KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE

OF

ECONOMIC RESEARCH

Discussion Paper No.827

"A non-parametric and entropy based analysis
of the relationship between the VIX and S&P500"

D.E. Allen, A. Kramadibrata, R. Powell,
M. McAleer and A.K. Singh

August 2012

KYOTO UNIVERSITY

KYOTO, JAPAN

A non-parametric and entropy based analysis of the relationship between the VIX and S&P500

D.E. Allen[a,], A. Kramadibrata[a,], M. McAleer[b], R. Powell[a], A. K. Singh[a]

[a]*School of Accounting, Finance and Economics, Edith Cowan University*
[b]*Erasmus University Rotterdam*

Abstract

This paper features an analysis of the relationship between the S&P500 Index and the VIX using daily data obtained from both the CBOE website and SIRCA (The Securities Industry Research Centre of the Asia Pacific). We explore the relationship between the S&P500 daily continuously compounded return series and a similar series for the VIX in terms of a long sample drawn from the CBOE running from 1990 to mid 2011 and a set of returns from SIRCA's TRTH datasets running from March 2005 to-date. We divide this shorter sample, which captures the behaviour of the new VIX, introduced in 2003, into four roughly equivalent sub-samples which permit the exploration of the impact of the Global Financial Crisis. We apply to our data sets a series of non-parametric based tests utilising entropy based metrics. These suggest that the PDFs and CDFs of these two return distributions change shape in various subsample periods. The entropy and MI statistics suggest that the degree of uncertainty attached to these distributions changes through time and using the S&P500 return as the dependent variable, that the amount of information obtained from the VIX also changes with time and reaches a relative maximum in the most recent period from 2011 to 2012. The entropy based non-parametric tests of the equivalence of the two distributions and their symmetry all strongly reject their respective nulls. The results suggest that parametric techniques do not adequately capture the complexities displayed in the behaviour of these series. This has practical implications for hedging utilising derivatives written on the VIX, which will be the focus of a subsequent study.

Keywords: S&P500, VIX, Entropy, Non-Parametric Estimation, Quantile Regressions,

1. Introduction

In this paper we analyse the relationship between the S&P500 Index and the VIX. The Standard and Poor's website suggests that: "the S&P 500 has been widely regarded as the best single gauge of the large cap U.S. equities market since the index was first published in 1957. The index has over US$ 5.58 trillion benchmarked, with index assets comprising approximately US$ 1.31 trillion of this total. The index includes 500 leading companies in leading industries of the U.S. economy, capturing 75% coverage of U.S. equities". On any given day, the index value is the quotient of the total float-adjusted market capitalization of the index's constituents and its divisor. Continuity in index values is maintained by adjusting the divisor for all changes in the constituents' share capital after the base date. Clearly, this major index is a magnet for active and passive fund management and is a bell-weather for global investors. It has also spawned a vast array of derivatives and related hedging instruments. One such major instrument is the CBOE VIX which was introduced in 1993 as the CBOE Volatility Index, VIX, which was originally designed to measure the market's expectation of 30- day volatility implied by at-the-money S&P 100 Index (OEX) option prices.

Ten years later on September 22nd 2003, CBOE together with Goldman Sachs (See Demeterfi et al (1999)), updated the VIX to reflect a new way to measure expected volatility, one that continues

---

Email address: `d.allen@ecu.edu.au` (D.E. Allen)

to be widely used by financial theorists, risk managers and volatility traders alike. The new VIX is based on the S&P 500 Index (SPX SM), the core index for U.S. equities, and estimates expected volatility by averaging the weighted prices of SPX puts and calls over a wide range of strike prices. Given that it is an average of a strip of prices it is not model-based. Speculators can use VIX derivatives to trade on volatility risk whilst hedgers can use them to avoid exposure to volatility risk. However, as Demeterfi et al (1999) point out, for this strip of prices to succesfully capture expectations of realised variance, large jumps in prices have to be ruled out: "in practice, both the effects of jumps and the risks of log replication with only a limited strike range cause the strategy to capture a quantity different from the true realized variance of the stock price".

The statistical properties and the nature of the probability density functions capturing the behaviour of the S&P500 and the VIX are important issues for investors and hedgers alike. However, the nature of their two density functions has not been closely explored and this issue is the focus of this paper. We bring to bear non-parametric estimation procedures and entropy based measures to capture the nature of the individual probability density functions for the S&P500 and the VIX and their joint density functions, with particular attention paid to the tails of their respective distributions, to shed greater light on their investment and hedging capabilities, particularly in extreme market circumstances, as represented by the recent Global Financial Crisis (GFC).

The VIX itself is not a cash instrument and trades have to be done on derivatives written against it; Chang et al (2011) discuss some of these issues. The Chicago Board Options Exchange (2003) defines VIX as being a measure of the expected volatility of the S&P500 over the next 30-days. It follows that VIX futures prices should reflect the current expectation of what the expected 30-day volatility will be at a particular time in the future (on the expiration date). VIX futures of necessity converge to the spot price at expiration, yet it is possible to have significant disparities between the spot VIX and VIX futures prior to expiration. Speculators can trade on volatility risk with VIX derivatives, and adopt positions according to their expectations of whether volatility will increase or decrease in the future, while hedgers can hedge exposure to volatility risk using volatility derivatives. VIX futures of different maturities can be used to hedge VIX futures, and VIX options can also be hedged using VIX futures (see, for example, Sepp (2008)). Optimal hedge ratios can be calculated using consistently estimated dynamic conditional correlations (see, for example, Caporin and McAleer (2011)).

There has been considerable prior work featuring an analysis of the VIX and its associated derivatives and futures contracts. An approximate analytical VIX futures pricing formula is derived by Brenner et al. (2006) . The skewness in the implied volatilities of VIX options is examined by Sepp (2008). Huskaj (2009) calculates the VaR of VIX futures, whilst McAleer and Wiphatthanananthakul (2010) contrast the VIX with an exploration of the empirics of alternative simple expected volatility indexes. Chang et al. (2011) analyse the VaR of VIX futures under the Basel Accord, whilst Ishida et al. (2011) propose a new method for estimating continuous-time stochastic volatility (SV) models for the S&P500 stock index process.

2. Research methods and data

We commence by taking a broad view of the dataset before concentrating on the behaviour of the indices since September 2003, when the new version of the VIX was introduced. Our initial summary analyses utilise data taken from the CBOE website (http://www.cboe.com/micro/VIX/vixintro.aspx) where a daily price history of the VIX is available running back until 1990. Given that the VIX was introduced in 1993 the first segment consists of constructed prices prior to the actual publication of the VIX. Table 1 presents a summary of the characteristics of the S&P500 and the VIX plus their daily compounded logarithmic returns. A set of graphs depicts their behaviour.

