\mathbf{M}\|_2 = \|\mathbf{M} - \mathbf{0}\|_2$ ) as a measure of departure (distance) of  $\mathbf{M}$  from a null matrix ( $\mathbf{0}$ ) [28]. The bounds on this norm are given by

$$0 \leq \|\mathbf{M}\|_2 \leq \infty. \quad (31)$$

If  $\mathbf{M}$  is a row-normalised ICA unmixing matrix ( $\mathbf{W}$ ), then (as discussed earlier) it lies between  $\mathbf{W}_0$  and  $\mathbf{W}_1$ . Hence, the bounds on  $\mathbf{W}$  are

$$\|\mathbf{W}_0\|_2 \leq \|\mathbf{W}\|_2 \leq \|\mathbf{W}_1\|_2. \quad (32)$$

Using (19) and (20), we can write

$$\|\mathbf{P}\mathbf{I}\|_2 \leq \|\mathbf{W}\|_2 \leq \left\| \frac{1}{\sqrt{N}}\mathbf{K} \right\|_2 \quad (33)$$

which can be simplified to

$$1 \leq \|\mathbf{W}\|_2 \leq \sqrt{N}. \quad (34)$$

Rearranging terms in this inequality gives

$$0 \leq \frac{\|\mathbf{W}\|_2 - 1}{\sqrt{N} - 1} \leq 1 \quad (35)$$

which gives us the same coupling metric as in (30) and shows that the metric is normalised; that is,  $0 \leq \eta \leq 1$ . For real-valued  $\mathbf{W}$ ,  $\|\mathbf{W}\|_2$  can be written as

$$\|\mathbf{W}\|_2 = \sqrt{\lambda_{\max}(\mathbf{W}^T \mathbf{W})}, \quad (36)$$

where  $\lambda_{\max}(\mathbf{W}^T \mathbf{W})$  is the maximum eigenvalue of  $\mathbf{W}^T \mathbf{W}$ . The unmixing matrix obtained using most ICA algorithms, including *icadec*, suffers from row permutation and sign ambiguity problems; that is, the rows are arranged in a random order and the sign of elements in each row is unknown [21, 24]. We note that as  $\|\mathbf{W}\|_2$  is independent of the sign and permutations of the rows of  $\mathbf{W}$ , therefore our measure of information coupling straightaway addresses the

problems of ICA sign and permutation ambiguities. Also, as the metric's value is independent of the row permutations of  $\mathbf{W}$ , therefore it provides symmetric results. The information coupling metric is valid for all dimensions of the unmixing matrix,  $\mathbf{W}$ . This implies that information coupling can be easily measured in high-dimensional spaces.

It is possible to obtain a measure of uncertainty in our estimation of the information coupling measure. We make use of the BFGS quasi-Newton optimisation approach over the most probable skew-symmetric matrix,  $\mathbf{J}$ , of (9), from which estimates for the unmixing matrix,  $\mathbf{W}$ , can be obtained, and thence the coupling measure  $\eta$  calculated. We also estimate the Hessian (inverse covariance) matrix,  $\mathbf{H}$ , for  $\mathbf{J}$ , as part of this process. Hence, it is possible to draw samples,  $\mathbf{J}'$ , say from the distribution over  $\mathbf{J}$ , as a multivariate normal:

$$\mathbf{J}' \sim \mathcal{MN}(\mathbf{J}, \mathbf{H}^{-1}). \quad (37)$$

These samples can be readily transformed to samples in  $\eta$  using (9), (8), and (30), respectively. Confidence bounds (and here we use the 95% bounds) may then be easily obtained from the set of samples for  $\eta$  (in our analysis we use 100 samples).

**4.2. Computational Complexity.** The information coupling algorithm achieves computational efficiency by making use of the sliding-window based decorrelating manifold approach to ICA. Making use of the reciprocal cosh based ICA source model also results in significant computational advantages. We now take a look at the comparative computational complexity of the information coupling measure and three frequently used measures of statistical dependence, that is, linear correlation, rank correlation, and mutual information. For bivariate data ( $n_s$  samples long), for which these four measures are directly comparable, linear correlation and rank correlation have time complexities of order  $O(n_s)$  and  $O(n_s \log n_s)$ , respectively [29], while mutual information and information coupling scale as  $O(n_s^2)$  and  $O(n_s)$ , respectively (there have been various estimation algorithms proposed for efficient computation of mutual information; however, they all result in increased estimation errors and require careful selection of various user-defined parameters [30]) [31]. Hence, even though the time complexity of the information coupling measure is of the same order as linear correlation, it can still accurately capture statistical dependencies in non-Gaussian data streams and is a computationally efficient proxy for mutual information.

For  $N$ -dimensional multivariate data, direct computation of mutual information has time complexity of order  $O(n_s^N)$  compared to  $O(n_s N^3)$  for the information coupling measure. In high-dimensions, even an approximation for mutual information can be computationally very costly. For example, using a Parzen-window density estimator, the mutual information computational complexity can be reduced to  $O(n_s n_b^N)$ , where  $n_b$  is the number of bins used for estimation [32], which will incur a very high computational cost even for relatively small values of  $N$ ,  $n_b$ , and  $n_s$ . As a simple example, Table 1 shows a comparison of computation time (in seconds) taken by mutual information and information

TABLE 1: Example showing comparison of average computation time (in seconds) of mutual information and information coupling, when these measures are used to analyse bivariate data sets containing different number of samples ( $n_s$ ). The approach used to estimate mutual information is based on a Parzen window based algorithm, as described in [45]. The computational cost of this algorithm is dependent on the window-size ( $h$ ). The values of  $h$  used for the simulations are  $h = 20$  for  $n_s = 10^2$ , and  $h = 100$  for all other simulations. Results are obtained using a 2.66 GHz processor as an average of 100 simulations.

Computation time (sec)	$n_s = 10^2$	$n_s = 10^3$	$n_s = 10^4$	$n_s = 10^5$
Mutual information	0.0214	2.8289	17.0101	119.5088
Information coupling	0.0073	0.0213	0.0561	0.5543

coupling measures for analysing bivariate data sets of varying lengths. As expected, mutual information estimation using the Parzen window based approach (which is considered to be a relatively efficient approach to compute mutual information) becomes computationally very demanding with an increase in the number of samples of the bivariate data set. In contrast, the information coupling measure is computationally efficient, even when used to analyse very large high-dimensional multivariate data sets.

