

Exponential smoothing weighted correlations

F. Pozzi^{1,a}, T. Di Matteo^{1,2}, and T. Aste^{1,3}

¹ Department of Applied Mathematics, Research School of Physical Sciences, The Australian National University, 0200 Canberra, ACT, Australia

² Department of Mathematics, King's College London, The Strand, WC2R 2LS London, UK

³ School of Physical Sciences, University of Kent, Canterbury, CT2 7NH Kent, UK

Received 25 August 2011 / Received in final form 8 March 2012

Published online 4 June 2012 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2012

Abstract. In many practical applications, correlation matrices might be affected by the “curse of dimensionality” and by an excessive sensitiveness to outliers and remote observations. These shortcomings can cause problems of statistical robustness especially accentuated when a system of *dynamic* correlations over a running window is concerned. These drawbacks can be partially mitigated by assigning a structure of weights to observational events. In this paper, we discuss Pearson’s ρ and Kendall’s τ correlation matrices, weighted with an exponential smoothing, computed on moving windows using a data-set of daily returns for 300 NYSE highly capitalized companies in the period between 2001 and 2003. Criteria for jointly determining optimal weights together with the optimal length of the running window are proposed. We find that the exponential smoothing can provide more robust and reliable dynamic measures and we discuss that a careful choice of the parameters can reduce the autocorrelation of dynamic correlations whilst keeping significance and robustness of the measure. Weighted correlations are found to be smoother and recovering faster from market turbulence than their unweighted counterparts, helping also to discriminate more effectively genuine from spurious correlations.

1 Introduction

Many different measures of dependency between variables may be used in order to describe the interplay of different elements in a complex system, and the strength of their relationship. In this paper we focus on two of the most broadly used measures, namely the Pearson product-moment correlation ρ and the Kendall rank correlation τ . When dynamic correlations of time series are calculated over a running window Δt , then a single anomalous data-point today will cause a bias for the entire length of Δt and the effect will abruptly cease thereafter, causing a false perception of the chronological or logical order of events. By introducing a structure of weights the above-mentioned shortcomings can be partially mitigated, if not fully healed, and the sensitiveness of correlations to outliers greatly reduced [1–8].

In general, individual correlation coefficients between two variables are statistically significant if calculated on a sufficiently large number of observations – in fact, the variance of their distribution is in inverse relation with the sample size, Δt . But this general principle is not sufficient to ensure a well-conditioned N -by- N correlation matrix when N variables are considered and N is large. If the *system of correlations*, rather than *individual correlations*, is the focus of the analysis, then the sample size needs to

be at least of the same order of N , usually $\Delta t \gg N$ is required [9]. In fact, even a full-rank matrix may be affected by the curse of dimensionality¹ and may provide misleading insights [10,11]. Therefore, the choices of Δt and N are mutually constrained. The main novelty of this paper, with respect to the existing literature on financial correlations [12–16], concerns the study of τ as alternative to or concurrent with ρ coefficients, and the introduction of exponential weighting with the aim to obtain reliable full-rank dependency matrices where present observations weight more than past measurements, thus better characterizing the dynamics of the evolving dependency structure within the market.

Jointly determining optimal running window Δt and weight structure is critical for a fully informative correlation matrix. A set of optimal criteria is then needed in order to improve robustness to outliers and correctly measure the system of relations between coefficients. In fact, the choice of an appropriate Δt is crucial and can be often influenced by issues not strictly related to the

¹ The curse of dimensionality was first introduced by Bellman [18,19] in problems of optimization where finding the minimum of a function gets virtually intractable as the number of its dimensions increase. The concept has been extended to distance functions and nearest neighbor search: in high dimensional spaces the distances between different pairs become indiscernible [20–22].

^a e-mail: francesco.pozzi@anu.edu.au

analyst's purpose, such as data availability or the need of ensuring a matrix full rank. Although the idea of weighted correlations is not new [1–8,17], the study of optimal tuning of the weights, aimed at avoiding unwelcome side effects while maximizing significance and robustness, has been mostly overlooked in the literature. In this paper, we show that incautious weights can destabilize the system of correlations, raising the condition number of the matrix, deteriorating the structure of matrices's eigenvalues and distorting the distribution of coefficients. Instead, a wise choice of weights can reduce the autocorrelation of dynamic correlations, dampening undue effects of remote events on the present and contributing to discriminate spurious from genuine correlations.

This paper is organized as follows: in Section 2, matrix notations for Pearson and Kendall correlations are introduced and some of their properties are discussed and compared. In Section 3, weighted correlation matrices are introduced with particular emphasis on the exponential smoothing criterion. In Section 4, empirical results are shown for a system of 300 *NYSE* highly capitalized companies between 2001 and 2003. In Section 5, some criteria for determining optimal weights and running window are then discussed. In Section 6, conclusions are drawn and some ideas for future works outlined.

2 Dynamic measures of dependency

In this section we introduce Pearson and Kendall correlations and we discuss their advantages and disadvantages. In the following, individual observations of variable $i = \{1, 2, \dots, N\}$ at time $t = \{1, 2, \dots, T\}$ are denoted as y_t^i . We study moving windows containing $\Delta t \leq T$ consecutive observations. The time series is denoted as y^i , column vectors of matrix \mathbf{Y} , such that $[y_{si}^i] = y_s^i$ with $s = \{t - \Delta t + 1, t - \Delta t + 2, \dots, t\}$ and $i = \{1, 2, \dots, N\}$.

2.1 Pearson product-moment correlation coefficient

In 1895, Karl Pearson introduced his product-moment correlation coefficient between two variables, defined as [23–31]:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \quad (1)$$

where $\sigma_{ij} = \frac{1}{\Delta t} \sum_{t=1}^{\Delta t} (y_t^i - \bar{y}^i)(y_t^j - \bar{y}^j)$ is the covariance of the two variables, \mathbf{y}^i and \mathbf{y}^j ; $\sigma_i = \sqrt{\frac{1}{\Delta t} \sum_{t=1}^{\Delta t} (y_t^i - \bar{y}^i)^2}$ and $\sigma_j = \sqrt{\frac{1}{\Delta t} \sum_{t=1}^{\Delta t} (y_t^j - \bar{y}^j)^2}$ are their standard deviations; $\bar{y}^i = \frac{1}{\Delta t} \sum_{t=1}^{\Delta t} y_t^i$ and $\bar{y}^j = \frac{1}{\Delta t} \sum_{t=1}^{\Delta t} y_t^j$ are their sample means. The Pearson correlation matrix is positive semi-definite (see Appendix A).

Pearson correlation coefficients are simple to calculate and formally elegant. However, ρ has some important shortcomings, in particular:

- since it is devised to capture linear relations it might be misleading when used for nonlinear relations [32,33];

- as a consequence, it depends on the unit of measure of variables and it does change for nonlinear transformations²;
- since the correlation matrix rank is $\text{rank}(\rho) \leq \min(\Delta t - 1, N)$, the inverse of ρ does not exist when $\Delta t < N + 1$ ³;
- it is equally sensitive to each observation without any regard for their order: if the two variables are time series then their correlation is equally affected by events in the present and in the remote past;
- it is not robust to outliers;
- it might be spurious if the two variables are dependent on a confounding factor, a lurking variable that governs both;
- it is only defined when the variances of the corresponding variables are finite;
- it can get unreliable in presence of fat-tailed distributions.

2.2 Capturing nonlinear relations: Kendall's τ

Among all measures of concordance (or association or statistical dependence), Kendall's τ rank correlation [34–36] is probably the most important [7]: it can catch nonlinear qualitative monotonic relationships (without the need to calculate ranks); it does not depend on the scale of variables; the rank of a Kendall correlation matrix can be full, even if $\Delta t \leq N$; it is robust to outliers; it is defined even if the variables's variances are infinite; it is a distribution-free measure, i.e. not dependent on the statistical distributions of the variables.

Kendall's τ rank correlation measures the degree of similarity between two variables by counting the concordant and discordant pairs. Its 1938 first formulation, not counting for tied pairs (so called τ_A) can be written as [34,35]:

$$\tau_{A\ ij} = \frac{2}{\Delta t (\Delta t - 1)} \sum_{u=1}^{\Delta t-1} \sum_{v=u+1}^{\Delta t} d_{uv}^i d_{uv}^j \quad (2)$$

where $d_{uv}^k = \text{sgn}(y_u^k - y_v^k)$, $k = i, j$. Then, in 1948, Kendall [35] proposed a correction for tied pairs (i.e. cases where $d_{uv} = 0$). The corrected measure is commonly known as τ_B ⁴:

$$\tau_{B\ ij} = \frac{\sum_{u=1}^{\Delta t-1} \sum_{v=u+1}^{\Delta t} d_{uv}^i d_{uv}^j}{\sqrt{\left[\frac{1}{2}\Delta t (\Delta t - 1) - n^i\right] \left[\frac{1}{2}\Delta t (\Delta t - 1) - n^j\right]}} \quad (3)$$

where n^k is the total number of tied pairs for variable k . Since we only consider τ_B , in the following it is simply denoted as τ .

² For instance, in finance, it is well known that correlations are not invariant under transformations of the risk [1].

³ Note that, since each column vector has been subtracted by its sample mean, one degree of freedom is lost.

⁴ See chapter 3, p. 26, equation (3.3) in [35].

In Appendix B we prove that Kendall’s correlation matrix can be seen as a particular kind of Pearson’s ρ , generally characterized by greater rank and maintaining positive semi-definiteness.

Despite its desirable theoretical qualities, τ ’s computational complexity, $\mathcal{O}(\Delta t \log \Delta t)$ at best [37,38], is a limit to its use on large samples. Furthermore, τ is still equally sensitive to events in the present and in the past and it might be spurious if the two variables are dependent on latent variables.

3 Assigning weights to observations

In financial studies, assigning the same value to each observation in time is equivalent to considering recent and remote past as indifferent from an informational point of view. Such hypothesis is hardly sensible since most operators would judge the information from recent events as more valuable than from remote ones for both descriptive and forecasting purposes. In fact, finance is a discipline projected towards the future. In order to take into account and incorporate the operators’ preference for fresh information, we assign a structure of weights⁵, $\mathbf{w} \geq \mathbf{0}$, such that $w_u > w_v, \forall u > v$.

3.1 Pearson weighted correlation matrix

In order to compute weighted Pearson correlation coefficients we introduce a weight structure, $w_t \geq 0$, such that $\sum_{t=1}^{\Delta t} w_t = 1$, which operates on sample means, variances and covariances as follows:

$$\begin{aligned} \bar{y}_k^w &= \sum_{t=1}^{\Delta t} w_t y_t^k & \sigma_k^w &= \sqrt{\sum_{t=1}^{\Delta t} w_t (y_t^k - \bar{y}_k^w)^2} \\ \sigma_{ij}^w &= \sum_{t=1}^{\Delta t} w_t (y_t^i - \bar{y}_i^w) (y_t^j - \bar{y}_j^w) & \rho_{ij}^w &= \frac{\sigma_{ij}^w}{\sigma_i^w \sigma_j^w}. \end{aligned} \quad (4)$$

Pearson weighted correlation coefficients are very similar and enjoy the same properties as the corresponding unweighted coefficients. In particular they are positive semi-definite matrices (see Appendix C).

3.1.1 Exponential smoothing for pearson’s correlations

When time series are concerned, since the recent past is often more indicative of the near future than the remote past, the structure of weights should be an increasing function with small values for the past and large value for the present. A common functional form for these weights is

⁵ In the trivial case when the weight of each observation, $[y_s^i, y_s^j]$, reflects its empirical frequency in the series, the same formulae introduced here apply, as a simplified method of computing traditional measures by aggregating common elements.

the exponential function often referred to as “exponential smoothing”. This is the kind of weighting we use in this paper, however other types of expressions can be considered as well. We therefore define weights as

$$w_t = w_0 \exp\left(\frac{t - \Delta t}{\theta}\right), \forall t \in \{1, 2, \dots, \Delta t\} \quad (5)$$

where θ is the weights’ characteristic time, $\theta > 0$. By varying this parameter the exponential smoothing can be made very flexible. Note that the same function can be written as $w_t = w_0 e^{\alpha(t - \Delta t)}, \forall t \in \{1, 2, \dots, \Delta t\}$, where $\alpha = \frac{1}{\theta}$ is the exponential decay factor, with $\alpha \in \mathbb{R}$ and $\alpha \geq 0$.

Let us note that J.P. Morgan/Reuters [8] as well as Litterman and Winkelmann [1,2] use a different specification of weights:

$$w_t = w_0 (1 - \xi)^{(\Delta t - t)}, \forall t \in \{1, 2, \dots, \Delta t\} \quad (6)$$

which is equivalent to equation (5) once we write $\xi = 1 - e^{-\alpha}$ and $\xi \in [0, 1]$.

The constant w_0 in equation (5) can be obtained by observing that w is subject to the constraint defined in equation (C.1), i.e.

$$\sum_{t=1}^{\Delta t} w_0 e^{\alpha(t - \Delta t)} = 1. \quad (7)$$

As proved in Appendix E, by solving for w_0 , we find

$$w_0(\alpha) = \frac{1 - e^{-\alpha}}{1 - e^{-\alpha \Delta t}}. \quad (8)$$

A *MATLAB* code that computes Pearson’s correlation matrices, weighted with exponential smoothing, is given in Appendix F.

Note that $\lim_{\alpha \rightarrow 0^+} \frac{1 - e^{-\alpha}}{1 - e^{-\alpha \Delta t}} = \frac{1}{\Delta t}$ and $\lim_{\alpha \rightarrow +\infty} \frac{1 - e^{-\alpha}}{1 - e^{-\alpha \Delta t}} = 1$. When α is 0 (or, equivalently, when $\theta \rightarrow \infty$) then weights are uniform; on the contrary, when α is large then remote events become increasingly irrelevant and recent events become ever more important. Given a finite number of observations, extremely large values for α (low values for θ) make the correlation matrix statistically non significant and numerically unstable. In fact, θ is a characteristic time providing the distance of the latest significant events. See Appendix G for some counterintuitive limits.

