DOI:
10.1109/CEEC.2016.7835896

Document Version
Peer reviewed version

The Use of Biweight Mid Correlation to Improve Graph Based Portfolio Construction

Patrick Veenstra
Department of Informatics
King's College London
Strand, London WC2R 2LS
Email: patrick.veenstra@kcl.ac.uk

Colin Cooper
Department of Informatics
King's College London
Strand, London WC2R 2LS
Email: colin.cooper@kcl.ac.uk

Steve Phelps
Department of Informatics
King's College London
Strand, London WC2R 2LS
Email: steve.phelps@kcl.ac.uk

***Abstract*—An analysis of the correlation between the returns of different securities is of fundamental importance in many areas of finance, such as portfolio optimisation. The most commonly used measure of correlation is the Pearson correlation coefficient; however, this suffers from several problems when applied to data from the real world. We propose an alternative estimator — the Biweight Mid Correlation (Bicor) — as a more robust measure for capturing the relationship between returns. We systematically evaluate Bicor empirically using data from the FTSE 100 constituents, and show that it is more robust when compared with the Pearson correlation coefficient. Finally, we demonstrate that Bicor can be used to improve a graph-based method of portfolio construction. Specifically, we show that when treating the correlation matrix as an adjacency matrix for a graph and using graph centrality to construct portfolios, the use of Bicor leads to better performing portfolios.**

I. Introduction

There has been a significant body of work studying the cross correlation structure of stock returns, and looking at extracting useful information from it using graph-theoretic methods. One of the earliest such papers [1], investigates the hierarchical structure of the minimum spanning tree (MST) calculated from the Pearson correlation matrix between the log returns of stocks within the Dow Jones Industrial Average portfolio.

The procedure that was introduced can be described as follows. Given N stocks and a time window w containing T periods, we define the return vector $\mathbf{r}_{iw}$ of a stock $i \in 1..N$ to be the vector of log returns for that stock in time window w. The t^{th} element in $\mathbf{r}_{iw}$ is $r_{iw}(t) = \ln P_{iw}(t) - \ln P_{iw}(t-1)$, where $P_{iw}(t)$ is the price of asset i at time period t in window w, and $t \in 1..T$.

For any pair of stocks $i, j \in 1..N$, the Pearson correlation between their log returns in the period $1..T$ in window w is calculated as:

$$\rho_w(i,j) = \frac{\sum_{t=1}^{T}(r_{iw}(t) - \bar{\mathbf{r}}_{iw})(r_{jw}(t) - \bar{\mathbf{r}}_{jw})}{\sqrt{\sum_{t=1}^{T}(r_{iw}(t) - \bar{\mathbf{r}}_{iw})^2}\sqrt{\sum_{t=1}^{T}(r_{jw}(t) - \bar{\mathbf{r}}_{jw})^2}} \tag{1}$$

Where $\bar{\mathbf{r}}_{iw}$ is the sample mean for $\mathbf{r}_{iw}$, that is $\bar{\mathbf{r}}_{iw} = (1/T)\sum_{t=1}^{T} r_{iw}(t)$. If this is calculated for every pair $i, j \in 1..N$, then the $N \times N$ correlation matrix $\mathbf{A}^{(w)}$ for window w can be constructed, with elements $A_{ij}^{(w)} = \rho_w(i,j)$.

The correlation matrix can easily be transformed into an adjacency matrix $\mathbf{A}^{(w)'}$ by setting the diagonal to zero:

$$A_{ij}^{(w)'} = \begin{cases} A_{ij}^{(w)} & i \neq j \\ 0 & i = j \end{cases} \tag{2}$$

This then allows us to interpret the correlations as the edge weights of an undirected *graph*.

The goal of filtering this complete graph is to extract or identify the key structure of underlying correlations in the system. This filtering is sometimes performed on the distance matrix instead of the correlation matrix. The distance matrix $\mathbf{D}^{(w)}$ is typically constructed from the correlation matrix $\mathbf{A}^{(w)}$ by setting $D_{ij}^{(w)} = \sqrt{2(1 - A_{ij}^{(w)})}$.

The filtering and analysis of these graphs have been used to improve portfolio optimisation [2] and centrality based portfolio selection [3]. The evolution of the graphs through market crashes have shown how topological properties change before, during and after such a crash [4].

