MPRA
Munich Personal RePEc Archive

The Market-Based Probability of Stock Returns

Olkhov, Victor

Independent

6 February 2023

Online at https://mpra.ub.uni-muenchen.de/116234/
MPRA Paper No. 116234, posted 06 Feb 2023 14:24 UTC

The Market-Based Probability of Stock Returns

Victor Olkhov

Independent, Moscow, Russia

victor.olkhov@gmail.com

ORCID: 0000-0003-0944-5113

ABSTRACT

We show how time-series of random market trade values and volumes completely describe stochasticity of stock returns. We derive equation that links up returns with current and past trade values and show how statistical moments of the trade values and volumes determine statistical moments of stock returns. We estimate statistical moments of the trade values and volumes by the conventional frequency-based probability. However we believe that frequencies of stock returns don't define its probabilities as market and financial concepts. We present the market-based treatment of the probability of stock returns that defines average returns during "trading day" that completely match conventional notion of the weighted value return of the portfolio. We derive how statistical moments of the market trade values and volumes define approximations of the characteristic functions and probability density functions of stock returns. We derive volatility of stock returns, autocorrelations of stock returns, returns-volume and returns-price correlations through corresponding relations between statistical moments of the market trade values and volumes. The market-based probability of stock returns reveals direct dependence of statistical properties of stock returns on market trade randomness and economic uncertainty. Any reasonable forecasting of stock returns should be based on well-grounded predictions of the market trades and economic environment.

Keywords : stock returns, volatility, correlations, probability, market trades

JEL: C1, 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 financial support, valuable offers of research grants or proposals of worthy research/professor positions.

1. Introduction

Forecasts of stock returns are among the most appealing predictions for investors. Modelling and forecasting of the market price and stock returns define the core research issues of financial economics. Irregular behavior of the stock returns during almost any time interval from minutes to months causes usage of probability methods for their assessments. The spectrum of returns research is very diverse and any its review requires a special survey.

This paper is not a review of returns studies and we assume that readers are familiar enough with conventional methods, models and results. However, complexity of economic and financial processes allows consider the familiar problems from a new perspective. In this paper we present a new, market-based look on current studies of stock returns statistics. To explain the meaning of our approach we briefly note conventional methods and models that are currently in use for description of statistical properties of stock returns.

Actually, returns have a wide importance in economics and finance and Solow (1963) indicates, "that the central concept in capital theory should be the rate of return on investment". That Solow's remark arises a question: how the average rate of return should be assessed? We present the market-based look on statistics of stock returns. Forecasts of stock returns and volatility are developed in numerous studies (Ferreira and Santa-Clara, 2008; Diebold and Yilmaz, 2009; Jordà et al., 2019; Bryan et al., 2022). Investors are interested in expected returns and assessments of factors those impact expected returns play central role in studies (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; van Binsbergen and Koijen, 2015; Martin and Wagner, 2019). Irregular evolution of stock prices and returns makes probability theory a major tool for modelling returns. Investigations of probability distributions and correlation laws that may explain returns change are presented by (Kon, 1984; Campbell, Grossman and Wang, 1993; Davis, Fama and French, 2000; Llorente, et al., 2001; Dorn, Huberman and Sengmueller, 2008; Lochstoer and Muir, 2022). Description of expected returns are complemented by research of realized returns and volatility (Schlarbaum, Lewellen and Lease, 1978; Andersen, et al., 2001; Andersen and Bollerslev, 2006; McAleer and Medeiros, 2008; Andersen and Benzoni, 2009). Probability distributions of realized and expected returns time-series are studied by (Amaral, et al., 2000; Knight and Satchell, 2001; Tsay, 2005).

In our paper we study the market-based origin of stock returns probability. The widespread definition of returns probability is the conventional and generally accepted issue. "The probabilistic description of financial prices, pioneered by Bachelier." (Mandelbrot, et al., 1997). Starting with Bachelier (1900) who outlines probabilistic character of the price frequent change, it became routinely to consider frequency of price and returns values as a firm ground for their probabilistic description. Indeed, habitual frequency-based treatment of probability of returns $r(t_i)$ presented by time-series at time t_i, $i=1,..$ during time averaging interval Δ is determined by number m_r of terms $r(t_i)$ or frequency m_r/N that returns $r(t_i)$ take particular value $r(t_i)=r$. If the total number of terms $r(t_i)$ of the time-series during the averaging interval Δ equals N then probability $P(r)$ of returns r is assessed as:

$$P(r) \sim \frac{m_r}{N} \qquad (1.1)$$

Common frequency-based definition of *n-th* statistical moments $E[r^n(t_i)]$ of returns $r(t_i)$ during interval Δ or mathematical expectation $E[r^n(t_i)]$ of the *n-th* degree of returns $r^n(t_i)$ is assessed as:

$$E[\, r^n(t_i)] \sim \frac{1}{N} \sum_{i=1}^{N} r^n(t_i) \qquad (1.2)$$

That *frequency-based* assessment of probability of a random returns $r(t_i)$ presented by finite time-series $r(t_i)$ during the averaging interval Δ serves as ground for almost all probabilistic models of price and returns. That is absolutely correct and completely verified approach based on solid ground of probability theory (Shephard, 1991; Shiryaev, 1999; Shreve, 2004). We note it further as the frequency-based approach to probability of returns. The frequency-based origin of the probability of price and returns is so simple and conventional that it is almost never discussed as a particular and important issue.

However, simple issues often hide complex relations and the conventional frequency-based probability of returns is exactly such a case.

Actually, description of highly irregular time-series of stock returns as a standing alone, independent problem leaves no chance except adopt and use the frequency-based probability (1.1; 1.2). However, returns time-series are not independent. Returns are completely determined by corresponding price time-series and probabilistic properties of price time-series for sure impact random properties of returns. In its turn, randomness of price time-series is completely determined by the market trade stochasticity. Thus, to describe randomness of returns one should take into account stochastic properties of market trades. Distribution of random returns as an issue of financial economics should reflect properties of the random market price and stochasticity of the market trades. Moreover, we feel that there exists a strange collision, a paradox between the frequency-based descriptions of probability of stock returns and the conventional portfolio theory. Indeed, starting at least with

Markowitz (1952), portfolio returns are assessed as weighted by values of the securities those compose the portfolio. Contrary to that generally accepted view, frequency-based assessment (1.1; 1.2) of mean returns does not take into account the values of the underling market trades and that results in significant distinctions with the conventional portfolio theory.

Our pure theoretical paper develops statistical description of stock returns that takes into account and brings together both issues. We present statistical moments of returns in a way that completely coincides with logic and requirements of the portfolio theory and define average returns "as weighted by values of the securities". As well, our description of statistical moments of returns, correlation functions of returns time-series, direct expressions of approximate characteristic functions and probability density functions of stock returns demonstrate direct dependence on the market trade stochasticity and randomness of the market price. As well we use conventional frequency-based probability (1.1; 1.2) to assess *n-th* statistical moments of the market trade values and volumes. We show how frequency-based *n-th* statistical moments of the market trade values and volumes determine statistical moments of returns, correlations of returns, approximations of characteristic function and corresponding approximations of probability density function of returns.

In the next Section we introduce main notations. In Sec. 3 for convenience we briefly present the description of the market-based statistical moments of price. In Sec.4 we introduce main equation that determines dependence of stock returns on the market trade values and derive the value weighted expressions of *n-th* statistical moments of returns that satisfy requirements of the portfolio theory. In Sec 5 we consider the market-based autocorrelations of returns. In Sec 6 we discuss the market-based probability of returns that is determined by set on *n-th* statistical moments. Actually, any reasonable time averaging interval Δ contain only finite number of terms of trade time-series and thus one can assess only a finite number m of statistical moments of stock returns. For the finite number m of statistical moments we present m-approximations of the characteristic functions and corresponding m-approximations of the probability density functions of returns. Sec. 7 – Conclusion. Appendix A presents derivation of correlations of returns in terms of the market-based price statistical moments. In Appendix B we describe return-volume correlations. In Appendix C we show how statistical moments of the market values and volumes determine returns-price relations.

We assume that readers are familiar enough with conventional models of stock returns and have skills in probability theory, usage of statistical moments, characteristic functions and etc. This paper is not for novices and we propose that readers know or can find on their own definitions, notions and terms that are not given in the text.