An elegant approach to determine a theoretically sound value for the decay factor is described in [1]: if the joint distribution of variables is known, then the maximum likelihood principle can be applied. For example, if returns are found to be jointly distributed as a multinormal, then the following function can be maximized as a function of the parameter ξ in equation (6):

$$\begin{aligned} \xi^* \equiv \arg \max_{0 \leq \xi < 1} & \left[-\frac{\Delta t}{2} \ln |\Sigma(\xi)| \right. \\ & \left. - \frac{1}{2} \sum_{t=1}^{\Delta t} (\mathbf{y}_t - \bar{\mathbf{Y}}(\xi))^T \Sigma(\xi)^{-1} (\mathbf{y}_t - \bar{\mathbf{Y}}(\xi)) \right]. \end{aligned} \quad (9)$$

$\Sigma(\xi)$ is the weighted covariance matrix and is defined in equation (C.2). This is only feasible, though, if the correlation matrix has full rank and if, together with variances, it is statistically significant - which is often the main difficulty when covariance/correlation matrices are concerned.

3.2 Kendall weighted correlation matrix

In analogy with the Pearson weighted correlations discussed in Section 3.1, a weighted Kendall Correlation may be defined as follows [3–7]:

$$\tau_{ij}^w = \sum_{u=1}^{\Delta t-1} \sum_{v=u+1}^{\Delta t} w_{uv} d_{uv}^i d_{uv}^j \quad (10)$$

where $d_{uv}^k = \text{sgn}(y_u^k - y_v^k)$, w_{uv} is increasing on both its subscripts and

$$\sum_{u=1}^{\Delta t-1} \sum_{v=u+1}^{\Delta t} w_{uv} = 1. \quad (11)$$

In Appendix H we show that τ^w matrix is positive semi-definite.

3.2.1 Exponential smoothing for Kendall correlations

In analogy with Section 3.1.1, we use the exponential smoothing:

$$w_{uv} = w_0 \exp\left(\frac{u - \Delta t}{\theta}\right) \exp\left(\frac{v - \Delta t}{\theta}\right) \quad (12)$$

where $u > v$. Alternatively, by posing $\alpha = \frac{1}{\theta}$,

$$\begin{aligned} w_{uv} &= w_0 e^{\alpha(u-\Delta t)} e^{\alpha(v-\Delta t)} \\ &= w_0 e^{\alpha(u+v-2\Delta t)}. \end{aligned} \quad (13)$$

Then, the constraint can be written as

$$\sum_{u=1}^{\Delta t-1} \sum_{v=u+1}^{\Delta t} w_0 e^{\alpha(u+v-2\Delta t)} = 1 \quad (14)$$

so that, as proved in Appendix J,

$$w_0(\alpha) = \frac{(e^\alpha - 1)^2 (e^\alpha + 1)}{e^{2\alpha} (1 - e^{-\alpha\Delta t}) [1 - e^{-\alpha(\Delta t-1)}]} \quad (15)$$

which tends to a uniform weight when α tends to zero,

$$\lim_{\alpha \rightarrow 0^+} w_0(\alpha) = \frac{2}{\Delta t (\Delta t - 1)} \quad (16)$$

and instead it diverges when $\alpha \rightarrow \infty$.

The *MATLAB* codes that compute Pearson and Kendall weighted correlation matrices are given in Appendix F.

4 Empirical findings: systems of correlations in finance

In equity markets, correlations of returns, measured over a running window of given length, vary over time; typically, they tend to increase during highly volatile periods with larger increases during bear periods than during bull periods. Thus, among other uses, they are also an indicator of the herd effect affecting investors who collectively tend to overreact (overshoot) during financial crisis, when panic spreads through the market, as well as during optimistic periods, when investors collectively underestimate risks [39–49].

4.1 Spurious correlations and false perceptions of events

Let us first apply the dynamic study of dependency with exponentially smoothed correlation to a practical case where we aim to distinguish between ‘genuine’ correlations which are a consequence of common trends of two stocks and ‘spurious’ correlations that are caused by large fluctuations within the entire market. We have considered $N = 300$ *NYSE* highly capitalized companies over a period of three years, 2001–2003, corresponding to a total $T = 748$ market days. From returns obtained as logarithmic differences of daily quotations, we have calculated all $\frac{N(N-1)}{2} = 44850$ dynamic Pearson correlation coefficients using a running window of six months ($\Delta t = 126$ market days) and, for all times $t = \{\Delta t, \Delta t + 1, \dots, T\}$, we have considered the overall average, $\bar{\rho}_t$. On the left side of Figure 1, we illustrate the trend over time of $\bar{\rho}_t$ (curve denoted by downward-pointing triangles) together with the *NYSE Composite Index* (curve denoted by upward-pointing triangles). On the lower subplot, as a proxy variable of market turbulence, the number of companies recording high returns or severe losses (positive or negative axis) is plotted; high returns and severe losses are defined as returns more than 2σ over or below the general daily average.

We observe that the average market correlation is stationarily low in the subperiod from July 2001 to May 2002, in spite of intense market turbulence, included the 9/11 shock. Then it begins to rise very steeply until July 23rd when it suddenly shoots up and remains up for $\Delta t = 126$ days after which it suddenly falls without any apparent cause; hence it slowly rises back a bit until July 2003 when it starts falling rapidly and steadily, while the market is finally recovering from the shocks of the latest years. We have therefore a clear indication that, during the shocks, $\bar{\rho}_t$ increased significantly to fall back again once normality was re-established. However, the length of the running window ($\Delta t = 126$ days) generates the optical illusion of a sharp variation in January 2003, six months after the shock occurred in July 23rd of the previous year.

Not only correlation measures are not stable over time, but they exhibit strong co-movement with market events and a large time-variability [47,48,50–52]. From Figure 1 we see that the average correlation reacts to market shocks

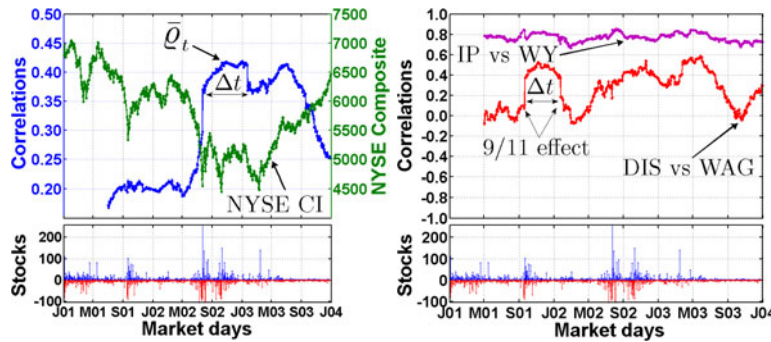


Fig. 1. (Color online) On the left, the downward-pointing triangles represent the average correlation ($\bar{\rho}_t$) for 300 stocks computed over a period of three years with a running window of six months ($\Delta t = 126$ days); the upward-pointing triangles represent the *NYSE Composite Index*. On the right, for $\Delta t = 84$ days, an example showing: the spurious correlation between *Walt Disney Corporation* and *Walgreen Company*, represented by the unstable curve at the bottom, *DIS* vs. *WAG*; the reliable strong correlation between *International Paper Company* and *Weyerhaeuser Company* represented by the stable curve at the top, *IP* vs. *WY*. The subplots report, as a proxy variable of market turbulence, the number of companies whose standardized returns are more than 2σ (positive axis) or less than -2σ (negative axis). Tick labels for the x axis are quarters of the year: “J” for “January”, “M” for “May”, “S” for “September”.

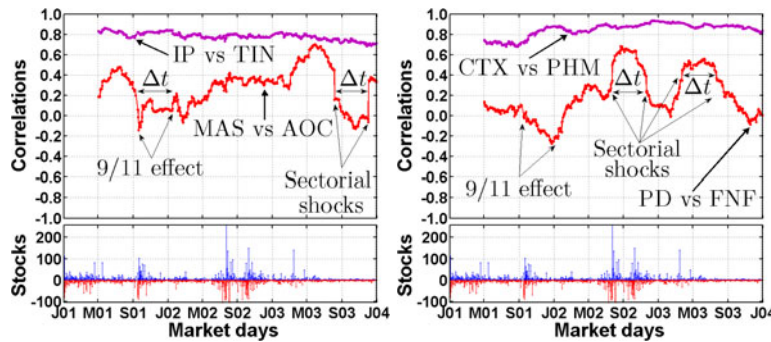


Fig. 2. (Color online) On the left, the spurious correlation between *Masco Corporation* and *AOC Corporation* (unstable curve at the bottom, *MAS* vs. *AOC*) and the reliable strong correlation between *International Paper Company* and *Temple-Inland Incorporated* (the stable curve at the top, *IP* vs. *TIN*). On the right, the spurious correlation between *Phelps Dodge Corporation* and *Fidelity National Financial Incorporated*, represented by the unstable curve at the bottom, *PD* vs. *FNF*; the reliable strong correlation between *Centex Corporation* and *Pulte Homes Incorporated* represented by the stable curve at the top, *CTX* vs. *PHM*. $\Delta t = 84$ days in both plots.

and is extremely volatile – revealing that most correlations might be just spurious. When market uncertainty increases, which is when an efficient diversification is needed more, equities become particularly correlated and make the task of diversification itself particularly difficult (herd effect) [1].

In Figure 1 (right panel) we report an example of what is likely to be a ‘spurious’ correlation between *Walt Disney Corporation* (Services, Broadcasting & Cable TV) and *Walgreen Company* (Services, Retail Drugs)⁶: this is a highly volatile curve where, at the beginning, the correlation is very low, then following the 9/11 financial crisis it increases sharply and it remains high for a period of $\Delta t = 84$ days, after which it goes back to its initial low values; then, again, during the financial instability of mid 2002 the correlation shoots up very significantly and finally moves back to a relatively low value, when

the turbulence is over. Let us stress that there is no apparent connection between the two companies which are operative in very different sectors of activity: this, together with its large variations and its prolonged sensitiveness to anomalous data-points, suggests that the correlation is spurious. On the contrary, the very stable curve on top shows the strong correlation between *International Paper Company* (Basic Materials, Paper & Paper Product) and *Weyerhaeuser Company* (Basic Materials, Forestry & Wood Products). In this case, the curve remains persistently well above the market average, no matter what shocks shake the market; of course the two companies operate in sectors which are strongly connected (paper is produced out of wood), so their correlation can be reasonably assumed to be genuine, accurate, reliable and stable over time.

Two more cases are reported in Figure 2. On the left, the unstable curve at the bottom is the ‘spurious’ correlation between *Masco Corporation* (Consumer Cyclical, Furniture & Fixtures) and *AON Corporation* (Financial,

⁶ In this and the next two examples, correlations have been calculated for $\Delta t = 84$ market days corresponding to four months.

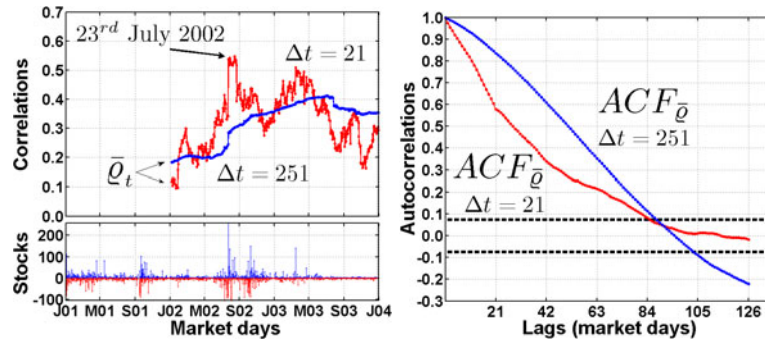


Fig. 3. (Color online) On the upper plot, on the left, the downward-pointing triangles represent the dynamic average correlation for a system of 300 stocks computed over a period of three years with a running window of twelve months ($\Delta t = 251$ days); the upward-pointing triangles represent the dynamic average correlation with a running window of one month ($\Delta t = 21$ days). On the lower plot, at each time t , the number of companies whose standardized returns is more than 2σ (positive axis) or less than -2σ (negative axis). On the right, the autocorrelation function for dynamic average correlations ($\bar{\rho}_t$) with $\Delta t = 21$ days and with $\Delta t = 251$ days vs. time lags from 1 to 126 days. The horizontal slashed lines are the upper and lower bounds for the null hypothesis.

Insurance Miscellaneous). Conversely, the stable curve at the top is the reliable strong correlation between *International Paper Company* and *Temple-Inland Incorporated* (Conglomerates). On the right, we report another ‘spurious’ correlation between *Phelps Dodge Corporation* (Basic Materials, Metal Mining) and *Fidelity National Financial Incorporated* (Financial, Insurance Property & Casualty) – the unstable curve at the bottom. A reliable strong correlation between *Centex Corporation* (Capital Good, Construction Services) and *Pulte Homes Incorporated* (Capital Good, Construction Services) is instead the stable curve at the top. Typically, we observe that outstanding market shocks greatly affect spurious correlations for a time lapse of Δt , while the same effect on reliable correlations is much weaker.

In general, correlations of return time series are characterized by high persistence and memory [41–43,53], which are also positively dependent on the number of observations. Figure 3, on the left, shows dynamic average correlations obtained for running windows $\Delta t = 21$ days, 1 month (highly volatile), and with $\Delta t = 251$ days, 12 months (more stable). Although the average has been calculated over all 44 850 coefficients of the correlation matrix, and for this reason its confidence bounds are very narrow, the effect due to the length of the running window is very evident from the figure in both cases.

Note that the shock recorded on the 23rd July 2002 is entirely reabsorbed in one month by the first curve, which reacts markedly on the day itself and after a period of $\Delta t = 21$ days; instead, in order to reabsorb the same shock, the second curve needs a period of $\Delta t = 251$ days. Note also the poor statistical significance of the correlation measured on the day when the data-point corresponding to the market shock exits from the running window. Each remarkable market shock, in fact, generates the illusion of a second similar shock after a period of Δt days: this shortcoming is particularly treacherous because it is additive and it makes it unpractical to interpret the real trend.

On the right of Figure 3 we plotted the autocorrelation functions for the linearly detrended time series of average

correlations⁷. We see that the persistence of $\bar{\rho}_t$ increases when the length of the running window is expanded. Note that for $\Delta t = 21$ days the autocorrelation has an homogeneous concave pattern between 1 and 21 lags and then a different pattern afterwards. For $\Delta t = 251$ days the autocorrelation is larger and the two functions cross each other after four months. Intriguingly they cross at a value that corresponds to the upper bound of the null hypothesis (dotted horizontal line).

4.2 Comparison between ρ and τ correlation matrices

As already argued in Section 2.2, Kendall τ correlation is a useful alternative to Pearson ρ . In the following, in order to better understand the advantages and disadvantages of these measures, we compare Pearson and Kendall correlations.