Table 1: Summary statistics, S&P 500, VIX, Daily S&P500 returns and daily VIX returns, Jan 1990 June 2011

	S&P500	S&P500 daily returns	VIX	VIX daily returns
Min	295.5	-0.0947	9.31	-0.3506
Median	1047	0.0005	18.88	-0.0031
Mean	939	0.00024	20.34	0.00043
Maximum	1560	0.1096	80.86	0.496
Variance		0.0001356015		0.0035794839

Figure 1 displays the time-series behaviour of the base series and the second figure graphs the logarithmic return series for the S&P500 and the VIX respectively for the period 2000 to June 2011.

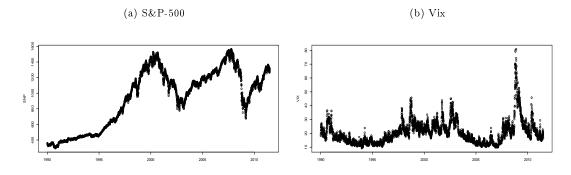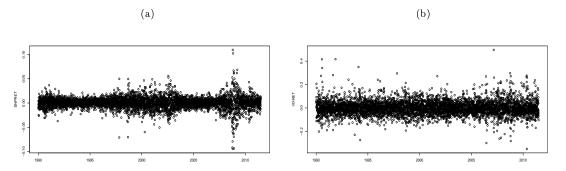Figure 1: Time series behaviour of the S&P500 and VIX series from Jan 1990 to June 2011

(a) S&P-500

(b) Vix

Figure 2: Time series behaviour of the S&P500 and VIX logarithmic return series from Jan 1990 to June 2011
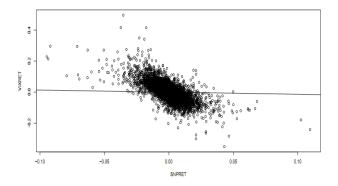
(a)

(b)

Edrington (1979) suggested that one way of assessing the hedging performance of one financial instrument in terms of another would be to run an Ordinary Least Squares (OLS) regression. In his case, as he was examining the relationship between spot and futures markets, he suggested the change in spot prices be regressed on the change in futures prices as a way of calculating the appropriate hedge ratio. The population coefficient of determination between the change in the cash price and the change in the future's price provides the optimum hedge ratio, and clearly the less noise in the relationship, or the higher the R^2 the more effective the hedge is likely to be. Now, Edrington (1979) predates the development of the theory of cointegration by Engle and Granger (1987). (See the discussion in Alexander (1999)). Ideally, the error-correction mechanism should be taken into account. However, for the current illustrative purposes, which are not the main thrust

of the paper, but merely a component of a background introduction, we shall introduce the results of an OLS regression of the daily continuously compounded returns on the S&P500 Index regressed on the daily continuously compounded return on the VIX, for the extended period above runnning from 1990 to June 2011. Clearly, hedge ratios are likely to change in market states, and this simplistic regression is a very blunt instrument but it will permit the development of an argument concerning the hedging and diversification properties of the two instruments. Table 2 above records the results of this simple regression, whilst figure 3 displays the fitted relationship.

Table 2: Ols regression of daily continuously compounded S&P500 returns on daily continuously compounded VIX returns

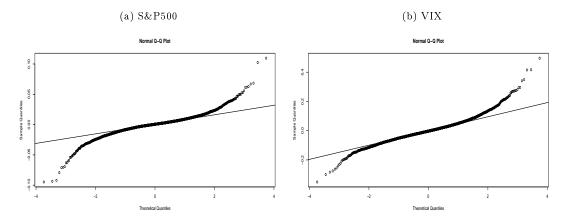|	Intercept	slope	
coefficient	0.000238	-0.13527	
Standard Error	0.0001138	0.001902	
t value	2.094*	-71.006***	
Adjusted R^2			0.4829
F Value			5056***
Significance: *** (1%), *(5%)			

Figure 3: Ols regression of daily continuously compounded S&P500 returns on daily continuously compounded VIX returns with fitted line

The regression results demonstrates a well-known fact, a significant negative relationship between S&P500 returns and VIX returns and therefore opportunities for hedging market portfolio returns. Whaley (2009) comments that: "the VIX was intended to provide an index upon which futures and options contracts on volatility could be written. The social benefits of trading volatility have long been recognized. The Chicago Board Options Exchange (CBOE) launched trading of VIX futures contracts in May 2004 and VIX option contracts in February 2006.

In 2008 the US stock market declined by more than a third from its previous level; a loss greater than any other since the era of the Great Depression. Global markets followed suit, fixed income offered no protection with a 5% gain yet volatility gained 81% in 2008. Portfolio diversification via VIX futures or options would have cushioned the impact the impact of the downturn. This brings us back to the issue of the appropriate hedge ratio and the previous regression analysis is based on assumptions of multivariate normality. Figure 4 displays QQplots of the returns on the S&P500 and the VIX against a theoretical normal distribution. The points should plot on the line along the quantiles if they follow a normal distribution. Clearly the tails of both distributions depart from normality.

Figure 4: QQ normal plots of the S&P500 returns and the VIX returns

(a) S&P500   (b) VIX

The departure from normality can be tested using a Cramer-von Mises test statistic which is an omnibus test for the composite hypothesis of normality (see Thode (2002)). The test statistic is shown in equation (1)

$$W = \frac{1}{12n} + \sum_{i=1}^{n}(p_{(i)} - \frac{2i-1}{2n}) \tag{1}$$

where $p_{(i)} = \Phi([x_i - \bar{x}]/s$. Here Φ is the cumulative distribution function of the standard normal distribution, and $\bar{x}$ and s are the mean and standard deviation of the data values. The results of tests of the normality of the two series are shown in Table 3.

Table 3: Cramer-von Mises tests of the normality of the return series for the S&P500 and the VIX for the period 1990 to June 2011

	S&P500 returns	VIX returns
Cramer-von Mise statistic	15.0744	5.6701
Probability value	0.00000	0.00000

Both show highly significant departures from normality. Hence-forward we will take this into account and apply non-parametric tests to our data sets.

