International Journal of Economics and Financial Issues

ISSN: 2146-4138

available at http: www.econjournals.com

International Journal of Economics and Financial Issues, 2025, 15(3), 50-58.

The Return Distribution of Stocks: A Dynamic Model

Joachim Kaldasch*

International School of Management, Hauptstraße 27, 10827 Berlin, Germany.
*Email: joachim.kaldasch@international-business-school.de

Received: 17 November 2024 **Accepted:** 03 March 2025 **DOI:** https://doi.org/10.32479/ijefi.18571

ABSTRACT

An analytical dynamic model for the return distribution of stocks is presented, based on the conservation equation of return-dependent shares of a stock. Two types of a return distribution are established: real and virtual. The real return distribution characterizes the current return distribution of shares of a stock, while the virtual return distribution is derived from traded price data. The main difference between both is that the real return distribution requires the mean holding time of shares. According to the theory, the real return distribution is stationary, featuring a double exponential distribution near mean return with power-law tails. In contrast, the virtual return distribution evolves over time and includes an additional Gaussian component around mean return. Comparison with empirical data validates the model's applicability to stocks satisfying the model's conditions.

Keywords: Return Distribution, Stocks, Stock Returns, Stock Market Dynamics
JEL Classifications: C02, C22, C46, C50, G12

1. INTRODUCTION

The stock market is a complex financial system characterized by non-linear interactions among traders and numerous unpredictable external factors. This work establishes an analytical return distribution derived from the dynamics governing the return of shares of a stock. The price return is defined as:

$$r(t)=\frac{p(t)-p(t_p)}{p(t_p)} \tag{1}$$

Where t indicates time, $p(t_p)$ is the purchase price at time step t_p and $p(t)$ is the current price. For small returns the relation can be also written as:[1]

$$r(t)\cong \ln\left(1+r(t)\right)=ln\left(\frac{p(t)}{p(t_p)}\right) \tag{2}$$

Returns are fundamental to asset pricing, portfolio management and financial risk analysis (Mantega and Stanley. 2000; Bouchaud and Potters, 2000; Malevergne and Sornette, 2006). Bachelier introduced the first model of stock returns, describing their evolution as an uncorrelated random Brownian walk (Bachelier, 1900). The central limit theorem suggests that the distribution of returns should therefore take the form of a Gaussian. However, Mandelbrot's pioneering analysis of cotton prices revealed that returns are non-Gaussian (Mandelbrot, 1963). He also found another interesting property: "time scaling." That is, the distributions of returns for various choices of a scaling time Δt, with $\Delta t=t-t_p$, ranging from 1 day up to 1 month have similar functional forms. Characteristics are a stable functional form and pronounced tails. Mandelbrot proposed that return distributions are consistent with a Lévy stable distribution (Lévy, 1937).

In the late 20th century, advancements in computer technology provided researchers with access to vast amounts of stock market data. This progress enabled precise statistical analysis of financial asset returns (Mantega and Stanley, 1994; Mantega and Stanley, 2000). Their research demonstrated that the central region of the return distribution can be described using a truncated Lévy

1 Dividends are neglected.

stable symmetric distribution or a Student t-distribution (Gu et al., 2008). The tails have the form of a power-law distribution (Plerou and Stanley, 2008; Stanley et al., 2008; Harris, 2017; Jacquier and Torricelli, 2020; Liu and Zheng, 2022). The universal nature of this return pattern is particularly remarkable, given the significant differences among financial markets. Financial returns are also a significant focus in financial mathematics and computer simulations (e.g. De Domenico et al., 2023; Li et al., 2023).

The dynamic model presented here is based on the conservation equation of the number of shares in the possession of stockholders *n(t,r)* at time step *t* and return *r*. The time evolution is based on the following assumptions:

(i) The total number of issued shares $\tilde{n}$ can be treated as constant for the time of investigation ΔT.
(ii) Shares are sold after an average holding time τ. Selling events occur independent of the current return and proportional to their number *n(t,r)*. When a share is sold, its return is reset to zero.
(iii) For small returns the return of a share performs small random jumps on the return scale denoted as the Brownian diffusion regime. This behaviour can be characterized by a constant diffusion parameter D_0.
(iv) For large returns $|r| > r_c$, where r_c is called crossover return, the diffusion parameter depends on current return *D(r)*, denoted as multiplicative Brownian diffusion regime. The diffusion parameter *D(r)* is proportional to r^2.[2]
(v) Average return of a stock evolves with a constant rate *v*.

The evolution of the return of a share can be understood as follows: When a share is purchased at $t=t_p$ the return is with (2) zero. During the holding time τ it evolves randomly on the return scale. For small returns governed by additive Brownian motion, for large returns $|r| > r_c$ by multiplicative Brownian motion. The chance that it is sold with return *r* is proportional to *n(t,r)* and occurs with rate $1/\tau$. When the share is sold, the return value instantly resets to $r=0$, and the random evolution begins anew. This paper considers two types of return distributions. The first, referred to as the real return distribution, describes the stationary return distribution of shares in the possession of the shareholders. The second, termed virtual return distribution, is derived from the stock's price evolution. It ignores the fact that the return is reset to zero after τ. Since Mandelbrot (1963), studying the virtual return distribution has become standard practice. Both distributions are established in Chapter 2. Chapter 3 compares the model with empirical stock data, followed by a discussion in Chapter 4.

