Management Studies, May-June 2023, Vol. 11, No. 3, 125-168
doi: 10.17265/2328-2185/2023.03.002

No-Arbitrage in Financial Economics: Solution of the Mystery of Implied Volatility and S&P 500 Volatility Index

Valery V. Shemetov

We have shown that classic works of Modigliani and Miller, Black and Scholes, Merton, Black and Cox, and Leland making the foundation of the modern asset pricing theory, are wrong due to misinterpretation of no arbitrage as the martingale no-arbitrage principle. This error explains appearance of the geometric Brownian model (GBM) for description of the firm value and other long-term assets considering the firm and its assets as self-financing portfolios with symmetric return distributions. It contradicts the empirical observations that returns on firms, stocks, and bonds are skewed. On the other side, the settings of the asset valuation problems, taking into account the default line and business securing expenses, BSEs, generate skewed return distributions for the firm and its securities. The Extended Merton model (EMM), taking into account BSEs and the default line, shows that the no-arbitrage principle should be understood as the non-martingale no arbitrage, when for sufficiently long periods both the predictable part of returns and the mean of the stochastic part of returns occur negative, and the value of the return deficit depends on time and the states of the firm and market. The EMM findings explain the problems with the S&P 500 VIX, the strange behavior of variance and skewness of stock returns before and after the crisis of 1987, etc.

Keywords: geometric Brownian model, Extended Merton model, business securing expenses, option and warrant pricing, corporate debt, default probability

Introduction

The motivation of this article is to trace how false ideas arising in one economic study penetrate into others; how these studies pass their results and false ideas further down like a baton in a relay race. At first, the papers are widely discussed and criticized if their results seem interesting enough, but if the economists fail to reveal any errors and the results look plausible, the logic of the original papers by and by gets wide recognition; it is used in many following studies, converting the original papers into seminal articles. Since that time, no orthodox economist questions the logic and methods of the seminal articles; they defend this logic aggressively against those who take risks to doubt it publicly. As a result, a whole branch of study goes in the wrong direction, poisoned with those ideas. Sometimes, false ideas from one sector of economics penetrate into another, and contamination continues, aggravating the state of economics even more. Of course, such things can happen to any science; we talk about economics just because of our professional preferences.

Here we consider the idea of no arbitrage, its interpretations, and their effects on financial economics, especially, asset valuation, credit risk estimation, the effect of debt on the firm value, etc. The no-arbitrage principle runs like this: "The term arbitrage refers to the possibility of making a trading gain with no chance of

Acknowledgment: I am infinitely thankful to my friend and colleague M. Rubinstein for valuable discussions, deep and constructive remarks, and an invariable interest in my work. Of course, all remaining errors are my full responsibility.

loss" (Financial Economics, Panjer, 1998). More specifically, if an initial value of an admissible portfolio is zero, $V(0) = 0$, then at time $t = 1$, $V(1) = 0$ with the unit probability. In a stochastic environment, the return can be presented as $V(1) - V(0) = [U(1) - U(0)] + [W(1) - W(0)]$, here $U(1) - U(0)$ is the predictable return and $W(1) - W(0)$ is the stochastic return. In these conditions, the no-arbitrage principle assumes the form: $U(1) - U(0) = 0$, and Mean$[W(1) - W(0)] = 0$. We say that an arbitrage opportunity is available if there is an admissible portfolio violating the no-arbitrage principle.

We begin our analysis with the famous Modigliani-Miller Propositions starting asset valuation studies (MMPs, Modigliani & Miller, 1958; 1963), where the authors use the no-arbitrage principle to prove their Theorem of Irrelevance that capital structure makes no effect on the firm value. The MMPs consider mean cash flows and use the no-arbitrage principle in a form: $V(1) - V(0) = 0$. We show that the MMPs conflict with another seminal paper by Merton (1974), which considers the firm value and the value of its security in essentially stochastic conditions. Therefore, Merton uses the no-arbitrage principle in the form: $U(1) - U(0) = 0$ and Mean$[W(1) - W(0)] = 0$. The last equality means that the stochastic variable $V(t)$ makes a martingale. Harrison and Kreps (1979) prove that the portfolio value makes a martingale if and only if it is equivalent to a self-financing portfolio, admitting no payments and no fund infusions. Thus, both the levered and unlevered firms in the MMPs make no payments; it points to an artificial construction of MMP premises. The same conclusion concerns the firm and its long-term security considered by Merton (1974). To stay consistent with the current understanding of no-arbitraging, Merton finally comes to the geometric Brownian model for the firm value, which does not allow any payment.

When the portfolio value makes a martingale, the portfolio's return distribution occurs symmetric. Therefore, the option pricing formulas (Black & Scholes, 1973), and the firm value, stock value, and zero-coupon bond value formulas (Merton, 1973; Merton, 1974; Black & Cox, 1976) are based on the normal distribution as the most "natural" one among other symmetric distributions. Using this logic, Black and Scholes (1973) comes to the option-pricing formula and Merton (1974) derives the equation for pricing any security whose value can be written as a function of the firm value and time. These results stimulate further development of Black-Scholes-Merton's (BSM) ideas: Black and Cox (1976) introduce the default line into financial mathematics, Leland (1994), integrating the BSM ideas with ideas of Black and Cox, works out mathematical foundations of the trade-off theory choosing the optimal debt leverage for the firm. All above-mentioned papers commit serious errors compromising their results, but Leland completes a logical circle: starting with no-arbitrage conditions, containing in methods and formulas of BSM and Black and Cox, he constructs a firm, which is an arbitraging machine. It is most wonderful that this feature of Leland's paper escapes the attention of economists. We will show that all these problems follow from the "classical" interpretation of the no-arbitrage principle. We call this interpretation the *martingale no-arbitrage* principle.

In contrast to theoretical estimations, empirical data give vast evidence that bond and stock returns are skewed. This fact definitely points to drawbacks of the theoretical models, but instead of revising the foundations of the financial risk theories, economists begin mending them with jump-diffusion processes, calibrated models, etc. However, the asymmetry in returns on the firm, stock, bond, etc. is already in Merton's equation (1974), describing the firm dynamics, and the absorbing boundary by Black and Cox (1976): the firm's payments decrease its returns, making them negatively skewed, and the absorbing boundary (the default line) accelerates the process. This asymmetry develops over time reducing the firm's value and the values of long-term assets issued by the firm. Therefore, the no-arbitrage conditions become $U(1)$ - $U(0) < 0$ and Mean[$W(1)$ - $W(0)$] < 0, meaning that both the predictable returns and the mean stochastic returns decrease over time due to their asymmetry. The shorter the time of observation, the lesser the deviation of the situation from the martingale no arbitrage. These conditions explain a relative success in option pricing due to the short life of options and makes short-term speculations quite a martingale process. We call this new interpretation of the no-arbitrage principle the *non-martingale no arbitrage* (understanding, however, that this is the only correct no-arbitrage principle). It brings a lot of inconveniences for long-term investors like big firms, mutual and pension funds, etc., but nobody can hide from reality; to survive and succeed, it is better to rely upon the bitter truth, than on sweet lies. Shemetov (2020) rejects the existence of no arbitraging, but the paper's logic proves that the author objects against the *martingale* no arbitraging. We see the future progress of financial economics with development of the Extended Merton model, which has explained some important empirical observations.

Here we briefly introduce the reader to the Extended Merton model (EMM, Shemetov, 2020; 2021), which we extensively use in this paper. Merton's equation for the firm development is:

$$dX = (\mu X - P)dt + CXdW,\ X(0) = X_0$$

where $X(t)$—firm's stochastic assets, $P > 0$—business securing expenses (BSEs), $P = FC + DP + TAX + DIV$, *FC*—fixed costs, *DP*—debt payments, *TAX*—taxes, *DIV*—dividends, all dollar per time unit; $P(t) = P_0\pi(t)$, $\pi(t)$—a piecewise continuous function of time, μ—the expected rate of return, *C*—volatility. We consider asset returns as $x = \ln(RX/P_0)$; using Ito's lemma, we come to:

$$dx = R(1 - \pi(t)e^{-x})dt + CdW \qquad \text{(i.1a)}$$

$$x(0) = x_0 = \ln(RX_0/P_0),\ R = \mu - C^2/2 \qquad \text{(i.1b)}$$

Equation (i.1a) represents an ordinary diffusion with a drift rate $DR(x,t) = R(1 - \pi(t)e^{-x})$ dependent on the location of Brownian particles on the *x*-axis and time. For the uniform payment, $\pi(t) \equiv 1$, the drift rate is: $DR(0) = 0$, $0 < DR(x) < R$ for $x > 0$, and $-\infty < DR(x) < 0$ for $x < 0$. When a part of the distribution of Brownian particles $V(x, t)$ gets below the line $x = 0$, its particles are transported to the negative infinity with an increasing drift rate, creating a deficit of particles below this line. The diffusion force compensates this deficit with particles from the upper part of the distribution, and the process continues until there are no more particles left above the line $x = 0$ (in economic applications, the process continues until the default probability equals unit).

The return distribution $V(x,t)$ satisfies the equation (V_y is a partial derivative over variable *y*):

$$V_t + R(1 - \pi(t)e^{-x})V_x - 0.5C^2V_{xx} + R\pi(t)e^{-x}V = 0 \qquad \text{(i.2a)}$$

starting its evolution from a normal distribution (the initial condition):

$$V(x,0) = N(x; H_0, \sigma_0^2),\ H_0 = \langle x(0)\rangle = \langle \ln[RX(0)/P_0]\rangle,\ \sigma_0^2 = \langle (x(0) - H_0)^2\rangle \qquad \text{(i.2b)}$$

To solve the problem, one must add boundary conditions for the return distribution, reflecting the fact that when the firm's assets intercept the boundary, the firm defaults (Black & Cox, 1976):

$$DL = \max\{\ln(RX_D/P_0), 0\} \qquad \text{(i.2c)}$$

line $DL = \ln(RX_D/P_0)$ corresponds to outstanding debt X_D (the exogenous default line). Line $DL = 0$ serves as a soft default line (the endogenous default line) because when $x < 0$, business activities inflict losses on the firm. If the problem setting does not include the firm's payments and the default line (e.g., Merton, 1974), the return distribution starting from the normal distribution remains normal, spreading with rate *C*. In actual cases with default line *DL* and payments, the mean return will rise or go down depending on the difference *RX - P*, and the return distribution decreases with rate *DPINT*(*t*), the intensity of default probability:

$$DPINT(t) = 2\int_{-\infty}^{DL(t)} V(x,t)dx \tag{i.3}$$

and default probability *DPR*(*t*) over the interval [0, *t*], *t*—the time of observation, is:

$$DPR(t) = \int_0^t DPINT(\tau)d\tau \tag{i.4}$$

The leak rate through the boundary is proportional to the part of distribution *V*(*x*, *t*) getting below the default line. The loss of mass (Brownian particles constituting the distribution) at the default line induces redistribution of the remaining distribution mass down to the default line. The distribution becomes negatively skewed. The cumulative effect of the diffusion spreads and the distribution distortion makes the negative tail grow heavier over time, increasing the default probability. So, the skewed return distribution *V*(*x*, *t*) is a *natural effect* of the firm dynamics, which does not need any additional mechanisms like a jump-diffusion process for its development.

The fact that asset returns are skewed is now well established (e.g., Simkowitz & Beedles, 1980; Singleton & Wingender, 1986; Badrinath & Chatterjee, 1988; Fortune, 1996; Harvey & Siddique, 2000). However, the nature and driving forces of asset skewness remain unclear to economists and investors unfamiliar with the EMM. The novel vision of the asset return development helps to improve understanding of some financial management problems. The EMM used for analysis of the firm's debt leverage demonstrates that debt *negatively* affects the returns and survival of the firm. The EMM reveals inconsistency in debt studies by Leland (1994); Leland and Toft (1996), etc., which make a quantitative basis for the trade-off theory.

Because of the continuous leak of Brownian particles at the default line, the mean of the stochastic part of returns is always negative, and stochastic returns never make a martingale. When the firm is far from default, the leak is low, making an illusion that returns make a martingale. When the firm comes closer to default, the leak increases noticeably, and the returns cannot be assumed a martingale anymore. That is why no-arbitrage acts in the form of the *martingale* no arbitrage in the market of the GBM firms (no leak at the default line) and becomes the non-martingale no arbitrage in the market of the EMM firms paying their BSEs (see Sections 3 and 4). So, all theoretical constructions admitting the martingale no arbitrage have a limited domain of validity. The non-martingale no-arbitrage principle in the market of firms paying their BSEs makes void the *risk-neutral approach* and *risk-neutral probabilities* (e.g., Cox et al., 1979; Harrison & Kreps, 1979; Harrison & Pliska, 1981) inapplicable to the analysis of credit risks.

To compute statistical moments of the firm's return like the mean, variance, skewness, etc., one must calculate the return distribution of the boundary problem (i.2) for the firms of the same prehistory and existing at time *t*. One can get this distribution $\hat{V}(x,t)$ from distribution $V(x,t)$ as:

$$\hat{V}(x,t) = V(x,t) - V(2DL - x, t) \tag{i.5}$$

determined in the interval of [DL, +∞) and meeting the condition of $\hat{V}(DL,t) = 0, t \in [0, +\infty)$. The moments of the return distribution $\hat{V}(x,t)$ are:

$$H(t) = \int_{DL}^{\infty} x\hat{V}(x,t)dx \ , \ Var(t) = \int_{DL}^{\infty} (x - H)^2\hat{V}(x,t)dx \tag{i.6}$$

$$SK(t) = \int_{DL}^{\infty} (x - H)^3 \hat{V}(x,t) dx$$

Parameters governing the evolution of the return distribution are: distance to default $H(t) - DL(t)$, variance $VAR(t)$, skewness $SK(t)$, the intensity of market shocks C^2, and stochastic ratio $RX(t)/P(t)$. One can add here a general economic motion influencing the firm development: a growing economy with higher return rates and lesser default risks, or a declining economy with lower return rates and higher default risks.

The rest of the article is organized as follows: in Section 1, we consider early empirical and theoretical results introducing the idea of a symmetric Brownian motion of stock and bond prices into financial economics and the first signs that the reality is more complicated. In Section 2, we consider Modigliani-Miller Propositions and the no arbitrage method of their proof; we show that MMPs are obtained if and only if both levered and unlevered firms can be modeled with self-financing portfolios admitting no payments and fund infusions. In Section 3, we discuss the Black-Scholes-Merton model and show how the classic interpretation of the no-arbitrage principle brings Black and Scholes to their formula for option pricing and lets Merton down in his formulas for warrant pricing, etc. We also discuss the implied volatility and its applications for estimating market volatility (the S&P 500 Volatility Index). Finally, we review theoretical improvements introduced into the BSM model after 1973 and the progress achieved in the description of advanced options. In Section 4, we analyze Merton's equation for pricing the firm and any security issued by this firm and demonstrate how Merton has come to the geometric Brownian equation as a descriptor of the firm's asset dynamics and the problems caused by this choice. In Section 5, we discuss the default line introduced by Black and Cox and its consequences for financial analysis. In Section 6, we consider the synthesis of ideas of Merton (1974) and Black and Cox (1976) used for analysis of the effect of debt on the firm value by Leland (1994), Leland and Toft (1996) and Leland's choice of the optimal debt leverage making the essence of the trade-off theory. Section 7 concludes the paper.

Early Empirical and Theoretical Results Supporting Idea of Brownian Motion of Stock Prices

What arguments make economists believe that stock market prices follow the Brownian motion? The main contribution has been made by two empirical papers by Kendall (1953) and Osborn (1959) and three theoretical papers using the ideas similar to that of Kendall and Osborn and providing the break-through in optimal portfolio allocation (Markowitz, 1952) and capital asset pricing (Sharp, 1964; Lintner, 1965).

The original objective of Kendall's study (1953) was to reveal regular cycles in market stock prices, but he has concluded that market prices follow a random walk. As he says, each data series appears to be "a wandering one, almost as if once a week the Demon of Chance draws a random number from a symmetrical population of fixed dispersion and adds it to the current price to determine the next week's price." (Kendall, 1953, p. 13). This result has been confirmed by Osborn who argues that:

> if $Y = \ln[P(t+\tau)/P_0(t)]$, where $P(t+\tau)$ and $P_0(t)$ are the price of the same random choice stock at random times $t+\tau$ and t, then the steady state distribution of Y is
>
> $$\varphi(Y) = (2\pi\sigma^2\tau)^{-1/2} \exp[-Y^2/(2\sigma^2\tau)],$$
>
> which is precisely the probability distribution for a particle in Brownian motion, if σ is the dispersion developed at the end of unit time. (Osborn, 1959, p. 145)

In other words, changes in stock price logarithms (asset returns) are independent random values following a normal distribution with a fixed variance.

Markowitz (1952) looks for the optimal portfolio allocation using two asset characteristics: their means and variances (co-variances). He considers the variance as a measure of asset risks. The mean-variance portfolio optimization supposes the normal distribution of asset returns, anticipating the results of Kendall and Osborne. In Markowitz's mean-variance portfolio optimization, the investor maximizes the mean portfolio return, keeping the variance fixed. Markowitz's theory is the first-ever scientific portfolio allocation method enthusiastically recognized by financial theorists and practitioners.

Mandelbrot (1963) and Fama (1965) cast the first doubts that stock prices follow the random walk model. Mandelbrot (1963) states that the empirical distributions of price changes conform better to the stable Paretian distributions than to its marginal form (the normal distribution). A typical representative of the stable Paretian distributions looks like a symmetric leptokurtic distribution with heavy tails. (In theory, their tails are so heavy that a typical distribution has an infinite variance). Fama (1965) collects statistics of daily changes in logarithmic prices that show that empirical frequency distributions of log prices for all stocks under study contain more relative frequency in their central part and more relative frequency in their extreme tails than it would be expected under the normal hypothesis. Our explanations of these results can be seen in Section 3.

By and by, economists have accumulated empirical data showing that not only asset returns but portfolio returns, too, are skewed. Simkowitz and Beedles (1980), Singleton and Wingender (1986), Badrinath and Chatterjee (1988), Fortune (1996), and Harvey and Siddique (2000) present empirical evidence of skewness in individual stock returns and market indexes in the US stock markets. Shemetov (2020; 2021), considering the firm valuation in settings with the firm's payments and the default line, shows that an initially normal distribution of asset returns becomes skewed over time. Those facts make economists reconsider methods for efficient portfolio allocation. Considering portfolio allocation with higher moments (skewness, kurtosis), Kraus and Litzenberger (1976) recommend selecting positively skewed assets to the portfolio. So, researchers looking for efficient portfolios begin to maximize the portfolio mean returns and skewness while keeping their variance fixed (Mencia & Sentana, 2009; Harvey et al., 2010; Adcock et al., 2012; Krueger, 2021). Opposing those views, Shemetov (2022) shows that high positive skewness comes with a high default probability, and the unconditional maximization of skewness dangerously increases portfolio risks. He also demonstrates that for skewed asset returns, the variance is no adequate measure of asset risks (see Section 3); thus, its minimization is useless. A proper measure of risks for skewed return distributions is the intensity of default probability. In portfolio optimization with higher moments, Shemetov (2022) recommends solving the following problem: *maximize the mean return and skewness while keeping the intensity of default probability under control*. Such an approach helps select both efficient and reliable portfolios.

The Capital Asset Pricing model, CAPM (Sharp, 1964; Lintner, 1965), developing Markowitz's ideas, relates the expected portfolio return $E(r_p)$ with the expected market return $E(r_M)$ and beta:

$$E(r_p) = r_f + \beta_p[E(r_M) - r_f]\,,\ \beta_p = cov(r_p, r_M)/var(r_M)$$

here r_f is the return on a risk-free asset. CAPM uses two basic assumptions:

(1) asset returns are normally distributed or

(2) investors have a quadratic utility function.

Notwithstanding these rigid limitations, practitioners widely apply the mean-variance theory for asset allocation because it provides an intuitively clear assessment of the relative merits of alternative portfolios. Moreover, one can span the mean-variance efficient frontiers with only two funds, which simplifies calculation and interpretation of these efficient frontiers. However, now economists know that asset returns are skewed, and the classic CAPM does not satisfy them anymore. An intensive search for the next generation of capital asset pricing models continues.