Pearson and Kendall measures usually provide quite different numerical outputs. We have calculated all dynamic correlation matrices for both ρ and τ , considering a running window of $\Delta t = 21$ market days and for all times $t = \{\Delta t, \Delta t + 1, \dots, T\}$. For each method, a total of $T - \Delta t + 1 = 728$ correlation matrices have been obtained, each matrix possessing $\frac{N(N-1)}{2} = 44850$ coefficients. On the left side of Figure 4 a low resolution scatter plot is reported, where the horizontal axis represents ρ coefficients and the vertical axis represents τ ’s: individual correlations vary considerably. Indeed, some very large and very small Pearson coefficients are associated to zero Kendall coefficients, and viceversa. Kendall correlations appear to persistently overestimate negative and underestimate positive Pearson correlations and the region has the shape of a parallelogram. In this figure we also plot for comparison the exact relation of $\tau = (2/\pi)\arcsin(\rho)$ (Greiner’s relation)

⁷ The autocorrelation function is defined as $ACF(h) = \frac{E[(X_t - \mu)(X_{t+h} - \mu)]}{\sigma_X^2}$ where here X_t is a correlation at time t minus the value of its linear best fit trend; h is the time lag and σ_X is the standard deviation of X .

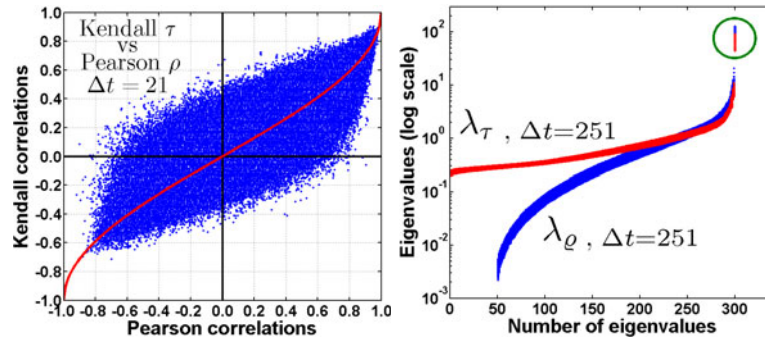


Fig. 4. (Color online) On the left, Kendall vs. Pearson correlations are represented for $N = 300$ NYSE stocks, with $\Delta t = 21$ days: Kendall correlations appear to persistently overestimate negative and underestimate positive Pearson correlations. The region has the unwonted shape of a parallelogram. On the right, the lower curves represent in logarithmic scale the eigenvalues of Pearson correlations (λ_ρ) computed with $\Delta t = 251$ days; the upper curves represent the eigenvalues of Kendall correlations (λ_τ), in logarithmic scale. Null eigenvalues for Pearson matrices have been discarded. Eigenvalues of the two measures have dissimilar structures. In the circle the largest eigenvalues of Pearson and Kendall matrices have been highlighted sharing the same values.

which holds for a large class of dependent signals with elliptical distributions [35,54,55].

The analysis of eigenvalues can provide a useful insight into the correlation matrices's structure. In fact, a large number of eigenvalues not significantly different from zero is indicative of a high noise level in the data. On the right of Figure 4, in logarithmic scale, eigenvalues are reported for Pearson and Kendall correlation matrices calculated with running windows of $\Delta t = 251$ days and for all times $t = \{\Delta t, \Delta t + 1, \dots, T\}$. For each method, a total of $T - \Delta t + 1 = 498$ correlation matrices have been obtained, each matrix associated to N eigenvalues. Eigenvalues are sorted in ascending order. Null eigenvalues, due to the singularity of Pearson matrices, have been discarded. In the figure, two distinct structures are easily recognizable: the set of 498 lower curves represents Pearson's eigenvalues, the set of 498 upper curves represents Kendall's eigenvalues, and the differences between the two structures are remarkable. Kendall lower eigenvalues, in fact, are generally much larger than the corresponding Pearson eigenvalues: about 49.4% of Pearson eigenvalues are smaller than the lowest of Kendall eigenvalues. In this example, none of the Kendall eigenvalues is null against $N - \Delta t + 1 = 300 - 250 = 50$ Pearson eigenvalues. Inside the circle on top, we have highlighted the largest eigenvalues – which retain most of the information – for both Pearson and Kendall matrices, and they are pretty similar. Pearson largest eigenvalue is never less than $0.210N$; Kendall's is never less than $0.146N$; the second largest eigenvalue is, respectively, never more than about $0.069N$ and $0.037N$. Let us recall that here we have $N = 300$. Similar results have been retrieved also using larger or smaller data sets. About 79.9% and 78.4% eigenvalues, respectively, are smaller than 1, showing in both cases a considerable amount of noise. Kendall eigenvalues are always convex in logarithmic scale while Pearson eigenvalues are initially concave and then convex⁸.

⁸ This result seems to be valid for more general cases, independently from the number of observations. In fact we have

Note that the Kendall correlation matrix has more structure than Pearson's and is also more robust from a statistical point of view: for this reason, τ_w might be suitable as shrinkage target for ρ_w , in a matrix shrinkage approach [56–61], so that the convex combination $\delta\rho_w + (1 - \delta)\tau_w$ might be more informational than τ_w from a quantitative point of view and at the same time more robust and with a superior structural endowment than ρ_w . The eigenvalues of the resulting *blend matrix* would be all positive. This is an intriguing approach that we are currently investigating.

In order to check the statistical significance of both ρ and τ correlations we have reported in Appendix I the distributions of the matrices's coefficients.

4.3 Effect of weights on empirical correlations

Weights have a very important smoothing effect on individual correlations, impacting also on the structure of correlation matrices, on the distribution of their coefficients and on the autocorrelations. In the following we assume weights with exponential smoothing according to equations (5) and (12), respectively for Pearson and Kendall methods.

4.3.1 Effects of weights on individual correlations

Figure 5 shows scatter plots of weighted vs. unweighted correlations for both Pearson (left) and Kendall (right) cases, calculated with a running window of $\Delta t = \{21, 251\}$ days and $\theta = \frac{\Delta t}{3}$. Individual correlations vary considerably but they are virtually symmetric with respect to the bisector line. In both cases the scatter plot assumes the shape of an “eye” with two cuspidal points at the extremities. Kendall weighted correlations slightly

verified it for many data-sets, both real and random, reaching systematically the same result.

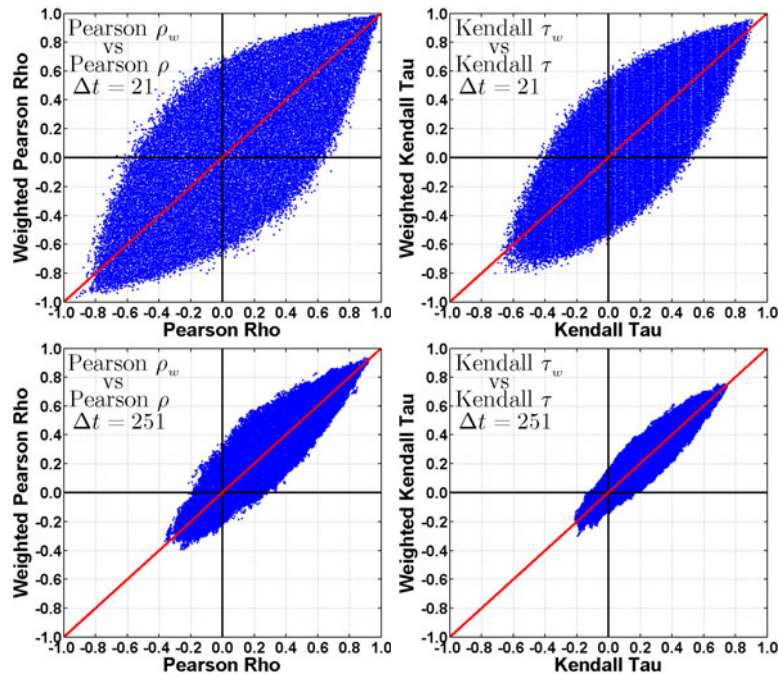


Fig. 5. (Color online) Weighted vs. unweighted correlations calculated with $\Delta t = \{21, 251\}$ days and $\theta = \frac{\Delta t}{3}$: the two measures are virtually symmetric with respect to the bisector line. On the left, Pearson's case has the shape of an eye with two cuspidal points at the extremities. On the right, Kendall's case, with the weighted measure slightly overestimating positive and underestimating negative values.

overestimate positive and underestimate negative values of the corresponding unweighted measure. The weights affect Pearson correlations more markedly than Kendall correlations - and this is due to the higher rank of Kendall matrix. When Δt is wider, the differences are reduced but still sizable.

4.3.2 Effects of weights on the matrix structure

The choice of parameter θ is crucial: if θ is too little all the weight is cast on recent events, individual correlations loose statistical significance, correlation matrices become numerically unstable and the distribution of their coefficients becomes distorted. On the other hand, if θ is too large, weights become uniform and, consequently, the weighting is ineffective.

In order to verify the impact of θ on the numerical stability of Pearson weighted correlation matrices, we check how the condition number⁹ reacts to decreasing values

⁹ The condition number of matrix A is defined as $\kappa(A) = \|A\| \cdot \|A^{-1}\|$ and is a measure of the loss in precision due to roundoff errors in Gaussian elimination. The higher the condition number is, the higher the matrix numerical instability becomes. If the condition number is much greater than 6.7×10^8 , caution is advised for subsequent computations. If $\kappa \propto 10^{16}$ the numerical instability is total and the matrix is ill-conditioned; if κ is small the matrix is well-conditioned. Given $\kappa \propto 10^k$, k is the expected number of digits of accuracy which would be presumably lost by solving the associated system of linear equations [62].

of the parameter. Let us recall that the condition number is defined for non-singular matrices and large values are indicative of numerical instability. However, since our Pearson correlation matrices are always singular, being $N > \Delta t$ for all lengths of the running windows considered, their condition number cannot be computed. Therefore, for $\Delta t = 251$ and for $\theta \in [\frac{\Delta t}{50}, \infty)$, we have reduced the singular matrices in row echelon form, detecting and keeping the linearly independent columns; then we have considered the largest full-rank sub-matrices and calculated their condition number. The results are reported on the left of Figure 6 where the vertical axis represents condition numbers in logarithmic scale, and the horizontal axis the ratio $\frac{\Delta t}{\theta}$: decreasing values of θ (higher values of $\frac{\Delta t}{\theta}$) are associated with an increasing deterioration of the matrix numerical stability. When θ is large enough, the slope is virtually flat. This suggests to refrain from excessively lowering the value of θ .

Since Kendall correlation matrix can be seen as a special case of Pearson correlation matrix the same considerations apply to it as well. However, Kendall matrices are usually endowed with a much higher rank and a much lower condition number than Pearson matrices¹⁰: therefore, the effect of θ on the condition numbers of the weighted Kendall correlation matrices for the cases

¹⁰ The superior numerical stability of Kendall correlation matrices can be explained by the fact that, usually, unless $N \gg \Delta t$, the following inequality holds: $\frac{\max(K, N)}{\min(K, N)} = \frac{\max(\Delta t(\Delta t - 1), 2N)}{\min(\Delta t(\Delta t - 1), 2N)} \gg \frac{\max(\Delta t - 1, N)}{\min(\Delta t - 1, N)}$.

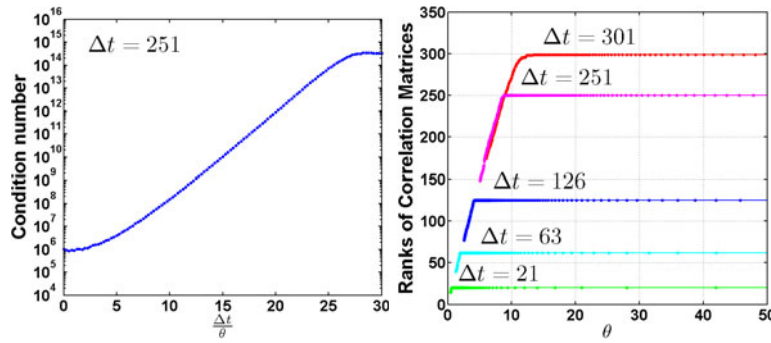


Fig. 6. (Color online) On the left, condition number of largest full-rank Pearson correlation sub-matrices as a function of parameter θ when $\Delta t = 251$. On the right, estimated ranks as a function of θ .

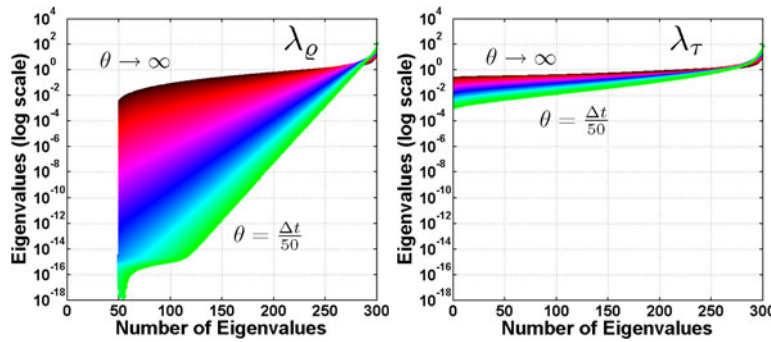


Fig. 7. (Color online) Changes of the Pearson (λ_ρ) and Kendall (λ_τ) average eigenvalues – respectively on the left and on the right – as a function of parameter θ with $\Delta t = 251$ days. The x axis reports the ordered index of eigenvalues sorted from the minimum to the maximum. The y axis reports the average (ordered) eigenvalues (in logarithm scale). Each curve is obtained as average of 498 vectors each representing 300 ordered eigenvalues of a correlation matrix calculated at given time and with fixed θ . The figure represents 501 of such curves, with different values of θ .

studied here, are negligible to all practical effects and, as such, they are not reported.

From a mere mathematical point of view, it is very unlikely that a loss of rank might ever occur as a consequence of an increase in parameter θ ; nonetheless, computationally, the *estimated* rank [62] can be lower than it should as a consequence of nearly null eigenvalues. On the right of Figure 6, the effect of θ on the estimated rank of Pearson correlation matrices is reported. Small values of θ cause a loss of the estimated rank for the matrices. There appears to be a linear relation between the full rank and the minimum θ that ensures it, with approximately rank $\approx \min(2 + 30\theta, \Delta t - 1, N)$. Computations have been performed in *MATLAB* using function “rank” with default tolerance.