2. Initial considerations

To study random properties of stock returns researchers investigate the market price time-series to assess returns with particular time shift. Fisher and Lorie (1964) consider "monthly closing prices of all common stocks on the New York Stock Exchange" and "daily high, low, and closing prices of all common stocks" as ground to derive returns time-series. Amaral et al. (2000) consider "trades and quotes (TAQ) database, and analyze 40 million records for 1000 US companies" and study "the probability distribution of returns over varying time scales| from 5 min up to 4 years". Andersen at al. (2001) mention that "five-minute return series are constructed from the logarithmic difference between the prices recorded at or immediately before the corresponding five minute marks". Overall we say that main attention is taken to select initial market price data and chose sample with closing prices, daily high or low prices "immediately before the corresponding five minute marks". Then "the logarithmic difference between the prices" or simple ratio of prices generates the time-series of stock returns. Samples of selected time-series of prices and corresponding samples of returns time-series can be as long as 5 years "for a total of 1,366 trading days" (Andersen et al., 2001) and even "35-year period 1962-96" (Amaral et al., 2000).

The common method of these studies: usage of price time-series to assess samples of returns time-series with selected time shift τ that can be equal 5 min, day, month etc. Duration of the samples of returns time series can be very long up to 5 years or even up to 35 years. Price time-series $p(t_i)$ allow determine returns $r(t_i,\tau)$ time-series in various forms:

$$r(t_i, \tau) = \frac{p(t_i)}{p(t_i-\tau)} \quad ; \quad r_c(t_i, \tau) = r(t_i, \tau) - 1 \quad ; \quad R(t_i, \tau) = \ln p(t_i) - \ln p(t_i - \tau) \quad (2.1)$$

Frequency-based studies (Amaral et al., 2000; Andersen et al., 2001) of returns time-series (2.1) give a lot of important results those uncover or seem to uncover the nature of stocks returns stochasticity. We say, "Seem to uncover" to underline existence of a different look on the random nature of stocks returns.

We believe that "frequency-based" investigations of returns time-series (2.1) alone are not sufficient for understanding and description of the economic nature of stock returns randomness. Investigation of the conventional frequency-based statistics of time-series (2.1) does not display the economic roots and properties of stock returns randomness and does not reveal impact of the market trade stochasticity. Moreover, as we mentioned above, conventional frequency-based treatment of returns time-series (2.1) does not match the common portfolio theory. Indeed, any introduction into the portfolio theory determines returns of the portfolio composed by N securities as weighted by "relative amount X_i invested in security *i*" (Markowitz, 1952). Actually, there is no difference between the assessments of returns of the portfolio composed by *N* securities and the assessments of the average returns as a result of *N* market trades during the averaging interval Δ that we note further as a "trading day". Indeed, one can consider each market trade during "trading day" as a separate "security" of the portfolio. Conventional frequency-based probability of returns considers 1000 trades with same returns *r* as much more probable than one trade with return *R*. However, if these 1000 trades were performed with the trade values $1 each, then they should have much less impact on the mean returns then one trade of $1 billion value with returns *R*. These considerations are absolutely similar to reasons those approve the choice of volume weighted average price (VWAP) (Berkowitz et al., 1988; Duffie and Dworczak, 2018) vs. frequency-based average price. Indeed, one trade of 10 million stocks with price p_1 gives much more impact on average price than 100 trades of 10 stocks each with price p_2. Thus, average returns during the "trading day" should be weighted average by "relative amount of securities values" absolutely in the same way as assessment of the portfolio's returns.

Frequency-based treatment of statistics of the returns time-series in current studies gives average returns according to (2.1) assessment for $n=1$. That contradicts with definition of portfolio returns (Markowitz, 1952). We introduce assessment of returns stochasticity that completely coincides with definition of the portfolio returns (Markowitz, 1952).

Below we show that to derive random properties of return time-series that reflect economic origin and impact of the market trade stochasticity on stock returns, one should complement conventional returns time-series (2.1) by time-series that describe corresponding market trade values and volumes.

In this paper we regard the market trade randomness as the origin of price and returns stochasticity. We show how statistical moments of the market trade values and volumes determine statistical moments of stock returns.

2.1. Main notations

We assume that initial time-series of the realized market transactions describe trade values $C(t_i)$, volumes $U(t_i)$ and prices $p(t_i)$ at time t_i. Time-series t_i determine initial discreteness of the problem and for simplicity we take the constant shift ε between moments t_i:

$$t_i - t_{i-1} = \varepsilon$$

Highly irregular market trade time-series are of little help for modelling asset pricing and returns on long-term horizon T>> ε. To describe statistical properties of returns and to model regular dynamics of mean returns at horizon *T* one should select particular time averaging interval Δ such that T> Δ > ε. We assume that the number N of terms of the market trade values $C(t_i)$ and volumes $U(t_i)$ time-series inside each interval Δ is sufficient to assess their statistical moments using conventional frequency-based probability similar to (1.1; 1.2). The choice of the averaging interval Δ determines time discreteness of the averaged variables. We note averaging interval Δ as a "trading day" or simply a "day" at t and Δ_k denotes the averaging interval or "trading day" that happened k days ago then "today" time t:

$$\Delta_k = \left[t_k - \frac{\Delta}{2}; t_k + \frac{\Delta}{2}\right] \quad ; \quad t_k = t - \Delta \cdot k \quad ; \quad k = 0, 1, 2, \dots \; ; \; \Delta_0 = \Delta \tag{2.2}$$

For convenience we denote as $t_{i,k}$ market trade time-series those belong to interval Δ_k:

$$t_{i,k} \in \Delta_k \quad ; \qquad \Delta = 2n \cdot \varepsilon \; ; \quad i = 1, ..N \tag{2.3}$$

Thus we obtain that

$$t_{i+1,k} - t_{i,k} = \varepsilon \; ; \quad t_{i,k} - t_{i,k+1} = \Delta = 2n \cdot \varepsilon \tag{2.4}$$

We consider moving averaging intervals with time shift less then Δ in Sec.5. We take instantaneous returns $r(t_i,\tau)$ with time shift τ as a simple ratio of price $p(t_i)$ at time t_i during "today" interval Δ to price $p(t_i- \tau)$:

$$r(t_i, \tau) = \frac{p(t_i)}{p(t_i - \tau)} \quad ; \; i = 1, ..N \; ; \tag{2.5}$$

As a time shift τ we consider multiple of ε and take $\tau=\varepsilon l$, $l=1,2,...$ Conventional instantaneous returns $r_c(t_i,\tau)$ take form:

$$r_c(t_i, \tau) = r(t_i, \tau) - 1 = \frac{p(t_i) - p(t_i - \tau)}{p(t_i - \tau)} \tag{2.6}$$

The knowledge of statistical properties of returns $r(t_i,\tau)$ (2.5) completely determine statistical properties of conventional returns $r_c(t_i,\tau)$ (2.6). To describe price and returns during the interval Δ_k that was k "trading days" ago one should replace time t_i in (2.5; 2.6) by $t_{i,k}$ (2.2-2.4). For simplicity we describe returns "today" at t during the interval Δ (2.2; 2.3).

3. The market-based statistical moments of price

Our description of the market-based probability of stock returns is alike to our consideration of the market-based asset price probability. For convenience and simplicity we briefly present below main results of the market-based price probability and refer to (Olkhov, 2021c; 2022a; 2022b) for all further details.

The initial reason for modelling the market-based price probability is simple. The market trade with value $C(t_i)$ and volume $U(t_i)$ at moment t_i determines the market price $p(t_i)$ as:

$$C(t_i) = p(t_i)U(t_i) \tag{3.1}$$

Equation (3.1) states that given statistical distributions of the market trade values $C(t_i)$ and volumes $U(t_i)$ determine statistical properties of the market price $p(t_i)$. One can't consider statistical properties of the price $p(t_i)$ independently of statistical properties of the time-series of market trade values $C(t_i)$ and volumes $U(t_i)$. We take market trade randomness as primary source of the market price stochasticity and describe random price properties as a result of random properties of market trades. We regard time-series of the market trade values $C(t_i)$, volumes $U(t_i)$ and price $p(t_i)$ as random variables during the interval Δ "today" and assume

$$t - \frac{\Delta}{2} \leq t_i \leq t + \frac{\Delta}{2} \quad ; \quad i = 1,..N \tag{3.2}$$

