The Market-Based Probability of Stock Returns

Olkhov, Victor

Independent

6 February 2023

The Market-Based Probability of Stock Returns

Victor Olkhov

Independent, Moscow, Russia

victor.olkhov@gmail.com

ORCID: 0000-0003-0944-5113

Abstract

This paper introduces a new economic, market-based probability of stock return that takes into account the impact of the size of the market's trade values and volumes. We define how the statistical moments of trade values and volumes determine the statistical moments of stock returns. To assess the statistical moments of the trade values and volumes, one should use conventional frequency-based probabilities. The market-based average return takes a form similar to Markowitz's definition of the weighted value return of the portfolio. We derive market-based volatility, autocorrelations of return, return-volume correlations, and return-price correlations as functions of the statistical moments of the trade values and volumes. We derive how a finite number of the statistical moments of the trade values and volumes determine the approximations of the characteristic functions and probability density functions of stock returns. To forecast the average stock return or volatility, one should predict the statistical moments of market trades. Our results are important for the largest investors and banks, economic and financial authorities, and all market participants.

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

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

The forecasts of asset prices and stock returns are among the most appealing. The irregular results of the market's trades generate fluctuating records of stock returns during almost any period. That makes the use of probabilistic methods almost inevitable. In this paper, we present a new, market-based look at the statistical description of stock returns. Our main result states that the market-based probability of stock return is determined by the statistical moments of market trade values and volumes but not by the frequencies of returns. To explain our approach, we briefly note some studies on return statistics.

Numerous papers study stock return (Ferreira and Santa-Clara, 2008; Diebold and Yilmaz, 2009; Jordà et al., 2019; Bryan 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; van Binsbergen and Koijen, 2015; Martin and Wagner, 2019). The irregular behavior of stock prices and returns makes probability theory a major tool for modelling returns. The probability distributions and correlation laws that can match the return change are studied (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). The description of the expected return is complemented by the research on the realized return and volatility (Schlarbaum, Lewellen, and Lease, 1978; Andersen, et al., 2001; Andersen and Bollerslev, 2006; McAleer and Medeiros, 2008; Andersen and Benzoni, 2009). The probability distributions of the realized and expected return time series are studied (Amaral, et al., 2000; Knight and Satchell, 2001; Tsay, 2005).

In our paper, we describe the market-based roots of the probability of stock returns. A widespread definition of probability is the conventional and generally accepted issue. Since Bachelier (1900), who outlined the probabilistic character of the price change, it become routine to consider the frequency of price and return values as the ground for their probabilistic description. Mostly, the stock return $r_s(t_i,\tau)$ (1.1) is taken as the ratio of the stock price $p(t_i)$ at time t_i to the price $p(t_i\text{-}\tau)$ at $t_i\text{-}\tau$ as:

$$r_s(t_i, \tau) = \frac{p(t_i)}{p(t_i-\tau)} - 1 \tag{1.1}$$

One can find the basic definitions in Section 2.1 and further. In our paper, we take the return $r(t_i,\tau)$ (1.2) as a simple price ratio:

$$r(t_i, \tau) = \frac{p(t_i)}{p(t_i-\tau)} \tag{1.2}$$

Obviously, the statistical moments of return $r(t_i,\tau)$ (1.2) completely define the statistical moments of $r_s(t_i,\tau)$ (1.1) (see 4.24). Hence, it is sufficient to describe the statistical moments of return $r(t_i,\tau)$ (1.2). The regular frequency-based probability $P(r)$ (1.3) of return time series $r(t_i,\tau)$ (1.2) at times t_i, $i=1,..N$ during the averaging interval Δ is assessed by the number m_r of terms that take a particular value $r(t_i,\tau)=r$. If the total number of terms of the time series during the averaging interval Δ equals N, then the probability $P(r)$ of return $r(t_i,\tau)=r$ is assessed as:

$$P(r) \sim \frac{m_r}{N} \tag{1.3}$$

The conventional frequency-based mathematical expectations $E[r^n(t_i,\tau)]$ of the *n-th* power of return $r^n(t_i,\tau)$ or the *n-th* statistical moments of return are assessed as:

