Can Statistical Models of Stock Returns "Explain" Empirical Regularities?

Koundouri, Phoebe and Kourogenis, Nikolaos and Pittis, Nikitas

2012

Can Statistical Models of Stock Returns "Explain" Empirical Regularities?

Phoebe Koundouri * Nikolaos Kourogenis† ‡ Nikitas Pittis§

Abstract

Statistical models are usually thought of as means for describing statistical regularities. Concerning stock returns, many empirical regularities have been documented in the literature together with their corresponding models. The main task of this paper is to investigate, under the prism of the philosophy of science, the conditions that a statistical model has to satisfy in order to be deemed as explanatory adequate for the existing regularities. We distinguish two alternative sets of criteria for the explanatory adequacy of a statistical model. The first one is given by the Deductive-Statistical model of explanation, put forward by Hempel (1962). The second set, which contains much stricter conditions than the first, corresponds to the Deductive-Probabilistic-Nomological model suggested by Railton (1978). It is shown that the two most important statistical models of stock returns, namely the multivariate GARCH model and the Factor Model with persistent betas, are D-S explanatory. It is also shown that the Factor Model partially satisfies the D-N-P conditions for explanatory adequacy whereas the GARCH model fails completely in this respect.

JEL Classification: C22, G10, G11, G12

Keywords: empirical regularities, stock returns, single factor model, autoregressive beta, statistical explanation.

1. Introduction

There exist two broad categories of empirical regularities: deterministic and statistical. Statistical regularities are described by statistical models. A single statistical model, M, usually describes more than one empirical regularity, that is a set S of empirical regularities $S_1, S_2, ..., S_n$. Following Ellis (1956), we define a regularity S_i to be of higher-order or more fundamental than another regularity S_j, if S_j can be deduced by S_i but not vice versa, that is when $S_i \implies S_j$ and $S_j \nRightarrow S_i$. In this case, S_j may be referred to as "derivative" regularity. Furthermore, two regularities are of the same order if each can be inferred from the other. It is important to note that the true relata in the aforementioned "$\implies$" relationships are the probabilistic interpretations, D_i and D_j, of S_i and S_j, respectively, rather than the regularities themselves. This in turn implies that the extent to which S_j is derived from S_i depends on the corresponding probabilistic interpretations D_i and D_j adopted by the statistician. For example, if S_j were interpreted by D_j' instead of D_j, then a proposition of the form $D_i \implies D_j'$ might not be valid. In such a case, S_j is not derived from S_i. This feature introduces some ambiguity to the extent that a given regularity is deduced from another one, since the necessary deducibility relationship may be obtained under one probabilistic interpretation but may fail under another.

The selection of the relevant theoretical interpretation at any given time depends on the "background theory" that prevails at this particular time. Historically, it has been observed that the same empirical regularity has derived alternative probabilistic/theoretical interpretations at different points in time (see, for example, Brewer and Lambert 200????). From now on, when we say that a regularity, S_j, is deduced from a broader regularity, S_i or that S_i entails or implies S_j, we shall mean that there exist (at least) two corresponding probabilistic interpretations D_i and D_j (in the sense defined above) respectively, such that $D_i \implies D_j$.

Next consider the set S= $\{S_1, S_2, ..., S_n\}$ to be the set of known regularities for the phenomenon of interest. Assume that there are two regularities, say S_1 and S_2 which are of higher-order than all the rest, namely $S_1 \Longrightarrow S_i$, $i = 3, 4, ..., n$ and $S_2 \Longrightarrow S_i$, $i = 3, 4, ..., n$. Moreover, assume that S_1 and S_2 are of the same order, that is $S_1 \Longleftrightarrow S_2$. There are two statistical models M_1 and M_2 constructed with the aim to describe S_1 and S_2, respectively. Put differently, M_1 and M_2 were born out of the probabilistic interpretations D_1 and D_2, respectively, of S_1 and S_2, respectively. This means that M_1 and M_2 are of purely empirical origins, since no "theoretical" subject-matter information is assumed to be involved in their conception. In other words, M_1 and M_2 were produced with the sole aim to "describe" (in a probabilistic sense) S_1 and S_2, respectively. However, despite their inductive/empirical origins, and descriptive aspirations, M_1 and M_2 can also be "explanatory" in a sense that will be defined below. Indeed, one of the main questions in this paper is whether the aforementioned relationships among the elements of S and the postulated origins of M_1 and M_2, imply that M_1 and M_2 are equivalent with respect to their *explanatory* power.