2. THE THEORY

2.1. The Real Return Distribution

The presented theory applies to a stock traded on a stock market and analysed over the time of investigation ΔT. Based on assumption (i) the total number of issued shares $\tilde{n}$ is a constant in this period. Each share of the stock is assigned with the current price *p(t)* and current return *r(t)*, related by (1) and (2). The function *n(t,r)* indicates the number of shares in the possession of shareholders at time step *t* with current return in the interval *r* and *r+dr*. Initially, all shares have zero return when issued at $t=t_0$. The initial condition can thus be written as:[3]

$$n(t_0, r) = \tilde{n}\delta(r) \tag{3}$$

The total number of shares can be obtained from:[4]

$$\tilde{n} = \int_{-\infty}^{\infty} n(r)\, dr \tag{4}$$

The return distribution *P(t,r)* of shares is defined by the probability density function (pdf) of the number of shares with return *r* at time step *t* by:

$$P(t,r) = \frac{n(t,r)}{\tilde{n}} \tag{5}$$

The average return is:

$$\bar{r}(t) = \int_{-\infty}^{\infty} P(t,r) r\, dr \tag{6}$$

And the variance:

$$\sigma^2(t) = \int_{-\infty}^{\infty} P(t,r)(r-\bar{r})^2\ dr \tag{7}$$

While *σ(t)* is the standard deviation.

The return dynamics of an ensemble of shares is given by the following conservation equation:[5]

$$\frac{\partial n(t,r)}{\partial t} = d(t,r) - s(t,r) - \frac{\partial j(t,r)}{\partial r} \tag{8}$$

The time-dependent evolution of *n(t,r)* is driven by three processes represented by the three terms on the right-hand side of this equation. The first two terms model the sales and purchase processes of shares at trading events. The number of shares in the possession of shareholders *n(t,r)* increases with purchase rate *d(t,r)*.[6] The number of shares decreases with the selling rate *s(t,r)*. The final term describes the time evolution of the shares on the return scale while they are in possession of the shareholders, based on the assumptions (iii)-(v).

According to assumption (ii) the selling rate is independent of current return and proportional to their number *n(t,r)*. Sales events occur at rate $1/\tau$, where τ indicates the average time of ownership of the shares. The selling rate has the form:

$$s(t,r) = \frac{1}{\tau} n(t,r) \tag{9}$$

The total number of shares is determined by:

$$\frac{d\tilde{n}}{dt} = \tilde{d}(t) - \tilde{s}(t) \tag{10}$$

2 That $D(r) \sim r^2$ follows from the Fokker-Planck-equation of the corresponding multiplicative Langevin equation.

3 The Dirac delta function is a distribution of the form: $\delta(r-r')=\infty$ for $r=r'$ and 0 for $r \neq r'$,normalized to one.

4 A tilde over variables indicates total numbers.

5 Such a relation is known as a convection-diffusion equation.

6 The function *d(r)* indicates demand.

With the total rates:

$$\tilde{d}(t) = \int_{-\infty}^{\infty} d(t,r)\,dr \tag{11}$$

and

$$\tilde{s}(t) = \int_{-\infty}^{\infty} s(t,r)\,dr \tag{12}$$

Inserting (9) in (12) yields with (4) for the total selling rate:

$$\tilde{s} = \frac{1}{\tau}\tilde{n} \tag{13}$$

Since the total number of shares $\tilde{n}$ is constant it follows from (10):

$$\tilde{s} = \tilde{d} \tag{14}$$

Purchase events are linked to a reset of a share's return to zero. Therefore, purchased shares start their return evolution at $r=0$. The demand rate can therefore be modeled with (13) and (14) as:

$$d(r) = \frac{1}{\tau}\tilde{n}\delta(r) \tag{15}$$

While the Dirac delta function ensures that the shares start their evolution at $r = 0$.

The last term in (8) accounts for the time evolution of shares on the return scale. The model assumes that shares undergo a random walk, with a mean flow characterized by the velocity v. The rate of motion $j(t,r)$ can thus be formulated as:

$$j(t,r) = \left(v - D(r)\frac{\partial n(t,r)}{\partial r} \right) \tag{16}$$

Where the mean velocity is defined by:

$$v = \frac{\bar{r}}{\tau} \tag{17}$$

The diffusion process is described by the diffusion parameter $D(r)$. Based on assumptions (iii) and (iv) $D(r)$ can be modelled as the sum of two terms:

$$D(r) \cong D_0 + D'\frac{r^2}{r'^2_c} \tag{18}$$

The diffusion parameter of multiplicative Brownian motion D' can be given as a multiple of the diffusion parameter D_0 as:

$$D' = \varepsilon D_0 \tag{19}$$

With the free parameter ε. Then $D(r)$ becomes:

$$D(r) = D_0\gamma(r) = D_0\left(1 + \frac{r^2}{r_c^2}\right) \tag{20}$$

Where $r_c^2 = r'^2_c/\varepsilon$. While for $|r| << r_c$, the second term in (20) can be neglected it becomes dominant for $|r| >> r_c$. The return r_c marks therefore the crossover between additive and multiplicative Brownian diffusion regimes.