Modigliani-Miller Propositions

The modern theories of capital structure and dividend policy decisions start with the outstanding series of papers by Modigliani and Miller (1958; 1961; 1963). Modigliani-Miller Propositions (hereafter MMPs) have been under intense scrutiny since their publication, and Miller (1988) remarks in his jubilee paper that first discussions were very hot. However, the MMP critique by and by calms down by the 90s, and the MMPs and the arbitrage method of their proof are now generally recognized (Bhattacharya, 1988). The main MMP results are, in brief, as follows. The MMP1 (1958) states that the capital structure does not affect the expected firm value in the perfect market. In other words, there is no optimal debt leverage in those conditions. The MMP2 (1961) argues that the dividend policy in the same conditions does not influence the expected firm value. The MMP1 and MMP2 shocked economists at the time of their publication because the choice of the capital structure and the dividend policy were considered crucial for the firm success. The MMP3 (1963) considers two firms: an unlevered firm and a levered firm identical to the unlevered one in all aspects but the capital structure. The theorem states that in the perfect market with corporate income taxes, the after-tax expected value of the levered firm equals the after-tax value of the unlevered firm plus tax deductions. This result gives theoretical underpinnings to the trade-off theory arguing that in the presence of corporate taxes and bankruptcy costs there is an optimal debt leverage maximizing the firm value.

Our objective is to test the MMP results in the light of the EMM (Shemetov, 2020; 2021) and show that the MMPs follow from the perfect market assumptions joined with an implicit assumption that the firm value meets the GBM distribution with no payments at all. We show that the MMP1 and MMP3 are false even when the firm's payments are proportional to the firm value, although the return distribution remains normal in that case. None of the MMPs holds good for the EMM firms with payments to be a function of time, and the return distribution becomes skewed. Because debt decreases the mean after-tax value, the mean after-tax value of the unlevered firm is higher, not lesser, than the mean after-tax value of the identical levered firm. It implies that the MMP3 and the trade-off theory following from it are wrong; hence, all conclusions and recommendations based on the trade-off theory are false.

The idea of proof of the MMPs consists of constructing an analog of the Marshallian industry for firms' cash flows and then applying the one-price principle to the market of perfect substitutes. Modigliani and Miller (1958) consider firms in the perfect market described with the assumptions:

(1) The firm value is determined only by the mean cash flow generated by the firm;

(2) All investors have full information about firms' cash flows; thus, the investors have homogenous expectations on corporate cash flows and their riskiness;

(3) There is an "atomistic" competition and no market friction of any kind. That implies, among other things, that at the market of corporate stocks and bonds (a) there are no agency costs, (b) bankruptcy entails no liquidation

costs, and (c) all investors, both individuals and institutions, can borrow at the same rate as corporations;

(4) The debt of firms and investors is riskless, so the interest rate of all debts is the risk-free rate for all possible amounts of debt;

(5) There are no corporate or personal income taxes;

(6) "All firms can be divided into 'equivalent return classes' (Miller (1988) calls them 'risk classes') such that the return on the shares issued by any firm in any given class is proportional to (and hence perfectly correlated with) the return on shares issued by any other firm in the same class" (Modigliani & Miller, 1958, p. 266). "All relevant properties of a share are uniquely characterized by specifying (1) the class to which it belongs and (2) its expected return." The authors claim that this assumption permits them "to classify firms into groups within which the shares of different firms are 'homogeneous,' that is, perfect substitutes for one another" (ibid).

The authors implicitly use one more assumption that the firm cannot default in its development because the firm space has no default line hitting which the firm defaults, and the firm cannot default on a special date like the maturity of its debt.

Assumption (1) looks strange taking account of the Markowitz mean-variance theory of efficient portfolio allocation (1952) recognized by 1958. This theory puts forward the hypothesis that the firm value is completely determined by its mean cash flow and risk represented by the cash flow variance. It seems natural to characterize the risk class with the expected return and variance. There is also a hypothesis supported by practical observations that the higher the firm's debt, the higher the risk. Discussing the effect of debt on the value of the firm's warrant, Merton (1973, p. 151) says "If the firm changed its capital structure by raising the debt/equity ratio, then the riskiness of the common stock would increase, and the warrant would become more valuable". Leland (1994) uses this relation as an established fact to show that the firm's return is a convex-up function of debt leverage. Using Markowitz's hypothesis and the dependence of the firm risk/variance on its leverage, one can conclude that the unlevered firm and the identical levered firm can never get into the same risk class, and, therefore, the firm value *depends* on its capital structure rejecting the MMP1.

The arbitrage proof of MMP1 can be explained by the rule of contraries. Let us consider two firms within the same risk class, an unlevered firm and the levered one identical to the first in all aspects but the capital structure. Suppose also to the contrary of MMP1 that the *value of the levered firm is higher* than that of the unlevered firm. Now investors, having in their portfolios shares of both levered and unlevered firms, can sell shares of the levered firm, buy cheaper shares of the unlevered one, and, borrowing enough debt, reproduce the capital structure of the levered firm in their portfolios but at a lower cost. The perfect market conditions guarantee that the cost of debt (the interest rate) remains the same for the levered firm and investors. So, this line of actions brings arbitrage profits to the investors. The arbitrage profit will exist until the values of levered and unlevered firms differ. To prevent arbitrage profits, the authors conclude that both levered and unlevered firms must have the same value, and the capital structure does not affect the firm value. However, there are two more possible conclusions in this line of reasoning: (1) the levered and unlevered firms identical in all aspects but the capital structure cannot get into the same risk class, and (2) the risk classes with the properties required by Assumption 6 do not exist. Merton's model (1974) supports the first alternative conclusion.

We test the MMP1 using Merton's model (1974), where the firm value is described by the equation:

$$dX = (\mu X - P)dt + CXdW, \; X(0) = X_0 \quad (2.1)$$
$$P = DP + DIV$$

Here $X(t)$ is the firm market value at time t, μ is the rate of instantaneous expected returns on the firm per unit time, P is the total dollar payouts by the firm per unit time to either its shareholders or liabilities-holders (dividend *DIV* or interest *DP* payments) per unit time, constant C^2 is the instantaneous variance of returns, W is a Wiener process.

Let the levered firm have its payments proportional to the firm value, $P = \delta X$, $0 < \delta < \mu$, while the unlevered firm pays nothing, $\delta = 0$. The firms are identical in all other respects: they have the same initial values X_0, expected returns μ, and the same intensity of shocks C. The model (2.1) describing dynamics of both firms in these conditions becomes the geometric Brownian model (GBM):

$$dX/X = (\mu - \delta)dt + CdW \ , \ X(0) = X_0 \tag{2.1a}$$

For the logarithmic variable $z = \ln(X/X_0)$, (Eq. 2.1a) by Ito's lemma transforms to:

$$dz = (R - \delta)dt + CdW, \ R = \mu - C^2/2 \tag{2.2}$$

and the mean return for the levered firm grows with the rate of expected return $R - \delta$:

$$\langle z_L(t) \rangle \equiv H_L(t) = (R - \delta)t$$

while the unlevered firm's return grows with the rate of expected return R:

$$\langle z_U(t) \rangle \equiv H_U(t) = Rt$$

One can see that $H_U(t) > H_L(t), t > 0$, and the firm value depends on its capital structure in the GBM with proportional payments.

What is the effect of taxes on the values of two identical levered and unlevered firms? Let payments of the unlevered firm consist of dividends only, $P_U = DIV = \delta_1 X$, $\delta_1 > 0$, while payments of the levered firm consist of dividends and debt payments, $P_L = DIV + DP = \delta_2 X$, $R > \delta_2 > \delta_1$. The returns $H_L(t)$ and $H_U(t)$ of both firms, described by Eq. (2.2), can be effectively considered as the returns of two *unlevered firms with no payments* $(P_{eL} = P_{eU} = 0)$ with different effective rates of return. Correspondingly, returns of the second firm, $H_L(t) \equiv \langle z_L(t) \rangle = (R - \delta_2)t$, $z = \ln(X/X_0)$, are lesser than returns of the first firm, $H_U(t) \equiv \langle z_U(t) \rangle = (R - \delta_1)t$. As an unlevered firm, the second firm has no right on the tax shield and pays its tax at the same rate as the first firm. Therefore, the after-tax mean value of the levered firm is *lesser* than the after-tax mean value of the unlevered one. This conclusion rejects the MMP3 (1963) in the GBM with payments proportional to the firm value. It means that debt *negatively affects* the after-tax mean value. So, all kinds of the trade-off theory, claiming after the MMP3 a positive effect of debt on the after-tax mean value, are wrong in the GBM, $P = \delta X$ (e.g., Kraus & Litzenberger, 1973; Leland, 1994; Ju et al., 2005; Frank & Goyal, 2007). All papers on the optimal capital structure, using the MMP3 and the GBM, $P = \delta X$ (e.g., Leland, 1994; Leland & Toft, 1996; Goldstein et al., 2001), are self-contradictory because the GBM, $P = \delta X$, is inconsistent with the MMPs.

We will not analyze the MMP2 (1961) here because it has already got a good share of criticism for its unrealistic assumptions (the perfect capital market, no taxes, the fixed investment policy, no risk of uncertainty, investors are indifferent between the dividends and the capital-gain income, etc.) and conclusions having no support in practice (see Baker & Powel, 1999; H. DeAngelo & L. DeAngelo, 2006; Dhanani, 2005).

Merton's model shows that the only way to obtain all MMPs in the frameworks of one model is to have absolutely no payments for the levered and unlevered firms. Because there are no dividend payments for both firms (*DIV* = 0), the dividend policy does not affect the firm value (MMP2, 1961). Because there are no debt payments (DP = 0), the capital structure has no influence on the firm value (MMP1, 1958). Because the levered firm does not pay for its debt but presumably enjoys the tax shield, its mean after-tax value is higher than the mean after-tax value of the unlevered firm due to tax deductions (MMP3, 1963). However, the MMP3 is a logical error. Because the levered firm does not pay for its debt, it is indistinguishable from the identical unlevered firm, and, therefore, *its tax shield must be zero*. We have that in the perfect market with corporate income taxes, the after-tax value of the levered firm equals the after-tax value of the unlevered firm (MMP1 with corporate taxes). The firm's payments make all MMPs wrong.

It is interesting that attempts to test M&M Invariance Theorem experimentally continue 60 years after the first publication of the theorem (Modigliani & Miller, 1958). Charness and Neugebauer (2019) investigate the invariance theorem in experimental asset markets, finding value-invariance for assets of identical risks (the levered assets and unlevered assets belong to the same risk class) when the returns to these assets are perfectly correlated. The authors do not motivate how (any) two different assets can have perfect correlation of their returns. Without such an explanation, their problem setting is unreal, and their results have neither theoretical no practical value.

The MMPs can be obtained if and only if both levered and unlevered firms can be modeled with self-financing portfolios admitting no payments and fund infusions. In the market of self-financing firms, the *martingale no-arbitrage principle* (see Section 3) is effective (Merton, 1973; Cox et al., 1979; Harrison & Kreps, 1979; Harrison & Pliska, 1981). The martingale no arbitrage is valid if and only if the mean stochastic return on the firm is zero. The mean stochastic return is zero if and only if the return distribution is symmetric. So, the firm's return distribution is symmetric in the MMPs. The last conclusion contradicts the facts observed in practice; we have a lot of reliable evidence that the firm's return distributions are skewed. The verdict on the MMPs is that the model is oversimplified to have any practical value. For long years of recognition, the MMPs stimulate the development of false intuition on the firm capital structure and continue to undermine logic and validity of all studies using them.

Nevertheless, the historical value of the MMPs as one of the first scientific financial models causes no doubts. The MMPs have set up the foundation and pointed the direction for capital structure theories for decades. The martingale no arbitrage principle and the MMPs serve as an essential argument in the seminal papers on option pricing (e.g., Black & Scholes, 1973; Merton, 1973), in the most influential article on valuation of corporate debt by Merton (1974), and in the paper on the joined effect of debt and taxes on the firm value by Ross (1985), and a more resonant series of articles on the same subject by Leland and his co-authors (Leland, 1994; Leland & Toft, 1996; Goldstein et al., 2001).

The Black-Scholes-Merton Model and Option Pricing

Theory of BSM Model

The pricing of financial derivatives is a crucial problem in portfolio management and investing. The Black-Scholes-Merton (BSM) model is the first universally recognized theory of rational option pricing on the long way of pricing various financial derivatives. The original BSM model offers the Black-Scholes (BS) formulas for pricing European call and put options that made options trading less gambling and more scientific. The BS formula relates the option price with the price of an underlying stock, the risk-free interest rate, the option exercise/strike price, its expiration time, and volatility and the return distribution of the underlying stock. All parameters, except volatility, are directly observable. This important feature of the BSM model favorably distinguishes it from its predecessors (e.g., Sprenkle, 1961; Samuelson, 1965; Samuelson & Merton, 1969; Chen, 1970). Empirical analysis confirms that the price estimates produced by the BS formula are often close to the prices observed in exchange. Many modified option pricing models are now at the disposal of option traders.

Scholes (2023) argues the option market prices can render an essential service to the theorists studying market risks and to the traders taking those risks into account in their operations: option prices can serve as a reconnaissance drone providing valuable information about the market risks in real time. It makes the basis for managing market risks proactively.

The main idea behind the method of option pricing is hedging the option by buying and selling the underlying stock in calculated proportions to eliminate risks.

The BS formula uses the following assumptions:

(a) The short-term interest rate is known and is constant through time.

(b) The stock price x follows a random walk in continuous time with constant variance rate v^2 proportional to the square of the stock price. Thus, the distribution of possible stock prices at the end of any finite interval is lognormal.

(c) The stock pays no dividends or other distributions.

(d) The option is "European" if it can only be exercised at its maturity, t^*.

(e) There are no transaction costs in buying or selling the stock or the option.

(f) It is possible to borrow any fraction of the price of a security to buy it or to hold it, at the constant short-term interest rate, r.

(g) There are no penalties to short selling. A seller who does not own a security will simply accept the price of the security from a buyer, and will agree to settle with the buyer on some future date by paying him an amount equal to the price of the security on that date.

The authors consider the European call option price, $w(x, t)$, as a function of the stock price x, time t, and the problem parameters. Black and Scholes (hereafter B&S) create a hedged portfolio consisting of a long position in the stock and a short position in the option. In this case, the value of that hedged position does not depend on the stochastic stock price but is a function of time and the problem parameters. B&S derive the equation:

$$w_t + rxw_x + 0.5v^2x^2w_{xx} - rw = 0 \tag{3.1a}$$

with a condition at the option maturity:

$$w(x, t^*) = \max[x - c, 0] \equiv [x - c]^+ \tag{3.1b}$$

A solution of the BS problem (3.1) is:

$$w(x,t) = xN(d_1) - ce^{-r(t^* - t)}N(d_2)$$

$$d_1 = [\ln\left(\frac{x}{c}\right) + (r + 0.5v^2)(t^* - t)]/(v\sqrt{t^* - t}) \tag{3.2}$$

$$d_2 = [\ln\left(\frac{x}{c}\right) + (r - 0.5v^2)(t^* - t)]/(v\sqrt{t^* - t})$$

c is the exercise price, t^*—the option expiration date (maturity), $N(x)$ is a cumulative normal distribution.

B&S consider also the pricing of European put options (options to sell). Denoting the value of a put option as $u(x, t)$, they show that the differential equation remains the same:

$$u_t + rxu_x + 0.5v^2x^2u_{xx} - ru = 0 \quad (3.3a)$$

but the boundary condition changes to:

$$u(x,t^*) = \max[c - x, 0] \equiv [c-x]^+ \quad (3.3b)$$

The solution to problem (3.3) is:

$$u(x,t) = -xN(-d_1) + ce^{-rt^*}N(-d_2) \quad (3.4)$$

where d_1 and d_2 are determined in Eq. (3.2). Concerning the behavior of the option price on the model parameters, B&S, referring Merton (1973), say that the option value increases continuously as any one of maturity t^*, interest rate *r*, or variance rate v^2 increases approaching the maximum value equal to the stock price.

B&S pay tributes to previous studies in warrant pricing (Sprenkle, 1961; Samuelson, 1965; Samuelson & Merton, 1969) and explain why those studies have failed to achieve satisfactory solutions. B&S indicate that the warrant analysis is much more complicated than the analysis of vanilla options because the warrant life is typically much longer (years) than the option life (a few months). Therefore, the warrant valuation must take into account many events that can happen during the warrant lifetime, such as dividend payments, changes in the strike price on specified dates, etc. However, B&S see the main effect of the warrant long life in substantial changes in the variance rate rather than in payments of the underlying firm.

B&S give examples of long-period business situations where, as they believe, the option pricing formula will successfully work for pricing long-term securities. At the same time, they understand that more complicated cases, when a company issues callable or convertible bonds, cannot be handled by using the option pricing equations (Eqs. (3.2) or (3.3)) because of significant changes in the variance rate of returns of the underlying firm during their long maturity. The EMM proves that the primary cause of the ineffectiveness of the BS formula for the valuation of long-term assets and securities is the development of distribution skewness caused by the firm's BSE payments and the default line.

Simultaneously with Black and Scholes, Merton (1973) published the most important paper in the option pricing literature. Paying tributes to the merits of the BS formula (the option price depends on observable parameters except for the market volatility), he remarks that "although their (B&S) derivation of the (option pricing) formula is intuitively appealing, such an important result deserves a rigorous derivation. … The rigorous derivation … gives insight into necessary conditions for the formula to obtain" (Merton, 1973, p. 161). Merton gives his alternative derivation of the BSM model with slightly relaxed assumptions.

(1) "Frictionless" markets: there are no transactions, costs, or differential taxes. Trading takes place continuously and borrowing and short-selling are allowed without restrictions. At that, the borrowing rate equals the lending rate.

(2) Stock price dynamics is described by the stochastic differential equation:

$$dS/S = \alpha dt + \sigma dz \quad (3.5)$$

where α is the instantaneous expected return on the common stock, σ^2 is the instantaneous variance of the return (non-stochastic, known function of time), and *dz* is a standard Gauss-Wiener process.

(3) Let P(τ) be the price of a riskless (in terms of default) discounted loan ("bond") which pays one dollar τ years from now. If the current and future interest rates are positive, then

$$1 = P(0) > P(\tau_1) > P(\tau_2) > \cdots >> P(\tau_n) \text{ for } 0 < \tau_1 < \tau_2 < \cdots < \tau_n$$

at a given point in calendar time. The dynamics of bond returns are described by the equation:

$$dP/P = \mu(\tau)dt + \delta(\tau)dq(t;\tau) \quad (3.6)$$

where μ is the instantaneous expected return, δ^2 is the instantaneous variance, and $dq(t;\tau)$ is a standard Gauss-Wiener process for maturity τ. At that process dq for one maturity τ is not perfectly correlated with process dq for another maturity T:

$$dq(t;\tau)dq(t;T) = \rho_{\tau T}dt \quad (3.7)$$

but $dq(s;\tau)dq(t;T) = 0$ for $s \neq t$, and $dq(s;\tau)dz(t) = 0$ for $s \neq t$.

Expected return μ may be stochastic, but instantaneous variance δ^2 is non-stochastic and independent of the level of P.

(4) All investors agree on the values of σ and δ, and on the distributional characteristics of dz and dq. It is not assumed that they agree on either α or μ.

(5) It is assumed the option price H is a function of the stock price S, riskless bond price P, exercise price E, and the length of time to expiration τ: $H(S, P, \tau; E)$.