The minimum spanning tree (MST) is one of the earlier methods of filtering the distance matrix [1]. A tree is a connected undirected graph with no cycles. If G is the graph created from the distance matrix, then a spanning tree T of G is a tree that includes all the vertices of G. The minimum spanning tree of G is the spanning tree that has the smallest sum of edge weights included in the tree. The planar maximally filtered graph (PMFG) is defined to be the graph constructed by connecting the most highly correlated stocks under the topological constraint of keeping the resulting graph planar [5], [6]. This is shown to be a less aggressive filtering compared to the MST and to contain more information, while retaining the hierarchical structure of the MST. Other studies [7] filter $\mathbf{A}^{(w)}$ by applying a threshold τ such that correlation values below this value are removed.

A natural next step is to look at the evolution over time of the correlation matrix and filtered versions thereof, as shown by [8]. In the notation introduced in this paper, that involves looking at the evolution of $\mathbf{A}^{(w)}$ for consecutive windows w. Given N stocks, each with k periods of data, we can split those k periods into M windows, each of size $T \leq k$. We can create multiple windows with a difference of δT periods between them. Note that if $\delta T < T$, then there is some overlap

between consecutive windows. The number of windows M will depend on the choice of T and δT and the total periods k available in the data.

This paper is structured as follows. In section II we summarise potential problems that could arise from using Pearson correlation to calculate the correlations between stock returns and demonstrate empirically how prevalent such issues are. In section III we define Biweight Mid Correlation (Bicor) as a more robust measure of correlation. The definition of Bicor involves a constant K which is ordinarily set to $K = 9$ as that has been shown to exhibit high efficiency, in general, across a variety of distributions. In section III-B we show the reasoning for this choice and in section IV we demonstrate that some different values of K provide further improvements in efficiency when looking at stock return data specifically. In section V we present the improvements in robustness achieved in calculating the correlations between stock returns for the FTSE 100 constituents by using Bicor. Section VI demonstrates the benefit of using Bicor instead of Pearson in graph based portfolio construction.

II. Empirical Single Observation Influence on Pearson

One important problem with Pearson correlation is its behaviour when applied to data having fat-tailed distributions or containing outliers. It is known that in the presence of such distributions and outliers, Pearson correlation is not *robust*. Given that we know the distribution of stock returns exhibit fat tails [9], [10], it is likely that this lack of robustness could lead to misleading correlations. In this section we formalise the notion of robustness, and analyse the robustness of the Pearson correlation using empirical data.

The *breakdown point* of an estimator is one of a variety of measures of statistical robustness. Intuitively, it is the proportion of observations that can be made arbitrarily large before the estimator provides an incorrect result. For example, the mean as an estimator of location has a breakdown point of 0 as even a single large observation will make the mean arbitrarily large also. In contrast, the median as an estimator of location has a breakdown point of 0.5 as up to half the observations can be made arbitrarily large without affecting the result [11].

An empirical influence curve can be used to visualise how the value of an estimator is affected by a single arbitrary observation. For some vector $\mathbf{x}$ of data, the empirical influence curve can be shown by setting (for some arbitrary l) x_l to an increasingly large (positive or negative) value, and then calculating the value for the estimator of interest on each of those instances. Plotting the estimator against a range of values for x_l will show its sensitivity to an increasingly large outlier. Examples of this can be found in [11].

In the following, we use a similar method to provide insight into the robustness of Pearson correlation for stock returns.

For every stock i in the FTSE 100 between 2003-03-21 and 2013-12-31, let $\mathbf{r}_{iw}$ be the vector of log daily returns for that stock in window w. Each window w contains $T = 60$ periods of daily returns (3 months of trading days). Let the step size $\delta T = 60$ days. Thus, we have $M = 43$ consecutive adjacent windows of 3 months per window. For each of these windows we have approximately 100 return vectors, $\mathbf{r}_{iw}$, $i \in 1..100$.

For every window $w \in 1..43$, let $\rho_w(i,j)$ be the Pearson correlation between stocks i, j in window w. For every pair i, j in every window w, we calculate how much $\rho_w(i,j)$ can change by the removal of a single datapoint in $\mathbf{r}_{iw}$ and $\mathbf{r}_{jw}$.