The return dependent evolution of the number of shares (8) can be written with (9), (15), (16) and (20) as:

$$\frac{\partial n(t,r)}{\partial t} = \frac{1}{\tau}\tilde{n}\delta(r) - \frac{1}{\tau}n(t,r) - v\frac{\partial n(t,r)}{\partial r} + D_0\frac{\partial}{\partial r}\left(\gamma(r)\frac{\partial n(t,r)}{\partial r}\right) \tag{21}$$

Scaling (21) by the total number of shares $\tilde{n}$, multiplying by τ and applying (5) and (17) a partial differential equation for the evolution of the real return distribution $P_r(t,r)$ can be established:

$$\frac{\partial P_r(t,r)}{\partial t}\tau = \delta(r) - P_r(t,r) + \bar{r}\frac{\partial P_r(t,r)}{\partial r} + k^2\frac{\partial}{\partial r}\left(\gamma(r)\frac{\partial P_r(t,r)}{\partial r}\right) \tag{22}$$

Where:

$$k = \sqrt{D_0\tau} \tag{23}$$

For mathematical simplicity the model of the real return distribution is confined to a time interval ΔT where $\bar{r} \ll 1$. The impact of the evolution of the mean return can then be neglected. The stationary solution of (22) for $r \neq 0$ becomes:

$$P_r(r) = k^2\frac{d}{dr}\left(\gamma(r)\frac{dP_r(r)}{dr}\right) \tag{24}$$

Under the condition that the crossover between both diffusion regimes is small, the stationary return distribution can be approximated by:

$$P_r(r) \cong \begin{cases} P_r'(r), |r| \ll r_c \\ P_r''(r), |r| \gg r_c \end{cases} \tag{25}$$

For $|r| << r_c$, the stationary return distribution (24) can be determined for $r \neq 0$ and $\gamma(r)=1$ from:

$$P_r'(r) \cong k^2\frac{d^2P_r'(r)}{d^2r} \tag{26}$$

The solution is given by:

$$P_r'(r) = Ae^{-\frac{r}{k}} + Be^{\frac{r}{k}} \tag{27}$$

With the free parameters A and B. The distribution can be normalized only if $B=0$ for $r>0$ and $A=0$ for $r<0$. For a continuous distribution is $A=B$ and therefore:

$$P_r'(r) \cong A\exp\left(-\frac{|r|}{k}\right) \tag{28}$$

For $|r| >> r_c$, the stationary return distribution (24) becomes:

$$P_r''(r) = k^2\frac{d}{dr}\left(\left(1 + \left(\frac{r}{r_c}\right)^2\right)\frac{dP_r''(r)}{dr}\right) \tag{29}$$

The differential equation can be solved with the ansatz:

$$P_r''(r) = \frac{Q}{|r|^{\alpha+1}} \tag{30}$$

Introducing the free parameters $Q>0$ and $\alpha>0$. The application of (30) in (29) yields:

$$P_r''(r) = \left(\frac{k^2}{r_c^2}(\alpha^2+\alpha) + \frac{k^2}{r^2}((\alpha+1)(\alpha+2))\right) P_r''(r) \quad (31)$$

The ansatz is valid if:

$$1 = \frac{k^2}{r_c^2}(\alpha^2+\alpha) + \frac{k^2}{r^2}((\alpha+1)(\alpha+2)) \quad (32)$$

For the considered range $|r| >> r_c$, the last term in (32) can be neglected compared to the first. In this approximation (32) can be used to determine the crossover return:

$$r_c \cong k\alpha\sqrt{(1+1/\alpha)} \quad (33)$$

To ensure a continuous distribution at r_c we demand:

$$P_r'(r_c) = P_r''(r_c) \quad (34)$$

The stationary real return distribution (25) can then be approximated with (28) and (30) by:

$$P_r(r) \cong \begin{cases} Ae^{-\frac{|r|}{k}}, |r| \le r_c \\ \dfrac{Q}{|r|^{\alpha+1}}, |r| > r_c \end{cases} \quad (35)$$

The free parameter can be obtained from (34):

$$Q = Ar_c^{\alpha+1}\exp\left(-\frac{r_c}{k}\right) \quad (36)$$

And the normalization condition:

$$A = \frac{1}{2k + 2\left(\dfrac{r_c}{\alpha} - k\right)\exp\left(-\dfrac{r_c}{k}\right)} \quad (37)$$

The stationary real return distribution (36) depends on two free parameters k and α. It is composed of a double exponential distribution for small returns $|r| \le r_c$, and a power-law tail resulting from multiplicative Brownian motion for large returns $|r| > r_c$.

2.2. The Virtual Return Distribution

The real return distribution $P_r(r)$ is not directly observable, as it requires all stockholders to disclose the current return on their shares to determine the mean holding time τ. However, traded price data $p(t)$ from the sales process of shares are readily available. A virtual return distribution $P_v(t,r)$ can be established from the price data for a time interval Δt, where Δt is referred to as scaling time. The return (2) is in this case evaluated from the sold price at time step t_0 and a later time step at $t_0+\Delta t$ according to:

$$r(\Delta t) = \ln\left(\frac{p(t_0+\Delta t)}{p(t_0)}\right) \quad (38)$$

While the time evolution is determined by:

$$t = t_0 + \Delta_t \quad (39)$$

The scaling time Δt is a period that serves as an analogy for the time of ownership τ. Thus:

$$\Delta t = \tau \quad (40)$$

Hence, from (23) follows for the virtual return distribution:

$$k = \sqrt{D_0 \Delta t} \quad (41)$$

The analysis of the return based solely on the price evolution implies that the reset of the return at purchase events is ignored. Consequently the differential equation (22) must be rewritten for the virtual return distribution as:

$$\frac{\partial P_v(t,r)}{\partial t} = -\frac{1}{\Delta t}P_v(t,r) - v\frac{\partial P_v(t,r)}{\partial r} + D_n\frac{\partial}{\partial r}\gamma(r)\frac{\partial P_v(t,r)}{\partial r} \quad (42)$$