**4.3. Capturing Market Dynamics.** To dynamically analyse information coupling in multivariate financial data streams, we need to make use of windowing techniques. Financial markets give rise to well-defined events, such as orders, trades and quote revisions. These events are irregularly spaced in *clock-time*; that is, they are asynchronous. Statistical models in *clock-time* make use of data aggregated over fixed intervals of time [33]. The time at which these events are indexed is called the *event-time*. Hence, for dynamic modelling, in *event-time* the number of data points can be regarded as fixed while time varies, while in *clock-time* the time period is considered to be fixed with variable number of data points. Although we may need adaptive windows in *clock-time*, we can use sliding-windows of fixed length in *event-time*. Using fixed length sliding-windows in *event-time* can be useful for obtaining consistent results when developing and testing different statistical models. Also, statistical models deployed for online analysis of financial data operate best in *event-time* as they often need to make decisions as soon as some new market information (such as quote update etc.) becomes available. Consider an online trading model making use of an adaptive window in *event-time*. At specific times of the day, for example, at times of major news announcements, trading volume can significantly increase. Hence, more data will be available to the algorithm and thus results obtained can be misleading [34]. Using a sliding-window of fixed length in *event-time* can overcome this problem. The length of the sliding-window needs to be selected appropriately. The financial application for which the model is being used is one of the factors which drives the choice of window length. As a general rule, for trading models a window of approximately the same size as the average time period between placing trades (inverse of trading frequency) is often used. This makes

it possible to accurately capture the rapidly evolving dynamics of the markets over the corresponding period, without being too long so as to only capture major trends or too short to capture noise in the data.

**4.4. Discussion.** The information coupling model offers us with multiple advantages when used to analyse multivariate financial data. Here we summarise some of the main properties the model, while the empirical results presented in the next section showcase some of its practical benefits.

- (i) The information coupling measure, a proxy for mutual information, is able to accurately pick up statistical dependencies in data sets with non-Gaussian distributions (such as financial returns).
- (ii) The information coupling algorithm is computationally efficient, which makes it particularly suitable for use in an online dynamic environment. This makes the algorithm especially attractive when dealing with data sampled at high frequencies. This is because with the ever-increasing use of high-frequency data, overcoming sources of latency is of utmost importance in a variety of applications in modern financial markets.
- (iii) It gives confidence levels on the information coupling measure. This allows us to estimate the uncertainty associated with the measurements.
- (iv) The metric provides normalised results; that is, information coupling ranges from 0 for decoupled systems to 1 for completely coupled systems. This makes it easier to analyse results obtained using the metric and to compare its performance with other similar measures of association. The metric also gives symmetric results.
- (v) The metric is valid for any number of signals in high-dimensional spaces; that is, it consistently gives accurate results irrespective of the number of time series between which information coupling is being computed. This makes it suitable for a range of financial applications.
- (vi) It is not data intensive; that is, it gives relatively accurate results even when a small sample size is used. This allows the metric to model the complex and rapidly changing dynamics of financial markets.
- (vii) It does not depend on user-defined parameters which can restrict its practical utility, as the evolving market conditions may require the parameters to be constantly updated, which may not be practical.

## 5. Results

We now present a set of synthetic and financial data examples showing the relative accuracy and practical utility of the information coupling measure. The following notations are used for different measures of statistical dependence in this

paper: ICA-based information coupling ( $\eta$ ), linear correlation ( $\rho$ ), rank correlation ( $\rho_R$ ), and normalised mutual information ( $I_N$ ). Normalisation of mutual information values ( $I$ ) is achieved using the following transformation [35]:

$$I_N = \sqrt{1 - \exp(-2I)}. \quad (38)$$

The financial data examples we present in this paper primarily make use of spot foreign exchange (FX) data. Spot price or spot rate is the price which is actually quoted for a currency transaction to take place. Return is the profit realised when a currency is traded. It is common practice to use the normalised log-returns of financial data in statistical analysis, instead of the raw data itself [36]. Using log-returns makes it possible to convert exponential problems into linear ones, thus significantly simplifying relevant analysis. A normalised log-returns data set, with a mean of zero and unit variance, can generally be regarded as a locally stationary process [37]. Therefore, many signal processing techniques meant solely for stationary processes can be successfully applied to the normalised log-returns time series in an adaptive environment. Denoting the mid-price, at time  $t$ , of a financial time series by  $P_t$ , the log-return value is given by

$$r_t = \ln \left( \frac{P_t}{P_{t-1}} \right). \quad (39)$$

The normalisation of log-returns is achieved by converting the data to a form such that it has a mean of zero and unit variance. This is easily achieved by removing the mean and dividing by the standard deviation of the data.

**5.1. Synthetic Data.** There is no single distribution which fits financial returns, especially those sampled at higher frequencies which tend to be highly non-Gaussian [2], although there have been attempts to model returns using a variety of distributions [38]. In this paper, we aim to capture the heavy-tailed, skewed, properties of financial returns using a Pearson type IV distribution [39, 40], which can be used to represent distributions with varying degrees of skewness and kurtosis and thus are useful for representing distributions of financial returns [41, 42]. The first four moments of the distribution can be uniquely determined by setting four parameters which characterise the distribution [43]. Until recently, due to its mathematical and computational complexity, this distribution has not been widely used in financial literature, although this is rapidly changing with advances in computational power and proposal of new, improved analytical methods [39, 42, 44].

To test accuracy of various measures of dependence, we need to generate coupled non-Gaussian synthetic data with known, predefined, correlation values. There is no straightforward way to simulate correlated random variables when their joint distribution is not known [46], as is the case with multivariate financial returns. One possible method that can be used to induce any desired predefined correlation between independent, randomly distributed variables, irrespective of their distributions, is commonly known as the Iman-Conover method, as presented in [47]. This method is based on

TABLE 2: Table showing accuracy of four measures of dependence, that is, information coupling ( $\eta$ ), linear correlation ( $\rho$ ), rank correlation ( $\rho_R$ ), and normalised mutual information ( $I_N$ ), when used to estimate the level of dependence in a coupled system with varying levels of true correlation ( $\rho_{TC}$ ).  $\rho_{TC}$  is induced in an independent bivariate, randomly distributed, system using the Iman-Conover method as described in the text. The dependence estimates, together with their standard deviation values, shown in the table are obtained using 1000 independent simulations using 1000 data points long data sets for each simulation. The last row of the table gives values for the mean absolute error (MAE).