Before we attempt to answer the question raised above, we must first define the sense in which a statistical model can be explanatory. 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 scientifically or formally (rather than subjectively) adequate?

The question of what constitutes an adequate statistical explanation of an empirical regularity has been central in the philosophy of science literature since 1948, year at which Hempel and Oppenheim published their seminal paper on the structure of explanation (see also Hempel 1962, 1965). The main point in the Hempelian view of explanation, is that explanation is achieved through *derivation*. More specifically an explanation of an empirical regularity is an argument to the effect that this regularity (expressed in the form of an exact probabilistic statement) is derived from other (more fundamental) empirical regularities (also expressed in the form of probabilistic statements) by means of probability theory. This model of explanation is usually referred to as the Deductive-Statistical (D-S) model of explanation. D-S explanations are special cases of the so-called "covering law" explanations, in the context of which a narrow empirical regularity is explained by being subsumed under a broader empirical regularity, with the latter being referred to as the covering law of the purported explanation. In the context of D-S, the statistical models M_1 and M_2 introduced above are both D-S explanatory for $\{S_3, S_4, ..., S_n\}$. Moreover, they are D-S equivalent. Indeed, as mentioned above $M_1(M_2)$ was constructed with the aim to represent in a specific parametric form the set of probabilistic properties $D_1(D_2)$ which was chosen to probabilistically interpret the empirical regularity $S_1(S_2)$. Since it has been assumed that $S_1 \Longrightarrow \{S_3, S_4, ..., S_n\}$ and $S_2 \Longrightarrow \{S_3, S_4, ..., S_n\}$ in the sense that $D_1 \Longrightarrow \{D_3, D_4, ..., D_n\}$ and $D_2 \Longrightarrow \{D_3, D_4, ..., D_n\}$, respectively we conclude that $M_1 \Longrightarrow \{D_3, D_4, ..., D_n\}$ and $M_2 \Longrightarrow \{D_3, D_4, ..., D_n\}$. This in turn implies that all the known empirical regularities $S_3, S_4, ..., S_n$ can be derived and hence explained by both models. Moreover, in terms of the higher order regularities S_1 and S_2 these models are equivalent since M_1 can be used as a "covering law" for deducing S_2 and, symmetrically, M_2 can play the role of covering law in a deductive argument terminating in S_1.

However, M_1 and M_2 may not be equivalent under a different set of stricter criteria for explanatory adequacy. These criteria are based on the thesis that all explanation must be causal, which in turn has produced the so-called "causal mechanistic" (C-M) models of explanation. A prominent member of the CM class is the Deductive-Nomological-Probabilistic (D-N-P) model of explanation suggested by Railton (1978, 1981). According to D-N-P, the mere subsumption of a narrow regularity, say S_3, under the broader regularity, say S_1, or equivalently under the statistical model M_1, does not constitute an explanation of S_3, unless M_1 is backed up with "a theoretical account of the mechanism(s) at work." More specifically, S_3 is explained by placing it within a web of "inter-connected series of law-based accounts of all the nodes and links in the causal network culminating in the explanandum, complete with a fully detailed description of the causal mechanisms involved in the theoretical derivations of all covering laws involved" (Railton 1981, pp. 174). This means that M_1 in itself cannot form the sole basis for a satisfactory explanation of S_3 unless M_1 is derivable from a theory accounting for the causal mechanism at work. Assume now that M_1 is indeed derivable from such a theory whereas M_2 is not. In such a case, the explanatory value of M_1 is higher than that of M_2 because M_1 as opposed to M_2 is derivable from a theoretical account of the chance mechanism at work. This enables M_1, but not M_2, to function as a link in the causal network culminating in the explanandum.

Given the definitions/criteria of statistical explanation introduced above, the types of explanation that we have obtained for S_3, in the context of the aforementioned example, are the following: (i) An M_1-dependent, D-N-P explanation for S_3. (ii) An M_1-dependent, D-S explanation for S_3. (iii) An M_2-dependent, D-S explanation for S_3. Note that due to the lack of deducibility of M_2 from a theoretical account of the chance mechanism at work, no M_2-dependent, D-N-P explanation for S_3 exists.

