Statistical Modeling of Stock Returns: A Historical Survey with Some Methodological Reflections

Koundouri, Phoebe and Kourogenis, Nikolaos and Pittis, Nikitas

2012

Statistical Modeling of Stock Returns: A Historical Survey with Some Methodological Reflections

Phoebe Koundouri,* Nikolaos Kourogenis,† Nikitas Pittis‡

Abstract

This paper aims at identifying the motivating forces that gave birth to the statistical models of asset returns since the beginning of the twentieth century. The major question addressed is: Where do statistical models of asset returns come from?" This central question encompasses a number of secondary ones: What do these models do? Do they explain or simply describe the empirical regularities of asset returns, identified at different historical periods? If explanation provides 'something', *over and above* description, then how can it be defined? Moreover, how is this reflected on explanatory versus descriptive models of asset returns? In the context of the models identified as explanatory, do these models offer an actual explanation for the regularities of interest or merely a potential explanation? Related to the last question, does the realism of the assumptions underlying the explanatory models matter? Has the literature adopted a realist or an instrumentalist attitude towards the explanatory models of asset returns? Our answers to these questions are being informed by our attempts to draw some analogies between the main issues concerning the statistical modelling of asset prices and those concerning the theoretical modelling of the Brownian motion in Physics.

*Department of International and European Economic Studies, Athens University of Economics and Business

†Department of Banking and Financial Management, University of Piraeus.

‡Department of Banking and Financial Management, University of Piraeus.

1 Introduction

In a series of papers, Box and his co-authors classified statistical models in two broad categories, the first including the so-called *empirical* or *interpolatory* models, and the second the *explanatory* or *mechanistic* ones (see Box and Hunter 1965, and Box and Draper 1987). Lehmann (1990) summarises the main differences between these two types as follows: "Empirical models are used as a guide to *action*, often based on forecasts of what to expect from future observations. In contrast, explanatory models embody the search for the basic mechanism underlying the process being studied; they constitute an effort to achieve *understanding*" (Lehmann 1990, p. 163). This classification, though intuitively appealing, leaves one fundamental question unanswered: under what objective conditions "understanding" is considered to be achieved? In other words, in the context of a statistical model what are the criteria that a purported explanation has to satisfy in order to be deemed as scientifically or formally (rather than subjectively) adequate?

Since 1948, year in which the Hempel and Oppenheim seminal paper on the nature of explanation was published, scientific explanation has been at the centre of a heated debate in the philosophy of science. Hempel and Oppenheim's paper makes the distinction between explanation of a single event and explanation of an empirical regularity. It also distinguishes between deterministic versus non-deterministic events, as well as between universal versus statistical regularities. Central to this paper is the thesis that explanation is an argument, either a deductive one, in the case of deterministic events and regularities (universal or statistical), or an inductive one, in the case of non-deterministic events. In either case, the explanation is achieved by rendering the explanandum (event or regularity) expected (either with certainty or with high inductive probability) by means of empirical laws and initial conditions (the explanans). This model of explanation, usually referred to as the "covering law" model, has been criticized by various authors during the last sixty years or so, but it still remains the starting point of any serious discussion on the nature of explanation.

The statistical modelling of returns of financial assets has attracted a great deal of research interest for more than a century. All the statistical models of asset returns that have been advanced so far, aim at either describing or explaining the empirical regularities exhibited by asset returns. The method by which a particular model is produced is heavily affected by whether description or explanation is the aim of the model. However, before the statistician develops a statistical model to either describe or explain an empirical regularity, she must first define clearly what this regularity is. This is not an easy task, especially in view of the fact that defining a regularity contains a great deal of subjective judgement. This judgement depends on the statistician's theoretical perceptions, as they are formed by the probabilistic concepts available while the regularity is being examined. In other words, an empirical regularity is not a purely objective property of the data, but it is partly defined by the manner in which the observer interprets the patterns in the data that she was able to discern. This subjectivity results in the empirical regularities being contingent, that is, the same regularity or the same pattern of behavior may be interpreted in radically different ways at different points in time.

The preceding discussion has identified two factors affecting the emergence of statistical models of asset returns: whether the model aspires to be explanatory or descriptive and how the empirical regularities that motivate the model have been identified and interpreted. The effects of these two factors on the statistical modelling of asset returns over time are the central theme of the present paper. Specifically, this paper aims at answering the following questions: Which factors motivated the birth of the various statistical models of asset returns over time? Or more simply, where have these models come from? What do these models do? Do they explain or simply describe the empirical regularities of asset returns that have been identified as such at each particular historical period? If explanation provides 'something', *over and above* description, then how can this be defined? Moreover, how is this reflected on explanatory versus descriptive models of asset returns? If the probabilistic properties/assumptions that make up an explanatory statistical model are the explanandum in a "covering-law" type of explanation, then what is the nature of the "laws" that are allowed to participate as premises in such arguments? Can these laws be of purely empirical origins or should they also be interpretable as "assumptions" of an empirically relevant theorem, with the latter being produced within a formal mathematical system? In the context of the models identified as explanatory, do these models offer an actual explanation for the regularities of interest or merely a potential explanation? Related to the last question, does the realism of the assumptions (explanans) underlying the explanatory models matter? To this end, has the literature adopted a realist or an instrumentalist attitude towards the explanatory models of asset returns? Apart from regularities, are the statistical models of asset returns capable of providing a formal explanation for singular events? For example, can these models explain why on 22/02/11 the price of Citigroup stock fell by more than 4%?

Our answers to these questions are being informed by our attempts to draw some analogies between the main issues concerning the statistical modelling of asset prices and those concerning the theoretical modelling of the Brownian motion in Physics. In particular, we show that despite the obvious differences of the subject matters, the two phenomena share some striking similarities concerning the underlying causal mechanisms that generate their observed behaviors. These similarities by being subjected to common probabilistic interpretations, give rise to models sharing the same theoretical structure.

The analysis of the origins and the nature of statistical models of stock returns can be conveniently organised around two basic hypothesis, namely the Independence hypothesis (IN) and the Normality hypothesis (N). In a nutshell, the Independence/Normality hypothesis states that the returns of any asset across time can be thought of as realisations of time-independent/normal random variables. It is no exaggeration to say that the question of whether or not asset returns exhibit Independence and/or Normality has motivated the production of almost the totality of asset returns models that have been suggested in the literature since the beginning of the twentieth century. The questions revolving around the Independence and Normality hypotheses, that will be dealt with in the sequel, include the following: To begin with, why IN and N were deemed to be interesting hypotheses to test? Was IN consistent with the view that prices are determined by means of economic laws? Related to this, what kind of economic theory was consistent with IN? Did the acceptance of IN result in a radical change in the theoretical paradigm? How was IN thought to be related to the so-called Efficient Market Hypothesis (EMH)? Were IN and N related in some way? More specifically, did IN imply to some extent N? What was the role of the central limit theorem (CLT) for the emergence of N as a property deducible from IN? In view of the empirical evidence against N, and in favor of leptokurtic distributions that has been accumulating since the early 1960s, what were the potential explanations for leptokurtosis? Related to this, what kind of limit theorems predicted non-normal limiting distributions? What were the implications of the alternative explanations for leptokurtosis for the theoretical paradigm existed at the time, and especially for the definition and measurement of financial risk? Put differently, was leptokurtosis interpreted as evidence of excess (compared to normal) instability of the economic system itself or was it simply viewed as a manifestation of our inability to conduct controlled experiments, namely to keep the experimental condition constant in repeated trials? Within the time span under study, can we identify a specific period in which the search for explanatory models was succeeded by that for descriptive ones, whose sole task was to capture the empirical properties of stock returns as perceived and interpreted at that time? Did the weaning of EMH from IN contribute to the switch from explanatory to descriptive statistical modelling?

This paper is organised as follows: Sections 2 and 3 analyze the origins and implications of IN and N, respectively in the field of economics and compare them with the corresponding ones for the case of Brownian Motion in the field of physics. Section 4 surveys the main theories of "explanation of empirical regularities" that have been advanced in the philosophy of science since the mid of the twentieth century. In view of these theories, it asks the question of how a statistical model explains an empirical regularity, or in what sense a statistical model can be explanatory? This section focuses on potential explanations of "normality of stock returns". It also examines the extent to which the proposed statistical models can explain the occurence of a single event, for example that on a particular date the daily returns of a specific stock were less that -4%. Section 5 surveys the empirical evidence against the normality hypothesis that started to accumulate in the early 1960s. It also compares the two main competing "explanations" for the presence of non-normal lepetokurtic return distributions, namely the "infinite-variance" and "finite-variance" ones that were put forward in the early 1960s and early 1970s, respectively. The anatomy of these explanations together with the origins and realism of their assumptions are also examined and their implications for the existing theoretical paradigm are discussed. Section 6 focuses on the relationship between the probabilistic properties of stock returns process and market efficiency. In particular, it presents the arguments that led to the abandonment of IN as a necessary condition for EMH and its replacement by the weaker condition of "martingale difference". It also analyzes the effects of this replacement on the validity of the two alternative explanations for leptokurtosis mentioned above. Section 7 analyzes the recent develpments in statistical modelling of asset returns which center around the interpretation of volatility clustring as second-order temporal dependence. This interpretation led to the birth of the so-called GARCH models that have dominated the finance literature for the last thirty years. The explanatory status of these models is also analyzed in detail. Section 8 concludes the paper.

2 Independence

In this section, we examine the origins and implications of the assumption of Independence for the returns process. As will be discussed below, the introduction of IN to financial economics was made by a non economist and initially caused embarassement and confusion to the profession. Soon afterwards, however, IN was interpreted in a radically different way and eventually became the flagship of the new theoretical paradigm in financial economics.

2.1 Empirical Motivation and Theoretical Justification

The period during which stock returns data was mainly interpreted as realizations from an Independent and Identically Distributed (IID) process covered approximately the period 1953-1982. This interpretation stemmed from the following two main sources. First, the publication of the first systematic study on the statistical properties of stock returns data by the eminent statistician Maurice Kendall, in 1953. The second source was of theoretical nature and had to do with the increasing awareness of the economic profession on the anticipatory nature of asset prices and the proper definition of economic rationality associated with it. Let us analyze these two sources in detail.

Kendall (1953) analyzed a set of 22 asset price series, observed at a weekly frequency, which included both commodity and stock prices. One of his major objectives was to determine whether these prices exhibited systematic versus random behavior over time. It must be noted that in Kendall's era, the concept of probabilistic independence was implicitly equated to that of serial non-correlation. By failing to produce evidence against non-zero serial correlation coefficients, Kendall concludes that "the price series was much less systematic than is generally believed". In fact he went so far as to declare that "The series looks like a "wandering" one, almost as if once a week the Demon of Chance drew a random number from a symmetrical population of fixed dispersion and added it to the current price to determine next week's price" (1953, p. 13).

Kendall's results appeared to be rather surprising at the time. The independence of successive commodity price changes was rather difficult to be meaningfully interpreted within the existing theoretical paradigm, since it seemed to defy fundamental economic "laws". Samuelson (1973) describes the situation as follows: "...there are the "fundamentalists" and economists who think that the future algebraic rise in the price of wheat will have something to do with possibly discernible patterns of what is going to happen to the weather in the plains states, the price of nitrogen fertilizer, the plantings of corn, and the fad for reducing diets." (1973, p. 5). And also: "The economists who served as discussants for Kendall's 1953 paper were outraged, as he expected them to be, at the notion that there is no economic law governing the wanderings of price, but rather only blind chance. Such nihilism seemed to strike at the very heart of economic science." (1973, p. 18). Kendall's results seemed to echo in the field of economic theory Bertrand's (1889) famous question "is not chance the antithesis of all laws?"

However, soon after the publication of Kendall's paper, the attitude of the economics profession towards the concept of independence of successive price changes changed dramatically. This was mainly due to the theoretical developments that took place in economic theory. In particular, the economic science experienced a radical switch from a state in which "independence" was inconsistent with the existing background theory, to a new state in which "independence" was dictated by the new theory. Samuelson (1965) offered a formal proof of the statement that "properly anticipated prices fluctuate randomly". This new theoretical developments, under the guidance and influence of economists such as Samuelson and Fama, were eventually developed to a brand new theoretical paradigm which became known under the rubric of Efficient Market Hypothesis (EMH). At the heart of EMH was the "anticipatory" nature of speculative prices and the assumption that investors process available information rapidly and accurately.

Specifically, assume that at time t a piece of information, ϕ_t, relevant for the price P_t of a particular asset becomes available. According to EMH, investors will instantaneously and accurately process ϕ_t, thus correctly extracting its full implications for the future price P_{t+1}. Consequently, they form *rational expectations*, $E(P_{t+1} \mid \phi_t)$ for P_{t+1}. Based on this expectation, the investors (assumed, for simplicity, to be risk-neutral) will act now, that is, at period t, bringing the current price P_t equal to $E(P_{t+1} \mid \phi_t)$. Moreover, investors will act this way not only for the particular piece of information ϕ_t, but for any piece of marginal information contained in the information set Φ_t. At period $t+1$, new information, ϕ_{t+1}, becomes available, and the new expectations $E(P_{t+2} \mid \phi_{t+1})$ are formed, which in turn determine the price P_{t+1}. The price change $P_{t+1} - P_t$ is thus determined only by the new information ϕ_{t+1} since the full effect of ϕ_t was already fully reflected in P_t. Since new information is by its very nature random, meaning that the events ϕ_{t+1} and ϕ_t are probabilistically independent, the successive stock price changes $P_{t+1} - P_t$ will also be independent.

Moreover, the subjective expectations $\mathcal{E}(P_{t+1} \mid \Phi_t)$ are assumed to be rational, in the Muthian sense, that is, they coincide with the true objective expectation $E(P_{t+1} \mid \Phi_t)$ (see Muth 1961). The initial argument in favour of rationality had the structure of a "reductio ad absurdum" argument. Specifically, it shows that non-rationality cannot last for long since this would imply unexploited profit opportunities. Roberts (1959, p. 7) justifies rationality through the independence of price changes as follows: "If the stock market behaved like a mechanically imperfect roulette wheel, people would notice the imperfections and, by acting on them, remove them". The assumption of rationality in the sense that $\mathcal{E}(P_{t+1} \mid \Phi_t) = E(P_{t+1} \mid \Phi_t)$ is crusial for the independence property of stock returns. In fact it is easy to show that non-rationality is sufficient for non-independence. As will be discussed below, Bachelier (1900) was the first to follow this line of reasoning, which eventually led him to conclude that stock prices follow an absolute Brownian motion[1].

An interesting question is "why did Kendall choose to test for independence in price changes?" Or this question may be reformulated as follows: "why did a statistician choose to *systematically* test for independence in economic price changes while no economist had done so until then?" This question brings us to the realm of the philosophy of science literature and specifically to the relation between observation and theory. Popper ap-

[1]His analysis, however, despite its brilliance, failed to account for a fundamental asymmetry in asset prices. Namely, the fact that given limited liability the price of an asset can never become negative although is unbounded from above. Osborne (1959) and Samuelson (1965) rectified this problem by recasting the previous arguments in terms of the logarithms $p_t = \log(P_t)$ of prices, thus giving rise to the so-called relative or geometric Brownian motion for stock prices. Because the logarithmic price change is just a continuous transformation of R_t, the geometric Brownian motion also implies that the stock returns process $\{R_t\}$ is independent.

proaches this relation as follows: "In science it is *observation* rather than perception which plays the decisive part. But observation is a process in which we play an intensely *active* part. An observation is a perception, but one which is planned and prepared. We do not "have"' an observation but we "make" an observation. An observation is always preceded by a particular interest, a question, or a problem - in short, by something theoretical." (1972, p. 342). He also states explicitly that "An observation always presupposes the existence of some system of expectations" (1972, p. 344). As far as the sate of economic theory before 1953 is concerned, the independence hypothesis did not pose any particular interest. In other words, there existed no "system of expectations" concerning the theoretical role of the independence hypothesis that would create some interest for the econometricians to test this hypothesis. On the contrary, in the context of the existing theory at the time, the independence hypothesis seemed to be meaningless if not absurd. Economic theory could not and did not direct empirical testing (experiments) towards this direction. On the other hand, Kendall, being a statistician did not face the same restrictions. In fact, for a statistician the property of "independence" together with "Gaussianity" examined below, were (and still are) a kind of "benchmark" properties, that is, fundamental probabilistic properties defining "randomness", against which any "systematic" behavior is assessed and measured. Indeed, it seems plausible that Kendall was guided to conduct the aforementioned tests for independence by a theory of a different kind, namely from probability (instead of economic) theory. This assumption seems to derive support by examining Kendall's own reaction to his results. Instead of expressing discomfort, he appears to be rather "amuzed" by his evidence on independence and the related property of Gaussianity: "To the statistician there is some pleasure in the thought that the symmetrical distribution reared its graceful head undisturbed amid the uproar of the Chicago wheat-pit" (1953, p. 13).