$\rho_{TC}$	$\eta$	$\rho$	$\rho_R$	$I_N$	$ \rho_{TC} - \eta $	$ \rho_{TC} - \rho $	$ \rho_{TC} - \rho_R $	$ \rho_{TC} - I_N $
0	$0.0046 \pm 0.0026$	$0.0038 \pm 0.0022$	$0.0147 \pm 0.0115$	$0.1851 \pm 0.0401$	0.0046	0.0038	0.0147	0.1851
0.1	$0.1079 \pm 0.0073$	$0.0917 \pm 0.0058$	$0.0942 \pm 0.0184$	$0.1687 \pm 0.0412$	0.0079	0.0083	0.0058	0.0687
0.2	$0.2145 \pm 0.0102$	$0.1862 \pm 0.0093$	$0.1899 \pm 0.0177$	$0.1131 \pm 0.0464$	0.0145	0.0138	0.0111	0.0869
0.3	$0.3176 \pm 0.0138$	$0.2812 \pm 0.0130$	$0.2863 \pm 0.0167$	$0.1644 \pm 0.0538$	0.0176	0.0188	0.0137	0.1356
0.4	$0.4166 \pm 0.0210$	$0.3764 \pm 0.0165$	$0.3835 \pm 0.0153$	$0.2787 \pm 0.0500$	0.0166	0.0236	0.0165	0.1213
0.5	$0.5135 \pm 0.0196$	$0.4720 \pm 0.0197$	$0.4818 \pm 0.0136$	$0.3819 \pm 0.0475$	0.0135	0.0280	0.0182	0.1181
0.6	$0.6070 \pm 0.0347$	$0.5688 \pm 0.0226$	$0.5815 \pm 0.0117$	$0.4788 \pm 0.0458$	0.0070	0.0312	0.0185	0.1212
0.7	$0.7011 \pm 0.0232$	$0.6670 \pm 0.0242$	$0.6830 \pm 0.0093$	$0.5728 \pm 0.0483$	0.0011	0.0330	0.0170	0.1272
0.8	$0.7936 \pm 0.0239$	$0.7676 \pm 0.0249$	$0.7864 \pm 0.0066$	$0.6652 \pm 0.0490$	0.0064	0.0324	0.0136	0.1348
0.9	$0.8864 \pm 0.0252$	$0.8713 \pm 0.0249$	$0.8919 \pm 0.0034$	$0.7595 \pm 0.0501$	0.0136	0.0287	0.0081	0.1405
1.0	$1.0000 \pm 0.0000$	$1.0000 \pm 0.0000$	$1.0000 \pm 0.0000$	$0.9601 \pm 0.0185$	0.0000	0.0000	0.0000	0.0399
MAE					0.0093	0.0201	0.0125	0.1163

inducing a known dependency structure in samples taken from the input independent marginal distributions using reordering techniques. The multivariate coupled structure obtained as the output can thus be used as the input data in various models of dependency analysis to test their relative accuracies.

We set parameters of the Pearson type IV distribution such that the coupled synthetic data we obtain using the Iman-Conover method has similar properties to financial returns. But first we need to consider some properties of financial returns which we want to mimic. Figures 1(a) and 1(b) show the distributions of two higher-order moments, that is, skewness ( $\gamma$ ) and kurtosis ( $\kappa$ ), for FX spot data sets sampled at three different frequencies. The plots are obtained using a sliding-window of length 50 data points, as an average of all G10 (Group of Ten) currency pairs, covering a period of 8 hours in the case of 0.25 second and 0.5 second sampled data and 2 years in case of 0.5 hour sampled data. The non-Gaussian (heavy-tailed, skewed) nature of the data is clearly visible. It is interesting to note that the kurtosis value almost never goes below three for any of the data sets, signifying the temporal persistence of non-Gaussianity. We now make use of the Iman-Conover method to induce varying levels of correlation between 1000 samples taken from independent, randomly distributed, Pearson type IV distributions. A 1000 sample data set makes it easier to accurately induce predefined correlations in the system as well as makes it possible to generate data with relatively accurate average kurtosis and skewness values, hence, allowing us to accurately capture the higher-order moments of financial returns. As an example, Figures 1(c) and 1(d) show distributions of the kurtosis and skewness of two variables for 1000 independent simulations. Note the similarity of these plots with the average of the corresponding distributions of higher-order moments for financial data, as presented in Figures 1(a) and 1(b). This shows the effectiveness of using synthetic data sampled from

a Pearson type IV distribution for capturing higher-order moments of financial returns.

Four different approaches are now used to estimate the level of dependence between the output coupled data. The process is repeated 1000 times for each level of *true correlation* ( $\rho_{TC}$ ). Table 2 presents the results obtained. The results show the accuracy of the information coupling measure when used to analyse non-Gaussian data. For this synthetic data example, on average, the information coupling measure was 53.7% more accurate than the linear correlation measure and 25.6% more accurate with respect to the rank correlation measure. The normalised mutual information provided the least accurate results.

We now extend this example by incorporating data dynamics. The same data generation process, as described above, is now used to construct a 32000 samples long bivariate data set in which the induced true correlation changes every 8000 time steps; that is,  $\rho_{TC} = 0.2$  when  $t = 1 : 8000$ ,  $\rho_{TC} = 0.4$  when  $t = 8001 : 16000$ ,  $\rho_{TC} = 0.6$  when  $t = 16001 : 24000$ , and  $\rho_{TC} = 0.8$  when  $t = 24001 : 32000$ . A 1000 data points wide sliding-window is used to dynamically measure dependencies in the data set. The resulting temporal information coupling plot, together with the 5th and 95th percentile confidence intervals, is presented in Figure 2(a). The four different coupling regions are clearly visible, together with the step changes in coupling after every 8000 time steps, showing ability of the algorithm to detect abrupt changes in coupling. The normalised empirical probability distributions over  $\eta$  for the four coupling regions are shown in Figure 2(b). Also plotted in the same figure are the normalised empirical pdfs for linear correlation ( $\rho$ ) and rank correlation ( $\rho_R$ ). The mutual information pdf is omitted for clarity as it gives relatively less accurate results, as presented in Table 2. It is interesting to see how the peaks of the  $\eta$  distribution correspond very closely to  $\rho_{TC}$  values, showing ability of the information coupling model to accurately capture statistical