Where the first term in (22), which accounts for the shift of the return to $r = 0$ at purchase events, is omitted. The solution of (42) is derived in Appendix A. The virtual return distribution can be approximated as:

$$P_v(r) \cong \begin{cases} \dfrac{C'}{\sqrt{4\pi}k}e^{-\frac{(r-\bar{r})^2}{4k^2}}, 0 \le |r| < k+\bar{r} \\ A'e^{-\frac{|r-\bar{r}|}{k}}, k+\bar{r} \le |r| \le r_c+\bar{r} \\ \dfrac{F'}{|r-\bar{r}|^{\alpha+1}}, |r| > r_c+\bar{r} \end{cases} \quad (43)$$

Where the free parameters F', C' and A' are determined in Appendix A by (A13), (A16) and (A17).

The model suggests that the virtual return distribution separates into three regions, symmetric around $\bar{r}$. A normal distribution is centered near mean return, followed by an exponential decay and a Pareto tail for large returns. The virtual return distribution contains three free parameters: $\bar{r}$, k, α. The range of the normal distribution increases with k. Since (41) implies that $k \sim \sqrt{\varnothing t}$, (43) is not stationary, but a function of the scaling time. For $\Delta t \to \infty$, the normal distribution around the center dominates the virtual return distribution.

For sufficiently small k the normal distribution becomes negligible, and the virtual return distribution (43) simplifies to a shifted Laplace distribution:

$$P_v(r) \cong \frac{1}{\beta\sigma}e^{-\beta\frac{|r-\bar{r}|}{\sigma}} \quad (44)$$

With:

$$\beta = \sqrt{2} \quad (45)$$

Where the standard deviation σ can be obtained from the empirical data by (A24). Introducing the scaling:

$$R = \frac{r-\bar{r}}{k} \quad (46)$$

With the condition:

$$\int_{-\infty}^{\infty} P_v(r)\,dr = \int_{-\infty}^{\infty} P_v(R)\,dR \tag{47}$$

The virtual return distribution (43) becomes:

$$P_v(R) \cong \begin{cases} A'e^{-\frac{R^2}{4}}, 0 \le |R| < 1 \\ A''e^{-R}, 1 \le |R| \le R_C \\ \dfrac{F''}{|R|^{\alpha+1}}, |R| > R_C \end{cases} \tag{48}$$

Where (33) leads with (46) to:

$$R_C = \alpha\sqrt{(1+1/\alpha)} \tag{49}$$

While:

$$A'' = \frac{1}{2\left(\sqrt{\pi}e^{-\frac{3}{4}}erf\left(\frac{1}{2}\right)+e^{-1}\right)} \tag{50}$$

and:

$$F'' = A''\, R_C^{\alpha+1}e^{-R_C} \tag{51}$$

With the scaled return R, the virtual return distribution becomes independent of Δt. The only free parameter is α.[7]

3. COMPARISON WITH EMPIRICAL DATA

To demonstrate the applicability of the model, the virtual return distribution of a stock is analysed. To be applicable, the model conditions (i)-(v) must be met. It implies that the parameters v, D_0 and α can be treated as constants. The standard deviation of a diffusion process can be written with (A24) for empirical data as:

$$\sigma = \sqrt{2D\Delta t} \tag{52}$$

Where D is denoted as the apparent diffusion parameter. If D is constant, the standard deviation of the empirical data evolves as $\sigma \sim \sqrt{\Delta t}$. The model parameter D_0 can then be obtained from (A26). The relation (17) suggests with (40) that the parameter v is constant if $\bar{r} \sim \varnothing t$. The only free parameter α must be fitted to the empirical return data.

An example, that satisfies the model conditions is the MasterCard (MA) stock traded at the NYSE. The historical price data are obtained from Yahoo finance historical data.[8] Investigated is the time interval ΔT=4403 days spanning from 25.5.2006 to 20.11.2023.[9] Displayed in Figure 1 are the standard deviation and the mean return (insert). The mean return $\bar{r}(\varnothing t)$ can be well approximated by a linear function (solid line) with $v \cong 1.05\times10^{-3}$

Figure 1: Displayed are the empirical data of the standard deviation $\sigma(\Delta t)$ (dots) as a function of the scaling time Δt for the stock MA (MasterCard) traded at the NYSE. The solid line is (52) with $D \cong 1.65\times10^{-4}$ per day. In the inset, the empirical data for the mean return $\bar{r}$ (Δt) (dots) are shown as a function of scaling time Δt. The solid line represents (17) applying (40) with $v \cong 1.05\times10^{-3}$ per day

