Orthant probabilities for robust correlation and structural performance enhancement

September 1, 2015.
Co-winner of the "Call for Papers 2015", Quant Congress USA, NY, July 14-15, 2015.
Published in abridged form as "Risk management for return enhancement", Risk magazine, February 2016.

Martin Anderson[1], Shan Chen[1], Jennifer Eichholz[1], Marc Lieberman[2], Mark Lundin[3], Vaida Maleckaite[1], Ryan Parham[1], Stephen Satchell[4], Mark Steed[1]

[1] Arizona Public Safety Personnel Retirement System
[2] Kutak Rock LLP
[3] Charles Schwab Investment Management[*]
[4] University of Sydney Business School, Discipline of Finance and University of Cambridge Trinity College

Abstract

Orthant probabilities applied in a two-dimensional framework are used to derive quadrant-conditional financial asset return correlations which fully capture both linear and non-linear components of co-variability. We investigate the potential for employing quadrant-conditional correlations in order to construct portfolios which generate a long-run average portfolio return which is more positive than long-run averages of individual assets' returns. Risk-based, but return-enhancing security selection applications involving assets co-variability criteria for investment management and high-frequency arbitrage trading are discussed.

Introduction

Correlation in the context of Markowitz's (1952) Modern Portfolio Theory is, in practice, often taken to mean application of Pearson's (1896) product-moment correlation coefficient; ρ as applies to a population, or r as applied to a sample. Mean and variance self-adjusting, invariant to multivariate scales, it serves as a powerful single-pass algorithm. ρ^2 (as estimated by r^2) efficiently measures that part of the variance of one variable that is explained linearly by another variable, however any non-linear dependence between variables goes unmeasured. Misinterpretation of what Pearsons' product-moment correlation actually estimates (as opposed to misestimation by the method itself) may lead to false conclusions in any implementation reliant on correlation. This is particularly relevant for analysis of financial securities, whose return distributions are rarely close to normal; their non-linearities routinely manifest themselves through distributional asymmetries (skewness) and narrow "peakedness" (kurtosis).

As with the variances and covariance from which Pearson's product-moment correlation coefficient is derived, the correlation between financial securities changes with time and

[*] The views and opinions expressed in this paper are those of the authors and do not necessarily reflect the views of Charles Schwab Investment Management, Inc. (CSIM) or its affiliates. The research and analysis described herein is not related to any strategies or activities currently used by CSIM.

presumably also with financial market conditions. Estimations of Pearson's r for different data samples produce differing results in ways which cannot be fully explained by statistical sampling fluctuations. It is possible that relevant information goes unmeasured, or that relevant information becomes smeared in a long-term average.
We discuss a correlation estimation method which gauges both linear and non-linear dependencies by ignoring distributional characteristics, relying instead on occupancy of the four quadrants of the Cartesian plane formed by the returns of two investments. A potentially profitable by-product of this method is that it involves the estimation of correlation in different regimes simultaneously, and we offer the conjecture that this allows a more microscopic view of what is typically gauged in practice by other methods as a correlation amalgam. Applied to financial securities, evidence is presented that this essentially risk-based estimation method includes return information which may allow portfolio performance improvement based on exploitable co-variability characteristics. From the perspective that a financial security may refer to an irreducible asset or itself be a portfolio of underlying assets, this discussion has important implications for security selection, portfolio construction, investment manager selection and asset allocation.

Orthant probabilities and orthant correlations

Given a simple portfolio of two assets whose marginal returns form a joint distribution, (X, Y), four performance outcomes are possible for price changes over a finite time period: Both assets' returns may be positive, the returns may contradict each other with the first being positive and the second negative, the returns may contradict each other with the first being negative and the second positive, or both assets' returns may be negative. For the bivariate normal distribution, two assets delivering random standard normal marginal distributions with zero means, the probability for each of these possibilities can be described separately through Sheppard's (1898, 1899) theorem of median dichotomy:

$$P_{11} = Prob[X > 0, Y > 0] = \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(\rho_{X,Y}) \quad , \tag{1}$$

$$P_{01} = Prob[X \leq 0, Y > 0] = \frac{1}{4} - \frac{1}{2\pi}\sin^{-1}(\rho_{X,Y}) \quad , \tag{2}$$

$$P_{00} = Prob[X \leq 0, Y \leq 0] = \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(\rho_{X,Y}) \quad , \tag{3}$$

$$P_{10} = Prob[X > 0, Y \leq 0] = \frac{1}{4} - \frac{1}{2\pi}\sin^{-1}(\rho_{X,Y}) \quad , \tag{4}$$

where $\rho_{X,Y}$ is the linear correlation between X and Y.

Equations (1-4) are often referred to as orthant normal probabilities. The term "orthant" is used in reference to a framework which can be expanded to n-dimensions, but for the two-dimensional case involving the returns of two financial securities there are four orthants or quadrants. The orthant probabilities arise from the conversion of the joint probability density function to polar coordinates and form one of the stronger links between probability theory and the vector and linear algebra descriptions of correlation.

Equations (1)-(4) apply to a mean=0 bivariate normal distribution. It can be shown (see Proof in Appendix A) that this also holds for the class of elliptical distributions which are mean zero and which can be described as conditionally scale bivariate normal, where the scale is a positive random variable with finite mean. This includes some members of the zero mean elliptical family with excess kurtosis, including the bivariate t-distribution and mixtures of normal distributions with the same covariance matrix. In these circumstances, rearranging equations (1)-(4) and solving for correlation terms results in four different but equivalent definitions of a single correlation. If the joint distribution is bivariate normal, these equivalent orthant correlations are also equal to Pearson's product-moment correlation coefficient.

For cases involving other distributions, solving Equations (1)-(4) for the correlation variable will result in implied orthant correlations which would be equal under normality/ellipticity. Rearranging for correlation terms then, expressions of implied quadrant correlations can be expressed conditionally on the quadrants of the Cartesian plane of the bivariate X, Y distribution:

$$\rho_{11} = \rho_{X.Y}|[X > 0, Y > 0] = \sin\left(2\pi\left(P_{11} - \tfrac{1}{4}\right)\right) \quad , \tag{5}$$

$$\rho_{01} = \rho_{X.Y}|[X \leq 0, Y > 0] = \sin\left(2\pi\left(-P_{01} + \tfrac{1}{4}\right)\right) \; , \tag{6}$$

$$\rho_{00} = \rho_{X.Y}|[X \leq 0, Y \leq 0] = \sin\left(2\pi\left(P_{00} - \tfrac{1}{4}\right)\right) \quad , \tag{7}$$

$$\rho_{10} = \rho_{X.Y}|[X > 0, Y \leq 0] = \sin\left(2\pi\left(-P_{10} + \tfrac{1}{4}\right)\right) \; . \tag{8}$$

Orthant Probability Testing

Before moving to a discussion on the interpretation and use of orthant, or quadrant correlations, we address the issue of statistical significance. A t-value for quadrant correlations is derived so that a t-test can be conducted and p-values inferred from Student's t-distribution. In addition, we provide exact tests for forecasting hypotheses concerning orthant probabilities.

Denoting
$P_{00} = Prob(X \leq 0, Y \leq 0) \quad ,$
$P_{11} = Prob(X > 0, Y > 0) \quad ,$
$P_{01} = Prob(X \leq 0, Y > 0) \quad ,$
$P_{10} = Prob(X > 0, Y \leq 0) \quad .$

It follows that:

$P_{00} + P_{01} + P_{10} + P_{11} = 1$.

If the probability density function, $pdf(x, y)$, is symmetric around (0,0), then $P_{00} = P_{11}$, $P_{10} = P_{01}$ and $P_{00} + P_{10} = \frac{1}{2}$.