Let $\mathbf{r}_{iw}^{(l)}$ and $\mathbf{r}_{jw}^{(l)}$ be the vectors $\mathbf{r}_{iw}$ and $\mathbf{r}_{jw}$ respectively with the datapoint at index $l \in 1..T$ removed. Note that this means the vectors $\mathbf{r}_{iw}^{(l)}$ and $\mathbf{r}_{jw}^{(l)}$ contain $T-1$ elements. Let $\rho_w^{(l)}(i,j)$ be defined as the Pearson correlation between $\mathbf{r}_{iw}^{(l)}$ and $\mathbf{r}_{jw}^{(l)}$:

$$\rho_w^{(l)}(i,j) = \frac{\sum_{t=1}^{T-1}(r_{iw}^{(l)}(t) - \bar{\mathbf{s}}_{\mathbf{iw}})(r_{jw}^{(l)}(t) - \bar{\mathbf{s}}_{\mathbf{jw}})}{\sqrt{\sum_{t=1}^{T-1}(r_{iw}^{(l)}(t) - \bar{\mathbf{s}}_{\mathbf{iw}})^2}\sqrt{\sum_{t=1}^{T-1}(r_{jw}^{(l)}(t) - \bar{\mathbf{s}}_{\mathbf{jw}})^2}} \tag{3}$$

Where $\bar{\mathbf{s}}_{\mathbf{iw}}$ is the sample mean for $\mathbf{r}_{iw}^{(l)}$. That is, $\bar{\mathbf{s}}_{\mathbf{iw}} = [1/(T-1)]\sum_{t=1}^{T-1} r_{iw}^{(l)}(t)$.

When for a given pair of stocks i, j in window w, the Pearson correlation $\rho_w^{(l)}(i,j)$ is calculated for every $l \in 1..T$, then the maximum change in Pearson correlation $\Delta_w(i,j)$ caused by the removal of a single observation is defined as:

$$\Delta_w(i,j) = \max_{l \in 1..T} |\rho_w(i,j) - \rho_w^{(l)}(i,j)| \tag{4}$$

That is, $\Delta_w(i,j)$ is the largest absolute change in the Pearson correlation between stocks i, j in window w, caused by the removal of a single data point in the vectors of returns for stocks i and j.

We found that between 2003 and 2013 in the 3 month windows of correlations, in 20% of all Pearson correlations, the change of a single observation in 3 months (60 datapoints) of data caused a change $\geq \pm 0.09$. In 5% of correlations, 1 datapoint could cause a change $\geq \pm 0.14$, and in 1 out of 100 Pearson correlations, a single data point was able to cause a change of $\geq \pm 0.21$.

This shows that a small proportion of stock correlations are significantly affected by just 1 data point; in this case by just 1 day in a 3 month period. In the following two sections we will introduce Biweight Mid Correlation as a more robust measure of correlation, and show to what extent it can remedy the problems discussed thus far.

III. Biweight Mid Correlation

A. Definition

In order to define Biweight Mid Correlation, it is helpful to first define Biweight Mid Variance.

The influence function of M-estimators of location can be used to define an entire class of measures of dispersion, details of which can be found in [11]. One such measure is the Biweight Mid Variance (Bivar).

As part of defining Bivar, it is helpful to set up two intermediary vectors. Firstly, for a random variable $\mathbf{X}$ with

median M_X and a constant K (the value of which will be discussed later), let the elements of $\mathbf{U}$ be:

$$U_i = \frac{X_i - M_X}{K \times \text{MAD}_X} \tag{5}$$

Where $\text{MAD}_X = \text{median}(|\mathbf{X} - M_X|)$, that is, the median of the absolute deviations from the median of $\mathbf{X}$. Notice that the magnitude of U_i is proportional to the distance between X_i and the median of $\mathbf{X}$.

Secondly, we define the elements of vector $\mathbf{a}$ to be:

$$a_i = \begin{cases} 1, & \text{if } |U_i| < 1 \\ 0, & \text{if } |U_i| \geq 1 \end{cases} \tag{6}$$

a_i will be 0 for any X_i more than K median absolute deviations away from the median of $\mathbf{X}$, and 1 otherwise.

We can now define Bivar as:

$$\hat{\zeta}_{bi}^2 = \frac{n \sum a_i (X_i - M_X)^2 (1 - U_i^2)^4}{(\sum a_i (1 - U_i^2)(1 - 5U_i^2))^2} \tag{7}$$

Given our earlier definitions of U_i and a_i, we can see that datapoints more than K median absolute deviations away from the median of $\mathbf{X}$ do not influence the Bivar. Furthermore, the influence of the remaining datapoints on the Bivar are weighted by their distance from the median of $\mathbf{X}$.