Using Ito's lemma, Merton writes an equation for changes in the option price:

$$dH = \beta H dt + \gamma H dz + \vartheta H dq$$
$$\beta = (0.5\sigma^2 H_{11} + \rho\sigma\delta SPH_{12} + 0.5\delta^2 H_{22} + \alpha SH_1 + PSH_2 - H_3)/H \quad (3.8)$$
$$\gamma = \sigma SH_1/H,\ \vartheta = \delta PH_2/H,\ H_1 = \frac{\partial H}{\partial S},\ H_2 = \frac{\partial H}{\partial P},\ H_3 = \frac{\partial H}{\partial t}$$

Then Merton forms a portfolio containing the common stock, the option, and riskless bonds with times to maturity equal to the option expiration τ, such that the aggregate investment in the portfolio is zero. This is achieved by using the proceeds of short sales and borrowing to finance long positions. Merton says nothing about how long the investor has to keep and watch this portfolio, but this time, t_p, matters. To estimate the option price, one must keep and watch the portfolio for the time comparable with the characteristic time of changes in the option price, $t_p \leq \tau$, because the portfolio dynamics must reflect the option price dynamics. During this time, the firm, which issued the underlying stock, makes no payments or makes a few payments, and its return distribution will change insignificantly, remaining almost the same. (The characteristic time of changes in the firm's returns is much greater than the characteristic time of changes in the option price). Correspondently, the expected stock price and the stock price distribution remain about the same. If W_1 is the number of dollars invested in the common stock, W_2 is the number of dollars invested in the option, and W_3 is the number of dollars invested in bonds, then $W_1 + W_2 + W_3 = 0$. For the instantaneous dollar return to the portfolio, dY, one has a relation:

$$dY = W_1\frac{dS}{S} + W_2\frac{dH}{H} + W_3\frac{dP}{P} = \quad (3.9)$$

$$[W_1(\alpha - \mu) + W_2(\beta - \mu)]dt + [W_1\sigma + W_2\gamma]dz + [W_2\vartheta - (W_1 + W_2)\delta]dq$$

Merton supposes that a hedging strategy exists ($W_j = W_j^*, j = 1, 2, 3$) making the coefficients at dz and dq always zero and the returns on that portfolio dY^* non-stochastic. Together with a zero return to the regular part of the portfolio change, it results in a linear system:

$$W_1^*(\alpha - \mu) + W_2^*(\beta - \mu) = 0$$
$$W_1^*\sigma + W_2^*\gamma = 0 \quad (3.10)$$
$$-W_1^*\delta + W_2^*(\vartheta - \delta) = 0$$

A non-trivial solution of this homogeneous system exists if and only if:

$$\frac{\beta-\mu}{\alpha-\mu}=\frac{\gamma}{\sigma}=\frac{\vartheta-\delta}{\delta} \tag{3.11}$$

Using relation (3.11) and definitions (3.8) of $\beta,\ \gamma$, and ϑ, Merton achieves the equation for the option price:

$$0.5(\sigma^2S^2H_{11}+2\rho\sigma SPH_{12}+\delta^2P^2H_{22})-H_3=0 \tag{3.12a}$$

The price $H(S, P, \tau; E)$ must also satisfy the boundary conditions:

$$H(0,P,\tau;E)=0 \tag{3.12b}$$

$$H(S,1,0;E)=\max(0,S-E) \tag{3.12c}$$

The existence of a strategy making the coefficients at dz and dq zero means that the returns to portfolio Y have the *martingale* property. We will call (3.10) the conditions of *martingale no arbitrage* (or *martingale arbitraging*). The portfolio is the martingale arbitraging one if and only if it is self-financing, meaning that the firm standing behind the portfolio makes no payments and receives no fund infusions. The spirit of self-financing imbues the entire Merton paper (see, for example, Eqs. (3.5) and (3.6)), and conditions (3.10) seem very natural to him. Merton concludes his derivation of (Eq. (3.12)) with the statement (p. 168) "We have derived the BS warrant pricing formula rigorously under assumptions weaker than they (B&S) postulate." One must remember that the formula for the long-term warrant pricing is derived for the market where the firms can default only at the debt maturity. In a real market where the firm pays its BSEs and can default at any moment causing the default of all its securities, the situation is quite different. We shall consider the case of actual market a bit later.

Next, Merton tries to extent the BSM model by including dividend payments. He analyzes the effect of dividends D on unprotected warrants in the case of a known and constant interest rate r. Using the same conditions (3.10), he comes to the equation for the warrant price $W(S, \tau; E)$:

$$0.5\sigma^2S^2W_{11}+(rS-D)W_1-W_2-rW=0 \tag{3.13}$$

subject to the boundary conditions, $W(0, \tau; E) = 0$, $W(S, 0; E) = \max(0, S - E)$ for a European warrant, and to additional arbitrage boundary condition $W(S, \tau; E) \geq \max(0, S - E)$ for an American warrant. Inclusion of dividend payments violates the self-financing condition which is necessary for the BSM model, and only practical usefulness can excuse this inconsistency. But the pricing model for a long-term warrant to be useful must take account of many requirements not included in the BSM model; therefore, model (3.13) and all inferences from it are, unfortunately, of a little value.

To make the problem setting more realistic, one must take into consideration the firm's payments and the default line introduced by Black and Cox (1976). It changes the situation drastically: Equations (3.12) may be good for short-term options, but they do not work for long-term warrants. To understand what conditions must be set for the hedging strategy ($W_j^*, j=1,2,3$), one should remember that the firm's market value develops due to two interrelated stochastic processes. The first and relatively slow one is the corporate asset development during the manufacturing and marketing goods by the firm. This process includes multiple payments and infusions of funds and is definitely not a self-financing process. It involves the growing distribution skewness and negative mean stochastic returns. The faster second process sets the market price on the firm's stock. We can prove (and will show it in a later paper) that this process follows the GBM if the firm asset value does not change. Thus, the stock market price has a symmetric return distribution, and a zero mean stochastic return. Conditions (3.10) are the conditions of hedging portfolio dY against the trading fluctuations. The cumulative effect of both processes makes the firm return distribution skewed and the mean stochastic return negative when pricing long-term assets and securities. At that, the managers controlling portfolio dY will see the coefficients at dz and dq to be close to zero at any moment because the firm mean returns decrease slowly.

The greater the payments and the longer their period, the sharper the effect of payments. Therefore, for short-term options, conditions (3.10) are almost fulfilled, but for long-term securities like warrants, shares of stock, bonds, etc. when the portfolio must be kept and watched for the times comparable with times of the firm's asset changes, conditions (3.10) are never satisfied. It explains the relative success of the BS option pricing formula and failure of the Merton formula for warrant pricing and his general equation for pricing any security (Merton, 1974; Section 4). The EMM (Shemetov, 2020) proves that the main cause making the BS formula ineffective for valuation of long-term securities is growing skewness in the firm's return distribution caused by BSE payments and the default line. The actual situation with the long-period hedging portfolio is

$$
\begin{gathered}
W_1^*(\alpha-\mu)+W_2^*(\beta-\mu)<0 \\
W_1^*\sigma+W_2^*\gamma<0 \\
-W_1^*\delta+W_2^*(\vartheta-\delta)<0
\end{gathered} \tag{3.14}
$$

We will call conditions (3.14) the *non-martingale* no arbitrage conditions to distinguish them from *martingale* no arbitrage conditions (3.10). The principal difference between these two no-arbitrage conditions is that the firm's return distribution satisfying conditions (3.10) and starting from a symmetric initial distribution remains always symmetric while the firm's return distribution meeting conditions (3.14) and starting from a symmetric initial distribution becomes more and more skewed over time. An important consequence is that the martingale arbitraging keeps the initial portfolio value safe, while the non-martingale arbitraging erodes (decreases) the initial portfolio value the more, the higher the payments and the longer their period. If $X = (X_n)$ is a martingale relative to filtration $F = (\mathcal{F}_n)_{n\geq 0}$:

$$E(X_n|\mathcal{F}_{n-1}) = X_{n-1}$$

and $X_n = x_1 + x_2 + \ldots + x_n$, $x_0 = 0$, then $x = (x_n)$ is a martingale difference with properties: x_n is $\mathcal{F}_n$—measurable, $E|x_n| < \infty$, and $E(x_n|\mathcal{F}_{n-1}) = 0$ (compare with conditions 3.10). On the other hand, if $X = (X_n)$ is a stochastic sequence with filtration $F = (\mathcal{F}_n)_{n\geq 0}$, $X_n = x_1 + x_2 + \ldots + x_n$, $x_0 = 0$, and $x = (x_n)$ has properties: x_n is $\mathcal{F}_n$—measurable, $E|x_n| < \infty$, and $E(x_n|\mathcal{F}_{n-1}) < 0$ (see conditions 3.14), then

$$E(X_n|\mathcal{F}_{n-1}) = X_{n-1} - \epsilon, \epsilon > 0$$

The value of ϵ depends on time and parameters of the firm and market: for short-term options $\epsilon \cong 0$, for stocks or bonds of unstable firms it can be fairly large. Without interventions from the outside, ε-value grows continuously. The arbitrage property of financial markets means impossibility to make money out of nothing. Formally, we say that a self-financing strategy provides an arbitrage opportunity if and only if $P(V_0 = 0) = 1$, $P(V_T \geq 0) = 1$ and $0 < P(V_T > 0) < 1$, here V_0 is an initial value of the portfolio, V_T is its value at time $T > 0$, $P(A)$ is a probability of event A. A market is (martingale) arbitrage free if there is no such strategy. For the firm (portfolio) paying its BSEs, if $P(V_0 = 0) = 1$, then $P(V_T < 0) = 1$. The market of the firms paying their BSEs is arbitrage free. To succeed in the non-martingale no arbitrage market, one must work hard; single speculation for sufficiently long time will certainly ruin one's wealth.

The system (3.14) cannot be solved by the way used by Merton; thus, Equation (3.12a) is inconsistent with the default line and cannot be used for pricing long-term securities like warrants, bonds, stocks, etc. Equation

(3.12a) can be used for pricing short-term securities in the presence of default line, but in this case, it converts into the BS equation. Equation (3.12a) can be used without any limitation for pricing any securities but in the artificial market, where the firm and its securities can default only at the maturity of the firm's debt.

Merton concludes his paper optimistically:

> As suggested by Black and Scholes and Merton, the model can be used to price the various elements of the firm's capital structure. Essentially, under conditions when the Modigliani-Miller theorem obtains, we can use the total value of the firm as a "basic" security and the individual securities within the capital structure (e.g. debt, convertible bonds, common stock, etc.) can be viewed as "options" or "contingent claims" on the firm and priced accordingly. (Merton, 1973, p. 178)

We would never mention this statement claiming too much generality to a particular problem of pricing short-term options if it were just evidence from the past. However, this statement has become a slogan aggressively proclaimed till now (e.g., Strebulaev & Whited, 2012, pp. 4-5; Sundaresan, 2013, p. 21), and the models based on the BSM model still appear in financial studies (e.g., Leland, 1994; Heston, 1993; Bates, 1996; Goldstein et al., 2001; Titman & Tsyplakov, 2007; Hugonnier et al., 2015) making sometimes bright but always practically useless contributions to financial literature.

Empiric Testing of BSM Model

Investigations of how well the BSM model works have started soon after the publication (Black & Scholes, 1973). We consider assessments of the model's validity following (Fortune, 1996), who compares the model's predictions with historical data to determine whether the predictions are accurate or not. For this study, he uses the data for the 1992-1994 period from the Chicago Board Options Exchange's Market Data Retrieval System. The author takes the S&P 500-stock index (SPX) as an underlying asset. The MDR reports the number of contracts traded, the time of the transaction, the premium paid, the option characteristics (put or call, expiration date, strike price), and the price of the underlying stock at its last trade. One can find the details of the data under study in the original paper.

The key unobservable parameter in the BSM model is volatility of returns of the underlying asset (SPX). The BSM model assumes that investors know the true standard deviation of returns over the term of the option. However, actual volatility is an unobservable variable whose estimate can be inferred post factum from the option premia observations. Measuring the option's stock price (S), its strike price (E), the remaining life of the option (τ - t), the riskless interest rate (r), and the dividend yield (q), one can estimate "implied volatility" initially suggested by Latane and Rendleman (1976). Implied volatility makes the predicted premium exactly equal to the actual premium.

Implied volatility reveals several problems in the BSM model**.** First, the average forecast error (actual volatility less implied volatility) for the empirical data is (-0.7283), with a t-statistic of (-8.22). It means that implied volatility is a biased estimate of actual volatility. Another problem concerns the residuals of the implied volatility regressed on the forecasting variables. For a good forecast, its residuals (the forecast errors) should not depend on any information available at the time the forecast is made. The residuals must be random and uncorrelated with the information available before the forecast. However, the author reports that F-statistic for the significance of the regression coefficients is 4.20 with a significance level of 0.2% and concludes that this is strong evidence of violation of the residual information test.

Another inference of the BSM model is that put options and call options identical in all aspects should have the same implied volatility and trade at the same premium. It follows from the arbitrage that enforces put-call parity. For each trading day in the 1992-1994 period, the differences between implied volatilities for the at-the-money puts and calls having the same expiration dates are computed. The results show that although puts sometimes have implied volatility less than calls, the norm is for the higher implied volatility and price for puts. So, the experimental data cast a shadow of doubts on the put-call parity.

An important factor affecting the option price is the price distribution of an underlying asset. B&S take the stock price distribution lognormal, but Simkowitz and Beedles (1980), Singleton and Wingender (1986), Badrinath and Chatterjee (1988), Fortune (1996), Harvey and Siddique (2000) argue that the distribution of changes in the logarithm of stock prices is skewed and has fatter tails than that of the normal distribution. Fortune (1996) takes a study of daily changes in the logarithm of the S&P 500 index for the period of 1980-1995. He reports the statistics for the pre-1987 crash period (January 2, 1980 to September 30, 1987), for the post-crash period (January 4, 1988 to March 31, 1995), and for the entire period (January 2, 1980 to March 31, 1995). The statistics for the stock returns during contiguous trading days presented by Fortune (1996, pp. 32-33): the mean return (percentage at annual rate) for the entire period is 21.48, for the pre-crash period 32.65, and for the post-crash period 8.29. The standard deviation of returns for the entire period is 12.58, for the pre-crash period 12.40, and for the post-crash period 11.27. Skewness of the return distribution for the entire period is 0.07, for the pre-crash period 0.21, and for the post-crash period -0.94. Kurtosis of the return distribution for the entire period is 7.33, for the pre-crash period 1.80, and for the post-crash period 8.32. Surprisingly that daily volatility of the SPX measured by standard deviation remains about the same both for the pre-crash period and the post-crash period, that skewness of daily returns changes from a positive value before the crash to a negative value after the crash, and that kurtosis of daily returns increases from a relatively low value before the crash to a high value after the crash. The relative frequency distribution of SPX returns does not fit the normal distribution of returns but shows some skewness.

Fortune explains the violation of the call-put parity by friction mechanisms involved in security trading, such as the uptick rule, premature termination of a short position, fees for stock lending, etc. He explains the non-normality of stock price changes by occasional shocks hitting stock prices, following the jump-diffusion theory of the development of the fat-tailed return distribution suggested by Merton (1976). Fortune makes no comment on the bias in the volatility forecast, on failure of the residual information test, on a surprising change of skewness in the crash of 1987. Here we offer our explanations to the observed phenomena.

The failure of the residual information test looks very natural because the BSM model does not use all information on the development of volatility, taking account of the diffusion expansion of the return distribution and missing the distribution distortion as a result of the firm’s payments.

The first question about the observed bias in the volatility forecast (actual volatility less implied volatility equals (-0.7283) is why the implied volatility is so higher than the actual volatility. Is it specific for the options of the chosen underlying asset (the S&P 500 Index) or timing? Fortune thinks this error is due to shortcomings of the BSM model, and we agree with him. The BSM model supposes that the return distribution on an underlying asset remains normal all the time with its variance growing over time with the same variance rate. However, the return variance of the firms paying BSEs grows due to a symmetric diffusion expansion with a given variance rate and a distortion of the return distribution due to developing skewness.

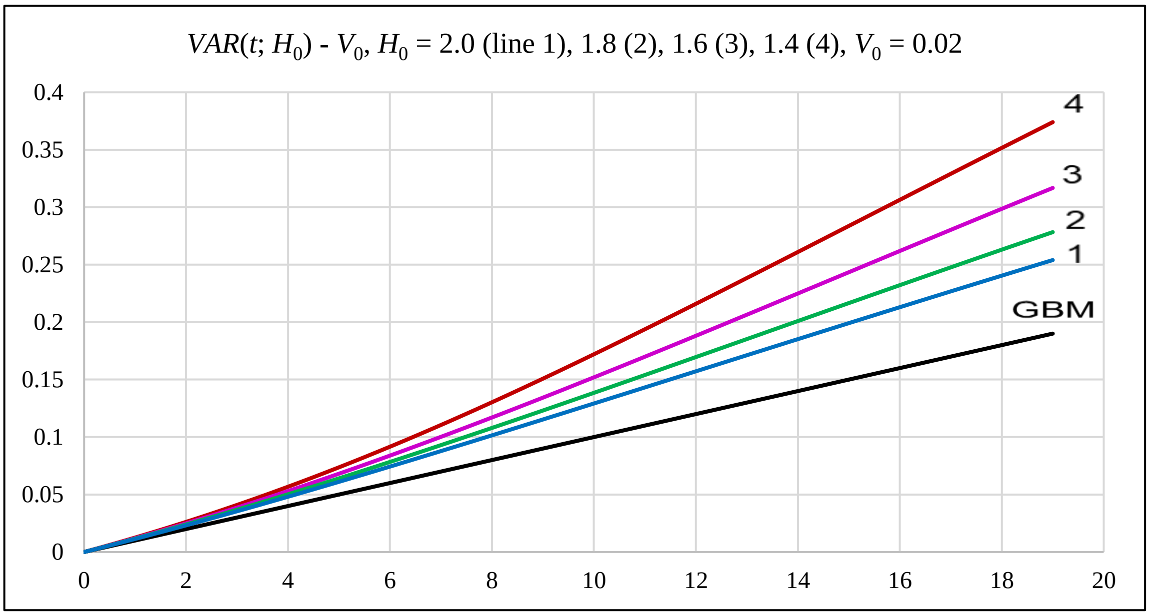

Figure 1. Variance $VAR(t; H_0)$ - VAR_0 as a function of time (years) for various initial conditions H_0 = 2.0 (line 1), 1.8 (line 2), 1.6 (line 3), 1.4 (line 4), and $VAR(GBM)$ - VAR_0 = C^2t (line GBM); VAR_0 = 0.02.

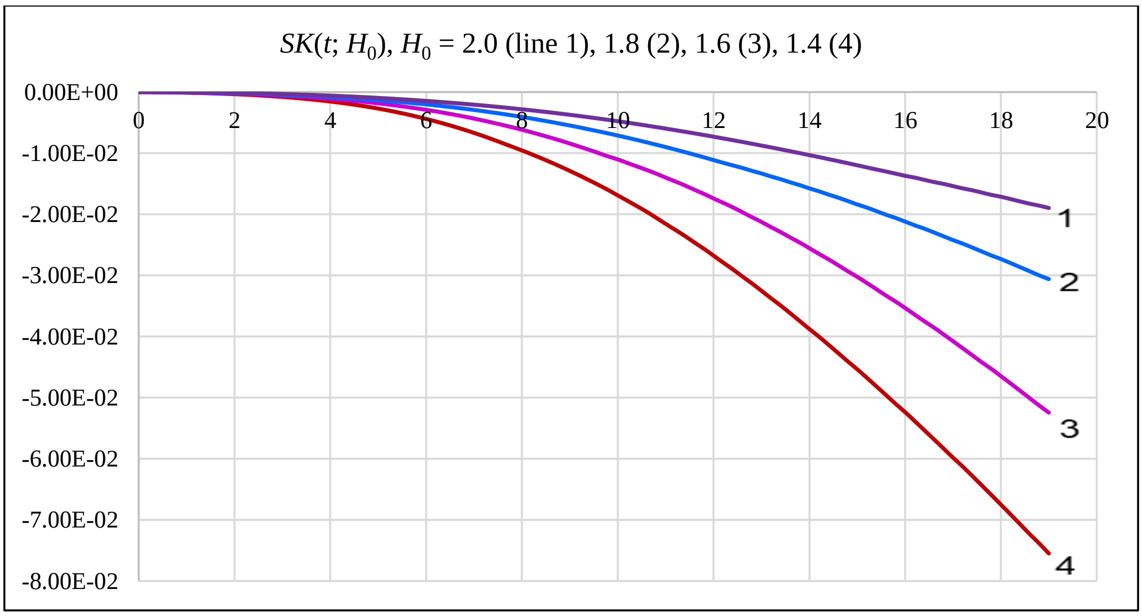

Figure 2. Skewness $SK(t, H_0)$ as a function of time (years) for various initial conditions H_0 = 2.0 (line 1), 1.8 (line 2), 1.6 (line 3), and 1.4 (line 4).

Here we present typical pictures of the firm's return distributions with statistical moments. The moments $VAR(t)$ and $SK(t)$ are defined by Equations (i.6), $DPR(t)$ by Equation (i.4).

