The Market-Based Probability of Stock Returns

Olkhov, Victor

Independent

25 November 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 describes the probability of stock returns through a description of the set of their market-based *n-th* statistical moments that depended on the *n-th* statistical moments of market trade values and volumes. We derive these relations as extensions of Markowitz's definition of a value weighted return of the portfolio. The average return of sale trades during the averaging period coincides with Markowitz's definition of portfolio return. We highlight the similarity between Markowitz's definition of portfolio return and the definition of average price as a volume weighted average price. We derive market-based volatility, autocorrelations of return, return-volume correlations, and return-price correlations as functions of the *n-th* statistical moments of the trade values and volumes. We derive how a finite number of the *n-th* 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, 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 irregular results of market trades cause fluctuating stock returns during almost any period. That makes the use of probabilistic methods almost inevitable. The correct choice of return probability is valuable for all investors. The conventional approach considers return probability to be proportional to the frequency of the return's events. However, top funds, banks, and large investors should take into account the impact of the deal's size statistics and consider the probability of return as a result of market trade randomness. We describe the dependence of statistical moments of market trade values and volumes on statistical moments of return. That highlights the direct ties between the randomness of market trade and stock returns.

Studies of stock returns are endless (Ferreira and Santa-Clara, 2008; Diebold and Yilmaz, 2009; Jordà et al., 2019; Kelly et al., 2022). The assessments of the factors that impact the expected return play a central role (Fisher and Lorie, 1964; Mandelbrot, Fisher, and Calvet, 1997; Campbell, 1985; Brown, 1989; Fama, 1990; Fama and French, 1992; Lettau and Ludvigson, 2003; Greenwood and Shleifer, 2013; van Binsbergen and Koijen, 2015; Martin and Wagner, 2019). The irregular behavior of stock prices and returns makes probability theory a major tool for modeling returns. The probability distributions and correlation laws that can match the return change are studied by Kon (1984) and (Campbell, Grossman, and Wang, 1993; Davis, Fama, and French, 2000; Llorente et al., 2001; Dorn, Huberman, and Sengmueller, 2008; Lochstoer and Muir, 2022). The description of the expected return is complemented by 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 are studied by Amaral et al. (2000), Knight and Satchell (2001), and Tsay (2005). That is only a tiny fraction of numerous studies of stock returns.

Since Bachelier (1900), who outlined the probabilistic character of the price change, it has become routine to consider the frequency of price and return values as the basis for their probabilistic description. In our paper, we take the return $r(t_i,\tau)$ (1.2) as a simple price ratio of price $p(t_i)$ at time t_i to price $p(t_i\text{-}\tau)$ in the past with time shift τ :

$$r(t_i,\tau) = \frac{p(t_i)}{p(t_i-\tau)} \quad (1.1)$$

The regular frequency-based probability $P(r)$ (1.2) of return time series $r(t_i,\tau)$ (1.1) 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:

$$P(r) \sim \frac{m_r}{N} \tag{1.2}$$

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:

$$E[\, r^n(t_i,\tau)] \sim \frac{1}{N} \sum_{i=1}^{N} r^n(t_i,\tau) \tag{1.3}$$

We use the symbol *"~"* to underline that (1.2; 1.3) should be treated as the assessments of the corresponding probabilities or mathematical expectations by a finite number N of the available terms of market time series during the averaging interval Δ. The frequency-based assessments of probability $P(r)$ (1.2) of return $r(t_i,\tau)$ presented by a finite time series $r(t_i,\tau)$ during the averaging interval Δ serve as a basis for almost all probabilistic models. That is the 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.

One should note that any particular sample of N terms of the irregular time series itself doesn't determine the averaging procedure or the correct probability distribution. The same series can be the result of random sampling of different random variables. The choice of an adequate averaging procedure that highlights the economic sense of the problem is the challenge for the researcher.