If we further assume that X, Y are bivariate elliptical with finite second moments, then it is well-known that a transformation of the sample correlation coefficient, R, based on a sample of N observations, is distributed as Student's t in particular,

t= $\frac{R\sqrt{N-2}}{\sqrt{1-R^2}}$ has a t-distribution with N-2 degrees of freedom. We present the following result.

Theorem.

To test if $P_{11} = Prob(X > 0, Y > 0) = \frac{1}{4}$; $P_{00} = Prob(X \leq 0, Y \leq 0) = \frac{1}{4}$; under the assumption of bivariate ellipticity with finite second moments. If $\widehat{P_{11}}$ is the sample estimator of $P_{11}$and $\widehat{P_{01}}$ is the sample estimator of P_{01}, then tan($2\pi(\widehat{P_{11}} - \frac{1}{4}))\sqrt{N-2}$ has a t-distribution with N-2 degrees of freedom and tan($2\pi(\frac{1}{4} - \widehat{P_{01}}))\sqrt{N-2}$ has a t-distribution with N-2 degrees of freedom.

Proof:

To test if $P_{11} = Prob(X > 0, Y > 0) = \frac{1}{4}$.

Using (1); $\frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(\rho_{X,Y}) = P_{11}$. In terms of sample data, R=sin($2\pi(\widehat{P_{11}} - \frac{1}{4})$)

.

Using the fact that

t= $\frac{R\sqrt{N-2}}{\sqrt{1-R^2}}$ has a t-distribution with N-2 degrees of freedom, we see that

t= tan($2\pi(\widehat{P_{11}} - \frac{1}{4}))\sqrt{N-2}$ has a t-distribution with N-2 degrees of freedom.

To test if $P_{00} = Prob(X < 0, Y < 0) = \frac{1}{4}$.

Using (1); $\frac{1}{4} - \frac{1}{2\pi}\sin^{-1}(\rho_{X,Y}) = P_{01}$. In terms of sample data, R=sin($2\pi(\frac{1}{4} - \widehat{P_{01}})$)

Using the fact that

t= $\frac{R\sqrt{N-2}}{\sqrt{1-R^2}}$ has a t-distribution with N-2 degrees of freedom, we see that

t= tan($2\pi(\frac{1}{4} - \widehat{P_{01}}))\sqrt{N-2}$ has a t-distribution with N-2 degrees of freedom.

We can apply the above results to test if the hit-rate, defined as the proportion of times that the forecasts and actual have the same sign. From the above result that to test if the hit-rate(HR) is 50%, we can define

HR=$P_{00} + P_{11}$=-$\frac{1}{2} + \frac{1}{\pi}\sin^{-1}(\rho_{X,Y})$ and so

t= tan($\pi(\widehat{HR} - \frac{1}{2}))\sqrt{N-2}$ has a t-distribution with N-2 degrees of freedom where $\widehat{HR}$ is the sample hit-rate.

Characteristics of orthant correlations

The four probability-dependent quadrant correlations of Equations (5) to (8) are shown graphically in two-dimensional rectangular Cartesian coordinate system of Figure 1, where the horizontal axis corresponds to the return of a first asset and the vertical axis corresponds to the return of a second asset. In the figure and subsequent text we follow the Cartesian convention of numbering quadrants one through four by starting with the upper right quadrant and moving counter-clockwise. If the joint probability distribution of assets' returns is zero-mean and random bivariate normal (therefore, median equal to mean) and uncorrelated, then returns from both assets occupy each of the four quadrants with equal probability of 0.25. An increase or decrease in occupancy will directly change (decreasing or increasing) that of some other quadrant as the sum of probabilities is one. In addition, any change in occupation or probability will necessarily affect quadrant correlations in the sinusoidal manner dictated by Equations (5)-(8), resulting either in an increase or decrease in correlation in a manner which may be proportional or inversely proportional to probabilities. Increases in probability will increase or decrease correlation only up to a limit (probability of 0.5) beyond which the derivative of the function and changes in correlation will change sign. We interpret a quadrant occupancy of greater than 50% of the joint distribution, or quadrant probability of greater than 0.5, as a region of asset combination infeasibility for which one of the assets is either so superior or so inferior that it should be taken alone as a portfolio itself or should be excluded from consideration as a portfolio choice. This commonly occurs when the median of one asset's marginal return distribution is extreme; either very positive or very negative.

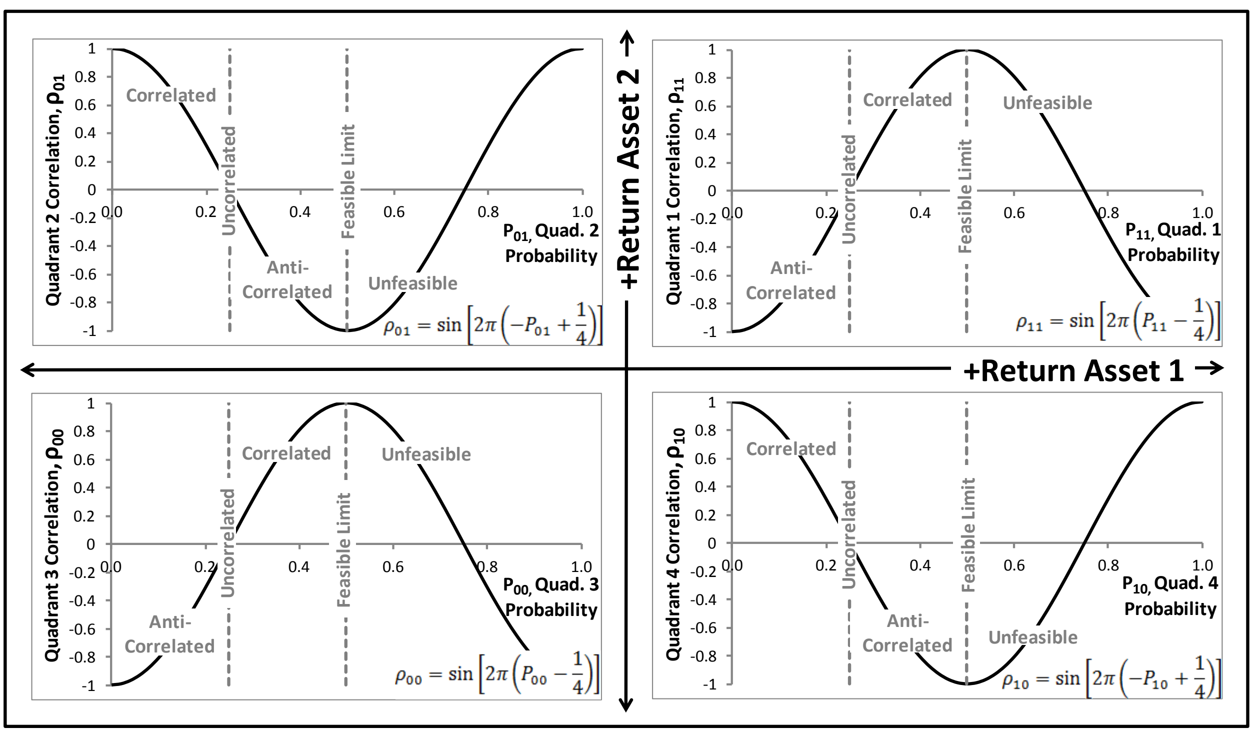

Figure 1: The relation between orthant correlations and orthant probabilities illustrated for the bivariate case of the return distributions of two financial assets, shown as a function of quadrant. A base case neutral condition might be considered that of a bivariate standard normal (mean zero, variance one) distribution for which occupancy in each quadrant would be

25%, translating to a probability of 0.25 and leading to an orthant correlation of zero for all quadrants. Occupancies which differ from this 25% occupancy base-case of occupancy or probability will result in either increases or decreases in correlation, depending on the quadrant involved.

