Market-Based Probability of Stock Returns

Olkhov, Victor

Independent

25 November 2023

Market-Based Probability of Stock Returns

Victor Olkhov

Independent, Moscow, Russia

victor.olkhov@gmail.com

ORCID: 0000-0003-0944-5113

Dec.15, 2024

Abstract

This paper describes the dependence of market-based statistical moments of returns on statistical moments and correlations of the current and past trade values. We use Markowitz's definition of value weighted return of a portfolio as the definition of market-based average return of trades during the averaging period. Then we derive the dependence of market-based volatility and higher statistical moments of returns on statistical moments, volatilities, and correlations of the current and past trade values. We derive the approximations of the characteristic function and the probability of returns by a finite number q of market-based statistical moments. To forecast market-based average and volatility of returns at horizon T, one should predict the first two statistical moments and correlation of current and past trade values at the same horizon. We discuss the economic reasons that limit the number of predicted statistical moments of returns by the first two. That limits the accuracy of the forecasts of probability of returns by the accuracy of the Gaussian approximations. To improve the reliability of large macroeconomic and market models like BlackRock's Aladdin, JP Morgan, and the U.S. Fed., the developers should use market-based statistical moments of returns.

Keywords : market statistical moments, volatility, correlations

JEL: C0, E4, F3, G1, G12

This research received no support, specific grant or financial assistance from funding agencies in the public, commercial or nonprofit sectors. We welcome valuable offers of grants, support and positions.

1. Introduction

Studies of stock returns are endless (Ferreira and Santa-Clara, 2008; Diebold and Yilmaz, 2009; Kelly et al., 2022). The assessments of the factors that impact the expected return play a central role (Fisher and Lorie, 1964; Mandelbrot, Fisher, and Calvet, 1997; Campbell, 1985; Brown, 1989; Fama, 1990; Fama and French, 1992; Lettau and Ludvigson, 2003; Greenwood and Shleifer, 2013; Martin and Wagner, 2019). The irregular behavior of stock prices and returns makes probability theory a major tool for modeling returns. The probability distributions and correlation laws that can match the random return changes are studied by Kon (1984) and (Campbell, Grossman, and Wang, 1993; Davis, Fama, and French, 2000; Llorente et al., 2001; Dorn, Huberman, and Sengmueller, 2008; Lochstoer and Muir, 2022). The description of the expected return is complemented by research on the realized return and volatility (Schlarbaum, Lewellen, and Lease, 1978; Andersen and Bollerslev, 2006; McAleer and Medeiros, 2008). The probability distributions of the realized and expected returns are studied by Amaral et al. (2000), Knight and Satchell (2001), and Tsay (2005). That is only a tiny fraction of the returns' studies.

The irregular values and volumes of market trades cause random fluctuations of stock returns during almost any period. Since Bachelier (1900), who highlighted the probabilistic nature of the price change, it has become routine to consider the frequency of the values of prices and returns as the basis for their probabilistic description. The average return $\rho(t,\tau;1)$ (1.2), which is determined by random time series of returns $r(t_i,\tau)$ (1.1) during the interval Δ (1.1), takes the form (1.2):

$$r(t_i,\tau) = \frac{p(t_i)}{p(t_i-\tau)} \quad ; \quad t-\frac{\Delta}{2} < t_i < t+\frac{\Delta}{2} \quad ; \quad i=1,\dots N \tag{1.1}$$

$$\rho(t,\tau;1) \sim \frac{1}{N}\sum_{i=1}^{N} r(t_i,\tau) \tag{1.2}$$

In (1.1), we consider return $r(t_i,\tau)$ as a ratio of a price $p(t_i)$ of market trade at time t_i to price $p(t_i-\tau)$ of trade at time $t_i-\tau$ in the past with the time shift τ. We use "~" to highlight that (1.2) gives the approximation of the average return $\rho(t,\tau;1)$ by the finite number N of terms of the time series. The frequency-based assessments of average return $\rho(t,\tau;1)$ (1.2) serve as a basis for almost all probabilistic models (Shephard, 1991; Shiryaev, 1999; Shreve, 2004), and we call them the frequency-based probability.

Let us consider a finite sample of N random returns $r(t_i,\tau)$, $i=1,..N,$ during Δ (1.1), as a sample of returns $r_i=r(t_i,\tau)$ of N securities that compose a portfolio. One of the most famous papers in financial economics and portfolio theory, "Portfolio Selection" by H. Markowitz

(1952), defines the portfolio return R (1.3) as the sum of returns r_i of the securities "weighted by the relative amount x_i invested in security i":

$$R = \sum_{i=1}^{N} r_i \cdot x_i \quad ; \quad \sum_{i=1}^{N} x_i = 1 \;\; ; \;\; x_i = \frac{X_i}{\sum_{i=1}^{N} X_i} \quad ; \;\; X_i \geq 0 \tag{1.3}$$

In (1.3), X_i denotes the "amount invested in security i" in the past. We consider all prices adjusted to the current time t. The paper by Markowitz (1952) is so famous that we see no reasons to add any comments or explanations to (1.3). However, we highlight a substantial parallel between Markowitz's definition of portfolio return R (1.3) and the definition of the average return (1.2) of N terms of time series during Δ (1.1). Both define the average return of returns' time series. Below we consider the similarity between the form of Markowitz's definition (1.3) and the form of the well-known definition of volume weighted average price (VWAP) that was proposed 36 years later (Berkowitz et al., 1988; Duffie and Dworczak, 2018). We consider Markowitz's definition of a portfolio's return as one that highlights the market-based origin of random properties of returns.

It is obvious that the differences between (1.2) and (1.3) are the results of different assumption and different approaches to the assessments of the average return of a given time series. One can easily see that if all the "amounts invested in security i" are constant, $X_i=constant$, $i=1,..N$, then "the relative amount invested in security i" $x_i=1/N$ and the portfolio return R (1.3) becomes equal to the average return $\rho(t,\tau;1)$ (1.2). Actually, there is almost no difference between the "amount invested in security i" X_i and the past value $C_0(t_i\text{-}\tau) = p(t_i\text{-}\tau)U(t_i)$ of the current trade volume $U(t_i)$ at price $p(t_i\text{-}\tau)$ of the past trade at time $t_i\text{-}\tau$. One can consider the definition of average return $\rho(t,\tau;1)$ (1.2) as one that neglects the impact of different "amounts invested in security i" X_i on average return.

We believe that top funds, banks, and investors, who perform multi-million market transactions and manage billion-valued portfolios, should care about the influence of the past values $C_0(t_i\text{-}\tau) = p(t_i\text{-}\tau)U(t_i)$ of current market trades on the average and volatility of returns. The predictions of market-based averages and volatilities of returns depend on forecasting of statistical moments and correlations of the current and past values of market trades. The use of the market-based statistical moments of returns is mandatory for the developers of large macroeconomic and market models like BlackRock's Aladdin, JP Morgan, and the U.S. Fed.

In Section 2, we introduce the trade return equation. In Section 3, we derive the dependence of the first four market-based statistical moments of returns on statistical moments and correlations of current and past trade values. In Section 4, we derive the expressions of return-value market-based correlations. In Section 5, we discuss how the dependence of market-based volatility of returns on statistical moments of current and past trade values limits the accuracy of forecasts of probability of return by Gaussian approximations. The conclusion is in Section 6. In Appendix A, we derive approximations of the characteristic functions and probability of returns by a finite number of statistical moments. In Appendix B we discus general definitions of market based statistical moments of returns. We calculate the 4th statistical moment through volatility of squares of returns and prove that such 4th statistical moment guarantees the non-negativity of kurtosis $Ku(t,\tau)$.

We assume that readers are familiar with conventional models of asset prices and stock returns and have skills in probability theory, statistical moments, characteristic functions, etc. All prices are adjusted to current time. We believe that readers know or can find on their own the definitions of terms that are not given in the text.