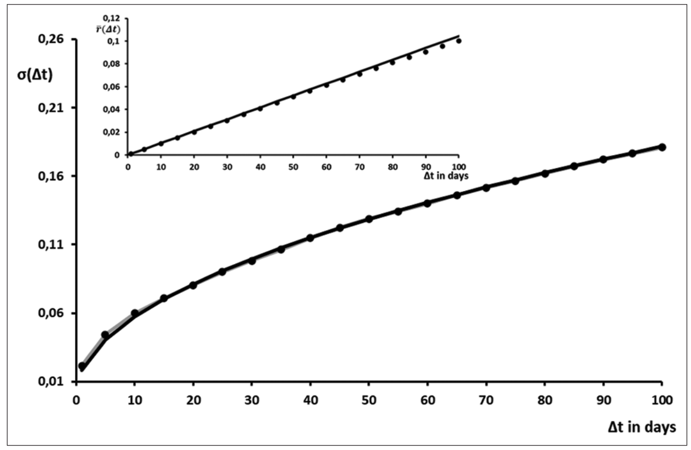
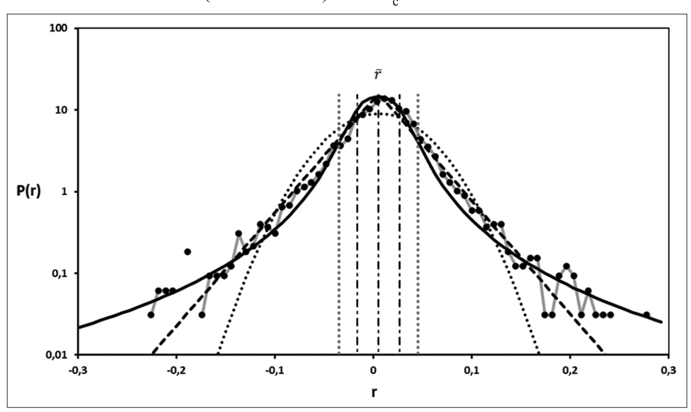


Figure 2: Displayed are the empirical return data for the stock MA for Δt=5 days (dots) with $\sigma=4.43\times10^{-2}$ and $\bar{r}=5.12\times10^{-3}$. The solid line represents the virtual return distribution (43) with $k=1.96\times10^{-2}$ and $\alpha=1.6$. Also displayed is the corresponding normal distribution (dotted line) and Laplace distribution (dashed line). Indicated is the normal distribution regime ($|r|=k$) (dash-dot-lines) and the crossover return (dotted lines) with $r_c=4.0\times10^{-2}$



per day. The standard deviation $\sigma(\Delta t)$ follows a square root dependence with $D \cong 1.65\times10^{-4}$ per day. The model is applicable to the evolution of this stock at least for Δt≤60 days.[10] The relation (A26) yields $D_0 \cong 7.7\times10^{-5}$ per day. Empirical studies suggest that the parameter α is not generally independent of Δt (Gopikrishnan et al., 1999). Following Alfonso et al. (2012) an average parameter α can be given by $\alpha \cong 1.6$.

Displayed in Figures 2-4 are the empirical return distributions (dots) together with the virtual return distribution (43) (solid line) in a semi-log-plot for Δt=5 *days*, Δt=20 *days* and Δt=60 *days*. The boundaries of the normal distribution, given by $|r|=k$, are indicated by dash-dot-lines, where k is obtained from (41). Between the dash-

7 In general, the parameter α may vary with Δt and differ between positive and negative returns. This case is not addressed here.

8 Stock splits are accounted for in the empirical price data.

9 https://finance.yahoo.com/

10 This is not generally the case for stocks traded on the NYSE.

Figure 3: Displayed are the empirical return data for the stock MA for Δt=20 days (dots) with $\sigma=8.04\times10^{-2}$ and $\bar{r}=2.02\times10^{-2}$. The solid line represents the virtual return distribution (43) with $k=3.92\times10^{-2}$ and α=1.6. Also displayed is the corresponding normal distribution (dotted line) and Laplace distribution (dashed line). Indicated is the range of the normal distribution ($|r|=k$) (dash-dot-lines) and the crossover return (dotted lines) with $r_c=8.0\times10^{-2}$

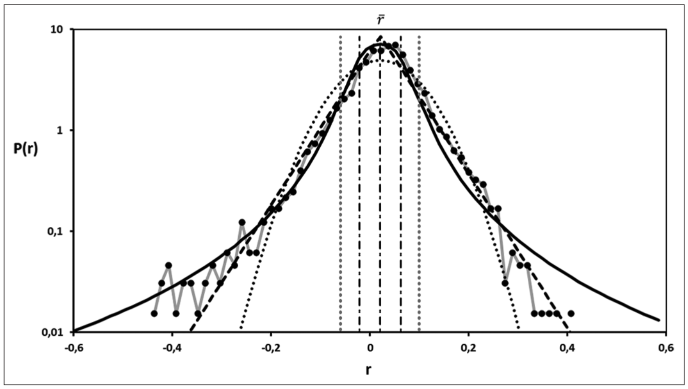
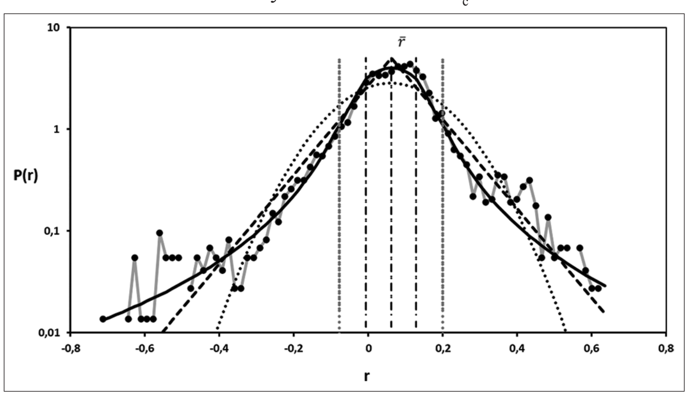