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 (1.2; 1.3). However, the time series of stock returns $r(t_i,\tau)$ are completely determined by the time series of stock prices, and hence the random properties of prices 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. Indeed, at least since Markowitz (1952), the return of the portfolio is determined as the value weighted of the securities that compose the portfolio. That is almost similar to the well-known definition of average price as the volume weighted average price (VWAP) (Berkowitz et al., 1988; Duffie and Dworczak, 2018). In this paper, we show that Markowitz' definition of a portfolio's return can be used as the definition of the average return of sale trades during the averaging period, and that is almost similar to the definition of average price as VWAP. We extend Markowitz's definition of a portfolio's return as the average return of sale trades and derive the dependence of the *n-th* statistical moments of return on the *n-th* statistical moments of market trade values and volumes. That dependence is similar to the dependence of the *n-th* statistical moments of price on the *n-th* statistical moments of trade values and volumes (Olkhov, 2021; 2022a). Our pure theoretical paper shows that the market origin of stock returns' stochasticity depends on the randomness of 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 the probability density functions of return depend on the statistical moments of market trade values and volumes. In turn, the statistical moments of market trade values and volumes are determined by the regular frequency-based probability (1.2; 1.3).

In Section 2, we discuss preliminary considerations. 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 dependence of the *n-th* statistical moments of return on statistical moments of market trade value and volume. 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. Conclusion in Section 7. 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 price-return relations.

We assume that readers are familiar 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 and terms that are not given in the text.

2. Preliminary considerations

Conventional frequency-based statistics of the return time series (1.1) do not describe the impact of the randomness of market trade size on stock returns. However, the return as a result of a big single trade value should contribute more to the average return during the averaging interval Δ than the returns as a result of several deals at small values.

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* sale trades during the averaging period *Δ*. 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). One can consider all sale trades during *Δ* as a "portfolio" and each particular trade as a "security" of the portfolio. Then, to assess the average return of trades during period *Δ* one should reproduce Markowitz' definition and weighted the time series of returns by their "relative value" in the past.

However, the frequency-based probability of return time series assumes that 1000 sale trades with the same return *r* are much more probable than one trade with a return *R*. Meanwhile, if these 1000 sale trades were made with an initial value of $1 each, then such trades should have much less impact on the mean return during *Δ* than a single sale trade of the initial value of $1 billion with a return of *R*. These reasons are similar to those that justify 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 shares at a price of p_1 has much more impact at the average price than 100 trades of 1 share each at a price of p_2.

There are strong financial parallels between the definition of the value weighed return of the portfolio (Markowitz, 1952) and the definition of VWAP (Berkowitz et al., 1988; Duffie and Dworczak, 2018). We use VWAP to introduce the market-based statistical moments of stock price (Olkhov, 2021; 2022a; 2022b). In this paper, we use Markowitz's definition of the value weighed return of the portfolio to introduce the market-based statistical moments of stock return.

We consider the time series of market trade value $C(t_i)$, volume $U(t_i)$, and price $p(t_i)$ during the averaging interval *Δ* as samples of random variables. For simplicity, we take the shift *ε* between t_i and t_{i-1} to be constant.

$$t_i - t_{i-1} = \varepsilon \tag{2.1}$$

We note as *Δ* the averaging interval "today" at *t*:

$$t - \frac{\Delta}{2} < t_i < t + \frac{\Delta}{2} \quad ; \quad i = 1, \dots N \tag{2.2}$$

We assume that the number *N* of the terms of the time series t_i of the trade values $C(t_i)$ and volumes $U(t_i)$ in interval *Δ* (2.2) is sufficient to assess the statistical moments using regular frequency-based probability (1.2; 1.3). We consider the moving average in Section 5. We assume that all prices are adjusted to present prices at time *t* and take the return $r(t_i,\tau)$ (1.1) with the time shift *τ* as a simple ratio of the price $p(t_i)$ at time t_i to the price $p(t_i\text{-}\tau)$ at time $t_i\text{-}\tau$

and take the time shift $\tau=\varepsilon l$, $l=1,2,...$ We assume that the time series of the trade values $C(t_i)$ and volumes $U(t_i)$ have terms t_i, $i=0,-1,-2,...$, so that $t_i-\tau = t_i-\varepsilon l = t_j$ for some t_j in the past.

The description of the market-based statistical moments of stock returns is similar to the description of the market-based statistical moments of price. We briefly present the main results that describe the market-based statistical moments of price and refer to (Olkhov, 2021; 2022a; 2022b) for further details.