One can equally describe a random variable by its probability density function, characteristic function and by set of *n-th* statistical moments (Shephard, 1991; Shiryaev, 1999; Shreve, 2004). For finite number N of terms of time-series during Δ we denote *n-th* statistical moments of the market trade value $C(t;n)$ and volume $U(t;n)$ using usual frequency-based (1.1;1.2) probability:

$$C(t;n) = E[C^n(t_i)] \sim \frac{1}{N}\sum_{i=1}^{N} C^n(t_i) \; ; \; U(t;n) = E[U^n(t_i)] \sim \frac{1}{N}\sum_{i=1}^{N} U^n(t_i) \; ; n = 1,2,. \tag{3.3}$$

We use $E[...]$ to denote mathematical expectation and ~ to remind that (3.3) is an assessment of *n-th* statistical moments by finite number N of terms of time-series.. That allows determine *n-th* statistical moments of price $p(t;n)$ that differ from frequency-based statistical moments assessed directly from price time-series $p(t_i)$ in a way alike to (1.2; 3.3). Let us take *n-th* degree of equation (3.1):

$$C^n(t_i) = p^n(t_i)U^n(t_i) \tag{3.4}$$

Relations (3.4) between *n-th* degrees of the market trade value $C^n(t_i)$, volume $U^n(t_i)$ and price $p^n(t_i)$ allow introduce (Olkhov, 2021c; 2022a) *n-th* price statistical moments $p(t;n)$ as a generalization of the well-known volume weighted average price (VWAP) (Berkowitz et al., 1988; Duffie and Dworczak, 2018). We define *n-th* statistical moments of price $p(t;n)$ as:

$$p(t;n) = E[p^n(t_i)] \sim \frac{1}{\sum_{i=1}^{N} U^n(t_i)} \sum_{i=1}^{N} p^n(t_i)U^n(t_i) \quad ; \quad n = 1,2, \ldots \tag{3.5}$$

From (3.3; 3.4) one can present relations (3.5) in equal forms (3.6):

$$p(t;n) \sim \frac{\sum_{i=1}^{N} C^n(t_i)}{\sum_{i=1}^{N} U^n(t_i)} = \frac{C(t;n)}{U(t;n)} \quad ; \quad C(t;n) = p(t;n)\, U(t;n) \tag{3.6}$$

It is obvious, that 1-st statistical moment of price $p(t;1)$ coincides with common VWAP. Statistical moments of price $p(t;n)$ completely describe its properties as a random variable during Δ (3.2). However, due to finite number of terms of time-series of the market trade values $C(t_i)$ and volumes $U(t_i)$, relations (3.3) assesses only finite number of *n-th* statistical moments. Thus, relations (3.5; 3.6) determine only finite number of price statistical moments and hence describe approximations of price characteristic function and probability density function only. For further details we refer to (Olkhov, 2021c; 2022a; 2022b). In Sec.6 we consider similar approximations of the characteristic function and probability density function of stock returns.

4. The market-based statistical moments of stock returns

The reasons for derivation of the market-based probability of stock returns are very similar to those found the market-based asset price probability model. Indeed, as we show above, the random properties of the stock price during *Δ* (3.2) are completely determined by the market trade value *C(t;n)* and volume *U(t;n)* statistical moments. Hence, stochastic moments of stock returns $r(t_{i,k},\tau)$ (2.5) should be determined by price statistical moments *p(t;n)* (3.5; 3.6) and thus by the market value *C(t;n)* and volume *U(t;n)* statistical moments. It seems reasonable that stochasticity of market trade should determine random properties of market price and returns. However, almost all studies, for example (Fisher and Lorie, 1964; Amaral, et al., 2000; Andersen et al., 2001) consider time-series of returns in terms of conventional frequency-based statistics. The frequency of returns during the "trading day" plays the primary and the only role in determining the average returns. In other words, returns related with small market trades are considered on equal basis with returns related with big market trades. However, conventional introductions to the portfolio theory at least since Markowitz (1952) define returns of the portfolio composed by *N* securities as returns weighted by "the relative amount invested in security *i" i=1,..N,* Markowitz (1952). Simply put, contribution into the portfolio's returns should be proportional to values of stocks weighted by their returns. Market trades during the "trading day" have numerous returns associated with different market trades values. So, why the assessments (Amaral, et al., 2000; Andersen et al., 2001) of average market trade returns based on frequency of the returns during the "trading day" ignores the values of corresponding market trades?

We introduce definition of the value weighted average returns that completely matches the portfolio theory and in the similar way determine all higher market-based *n-th* statistical moments of stock returns. We start with the same simple market trade equation (3.1; 3.4) at t_i during "today" and a time shift τ:

$$C(t_i) = p(t_i)U(t_i) \tag{4.1}$$

We take definition of returns $r(t_i,\tau)$ (2.5) and transform (4.1) as:

$$C(t_i) = \frac{p(t_i)}{p(t_i-\tau)}\, p(t_i-\tau)U(t_i) \tag{4.2}$$

We denote $C_a(t_i,\tau)$ as market trade value determined by market price $p(t_i\text{-}\tau)$ "adjusted" to trading volume $U(t_i)$ during *Δ* as:

$$C_a(t_i,\tau) \equiv p(t_i-\tau)U(t_i) \tag{4.3}$$

Thus equation (4.2) takes form of (4.4) that determines the returns $r(t_i,\tau)$ (2.5) through the market trade value $C(t_i)$ "today" and "adjusted" market trade value $C_a(t_i,\tau)$ (4.3):

$$C(t_i) = r(t_i,\tau)\, C_a(t_i,\tau) \tag{4.4}$$

Equation (4.4) has a simple interpretation in terms of the conventional portfolio theory. Let us consider each market trade *i*, $i=1,2,..\ N$ during Δ (3.2) as a deal with a particular "security *i*" of the portfolio. Equation (4.4) demonstrates that the "security *i*" of the portfolio has value $C_a(t_i,\tau)$ at time t_i-τ and since time τ at moment t_i due to returns $r(t_i,\tau)$ of that particular "security *i*" its value equals $C(t_i)$. Equation (4.4) generates relations that determine all *n-th* statistical moments of returns in a way that match the portfolio theory.

Similar to (3.4), *n-th* degree of (4.4) results

$$C^n(t_i) = r^n(t_i,\tau)\, C_a^n(t_i,\tau) \quad ; \quad n = 1,2,\ldots \tag{4.5}$$

We introduce *n-th* statistical moments $C_a(t,\tau;n)$ of the "adjusted" trade values $C_a(t_i,\tau)$ similar to (3.3):

$$C_a(t,\tau;n) \equiv E[C_a^n(t_i,\tau)] \sim \frac{1}{N}\sum_{i=1}^{N} C_a^n(t_i,\tau) \tag{4.6}$$

Similar to equations (3.1; 3.4) one can state that equations (4.4; 4.5) prohibit independent description of random properties of the *n-th* degree of the market trade value $C^n(t_i)$, "adjusted" market trade value $C_a(t_i,\tau)$ (4.3) and returns $r^n(t_i,\tau)$ with time shift τ. Given *n-th* statistical moments $C(t;n)$ of the market trade value $C(t_i)$ and *n-th* statistical moments $C_a(t,\tau;n)$ of the "adjusted" market trade value $C_a(t_i,\tau)$ (4.3) completely determine *n-th* statistical moments $r(t,\tau;n)$ of the returns $r(t_i,\tau)$. From (4.5; 4.6) and (3.3) similar to (3.5; 3.6) we introduce *n-th* statistical moments of returns $r(t,\tau;n)$ at day *t* with the time shift τ as:

$$r(t,\tau;n) \equiv E[r^n(t_i,\tau)] \sim \frac{1}{\sum_{i=1}^{N} C_a^n(t_i,\tau)} \sum_{i=1}^{N} r^n(t_i,\tau) C_a^n(t_i,\tau) \tag{4.7}$$

$$r(t,\tau;n) \sim \frac{\sum_{i=1}^{N} r^n(t_i,\tau) C_a^n(t_i,\tau)}{\sum_{i=1}^{N} C_a^n(t_i,\tau)} = \frac{\sum_{i=1}^{N} C^n(t_i)}{\sum_{i=1}^{N} C_a^n(t_i,\tau)} = \frac{C(t;n)}{C_a(t,\tau;n)} \tag{4.8}$$

$$C(t;n) = r(t,\tau;n) C_a(t,\tau;n) \tag{4.9}$$

For $n=1$ relations (4.7- 4.9) describe value weighted average returns $r(t,\tau;1)$ (VaWAR):

$$r(t,\tau;1) \sim \frac{1}{\sum_{i=1}^{N} C_a(t_i,\tau)} \sum_{i=1}^{N} r(t_i,\tau) C_a(t_i,\tau) = \frac{C(t;1)}{C_a(t,\tau;1)} \tag{4.10}$$

VaWAR $r(t,\tau;1)$ (4.10) is alike to well-known expression of VWAP $p(t;1)$ (3.5) for $n=1$ (Berkowitz et al., 1988; Duffie and Dworczak, 2018):

$$p(t;1) = E[p(t_i)] \sim \frac{1}{\sum_{i=1}^{N} U(t_i)} \sum_{i=1}^{N} p(t_i) U(t_i) = \frac{C(t;1)}{U(t;1)} \tag{4.11}$$

We underline that VaWAR $r(t,\tau;1)$ (4.10) almost coincides with definition of the returns of the portfolio (Markowitz, 1952) composed by *i* securities, $i=1,\ldots N$ with values $C_a(t_i,\tau)$ at times t_i-τ and returns $r(t_i,\tau)$ of each "security i". The only distinction – at this stage we don't take into account depreciations of the "securities".