2.1. Entropy based measures

One attractive set of measures that are distribution free are based on entropy. The concept of entropy originates from physics in the 19th century; the second law of thermodynamics stating that the entropy of a system cannot decrease other way than by increasing the entropy of another system. As a consequence, the entropy of a system in isolation can only increase or remain constant over time. If the stock market is regarded as a system, then it is not an isolated system: there is a constant transfer of information between the stock market and the real economy. Thus, when information arrives from (leaves to) the real economy, then we can expect to see an increase (decrease) in the entropy of the stock market, corresponding to situations of increased (decreased) randomness.

Most often, entropy is used in one of the two main approaches, either as Shannon Entropy − in the discrete case − or as Differential Entropy − in the continuous time case. Shannon Entropy quantifies the expected value of information contained in a realization of a discrete random variable. Also, is a measure of uncertainty, or unpredictability: for a uniform discrete distribution, when all the values of the distribution have the same probability, Shannon Entropy reaches his maximum. Minimum value of Shannon Entropy corresponds to perfect predictability, while higher values of

Shannon Entropy correspond to lower degrees of predictability. The entropy is a more general measure of uncertainty than the variance or the standard deviation, since the entropy depends on more characteristics of a distribution as compared to the variance and may be related to the higher moments of a distribution.

Secondly, both the entropy and the variance reflect the degree of concentration for a particular distribution, but their metric is different; while the variance measures the concentration around the mean, the entropy measures the diffuseness of the density irrespective of the location parameter. In information theory, entropy is a measure of the uncertainty associated with a random variable. The concept developed by Shannon (1948) entropy, which quantifies the expected value of the information contained in a message, frequently measured in units such as bits. In this context, a 'message' means a specific realization of the random variable. The USA National Science Foundation workshop (2003) pointed out that the; "Information Technology revolution that has affected Society and the world so fundamentally over the last few decades is squarely based on computation and communication, the roots of which are respectively Computer Science (CS) and Information Theory (IT)". Shannon.(1948) provided the foundation for information theory. In the late 1960s and early 1970s, there were tremendous interdisciplinary research activities from IT and CS, exemplified by the work of Kolmogorov, Chaitin, and Solomonoff, with the aim of establishing the algorithmic information theory. Motivated by approaching the Kolmogorov complexity algorithmically, A. Lempel (a computer scientist), and J. Ziv (an information theorist) worked together in later 1970s to develop compression algorithms that are now widely referred to as Lempel-Ziv algorithms. Today, these are a de facto standard for lossless text compression; they are used pervasively in computers, modems, and communication networks. Shannon's entropy represents an absolute limit on the best possible lossless compression of any communication, under certain constraints: treating messages to be encoded as a sequence of independent and identically-distributed random variables Shannon's source coding theorem shows that, in the limit, the average length of the shortest possible representation to encode the messages in a given alphabet is their entropy divided by the logarithm of the number of symbols in the target alphabet. For a random variable X with $n\,outcomes$, $\{x_i : i = 1, ...n\}$ the Shannon entropy is defined as:

$$H(X) = -\sum_{i=1}^{n} p(x_i)log_b p(x_i) \tag{2}$$

where $p(x_i)$is the probability mass function of outcome x_i. Usually logs to the base 2 are used when we are dealing with bits of information. We can also define the joint entropy of two random variables as follows:

$$H[X, Y] = -\sum_{x \in \chi} \sum_{y \in \gamma} Pr(x, y)log_2(Pr(x, y)) \tag{3}$$

The joint entropy is a measure of the uncertainty associated with a joint distribution. Similarly, the conditional entropy can be defned as:

$$H[X \mid Y] = -\sum_{x \in \chi} \sum_{y \in \gamma} Pr(x, y)log_2 Pr(x \mid y) \tag{4}$$

Where the conditional entropy measures the uncertainty associated with a conditional probability. Clearly, a generalised measure of uncertainty has lots of important implications across a wide number of disciplines. In the view of Jaynes (1957), thermodynamic entropy should be seen as an application of Shannon's information theory. Jaynes (2003) gathers various threads of modern thinking about Bayesian probability and statistical inference, develops the notion of probability theory as extended logic and contrasts the advantages of Bayesian techniques with the results of other approaches. Golan (2002) provides a survey of information-theoretic methods in econometrics, to examine the connecting theme among these methods, and to provide a more detailed summary and synthesis of the sub-class of methods that treat the observed sample moments as stochastic. Within the above objectives, this review focuses on studying the interconnection between information theory, estimation, and inference. Granger, Massoumi and Racine (2004) applied estimators based

on this approach as a dependence metric for nonlinear processes. Pincus (2008) demonstrate the utility of approximate entropy (ApEn), a model-independent measure of sequential irregularity, via several distinct applications, both empirical data and model-based. He also considers cross-ApEn, a related two-variable measure of asynchrony that provides a more robust and ubiquitous measure of bivariate correspondence than does correlation, and the resultant implications for diversification strategies and portfolio optimisation. A theme further explored by Bera and Park (2008). Sims (2005) discusses information theoretic approaches that have been taken in the existing economics literature to applying Shannon capacity to economic modelling, whilst both critiquing existing models and suggesting promising directions for further progress.

Usually, the variance is the central measure in the risk and uncertainty analysis in financial markets. However, the entropy measure can be used as an alternative measure of dispersion, and some authors consider that the variance should be interpreted as a measure of uncertainty with some precaution [see, e.g. Maasoumi (1993) and Soofi (1997)]. Ebrahimi, Maasoumi and Soofi (1999) examined the role of the variance and entropy in ordering distributions and random prospects, and concluded that there is no general relationship between these measures in terms of ordering distributions. They found that, under certain conditions, the ordering of the variance and entropy is similar for transformations of continuous variables, and show that the entropy depends on many more parameters of a distribution than the variance. Indeed, a Legendre series expansion shows that the entropy is related to higher-order moments of a distribution and thus, unlike the variance, could offer a better characterization of $p_X(x)$ since it uses more information about the probability distribution than the variance [see Ebrahimi *et al.* (1999)].

Maasoumi and Racine (2002) argue that when the empirical probability distribution is not perfectly known, then entropy constitutes an alternative measure for assessing uncertainty, predictability and also goodness-of-fit. It has been suggested that entropy represents the disorder and uncertainty of a stock market index or a particular stock return series, since entropy has the ability to capture the complexity of systems without requiring rigid assumptions that may bias the results obtained.