As far as stock returns are concerned there are currently certain statistical regularities which are widely recognized as "stylized facts", since they appear to be common across many different markets, assets and time periods. These regularities may be classified in two broad categories: The first category, hereafter "Regularities of Type-I" (RT-I), refers to the individual temporal behaviour of each returns series $R_{i,t}$ $i = 1, 2, ..., n$. The most important regularities in RT-I are the following: (i) Fast Mean Reversion (FMR), that is the tendency of stock returns to revert to an average value quite rapidly. Put it differently, the degree of persistence of returns' deviation from this average value is very low, if not zero. (ii) Volatility Clustering (VC), that is the fact that "large (price) changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes" (Mandelbrot 1963, *pp. ???*). (iii) Empirical Leptokurtosis (EL), namely the empirical distributions of stock returns are characterised by heavy tails with positive excess kurtosis. (iv) Empirical Aggregational Gaussianity (EAG), that is the fact that the degree of leptokurtosis in the empirical distributions tends to diminish as the return horizon increases. (v) Empirical Aggregational Independence (EAI), that is the observation that the volatility clustering effects tend to disappear as the returns horizon increases or equivalently, as we move from higher to lower frequencies (e.g. from daily to quarterly observations).

The probabilistic interpretations/descriptions of these empirical regularities have taken the following forms: FMR is described by assuming that the stochastic sequence $\{R_{i,t}\}$ is martingale difference (MD) with respect to its own past history. VC is usually interpreted as "dynamic conditional heteroscedasticity" (DCH) which is a specific type of non-linear temporal dependence of $\{R_{i,t}\}$, arising through the conditional variance. EL has a natural interpretation in terms of theoretical leptokurtosis (TL) of the (stationary) distributions of the random variables $R_{i,t}$. EAG is interpreted as a tendency of the aggregate random variables, $R_{i\tau}(k) = \sum_{l=1}^{k} R_{i,t-k+l}$ to converge in law to the Normal distribution as the returns' horizon k increases (AG). Finally, EAI is described by the probabilistic property that the random variables $R_{i\tau}(k)$ and $R_{is}(k)$, $\tau \neq s$ tend to be independent as the returns horizon k increases (AI).

Among the above mentioned regularities and their theoretical intepretations, VC/DCH proved to be of higher-order than the rest. This is due to the fact that VC/DCH motivated the birth of the Martingale-Difference GARCH model (MD-GARCH) in the context of which EL/TL, EAG/AG and EAI/AI are derived, whereas FMR/MD is imposed (see, for example, Engle 1982, Bollerslev 1987 for $DCH \Longrightarrow TL$ and, for example, Diebold 1988, Drost and Nijman 1993 and Meddahi and Renault 2004 for $DCH \Longrightarrow AG$ and $DCH \Longrightarrow AI$).

The second category, hereafter referred to as "Regularities of Type-II" (RT-II) refers to the joint temporal behaviour of all returns series or to the joint behaviour between each returns series $R_{i,t}$ and another factor (or factors). The most important regularity in RT-II is the fact that the stock prices tend to move together over time. In other words, stock returns appear to be positively correlated temporally. Roll and Ross (1980) refer to the common variability of stock returns as “the single most widely-acknowledged empirical regularity” (1980, pp. 1073). Moroever, the degree of these comovements do not appear to be constant over time, but rather exhibit "clustering" patterns, similar to VC (see for example, Christodoulakis 2001). This empirical regularity will be referred to as "covariation clustering", abbreviated as CC. An obvious probabilistic interpretation of CC is to assume that the conditional covariance matrix Σ_t of the (stationary) random vector $\mathbf{R}_t = [R_{1,t}, R_{2,t}, ..., R_{n,t}]^\top$ is a function of the past history of $\mathbf{R}_t$, that is it exhibits dynamic conditional heteroskedasticity. This interpretation has given rise to the so-called multivariate MD-GARCH models (M-MD-GARCH).