We outline interesting relations between VaWAR $r(t,\tau;1)$ (4.10) and VWAP $p(t;1)$ (4.11). Indeed, from (4.7; 4.8) for VaWAR $r(t,\tau;1)$ (4.10) obtain:

$$r(t,\tau;1)=\frac{C(t;1)}{C_a(t,\tau;1)} \tag{4.12}$$

Let us mention that *n-th* degree of (4.3) gives:

$$C_a^n(t_i,\tau)\equiv p^n(t_i-\tau)U^n(t_i) \tag{4.13}$$

From equation (4.13) and similar to (3.5; 3.6) obtain *n-th* statistical moments $p_a(t,\tau;n)$ of price $p(t_i$-$\tau)$ at time t_i-τ "adjusted" to volumes $U(t_i)$ traded at t_i during Δ "today":

$$p_a(t,\tau;n)\equiv E[p_a^n(t_i,\tau)]\sim\frac{1}{\sum_{i=1}^N U^n(t_i)}\sum_{i=1}^N p^n(t_i-\tau)U^n(t_i) \tag{4.14}$$

From (3.3; 3.4) present (4.14) in equal forms:

$$p_a(t,\tau;n)=\frac{C_a(t,\tau;n)}{U(t;n)}\quad;\quad C_a(t,\tau;n)=p_a(t,\tau;n)U(t;n) \tag{4.15}$$

It is obvious that for all $n=1,.2,..$ equation (4.15) results zero correlations between *n-th* degrees of trade volume and "adjusted" price:

$$corr_{paU}(t,\tau;n|t;n)\equiv E[p^n(t_i,\tau)U^n(t_i)]-E[p^n(t_i,\tau)]E[U^n(t_i)]=0$$

However, similar to the market-based price probability, time-series $U(t_i)$ and $p(t_i$-$\tau)$ are not statistically independent. For example (Olkhov, 2021c; 2022a), one can easy assess correlation between time-series of $p(t_i$-$\tau)$ and $U^2(t_i)$:

$$corr_{paU^2}(t,\tau;1|t;2)=corr_{CaU}(t,\tau;1|t;1)-p_a(t,\tau;1)\sigma_U^2(t)$$

Here volatility $\sigma^2_U(t)$ of trade volumes takes form

$$\sigma_U^2(t)=U(t;2)-U^2(t;1)$$

and:

$$corr_{paU^2}(t,\tau;1|t;2)\equiv E[p(t_i,\tau)U^2(t_i)]-E[p(t_i,\tau)]E[U^2(t_i)]$$

$$E[p(t_i,\tau)U^2(t_i)]=E[C_a(t_i,\tau)U(t_i)]=C_a(t,\tau;1)U(t;1)+corr_{CaU}(t,\tau|t)$$

$$E[C_a(t_i,\tau)U(t_i)]\sim\frac{1}{N}\sum_{i=1}^N C_a(t_i,\tau)U(t_i)$$

For $n=1$ relations (4.14; 4.15) define volume weighted average price $p_a(t,\tau;1)$ (VWAPa) at time t-τ "adjusted" to volumes $U(t_i)$. From (4.11-4.15) obtain for VaWAR $r(t,\tau;1)$ returns:

$$r(t,\tau;1)=\frac{C(t;1)}{C_a(t,\tau;1)}=\frac{p(t;1)}{p_a(t,\tau;1)}\quad;\quad p(t;1)=r(t,\tau;1)\,p_a(t,\tau;1) \tag{4.16}$$

We obtain, that the market-based value weighted average returns $r(t,\tau;1)$ with a time shift τ equal ratio of VWAP $p(t;1)$ (3.6) for $n=1$, today at t to volume weighted average price $p_a(t,\tau;1)$ ($VWAP_a$) (4.14; 4.15) at time t-τ "adjusted" to volumes $U(t_i)$ traded today.

From (4.9; 4.15) obtain similar relation for all *n-th* statistical moments of returns:

$$r(t,\tau;n)=\frac{p(t;n)}{p_a(t,\tau;n)} \quad ; \quad p(t;n)=r(t,\tau;n)\,p_a(t,\tau;n) \tag{4.17}$$

We underline that VWAP $p(t;1)$ is a price $p(t_i)$ "today" weighted by volumes $U(t_i)$ and it differs from VWAPa $p_a(t,\tau;1)$ that is price $p(t_i-\tau)$ at $t_i-\tau$ weighted by volumes $U(t_i)$.

Relations (4.7; 4.8; 4.17) allow derive the market-based volatility $\sigma^2_r(t,\tau)$ of returns. The market-based volatility $\sigma^2_r(t,\tau)$ of returns $r(t_i,\tau)$ is determined by *2-d* statistical moment $r(t,\tau;2)$ of returns (4.17) for $n=2$. From (4.8-4.10) and (3.3; 4.6) obtain:

$$r(t,\tau;2)\equiv E[r^2(t_i,\tau)]=\frac{C(t;2)}{C_a(t,\tau;2)}=\frac{p(t;2)}{p_a(t,\tau;2)} \tag{4.18}$$

Hence, volatility of returns $\sigma^2_r(t,\tau)$ "today" at t with time shift τ takes form:

$$\sigma_r^2(t,\tau)\equiv E[(r(t_i,\tau)-r(t,\tau;1))^2]=r(t,\tau;2)-r^2(t,\tau;1) \tag{4.19}$$

$$\sigma_r^2(t,\tau)=\frac{C(t;2)}{C_a(t,\tau;2)}-\frac{C^2(t;1)}{C_a^2(t,\tau;1)}=\frac{p(t;2)}{p_a(t,\tau;2)}-\frac{p^2(t;1)}{p_a^2(t,\tau;1)} \tag{4.20}$$

Let us take volatility $\sigma^2_C(t)$ of the trade value and volatility of the adjusted trade value $\sigma^2_{Ca}(t,\tau)$ as:

$$\sigma_C^2(t)=C(t;2)-C^2(t;1) \quad ; \sigma_{Ca}^2(t,\tau)=C_a(t,\tau;2)-C_a^2(t,\tau;1) \tag{4.21}$$

then volatility $\sigma^2_r(t,\tau)$ of returns (4.20) takes form:

$$\sigma_r^2(t,\tau)=\frac{\sigma_C^2(t)C_a^2(t,\tau;1)-\sigma_{Ca}^2(t,\tau)C^2(t;1)}{C_a^2(t,\tau;1)C_a(t,\tau;2)} \tag{4.22}$$

The similar relations present volatility of returns through volatilities of prices (App. A.6):

$$\sigma_r^2(t,\tau)=\frac{\sigma_p^2(t)p_a^2(t,\tau;1)-\sigma_{pa}^2(t,\tau)p^2(t;1)}{p_a^2(t,\tau;1)p_a(t,\tau;2)} \tag{4.23}$$

$$\sigma_p^2(t)=p(t;2)-p^2(t;1) \quad ; \sigma_{pa}^2(t,\tau)=p_a(t,\tau;2)-p_a^2(t,\tau;1)$$

Expression (4.20-4.23) ties down volatility $\sigma^2_r(t,\tau)$ of returns with volatilities of trade volumes or volatilities of market-based prices (Olkhov, 2021c; 2022a; 2022b).