To estimate entropy we used the 'entropy package' in the R library. Hausser and Strimmer (2009), we draw on their account to explain how they develop their estimators: to define the Shannon entropy, consider a categorical random variable with alphabet size p and associated cell probabilities $\theta_1, ..., \theta_p$, with $\theta_k > 0$ and $\sum_k \theta_k = 1$. If it is assumed that p is fixed and known. In this case Shannon entropy in natural units is given by:

$$H = -\sum_{k=1}^{p} \theta_k log(\theta_k) \tag{5}$$

In practice the underlying probablity mass function is unknown and therefore H and θ_k need to be estimated from observed cell counts from the sample used $y_k \geq 0$. A commonly used estimator of entropy is the maximum likelihood estimator (ML) which is given by:

$$\hat{H}^{ML} = -\sum_{k=1}^{p} \hat{\theta}_k^{ML} log(\hat{\theta}_k^{ML}) \tag{6}$$

This is formed by plugging in the ML frequency estimates

$$\hat{\theta}_k^{ML} = \frac{y_k}{n} \tag{7}$$

into equation (5), with $n = \sum_{k=1}^{p} y_k$ being the total number of counts.

2.1.1. Maximum Likelihood Estimation

The multinomial distribution is used to make the connection between observed counts y_k and frequencies θ_k.

$$Prob(y_1, ..., y_p) = \frac{n!}{\prod_{k=1}^{p} y_k!} \prod_{k=1}^{p} \theta_k^{y_k} \tag{8}$$

Note that $\theta_k > 0$ otherwise the distribution is singular. By contrast there may be zero counts y_k. The ML estimator of θ_k maximizes the right hand side of equation (8) for fixed y_k, leading to the observed frequencies $\hat{\theta}_k^{ML} = \frac{y_k}{n}$ with variances $Var(\hat{\theta}_k^{ML}) = \frac{1}{n}\theta_k(1 - \theta_k)$ and Bias $(\hat{\theta}_k^{ML}) = 0$ as $E(\hat{\theta}_k^{ML}) = \theta_k$

2.1.2. Miller-Madox Estimator

Even though θ_K^{ML} is unbiased, the plug in entropy estimator $\hat{\theta}^{ML}$ is not. First order bias correction leads to:

$$\hat{H}^{MM} = \hat{H}^{ML} + \frac{m > 0 - 1}{2n} \tag{9}$$

where $m > 0$ is the number of cells with $y_k > 0$. This is temed the Miller-Madow estimator, see Miller (1955).

2.1.3. Bayesian Estimators

Bayesian regularization of the cell counts may lead to improvements over ML estimates. The Dirichlet distribution with parameters $a_1, a_2, ..., a_p$ as prior, the resulting posterior distribution is also Dirichlet with mean

$$\hat{\theta}_k^{Bayes} = \frac{y_k + a_k}{n + A}$$

where $A = \sum_{k=1}^p a_k$. The flattening constants a_k play the role of pseudo counts (compare with equation (7)), therefore A may be interpreted as the $a\,priori$ sample size.

$$\hat{H}^{Bayes} = -\sum_{k=1}^p \hat{\theta}_k^{Bayes} log(\hat{\theta}_k^{Bayes}) \tag{10}$$

2.1.4. The Chao-Shen Estimator

The Chao-Shen (2003) estimator utilises the Horvitz-Thompson (1952) estimator combined with the Good-Turing correction (see Good (1953) and Orlitsky et al, (2003)), in the utilisation of the empirical cell probabilities to the issue of entropy estimation. The frequency estimates after application of the Good-Turing correction are:

$$\hat{\theta}_k^{GT} = (1 - \frac{m_1}{n})\hat{\theta}_k^{ML}$$

where m_1 represents the number of singletons, in effect cells with $y_k = 1$. If this is combined with the Horvitz-Thompson estimator the result is:

$$\hat{H}^{CS} = -\sum_{k=1}^p \frac{\hat{\theta}_k^{GT} log\hat{\theta}_k^{GT}}{(1 - (1 - \hat{\theta}_k^{GT})^n)} \tag{11}$$

2.1.5. Mutual information

One attraction of entropy-based measures is that they can relax the linearity assumption and capture nonlinear associations amongst variables. The starting point is to capture the mutual information between pairs of variables $MI(X, Y)$. The mutual information is the Kullback-Leibler distance from the joint probability density to the product of the marginal probability densities:

$$MI(X, Y) = E_{f(x,y)} \left\{ log\frac{f(x, y)}{f(x)f(y)} \right\} \tag{12}$$

The measure mutual information (MI) is always non-negative, symmetric, and equals zero only if X and Y are independent. In the case of normally distributed variables MI is closely related to the Pearson Correlation coefficient.

$$MI(X,Y) = -\frac{1}{2}log(1-\rho^2)$$

The entropy representations is:

$$MI(X,Y) = H(X) + H(Y) - H(X,Y) \tag{13}$$

This shows that MI can be computed from the joint and marginal entropies of the two variables.

Some preliminary results

Table 4 presents some of the common methods adopted.

Table 4: Common choices for the parameters of the Dirichlet prior in the Bayesian estimators of cell frequencies, and corresponding entropy estimators

a_k	Cell frequency prior	Entropy estimator
0	no prior	maximum likelihood
1/2	Jeffreys prior (Jeffreys, 1946)	Krichevsky and Trofimov (1981)
1	Bayes-Laplace uniform prior	Holste et al (1998)
1/p	Perks prior (Perks, 1946)	Sh$\ddot{u}$rmann and Grassberger (1996)
$\sqrt{n}/p$	minmax prior (Trybula, 1958)	

Source:Hausser and Strimmer (2009)

We used our returns series for the S&P500 and the VIX and estimated their entropy and cross-entropy. The results are shown in Table 5.

Table 5: Entropy and MI statistics for S&P500 returns and VIX returns 1990-June 2011

	S&P500 returns	VIX returns
Maximum Likelihood Estimate	2.1434	3.8269
Miller-Madow Estimator	2.1462	3.8367
Jeffrey's prior	2.1598	3.8846
Bayes-Laplace	2.1747	3.9322
SG	2.1444	3.8279
Minimax	2.1983	3.8780
ChaoShen	2.1494	3.8424
Mutual Information MI		
MI Dirichlet (a = 0)	0.3535395	
MI Dirichlet (a = 1/2)	0.3465671	
MI Empirical (ML)	0.3535395	