Table 1 presents the results of application of both Pearson's product-moment correlation and the orthant correlations to joint distributions X, Y where X is standard normal and Y is defined variously either as independent from X or defined jointly with X. In the first two lines of the table ($Y = \pm 1$) X is a mean-zero random normal distribution with unit variance, Y is a constant and X and Y are clearly independent. For rows three and four of Table 1 ($Y = \pm X$), X is a standard normal, $N(0,1)$, and X and Y are linearly dependent. For rows five and six of Table 1 ($Y = \pm X^2$) X is a standard normal and X and Y are not linearly dependent, but are non-linearly dependent as defined in the first column. In rows seven and eight, X and Y are both standard normal, $N(0,1)$, with linear correlation between them (gauged by Pearson's product-moment correlation) enforced to be zero (in row seven) and 0.5 (in row eight).

	Pearson's ρ	$\boldsymbol{\rho_{11}}$ (quad. 1)	$\boldsymbol{\rho_{01}}$ (quad. 2)	$\boldsymbol{\rho_{00}}$ (quad. 3)	$\boldsymbol{\rho_{10}}$ (quad. 4)
$\boldsymbol{Y = 1}$	0	1	-1	-1	1
$\boldsymbol{Y = -1}$	0	-1	1	1	-1
$\boldsymbol{Y = X}$	1	1	1	1	1
$\boldsymbol{Y = -X}$	-1	-1	-1	-1	-1
$\boldsymbol{Y = X^2}$	0	1	-1	-1	1
$\boldsymbol{Y = -X^2}$	0	-1	1	1	-1
$\boldsymbol{Y = N(0,1)}$	0	0	0	0	0
$\boldsymbol{Y = N(0,1)}$	0.5	0.5	0.5	0.5	0.5

Table 1: Results of various benchmark cases of dependency between two distributions, X and Y, comparing the Pearson's product-moment correlation coefficient to the four quadrant orthant correlations for the bivariate joint distribution. The ability of orthant correlations to fully capture non-linearities, in comparison with Pearson's product-moment linear correlation, is especially apparent in the classic textbook case of $Y = X^2$ where there is full dependence between X and Y but no linear correlation exists.

For the case of $Y = 1$, from Table 1, Y is fully independent of X and Pearson's product-moment correlation coefficient reflects this with an estimation of zero. However, the orthant correlations consider the relation X, Y in the separate and isolated perspective of each of four quadrants, providing a more microscopic view which can be less readily open to interpretation. In this case, the occupancies or measured probabilities in quadrants one to four ($P_{11}, P_{01}, P_{00}, P_{10}$) are 0.5, 0.5, 0 and 0, respectively. As a result, correlation in the first quadrant, ρ_{11}, is equal to one due to occupancy in that quadrant of 50% of the distribution. Examination of Figure 1 reveals this to be 25% more than would be the case if X, Y were independent (in which case occupancy and probability would be just 25%). The second quadrant has an equally high occupancy (50%), but these samples fall into a category of opposite signed samples ($-X, +Y$) leading, unlike quadrant one, to full anti-correlation for ρ_{01}, as occupation occurs in numbers greater than would be the case for random and independent median zero joint distributions (or 25%). Absence of any observations in quadrant three (corresponding to negative returns for both assets) is interpreted by the orthant relations as full anti-correlation for ρ_{00}. If quadrant three was highly occupied then the conclusion would be the opposite. Likewise in quadrant four, the absence of any observations is interpreted not in terms of what is present, but in terms of what is missing compared to a joint distribution for which the marginal medians are equal to zero (occupancy of 25% or probability of 0.25), and the result is full correlation for ρ_{10}.

The $Y = X$ case of Table 1 results in full correlation in all quadrants which is both logical and in agreement with Pearson's product-moment correlation coefficient.

The continuous function $Y = X^2$ from Table 1 is a textbook example illustrating that linear correlation does not equate with dependence. In this case, full dependence between X and Y is obvious and yet Pearson's product-moment correlation coefficient (capturing only linear dependence) is equal to zero. Quadrant correlations for this case are non-zero, but are also equivalent to quadrant correlation results for the fully independent case of $Y = 1$. Quadrant correlations gauge the linear component of dependence as well as non-linear, and when gauging correlation they make no distinction between functional or distributional forms, even distributions that do not exist or are more exotic in character.

Table 1 also includes two cases for which both X and Y are standard normal distributions (mean zero, variance one) and with Pearson's product-moment correlation enforced for the separate cases of zero and 0.5. Results illustrate that orthant correlations revert to Pearson's product moment correlation coefficient, which fully captures linear dependence, for standard-normal joint distributions. All four quadrant correlations are equal to each other for distributions whose medians are equal to their means, but for bivariate standard normal distributions they are also equal to Pearson's product-moment correlation coefficient. In any other circumstances quadrant correlations may produce results which are close to Pearson's correlation, but will almost never be exactly equal, typically as the result of non-linear dependencies between joint distributions which Pearson's linear correlation does not detect.

In the presence of skewness and kurtosis

Orthant or quadrant-conditional correlations in the presence of imposed skewness and kurtosis and fixed liner correlations are shown in Figure 2. The Vale-Maurelli (1983) implementation of Fleishman's (1978) method was used to generate bivariate joint distributions with pre-specified first through fourth marginal distribution moments, as well as Pearson product-moment linear correlations between the two. A multivariate SAS/IML 6.0 implementation of the Vale-Maurelli (1983) method, designed and incorporated in SAS by Wicklin (2013) was used for this purpose. The first marginal distribution was specified to be standard normal, N(0,1), zero skewness and zero excess kurtosis. In Figure 2a, the second marginal distribution was specified as zero mean, unit variance, excess kurtosis of 10 and varying degrees of skewness. A fixed value of 10 for kurtosis, γ_4, was selected in order to allow a wide breadth of skewness, γ_3, since $\gamma_4 \geq \gamma_3^2 + 1$ (see Stuart and Ord (1994) and Vargo et al. (2010)). In Figure 2b, the second marginal distribution was specified as zero mean, unit variance, zero skewness and varying degrees of Kurtosis. In both cases, joint distributions were pre-specified to have fixed linear Pearson's product-moment correlation coefficients of 0.5 and all pre-specified values were enforced to five significant figures.