2. Returns equation

Let us define the value $C(t_i)$, volume $U(t_i)$, and price $p(t_i)$ of market trade at time t_i that follow the trivial price equation (2.1):

$$C(t_i) = p(t_i)U(t_i) \tag{2.1}$$

We assume that the time interval ε between the trades at time t_i and t_{i+1} is constant, ε = const and that the number $N>>1$ of the terms of the trade time series during Δ (1.1) is sufficiently large to assess the statistical moments of trade value and volume using regular frequency-based probability (1.2). The *n-th* statistical moments of trade value $C(t;n)$ (2.2) and volume $U(t;n)$ (2.3) during the averaging Δ (1.1) take the form:

$$C(t;n) \sim \frac{1}{N}\sum_{i=1}^{N} C^n(t_i) \quad ; \quad C_\Sigma(t;n) = N \cdot C^n(t;1) \;\; ; \;\; n = 1,2,.. \tag{2.2}$$

$$U(t;n) \sim \frac{1}{N}\sum_{i=1}^{N} U^n(t_i) \quad ; \quad U_\Sigma(t;n) = N \cdot U^n(t;1) \tag{2.3}$$

As $C_\Sigma(t;n)$ (2.2) and $U_\Sigma(t;n)$ (2.3) we denote total sum of N terms of time series of the n-th degrees of trade values $C^n(t_i)$ and volumes $U^n(t_i)$ during Δ (1.1) respectively. Relations (2.1-2.3) allow present the VWAP $p(t;1)$ (2.4) during Δ (1.1):

$$p(t;1) = \frac{1}{U_\Sigma(t;1)}\sum_{i=1}^{N} p(t_i)U(t_i) = \sum_{i=1}^{N} p(t_i)u(t_i,t;1) = \frac{C(t;1)}{U(t;1)} = \frac{C_\Sigma(t;1)}{U_\Sigma(t;1)} \tag{2.4}$$

$$u(t_i,t;1) = \frac{U(t_i)}{U_\Sigma(t;1)} \tag{2.5}$$

Relations (2.4; 2.5) highlight that VWAP $p(t;1)$ (2.4) is a ratio of total value $C_\Sigma(t;1)$ (2.2) to total volume $U_\Sigma(t;n)$ (2.2), or ratio of average value $C(t;1)$ (2.2) to average volume $U(t;1)$ (2.2), or sum of prices $p(t_i)$ weighted by the relative volumes $u(t_i,t;1)$ of market trades during Δ (1.1). That form is almost the same as the form of Markowitz's definition of the portfolio's return R (1.3), but weighted by the relative amounts x_i invested into security i. One can consider VWAP $p(t;1)$ (2.4) as the definition of the average price of N securities that compose the portfolio. The dualism between VWAP and Markowitz's definition of the portfolio's return R (1.3) is almost obvious: both have meaning of average values as for portfolio of N securities, as for the N terms of time series of market trades during the averaging interval Δ (1.1). To enhance and underline the similarity between them we use (1.1) and transfer the price equation (2.1) into return equation (2.6).

$$C(t_i) = p(t_i)U(t_i) = \frac{p(t_i)}{p(t_i-\tau)}p(t_i-\tau)U(t_i) = r(t_i,\tau)C_0(t_i,\tau)$$

$$C(t_i) = r(t_i,\tau)C_0(t_i,\tau) \quad ; \quad C_0(t_i,\tau) = p(t_i-\tau)U(t_i) \tag{2.6}$$

Function $C_0(t_i,\tau)$ in (2.6) have meaning of the past value of the current trade volume $U(t_i)$ at trade price $p(t_i\text{-}\tau)$ at time $t_i\text{-}\tau$ in the past. Its meaning almost completely equal to the meaning of "*the amount invested in the security i*" that is used by Markowitz. Let us define the n-th statistical moments of the past values $C_0(t,\tau;n)$ (2.7) and weight function of "relative past values" $z(t_i,\tau;n)$ (2.8)

$$C_o(t,\tau;n) \sim \frac{1}{N}\sum_{i=1}^{N} C_0^n(t_i,\tau) \quad ; \quad C_{\Sigma o}(t,\tau;n) = N \cdot C_o(t,\tau;n) \tag{2.7}$$

$$z(t_i,\tau;n) = \frac{C_0^n(t_i,\tau)}{\sum_{i=1}^{N} C_0^n(t_i,\tau)} \quad ; \quad \sum_{i=1}^{N} z(t_i,\tau;n) = 1 \tag{2.8}$$

The use of (2.6-2.8) allow present Markowitz's definition of return $r(t,\tau;1,1)$ of the portfolio composed of N securities with return $r(t_i,\tau)$ (1.1; 2.1) similar to the form VWAP (2.4; 2.5):

$$r(t,\tau;1,1) = \frac{1}{C_{\Sigma o}(t,\tau;1)}\sum_{i=1}^{N} r(t_i,\tau)C_o(t_i,\tau) = \frac{C(t;1)}{C_o(t,\tau;1)} = \frac{C_\Sigma(t;1)}{C_{\Sigma o}(t,\tau;1)} \tag{2.9}$$

$$r(t,\tau;1,1) = \sum_{i=1}^{N} r(t_i,\tau)z(t_i,\tau;1) \tag{2.10}$$

To highlight the alikeness of the forms of VWAP $p(t;1)$ (2.4) and average return $r(t,\tau;1,1)$ (2.9; 2.10) we denote it further as Value Weighted Average Return (VaWAR). We use the strong financial parallels between VaWAR $r(t,\tau;1,1)$ (2.9; 2.10) and VWAP $p(t;1)$ (2.4) and the identity of the forms of the price equation (2.1) and return equation (2.6) to introduce the market-based statistical moments of returns. The equation (2.6) is the basic for derivation of market-based statistical moments of returns. The similarity between price equation (2.1) and return equation (2.6) allows follow the results and description of market-based statistical moments of prices (Olkhov, 2022a; 2022b) and we refer there for details.

Let us consider the *n-th* degree of return equation (2.6):

$$C^n(t_i) = r^n(t_i,\tau)\, C_o^n(t_i,\tau) \quad ; \quad n = 1,2,\ldots \tag{2.11}$$

The equations (2.11) for $n=1,2,..$ prohibit the independent description of random properties of the *n-th* degrees of the current trade value $C^n(t_i)$, past value $C_o^n(t_i,\tau)$ (2.6), and return $r^n(t_i,\tau)$. The averaging over the weight functions $z(t_i,\tau;n)$ (2.8), for each $n=1,2,..$, determines new independent averaging procedure. One can consider the averaging of the *m-th* degree of returns $r^m(t_i,\tau)$ by the weight function $z(t_i,\tau;n)$ (2.8) and define the *m-th* statistical moments of returns $r(t,\tau;m,n)$ (2.12):

$$r(t,\tau;m,n) = \sum_{i=1}^{N} r^m(t_i,\tau) z(t,\tau;n) = \frac{1}{\sum_{i=1}^{N} C_o^n(t_i,\tau)} \sum_{i=1}^{N} r^m(t_i,\tau) C_o^n(t_i,\tau) \quad (2.12)$$

If $n=m$, then statistical moments of returns $r(t,\tau;n,n)$ (2.13) take the form:

$$r(t,\tau;n,n) = \frac{C(t;n)}{C_o(t,\tau;n)} = \frac{C_\Sigma(t;n)}{C_{\Sigma o}(t,\tau;n)} \quad (2.13)$$

Let us denote the average of the *n-th* degree of price $p^n(t_i)$ that is averaged over the weight function $u(t_i,t;n)$ (2.14) as the n-th VWAP $p(t;n)$ (2.15):

$$u(t_i,t;n) = \frac{U^n(t_i)}{U_\Sigma(t;n)} \quad (2.14)$$

$$p(t;n) = \sum_{i=1}^{N} p^n(t_i) u(t_i,t;n) = \frac{1}{U_\Sigma(t;n)} \sum_{i=1}^{N} p^n(t_i) U^n(t_i) = \frac{C(t,\tau;n)}{U(t;n)} \quad (2.15)$$

Let us consider the n-th degree of (2.6) as "past price" equation (2.16):

$$C_0^n(t_i,\tau) = p^n(t_i-\tau) U^n(t_i) \quad (2.16)$$

Similar to (2.15) derive the n-th VWAP of the past price $p(t\text{-}\tau;n)$ (2.17) averaged by the weight functions $u(t_i,t;n)$ (2.14):

$$p(t-\tau;n) = \sum_{i=1}^{N} p^n(t_i-\tau) u(t_i,t;n) = \frac{1}{U_\Sigma(t;n)} \sum_{i=1}^{N} p^n(t_i-\tau) U^n(t_i) = \frac{C_o(t,\tau;n)}{U(t;n)} \quad (2.17)$$

The definitions of current *n-th* VWAP $p(t;n)$ (2.15) and past price $p(t\text{-}\tau;n)$ (2.17) presents the n-th statistical moments of returns $r(t,\tau;n,n)$ (2.13) as (2.18):

$$r(t,\tau;n,n) = \frac{C(t;n)}{C_o(t,\tau;n)} = \frac{C(t;n)}{U(t;n)} \frac{U(t;n)}{C_o(t,\tau;n)} = \frac{p(t;n)}{p(t-\tau;n)} \quad (2.18)$$

$$C(t;n) = p(t;n) U(t;n) \quad ; \quad C_o(t,\tau;n) = p(t-\tau;n) U(t;n) \quad (2.19)$$

Equations (2.19) highlight the dependence of the *n-th* statistical moments of current values $C(t;n)$ (2.2) and past values $C_o(t,\tau;n)$ (2.7) on the *n-th* statistical moments $U(t;n)$ (2.3) of trade volumes and on n-th VWAP of current $p(t;n)$ (2.15) and past prices $p(t\text{-}\tau;n)$ (2.17).