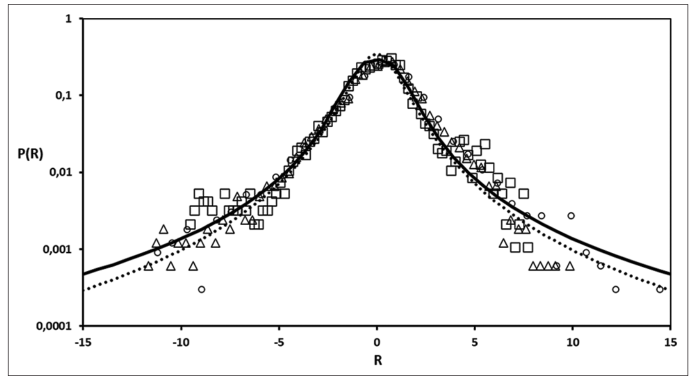
Figure 4: Displayed are the empirical return data for the stock MA for Δt=60 days (dots) σ=0.14 and $\bar{r}=6.12\times10^{-2}$. The solid line represents the virtual return distribution (43) with $k=6.79\times10^{-2}$ and α=1.6. Also displayed is the corresponding normal distribution (dotted line) and Laplace distribution (dashed line). The range of the normal distribution is marked between dash-dot lines, while the crossover return is indicated by dotted lines with r_c=0.14



dot-lines and dotted lines the return distribution can be described by an exponential distribution. For $|r|\geq r_C$, the distribution exhibits a power-law tail, while r_c is indicated by dotted lines. Also shown are the corresponding Laplace (dashed line) and normal distribution (dotted line) with equivalent σ and $\bar{r}$. While the Laplace distribution predicts a pronounced central peak at mean return, the normal distribution is much broader than the empirical data. Both are not appropriate to describe the empirical return distribution.

Figure 5 displays the empirical return data scaled according to (46) for the considered scaling times. As suggested by the model, the scaled empirical data collapse onto a single distribution. For this stock the scaled virtual return distribution (48) can be well described with α=1.6 (solid line). Previous studies have proposed alternative approaches for fitting the return distribution, such as using a truncated Lévy stable symmetrical distribution or a Student-t distribution (Mantega and Stanley (1994), Gu et al.,

Figure 5: The scaled empirical return data for the stock MA are shown for Δt=5 days (circles), Δt=20 days (triangles) and Δt=60 days (squares). The solid line depicts the scaled virtual return distribution (48) with α=1.6. The dotted line is a Student-t distribution with the degree of freedom n=2



2008). For comparison, a Student-t distribution with degree of freedom n=2 is displayed in Figure 5 (dotted line), providing a similar good fit to the scaled empirical return data.

4. DISCUSSION AND CONCLUSION

The paper introduces a dynamic model for the return distribution of stocks, based on the conservation equation of return-dependent shares. The model suggests that stock returns satisfying the model conditions (i)-(v) show next to mean motion an additive Brownian random walk for small returns and multiplicative Brownian motion for large returns. The model derives two return distributions: real and virtual. The real return distribution accounts for the reset of a share's return to zero associated with sales events. It features a symmetric double exponential distribution around mean return and power-law tails for large returns.

The virtual return distribution can be established from easily accessible price data from purchase events of a stock. The return is evaluated over a scaling time Δt that simulates the holding time τ. By disregarding the reset of the return at purchase events, an approximate virtual return distribution can be established. This distribution comprises a normal distribution around mean return, followed by an exponential decline and a power-law tail for large returns. The theory suggests that the return distribution gradually converges to a Gaussian distribution after a sufficiently long scaling time Δt. This result agrees with empirical findings of Wu (2006) and Liu and Zheng (2022).

The applicability of the model is demonstrated for the MA (MasterCard) stock traded at the NYSE, fulfilling the model conditions. In previous research the tails of the distribution were found to follow a power-law density $P(r)\sim r^{-(\alpha+1)}$, with $\alpha\cong3$ (e.g. Amaral et al., 2000) and $\alpha\cong1.6$ (Alfonso et al., 2012). The application of the model to the tails of the return of the MasterCard stock with $\alpha\cong1.6$ exhibits good agreement with the empirical data. Also, a Student-t distribution with degree of freedom n=2 fits the data well, as proposed e.g. by Mantega and Stanley (1994) and Gu et al. (2008).

In contrast to most contemporary studies, Matia et al. (2004) observed an exponential distribution of the form $P(r)\sim e^{-\beta|r|/\sigma}$

for daily returns of the 49 largest stocks in the Indian stock market. They reported values of $\beta = 1.51$ for the negative and $\beta = 1.34$ for the positive tail. The model presented can explain this result for the case that the contributions from the normal distribution around the center and the Pareto tail are negligible. Then the virtual return distribution can be approximated by the Laplace distribution (44) with $\beta = \sqrt{2}$, which is close to the quantities found by Matia et al. While they attributed the empirical outcome to the maturity of the financial market, the model suggests that it may rather stem from the selection of stocks analysed by the investigators.