Another empirical regularity in RT-II is based on the observation that stock returns tend to respond to (unanticipated) changes in one or more variables, such as the market portfolio (more precisely a proxy of it) or certain macroeconomic and/or financial variables/factors. This regularity will be referred to as "common factor" regularity (CF). Sharpe (1964) refers implicitly to this regularity as follows: "it is common practice for investment councelors to accept a lower expected return from defensive securities (those which respond little to changes in the economy) than they require from aggressive securities (which exhibit significant response)" (1964, pp. 442). In the simplest case of a single factor, X_t, the probabilistic interpretation of CF takes the form of a linear regression model, $R_{i,t} = a_i + \beta_i M_t + u_{i,t}$, where $M_t = X_t - E(X_t \mid \Phi_{t-1})$. Concerning the identity of X_t, the majority of empirical studies have identified the single factor with (a proxy of) the return on the "market portfolio", following the theoretical suggestions of the Capital Asset Pricing Model (CAPM). The error term, $u_{i,t}$, is assumed to be a zero-mean i.i.d. process with finite variance, satisfying the condition $E\left(u_{i,t} \mid M_t\right) = 0$, for $i = 1, 2, ..., n$.

The slope coefficient, β_i, is interpreted as a measure of the systematic risk of the stock i, and is usually referred to as the "beta coefficient", or simply the "beta" of the stock i. In the context of this model, the returns on all the existing assets, $i = 1, 2, ..., n$, are related only through M_t. This assumption amounts to the covariance matrix, Σ_u, of $u_{i,t}$, $i = 1, 2, ..., n$, being diagonal, that is, $Cov(u_{i,t}, u_{j,t}) = 0$ for $j \neq i$. Following the relevant literature, this particular description of CF will be referred to as the single factor model (SFM).

Initially, an implicit assumption in the CF regularity mentioned above was that the degree of response of each stock to changes in the factor, was constant over time. Therefore, the corresponding description of this degree of response in the context of SFM took the form of a time-invariant beta. However, more detailed statistical analysis showed that the estimates of beta were not constant over time. As a result the CF regularity was replaced by a more general regularity according to which stock returns tend to respond to changes in the factor with the degree of this response changing over time. Some researchers took the view that the variation in the degree of response is random (see, for example, Blume 1971, 1975, Fabozzi and Francis 1977) whereas others believed that this variation exhibits signs of temporal persistence (see, for example, Fisher and Kamin 1985, Sunder 1980, Bos and Newbold 1984, Collins, Ledolter and Rayburn 1987, Andersen et al. 2005, and Jostova and Philipov 2005). As a result, the original CF regularity was replaced, by the the more general CFP one, which states that "the degree of response of each stock to changes in the factor changes over time in a persistent fashion". CFP may be probabilistically described by a single factor model in which the stochastic sequence $\{\beta_{it}\}$ is assumed to be a first-order autoregressive process, $\beta_{i,t} = \varphi_i \beta_{i,t-1} + \varepsilon_{i,t}$. The resulting SFM model will be referred to as SFM-AR.

By noting that VC is a special case of CC, we may collect the aforementioned empirical regularities of stock returns in the following set S_R,

$$\mathcal{S}_R = \{CC, CFP, FMR, EL, EAG, EAI\}.$$

As mentioned above, CC motivated the birth of M-MD-GARCH, whereas CFP gave rise to SFM-AR. The main task of this paper is to compare the explanatory power of M-MD-GARCH with that of SFM-AR, under the D-S and D-N-P criteria of explanatory adequacy. The paper is organised as follows: Section 2 begins with a brief outline of the basic features of the D-S model of statistical explanation. Then it proceeds to critically review the existing econometric results demonstrating that M-MD-GARCH entails FMR (tautologically) as well as EL, EAG, and EAI. This model also implies CFP in the case that the single factor is the returns on the market portfolio (see Bollerslev, Engle and Wooldridge 1988). Section 3 proves that SFM-AR also entails FMR, EL, EAG, and EAI. This section also shows that SFM-AR produces deductively CC, regardless of whether the risk factor is the market index or any other variable. The combined results of Sections 2 and 3 lead to the conclusion that M-MD-GARCH and SFM-AR are D-S explanatory equivalent. Section 4 examines whether this equivalence carries forward to the case in which the stricter D-N-P criteria for explanatory adequacy are adopted. This section starts with a brief presentation of the D-N-P model of explanation and then proceeds to compare this model with D-S. Then it examines the extent to which M-MD-GARCH and SFM-AR satisfy the D-N-P conditions of explanatory adequacy. It is shown that SFM-AR dominates M-MD-GARCH in terms of the D-N-P criteria, although it falls short of achieving the ideal explanatory text, with the latter being defined as the upper limit of a D-N-P explanation. This section argues that one of the main reasons for the partial failure of SFM-AR to satisfy the D-N-P criteria is that this model does not reveal the identity of the risk factors. A special version of SFM-AR which stems from CAPM fares better with respect to the issue of risk factor identification, although other problems pertaining to the causal interpretation of the identified factor are present. Section 5 examines whether the true causal risk factors can be identified empirically. Here, the basic obstacle towards achieving this task lies in the general difficulty of reducing causal relationships to statistical correlations. Various attempts to employ extra-statistical information for unravelling causal connections, such as the adoption of the principle of temporal priority of causes over their effects, are discussed. Section 6 concludes the paper.