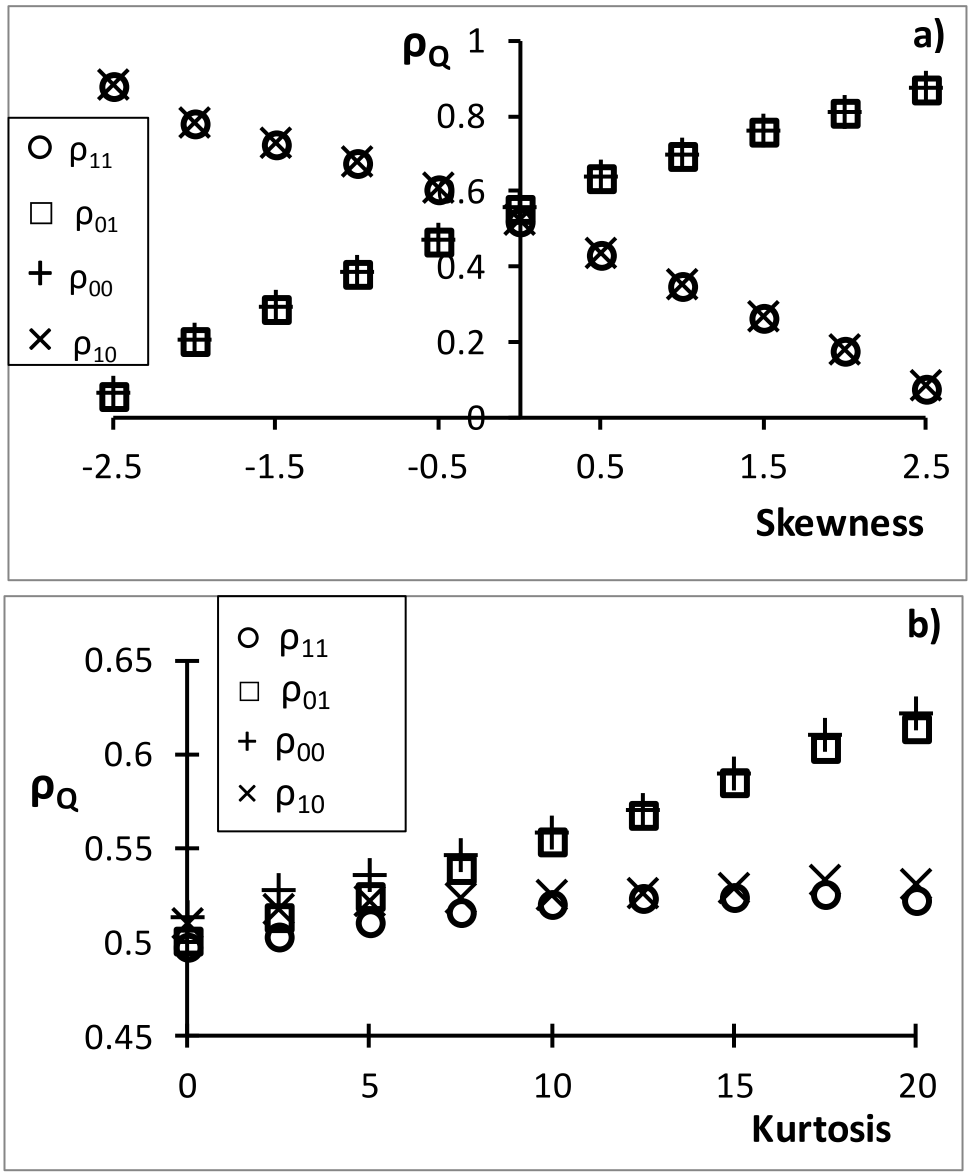

Figure 2: Orthant correlations of the four quadrants of joint bivariate distributions for various levels of non-linearity. For sub-figure a) Orthant correlations are shown as a function of skewness between a standard normal, N(0,1) distribution (with zero skewness and zero excess kurtosis) and a distribution with mean zero, variance of one, excess kurtosis of 10 and skewness varying from -2.5 to 2.5. In all cases, linear correlation as measured by Pearson's product-moment correlation coefficient is pre-specified and fixed for each case of skewness at 0.5. Quadrants two and three correlations, ρ_{01} and ρ_{00}, appear linearly related to skewness and quadrants one and four, ρ_{11} and ρ_{10}, inversely, linearly related. For sub-figure b) Orthant correlations are shown between a standard normal, N(0,1) distribution (with zero skewness and zero excess kurtosis) and a distribution with mean zero, variance of one, skewness of zero and various levels of excess kurtosis between 0 and 20. In all cases, linear correlation as measured by Pearson's product-moment correlation coefficient is fixed to 0.5.

Orthant correlation's quadrant sensitivity to skewness in Figure 2a is significant, with first and fourth quadrants correlations being inversely related and second and third quadrants being directly related. Direct sensitivity to kurtosis is present in Figure 2b for all four quadrants, but appears to be of a lower magnitude compared to skewness for the regions of skewness and kurtosis which were tested. We note that the sensitivities of orthant correlations to both skewness and kurtosis are likely also dependent on a number of other factors such as the distribution first and second moments and the linear dependence (in these tests imposed at 0.5) of joint distributions.

Quantifying the complementarity of asset characteristics

The absolute and relative benefit of holding assets is primarily apparent through their risk and return characteristics. A secondary and joint characteristic is the correlation between assets held, or the extent to which their variability is related. A commonly employed method for estimation of correlation is through Pearson's product-moment correlation coefficient, ρ, which can be estimated for a bivariate sample by Pearson's r. Pearson's product-moment correlation coefficient gauges only the part of variability in one asset which is linearly dependent on the variability of another. In the absence of non-linearities, or even in the presence of low levels of non-linearities, typically made apparent through distributional skewness and kurtosis, Pearson's product-moment correlation is a practical and powerful metric for optimizing portfolio "fit" between assets. Unfortunately, linearity is rarely the case with financial return distributions, which more often than not exhibit skewness and excess kurtosis. In such cases any non-linear co-variability will go unquantified by Pearson's product-moment correlation, leading to underestimation of dependence. Orthant correlations present at least two advantages: They are agnostic to distributional form and, in addition, may present an opportunity for gauging co-variability on a more microscopic scale. The exploitation of both of these advantages may find applications in portfolio construction and higher frequencies of arbitrage securities trading.

Figure 1 provides one possible starting point for the practical application of orthant correlations and its examination in the context of what is beneficial or detrimental to the goals of a practical application can act as a guide. From the perspective of a simple two-asset portfolio, occupancy of greater than 25% (the base case of a random-normal bivariate with zero median) of joint distributions is typically judged as desirable for the first quadrant of their joint distribution. Occupancy in the first quadrant, P_{11}, reflects positive returns in both assets and Figure 1 indicates that an occupancy of greater than 25% necessarily involves an increase of the correlation ρ_{11} between the two assets. For the purposes of this exercise, further making an assumption that the first asset is a fixed decision (for example, a previously existing, long-term portfolio investment whose sale is undesirable), higher occupancy in quadrant two, P_{01}, is also a desirable outcome; the second asset return, to some extent, compensating for the negative return of the first fixed asset. For quadrant two, such an increase in occupancy is linked with a decrease in the quadrant correlation ρ_{01} (see quadrant two of Figure 1). Occupancy of quadrant three, P_{00}, is the least desirable result (both assets delivering negative returns); a decrease below a probability of 25% is desirable and quadrant three of Figure 1 indicates that such a shift in occupancy and probability decrease would be linked with a decrease in the correlation, ρ_{00}, for that particular quadrant. Again assuming that holding asset one is a given constraint, occupancy in quadrant four of Figure 1 would by preference be lower than 25% and a decrease in occupancy and probability in quadrant four, P_{10}, is linked to an increase in the correlation,

ρ_{10}, for that quadrant. Differing arguments and conclusions might be made for differing applications and perspectives, but the conclusions of these particular arguments are summarized in Table 2.

Quadrant	**Desirable Probability Change From $P_Q = 0.25$**	**Resulting Correlation Change From $\rho_Q = 0$**
Q1 ($R_{Asset\,1} > 0, R_{Asset\,2} > 0$)	Increase	Increase
Q2 ($R_{Asset\,1} \leq 0, R_{Asset\,2} > 0$)	Increase	Decrease
Q3 ($R_{Asset\,1} \leq 0, R_{Asset\,2} \leq 0$)	Decrease	Decrease
Q4 ($R_{Asset\,1} > 0, R_{Asset\,2} \leq 0$)	Decrease	Increase

Table 2: Summary of assessment of quadrant orthant correlation desirable changes for the particular case of a two asset portfolio for which holding the first asset is a fixed assumption and the second asset is a candidate for portfolio addition. The rationale for "desirability" of change is likely dependent on the particular application in mind and the rationale for this particular case is described in the text.