2.2. Non-parametric estimation

Hayfield and Racine (2008) comment that: "the appeal of nonparametric methods, for applied researchers at least, lies in their ability to reveal structure in data that might be missed by classical parametric methods". Their "NP" package available in R provides a flexible underlying the treatment of kernel methods in the presence of a mix of categorical and continuous data and lies in the use of what they call 'generalized product kernels'. They develop their approach as follows: suppose you are confronted with discrete data $X^d \in S^d$ where S^d denotes the support of X^d. If we use x_s^d and X_{is}^d to denote the *sth* component of x^d and $X_i^d \in \{0, 1, ..., c_s - 1\}$, we define a discrete univariate kernel function
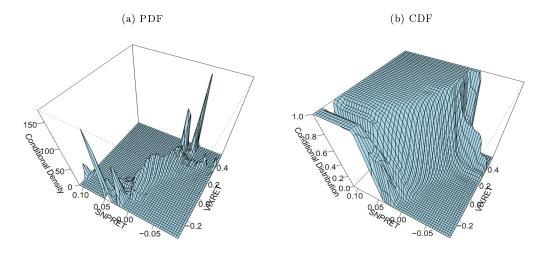
$$l^u(X_{is}^d, x_s^d, \lambda_s) = \left\{ \begin{array}{ll} 1 - \lambda_s & if \ X_{is}^d = x_s^d \\ \lambda_s/(c_s - 1) & if \ X_{is}^d \neq x_s^d \end{array} \right\} \tag{14}$$

Note that when $\lambda_s = 0$, $l(X_{is}^d, x_s^d, 0) = 1(X_{is}^d = x_s^d$ becomes an indicator function, and if $\lambda_s = (c_s - 1)/c_s$ then $l(X_{is}^d = x_s^d, \frac{c_s-1}{c_s}) = 1/c_s$ is a constant for all values of X_{is}^d and x_s^d. The range of λ_s is $[0, (c_s - 1)/c_s]$. Observe that these weights sum to one. If a variable was continuous, you could use, the second order Gaussian kernel, namely

$$w(x^c, X_i^c, h) = \frac{1}{\sqrt{2\pi}} \, exp \left\{ -\frac{1}{2} \left(\frac{X_i^c - x^c}{h} \right)^2 \right\} \tag{15}$$

Observe that these weights integrate to 1.

Figure 5: Nonparametric conditional PDF and CDF estimates of the joint distribution of S&P500 returns and VIX returns 1990-June 2011

(a) PDF   (b) CDF

2.2.1. Kernel estimation of a conditional quantile

Once we can estimate conditional CDFs such as those shown in Figure 5, we can then estimate conditional quantiles, (see the discussion in Racine (2008)). That is, having estimated the conditional CDF we simply invert it at the desired quantile as described in equation (15). A conditional αth quantile of a conditional distribution function $F(\cdot \mid x)$ is defined by ($\alpha \in (0, 1)$)

$$q_\alpha(x) = inf\left\{y : F(y \mid x) \geq \alpha\right\} = F^{-1}(\alpha \mid x) \tag{16}$$

or equivalently, $F(q_\alpha(x) \mid x) = \alpha$. The conditional quantile function $q_\alpha(x)$ can be directly estimated by inverting the previously estimated conditional CDF function. In effect:

$$q_\alpha(x) = inf\left\{y : \hat{F(y \mid x)} \geq \alpha\right\} \equiv F^{-1}(\hat{\alpha} \mid x) \tag{17}$$

The results of this are shown in table .

2.2.2. Testing the Equality of Univariate Densities

Massoumi and Racine (2002) suggest a metric entropy which is useful for testing for the equality of densities for two univariate random variables X and Y. They compute the nonparametric metric entropy (normalised Hellinger of Granger *et* al (2004)). This tests the nul of equality of two univariate density functions. For continuous variables they suggest the construction of:

$$S_p = \frac{1}{2} \int (f_1^{1/2} - f_2^{1/2})^2 dx$$

Table 7: Entropy measures VIX continuously compounded daily returns March 2005 to May 2012

	VIX 2005-06	VIX 2007-08	VIX 2009-10	VIX 2011-12
Maximum Likelihood Estimate	3.6441	4.0114	3.7744	3.8723
Miller-Madow Estimator	3.7045	4.0911	3.8421	3.9669
Jeffrey's prior	3.9220	5.3321	4.0743	4.2843
Bayes-Laplace	4.0811	5.7334	4.2444	4.4751
SG	3.6541	4.0278	3.7835	3.8861
Minimax	3.7787	4.2524	3.9054	4.0306
ChaoShen	3.7086	4.1006	3.8539	3.9573

$$= \frac{1}{2} \int \left(1 - \frac{f_2^{1/2}}{f_1^{1/2}} \right)^2 dF_1(x)$$

where $f_1 = f(x)$ and $f_2 = f(y)$ are the marginal densities of the random variables X and Y. If X and Y are discrete or categorical, then the integration can be replaced by summing over all possible outcomes.

3. Main Results

3.1. Entropy metrics

We focus our attention on a time frame during which the current VIX has been in operation. We obtained our data from SIRCA's TRTH data and obtained daily closing prices for the VIX beginning in March 2005 and running through until May 2012. The data was divided into roughly two year periods as we wanted to explore how these measures varied during times of crisis. Table (5) presents the entropy measures for our two daily continuously compounded return series; S&P500 and VIX for four roughly one and a half to two year blocks running from: March 9th 2005 until 29th December 2006, January 3rd 2007 to 31st December 2008, January 2nd 2009 to 31st December 2010, and January 3rd 2011 to 1st May 2012.

Table 6: Entropy measures S&P500 continuously compounded daily returns March 2005 to May 2012

	S&P500 2005-06	S&P500 2007-08	S&P500 2009-10	S&P500 2011-12
Maximum Likelihood Estimate	1.6518	2.4165	2.2611	2.1859
Miller-Madow Estimator	1.6596	2.4454	2.2790	2.2211
Jeffrey's prior	1.6610	2.5364	2.2956	2.2516
Bayes-Laplace	1.6698	2.6299	2.3262	2.3051
SG	1.6542	2.4243	2.2649	2.1935
Minimax	1.6967	2.545	2.331	2.2920
ChaoShen	1.6527	2.4781	2.2809	2.2282
Mutual Information MI S&P500 and VIX				
MI Dirichlet (a = 0)	0.09489	0.1941	0.1927	0.2055
MI Dirichlet (a = 1/2)	0.08485	0.1693	0.1702	0.1746
MI Empirical (ML)	0.09489	0.1941	0.1927	0.2055

The entropy measures for the VIX are shown in Table 7

The results in Tables (6) and (7) are consistent and intuitive. Entropy measures capture changes in uncertainty, and clearly for both the S&P500 and the VIX the maximum uncertainty, in terms of entropy measures, the highest minimum estimates of 2.42 for the S&P500 continuously compounded daily returns, or 4.01 for the VIX continuously compounded daily returns were recorded in the two

year interval 2007-2008. Given that this captures the start of the Global Financial Crisis (GFC), this is not surprising. The Mutual Information (MI) measure results are a bit more surprising. Mutual information measures the information that X and Y share: it measures how much knowing one of these variables reduces uncertainty about the other. The measure at the foot of Table (6) suggests that the amount of information revealed by the VIX's continuously compounded daily returns about the S&P500 continuously compounded returns increases over the two year time frames examined and records its highest levels on two metrics of over 0.20 in the period 2011-2012.