The radical change of the theoretical attitude towards IN, analyzed above, was accompanied by a similar change to the frequency and clarity of observational statements concerning the detection of independence in empirical data. Indeed, under the guidance of the new theory of efficient markets, economists started a battery of statistical tests for independence which were not limited to estimating serial correlation coefficients (see, for example, Alexander 1961, Moore 1962). Other studies tested the independence hypothesis by equating the concept of independence with that of unpredictability, and examining the extent to which professional fund managers suceed in generating systematically abnormal returns (see Jensen 1968). Relatively soon a solid body of evidence in favor of the independence hypothesis was accumulated. Fama (1965) went so far as to declare "I know of no study in which standard statistical tools have produced evidence of important dependence in series of successive price changes" (1965, p. 57). This statement, however, seems to exaggerate on the level of aggreement that was achieved in the empirical literature with respect to the empirical validity of the independence hypothesis, since there were quite a few studies in which the evidence for independence was at best mixed (see, for example, Houthakker 1961, Cootner 1962, and Steiger 1964).

In concluding this section, it is of some historical interest to note that some studies suggesting that the independence hypothesis was consistent with rationality, existed in the literature even before 1953. Had the theoretical economists paid more attention to the studies by Working (1934), Taussig (1921), and especially Bachelier (1900), they would have been protected from the embarassement caused by Kendall's results. To this end, an important study on the anticipatory nature of asset prices, published in 1958 by Working, seems to have paved up the way for the emergence of the efficient market hypothesis.

2.2 Analogies with Brownian Motion in Physics: Part I.

The state of the economic theory during the period that lasted between the time of publication of Kendall's results and the time in which the efficient market hypothesis was sufficently articulated, that is, the period between 1953 and the early 1960s, was largely reminiscent of the situation that prevailed in theoretical physics in the period covering about fifteen years before and three years after Einstein's formulation of a quantitative theory for the Brownian motion in 1905. In both cases, an initial confusion concerning the interpretation of a set of observed regularities was succeeded by a neat and elegant explanation of them in the light of a newly arrived theory.

The phenomenon of Brownian movement or motion, that referred to the continuous erratic movements of sufficiently small particles suspended in a fluid, was observed for the first time in a systematic manner by Robert Brown in 1827. For a long period after Brown's discovery, Brownian motion attracted little attention by both theoretical and experimenal physicists. Some tentative explanations for it, which were eventually dismissed, included the one supporting the view that Brownian movement was analogous to the one performed by dust particles under the influence of air currents or the one suggesting that the movement is produced by forces external to the fluid, whithin which the particle was suspended (see Perrin 1909 for a survey of such explanations).

There was also considerable confusion about the basic properties or characteristics of the Brownian movement and the factors that are likely to affect it. The main source of the confusion was the lack of any guidance about what exactly was the object of observation. Granted, the facts should be objectively stated before any theorization of them was attempted, but what exactly were the *relevant* facts? How was the design of any given experiment affected by the theoretical dispositions of the experimenter? Would a physicist believing that the Brownian motion was caused by external perturbative factors observe and measure the same aspects of the motion with another physicist holding the view that the observed motion was due to molecular agitation? Brush (1968) comments on these issues as follows: "One can hardly find a better example in the theory of science of the complete failure of experiments and observation, unguided (until 1905) by theory, to unearth the simple laws governing a phenomenon" (1968, p. 23). In a similar tone, Maiocchi (1990) remarks: "This muddle of irreconcilable assertions illustrates how difficult it is to make a meaningful and conclusive scientific observation and, as a result, how any inductive conception, which claims to start from an empirical base in order to then construct theories of some importance, is unsustainable." (1990, p. 260).

Perrin (1909) argues that the first who ascribed Brownian motion to molecular agitation was Wiener in 1863. Similar explanations were put forward by Fathers Delsaulx and Carbonnelle during the period 1877- 1880, which, like Wiener's, remained largely unnoticed. On the contrary, Gouy's research in 1888 aroused great interest on the subject. As is well known, the decisive step in the theoretical explanation of Brownian motion was made by Einstein (1905), (and independently by von Smoluchowski 1906) who derived the mathematical laws governing the motion of a *free* particle on the grounds of the kinetic-molecular theory. Einstein's theoretical work had some important empirical implications. Specifically, it made clear that the relevant aspect of the motion that had to be measured was not the velocity of the particle but rather its *displacement* or more accurately the mean square value of this displacement. Indeed, it was this change of interest with respect to the object of observation that enabled Perrin in 1908 to offer Brownian motion its definitive confirmation. Maiocchi (1990) comments on this change in the object of observation effected by Einstein's theory as follows: "While previously the attempt had always been to estimate the length of the trajectory actually traversed by a particle, Einstein's theory deals with the displacement effected in a given time, i.e. the intervening distance between the points of departure and arrival, *independently of the path followed*." (1990, p. 263)

The keywords in the above statement are the italicized ones. Indeed, as in the case of price changes, similarly in the case of particle displacements, the key underlying concept is that of independence. As in the context of the efficient market theory, independence of price changes arises from the rational forward looking behaviour of economic agents, similarly in the context of Einstein's theory, independence of particle displacements arises from specific governing laws on the displacement of molecules (the molecular nature of matter). Perrin (1909) refers to independence of the displacements as "the most striking feature of the Brownian movement" (1909, p. 5). Indeed, as surprising and counterintuitive the idea of independent price changes initially appeared to economists, equally surprising the idea of independent displacements appeared to physicists. However, in both cases, the initial surprise was eventually followed by the realization that the independent price-changes/displacements were exactly what one should expect in the context of the appropriate theories. In the case of economic random walk, asset prices move over time because of their "continuous bombardement" with independent pieces of new information that become available to the investors. In the case of physical Brownian motion, particles suspended in a fluid move because of their "continuous bombardement" with independent non-coordinated molecules in agitation. Fama (1963) remarks: "At any point in time there will be many items of information available. Thus price changes between transactions will reflect the effects of many different bits of information" (p. 425). In a similar manner, Perrin (1916) makes the following statement: "Every granule suspended in a fluid [...] is being stuck continually by the molecules in its neighbourhood and receives impulses from them that do not in general exactly counterbalance each other; consequently it is tossed hither and thither in an irregular fashion" (1916, p. 86).

It must be noted that the similarities between the two theoretical developments in economics and physics discussed above, do not end in the fundamental role of the independence hypothesis ascribed to by both these theories. More striking similarities characterizing these two apparently quite diverse phenomena will be identified and analyzed below. For the moment, it suffices to add that as Einstein's theory made clear that the basic object for observation were the successive displacements of particles, similarly the efficient market theory suggested that the central object of observation must be the successive changes of prices.

3 Normality

In this section, we first examine the origins and implications of the assumption of Normality for the returns process and second, we compare them with the corresponding ones for the case of physical Brownian motion. We also analyze the connections of N with IN in both the economics and the physics cases.

3.1 Empirical Motivation and Theoretical Justification

Once the independence of asset returns was recognized to be a consequence of economic rationality, the property of Gaussianity of the distribution of returns immediately followed. This in turn was due to the probabilistic background theory available at that period, which was dominated by the Central Limit Theorem (CLT). More specifically, a consequence of the idea that an asset price is continuously bombarded by independent news is that (logarithmic) price change within a given interval, say a day, is the sum of the elementary returns from transaction to transaction occured within this interval. To this end Osborne (1959) argues as follows: "This nearly normal distribution in the changes of logarithm of price changes suggests that it may be a consequence of many independent random variables contributing to the changes in values. The normal distribution arises in many stochastic processes involving large numbers of independent variables, and certainly the market place should fulfill this condition, at least" (Osborne 1959, p. 151). Formally, the random variable R_t, denoting the returns of day t, may be thought of as the sum of the elementary rates of return ξ_{tj} in that day:

$$R_t = \sum_{j=1}^{n} \xi_{tj} \tag{1}$$

where n denotes the number of transactions in day t. The assumption of independence of the random variables ξ_{tj} together with some additional moment conditions, such as those in Feller (1935), allowed the application of the Central Limit Theorem (CLT), according to which the limiting distribution of R_t is the normal. Fama (1963) states this reasoning as follows: "If the price changes from transaction to transaction are independent, identically distributed random variables with finite variance and if transactions are fairly uniformly spaced through time, the central-limit theorem leads us to believe that price changes across differencing intervals such as a day, a week, or a month will be normally distributed since they are simple sums of the changes from transaction to transaction" (1963, p. 420). Normality was formally proved (in continuous time) by Osborne (1959) and Bachellier (1900).

Studies supporting the normality (or approximate normality) hypothesis for stock returns include Osborne (1959) and Larson (1960). Kendall's (1953) results were also supportive for the normality hypothesis in the price, rather than logarithmic price, changes. However, all these authors expressed with one way or another some reservations concerning the extent to which the normal distribution does in fact adequately fit the data. Osborne refers to the empirical distributions as "nearly normal" (1959, p. 129). Larson expressed more serious doubts by arguing: "The distribution ...has mean near zero, and is symmetrical and very nearly normally distributed for the central 80 per cent of the data, but there is an excessive number of extreme values. Also, some of these are quite extreme, being 8 or 9 standard deviations from the mean" (1960, p. 318). Although Kendall himself stated clearly that distributions look "very much like a normal form" (1953, p. 23), he nevertheless identified cases in which "The distributions are accordingly rather leptokurtic" (1953, p. 13).

3.2 Empirical Puzzles, Evidence of Non-Normality and the Role of Central Limit Theorem

Apart from the aforementioned comments on the approximate character of the normal distribution as a model for asset returns, there were studies whose results were substantially more negative for the normality hypothesis. For example, in commenting on Osborne's (1959) results, Alexander (1961) argues as follows: "But Osborne did not rigorously test the normality of the distribution. A rigorous test, for example the application of the chi-square test to some of the data used by Osborne, would lead us strongly to dismiss the hypothesis of normality" (1961, p. 16). This type of non-favorable results for the "'normality hypothesis" caused some confusion to the newly established efficient market paradigm.

Such confusion was completely absent in the period following Perrin's experimental results on the physical Brownian motion. Indeed, in the case of Physics, the normality hypothesis derived the strongest possible empirical support. What are the reasons for the presence of such a different level of empirical support between the normality hypothesis for the case of asset returns and the same hypothesis for the case of displacements of a particle under Brownian motion? In order to identify the origins of this difference we must first analyze in some detail the reasoning by which the normality hypothesis emerged in each of the two cases under consideration. Put differently, in order to explain why the normality hypothesis enjoyed sufficient empirical support in the case of the displacements of Brownian particles but not in the case of asset reurns, we must follow the assumptions that led to this hypothesis in each of the aforementioned cases, and identify which one(s) are likely to have failed in the case of asset returns.

First it must be noted that in the case of the displacements of a Brownian particle, normality can be obtained by using CLT arguments exactly as in the case of asset returns. Once independence of displacements is assumed, the normal distribution emerges quite naturally as the relevant probabilistic description of displacements. More specifically, in a discrete-time setting, assume that ξ_j denotes instead of the elementary returns in transaction j, the "elementary displacement" of a particle at step j, projected onto the horizontal axis. Then, $S_n = \sum_{j=1}^{n} \xi_j$ is the displacement of the particle from its starting point to its present state. As in the case of stock returns, the independence of the ξ_j's, together with some additional "moment conditions" on the ξ_j's allows for the application of CLT, which in turn implies that S_n, properly standardized, converges to $N(0,1)$. Kac (1947) has developed a discrete-time approach to Brownian motion, with the latter being treated as a "discrete random walk".

However, it is of some historical interest to note that in the beginning of the twentieth century, when Bachelier (1900) and Einstein (1905) independently arrived at the normality hypothesis, the central limit theorem was in its infancy. In fact, the relevant theorem was not even called "central" at the time. This name was given by Polya in 1924 to emphasize the central role of this theorem in the probability theory. At the beginning of the twentieth century, the most widely known version of CLT was that of De Moivre - Laplace, which could apply only to Bernoulli IID random variables. Although, the so-called "St Petersburg School" had already begun since 1887 with Chebyshev to extend CLT in various directions, including the case of general random variables, ξ_j, with infinite support, these results were not widely known at the time, especially among researchers outside the field of probability theory. Indeed in 1901, that is, one year after the publication of Bachelier's thesis and four years before the publication of Einstein's paper, Lyapounov (1901) had succeeded in producing a precise set of sufficient moment conditions under which CLT holds, as well as identifying the exact rate of convergence to normality. In the absence of knowledge of these results, however, both Einstein and Bachelier arrived at the normality hypothesis through different routes. Nevertheless, as will be shown below, the structure of their arguments largely resembles that of CLT.

3.3 Analogies with Brownian Motion in Physics, Part II: Einstein's Derivation of Normality.

In his analysis, Einstein (1905) introduces explicitly the independence hypothesis right from the start: "Evidently it must be assumed that each single particle executes a movement which is independent of the movement of all other particles; the movements of one and the same particle after different intervals of time must be considered as mutually independent processes, so long as we think of these intervals of time as being chosen not too small." (1905, p, 556, english translation: pp. 12-13)

The first thing to note in the above statement is Einstein's convinction that in order for the independence hypothesis to be empirically valid, the time interval, τ, between any two successive observations of the movements of the particle should not be *too small.* Why did Einstein impose sush a requirement, especially in view of the fact that he had no immediate or indirect experimental observations on that matter? It is well known that Einstein's ability in conducting "thought experiments" was unsurpassed. The above mentioned requirement seems to be another manifestation of this ability. More specifically, Einstein imposes the condition of τ being not too small in order to eliminate the possibility that the observed particle displacements exhibit any kind of *memory.* Put differently, if τ were extremely small, then the current position of the particle may be partly determined by its previously observed position, or in other words its current displacement may be probabilistically *dependent* on its previous one.

In what follows, we adopt Einstein's notation, namely τ denotes a very small (but not too small) time interval, n denotes the total number of particles suspended in the liquid and Δ denotes the change in the x-coordinates of the single particle that occurs within the interval τ. Einstein is interested in deriving the probability law for Δ. As a first step towards this direction, he assumes that the number dn of the particles which experience a displacement which lies between Δ and $\Delta + d\Delta$ within the time interval τ is of the form

$$dn = n\phi(\Delta)d\Delta$$

where

$$\int_{-\infty}^{\infty} \phi(\Delta)d\Delta = 1.$$

Next, he makes some important assumptions on the density function ϕ. Specifically, he assumes that

E1a: "ϕ only differs from zero for very small values of Δ" (1905, p. 556).

This assumption seems to be motivated by Einstein's physical considerations and especially by the belief that it is unlikely that the particles travel a long distance in a very small time, or in other words, that the movement of the particles exhibits discontinuities. E1a can be restated as follows:

E1b: Brownian paths are continuous.

Assumptions E1a and E1b imply that if Δ is not very small, the number of particles which exhibit a change in their x-coordinate larger than Δ within the time interval τ is practically zero. This fact, when translated to frequencies, implies that when Δ is not very small ϕ is almost zero. In modern probability language E1a may be interpreted as representing some kind of moment conditions on the random variable Δ. For example, it could be taken to imply that:

E1c: The random variable Δ has finite moments of any order.

This is a rather strict assumption. The following weaker assumption may ensure the continuity of Brownian paths:

E1d: The random variable Δ has finite variance.

It must be noted that all the four assumptions mentioned above may be thought of as the mathematical counterparts of the empirical hypothesis E1: "particle displacements cannot be arbitrarily large". It must also be noted that not many years after the publication of Einstein's paper, assumption E1d was found to be sufficient (under independence) to produce CLT. Moreover, as will be discussed in length below, the equivalent of E1d for asset returns was found to be at the center of a heated debate among economists during the 1960s. Finally, Einstein assumes $\phi(\Delta) = \phi(-\Delta)$, that is ϕ is symmetric.