$$E[\, r^n(t_i,\tau)] \sim \frac{1}{N} \tag{1.4}$$

We use the symbol *"~"* to underline that (1.3; 1.4), and the relations below should be treated as the assessments of the corresponding probabilities or mathematical expectations of the left-hand side by a finite number N of the available terms of market time series during the averaging interval Δ. The *frequency-based* assessments of probability $P(r)$ (1.3) of return $r(t_i,\tau)$ presented by a finite time series $r(t_i,\tau)$ during the averaging interval Δ serve as a ground for almost all probabilistic models. That is the absolutely correct and verified approach based on the solid ground of probability theory (Shephard, 1991; Shiryaev, 1999; Shreve, 2004). We note it further as the frequency-based probability of stock return. The frequency-based origin of the probability of return is so simple and widespread that it is almost never discussed as a special and important issue.

However, simple issues often hide complex relationships and the familiar, regular frequency-based probability of stock return is exactly such a case.

Actually, the description of a highly irregular time series of stock returns as a standing-alone, independent problem leaves no chance except using frequency-based probability (1.3; 1.4). However, the time series of stock return $r(t_i,\tau)$ depends on the market and economic environment. Returns are completely determined by the time series of stock prices, and hence the probabilistic properties of prices certainly impact the random properties of returns. In turn, the stochasticity of market trade completely determines the randomness of stock prices. To describe the random properties of stock returns, one should take into account the stochastic properties of market trades. Moreover, we feel that there exists a strange collision, a paradoxical contradiction between the frequency-based probability (1.3) of stock returns and regular portfolio theory. Indeed, at least since the paper by Markowitz (1952), the return of the portfolio is determined as a weighted average of the values of the securities that compose the portfolio. Contrary to that generally accepted model, the frequency-based assessments (1.3; 1.4) of the average return don't take into account the trade values, which cause distinctions with conventional portfolio theory.

Our pure theoretical paper develops a market-based statistical description of stock return. We show that the market origin of stock returns' stochasticity results in the fact that probabilities of return are not assessed by their frequencies but depend on the size of the values and volumes of the market trades. We derive how the statistical moments of stock return, the approximations of characteristic functions, and probability density functions of return depend on the statistical moments of market trade values and volumes. In turn, we assume that the statistical moments of market trade values and volumes are determined by the regular frequency-based probability (1.3; 1.4).

In the next Section, we introduce the main notations. In Section 3, we briefly describe the market-based statistical moments of price. In Section 4, we introduce the equation that links up the trade values and return and we derive the *n-th* statistical moments of return. In Section 5, we consider the market-based autocorrelations of return. Actually, any reasonable averaging interval Δ contains only a finite number of terms of the market trade time series, and thus, one can assess only a finite number m of statistical moments of stock return. In Sec. 6, we show how a finite number m of the statistical moments of stock return defines the m-approximations of characteristic functions and probability density functions of return. Section 7 - Conclusion. In Appendix A, we describe the dependence of correlations of return on market-based price statistical moments. In Appendix B, we describe return-volume correlations. In Appendix C, we show how the statistical moments of trade 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, statistical moments, characteristic functions, etc. We propose that readers know or can find on their own the definitions, notions, and terms that are not given in the text.

2. Initial considerations

The market price time series serves as the origin for the assessments of return. 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 the time series of return. Amaral et al. (2000) consider the "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 et 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, the researchers selected the initial market price data and chose the samples with closing prices, daily highs or lows prices "immediately before the corresponding five-minute marks." Then "the logarithmic difference between the prices" or a simple ratio of prices generates the time series of stock return. Samples of the selected time series of prices and the corresponding samples of return time series can be as long as 5 years "for a total of 1,366 trading days" (Andersen et al., 2001) and even a "35-year period from 1962-96" (Amaral et al., 2000). All referred authors investigate the time series of stock returns using conventional frequency-based probability (1.3; 1.4). Frequency-based studies (Amaral et al., 2000; Andersen et al., 2001) of the return time series (2.1) give important results that uncover, or seem to uncover, the nature of stock returns' stochasticity. We say "seem to uncover" to mention the existence of a different look at the random nature of return.