In conclusion, the presented dynamic theory of the return distribution of stocks offers valuable insights into the relationship between return dynamics and the corresponding return distribution.

REFERENCES

Alfonso, L., Mansilla, R., Terrero-Escalante, C.A. (2012), On the scaling of the distribution of daily price fluctuations in the Mexican financial market index. Physica A, 391, 2990-2996.

Amaral, N.F., Plerou, V., Gopikrishnan, P., Meyer, M., Stanley H.E. (2000), The distribution of returns of stock prices. International Journal of Theoretical and Applied Finance, 3, 365-369.

Bachelier, L. (1900), Théorie de la Spéculation. Ph.D. Thesis. Paris, France: University of Paris.

Bouchaud, J.P., Potters, M. (2000), Theory of Financial Risks: From Statistical Physics to Risk Management. Cambridge, UK: Cambridge University Press.

De Domenico, F., Livan, G., Montagna, G., Nicrosini, O. (2023), Modeling and simulation of financial returns under non-Gaussian distributions. Physica A: Statistical Mechanics and its Applications, 622, 128886.

Gopikrishnan, P., Plerou, V., Amaral, L.A.N., Meyer, M., Stanley, H.E. (1999), Scaling of the distribution of fluctuations of financial market indices. Physical Review E, 60, 5305-5316.

Gu, G.F., Chen, W., Zhou, W.X. (2008), Empirical distributions of Chinese stock returns at different microscopic timescales. Physica A: Statistical Mechanics and its Applications, 387, 495-502.

Harris, E.D. (2017), The distribution of returns. Journal of Mathematical Finance, 7, 769-804.

Lévy, P. (1937), Theorie de l'Addition des Variables Aleatoires. Paris: Gauthier-Villars.

Li, Y.C., Huang, H.Y., Yang, N.P., Kung, Y.H. (2023) Stock market forecasting based on spatiotemporal deep learning. Entropy (Basel), 25, 1326.

Liu P., Zheng Y. (2022) Precision measurement of the return distribution property of the Chinese stock market index. Entropy, 25(1), 36.

Malevergne, Y., Sornette, D. (2006), Extreme Financial Risks: From Dependence to Risk Management. Berlin, Heidelberg, Germany: Springer.

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

Mantega, R.N., Stanley, H.E. (1994), Stochastic process with ultraslow convergence to a gaussian: The truncated Lévy flight. Physical Review Letters, 73, 2946-2949.

Mantega, R.N., Stanley, H.E. (2000) An Introduction to Econophysics: Correlations and Complexity in Finance. Cambridge, UK: Cambridge University Press.

Matia, K., Pal, M., Salunkay, H., Stanley, H.E. (2004), Scale-Dependent Price Fluctuations for the Indian Stock Market; Europhys. Lett., 66, 909-914.

Plerou, V., Stanley, H.E. (2008), Stock return distributions: Tests of scaling and universality from three distinct stock markets. Physical Review E, 77, 037101.

Stanley, H.E., Plerou, V., Gabaix, X. (2008) A statistical physics view of financial fluctuations: Evidence for scaling and universality. Physica A: Statistical Mechanics and its Applications, 387, 3967-3981.

Wu, L. (2006), Dampened power law: Reconciling the tail behaviour of financial security returns. The Journal of Business, 79, 1445-1473.

APPENDIX A

Derivation of the Virtual Return Distribution

We start analysing the return dynamics by disregarding the convection term in (42). The corresponding return distribution is denoted $P_0(t,r)$ and has the form:

$$\frac{\partial P_0(t,r)}{\partial t} \cong -\frac{1}{\Delta t} P_0\left(t,r\right) + D_0 \frac{\partial}{\partial r}\gamma(r)\frac{\partial P_0\left(t,r\right)}{\partial r} \tag{A1}$$

Introduced is the scaled return:

$$r' = \frac{r}{\rho} \tag{A2}$$

where the parameter $\rho > 0$ is chosen such that r' is of the order one for the considered return region. The evolution of (A1) can be investigated by the dimensionless number:

$$K = \sqrt{\frac{D_0 \Delta t}{\rho^2}} = \frac{k}{\rho} \tag{A3}$$

Applying (41). The parameter K quantifies the relation between the time of ownership Δt compared to the diffusive time ρ^2/D_0. (A1) becomes with (A2) and (A3):

$$\frac{\partial P_0(t,r')}{\partial t}\Delta t = -P_0\left(t,r'\right) + K\frac{\partial}{\partial r'}\gamma(r)\frac{\partial P_0\left(t,r'\right)}{\partial r'} \tag{A4}$$

For fixed Δt the magnitude of K depends only on ρ. Two cases can be distinguished.