One can see, that if all past values $C_o(t_i,\tau)$ (2.6) of all N trades at t_i during Δ (1.1) are constant, then the weight functions $z(t_i,\tau;n)$ (2.8) for all $n=1,2,..$ equal to $z(t_i,\tau;n)=1/N$ and all statistical moments of return take the form similar to (2.2; 2.3) that is determined by frequency-based averaging. The assumption that all past values $C_o(t_i,\tau)$=const almost equals to the assumptions that all current trade volumes $U(t_i)=const$ and past prices $p(t_i\text{-}\tau)=constant$ during Δ (1.1). That rather exotic case is far from any market reality. The consideration of the influence of the observed random trade volumes $U(t_i)$ and random or irregular past prices $p(t_i\text{-}\tau)$ on the assessments of the *n-th* statistical moments of returns $r(t,\tau;n,n)$ (2.13) during Δ (1.1) requires mandatory use of (2.12-2.19).

Now let us consider the derivation of *market-based statistical moments* of returns.

3. Market-based statistical moments of returns

The averaging (2.12) by the different weight functions $z(t_i,\tau;n)$ (2.8) determines the sets of the *m-th* statistical moments $r(t,\tau;m,n)$ (2.12) of returns that could be non-consistent for different $n=1,2,..$. Simply speaking, the set of the statistical moments $r(t,\tau;n,n)$ (2.12) of returns for different n could result in negative volatility. To select the self-consistent set of market-based statistical moments of return, one should prove non-negativity of each statistical moment and non-negativity of their even central moments. In this section we derive the first four self-consistent market-based statistical moments of returns. In Section 5, we explain why that is more than sufficient for forecasting the market-based probability of returns. The first four market-based statistical moments describe the approximations of the characteristic function and probability of returns as a random variable (App. A.).

We consider VaWAR $r(t,\tau;1,1)$ (2.9; 2.10) as the 1st market-based statistical moment $h(t,\tau;1)$ and use such a notation to distinguish it from statistical moments $r(t,\tau;m,n)$ (2.12). We denote market-based mathematical expectation $E_m[..]$ and define:

$$E_m[r(t_i,\tau)] = h(t,\tau;1) = r(t,\tau;1,1) \tag{3.1}$$

$$h(t,\tau;1) = \frac{1}{C_{\Sigma o}(t,\tau;1)} \sum_{i=1}^{N} r(t_i,\tau) C_o(t_i,\tau) = \frac{C(t;1)}{C_o(t,\tau;1)} = \frac{C_\Sigma(t;1)}{C_{\Sigma o}(t,\tau;1)} \tag{3.2}$$

To justify the choice (3.1; 3.2), we refer to Markowitz's definition of portfolio return (1.3) and consider it as the one that supports the economic sense of (3.2). The second argument in favor of (3.2) is that VaWAR (3.2) almost completely reproduces the economic meaning of VWAP. One can consider Markowitz's definition of a portfolio return as the origin for the definition of VWAP. Both have the same economic meaning and the same form.

3.1. Market-based volatility of returns

We define market-based volatility of returns $\sigma^2(t,\tau|1)$ (3.4) as the market-based mathematical expectation of the square of variations of return $\delta^2 r(t_i,\tau)$ (3.3) near $h(t,\tau;1)$:

$$\delta r(t_i,\tau) \ = \ r(t_i,\tau) - h(t,\tau;1) \tag{3.3}$$

$$\sigma^2(t,\tau|1) = E_m[\delta^2 r(t_i,\tau)] = h(t,\tau;2) - h^2(t,\tau;1) \tag{3.4}$$

The index 1 in the notation of volatility $\sigma^2(t,\tau|1)$ (3.4) highlights that it is volatility of the 1st degree of returns. The 2nd statistical moment $h(t,\tau;2)$ of returns has the form (3.5):

$$E_m[r^2(t_i,\tau)] = h(t,\tau;2) \geq 0 \tag{3.5}$$

However, to calculate (3.4) and (3.5) in a self-consistent form, one should guarantee that $\sigma^2(t,\tau|1) \geq 0$. To do that we average $\delta^2 r(t_i,\tau)$ (3.3) by the weight function $z(t_i,\tau;2)$ (2.8) that is determined by the relative amounts of squares of past trade values $C_o^2(t_i,\tau)$ and the equation (2.11) for $n=2$. Thus we define market-based volatility of returns $\sigma^2(t,\tau|1)$:

$$\sigma^2(t,\tau|1) = \sum_{i=1}^{N} \delta^2 r(t_i,\tau)\, z(t,\tau;2) = \sum_{i=1}^{N} (r(t_i,\tau) - h(t,\tau;1))^2 z(t,\tau;2) \qquad (3.6)$$

The form of the weight function $z(t_i,\tau;2)$ (2.8) guarantees that market-based volatility of return $\sigma^2(t,\tau|1)\geq 0$ is non-negative and (3.4) guarantees non-negativity of $h(t,\tau;2)\geq 0$. To calculate (3.6), we use relations (2.12):

$$\sigma^2(t,\tau|1) = r(t,\tau;2,2) - 2r(t,\tau;1,2)h(t,\tau;1) + h^2(t,\tau;1) \qquad (3.7)$$

From (2.12; 2.13), obtain:

$$r(t,\tau;2,2) = \frac{C(t;2)}{C_o(t,\tau;2)}$$

From (2.12) and returns' equation (2.6), obtain:

$$r(t,\tau;1,2) = \frac{1}{C_o(t,\tau;2)} \frac{1}{N} \sum_{i=1}^{N} r(t_i,\tau) C_o^2(t_i,\tau) = \frac{1}{C_o(t,\tau;2)} \frac{1}{N} \sum_{i=1}^{N} C(t_i) C_o(t_i,\tau)$$

We denote usual frequency-based mathematical expectation as $E[..]$ and present:

$$E[C(t_i)C_o(t_i,\tau)] = \frac{1}{N} \sum_{i=1}^{N} C(t_i)C_o(t_i,\tau) = C(t;1)C_o(t,\tau;1) + corr\{C(t), C_o(t,\tau)\}$$

We omit simple transformations and obtain the expression for the market-based volatility of return $\sigma^2(t,\tau)$:

$$\sigma^2(t,\tau|1) = \frac{\Omega^2(t|1) + h^2(t,\tau;1)\Phi^2(t,\tau|1) - 2h(t,\tau;1)corr\{C(t), C_o(t,\tau)\}}{C_o(t,\tau;2)} \qquad (3.8)$$

In (3.8), we denote volatility $\Omega(t|1)$ (3.9) of the current trade value and volatility $\Phi(t,\tau|1)$ (3.10) of the past trade value:

$$\Omega^2(t|1) = E[(C(t_i) - C(t;1))^2] = C(t;2) - C^2(t;1) \qquad (3.9)$$

$$\Phi^2(t,\tau|1) = E[(C_o(t_i,\tau) - C_o(t,\tau;1))^2] = C_o(t,\tau;2) - C_o^2(t,\tau;1) \qquad (3.10)$$

The notation $\Omega(t|1)$ highlights that (3.9) denotes volatility of the 1st degree of trade value. From (3.4; 3.5) and (3.8), obtain the expression of the market-based 2nd statistical moment of returns $h(t,\tau;2)$:

$$h(t,\tau;2) = \frac{C(t;2) + 2h^2(t,\tau;1)\Phi^2(t,\tau|1) - 2h(t,\tau;1)corr\{C(t), C_o(t,\tau)\}}{C_o(t,\tau;2)} \qquad (3.11)$$

The above relations reveal the direct dependence of the market-based volatility $\sigma^2(t,\tau|1)$ (3.8) and the 2nd statistical moment $h(t,\tau;2)$ (3.11) of return on the first two frequency-based statistical moments, volatilities, and correlations of the current $C(t_i)$ and past $C_o(t_i,\tau)$ trade values. In turn, the relations (2.19) highlight the dependence of the first two statistical moments of past values $C_o(t,\tau;1)$ and $C_o(t,\tau;2)$ on statistical moments $U(t;1)$ and $U(t;2)$ of trade volumes and VWAP past prices $p(t\text{-}\tau;1)$ and $p(t\text{-}\tau;2)$ (2.17) with respect to the relative amount of current trade volumes $u(t_i,t;1)$ and $u(t_i,t;2)$ (2.14).