Next, Einstein defines $f(x,t)$ to be the density of particles at time t with coordinate of the x-axis equal to x. This definition implies that for $n = 1$, and for $t = \tau$, the functions $f(\cdot,\tau)$ and $\phi(\cdot)$ coincide. In general, for $n = 1$, the quantity $\int_x^y f(z,t)dz$ corresponds to the probability that the change in the x-coordinate of a particle, from time t_0 to time $t_0 + t$, lies in $[x, y]$, for every t_0. Moreover, this means that $f(x,t)dx$ is the number of particles whose x-coordinate has increased between $t = 0$ and $t = t$ by a quantity, which lies between x and $x + dx$. Therefore,

$$f(x,0) = 0 \text{ and } \int_{-\infty}^{\infty} f(x,t)dx = n \tag{2}$$

At this point Einstein is ready to ask the important question: What is the distribution of particles at time $t + \tau$ given the distribution at time t? By using assumption E1a, Einstein obtains the following equation:

$$\frac{\partial f}{\partial t} = D\frac{\partial^2 f}{\partial x^2}, \tag{3}$$

which is known as "the heat equation in one dimension" or "the diffusion equation". The solution of (3) under conditions (2) is

$$f(x,t) = \frac{n}{\sqrt{4\pi Dt}}e^{-\frac{x^2}{4Dt}}. \tag{4}$$

Note that for $n = 1$ the above solution refers to the case of a single particle. Using physical arguments, Einstein showed that the coeffcient D, known as the "diffusion coefficient", which enters the expression for the second moment of $f(x,t)$ is related to a set, Θ, of some measurable physical magnitudes, as well as to the so-called Avogadro's number, N (the mean number of molecules per gram molecule), that is

$$D = g(N, \Theta). \tag{5}$$

From (4), Einstein concluded that the mean squared displacement $\overline{s}^2$ (the second moment) of the particle is given by

$$\overline{s}^2 = 2Dt. \tag{6}$$

The last equation, which is a consequence of the normality result (4), has some important empirical implications: (i) The average distance, $\overline{s}$, that a Brownian particle travelled along the x-axis is proportional to the square root of time. (ii) By replacing D in (6) with (5) and solving for N, Einstein derived an expression for N in terms of $\overline{s}$, Θ and t. This specific implication of normality, namely the possibility to determine N by means of

$\overline{s}$, Θ and t, formed the basis of Perrin's experimental results. These results demonstrated a close agreement between the experimental values of N and those predicted by the theory, which in turn rendered very strong support for the normality hypothesis. In fact this support was so impressive that automatically led to an almost universal agreement, not only about the empirical relevance of the normality hypothesis but, more importantly, on the basic hypothesis that produced this property, namely the molecular nature of matter.

3.4 Bachelier and the Normality of Asset Returns

In 1900, L. Bachelier presented his famous thesis with title "Theory.of Speculation". In his thesis, Bachelier attempted to produce a probabilistic model for stock price changes. He made the following initial assumptions:

B1: At any moment, the market does not believe that any stock price will go up or down.

B2: The mathematical expectation of the speculator is zero.

B3: The real price is the one that the market thinks as the most probable.

Assumptions B1 and B2 imply that both the market and any spaculator cannot predict the changes in the stock prices. These initial assumptions enable Bachelier to include in his treatment the following assumption**:**

B4: Stock prices have independent and identically distributed increments: "The probability that a price at time t lies between x and $x + dx$ given that y is the price at time $s < t$, is a function only of $x - y$ and $t - s$."

Bachelier defines by $p_{x,t}dx$ the probability that the price at time t lies between x and dx. Then, he explains what he aims at: "We are searching for the probability that price z is reached at time $t_1 + t_2$ and the price was x at time t_1." Bachelier identifies this probability with the conditional probability,

$$p_{x,t_1}dx \cdot p_{z,t_1+t_2}|_{(x,t_1)}dz$$

By employing B4, Bachelier obtains

$$p_{z,t_1+t_2}|_{(x,t_1)}dz = p_{z-x,t_2}dz$$

Hence the wanted probability he seeks equals to

$$p_{x,t_1}dx \cdot p_{z-x,t_2}dz$$

Integrating over $\mathbb{R}$, he finally obtains the probability that price z is reached at time t_1+t_2:

$$p_{z,t_1+t_2}dz = \int_{-\infty}^{\infty} p_{x,t_1}p_{z-x,t_2}dxdz$$

Last equation describes a required property of densities, namely

$$p_{z,t_1+t_2} = \int_{-\infty}^{\infty} p_{x,t_1}p_{z-x,t_2}dx \tag{7}$$

The derivation of a general expression that can describe any function satisfying (7), appears to be very difficult. One property that such a solution should satisfy is that of "stability under addition". The task of solving (7) would become much easier if one were willing to assume some moment-like conditions, such as E1a or E1b, assumed by Einstein. However, Bachelier did not realize the need for imposing such conditions. On the contrary, he proceeds with the solution of (7) by simply observing that this condition is satisfied by a very specific function of the form

$$p = Ae^{-B^2x^2} \tag{8}$$

where A and B are functions of t and

$$\int_{-\infty}^{\infty} pdx = 1,$$

therefore $B = A\sqrt{\pi}$. The last relationship between the parameters B and A makes clear that (8) represents the normal density function.

The preceding analysis implies that Bachelier, as opposed to Einstein, arrives at the normality hypothesis in a rather arbitrary way. Indeed, he failed to diagnose the need for making an empirical assumption similar to E1 that would have led to its theoretical counterparts E1c or E1d. Such an assumption should have stated that "stock prices are unlikely to change too much in a very short time". In the absense of any moment-like conditions, such as E1c or E1d, equation (7), is satisfied not only by (8) but by a whole class of density functions. As Paul Levy showed, almost twenty five years after Bachelier's results, the members of this class, usually referred to as the Levy-Pareto stable additive class, do not have finite second moments. The only member in this class which possesses finite variance is the normal. As a result all the members in the Paretian class, with the exception of the normal, do not produce continuous paths (see

Samuelson 1973, for similar arguments against Bachelier's solution). Nevertheless, despite the aforementioned limitation, Bachelier's thesis is full of pioneering results. In them we find a derivation of the diffusion equation (3). To obtain equation (3), Bachelier explicitly applies the technique used for the derivation of the equation of heat diffusion to the relevant probabilities.

As shown above, the methods by which Einstein and Bachelier arrived at the normality hypothesis are quite different than those developed later in the field of probability theory, under the rubric "central limit theory". However, the sufficient conditions imposed by Einstein, namely, independence of the summands together with some implicit conditions on their moments, are similar in nature to those obtained by Lyapounov (1901) in the context of the first CLT of the modern times. This type of direct CLT-related arguments for establishing the normality hypothesis for asset returns were employed for the first time by Osborne (1959). Moreover, Bachelier's failure to realise that (7) has a multiplicity of solutions, all of which - with the notable exception of the normal - have infinite variance, will persist in the economics literature until 1963, year at which Mandelbrot made known to the economics literature the non-Gaussian limit theorems obtained by Paul Levy in the mid 1920s (see below).

4 Explanation of Normality: What does it amount to?

As analyzed above, the Normality hypothesis of returns was a major pillar of the Brownian Motion model for the logarithm of asset prices. This model in its discrete time version may be referred to as the NIID model for asset returns. The question which arises at this point is the following: How does the NIID model *explain* the regularity or the stylized fact (perceived at the time) that the empirical distributions of asset returns are *usually* normal? More generally, how a statistical model explains an empirical regularity or in what sense a statistical model can be explanatory?

4.1 Deductive Nomological, Deductive Statistical and Inductive Statistical Explanations

As mentioned above, our central question in this section is "what constitutes an explanation of an empirical regularity?". In their widely known model of Deductive-Nomological (DN) explanation, Hempel and Oppenheim (1948) argue that an explanation of an empirical regularity is achieved by showing that it can be deduced (or follows with necessity) from a broader regularity or, in other words from one or more universal/statistical laws (and initial conditions in some cases). This means that an explanation of an empirical regularity is an argument to the effect that the regularity to be explained (the explanandum) was to be expected by reason of certain explanatory facts (the explanans) which include at least one more general regularity. Put it slightly differently, the explanation of the regularity of interest amounts to sub-suming it under a broader empirical law (or laws), which is usually referred to as the covering law(s). In the context of the DN model, the characteristic feature of explanation is that the explanadum is deducible from the explanans. Hempel draws the distinction between "actual" and "potential" explantion in an attempt to distinguish between "the logical structure of explanatory arguments" and the empirical validity of the explanans (see Hempel, 1965). In other words, a potential explanation is "a valid argument such that if it were also sound, it would explain the *explanandum*." (Psillos 2002, p. 221). On the other hand, an actual explanation is a valid sound (empirically adequate) argument that does explain the explanandum.

In the case under study, let us first define precisely the empirical regularity to be explained. First, consider the following statement, UN: "The empirical distributions of stock returns are Normal". This is a universal statement since it states that for every stock the empirical distribution of data $(R_1, R_2, ..., R_T)$ on returns of this stock is normal. UN may be contrasted with the following statistical statement, SN1: "The empirical distributions of stock returns are *usually* Normal". SN1 implies that for most stocks q, their empirical distributions are normal. SN1 differs from UN in that it allows for cases in which stocks have non-normal empirical distributions. SN1 may become more precise by stating the probability for a randomly selected empirical distribution to be normal. For example consider, SN2: "The probability for the empirical distribution of returns of a (randomly selected) stock to be Normal is p". In the context of the covering-law model, an explanation of UN is of the DN type, that is such an explanation is achieved when UN is deduced from one or more universal laws. On the other hand, an explanation of either SN1 or SN2 is usually referred to as Deductive-Statistical (DS) explanation and amounts to deducing SN1 or SN2 from one or more statistical laws (see also Hempel 1965).

Let us start with the universal regularity UN. The structure of a DN explanation of UN may be schematized in terms of the following scheme, S1:

L1: For all R_t and all t, if R_t were the returns of a financial asset at t, then R_t is the sum of n_t elementary returns ξ_{tj} that occured at t, that is $R_t = \sum\limits_{j=1}^{n_t} \xi_{tj}$.

L2: For all n_t and all t, if n_t were the number of elementary returns that occured at t, then n_t does not differ "substantially" from a large but fixed number n.

L3: For all $\left\{\xi_{tj}\right\}_t$ and all t, if $\left\{\xi_{tj}\right\}_t$ were the set of elementary returns at t, then each member of this set does not affect and is not affected by any other member of either this set $\left\{\xi_{tj}\right\}_t$ or of any other set $\left\{\xi_{sj}\right\}_s$, $t \neq s$, that is all the ξ_{tj}'s are "independent".

L4: For all $\left\{\xi_{tj}\right\}_t$ and all t, if $\left\{\xi_{tj}\right\}_t$ were the set of elementary returns at t, then each member of this set may be thought of as a random draw from a common distribution, that is all the ξ_{tj}'s are "identically distributed".

L5: For all $\left\{\xi_{tj}\right\}_t$ and all t, if $\left\{\xi_{tj}\right\}_t$ were the set of elementary returns at t, then each member of this set cannot be arbitrarily large.

UE1: "For all R_t and all t, if R_t were the returns of a financial asset at t, then R_t may be thought of as a random draw from a normal distribution.

Remarks:

(i) The formulation of the laws L1-L5 intentionally uses only empirical concepts. This is why the terms "independent" and "identically distributed" were put in inverted commas, to be distinguished from the theoretical concepts of independent and identically distributed random variables that will be employed at a later stage.

(ii) The empirical concept of "identically distributed" is difficult to visualise. One different way to describe this is that under any random reordering of the elementary returns, at any sub-interval, produces the same relative frequencies as n goes to ∞.

(iii) L1, L2 and L5 can be thought of as universal generalizations referring to events occuring at the same point in time t. On the other hand, L3 and L4 are universal generalisations referring to both events occuring at the same t and events occuring at different points in time (t and s). Of course, the term "successive" in the aforementioned "successive price changes" imply a further discretization of time within t, in which case L3 and L4 refer solely to events occuring at different points in time.

(iv) Apart from UE1, L1-L5 also entail the "independence" of R_t with any other return R_s occuring at s, with $s \neq t$. It also implies that all the R_t's are "identically distributed" in the sense defined above, given that for each t the number of elementary returns is about the same.

The five laws, L1-L5 produce deductively UE1. However UE1 is not equivalent to UN - the universal regularity to be explained. Indeed, a data set $(R_1, R_2, ..., R_T)$ in which each R_i, $i = 1, 2, ..., T$ is a random draw from a common normal distribution does not always yield a normal empirical distribution (histogram), especially when T is small. As a result UN cannot be explained by L1-L5. On the other hand, SN1 and SN2 can be deductively produced by L1-L5. More precisely UE1, which states that the returns R_t of every financial asset can be thought of as "independent" realisations from a common normal distribution, implies that any data set $(R_1, R_2, ..., R_T)$ of returns on a financial asset can be thought of as a data set of independent realisations from a common normal distribution. This in turn implies the following statistical law, SN2a: "The probability for the empirical distribution of a set of returns $(R_1, R_2, ..., R_T)$ of a financial asset to be Normal is p_T, with $p_T \rightarrow 1$ as $T \rightarrow \infty$", where by "Normal" we mean that the corresponding relative frequencies satisfy an empirical/statistical criterion which confirms that a sample is being drawn from the Normal distribution. More specifically, SN2a is the explanandum in the following DS argument:

SL6: The probability for the empirical distribution of a set $(x_1, x_2, ..., x_T)$ of independent realisations from a common normal distribution to be normal is p_T, with $p_T \to 1$ as $T \to \infty$.

UE1a: For all $(R_1, R_2, ..., R_T)$, if $(R_1, R_2, ..., R_T)$ were a set of returns on a financial asset, then $(R_1, R_2, ..., R_T)$ may be thought of as a set of independent realisations from a common normal distribution.

——————————————————————————— $[p_T]$

SN2a: The probability for the empirical distribution of a set of returns $(R_1, R_2, ..., R_T)$ of a financial asset to be Normal is p_T, with $p_T \to 1$ as $T \to \infty$.

Remark

SN2a can play the role of a covering statistical law for the explanation of the specific event E2: "The empirical distribution F_q of returns $(R_{q1}, R_{q2}, ..., R_{qT})$ of stock q is Normal". This explanation has the structure of the so-called Inductive-Statistical (IS) explanation of a particular non-deterministic event put forward by Hempel (1965). An IS explantion is still an argument (albeit not a deductive one) that makes the occurence of the event highly probable (although not certain). For the present case, the IS explanation of E2 takes the following form:

SN2a: "The probability for the empirical distribution of returns $(R_1, R_2, ..., R_T)$ of a financial asset to be Normal is p_T, with $p_T \to 1$ as $T \to \infty$"

IC: F_q is the empirical distribution of returns $(R_{q1}, R_{q2}, ..., R_{qT})$ of stock q.

——————————————————————————— $[p_T]$

E2: The empirical distribution F_q is Normal.

Note that $[p_T]$ is not a statistical probability but the inductive probability confered to the conclusion by the explanans, or in other words, the degree of inductive support.

In the explanatory scheme outlined above, we claim that the explanandum statement, UE1, is produced deductively from the five laws, L1 - L5. This claim calls for an explanation. More specifically, in view of the fact that it is far from obvious how L1-L5 entail UE1, the question which naturally arises is "how has the alleged deduction been achieved? The answer to this question, as it might have been clear from the analysis of the previous sections, is "via the central limit theorem". In order to exploit the deductive power of CLT, however, we must first make a transition from the "empirical concepts" present in the above scheme to probabilistic concepts present in CLT. In other words, an empirically interpreted theoretical probabilistic system has to be adopted. As a first step towards this direction, the elementary returns ξ_{tj} are interpreted as random variables defined on an abstract probability space $(\Omega, \mathcal{F}, P)$, which in turn allows for relative frequencies or "limiting" relative frequencies to be interpreted as probabilities defined on $\mathcal{F}$[2]. Next, the hypothesized or observed empirical properties characterizing the elementary returns, are translated in terms of probabilistic properties of the random variables ξ_{tj}. For example, the empirical assumption/observation that price changes from transaction to transaction are caused by "independent" factors, is translated into the probabilistic assumption of independence of the random variables ξ_{tj}. The empirical assumption/observation that elementary returns are not arbitrarily large is translated into the probabilistic assumption that the random variables ξ_{tj} have finite variances. The mapping between empirical and probabilistic concepts is completed, when all the laws, L1-L5 have been re-stated in terms of their probabilistic counterparts. Once this transformation is finished, we are ready to exploit the mathematical structure of our probabilistic system in order to produce theorems, such as CLT. These theorems, when re-interpreted back to the empirical world, take the form of new empirical statements, such as UE1, that may be thought of as being derived from the original ones.