We believe that the frequency-based analysis of return's (1.2) time series is not sufficient for understanding the market nature of return's stochasticity. Conventional frequency-based statistics of the time series of return (1.2) do not display the economic roots of the stock return's randomness and do not reveal the impact of market stochasticity. Moreover, conventional frequency-based analysis of return time series (2.1) does not match portfolio theory. Indeed, the return of the portfolio, which is composed of N securities, is determined as a weighted average of the "relative amount X_i invested in security i" (Markowitz, 1952). Actually, there is almost no difference between the assessment of the return of the portfolio composed of N securities and the assessment of the average return of N trades with a selected stock over the period Δ. Indeed, one can consider each market trade during Δ as a separate "security" of the portfolio. Conventional frequency-based probability of return assumes that 1000 trades with the same return r are much more probable, than one trade with a return R. However, if these 1000 trades were made with a trade value of $1 each, then such trades should have much less impact on the mean return during the "trading day" Δ than a singe trade of $1 billion in value with a return of R. These considerations are absolutely similar to the reasons given for approving the volume-weighted average price (VWAP) (Berkowitz et al., 1988; Duffie and Dworczak, 2018) vs. the frequency-based average price. Indeed, a single trade of 100 million stocks at a price p_1 has much more impact at the average price than 100 trades of 1 stock each at a price p_2. Thus, the average return during the “trading day” Δ should be weighted average by the “relative amount of securities values” absolutely in the same way as the return of the portfolio. We define the market-based average return as the one that matches the weighed value return of the portfolio (Markowitz, 1952).

To describe the random properties of return that reflect the impact of market stochasticity, one should the complement conventional time series of return (2.1) with the time series of the corresponding market trade values and volumes.

2.1. The main notations

Assume that time series describe market trade value $C(t_i)$, volume $U(t_i)$, and price $p(t_i)$ at a time t_i. The time-series at time t_i determines the initial discreteness of the problem. For simplicity, we take the shift ε between moments t_i to be constant.

$$t_i - t_{i-1} = \varepsilon \qquad (2.1)$$

The highly irregular market trade time series are of little help for the model price and return on the long-term horizon T>> ε. To describe the statistical properties of return and model the regular dynamics of the mean return or volatility at horizon T, one should select a particular time averaging interval Δ, such that T> Δ >> ε. We assume that the number N of the terms of the time series of the trade values $C(t_i)$ and volumes $U(t_i)$ inside each interval Δ is a constant and is sufficient to assess the statistical moments using the regular frequency-based probability similar to (1.3; 1.4). We note the interval Δ as a “trading day” at t, and Δ_k denotes the averaging interval that happened k days in the past:

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

For convenience, we denote as $t_{i,k}$ the time series that belong to the interval Δ_k:

$$t_{i,k} \in \Delta_k \quad ; \qquad i = 1, .. N \qquad (2.3)$$

Thus, due to (2.2) and (2.3) we take that

$$t_{i+1,k} - t_{i,k} = \varepsilon \ ; \quad t_{i,k} - t_{i,k+1} = \Delta \qquad (2.4)$$

We consider the moving averaging intervals with the time shift less than Δ in Section 5. We take the instantaneous return $r(t_i,\tau)$ with the time shift τ as a simple ratio of the price $p(t_i)$ at time $t_i = t_{i,0}$, $t_{i,0} \in \Delta_0 = \Delta$ today to the price $p(t_i - \tau)$. We assume that all prices $p(t_{i,k})$ at time $t_{i,k}$ in all intervals Δ_k (2.3) are adjusted to present prices at time t.

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

We assume that the time shift τ is a multiple of ε (2.1) and take $\tau = \varepsilon l$, $l = 1, 2, \ldots$ The regular instantaneous stock return $r_s(t_i,\tau)$ (1.1) takes the form:

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

The statistical moments of return $r(t_i,\tau)$ completely determine the statistical moments of the conventional stock returns $r_s(t_i,\tau)$ (see 4.24). To describe the price and return over the interval Δ_k that was k "trading days" in the past one should replace time t_i in (2.5; 2.6) by $t_{i,k}$ (2.2-2.4). For simplicity, below we describe return during Δ (2.2; 2.3).