Conclusions from Table 2 might be quantified in a simple linear combination of orthant correlations as in Equation (9).

$$\Delta\rho_Q = \rho_{11} - \rho_{01} - \rho_{00} + \rho_{10} \quad . \tag{9}$$

As the quadrant correlations are bounded by +/-1, $\Delta\rho_Q$will be bounded by +-4, or $\Delta\rho_Q$ might be divided by four, making it bounded by +/-1 for greater ease of interpretation. Figure 3 provides insight into the sensitivity of $\Delta\rho_Q$ defined by Equation (9) to skewness and kurtosis. Returns of a first asset are distributed as a standard normal with Pearson's product-moment correlation enforced to be 0.5 with the second asset. In Figure 3a the second asset has a fixed excess kurtosis of 10, mean zero, variance of one and varying degrees of skewness. The inverse linear relation between $\Delta\rho_Q$ and skewness indicates that a negatively skewed second asset would be a more optimal portfolio combination, at least with an asset whose returns are distributed as a standard normal. In Figure 3b, returns of the second asset also have mean zero, variance one and an enforced linear correlation of 0.5 with the first asset, but zero skewness and varying degrees of kurtosis, revealing a non-linear and strictly negative relation between $\Delta\rho_Q$ and positive kurtosis.

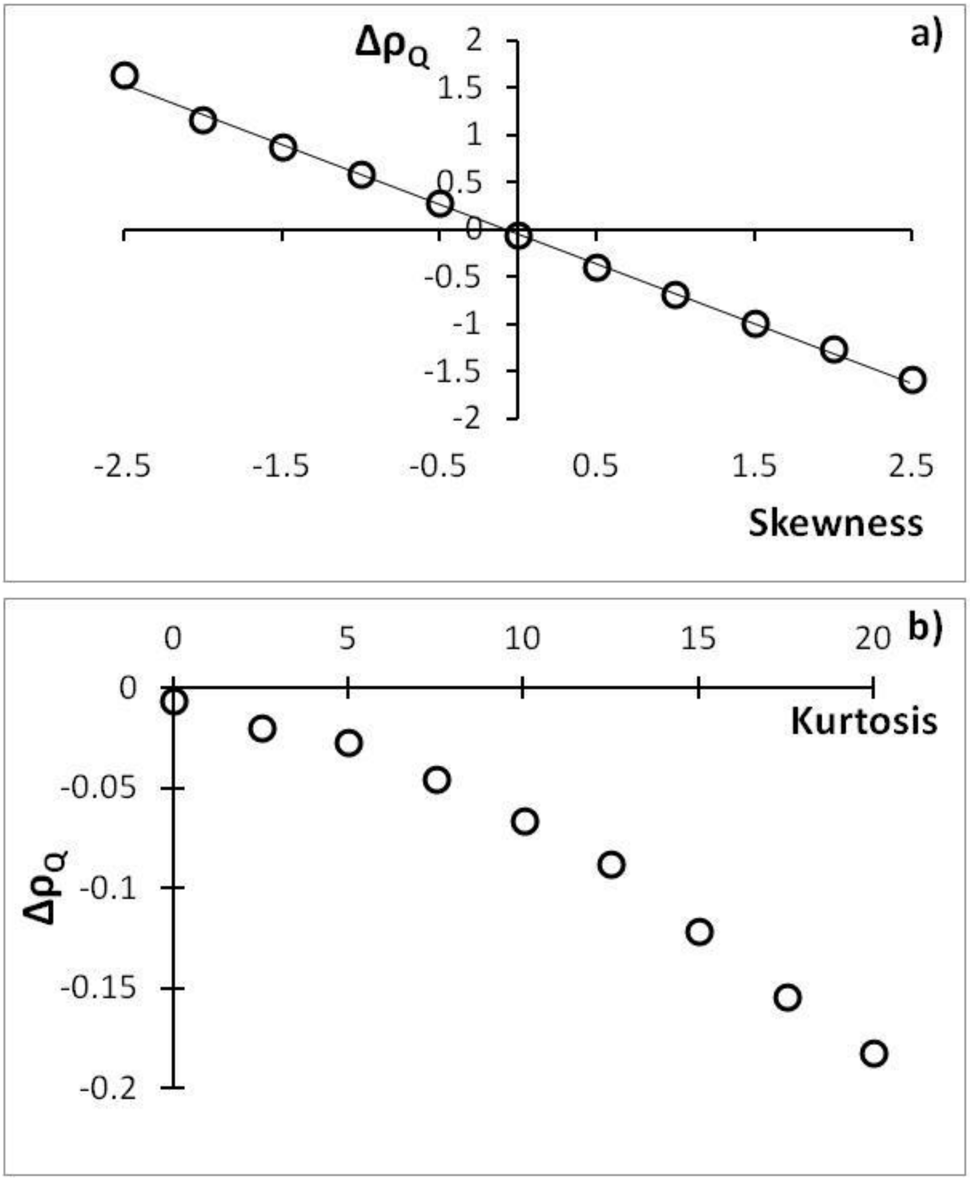

Figure 3: Delta-rho-Q as defined by Equation (9), as a function of excess kurtosis and skewness for the case of a bivariate joint distribution of assets' returns in which the first asset is as standard normal distribution with mean zero, unit variance, zero skewness and zero kurtosis. In all cases linear correlation as measured by the Pearson's product-moment correlation coefficient was fixed to 0.5. a) The second distribution has zero mean, unit variance and excess kurtosis fixed at 10 while skewness is varied. b) The second distribution as zero mean, unit variance and skewness fixed at zero while excess kurtosis is varied.

Figure 4 shows $\Delta\rho_Q$ as a function of mean return for a portfolio of two assets, along with a best fit line which gauges the linear dependence between the two. As part of a Monte Carlo study, 100,000 random joint bivariate events were generated for two standard normal, N(0,1), distributions and linear correlation (Pearson's r) was enforced to be 0.5. The joint return history was then sub-divided into 1,000 non-overlapping subsets of 100 samples each. Means (=0), variances (=1), skewness (=0), excess kurtosis (=0) for both distributions, and Pearson's product-moment correlation (=0.5) were forced to a precision of five significant figures for each subset of data addressed separately. The mean of the sample distribution of 1,000 equally weighted portfolio combination assets' returns was 1×10^{-17} and its standard deviation was 0.08548. The mean value of $\Delta\rho_Q{}'s$ was -0.0066 with a standard deviation of 0.5170. However, division of the total population into 1,000 smaller groups of 100 observations each allows for sampling fluctuations to occur and these lead to local variations in linear correlation from 0.5, non-zero portfolio returns and non-zero $\Delta\rho_Q$ values for the subsamples, even while the full sample of data has distribution mean value and $\Delta\rho_Q$ equivalent to zero and linear correlation of 0.5 to within five significant figures.