We have also evaluated the eigenvalues of the correlation matrices with $\Delta t = 251$ and several values of θ in the range $[\frac{\Delta t}{50}, \infty)$. For all θ , we have computed the average eigenvalue over $T - \Delta t + 1$ dynamic correlation matrices. The results are illustrated in Figure 7 for Pearson (left) and Kendall (right) correlations. For Pearson correlations we observe that the original curve rotates downwards anti-clockwise so that, for large values of θ , only few eigenvalues are significant while all others are null or negligible. Conversely for Kendall eigenvalues (on the right) the associated eigenvalues do not suffer from the same sensitiveness to decreasing values of θ . This is a consequence of the fact that this correlation is computed over a very large number

of differences, namely $\frac{\Delta t(\Delta t - 1)}{2} = 31375$, therefore it has full rank with all positive eigenvalues.

4.3.3 Effects of weights on the distribution of coefficients

The choice of the values of Δt and θ has a strong effect on the statistical distribution of the correlation coefficients. In particular the distribution tends to a Gaussian when both θ and Δt are large. Conversely it deviates from the normal behavior becoming increasingly irregular when θ decreases. The distributions for various θ and Δt are reported in Appendix I. Let us here focus on the standard deviations as a function of parameter θ .

On top of Figure 8 are reported the standard deviations for Pearson (left) and Kendall (right) correlation coefficients when $\Delta t = 251$, computed for various θ , for $T - \Delta t + 1 = 498$ dynamic matrices. The thick curve in the middle corresponds to the average value of the standard deviations over the 498 dynamic matrices. We find that in the interval $[\frac{\Delta t}{3}, \infty)$ the general standard deviation has an almost stationary or a slightly decaying convex trend; in the $[\frac{\Delta t}{7}, \frac{\Delta t}{3}]$ interval, the curve is almost stationary or slightly increasing; after that, the curve is generally increasing and concave. A minimum is present in the $[\frac{\Delta t}{3}, \infty)$ interval and an inflection point can be individuated at about $\theta \approx \frac{\Delta t}{7}$.

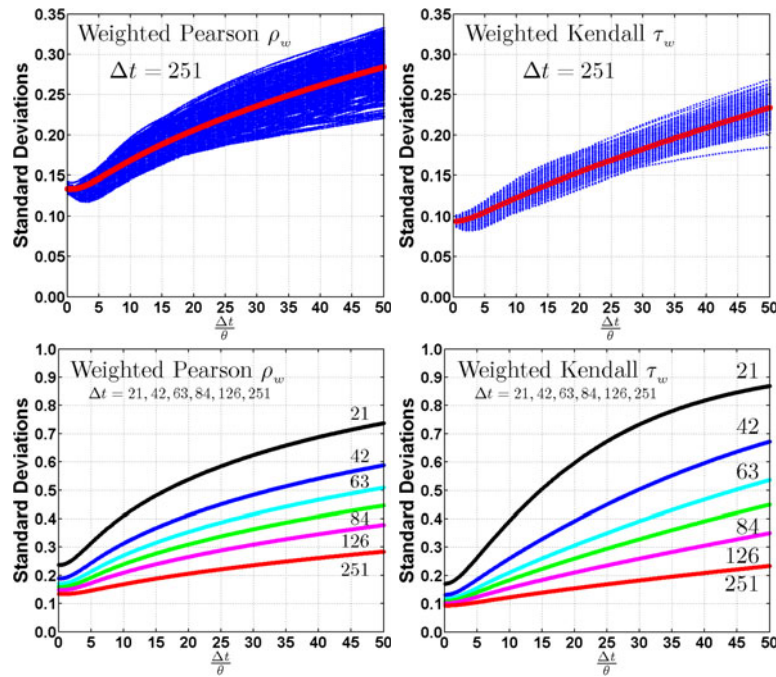


Fig. 8. (Color online) Standard deviation of weighted correlation coefficients as a function of parameter θ when $\Delta t = 251$ (Pearson on the left, Kendall on the right). On the top are reported $T - \Delta t + 1 = 498$ curves corresponding to all the dynamic matrices that can be computed over the period T . The thick line is the average standard deviation per each θ over these 498 curves. At the bottom the average standard deviations for $\Delta t = \{21, 42, 63, 84, 126, 251\}$ are reported.

Similar behaviors are observed for the average standard deviations computed for $\Delta t = \{21, 42, 63, 84, 126, 251\}$, which are reported on the bottom of Figure 8. We see that for all Δt the standard deviations persistently exhibit similar trends; the curves for the Kendall method (right) are characterized by lower levels at Δt and, viceversa, higher levels at smaller Δt . As observed before, the effect of weights is generally magnified by narrowing running windows: in fact, the curves shift fast upwards as Δt gets narrower.

The median's range for minimum points is in $\frac{\Delta t}{\theta} \in [0, \frac{1}{2}]$. The 95% confidence interval is about $\frac{\Delta t}{\theta} \in [0, \frac{7}{2}]$, with its upper bound increasing as Δt increases.

The median's range for inflection points is in $\frac{\Delta t}{\theta} \in [\frac{9}{2}, 7]$. The upper bound of the 95% confidence interval reaches as far as $\frac{\Delta t}{\theta} \approx 15$ and increases as Δt increases.

Confidence intervals of individual coefficients

In order to further assess the effect of Δt and θ on the statistical properties of the weighted correlation coefficients, we measure the uncertainty associated to these coefficients by estimating the width of their 100(1 - η)% confidence intervals by using bootstrap resampling. In particular we have applied the following procedure¹¹:

1. draw each sample by extracting time units with uniform probability, with replacement;

¹¹ Note that, in the special case when weights are the empirical frequency of the data, the correct resampling would be:

2. assign to each selected time unit its corresponding weight, e.g. according to equation (5) or (12);
3. re-normalize weights to one.

For all times $t = \{\Delta t, \Delta t + 1, \dots, T\}$, for both $\Delta t = 21$ and $\Delta t = 251$, we have randomly extracted 1000 samples of size Δt . For each sample, we have calculated the coefficients of weighted correlations. Then, for all coefficients, we have measured the width of the 95% confidence intervals ($CIW_{95\%}$). Figure 9 shows that the confidence intervals get increasingly wide for decreasing values of θ . When $\Delta t = 21$ the curves shift rapidly upwards, towards the maximum limit of 2. In Appendix K we verify this procedure by applying it to a multivariate Gaussian signal.

Given the effect of parameters θ and Δt over the distribution of correlation coefficients and the individual distributions of coefficients, calibrating the desired effect depends strongly on the purpose of specific analyses. In fact, a high standard deviation caused by a mighty smoothing can at the same time spoil the informative content of the measure or help differentiate individual coefficients. While excessively high values of θ are virtually useless, indiscernible from unweighted quantities to all practical effects, certainly very low values make

1. draw each sample by extracting time units with probability proportional to the assigned weights, e.g. according to equation (5) or (12), with replacement;
2. assign to each selected time unit a weight proportional to its frequency in the corresponding sample;
3. re-normalize weights to one.

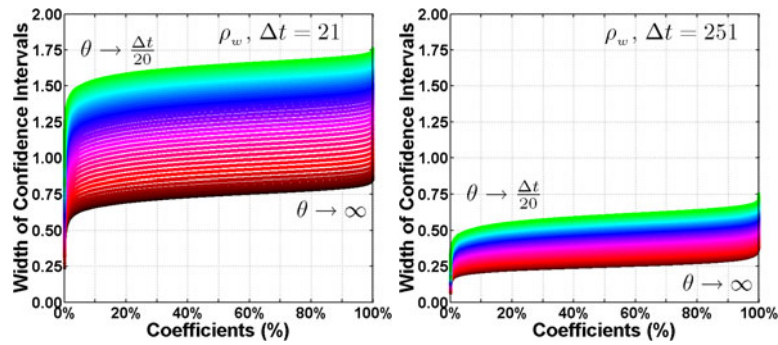


Fig. 9. (Color online) Width of 95% confidence intervals for all coefficients of Pearson weighted correlation matrices (based on a bootstrap sample of 1000 drawing from each time unit). General averages are displayed for $\Delta t = 21$ (left) and $\Delta t = 251$ (right). Weights vary from $\theta \rightarrow \infty$ to $\theta = \frac{\Delta t}{20}$. For each θ the intervals have been sorted in ascending order with respect to their width.

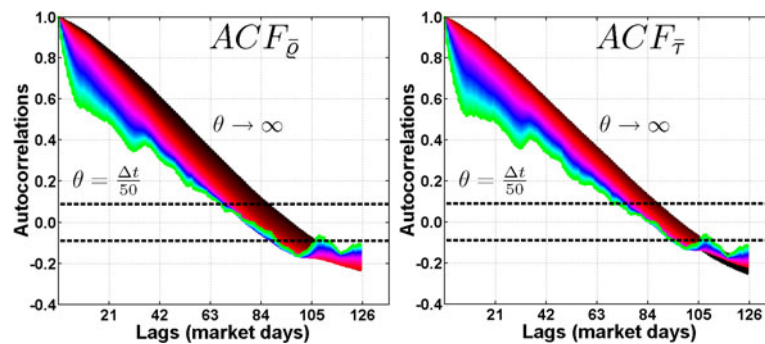


Fig. 10. (Color online) Autocorrelation functions for average weighted correlations computed for a running window of $\Delta t = 251$ days and different values of θ in the interval $[\frac{\Delta t}{50}, \infty)$. Pearson’s case is on the left, $ACF_{\bar{\rho}}$, Kendall’s on the right, $ACF_{\bar{\tau}}$. In the color, online version, the colors vary from black to red, magenta, blue, cyan and green accompanying the downwards shift of the curves as θ decreases. Horizontal dotted lines delimit approximate upper and lower confidence bounds for the null hypothesis, assuming that the series is a $MA(0)$ process. As θ gets shorter, the autocorrelation becomes statistically non-significant at a faster rate; negative values associated with extended lags are statistically significant but weak and possibly negligible.

the correlation matrix meaningless and consequently risky (see Appendix G for an extreme example). Generally, a cautious approach would suggest to favour the former rather than the latter of the two hazards.

4.3.4 Effect of weights on autocorrelations

We observe that weights have a strong effect on the autocorrelations contributing effectively to make a regular and convex autocorrelation function of correlation coefficients. Figure 10 reports the autocorrelation function of the average weighted dynamic correlations ($\bar{\rho}_t$ and $\bar{\tau}_t$) calculated with linear detrending for different values of θ and $\Delta t = 251$ (Pearson’s case on the left, Kendall’s on the right). When $\theta \rightarrow \infty$ (uniform weights), the autocorrelation function is at its highest level; then, as θ decreases, the autocorrelation function shifts increasingly downwards. Horizontal dotted lines delimit approximate upper and lower confidence bounds for the null hypothesis, assuming that the series is a $MA(0)$ process: it takes at least 63 lags for the lowest curve to fall below the upper bound while the highest curve crosses the same bound after about 84 lags. So, one of the beneficial effects of lowering θ , in agreement with results reported above, is

to reduce the improper trail of the remote past over the present.

4.3.5 Effect of weights on spurious correlations

The increase in temporal resolution associated with the exponential weights can be used to differentiate between ‘genuine’ and ‘spurious’ correlations. For instance, Figure 11 (top) shows how weighted Pearson average correlations, calculated with $\Delta t = 21$ days and $\theta = \frac{\Delta t}{3} = 7$ days, are scarcely sensitive to shocks occurred in the remote past and how, on the contrary, they record much higher peaks at present shocks. While normal – unweighted – Pearson average correlation (top left) is very sensitive to the shock of the 23rd of July 2002 on the day itself and after $\Delta t = 21$ days, remaining persistently on an unusual large value during the time interval, the weighted average correlation (top right) is very sensitive only on the day of the shock itself, recovering fast in the subsequent days.

At the bottom of Figure 11 differences between weighted and unweighted correlations for a selection of cases are reported for a running window of four months, $\Delta t = 84$ days, and $\theta = \frac{\Delta t}{3} = 28$ days. Highly volatile

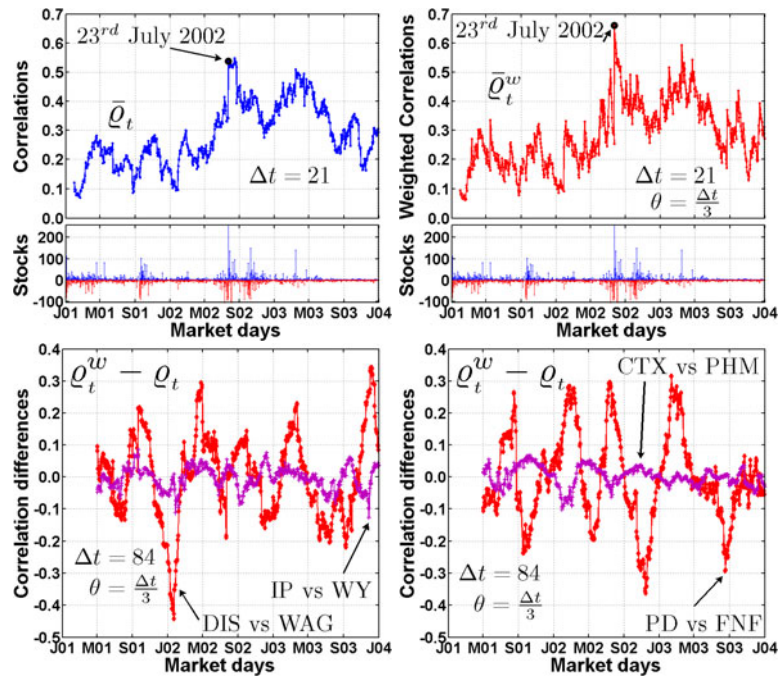


Fig. 11. (Color online) On the left, on top, the curve represents the dynamic average of Pearson correlations for 300 *NYSE* stocks computed with a running window of $\Delta t = 21$ days; on the right the curve represents the dynamic average of weighted correlations with Exponential Smoothing and $\theta = 7$. Patterns of the two measures are similar but weighted correlations are more sensitive with respect to the present and do not react significantly to shocks occurred in the past. Below, the difference between weighted and unweighted correlations: on the left, the spurious correlations between *Walt Disney Corporation* and *Walgreen Company* (volatile curve with diamond markers, *DIS* vs. *WAG*) and the strong correlation between *International Paper Company* and *Weyerhaeuser Company* (stable curve with “plus” markers, *IP* vs. *WY*); on the right the spurious correlation between *Phelps Dodge Corporation* and *Fidelity National Financial Incorporated* (volatile curve with diamond markers, *PD* vs. *FNF*) and the strong correlation between *Centex Corporation* and *Pulte Homes Incorporated* (stable curve with “plus” markers, *CTX* vs. *PHM*). When correlations are spurious, the difference between weighted and unweighted coefficients is very unstable over time. Correlations, for these examples, have been computed with $\Delta t = 84$ days and $\theta = \frac{\Delta t}{3} = 28$ days.

curves correspond to ‘spurious’ correlations between *Walt Disney Corporation* and *Walgreen Company* (left) and between *Phelps Dodge Corporation* and *Fidelity National Financial Incorporated* (right). Stable curves fluctuating between the zero line represent instead ‘genuine’ correlations between *International Paper Company* and *Weyerhaeuser Company* (left) and between *Centex Corporation* and *Pulte Homes Incorporated* (right). We see that when correlations are spurious, the introduction of exponential smoothing weights produces large differences, whereas when correlations are genuine the oscillations are less pronounced.