5. The market-based autocorrelations of stock returns

Above considerations can be generalized for moving averaging intervals. Below we take moving averaging interval Δ_{k-1} (2.2; 2.3) with time shift εj to the previous time interval Δ_k:

$$t_k-t_{k+1}=\varepsilon j \quad ; \quad t_{i,k}-t_{i,k+1}=\varepsilon j \quad ; \quad j=1,2,.. \tag{5.1}$$

$$t_{i,k}\in\Delta_k=\left[\,t_k-\frac{\Delta}{2},t_k+\frac{\Delta}{2}\right]; \quad i=1,..N; \; k=0,1,.. \tag{5.2}$$

Moving averaging intervals (5.1; 5.2) allow describe autocorrelations of returns with time shift multiple of εj that can be less $\varepsilon j<\Delta$ than the averaging interval Δ:

$$t_{i,k}-t_{i,k+1}=\varepsilon j<\Delta \tag{5.3}$$

Let us show, how moving averaging intervals (5.1-5.3) allow describe the market-based autocorrelations between stock returns. The derivation of correlations of returns follows description of the market-based price correlations (Olkhov, 2022c) and we refer there for details. Let us take returns $r(t_i,\tau)$ (2.5) "today" at t with time shift τ and returns $r(t_{i,2},\tau_2)$ at "day" t_2 ago with time shift τ_2 so, that time shift between t_i and $t_{i,2}$ equals λ:

$$t_i - t_{i,2} = \lambda \;\; ; \;\; \lambda = \varepsilon j \;\; ; \;\; j = 0,1,2,.. \tag{5.4}$$

Time shift λ (5.4) can be less than the time shift τ and even equals zero. Let us describe autocorrelations $corr_r(t,\tau|t_2,\tau_2)$ between returns $r(t_i,\tau)$ and $r(t_{i,2},\tau_2)$

$$corr_r(t,\tau|t_2,\tau_2) = E\big[r(t_i\,,\tau)r\big(t_{i,2}\,,\tau_2\big)\big] - E[r(t_i\,,\tau)]\,E\big[r\big(t_{i,2}\,,\tau_2\big)\big] \tag{5.5}$$

From (4.10; 4.12; 4.16; 4.17) obtain expressions for average returns $r(t,\tau;1)$ and $r(t_2,\tau_2;1)$:

$$r(t,\tau;1) \equiv \; E[r(t_i\,,\tau)] = \frac{C(t;1)}{C_a(t,\tau;1)} = \frac{p(t;1)}{p_a(t,\tau;1)} \tag{5.6}$$

$$r(t_2,\tau_2;1) \equiv \; E[r(t_{i2}\,,\tau_2)] = \frac{C(t_2;1)}{C_a(t_2,\tau_2;1)} = \frac{p(t_2;1)}{p_a(t_2,\tau_2;1)} \tag{5.7}$$

We denote as $C_a(t,\tau;1)$ and $p_a(t,\tau;1)$ (4.3; 4.14; 4.15) – mean trading value and price at "day" t - τ "adjusted" to trading volumes at day t. Respectively, $C_a(t_2,\tau_2;1)$ and $p_a(t_2,\tau_2;1)$ denote mean trading value and price at "day" t_2-τ_2 "adjusted" to trading volumes at day t_2.

To describe autocorrelations of returns $corr_r(t,\tau|t_2,\tau_2)$ one should describe mathematical expectation of their product in (5.5). To do that let us take equation (4.4) that describes returns and repeat it for times t_i with time shift τ and $t_{i,2}$ with time shift τ_2

$$C(t_i) = r(t_i,\tau)\;C_a(t_i,\tau) \quad ; \quad C\big(t_{i,2}\big) = r\big(t_{i,2},\tau_2\big)\;C_a\big(t_{i,2},\tau_2\big) \tag{5.8}$$

The product of equations (5.8) gives equation that allows describe mathematical expectation of product of stock returns:

$$C(t_i)C\big(t_{i,2}\big) = r(t_i,\tau)r\big(t_{i,2},\tau_2\big)\;\;C_a(t_i,\tau)C_a\big(t_{i,2},\tau_2\big) \tag{5.9}$$

We denote mathematical expectations of products of trade values (5.10; 5.11) using conventional frequency-based probability (1.1; 1.2) :

$$C(t;t_2) \equiv E\big[C(t_i)C\big(t_{i,2}\big)\big] \sim \frac{1}{N}\sum_{i=1}^{N} C(t_i)C\big(t_{i,2}\big) \tag{5.10}$$

$$C_a(t,\tau;t_2,\tau_2) \equiv E\big[C_a(t_i,\tau)C_a\big(t_{i,2},\tau_2\big)\big] \sim \frac{1}{N}\sum_{i=1}^{N} C_a(t_i,\tau)C_a\big(t_{i,2},\tau_2\big) \tag{5.11}$$

We define mathematical expectations of product of returns similar to (4.10; 4.11):

$$r(t,\tau;t_2,\tau_2) \equiv E\big[r(t_i,\tau)r\big(t_{i,2},\tau_2\big)\big] \tag{5.12}$$

$$r(t,\tau;t_2,\tau_2) = \frac{1}{\sum_{i=1}^{N} C_a(t_i,\tau)C_a\big(t_{i,2},\tau_2\big)}\sum_{i=1}^{N} r(t_i,\tau)r\big(t_{i,2},\tau_2\big)C_a(t_i,\tau)C_a\big(t_{i,2},\tau_2\big) \tag{5.13}$$

$$r(t,\tau;t_2,\tau_2) = \frac{C(t;t_2)}{C_a(t,\tau;t_2,\tau_2)} \tag{5.14}$$

Now we outline that relations (5.10; 5.11) allow present (5.14) through autocorrelations $corr_C(t|t_2)$ between trading values $C(t_i)$ and $C(t_{i,2})$ and $corr_{Ca}(t,\tau|t_2,\tau_2)$ between trading values $C_a(t_i,\tau)$ and $C_a(t_{i2},\tau_2)$.

$$E\left[C(t_i)C\left(t_{i,2}\right)\right] = E[C(t_i)]\, E\left[C\left(t_{i,2}\right)\right] + corr_C(t|t_2)$$

$$E\left[C_a(t_i,\tau)C_a\left(t_{i,2},\tau_2\right)\right] = E[C_a(t_i,\tau)]\, E\left[C_a\left(t_{i,2},\tau_2\right)\right] + corr_{Ca}(t,\tau|t_2,\tau_2)$$

We underline that autocorrelations $corr_C$ and $corr_{Ca}$ are significantly different. Function $corr_C(t|t_2)$ describes correlation (5.15) between trading values $C(t_i)$ during trading "days" t and values $C(t_{i,2})$ during "day" t_2 and thus is a function of two times t and t_2. The same time function $corr_{Ca}(t,\tau|t_2,\tau_2)$ describes correlation between "adjusted" values $C_a(t_i,\tau)$ and $C_a(t_{i2},\tau_2)$ and hence is a function of all four times $t,\tau;t_2,\tau_2$.

$$C(t;t_2) \equiv C(t;1)C(t_2;1) + corr_C(t|t_2) \tag{5.15}$$

$$C_a(t,\tau;t_2,\tau_2) \equiv C_a(t,\tau;1)C_a(t_2,\tau_2;1) + corr_{Ca}(t,\tau|t_2,\tau_2) \tag{5.16}$$

Substituting (5.15; 5.16) into (5.5) and (5.6; 5.7) and simple transformations give

$$corr_r(t,\tau|t_2,\tau_2) = \frac{corr_C(t|t_2) - r(t,\tau;1)r(t_2,\tau_2;1)corr_{Ca}(t,\tau|t_2,\tau_2)}{C_a(t,\tau;t_2,\tau_2)} \tag{5.17}$$

If $t_2=t$ and $\tau_2=\tau$ than $corr_r(t,\tau|t,\tau)$ (5.17) coincides with volatility $\sigma^2_r(t,\tau)$ of returns (4.19; 4.20) and coincides with (4.22).

$$\sigma_r^2(t,\tau) = \frac{\sigma_C^2(t)C_a^2(t,\tau;1) - \sigma_{Ca}^2(t,\tau)C^2(t;1)}{C_a^2(t,\tau;1)C_a(t,\tau;2)}$$

Correlation $corr_r(t,\tau|t_2,\tau_2)$ (5.17) equals zero if

1. $corr_r(t,\tau|t_2,\tau_2) = 0 \quad if \quad r(t,\tau;1)r(t_2,\tau_2;1) = \frac{corr_C(t|t_2)}{corr_{Ca}(t,\tau|t_2,\tau_2)}$
2. $corr_r(t,\tau|t_2,\tau_2) = 0 \quad if \quad corr_C(t|t_2) = corr_{Ca}(t,\tau|t_2,\tau_2) = 0$