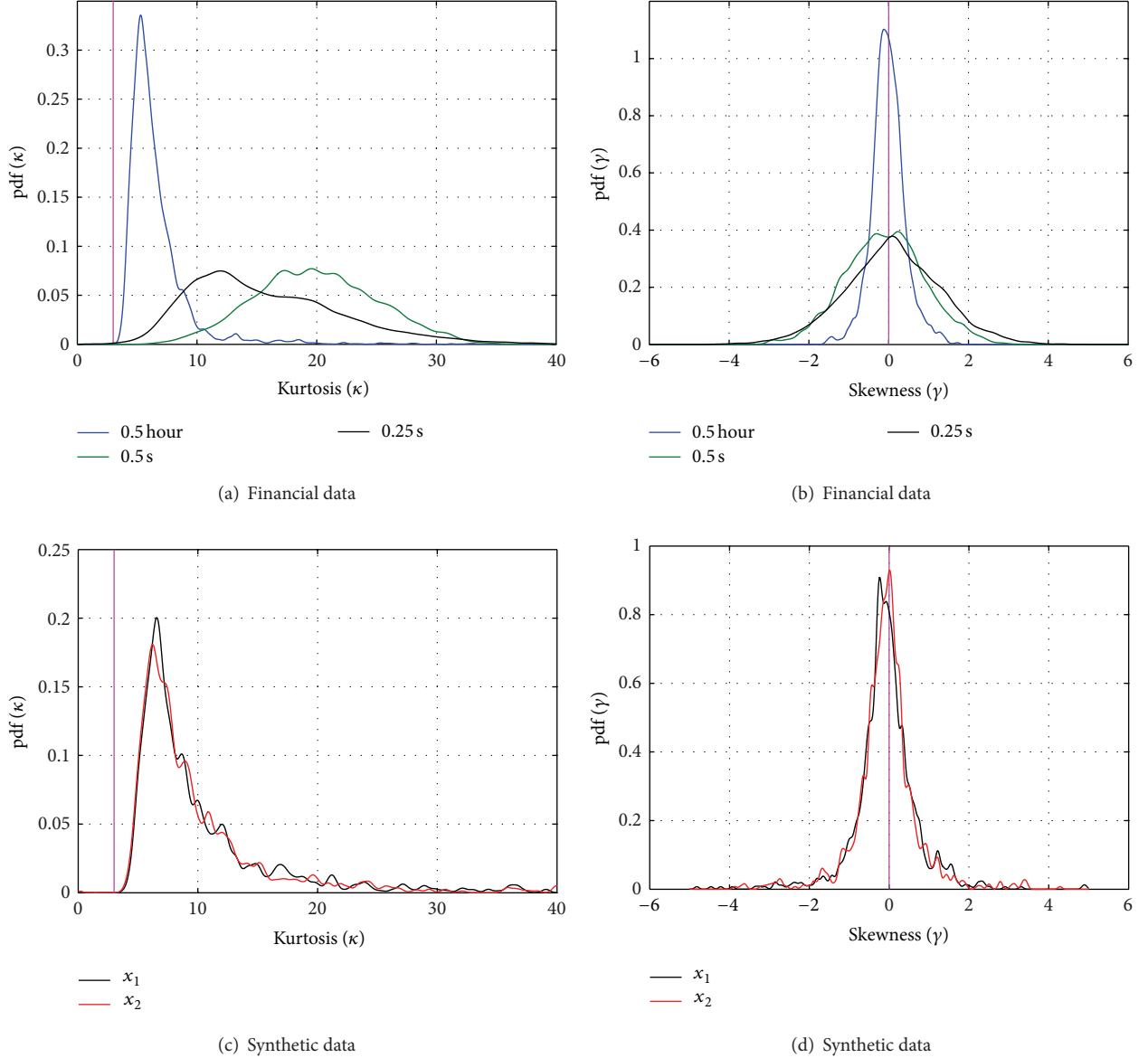


FIGURE 1: (a, b) Plots showing the normalised pdfs of (a) kurtosis ( $\kappa$ ) and (b) skewness ( $\gamma$ ), obtained using a sliding-window of length 50 points, as an average of all G10 currency pairs, covering a period of 8 hours in case of 0.25 second and 0.5 second sampled data and 2 years in case of 0.5 hour sampled data. The plots clearly show the heavy-tailed, skewed, nature of FX returns at all three sampling frequencies. (c, d) An example of the normalised pdf plots showing the average kurtosis and skewness values, respectively, for synthetic data generated using Pearson type IV distributions. The data has properties similar to the average of the distributions presented in (a) and (b). The vertical lines indicate the kurtosis ( $\kappa = 3$ ) and skewness ( $\gamma = 0$ ) values for a Gaussian distribution.

dependencies in a dynamic environment. The least accurate measure in this example is the linear correlation.

**5.2. Financial Data.** We first present a simple application of the information coupling algorithm to a section of 0.5 second sampled FX spot log-returns data set (in this paper, all currencies are referred by their standardised international three-letter codes, as described by the ISO 4217 standard. For the currencies mentioned in this paper, the three-letter codes are USD (U.S. dollar), EUR (Euro), JPY (Japanese yen), GBP

(British pound), CHF (Swiss franc), and AUD (Australian dollar)). Figure 3 shows the variation of information coupling and linear correlation with time for EURUSD and GBPUSD. The results are obtained using a 5-minute wide sliding-window. We note that the two measures of dependence frequently give different results, which reflects on the inability of linear correlation to capture dependencies in non-Gaussian data streams. We also note that dependencies in FX log-returns exhibit rapidly changing dynamics, often characterised by regions of quasi-stability punctuated by abrupt