Figure 1 shows the variance development caused by pure diffusion (the line GBM; the volatility rate squared is 0.01, $VAR(t)$ - VAR_0 = C^2t), and four variance lines generated by the EMM for different values of parameter $H_0 = \langle x(0)\rangle = \langle \ln[RX(0)/P_0]\rangle$, here R = 0.10 is the expected rate of return on firm's assets, P_0 is the rate of firm's payments, $X(0)$ is the initial distribution of the firm's assets (normal).

Parameter H_0 shows the height of the firm's mean relative returns over the default line (in this example, *DL* = 0, no debt) and reflects the firm stability: the lesser H_0, the faster grows the default probability in a given time interval. The firms with H_0 = 2.0, 1.8, 1.6, and 1.4 are steady firms. See details in Shemetov (2020; 2022). The difference between the EMM-variance and the GBM-variance appears due to the effect of payments on the distribution distortion. It is clear that to achieve the same variance value in the BS model, one must use a greater variance rate:

$$C_{eff}(t) = \sqrt{\frac{dVAR}{dt}} > C$$

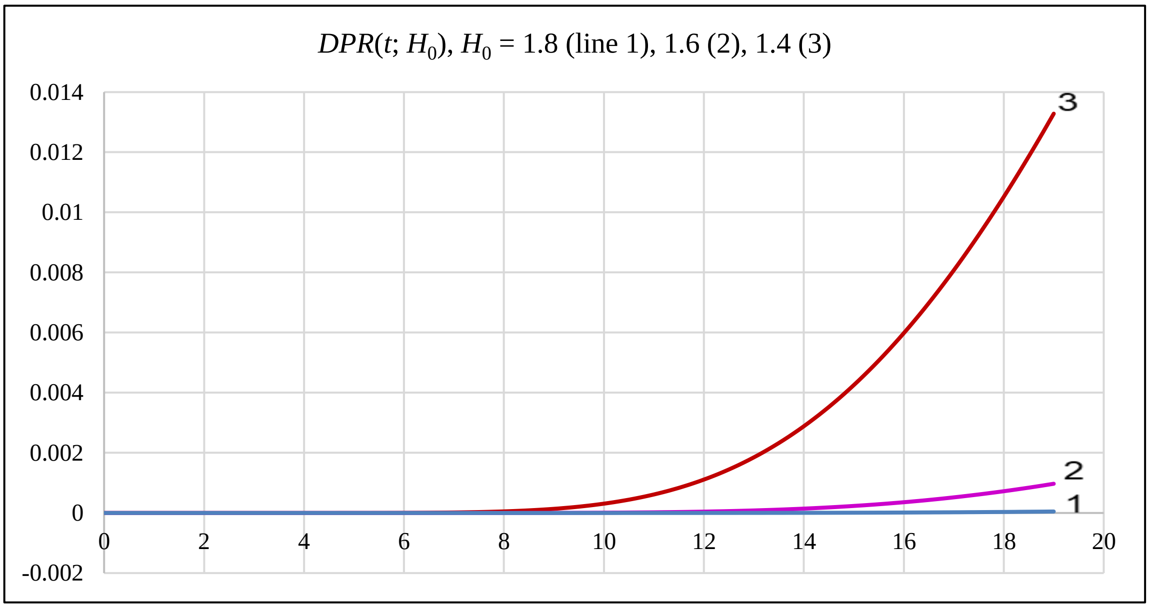

Figure 3. Default probability $DPR(t; H_0)$ as a function of time (years) for various initial conditions H_0 = 1.8 (line 1), 1.6 (2), and 1.4 (3); VAR_0 = 0.02.

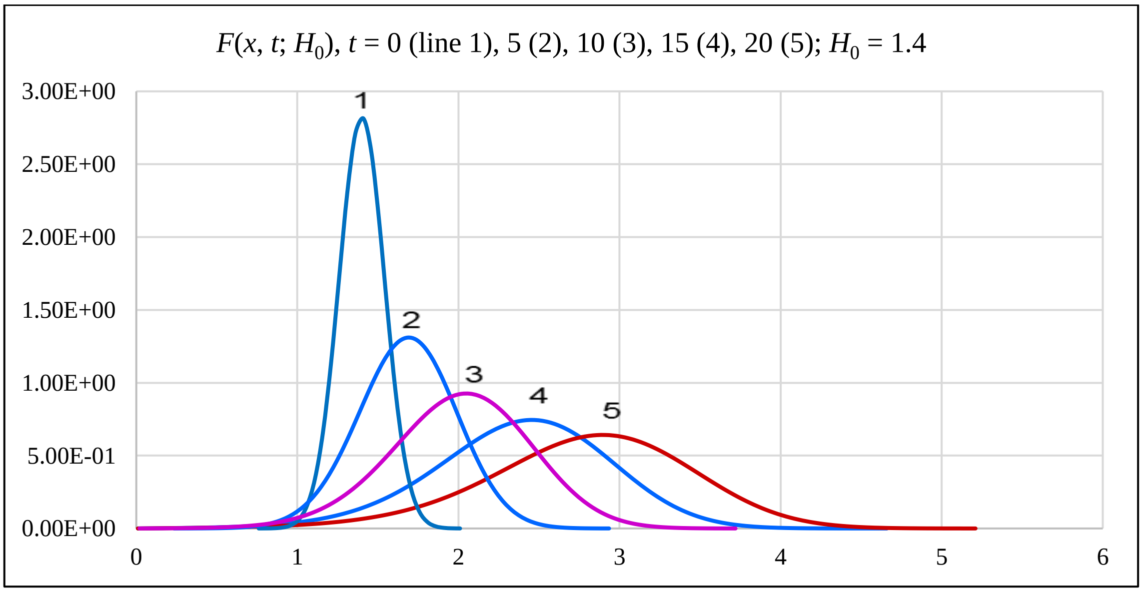

Figure 4. Evolution of the return distribution $F(x, t; H_0)$ for H_0 = 1.4 and t = 0 (line 1), 5 (line 2), 10 (3), 15 (4), 20 (5); time is measured in years, VAR_0 = 0.02. Mind the development of negative skewness.

It explains the negative difference between the EMM variance rate and the BSM variance rate for implied volatility and shows that the EMM variance rate is closer to the actual rate. Concerning the question what will happen to the bias of the volatility forecast if one takes another set of firms instead of the SPX, we can assume that the bias should increase because the top S&P 500 firms are expected to be more stable (have greater H_0) compared to the firms from the middle of the exchange list, but of course, it must be verified.

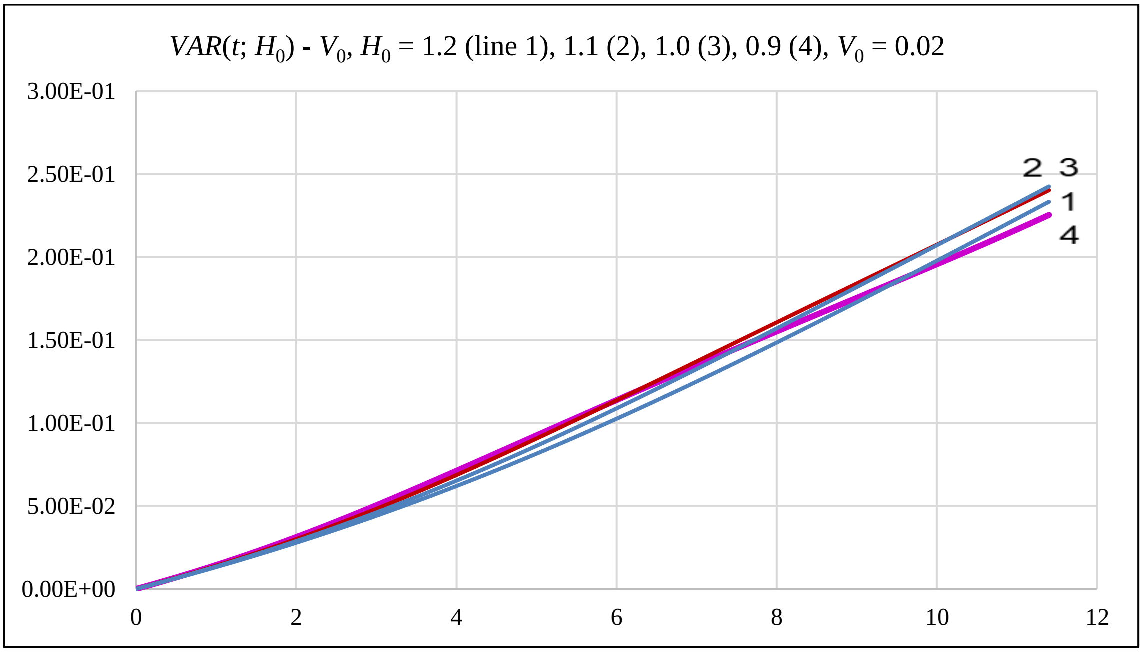

Figure 5. Variance $VAR(t; H_0)$ - VAR_0 as a function of time (years) for various initial conditions H_0 = 1.2 (line 1), 1.1 (line 2), 1.0 (line 3), and 0.9 (line 4); VAR_0 = 0.02.

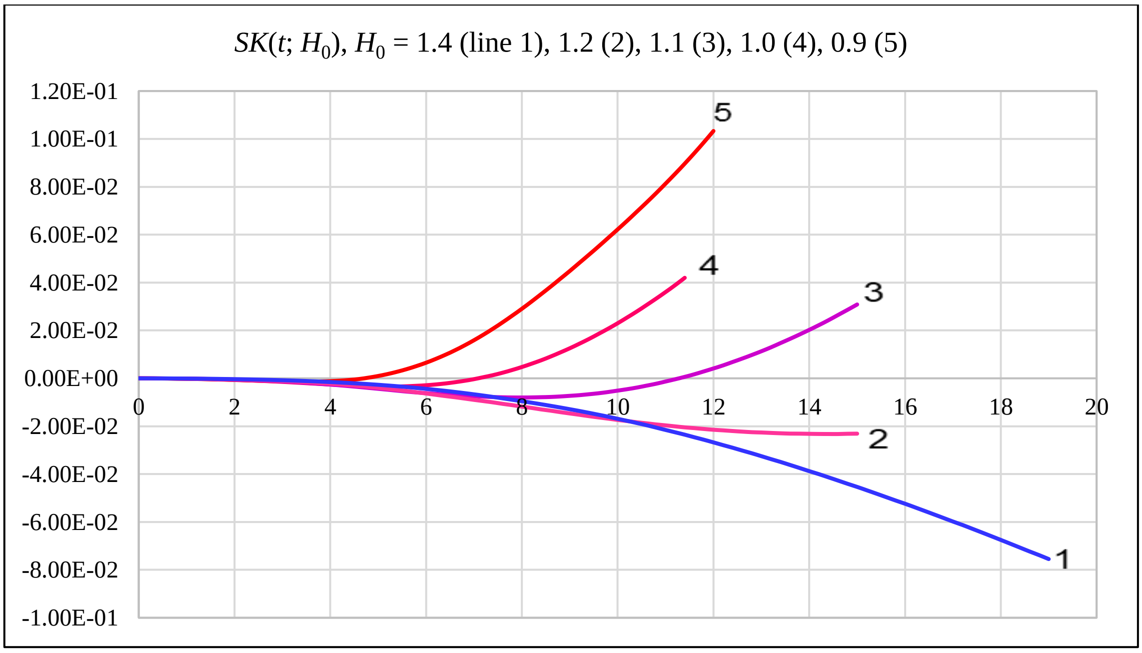

Figure 6. Skewness $SK(t; H_0)$ as a function of time (years) for various initial conditions H_0 = 1.4 (line 1), 1.2 (line 2), 1.1 (line 3), 1.0 (line 4), and 0.9 (line 5); VAR_0 = 0.02.

Figure 2 demonstrates that steady firms (H_0 = 2.0, 1.8, 1.6, 1.4) have low to moderate negative skewness and a low default probability (Figure 3); Figure 4 shows the development of a negatively skewed distribution from an initially symmetric (normal) distribution of relative returns for a steady firm.

Figure 6 displays that for less steady firms (H_0 = 1.2, 1.1, 1.0, 0.9), their skewness changes from negative to positive values and their default probability dangerously increases (Figure 7); the lesser H_0, the greater the default probability. Figure 8 illustrates the development of a positively skewed distribution from an initially symmetric distribution.

Now we are ready to explain the transition of the cumulative return distribution of the top S&P 500 from positive skewness (0.21 before the market crash) to negative (-0.94 after the crash). The high positive skewness indicates that before the crash, there were a lot of unstable firms among the top S&P 500 (Figure 6). The negative skewness is specific for more steady firms (Figure 2). After the crash, unstable firms drop out from the top S&P 500, and new, more steady firms, who survived the crash, get in the top S&P 500. The great concern causes the cumulative skewness of the top S&P 500 close to zero (0.07) over the entire period of observation, giving evidence that even in a more or less "normal environment" there are a lot of firms of low stability among the top S&P 500.

Figures 1 and 5 display clearly that variance can serve as a measure of volatility for steady firms only. For unstable positively skewed firms, variance is no good measure of volatility: the variances of the unsteady firms collect themselves in a tight bunch (Figure 5), while their default probabilities vary significantly (Figure 7). These pictures explain the paradoxical fact marked by Fortune that the volatility represented by variance does not change significantly in the crash: 12.40 before the crash and 11.27 after, while the true volatility characterizing the default probability is much higher before the crash relaxing somewhat after it. Shemetov (2022) shows that the true measure of volatility and risk is the intensity of default probability, but it is unobservable variable for which one needs a good observable proxy to obtain.

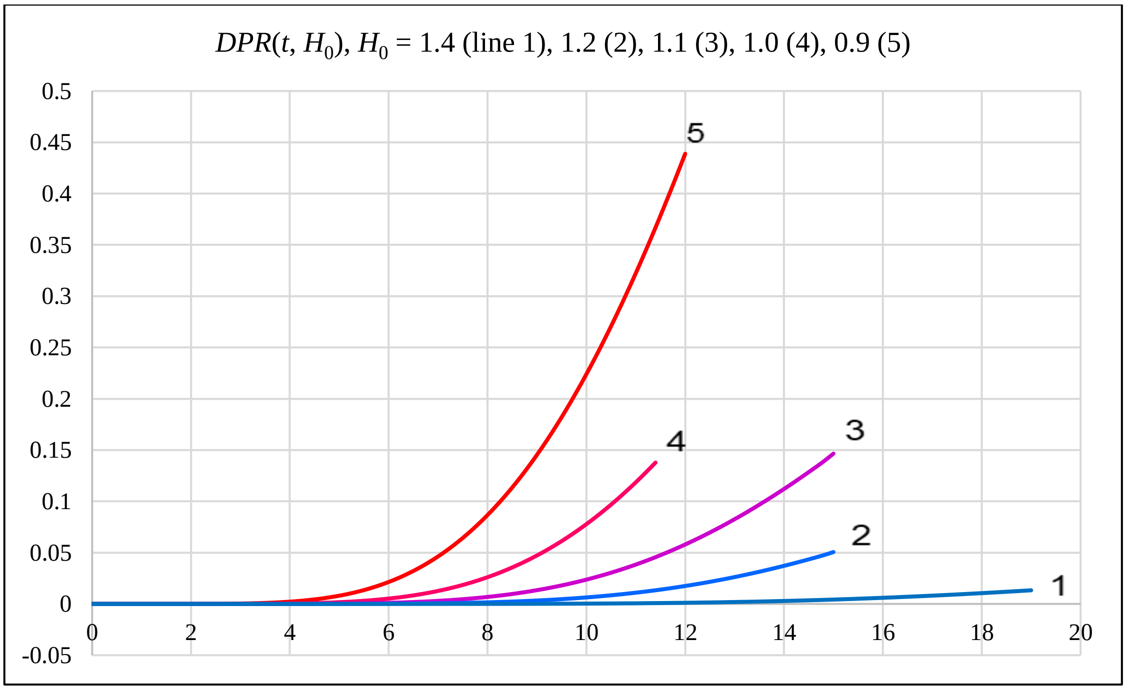

Figure 7. Default probability $DPR(t; H_0)$ as a function of time (years) for various initial conditions H_0 = 1.4 (line 1), 1.2 (line 2), 1.1 (line 3), 1.0 (line 4), and 0.9 (line 5); VAR_0 = 0.02.

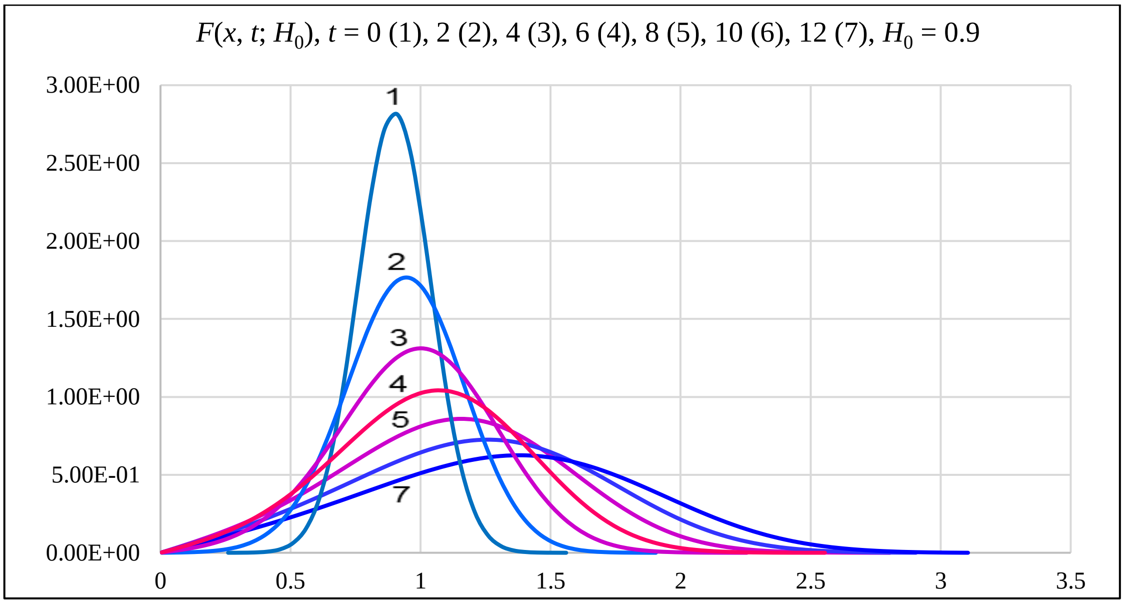

Figure 8. Evolution of the return distribution $F(x, t; H_0)$ for $H_0 = 0.9$ and $t = 0$ (line 1), 2 (line 2), 4 (line 3), 6 (line 4), 8 (line 5), 10 (line 6), 12 (line 7); $VAR_0 = 0.02$. Mind the development of positive skewness.

These facts explain the shortcomings of the S&P 500 Volatility Index (VIX), developed to be an efficient predictor of risk. Since the crisis of 2008-2009, the VIX is published in financial news along with stock market indices. It is considered as an overall barometer of the market risk weather (Kownatzki, 2016). The VIX concept uses a weighted average of implied volatility to find the variance rate as a measure of risk. The author asks how good the VIX is as a predictor of market risks. His empirical study embraces the period of January 1990 until December 2014 and traces the VIX moments (the mean, standard deviation, skewness, and kurtosis) and percentile ranges for daily closing levels, correlation of concurrent levels of the VIX and actual volatility, etc. Kownatzki compares the VIX moments with the moments of actual volatility and concludes that the VIX is unsuitable for many risk management applications because the VIX systematically overestimates actual volatility in non-crisis periods (compare it with the negative bias in volatility estimates reported by Fortune) and underestimates actual volatility in times of financial crises (because variance is no measure for volatility of unsteady firms, Figures 5 and 7). The author's empirical conclusions and our theoretical results mutually support each other. It seems natural that the next S&P 500 volatility index will be founded on the EMM as a more precise and accurate model for estimating market volatility.

Figures 2 and 6 demonstrate that skewness of the return distribution is a natural property of the firms paying their BSEs. There is no need to patch the BSM model introducing poorly motivated jumps in the GBM to achieve the desired skewness. One should remember that jump statistics must change over time following a special law which poses additional difficulties for jump-diffusion processes (JDPs) used to describe the return distributions of real firms.