Analogous examples with Kendall correlations are reported in Figure 12: on top on the left, average Kendall correlations, on the right weighted Kendall average correlations, calculated with $\Delta t = 21$ days and $\theta = \frac{\Delta t}{3} = 7$ days. At the bottom of Figure 12, we report two examples of correlations, represented by volatile curves with “diamond” markers: (left), *News Corporation Limited* (Services – Printing & Publishing) and *Federated department stores* (Services – Retail Department & Discount); (right) *Omnicom Group* (Services – Advertising) vs. *Federated Department Stores* (Services – Retail Department & Discount). These couples of firms have no evident economic connections and they are presumably

connected by ‘spurious’ correlations and the differences between weighted and unweighted correlations are large. Conversely, the stable curves at the bottom of Figure 12, identified by “plus” markers, are two examples of ‘genuine’ correlations: (left) *International Paper Company* (Basic Materials – Paper & Paper Products) vs. *Weyerhaeuser Company* (Basic Materials – Forestry & Wood Products); (right) *Occidental Petroleum Corporation* (Energy – Oil & Gas Operations) vs. *Unocal Corporation* (Energy – Oil & Gas Operations). In both cases the companies are strongly connected from an economic point of view and the differences between weighted and unweighted correlations are small.

5 Some criteria for choosing θ

Let us recall that θ is the weights’ characteristic time and it represents the horizon in the past relevantly contributing to the correlation’s measure. The choice of its size depends on the aim of the investigation. We can generally affirm that small values of θ cause some *adverse* effects such as: an increase in the condition number of the correlation matrix; a fall of most eigenvalues to negligible figures; a fall in the estimated rank of the correlation matrix;

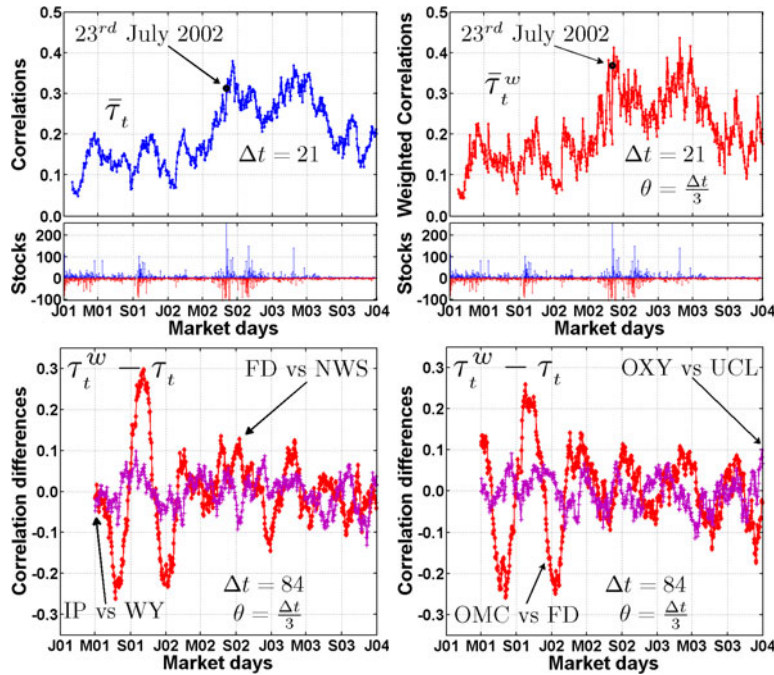


Fig. 12. (Color online) On top, on the left, dynamic average of Kendall correlations calculated with $\Delta t = 21$; on the right the dynamic average of weighted Kendall correlations with exponential smoothing, $\theta = \frac{\Delta t}{3} = 7$ days. At the bottom, with $\Delta t = 84$ days and $\theta = \frac{\Delta t}{3} = 28$ days, the differences between weighted and unweighted Kendall correlations. Spurious correlations: (left) *News Corporation Limited (NWS)* and *Federated Department Stores (FD)*; (right) *Omnicom Group (OMC)* and *Federated Department Stores (FD)*. Genuine correlations: (left) *International Paper Company (IP)* and *Weyerhaeuser Company (WY)*; (right) *Occidental Petroleum Corporation (OXY)* and *Unocal Corporation (UCL)*.

a distortion of the coefficients' distribution and in particular an increase in their standard deviations. Nevertheless, at the same time, small values of θ also cause desirable effects such as: a less persistent autocorrelation function and a far less undue influence of remote events over the understanding of current events. Consequently, a trade-off exists between avoiding adverse effects and achieving desirable effects. Therefore the choice of θ should be the result of a multi-objective optimization process.

In the following we discuss some criteria that can be taken into consideration when choosing the values of θ for a given Δt .

From the analysis of the condition number, reported in Figures 6, we see that, when $\Delta t = 251$ days, it is rather stationary for large values of θ , but it suddenly grows fast when θ is shortened. So a first criterion that may help determine θ is to set a threshold to the condition number. For instance a threshold at 10^8 in the example in Figures 6 will yield to $\Delta t/\theta \approx 10$ (i.e. $\theta \approx 25$ days, about a month).

From Figures 7, we see that, by increasing θ , the lowest non-null eigenvalues fall steeply to negligible values. Then a second concurrent criterion might be to set a threshold on the minimum non-null eigenvalue so to obtain a structure that is not excessively noisy. For instance in the example reported in the figure, $\lambda_{min} \geq 10^{-4}$ corresponds to a value of $\Delta t/\theta \approx 8$ and instead $\lambda_{min} \geq 10^{-8}$ corresponds to a value of $\Delta t/\theta \approx 19$ (i.e. respectively $\theta \approx 31$ and 13 days).

While high condition numbers and null eigenvalues are clearly not advisable, whether a high or low standard deviation of correlation coefficients is desirable or not is less evident. Two contrasting aspects can be outlined: on one hand, it is wise to be cautious and reduce θ only the strictly necessary to avoid an excessive influence of recent events in the computation and an excessive distortion of the distribution of coefficients; on the other hand the more diverse the coefficients are (i.e. the higher their standard deviations), the easier the task of differentiating them, which is an element in favor of strongly reducing θ . Therefore two alternative criteria might be helpful, depending on research goals, and integrating the previous criteria:

1. Choose the highest value of θ with $\theta \leq \Delta t$ such that the least increase of coefficients' standard deviation (or the greatest decrease of it) is caused. Applied to the examples discussed in this paper, this corresponds to $\Delta t/\theta \approx 3$ (i.e. $\theta \approx 84$ days).
2. Choose θ in the proximity of the inflection point (as in Fig. 8) so as to maximize the increase rate of the coefficients' standard deviation, and restrain from reducing θ further. Applied to the examples discussed in this paper this corresponds to $\Delta t/\theta \approx 7$ (i.e. $\theta \approx 36$ days).

Let us note that, in the previous examples ($\Delta t = 251$ days), the increase in the standard deviation is relatively small, with average standard deviation ranging between 0.14 to 0.29 for Pearson coefficients and between

0.10 to 0.23 for Kendall coefficients, with $\frac{\Delta t}{\theta} = [3, 50]$ (see Fig. 9).

It should be also reminded that low values of θ and Δt cause an increase in the variance of individual coefficients so that the accuracy of the measures might be compromised, however, in the examples studied here, this effect appears to be relatively small.

6 Conclusion

Weighted correlations can fix some of the typical shortcomings affecting unweighted correlations – especially in the field of time series analysis. In particular, a structure of weights can make correlations more robust with respect to outliers, anomalous or remote data-points. The exponential smoothing criterion has the added advantage of significantly reducing the persistence/memory of dynamic average correlations which, by construction, is usually high. Furthermore, extending Δt , while at the same time reducing θ , might improve the correlation matrix numerical stability (rank and/or condition number of the largest full-rank sub-matrix). This strategy can thus help mitigate the curse of dimensionality by allowing the use of more data-points, when these are available, without the inconvenience of conferring remote events more importance than they are worth of for the research specific goals. Finally, comparing weighted vs unweighted measures can also help diagnosing spuriousness in the correlations.

However, assigning weights to observations is a delicate procedure potentially originating side effects which might offset other benefits. For this reason, several criteria for determining the best weights have been discussed. In particular, weights should be chosen having regard for the numerical stability and statistical robustness of the resulting correlation matrices: they should not cause excessive distortions in the distribution of coefficients, nor cause a collapse in that of eigenvalues, nor decrease the estimated rank (i.e. the number of estimated non-null eigenvalues) or increase excessively the condition number of the largest full-rank sub-matrices.

In general, unless $N \gg \Delta t$, due to its higher numerical stability, Kendall's correlation matrix is much less affected by these side effects than Pearson's.

In our analysis, we have found that $\theta = \frac{\Delta t}{3}$ is a reasonable choice for our data-set when Δt is sufficiently large, e.g. $\Delta t \in [84, 251]$: in this interval $\theta = \frac{\Delta t}{3}$ does not appear to overly affect the distribution of coefficients and the distribution of eigenvalues for neither Pearson nor Kendall correlations; although not enough to make convex the autocorrelation function of average correlations, it contributes to lower the curve; furthermore, it does not affect the numerical stability of the resulting correlation matrices: the condition number of the largest full rank sub-matrices does not increase significantly; finally, the estimated ranks of correlation matrices are not affected.

Future works will be dedicated to investigate how different weights, including an hyperbolic specification for the weights, affect properties and structure of graphs used for filtering correlation matrices, such as Minimum

Spanning Trees and Planar Maximally Filtered Graphs, on the same research line of previous works [63–67].

This work was partially supported by COST MP0801 project. Many thanks to the referees of this paper for their very useful comments.

Appendix A: Positive semi-definiteness of ρ

A useful notation allows us to write Pearson correlations in matrix form. Indeed, for a given matrix \mathbf{Y} , Pearson correlations can be written as:

$$\rho = \mathbf{D}^{-\frac{1}{2}} \boldsymbol{\Sigma} \mathbf{D}^{-\frac{1}{2}} \quad (\text{A.1})$$

with

$$\boldsymbol{\Sigma} = \mathbf{Y}^T (\mathbf{I}_{\Delta t} - \mathbf{U}_{\Delta t})^T (\mathbf{I}_{\Delta t} - \mathbf{U}_{\Delta t}) \mathbf{Y} \quad (\text{A.2})$$

and

$$\mathbf{D} = \boldsymbol{\Sigma} \circ \mathbf{I}_N \quad (\text{A.3})$$

where $[\dots]^T$ denotes the transpose, $\mathbf{U}_{\Delta t}$ denotes a square matrix of 1s of order Δt , \mathbf{I}_N the identity matrix of order N , $\mathbf{I}_{\Delta t}$ the identity matrix of order Δt , the symbol “ \circ ” is the Hadamard product [68–77]¹². $\boldsymbol{\Sigma}$ is the covariance matrix. \mathbf{D} is a matrix containing sample variances on the diagonal and zero elsewhere, such that $[d_{ii}] = \sigma_i^2$ and $[d_{ij}]_{i \neq j} = 0$; $\mathbf{D}^{-\frac{1}{2}}$ is a matrix containing reciprocals of sample standard deviations on the diagonal and zero elsewhere.

We see that, since $\boldsymbol{\Sigma}$ is a quadratic form of the type $\mathbf{X}^T \mathbf{X}$, it must be a positive semi-definite matrix; \mathbf{D} is an element-wise product of positive semi-definite matrices (it is a diagonal matrix of variances, which are always non-negative), so \mathbf{D} also must be a positive semi-definite matrix¹³. Given that ρ is a product of positive semi-definite matrices, it is also positive semi-definite.

Note that, considering standardized variables, $z_t^i = \frac{y_t^i - \bar{y}^i}{\sigma_i}$, Pearson's correlations can be written as $\rho_{ij} = \frac{\mathbf{z}^{iT} \mathbf{z}^j}{\Delta t}$ and the Euclidean distance between the two standardized variables is related to the correlation coefficient as:

$$\begin{aligned} d_{ij} &= \sqrt{\frac{1}{\Delta t} \sum_{s=1}^{\Delta t} (z_s^i - z_s^j)^2} \\ &= \sqrt{\frac{1}{\Delta t} \sum_{s=1}^{\Delta t} (z_s^i)^2 + \frac{1}{\Delta t} \sum_{s=1}^{\Delta t} (z_s^j)^2 - \frac{2}{\Delta t} \sum_{s=1}^{\Delta t} z_s^i z_s^j} \\ &= \sqrt{2(1 - \rho_{ij})} \end{aligned} \quad (\text{A.4})$$

which is defined in the interval $[0, 2]$ [78].

¹² The Hadamard product (or Schur product) of $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$, denoted as $\mathbf{C} = \mathbf{A} \circ \mathbf{B}$, is the element-wise product such that $[c_{ij}] = [a_{ij} b_{ij}]$. The Hadamard product allows us to prove the positive semi-definiteness of Pearson and Kendall correlation matrices, both weighted and unweighted.

¹³ Schur [69] proved that if \mathbf{A} and \mathbf{B} are positive semi-definite then also $\mathbf{C} = \mathbf{A} \circ \mathbf{B}$ is positive semi-definite.