An interesting question, which has attracted a lot of attention in the philosophy of science literature is whether the deduction of UE1 from its empirical premises L1-L5 can be achieved without the mediation of CLT. The relevant debate revolves around the so-called "indispensability argument" (see, for example, Melia 2000).

In order to appreciate further the explanatory power of the CLT-based explanation of normality of stock returns, hereafter referred to as CLT-EX, let us consider the following hypothetical situation: There is a person, P, who has just examined the histogram, HA, of daily returns of stock A (over a sufficiently long period of time) and found convincing evidence of normality. This person seeks an explanation for this empirical pattern exhibited by stock A. He consults a rather experienced econometrician who replies to him by saying that "It comes as no surprise to me, histograms of daily stock returns are usually normal". Or in a more precise form: "...,the probability that a histogram of daily stock returns is normal is p (with p sufficiently large)" Has P received a satisfactory explanation for the normality of HA? Consider the following IS explanation for the normality of HA, hereafter referred to as N-EX:

EL1A: "The probability that a histogram of daily stock returns is normal is p".

IC1: HA is a histogram of daily stock returns

———————————————————————— [p]

EXP: HA is normal.

[2]Whether relative frequencies satisfy the conditions of Kolmogorov's axiomatic system in order to be called "probabilities" is a debatable isue. For example Van Fraassen (1979) shows that relative frequencies do not satisfy countable additivity.

The questions to be discussed are the following: Should P be satisfied with the N-EX explanation for the normality of HA? How does N-EX compare with CLT-EX? The answers to these questions largely depend on what the origins of EL1A in N-EX are. If EL1A has emerged out of "inductive enumeration", then N-EX although satisfies the Hempelian criteria for an adequate IS explanation it cannot be regarded as a satisfactory scientific explanation of the normality of HA. In other words, the fact that N-EX renders EXP expectable with high probability is irrelevant for the scientific adequacy of this explanation. Woodward (2003) argues that explanations such as N-EX are not regarded by anyone, with the exception of philosophers, as serious scientific explanations "or at least they are not advanced as such in scientific textbooks and monographs" (2003, p. 190). Therefore, the inferiority of N-EX relative to CLT-EX lies in the fact that in the former as opposed to the latter explanation, the covering statistical law is inductively produced by simply enumerating the instances of observed normal histograms. In other words, in N-EX the covering law itself is left unexplained. On the other hand, in CLT-EX the covering law SN2a is the explanandum of a DS explanation whose explanans are SL6 and UE1a. Furthermore, UE1a is the explanandum of a DN explanation whose explanans are L1-L5.

4.2 Deductive-Nomological-Probabilistic Explanations

The superiority of CLT-EX can be further appreciated within the context of the Deductive-Nomological-Probabilistic (D-N-P) model proposed by Railton (1978, 1981). Railton rejects the basic thesis of the I-S model, namely that the explanadum event must be expectable with high probability. Instead, he puts forward the idea that a probabilistic event (or an empirical regularity) can only be explained in terms of the *mechanism* that produced this event (or regularity). "The goal of understanding the world is a theoretical goal, and if the world is a machine - a vast arrangement of nomic connections - then our theory ought to give us some insight into the structure and workings of the mechanism, above and beyond the capability of predicting and controlling its outcomes." (1978, p. 208). In other words, an IS explanation which employs a statistical law such as EL1A is unsatisfactory (even if EL1A is true) unless the statistical law is backed up with "an account of the mechanism(s) at work". In other words, the statistical law in itself cannot form the basis for a satisfactory explanation of the explanadum event, unless L is "derivable from our theory without appeal to particular facts." (1978, p. 215). As Railton puts it, D-N-P explanations "subsume a fact in the sense of giving *a D-N account* of the chance mechanism responsible for it, and showing that our theory implies the existence of some physical probability, *however small*, that this mechanism will produce the explanandum in the circumstances given." (1978, p. 209, emphasis added). In this respect Railton seems to follow Jeffrey (1969) who in his criticism of the IS explanation argues: "... in the statistical case I find it strained to speak of knowledge why the outcome is such-and-such. I would rather speak of *understanding the process*, for the explanation is the same no matter what the outcome: it consists of a statement that the process is a stochastic one, following such-and-such a law." (1969, p. 109). It is evident that in Jeffrey's as well as in Railton's views, the burden of explanation lies in "understanding the process" or "providing an account of the chance mechanism at work", respectively.

It must be noted that Railton's account of probabilistic explanation allows partial or incomplete explanations to qualify as adequate. For example, CLT-EX may be thought of as an incomplete explanation of the normality of HA since the origins of some of the laws L1-L5 are not specified. For example, one may ask why the elementary stock returns are independent or why cannot be arbitrarily large? What parts of the mechanism at work are responsible for L3 or L5? If these questions are left unanswered, then the offered CLT-EX explanation does not illuminate all the explanatory text but only parts of it , that is it furnishes explanatory information. In the limiting case, in which all the explanatory information is taken into account, the corresponding explanation is called "ideal" explanation.

Remark

Concerning, the origins of L3, we may appeal to the efficient market hypothesis which explains why elementary returns are independent. Put differently, EMH may be used to fill parts of the ideal explanatory text for the explanation of the normality of HA.

4.3 Counterfactual Dependencies

Another feature of CLT-EX that contributes to its superiority over N-EX is that it allows for "counterfactual dependencies". Woodward (2003) argues that explanations such as CLT-EX possess a distinctive explanatory feature that is missing from the N-EX variety. He identifies this feature by comparing two alternative types of explanation: the first one, referred to as (5.1.3)-(5.1.4) in Woodward's text, is the type of theoretical explanation which is typically found in scientific disciplines such as physics or economics; it largely corresponds to our CLT-EX. The second type of explanation, referred to as (5.1.1)-(5.1.2) in Woodward's text, has a structure identical to that of N-EX. His arguments on the explanatory superiority of CLT-EX over N-EX are summarised as follows (the reader may replace "(5.1.3)-(5.1.4)" with "CLT-EX" and "(5.1.1)-(5.1.2)" with "N-EX" in the following paragraph): "Explanation is a matter of exhibiting systematic patterns of counterfactual dependence. Not only can the generalizations cited in (5.1.3)-(5.1.4) be used to show that the explananda of (5.1.3)-(5.1.4) were to be expected, given the initial and boundary conditions that actually obtained, but they also can be used to show how these explananda would *change* if these initial and boundary conditions had changed in various ways. (5.1.3)-(5.1.4) locate their explananda within a space of alternative possibilities and show us how which of these alternatives is realized systematically depends on the conditions cited in their explanans. They do this by enabling us to see how, if these initial conditions had been different or had changed in various ways, various of these alternative possibilities would have been realized instead. Put slightly differently, the generalizations cited in (5.1.3)-(5.1.4) are such that they can be used to answer a range of counterfactuals questions about the conditions under which their explananda would have been different (what–if– things-had-been-different or w-questions, for short). In this way, (5.1.3) – (5.1.4) give us a sense of the range of conditions under which their explananda hold and of how, if at all, those explananda would have been different if the conditions in (5.1.3) – (5.1.4) had been different. It is this sort of information that enables us to see that (and how) the conditions cited in the explanans of (5.1.3) – (5.1.4) are explanatorily relevant to these explananda." (p. 191, 2003).

The distinctive explanatory feature that characterizes CLT-EX is exactly its ability to answer counterfactual or w-questions. For example, what would have been the distribution of stock returns if elementary transactions were allowed to be arbitrarily large? Also what would happen to the same distribution if the number of transactions over time exhibited substantial variation? The structure of the CLT-EX type of explanation can indeed answer such questions. This is because the counterfactual questions raised above correspond to different assumptions on the random variables ξ_{tj} and the number of transactions n_t. As will be discussed in the following sections, these alternative sets of assumptions formed the basis on which the statistical modelling of stock returns was developed in the period between early 1960s and late 1970s.

4.4 Explanations of Single Events

So far the analysis was mainly centered on explaining empirical regularities[3]. Let us now examine how the aforementioned explanatory accounts work in the case of single events. More specifically, we shall investigate whether and how the NIID model, as derived above, explains the occurence of a single event E, with special emphasis on the case in which E is a low-probability event. For example, consider the event E :"The returns of the stock q at $t = t_1$ were lower than -4%" or in an equivalent form, E :"$R_{q,t_1} < -0.04$". UE1 implies

[3]We have also discussed indirectly expanations of a single event, for the case in which the event refers to the observation of a regularity in a singular dataset. For example, consider the event E: The histogram of this specific dataset was found to be (approximately) normal.

the following statistical law: UE1*: "The probability of the returns R_t of a financial asset at time t to be less than x, $x \in \mathbb{R}$, is given by the normal law with mean and variance equal to μ and σ^2, respectively." Let us first try the following IS explanation, using UE1* as the statistical covering law:

UE1*: "The probability of the returns R_t of a financial asset at time t to be less than x, $x \in \mathbb{R}$, is given by the normal law with mean and variance equal to μ and σ^2, respectively." Put differently, $P_{(\mu,\sigma^2)}(R_t < x) = p_x$

IC2: R_{q,t_1} is the returns of the financial asset at t_1.

IC3: The mean μ_q and variance σ_q^2 of the distribution of R_{q,t_1} are equal to $\widehat{\mu}_q$ and $\widehat{\sigma}_q^2$, respectively, where $\widehat{\mu}_q$ and $\widehat{\sigma}_q^2$ are consistent estimates of μ_q and σ_q^2, respectively.

———————————————————————— $[p_{-0.04}]$

E : "$R_{q,t_1} < -0.04$"

Does this argument satisfy the conditions for an adequate IS explanation of E? The answer is no. For any sensible values of μ_q and σ_q^2, the inductive probability $p_{-0.04}$ is prohibitively small. In other words, the probability confered to the conclusion by the explanans or the degree of inductive support of E is too small. Nonetheless, the event E, expectable or not, has actually occured and has equal rights to enjoy some kind of explanation with any high-probability event that has also occured. To this end, Salmon (1998) raises the following question: "Can events whose probabilities are low be explained?" (1998, p. 97). In particular, he puts the question as follows: "If some events are probable, without being certain, others are improbable. If a coin has a strong bias for heads, say 0.9, then tails has a nonvanishing probability, and a small percentage of the tosses will in fact result in tails. It seems strange to say that the results of tosses in which the coin lands heads-up can be explained, while the results of those tosses of the very same coin in which tails show are inexplicable. To be sure, the head-outcomes far outnumber the tail-outcomes, *but is it not an eccentric prejudice that leads us to discriminate against the minority, condemning its members to the realm of inexplicable*?"

The preceding discussion suggests that an explanation of E cannot be of the IS variety. However, an explanation of E may be achieved by Railton's D-N-P model. As already mentioned, D-N-P does not view a high inductive probability as a necessary condition for a satisfactory explanation of a chance event. As mentioned above, the D-N-P model of explanation includes the following ingredients: (i) A statistical law governing the behavior of the explanandum event. In the case under study, the statistical law and the explanandum event are UE1* and E, respectively. (ii) A set of initial conditions, which in our case are IC2 and IC3. (iii) Derivation of the statistical law from our theoretical account of the mechanism at work,. In the case under study, the derivation of UE1* has been achieved in terms of L1-L5, that is $[(L1 - L5) \Longrightarrow UE1^*]$.

Remark

To be precise, the three elements mentioned above costitute an explanatory account for the *probability of the explanadum event*, in our case for $P_{(\mu_q,\sigma_q^2)}(R_{q,t_1} < -0.04)$ but not for the explanadum event itself. To allow for the explanation of the event we must supply the following: (iv) A *parenthetic addendum*, PA, to the effect that the explanadum event did occur.

According to the D-N-P explanation of E, outlined above, the explanatory account EA=(UE1*,(IC2,IC3),[$(L1 - L5) \Longrightarrow UE1^*$] ,PA) does not explain why E *had to* take place, nor does it explain why E *could be expected to* take place (see Railton 1978, p. 216). All that EA does is to explain why the low-probability event E actually took place.

In concluding this section, it is worth comparing the explanation of E discussed above with the following more pragmatic explanation, PE, of E given by a typical market analyst: "E occured because the level of the industrial production index that was announced at t_1 was much lower than expected". In other words, the price of asset q fell by more than 4% because of a large negative surprise in the industrial production index. How does the D-N-P explanation of E compare to PE? Put it differently, does D-N-P accomodate in any way PE? The answer is that D-N-P accomodates only partially and indirectly PE, through the EMH-based explanatory account of why the elementary returns are independent. More specifically, in the context of our example, the industrial production index may be thought of as a "fundamentals" variable, X_t, the current value of which affects the expectations for next period's stock price, thus determining the current price by the mechanism described in Section 2. In order to achieve an explanation of E which utilizes directly information about other variables (such as the industrial production index) we must employ models different than the NIID model. Such models have been put forward in the literature under the rubric "factor models for stock returns" (see, for example, Ross, 1977). Investigating how these models explain E is tantamount to analyzing how linear regression models can, in general, be explanatory. Such an analysis will take us outside the scope of the present paper[4]. In the present context, suffice it to say that the explanatory nature of these models depends on whether the employed factors are probabilistically causally relevant for the explanandum event. To this end, the Statistical-Relevance model for explanation, SR, suggested by Salmon (1971) may serve as a good starting point for developing the relevant arguments. More specifically, in the context of SR, an explanation of an event is no longer an argument but rather "an assemblage of factors relevant to the occurrence or nonoccurence of the event to be explained, along with the associated probability values" (Salmon, 1998, p.108). This model suggests that an explantion of E has been achieved when the probability of E conditional

[4]The explanatory status of this class of models is analyzed in detail in Koundouri et. al. 2012.

on all the relevant factors for E (industrial production, for example) has been obtained, regardless of whether this probability is high or low. This probability is interpreted as the relative frequency of E within the "broadest homogeneous" reference class, that is the reference class which is generated by partitioning the initial class (the one in terms of which the aforementioned low probability $p_{-0.04}$ is defined) solely by relevant conditions.

5 Non-Normal Distributions

As already mentioned in the beginning of the previous section, the empirical adequacy of the normality hypothesis for asset returns at the early 1960s did not enjoy the same degree of support with the normality hypothesis for the displacements of Brownian particles at the late 1900s. The situation prevailed at the time may be summarized as follows: A new theory for economic rationality, namely the efficient market theory, entailed the independence hypothesis of stock returns. This hypothesis, in turn, implied the Gaussianity of stock returns, via CLT. On the empirical front, researchers faced a rather embarassing situation in which the basic hypothesis, namely independence, enjoyed sufficient empirical support, but its immediate consequence, namely Gaussianity, was clearly lacking in this respect.

These thoughts lead us quite naturally to ask the following questions: In view of the fact that returns, R_t, are sums of independent random variables as indicated by (1) why does the CLT not work? One possible answer to this question might be that the number of summands, n_t, is not sufficiently large to ensure a close approximation of the distribution of R_t to the normal. This argument might have been convincing in the case that the object of observation was not the daily or weekly returns but rather five or ten minute returns. In fact, within a day, let alone a week, the number n_t of transactions (at least for highly liquid assets) is so large that the aforementioned argument completely lacks any empirical relevance.

5.1 Empirical Motivation and Theoretical Justification

It must be noted that the issue of the “small number of summands” raised above becomes quite important for the case of the physical Brownian motion when the time t, elapsed between the beginning of the motion and the current state of the particle is small. This issue was first raised by Einstein himself, and was dealt with in a definitive way by Ornstein (1917) and independently by Furth (1920). These authors derived a generalisation of (6) for all times, which states that the mean square displacement is in general a non-linear function of time, reducing to (6) when t is large.