Fortune describes a relative frequency distribution for the top S&P 500 as "roughly symmetric (with signs of skewness), and very leptokurtic with minor fat tails" (Fortune, 1996, p. 36), supporting earlier findings of Mandelbrot (1963) and Fama (1965). The author explains the phenomenon by jumps in the stock prices. However, the frequency distribution of the S&P 500 demonstrates significant changes in skewness from positive before the market crash in 1987 to negative after it. This fast change in the direction of random jumps has no explanation in the frameworks of the JDP model of stock prices. Figures 2 and 6 show that the firm's return distribution could be skewed negatively or positively, and zero skewness can appear only for a short time. Therefore, the rough symmetry of the S&P 500 distribution of the logarithm of stock prices can be explained by "diversification" of positively and negatively skewed distributions of individual firms in the top S&P 500. Before the 1987-crash, the number of positively skewed firms in the top S&P 500 prevails, provoking market instability. After the crash, when the unstable firms dropped out from the top, the top S&P 500 consist mostly of negatively skewed firms which survived the crash, and the market starts another cycle of its functioning.

The cumulative return distribution of the top S&P 500 can serve as a barometer that shows the increasing general tension in the market. However, the market crash depends on return distributions of individual firms. All firms combine into business networks through their business relations; no one is by itself. On the eve of a crash, many firms in the market have noticeably skewed return distributions of low stability. When a weak link in a network fails and the firm defaults due to market fluctuations, the tension in the remaining network increases, making the other firms in this network less stable. Another default within the same network, increasing its strain even more, can cause the default in the entire network, and the market crash develops without any visible cause as an avalanche, purging unstable firms from the market.

The statistics presented by Fortune and interpreted in the light of the EMM convincingly demonstrate that the BS formula can predict correct option prices with sufficient precision and accuracy only occasionally because the return distributions even in the top S&P 500 are skewed. It is good if they are *negatively* skewed that guarantees some market stability. It is much worse when the return distributions are *positively* skewed because such firms are in a tensed unstable state, but the traders and public estimate their shares of stock as good and reliable ones judging them by the stock variance.

Progress in Option Pricing After 1973

Since the publication by Black and Scholes (1973) and Merton (1973), the option/warrant valuation theory has made an impressive progress. Researchers react to the shortcomings of the BSM model discussed above by relaxing its rigid assumptions. Merton (1976) considers the problem of option pricing, when the price of an underlying stock is not continuous but can make random jumps, following the Poisson distribution with jump heights distributed normally. Merton derives a closed-form solution for that model, introducing the class of jump-diffusion processes (JDPs) in option pricing. Kou (2002) develops this approach further, suggesting more flexible JDPs whose jumps in the prices of underlying stocks have a Poisson distribution, and jump heights have a double-exponential distribution. Schwartz (1977) first applies numeric methods to solve option/warrant valuation problems for which closed-form solutions are unavailable, like valuation of an American warrant on a stock which pays discrete dividends. Heston (1993) attacks the constant volatility of the BSM model, introducing the class of stochastic volatility models in option pricing. He derives a closed-form solution for a European call option. Bates (1996) makes the next step and suggests a more general class of the option pricing models with stochastic volatility and jumps in the prices of underlying stocks.

At the same time, the variety of options in use in financial practice continuously expands. Although the greatest part of all options traded make European and American options, there exist more complicated options whose pricing is impossible without using numeric methods. We have no intension to give a comprehensive review of exotic options and constrain ourselves with three of them: the Bermuda, Asian, and Swing options. As one knows, holders of European options can exercise their rights at the option expiration date only. Holders of

American options can exercise their rights at any time before the expiration date, and, therefore, they meet the optimal stopping time problem. Holders of Bermuda options can exercise their rights on some specified dates until the expiration, having to solve another optimal stopping problem.

Sometimes the payoff function of an option depends on the path of an underlying asset price and its value at the expiration date. The Asian option is an example of such path-dependent options because its payoff function is determined by the mean price of an underlying asset over a specific time interval (e.g., Kemna & Vorst, 1990; Foufas & Larson, 2008). Asian options are traded mostly on oil products, agricultural commodities, and currencies.

Swing options are commonly used in the energy market. Widely varying demands for energy consumption and limited facilities for its storage make energy prices highly volatile. To control energy expenses, consumers use various financial tools to reduce market risks related to energy volatility, such as forwards, futures, swaps, and swing options. A swing option secures the option holder with an instrument to control the delivery timing and the amount of energy delivered. A swing option contract contains a base load agreement representing a set of forward contracts with different expiry dates. Each forward contract offers a fixed amount of a commodity. At each expiry date, the option holder has the right to purchase an excess amount of the commodity or decrease the base load volume. Thus, the amount of the commodity purchased at a predetermined price (the strike price) by the option holder can swing within a range. If the swing is positive/negative, the option exercised by the holder at an opportunity time is called upswing/downswing. So, the upswing is a buy option, and a downswing is a sell option. In other words, a swing option includes a set of pure forward contracts and a fixed number of exercise rights for buying or selling the commodity. Usual restrictions in the swing option are: (a) the total number of upswings and downswings is limited, (b) any two points of the right exercising are separated by a minimum waiting time, (c) if the overall volume of energy purchased during the option life exceeds a predetermined quantity, the option holder can be penalized. Dahlgren and Korn (2005) have developed a continuous time model for swing option pricing based on the BSM model and dynamic programming. One can find a mathematical framework for swing options in the papers by Carmona and Dayanik (2008), and Carmona and Touzi (2008).

Assessing the progress in theoretical extensions of the BSM model and its numerous applications in the sphere of exotic options, one must remember that all these developments are superstructures on the BSM basement, and, therefore, they inherit all the shortcomings of the BSM model. A lack of understanding of a financial instrument (an option, a share of stock, etc.) and its behavior might create a false feeling of assurance in a dangerous position leading to unexpected losses.

Merton's General Equation for Any Security Pricing

Merton extends the Black-Scholes model believing that "while options are highly specialized and relatively unimportant financial instruments, […], the same basic approach could be applied in developing a pricing theory for corporate liabilities in general" (Merton, 1974, p. 449). Merton claims the development of a general model for pricing any financial instrument whose value can be written as a function of the firm value and time. To demonstrate an application of this universal model to the valuation of a specific security, he selects a zero-coupon bond.

The Black-Scholes-Merton pricing model (BSM) is based on the following assumptions:

(1) There are no transactions costs, taxes, or problems with indivisibilities of assets.

(2) There are a sufficient number of investors with comparable wealth levels so that each investor believes that he can buy and sell as much of an asset as he wants at the market price.

(3) There exists an exchange market for borrowing and lending at the same rate of interest.

(4) Short-sales of all assets, with full use of the proceeds, is allowed.

(5) Trading in assets takes place continuously in time.

(6) The Modigliani-Miller theorem that the value of the firm is invariant to its capital structure is obtained.

(7) The term structure is "flat" and known with certainty. The price of a riskless discount bond which promises a payment of one dollar at time *t* in the future is $P(t) = exp(-rt)$ where *r* is the (instantaneous) riskless rate of interest, the same for all time.

(8) The dynamics for the value of the firm, *V*, through time can be described by a diffusion-type stochastic process with stochastic differential equation:

$$dV = (\alpha V - C)dt + \sigma V dz$$

where α is the instantaneous expected rate of return on the firm per unit time, *C* is the total dollar payout by the firm per unit time to either its shareholders or liabilities-holders, σ^2 is the instantaneous variance rate of returns on the firm per unit time, dz is a standard Gauss-Wiener process.

(9) There exists a security whose market value, *Y*, can be presented as a function of the firm value and time, $Y = F(V, t)$. Merton writes the dynamics of this security's value as:

$$dY = (\alpha_Y Y - C_Y)dt + \sigma_Y Y dz_Y$$

where α_Y is the instantaneous expected rate of return per unit time on this security, C_Y is the dollar payout per unit time to this security, ${\sigma_Y}^2$ is the instantaneous variance rate of returns per unit time, dz_Y is a standard Gauss-Wiener process.

(10) The firm defaults if its value *V* is less than the outstanding debt *D* at the date of debt maturity *T*: $V_T \leq D$. Default is observable only at the debt maturity date since only at that time it is checked whether the firm can discharge its debt or not.

To derive a general model for pricing financial instruments whose value *Y* can be presented as a function of the firm value and time, $Y = F(V, t)$, Merton repeats his technique (1973) of a three-security portfolio containing the firm, the particular security, and riskless debt such that the aggregate investment in the portfolio is zero. The author introduces three variables W_1, W_2, and $W_3 = -(W_1 + W_2)$, where W_1 is the number of dollars of the portfolio invested in the firm, W_2 is the number of dollars invested in the security, and W_3 is the number of dollars invested in the riskless debt. The instantaneous dollar return to the portfolio, *dx*, is then:

$$dx = W_1 \frac{dV + Cdt}{V} + W_2 \frac{dY + C_Y dt}{Y} + W_3 r dt \quad (4.1)$$

$$= [W_1(\alpha - r) + W_2(\alpha_Y - r)]dt + (W_1\sigma + W_2\sigma_Y)dz$$

The author looks for a triple W_1^*, W_2^*, and W_3^* for which the zero-value portfolio is a hedging one:

$$W_1^*\sigma + W_2^*\sigma_Y = 0 \quad (4.2a)$$

$$W_1^*(\alpha - r) + W_1^*(\alpha_Y - r) = 0 \quad (4.2b)$$

A nontrivial solution to this system exists if and only if:

$$\frac{\alpha - r}{\sigma} = \frac{\alpha_Y - r}{\sigma_Y} \quad (4.3)$$

Substituting expressions for α_Y and σ_Y:

$$\alpha_Y Y = \alpha_Y F = F_t + (\alpha V - C)F_V + \frac{1}{2}\sigma^2 V^2 F_{VV} + C_Y \tag{4.4a}$$

$$\sigma_Y Y = \sigma_Y F = \sigma V F_V \tag{4.4b}$$

Merton comes to the equation which must be satisfied by any security whose value can be written as a function of the firm value and time:

$$F_t + (rV - C)F_V + 0.5\sigma^2 V^2 F_{VV} - rF + C_Y = 0 \tag{4.5}$$

The no-arbitrage conditions (Eq. (4.2)) are an exact analogue of the martingale no-arbitrage conditions of (3.10). We remind that Eq. (4.5) is derived in the case when the firm can default at the debt maturity only.

Merton applies Eq. (4.5) for pricing a zero-coupon bond, $F(V,t)$, setting $C_Y = 0$ because there are no coupon payments and $C = 0$ because the firm cannot issue any new senior or equivalent rank claims on the firm nor it can pay cash dividends or do share repurchase prior to maturity date T of debt B. Using time τ measured from the maturity T to a current moment t: $\tau = T - t$, he comes to the equation:

$$-F_\tau + rVF_V + 0.5\sigma^2 V^2 F_{VV} - rF = 0 \tag{4.6a}$$

with the initial and boundary conditions:

$$F(V, \tau = 0) = \min(V, B) \tag{4.6b}$$

$$F(V, \tau) \leq V \tag{4.6c}$$

The author deduces the equation for the value of equity, $f(V,t)$, considering Eq. (4.6a) and the balance $V = F(V,\tau) + f(V,\tau)$:

$$-f_\tau + rVf_V + 0.5\sigma^2 V^2 f_{VV} - rf = 0 \tag{4.7a}$$

with the same boundary condition and the initial condition:

$$f(V, 0) = \max[0, V - B] \tag{4.7b}$$

The problem (Eq. (4.7)) is identical to the Black-Scholes equation for the European call option on a non-dividend-paying common stock where the firm value V in the problem (4.7) corresponds to the stock price and B corresponds to the exercise price. Thus, the equity value is determined as:

$$f(V,\tau) = V\Phi(x_1) - Be^{-r\tau}\Phi(x_2) \tag{4.8a}$$

$$\Phi(x) = (2\pi)^{-1/2}\int_{-\infty}^{x} \exp(-z^2/2)dz \tag{4.8b}$$

$$x_1 = [\ln(V/B) + (r + 0.5\sigma^2)\tau\,]/\sigma\sqrt{\tau} \tag{4.8c}$$

$$x_2 = x_1 - \sigma\sqrt{\tau} \tag{4.8d}$$

Merton argues that precisely these boundary condition specifications for the general equation (Eq. (4.5)) distinguish one security from another (e.g., debt from equity). The idea that the contingent claim on the firm's cash flows is equivalent to a call on the cash flows with the exercise price equal to the face value of the debt becomes dominating in financial economics since that time (e.g., Ross, 1985; Leland, 1994; Strebulaev & Whited, 2012; Sundaresan, 2013).

However, there is a dramatic difference between options, on the one hand, and corporate liabilities, such as stocks or bonds, on the other. The option is a short-term financial instrument whose existence is guaranteed within its expiration period, making typically a few months. The short expiration period makes the option insensitive to slow changes in the state of the underlying firm. The stocks and bonds are long-term instruments which can default at any moment. Considering zero-coupon bonds independently of the firm's state, Merton naturally comes to a result identical to that of Black and Scholes.

Conditions (4.2) describe the case of short-term securities like options when the firm's return does not change. Considering the behavior of long-term objects like a firm or its stocks or bonds in the market with default lines and firms paying their BSEs, one cannot expect that conditions (4.2) will be met at least to some satisfactory degree. The true no-arbitrage conditions for long-term objects are:

$$W_1^*\sigma + W_2^*\sigma_Y < 0 \tag{4.9a}$$

$$W_1^*(\alpha - r) + W_1^*(\alpha_Y - r) < 0 \tag{4.9b}$$

and Equation (4.5) cannot be derived using Merton's logic. It is important to emphasize that the general equation for pricing any security is valid in a special market where the firm can default only at the maturity of its debt. For more realistic markets with the default lines and firms paying their expenses, Equation (4.5) is effective for pricing short-term securities in whose lifetimes the underlying firm makes no payments. In this case, Equation (4.5) reduces to the BS equation. The general Equation (4.5) cannot be applied for pricing the long-term securities in the market with the default line and firm payments.

There are vast empirical data that asset distributions and portfolios composed of those assets are skewed (e.g., Simkowitz & Beedles, 1980; Singleton & Wingender, 1986; Badrinath & Chatterjee, 1988; Fortune, 1996; Harvey & Siddique, 2000), and this skewness is due to payments made by firms in their business activities. Merton needs Eq. (4.2) both for economic and mathematical reasons, but these equations significantly restrict the setting of his original problem ($C = C_Y = 0$) and, consequently, its solution. For $C = C_Y = 0$, Eq. (4.5) becomes exactly the Black-Scholes equation. To understand this result, one must remember that the approach used by Black, Scholes, Merton, Samuelson, and other researches before them is limited with the case of self-financing portfolios. This approach has provided a relative success in the case of short-term, no-payment European call and put options but failed in the case of long-term securities like stocks and bonds, in a life time of which the underlying firm makes multiple payments.

It is interesting that Merton assumes the MMP1 obtains (A6, the firm value is invariant to its capital structure) just for convenience, to simplify the derivation and the structure of his general security pricing equation (Eq. (4.5)). It is clear from our analysis that the MMP1 is consistent with this approach: the MMP1 is true if and only if the firm makes no payments, and in Merton's approach, one has $C = C_Y = 0$ because of Eq. (4.2). No wonder that Merton succeeds in proving the MMP1 theorem with bankruptcy sharing the honor of success with Stiglitz (1969).

Merton just cannot come to the general equation for security pricing other than the Black-Scholes equation using "natural" conditions (Eq. (4.2)). Since that time, the basic model for the firm value, the self-financing geometric Brownian model (GBM):

$$dV/V = \alpha dt + \sigma dz \tag{4.10}$$

has been generally recognized as the model describing various contingent claims instead of Merton's original equation ($C \neq \delta V, 0 \leq \delta < \alpha$):

$$dV = (\alpha V - C)dt + \sigma V dz \tag{A.8}$$

However, the choice that Merton has made between models (A.8) and (4.10) brings a lot of problems. Firstly, it meets an objection from the theory of differential equations. This theory proves that if two differential equations differ in their structures (like Eq. (A.8) and Eq. (4.10)), they must have different solutions, and solutions of (Eq. (A.8)) can never be the same as solutions of (Eq. (4.10)). (Different solutions for each of these two equations come from different boundary conditions reflecting specific features of securities under study.) The contradiction between the theory of differential equations and Merton's conclusion on isomorphic correspondence between almost any corporate liability and options definitely points to an inconsistence in logic of derivation of the general equation for pricing any security.

Secondly, the GBM describes the self-financing firms; this property makes the GBM unsuitable for analysis of real firms paying their taxes, dividends, debts, etc. To reconcile the GBM with the firm's payments, researchers sometimes invent exotic funding sources and methods (see examples in Sections 5 and 6), which, of course, decrease reliability of their conclusions and recommendations.

Thirdly, the GBM generates solutions with normally distributed returns and default probabilities much lesser than probabilities observed in practice. To decrease the difference between theoretical probabilities and probabilities observed in practice, three main classes of heuristic models have been suggested. The first class consists of so-called calibrated models like the Moody's KMV (Bohn, 2006), matching GBM characteristics (e.g., the distance-to-default) with databases of actual defaults. The model is used for estimating default risks at a one-year horizon. Shemetov (2022) presents a detailed discussion of calibrated models and shows that those models might deliver correct estimations only occasionally. The jump-diffusion processes (JDPs) make the second class of heuristic models that bring theoretical default probabilities closer to the observed probabilities. Here the GBM is supplemented with jumps having a Poisson distribution and their lengths having a normal distribution (Merton, 1976) or a double-exponential distribution (Kou, 2002). The endogenous process of developing skewness in the firm's returns is simulated with an exogenous process of jumps in the stock prices. To tune JDP statistics to a particular firm, one must use the data averaged over market (with another calibrated procedure). Therefore, the likelihood of a success in describing correctly the return distribution of a specific firm by the JDP model is very low. The third class embraces the GBMs with stochastic volatility. These models meet the same problem as the JDP models: volatility statistics are unknown and must be extracted from averaged market data.

An outstanding Merton's contribution to the financial risk literature consists of his brilliant analysis of the security pricing problem and the axiomatic Equation (A8) describing the balance of the firm's cash flows in stochastic conditions. The axiomatic approach, verified by extensive scientific observations, experiments, and practice, is widely used in modern science. Merton's Equation (A8), joined with the default line, could make a solid basis for a new period of financial economics, but, alas, the default line appears in financial theory two years later. Following the logic of his problem setting (no default line), the author uses the no-arbitrage principle as the *martingale* no arbitrage, which has confined the space of possible distributions to symmetric ones only. Symmetric distributions are consistent with the MMPs, and keeping the relation of his model with the MMP1 (Assumption 6), Merton does not decrease the generality of his solution. However, the Equation (A8) is consistent with the martingale no-arbitrage principle and the MMPs for zero payments only. In a difficult choice between the martingale no arbitrage and Equation (A8), Merton prefers to reduce his model to the GBM. Under pressing of Merton's crystal logic, economists have recognized the GBM as an adequate model for the financial risk analysis. The undisputable merits of the GBM are that it often helps to achieve intuitively clear closed-form solutions, and that GBM problems can be analyzed using risk-neutral approach, which significantly simplifies risk analysis. These merits appeal to the majority of economists, and the GBM soon becomes the dominating model for estimating financial risks.

Default Line Appears in Asset Pricing

Black and Cox (1976) continue the development of bond valuation models started by Merton (1974). They argue that Merton's assumption that a firm can only default at debt maturity is far from reality. They suggest a threshold, triggering default when the firm's assets hit the threshold line (the default line). Black and Cox (hereafter B&C) repeat all assumptions made by Merton and build up their model using the general equation for security pricing (Merton, 1974):

$$F_t + (rV - P)F_V + 0.5\sigma^2 V^2 F_{VV} - rF + Q = 0 \tag{5.1}$$

here *F* is a generic label for any of the firm's securities, *V* is the firm value, *P*(*V*, *t*) is the net total payout made or inflow received per unit of time, *Q*(*V*, *t*) is the payout received or payment made by security *F*, *r* is the expected interest rate, σ^2—the instantaneous variance rate of the firm's returns, *t*—time. If the firm has outstanding equity and a single bond issue with a promised final payment of *D*, then, at the time of bond maturity, *T*, the bonds will have the value *min* (*V*, *D*), and the stock will have the value *max* (*V* - *D*, 0). The authors consider the effect of indenture agreements introducing the boundary $DL(t)$ at which the firm will be reorganized (the absorbing barrier or the default line). This barrier has a profound effect: the firm risks to default at any moment, and the default probability grows over time. The firm's payments and the absorbing boundary break the spatial symmetry of the Merton's setting, reflecting the asymmetry of the firm's return distribution (the distribution skewness). We remind the reader that (Eq. (5.1/4.5)) has been derived for the problem settings without default line, when the firm can default at the debt maturity only. Equation (5.1/4.5) cannot be transferred mechanically to the B&C problem setting.

