Measurement of current market correlations based on ensemble statistics

Jack Sarkissian, Algostox Trading

Joel Sebold, Refinitiv[1]/Thomson Reuters

Abstract

We employ averaging over statistical ensemble of assets to derive an index characterizing the level of correlations in a financial market – the $eCORR$ index. This index does not require lengthy historical data and reacts immediately to any changes in correlations. Study of statistical properties of $eCORR$ for US equity markets reveals how volatility is distributed between the common part and the part specific to individual equities. It also allows to demonstrate and quantify the correlation-drawdown hysteresis effect. The $eCORR$ index promises to be useful for early detection of market correlations, managing risk concentrations and maintaining portfolio diversification.

1. Introduction

Stochastic financial models treat price dynamics as a random process. Depending on the modeling goals, that may work well for an individual asset. However, as soon as assets are combined to describe a market,

[1] Refinitiv is formerly the Financial and Risk business of Thomson Reuters

one cannot overlook correlations between them. Correlation plays an important role in overall price evolution and is one of the main components responsible for financial dynamics.

Market correlations can be caused by various factors. For example, gold mining companies will be correlated because their revenue depends on price of gold. Corporate bond market may be correlated because it is affected by monetary policy, inflation and other factors. Equity markets may be correlated because traders use market indices as a gauge for making trading decisions.

Traders and asset managers know from experience that correlations play a subtle but important role in financial management. Higher correlations reduce diversification and expose portfolios to concentration risk. Correlations are known to spike in times of market corrections or financial crises and go hand in hand with them. Overlooking the role of correlations in portfolio management may lead to unexpected losses, particularly large losses, or even to simple violations of investment declarations. Market correlations can also affect pricing, particularly pricing of derivative instruments. Overall, there are many reasons why staying informed of the current market correlation levels is important.

Traditional way of calculating pairwise correlations γ_{ij} for two assets is through time averaging product of their returns:

$$\gamma_{ij} = \frac{< r_i r_j >_t}{\sigma_i \sigma_j}$$

Here r_i and r_j and the returns corresponding to the i-th and j-th assets, σ_i and σ_j are their respective standard deviations, and $\langle \ldots \rangle_t$ stands for time averaging operation. Once the pairwise correlations have been measured, the overall market correlation can be calculated as their average

$$Correlation = \langle \gamma_{i \neq j} \rangle_{ij} \quad (1)$$

where now $< \cdots >_{ij}$ stands for across all assets. Even though this method follows the direct definition of correlation, it encounters serious difficulties that make it hard to use in practice. Statistical validity requires the observation period to be large. However, large period diminishes the quantity's sensitivity to new data. As a result, correlations obtained with Eq. (1) are either smooth but delayed (for longer observation periods) or have faster response but lack statistical validity (for shorter observation period).

Given these difficulties, use of correlations in finance reduces to technical exercise that can only be useful under normal market conditions when correlations just fluctuate around their average values. However, as soon as the market picture changes, the reliability of models involving correlations drops significantly, and may even become counterproductive, leading to losses.

The new approach, introduced in [1], has proven to be effective in coping with these difficulties. According to this approach, when calculating statistical properties of the markets, the averaging over time is replaced by averaging over the ensemble of assets, comprising the market. To the extent that the market can be considered to be an ergodic system, the two averages should converge. It has been pointed out that volatility indices based on ensemble averaging approach – the $eVOL$ and $eVAR$ – provide more relevant measures of current market volatility than the traditional historical and implied volatility measures [1, 2].

In this paper we use the ensemble averaging method to calculate the level of current correlations in the markets. Prior to doing it, let us estimate their typical values and how they come across to affect the markets.