To predict the market-based volatility $\sigma^2(t,\tau|1)$ (3.8) of returns, one should forecast the statistical moments and correlations of the values and volumes of market trades. Any predictions of market-based volatility $\sigma^2(t,\tau|1)$ of returns, which ignore the dependence on statistical moments and correlations of the values, volumes, and past prices (2.17; 2.19; 3.8-3.10), could be untrustworthy. It is particularly important for top funds, banks, and investors, who perform multi-million market transactions and manage billion-valued portfolios, and for the developers of large macroeconomic and market models like BlackRock's Aladdin, JP Morgan, and the U.S. Fed. The randomness of market trades is the origin of return's volatility, and its influence must be taken into account.

3.2. The 3rd market-based statistical moment of returns

To guarantee non-negativity of the 3rd statistical moment $h(t,\tau;3)$ of returns we define it as the average (2.13) over the weight function $z(t_i,\tau;3)$ (2.8):

$$h(t,\tau;3) = r(t,\tau;3,3) = \frac{1}{C_o(t,\tau;3)}\frac{1}{N}\sum_{i=1}^{N} r^3(t_i,\tau)C_o^3(t_i,\tau) = \frac{C(t;3)}{C_o(t,\tau;3)} \geq 0 \quad (3.12)$$

If one takes that all past values are constant, $C_o(t_i,\tau)=constant$ during the averaging interval Δ (1.1), then the 3rd statistical moment $h(t,\tau;3)$ (3.12) take the form that is determined by the frequency-based assessments of time series of returns.

3.3. The 4th market-based statistical moment of returns

The 4th market-based statistical moment $h(t,\tau;4)\geq 0$ of return should be non-negative and it should guarantee that two even market-based central statistical moments $E_m[(r(t_i,\tau)-h(t,\tau;1))^4]$ and $E_m[(r^2(t_i,\tau)-h(t,\tau;2))^2]$ are non-negative (App. B). To fulfill that we determine the volatility of squares of returns $\sigma^2(t,\tau|2)=E_m[(r^2(t_i,\tau)-h(t,\tau;2))^2]$ as average over the weight function $z(t_i,\tau;4)$ (2.8):

$$\sigma^2(t,\tau|2) = E_m\left[\left(r^2(t_i,\tau) - h(t,\tau;2)\right)^2\right] = h(t,\tau;4) - h^2(t,\tau;2) \geq 0 \quad (3.13)$$

$$\sigma^2(t,\tau|2) = \sum_{i=1}^{N}(r^2(t_i,\tau) - h(t,\tau;2))^2 z(t,\tau;4) \quad (3.14)$$

$$\sigma^2(t,\tau|2) = r(t,\tau;4,4) - 2r(t,\tau;2,4)h(t,\tau;2) + h^2(t,\tau;2) \quad (3.15)$$

From (2.12; 2.13), obtain:

$$r(t,\tau;4,4) = \frac{C(t;4)}{C_o(t,\tau;4)}$$

From (2.12) and (2.6), obtain:

$$r(t,\tau;2,4) = \frac{1}{C_o(t,\tau;4)}\frac{1}{N}\sum_{i=1}^{N} r^2(t_i,\tau)C_o^4(t_i,\tau) = \frac{1}{C_o(t,\tau;4)}\frac{1}{N}\sum_{i=1}^{N} C^2(t_i)C_o^2(t_i,\tau)$$

We denote usual frequency-based mathematical expectation as $E[..]$ and present:

$$E[C^2(t_i)C_o^2(t_i,\tau)] = \frac{1}{N}\sum_{i=1}^{N} C^2(t_i)C_o^2(t_i,\tau) = C(t;2)C_o(t,\tau;2) + corr\{C^2(t), C_o^2(t,\tau)\}$$

We omit simple transformations and obtain the expression for the market-based volatility of squares $\sigma^2(t,\tau|2)$ (3.16) and for the 4th statistical moment $h(t,\tau;4)\geq 0$ (3.17) that is non-negative due to (3.13).

$$\sigma^2(t,\tau|2) = \frac{\Omega^2(t|2)+h^2(t,\tau;2)\Phi^2(t,\tau|2)-2h(t,\tau;2)corr\{C^2(t_i),C_0^2(t_i,\tau)\}}{C_o(t,\tau;4)} \geq 0 \quad (3.16)$$

$$h(t,\tau;4) = \frac{C(t;4)+2h^2(t,\tau;2)\Phi^2(t,\tau|2)-2h(t,\tau;2)corr\{C^2(t_i),C_0^2(t_i,\tau)\}}{C_o(t,\tau;4)} \geq 0 \quad (3.17)$$

Functions $\Omega(t|2)$ (3.18) and $\Phi(t,\tau|2)$ (3.19) denote volatility of squares $C^2(t_i)$ of current and past $C_o^2(t_i,\tau)$ trade values:

$$\Omega^2(t|2) = E[(C^2(t_i) - C(t;2))^2] = C(t;4) - C^2(t;2) \quad (3.18)$$

$$\Phi^2(t,\tau|2) = E[(C_0^2(t_i,\tau) - C_o(t,\tau;2))^2] = C_o(t,\tau;4) - C_o^2(t,\tau;2) \quad (3.19)$$

In App. B. we prove that the 4th central statistical moment that is usually noted as Kurtosis $Ku(t,\tau)$ (3.20; 3.21) is non-negative:

$$Ku(t,\tau)\sigma^4(t,\tau|1) = E_m\left[\left(r(t_i,\tau) - h(t,\tau;1)\right)^4\right] \quad (3.20)$$

$$Ku(t,\tau)\sigma^4(t,\tau|1) = h(t,\tau;4) - 4h(t,\tau;3)h(t,\tau;1) + 6h(t,\tau;2)h^2(t,\tau;1) - 3h^4(t,\tau;1) \quad (3.21)$$

The non-negativity of $Ku(t,\tau)$ (3.21) proves that the definitions of the first four market-based statistical moments $h(t,\tau;n)$, $n=1,2,3,4$ are self-consistent.

3.4. Higher market-based statistical moments of returns

In App. B. we derive higher market-based statistical moments of returns. The odd statistical moments $h(t,\tau;2k-1)$ (3.22; B.12):

$$h(t,\tau;2k-1) = \frac{C(t;2k-1)}{C_o(t,\tau;2k-1)} \geq 0 \quad ; \quad k = 1,2,3,\ldots \quad (3.22)$$

The even statistical moments $h(t,\tau;2k)$ (3.23; B.13-B.16) take the form:

$$h(t,\tau;2k) = \frac{C(t;2k)+2h^2(t,\tau;k)\Phi^2(t,\tau|k)-2h(t,\tau;k)corr\{C^k(t_i),C_0^k(t_i,\tau)\}}{C_o(t,\tau;2k)} \geq 0 \quad ; \quad k = 1,2,3,\ldots \quad (3.23)$$

The relations (3.22; 3.23) and (B.12-B.18) define market-based statistical moments $h(t,\tau;n)$ of return for $n=1,2,..q<N$. The number q of statistical moments determines the accuracy of q-approximation of the characteristic function (A.4-A.6) and the probability (A.2) of returns that is limited by the number N of market trades during the averaging interval Δ (1.1). However, as we discuss in Section 5., the economic-based reasons limit the number of predicted statistical moments of current and past trade values and hence the number of predicted of statistical moments of returns by the first two. That limits the accuracy of economic-based forecasts of market-based probability of returns by the accuracy of the Gaussian approximations.

4. Correlations

We highlight that correlations between two economic time series could be assessed in different manners. Two time series themselves don't determine the method for the assessments of their correlation. The simplest way is to consider two time series independently from other economic factors and use the conventional frequency-based expressions of correlation between two time series (Campbell et al., 1993). However, one should take into account the market-based origin of return. To use the correct statistical approach for the assessments of correlations of two time series, one should follow their economic and market-based logic. Below we consider correlations between return, past values, and past price that cause from the market-based approach to the description of random time series of returns, prices, and trade values.