B&C revisit Merton's valuation of a zero-coupon bond (1974) when the firm can default only on debt maturity, applying the general approach with indenture agreements. B&C consider the effect of safety covenants on the value and behavior of the bond. The safety covenants are contractual provisions which give the bondholders the right to bankrupt or force reorganization of the firm if it falls short of agreed standards. The authors analyze the case when stockholders are allowed to receive a continuous dividend payment proportional to the firm value, aV. The bond value *B*(*V*, *t*) satisfies the model:

$$B_t + (r - a)VB_V + 0.5\sigma^2 V^2 B_{VV} - rB = 0 \tag{5.2a}$$

$$B(V, T) = \min(V, D) \tag{5.2b}$$

$$B(DL(t), t) = DL(t) \tag{5.2c}$$

The stock value, *S*(*V*, *t*), satisfies another model:

$$S_t + (r - a)VS_V + 0.5\sigma^2 V^2 S_{VV} - rS + aV = 0 \tag{5.3a}$$

$$S(V, T) = \max(V - D, 0) \tag{5.3b}$$

$$S(DL(t), t) = 0 \tag{5.3c}$$

B&C find the probability for $V \geq K$ and the valuation formula for bond *B* and stock *S* in the form of the direct signal determined by the BS formula (Black & Scholes, 1973) for European call option and the BS signal reflected from the absorbing boundary. We do not reproduce their results here because these results will take a lot of space, and the BS option pricing formula does not cover the case of bond/stock pricing in the B&C problem setting.

Using the same approach, B&C study the firm that has interest paying bonds outstanding. Understanding self-financing restrictions of the GBM, B&C underline the importance of methods of raising the money to make payments to bondholders. A method used in theoretical studies of interest paying bonds (but not in practice!) is to sell an asset to get the funds and make those payments. However, many bonds have contractual provisions limiting the extent of such sales. Focusing on the effect of these restrictions, B&C suggest issuing new securities to raise funds for payment of interests and dividends. To protect the value of their claim, the bondholders require that the new securities be equity or subordinated bonds. The authors concentrate their attention on the subordinated bonds, when the indenture agreement subordinates the claims of holders of junior bonds to the claims of holders of senior bonds. At the maturity date of the bonds, payments to the holders of junior bonds are made only after the full promised payment to the holders of senior bonds. The B&C results in valuation of the subordinate bonds paying interest need independent verification because they are achieved in the frameworks of Merton's model unsuitable for this case. B&C use the following model to describe the bond value:

$$B_t + rVB_V + 0.5\sigma^2V^2B_{VV} - rB + \sum_{j=1}^{n} c_j\delta(t - t_j) = 0 \tag{5.4a}$$

$$B(V, T) = \min(V, D) \tag{5.4b}$$

$$B(DL(t), t) = DL(t) \tag{5.4c}$$

Here c_j is the jth interest payment made at time t_j, n is total number of such payments, and $\delta(\cdot)$ is the Dirac delta function. The authors explain how a solution to this problem can be obtained by the recursive technique developed for the problem (5.2). However, they want to give a better perspective on the behavior of *B*. For that objective they consider the case of a perpetual bond ($B_t = 0$) with continual interest payments of *c* per unit time. It reduces problem (5.4) to:

$$0.5\sigma^2V^2B_{VV} + rVB_V - rB + c = 0 \tag{5.5a}$$

$$B(DL) = DL \tag{5.5b}$$

B&C solve the problem (5.5) and analyze its solution. They show that function $B(V, c)$ is an increasing concave function of V and c and a decreasing function of σ^2, etc. However, this solution and information about its properties are of a low value because the valuation problem involving the default line and firms paying BSEs has no time-independent solution. In B&C problem setting, the firm has a non-zero probability to default at any moment. From the point of view of physical systems, a model without the absorbing boundary is a conservative system whose number of Brownian particles remains the same all the time (the firms making the market have an infinite longevity). A model with the absorbing boundary is an open system continuously losing its particles at this boundary (the firms default and leave the market). With no inflow, the open system can exist only for a limited period of time. This fact rejects the existence of perpetual bonds, stocks, warrants, etc. in the market where the firms pay their BSEs and are subjects to default at any time. Perpetual securities can appear in the imaginary market of self-financing firms consistent with the BSM model.

To analyze the effect of indenture agreements on valuation of a zero-coupon bond, B&C use Merton's general equation for asset pricing, which is, effectively, the Black-Scholes equation for option pricing. Therefore, their bond and stock valuation formulas are wrong. Nevertheless, Black and Cox have made a major contribution to the asset pricing literature. Their default line makes the problem definition in security pricing more realistic compared against the problem setting suggested by Merton (1973; 1974), bringing us closer to the true security valuation theory.

Recipe for Achieving Optimal Capital Structure and Trade-Off Theory

The story of search for the optimal capital structure starts with Modigliani-Miller Propositions (the MMPs, Modigliani & Miller, 1958; 1963) stating that in the perfect market with taxes, the higher the firm's debt, the greater the firm value. Kraus and Litzenberger (1973) suggest a single-period valuation model of a firm paying corporate taxes and bankruptcy penalties in a complete market. They conclude that the firm market value, in general, is not a concave function of debt leverage. Scott (1976) works out a multi-period model to determine the optimal capital structure, taking into account tax benefits related to debt and bankruptcy costs. He supports the conclusion of Kraus and Litzenberger (1973) that the firm's market value is not necessarily a concave function of debt leverage.

Discussing the optimal capital structure and the effect of debt on the firm value, one cannot pass by the famous article by Leland (1994) that won the first-ever Stephen A. Ross Prize (2008) for its achievements "in explaining underpinnings of corporate finance and capital markets" (Newswise, 2008). Leland's objective (1994) is to estimate the corporate debt value and determine the optimal capital structure of the firm paying corporate taxes and bankruptcy costs. The author follows a capital structure theory suggested by Modigliani and Miller (1958; 1963), expanding it with arguments of Kraus and Litzenberger (1973) that as the debt leverage increases, the tax advantage of debt is offset by the firm's losses related to increasing risks of bankruptcy. (The thesis that growing debt increases default probability cannot be proved within the GBM frameworks, but it is supported by many empirical studies.) Leland considers two bankruptcy scenarios. The first is the endogenous bankruptcy triggered by the firm management, when it cannot raise sufficient equity capital to meet the firm's debt obligations. The second scenario is the exogenous bankruptcy triggered, when the firm value falls to the debt principal value (the firm value hits the default line).

The program of Leland's study includes the questions most important for financial theory and practice:

- How do yield spreads on corporate debt depend on leverage, firm risk, taxes, payouts, protective covenants, and bankruptcy costs?
- Do high-risk bond values behave in qualitatively different ways than investment-grade bond values?
- What is the optimal value of leverage, and how does it depend on risk-free interest rates, firm risks, taxes, protective covenants, and bankruptcy costs?
- How does a positive net-worth covenant affect the potential for agency problems between bondholders and stockholders?
- When can debt renegotiation be expected prior to bankruptcy, and can renegotiation achieve results that debt repurchase cannot?

To achieve these ambitious goals, Leland applies Merton's general equation for security pricing supplemented with Black-Cox's default line. The author uses all Merton's assumptions (see Section 4), plus a new one that capital structure decisions, once made, remain static. He needs this assumption for technical reasons. To justify the time-independent level of debt, the author refers to Modigliani and Miller (1958), Merton (1974), Black and Cox (1976), and Brennan and Schwartz (1978), who consider debt of infinite maturity and argues that for debt of sufficiently long maturity, the relative value of principal becomes low and can be neglected compared against the total debt value. Another case of time-independent environment uses the following scheme. When the debt matures, it is rolled over at a fixed interest rate unless terminated when the firm value hits the default line.

One must remember that Merton's general equation for security pricing cannot be used with the default line, and the time-independent environment is possible only in the BSM market of self-financing firms.

Leland's model includes the firm and security representing a claim on the firm, which continuously pays for a nonnegative coupon per instant of time when the firm is solvent. To satisfy the payment restrictions of the GBM (the firm must be self-financing, that is, make no payment or receive no funds from outside), Leland suggests that the firm finances the net cost of this coupon by selling additional equity from outside of the firm's portfolio to be consistent with the bond covenants that restrict firms from selling assets. The artificial nature of this scheme is obvious.

Leland argues that a security value depending on the firm value and time follows Merton's general equation with boundary conditions determined by payments at debt maturity or by payments in bankruptcy, should it happen before the maturity. Because the closed-form solution of this problem is unknown, Leland looks for its time-independent solution, tending time to infinity. He comes to a closed-form solution for this marginal case writing explicit equations for the firm's debt, equity, and the firm value equal to the firm's asset value, plus the tax deduction of coupon payments, less the value of bankruptcy costs. The last relation makes the quantitative basis of the trade-off theory, which Leland uses to determine the optimal asset structure for the exogenous default. If the firm management can choose the moment of default, they can do it, maximizing the firm's equity at that moment (the endogenous default). Using his powerful formalism, Leland presents a multilateral analysis of various debt covenants and their influence on debt variables. However, his optimal asset structure occurs extremely high (75%-90%), which definitely indicates that the author has missed important phenomena in his analysis. Authors of succeeding articles, trying to bring the optimal debt level to the levels observed in practice, supplement Leland's model with various mechanisms, relaxing the perfect market conditions, like different kinds of friction, dynamic borrowing, etc.

Leland (1994) starts his analysis of the corporate debt value accepting Merton's assumptions (A.1-A.9, Section 4). He describes dynamics of the firm value, *V*, with the GBM:

$$dV/V = \mu dt + \sigma V dW \tag{6.1}$$

where μ is the expected rate of return, σ^2 is the instantaneous variance of the return per unit of time, dW is a standard Gauss-Wiener process. To use Equation (6.1), Leland assumes (A.10) that "any net cash outflows associated with the choice of leverage must be financed by selling additional equity" (Leland, 1994, p. 1217). At that, this equity must be external to the firm because "bond covenants restrict firms from selling their assets."

The value of a claim $F(V, t)$ on the firm that continuously pays a nonnegative coupon, C, per instant of time when the firm is solvent, Leland describes with Merton's general equation for security pricing:

$$F_t + rVF_V + 0.5\sigma^2 V^2 F_{VV} - rF + C = 0 \tag{6.2}$$

with boundary conditions determined by payments at maturity, and by payments in bankruptcy should it happen prior to maturity. We have shown (Section 4) that Merton's general equation is effective for pricing short-term securities only when no arbitrage assumes the form of the martingale no arbitrage (see Eqs. (4.1-4.6) and (4.9) and a corresponding discussion). So, the coupon payment in Eq. (6.2) must be zero ($C = 0$), and Eq. (6.2) is just the Black-Scholes equation. The general equation for pricing long-term securities in the Leland's problem setting does not exist.

The author remarks that there is no closed-form solution for Eq. (6.2) for arbitrary boundary conditions. Thus, he decides to look for the time-independent solution when $F_t = 0$ and the claim value depends explicitly on the firm value only, $F(V)$ (he makes a marginal transition with $t \to \infty$):

$$rVF_V + 0.5\sigma^2V^2F_{VV} - rF + C = 0 \quad (6.3)$$

We have already explained that this transition is possible in the BSM environment only; on the contrary, in the market where firms pay their taxes, dividends, debts, etc. and can default at any time, any firm has finite longevity and $C = 0$. On the other side, if one admits that the coupon payment is non-zero, $C > 0$, then one needs an infinite-value asset for coupon payments during the infinite time interval, which is absurd.

Leland, following Black and Cox (1976), writes a general solution of (6.3) as:

$$F(V) = A_0 + A_1V + A_2V^{-X} \quad (6.4)$$

$$X = 2r/\sigma^2$$

with constants A_0, A_1, and A_2 to be determined by boundary conditions. Having the general solution for pricing any time-independent claim, Leland turns to examining specific securities determined by their boundary conditions.

He denotes the debt value as $D(V, C)$, the level of asset value at which bankruptcy is declared as V_B, the fraction of value which is lost to bankruptcy as α ($0 \le \alpha \le 1$), leaving stockholders with nothing and debtholders with a value of $(1 - \alpha)V_B$. So, boundary conditions for the debt value are:

at $V = V_B$, $D(V, C) = (1 - \alpha)V_B$, and as $V \to \infty$, $D(V, C) \to C/r$. Leland derives for the debt value:

$$D(V, C) = C/r + [(1 - \alpha)V_B - C/r](V/V_B)^{-X} \quad (6.5)$$

Next, he writes the boundary conditions for bankruptcy costs, $BC(V, C)$, as: at $V = V_B$, $BC(V, C) = \alpha V_B$, and as $V \to \infty$, $BC(V, C) \to 0$. The solution of Equation (6.4) for these boundary conditions is:

$$BC(V, C) = \alpha V_B(V/V_B)^{-X} \quad (6.6)$$

Then Leland estimates the value of tax benefits, $TB(V, C)$, associated with debt financing. The tax benefit meets the boundary conditions: at $V = V_B$, $TB(V, C) = 0$, and as $V \to \infty$, $TB(V, C) \to \tau C/r$; here τC is the tax-sheltering value when the firm is solvent. The tax-benefit function is:

$$TB(V, C) = (\tau C/r)[1 - (V/V_B)^{-X}] \quad (6.7)$$

The author remarks that tax benefits are an increasing, strictly concave function of V. Leland argues that the total value of the firm, $v(V, C)$, consists of three components: the firm's asset value, plus the value of tax deduction of coupon payments, less the value of bankruptcy costs:

$$v(V, C) = V + TB(V, C) - BC(V, C) \quad (6.8)$$

and determines the value of the firm's equity as:

$$E(V, C) = v(V, C) - D(V, C) \quad (6.9)$$

Equation (6.8) reveals that the firm pays nothing for debt coupons, else this equation will include (infinite) expenses for coupon payments. On the other side, when the firm pays nothing for its debt but receives tax returns for debt payments, it becomes a *classic arbitrage machine* generating arbitrage profits. This circumstance explains the surprising result that the greater the debt, the greater the total firm value (compare it to the MMP3, Section 2), and that the firm gets full benefits when its debt leverage is 100 percent if there is no bankruptcy cost. The bankruptcy costs make the total value function a concave down function, securing the existence of optimal leverage. The arbitrage catastrophe makes void all presented results. To the best of our knowledge, this catastrophe remains unnoticed by economists till now (June 2023).

Equation (6.8) makes the quantitative basis of the trade-off theory. With its help, Leland determines the optimal asset structure for the exogenous default boundary V_B. He also puts forward the idea of the endogenous default and evaluates the optimal value of the default boundary, maximizing the firm's equity at the time of default, if the firm management can choose the moment of default. The author gives the most detailed analysis of the behavior of bond prices and optimal debt-equity ratios, the asset value, risk, taxes, interest rates, bond covenants, payout rates, and bankruptcy costs change. Unfortunately, all these results are false. Errors in Leland's method and trade-off theory make void all following studies using them, like Hackbarth et al. (2007) analyzing the optimal structure of debt, Hackbarth et al. (2006) considering the effect of macroeconomic conditions on dynamic capital structure choice, Hackbarth and Mauer (2010) studying the interaction between financing and investment decisions with an extension of Leland's model (1994), etc.

Analyzing the optimal leverage, the author comes to a conclusion that

> leverage of about 75 to 90 percent is optimal for firms with low-to-moderate levels of asset value risk and moderate bankruptcy costs. Even firms with high risks and high bankruptcy costs should leverage on the order of 50 to 60 percent when the effective tax rate is 35 percent. (Leland, 1994, p. 1230)

The authors of the following papers, accepting Leland's model as a base, try to decrease his extremely high optimal leverage by introducing various kinds of friction in their models. However, market friction cannot change a sign of the debt effect on the firm value from the positive one (the MMP3, Leland, 1994) to the negative one (the EMM; the GBM, $P = \delta X$); it only decreases the "optimal" leverage and tax benefits of debt. A series of papers on the optimal capital structure based on the GBM (e.g., Leland & Toft, 1996; Goldstein et al., 2001; Strebulaev, 2007; Titman & Tsyplakov, 2007; Hugonnier et al., 2015) convincingly demonstrate it. These papers do achieve optimal debt levels comparable to that observed in practice, but because they replicate all errors of Leland's model, their seeming success increases misunderstanding of the effect of corporate debt on the firm value, confusing their readers even more.

The next paper on the optimal capital structure by Leland and Toft (1996) examines the effect of debt size and maturity on bond prices, credit spreads, and the optimal capital structure. To study the debt of finite debt maturity, the authors following (Merton, 1974; Black & Cox, 1976; Brennan & Schwartz, 1978) describe the firm's productive asset, V, by continuous diffusion process with constant proportional volatility σ:

$$dV/V = [\mu(V,t) - \delta]dt + \sigma V dz \tag{6.10}$$

where $\mu(V,t)$ is the total expected rate of return on asset value V, δ is the constant fraction of the value paid out to security holders, and dz is a standard Gauss-Wiener process. The process continues until V falls to a default-triggering value V_B. The authors consider a bond issue with maturity t from the present, which has principal $p(t)$ and continuously pays a constant coupon flow $c(t)$. In the event of bankruptcy, the debt of maturity t receives the fraction $\varrho(t)$ of asset value V_B. Introducing the density of the first passage time τ from V to V_B, $f(\tau; V, V_B)$, and its cumulative distribution, $F(\tau; V, V_B)$, and using the risk-neutral technique, the authors get the equation for the debt value of maturity t:

$$d(V;V_B,t) = \int_0^t e^{-r\tau}c(t)[1 - F(\tau;V,V_B)]d\tau + e^{-rt}p(t)[1 - F(t;V,V_B)]$$
$$+ \int_0^t e^{-r\tau}\rho(t)V_B f(\tau;V,V_B)d\tau \tag{6.11}$$

The first term in Eq. (6.11) represents the discounted expected value of the coupon flow paid at time τ with the probability 1 - $F(\tau)$, the second term shows the expected discounted value of principal repayment, and the third term represents the expected discounted value of the asset fraction which goes to debt of maturity t if bankruptcy occurs.

The authors do not discuss the sources and methods of financing coupon flow $c(t)$, although those sources and methods are of great importance for the GBM problems. However, one can restore the source and method of firm financing, taking account of the authors' evidence that Formula (6.11) for the debt value of maturity t converts to Formula (6.5) for the debt value of infinite maturity when $t \to \infty$ and $\rho = 1 - \alpha$. It becomes clear that the method of the coupon flow financing remains the same as it is in Leland (1994, p. 1217): "any net cash outflows associated with the choice of leverage must be financed by selling additional equity". At that, this equity must be external to the firm because "bond covenants restrict firms from selling their assets." This financing method makes the firm an arbitraging machine generating arbitrage profits and, correspondently, makes void all results by Leland and Toft (1996). One should remember that the risk-neutral technique is equivalent to the BSM model considering self-financing firms only. Therefore, all results derived within this framework have very little to do with the market of real firms paying their BSEs.

Debt and Its Effect on the Firm's Value in EMM Framework

To illustrate the effect of debt on the firm development in the EMM frameworks (Shemetov, 2021), we consider the following settings. Suppose there is a business promising an expected annual rate of return of μ percent on the firm's assets, to enter which, a firm must have assets of no less than 1000 dollar units. Suppose also that a firm has asset X = 1000 and enters this business immediately (the unlevered firm). The unlevered firm pays its BSEs in the form of fixed costs only, $P_U = P_0$. We do not include corporate taxes and dividend payments at this stage, and we also suppose the continuous mode of BSE payments. Another firm, identical to the first one in all aspects but the size of assets, has assets of X = 1000 - A units. To enter the business, the firm borrows A units of capital for T_m years at the annual interest rate of r percent (the levered firm). The total debt of this firm is $X_D = Aexp(rT_m)$. Suppose also that the debt is discharged with a constant flow $DP = (A/T_m)exp(rT_m)$. The total BSEs for the levered firm are now $P_L = P_0 + DP = P_0(1+\beta)$, $\beta = (A/(P_0T_m))exp(rT_m)$.