One way that correlations present themselves in equity markets is that volatility of equity indices is much larger than it should be if the markets were uncorrelated. For example, an average DJI component stock has volatility of daily returns around 1.7%. Given that there are 30 components in DJI, the index volatility would have been 0.3% if its components were uncorrelated. That number would have been 0.07% for S&P 500, which has 500 components. I the meantime, we know that average DJI and S&P 500 volatilities are approximately the same and are around 1.1%. Comparing index volatilities with average volatilities of their components, we can deduce that the implied market correlation is about 0.45.

The ergodic property is essential if you want to replace time averaging with ensemble averaging [3-5]. In that sense, this paper addresses the opposite issue - market non-ergodicity - since it's the correlations that destroy ergodicity. Correlations prevent the statistical system from exploring its phase space in its entirety. They serve as an order parameter that measures the degree of non-ergodicity of the system. This paper therefore will address not only the question how to measure current market correlations, but also the degree to which the markets can be considered ergodic.

2. Ensemble correlation: the $eCORR$ index

Following the described considerations, let us now try to calculate correlations across the statistical ensemble of assets. We will rely on assumption that the returns of all assets r_i, observed over the time step represent all possible states, in which the market could evolve over that time step. If we use these returns to compose a matrix

$$M_{ij} = r_i r_j$$

then the variance of returns would be represented by the average of its diagonal elements:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} r_i^2$$

and covariance would be represented by the average of the off-diagonal elements:

$$Cov = \frac{1}{N(N-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{N,N} r_i r_j$$

Taking the ratio of the two, let us define a new index:

$$eCORR = \frac{\langle r_i r_j \rangle_{i,j}}{\langle r_i^2 \rangle_i} = \frac{Cov}{\sigma^2}$$

If our paradigm is correct, the $eCORR$ index must relate to the level of market correlation. Indeed, if all returns r_i were independent and random, the value of $eCORR$ would only be limited by the statistical error. For N components that error would be $1/N$, about 3% for a 30-component index. If, however, all r_i were equal, the value of $eCORR$ would be 1. Additionally, it does not matter which sign of r_i is more prevalent, since $eCORR$ is made of their product $r_i r_j$. We can therefore say that $eCORR$ varies from 0 to 1 and is related to market correlations. The important feature of $eCORR$ is that it measures market correlations observed over a single time period without averaging over many time periods.