It has been shown empirically that $\hat{\zeta}_{bi}^2$ has a breakdown point of 0.5, but a formal proof does not yet exist [12]. Bivar has been shown to be a good choice as a robust dispersion measure in practice, as it has a high tri-efficiency compared to many other measures of dispersion when the constant $K = 9$ is chosen [13] [14]. We will define tri-efficiency and discuss the choice of $K = 9$ further in section III-B.

We can now define the Biweight Mid Covariance (Bicov) and Biweight Mid Correlation (Bicor). Let $\mathbf{X}$ and $\mathbf{Y}$ be two random variables and let M_X, M_Y and MAD_X, MAD_Y be the medians and median absolute deviations respectively for $\mathbf{X}$ and $\mathbf{Y}$. Similarly to how we constructed Bivar, let $\mathbf{U}$, $\mathbf{V}$, $\mathbf{a}$ and $\mathbf{b}$ be intermediary vectors, with their elements defined as:

$$\begin{aligned} U_i &= (X_i - M_X)/(K \times \text{MAD}_X) \\ V_i &= (Y_i - M_Y)/(K \times \text{MAD}_Y) \end{aligned}$$

$$a_i = \begin{cases} 1, & \text{if } |U_i| < 1 \\ 0, & \text{if } |U_i| \geq 1 \end{cases} \quad b_i = \begin{cases} 1, & \text{if } |V_i| < 1 \\ 0, & \text{if } |V_i| \geq 1 \end{cases} \tag{8}$$

Where K is a constant as before. Furthermore, let the elements of vectors $\mathbf{p}$ and $\mathbf{q}$ be:

$$\begin{aligned} p_i &= a_i(1 - U_i^2) \\ q_i &= b_i(1 - V_i^2) \end{aligned} \tag{9}$$

Using the above vectors, equations (10) and (11) define the Biweight Mid Covariance and Biweight Mid Correlation respectively.

$$s_B(X, Y) = \frac{n \sum p_i^2 (X_i - M_x) q_i^2 (Y_i - M_y)}{[\sum p_i (1 - 5U_i^2)][\sum q_i (1 - 5V_i^2)]} \tag{10}$$

$$r_B(X, Y) = s_B(X, Y) / \sqrt{s_B(X, X) s_B(Y, Y)} \tag{11}$$

The choice of parameter K will be discussed next.

B. Bicor Parameter K

In 1985, Lax [13] compared 150 methods of estimating measures of dispersion, including Bivar with various values for the parameter K. A more recent reproduction in 2008 [14] of the Lax study added more dispersion measures to the comparison and made use of modern computational techniques.

In these studies, the robustness of the measures of dispersion was tested on three distributions. These "corner" distributions are considered to represent important cases in sampling, details of which can be found in [15].

The tri-efficiency of each measure of dispersion under investigation was calculated as follows. Let $\hat{\zeta}_i^2$ be the i^{th} measure of dispersion investigated. Each corner distribution d was sampled k times; and every estimator $\hat{\zeta}_i^2$ was then calculated for each of those samples.

Let $\mathbf{X}_i^{(d)}$ be the k-sized vector containing the value of dispersion estimator i on corner distribution d for the k samples.

For a given corner distribution d, let:

$$V_{min}^{(d)} = \min_i \left(\text{Var}(\ln(\mathbf{X}_i^{(d)})) \right) \tag{12}$$

Where Var calculates variance. That is, for every estimator i, the variance of the log of estimator i for the k samples of distribution d is $\text{Var}(\ln(\mathbf{X}_i^{(d)}))$. $V_{min}^{(d)}$ is set to the minimum variance found out of all the estimators.

The efficiency of each estimator i for distribution d is:

$$E_i^{(d)} = 100 \left(\frac{V_{min}^{(d)}}{\text{Var}(\ln(\mathbf{X}_i^{(d)}))} \right) \tag{13}$$

Each estimator has three efficiency values, one for each of the three corner distributions. The smallest of those three values is defined to be the tri-efficiency of the estimator:

$$\text{Tri}_i = \min_d (E_i^{(d)}) \tag{14}$$

In previous studies, the most efficient measure of dispersion has been shown to be the Biweight Mid Variance with $K = 9$, having a tri-efficiency of 85.8 [13] [14].