Relations (5.17) uncover dependence of returns correlation $corr_r(t,\tau|t_2,\tau_2)$ on correlations $corr_C(t|t_2)$ (5.15) between trade values $C(t_i)$ and $C(t_{i,2})$ and on correlation $corr_{Ca}(t,\tau|t_2,\tau_2)$ (5.16) between "adjusted" trade values $C_a(t_i,\tau)$ and $C_a(t_{i,2},\tau_2)$. If $t_2=t$ then

$$corr_r(t,\tau|t,\tau_2) = \frac{\sigma_C^2(t) - r(t,\tau;1)r(t,\tau_2;1)corr_{Ca}(t,\tau|t,\tau_2)}{C_a(t,\tau;t,\tau_2)}$$

if $corr_{Ca}(t,\tau|t,\tau_2)=0$ then (see App.A)

$$corr_r(t,\tau|t,\tau_2) = \frac{\sigma_C^2(t)}{C_a(t,\tau;1)C_a(t,\tau_2;1)} = \frac{\sigma_p^2(t)}{p_a(t,\tau;1)p_a(t,\tau_2;1)} \tag{5.18}$$

One can derive relations similar to (5.17) through price correlations (A.6 - App. A). Relations (5.18) demonstrate correlations of returns at t today with different time shifts τ and τ_2. Even in the absence of correlations between "adjusted" trade values, correlations $corr_r(t,\tau|t,\tau_2)$ (5.18) of returns $r(t_i,\tau)$ "today" at t with time shift τ and returns $r(t_i,\tau_2)$ with time shift τ_2, in the main are determined by price volatility $\sigma^2_p(t)$ "today".

The market-based returns-volume autocorrelations $corr_{rU}(t,\tau|t_2)$ between returns "today" at t with time shift τ and trade volumes at "day" t_2 are derived in (B.6; App. B).

$$corr_{rU}(t,\tau|t_2) = \frac{corr_{CU}(t|t_2)}{C_a(t,\tau;1)} = \frac{corr_{CU}(t|t_2)}{p_a(t,\tau;1)U(t;1)}$$

The market-based returns-price autocorrelations $corr_{rp}(t,\tau|t)$ can be expressed through value-volume $corr_{CaU}(t,\tau|t)$, volatility of value $\sigma^2_C(t)$ (see C.12; App.C):

$$corr_{rp}(t,\tau|t) = \frac{\sigma_C^2(t) - r(t,\tau;1)p(t;1)corr_{C_aU}(t,\tau|t)}{C_aU(t,\tau|t)}$$

It is clear that various consequences of the above relations require further investigations, but that is not the subject of the current paper.

6. The market-based probability of returns

In this section we consider the market-based probability of stock returns that results from assessments of the statistical moments (4.7-4.9) of returns. Our derivation is parallel to description of the market-based price probability and we refer (Olkhov, 2021c; 2022a) for further details. As we mentioned above the set of *n-th* statistical moments completely describes properties of a random variable and determines its characteristic function (Shephard, 1991; Shiryaev, 1999; Shreve, 2004). Characteristic function $R(t,\tau;x)$ determined by *n-th* statistical moments $r(t,\tau;n)$ (4.7-4.9) of returns, (3.3; 4.6) or (3.6; 4.14; 4.15) is given by Taylor series as :

$$R(t,\tau;x) = 1 + \sum_{n=1}^{\infty} \frac{i^n}{n!} r(t,\tau;n)\, x^n \tag{6.1}$$

$$r(t,\tau;n) = \frac{C(t;n)}{C_a(t,\tau;n)} = \frac{p(t;n)}{p_a(t,\tau;n)} = \frac{d^n}{(i)^n dx^n} R(t,\tau;x)|_{x=0} \tag{6.2}$$

In (6.1;6.2), we note i as imaginary unit and $i^2=-1$. Relations (6.1; 6.2) completely determine random properties of returns during the averaging interval Δ (2.3; 2.4). However, the finite number of market trades during the averaging interval Δ results that only finite number m of the *n-th* statistical moments of stock returns can assessed. Finite number m of statistical moments of stock returns $r(t,\tau;n)$ (4.7-4.9) determine only m-approximation of the characteristic function $R_m(t,\tau;x)$ of returns and m-approximation of the probability density function of returns as Fourier transform of the characteristic function. Usage of finite Taylor series (6.1) is not convenient to get Fourier transform and we replace m-approximation of the characteristic functions $R_m(t,\tau;x)$ as Taylor series (6.3)

$$R_m(t,\tau;x) = 1 + \sum_{n=1}^{m} \frac{i^n}{n!} r(t,\tau;n)\, x^n \tag{6.3}$$

by integrable exponential characteristic function Q_m:

$$Q_m(t,\tau;x) = \exp\left\{\sum_{n=1}^{m} \frac{i^j}{n!}\, a_n(t,\tau;n)\, x^n - b\, x^{2q}\right\} \quad ; \quad m = 1,2,..; \quad b \geq 0; \quad 2q > m \quad (6.4)$$

Functions $a_n(t,\tau;n)$ can be obtained in recurrent series from requirements (6.2):

$$\frac{d^n}{(i)^n dx^n} Q_m(t,\tau;x)|_{x=0} = r(t,\tau;n) = \frac{C(t;n)}{C_a(t,\tau;n)} = \frac{p(t;n)}{p_a(t,\tau;n)} \quad ; \quad n = 1,..m \quad (6.5)$$

Relations (6.4) guaranties existence of m-approximation of returns probability density function $\mu_m(t,\tau;r)$ as Fourier transform of (6.4):

$$\mu_m(t,\tau;r) = \frac{1}{\sqrt{2\pi}} \int dx\, Q_m(t,\tau;x) \exp(-ixr) \quad (6.6)$$

$$r(t,\tau;n) = \frac{C(t;n)}{C_a(t,\tau;n)} = \frac{p(t;n)}{p_a(t,\tau;n)} = \int dr\ r^n \mu_m(t,\tau;r) \quad ; \quad n \leq m \quad (6.7)$$

For $n=2$ approximation of the returns characteristic function $Q_2(t,\tau;x)$ takes form

$$Q_2(t,\tau;x) = \exp\left\{i\, r(t,\tau;1)x - \frac{\sigma_r^2(t,\tau)}{2} x^2\right\} \quad (6.8)$$

The market-based mean returns $r(t,\tau;1)$ (4.16) and returns volatility $\sigma^2_r(t,\tau)$ (4.20) determine 2-approximation of the characteristic function $Q_2(t,\tau;x)$ of returns (6.8). Corresponding 2-approximation of the returns probability density function $\mu_2(t,\tau;r)$ take usual Gaussian form

$$\mu_2(t,\tau;r) = \frac{1}{(2\pi)^{\frac{1}{2}} \sigma_r(t,\tau)} \exp\left\{-\frac{\left(r - r(t,\tau;1)\right)^2}{2\sigma_r^2(t,\tau)}\right\} \quad (6.9)$$

Simplicity of (6.9) is compensated by requirement to assess second statistical moments (4.16; 4.18; 4.20) of returns. Assessments of higher n-th statistical moments $n=3,4,..$ of the trade values $C(t;n)$ (3.3) and $C_a(t,\tau;n)$ (4.6) and corresponding assessments of statistical moments of returns allow derive higher approximations of the characteristic function $Q_m(t,\tau;x)$ (6.4; 6.5) of returns and approximations of its probability density function $\mu_m(t,\tau;r)$.

7. Conclusion

We present theoretical model that concludes that the randomness of time-series of the market trade values and volumes completely describe the stochasticity of stock returns.

The market-based approach to probability of stock returns is grounded on a simple issue. We presume that 1000 identical returns r related with trade values \$1 should have much less impact on the average returns and on probability of returns than one return R of one \$100 million deal. Since (Markowitz, 1952) definition of the portfolio returns is taken as weighted value of N securities of the portfolio. Thus any market-based probability of the stock returns should support the same definition of the returns, averaged during the interval Δ. However, conventional frequency-based descriptions of the stock returns stochasticity assess 1000 deals with return r as 1000 times more probable then 1 deal with return R, not taking into account the values of corresponding market trades. That seems not too fair, especially for investors and market traders who manage \$50 billion portfolios and make \$100 million deals.