3. The market-based statistical moments of price

The description of the market-based probability of stock returns is similar to the description of the market-based asset price probability. We briefly present the main results of the market-based price probability and refer to Olkhov (2021; 2022a; 2022b) for further details. The trade value $C(t_i)$ and volume $U(t_i)$ at time t_i determine the trade price $p(t_i)$ as follows:

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

The simple equation (3.1) states that the given statistical distributions of the market trade values $C(t_i)$ and volumes $U(t_i)$ determine the statistical properties of the market price $p(t_i)$. One cannot consider the statistical properties of the price $p(t_i)$ independently of the statistical properties of the market trade values $C(t_i)$ and volumes $U(t_i)$. We take the market's trade randomness as the origin of the price stochasticity and show how trade statistical moments determine the statistical moments of the price. We take the time series of the trade values $C(t_i)$, volumes $U(t_i)$ and price $p(t_i)$ as random variables during Δ and assume:

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

One can equally describe a random variable by its probability density function, characteristic function, and the set of the *n-th* statistical moments (Shephard, 1991; Shiryaev, 1999; Shreve, 2004). For a finite number N of terms of the time series during Δ, we assess the *n-th* statistical moments for $n=1,2,...$ of the market trade value $C(t;n)$ and volume $U(t;n)$ using the regular frequency-based probability (1.3;1.4):

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

We denote mathematical expectation as $E[...]$ and indicate that the symbol "~" denotes an assessment, an estimator of the *n-th* statistical moments. To determine the *n-th* statistical moments of price $p(t;n)$, let us take the *n-th* power of equation (3.1):

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

The relations (3.4) between the *n-th* power of the market trade value $C^n(t_i)$, volume $U^n(t_i)$, and price $p^n(t_i)$ allow defining the *n-th* statistical moments $p(t;n)$ of price (Olkhov, 2021; 2022a) alike to the well-known volume weighted average price (VWAP) (Berkowitz et al., 1988; Duffie and Dworczak, 2018):

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

From (3.3; 3.4) one can present the relations (3.5) as (3.6):

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

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

4. The market-based statistical moments of stock return

The derivation of the market-based probability of stock return is similar to the derivation of the market-based asset price probability. We start with the market equations (3.1; 3.4):

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

We take the return $r(t_i,\tau)$ (2.5) with a time shift τ and transform (4.1) as follows:

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

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

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

The equation (4.2) takes the form of (4.4), which determines the stock return $r(t_i,\tau)$ (2.5) through the market trade value $C(t_i)$ and the "adjusted" trade value $C_a(t_i,\tau)$ (4.3):

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

The equation (4.4) has a simple interpretation in the terms of conventional portfolio theory. For *i*, *i=1,2,.. N*, one can consider (4.4), as the relations between the value $C_a(t_i,\tau)$ which due to the return $r(t_i,\tau)$ takes the value $C(t_i)$. Similar to (3.4), the *n-th* of (4.4) gives:

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

We determine the *n-th* statistical moments $C_a(t,\tau;n)$ of the "adjusted" values $C_a(t_i,\tau)$ as (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 the equations (3.1; 3.4), one can state that the equations (4.4; 4.5) prohibit the independent description of the random properties of the *n-th* power of the trade value $C^n(t_i)$, "adjusted" value $C_a(t_i,\tau)$ (4.3), and return $r^n(t_i,\tau)$. The given *n-th* statistical moments $C(t;n)$ of the trade value $C(t_i)$ and the statistical moments $C_a(t,\tau;n)$ (4.6) of the "adjusted" value $C_a(t_i,\tau)$ (4.3) determine the *n-th* statistical moments $r(t,\tau;n)$ of return. From (4.5; 4.6) and (3.3), similar to (3.5; 3.6), we determine the *n-th* statistical moments of return $r(t,\tau;n)$:

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

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

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

The dependence of the statistical moments (4.7-4.9) of return on the statistical moments of the trade values $C(t;n)$, $C_a(t,\tau;n)$ (3.3; 4.6; 4.14; 4.15), and volumes $U(t;n)$ (3.3; 4.14; 4.15), underlines the impact of the size of the value and volume of market trades on the market-based probability of stock returns. Accounting for this dependency is necessary for the largest investors, traders, and banks - all those, who perform the major market transactions. Taking into account the dependence (4.7-4.9) of the statistical moments of return on the statistical moments of market trades is important for those who assess and manage macroeconomic, financial and market trends: the Central banks, economic, and financial authorities.