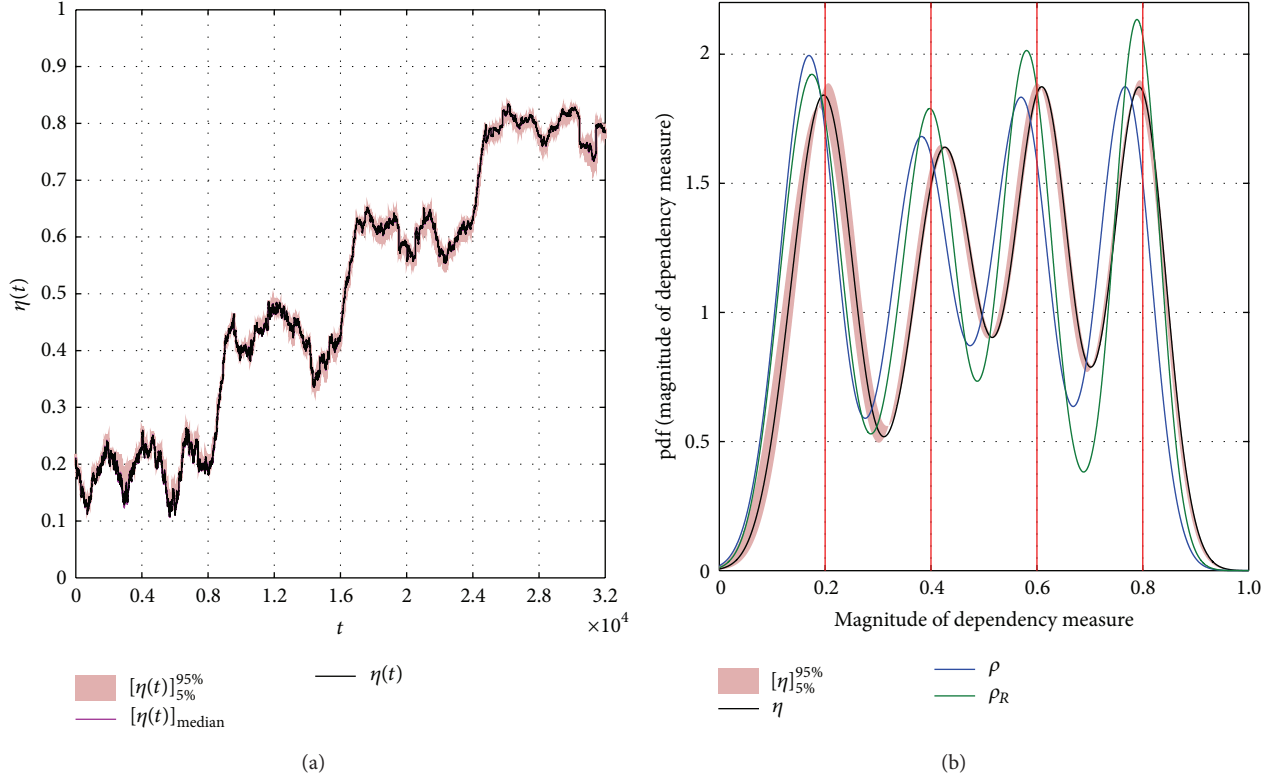


FIGURE 2: (a) Temporal variation of information coupling,  $\eta(t)$ , plotted as a function of time. Step changes in coupling are visible after every 8000 time steps; that is,  $\rho_{TC} = 0.2$  when  $t = 1:8000$ ,  $\rho_{TC} = 0.4$  when  $t = 8001:16000$ ,  $\rho_{TC} = 0.6$  when  $t = 16001:24000$ , and  $\rho_{TC} = 0.8$  when  $t = 24001:32000$ . Also plotted are the median,  $[\eta(t)]_{\text{median}}$ , and the 5% and 95% confidence interval contours,  $[\eta(t)]_{5\%}^{95\%}$ . (b) Normalised empirical pdf plots for information coupling ( $\eta$ ), linear correlation ( $\rho$ ), and rank correlation ( $\rho_R$ ). The vertical lines represent the true correlation ( $\rho_{TC}$ ) values. The relative accuracy of the information coupling measure is evident from these results.

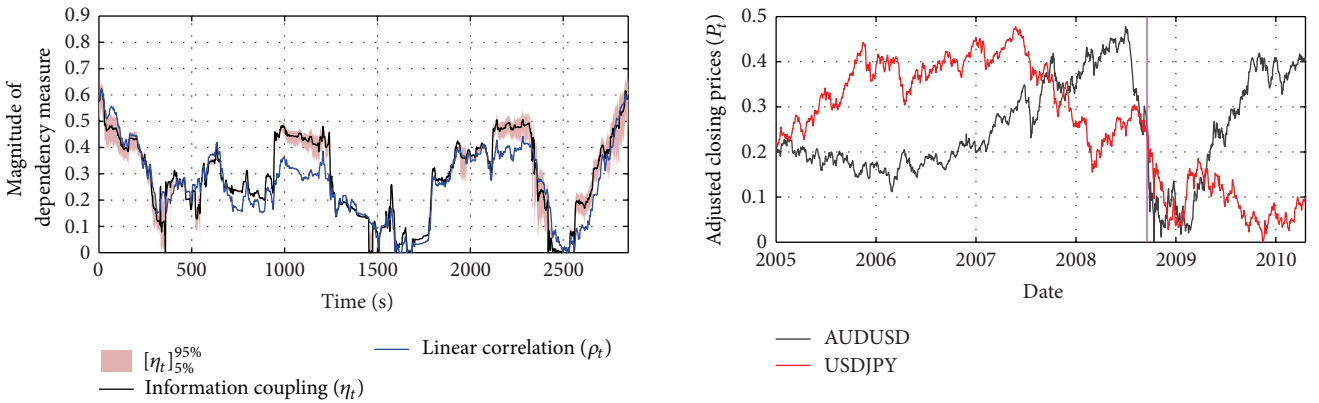


FIGURE 3: Information coupling ( $\eta_t$ ) and linear correlation ( $\rho_t$ ) plotted as a function of time for a section of 0.5 second sampled EURUSD and GBPUSD log-returns data set. A 5-minute wide sliding-window is used to obtain the results.

FIGURE 4: Daily closing mid-prices ( $P_t$ ) of AUDUSD and USDJPY. The two plots are adjusted for ease of comparison. The vertical line corresponds to the September 2008 financial meltdown.

changes. In [48] we show that these regions of persistence in statistical dependence may be captured using a hidden Markov ICA model, which is a hidden Markov model with an ICA observation model. The information thus obtained can be useful for a range of financial applications, such as

finding regions of financial market volatility clustering and persistence [49].