The equation for the return distribution of *the unlevered firm* $V(x, t)$ is ($x = \ln(RX/P_0)$, $N(x; H_0, \sigma_0^2)$ is a normal function):

$$V_t + R(1 - e^{-x})V_x - 0.5C^2V_{xx} + Re^{-x}V = 0 \quad (6.12a)$$

$$V(x,0) = N(x; H_0, \sigma_0^2),\ H_0 = \langle x_0 \rangle = \langle \ln(RX_0/P_0) \rangle,\ \sigma_0^2 = \langle (x - H_0)^2 \rangle \quad (6.12b)$$

$$V(DL, t) = 0,\ DL = 0;\ R = \mu - C^2/2 \quad (6.12c)$$

Distribution moments $H(t)$, $VAR(t)$, $SK(t)$ are defined by Equations (i.6), $DPINT(t)$ and $DPR(t)$ by Equations (i.3) and (i.4).

The equation for the return distribution $U(x, t)$ of *the levered firm* is:

$$U_t + R[1 - (1+\beta)e^{-x}]U_x - 0.5C^2U_{xx} + R(1+\beta)e^{-x}U = 0 \quad (6.13a)$$

$$U(x, 0) = N(x; H_0, \sigma_0^2) \quad (6.13b)$$

$$U(DL(t), t) = 0,\ DL(t) = \max[0, \ln(RX_D(t)/P_0)] \quad (6.13c)$$

$$X_D(t) = X_D(0)(1 - t/T_m),\ 0 \le t \le T_m \tag{6.13d}$$
$$X_D(t) = 0,\ t > T_m$$

For demonstration of the debt effect on the firm value and stability, we select steady firms with $H_0 = 2.0$, $\sigma_0^2 = 0.02$, $R = 0.10$, $C^2 = 0.01$.

We start our demonstration with a levered firm taking a small-size loan of $A = 50$ units, $r = 0.05$, with debt maturity $T_m = 3$ (the initial debt leverage is 0.0576, $\beta = 0.4811$), and compare its results against the results of unlevered firm. It is easy to verify that for loan of $A = 50$, $T_m = 3$, the default line is zero, $DL = 0$. The slope of the mean return $H(t)$ (Figure 9) shows the effective rate of return on the firm's assets after all payments. For example, the effective rate of return of the unlevered firm in Figure 9 is 0.087. As one can see, the mean return of the levered firm has a lesser slope during the first three years of the firm development because of higher payments in this period. When the payments return to the payment level of the unlevered firm, the slope of $H(t)$ of the levered firm rises close to that of $H(t)$ of the unlevered firm. However, the mean return $H(t)$ and its slope of the levered firm remains lesser than that of the unlevered firm for all time.

The variance $VAR(t)$ (Figure 10) of the levered firm grows faster in the interval of debt maturity than the variance of the unlevered firm. After discharging the debt, the growth rate of the variance declines, but $VAR(t)$ of the levered firm remains greater than $VAR(t)$ of unlevered firm. The development of skewness (Figure 11) demonstrates the same behavior: $SK(t)$ of the levered firm grows faster during the maturity period than $SK(t)$ of the unlevered firm. Then the growth rate declines, but $SK(t)$ of the levered firm remains greater by absolute value than $SK(t)$ of the unlevered firm.

Due to the choice of steady firms ($H_0 = 2.0$, $\sigma_0^2 = 0.02$) and low debt, the intensity of default probability (Figure 12) of the levered firm is about 10-15 times greater than the intensity of unlevered firm, but still remains low. The general conclusion from the graphs is that the debt does affect the firm development, and its effect is negative. The variance, skewness, default probability, and its intensity of the levered firm grow faster than the similar variables of the unlevered firm. On the other side, the mean return grows slower than that of the unlevered firm. The effect of a small debt on a steady firm is small.

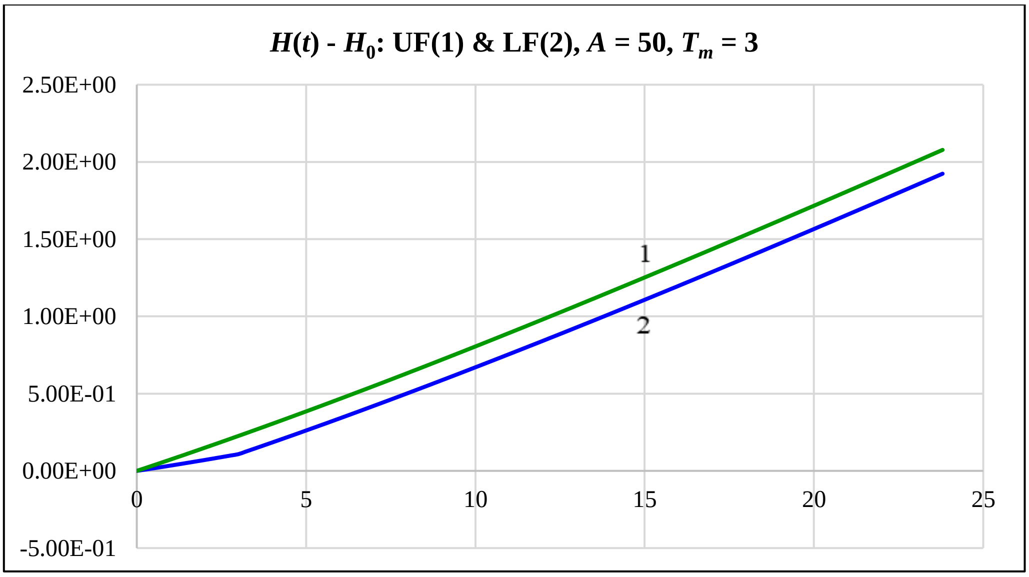

Figure 9. Evolution of mean returns $H(t)$ - H_0 for unlevered firm (line 1) and levered firm (line 2), $A = 50$, $T_m = 3$.

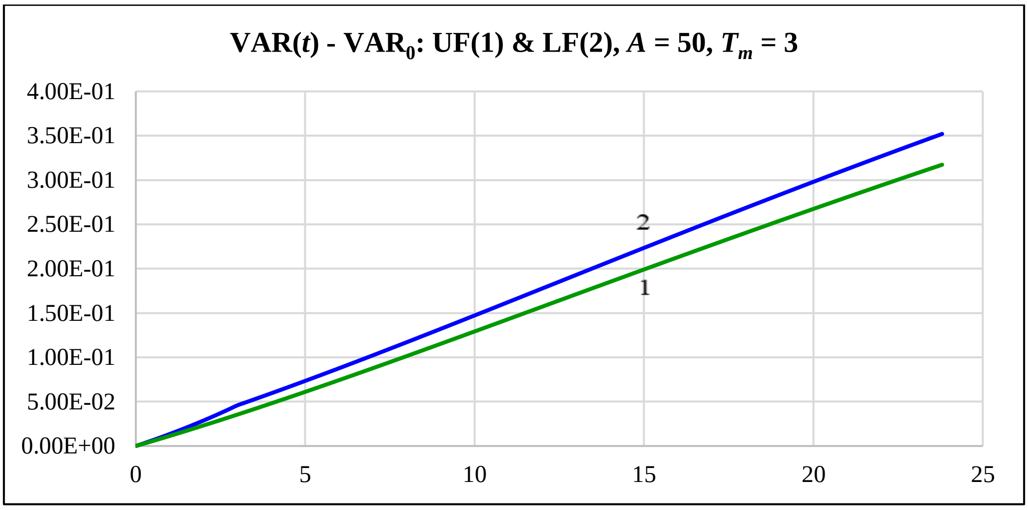

Figure 10. Variances VAR(t) - VAR_0 for unlevered firm (line 1) and levered firm (line 2), A = 50, T_m = 3.

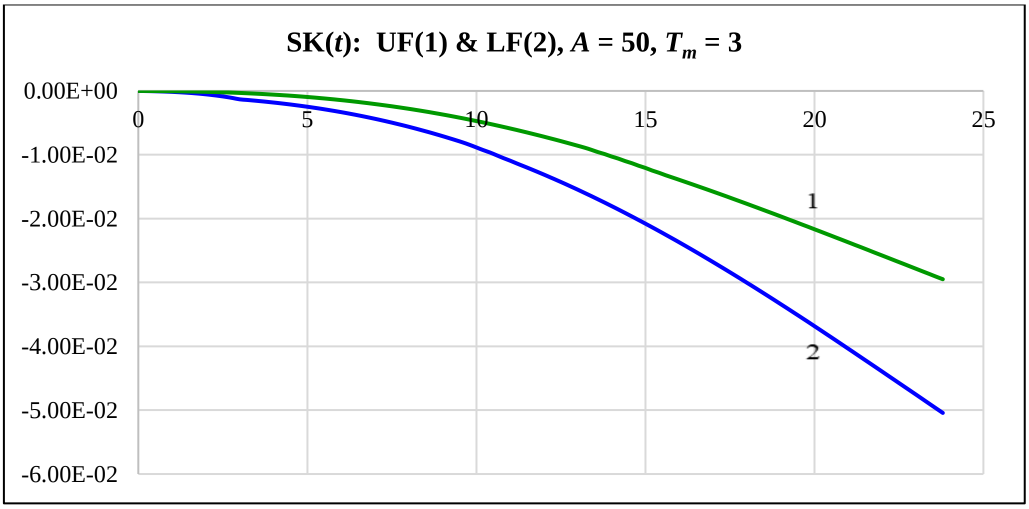

Figure 11. Evolution of skewness SK(t) for unlevered firm (line 1) and levered firm (line 2), A = 50, T_m = 3.

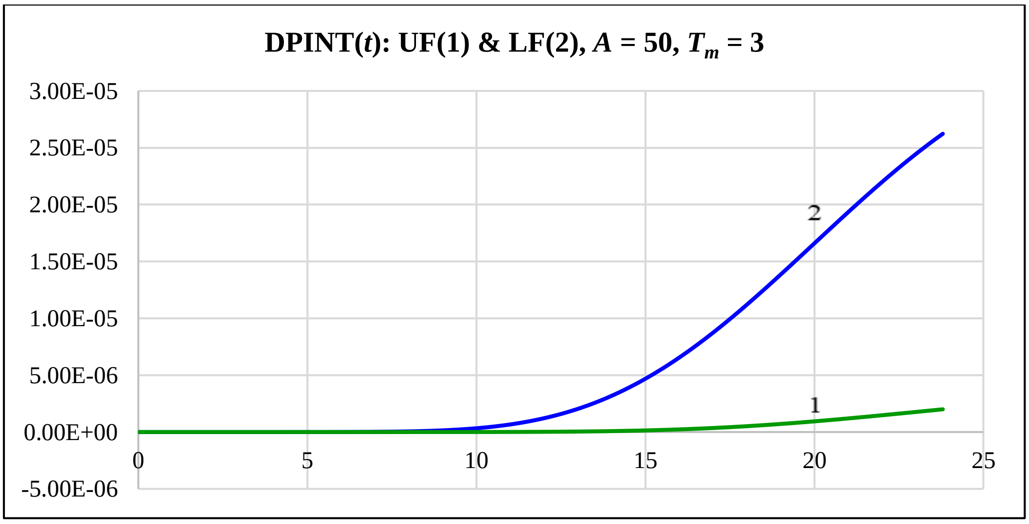

Figure 12. Intensities of default probability DPINT(t) for unlevered firm (line 1) and levered firm (line 2), A = 50, T_m = 3.

Now we consider the effects of a medium-size loans A = 150, r = 0.05, with maturities T_m = 10 and 15 years (the initial debt leverages make 0.2254 and 0.2720). Other problem parameters remain the same: R = 0.10, C^2 = 0.01, $\sigma_0^2 = 0.02$, H_0 = 2.0. The first distinction between a small-size debt and a medium-size debt is a non-zero level of the default line (Figure 13). The time, when the default line descends from the top to zero, is t_{DL0} = 8.671, and β = 1.580 for the debt maturity of T_m = 15. The BSE payments of the levered firm are 2.580 times greater than the BSE payments of the unlevered firm over debt maturity, and both firms pay the same BSEs outside the maturity. Those high payment rates explain low slopes of the mean returns in Figure 14 (lines 2 and 3) and the fast growth of the intensities of default probability in Figure 15.

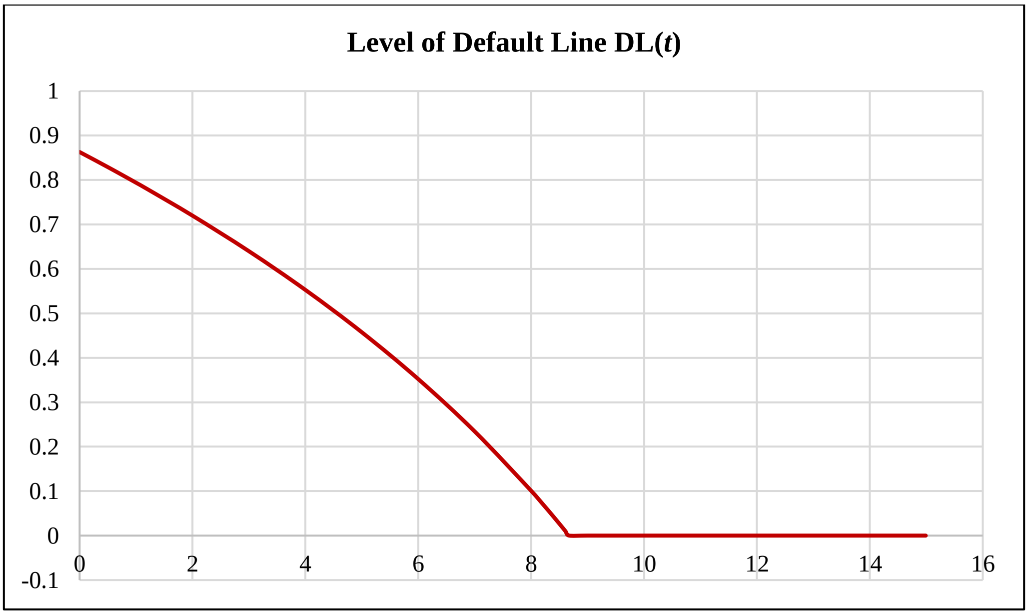

Figure 13. Level of default line DL(t) of levered firm with debt parameters A = 150, T_m = 15, t_{DL0} = 8.671.

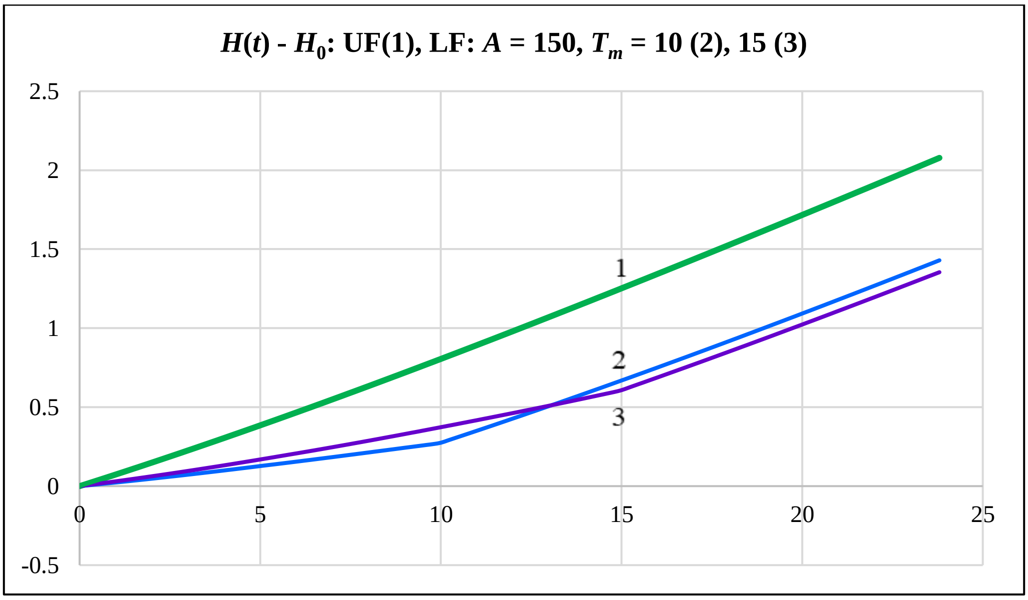

Figure 14. Evolution of mean return $H(t)$ - H_0 for unlevered firm (line 1) and two levered firms, A = 150, T_m = 10 (line 2), T_m = 15 (line 3).

When the rate of payments of the levered firm decreases to the unlevered firm level, the slope of function $H(t)$ returns close to its unlevered value (Figure 14). As one can see in Figures 14-16, the effect of a medium-size debt is not small, it influences the firm development for the rest of its life, and the longer the maturity, the greater the effect. (Actually, for any firm state and debt value acceptable for the firm, there is an optimal maturity providing the minimum default probability. However, this optimum is not critical, and there is an interval of maturities securing a suboptimal mode of debt discharging.) The default probability for a medium-size debt increases to levels of 0.10-0.15 in 25 years, while the default probability of the unlevered firm is about 10^{-5} for the same period. Any payment, including taxes and dividends, decreases the firm's profitability and survivability.

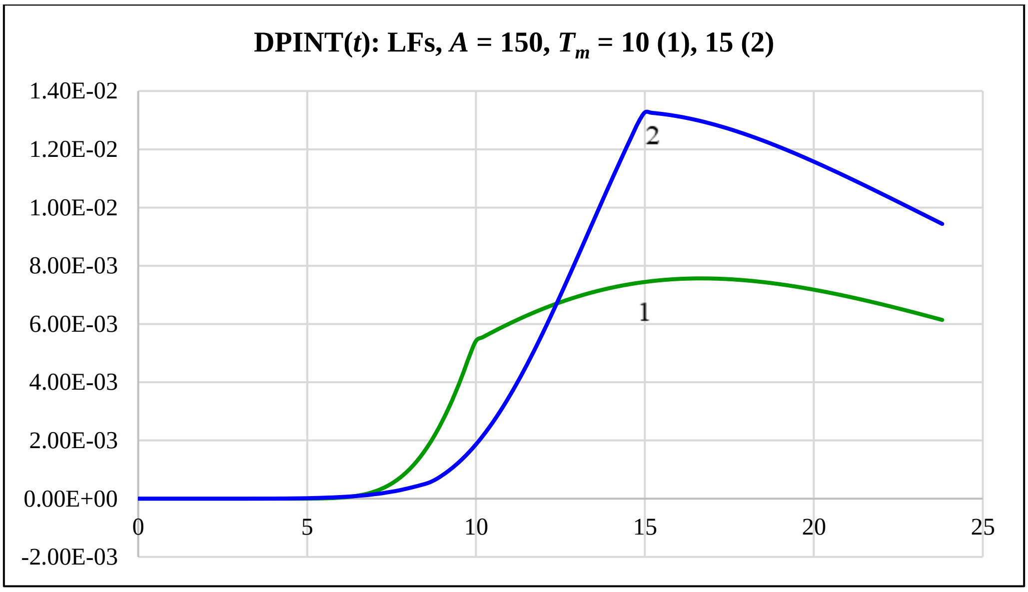

Figure 15. Intensities of default probability DPINT(t) for two levered firms with $A = 150$, $r = 0.05$ and maturities T_m = 10 (line 1) and 15 years (line 2).