We use conventional frequency-based probability to assess statistical moments of the market trade values and volumes. Finite number terms of time-series of trades during any particular interval Δ results that only finite number m of trade statistical moments can assessed. That results that only m-approximations of the characteristic function and probability density function of stock returns can be assessed. We take simple equation (3.1) which links up trade value, volume and price at time t_i and transform it into equation (4.4) on returns which binds trade value at t_i, returns at t_i with time shift τ and trade value at t_i-τ. Equation (4.4) allows define *n-th* statistical moments of returns similar to definition of the market-based *n-th* statistical moments of asset price (Olkhov, 2021c; 2022a). We underline parallels between values of N securities those compose the portfolio and N market trades during the "trading day" which define average returns. Usage of equation on returns (4.4) and dependence of statistical moments of stock returns on market trade statistical moments allow derive value weighted average returns (VaWAR) that coincide with definition of the portfolio's return (Markowitz, 1952). Further we derive expressions for returns volatility, time autocorrelation of returns, return-volume and return-price correlations as functions of corresponding statistical moments of the market trade values and volumes.

However, our consideration of the market-based probability of stock returns reveals a lot of hidden complexity. Even apparent simplicity of usual Gaussian distribution (6.9) hides tough problems related with volatility of returns $\sigma^2_r(t,\tau)$ (4.20; 4.22; 4.23). Indeed, volatility of returns $\sigma^2_r(t,\tau)$ depends on *2-d* statistical moments of the trade values $C(t;2)$ (3.3) and $C_a(t,\tau;2)$ (4.6). Thus volatility of returns $\sigma^2_r(t,\tau)$ links the "simple" problem of forecasting Gaussian distribution (6.9) of returns with completely new and undiscovered problem of description and prediction of the market and economic variables composed by squares of corresponding variables of economic agents. We call that latent and still invisible problem as the second order economic theory (Olkhov, 2021a; 2021b). Indeed, current economic models attempt describe relations between macroeconomic variables composed by sums of agent's variables, by sums of variables of the 1-st degrees. For example, macroeconomic investments, credits, consumption are composed as sums (without duplicating) of the 1-st degrees of investments, credits, consumption made by economic agents. Description of the *2-d* statistical moments of the trade values $C(t;2)$ (3.3) arises the problem of modelling relations between macroeconomic variables composed as sums of squares of corresponding agent's variables. For example: macroeconomic 2-investments composed as sums of squares of investments made by separate economic agents. On one hand such *2-d* order economic variables open the way for description of volatility of macroeconomic variables - volatility of investments, credits, consumption. On the other hand that arises tough problems of econometric assessments and theoretical description of evolution and mutual interactions of these *2-d* order macroeconomic variables. That is the high price for understanding the covert and secret relations that conduct the market-based volatility of stock returns.

Note one effect else. Forecasts of probability distribution of stock returns at horizon T often serve as hedging tool and ground for Value-at-Risk (VAR) methods that help investors reduce their losses due to market-based returns fluctuations. Our paper demonstrates that the market-based probability of stock returns completely depends on statistical moments of the trade values and volumes. The more statistical moments of the market trades could be predicted explicitly at horizon T – the more accuracy of probability of returns could be achieved. Thus, to get high accuracy of predictions of probability of returns at horizon T one should derive high accuracy forecasting of many statistical moments of trade values and volumes at same horizon T. But that lucky person who would be able to predict exactly statistical moments of the market trade values and volumes at horizon T could benefit of managing alone the future market trades much more, then modelling VAR. However, the secret of long-term predictions of statistical moments of the market trades is hidden probably along with treasuries of Ali Baba's cave.

The market-based probability of stock returns reveals direct dependence of statistical properties of stock returns on the market trade randomness and economic uncertainty. To forecast returns trends or statistical moments one should predict statistical moments of the market trade values and volumes. Further specification and investigation of different consequences of the above relations and econometric assessments of the market-based statistical moments of stock returns require a lot of extra studies. And that are really interesting and though problems for the future.

Appendix A

Correlations of returns depend on correlations of prices

Let us derive how relations similar to (5.17) depend on price correlations (Olkhov, 2022c). To do that let us multiply equation (3.1) at time t_i by same equation at time $t_{i,2}$ and obtain:

$$C(t_i)C(t_{i,2}) = p(t_i)p(t_{i,2})U(t_i)U(t_{i,2}) \tag{A.1}$$

Equation (A.1) is similar to (5.9). Let us determine mathematical expectation of product of trade volumes in (6.1) and trade volume correlations similar to (5.10) and (5.15):

$$U(t;t_2) \equiv E\left[U(t_i)U\left(t_{i,2}\right)\right] \sim \frac{1}{N}\sum_{i=1}^{N} U(t_i)U\left(t_{i,2}\right) \tag{A.2}$$

$$corr_U(t|t_2) \equiv U(t;t_2) - U(t;1)U(t_2;1) \tag{A.3}$$

Average of product of prices in (A.1) and their correlations have similar form:

$$p(t;t_2) \equiv E\left[p(t_i)p\left(t_{i,2}\right)\right] \quad ; \quad corr_p(t|t_2) \equiv p(t;t_2) - p(t;1)p(t_2;1) \tag{A.4}$$

Same considerations that allows derive VWAP (3.5; 3.6), V$_a$WAR (4.10; 4.11) and (5.10-5.14) give definition of (A.4) as:

$$p(t;t_2) = \frac{1}{\sum_{i=1}^{N} U(t_i)U\left(t_{i,2,}\right)}\sum_{i=1}^{N} p(t_i)p\left(t_{i,2}\right)U(t_i)U\left(t_{i,2}\right) = \frac{C(t;t_2)}{U(t;t_2)} \tag{A.5}$$

We define the product of "adjusted" prices with time shifts τ and τ_2 in the same way:

$$C_a(t_i,\tau)C_a\left(t_{i,2},\tau_2\right) = p(t_i-\tau)p\left(t_{i,2}-\tau_2\right)U(t_i)U\left(t_{i,2}\right)$$

From (5.16) and (6.2) obtain:

$$p_a(t,\tau;t_2,\tau_2) = \frac{1}{\sum_{i=1}^{N} U(t_i)U\left(t_{i,2}\right)}\sum_{i=1}^{N} p(t_i-\tau)p\left(t_{i,2}-\tau_2\right)U(t_i)U\left(t_{i,2}\right) = \frac{C_a(t,\tau;t_2,\tau_2)}{U(t;t_2)}$$

$$corr_{pa}(t,\tau|t_2\tau_2) = p_a(t,\tau;t_2\tau_2) - p_a(t,\tau;1)p_a(t_2,\tau_2;1)$$

That allow present (5.14) as

$$r(t,\tau;t_2,\tau_2) = \frac{C(t;t_2)}{C_a(t,\tau;t_2,\tau_2)} = \frac{C(t;t_2)}{U(t;t_2)}\frac{U(t;t_2)}{C_a(t,\tau;t_2,\tau_2)} = \frac{p(t;t_2)}{p_a(t,\tau;t_2,\tau_2)}$$

From (4.17) obtain:

$$corr_r(t,\tau|t_2,\tau_2) = \frac{p(t;t_2)}{p_a(t,\tau;t_2,\tau_2)} - \frac{p(t;1)}{p_a(t,\tau;1)}\frac{p(t_2;1)}{p_a(t_2,\tau_2;1)}$$

$$corr_r(t,\tau|t_2,\tau_2) = \frac{p_a(t,\tau;1)p_a(t_2,\tau_2;1)corr_p(t|t_2) - p(t;1)p(t_2;1)corr_{pa}(t,\tau|t_2,\tau_2)}{p_a(t,\tau;t_2,\tau_2)p_a(t,\tau;1)p_a\left(t_2,\tau_2;1\right)}$$

If t_2=t and τ_2= τ

$$corr_r(t,\tau|t,\tau) = \sigma_r^2(t,\tau) = \frac{p_a^2(t,\tau;1)\sigma_p^2(t) - p^2(t;1)\sigma_{pa}^2(t,\tau)}{p_a(t,\tau;2)p_a^2(t,\tau;1)} \tag{A.6}$$

$$\sigma_r^2(t,\tau) = \frac{\sigma_p^2(t) - r^2(t,\tau;1)\sigma_{pa}^2(t,\tau)}{p_a(t,\tau;2)} \quad ; \quad \frac{\sigma_p^2(t)}{\sigma_{pa}^2(t,\tau)} > r^2(t,\tau;1)$$

Appendix B

Return–volume correlation

We underline that the choice of the averaging procedure primary and substantially determines the return-volume correlation. Campbell, Grossman and Wang (1993) present the frequency-based approach to return-volume correlations. The market-based approach to probability of stock returns presented in our paper gives another look at the same problem and allows present return-volume correlations in a simple form.