Appendix B: Kendall's τ matrix notation and its rank

Let us first show that Kendall's τ coefficients can be written in matrix notation.

Posed $K = \binom{\Delta t}{2} = \frac{\Delta t(\Delta t - 1)}{2}$, let \mathbf{S} be a K -by- N matrix such that $s_{ki} = \text{sgn}(y_t^i - y_{t-h}^i)$ and each index $k \in \{1, 2, \dots, K\}$ is associated to time $t \in \{2, 3, \dots, \Delta t\}$ and interval $h \in \{1, \dots, t - 1\}$ through the relation $k(t, h) = \frac{1}{2}t(t - 1) - h + 1$, so that $t(k) = \left\lfloor \frac{1}{2} + \frac{1}{2}\sqrt{1 + 8(k - 1)} \right\rfloor + 1$, where $\lfloor \bullet \rfloor$ denotes the floor function, and $h[k, t(k)] = \frac{1}{2}t(k)(t(k) - 1) - k + 1$. Then τ can be expressed as

$$\tau = \mathbf{D}^{-\frac{1}{2}} \boldsymbol{\Sigma} \mathbf{D}^{-\frac{1}{2}} \quad (\text{B.1})$$

with

$$\boldsymbol{\Sigma} = \mathbf{S}^T \mathbf{S} \quad (\text{B.2})$$

and

$$\mathbf{D} = \boldsymbol{\Sigma} \circ \mathbf{I}_K. \quad (\text{B.3})$$

Therefore, τ can be seen as a Pearson's ρ correlation applied to \mathbf{S} ¹⁴. The rank of the Kendall Correlation Matrix is no more than the minimum between K and N :

$$\text{rank}(\tau) \leq \min \left\{ \frac{\Delta t(\Delta t - 1)}{2}, N \right\} \quad (\text{B.4})$$

so that, given N variables, in order to obtain a full rank matrix, the number of required observations should be such that:

$$\begin{aligned} \Delta t &\geq \left\lceil \frac{1}{2} + \sqrt{2N + \frac{1}{4}} \right\rceil = \left\lceil \frac{1}{2} + \sqrt{2N} \right\rceil \\ &= \left\lceil \frac{1}{2} + \sqrt{2N} + 1 \right\rceil = \text{round}(\sqrt{2N} + 1) \end{aligned} \quad (\text{B.5})$$

where $\lceil \bullet \rceil$ denotes the ceiling function and $\text{round}(\bullet)$ denotes the rounding to the nearest integer¹⁵. It follows that:

$$\Delta t > \text{round}(\sqrt{2N}). \quad (\text{B.6})$$

The proof for equalities in equation (B.5) are reported afterwards in Appendix D.

Appendix C: Positive semi-definiteness of ρ_w

Pearson weighted correlation matrices enjoy the same properties as their corresponding unweighted matrices: they are positive semi-definite matrices and this can be easily proved by representing them in matrix notation.

¹⁴ Note, though, that \mathbf{S} has not been subtracted by its sample mean, so no degree of freedom has been lost. This is the reason why the rank of τ can be $\frac{\Delta t(\Delta t - 1)}{2}$ when N is larger.

¹⁵ Properties and some proofs of similar diophantine equations involving floor and ceiling functions can be found in [79–81].

Let \mathbf{w} be a Δt -by-1 vector with non-negative elements such that $\mathbf{w} \geq \mathbf{0}$; let $\mathbf{u}_{\Delta t}$ be a Δt -by-1 vector of 1s; then condition

$$\mathbf{w}^T \mathbf{u}_{\Delta t} = 1 \quad (\text{C.1})$$

must be satisfied. Let $\bar{\mathbf{Y}} = \mathbf{w}^T \mathbf{Y}$ be the 1-by- N vector of weighted sample means. Let us introduce a vector \mathbf{w}^* whose elements are the square roots of \mathbf{w} 's elements, such that $w_t^* = \sqrt{w_t}$, $\forall t$; since $\mathbf{w} \geq \mathbf{0}$, this is also true for \mathbf{w}^* . Let \mathbf{u}_N be an N -by-1 vector of 1s. Then the weighted covariance matrix can be written as:

$$\begin{aligned} \boldsymbol{\Sigma}_w &= \left[(\mathbf{Y} - \mathbf{u}_{\Delta t} \bar{\mathbf{Y}}) \circ (\mathbf{w}^* \mathbf{u}_N^T) \right]^T \\ &\quad \times \left[(\mathbf{Y} - \mathbf{u}_{\Delta t} \bar{\mathbf{Y}}) \circ (\mathbf{w}^* \mathbf{u}_N^T) \right] \end{aligned} \quad (\text{C.2})$$

$\boldsymbol{\Sigma}_w$ is a quadratic form so it must be a positive semi-definite symmetric matrix. Let \mathbf{I}_N be an N -by- N identity matrix. By using the Hadamard product again, let $\mathbf{D}_w = \boldsymbol{\Sigma}_w \circ \mathbf{I}_N$ be a square diagonal positive semi-definite matrix whose elements along the diagonal are weighted sample variances; then, finally the weighted correlation matrix is:

$$\rho_w = \mathbf{D}_w^{-\frac{1}{2}} \boldsymbol{\Sigma}_w \mathbf{D}_w^{-\frac{1}{2}} \quad (\text{C.3})$$

where $\mathbf{D}_w^{-\frac{1}{2}}$ is a matrix containing reciprocals of weighted sample standard deviations on the diagonal and zero elsewhere. Since it can be written as a product of positive semi-definite matrices and it can be written as a quadratic form, the weighted correlation matrix must be symmetric and positive semi-definite.

Note that ρ_w , as Pearson correlation matrix, is invariant to arbitrary affine transformations so that, if the unit system of each column of \mathbf{Y} changes as $z^i = a^i y + b^i$ where parameters $a^i, b^i \in \mathbb{R}$ and $a^i > 0$, $\forall i$, then the weighted correlation remains unchanged. In fact, as equation (C.3) makes it clear, ρ_w is still a linear operator.

Appendix D: Equalities of equations in (B.5)

In order to obtain a full rank Kendall Correlation Matrix, the following necessary – but not sufficient – condition needs to be satisfied:

$$\text{rank}(\tau) \leq \min \left\{ \frac{\Delta t(\Delta t - 1)}{2}, N \right\}. \quad (\text{D.1})$$

The minimum number of observations that is necessary for a full rank matrix is obtained when $\frac{\Delta t(\Delta t - 1)}{2} = N$ so that $\Delta t(\Delta t - 1) - 2N = 0$ which is solved for $\Delta t = \frac{1}{2} + \sqrt{2N + \frac{1}{4}}$. Since $\Delta t \in \mathbb{N}$, the ceiling of the right-hand side of the equation must be considered instead, therefore $\Delta t \geq \left\lceil \frac{1}{2} + \sqrt{2N + \frac{1}{4}} \right\rceil$. Now, we prove that

$$\left\lceil \frac{1}{2} + \sqrt{2N + \frac{1}{4}} \right\rceil = \left\lceil \frac{1}{2} + \sqrt{2N} \right\rceil = m \quad (\text{D.2})$$

where $m \in \mathbb{N}$. When $N = 1$ the equality certainly holds and $m = 2$. So we are after solutions for $m \geq 2$. From left-hand side of equation (D.2), after removing the ceiling operator, follows that

$$\begin{cases} \frac{1}{2} + \sqrt{2N + \frac{1}{4}} \leq m \\ \frac{1}{2} + \sqrt{2N + \frac{1}{4}} + 1 > m \end{cases} \Rightarrow \begin{cases} 2N + \frac{1}{4} \leq (m - \frac{1}{2})^2, m \geq \frac{1}{2} \\ 2N + \frac{1}{4} > (m - \frac{3}{2})^2, m \geq \frac{3}{2} \end{cases} \Rightarrow \begin{cases} N \leq \frac{m(m-1)}{2} \\ N > \frac{m^2-3m+2}{2} \end{cases} \Rightarrow \begin{cases} N \leq \frac{m(m-1)}{2} \\ N \geq \frac{m^2-3m+4}{2} \end{cases}.$$

In order to verify that $N > \frac{m^2-3m+2}{2}$ implies $N \geq \frac{m^2-3m+4}{2}$ it suffices to note that

- $m^2 - 3m$ is even, so that, since m is a positive integer, $\frac{m^2-3m+2}{2}$ is a non-negative integer, so N cannot be less than $\frac{m^2-3m+2}{2} + 1$;
- when $N = 4$, then $m = 4$ and $\frac{m^2-3m+4}{2} = 4$, so the “>” sign must change into “ \geq ” at least in this case.

So, from left-hand side of equation (D.2) we obtain:

$$\frac{m^2 - 3m + 4}{2} \leq N \leq \frac{m(m-1)}{2}. \quad (\text{D.3})$$

Analogously, from right-hand side of equation (D.2), after removing the ceiling operator, follows that

$$\begin{cases} \frac{1}{2} + \sqrt{2N} \leq m \\ \frac{1}{2} + \sqrt{2N} + 1 > m \end{cases} \Rightarrow \begin{cases} 2N \leq (m - \frac{1}{2})^2, m \geq \frac{1}{2} \\ 2N > (m - \frac{3}{2})^2, m \geq \frac{3}{2} \end{cases} \Rightarrow \begin{cases} N \leq \frac{m(m-1)}{2} + \frac{1}{8} \\ N > \frac{m^2-3m+2}{2} + \frac{1}{8} \end{cases} \Rightarrow \begin{cases} N \leq \frac{m(m-1)}{2} \\ N \geq \frac{m^2-3m+4}{2} \end{cases}.$$

The last pair of implications are straightforward and are due to the fact that $m(m-1)$ and $m^2 - 3m + 2$ are even. So the interval in equation (D.3) is found solving both sides of equation (D.2) and this completes the proof. In order to verify that the solution is exact it is straightforward to note that, by substituting the bounds found in equation (D.3), one at a time, back in each side of equation (D.2), the relation still holds in all of the four resulting cases.

Appendix E: Proof of equation (8)

By solving for w_0 , we find $w_0(\alpha) = \frac{1}{\sum_{t=1}^{\Delta t} e^{\alpha(t-\Delta t)}}$. Note

that $\sum_{t=1}^{\Delta t} e^{\alpha(t-\Delta t)} = \frac{e^\alpha}{e^\alpha - 1} [1 - e^{-\alpha\Delta t}]$, hence equation (8)

follows. In fact:

$$\begin{aligned} \sum_{t=1}^{\Delta t} e^{\alpha(t-\Delta t)} &= e^{-\alpha\Delta t} \sum_{t=1}^{\Delta t} (e^\alpha)^t = e^{-\alpha\Delta t} \left[\frac{1 - (e^\alpha)^{\Delta t+1}}{1 - (e^\alpha)} - 1 \right] \\ &= \frac{e^{-\alpha\Delta t}}{1 - e^\alpha} [e^\alpha - e^{\alpha\Delta t+\alpha}] = \frac{e^\alpha}{e^\alpha - 1} [1 - e^{-\alpha\Delta t}]. \end{aligned}$$

Appendix F: MATLAB code for weighted correlations

In this paper computations have been made in *MATLAB* using the following codes, respectively for Pearson and Kendall correlation matrices.

Pearson Weighted Correlation Matrix

```
function R = weightedcorr(Y, w)
% dt: number of observations in window;
% N: number of variables
[dt, N] = size(Y);
% Remove weighted mean
temp = Y - repmat(w * Y, dt, 1);
% Weighted Covariance Matrix
temp = temp' * (temp .* repmat(w, 1, N));
% Must be exactly symmetric
temp = 0.5 * (temp + temp');
% Variances
R = diag(temp);
% Matrix of Weighted Correlation Coefficients
R = temp ./ sqrt(R * R');
```

where \mathbf{Y} is a dt -by- N array and \mathbf{w} is a dt -by-1 array such that $\text{sum}(\mathbf{w}) = 1$.

Kendall Weighted Correlation Matrix

```
function tau = kendalltau(Y, w)
% dt: number of observations in window;
% N: number of variables
[dt, N] = size(Y);
% Indexes for all dt * (dt - 1) / 2 combinations
without repetition:
% {1, 2}, {1, 3}, ..., {1, dt}, {2, 3}, {2, 4}, ...,
{2, dt}, ..., {dt - 1, dt}
[i2, i1] = find(tril(ones(dt, 'uint8'), -1));
% Signs of differences between variables
at time i2 and at time i1
tau = sign(Y(i2, :) - Y(i1, :));
% Number of concordant/discordant pairs (weighted)
tau = tau' * (tau .* repmat(w, 1, N));
% Must be exactly symmetric
tau = 0.5 * (tau + tau');
% Number of pairs minus number of ties (weighted)
temp = diag(tau);
% Matrix of Weighted Correlation Coefficients
tau = tau ./ sqrt(temp * temp');
```


where \mathbf{Y} is a dt -by- N array and \mathbf{w} is a $\frac{dt(dt-1)}{2}$ -by-1 array such that $\text{sum}(\mathbf{w}) = 1$. To be compatible with the function, indices for \mathbf{w} must be the same as in [i2, i1].

Note that the code for Kendall correlations, unless dt is extremely large, is generally faster than *MATLAB* native *corr* function; nevertheless, since it stores a $\frac{dt(dt-1)}{2}$ -by- N matrix, it will generate an ‘‘Out of Memory’’ error message if the product $N \frac{dt(dt-1)}{2}$ is very large.

The following codes generate the exponential weights for Pearson and Kendall correlations, compatible with their relative functions above.

Exponential Weights (Pearson)

```
function w = expweights(dt, theta)
% Calculate constant w0
w0 = (1 - exp(-1 / theta)) / (1 - exp(-dt / theta));
% Calculate exponential weights w, for Pearson
w(:, 1) = w0 * exp(((1:dt) - dt) / theta);
% Ensure sum of weights is 1
w = w / sum(w);
```

where θ is parameter θ in the text.