There have been numerous academic studies on the causes and effects of the 2008 financial crisis [50, 51]. However, very few of these have focused on the impact of the crisis on interdependencies in the global FX market. Here we present a set of examples which give a unique insight into the

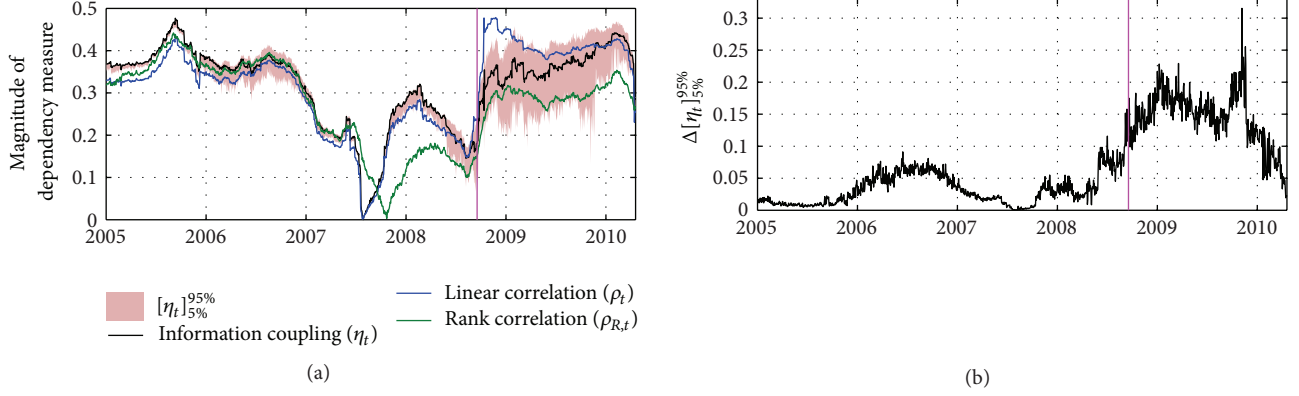


FIGURE 5: (a) Three approaches used to measure temporal dependencies between AUDUSD and USDJPY log-returns. (b) Range of confidence intervals ( $\Delta[\eta]_{5\%}^{95\%} = \eta_{95\%} - \eta_{5\%}$ ) plotted as a function of time, showing the uncertainty associated with the information coupling measurements. The vertical lines correspond to the September 2008 financial meltdown.

effects of the crisis on the nature of statistical dependencies in the spot FX market. Accurate estimation of dependencies at times of financial crises is of utmost importance, as these estimates are used by financial practitioners for a range of tasks, such as rebalancing portfolios, accurately pricing options, and deciding on the level of risk-taking. We first present an application of the information coupling model for detecting temporal changes in dependencies in bivariate FX data streams at times of financial crises. Figure 4 shows the adjusted daily closing mid-prices ( $P_t$ ) for AUDUSD and USDJPY from January 2005 till April 2010. The two plots clearly show an abrupt change in the exchange rates in September-October 2008. This was caused at the height of the 2008 global financial crisis due to the unwinding of carry trades [52]. Figure 5(a) displays three plots showing the temporal variation of information coupling ( $\eta_t$ ), linear correlation ( $\rho_t$ ), and rank correlation ( $\rho_{R,t}$ ) between AUDUSD and USDJPY log-returns. The plots are obtained using a six-month long sliding-window. We notice the rise in uncertainty of the information coupling measure (Figure 5(b)) right before the crash, with uncertainty decreasing gradually thereafter; this information may be useful to systematically predict upheavals in the market, although we do not carry out this study in detail in this paper. Information about the level of uncertainty can be used as a measure of confidence in the information coupling values and can be useful in various practical decision making scenarios, such as deciding on the capital to deploy for the purpose of trading or selecting stocks (or currencies) for inclusion in a portfolio. As daily sampled data is generally less non-Gaussian than data sampled at higher frequencies, therefore, the three plots in Figure 5(a) are somewhat similar during certain time periods. However, right after the September 2008 crash, the plots significantly deviate from each other. We believe that this is because the nature of the data, in particular its level of non-Gaussianity, has changed. As shown in Figure 6, the distance measure,  $(\eta_t - |\rho_t|)^2$ , between information coupling and linear correlation closely matches the non-Gaussianity of the data under consideration. The two plots are adjusted for ease of comparison. The degree of non-Gaussianity is calculated

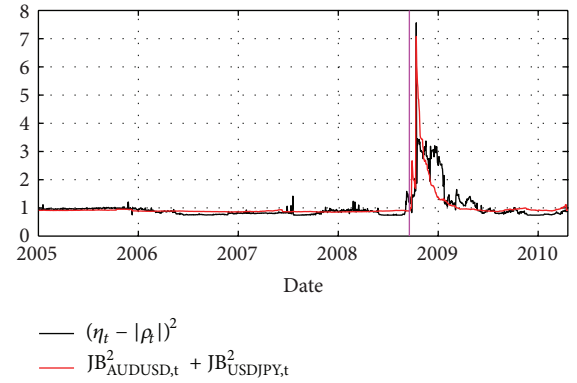


FIGURE 6: Difference between information coupling and linear correlation plotted as a function of time. Also plotted is a measure of non-Gaussianity of the two time series as defined by (40). The two plots are adjusted for ease of comparison. The vertical line corresponds to the September 2008 financial meltdown.

using the multivariate Jarque-Bera statistic ( $JB_{MV}$ ) which we define for a  $N$ -dimensional multivariate data set as

$$JB_{MV} = \sum_{j=1}^N \left[ \frac{n_s}{6} \left( \gamma_j^2 + \frac{(\kappa_j - 3)^2}{4} \right) \right]^2, \quad (40)$$

where  $n_s$  is the number of data points (in this case the size of the sliding-window),  $\gamma$  is the skewness of the data under analysis, and  $\kappa$  is its kurtosis. This shows that relying solely on correlation measures to model dependencies in multivariate financial time series, even when using data sampled at relatively lower frequencies, can potentially lead to inaccurate results. In contrast, the information coupling model takes into account properties of the data being analysed, resulting in an accurate approach to measure statistical dependencies.

We now show utility of the information coupling model for analysing multivariate statistical dependencies. Figure 7 shows the temporal variation of information coupling between four major liquid currency pairs (EURUSD,

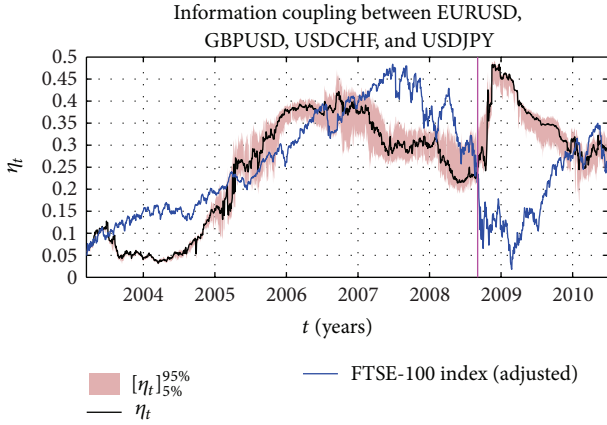


FIGURE 7: Information coupling between EURUSD, GBPUSD, USDCHE, and USDJPY log-returns plotted as a function of time. Also plotted is the FTSE-100 (Financial Times Stock Exchange 100) index for the corresponding time period, which has been adjusted for ease of comparison. The plot clearly shows an abrupt upward shift in coupling between the four currency pairs right at the time of the September 2008 financial meltdown, with gradual decrease in coupling over the next year. We also notice an increase in the uncertainty associated with the information coupling measure before the 2008 crash. The increase in dependence of financial instruments in times of financial crises has been observed for other asset classes as well [53]. Our unique example, showing the dynamics of multivariate dependencies within the spot FX space, provides further insight into the nature of interdependencies in times of financial crisis.