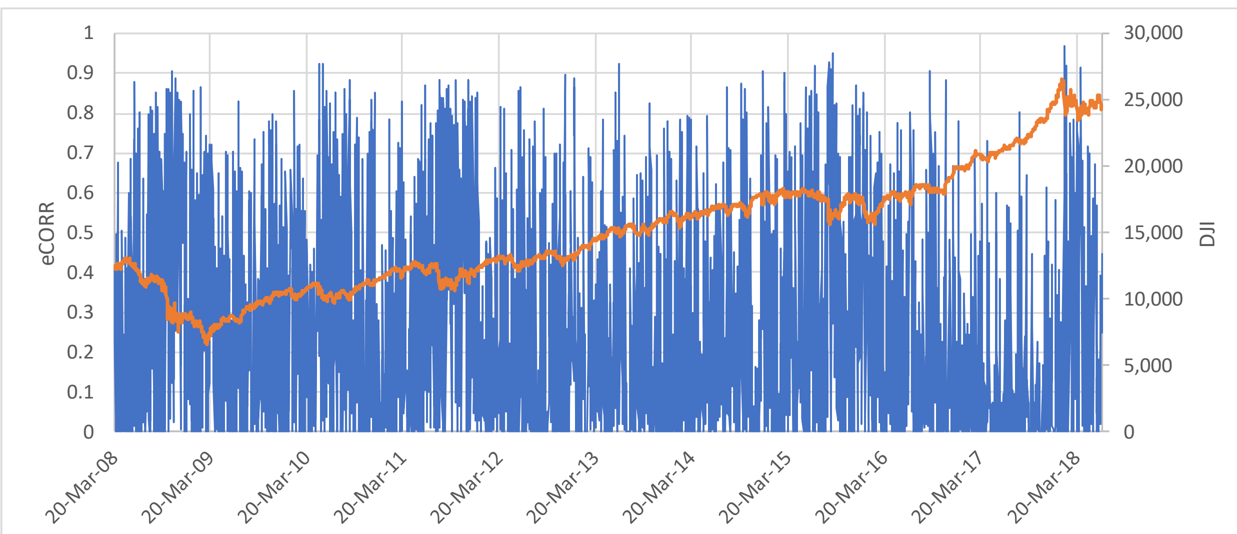

Fig. 1. Dynamics of 1-day DJI $eCORR$ index (computed from 1-day returns of DJI components) from March 20, 2008 to June 25, 2018.

Following its definition, an $eCORR$ index must specify the reference time interval, on which it is based. For example, an index based on 1-day returns must reference 1 day. It must also reference the composition that went into its calculation. For example, an $eCORR$ index based on the components of S&P 500 index must reference that index. As an example, the dynamics of 1-day DJI $eCORR$ index is shown in Fig. 1 over the time interval from March 20, 2008 to June 25, 2018.

3. Correlated markets

To relate the $eCORR$ index to the markets, we need to use a model of such markets. One way to describe correlated markets is to split individual asset returns into common and specific portions:

$$r_i = \frac{ds_i}{s_i} = \sigma\sqrt{\rho}\, dZ + \sigma\sqrt{1-\rho}\, dz_i \tag{2}$$

Here dZ and dz_i are normally distributed random variables, of which the dZ corresponds to common factor, same for all components, dz_i corresponds to specific factor, and we assumed that all assets have the same volatility σ. Then, the time average pairwise correlations can be computed easily:

$$\gamma_{ij} = \frac{< r_i r_j >_t}{\sigma^2} = \rho + (1-\rho)\delta_{ij} \tag{3}$$

In other words, $\gamma_{ii} = 1$ and $\gamma_{i\neq j} = \rho$, showing that in this model any two assets with have correlation coefficient ρ.

Let us now compose an N-component market index:

$$I = \sum_{i=1}^{N} w_i\, s_i$$

Its returns are then equal

$$R = \sum_{i=1}^{N} w_i\, r_i$$

Each index component is correlated to the index with the same correlation coefficient ρ:

$$\frac{< r_i R >}{\sigma^2} = \rho + (1-\rho) w_i \approx \rho$$

since $w_i \sim \frac{1}{N}$. Similarly, standard deviation of index's returns is $\sqrt{\rho}$ times smaller that of individual assets:

$$\frac{< R^2 >}{\sigma^2} = \rho + (1-\rho) \sum_{i=1}^{N} w_i^2 \approx \rho$$

Note, that this result does not depend on the number of index components N – just as we have with DJI and S&P 500.

Going back to Eq. (1) and combining it with Eq. (3), we have:

$$Correlation = \langle \gamma_{i \neq j} \rangle_{ij} = \rho$$

In this formula, we determined correlations by first performing time averaging, and then ensemble averaging. Changing the order of averaging, we still have

$$Correlation = \frac{\langle \langle r_i r_j \rangle_{i,j} \rangle_t}{\langle \langle r_i^2 \rangle_i \rangle_t} = \rho \qquad (4)$$

This demonstrates that if the number of assets in the market is sufficient to guarantee statistical validity, correlation measurements based on ensemble averaging and time averaging must converge. The $eCORR$ index comes from factoring out the elements involved in a single time step:

$$eCORR = \frac{\langle r_i r_j \rangle_{i,j}}{\langle r_i^2 \rangle_i}$$

To practically compute the $eCORR$ index we will rewrite Eq. (2) for a single time step:

$$r_i = \alpha + \sigma_e \, dz_i \qquad (5)$$

where $\alpha = \sigma\sqrt{\rho}\, dZ$ is the common factor realized over the time step under consideration and $\sigma_e = \sqrt{1-\rho}\, \sigma$ corresponds to the $eVOL$ index in [1]. We deliberately relabeled the variables to avoid distraction by their fluctuations in time. We are currently considering only one step in time, and all information is related to that time step. Common factor α also approximately equal to the index return:

$$\alpha \approx R$$

Computing the $eCORR$ index with this model gives

$$eCORR = \frac{\alpha^2}{\alpha^2 + \sigma_e^2}$$

or approximately

$$eCORR \approx \frac{R^2}{R^2 + \sigma_e^2}$$

It's easy to see now that the $eCORR$ index measures the fraction of return variance attributed to the common factor out of its total value.