3. The market-based statistical moments of price

To some extent, the market-based statistical moments of price are determined by generalization of the definition of average price as a VWAP (Berkowitz et al., 1988; Duffie and Dworczak, 2018). Indeed, VWAP describes the average price, or the 1-st statistical moment of price, that is considered a random variable during the averaging interval. However, the 1-st statistical moment doesn't describe all the statistical properties of a random variable. To describe a random variable, one should describe the set of n-th statistical moments for n=1,2,… We introduce market-based *n-th* statistical moments of price that maintain the structure of VWAP. Let us consider the trade value $C(t_i)$ and volume $U(t_i)$ at time t_i that determine the price $p(t_i)$:

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

The simple equation (3.1) states that the given statistical distributions of the trade values $C(t_i)$ and volumes $U(t_i)$ as random variables during Δ determine the statistical properties of the random 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 market trade randomness as the origin of the price stochasticity and show how statistical moments of trade value $C(t_i)$ and volume $U(t_i)$ determine the statistical moments of the price. We take the trade values $C(t_i)$, volumes $U(t_i)$, and price $p(t_i)$ as random variables during Δ. One can equally describe the statistical properties of 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 (2.2), we assess the *n-th* statistical moments for $n=1,2,...$ of market trade value $C(t;n)$ and volume $U(t;n)$ using the regular frequency-based probability (1.3):

$$C(t;n) \equiv E[C^n(t_i)] \sim \frac{1}{N}\sum_{i=1}^{N} C^n(t_i) \quad ; \quad U(t;n) \equiv E[U^n(t_i)] \sim \frac{1}{N}\sum_{i=1}^{N} U^n(t_i) \tag{3.2}$$

To determine the *n-th* statistical moments $p(t;n)$ of price (3.4), take the *n-th* power of (3.1):

$$C^n(t_i) = p^n(t_i)U^n(t_i) \tag{3.3}$$

The relations (3.3) 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 to define the *n-th* statistical moments *p(t;n)* (3.4) of price (Olkhov, 2021; 2022a) alike to the well-known VWAP (Berkowitz et al., 1988; Duffie and Dworczak, 2018):

$$p(t;n) \equiv E[p^n(t_i)] = \frac{1}{\sum_{i=1}^{N} U^n(t_i)} \sum_{i=1}^{N} p^n(t_i) U^n(t_i) \quad ; \quad n = 1,2,\ldots \tag{3.4}$$

From (3.2; 3.3) one can present the relations (3.4) as (3.5):

$$p(t;n) = \frac{\sum_{i=1}^{N} C^n(t_i)}{\sum_{i=1}^{N} U^n(t_i)} = \frac{C(t;n)}{U(t;n)} \quad ; \quad C(t;n) = p(t;n)\, U(t;n) \tag{3.5}$$

The economic justification of (3.4; 3.5) is rather simple. The VWAP *p(t;1)* (3.5) equals the ratio of the sum of value of trades $\sum_{i=1}^{N} C(t_i)$ during period *Δ* to the sum of volume $\sum_{i=1}^{N} U(t_i)$ of these trades. Absolutely the same is valid for the average of the n-th power of price or the n-th statistical moment of price *p(t;n)* (3.4). It is equal the ratio of the sum of the n-th power of each trade value $\sum_{i=1}^{N} C^n(t_i)$ to the sum of the n-th power of the volume of these trades $\sum_{i=1}^{N} U^n(t_i)$. It is a direct consequence of equations (3.1) and (3.3) that determine the n-th power of price $p^n(t_i)$ through the n-th powers of trade value $C^n(t_i)$ and volume $U^n(t_i)$.

Obviously, the first statistical moment *p(t;1)* of price coincides with VWAP. The set of the *n-th* statistical moments of price *p(t;n)* completely describes its properties as a random variable during *Δ*. However, due to a finite number of terms *N* in the time series of the market trade values $C(t_i)$ and volumes $U(t_i)$, relation (3.2) assesses only a finite number of the *n-th* statistical moments. Thus, the relations (3.4; 3.5) 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 statistical moments of stock return is similar to the derivation of the market-based statistical moments of price. We consider the returns of all sale trades during the averaging interval *Δ*. We start with the market equations (3.1; 3.3):

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

We take the return $r(t_i,\tau)$ (1.1) 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 $C_o(t_i,\tau)$ (4.3) as the original value determined by the market price $p(t_i\text{-}\tau)$ in the past at $t_i\text{-}\tau$ and by the same trading volume $U(t_i)$ at t_i:

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

The equation (4.2) takes the form of (4.4), which determines the stock return $r(t_i,\tau)$ (1.1) through the market trade value $C(t_i)$ and the original value $C_o(t_i,\tau)$ (4.3):

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

The equation (4.4) on return $r(t_i,\tau)$ has a form that is identical to the form of the price equation (3.1). The equation (4.4) has a simple interpretation in terms of Markowitz's portfolio theory. For $i=1,2,..\ N$, one can consider (4.4), as the relation between the original value of the "security" $C_o(t_i,\tau)$, the return $r(t_i,\tau)$ and the current value $C(t_i)$. Similar to (3.3), the *n-th* power of (4.4) gives:

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

Similar to (3.2) we determine the *n-th* statistical moments $C_o(t,\tau;n)$ (4.6) of the original values $C_o(t_i,\tau)$:

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

Similar to the equations (3.1; 3.3), 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 $C^n(t_i)$, original value $C_o(t_i,\tau)$ (4.3), and return $r^n(t_i,\tau)$. The given *n-th* statistical moments $C(t;n)$ (3.2) of the value and the *n-th* statistical moments $C_o(t,\tau;n)$ (4.6) of the original value determine the *n-th* statistical moments $r(t,\tau;n)$ of return. From (4.5; 4.6) and (3.2), similar to (3.4; 3.5), we determine the *n-th* statistical moments of return $r(t,\tau;n)$:

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

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

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

The economic justification of (4.7-4.9) is completely the same as the justification for the statistical moments of price (3.4; 3.5). The return equations (4.4; 4.5) have the same form as the price equations (3.1; 3.3). The average return $r(t,\tau;1)$ at time t with time shift τ equals the ratio of the sum of the sale trades $\sum_{i=1}^{N} C(t_i)$ at time t during Δ to the sum of the original trades $\sum_{i=1}^{N} C_o(t_i,\tau)$ at time t-τ, and that completely matches Markowitz's definition of the portfolio return. Simply speaking, portfolio return is the average return $r(t,\tau;1)$ of the securities that compose the portfolio. One can consider Markowitz's definition of portfolio return as the origin for the definition of average price as VWAP. Both have the same economic meaning and the same structure. The definitions of the average of the *n-th* power of return or the *n-th* statistical moments of return $r(t,\tau;n)$ for $n=2,3,..$ (4.7-4.9) reproduce the definitions of the n-th statistical moments of price *p(t;n)* (3.4; 3.5) and follow the same logic as Markowitz's definition. The *n-th* statistical moments of return *r(t,τ;n)* equals the ratio of the sum of the n-th power of the sale trade value $\sum_{i=1}^{N} C^n(t_i)$ at time *t* to the sum of the *n-th* power of the original values $\sum_{i=1}^{N} C_o^n(t_i,\tau)$ at time *t-τ.* These definitions conform to the structure and meaning of the VWAP and Markowitz's definition of the portfolio return.

The dependence of the statistical moments *r(t,τ;n)* (4.7-4.9) of return on the statistical moments of the trade values *C(t;n)*, $C_o(t,\tau;n)$ (3.2; 4.6), and volumes *U(t;n)* (3.2), highlights the impact of the random size of the value and volume of trades on the market-based probability of stock returns. The largest investors, traders, banks, and financial authorities - all those who perform the major market transactions and forecast macroeconomic, financial, and market trends should take into account the impact of the random size of the market trades on statistical moments of stock return.

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

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

We repeat, that VaWAR *r(t,τ;1)* (4.10) of all sale trades during the averaging interval *Δ* coincides with Markowitz's (1952) definition of return of the portfolio composed of *N* securities with the original values $C_o(t_i,\tau)$ at the times t_i-τ and returns $r(t_i,\tau)$ of each "security *i"*. The same time (4.10) takes the form alike to the well-known expression of VWAP *p(t;1)* (3.4) for *n=1*:

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

We highlight the relations between VaWAR *r(t,τ;1)* (4.10) and VWAP *p(t;1)* (4.11). Indeed, from (4.7; 4.8) for VaWAR *r(t,τ;1)* (4.10), obtain:

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

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

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

From the equation (4.13), and similar to (3.4; 3.5), obtain the *n-th* statistical moments $p_o(t,\tau;n)$ of original price $p(t_i-\tau)$ at time t_i-τ determined by volumes $U(t_i)$ traded at time t_i:

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

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

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

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

Indeed:

$$E[p_o^n(t_i,\tau)U^n(t_i)] = C_o(t,\tau;n) = p_o(t,\tau;n)U(t;n) = E[p_o^n(t_i,\tau)]E[U^n(t_i)]$$

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 $U^2(t_i)$:

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

$$E[p_o(t_i,\tau)U^2(t_i)] = E[C_o(t_i,\tau)U(t_i)] = C_o(t,\tau;1)U(t;1) + corr_{C_oU}(t,\tau|t)$$

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

$$corr_{p_oU^2}(t,\tau;1|t;2) = corr_{C_oU}(t,\tau;1|t;1) - p_o(t,\tau;1)\sigma_U^2(t)$$

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 original price $p_o(t,\tau;1)$ (VWAPa) at time $t\text{-}\tau$. From (4.11-4.15), obtain for VaWAR $r(t,\tau;1)$:

$$r(t,\tau;1) = \frac{p(t;1)}{p_o(t,\tau;1)} \qquad ; \qquad p(t;1) = r(t,\tau;1)\, p_o(t,\tau;1) \tag{4.16}$$

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 original price $p_o(t,\tau;1)$ (VWAP_a) (4.14; 4.15). From (4.9; 4.15), obtain similar relation for all *n-th* statistical moments of return:

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

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

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

The volatility of return $\sigma^2{}_r(t,\tau)$ at t with time shift τ takes the form:

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

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

Consider the volatility $\sigma^2{}_C(t)$ of the value at t and the volatility $\sigma^2{}_{Co}(t,\tau)$ of the original value at $t\text{-}\tau$:

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

Then the volatility $\sigma^2{}_r(t,\tau)$ of return (4.20) takes the form:

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

The similar relations describe the volatility of return (App. A.6) via the price volatilities:

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

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

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).

5. The market-based autocorrelations of stock return

To describe the autocorrelations of stock returns, let us consider the above for the case of moving averaging intervals. Let us take that the interval Δ_{k-1} in the past contains the same number N of time series $t_{i,k-1}$. Let us take the time shift εj between Δ_k and the previous interval Δ_{k-1}:

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

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

We assume that the interval $\Delta_0=\Delta$ (2.1; 2.2) describes the current time $t=t_0$. All intervals Δ_k, $k=-1,-2,..$ describe the past. Moving averaging intervals (5.1; 5.2) allow to describe the autocorrelations of returns with the time shift εj that can be less than the interval Δ:

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

Consider return $r(t_i,\tau)$ (1.1) 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. Time $t_{i,2}$ can belong to any interval Δ_k, $k=-1,-2,...$ in the past. Let us take that:

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

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:

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

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

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

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

$C_o(t,\tau;1)$ and $p_o(t,\tau;1)$ (4.3; 4.14; 4.15) denote the original average value and price at t-τ. Respectively, $C_o(t_2,\tau_2;1)$ and $p_o(t_2,\tau_2;1)$ denote the original average value and price at t_2-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-τ and at time $t_{i,2}$-τ_2:

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

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

$$C(t_i)C(t_{i,2}) = r(t_i,\tau)r(t_{i,2},\tau_2)\ C_o(t_i,\tau)C_o(t_{i,2},\tau_2) \quad (5.9)$$

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

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

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

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

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

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

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

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 $corr_{Co}(t,\tau|t_2,\tau_2)$ between $C_o(t_i,\tau)$ and $C_o(t_{i2},\tau_2)$.

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

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

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 $corr_{Co}(t,\tau|t_2,\tau_2)$ describes the correlation between the original values $C_o(t_i,\tau)$ and $C_o(t_{i2},\tau_2)$ at time t-τ and t_2-τ_2.

$$C(t;t_2) \equiv C(t;1)C(t_2;1) + corr_C(t|t_2) \quad (5.15)$$

$$C_o(t,\tau;t_2,\tau_2) \equiv C_o(t,\tau;1)C_o(t_2,\tau_2;1) + corr_{Co}(t,\tau|t_2,\tau_2) \quad (5.16)$$

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

$$corr_r(t,\tau|t_2,\tau_2) = \frac{corr_C(t|t_2) - r(t,\tau;1)r(t_2,\tau_2;1)corr_{Co}(t,\tau|t_2,\tau_2)}{C_o(t,\tau;t_2,\tau_2)} \quad (5.17)$$

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).