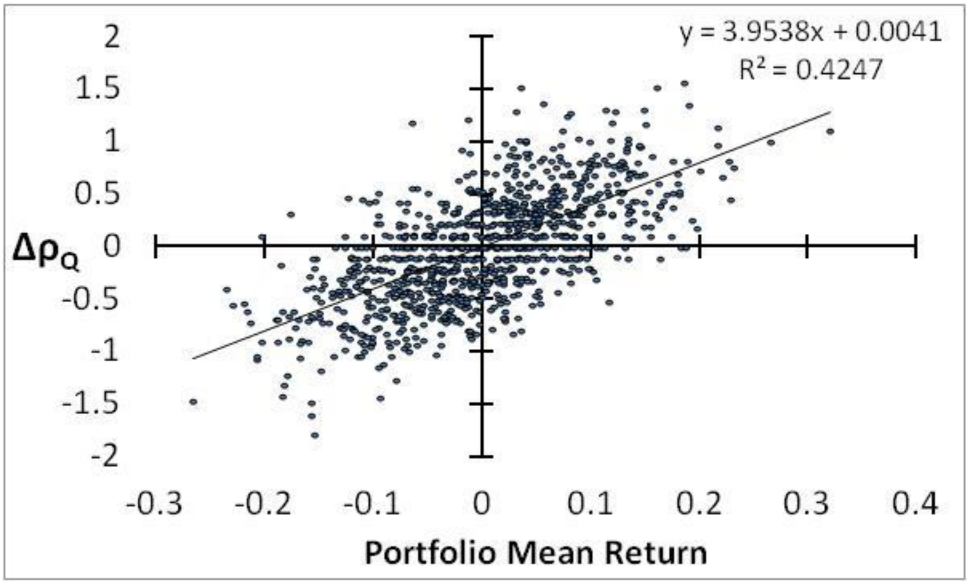

Figure 4: Delta-rho-Q as defined by Equation (9), as a function of the mean value of portfolio returns for an equally weighted combination of assets. Each of the 1,000 Individual data points consists of a non-overlapping subset of 100 joint samples from a bivariate normal, N(0,1), distribution (with zero skewness and zero excess kurtosis) and linear correlation (as measured by Pearson's product-moment correlation coefficient) was fixed to 0.5 for the entire sample of 100,000 random joint samples. According to a least-squares best fit to a line (shown in the figure) and also as a result of t-tests, there are reliable indications of a significant relation between delta-rho-Q and portfolio returns.

Figure 4 shows the relation between $\Delta\rho_Q$ and finite period portfolio returns. A linear fit to the data indicates a strong linear dependence between the two ($R^2 = 0.4247$, or Pearson's linear correlation of 0.65). The average of portfolio returns jointly occurring with $\Delta\rho_Q$>0 was 0.0509, whereas the average of portfolio returns jointly occurring with $\Delta\rho_Q < 0$ was -0.0485. Assuming that $\Delta\rho_Q$for a future investment period is known, it is theoretically possible to use the orthant correlation information of $\Delta\rho_Q$as a return-enhancing information source. For example, a simple investment strategy of holding cash conditional on $\Delta\rho_Q < 0$, and holding an equal combination of the two assets conditional on $\Delta\rho_Q > 0$ would produce positive portfolio returns over the long-term, even though assets' returns over the long-term would be zero. Going further, a simple investment strategy of short selling both assets for $\Delta\rho_Q < 0$, and holding an equal combination of the two assets when $\Delta\rho_Q > 0$ would produce an even more positive return result. For a limited time interval, it is not unusual for an asset combination to produce a weighted average portfolio combination return which is greater than either of the two assets. If the returns of two equally weighted assets are distributed as uncorrelated random standard normals, this would occur in 25% of joint samples. It is this level of probability, 25%, which has been referred to previously in this discussion as a "base case" or neutral condition. Unfortunately, a separate 25% of samples would also result in the opposite condition of a portfolio return more negative than either asset and at issue is whether or not the long-term average of weighted returns can be systematically improved in comparison with the long-term average of weighted assets' distributions expectation values, conditional on information related to assets co-variability. This might be considered unusual, and we note that systematically positive deviation of portfolio return from the weighted average of assets' returns only occurs in the presence of future return information which can be employed to guide either relative asset selection in a context of strategic asset allocation or tactical asset weighting in a context of active management.

T-tests were performed in an attempt to gauge the confidence with which it can be concluded that the mean or expected values of distributions' portfolio mean returns as shown in Figure 4, but conditional on $\Delta\rho_Q < 0$ and $\Delta\rho_Q > 0$, actually differ. Both pooled and unpooled t-tests were performed on the negative, $Z_{1,i} = 0.5 \times (\overline{X_\iota} + \overline{Y_\iota})|(\Delta\rho_Q < 0)$, and positive, $Z_{2,i} = 0.5 \times (\overline{X_\iota} + \overline{Y_\iota})|(\Delta\rho_Q > 0)$, conditional distributions of mean return values of the $i = 1..1{,}000$ subsets of portfolio combinations of asset one and asset two return distributions, X and Y. Subset portfolio mean returns were composed of equally weighted asset returns which is equivalent to an equally weighted average of their means, $\overline{X_\iota}$ and $\overline{Y_\iota}$. Due to the fact that a limited sample of 100 values were used for each of the 1,000 data points in Figure 4, some granularity is apparent in the figure at values of $\Delta\rho_Q$ closer to zero. For the same reason, 71 out of 1,000 data points actually resulted in $\Delta\rho_Q = 0$. Of the remaining 929 non-zero values, 477 out of 1,000 samples in Figure 4 had $\Delta\rho_Q < 0$ and 452 samples had $\Delta\rho_Q > 0$. The underlying assumption of t-tests on the conditional distributions, Z_1 and Z_2, was that the means of the two distributions are the same, or compatible. A pooled Sasabuchi (1988a, 1988b) t-test (which assumed equal variances of the two distributions) resulted in a t-value of 21.23 (with 927 degrees of freedom) and corresponding p-value of 0.0001. The t-value is related to the size of the difference between the means for the two samples under comparison and the p-value is the probability of obtaining a t-value at this extreme or greater under the assumption of equivalent means. An unpooled Satterthwaite (1946) style t-test (which assumes unequal variances) was performed and resulted in a t-value of 21.22 (with 922.04 degrees of freedom) and corresponding p-value of 0.0001. Finally a Cochran and Cox (1950) style unpooled t-test (assuming unequal variances) was performed and resulted in a t-value of 21.22 and corresponding p-value of 0.0001. The number of degrees of freedom for the Cochran and Cox t-test is undefined when the number of samples from each pair is unequal, which was the case here. Separately, a Steel and Torrie (1980) equality of variances test was performed and revealed a two-tailed F-value of 1.04 with a probability to be greater than F of 0.6748, which we interpret as an indication that there is insufficient evidence of unequal variances between Z_1 and Z_2. Examination of skewness and kurtosis in Z_1 and Z_2 revealed only slight non-normality and this was confirmed by visual examination of Q-Q plots of data quantiles versus normal distribution theoretical quantiles. All three t-tests resulted in highly significant p-values, supporting the conclusion of a significant difference between the means for the return distributions conditional about zero on $\Delta\rho_Q$.

There is a strong assumption behind the conclusion that quadrant correlations can be employed to produce a long-term average investment return which is greater than any of the individual asset long-term averages of returns. The assumption is that there exists perfect, or more realistically significant, knowledge of the quadrant correlations of investment sub-periods before those periods begin. The ability to forecast quadrant correlations, for example by taking recent historical values as forecasts of the near-term future values, remains for us an open question. However, we do surmise that quadrant correlation values, as we presume is the case for assets variance and covariance, to which correlation is closely related, are more persistent and predictable than future security performance or mean return expectation values if only based on the observation that the former tends to vary much less (and more slowly) than the latter.

For comparison, Pearson's product-moment correlation coefficient, r, was estimated for the same subsets of simulation data from Figure 4 and results are shown as a function of portfolio subset average return in Figure 5. The average of portfolio returns under the condition $r > 0.5$ was 0.0005 and for $r < 0.5$ was -0.0006, providing some indication that little or no portfolio return information is reflected by linear correlation. Figure 5 also shows the results of an unsuccessful attempt to perform a least-squares regression fit to a straight line and confirms that no significant, direct relation presents itself. The source of the relatively strong relation between portfolio returns and orthant correlations, as quantified by $\Delta\rho_Q$ and as shown in Figure 4, likely stems from the sensitivity of orthant correlations to non-linearities, or from the more granular information afforded by quadrant-dependent correlations, or both sources combined.