4.1 Correlation between Returns and Past Values

Let us derive market-based expressions of correlation between returns $r(t_i,\tau)$ and past trade values $C_o(t_i,\tau)$. From (2.2; 2.6; 2.7; 2.9; 3.1), obtain that the correlation $corr\{r(t,\tau),C_o(t,\tau)\}$ between return $r(t,\tau)$ and past trade values $C_o(t,\tau)$ during Δ (1.1) equals to zero:

$$corr\{r(t,\tau),C_o(t,\tau)\} = E[r(t_i,\tau)C_o(t_i,\tau)] - E_m[r(t_i,\tau)]\,E[C_o(t_i,\tau)] \tag{4.1}$$

From (2.6), obtain:

$$E[r(t_i,\tau)C_o(t_i,\tau)] = E[C(t_i)] = C(t;1)$$

Hence from (2.7; 2.9; 3.1), obtain:

$$corr\{r(t,\tau),C_o(t,\tau)\} = C(t;1) - h(t,\tau;1)\,C_o(t,\tau;1) = 0 \tag{4.2}$$

4.2 Correlation between Returns and Squares of Past Values

The expression of correlation $corr\{r(t,\tau),C_o^2(t,\tau)\}$ between returns $r(t_i,\tau)$ and squares of the past trade values $C_o^2(t_i,\tau)$ takes the form:

$$corr\{r(t,\tau),C_o^2(t,\tau)\} = E[r(t_i,\tau)C_o^2(t_i,\tau)] - E_m[r(t_i,\tau)]\,E[C_o^2(t_i,\tau)]$$

From (2.6):

$$E[r(t_i,\tau)C_o^2(t_i,\tau)] = E[C(t_i)C_o(t_i,\tau)] = C(t;1)C_o(t,\tau;1) + corr\{C(t),C_o(t,\tau)\}$$

$$E[C(t_i)C_o(t_i,\tau)] = \frac{1}{N}\sum_{i=1}^{N} C(t_i)C_o(t_i,\tau)$$

Hence, correlations $corr\{r(t,\tau),C_o^2(t,\tau)\}$ during Δ (1.1) takes the form:

$$corr\{r(t,\tau),C_o^2(t,\tau)\} = C(t;1)C_o(t,\tau;1) - h(t,\tau;1)\,C_o(t,\tau;2) + corr\{C(t),C_o(t,\tau)\}$$

From (3.2; 3.10), obtain:

$$corr\{r(t,\tau),C_o^2(t,\tau)\} = corr\{C(t),C_o(t,\tau)\} - h(t,\tau;1)\,\Phi^2(t,\tau) \tag{4.3}$$

Relations (4.3) reveal the dependence of the market-based $corr\{r(t,\tau),C_o^2(t,\tau)\}$ between returns $r(t,\tau)$ and squares of the past trade values $C_o^2(t,\tau)$ on the correlations $corr\{C(t),C_o(t,\tau)\}$ between current $C(t)$ and past $C_o(t,\tau)$ trade values.

4.3 Correlation between Returns and Past Prices

Correlation $corr\{r(t,\tau),p(t-\tau)\}$ between returns $r(t_i,\tau)$ and past prices $p(t_i-\tau)$ takes the form:

$$corr\{r(t,\tau),p(t-\tau)\} = E[r(t_i,\tau)p(t_i-\tau)] - E_m[r(t_i,\tau)]\,E_m[p(t_i-\tau)] \qquad (4.4)$$

$$E_m[r(t_i,\tau)p(t_i-\tau)] = E_m[p(t_i)]$$

$$corr\{r(t,\tau),p(t-\tau)\} = E_m[p(t_i)] - E_m[r(t_i,\tau)]\,E_m[p(t_i-\tau)]$$

From (2.4; 2.17) and (2.18), for $n=1$, obtain:

$$corr\{r(t,\tau),p(t-\tau)\} = p(t;1) - h(t,\tau;1)\,p(t-\tau;1) = 0$$

The market-based correlation $corr\{r(t,\tau),p(t-\tau)\}$ (4.4) between returns $r(t_i,\tau)$ and past prices $p(t_i-\tau)$ likewise correlation $corr\{r(t,\tau),C_o(t,\tau)\}$ (4.1) between returns $r(t_i,\tau)$ and past trade values $C_o(t_i,\tau)$ equals to zero.

5. Economic complexity limits the number of predicted statistical moments

In this section we briefly explain the economic factors that limit the predictions of market-based statistical moments by the first two. In turn, that limits the accuracy of any forecasts of probability of returns by Gaussian approximations.

The randomness of market trades is the origin of these limitations. Indeed, the 1st statistical moments $C(t;1)$ and $U(t;1)$ (2.4) depend on the sums of the 1st degrees of the values and volumes of market trades during the averaging interval Δ. Hence, the market-based average return $h(t,\tau;1)$ (3.1) as well as VWAP $p(t;1,1)$ (2.14) also depend on sums of the 1st degrees of the values and volumes of market trades during the averaging interval Δ. We consider market trade as the only origin of change of macroeconomic variables. Any economic regulations, market restrictions, and economic or monetary policies have the impact on macroeconomic variables only after at least one market transaction under these policies occurs. All macroeconomic variables, such as investment and consumption, credit and sales, etc., depend on the values or volumes of corresponding transactions during Δ. We denote variables that are determined by sums of values or volumes of market trades as the 1st order variables. Actually, the market data about all market trades during the averaging interval Δ (1.1) that are required for the assessments of the changes of macroeconomic variables during Δ (1.1) can be absent. Current econometric methodologies (Fox et al., 2019) use available economic data to estimate macroeconomic variables. The collisions of market trade data with econometric methodologies result in additional uncertainty of econometric assessments. Macroeconomic theories use econometric assessments and describe the relations between macroeconomic variables. We call economic models that describe the dependence between economic and financial variables that are composed of the sums of the 1st degrees of the values or volumes of market trades as the 1st order theories.

However, markets essentially depend on the volatility of market prices and returns. As we show, market-based volatility of prices (Olkhov, 2022a), volatility $\sigma^2(t,\tau)$ (3.8) of returns, and the 2-d statistical moment $h(t,\tau;2)$ (3.9) of returns depend on sums of ***squares*** of values and volumes of market trades. That dependence on ***squares*** *of values and volumes* introduces macroeconomic variables that are significantly different from the 1st order variables composed of sums of values or volumes. The volatilities and 2nd statistical moments of the values and volumes of market trades, prices, returns, and macroeconomic variables that depend on them establish the set of variables that complement the 1st order variables. Actually, any econometric assessments of macroeconomic variables are not "exact," and the uncertainty of observations of macroeconomic variables can be described by volatilities that depend on 2nd statistical moments of corresponding values or volumes of market trades (Olkhov, 2021b; 2023a; 2023b). Simply speaking, one can consider macroeconomic variables during the averaging interval Δ as random variables that are determined by random values or volumes of market trades. The averages of random macroeconomic variables depend on sums of values or volumes of trades and establish the set of the conventional macroeconomic variables of the 1st order. However, the volatilities of random macroeconomic variables, prices, and returns depend on the sums of ***squares*** of values and volumes of market trades. It is obvious that macroeconomic evolution, financial markets, and business cycles essentially depend on the mutual relations between the 1st and 2nd order macroeconomic variables. However, current econometric estimates only consider macroeconomic variables of the 1st order. Econometric assessments of volatilities of values and volumes of market trades and market-based volatilities of price and returns (3.8) are absent. Modern economic theories don't describe mutual dependence of the 1st and 2nd order variables. That makes it impossible to develop economic-based forecasts of 2nd statistical moments of trade values and volumes, prices, and returns. One can use market records and assess the current values of the 3rd and 4th statistical moments of trade values, volumes, prices, and returns (3.16; 3.17) that depend on sums of 3rd and 4th degrees of values and volumes. However, economic predictions of the 3rd and 4th statistical moments require econometric assessments and economic models that describe the evolution of such variables. All of that is absent now. The nearest, but still far goal is the description of the 1st and 2nd order variables. Till then, the economic-founded predictions of the market-based statistical moments of trades, prices, returns, and macroeconomic variables are limited by the first two. Even the forecasts of the 2nd statistical moments up now have almost no economic ground. Hence, the forecasts of their probabilities are limited by Gaussian approximations.