4. Properties of the $\boldsymbol{eCORR}$ index

Now that we have related the $eCORR$ index to market quantities, let's measure some market characteristics. We will use the daily Dow Jones Industrial Average Index (DJI) data for the timeframe from March 20, 2008 to June 25, 2018.

Response time of $eCORR$ vs Historical correlation

The best way to study properties of correlation measures is by building a market model with Eq. (2) and applying the measures on it. Since we set the correlations in the model, we can judge the relevance degrees of various measures.

To gauge time response we generate 100 steps of 30 returns r_i in Eq. (2) with real correlation ρ jumping from 0.1 to 0.9 over only 3-4 steps. That real correlation is represented with a red line in Fig. 2. We then measure the $eCORR$ and the 10-step historical correlation, also shown in the figure. Even though $eCORR$ appears to be fluctuating with a large amplitude, it's apparent that it hits the 0.9 level immediately, while the historical correlation lags over about 10 steps.

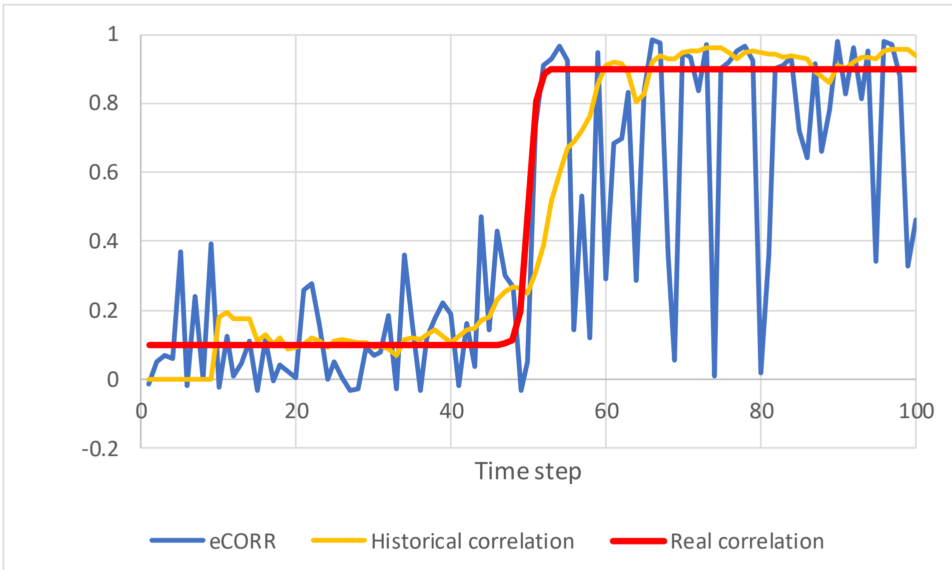

Fig. 2. Response time of $eCORR$ vs. historical correlation. Historical correlation takes time to adjust to new levels of correlation, while $eCORR$ responds immediately.