The preceding discussion suggests that if the much desired reconcilation of independence with non-normality of asset returns were to be achieved, the reasons that caused failure of CLT (in the presence of independence) had to be identified. This identification depended on the following two questions: (i) Did the state of the art in probability theory at the beginning of the 1960s provide results showing the conditions under which a sum of independent random variables converges to a non-normal distribution? (ii) Were the econometricians at the time aware of such results? As will be shown below, the answer to the first question is affirmative whereas that of the second question is negative. In fact, probability theory had already identified at least two cases in which the limiting distribution of a sum of independent random variables is a non-normal distribution. The first of these cases is the one developed by Levy (1925) and, as already mentioned, is the main reason why Bachelier's solution of (7) was incomplete. This case was adopted in 1963 by Benoit Mandelbrot. However, the reconcilation between independence and non-normality put forward by Mandelbrot, did not come without a price for the existing theoretical paradigm.

5.2 Mandelbrot's Stable Paretian Hypothesis

In 1963, Benoit Mandelbrot expressed forcefully and without any reservations the argument that the distribution of stock returns was not Gaussian. Moreover, Mandelbrot offered an elegant explanation of the observed leptokurtosis of R_t, which - importantly- was consistent with the theoretically desirable independence hypothesis. However, as will be discussed below, Mandelbrot's interpretation did not come without any cost for the existing theoretical paradigm. Specifically, Mandelbrot argued that the only assumption that had to be made in order to obtain leptokurtosis is that (the independent) ξ_{tj}s have infinite variance. Mandelbrot was aware of the probability theory results of Levy (1925) according to which a sequence of independent and identically distributed sequence of random variables converges-in-law to the so-called Stable Paretian family with characteristic exponent α, a prominent member of which is the Gaussian distribution for which $\alpha = 2$. Quite importantly, the normal is the only distribution in this family with a finite variance (see also Khintchine 1933). The extent to which this sequence converges to the Gaussian or some other member of the Stable Paretian family, depends on whether the random variables of this sequence possess finite second moments. Consequently, in order for Mandelbrot to derive the desired result (convergence to a non-Gaussian leptokurtic Stable distribution) he had to abandon the assumption that the ξ_{tj}s have finite variances.

More specifically, the limit theorem employed by Mandelbrot refers to a sequence of independent random variables, in the case under study $\xi_{t1}, \xi_{t2}, ...$, for which (i) $Var(\xi_{tj}) =$

∞ and (ii) the so-called Pareto-Doebkin-Gnedenko conditions on the tail behaviour of the ξ_{tj}s hold. These conditions require that for $x > 0$, $F_X(x) = 1 - (1 + e_1(x)) \left(\frac{x}{x_1}\right)^{-a}$, where $e_1(x) \to 0$ as $x \to \infty$, and for $x < 0$, $F_X(x) = (1 + e_2(x)) \left(\frac{x}{x_2}\right)^{-a}$, where $e_2(x) \to 0$ as $x \to -\infty$. Conditions (i) and (ii) are necessary and sufficient for the validity of the stable central limit theorem, that is, for the distribution function of the random variables $\xi_{t1}, \xi_{t2}, ...$ to belong to the domain of attraction (DA) of a stable law, G_a.

The preceding discussion implies that the probability theory available at the time had already produced the necessary theoretical results for a potential explanation of the observed leptokurtosis. More importantly, this explanation did not have to sacrifice the independence hypothesis. The only change in the set of the existing assumptions that had to be made is to replace the finite variance assumption of the elementary stock returns with the infinite-variance one. To this end, Cootner (1964) speaks apologetically on behalf of the financial economists of the time about "... our guilt at our failure to appreciate the possibility of non-Gaussian central limit theorems ..." (p. 413).

Mandelbrot's explanation of the leptokurtosis of stock returns had a structure identical to the CLT-EX explanation of normality analyzed above. In fact, Mandelbrot maintained the laws L1-L4 and replaced only L5 with the following:

L5M: For all $\left\{\xi_{tj}\right\}_t$ and all t, if $\left\{\xi_{tj}\right\}_t$ were the set of elementary returns at t, then each member of this set can be arbitrarily large.

The replacement of L5 by L5M resulted in a drastic change in the explanandum. Instead of UE1, L1-L4 plus L5M deductively produce the following UE1M.

UE1M: "For all R_t and all t, if R_t were the returns of a financial asset at t, then R_t may be thought of as a random draw from a stable Paretian distribution.

Remark:

The case analyzed above may be thought of as an instance of how a satisfactory scientific explanation exhibits the virtue of "counterfactual dependencies".

How inoccuous was this change for the existing theoretical paradigm? What were the implications of the infinite-variance assumption for the concept and measures of financial risk? These questions will be dealt with in detail in subsequent sections. For the moment, suffices to say that economists did not rush to embrace Mandelbrot's interpretation, despite the fact that this interpetation left the independence assumption intact. Cootner (1964) summarises the discomfort that the infinite-variance assumption caused to the academic community as follows: "Mandelbrot, like Prime Minister Churchill before him, promises us not utopia but blood, sweat, toil and tears. If he is right, almost all of our statistical tools are obsolete - least squares, spectral analysis, workable maximum likelihood solutions, all our established sample theory, closed distribution functions. Almost without exceptions, past econometric work is meaningless" (Cootner, 1964, p. 337). Although Mandelbrot had succeeded in producing an empirically adequate model for stock returns, his model was not fully compatible with the emerging paradigm of "efficient markets with controllable risk" and its adoption would have meant the collapse of a substantial part of the paradigm itself. This was a rather unwelcome outcome as "surely, before consigning centuries of work to the ash pile, we should like to have some assurance that all our work is truly useless. If we have permitted ourselves to be fooled as long as this into believing that the Gaussian assumption is a workable one, is it not possible that the Paretian revolution is similarly illusory?" (Cootner 1964, p. 337).

It is worth noting that apart from leptokurtosis Mandelbrot was the first to detect another empirical regularity of asset returns, namely that "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes." (Mandelbrot, 1963, p. 418). This regularity, usually referred to as "volatility clustering" went largely unnoticed for almost two decades, that is, between the early 1960s and early 1980s. This effect was usually referred to as "curious behaviour of volatility" and was thought of as another manifestation of the "infinite variance" effect or, alternatively, as a symptom of the non-uniformity of transactions over time (see below). As will be discussed in the last section of the paper, a re-interpretation of this regularity formed the basis for the emergence of a new statistical paradigm for describing asset returns, at the heart of which was the concept of "conditional heteroscedasticity".

5.2.1 The "Silent" Decade: 1963-1973

Despite Cootner's objections mentioned above, Mandelbrot's clear statement on the presence of leptokurtosis in the empirical distributions of stock returns, along with his interpretation of this leptokurtosis delivered a severe blow against the "approximate normality" view, which prevailed in the literature up to 1963. His analysis forced the econometricians to look the naked truth about leptokurtosis and recognise it as a clear and distinct empirical regularity of stock returns. In his comment on Mandelbrot's paper, Cootner (1964) describes the situation as follows: "Dr Mandelbrot's series of papers on the application of Paretian distributions to economic phenomena has forced us to face up in a substantive way to those *uncomfortable* empirical observations that there is little doubt most of us have had to sweep under the carpet up to now." (1964, p. 333). As a result, econometricians chose to remain silent on this issue until they were able to come up with an explanation of this leptokurtosis that does not have to abolish the finite-variance hypothesis. In fact, between 1963 and 1973, the only published papers on the type of stock returns distribution were by Fama (1965), Fama and Blume (1966), Teichmoeller (1971) and Officer (1972), all of which advancing further either theoretically or empirically the Stable Paretian hypothesis.

The situation described above seems to suggest that the majority of the economists at the time chose to remain silent on the issue of distribution of asset returns, hoping that a finite-variance explanation of leptokurtosis is "out there" waiting to be found. Methodologically, this situation seems to conform to what Stanford (2006) termed as "the problem of unconceived alternatives". According to this view, at the current stage of scientific progress, in which an existing theory is endorsed by the majority of the scientific community, there are radically different theories that remain unconceived, which are at least as well-supported by the empirical evidence as the current theory. Eventually, one or more of these alternative theories are formulated, found empirically adequate and gain the acceptance of (some parts of) the scientific community. As a result, the old theory becomes obsolete despite the fact that once had enjoyed empirical support. This in turn implies that every theory or hypothesis, including the infinite-variance one, is in principle "underdetermined" by any kind of finite empirical evidence. As will be discussed below, such a radically alternative theory of leptokurtosis was eventually "conceived" almost ten years after Mandelbrot's infinite-variance theory.

5.3 A "Finite-Variance" Explanation of Leptokurtosis

The long awaited "finite-variance" explanation of leptokurtosis finally came in the beginning of 1970s with the works of Praetz (1972) and Clark (1973) although the roots of the basic idea can be traced in Press (1968). These authors followed a line of reasoning similar to that of Mandelbrot. In particular, they also recognised the fact that since R_t is the sum of independent random variables ξ_{tj}, the non-normality of the distribution of R_t implies failure of CLT. But instead of assuming that the CLT failure was due to the non-finiteness of the variance of ξ_{tj}, they put forward the hypothesis that CLT failure was due to the *randomness* of the number of summands, n. Let us analyze Clark's ideas in slightly more detail: As Mandelbrot employed non-standard limit theorem results to interpret the observed leptokurtosis, so did Clark. However, instead of employing Paretian limit theorems, Clark employed limit theorems for *random* sums of random variables, that had appeared in the probability literature since 1948. More specifically, for each t, let $\{\xi_{t,j}\}_{i\geq 1}$ be an iid sequence of random variables with finite $E(\xi_{t,j}) = \mu_\xi$ and $Var(\xi_{t,j}) = \sigma_\xi^2 > 0$. Moreover, let $\{N_n\}_{n\geq 1}$ be a sequence of non-negative, integer-valued random variables. The random sum process is defined as

$$R_{t,N_n} = \sum_{j=1}^{N_n} \xi_{t,j}. \tag{9}$$

The question that was raised in the relevant probability theory was the following: Under what conditions does the properly normed and centered random sum, R_{t,N_n}, converge in law to some random variable, Z, and, further, under what additional conditions is Z distributed as $N(0,1)$. Robbins (1948) obtained sufficient conditions for the convergence in law of the properly centered and normed sequence, R_{t,N_n}, to the normal distribution, under the assumption that N_n is independent of the summands, $\xi_{t,1}, \xi_{t,2}, ...$ Renyi (1960) and Blum, Hanson and Rosenblatt (1963) derived sufficient conditions that are similar to those of Robbins (1948) without the assumption of independence between N_n and the summands. Whether the properly centered and normed sequence, R_{t,N_n} converges to the $N(0,1)$ or not depends on the variability of N_n around n as n increases. Specifically, if N_n exhibits a "substantial rate of variation" around n, in the sense that the condition

$$p \lim_{n \longrightarrow \infty} \frac{N_n}{n} = 1 \tag{10}$$

is violated, then the limiting distribution of the properly centered and normed sequence, R_{t,N_n} is not the standard normal and depends on the distribution of N_n. Clark assumes that $N_n = [Zn]$ where Z is a random variable with mean 1 and variance $\Gamma > 0$ and $[]$ denoting "the largest integer less than". It is easy to show that under the aforementioned assumptions on the random variables $\xi_{t,j}$, and for any given realization of Z, we have that

$$\frac{1}{\sigma_\xi \sqrt{n}} R_{t,N_n} = \sqrt{\frac{[nZ]}{n}} \frac{1}{\sigma_\xi \sqrt{[nZ]}} \sum_{j=1}^{[nZ]} \xi_{t,j} \xrightarrow{L} N(0, Z).$$

We observe that the variance of the limit distribution is the random variable Z, hence the unconditional limiting distribution is a mixture of normals. Such a distribution may well be leptokurtic. Clark (1973) justified the appeal to random limit theory on the grounds that transactions are not spread uniformly across time but instead display substantial variation.

The concept of "substantial variation" is central in delivering non-normal limiting distributions. Put differently, randomness of N_n *per se* does not ensure convergence to a non-normal distribution. As mentioned above, if (10) is satisfied then the limiting distribution of the properly centered and normed sequence, R_{t,N_n} is $N(0,1)$. The importance of (10) calls for a further ellaboration of the concept of substantial variation. More specifically, condition (10) imposes restrictions on the variability of N_n around n, as n increases. This property, referred to as "theoretical variation" refers to the relationship, $g(n)$, between the variance of N_n and n, that is $Var(N_n) = g(n)$. It is the functional form of $g(n)$ that affects the convergence properties of the random sum sequence. Koundouri and Kourogenis (2011) assume that N_n is defined by

$$N_n = n^r U + n, \text{ for some } 0 \leq r \leq \frac{1}{2} \tag{11}$$

where U is a random variable with $E(U) = 0$ and $Var(U) = c > 0$. This assumption implies the following function $g()$:

$$g(n) = Var(N_n) = cn^{2r} \ . \tag{12}$$

It can be shown that for $0 \leq r < \frac{1}{2}$, that is, when theoretical variation is moderate, the properly centered and normed sequence R_{N_n} converges to $N(0,1)$. On the other hand, for $r = \frac{1}{2}$, that is, when theoretical variation is substantial, the limiting distribution is a mixture of normals.

Once a "finite-variance" explanation of leptokurtosis was available in terms of substantial variation in the number of elementary transactions across days, the empirical literature on the type of stock returns distribution took off. Blattberg and Gonedes (1974) assume that Z^{-1} follows a gamma-2 distribution which implies that the resulting limiting distribution of $\frac{1}{\sigma_\xi \sqrt{n}} R_{t,N_n}$ is the student. Kon (1984) offers evidence in favor of the assumption that the stock returns distribution is a discrete mixture of normals.

As was the case with Mandelbrot's, Clark's explanation of the leptokurtosis also had a structure identical to CLT-EX. Specifically, Clark maintained L1, L3, L4 and L5 but replaced only L2 with the following:

L2C: For all n_t and all t, if n_t were the number of elementary returns that occured at t, then n_t may exhibit substantially variation across t.

Substituting L2 for L2C while maintaining L1, L3, L4 and L5 results in,

UE1C: "For all R_t and all t, if R_t were the returns of a financial asset at t, then R_t may be thought of as a random draw from a Mixed Normal distribution.

5.4 Realism of the Assumptions: Methodological Implications, Theoretical Consequences and Empirical Testing

The main difference between Mandelbrot's and Clark's models concerns the questions of how "large" elementary returns and how "volatile" the number of transactions over time can be. Mandelbrot's main assumption is the following: M1: "elementary returns can be arbitrarily large". He also makes an implicit assumption claiming that the number of transactions is constant over time. On the other hand, Clark's fundamental assumption can take the following form: C1: "the number of transactions varies substantially across time", while his maintained assumption is that elementary returns cannot be arbitrarily large.

In this section we will be concerned with the following five issues/questions, that aim at comparing various aspects of Mandelbrot and Clark's explanations of leptokurtosis:

(i) The first issue that needs to be addressed at the outset is that Mandelbrot's explanation of leptokurtosis, as opposed to that of Clark, seems to be somewhat circular. Specifically, by assuming M1, Mandelbrot in fact assumes that the elementary returns ξ_{tj} (being arbitrarily large) are leptokurtic. This means that Mandelbrot assumes leptokurtosis while aiming at explaining it. This feature is absent in Clark's explanation since he does not make any moment or distributional assumptions, implicit or explicit, on the elementary returns.

(ii) Are the assumptions M1 and C1, introduced in the previous section, directly testable? One may argue that both M1 and C1 are not sufficiently precise conditions so that can be tested empirically, because of the presence of the vague terms "arbitrarily" and "substantially" in M1 and C1 respectively. However, the probabilistic interpretations of M1 and C1, namely, M1P: "the random variables ξ_{tj} are Stable Paretian with characteristic exponent $\alpha < 2$" and C1P: "Condition (10) fails", respectively, are clear and unambiguous mathematical statements. As a result, M1P and C1P can be used in order to develop formal tests for M1 and C1, conditional on the assumption that M1 and C1 are adequately represented by M1P and C1P, respectively. For example, if data on elementary returns, ξ_{tj}, were available, one could design a statistical test for the hypothesis that the characteristic exponent α of ξ_{tj} is equal to 2, against the alternative hypothesis $\alpha < 2$. Similarly, if time series data on the number of transactions were available, one could use equation (11) in order to develop a test for the hypothesis $r = \frac{1}{2}$ against the alternative $r < \frac{1}{2}$.