IV. Empirical Efficiency on Stock Return Data

We can apply the same methodology as [13] to find the parameter K that leads to the most efficient version of Bivar when applied to our stock return data. Instead of using the three corner distributions, we use empirical stock return distributions to determine the efficiency of dispersion measures.

The measures of dispersion $\hat{\zeta}_i^2$ we will consider are Variance σ^2 and Bivar $\hat{\zeta}_{bi}^2$ with parameters $K \in \{1..20\}$.

Our analysis is on the return of all stocks in the FTSE 100 between 2003 and 2013 in windows of size $T = 1$ year with step size $\delta T = 6$ months. This leads to $M = 20$ windows. Each window w contains 100 distributions, one for each stock. Each of those distributions is sampled 1000 times.

Let $\mathbf{X}_i^{(ws)}$ be the vector of size 1000, containing the values for dispersion estimator i for the 1000 samples of the return distribution for stock s in window w.

Let $V_{min}^{(ws)}$ be:

$$V_{min}^{(ws)} = \min_i \left(\mathrm{Var}(\ln(\mathbf{X}_i^{(ws)})) \right) \tag{15}$$

We define $E_i^{(ws)}$ as the efficiency of dispersion measure i in window w for stock s.

$$E_i^{(ws)} = \frac{V_{min}^{(ws)}}{\mathrm{Var}(\ln \mathbf{X}_i^{(ws)})} \tag{16}$$

We define the minimum window efficiency $J_i^{(w)}$ for an estimator i in window w to be:

$$J_i^{(w)} = \min_s E_i^{(ws)} \tag{17}$$

The distribution of minimum window efficiency $J_i^{(w)}$ in the $M = 20$ windows for all 20 estimators is shown in figure 1.

This shows that Biweight Mid Variance with $K = 12$ or $K = 13$ appears to be the most efficient estimator to use for our stock data. Figure 2 shows that this result is relatively constant over time from 2003 to 2013.

In the remainder of this paper, we show how Bicor $K = 9$ and Bicor $K = 13$ compare to Pearson correlation. In the next section, we show the improved single observation influence.

V. Improved Empirical Robustness in Stock Return Data

In Section II we investigated the empirical single observation influence on the Pearson correlation between all pairs of stocks i, j in the FTSE 100 between 2003 and 2013. If we reproduce this analysis for Bicor with $K = 9$ and $K = 13$, we find a much improved situation as shown in figure 3 and in table I. When Bicor is used to calculate the correlation between stock returns, the maximum influence that a single daily return has on the value of the 3-month correlation is reduced. Hence, the use of Bicor is more robust to outliers. Note that Bicor with $K = 13$ has a higher single observation influence than Bicor with $K = 9$. This is due to the fact that when $K = 9$, observations 9 median absolute deviations away from the median are not considered, but when $K = 13$, this increases to 13.

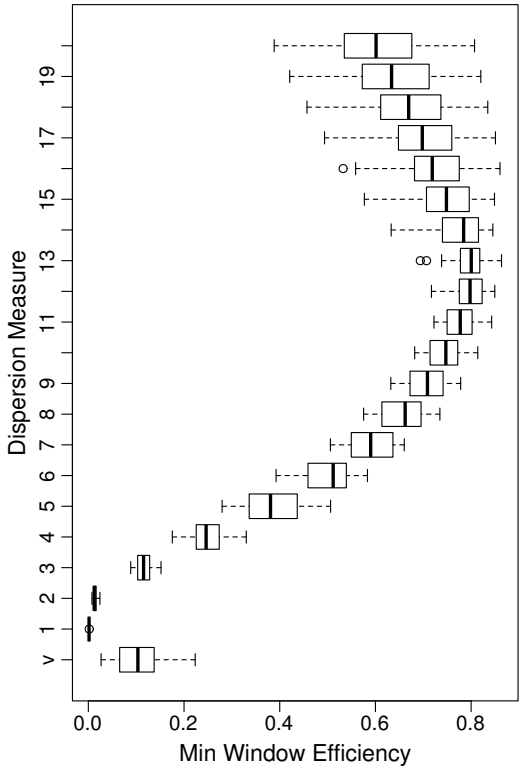

Fig. 1: The distribution of minimum window efficiency for variance (v) and the 20 Bivar measures with $K = 1..20$.