This leads to consideration of the nature of the two probability density functions for the two return series. We have generated the non parametric conditional PDF and CDF conditioning the S&P500 returns on the VIX returns for our roughly two-year sample intervals. We suspect that the relationship may not remain constant through time. The results are shown in Figure (6)

3.2. Non parametric conditional PDF and CDF estimation

Figure 6: Conditional Density Plots PDF and CDF for S&P500 and VIX returns 2005-2006 for our sample intervals
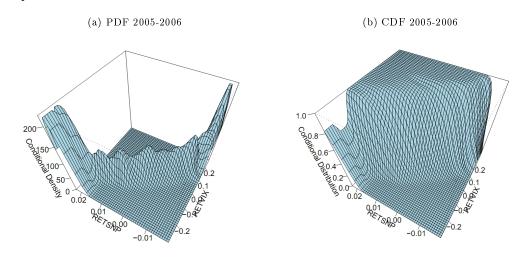
(a) PDF 2005-2006   (b) CDF 2005-2006

Figure 7: Conditional Density Plots PDF and CDF for S&P500 and VIX returns 2005-2006 for our sample intervals

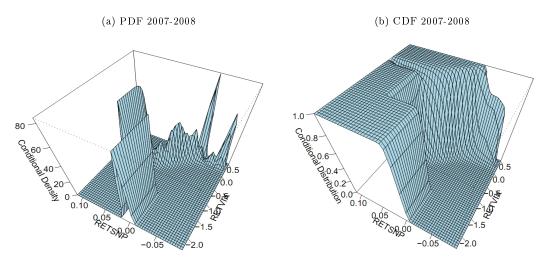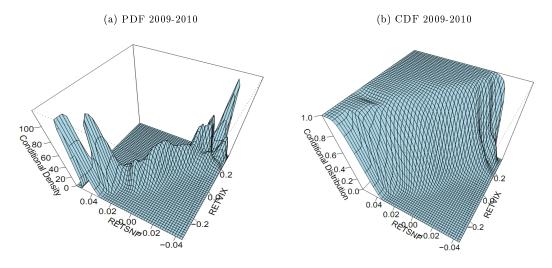(a) PDF 2007-2008   (b) CDF 2007-2008

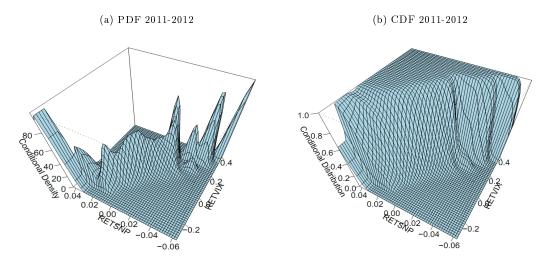Figure 8: Conditional Density Plots PDF and CDF for S&P500 and VIX returns 2009-2010 for our sample interval

(a) PDF 2009-2010   (b) CDF 2009-2010

Figure 9: Conditional Density Plots PDF and CDF for S&P500 and VIX returns 2011-2012 for our sample interval

(a) PDF 2011-2012   (b) CDF 2011-2012

3.3. *Quantile regression analysis of hedge ratios*

The relationship between the S&P500 and the VIX can be further explored by application of quantile regression. This will help to shed light on whether hedging ratios are likely to be constant across the two distributions. Quantile Regression is modelled as an extension of classical OLS (Koenker and Bassett, (1978)), in quantile regression the estimation of conditional mean as estimated by OLS is extended to similar estimation of an ensemble of models of various conditional quantile functions for a data distribution. In this fashion quantile regression can better quantify the conditional distribution of $(Y|X)$. The central special case is the median regression estimator that minimizes a sum of absolute errors. We get the estimates of remaining conditional quantile functions by minimizing an asymmetrically weighted sum of absolute errors, here weights are the function of the quantile of interest. This makes quantile regression a robust technique even in presence of outliers. Taken together the ensemble of estimated conditional quantile functions of

$(Y|X)$ offers a much more complete view of the effect of covariates on the location, scale and shape of the distribution of the response variable.

For parameter estimation in quantile regression, quantiles as proposed by Koenker and Bassett (1978) can be defined through an optimization problem. To solve an OLS regression problem a sample mean is defined as the solution of the problem of minimising the sum of squared residuals, in the same way the median quantile (0.5%) in quantile regression is defined through the problem of minimising the sum of absolute residuals. The symmetrical piecewise linear absolute value function assures the same number of observations above and below the median of the distribution. The other quantile values can be obtained by minimizing a sum of asymmetrically weighted absolute residuals, (giving different weights to positive and negative residuals). Solving

$$min\xi \varepsilon R \sum \rho_\tau(y_i - \xi) \tag{18}$$

Where $\rho_\tau(.)$ is the tilted absolute value function. Taking the directional derivatives of the objective function with respect to ξ (from left to right) shows that this problem yields the sample quantile as its solution.to obtain an estimate of the conditional median function, the scalar ξ in the first equation is replaced by the parametric function $\xi(x_t, \beta)$ and τ is set to $1/2$. The estimates of the other conditional quantile functions are obtained by replacing absolute values by $\rho_\tau(.)$ and solving

$$min_{\mu \varepsilon R^p} \sum \rho_\tau(y_i - \xi(x_i, \beta)) \tag{19}$$

The resulting minimization problem, when $\xi(x, \beta)$ is formulated as a linear function of parameters, and can be solved very efficiently by linear programming methods. The slope coefficients from the results of quantile regression of the daily continuously compounded S&P500 returns and the daily continuously compounded VIX returns are shown in Figures (10) and (11) All the slope coefficients are significant at a 1% level and all are negative, confirming the utility of the conceptual employment of VIX returns as a hedge against movements in S&P500 returns. In addition, the quantile regression slope coefficients vary considerably across the deciles, and also the nature of this variation is not consistent. In the first period hedge ratios become less negative by the decile, whilst in the second period the reverse occurs and they become more negative as they progress across the deciles. In the third period they continually increase, whilst in the fourth they steadily increase until the 7th decile and then become progressively more negative. These results are suggestive of the difficulty of setting up precise hedges incorporating volatility changes. In practice it is more difficult than this suggests, as instruments written against the VIX, options or futures, have to be used in practice. (This will be the subject of a subsequent paper).