$eCORR$ and market index return

Although model Eq. (5) is reasonable, it is incomplete. It assumes constant volatility, while we are actually aiming to describe the dynamics of time series. Market data points that when index return is large, volatility increases [1]. Let us include this dependence in the model:

$$r_i = \alpha + \sigma_e(\alpha)\, dz_i$$

Although the underlying processes are stochastic in nature, the dependence of variance (σ^2) on return can be taken as smooth. We can include it by taking a Taylor series expansion

$$\sigma_e^2(\alpha) \approx v_0 + v_1\alpha + v_2\alpha^2$$

Then, we can approximate the $eCORR$ with the following function:

$$eCORR \approx f(R) = \frac{R^2}{v_0 + v_1 R + (1 + v_2)R^2} \tag{6}$$

Fig. 3 shows a scatterplot of 1-day DJI $eCORR$ values against the 1-day DJI returns. One can notice its asymmetric nature, showing that $eCORR$ is more responsive to negative returns than to positive returns. Its asymptotic value for large market moves is below 1, which is due to $v_2 > 0$.

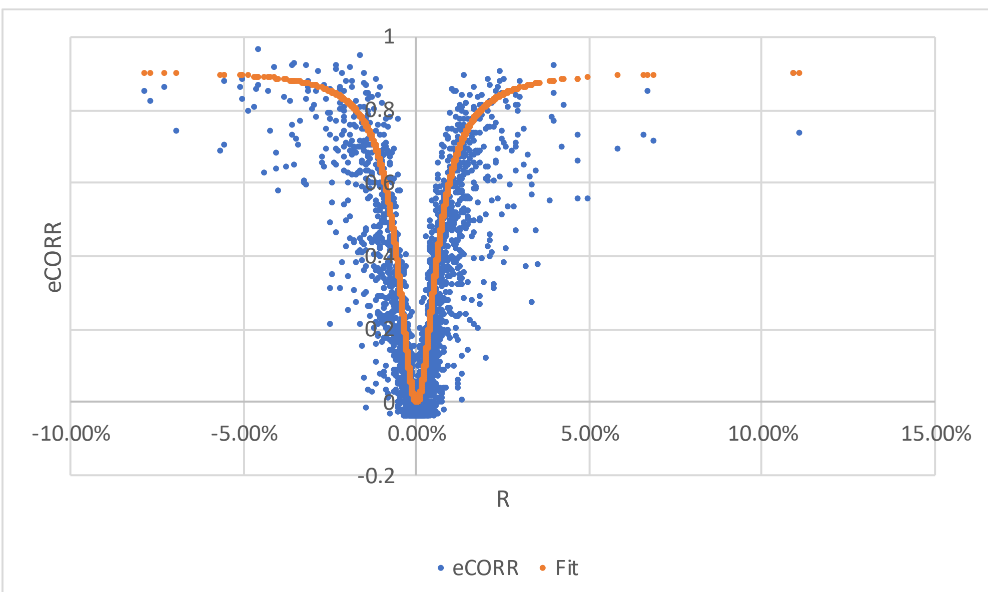

Fig. 3. 1-day DJI $eCORR$ values against the 1-day DJI returns

for the time interval from March 20, 2008 to June 25, 2018.

Fitting was performed with error elements weighted by R_i^2 to make sure that the scarce tail data made adequate contributions:

$$\epsilon = \sum_i [f(R_i) - eCORR_i]^2 \, R_i^2$$

Fitting parameters of Eq. (6) were found to be equal $v_0 = 0.0039\%$, $v_1 = 0.00027$, and $v_2 = 1.26$. The corresponding chart of $\sigma_e(R)$ is shown in Fig. 4. Interestingly, the asymmetric part v_1 is insignificantly small, in accordance with [1].