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

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

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

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

As we mentioned in Section 2, we consider all stock prices at any time adjusted to the present. Thus, VaWAR $r(t,\tau;1)$ (4.10) coincides with the definition of the return of the portfolio (Markowitz, 1952) composed of N securities with the values $C_a(t_i,\tau)$ at the times t_i-τ and returns $r(t_i,\tau)$ of each "security *i*".

We outline the 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 the *n-th* power of (4.3) gives:

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

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

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

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

For all $n=1,.2,..$ (4.14; 4.15), result in zero correlations between the *n-th* powers of the trade volume and the "adjusted" price:

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

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

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

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

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

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

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

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

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

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

We obtain, that the market-based VaWAR $r(t,\tau;1)$ (4.16) equals the ratio of VWAP $p(t;1)$ (3.6) to the volume weighted average price $p_a(t,\tau;1)$ ($VWAP_a$) (4.14; 4.15). From (4.9; 4.15), obtain similar relation for all *n-th* statistical moments of return:

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

From (4.17), obtain the *2-d* statistical moment $r(t,\tau;2)$ of return (4.18):

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

The volatility of return $\sigma^2_r(t,\tau)$ at t with time shift τ takes the 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 the volatility $\sigma^2_C(t)$ of the trade value and of the "adjusted" trade value $\sigma^2_{Ca}(t,\tau)$:

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

Then the volatility $\sigma^2_r(t,\tau)$ of return (4.20) takes the 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 describe the volatility of return (App. A.6) via the price volatilities:

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

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

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

The statistical moments $r_s(t,\tau;n)$ of return $r_s(t_i,\tau)$ (1.1) are determined by $r(t,\tau;n)$ (4.7-4.9):

$$r_s(t,\tau;n) \equiv E[r_s^n(t_i,\tau)] = E[(r(t_i,\tau)-1)^n] \sim \sum_{k=0}^{n} \frac{(-1)^{n-k}\; n!}{k!(n-k)!} r(t,\tau;n) \tag{4.24}$$

$$r(t,\tau;0) \equiv E[r^0(t_i,\tau)] \equiv 1$$

The equations (3.6) and (4.9) establish the relations (4.25; 4.26) between the statistical moments of price $p(t;n)$ and stock returns $r(t,\tau;n)$:

$$p(t;n)\,U(t;n) = r(t,\tau;n)C_a(t,\tau;n) \tag{4.25}$$

$$p(t;n) = r(t,\tau;n)\,\frac{C_a(t,\tau;n)}{U(t;n)} \tag{4.26}$$

The relations (4.25; 4,26) outline the equal predictability of the market-based statistical moments of price and stock return.

5. The market-based autocorrelations of stock return

The above relations can be generalized to the case of moving averaging intervals. We take the interval Δ_{k-1} (2.2; 2.3) with the time shift εj to the previous interval Δ_k:

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

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

Moving averaging intervals (5.1; 5.2) allow to describe the autocorrelations of returns with the time shift multiple of εj that can be less than the interval Δ:

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

Let us take return $r(t_i,\tau)$ (2.5) with the time shift τ and return $r(t_{i,2},\tau_2)$ in the past with the time shift τ_2 so that the time shift between t_i and $t_{i,2}$ equals to λ:

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

The time shift λ (5.4) can be less than the time shift τ and even equal zero. The autocorrelations $corr_r(t,\tau|t_2,\tau_2)$ between returns $r(t_i,\tau)$ and $r(t_{i,2},\tau_2)$ take the form:

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

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

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

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

$C_a(t,\tau;1)$ and $p_a(t,\tau;1)$ (4.3; 4.14; 4.15) denote the "adjusted" average value and price at t with the shift τ. Respectively, $C_a(t_2,\tau_2;1)$ and $p_a(t_2,\tau_2;1)$ denote the "adjusted" average value and price at t_2 with the shift τ_2. To describe the autocorrelations $corr_r(t,\tau|t_2,\tau_2)$ one should assess mathematical expectation of their product in (5.5). Let us take the equation (4.4) at time t_i with the shift τ and at time $t_{i,2}$ with the time shift τ_2:

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

The product of equations (5.8) gives the equation (5.9):

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

We denote mathematical expectations of products of the trade values (5.10; 5.11) using the regular frequency-based probability (1.3; 1.4):

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

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

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

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

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

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

We highlight that (5.10; 5.11) allows present (5.14) through the autocorrelations $corr_C(t|t_2)$ between $C(t_i)$ and $C(t_{i,2})$ and $corr_{Ca}(t,\tau|t_2,\tau_2)$ between $C_a(t_i,\tau)$ and $C_a(t_{i2},\tau_2)$.

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

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

The function $corr_C(t|t_2)$ describes the correlation (5.15) between the trading values $C(t_i)$ during "day" t and the values $C(t_{i,2})$ during "day" t_2 and depends on two times t and t_2. The function $corr_{Ca}(t,\tau|t_2,\tau_2)$ describes the correlation between the "adjusted" values $C_a(t_i,\tau)$ and $C_a(t_{i2},\tau_2)$ and depends on four times t, τ, t_2, τ_2.

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

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

Substitute (5.15; 5.16) into (5.5-5.7), and, after simple transformations, obtain:

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

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

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

The relations (5.17) display the dependence of the correlation $corr_r(t,\tau|t_2,\tau_2)$ on the correlation $corr_C(t|t_2)$ (5.15) between the values $C(t_i)$ and $C(t_{i,2})$ and on the correlation $corr_{Ca}(t,\tau|t_2,\tau_2)$ (5.16) between the "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 the relations similar to (5.17) through the price correlations (A.6 - App. A). The relations (5.18) demonstrate the correlation of returns at t today with a different time shifts τ and τ_2. Even in the absence of the correlation between the "adjusted" trade values, the correlation $corr_r(t,\tau|t,\tau_2)$ (5.18) of return $r(t_i,\tau)$ at t with a time shift τ and return $r(t_i,\tau_2)$ with a time shift τ_2, in the main are determined by the price volatility $\sigma^2_p(t)$ at t. The market-based return-volume correlation $corr_{rU}(t,\tau|t_2)$ between the return at t with time shift τ and the trade volume at 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 return-price correlation $corr_{rp}(t,\tau|t)$ can be expressed through the value-volume correlation $corr_{CaU}(t,\tau|t)$ and the volatility of value $\sigma^2_C(t)$ (see C.12; App.C):

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

6. The market-based probability of stock return

In this section, we consider the market-based probability of stock returns determined by the statistical moments (4.7-4.9). Our derivation is parallel to the description of market-based price probability (Olkhov, 2021; 2022a). The set of the *n-th* statistical moments completely describes the properties of a random variable (Shephard, 1991; Shiryaev, 1999; Shreve, 2004) and determines its characteristic function $R(t,\tau;x)$ as Taylor series:

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

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

In (6.1; 6.2), we take i as an imaginary unit and $i^2=-1$. The relations (6.1; 6.2) determine the random properties of return during Δ (2.3; 2.4). However, the finite number of market trades during Δ results in only a finite number m of the n-th statistical moments of stock return can be assessed. Different random variables can have the same first m statistical moments. The given m of statistical moments of stock return $r(t,\tau;n)$ determine only the m-approximations of the characteristic function $R_m(t,\tau;x)$:

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

The finite Taylor series (6.3) is not too convenient to calculate the Fourier transform to get the m-approximation of probability density function of return. We replace (6.3) by the integrable characteristic function Q_m with the same first m statistical moments:

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

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

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

Relations (6.4) guarantee the existence of the Fourier transform (6.6) that defines the m-approximation of the probability density function $\mu_m(t,\tau;r)$ of return:

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

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

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

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

The market-based average return $r(t,\tau;1)$ (4.16) and return volatility $\sigma^2_r(t,\tau)$ (4.20) determine the 2-approximation of the characteristic function $Q_2(t,\tau;x)$ (6.8). Correspondingly, the Gaussian approximation of the return probability $\mu_2(t,\tau;r)$ takes the known form:

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

The simplicity of (6.9) is compensated by the requirement to assess the second statistical moments (4.16; 4.18; 4.20) of return. The assessments of the 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 the assessments of the statistical moments of return allow to derive the higher approximations of the characteristic functions $Q_m(t,\tau;x)$ (6.4; 6.5) and probability density functions $\mu_m(t,\tau;r)$ of return.

7. Conclusion

Our paper gives pure theoretical considerations showing that the randomness of market trade values and volumes completely describes the stochasticity of stock returns. We show that the statistical moments of the trade values and volumes determine the statistical moments of return. That uncovers the important dependence of the market-based probability of stock returns on the size of the values and volumes of the market deals. The largest investors, trades, and banks – who manage substantial portfolios and perform major market transactions - should take our results into account. The use of the market-based probability of stock returns is important for the correct assessments and forecasts of the market, financial, and macroeconomic trends by the Central banks, economic, and financial authorities.

The forecasts of the probability of stock returns at horizon *T* often serve as a ground for Value-at-Risk (VAR) methods that help hedge funds secure their portfolios from the risks of unexpected losses due to market fluctuations of return. This paper shows that the more statistical moments of the market trades could be explicitly predicted at the horizon *T*, the more accuracy of the probability of returns could be achieved. But that lucky person who predicts exactly many statistical moments of the trade values and volumes at the horizon *T* would benefit much more from managing the future markets alone than from modelling VAR. The financial institutions that use VaR in their practices should take our results into account.

Appendix A

Correlation of returns depend on correlation of prices

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

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

The equation (A.1) is similar to (5.9). We determine mathematical expectation of product of the trade volumes in (6.1) and the trade volume correlation similar to (5.10) and (5.15):

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

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

The average of the product of prices in (A.1) and their correlation have similar form:

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

The same considerations that allow derive VWAP (3.5; 3.6), V_aWAR (4.10; 4.11), and (5.10-5.14) give the 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 the “adjusted” prices with the 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 correlation $corr_{pa}(t,\tau|t_2,\tau_2)$ of “adjusted” prices:

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

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

That allows 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 correlation of return:

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

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

If $t_2=t$ and $\tau_2=\tau$

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

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

Appendix B

The return–volume correlation

The choice of the averaging procedure is primary, and it substantially determines the return-volume correlation. Campbell, Grossman, and Wang (1993) present a frequency-based approach to the return-volume correlation. The market-based approach to the probability of stock return gives another look at the same problem and allows the return-volume correlation to be presented in a simple form. To assess the return-volume correlation, take the equation (4.4) at a time t_i with the time shift τ and multiply it by the trade volume $U(t_{i,2})$ at $t_{i,2}$:

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

Similar to the above, define:

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

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

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

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

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

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

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

Hence:

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

Appendix C

The price-return relations

In this Appendix we show, how the market-based approach describes mathematical expectations and correlation between the *n-th* powers of return and the *m-th* powers of price time series for *n,m=1,2,..* Take the equation on the *n-th* power of return $r^n(t_i,\tau)$ with the time shift τ (4.5) and multiply it by the equations on the *m-th* power of price $p^m(t_{i,2})$ (3.4) at $t_{i,2}$

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

Similar to (3.3; 3.5; 4.6), we define:

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

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

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

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

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

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

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

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

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

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

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

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

To simplify the notations, we omit the notation of *1* in correlation when the powers $n=m=1$. For $t_2=t$ and $n=m=1$ obtain the correlation between return and price:

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

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, 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. (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, NBER, Cambridge, WP 18686, 1-51

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. (2021). Three Remarks On Asset Pricing, SSRN WP3852261, 1-24

Olkhov, V. (2022a). The Market-Based Asset Price Probability, MPRA WP115382, 1-21

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

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