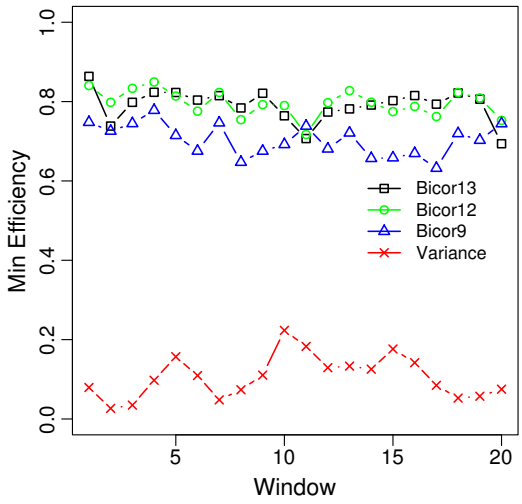

Fig. 2: Consistency of Minimum Window Efficiency Over Time for Variance and Bicor with various K.

In the previous section, we have shown Bivar, the dispersion estimator used for Bicor, to be a more efficient estimator than variance. In this section, we have shown that Bicor is more robust than Pearson correlation in terms of single observation

	50%	80%	90%	95%	99%
Pearson	0.06	0.09	0.11	0.14	0.21
Bicor $K = 9$	0.05	0.06	0.07	0.07	0.08
Bicor $K = 13$	0.05	0.07	0.08	0.09	0.11

TABLE I: Quantiles of the largest absolute change in the Pearson, Bicor with $K = 9$ and Bicor with $K = 13$ correlation between the returns of all stocks i, j in all windows w, caused by the removal of a single data point in the vectors of returns for stocks i and j in window w.

influence. In the next section we show that these improved properties lead to better graph based portfolio construction.

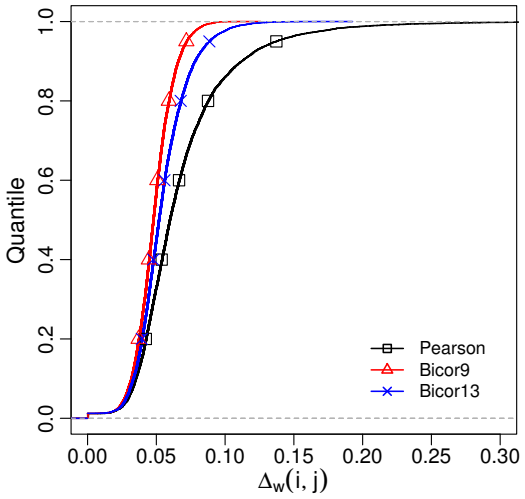

Fig. 3: The empirical CDF of maximum change in correlation between daily stock returns, caused by the removal of 1 observation in 3 month windows.

VI. Improved Peripheral Stock Portfolios

Here we show that the use of Biweight Mid Correlation can improve the selection of stocks based on network centrality [16]. A previous study [3] shows that when the Pearson correlation matrix of the stocks is transformed to a distance matrix and then filtered using MST or PMFG, portfolios built with lower centrality stocks perform better than those made up of high centrality stocks. Here we show a similar result and demonstrate how this can be further improved by using Bicor to measure correlation between stock returns.

Let $\mathbf{A}_w^{(P)}$ be the Pearson correlation matrix in window w for the constituents of the FTSE 100. Let $\mathbf{A}_w^{(BK)}$ be the correlation matrix for Bicor with parameter K. We have calculated these matrices for windows w of size $T = 12$ months every 1 month between 2003 and 2013.

Both correlation matrices are transformed to a distance matrix and then filtered using the PMFG method described in the introduction [5]. Let $G_w^{(P)}$ be the graph obtained by transforming $\mathbf{A}_w^{(P)}$ to a distance matrix and then filtered using the PMFG method. That is, $G_w^{(P)}$ is the planar graph that contains the edges that represent the highest correlations, under the constraint that the graph is planar. Let $G_w^{(BK)}$ be the equivalent for $\mathbf{A}_w^{(BK)}$.

The closeness centrality of a vertex (stock) v in a graph is defined as:

$$C_v = \left(\sum_{i \neq v} d(v, i) \right)^{-1} \tag{18}$$

Where $d(v, i)$ is the shortest distance from vertex (stock) v to a vertex (stock) i in the graph. This closeness centrality can be calculated for every stock in the graphs $G_w^{(P)}$ and $G_w^{(BK)}$. A portfolio size m can be chosen, and then based on the closeness centrality, equally weighted portfolios made up of the m most peripheral stocks and m most central stocks can then be constructed from the two graphs.