Figure 10: Quantile regression slope coefficients of S&P500 daily returns for sample sub-periods regressed in daily continuously compounded VIX returns, sub-samples 2005-2006 and 2007-2008

(a) Quantile regression slope coefficients by decile 2005-2006 with error bands

(b) Quantile regression slope coefficients by decile 2007-2008 with error bands
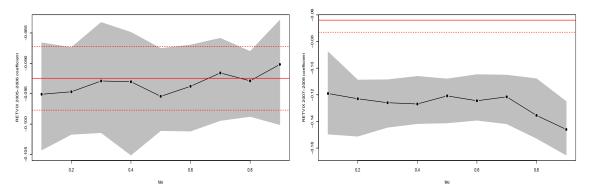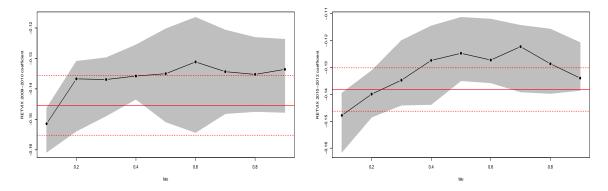

Figure 11: Quantile regression slope coefficients of S&P500 daily returns for sample sub-periods regressed in daily continuously compounded VIX returns, sub-samples 2009-2010 and 2011-2012

(a) Quantile regression slope coefficients by decile 2009-2010 with error bands

(b) Quantile regression slope coefficients by decile 2009-2010 with error bands

3.4. Non-parametric tests of density equalities

Our final set of tests feature an a non-parametric analysis of the equivalence of the density functions of our two sets of return distributions derived from the S&P500 and the VIX. We apply Massoumi and Racine (2002) metric entropy test of the equality of the densities of two random variables, in this case, the continuously compounded daily S&P500 returns and the daily continuously compounded VIX returns for our four subsample periods from 2005 to 2012. The bootstrap test is conducted by sampling with replacement from the pooled empirical distribution of the return series. We also conduct a non-parametric test of the symmetry of the two distributions of returns. This is a non-parametric metric entropy discussed in Massoumi and Racine (2009) which tests the null of symmetry using the densities/probabilities of the data series and rotated data series. Bootstrapping is conducted by resampling from the empirical distribution of the pooled data and rotated data. The results are shown in Tables (8) and (9).

Table 8: Entropy density equality tests for sub-samples 2005-2006, 2007-2008, 2009-2010, and 2011-2012 for daily continuously compounded S&P500 returns and VIX returns

	2005-2006	2007-2008	2009-2010	2011-2012
Consistent Univariate Entropy Density Equality Test				
Test Statistic 'Srho'	0.4660	0.3017	0.3219	0.3817
Probability	2.22e-16 ***	2.22e-16 ***	2.22e-16 ***	2.22e-16 ***
***Significant at 1% level				

Table 9: Entropy density asymmetry tests sub-samples 2005-2006, 2007-2008, 2009-2010, and 2011-2012 for daily continuously compounded S&P500 returns and VIX returns

	2005-2006	2007-2008	2009-2010	2011-2012
Consistent entropy asymmetry test				
S&P500 Test Statistic 'Srho'	0.00419	0.034547	0.01239	0.01904
Probability	0.52525	0.0***	0.19191	0.09090
VIX Test Statistic 'Srho'	0.00923	0.01278	0.02485	0.03018
Probability	0.1818	0.15151	0.0 ***	0.0 ***
***Significant at 1% level				

The results of the entropy based density equality tests shown in Table (8) clearly reject the null that the distributions of the daily returns on the S&P500 series and the VIX series in our four subsample periods are drawn from the same distribution and all the results are highly significant at the 1% level. The entropy based tests of asymmetry results in Table (9) also reject the null of symmetry at the 1% level in the cases of the S&P500 in the period 2007-2008, and the VIX in 2009-2010, and 2011-2012. These results also support our examination of the conditional density functions of the two series and Figures (6) to (9) which strongly suggested the existence of assymetry and non-equality of the two sets of return distributions in our four sub sample periods.

4. Conclusion

We have examined the relationship between the S&P500 daily continuously compounded return series and a similar series for the VIX in terms of a long sample drawn from the CBOE website running from 1990 to mid 2011 and a set of returns from SIRCA's TRTH datasets running from March 2005 to-date which we divided into four roughly equivalent sub-samples and subjected to a series of non-parametric based tests utilising entropy based metrics. These suggested that the PDFs and CDFs of these two return distributions change shape in various subsample periods. The entropy and MI statistics suggest that the degree of uncertainty attached to these distributions changes through time and using the S&P500 return as the dependent variable, the amount of information obtained from the VIX also changes with time, though it increased in the most recent period from 2011 to 2012. The entropy based non parametric tests of the equivalence of the two distributions and their symmetry strongly reject their respective nulls in the case of the first tes and in certain periods for the second test.

The implication of these non-parametric tests is that the relationship between the S&P500 and the VIX is complex and subject to change in different time periods. This implies that portfolio risk-hedging using the two series will be complex and difficult to calibrate. This is compounded by the fact that the VIX is not a cash insrument and derivates written against the VIX such as options and futures have to be utilised for practical hedging purposes and therefore add a second layer of complexity. The analysis of which will be the subject of a companion paper.

References

[1] Alexander, C. (1999) "Optimal Hedging Using Cointegration", *Philosophical Transactions of the Royal Society*, London, Series A, **357**, : 2039-2058

[2] Bera, A.K. and S. Y. Park (2008) "Optimal Portfolio Diversification Using the Maximum Entropy Principle", *Econometric Reviews*, 27:4-6, 484-512.

[3] Brenner, M., E.Y. Ou and J.E. Zhang (2006), Hedging volatility risk, *Journal of Banking and Finance*, 30, 811-821.

[4] Caporin, M. and M. McAleer (2011), Do we really need both BEKK and DCC? A tale of two multivariate GARCH models, to appear in *Journal of Economic Surveys*.

[5] Chang, C.L., J-A, Jiminez-Martin, M. McAleer, T. Perez-Amaral, (2011) "The Rise and Fall of S&P Variance Futures", Working Paper, Department of Quantitative Economics, Complutense University of Madrid.