Exponential Weights (Kendall)

```
function w = kendallexpweights(dt, theta)
% Calculate constant w0
w0 = (exp(1/theta) + 1) * ((exp(1/theta) - 1)^2) /
      exp(2/theta) / (1 - exp(-dt/theta)) /
      (1 - exp((1 - dt)/theta));
% Indexes for all dt * (dt - 1) / 2 combinations
% without repetition:
% {1, 2}, {1, 3}, ..., {1, dt}, {2, 3}, {2, 4}, ...,
% {2, dt}, ..., {dt - 1, dt}
[i2, i1] = find(tril(ones(dt, 'uint8'), -1));
% Calculate exponential weights w, for Kendall
w(:, 1) = w0 * exp((i1 + i2 - 2 * dt) / theta);
% Ensure sum of weights is 1
w = w / sum(w);
```

Appendix G: Sign of weighted correlation when $\theta \rightarrow 0$

Note that, when $\theta \rightarrow 0$, only the last event remains significant and $\lim_{\alpha \rightarrow +\infty} \mathbf{w}(\alpha) = (0, 0, \dots, 0, 1)^T$ while $\lim_{\alpha \rightarrow +\infty} \rho_{ij}^w[\mathbf{w}(\alpha)] = \pm 1$, where the sign of the corresponding unweighted correlation is *not* necessarily preserved. In order to prove that signs are not preserved at their limits, a single counter-example suffices: let $\mathbf{y}^1 = [0, 1, 2]$ and $\mathbf{y}^2 = [0, 2, 1]$; then it is straightforward to verify that the unweighted correlation is equal to $\frac{1}{2}$ while the limit of the weighted correlation, when we allow $\alpha \rightarrow +\infty$, is equal to -1 . In fact, it is easy to prove that, in our counter-example, after some factorizations and simplifications,

such limit can be written as:

$$\begin{aligned} \lim_{\alpha \rightarrow +\infty} \rho_{12}^w[\mathbf{w}(\alpha)] &= \lim_{\alpha \rightarrow +\infty} \frac{\sigma_{12}^w[\mathbf{w}(\alpha)]}{\sigma_1^w[\mathbf{w}(\alpha)] \sigma_2^w[\mathbf{w}(\alpha)]} \\ &= \lim_{\alpha \rightarrow +\infty} \frac{e^{-\alpha} (2e^{-2\alpha} + 2e^{-\alpha} - 1)}{(e^{-2\alpha} + e^{-\alpha} + 1)^2} \\ &= \lim_{\alpha \rightarrow +\infty} \frac{\sqrt{\frac{e^{-\alpha}(e^{-2\alpha} + 4e^{-\alpha} + 1)}{(e^{-2\alpha} + e^{-\alpha} + 1)^2}} \sqrt{\frac{e^{-\alpha}(4e^{-2\alpha} + e^{-\alpha} + 1)}{(e^{-2\alpha} + e^{-\alpha} + 1)^2}}}{2e^{-2\alpha} + 2e^{-\alpha} - 1} = -1. \end{aligned}$$

Appendix H: Positive semi-definiteness of $\tau_{\mathbf{w}}$

By using the matrix notation for the weighted Kendall correlations and defining $K = \binom{\Delta t}{2} = \frac{\Delta t(\Delta t - 1)}{2}$, and \mathbf{w} a K -by-1 vector with non-negative elements: $\mathbf{w} \geq \mathbf{0}$; let $\mathbf{u}_{\mathbf{K}}$ be a K -by-1 vector of 1s; then condition

$$\mathbf{w}^T \mathbf{u}_{\mathbf{K}} = 1 \quad (\text{H.1})$$

must be satisfied. Let us introduce a vector \mathbf{w}^* whose elements are the square roots of \mathbf{w} 's elements and, since $\mathbf{w} \geq \mathbf{0}$, this is also true for \mathbf{w}^* . Let $\mathbf{u}_{\mathbf{N}}$ be an N -by-1 vector of 1s. Then, recalling matrix \mathbf{S} , from Appendix B, we can write the Weighted Kendall Correlation Matrix as

$$\tau_{\mathbf{w}} = \mathbf{D}_{\mathbf{w}}^{-\frac{1}{2}} \boldsymbol{\Sigma}_{\mathbf{w}} \mathbf{D}_{\mathbf{w}}^{-\frac{1}{2}} \quad (\text{H.2})$$

where

$$\boldsymbol{\Sigma}_{\mathbf{w}} = \left[\mathbf{S} \circ \left(\mathbf{w}^* \mathbf{u}_{\mathbf{N}}^T \right) \right]^T \left[\mathbf{S} \circ \left(\mathbf{w}^* \mathbf{u}_{\mathbf{N}}^T \right) \right] \quad (\text{H.3})$$

and

$$\mathbf{D}_{\mathbf{w}} = \boldsymbol{\Sigma}_{\mathbf{w}} \circ \mathbf{I}_{\mathbf{N}} \quad (\text{H.4})$$

$\tau_{\mathbf{w}}$ can be seen as a special kind of Weighted Pearson correlation applied to matrix \mathbf{S} ¹⁶ and it has the same properties as in equation (B.1) – it is symmetric, positive semi-definite, with rank no more than the minimum between K and N .

Appendix I: Coefficients' distribution for ρ and τ

The statistical significance of both Pearson and Kendall correlations is greatly affected by the choice of the length of the running window, Δt , as are the distributions of the matrix coefficients. We have calculated dynamic correlations for $\Delta t = \{21, 42, 63, 84, 126, 251\}$ market days, corresponding to about 1, 2, 3, 4, 6, 12 months. For each running window and for all times $t = \{\Delta t, \Delta t + 1, \dots, T\}$ and for both methods, a total of $T - \Delta t + 1 = \{728, 707, 686, 665, 623, 498\}$ matrices have been computed. Then the distributions of all coefficients associated to a given Δt have

¹⁶ It has to be reminded, though, that \mathbf{S} is not centered, as it has not been subtracted by its sample mean.

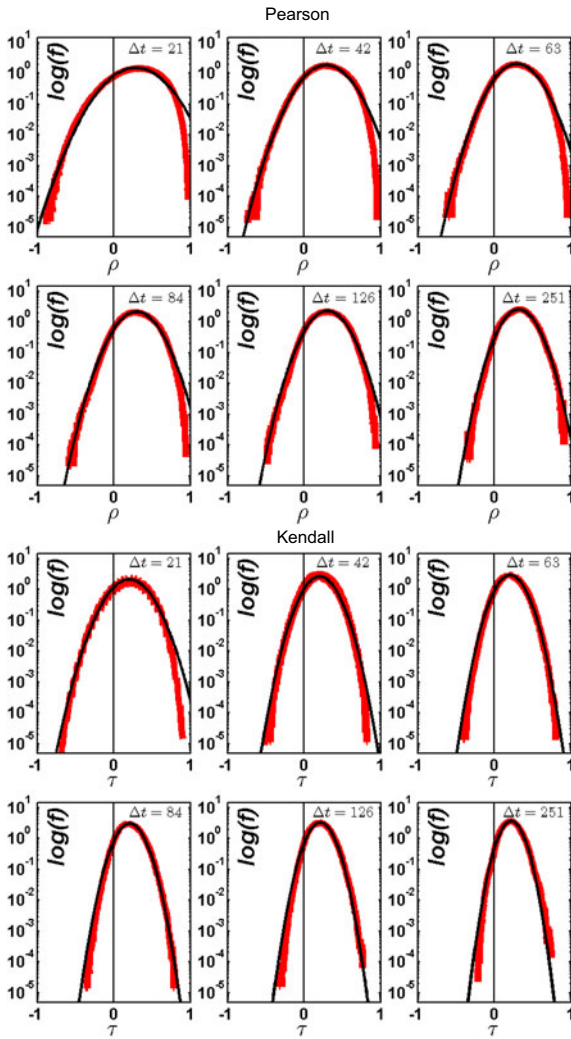


Fig. I.1. (Color online) Empirical logarithmic distributions of Pearson ρ and Kendall τ correlations computed for different values of the length of the running window Δt . For each case, $T - \Delta t + 1$ dynamic correlation matrices have been considered. The smooth curves represent the Gaussian distribution and the thick curves the empirical distributions. Pearson distributions associated to small running windows are greatly irregular with respect to the Gaussian distribution.

been produced. The results are illustrated in Figure I.1 where the logarithm of densities have been reported on vertical axes and coefficients on horizontal axes. The thick curves represent the empirical distributions while the thin curves represent the Gaussian distributions with empirical mean-variance. For both methods, as Δt increases, the distributions get increasingly closer to a Gaussian, the sample means become larger and the standard deviations smaller. Pearson correlation coefficients appear far more irregular, though.

The distribution of Pearson weighted correlation coefficients has been calculated with $\Delta t = 251$ days and different values of θ : on top of Figure I.2 it is shown that, as θ decreases, the distribution gets increasingly irregular and the variance increases. The analogous plots for

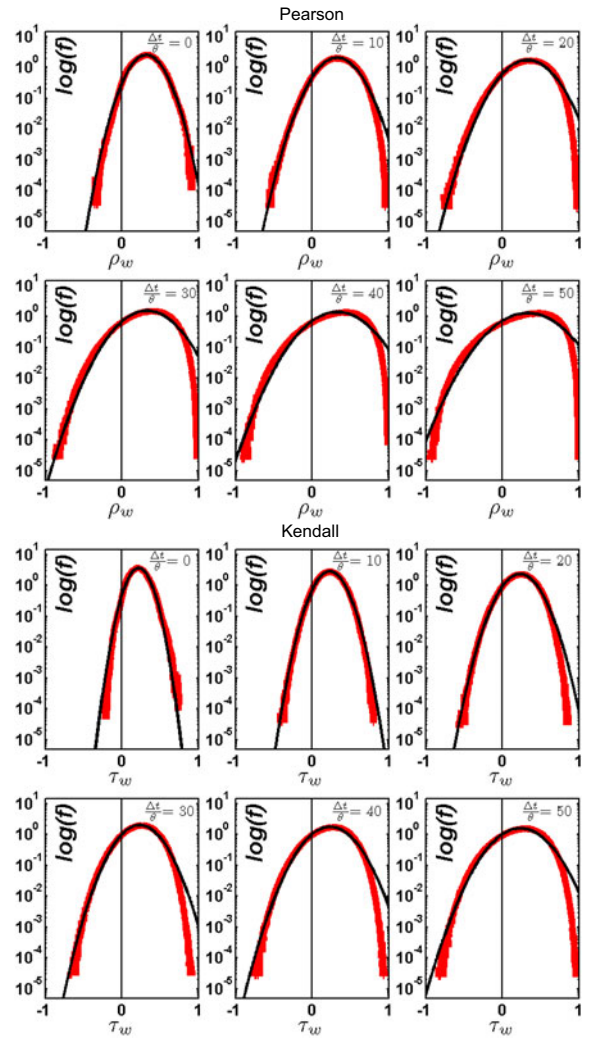


Fig. I.2. (Color online) At the top, empirical logarithmic distribution of Pearson weighted correlations computed for a running window of $\Delta t = 251$ days and for different values of θ . For each case, $T - \Delta t + 1 = 498$ dynamic correlation matrices have been considered. At the bottom, the analogous for Kendall weighted coefficients is shown.

Kendall weighted coefficients are reported at the bottom of the same Figure.

Appendix J: Proof of equations (15) and (16)

Since $w_{uv} = w_{vu}$ and $w_{uu} = 0 \forall u \in \{1, 2, \dots, \Delta t\}$, note that the constraint can be equivalently written as

$$\sum_{u=1}^{\Delta t} \sum_{v=1}^{\Delta t} w_0 e^{\alpha(u+v-2\Delta t)} - \sum_{u=1}^{\Delta t} w_0 e^{2\alpha(u-\Delta t)} = 2. \quad (\text{J.1})$$

Equation (J.1) can be written as

$$w_0 = \frac{2}{\sum_{u=1}^{\Delta t} \sum_{v=1}^{\Delta t} e^{\alpha(u+v-2\Delta t)} - \sum_{u=1}^{\Delta t} e^{2\alpha(u-\Delta t)}}. \quad (\text{J.2})$$

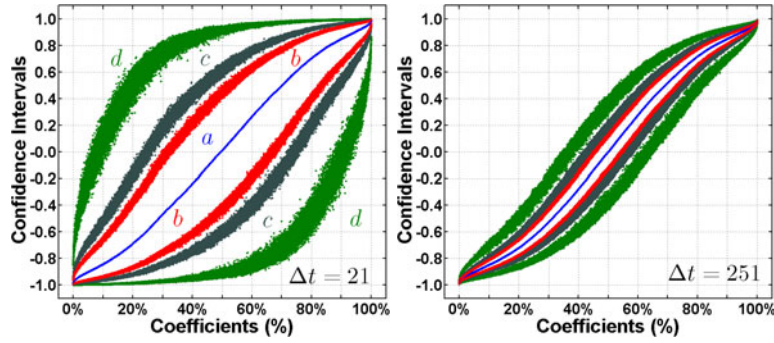


Fig. K.1. (Color online) 95% confidence intervals for all coefficients of Pearson Weighted Correlation Matrices (based on 1000 artificial series whose i th row is distributed as $\mathcal{N}(\mathbf{0}, \Sigma)$). The case with $\Delta t = 21$ is on the left; the case with $\Delta t = 251$ is on the right. The smooth line, a , represents the theoretical (known) correlations; the curves denoted as b represent the 95% confidence interval with $\theta \rightarrow \infty$; c is obtained for $\theta = \frac{\Delta t}{5}$ and d for $\theta = \frac{\Delta t}{10}$. The order of the curves is the same in the two plots.

$$\begin{aligned}
 \lim_{\alpha \rightarrow 0^+} w_0(\alpha) &= \lim_{u \rightarrow 1^+} \frac{D_u [(u-1)^2 (u+1)]}{D_u [u^2 (1-u^{-\Delta t}) (1-u^{1-\Delta t})]} = \lim_{u \rightarrow 1^+} \frac{3u^2 - 2u - 1}{2u + (\Delta t - 3)u^{2-\Delta t} + (\Delta t - 2)u^{1-\Delta t} + (3 - 2\Delta t)u^{2(1-\Delta t)}} \\
 &= \lim_{u \rightarrow 1^+} \frac{D_u [3u^2 - 2u - 1]}{D_u [2u + (\Delta t - 3)u^{2-\Delta t} + (\Delta t - 2)u^{1-\Delta t} + (3 - 2\Delta t)u^{2(1-\Delta t)}]} \\
 &= \frac{6 - 2}{2 + (\Delta t - 3)(2 - \Delta t) + (\Delta t - 2)(1 - \Delta t) + (3 - 2\Delta t)(2 - 2\Delta t)} = \frac{2}{\Delta t(\Delta t - 1)} \tag{J.3}
 \end{aligned}$$