For each portfolio p, for the $t = 1..250$ days after building the portfolio, we calculate the return per unit risk, $R_t^{(p)}/\sigma_t^{(p)}$, where $R_t^{(p)} = (P_t^{(p)} - P_1^{(p)})/P_1^{(p)}$ and $\sigma_t^{(p)}$ is the standard deviation of the daily returns for portfolio p between period 1 and t. $P_t^{(p)}$ is the value of the portfolio at time t, which is calculated as the sample mean of the price of the constituent stocks within the portfolio.

Figure 4 shows the mean performance of the portfolios of the 14 most peripheral and 14 most central stocks in the Pearson and Bicor $K = 9$ PMFG filtered graphs in the windows between 2003 and 2013. This confirms earlier findings [3] that portfolios built from periphery stocks perform better than portfolios built from central stocks. Furthermore it shows that this gap is widened when using Bicor.

While figure 4 shows the portfolio based on Bicor with $K = 9$, note that the portfolios based on the Bicor matrix where Bicor was calculated with $7 \leq K \leq 16$ all performed better than the Pearson based periphery portfolios when the portfolio size was 12 or greater.

There are many choices for K when calculating the Biweight Mid Correlation, and many possible choices for the portfolio size m. Figure 5 shows the relationship between the combinations of Bicor K and portfolio size m, and the mean portfolio performance 250 trading days after portfolio construction. This demonstrates that on average, the optimal portfolio made up of FTSE 100 stocks, is the one consisting of the $m = 14$ least central stocks in the PMFG based on the Bicor $K = 9$ correlation matrix. Portfolios containing the 14 most peripheral stocks in the PMFG of $\mathbf{A}_w^{(B9)}$ have on average a better return to risk ratio 250 trading days after portfolio construction than the market portfolio over the same period ($p = 0.047$).

VII. Conclusion and Future Work

We have shown that Biweight Mid Variance is a more robust choice of dispersion metric when applied on stock return data, and Biweight Mid Correlation a more robust correlation metric. The definition of Bivar includes a constant K, the usual choice for this being $K = 9$. While Bivar with $K = 9$ is

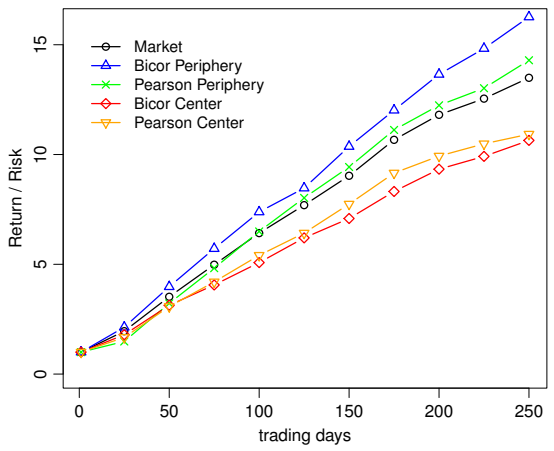

Fig. 4: Return/Risk performance of various portfolios based on centrality measures of the PMFG filtered graphs of the Pearson and Bicor correlation matrix.

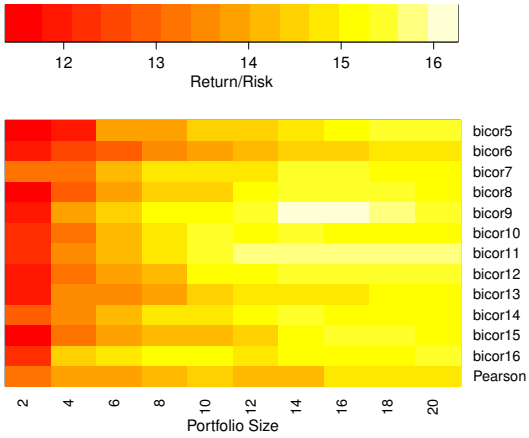

Fig. 5: Mean Return/Risk performance of portfolios of various sizes m, constructed from the m least central stocks in the PMFG based on the Pearson and Bicor correlation matrices.

significantly more efficient than variance, it is not the optimal choice when applied to stock return data. Our experiments revealed that Bivar with $K = 13$ is the most efficient choice. In any case, whether the usual $K = 9$ or $K = 13$ is chosen, Bivar is a significantly more efficient dispersion estimator than Variance.