6. Conclusion

The irregular time series of stock returns themselves don't uniquely determine the averaging procedure and the probability distribution. The conventional approach considers return time series as a stand-alone sample of a random variable, and that results in a frequency-based assessment of the average return (1.2). However, the market nature of the randomness of stock returns implies that statistical moments of returns should depend on statistical moments of market trades. The market-based origin of statistical moments of returns results in their dependence on statistical moments and correlations of current and past trade values. If one assumes that all past values are constant, then market-based statistical moments of returns take the form that is determined by the usual frequency-based assessments of statistical moments of random time series of returns.

The choice between frequency-based and market-based assessments of statistical moments of returns is determined by the habits and goals of investors. The largest investors, traders, and banks, who manage billion-valued portfolios and perform multi-million market transactions, should care about the impact of the randomness of trade values and volumes on the statistical moments of returns.

To predict market-based statistical moments of stock return, one should forecast statistical moments and correlations of current and past trade values. That complicates the assessment of market-based statistical moments of return but highlights direct market ties of trade stochasticity and the randomness of stock returns. The use of the market-based statistical moments of returns is mandatory for the developers of large macroeconomic and market models like BlackRock's Aladdin, JP Morgan, and the U.S. Fed.

Frequency-based assessments of return statistics are simpler, familiar, and may follow the expectations of most investors. These expectations can influence investment decisions and thus impact market trade stochasticity and, hence, the randomness of stock returns. However, the forecasts of the frequency-based properties of random returns have almost no ties with predictions of market trade statistical moments. Thus, frequency-based forecasts have low reliability and a poor economic basis.

Eventually, the description of the random properties of stock returns as a result of market trades requires the use of market-based statistical moments.

Appendix A. Approximations of the characteristic functions and probability measures of random returns by a finite set of statistical moments

We consider stock return as a random variable during the averaging interval *Δ* (1.1). The random variable can equally be described by the set of statistical moments, characteristic function *F(t,τ;x)* (A.1), and probability measure *μ(t,τ;r)* (A.2) (Shephard, 1991; Shiryaev, 1999; Shreve, 2004). The Taylor series expansion of the market-based characteristic function *F(t,τ;x)* presents it through the set of market-based *n-th* statistical moments *h(t,τ;n)*:

$$F(t,\tau;x) = 1 + \sum_{n=1}^{\infty} \frac{i^n}{n!} h(t,\tau;n)\, x^n \tag{A.1}$$

$$\mu(t,\tau;r) = \frac{1}{\sqrt{2\pi}} \int F(t;x) \exp(-ixr)\, dx \tag{A.2}$$

$$h(t,\tau;n) = \frac{d^n}{(i)^n dx^n} F(t,\tau;x)|_{x=0} = \int r^n \mu(t,\tau;r)\, dr \quad ; \quad \int \mu(t,\tau;r)\, dr = 1 \tag{A.3}$$

In (A.1-A.3), i is the imaginary unit. For simplicity, we take the returns as a continuous random variable during *Δ* (1.1). A finite number q of the statistical moments *h(t,τ;n)*, *n=1,2,..q,* determines the q-approximation of the price characteristic function $F_q(t,\tau;x)$ (A.4):

$$F_q(t,\tau;x) = 1 + \sum_{n=1}^{q} \frac{i^n}{n!} h(t,\tau;n)\, x^n \tag{A.4}$$

Taylor expansion (A.4) is not too useful to derive the Fourier transform (A.2), and to obtain a q-approximation of the price probability measure $\mu_q(t,\tau;r)$, we consider the approximation of price characteristic functions $G_q(t,\tau;x)$ (A.5):

$$G_q(t,\tau;x) = exp\left\{\sum_{n=1}^{q} \frac{i^n}{n!}\, b(t,\tau;n)\, x^n - B\, x^{2Q}\right\} \; ; \; q = 1, 2, \ldots ; \; q < 2Q \; ; \; B > 0 \tag{A.5}$$

and require that $G_q(t,\tau;x)$ (A.5) obey relations (A.3):

$$h(t,\tau;n) = \frac{d^n}{(i)^n dx^n} G_q(t,\tau;x)|_{x=0} \quad ; \quad n \le q \tag{A.6}$$

Relations (A.6) define functions *b(t,τ;n)* in (A.5) through market-based statistical moments *h(t,τ;n)*, $n \le q$. The terms Bx^{2Q}, $B>0$, and $2Q>q$ don't impact relations (A.3; A.6) but guarantee the existence of the probability measures $\mu_q(t,\tau;r)$ as the Fourier transform (A.2) of the characteristic functions $G_q(t,\tau;x)$ (A.5). The uncertainty of $B>0$ and the degree $2Q>q$ in (A.5) highlight the well-known fact that the first q statistical moments don't explicitly determine the characteristic function and probability measure of a random variable. Relations (A.5) describe the set of characteristic functions $G_q(t,\tau;x)$ with different $B>0$ and $2Q>q$ and the corresponding set of probability measures $\mu_q(t,\tau;r)$ that match (A.2; A.5; A.6).

For $q=1$, the approximate characteristic function $G_1(t,\tau;x)$ and probability $\mu_q(t,\tau;r)$:

$$G_1(t,\tau;x) = exp\{i\, b(t,\tau;1)x\} \;\; ; \; h(t,\tau;1) = -i\frac{d}{dx}G_1(t,\tau;x)|_{x=0} = b(t,\tau;1) \quad \text{(A.7)}$$

$$\mu_1(t,\tau;r) = \int dx\;\; G_1(t,\tau;x)\, exp(-irx) = \delta\big(r - \; b(t,\tau;1)\big) \quad \text{(A.8)}$$

For $q=2$, the approximation $G_2(t,\tau;x)$ describes the Gaussian probability measure $\mu_2(t,\tau;r)$:

$$G_2(t,\tau;x) = exp\left\{i\, b(t,\tau;1)x - \frac{b(t,\tau;2)}{2}x^2\right\} \quad \text{(A.9)}$$

One can show that:

$$h(t,\tau;2) = -\frac{d^2}{dx^2}G_2(t,\tau;x)|_{x=0} = b(t,\tau;2) + b^2(t,\tau;1)$$

$$b(t,\tau;2) = h(t,\tau;2) - b^2(t,\tau;1) = \sigma^2(t,\tau|1) \quad \text{(A.10)}$$

The coefficient $b(t,\tau;2)$ equals the market-based volatility $\sigma^2(t,\tau|1)$ of returns (3.7; 3.8), and the Fourier transform (A.2) for $G_2(t,\tau;x)$ (A.9) gives the Gaussian price probability $\mu_2(t,\tau;r)$:

$$\mu_2(t,\tau;r) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma^2(t,\tau|1)} exp\left\{-\frac{(r-b(t,\tau;1))^2}{2\sigma^2(t,\tau|1)}\right\} \quad \text{(A.11)}$$

One can consider also non-Gaussian approximations of the characteristic function $G_2(t,\tau;x)$:

$$G_2(t,\tau;x) = exp\left\{i\, b(t,\tau;1)x - \frac{b(t,\tau;2)}{2}x^2 - B\, x^{2Q}\right\} \quad \text{(A.12)}$$

For $q=3$, the approximation $G_3(t,\tau;x)$ of characteristic function has the form:

$$G_3(t,\tau;x) = exp\left\{i\, b(t,\tau;1)x - \frac{b(t,\tau;2)}{2}x^2 - i\,\frac{b(t,\tau;3)}{6}x^3 - B\, x^{2Q}\right\} \quad \text{(A.13)}$$

$$h(t,\tau;3) = i\frac{d^3}{dx^3}G_3(t;x)|_{x=0} = b(t,\tau;3) + 3b(t,\tau;2)\sigma^2(t,\tau|1) + \; b^3(t,\tau;1)$$

$$b(t,\tau;3) = E_m\left[\big(r - b(t,\tau;1)\big)^3\right] = Sk(t,\tau)\sigma^3(t,\tau|1) \quad \text{(A.14)}$$

The coefficient $b(t,\tau;3)$ (A.14) depends on the market-based skewness $Sk(t,\tau)$ (3.13-3.15) of returns and describes the asymmetry of the probability from the normal distribution.