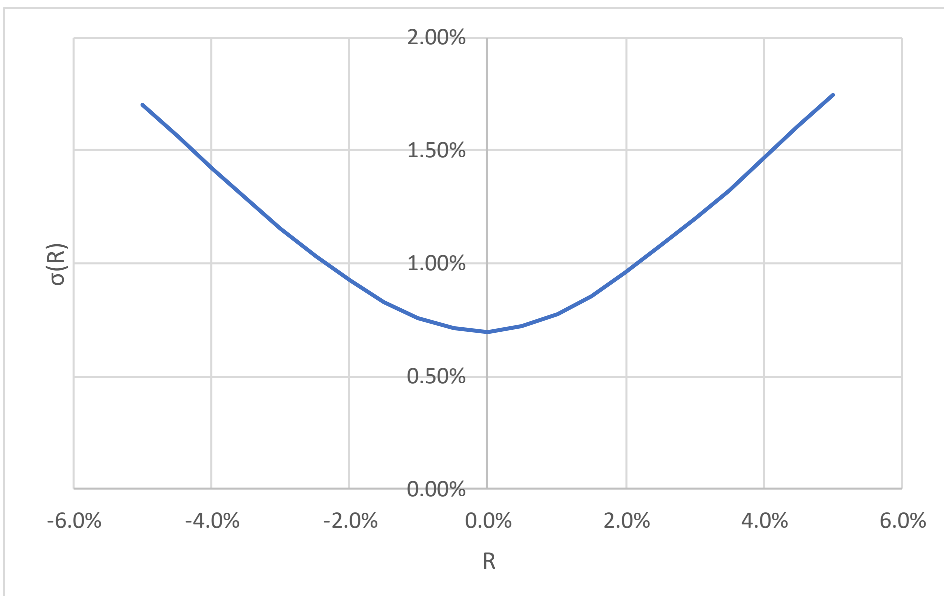

Fig. 4. Dependence of specific volatility on index return $\sigma_e(R) \approx \sqrt{v_0 + v_1 R + v_2 R^2}$

eCORR and market drawdown

Market correlations are directly related to drawdowns. While drawdown is a qualitative characteristic describing sudden large drops market valuation, let us choose a formal definition for the purposes of this paper. First of all, *Market Drawdown* as a quantity is

$$D(t) = -\min\left[\frac{I(t)}{I_0} - 1, \qquad 0\right],$$

where $I(t)$ is the current value of market index and I_0 is its value when the drawdown process began, so that $D(t) \geq 0$. *Drawdown Process* begins at time t_i when $D(t_i) \geq 5\%$ and $D(t_{i-1}) = 0$, to align the definition with large sudden market drops.

It is a known observation among traders, asset managers and risk managers that as drawdown accumulates, correlations take higher values. They relax quickly after the drawdown passes its maximum point. This effect can be demonstrated schematically with diagram in Fig. 5. The blue line on the diagram demonstrates a market entering the drawdown, and the orange line demonstrates a market exiting it. Let us see how $eCORR$ relates to this effect.

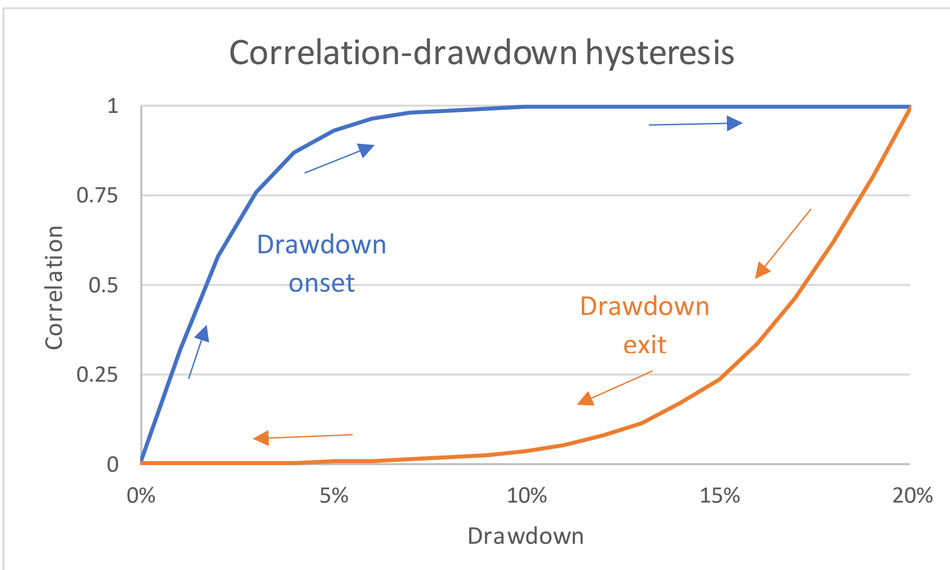

Fig. 5. Diagram demonstrating hysteresis between correlation and drawdown.