We demonstrated that existing methods of graph based portfolio construction can be improved by using Bicor as the correlation metric. In our study on the FTSE 100 stocks between 2003 and 2013, we have found the optimal portfolio to be that made up of the 14 most peripheral stocks in the PMFG of the Bicor $K = 9$ correlation matrix. However, many choices of K lead to improved portfolios when compared to the market portfolio or the peripheral portfolio based on the Pearson correlations.

It is noted that while Bicor $K = 9$ was the optimal choice for peripheral portfolio performance, this choice of constant was not shown to be the most statistically efficient for the Bivar dispersion measure on which Bicor is based.

A topic that requires further analysis in a future study is the relationship between the graph centrality based portfolio performance and models of returns such as the single index model. It would be interesting to understand how market risk β and individual risk α relate to centrality measures in these correlation matrices.

Finally, there is scope for a comprehensive study of the various methods of filtering the correlation matrix. An in depth analysis on what information the filtering methods retain, remove or uncover, is needed. Currently, the filtering methods are primarily judged by their usefulness when applied.

Acknowledgment

We gratefully acknowledge the EPSRC Doctoral Training Grant for funding support. The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

References

[1] R. N. Mantegna, "Hierarchical structure in financial markets," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 11, no. 1, pp. 193–197, 1999.

[2] V. Tola, F. Lillo, M. Gallegati, and R. N. Mantegna, "Cluster analysis for portfolio optimization," *Journal of Economic Dynamics and Control*, vol. 32, no. 1, pp. 235–258, 2008.

[3] F. Pozzi, T. Di Matteo, and T. Aste, "Spread of risk across financial markets: better to invest in the peripheries," *Scientific reports*, vol. 3, 2013.

[4] T. Aste, W. Shaw, and T. Di Matteo, "Correlation structure and dynamics in volatile markets," *New Journal of Physics*, vol. 12, no. 8, p. 085009, 2010.

[5] T. Aste, T. Di Matteo, and S. Hyde, "Complex networks on hyperbolic surfaces," *Physica A: Statistical Mechanics and its Applications*, vol. 346, no. 1, pp. 20–26, 2005.

[6] M. Tumminello, T. Aste, T. Di Matteo, and R. N. Mantegna, "A tool for filtering information in complex systems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 30, pp. 10 421–10 426, 2005.

[7] W.-Q. Huang, X.-T. Zhuang, and S. Yao, "A network analysis of the Chinese stock market," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 14, pp. 2956–2964, 2009.

[8] J.-P. Onnela, A. Chakraborti, K. Kaski, and J. Kertesz, "Dynamic asset trees and Black Monday," *Physica A: Statistical Mechanics and its Applications*, vol. 324, no. 1, pp. 247–252, 2003.

[9] B. Mandelbrot, "The Variation of Certain Speculative Prices," *The Journal of Business*, vol. 36, no. 4, pp. 394–419, 1963.

[10] R. N. Mantegna and H. E. Stanley, "Modeling of financial data: comparison of the truncated Lévy flight and the ARCH (1) and GARCH (1, 1) processes," *Physica A: Statistical Mechanics and its Applications*, vol. 254, no. 1, pp. 77–84, 1998.

[11] R. R. Wilcox, *Introduction to robust estimation and hypothesis testing*. Academic Press, 2012.

[12] K. M. Goldberg and B. Iglewicz, "Bivariate extensions of the boxplot," *Technometrics*, vol. 34, no. 3, pp. 307–320, 1992.

[13] D. A. Lax, "Robust estimators of scale: Finite-sample performance in long-tailed symmetric distributions," *Journal of the American Statistical Association*, vol. 80, no. 391, pp. 736–741, 1985.

[14] J. A. Randal, "A reinvestigation of robust scale estimation in finite samples," *Computational Statistics & Data Analysis*, vol. 52, no. 11, pp. 5014–5021, 2008.

[15] J. A. Randal and P. J. Thomson, "Maximum likelihood estimation for Tukey's three corners," *Computational statistics & data analysis*, vol. 46, no. 4, pp. 677–687, 2004.

[16] M. Newman, *Networks: an introduction*. Oxford University Press, 2010.