(iii) Can we distinguish between M1 and C1 not by direct testing procedures, as the ones described above, but by means of their implications? Leptokurtosis alone does not offer much help towards this direction since both the aforementioned properties entail leptokurtosis. Additional empirical implications for each of the two competing assumptions should be identified, thus enabling an indirect comparison between M1 and C1. More specifically, if we were able to obtain that M1 entails leptokurtosis plus "property A" whereas C1 entails leptokurtosis plus "property B" with A being distinctly different (incompatible) with B, then a comparison between M1 and C1 would have been possible. This type of indirect testing between M1 and C1 has been attempted in the literature. Specifically, apart from leptokurtosis, M1 and C1 entail the properties of "stability under addition" and "Aggregational Gaussianity", (AG) respectively. The empirical implications of these properties are quite different. Stability under addition implies that the returns distributions for various frequencies of observation (daily, weekly, monthly etc) must exhibit the same degree of leptokurtosis. On the other hand, AG implies that the returns distribution comes closer to the normal as the frequency of observation decreases. The evidence from these tests seem to support the AG property, which in turn lends support to C1 over M1.

(iv) Is the aforementioned task, namely, examining the *realism* of M1 or C1 a useful or even meaningful endeavor at all? For example, a proponent of the "instrumentalist" view of the goals of science may protest against this project by saying that the realism of a given assumption is not important so long as this assumption contributes in making correct and useful predictions for the phenomenon of interest. In the field of economics, this view was forcefully put forward by Milton Friedman (1953) in his defense against the accusation that the fundamental assumptions of neoclassical economic theory were clearly "unrealistic". Friedman's view may be roughly summarized in the following paragraph: "Truly important and significant hypotheses will be found to have "assumptions" that are wildly inaccurate descriptive representations of reality, and, in general, the more significant the theory, the more unrealistic the assumptions (in this sense).The reason is simple. A hypothesis is important if it "explains" much by little, that is, if it abstracts the common and crucial elements from the mass of complex and detailed circumstances surrounding the phenomena to be explained and permits valid predictions on the basis of them alone. To be important, therefore, a hypothesis must be descriptively false in its assumptions" (Friedman, 1953, p. 14). By transferring Friedman's arguments to the case at hand, one could say that it is indifferent whether M1 or C1 are true, since they both predict a leptokurtic distribution for asset returns. To the extent that leptokurtosis is all that matters, both M1 and C1 are equally useful no matter how "unrealistic" or "vague" they might be. If one insists on deciding which of the two, possibly unrealistic, assumptions to adopt, she will have to rely on criteria that do not assess the realism of the aforementioned assumptions- for example, criteria of "simplicity". In any case, the instrumentalist may strengthen her case by appealing to the "problem of unconceived alternatives" mentioned in a previous section. More specifically, according to this view there is no point in trying to establish which of M1 or C1 is true because, most likely, neither of them is. The fundamental problem of "recurrent theory underdetermination", analyzed by Stanford (2006), means that neither M1 nor C1 are entitled to claim the role of the accurate description of the deep structure of reality, since in due course a hitherto unconceived alternative, A1, will be put forward, which will be found to be at least as empirically adequate as M1 or C1, and which will eventually replace both of them. This alternative will be such that "A1 plus not-M1 plus not-C1" entails leptokurtosis. Hence, we should not put any special epistemic weight on the distinction between M1 and C1.

This negative attitude towards examining the realism of the premises that led to a specific conclusion (in the present case leptokurtosis) is translated into a negative attitude towards the concept of explanation itself. Indeed, Friedman has stated clearly that economists should not seek explanations but only useful predictions. This attitude means that the whole idea of finding an explanatory model for asset returns is futile and therefore we should focus our efforts towards specifying a descriptive (or in Box's terms an empirical) model instead. In such a case the asset returns data should be interpreted as realizations of a "nameless stochastic process" (Spanos, 2006) while the econometrician's objective should be confined to selecting the statistical model solely in terms of "goodness-of-fit" criteria[5].

(v) Were M1 and C1 empirically motivated? Did M1 and C1 come first and then were translated into their probabilistic countrparts M1P and C1P, respectively, or the other way round? In other words, it seems quite possible that M1P and C1P were conceived before M1 and C1, respectively, since these probabilistic assumptions are all that was required for non classical limit theorems to work. For example, without claiming that we can read anyone's mind, it seems quite likely that Mandelbrot first realised that what was required to produce leptokurtosis was M1P. Once he realised that, he adopted M1P simply because "it worked mathematically", and then translated M1P into its nominalistic counterpart M1 at a later time. In other words, he developed a "theory" out of a "theorem" by simply translating the conditions of the mathematical theorem into empirical laws. In doing this, he chose among alternative "realities" the one that fit his (unique) model rather than among alternative models the one that fit the (unique) reality. As Nagel (1929) puts it "The history of thought is replete with attempts to lay the universe on the procrustean bed of one of several elements of analysis" (1929, p. 169). These considerations further support the view that both M1 and C1 emerged simply as "assumptions that work" in accounting for leptokurtosis and as such they raise no serious claims to realism. Of course similar criticisms may apply to Einstein's motivation of assumptions E1a-E1d. Were these assumptions motivated by the mathematical need to eliminate the higher-order terms from (3)? Or did the empirical condition E1 emerge first out of physical considerations which rendered "arbitrarily large particle displacements" impossible and then E1a-E1d emerged simply as the mathematical counterparts of E1?

It must be noted that at least until the begininning of the 1980s the literature did not adopt the aforementioned instrumentalist or neo-instrumentalist views. On the contrary, there was a considerable effort in the literature to empirically document the realism of

[5]To this end, Spanos (2006) defines a statistical model as "an internally consistent set of probabilistic assumptions aiming to provide an idealized probabilistic description of the stochastic mechanism that gave rise to the observed data" (2006, p. 98). In a similar vein, Lehmann (1990) suggests that the specification of a statistical model should make no use of substantive subject-matter information if a general approach to statistical modelling is to be attained.

C1 over M1. The reason behind this movement lies in the unfavorable or, in some cases, detrimental implications of M1 for the theoretical paradigm prevailing at the time. Fama (1963) identifies the following implications: (i) EMH implies that any price changes is the result of the reaction of rational investors to the arrival of new information. As a result, the arbitrarily large price changes implied by M1 are caused (under EMH) by arbitrarily large changes in the economic factors affecting prices. This, in turn, implies that under M1 economy as a whole is a much more volatile system than would be the case under a finite-variance alternative (such as C1). (ii) Asset markets under M1 are inherently more risky than under a finite-variance alternative. Fama argues that this is the reason behind the apparent reluctance of investors to invest a larger proportion of their wealth in speculative assets than the proportion which is actually observed. (iii) The infinite-variance implication of M1 means that the measures of risk that had been adopted by the literature, which are defined in terms of the second moments of the returns distribution, were clearly inappropriate and had to be replaced by new ones. This in turn implies that the landmarks of finance theory at the time, such as the portfolio theory developed by Markowitz (1952) and the Capital Asset Pricing Model (CAPM) developed by Treynor (1962), Sharpe (1964), Lintner (1965a and 1965b), Mossin (1966), etc., had to be reformulated in terms of the new measures of risk. Moreover, the robustness of their theoretical predictions with respect to these new measures of risk had to be examined. In short, the implications of M1 appeared to be pervasive to almost all the major theoretical concepts and models of the finance theory that had been developed up to that point. To this end, Fama (1965b) shows that under M1 one of the main implications of portfolio theory, namely that diversification reduces portfolio risk, is weakened or even reversed. Indeed, in a Stable Paretian market with the characteristic exponent α being less than unity, increasing diversification causes the portfolio risk (as measured by the scale parameter of the Stable Paretian distribution) to increase.

The preceding discussion suggests that if taken on face value, M1 and C1 are radically different assumptions describing diametrically different economic systems. M1 implies that economy as a whole is inherently riskier than it would be under the finite-variance alternative. In other words, M1 strikes directly to the DNA of the free economy, thus having far-reaching ontological and epistemological consequences. On the contrary, C1 is a rather inauccuous assumption which may be thought of as expressing our inability to conduct "controlled experiments" with fixed experimental conditions. Indeed, if the econometrician could control the uniformity of transactions the same way as the physicist can control the uniformity of the surrounding fluid (by keeping, for example, the temperature and the viscosity of the fluid constant across her measurements) then the normality of the returns distribution would have been obtained. To this end, it is interesting to note that Einstein's result on the normality of the distribution of particle displacements no longer holds in the case of a non-uniform fluid. Chapman (1928) demonstrates that if the temperature, composition or any other property affecting the diffusion coefficient D are not uniform then the emerging probability density function $P(x_0 \mid x; t)$ is not Gaussian. In fact, this distribution contains additional terms which in certain cases imply leptokurtosis.

6 Market Efficiency does not Require Independence: Martingale Difference

On the finance theory front, a major contribution took place around the mid 1960s, regarding the probabilistic properties of the returns generating process that are required for market efficiency. In particular, Samuelson (1965) and Fama (1965) advanced the idea that independence is too strong a property for market efficiency; instead, the much weaker concept of martingale difference for stock returns is all that is required for market to be efficient. As Samuelson (1973) puts it, the random walk model should be generalized to "an unbiased profitless-in-the-mean fair game" (1973, p. 12). It is worth noting that Mandelbrot (1966) expressed the same view. As already explained in some detail above, market efficiency means that prices at any time t fully reflect all information available at t. If some new information appears in the market at time t, then this information is processed by the market instantaneously and accurately so that this particular piece of information is incorporated into the (logarithm of) current price, p_t, without delay. As a result, since all the available information, Φ_t, has already been incorporated in p_t, the best forecast, $E(p_{t+1} \mid \Phi_t)$, for tomorrow's price p_{t+1} is the current price p_t itself, that is,

$$E(p_{t+1} \mid \Phi_t) = p_t. \tag{13}$$

Equation (13) constitutes the definition of a martingale sequence. From this, it follows that

$$E(R_{t+1} \mid \Phi_t) = E(p_{t+1} \mid \Phi_t) - p_t = 0 \tag{14}$$

which in turn implies that the returns form a martingale difference sequence, put differently, a fair game. Since a martingale difference may exhibit temporal dependence arising through higher moments, it is obvious that market efficiency does not require independence of the returns process, although of course, is consistent with it.

It must be noted that (14) defines efficient markets with risk-neutral investors. If the investors are risk averse, they would require compensations, r_t and ρ_t for the time value of money and the (systematic) risk, respectively. In the case that r_t and ρ_t are both constant over time, i.e. $r_t = r > 0$ and $\rho_t = \rho > 0$, then (14) becomes

$$E(R_{t+1} \mid \Phi_t) = r + \rho \tag{15}$$

which in turn implies that the returns process is a submartingale difference sequence. However, if the time-invariance of r_t or ρ_t does not hold then the martingale (or submartingale) difference property of stock returns may also fail to hold. For example when the risk premium, ρ_t, is time-varying and serially correlated, then $E(R_{t+1} \mid \Phi_t)$ depends on Φ_t in a systematic way thus violating the martingale-difference property of $\{R_t\}$ regardless of whether market is efficient or not. This in turn implies that the martingale-difference property is not necessary for market efficiency. This is the point of departure between the concepts of "martingale difference" and "market efficiency". In other words, the martingale difference is a joint hypothesis consisting of two individual hypotheses, namely, market efficiency and a specific form of risk aversion. Campbell, Lo and MacKinlay (1997) state this problem as follows: "First, any test of efficiency must assume an equilibrium model that defines normal security returns. If efficiency is rejected, this could be because the market is truly inefficient or because an incorrect equilibrium model has been assumed. This joint hypothesis problem means that market efficiency as such can never be rejected." (1997, p.24).

6.1 Distributional Implications of Non-Independence

The replacement of the concept of independence with that of martingale difference in the definition of market efficiency had some important consequences for the future development of the empirical modelling of stock returns. Independence was no longer a theoretical pre-requisite for market efficiency; an efficient market may well be characterized by non-independent returns. As a result, econometricians turned their attention to examining this possibility, namely the extent to which non-linear dependence is present in stock returns series. Specifically, a martingale difference sequence may well be dependent, with dependence arising via higher moments. Mandelbrot as far back as 1966 had anticipated these developments by saying: "It should also be stressed that the distribution of $Z(t+T)$ (future price), conditioned by known values of $Z(t)$ (current price) and of the $Z(t_i^0)$ (past prices), may very well depend upon the past values $Z(t_i^0)$: the expectation alone is unaffected by the $Z(t_i^0)$." (1966, p. 244). As will become clear in the next section, Mandelbrot's intuition took the form of "conditional heteroscedastic" models whose various formulations have dominated the financial econometrics literature since the early 1980s.

However, the abandonment of independence for the shake of maringale difference presented some issues with respect to the distribution of stock returns. The three major sets of results on the limiting distribution of stock returns presented above, namely Gaussian, Stable Paretian and Mixture-of-Normals, relied on the assumption of independence of the elementary price increments process $\{\xi_{tj}\}_{j\geq 1}$. The question which naturally arises at this point is whether these results are still valid when $\{\xi_{tj}\}_{j\geq 1}$ is not independent but merely martingale difference. Despite its obvious significance, this question was not dealt with by any of the major advocates of the aforementioned three schools of thought. This is not surprising since results on limit theorems for dependent sequences started to appear in the probability literature only around the mid 1950s. Marsaglia (1954) proved a CLT for m-dependent sequences with bounded variances. Billingsley (1961) proved a CLT, tailor-made for the new "martingale" view of market efficiency, for stationary and ergodic martingale difference sequences (see also Brown 1971). Ibragimov (1962) launched a new era for CLTs, those concerning asymptotically independent stochastic sequences usually referred to as "mixing sequences" (see, for example, Ibragimov 1975, Hall and Hayde 1980, Herrndorf 1984). These limit theorems allow $\{\xi_{tj}\}_{j\geq 1}$ to exhibit various forms of (weak) dependence as long as it is asymptoticly independent. The amount of dependence in $\{\xi_{tj}\}_{j\geq 1}$ that these limit theorems allow, depends on the moment conditions imposed on the ξ_{tj}'s. The higher the order of the required moment conditions is the "more dependent"' the sequence $\{\xi_{tj}\}_{j\geq 1}$ is allowed to be. However, the extention of these CLT results to the case of infinite variance sequences belonging to the domain of attraction of a stable non-Gaussian law (the Mandelbrot case) proved to be a substantially more difficult problem (see Bartkiewicz et. al. 2011).

Overall, both the theoretical considerations concerning market efficiency and the probability theory results concerning convergence-in-law of dependent sequences, paved the way for the current probabilistic interpretations of stock returns data. These interpretations view stock returns as realizations of non-linearly dependent stochastic sequences, with this temporal dependence arising through higher conditional moments.