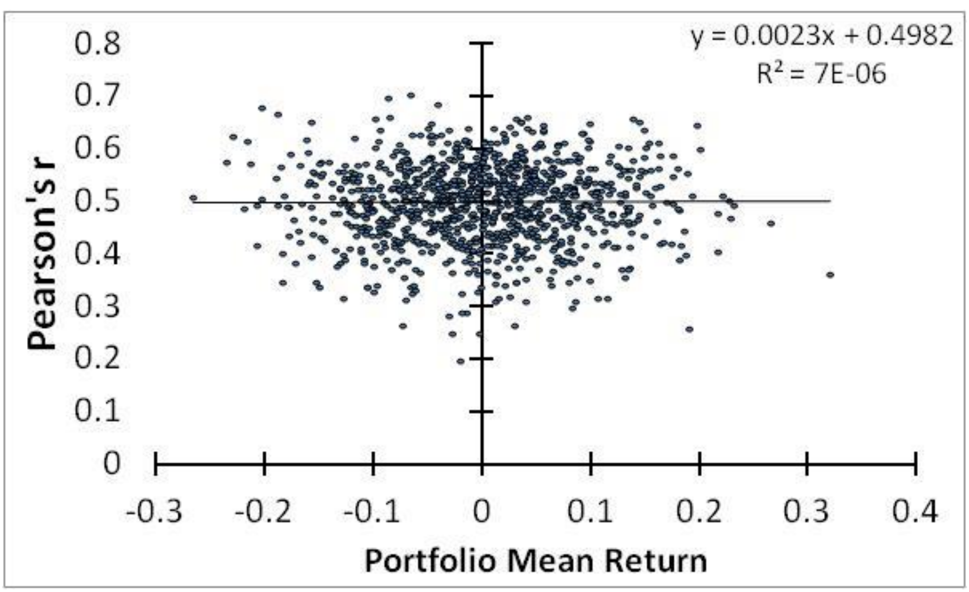

Figure 5: Linear correlation as measured by Pearson's product-moment correlation coefficient, as a function of the mean value of portfolio returns for an equally weighted combination of assets. Each of the 1,000 Individual data points consists of a non-overlapping subset of 100 joint samples from a bivariate normal, N(0,1), distribution (with zero skewness and zero excess kurtosis) and linear correlation (as measured by Pearson's product-moment correlation coefficient) was fixed to 0.5 for the entire sample of 100,000 random joint samples. Both by eye and according to a least-squares best fit to a line (shown in the figure) there is no indication of a relation between Pearson's r and portfolio returns.

Conclusions

Orthant probabilities have been applied in order to gauge correlation of financial assets' returns in a manner which is robust to the non-linearities which often manifest themselves through skewness and kurtosis, while also opening a more microscopic view on co-variability under differing (for example rising versus falling) market conditions. The case of a simple, two-asset investment portfolio has been used to demonstrate the potential for application of orthant correlations in investment management. A strategy involving investment in the two assets simulated for the results shown in Figure 4 when $\Delta\rho_Q > 0$, but remaining in cash otherwise, produced a long-run portfolio return greater than zero, with a convincing level of confidence, while both the long-run averages of assets' returns, and their weighted average, were zero. This result could only be achieved through application of new information which directly explains assets' returns, information which is not apparent with the application of Pearson's product-moment correlation. We preliminarily conclude that the upper limit for the potential of this method, paradoxically based in risk management, is to systematically deliver higher long-term portfolio returns than the long-term returns of individual assets themselves taken separately.

In demonstrating this upper limit potential, a strong assumption involving knowledge of future orthant correlations was made, but it can be argued that the forecastability of correlation is at least as tenable as, or even more easily attainable than, the forecasting of asset returns which in practice is typically implicit (either quantitatively or qualitatively) in asset selection. This discussion was limited to the two-dimensional framework of quadrants and we see no closed form solution involving more than four orthants. Therefore, we note that more work is likely required in order to achieve a multivariate extrapolation to larger portfolios. We view orthant correlation as a unique quantifier with a potential for portfolio construction which is complementary to return expectation value analysis.

Some effort has been made to illustrate the character of correlations, particularly in the presence of skewness and excess kurtosis. We view these efforts as a starting point in research and development and future study is called for. It is noted that much of our current Monte Carlo simulation has been performed under a forced linear correlation value of 0.5 (as gauged by Pearson's product-moment correlation); orthant correlation in the full correlation space merits further investigation. This is particularly relevant for the case of negative correlation which presumably exists with high prevalence in long-short security combinations. Likewise, the simple orthant correlations combination, $\Delta\rho_Q$, as defined in Equation (9), is viewed as one of many possible. We currently perceive that the manner in which orthant correlations are most optimally applied will be dependent on the application in mind.

In this discussion, we have addressed the case of a fixed investment portfolio in the context of a search for a new portfolio candidate asset. In one straightforward scenario, quantifiers such as the $\Delta\rho_Q$ defined here, when estimated separately for various candidates, have a potential for adding value in a comparison-based asset selection process. In addition to asset selection and portfolio construction, we also note the potential for application of orthant correlations in the realm of high frequency arbitrage trading where bivariate or multivariate combinations of highly liquid assets might be selected or weighted according to quantifiers related to $\Delta\rho_Q$ in order to produce a systematic positive return largely or fully unrelated to security return analysis.

Appendix A

Proof of application of Sheppard's theorem to the bivariate elliptical

There exist elegant and simple specific profs of Sheppard's theorem based on the geometry of the circle; we shall present a more complex argument which has the definite benefit of allowing us to consider more general distributions. As in the main text, we shall use the notation:

$$P_{00} = Prob(X \leq 0, Y \leq 0) \quad ,$$
$$P_{11} = Prob(X > 0, Y > 0) \quad ,$$
$$P_{01} = Prob(X \leq 0, Y > 0) \quad ,$$
$$P_{10} = Prob(X > 0, Y \leq 0) \quad .$$

It follows that:

$$P_{00} + P_{01} + P_{10} + P_{11} = 1 \quad .$$

If the probability density function, $pdf(x, y)$, is symmetric around $(0,0)$, then $P_{00} = P_{11}$, $P_{10} = P_{01}$ and $P_{00} + P_{10} = \frac{1}{2}$.

Define :

$$sgn(x) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{e^{itx} dt}{it}$$
$$= 1, \ if\ x > 0$$
$$= 0, \ if\ x = 0$$
$$= -1, \ if\ x < 0 \quad .$$

For proof see Kendall (1975).

Also, let $S = E(sgn(x) \cdot sgn(y))$

$$= P_{00} + P_{11} - P_{10} - P_{01} \quad .$$

Under symmetry about (0,0),

$$S = 2P_{00} - 2P_{10} \quad ,$$

so, $P_{00} = \frac{1}{4}(1 + S)$ and

$$P_{01} = \frac{1}{4}(1 - S) \quad . \qquad \text{(A.1)}$$

In order to compute

$$S = \frac{1}{\pi^2} \int_{-\infty}^{+\infty} \frac{dt_1}{it_1} \int_{-\infty}^{+\infty} \frac{dt_2}{it_2} \varphi(t_1, t_2),$$

where $\varphi(t_1, t_2) = E\left(e^{it_1x+it_2y}\right)$,

let $\binom{x}{y} \sim N\left(\binom{0}{0}\begin{pmatrix}1 & \rho \\ \rho & 1\end{pmatrix}\right)$, then

$\varphi(t_1, t_2) = E\left(e^{-1/2(t_1^2+2\rho t_1 t_2+t_2^2)}\right)$.

(We note that, in the main body of the text we use the notation $\rho_{X,Y}$ in order to call attention to a correlation which may be conditional on X and Y. For the sake of better ease of reading, this notation has been dropped in the context of this proof as it serves no functional purpose.)