GBPUSD, USDCHE, and USDJPY). The results are obtained using daily log-returns for a seven-year period and a six-month long sliding-window. Also plotted on the same figure is the FTSE-100 (Financial Times Stock Exchange 100) index for the corresponding time period, which has been adjusted for ease of comparison. The plot clearly shows an abrupt upward shift in coupling between the four currency pairs right at the time of the September 2008 financial meltdown, with gradual decrease in coupling over the next year. We also notice an increase in the uncertainty associated with the information coupling measure before the 2008 crash. The increase in dependence of financial instruments in times of financial crises has been observed for other asset classes as well [53]. Our unique example, showing the dynamics of multivariate dependencies within the spot FX space, provides further insight into the nature of interdependencies in times of financial crisis.

## 6. Conclusions

We present an ICA-based approach to dynamically measure information coupling, as a proxy for mutual information. This approach makes use of ICA as a tool to capture information in the tails of the underlying distributions and is suitable for efficiently and accurately measuring statistical dependencies between multiple non-Gaussian signals. As far as we know, this is the first attempt to quantify multivariate dependencies using information encoded in the ICA unmixing matrix. Our proposed information coupling model has multiple other benefits associated with its practical use. It provides a framework for estimating confidence bounds on the information coupling metric, can be efficiently used to directly model dependencies in high-dimensional spaces, and gives normalised, symmetric results. It has the added advantage of not depending on any user-defined parameters and is not data intensive; that is, it can be used even with relatively small data sets without significantly affecting its

performance, an important requirement for analysing data streams with rapidly changing dynamics such as financial returns. The model makes use of a sliding-window based decorrelating manifold approach to ICA, with a reciprocal cosh source model, to infer the ICA unmixing matrix, which results in increased accuracy and efficiency of the algorithm. This gives the information coupling model the computational complexity similar to that of linear correlation with the accuracy of mutual information.

## Acknowledgments

The authors are grateful to the Oxford-Man Institute of Quantitative Finance for help in supporting this research. The first author would also like to thank Exeter College (University of Oxford) for funding.

## References

- [1] D. Berg and K. Aas, "Models for construction of multivariate dependence—a comparison study," *The European Journal of Finance*, vol. 15, no. 7-8, pp. 639–659, 2009.
- [2] M. M. Dacorogna, *An Introduction to High-Frequency Finance*, Academic Press, New York, NY, USA, 2001.
- [3] J. Hlinka, M. Paluš, M. Vejmelka, D. Mantini, and M. Corbetta, "Functional connectivity in resting-state fMRI: is linear correlation sufficient?" *NeuroImage*, vol. 54, no. 3, pp. 2218–2225, 2011.
- [4] S. J. Devlin, R. Gnanadesikan, and J. R. Kettenring, "Robust estimation and outlier detection with correlation coefficients," *Biometrika*, vol. 62, no. 3, pp. 531–545, 1975.
- [5] A. R. Cowan, "Nonparametric event study tests," *Review of Quantitative Finance and Accounting*, vol. 2, no. 4, pp. 343–358, 1992.
- [6] G. J. Glasser and R. F. Winter, "Critical values of the coefficient of rank correlation for testing the hypothesis of independence," *Biometrika*, vol. 48, no. 3-4, pp. 444–448, 1961.
- [7] P. Embrechts, F. Lindskog, and A. McNeil, "Modelling dependence with copulas and applications to risk management," in *Handbook of Heavy Tailed Distributions in Finance*, vol. 1, chapter 8, pp. 329–384, 2003.
- [8] E. Jondeau, S. H. Poon, and M. Rockinger, *Financial Modeling Under Non-Gaussian Distributions*, Springer, New York, NY, USA, 2007.
- [9] J. D. Fermanian and O. Scaillet, "Some statistical pitfalls in copula modeling for financial applications," in *Capital Formation, Governance and Banking*, p. 59, 2005.
- [10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, New York, NY, USA, 2006.
- [11] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical Review E*, vol. 69, no. 6, Article ID 066138, 16 pages, 2004.
- [12] S. Khan, S. Bandyopadhyay, and A. R. Ganguly, "Computing mutual information based nonlinear dependence among noisy and finite geophysical time series," in *Proceedings of the American Geophysical Union, Fall Meeting*, 2005, abstract no. NG22A-03.
- [13] M. M. V. Hulle, "Edgeworth approximation of multivariate differential entropy," *Neural Computation*, vol. 17, no. 9, pp. 1903–1910, 2005.