The term at the denominator, H , can be expanded and simplified:

$$\begin{aligned}
 H &= e^{-2\alpha\Delta t} \sum_{u=1}^{\Delta t} (e^\alpha)^u \sum_{v=1}^{\Delta t} (e^\alpha)^v - e^{-2\alpha\Delta t} \sum_{u=1}^{\Delta t} (e^{2\alpha})^u \\
 &= e^{-2\alpha\Delta t} \left\{ \left[\frac{1 - (e^\alpha)^{\Delta t+1}}{1 - (e^\alpha)} - 1 \right] \left[\frac{1 - (e^\alpha)^{\Delta t+1}}{1 - (e^\alpha)} - 1 \right] \right. \\
 &\quad \left. - \left[\frac{1 - (e^{2\alpha})^{\Delta t+1}}{1 - (e^{2\alpha})} - 1 \right] \right\} \\
 &= e^{-\alpha\Delta t} \left[\frac{e^\alpha - (e^\alpha)^{\Delta t+1}}{1 - (e^\alpha)} \right] e^{-\alpha\Delta t} \left[\frac{e^\alpha - (e^\alpha)^{\Delta t+1}}{1 - (e^\alpha)} \right] \\
 &\quad - e^{-2\alpha\Delta t} \left[\frac{(e^{2\alpha}) - (e^{2\alpha})^{\Delta t+1}}{1 - (e^{2\alpha})} \right] \\
 &= e^{2\alpha} \left(\frac{1 - e^{-\alpha\Delta t}}{e^\alpha - 1} \right)^2 - e^{2\alpha} \left(\frac{1 - e^{-2\alpha\Delta t}}{e^{2\alpha} - 1} \right) \\
 &= e^{2\alpha} \left[\frac{(1 - e^{-\alpha\Delta t})^2}{(e^\alpha - 1)^2} - \frac{(1 + e^{-\alpha\Delta t})(1 - e^{-\alpha\Delta t})}{(e^\alpha + 1)(e^\alpha - 1)} \right] \\
 &= e^{2\alpha} \left(\frac{1 - e^{-\alpha\Delta t}}{e^\alpha - 1} \right) \\
 &\quad \times \left[\frac{(1 - e^{-\alpha\Delta t})(e^\alpha + 1) - (1 + e^{-\alpha\Delta t})(e^\alpha - 1)}{(e^\alpha + 1)(e^\alpha - 1)} \right] \\
 &= e^{2\alpha} \left(\frac{1 - e^{-\alpha\Delta t}}{e^\alpha - 1} \right) \left\{ \frac{2[1 - e^{-\alpha(\Delta t-1)}]}{(e^\alpha + 1)(e^\alpha - 1)} \right\} \\
 &= \frac{2e^{2\alpha}(1 - e^{-\alpha\Delta t})[1 - e^{-\alpha(\Delta t-1)}]}{(e^\alpha - 1)^2(e^\alpha + 1)}.
 \end{aligned}$$

Substituting the expression of H back at the denominator of equation (J.2), and simplifying, equation (15) is obtained.

Note also that, when α tends to zero, the limit of weights tends to

see equation (J.3) above

where $D_u[\dots]$ denotes the derivative with respect to u ; then the l'Hôpital rule has been applied. By substituting equation (J.3) in equation (13) uniform weights are obtained, as in equation (16).

Appendix K: Bootstrapping procedure over a multivariate gaussian process

We have generated random variables distributed as a multivariate Gaussian, $\mathcal{N}(\mathbf{0}, \Sigma)$, with zero mean, unitary standard deviation and a given correlation matrix. Specifically, for $\Delta t = \{21, 251\}$, we have generated 1000 Δt -by-300 matrices and for each we have calculated the weighted correlations with $\frac{\Delta t}{\theta} = \{0, 5, 10\}$. Then for each coefficient we have calculated the percentiles $\{2.5, 97.5\}$. The results are illustrated in Figure K.1 where the case with $\Delta t = 21$ is on the left and the case with $\Delta t = 251$ is on the right: the smooth line (denoted as a) represents the theoretical coefficients which, in the example, are known; the narrower confidence interval (denoted as b) is obtained with $\theta \rightarrow \infty$; c is obtained for $\theta = \frac{\Delta t}{5}$ and d is obtained for $\theta = \frac{\Delta t}{10}$. The order of the curves is the same in the two plots. It is evident from the two plots that lowering θ causes an increase in the uncertainty associated to the estimate of individual correlation coefficients.

References

1. A. Meucci, *Risk and Asset Allocation* (Springer-Verlag Berlin Heidelberg, 2005)
2. R. Litterman, K. Winkelmann, Estimating covariance matrices, in *Goldman Sachs, Risk Management Series* (1998)
3. G.S. Shieh, *Stat. Probab. Lett.* **39**, 17 (1998)
4. I.A. Salama, D. Quade, A nonparametric comparison of the structure of two multiple-regression prediction situations, in *Inst. of Statist. Mimeo Series* (Univ. North Carolina, Chapel Hill, 1981), Vol. 1325
5. I.A. Salama, D. Quade, *Commun. Stat. Theory Methods* **11**, 1185 (1982)
6. D. Quade, I.A. Salama, A Survey of Weighted Rank Correlation, in *Order Statistics and Nonparametrics: Theory and Applications*, edited by P.K. Sen, I.A. Salama (Elsevier, Amsterdam, 1992), pp. 213–224
7. G.S. Shieh, Correlation, Weighted, in *Encyclopedia of Statistical Sciences* (John Wiley & Sons, Inc., 2006)
8. J.P. Morgan/Reuters, *RiskMetricsTM – Technical Document*, 4th edn. (New York, 1996)
9. S. Pafka, M. Potters, I. Kondor, Exponential Weighting and Random-Matrix-Theory-Based Filtering of Financial Covariance Matrices for Portfolio Optimization, [arXiv:cond-mat/0402573v1\[cond-mat.stat-mech\]](https://arxiv.org/abs/cond-mat/0402573v1)
10. T. Hayashi, Y. Nakahiro, *Bernoulli* **11**, 359 (2005)
11. P. Malliavin, M.E. Mancino, *Finance and Stochastics* **6**, 49 (2002)
12. S. Drożdż, J. Kwapień, F. Grümmer, F. Ruf, J. Speth, *Physica A* **299**, 144 (2001)
13. S. Drożdż, F. Grümmer, F. Ruf, J. Speth, *Physica A* **294**, 226 (2001)
14. J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, A. Kanto, *Phys. Rev. E* **68**, 056110 (2003)
15. L. Bauwens, S. Laurent, J.V.K. Rombouts, *J. Appl. Econ.* **21**, 79 (2006)
16. R. Engle, *J. Bus. Econ. Stat.* **20**, 339 (2002)
17. J. Svensson, *Physica A* **385**, 621 (2007)
18. R.E. Bellman, *Dynamic Programming* (Princeton University Press, New Jersey, 1957)
19. R.E. Bellman, *Adaptive Control Processes: A Guided Tour* (Princeton University Press, 1961)
20. K. Beyer, J. Goldstein, R. Ramakrishnan, U. Shaft, When is ‘Nearest Neighbor’ Meaningful?, *Proceedings of 7th International Conference on Database Theory* (1999), Vol. 1540, pp. 217–235
21. R.B. Marimont, M.B. Shapiro, *Journal of the Institute of Mathematics and its Applications* **24**, 59 (1979)
22. E. Chavez et al., *ACM Comput. Surv.* **33**, 273 (2001)
23. S.M. Stigler, *Stat. Sci.* **4**, 73 (1989)
24. J. Lee Rodgers, W. Alan Nicewander, *Am. Stat.* **42**, 59 (1988)
25. F. Galton, *J. Anthropol. Inst.* **15**, 246 (1885)
26. K. Pearson, *Roy. Soc. Proc.* **58**, 241 (1895)
27. K. Pearson, *Biometrika* **13**, 25 (1920)
28. T.C. Cook, D.T. Campbell, *Quasi-Experimentation* (Houghton Mifflin, Boston, 1979)
29. H.M. Walker, *Studies in the History of Statistical Method* (Williams & Wilkins, Baltimore, 1929)
30. H.L. Seal, *Biometrika* **54**, 1 (1967)
31. C. Darwin, *The Variation of Animals and Plants under Domestication* (John Murray, London, 1868)
32. A. Kraskov, H. Stögbauer, P. Grassberger, *Phys. Rev. E* **69**, 066138 (2004)
33. A. Kraskov, H. Stögbauer, R.G. Andrzejak, P. Grassberger, *Europhys. Lett.* **70**, 278 (2005)
34. M.G. Kendall, *Biometrika* **30**, 81 (1938)
35. M.G. Kendall, *Rank Correlation Methods* (Charles Griffin & Co. Ltd., London, 1948)
36. W.H. Kruskal, *J. Am. Stat. Assoc.* **53**, 814 (1958)
37. W.R. Knight, *J. Am. Stat. Assoc.* **61**, 436 (1966)
38. D. Christensen, *Computational Statistics* **20**, 51 (2005)
39. V.D. Skintzi, A.-P.N. Refenes, *J. Futures Mark.* **25**, 171 (2005)
40. Y.H. Cho, R. Engle, Time-Varying Betas and Asymmetric Effects of News: Empirical Analysis of Blue Chip Stocks, Working paper No. 7730, National Bureau of Economic Research, Cambridge, 1999
41. T. Andersen, T. Bollerslev, F. Diebold, H. Ebens, *J. Financ. Econ.* **61**, 43 (2001)
42. T. Andersen, T. Bollerslev, F. Diebold, P. Labys, *J. Am. Stat. Assoc.* **96**, 42 (2001)
43. E. Sheedy, *J. Int. Financ. Mark. Inst. Money* **8**, 59 (1998)
44. B. Solnik, C. Bourcell, Y. Le Fur, *Financ. Anal. J.* **5**, 17 (1996)
45. L. Ramchand, R. Susmel, *J. Empir. Finance* **5**, 397 (1998)
46. A. Ang, J. Chen, *J. Financ. Econ.* **63**, 443 (2002)
47. C.B. Erb, C.R. Harvey, E. Viskanta, *Financ. Anal. J.* **6**, 32 (1994)
48. F. Longin, B. Solnik, *J. Int. Money Financ.* **14**, 3 (1995)
49. K.J. Forbes, R. Rigobon, *J. Financ.* **57**, 2223 (2002)
50. G.M. von Fustenber, B.N. Jeon, N.G. Mankiw, R.J. Shiller, *Papers on Economic Activity* **1**, 125 (1989)
51. P.D. Koch, T.W. Koch, *J. Int. Money Financ.* **10**, 231 (1991)
52. C. Borghesi, M. Marsili, S. Miccichè, *Phys. Rev. E* **76**, 026104 (2007)
53. M. Lundin, M.M. Dacorogna, U.A. Muller, Correlation of high frequency financial time series, in *The financial markets tick-by-tick*, edited by P. Lequeux (Wiley, London, 1999)
54. M.G. Kendall, *Biometrika* **36**, 177 (1949)
55. R. Newson, *The Stata Journal* **1**, 1 (2001)
56. C. Stein, Inadmissibility of the usual estimator for the mean of a multivariate distribution, in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, edited by J. Neyman (Univ. California Press, Berkeley, 1956), pp. 197–206
57. O. Ledoit, M. Wolf, *J. Portfolio Manag.* **30**, 110 (2004)
58. J. Schäfer, K. Strimmer, *Statistical Applications in Genetics and Molecular Biology* **4** (2005)
59. D.J. Disatnik, S. Benninga, *J. Portfolio Manag.* **33**, 55 (2007)
60. J.-P. Bouchaud, M. Potters, *Financial Applications of Random Matrix Theory: a short review* (2009), [arXiv:0910.1205v1](https://arxiv.org/abs/0910.1205v1)
61. C.C.Y. Kwan, *Spreadsheets in Education (eJSiE)* **4** (2011)
62. E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, D. Sorensen, *LAPACK User’s Guide*, 3rd edn. (SIAM, Philadelphia, 1999)
63. F. Pozzi, T. Aste, G. Rotundo, T. Di Matteo, *Proc. SPIE* **6802**, E68021 (2008)

64. F. Pozzi, T. Di Matteo, T. Aste, [Advances in Complex Systems](#) **11**, 927 (2008)
65. T. Di Matteo, F. Pozzi, T. Aste, [Eur. Phys. J. B](#) **73**, 3 (2009)
66. F. Pozzi, T. Aste, W. Shaw, T. Di Matteo, The use of topological quantities to detect hierarchical properties in financial markets: the Financial sector in NYSE, [Proceedings of 10th WSEAS international conference on Mathematics and computers in business and economics, Recent Advances in Computer Engineering](#) (2009), pp. 301–304
67. T. Aste, W. Shaw, T. Di Matteo, [New J. Phys.](#) **12**, 085009 (2010)
68. J. Hadamard, *Leçons sur la Propagation des Ondes et les Équations de l'Hydrodynamique* (Hermann, Paris, 1903)
69. J. Schur, [J. Reine Angew. Math.](#) **140**, 1 (1911)
70. P.R. Halmos, *Finite-Dimensional Vector Spaces* (Princeton University Press, 1948)
71. G.P.H. Styan, [Linear Algebra Appl.](#) **6**, 217 (1973)
72. M. Marcus, Basic Theorems in Matrix Theory, in *National Bureau of Standards, Applied Mathematics Series* (US Government Printing Office, Washington, D.C., 1960), Vol. **57**
73. L. Mirsky, *An Introduction to Linear Algebra* (Oxford University Press, 1955)
74. R.A. Horn, C.R. Johnson, *Topics in Matrix Analysis* (Cambridge University Press, 1994), Chap. 5, pp. 299–381
75. S. Barnett, *Matrices, Methods and Applications* (Oxford University Press, New York, 1990)
76. R.A. Beezer, *A First Course in Linear Algebra*, 1.08 edn. (2007)
77. E. Million, *The Hadamard Product* (Creative Commons, 2007)
78. R.N. Mantegna, [Eur. Phys. J. B](#) **11**, 193 (1999)
79. S.I. Ramanujan, *Collected Papers* (Cambridge University Press, Cambridge, 1927)
80. D.E. Knuth, The art of computer programming ([Addison Wesley Longman, 1997](#)), 3rd edn., Vol. 1
81. R.L. Graham, D.E. Knuth, O. Patashnik, *Concrete Mathematics*, 2nd edn. ([Addison-Wesley, 1994](#))