For small returns $\rho << k$, i.e. $K \to \infty$, the first term on the r.h.s. in (A4) can be neglected compared to the second. Therefore (A4) can be approximated in this case by:

$$\frac{\partial P_0(t,r)}{\partial t} \cong D_0 \frac{\partial^2 P_0(t,r)}{\partial r^2} \tag{A5}$$

For small returns, the Brownian diffusion process dominates.[11] Since (A5) is a standard diffusion equation, it can be solved with the ansatz:

$$P_0(t,r) = C_0 t^{-\frac{1}{2}} e^{-\frac{r^2}{4D_0 t}} \tag{A6}$$

With the initial condition (3) at $t_0=0$:

$$P_0(t_0, r) = \delta(r) \tag{A7}$$

We obtain:

$$P_0(t,r) = \frac{C'}{\sqrt{2\pi}\sigma_0} e^{-\frac{r^2}{2\sigma_0^2}} \tag{A8}$$

While

$$\sigma_0 = \sqrt{2D_0 t} \tag{A9}$$

And

$$C_0 = \frac{C'}{\sqrt{4\pi D_0}} \tag{A10}$$

For $\rho>k$, i.e. $K>1$, a stationary solution exists, which can be derived from (A4):

$$P_0(r) = k^2 \frac{\partial}{\partial r} \gamma(r) \frac{\partial P_0(r)}{\partial r} \tag{A11}$$

This relation is equivalent to (24). The stationary distribution can therefore be written with (35) as:

$$P_0(r) = \begin{cases} A' e^{-\frac{|r|}{k}}, r_k \le |r| \le r_c \\ \dfrac{F'}{|r|^{\alpha+1}}, |r| > r_c \end{cases} \tag{A12}$$

With the free parameters α, A' and F'. The crossover return r_c is determined by (33). From (34) follows:

$$F' = A' r_c^{\alpha+1} \exp\left(-\frac{r_c}{k}\right) \tag{A13}$$

The transition between the solutions (A8) and (A12) occurs at $K \cong 1$, that is at:

$$\rho \cong k \tag{A14}$$

Thus, the virtual return distribution can be approximated with (A8) and (A12) as:

$$P_0(t,r) \cong \begin{cases} \dfrac{C'}{\sqrt{2\pi}\sigma_0} e^{-\frac{r^2}{2\sigma_0^2}}, 0 \le |r| < k \\ A' e^{-\frac{|r|}{k}}, k \le |r| \le r_c \\ \dfrac{F'}{|r|^{\alpha+1}}, |r| > r_c \end{cases} \tag{A15}$$

While σ_0 is with (A9) a function of t and k with (41) a function of Δt.

11 This is the case for $k<r_c$.

For a continuous distribution (A8) and (A12) are equal at k, which leads for $k< r_c$ to:

$$C' = A'\sqrt{2\pi}\sigma_0 \exp\left(\frac{k^2}{2\sigma_0{}^2}-1\right) = A'\sqrt{4\pi}\; ke^{-\frac{3}{4}} \tag{A16}$$

From the normalization condition follows:

$$A' = \left(\begin{array}{l} 2k\left(\sqrt{\pi}e^{-\frac{3}{4}}\mathrm{erf}\left(\frac{1}{2}\right)+e^{-1}\right)+ \\ 2k^2 exp\left(-\alpha\sqrt{\left(1+\frac{1}{\alpha}\right)}\right)\frac{\left(\alpha\sqrt{\left(1+\frac{1}{\alpha}\right)}\right)^2}{\alpha-1} \end{array}\right)^{-1} \tag{A17}$$

The convection term can be incorporated by expressing the virtual return distribution as:[12]

$$P(t,r) = P_0(t, r\text{-}vt) \tag{A18}$$

Applying the chain rule leads to:

$$\frac{\partial P(t,r)}{\partial t} = \frac{\partial P_0(t,r)}{\partial t} - v\frac{\partial P_0(t,r)}{\partial r} \tag{A19}$$

Thus (A1) becomes:

$$\frac{\partial P(t,r)}{\partial t} = -\frac{1}{\Delta t}P_0(t,r) - v\frac{\partial P_0(t,r)}{\partial r} + D_0\frac{\partial}{\partial r}\gamma(r)\frac{\partial P_0(t,r)}{\partial r} \tag{A20}$$

Which is equivalent to (42).[13] Hence, (A18) solves the differential equation (42). With (A15) and (A18) the virtual return distribution can finally be written as:

$$P_v(t,r) \cong \begin{cases} \dfrac{C'}{\sqrt{2\pi}\sigma_0}e^{-\frac{(r-\bar{r})^2}{2\sigma_0{}^2}}, 0 \le |r| < k+\bar{r} \\ A'e^{-\frac{|r-\bar{r}|}{k}}, k+\bar{r} \le |r| \le r_c+\bar{r} \\ \dfrac{F'}{|r-\bar{r}|^{\alpha+1}}, |r| > r_c+\bar{r} \end{cases} \tag{A21}$$

By choosing $t_0=0$, (39) suggests that $t=\Delta t$. Therefore (A9) turns with (41) into:

$$\sigma_0 = \sqrt{2D_0\Delta t} = \sqrt{2}k \tag{A22}$$

Note that the variance of the virtual return distribution (A21) is with (7) determined by:

$$\sigma^2(t) = \int_{-\infty}^{\infty} P_v(t,r)(r-\bar{r})^2\, dr \tag{A23}$$

Introducing the apparent diffusion parameter obtained from empirical data:

$$D = \frac{\sigma^2(t)}{2\Delta t} \tag{A24}$$

The diffusion parameter of the Brownian diffusion regime can be evaluated with (A23) from:

$$D_0 = D\left(1+\frac{5}{2}e^{-1} - \left(\frac{R_C^2}{2}+R_C+1\right)e^{-R_C} + A'R_C^3e^{-R_C}\right)^{-1} \tag{A25}$$

Applying (49) and (50).

12 Since (A18) is an approximation derived from (A17), the distribution is not strictly normalized to one. The error is of the order k^2.

13 The evolution of the mean return can be interpreted as a moving reference frame with velocity v.