- [14] L. B. Almeida, "Linear and nonlinear ICA based on mutual information," in *Proceedings of the IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium (AS-SPCC '00)*, pp. 117–122, IEEE, 2000.
- [15] A. D. Back and A. S. Weigend, "A first application of independent component analysis to extracting structure from stock returns," *International Journal of Neural Systems*, vol. 8, no. 4, pp. 473–484, 1997.
- [16] E. Oja, K. Kiviluoto, and S. Malaroiu, "Independent component analysis for financial time series," in *Proceedings of the IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium (AS-SPCC '00)*, pp. 111–116, IEEE, 2000.
- [17] A. Hyvärinen, "Survey on independent component analysis," *Neural Computing Surveys*, vol. 2, no. 4, pp. 94–128, 1999.
- [18] C. J. Lu, T. S. Lee, and C. C. Chiu, "Financial time series forecasting using independent component analysis and support vector regression," *Decision Support Systems*, vol. 47, no. 2, pp. 115–125, 2009.
- [19] H. Joe, "Relative entropy measures of multivariate dependence," *Journal of the American Statistical Association*, vol. 84, no. 405, pp. 157–164, 1989.
- [20] M. Novey and T. Adali, "Complex ICA by negentropy maximization," *IEEE Transactions on Neural Networks*, vol. 19, no. 4, pp. 596–609, 2008.
- [21] S. Roberts and R. Everson, *Independent Component Analysis: Principles and Practice*, Cambridge University Press, New York, NY, USA, 2001.
- [22] J. Palmer, K. Kreutz-Delgado, and S. Makeig, "Super-Gaussian mixture source model for ICA," in *Independent Component Analysis and Blind Signal Separation*, pp. 854–861, 2006.
- [23] R. A. Choudrey and S. J. Roberts, "Variational mixture of Bayesian independent component analyzers," *Neural Computation*, vol. 15, no. 1, pp. 213–252, 2003.
- [24] R. Everson and S. Roberts, "Independent component analysis: a flexible nonlinearity and decorrelating manifold approach," *Neural Computation*, vol. 11, no. 8, pp. 1957–1983, 1999.
- [25] R. E. Kass, L. Tierney, and J. B. Kadane, "Laplace's method in Bayesian analysis," in *Statistical Multiple Integration: Proceedings of the AMS-IMS-SIAM Joint Summer Research Conference [on Statistical Multiple Integration] Held at Humboldt University, June 17–23, 1989*, vol. 115, p. 89, American Mathematical Society, 1991.
- [26] W. Addison and S. Roberts, "Blind source separation with non-stationary mixing using wavelets," in *Proceedings of the ICA Research Network Workshop*, The University of Liverpool, 2006.
- [27] N. J. Higham, "Matrix nearness problems and applications," in *Applications of Matrix Theory*, 1989.
- [28] K. B. Datta, *Matrix and Linear Algebra*, PHI Learning, 2004.
- [29] K. Boudt, J. Cornelissen, and C. Croux, "The Gaussian rank correlation estimator: robustness properties," *Statistics and Computing*, vol. 22, no. 2, pp. 471–483, 2012.
- [30] D. Evans, "A computationally efficient estimator for mutual information," *Proceedings of the Royal Society A*, vol. 464, no. 2093, pp. 1203–1215, 2008.
- [31] K. Torkkola, "On feature extraction by mutual information maximization," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '02)*, vol. 1, pp. I/821–I/824, IEEE, May 2002.
- [32] K. Nagarajan, B. Holland, C. Slatton, and A. D. George, "Scalable and portable architecture for probability density function estimation on FPGAs," in *Proceedings of the 16th IEEE Symposium on Field-Programmable Custom Computing Machines (FCCM '08)*, pp. 302–303, IEEE Computer Society, April 2008.
- [33] C. G. Bowsher, "Modelling security market events in continuous time: intensity based, multivariate point process models," *Journal of Econometrics*, vol. 141, no. 2, pp. 876–912, 2007.
- [34] B. Peiers, "Informed traders, intervention, and price leadership: a deeper view of the microstructure of the foreign exchange market," *Journal of Finance*, vol. 52, no. 4, pp. 1589–1614, 1997.
- [35] A. Dionisio, R. Menezes, and D. A. Mendes, "Mutual information: a measure of dependency for nonlinear time series," *Physica A*, vol. 344, no. 1–2, pp. 326–329, 2004.
- [36] J. Y. Campbell, A. W. Lo, A. C. MacKinlay, and R. F. Whitelaw, "The econometrics of financial markets," *Macroeconomic Dynamics*, vol. 2, no. 4, pp. 559–562, 1998.
- [37] G. Koutmos, C. Negakis, and P. Theodossiou, "Stochastic behaviour of the Athens stock exchange," *Applied Financial Economics*, vol. 3, no. 2, pp. 119–126, 1993.
- [38] D. M. Guillaume, M. M. Dacorogna, R. R. Davé, U. A. Müller, R. B. Olsen, and O. V. Pictet, "From the bird's eye to the microscope: a survey of new stylized facts of the intra-daily foreign exchange markets," *Finance and Stochastics*, vol. 1, no. 2, pp. 95–129, 1997.
- [39] R. Cheng, "Using pearson type IV and other Cinderella distributions in simulation," in *Proceedings of the Winter Simulation Conference (WSC '11)*, pp. 457–468, IEEE, 2011.
- [40] Y. Nagahara, "The PDF and CF of Pearson type IV distributions and the ML estimation of the parameters," *Statistics and Probability Letters*, vol. 43, no. 3, pp. 251–264, 1999.
- [41] A. Shephard, *Pearson IV*, Institute for Fiscal Studies, University College London, 2008.
- [42] S. Stavroyiannis, I. Makris, V. Nikolaidis, and L. Zarangas, "Econometric modeling and value-at-risk using the Pearson type IV distribution," *International Review of Financial Analysis*, vol. 22, pp. 10–17, 2012.
- [43] M. Kendall, A. Stuart, and J. K. Ord, *Kendall's Advanced Theory of Statistics*, Charles Griffin, 1987.
- [44] R. Willink, "A closed-form expression for the pearson type IV distribution function," *Australian and New Zealand Journal of Statistics*, vol. 50, no. 2, pp. 199–205, 2008.
- [45] A. Venelli, "Efficient entropy estimation for mutual information analysis using B-splines," in *Information Security Theory and Practices. Security and Privacy of Pervasive Systems and Smart Devices*, pp. 17–30, 2010.
- [46] C. G. Park and D. W. Shin, "An algorithm for generating correlated random variables in a class of infinitely divisible distributions," *Journal of Statistical Computation and Simulation*, vol. 61, no. 1–2, pp. 127–139, 1998.
- [47] R. L. Iman and W. J. Conover, "A distribution-free approach to inducing rank correlation among input variables," *Communications in Statistics-Simulation and Computation*, vol. 11, no. 3, pp. 311–334, 1982.
- [48] N. Shah and S. Roberts, "Hidden Markov independent component analysis as a measure of coupling in multivariate financial time series," in *Proceedings of the ICA Research Network International Workshop*, Liverpool, UK, 2008.
- [49] A. Rossi and G. M. Gallo, "Volatility estimation via hidden Markov models," *Journal of Empirical Finance*, vol. 13, no. 2, pp. 203–230, 2006.
- [50] J. Crotty, "Structural causes of the global financial crisis: a critical assessment of the 'new financial architecture'," *Cambridge Journal of Economics*, vol. 33, no. 4, pp. 563–580, 2009.

- [51] J. A. Murphy, "An analysis of the financial crisis of 2008: causes and solutions," *Social Science Research Network*, 2008.
- [52] M. Pojarliev and R. M. Levich, "Detecting crowded trades in currency funds," *Financial Analysts Journal*, vol. 67, no. 1, pp. 26–39, 2011.
- [53] L. Sandoval and I. D. P. Franca, "Correlation of financial markets in times of crisis," *Physica A*, vol. 391, no. 1, pp. 187–208, 2012.