Hence

$$S = \frac{1}{\pi^2}\int_{-\infty}^{+\infty}\frac{dt_1}{it_1}\int_{-\infty}^{+\infty}\frac{dt_2}{it_2}e^{-1/2(t_1^2+2\rho t_1 t_2+t_2^2)}. \quad \text{(A.2)}$$

Now

$$\frac{\partial S(\rho)}{\partial \rho} = \frac{1}{\pi^2}\iint_{-\infty}^{+\infty} e^{-1/2(t_1^2+2\rho t_1 t_2+t_2^2)}\,\partial t_1 \partial t_2,$$

and $S(0) = 0$, since $P_{00} = P_{11} = P_{01} = P_{10} = \frac{1}{4}$.

Also, from the properties of the normal distribution:

$S'(\rho) = \frac{2}{\pi\sqrt{1-\rho^2}}$,

so integrating w.r.t. ρ gives:

$$S(\rho) = \frac{2\sin^{-1}(\rho)}{\pi} , \quad \text{(A.3)}$$

since $\frac{\partial \sin^{-1}\rho}{\partial \rho} = \frac{1}{\sqrt{1-\rho^2}}$.

Thus $P_{00} = \frac{1}{4}\left(1 + \frac{2}{\pi}\sin^{-1}(\rho)\right)$

$= \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(\rho)$.

And $P_{01} = \frac{1}{4} - \frac{1}{2\pi}\sin^{-1}(\rho)$ from $(A.1)$ and $(A.3)$.

Let τ be our positive stochastic scale factor which we condition on. It is intuitively clear that:

If the conditional p.d.f is denoted $\left(\begin{matrix}x\\y\end{matrix}\middle|\tau\right) \sim N\left(0, \tau\begin{pmatrix}1 & \rho \\ \rho & 1\end{pmatrix}\right)$, (A.4)

then: $Prob(X > 0, y > 0|\tau) = \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(\rho)$

$\therefore Prob(X > 0, y > 0) = \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(\rho)$.

This can be further demonstrated by going through the steps from (A.2) onwards where (A.2) becomes :

$$S = \frac{1}{\pi^2}\int_{-\infty}^{+\infty}\frac{\partial t_1}{it_1}\int_{-\infty}^{+\infty}\frac{\partial t_2}{it_2}e^{-\tau/2(t_1^2+2\rho t_1 t_2+t_2^2)}.$$

Extensions

We have shown that Sheppard's theorem holds for a subset of the zero mean elliptical family described by (A.4). We now consider potential extensions.

Consider a mixture model for which, after standardizing data, we can write the probability density function as:

$$pdf(x, y) = \sum_{j=1}^{m} a_j N\left(\binom{0}{0}, \Omega_j\right) \quad ,$$

Where $a_j \geq 0$, $\sum_{j=1}^{m} a_j = 1$, and $\Omega_j = \begin{pmatrix} 1 & \rho_j \\ \rho_j & 1 \end{pmatrix}$ and $N\left(\binom{0}{0}, \Omega_j\right)$, means distributed as bivariate normal with mean $\binom{0}{0}$ and covariance matrix, Ω_j.

This assumes that all of the m subpopulations have the same mean vector and same variances, but that they differ in correlation, thus they lie outside of the elliptical family. Now their characteristic function is:

$$\varphi(t_1, t_2) = \sum_{j=1}^{m} a_j e^{-1/2\left(t_1^2+2\rho_j t_1 t_2+t_2^2\right)} \quad , \qquad \text{(A.5)}$$

and computing S as before and differentiating w.r.t $\rho_1 \cdots \rho_m$ we find that:

$$\frac{\partial^m S(\rho_1 \cdots \rho_m)}{\partial\rho_1 \cdots \partial\rho_m} = \sum_{j=1}^{m}\left(\frac{2a_j}{\pi\sqrt{1-\rho_j^2}}\right).$$

Nothing that $S(0,0 \cdots 0)$=0 and integrating over $\rho_1 \cdots \rho_m$ we see that:

$$S(\rho_1 \cdots \rho_m) = 2\sum_{j=1}^{m}\frac{a_j \sin^{-1}(\rho_j)}{\pi}$$

and

$$P_{00} = \frac{1}{4} + \frac{1}{2\pi}\sum_{j=1}^{m} a_j \sin^{-1}(\rho_j) \quad ,$$

$$P_{01} = \frac{1}{4} - \frac{1}{2\pi}\sum_{j=1}^{m} a_j \sin^{-1}(\rho_j) \quad .$$

As before, if $\rho_j \to 1$ for $j = 1 \cdots m$, then $\sin^{-1}(\rho_j) \to \frac{\pi}{4}$.

Thus, since $\sum_{j=1}^{m} a_j$=1, $P_{00} = \frac{1}{2}$ and $P_{01} = 0$

It is also clear that this can be extended to consider cases where each subpopulation satisfies assumption (A.4); this now covers a fairly large family of distributions whose orthant probabilities can be computed explicitly. Also, since the characteristic function (A.5) is real, it means that the probability density function is symmetric about $(0,0)$ and hence its medians and means are both zero.

Acknowledgements

The authors would like to extend thanks to Professor of Econometrics Theo Dijkstra, University of Groningen Faculty of Economics and Business, for useful insights, comments and discussion. The authors also thank Della Hieb of the Arizona Public Safety Personnel Retirement System for editorial assistance.

References

Cochran, W.G. and Cox, G.M. (1950), *Experimental Designs*, New York: John Wiley & Sons.
Fleishman, A.I. (1978), A method for simulating non-normal distributions, Psychometrika, **43**, 51-532.
Kendall, M.G. (1975), *Rank Correlation Methods*, 4th ed, 2nd impression, Charles Griffin and Company, London & High Wycombe, p. 137.
Markowitz, H. (1952), Portfolio Selection, Journal of Finance, **7**(1), March, pp. 77-91.
Pearson, K. (1896), "Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia" Philos. Trans. Royal Soc. London Ser. A , **187**, pp. 253–318.
Sasabuchi, S. (1988a), "A Multivariate Test with Composite Hypotheses Determined by Linear Inequalities When the Covariance Matrix Has an Unknown Scale Factor", Memoirs of the Faculty of Science, Kyushu University, Series, A, **42**, 9-19.
Sasabuchi, S. (1988b), "A Multivariate Test with Composite Hypotheses When the Covariance Matrix is Completely Unknown", Memoirs of the Faculty of Science, Kyushu University, Series, A, **42**, 37-46.
Satterthwaite, F.W. (1946), "An Approximate Distribution of Estimates of Variance Components", Biometric Bulletin, **2**, 110-114.
Sheppard, W. (1898), "On the Geometrical Treatment of the 'Normal Curve' of Statistics, with Especial Reference to Correlation and to the Theory of Error." Proc. Roy. Soc. London, vol lxii, pp. 171-173.
Sheppard, W. (1899) "On the application of the theory of error to cases of normal distributions and normal correlations." Phil. Trans., vol. xcxii A, pp. 101-167.
Steel, R.G.D. and Torrie, J.H. (1980), *Principles and Procedures of Statistics*, Second Edition, New York: McGraw-Hill.
Stuart, A. and Ord, K. (1994), *Kendall's Advanced Theory of Statistics*, Vol. 1. New York: Hodder Arnold.

Vale, C.D. and Maurelli, V.A. (1983), “Simulating multivariate non-normal distributions”, Psychometrika, **48**, 465-471.
Vargo, E., Pasupathy, R. and Leemis, L. (2010), Journal of Quality Technology, **42**, Num. 3, July, 2010, 276-286.
Wicklin, R. (2013), *Simulating Data with SAS*, Cary: SAS Institute.