7 The New Era: Conditional Heteroscedasticity and Non-Linear Dependence

The 1980s witnessed the birth of a new class of statistical models for asset returns which were motivated by the realisation that asset return series exhibit non-linear temporal dependence. More specifically, the major breakthrough of this decade was a probabilistic re-interpretation of the volatility clustering effect, first observed by Mandelbrot as far back as 1963. In this decade, the "curious behaviour of volatility" observed by previous authors was re-interpreted in a fundamentally different way. Instead of seeing it as a manifestation of "infinite-variance" or "non uniformity of transactions over time", the new view interpreted the "volatility clustering" effect as temporal non-linear dependence arising from the conditional variance. In other words, the observed data could have been produced by a strictly (or even second-order) stationary process which exhibits conditional heteroscedasticity. The ARCH model of Engle (1982) and its extentions (see the GARCH(p,q) model of Bollerslev 1986, etc.) offered a convenient way for describing such processes. More specifically, the well-known MD-GARCH(1,1) model takes the following form:

$$\begin{aligned} R_t &= c + \varepsilon_t \qquad (16) \\ \varepsilon_t &= h_t \nu_t \\ h_t^2 &= a_0 + a_1 h_{t-1}^2 + a_2 \varepsilon_{t-1}^2, \;\; a_0 > 0, \;\; a_1 \geq 0, \;\; a_2 \geq 0 \\ \nu_t &\sim IID(0, \sigma_\nu^2) \end{aligned}$$

The process $\{R_t - c\}$ where $\{R_t\}$ is described by **(16)** is martingale difference**.** Early attempts to investigate the probabilistic properties of the process defined by (16) focused mainly on (i) showing that a GARCH process is leptokurtic and (ii) establishing the conditions under which this process is covariance stationary. The first results showed that $\{R_t\}$ is second-order stationary if $a_1 + a_2 < 1$, in which case the unconditional variance of R_t exists and is equal to $a_0/(a_1 + a_2)$. However, the estimates of these parameters were found to be in the viscinity of the unit root area, that is they point towards that $a_1 + a_2 \simeq 1$. These estimates gave rise to the so-called Integrated GARCH process (IGARCH), that is a process described by (16) with $a_1 + a_2 = 1$. This process is clearly not covariance stationary since the unconditional variance is infinite although it is still strictly stationary and ergodic (see Nelson 1990). The near to unit root estimates of the conditional variance revived the debate on the "infinite-variance" issue. In fact, the "infinite-variance" problem, which came out of the front door with Clark's explanation re-emerged in the context of the IGARCH model from the rear window. Was Mandelbrot right? Should the presence of a unit root in the conditional variance be interpreted as supporting evidence - obtained from a brand new statistical method - for the Mandelbrotian infinite variance hypothesis. The answer to this question is a definitive "no". Using an intermediate result of Nelson (1990), Kourogenis and Pittis (2008) showed that the unconditional variance of an IGARCH process is "barely infinite", meaning that all the moments with order less than two exist! In the context of (16) with $a_1 + a_2 = 1$ the barely infinite variance hypothesis is stated as $E|R_t|^{\delta} < \infty$ for every $0 \leq \delta < 2$. The difference between the "barely infinite variance" IGARCH process defined by (16) with $a_1 + a_2 = 1$ and the independent Stable Paretian process proposed by Mandelbrot is huge as far as their asymptotic properties are concerned. More specifically, as will be discussed below, in spite of having (barely) infinite variance, an IGARCH process is in the domain of the attraction of the normal law.

It is worth noticing that the "objective" regularities exhibited by stock returns do not seem to have changed in any fundamental way from the beginning of the twentieth century until today. In other words, the fundamental empirical regularities namely, leptokurtosis of empirical distributions (histograms), very low or zero sample autocorrelation coefficients, and volatility clustering seem to characterise high-frequency asset returns data for any (sufficiently long) sub-sample of this period. What changed drastically, was the probabilistic interpretation of these regularities. From the Paretian IID interpetation of Mandelbrot or the Mixed-Normal IID interpretation of Clark the literature took a sharp turn in adopting the Martingale-Difference GARCH (MD-GARCH) interpretation of Engle and Bollerslev. The implications of this change for the ability of MD-GARCH to explain (in a formal sense) the observed regularities are analyzed below.

7.1 The Explanatory Status of the MD-GARCH Model

As analyzed in the previous section, the main explanatory virtue of both Mandelbrot and Clark's models stems from the fact that the empirical regularities of observed asset returns, R_t, were deduced from fundamental laws (assumptions) governing the behaviour of the elementary returns ξ_{tj}. In other words, the covering law dictating the behaviour of R_t did not emerge inductively, that is they were not inferred from the observed properties of R_t, but rather deductively from "first principles" concerning the properties of the constituent parts of the chance mechanism at work. This is the reason why these models were found to (partly) satisfy the conditions of the Deductive-Nomological-Probabilistic model of explanation. On the contrary, MD-GARCH emerged from the probabilistic interpretation of the regularities exhibited by the R_t's themselves, without any attempt to account for the chance mechanism at work. In other words, the birth of MD-GARCH conforms to the "narrowly inductivist view" according to which hypotheses should be inductively inferred from the available evidence (see, Hempel 1965 for a critique of this view). More specifically, we may distinguish two questions in the context of the MD-GARCH model, whose answers will determine the explanatory value of this model. These are the following:

(i) What are the probabilistic assumptions, G1P, for the random variables ξ_{tj}, representing elementary returns, on the basis of which it can be deduced that the daily stock returns process, $\{R_t\}_{t\geq1}$, where $R_t = \sum_{j=1}^{n_t} \xi_{tj}$ and n_t is the number of transactions for day t, is MD-GARCH?

(ii) What are the empirical regularities, G1, exhibited by the elementary returns, whose probabilistic interpretation may take the form of G1P?

A tentative (although circular) answer to (i) is to assume that the process $\{\xi_{tj}\}_{j\geq1}$ is itself GARCH. For example G1P: $\{\xi_{tj}\}_{j\geq1} \sim GARCH(1,1)$. Concerning (ii), the obvious empirical counterpart, G1 of G1P is: G1: "elementary returns exhibit volatility clustering". This type of argument reminds that of Mandelbrot who demonstrated that R_t is leptokurtic (Stable Paretian with infinite variance) by assuming that the ξ_{tj}'s are Stable Paretian with infinite-variance (leptokurtic). In the case of Mandelbrot, the argument worked mainly due to the "stability-under-addition" property analyzed above. Does a similar property hold for the case of GARCH processes? In other words, are GARCH processes stable under addition? The answer to this question is, in general, negative under the definition of a strong GARCH process. Nevertheless, under the less demanding definition of a "weak GARCH" process, Drost and Nijman (1993) proved closedness under addition. Note, however, that the results of Drost and Nijman do not cover the IGARCH case.

The preceding discussion leads to the following conclusions:

(i) The MD-GARCH model may be thought of as a representation of the empirical regularity of volatility clustring observed in asset returns. This regularity entails a narrower regularity, namely leptokurtosis of asset returns. As a result, a Deductive-Statistical explanation of leptokurtosis is obtained in the context of MD-GARCH.

(ii) The origins of MD-GARCH are blurred. The fathers of this model did not deduce it from some more fundamental laws concerning the behaviour of the elementary returns, ξ_{tj}. To this end, MD-GARCH lacks an account for the chance mechanism at work and therefore it cannot provide a D-N-P explanation of leptokurtosis.

Can MD-GARCH provide D-S explanations for other empirical regularities of R_t such as Aggregational Gaussianity and Aggregational Independence (AI), with the latter being the tendency of returns to become time-independent as we move from higher to lower frequencies of observation? To this end, we may distinguish the following two cases:

(i) Assume first that R_t represents the daily returns of a specific asset and $\{R_t\}_{t\geq1}$ is indeed a second-order stationary process with finite variance. In such a case the monthly returns, R_τ, being the sum of the daily returns of the corresponding month are likely to be approximately normally distributed provided that $\{R_t\}_{t\geq1}$ satisfies some memory conditions describing asymptotic independence, such as mixing conditions. To this end, Carrasco and Chen (2002) and Francq and Zakoian (2006) showed that a second-order stationary GARCH process is $\beta-$mixing with exponential mixing rate. Since $\beta-$mixing implies $\alpha-$mixing, we may appeal to the CLT for strong mixing processes, first proved by Ibragimov (1962), to conclude that the R_t's are in the domain of attraction of the normal law; hence R_τ is approximately normally distributed.

(ii) The case of IGARCH is substantially different. This is due to the fact that when the unconditional variance of $\{R_t\}_{t\geq 1}$ is infinite, this sequence may not belong to the domain of attraction of the normal law. As analyzed above, the infinite-variance problem has been analyzed by Mandelbrot (1963) in the context of independent sequences. However, Mandelbrot does not assume that the infinite variance arises from intensive second-order temporal dependence, such as IGARCH. Instead, he assumes that the elementary returns process belongs to the domain of attraction of a stable law with index a, $0 < a < 2$, thus excluding the normal distribution from the class of potential limit distributions. Under IGARCH, however, we face a completely different situation: If the sequence $\{R_t\}_{t\geq 1}$ has finite second moments of order δ, $0 \leq \delta < 2$, then the normal distribution may well be the limit of the (properly standardized) partial sums of $\{R_t\}_{t\geq 1}$. This assertion stems from a limit theorem due to Bradley (1988) which states that under some weak conditions, $\rho-$mixing sequences with barely infinite variance belong to the (non-normal) domain of attraction of the normal distribution. Peligrad (1990) obtains a similar result for $\phi-$mixing processes. These results suggest that an IGARCH process may obey the CLT. A final answer to this question was given very recently by Zhang and Lin (2012). Specifically, Zhang and Lin proved that for a general class of GARCH models, that covers the case of IGARCH, the central limit theorem holds.

Let us now turn our attention to Asymptotic Independence. The preceding discussion suggests that starting with the assumption that $\{R_t\}_{t\geq 1}$ is either a GARCH(1,1) or even an IGARCH(1,1) process we are led to the conclusion that R_τ is approximately normal. This property (AG) also has implications for the dependence properties of the sequence $\{R_\tau\}_{\tau\geq 1}$. More specifically, if leptokurtosis is a manifestation of GARCH effects, then the decrease of leptokurtosis, implied by AG, means that the GARCH effects are diminishing as we move from higher to lower frequencies. Hence, Aggregational Independence emerges. This heuristic argument was formally proved by Diebold (1988), who showed that GARCH effects dissapear as the sampling interval tends to infinity.

To sum-up: MD-GARCH can offer D-S explanations for the empirical regularities of leptokurtosis, Aggregational Gaussianity and Aggregational Independence of asset returns (see also Koundouri et. al. 2012 for an extentive discussion). However, due to the absence of any theoretical derivation of MD-GARCH from assumptions concerning the elementary returns ξ_{tj}, that is, due to the failure of this model to give any insight into the structure and workings of the chance mechanism at work, MD-GARCH fails to provide explanations of the D-N-P type for the above mentioned regularities. Indeed, in order for MD-GARCH to satisfy the conditions of D-N-P model of explanation, it should have been derived from some kind of theory about the properties of elementary returns, "without appeal to particular facts". Only then, an "understanding of the process" by which returns are generated would have been achieved.

8 Conclusions

One of the most interesting problems in the philosophy of science is that of finding criteria that define adequate statistical explanations (either of single events or empirical regularities). In this paper we critically reviewed and analyzed the extensive literature on asset returns, since the beginning of the twentieth century, with the aim of distinguishing between explanatory and descriptive asset returns models. In particular, we identified those models that satisfy the criteria of explanatory adequacy, set forth by alternative theories of scientific explanation. We focused primarily on explanations of empirical regularities, rather than those of single events, since this type of explanation seems to motivate all the statistical models of asset returns that aspire to be explanatory.

The statistical modelling of asset returns between the late 1950s and the early 1980s was revolving around three interconnected axes. First, an empirical regularity, $\mathcal{R}$, was detected. Second, $\mathcal{R}$ was given a probabilistic intepretation in terms of a set, $\mathcal{P}$, of properties of a sequence $\{R_t\}_{t\geq 1}$ of random variables, Note that during the aforementioned period, the same empirical regularity was given alternative probabilistic interpretations. Third, a statistical model, $\mathcal{M}$, that accounts for $\mathcal{P}$ was suggested. Whether or not $\mathcal{M}$ *explains* $\mathcal{R}$ depends on the way by which $\mathcal{M}$ is produced. Specifically, if $\mathcal{M}$ is derived from a theoretical account of the chance mechanism at work, then $\mathcal{M}$ satisfies the conditions for explanatory adequacy imposed by the Deductive-Nomological-Probabilistic model of explanation. In such a case, $\mathcal{M}$ is deemed to be explanatory. On the other hand, if $\mathcal{M}$ is inductively inferred from the available data, without having any theoretical underpinnings, $\mathcal{M}$ is deemed to be descriptive. Nevertheless, $\mathcal{M}$ can still play the role of a covering law in a D-S explanation of an empirical regularity $\mathcal{R}'$, provided that $\mathcal{R}'$ is narrower than $\mathcal{R}$.

Our critical examination of the origins of the statistical models of stock returns during the aforementioned period, showed that these models, with their leading examples being Osborne's NIID (1959), Mandelbrot's Stable-Paretian (1963) and Clark's mixture-of-normals (1973), enjoy a sufficiently high D-N-P explanatory status. In addition, the explanatory value of these models is further enhanced by their ability to answer counterfactual questions, that is questions about the conditions under which the empirical regularities (explananda) generated by these models would have been different.

A common characteristic of all the aforementioned authors - on which the explanatory feature of their models was based - was their insistence on deriving their models for observable returns from alternative sets of primitive assumptions concerning the behaviour of elementary returns ξ_{tj}. In other words, each of these authors derived his corresponding model from his own theoretical account of the returns generating mechanism. The main differences among the aforementioned theoretical accounts center around the properties of elementary transactions ξ_{tj}. Osborne assumed that ξ_{tj} cannot be arbitrarily large and that the number of transactions across time is constant. Mandelbrot retained the constancy of transactions over time but he took the radical view that elementary transactions can be arbitrarily large, a direct implication of which is that the variance of stock returns is infinite. Finally, Clark in his attempt to salvage the finite-variance hypothesis, he elevated the hypothesis of "the substantial temporal variation of the number of transactions" as the most fundamental one concerning the generation of stock returns. Osborne's explanation - being in the spirit of the original explanation of Bachelier - aimed at explaining the empirical regularities of asset returns identified at the time, namely "independence" and "normality". On the other hand, both Mandelbrot and Clark focused on explaining the "new" empirical regularity identified by the beginning of the 1960s, namely leptokurtosis of the asset return distributions. The deductive power in all the aforementioned explanations stemmed from limit theorems. Osborne employed the classic central limit theorem, Mandelbrot used the limit theorems for sequences of random variables with infinite variance and Clark utilized limit theorems for random sums of random variables.

The realism of the aforementioned assumptions was the subject of heated debate between the few economists who supported Mandelbrot's explanation and those (the majority) who supported Clark's. The reason is that Mandelbrot's interpretation had unpleasant implications for almost all the major theoretical concepts and models of the finance theory that existed at that time. On the other hand, Clark's interpretation identified the origins of leptokurtosis with the inability of the experimenter to conduct controlled experiments, that is to keep the number of transactions approximately constant over time.

The paper maps interesting analogies, developing over time, between the central issues in the statistical modelling of asset prices and those concerning the theoretical modelling of the Brownian motion in Physics. Our arguments focus on the fundamental works of Bachelier (1900) and Einstein (1905) and identify, for both disciplines, how Brownian motion was introduced from the corresponding sets of initial assumptions. We have shown that despite the obvious differences of the subject matters, the development of the respective literatures share some striking similarities concerning the underlying causal mechanisms that generate their observed behaviors. Because these similarities were subjected to common probabilistic interpretations, they gave rise to models with the same theoretical structure.

The arrival of the 1980s witnessed new developments in the statistical modelling of asset returns. The recognition that the Efficient Market Hypothesis requires returns to be just a martingale difference process, led to removal of the need for independence in returns. In the early 1980's, the assumption of martingale difference combined with the empirical evidence of non-constant volatility, gave rise to the conditionally heteroskedastic models (GARCH) by Engle and Bollerslev. These models were motivated mostly by an attempt to describe the stylized facts of assets returns, rather than an attempt to explain their generating mechanism. The rise of GARCH models marks the prevalence of the statistical-inductive approach over the explanatory-deductive one. Nevertheless, the GARCH literature has made important progress towards studying theoretical properties of these models, such as closeness under temporal aggregation and infinite divisibility. The introduction of the so-called "weak" GARCH models and continuous time processes GARCH models, are examples of such progress.

In this paper we have presented and explained the motivating forces and dynamic evolution of the scientific formalization of asset returns processes, since the beginning of the 20th century. We have provided an in-depth analysis of the similarities and differences, as well as the strengths and weaknesses, between the various modelling approaches of asset returns, both explanatory and descriptive ones. Of course, the evolution of the modelling of asset returns does not end; the hunt for a model which comes closer to Railton's ideal explanatory text is indeed unended! Our detailed analysis of the history of this evolution seems to uncover at least one source of inspiration for the newcoming models: the need for the joint exploitation of both substantive and statistical information in the specification of these models. As Wold had insightfully remarked forty years ago, "the construction process (of models) alternates several times between the empirical and theoretical sides, building up the model by layer after layer of empirical and theoretical knowledge." (1969 pp. 437).

References

[1]L. S. Ornstein, “On the Brownian motion,” Proc. Royal Acad. Amsterdam 21, 96–108 (1917)

[2]Pólya, G., 1920, Ueber den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem, Math. Zeit. vol. 8 (1920) pp. 171-181. (Change the year in the text)

[3]Fürth, R., (1920) Brownian motion with directional persistence. With applications to the movement of living infusoria. Zeitschrift für Physik, 2:244–256 (in German).