To assess returns-volume correlations let us take equation (4.4) at time t_i with time shift τ and multiply it by trade volume $U(t_{i,2})$ at $t_{i,2}$. Then obtain:

$$C(t_i)U(t_{i,2}) = r(t_i,\tau)U(t_{i,2})C_a(t_i,\tau) \quad \text{(B.1)}$$

Similar to above define:

$$CU(t,t_2) \equiv E\left[C(t_i)U(t_{i,2})\right] \sim \frac{1}{N}\sum_{i=1}^{N} C(t_i)U(t_{i,2}) \quad \text{(B.2)}$$

$$rU(t,\tau;t_2) \equiv E\left[r(t_i,\tau)U(t_{i,2})\right] = \frac{1}{\sum_{i=1}^{N} C_a(t_i,\tau)} \sum_{i=1}^{N} r(t_i,\tau)U(t_{i,2})C_a(t_i,\tau) = \frac{CU(t,t_2)}{C_a(t,\tau;1)} \quad \text{(B.3)}$$

$$corr_{rU}(t,\tau|t_2) = E\left[r(t_i,\tau)U(t_{i,2})\right] - E[r(t_i,\tau)]E\left[U(t_{i,2})\right] \quad \text{(B.4)}$$

From (3.3; B.2; B.3; 4.16) obtain

$$corr_{rU}(t,\tau|t_2) = \frac{CU(t,t_2)}{C_a(t,\tau;1)} - \frac{C(t;1)}{C_a(t,\tau;1)}U(t_2;1) \quad \text{(B.5)}$$

From (B.2) and (3.6) obtain:

$$CU(t,t_2) = C(t;1)U(t_2;1) + corr_{CU}(t|t_2)$$

Hence

$$corr_{rU}(t,\tau|t_2) = \frac{corr_{CU}(t|t_2)}{C_a(t,\tau;1)} = \frac{corr_{CU}(t|t_2)}{p_a(t,\tau;1)U(t;1)} \quad \text{(B.6)}$$

Appendix C

Price-return relations

In this Appendix we show, how our market-based method describes mathematical expectations and correlations between *n-th* degrees of returns and *m-th* degrees of prices for *n,m=1,2,..* Let us take equation on *n-th* degree of returns $r^n(t_i,\tau)$ with time-shift τ (4.5) and multiply it by equations on *m-th* degree of price $p^m(t_{i,2})$ (3.4) at $t_{i,2}$

$$C^n(t_i)C^m(t_{i,2}) = r^n(t_i,\tau)p^m(t_{i,2})\, C_a^n(t_i,\tau)U^m(t_{i,2}) \quad (C.1)$$

Similar to (3.3; 3.5; 4.6) we introduce:

$$C(t;n|t_2;m) \equiv E\left[C^n(t_i)C^m(t_{i,2})\right] \sim \frac{1}{N}\sum_{i=1}^{N} C^n(t_i)C^m(t_{i,2}) \quad (C.2)$$

$$C_aU(t,\tau;n|t_2;m) \equiv E\left[C_a^n(t_i,\tau)U^m(t_{i,2})\right] \sim \frac{1}{N}\sum_{i=1}^{N} C_a^n(t_i,\tau)U^m(t_{i,2}) \quad (C.3)$$

$$rp(t,\tau;n|t_2;m) \equiv E\left[r^n(t_i,\tau)p^m(t_{i,2})\right] \quad (C.4)$$

$$rp(t,\tau;n|t_2;m) = \frac{1}{\sum_{i=1}^{N} C_a^n(t_i,\tau)U^m(t_{i,2})}\sum_{i=1}^{N} r^n(t_i,\tau)p^m(t_{i,2})\, C_a^n(t_i,\tau)U^m(t_{i,2}) \quad (C.5)$$

$$rp(t,\tau;n|t_2;m) = \frac{C(t;n|t_2;m)}{C_aU(t,\tau;n|t_2;m)} \quad (C.6)$$

$$C(t;n|t_2;m) = C(t;n)C(t_2;m) + corr_C(t;n|t_2;m) \quad (C.7)$$

$$C_aU(t,\tau;n|t_2;m) = C_a(t,\tau;n)U(t_2;m) + corr_{C_aU}(t,\tau;n|t_2;m) \quad (C.8)$$

$$rp(t,\tau;n|t_2;m) = r(t,\tau;n)p(t_2;m) + corr_{rp}(t,\tau;n|t_2;m) \quad (C.9)$$

From (3.3; 3.6; 4.6; 4.8) and (C.2-C.9) obtain:

$$corr_{rp}(t,\tau;n|t_2;m) = \frac{C(t;n|t_2;m)}{C_aU(t,\tau;n|t_2;m)} - \frac{C(t;n)}{C_a(t,\tau;n)}\frac{C(t_2;m)}{U(t_2;m)} \quad (C.10)$$

$$corr_{rp}(t,\tau;n|t_2;m) = \frac{C(t;n)C(t_2;m)+corr_C(t;n|t_2;m)}{C_a(t,\tau;n)U(t_2;m)+corr_{C_aU}(t,\tau;n|t_2;m)} - \frac{C(t;n)C(t_2;m)}{C_a(t,\tau;n)U(t_2;m)}$$

$$corr_{rp}(t,\tau;n|t_2;m) = \frac{corr_C(t;n|t_2;m) - r(t,\tau;n)p(t_2;m)corr_{C_aU}(t,\tau;n|t_2;m)}{C_aU(t,\tau;n|t_2;m)} \quad (C.11)$$

To simplify notations we omit notations *1* in correlations when degrees *n=m=1*. For $t_2=t$ and *n=m=1* obtain correlations between returns and prices as

$$corr_{rp}(t,\tau|t) = \frac{\sigma_C^2(t) - r(t,\tau;1)p(t;1)corr_{C_aU}(t,\tau|t)}{C_aU(t,\tau;1|t;1)} \quad (C.12)$$

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., Bollerslev, T., Diebold, F.X, Ebens, H. (2001). The Distribution of Realized Stock Return Volatility, Journal of Financial Economics, 61, 43-76

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

Andersen, T. and L. Benzoni, (2009). Realized Volatility, 555-570, in Andersen, T., Davis, R., Kreiß, J-P. and T. Mikosch, Handbook of Financial Time Series, Springer-Verlag Berlin Heidelberg, 1-1031.

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

van Binsbergen, J.H. and R. Koijen, (2015). The Term Structure Of Returns: Facts And Theory, NBER WP 21234, Cambridge, 1-38

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, Cambridge, 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 WP18686, NBER, Cambridge, 1-52

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. (2017). Concepts and Methods of the U.S. National Income and Product Accounts. BEA, US.Dep. Commerce, 1-447

Greenwood, R. and A. Shleifer, (2013). Expectations Of Returns And Expected Returns,

Jordà, O., Knoll, K., Kuvshinov, D., Schularick, M. and A.M.Taylor, (2019). The Rate Of Return On Everything, The Quarterly J. Economics, 134 (3), 1225–1298, https://doi.org/10.1093/qje/qjz012

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

Martin, I. (2021). On the Autocorrelation of the Stock Market, J. Financial Econometrics, 19(1), 39–52

NYSE, (2022). TAQ Products Client Specification, 1-58, https://www.nyse.com/publicdocs/nyse/data/TAQ_Pillar_Products_Client_Spec_v2.3i.pdf

(Olkhov, V., (2021a). Price, Volatility and the Second-Order Economic Theory, ACRN Jour. Finance and Risk Perspectives, 10, 139-165

Olkhov, V., (2021b). Theoretical Economics and the Second-Order Economic Theory. What is it? , MPRA WP110893, 1-14, https://mpra.ub.uni-muenchen.de/110893/

Olkhov, V., (2021c). Three Remarks On Asset Pricing, SSRN WP3852261, 1-24, https://ssrn.com/abstract=3852261

Olkhov, V. (2022a). The Market-Based Asset Price Probability, MPRA WP115382, 1-21, https://mpra.ub.uni-muenchen.de/115382/

Olkhov, V. (2022b). Price and Payoff Autocorrelations in the Consumption-Based Asset Pricing Model, SSRN, WP 4050652, 1- 18

Olkhov, V. (2022c). Introduction of the Market-Based Price Autocorrelation, MPRA, WP112003, 1-13, https://mpra.ub.uni-muenchen.de/112003/

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

Solow, R. (1963). Capital Theory And The Rate Of Return. North-Holland Publ.C., Amsterdam, 1-96

Tsay, R.S. (2005). Analysis of Financial Time Series, J.Wiley&Sons, Inc., New Jersey, 1-638