Due to stochastic nature of the process, it’s not easy to trace individual loops of $eCORR$ on its diagram. We can, however, measure the hysteresis effect indirectly by taking the sum

$$H(t) = \sum_{i=1}^{N:t=\tau_N} eCORR(\tau_i)[D(\tau_i) - D(\tau_{i-1})]$$

As the market enters a drawdown we should see this quantity increase above its previous levels. This increase should be followed by a decrease as the market comes out of drawdown. When the loop is completed, the resulting $H(t)$ should be larger than its value before the drawdown. Dynamics of $H(t)$ is shown for 1-day DJI $eCORR$ in Fig. 6, where we defined drawdown as a market event, in which the market loses at least 5% of its value in the first day.

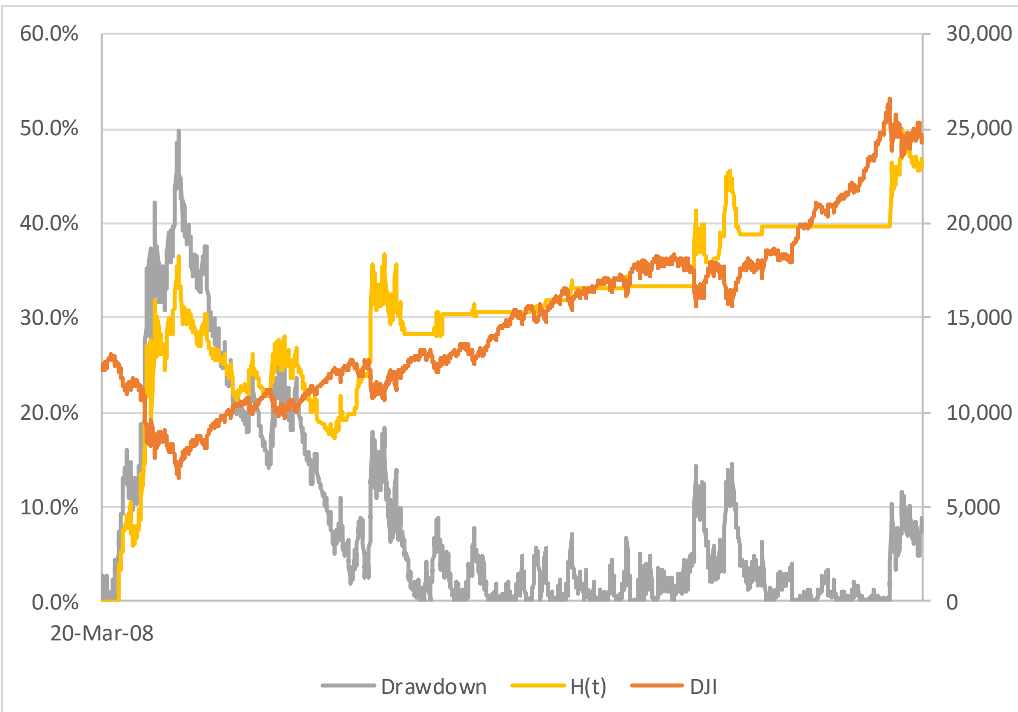

Fig. 6. Dynamics of $H(t)$ for 1-day DJI $eCORR$ evolution.

5. Discussion

The $eCORR$ index contains an amount of paradox in it. On one hand it is based on ensemble averaging, which can only replace the time average if the system is ergodic. On the other hand, the $eCORR$ itself is a measure of the degree of market ergodicity. When correlations are low the selection between time or ensemble averaging is unimportant. But when correlations are large the $eCORR$ becomes the only reliable measure of current market correlations.