[4]Wiener, L. C., 1863, “Erklärung des atomistischen Wesens des tropfbarflüssigen Körperzustandes und Bestätigung desselben durch die sogenannten Molekularbewegungen” Annalen der Physik und Chemie, 118 (1863), 79–94.

[5]Alexander, S.S. (1961) Price Movements in Speculative Markets: Trends or Random Walks. Industrial Management Review, Vol. 2 (1961), pp. 7-26.

[6]Antypas, Antonios , Phoebe Koundouri and Nikolaos Kourogenis, 2011, Aggregational Gaussianity and Barely Infinite Variance in Financial Returns, University of Piraeus Working Paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2120104 .

[7]Bachelier, L., 1900, Theorie de la speculation - Gauthier-Villars (Theory of speculation in: P. Cootner, ed., 1964, The random character of stock market prices L Bachelier - 1900 - MIT Press, Cambridge, Mass).

[8]Bartkiewicz, Katarzyna, Adam Jakubowski, Thomas Mikosch and Olivier Wintenberger, Stable limits for sums of dependent infinite variance random variables, Probability Theory and Related Fields 150, 3, pp. 337-372

[9]Bertrand, J. - 1889 Calcul des probabilités - Gauthier-Villars Paris

[10]Billingsley, Patrick (1961), The Lindeberg-Levy Theorem for Martingales, Proceedings of the American Mathematical Society , Vol. 12, No. 5 (Oct., 1961), pp. 788-792 .

[11]Blattberg, R.C., and N. Gonedes. 1974. "A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices." Journal of Business 47:244-280.

[12]Blum, J.R., D.L. Hanson, and J.I. Rosenblatt. 1963. "On the central limit theorem for the sum of a random number of independent random variables." Probability Theory and Related Fields 1:389-393.

[13]Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31, 3, pp. 307–327.

[14]Box, G.E.P., Draper, N.R., 1987, Empirical Model-building and Response Surfaces. Wiley, New York.

[15]Box, G E P, Hunter, W G, The experimental study of physical mechanisms. Technometrics 7, 1965

[16]Bradley, R. C. (1988). A central limit theorem for stationary ρ-mixing sequences with infinite variance. The Annals of Probability 16:313–332.

[17]Brown, B. M. (1971), Martingale Central Limit Theorems, The Annals of Mathematical Statistics , Vol. 42, No. 1 (Feb., 1971), pp. 59-66

[18]Brown, R., 1827, A Brief account of microscopical observations made in the months of June, July and August 1827 on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies.

[19]Brush, Stephen G. 1968, A History of Random Processes: I. Brownian Movement from Brown to Perrin, Archive for History of Exact Sciences , Vol. 5, No. 1 (6.8.1968), pp. 1-36

[20]Campbell, John Y., Andrew Wen-Chuan Lo, Archie Craig MacKinlay, 1997, The econometrics of financial markets, Princeton University Press, Princeton, New Jersey.

[21]Carrasco, M., Chen, X., 2002. Mixing and moment properties of various GARCH and stochastic volatility models. Econometric Theory 18, 17-39.

[22]Chapman, S. (1928), On the Brownian Displacements and Thermal Diffusion of Grains Suspended in a Non-Uniform Fluid, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character , Vol. 119, No. 781 (May 1, 1928), pp. 34-54

[23]Clark, Peter K. 1973, A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices, Econometrica , Vol. 41, No. 1 (Jan., 1973), pp. 135-155.

[24]Cootner, P. H., (1962), "Stock Prices: Random vs. Systematic Changes", Industrial Management. Review, vol. 3, spring, pp.24-45

[25]Cootner, Paul H. (1964). The random character of stock market prices. MIT Press. ISBN 9780262030090

[26]Diebold, F. X. (1988): Empirical Modeling of Exchange Rates. New York: Springer Verlag.

[27]Drost, F., Nijman, T., 1993. Temporal aggregation of GARCH processes. Econometrica 61, 909-27.

[28]Einstein, A., 1905, Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, Annalen der Physik, Vol. 17 , pages 549–560.

[29]Engle, Robert F. (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, Econometrica , Vol. 50, No. 4 (Jul., 1982), pp. 987-1007.

[30]Fama, Eugene F. 1965b, Portfolio Analysis in a Stable Paretian Market, Management Science , Vol. 11, No. 3, Series A, Sciences (Jan., 1965), pp. 404-419

[31]Fama, Eugene F. , 1965a, Random Walks in Stock Market Prices, Financial Analysts Journal Vol. 21, No. 5 (Sep. - Oct., 1965), pp. 55-59

[32]Fama, Eugene F. 1963, Mandelbrot and the Stable Paretian Hypothesis, The Journal of Business , Vol. 36, No. 4 (Oct., 1963), pp. 420-429

[33]Fama, Eugene F. 1965. The Behavior of Stock-Market Prices, Journal of Business , Vol. 38, No. 1 (Jan., 1965), pp. 34-105

[34]Fama, Eugene F., and Marshall E. Blume, 1966, Filter Rules and Stock-Market Trading, The Journal of Business , Vol. 39, No. 1, Part 2: Supplement on Security Prices (Jan., 1966), pp. 226-241.

[35]Feller, W. 1935, Über den zentralen Genzwertsatz der Wahrscheinlichkeitsrechnung, Math. Z. 40, 521-559.

[36]Francq, C., Zakoian, J.M., 2006. Mixing properties of a general class of GARCH(1,1) models without moment assumptions on the observed process. Econometric Theory 22, 815-834.

[37]Friedman, Milton, 1953, Essays in Positive Economics, The University of Chicago Press, Chicago.

[38]Gouy, M. (1888) Note sur le mouvement brownien, Journal de Physique Théorique et Appliquée Vol. 7 No. 1 p. 561.

[39]Hall, P., Heyde, C. C. (1980). Martingale Limit Theory and Its Application. New York: Academic Press.

[40]Hempel, C., 1965, Aspects of Scientific Explanation and Other Essays in the Philosophy of Science, New York: Free Press.

[41]Hempel, D. G., Oppenheim, P., 1948, Studies in the Logic of Explanation, Philosophy of Science 15, 135-175

[42]Herrndorf, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables. The Annals of Probability 12:141–153.

[43]Houthakker, H. S. (1961). Systematic and Random Elements in Short-Term Price Movements, American Economic Review 51, 164-172

[44]Ibragimov, I. A. (1962). Some limit theorems for stationary processes. Theory Prob. Appl. 7:349–382.

[45]Ibragimov, I.A. (1975), A note on the central limit theorem for dependent random variables, Teor. Veroyatnost. i Primenen., 1975, Volume 20, Issue 1, Pages 134–140.

[46]Jeffrey, R. C. 1969, Statistical explanation vs Statistical inference, in N. Rescher (Editor): Essays in Honor of Carl G. Hempel: A Tribute on the Occasion of his Sixty-Fifth Birthday (Synthese Library), Springer.

[47]Jensen, Michael C. The Performance of Mutual Funds in the Period 1945-1964 The Journal of Finance Vol. 23, No. 2, Papers and Proceedings of the Twenty-Sixth Annual Meeting of the American Finance Association Washington, D.C. December 28-30, 1967 (May, 1968), pp. 389-416

[48]Kac, Mark, 1947, Random Walk and the Theory of Brownian Motion, The American Mathematical Monthly , Vol. 54, No. 7, Part 1 (Aug. - Sep., 1947), pp. 369-391

[49]Kendall M. G. , 1953, The Analysis of Economic Time-Series-Part I: Prices, Journal of the Royal Statistical Society. Series A (General) Vol. 116, No. 1 (1953), pp. 11-34.

[50]Khintchine, A. 1933, Asymptotische Gesetze der Wahrscheinlichkeitsrechnung, Ergebnisse der Mathematik, vol. 2, no. 4, Berlin.

[51]Kon, Stanley J., 1984, Models of Stock Returns–A Comparison, The Journal of Finance , Vol. 39, No. 1 (Mar., 1984), pp. 147-165.

[52]Koundouri, Phoebe & Nikolaos Kourogenis, 2011. "On the Distribution of Crop Yields: Does the Central Limit Theorem Apply?," American Journal of Agricultural Economics, Agricultural and Applied Economics Association, vol. 93(5), pages 1341-1357.

[53]Koundouri, P., Kourogenis, N., and N. Pittis, 2012. "Can Statistical Models of Stock Returns "Explain" Empirical Regularities?", University of Piraeus Working Paper.

[54]Kourogenis, N., Pittis, N., 2008. Testing for a unit root under errors with just barely infinite variance. Journal of Time Series Analysis 29, 1066-1087.

[55]Larson, A. B. 1960. Measurement of a Random Process in Futures Prices. Food Research Institute Studies 1: 313–324.

[56]Lehmann E. L. , 1990, Model Specification: The Views of Fisher and Neyman, and Later Developments, Statistical Science Vol. 5, No. 2 (May, 1990), pp. 160-168.

[57]Levy, Paul (1925) Calcul des Probabilités, Paris: Gauthier-Villars.

[58]Lintner, J. (1965a). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. Review of Economics and Statistics 47, 13-37.

[59]Lintner, J. (1965b). Security Prices, Risk, and Maximal Gains From Diversification. Journal of Finance 20, 587-615.

[60]Lyapounof, A.M. (1901), "Nouvelle forme du théorème sur la limite de probabilité", Mem. Acad. Sci. St. Petersburg 8, 1-24

[61]Maiocchi, Roberto, 1990, The case of Brownian motion, The British Journal for the History of Science (1990), 23 : pp 257-283

[62]Mandelbrot, Benoit, (1963) "The Variation of Certain Speculative Prices," Journal of Business 36 (Oct. 1963), 394-419.

[63]Mandelbrot, Benoit 1966, Forecasts of Future Prices, Unbiased Markets, and "Martingale" Models, The Journal of Business , Vol. 39, No. 1, Part 2: Supplement on Security Prices (Jan., 1966), pp. 242-255

[64]Markowitz, Harry 1952, Portfolio Selection, The Journal of Finance , Vol. 7, No. 1 (Mar., 1952), pp. 77-91.

[65]Marsaglia, G. (1954). Iterated limits and the central limit theorem for dependent variables. Proc. Amer. Math. Soc. 6:987–991.

[66]Melia, J. (2000). Weaseling away the indispensability argument, Mind 109 (435): 455-480.

[67]Moore 1962 Regression As An Analytical Tool P. G. Moore Journal of the Royal Statistical Society. Series C (Applied Statistics)Vol. 11, No. 2 (Jun., 1962), pp. 106-119.

[68]Mossin, J. (1966). Equilibrium in a Capital Asset Market. Econometrica 34, 768-783.

[69]Muth, John F. , 1961, Rational Expectations and the Theory of Price Movements, Econometrica Vol. 29, No. 3 (Jul., 1961), pp. 315-335

[70]Nagel, Ernest, 1929, Nature and Convention, The Journal of Philosophy , Vol. 26, No. 7 (Mar. 28, 1929), pp. 169-182

[71]Nelson, Daniel B. (1990), Stationarity and Persistence in the GARCH(1,1) Model, Econometric Theory , Vol. 6, No. 3 (Sep., 1990), pp. 318-334

[72]Officer, R. R. 1972, The Distribution of Stock Returns, Journal of the American Statistical Association , Vol. 67, No. 340 (Dec., 1972), pp. 807-812

[73]Osborne, M. F. M. , 1959, Brownian Motion in the Stock Market, Operations Research Vol. 7, No. 2 (Mar. - Apr., 1959), pp. 145-173

[74]Peligrad, M. (1990). On Ibragimov–Iosifescu conjecture for φ-mixing sequences. Stochastic Processes and Their Applications 35:293–308.

[75]Perrin, J. (1908), L'agitation moléculaire et le mouvement brownien, Comptes Rendus (Paris) 146, p. 967.

[76]Perrin, J., 1909, Mouvement brownien et réalité moléculaire, Annales de Chimie et de Physique, Vol. 18 (1909), pp. 5-104

[77]Perrin, Jean, 1916, Atoms, D. Van Nostrand company.

[78]Popper, Karl , 1972, Objective Knowledge publ. Clarendon Press, Oxford, UK

[79]Praetz, Peter D. 1972, The Distribution of Share Price Changes, The Journal of Business , Vol. 45, No. 1 (Jan., 1972), pp. 49-55

[80]Press, S. J. (1968). A compound events model for security prices. Journal of Business, 40, 317-335.

[81]Psillos, Stathis 2002, Causation and Explanation, Acumen & McGill-Queens U.P.

[82]Railton, Peter 1978, A Deductive-Nomological Model of Probabilistic Explanation, Philosophy of Science , Vol. 45, No. 2 (Jun., 1978), pp. 206-226

[83]Railton, Peter 1981, Probability, Explanation, and information, Synthese , Vol. 48, No. 2, Probabilistic Explanation, Part I (Aug., 1981), pp. 233-256

[84]Renyi, A. 1960. "On the central limit theorem for the sum of a random number of independent random variables." Acta Mathematica Hungarica 11:97-102.

[85]Robbins, H. 1948. "On the Asymptotic Distribution of the Sum of a Random Number of Random Variables." Proceedings of the Natiinal Academy of Sciences 34:162-163.

[86]Roberts, Harry V. , 1959, Stock-Market "Patterns" and Financial Analysis: Methodological Suggestions, Journal of Finance Vol. 14, No. 1, pp. 1-10.

[87]Ross, Stephen A. 1977, The arbitrage theory of capital asset pricing, Journal of Economic Theory 13, 341-360.

[88]Salmon, W.C. 1971, Statistical explanation and Statistical Relevance, University of Pittsburgh Press.

[89]Salmon, W.C. 1998, Causality and Explanation, Oxford University Press (New York, Oxford).

[90]Samuelson Paul A. , 1973, Mathematics of Speculative Price, SIAM Review Vol. 15, No. 1 (Jan., 1973), pp. 1-42

[91]Samuelson, Paul A., (1965) Proof that Properly Anticipated Prices Fluctuate Randomly, Industrial Management Review, Vol. 6, pp. 41-49.

[92]Samuelson, Paul A. 1973, Mathematics of Speculative Price, SIAM Review , Vol. 15, No. 1 (Jan., 1973), pp. 1-42

[93]Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. Journal of Finance 19, 425-442.

[94]Spanos, Aris, 2006, Where Do Statistical Models Come from? Revisiting the Problem of Specification, Lecture Notes-Monograph Series , Vol. 49, Optimality: The Second Erich L. Lehmann Symposium (2006), pp. 98-119

[95]Stanford, P. Kyle 2006, Darwin's Pangenesis and the Problem of Unconceived Alternatives, British Jnl. for the Philosophy of Sci. Volume 57, Issue 1, pp. 121-144

[96]Stegmüller, W., 1973, Carnap's Normative Theory of Inductive Probability, Studies in Logic and the Foundations of Mathematics, Volume 74, 1973, Pages 501-513.

[97]Steiger, W (1964). A test of nonrandomness in stock price changes, in Cootner, Paul H. ed. The random character of stock market prices.

[98]Taussig, F. W., 1921, Is Market Price Determinate?, The Quarterly Journal of Economics , Vol. 35, No. 3 (May, 1921), pp. 394-411

[99]Teichmoeller, John 1971, A Note on the Distribution of Stock Price Changes, Journal of the American Statistical Association , Vol. 66, No. 334 (Jun., 1971), pp. 282-284.

[100]Treynor, J. (1962). Toward a Theory of Market Value of Risky Assets. Published in ed. R. A. Korajczyk: Asset Pricing and Portfolio Performance: Models, Strategy and Performance Metrics, Risk Books, 1999.

[101]von Smoluchowski, M. 1906, Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen, Annalen der Physik, SER4,T21, pp. 756-780.

[102]Woodward, James 2003, Making Things Happen: A Theory of Causal Explanation, Oxford Studies in the Philosophy of Science, Oxford University Press.

[103]Working, Holbrook, 1934, A Random-Difference Series for Use in the Analysis of Time Series, Journal of the American Statistical Association 29, 11-24.

[104]Working, Holbrook, 1958, A Theory of Anticipatory Prices, The American Economic Review , Vol. 48, No. 2, Papers and Proceedings of the Seventieth Annual Meeting of the American Economic Association (May, 1958), pp. 188-199

[105]Zhang, R.-M., Lin, Z.-Y., 2012. Limit theory for a general class of GARCH models with just Barely infinite variance. Journal of Time Series Analysis 33:1, 161-174.