If $q=4$, then the approximations of characteristic function $G_4(t;x)$ depends on the choice of $B>0$ and the degree $2Q>4$:

$$G_4(t,\tau;4) = exp\left\{i\, b(t,\tau;1)x - \frac{b(t,\tau;2)}{2}x^2 - i\,\frac{b(t,\tau;3)}{6}x^3 + \frac{b(t,\tau;4)}{24}x^4 - Bx^{2Q}\right\} \;\; ; \; 2Q > 4 \quad \text{(A.15)}$$

Simple, but long calculations give:

$$b(t,\tau;4) = h(t,\tau;4) - 4h(t,\tau;3)h(t,\tau;1) + 12h(t,\tau;2)h^2(t,\tau;1) - 6h^4(t,\tau;1) - 3h^2(t,\tau;2)$$

$$b(t,\tau;4) = E_m\left[\big(r - b(t,\tau;1)\big)^4\right] - 3E_m^2\left[\big(r - b(t,\tau;1)\big)^2\right]$$

Kurtosis $Ku(t,\tau)$ of return (B.1) describes the distinction of the tails of return probability measure $\mu_4(t,\tau;r)$ from the tails of a normal distribution.

$$Ku(t,\tau)\sigma^4(t,\tau|1) = E_m\left[\big(r - b(t,\tau;1)\big)^4\right] \quad ; \quad b(t,\tau;4) = [Ku(t,\tau) - 3]\sigma^4(t,\tau|1)$$

Appendix B. Non-negativity of kurtosis *Ku(t,τ)* and definitions of higher statistical moments of returns

Let us consider the expression of kurtosis *Ku(t,τ)* (3.20; 3.21):

$$Ku(t,\tau)\sigma^4(t,\tau|1) = h(t,\tau;4) - 4h(t,\tau;3)h(t,\tau;1) + 6h(t,\tau;2)h^2(t,\tau;1) - 3h^4(t,\tau;1) \quad \text{(B.1)}$$

In this Appendix we prove that market-based statistical moments *h(t,τ;1)* (3.1;3.2), *h(t,τ;2)* (3.4; 3.11), *h(t,τ;3)* (3.12), *h(t,τ;4)* (3.13; 3.3.17) determine non-negative kurtosis *Ku(t,τ)* (B.1). Let us consider volatility of squares $\sigma^2(t,\tau|2)$ (3.13; 3.16) of returns and present *h(t,τ;4)* as:

$$h(t,\tau;4) = \sigma^2(t,\tau|2) + h^2(t,\tau;2) \quad \text{(B.2)}$$

Let us use (3.4) and present *h(t,τ;2)* as (B.3):

$$h(t,\tau;2) = \sigma^2(t,\tau|1) + h^2(t,\tau;1) \quad \text{(B.3)}$$

Let us substitute (B.2; B.3) into (B.1) and obtain:

$$Ku(t,\tau)\sigma^4(t,\tau) = \sigma^2(t,\tau|2) + \sigma^4(t,\tau|1) + 8\sigma^2(t,\tau|1)h^2(t,\tau;1) + 4h^4(t,\tau;1) - 4h(t,\tau;3)h(t,\tau;1) \quad \text{(B.4)}$$

Let us present 3rd statistical moment *h(t,τ;3)* as:

$$h(t,\tau;3) = E_m\left[r^3(t_i,\tau)\right] = h(t,\tau;1)h(t,\tau;2) + corr\{r(t_i,\tau), r^2(t_i,\tau)\} \quad \text{(B.5)}$$

$$corr\{r(t_i,\tau), r^2(t_i,\tau)\} = \sigma(t,\tau|1)\sigma(t,\tau|2)\frac{corr\{r(t_i,\tau), r^2(t_i,\tau)\}}{\sigma(t,\tau|1)\sigma(t,\tau|2)} \quad \text{(B.6)}$$

By the Cauchy-Schwartz-Bunyakovskii inequality (Shiryaev, 1999, p.123), the coefficient of correlation of returns is less than a unit:

$$-1 \le \frac{corr\{r(t_i,\tau), r^2(t_i,\tau)\}}{\sigma(t,\tau|1)\sigma(t,\tau|2)} \le 1 \quad ; \quad corr\{r(t_i,\tau), r^2(t_i,\tau)\} \le \sigma(t,\tau|1)\sigma(t,\tau|2)$$

Hence, for 3rd statistical moment *h(t,τ;3)* obtain inequality:

$$h(t,\tau;3) \le h(t,\tau;1)h(t,\tau;2) + \sigma(t,\tau|1)\sigma(t,\tau|2) \quad \text{(B.7)}$$

Let us substitute (B7) into (B.4):

$$Ku(t,\tau)\sigma^4(t,\tau) \ge \sigma^2(t,\tau|2) + \sigma^4(t,\tau|1) + 8\sigma^2(t,\tau|1)h^2(t,\tau;1) + 4h^4(t,\tau;1) - 4h(t,\tau;2)h^2(t,\tau;1) - 4h(t,\tau;1)\sigma(t,\tau|1)\sigma(t,\tau|2)$$

Now use (B.3) and obtain:

$$Ku(t,\tau)\sigma^4(t,\tau) \ge \sigma^2(t,\tau|2) + \sigma^4(t,\tau|1) + 8\sigma^2(t,\tau|1)h^2(t,\tau;1) + 4h^4(t,\tau;1) - 4\sigma^2(t,\tau|1)h^2(t,\tau;1) - 4h^4(t,\tau;1) - 4h(t,\tau;1)\sigma(t,\tau|1)\sigma(t,\tau|2)$$

$$Ku(t,\tau)\sigma^4(t,\tau) \ge \sigma^2(t,\tau|2) + \sigma^4(t,\tau|1) + 4\sigma^2(t,\tau|1)h^2(t,\tau;1) - 4h(t,\tau;1)\sigma(t,\tau|1)\sigma(t,\tau|2)$$

Last inequality takes the form:

$$Ku(t,\tau)\sigma^4(t,\tau) \ge \sigma^4(t,\tau|1) + [\sigma(t,\tau|2) - 4h(t,\tau;1)\sigma(t,\tau|1)]^2 \ge 0 \quad \text{(B.8)}$$

Inequality (B.8) proves that kurtosis *Ku(t,τ)* (B.1) that is determined by market-based statistical moments *h(t,τ;1)* (3.1;3.2), *h(t,τ;2)* (3.4; 3.11), *h(t,τ;3)* (3.12), *h(t,τ;4)* (3.13; 3.3.17) is non-negative, these statistical moments are self-consistent and determine approximation of characteristic function (A.4) and probability.

The dependence of market-based statistical moments *h(t,τ;1)* (3.1;3.2), *h(t,τ;2)* (3.4; 3.11), *h(t,τ;3)* (3.12), *h(t,τ;4)* (3.13; 3.3.17) on statistical moments and correlations of current and past trade values projects the dependence of higher market-based statistical moments. For convenience we reproduce the dependence of *h(t,τ;1)*, *h(t,τ;2)*, *h(t,τ;3)*, *h(t,τ;4)*:

$$h(t,\tau;1) = \frac{C(t;1)}{C_o(t,\tau;1)} \quad ; \quad h(t,\tau;3) = \frac{C(t;3)}{C_o(t,\tau;3)} \tag{B.9}$$

$$h(t,\tau;2) = \frac{C(t;2)+2h^2(t,\tau;1)\Phi^2(t,\tau|1)-2h(t,\tau;1)corr\{C(t),C_o(t,\tau)\}}{C_o(t,\tau;2)} \tag{B.10}$$

$$h(t,\tau;4) = \frac{C(t;4)+2h^2(t,\tau;2)\Phi^2(t,\tau|2)-2h(t,\tau;2)corr\{C^2(t_i),C_0^2(t_i,\tau)\}}{C_o(t,\tau;4)} \tag{B.11}$$

We highlight the similarity of the dependence of *h(t,τ;1)* and *h(t,τ;3)* and of the dependence of *h(t,τ;2)* and *h(t,τ;4)*. We consider that alikeness as ground to the definitions of higher market-based statistical moments. We determine the odd statistical moments (B.12) by odd weight functions $z(t_i,\tau;2k-1)$ (2.8):

$$h(t,\tau;2k-1) = \frac{1}{C_o(t,\tau;2k-1)1}\frac{1}{N}\sum_{i=1}^{N} r^{2k-1}(t_i,\tau)\, C_0^{2k-1}(t_i,\tau) = \frac{C(t;2k-1)}{C_o(t,\tau;2k-1)} \tag{B.12}$$

$$k = 1,2 \dots$$

We determine the even statistical moments *h(t,τ;2k)* by even weight functions $z(t_i,\tau;n)$ (2.8):

$$h(t,\tau;2k) = \sigma^2(t,\tau|k) + h^2(t,\tau;k) \quad ; \quad k = 1,2,\dots \tag{B.13}$$

$$\sigma^2(t,\tau|k) = E_m[(r^k(t_i,\tau) - h(t,\tau;k))^2] = \sum_{i=1}^{N}(r^k(t_i,\tau) - h(t,\tau;k))^2 z(t,\tau;2k) \tag{B.14}$$