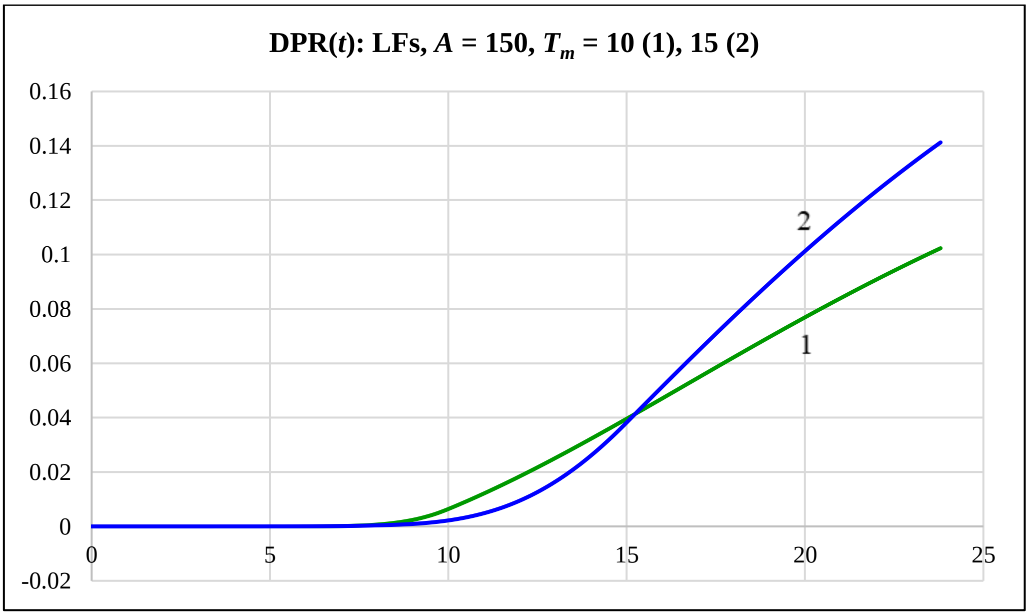

Figure 16. Default probabilities DPR(t) for two levered firms with $A = 150$, $r = 0.05$ and maturities T_m = 10 (line 1) and 15 years (line 2).

The model presented above does not include taxes and tax returns, and, therefore, cannot be directly compared against the trade-off theory. However, according to this theory, if there are no taxes, corporate debt makes no effect on the firm's value and returns. On the contrary, the EMM shows that debt makes a negative effect on the firm's value and returns. At that, the greater the debt, the stronger the effect of debt; it never disappears and can be detected many years after the debt maturity by reduced mean returns and an increased default probability. The negative effect of moderate debt develops slowly; it makes it difficult to see the relation between the deteriorating state of the firm at present and the debt financing decision made 10 years ago. In our opinion, debt financing should be considered as a strong medicine having serious collateral effects. It must be prescribed very cautiously taking into account possible distant complications.

The trade-off theory assumes that capital structure can be optimized with management weighting up the relative advantage of the tax-shield benefits of debt against the increased likelihood of incurring debt-related bankruptcy costs. Leland (1994) tries to develop a quantitative method of finding the optimal debt leverage in this problem. We demonstrate that the optimal capital structure calculated by Leland is an imaginary result of the BSM model for firms with no payments, Leland's unrealistic method of financing the firm's expenses, and his mathematically illegal transition to the time-independent equation. So, the trade-off theory, supposing debt neutrality to the firm value when there is no tax and debt benefits in the presence of taxes, is false. Sarkar and Zapatero (2003) present empirical evidence of the strong inverse relationship between profitability and leverage, supporting the EMM conclusions. On the other side, there is a wide flow of the papers based on the MMP3 by Modigliani and Miller (1963) and Leland (1994), defending the trade-off theory (e.g., Brealey & Myers, 1996, pp. 474-509; Leland & Toft, 1996; Goldstein et al., 2001; Fama & French, 2002; Strebulaev, 2007; Titman & Tsyplakov, 2007; Hugonnier et al., 2015; etc.). When the economists cannot come to a general opinion on the subject, the practitioners have no reliable recommendations on the debt leverage choice. Empiric studies by Beattie, Goodacre, and Thomson (2006) show that about half of the respondent firms seek to maintain a target debt level or a target range consistent with the trade-off theory. We explain this influence of trade-off theory on financial decision made by prolonged psychological suggestion, which financial theorists exert on corporate management. It is another illustration of a propaganda thesis that a lie repeated many times becomes (almost) the truth in the public opinion.

Does it mean that debt always makes negative effects on the firm, its value, and stability and must be avoided by all means? Not really. Shemetov (2021) shows that when the firm buys on credit a new technology increasing its future returns, it can be beneficial for the firm; the result of that project depends on details of the contract agreement, the state of the firm, and market.

Conclusion

Analyzing the seminal papers of Modigliani and Miller (1958; 1963), Black and Scholes (1973), Merton (1973; 1974), Black and Cox (1976), Leland (1994), and Leland and Toft (1996), we have shown that all of them are wrong. Because these classic works lay the foundation for the following financial studies, a significant branch of financial economics based on these papers is now in a crisis: no conclusions except the most general ones are correct, and no recommendation is reliable. To be specific, we have shown theoretically and Fortune (1996) empirically that the Black-Scholes option pricing formula cannot be sufficiently precise and accurate because the return distributions of the underlying stocks are skewed even for the top S&P 500, while Black and Scholes assume the stock return distribution to be normal. As a consequence, all theoretical extensions of the Black-

Scholes formula (Merton, 1976; Kou, 2002; Heston, 1993) and theoretical studies of exotic options based on this formula inherit its shortcomings.

We have shown theoretically and Kownatzki (2016) empirically that the S&P 500 Volatility Index (VIX) using a weighted average of implied volatility is unsuitable for many risk management applications because the VIX systematically overestimates actual volatility in non-crisis periods and underestimates actual volatility in times of financial crises (when the variance ceases to be a good measure for the volatility of unsteady firms). The volatility estimations made with the EMM are more adequate than the current volatility estimations made by the VIX and could be recommended for the next version of the Volatility Index. We have also explained a strange behaviour of the variance and skewness of stock returns before and after the crisis of 1987.

We have shown that Merton's general equation for security pricing is effective only when the firm can default at debt maturity. In a more realistic market where the firm pays its BSEs and theoretically can default at any moment, Merton's general equation is equivalent to the Black-Scholes equation, which is good for pricing short-term securities and unsuitable for pricing long-term securities and assets. As a result, all succeeding stock and bond pricing formulas are wrong, and the stock and bond traders nowadays are close to a position of traders at the dawn of security trading with very few theoretical reference points.

In contrast to the trade-off theory looking for the optimal capital structure under the MMP3 (Modigliani & Miller, 1963) that argues debt brings tax benefits to the firm, we have shown that debt negatively affects the firm's value, returns, and stability supporting the empirical study by Sarkar and Zapatero (2003). All papers teaching how to choose debt leverage in static or dynamic conditions (e.g., Leland, 1994; Leland & Toft, 1996; Goldstein et al., 2001; Strebulaev, 2007) are false and can only damage the firms if they take risks to use their theoretical recommendations as a practical guidance.

How could it happen that so many bright minds occurred to be so wrong? The error is so subtle that nobody can doubt that this idea is not the absolute truth. We show that the error hides in the interpretations of no arbitrage. Modigliani and Miller, Samuelson, Black and Scholes, and Merton interpret the no-arbitrage property of financial markets that no one can make money out of nothing ("no free lunches") neither regularly, no through stochastic speculations. One can say that a self-financing strategy provides an arbitrage opportunity if and only if $P(V_0 = 0) = 1$, $P(V_T \geq 0) = 1$, and $0 < P(V_T > 0) < 1$, here V_0 is an initial portfolio value, V_T is its value at time $T > 0$. A market is arbitrage free if there is no such strategy, understanding that $P(V_T = 0) = 1$ (having no capital, one can make no profit and no debt over time T). We call this interpretation the *martingale* no arbitrage. The martingale no arbitrage puts strong constraints on the return distributions of portfolio assets: they must be symmetric. That restriction causes no problem for pricing short-term securities like options, but it lays heavily on pricing long-term assets and securities. The martingale interpretation of the no-arbitrage principle made Merton choose the geometric Brownian model (GBM) as a descriptor of the firm and its long-term securities. In its turn, this choice leads to conflicts with observable facts: (1) the theoretical return distribution is always normal while in practice it is skewed; (2) theoretical default probabilities are always much less than that in practice; (3) the GBM describes the self-financing firms only; it makes the GBM unsuitable for analysis of real firms paying their taxes, dividends, debts, etc. Economists respond to this challenge by mending the theory with calibrated models, jump-diffusion processes, stochastic volatility processes, etc., to bring the predicted default probability closer to the probabilities observed in practice and add skewness to the return distributions of stocks and underlying firms.

The true solution to those problems consists in rejection of the martingale interpretation of the no-arbitrage principle. A qualitative analysis and a numeric solution of the original Merton equation for the firm with payments

(Merton, 1974) in the presence of Black-Cox's absorbing boundary (the default line) shows that the development of a skewed return distribution is a natural consequence of payments of the firm's expenses (the firm stops to be self-financing). Because of a continuous mass leak at the default line, the market of firms paying their expenses (BSEs) has no martingale property. The rate of mean returns of those firms continuously decreases, reducing future values of the firms and their securities. This leak runs slowly (its characteristic time is about a year); therefore, it affects the pricing of long-term assets, securities, and portfolios. Mind that the decrease in mean portfolio returns has no relation with market speculations; it is due to the slow degradation of the long-term assets constituting the portfolio. These changes in the firm state and the states of its securities correspond to organism aging in biology. Investors keeping long-term portfolios should regularly re-estimate their assets and portfolios as a whole. The short-time market speculations (one day, week, month) with any securities is a good martingale process making no changes in their activities. The new situation presents additional problems to long-term investors such as big firms, mutual and pension funds, saving and investment funds, etc., which now have to estimate and re-estimate assets in their portfolios, taking into account asset aging.

We hope we managed to show that an important branch of financial economics is in a crisis. It is bad news for those who have contributed to the current state of economics and those who have to solve practical problems with very few theoretical reference points left (but with no false recipes, too!). The good news is we know that this branch of financial economics must be reconstructed on new principles, and we know some of them. We hope that there are a lot of capable and creative economists who will do their best to fill theoretical gaps as soon as possible. The first step on this way could be organizing a cooperation of those who can and will make their contribution.

References

Adcock, C., Eling, M., & Loperfido, N. (2012). Skewed distributions in finance and actuarial science. *European J. of Finance, 21*(13), 1-29.

Badrinath, S., & Chatterjee, S. (1988). On measuring skewness and elongation in common stock return distribution: The case of market index. *J. of Business, 61*(4), 451-472.

Baker, K., & Powel, G. (1999). How corporate managers view dividend policy. *Quart. J. of Business and Economics, 38*(1), 17-35.

Bates, D. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche mark options. *Review of Financial Studies, 9*(1), 69-107.

Beattie, V., Goodacre, A., & Thomson, S. (2006). Corporate financing decisions: UK survey evidence. *J. of Business Finance and Accounting, 33*(9-10), 1402-1434.

Bhattacharya, S. (1988). Corporate finance and the legacy of Miller and Modigliani. *J. of Economic Perspectives, 2*(4), 135-147.

Black, F., & Cox, J. (1976). Valuing corporate securities: Some effects of bond indenture provisions. *J. of Finance, 31*(2), 351-367.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *J. of Political Economy, 81*(3), 637-654.

Bohn, J. (2006). Structural modelling in practice. Moody's. Retrieved from http://web-docs.stern.nyu.edu/salomon/docs/Credit2006/Bon.pdf

Brealey, R., & Myers, S. (1996). *Principles of corporate finance*. New York: McGraw-Hill.

Brennan, M., & Schwartz, E. (1978). Corporate income taxes, valuation, and the problem of optimal capital structure. *J. of Business, 51*(1), 103-114.

Carmona, R., & Dayanik, S. (2008). Optimal multiple-stopping of linear diffusions. *Mathematics and Operations Research, 33*(2), 446-460.

Carmona, R., & Touzi, N. (2008). Optimal multiple stopping and valuation of swing options. *Mathematical Finance, 18*(2), 239-268.

Charness, G., & Neugebauer, T. (2019). A test of the Modigliani-Miller invariance theorem and arbitrage in experimental asset markets. *J. of Finance, 74*(1), 493-529.

Chen, A. (1970). A model of warrant pricing in a dynamic market. *J. of Finance, 25*(5), 1041-1059.

Cox, J., Ross, S., & Rubinstein, M. (1979). Option pricing: A simplified approach. *J. of Financial Economics, 7*(3), 229-263.

Dahlgren, M., & Korn, R. (2005). A swing option on the stock market. *Int. J. of Theoretical and Applied Finance, 8*(1), 123-139.

DeAngelo, H., & DeAngelo, L. (2006). The irrelevance of the MM dividend irrelevance theorem. *J. of Financial Economics, 79*(2), 293-315.

Dhanani, A. (2005). Corporate dividend policy: The views of British financial managers. *J. of Business Finance & Accounting, 32*(7, 8), 1625-1672.

Fama, E. (1965). The behavior of stock-market prices. *J. of Business, 38*(1), 34-105.

Fama, E., & French, K. (2002). Testing trade-off and pecking order predictions about dividends and debts. *Review of Financial Studies, 15*(1), 1-33.

Fortune, P. (1996). Anomalies in option pricing: The Black-Scholes model revisited. *New England Economic Review, 3*, 17-40.

Foufas, G., & Larson, M. (2008). Valuing Asian options using the finite element method and duality techniques. *J. of Computational and Applied Mathematics, 222*(1), 144-158.

Frank, M., & Goyal, V. (2007). Trade-off and Peking order theories of debt. In B. E. Eckbo (ed.), *Handbook of Corporate Finance: Empirical Corporate Finance (Handbooks in Finance Series)*, North Holland: Elsevier, pp. 135-202.

Goldstein, R., Ju, N., & Leland, H. (2001). An EBIT-based model of dynamic capital structure. *J. of Business, 74*(4), 483-512.

Hackbarth, D., & Mauer, D. (2010). Optimal priority structure, capital structure, and investment. *Review of Financial Studies, 25*(3), 747-796.

Hackbarth, D., Hennessy, C., & Leland, H. (2007). Can the trade-off theory explain debt structure? *Review of Financial Studies, 20*(5), 1389-1428.

Hackbarth, D., Miao, J., & Morellec, E. (2006). Capital structure, credit risk, and macroeconomic conditions. *J. of Financial Economics, 82*(3), 519-550.

Harrison, J., & Kreps, D. (1979). Martingale and arbitrage in multiperiod securities markets. *J. of Economic Theory, 20*(3), 381-408.

Harrison, J., & Pliska, R. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications, 11*(3), 215-260.

Harvey, C., & Siddique, A. (2000). Conditional skewness in asset pricing tests. *J. of Finance, 55*(3), 1263-1295.

Harvey, C., Liechty, J., Liechty, M., & Mueller, P. (2010). Portfolio selection with higher moments. *Quantitative Finance, 10*(5), 469-485.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies, 6*(2), 327-344.

Hugonnier, J., Malamud, S., & Morellec E. (2015). Credit market frictions and capital structure dynamics. *J. of Economic Theory, 157*(C), 1130-1158.

Ju, N., Parrino, R., Poteshman, A., & Weisbach, M. (2005). Horses and rabbits? Trade-off theory and optimal capital structure. *Journal of Financial and Quantitative Analysis*, *40*(2), 259-281.

Kemna A., & Vorst, A. (1990). A pricing method for options based on average asset values. *J. of Finance and Banking, 14*(1), 113-129.

Kendall, M. (1953). The analysis of economic time-series—Part I: Prices. *J. of the Royal Statistical Society, 116*(1), 11-34.

Kou, S. (2002). A jump-diffusion model for option pricing. *Management Science, 48*(8), 1086-1101.

Kownatzki, C. (2016). How good is the VIX as a predictor of market risk? *J. of Accounting and Finance, 16*(6), 39-60.

Kraus, A., & Litzenberger, R. (1973). A state-preference model of optimal financial leverage. *J. of Finance, 28*(4), 911-926.

Kraus, A., & Litzenberger, R. (1976). Skewness preference and the valuation of risk assets. *Journal of Finance, 31*(4), 1085-1100.

Krueger, J. (2021). Nonparametric portfolio efficiency management with higher moments. *Empirical Economics*, *61*(3), 1435-1469.

Latane, H., & Rendleman, R. (1976). Standard deviations of stock price ratios implied in option pricing. *J. of Finance, 31*(2), 369-381.

Leland, H. (1994). Corporate debt value, bond covenants, and optimal capital structure. *J. of Finance, 49*(4), 1213-1252.

Leland, H., & Toft, K. (1996). Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. *J. of Finance, 51*(3), 987-1019.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics, 47*(1), 13-37.

Mandelbrot, B. (1963). The variation of certain speculative prices. *J. of Business, 36*(4), 394-419.

Markowitz, H. (1952). Portfolio selection. *J. of Finance, 7*(1), 77-91.

Mencia, J., & Sentana, E. (2009). Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation. *Econometrics, 153*(2), 105-121.

Merton, R. (1973). Theory of rational option pricing. *Bell J. of Economics and Management Science, 4*(1), 141-183.

Merton, R. (1974). On the pricing of corporate debt: The risk structure of interest rates. *J. of Finance, 29*(2), 449-470.

Merton, R. (1976). Option pricing when underlying stock returns are discontinuous. *J. of Financial Economics, 3*(1-2), 125-144.

Miller, M. (1988). The Modigliani-Miller propositions after thirty years. *J. of Economic Perspectives, 2*(4), 99-120.

Miller, M., & Modigliani, F. (1961). Dividend policy, growth, and the valuation of shares. *J. of Business, 34*(4), 411-433.

Modigliani, F., & Miller, M. (1958). The cost of capital, corporation finance and the theory of investment. *American Economic Review, 48*(3), 261-297.

Modigliani, F., & Miller, M. (1963). Corporate income taxes and the cost of capital: A correction. *American Economic Review, 53*(3), 433-434.

Osborn, M. (1959). Brownian motion in the stock market. *Operations Research, 7*(2), 145-173.

Panjer, H. (Ed.). (1998). *Financial economics: With applications to investments, insurance, and pensions*. Schaumburg, Illinois: The Actuarial Foundation.

Ross, S. (1985). Debt and taxes and uncertainty. *J. of Finance, 60*(3), 637-657.

Samuelson, P. (1965). Rational theory of warrant pricing. *Industrial Management Review, 6*(2), 13-39.

Samuelson, P., & Merton, M. (1969). A complete model of warrant pricing that maximizes utility. *Industrial Management Review, 10*(2), 17-46.

Sarkar, S., & Zapatero, F. (2003). The trade-off model with mean reverting earnings: Theory and empirical tests. *Economic Journal, 113*(490), 834-860.

Scholes, M. (2023). Using option pricing information to time diversify portfolio returns. Ch. 1: 1-15. In D. Gershon (Ed.), *World scientific lecture notes in finance (Vol. 6), options—45 years since the publication of the Black-Scholes-Merton model*. Jerusalem: The Gershon Fintech Center Conference.

Schwartz, E. (1977). The valuation of warrants: Implementing a new approach. *J. of Financial Economics, 4*(1), 79-93.

Scott, J. (1976). A theory of optimal capital structure. *Bell J. of Economics, 7*(1), 33-54.

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *J. of Finance, 19*(3), 425-442.

Shemetov, V. (2020). Continuous-time models for firm valuation and their collateral effect on risk-neutral probabilities and non-arbitraging principle. *Management Studies, 8*(3), 191-214.

Shemetov, V. (2021). Effect of corporate debt on firm value. *Management Studies, 9*(5), 368-411.

Shemetov, V. (2022). Some financial problems in the light of EMM results: Asset pricing and efficient portfolio allocation. *Management Studies, 10*(5), 294-324.

Simkowitz, M., & Beedles, W. (1980). Asymmetric stable distributed security returns. *J. of American Statistical Association, 75*(No. 370), 306-312.

Singleton, J., & Wingender, J. (1986). Skewness persistence in common stock returns. *Financial and Quantitative Analysis, 21*(3), 335-341.

Sprenkle, C. (1961). Warrant prices as indicators of expectations and preferences. *Yale Economic Essays, 1*, 178-231.

Stiglitz, J. (1969). A re-examination of the Modigliani-Miller theorem. *American Economic Review, 59*(5), 784-793.

Strebulaev, I. (2007). Do tests of capital structure theory mean what they say? *J. of Finance, 62*(4), 1747-1787.

Strebulaev, I., & Whited, T. (2012). Dynamic models and structural estimation in corporate finance. *Foundations and Trends in Finance, 6*, 1-163.

Sundaresan, S. (2013). A review of Merton's model of the firm's capital structure with its wide applications. *Annual Review of Financial Economics, 5*(1), 21-41.

Titman, S., & Tsyplakov, S. (2007). A dynamic model of optimal capital structure. *Review of Finance, 11*(3), 401-451.

University of California, Berkeley Haas School of Business. (2008). Business professor Hayne Leland honored with $100K financial economics prize. Retrieved from https://www.newswise.com/articles/business-professor-hayne-leland-honored-with-100k-financial-economics-prize