The physical meaning of $eCORR$ is to show what fraction of the current market move is caused by the common fluctuation out of the total fluctuation. It carries important information about the market as a whole and it does so without a time delay. As a measure of the market as a whole, $eCORR$ can be applied for portfolio management or for risk management.

The $eCORR$ index can be used by portfolio managers to point to decreasing level of diversification. The mere fact that portfolios contain many assets from different market sectors does not always mean that portfolios are diversified. At times when common factor becomes large, portfolio managers might want to support their level of diversification by reducing exposure to that common factor, for example by short-selling the index or buying puts on it while paying a small cost of risk for it.

Given its fast response to drawdowns, systematically high levels of $eCORR$ tend to point at a forming drawdown, while relaxation of $eCORR$ from the high levels tends to point at the end of a drawdown. Risk managers might want to react to these signs by enforcing stricter risk limits.

The $eCORR$ index can be defined for various asset classes as long as there are enough assets in the class to make it an ensemble. Equity markets are a straightforward choice. The index can be applied to the entire market or different market sectors.

It may be argued that the definition of $eCORR$ must include capitalization weights. After all, the ensemble average based volatility index $eVOL$ does contain that. Similar definition for correlations would read:

$$eCORR_w = \frac{1}{\sigma^2} \sum_{\substack{i,j=1 \\ i \neq j}}^{N,N} w_i w_j \, r_i r_j$$

The largest contribution to either of these quantities is made by the assets with largest capitalization. Since the $eVOL$ index is made of nonnegative terms, keeping only the large-cap terms and omitting the small cap contributions would result in underestimating the $eVOL$ index. It would, however, not invalidate it or its behavior. Unlike $eVOL$, the $eCORR$ index is made of both positive and negative terms. The weighted index would then primarily express the correlation level of a limited number of assets with largest capitalization and contain a large statistical error. This is why we have to refrain from weighing the returns in $eCORR$ index.

So, how ergodic are the markets after all? Can we safely switch between time and ensemble averaging to get the same results faster? The answer for US equity markets is contained in historical values of $eCORR$. Historically these values ranged anywhere from 0 to 1 averaging 0.26. That means that on average only 1/4 of equity variance is caused by the common factors, leaving 3/4 for individual factors. That is not a very large number for most finance applications. However, we should be careful interpreting the results when $eCORR$ is close to 1.

Summarizing, we showed how application of physics methodologies can help extract useful information about financial markets. Like any methodology this one has to be applied carefully to mean only what it is designed to mean. It remains unknown why a simple and straightforward method such as ensemble averaging was overlooked in finance for so many decades.

6. References

[1] J. Sarkissian, "Fast measurement of market volatility using ensemble averaging" (July 20, 2016). Available at SSRN: http://ssrn.com/abstract=2812353

[2] P. Jackson, J. Sebold and J. Sarkissian, "Market Voice: Past Volatility, Future Volatility: What About Current Volatility?", (June 7, 2018), Market Voice, Thomson Reuters

[3] L.D. Landau, E.M. Lifshitz, “Statistical Physics, Third Edition, Part 1: Volume 5 (Course of Theoretical Physics, Volume 5) 3rd Edition”, Butterworth-Heinemann, (1980)

[4] Ya.B. Zeldovich, A.D. Myshkis, ”Elements of Mathematical Physics: Noninteracting Particles”, Nauka (1973)

[5] G.G. Malinetskiy, A.B. Potapov, “Modern Problems of Nonlinear Mechanics”, Editorial URSS (2000)

provided for illustrational purposes to aid comprehension of the material contained in this document. No assurances can be made that any information contained or implied herein including aims, expectations, activities, strategies, assumptions, and goals were or will be realized or that it will continue to be valid at all or in the same manner as is currently described herein.