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

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 $corr_{Co}(t,\tau|t_2,\tau_2)$ (5.16) between the original values $C_o(t_i,\tau)$ and $C_o(t_{i,2},\tau_2)$. If $t_2=t$ then:

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

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

$$corr_r(t,\tau|t,\tau_2) = \frac{\sigma_C^2(t)}{C_o(t,\tau;1)C_o(t,\tau_2;1)} = \frac{\sigma_p^2(t)}{p_o(t,\tau;1)p_o(t,\tau_2;1)} \quad (5.18)$$

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 original 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).

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

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

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

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, *n=1,2,…* 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:

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

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

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 $R_m(t,\tau;x)$:

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

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 approximate (6.3) by the integrable characteristic function Q_m, which has the same first *m* statistical moments:

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

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

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

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:

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

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

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

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

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:

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

The simplicity of (6.9) is compensated by the requirement to assess the 2-d statistical moments (4.8; 4.9; 4.16; 4.18; 4.20) of return. To do that one should assess the 2-d statistical moments of trade value $C(t;2)$ (3.2) and original value $C_o(t;2)$ (4.6). The assessments of the higher *n-th* statistical moments $n=3,4,..$ of the trade values $C(t;n)$ (3.3) and $C_o(t,\tau;n)$ (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

The irregular time series of stock returns themselves don't uniquely determine the choice of the averaging procedure and the probability distribution. The conventional approach considers return time series as a standing-alone sample of a random variable, and that results in a frequency-based (1.2; 1.3) assessment of the random properties of return. However, the market nature of the stochasticity of stock returns implies that statistical moments of return should depend on statistical moments of market trades. The market-based origin of return statistics results in the description of return correlations, return-volume correlations, and return-price correlations as functions of trade value and volume statistical moments and correlations, which differ from the frequency-based assessments.

The choice between frequency-based and market-based assessments of stock return statistical moments is determined by the habits and goals of investors. The largest investors, traders, and banks that manage substantial portfolios and major market transactions should take into account the impact of trade value and volume size randomness on the statistical properties of return as described in our paper. To predict market-based statistical moments of stock return, they should forecast statistical moments of trade values and volumes. That complicates the assessment of return statistical moments but highlights direct ties between market trade stochasticity and the randomness of stock returns.

Frequency-based assessment of return statistics is more simple, familiar, and may follow most investors' expectations. These expectations can influence investment decisions and thus impact market trade stochasticity and, hence, the randomness of stock returns.

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

Appendix A

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 (A.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 original prices with the time shifts τ and τ_2 in the same way:

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

From (5.16) and (6.2), obtain correlation $corr_{po}(t,\tau|t_2,\tau_2)$ of original prices:

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

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

That allows present (5.14) as

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

From (4.17), obtain correlation of return:

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

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

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

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

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

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 presents a simple form of the return-volume correlation. 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 $t_{i,2}$:

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

Similar to the above, define:

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

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

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

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

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

From (B.2) and (3.5), obtain:

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

Hence:

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

Appendix C

The price-return relations

Here 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 $p^m(t_{i,2})$ (3.3) at $t_{i,2}$

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

Similar to (3.2; 3.4; 4.6), we define:

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

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

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

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

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

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

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

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

From (3.2; 3.5; 4.6; 4.8) and (C.2-C.9) obtain:

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

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

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

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:

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

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.

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

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

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

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

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/

Schlarbaum, G.G., Lewellen, W.G. and R.C. Lease, (1978). Realized Returns on Common Stock Investments: The Experience of Individual Investors, J. of Business, 51(2), 299-325

Shephard, N.G. (1991). From Characteristic Function to Distribution Function: A Simple

Framework for the Theory. *Econometric Theory*, 7 (4), 519-529

Shiryaev, A.N. (1999). Essentials Of Stochastic Finance: Facts, Models, Theory. World Sc. Pub., Singapore. 1-852

Shreve, S. E. (2004). Stochastic calculus for finance, Springer finance series, NY, USA

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