The simple transformations similar to (3.13-3.17) give the expression for the market-based volatility $\sigma^2(t,\tau|k)$ (B.15) of the *k*-th degree of returns $r^k(t_i,\tau)$ and for the 2k[th] statistical moment $h(t,\tau;2k)\geq 0$ (B.16) of returns that are non-negative due to (B.13).

$$\sigma^2(t,\tau|k) = \frac{\Omega^2(t|k)+h^2(t,\tau;k)\Phi^2(t,\tau|k)-2h(t,\tau;k)corr\{C^k(t_i),C_0^k(t_i,\tau)\}}{C_o(t,\tau;2k)} \geq 0 \quad ; \quad k = 1,2,3,.. \tag{B.15}$$

$$h(t,\tau;2k) = \frac{C(t;2k)+2h^2(t,\tau;k)\Phi^2(t,\tau|k)-2h(t,\tau;k)corr\{C^k(t_i),C_0^k(t_i,\tau)\}}{C_o(t,\tau;2k)} \geq 0 \tag{B.16}$$

Functions $\Omega(t|k)$ (B.17) and $\Phi(t,\tau|k)$ (B.18) in (B.15; B.16) denote volatility of the *k-th* degree $C^k(t_i)$ of current and past $C_o^k(t_i,\tau)$ market trade values:

$$\Omega^2(t|k) = E[(C^k(t_i) - C(t;k))^2] = C(t;2k) - C^2(t;k) \tag{B.17}$$

$$\Phi^2(t,\tau|k) = E[(C_0^k(t_i,\tau) - C_o(t,\tau;k))^2] = C_o(t,\tau;2k) - C_o^2(t,\tau;k) \tag{B.18}$$

The relations (B.12-B.18) determine the set of market-based statistical moments of returns $h(t,\tau;n)$ $n=1,2,3,\ldots q$, during the averaging interval Δ (1.1). The number N of market trades during Δ limits the number $q<N$ of statistical moments and thus limits the accuracy of the approximations of the characteristic function (A.4-A.6) and the probability of returns. To prove the self-consistency of statistical moments $h(t,\tau;n)$ $n=1,2,3,\ldots q,$ one should prove the non-negativity of all even central statistical moments of returns that are determined by (B.12-B.16). We believe that the use of the relations similar to (B.2; B.3) and (B.5-B.7) would prove by induction the non-negativity of even central statistical moments of returns, but we don't give the general proof here.

References

Amaral, L.I., Plerou, V., Gopikrishnan, P., Meyer, M. and H. E. Stanley, (2000). The Distribution of Returns of Stock Prices, Int.J.Theoretical and Applied Finance, 3(3), 365-369

Andersen, T. and T. Bollerslev, (2006). Realized Return Volatility, Asset Pricing, and Risk Management, NBER Reporter Online, NBER, Cambridge, 7-10

Bachelier, L., (1900). Théorie de la speculation, Annales scientifiques de l'É.N.S. 3e série, **17**, 21-86

Berkowitz, S.A., Dennis, E., Logue, D.E., Noser, E.A. Jr. (1988). The Total Cost of Transactions on the NYSE, *The Journal of Finance*, 43, (1), 97-112

Brown, S.J. (1989). The Number of Factors in Security Returns, J. Finance, 44(5), 1247-1262

Campbell, J. (1985). Stock Returns And The Term Structure, NBER WP1626, 1-53

Campbell, J., Grossman, S.J. and J.Wang, (1993). Trading Volume And Serial Correlation In Stock Returns, Quatr. Jour. Economics, 108 (4), 905-939

Davis, J.L., Fama, E.F. and K. R. French, (2000). Characteristics, Covariances, and Average Returns: 1929 to 1997, J. Finance, 55(1), 389-406

Diebold, F.X. and K. Yilmaz, (2009). Measuring Financial Asset Return And Volatility Spillovers, With Application To Global Equity Markets, The Economic J., 119, 158–171

Dorn, D., Huberman, G. and P. Sengmueller, (2008). Correlated Trading and Returns, J. Finance, 63(2), 885-920

Duffie, D. and P. Dworczak, (2018). Robust Benchmark Design, NBER WP 20540, 1-56

Fama, E.F. (1990). Stock Returns, Expected Returns, and Real Activity, J. Finance, 45(4), 1089-1108

Fama, E.F. and K. R. French, (1992). The Cross-Section of Expected Stock Returns, J.Finance, 47 (2), 427-465

Ferreira, M.A. and P. Santa-Clara, (2008). Forecasting Stock Market Returns: The Sum Of The Parts Is More Than The Whole, WP 14571, NBER, Cambridge, 1-34

Fisher, L. and J. H. Lorie, (1964). Rates Of Return On Investments In Common Stocks, J. Business, 37(1), 1-21

Fox, D.R., et al. (2019). Concepts and Methods of the U.S. National Income and Product Accounts. BEA, Dep. Commerce, US, Chapters 1-13, 1- 449

Greenwood, R. and A. Shleifer, (2013). Expectations Of Returns And Expected Returns, NBER, Cambridge, WP 18686, 1-51

Kelly, B.T., Malamud, S. and K. Zhou, (2022). The Virtue Of Complexity In Return

Prediction, WP30217, NBER, Cambridge, 1-127
Knight, J. and S. Satchell, (Ed). (2001). Return Distributions In Finance, Butterworth-Heinemann, Oxford, 1-328
Kon, S.J. (1984). Models of Stock Returns-A Comparison, J.Finance, 39(1), 147-165
Lettau, M. and S. C. Ludvigson, (2003). Expected Returns And Expected Dividend Growth, WP 9605, NBER, Cambridge, 1-48
Llorente, G., Michaely R., Saar, G. and J. Wang. (2001). Dynamic Volume-Return Relation of Individual Stocks. NBER, WP 8312, Cambridge, MA., 1-55
Lochstoer, L.A. and T. Muir, (2022). Volatility Expectations and Returns, J. Finance, 77 (2), 1055-1096
McAleer, M. and M. C. Medeiros, (2008). Realized Volatility: A Review, Econometric Reviews, 27(1–3), 10–45
Mandelbrot, B., Fisher, A. and L. Calvet, (1997). A Multifractal Model of Asset Returns, Yale University, Cowles Foundation Discussion WP1164, 1-39
Markowitz, H. (1952). Portfolio Selection, J. Finance, 7(1), 77-91
Martin, I. and C. Wagner, (2019). What Is the Expected Return on a Stock?, J. Finance, 74(4), 1887-1929
Olkhov, V. (2021a). Three Remarks On Asset Pricing, SSRN WPS 3852261, 1-20
Olkhov, V. (2021b). Theoretical Economics and the Second-Order Economic Theory. What is it?, MPRA WP120536, 1-13
Olkhov, V. (2022a). Market-Based Asset Price Probability, SSRN WPS 4110345, 1-18
Olkhov, V. (2022b). Market-Based Price Autocorrelation, SSRN WP 4035874, 1-13
Olkhov, V. (2023a). Economic Complexity Limits Accuracy of Price Probability Predictions by Gaussian Distributions, SSRN WPS 4550635, 1-23
Olkhov, V. (2023b). Theoretical Economics as Successive Approximations of Statistical Moments, SSRN WPS 4586945, 1-17
Schlarbaum, G.G., Lewellen, W.G. and R.C. Lease, (1978). Realized Returns on Common Stock Investments: The Experience of Individual Investors, J. of Business, 51(2), 299-325
Shephard, N.G. (1991). From Characteristic Function to Distribution Function: A Simple Framework for the Theory. *Econometric Theory*, 7 (4), 519-529
Shiryaev, A.N. (1999). Essentials Of Stochastic Finance: Facts, Models, Theory. World Sc. Pub., Singapore. 1-852
Shreve, S. E. (2004). Stochastic calculus for finance, Springer finance series, NY, USA
Tsay, R.S. (2005). Analysis of Financial Time Series, J.Wiley&Sons, Inc., New Jersey, 1-638