[6] Chaitin, G. J. (1969) "On the Simplicity and Speed of Programs for Computing Infinite Sets of Natural Numbers", *Journal of the ACM* 16 (3): 407

[7] Chao,A., and T. J. Shen (2003) "Nonparametric estimation of Shannon's index of diversity when there are unseen species", *Environ. Ecol. Stat.*, 10:429–443, 2003.

[8] Demeterfi, K., E. Derman, M. Kamal, and J. Zou, (1999) "More Than You Ever Wanted To Know About Volatility Swaps", Quantative Strategies Research Notes, Goldman Sachs.

[9] Ebrahimi, N., Maasoumi, E., Soofi, E. (1999) Journal of Econometrics, 90, 2, 317-336.

[10] Edrington, L., 1979, The Hedging Performance of the New Futures Markets, *Journal of. Finance*, 34: 157-70.

[11] Engle, R. F., and C.W.J. Granger, (1987) "Co-integration and error correction: Representation, estimation and testing", *Econometrica*, 55(2), 251-276

[12] .A. Golan, "Information and Entropy Econometrics – Editor's view," *Journal of Econometrics*, 107, 1-2 (2002), 1-15.

[13] Golan. A and E. Maasoumi (2008) "Information Theoretic and Entropy Methods: An Overview", *Econometric Reviews*, 27:4-6, 317-328.

[14] Good, I. J. (1953) "The population frequencies of species and the estimation of population parameters", *Biometrika*, 40:237–264.

[15] Granger, C., Maasoumi, E. and Racine, J. (2004), 'A dependence metric for possibly nonlinear time series', *Journal of Time Series Analysis* 25(5), 649{669.

[16] Hayfield,T., and J.S. Racine (2008) "Nonparametric Econometrics: The np Package", Journal of Statistical Software, 27,5, 1-32.

[17] Hausser, J., and K. Strimmer, (2012) *Package 'entropy'*, Repository CRAN, http://www.r-project.org/

[18] Hausser, J., and K. Strimmer, (2009) "Entropy Inference and the James-Stein Estimator, with Application to Nonlinear Gene Association Networks", Journal of Machine Learning Research, 10, 1469-1484.

[19] Holste, D., I. Große, and H. Herzel, (1998) "Bayes' estimators of generalized entropies", *J. Phys. A: Math. Gen.*, 31:2551–2566.

[20] Horvitz, D.G. and D. J. Thompson, (1953) "A generalization of sampling without replacement from a finite universe", *J. Amer. Statist. Assoc.*, 47:663–685.

[21] Huskaj, B. (2009), A value-at-risk analysis of VIX futures: Long memory, heavy tails, and asymmetry. Available at SSRN: http://ssrn.com/abstract=1495229.

[22] Ishida, I., M. McAleer and K. Oya (2011), Estimating the leverage parameter of continuous-time stochastic volatility models using high frequency S&P500 and VIX, *Managerial Finance*, 37, 1048-1067.

[23] Jaynes, E. T. (2003) *Probability Theory: The Logic of Science.* Cambridge University Press, ISBN 0-521- 59271-2.

[24] Jeffreys, H., (1946) "An invariant form for the prior probability in estimation problems," *Proc. Royal. Society (Lond.) A*, 186:453–461.

[25] Koenker, R. W., & Bassett, G. Jr. (1978) "Regression Quantiles" *Econometrica 46* (1), 33-50.

[26] Koenker, R. (2005). *Quantile Regression*, Econometric Society Monograph Series: Cambridge University Press.

[27] Kolmogorov, A.N. (1965). "Three Approaches to the Quantitative Definition of Information", *Problems Inform. Transmission* 1 (1): 1–7.

[28] Kolmogorov, A.N. and V. A. Uspensky, (1987) "Algorithms and randomness", SIAM J. *Theory of Probability and Its Applications*, vol. 32 389-412.

[29] Krichevsky, R.E., and V. K. Trofimov, (1981) "The performance of universal encoding". *IEEE Trans. Inf. Theory,* 27:199–207.

[30] Maasoumi, E. (1993) "A Compendium to Information Theory in Economics and Econometrics", *Econometric Reviews,* , 12, 2, 137-181

[31] Maasoumi, E., and Racine, (2002). "Entropy and Predictability of Stock Market Returns," *Journal of Econometrics*, 107(2), 291–312.

[32] McAleer, M. and C. Wiphatthanananthakul (2010), A simple expected volatility (SEV) index: Application to SET50 index options, *Mathematics and Computers in Simulation*, 80, 2079- 2090.

[33] Miller, G.A. (1955) "Note on the bias of information estimates", In H. Quastler, editor, *Information Theory in Psychology II-B*, pages 95–100. Free Press, Glencoe, IL.

[34] Orlitsky, A,. N. P. Santhanam, and J. Zhang, (2003) "Always Good Turing: asymptotically optimal probability estimation", *Science,* 302:427–431, 2003.

[35] Perks W., (1947) "Some observations on inverse probability including a new indifference rule", *J. Inst. Actuaries*, 73:285–334.

[36] Pincus, S. (2008) "Approximate Entropy as an Irregularity Measure for Financial Data", *Econometric Reviews*, 27:4-6, 329-362

[37] Racine, J.S. (2008) "Nonparametric Econometrics: A Primer", *Foundations and Trends in Econometrics*, 3, 1, 1–88

[38] Sepp, A. (2008), VIX option pricing in a jump-diffusion model, *Risk Magazine*, April, 84-89.

[39] Shannon, C.E. (1948) "A Mathematical Theory of Communication", *The Bell System Technical Journal*, Vol. 27, pp. 379–423, 623–656.

[40] Schürmann, T., and P. Grassberger, (1996) "Entropy estimation of symbol sequences", *Chaos,* 6:414–427.

[41] Sims, C.A. (2005) "Rational Inattention: A Research Agenda", Deutsche Bundesbank Discussion Paper Series 1: Economic Studies No 34/2005

[42] Soofi, E.. (1997) Information Theoretic Regression Methods. In: Advances in Econometrics - Applying Maximum Entropy to Econometric Problems, Fomby, T. and R. Carter Hill eds. Vol. 12. Jai Press Inc., London.

[43] Solomonoff, R. (1960) "A Preliminary Report on a General Theory of Inductive Inference", Report V-131 (Cambridge, Ma.: Zator Co.). revision, Nov., 1960.

[44] Thode Jr., H.C. (2002) *Testing for Normality*, Marcel Dekker, New York.

[45] Trybula, S. (1958) "Some problems of simultaneous minimax estimation", *Ann. Math. Statist.*, 29:245–253.

[46] Whaley, R.E. (2009) "Understanding the VIX", *The Journal of Portfolio Management,* 35, 3: 98-105