1 Deductive Statistical Explanations of the M-MD-GARCH Model

As already mentioned in the Introduction, the Hempelian D-S model of explanation identifies explanation with derivation. This model "...is used to explain a statistical regularity by showing that it follows with necessity from one or more statistical laws (and initial conditions in some cases)." (Salmon 1984, pp. 295). Kinoshita (1990) comments on the issue of regularity explanation as follows: "A regularity explanation does not *amplify* the nature of a particular regularity, but rather *orients* the regularity relative to other regularities. Regularity explanations show a regularity to be reasonable or proper by showing it to be a special case of one or more (more comprehensive) regularities." (Kinoshita, 1990, pp. 301). Specifically, an explanation of S_j orients this regularity within a complex hierarchy of regularities by showing that S_j is a special case or manifestation of S_i. Friedman (1974) defines explanation in terms of unification or conceptual economy. If the number of empirical regularities that have to be assumed as "brute" is minimized, then our understanding of the phenomenon is increased. For example, if S_i and S_j are two different regularities, then the case in which S_i implies S_j achieves a higher order of understanding, than the case in which S_i and S_j are independent. Friedman (1974, pp. 15) argues: "I claim that this is the crucial property of scientific theories we are looking for; this is the essence of scientific explanation - science increases our understanding of the world by reducing the total number of independent phenomena that we have to accept as ultimate or given. A world with fewer independent phenomena is, other things being equal, more comprehensible than with more". Hempel, himself appraises the role of unification in explanation as follows: "What scientific explanation, especially theoretical explanation aims at is not [an] intuitive and highly subjective kind of understanding, but an objective kind of insight that is achieved by a *systematic unification*, by exhibiting the phenomena as *manifestations* of common, underlying structures and processes that conform to specific, testable, basic principles" (Hempel, 1966, pp. 83, **emphasis added**???????????).

The precise situation in which a regularity S_i explains another regularity S_j in the D-S sense may be described as follows: (i) The empirical regularity, S_i, is detected. (ii) S_i derives a probabilistic interpretation, say D_i. (iii) D_i motivates or inspires the creation of a statistical model M_1. For example, M_1 may be thought of as an attempt to express D_i in an explicit parametric form. (iv) M_1 is subject to theoretical analysis which reveals that M_1 exhibits an additional probabilistic property, D_j, over and above D_i, that is $M_1 \Longrightarrow D_j$, with $D_j \neq D_i$. (v) Assume that there exists an empirical regularity S_j whose probabilistic interpretation is D_j. (i)-(v) imply that (a) S_i "explains" S_j in the D-S sense, and (b) M_1 is D-S explanatory for S_j.

Since M-MD-GARCH is an extention of the univariate MD-GARCH, we shall begin our discussion focusing on the latter model. Let us consider a market with n assets (stocks) and let $R_{i,t}$ be the one-period continuously compounded return on an individual stock, defined as $R_{i,t} = p_{i,t} - p_{i,t-1}$, where $p_{i,t}$ is the natural logarithm of the price of the particular stock. By dropping the subscript, i, for notational economy, the simplest MD-GARCH model (MD-GARCH(1,1)) takes the following form:

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

The process $\{R_t - c\}$ where $\{R_t\}$ is described by **(1)** is martingale difference**.** This model was motivated by the empirical regularity of volatility clustring (VC) and its probabilistic interpretation, as dynamic conditional heteroskedasticity (DCH). It is worth noting that there have been historical eras in which VC had derived probabilistic interpretations different from DCH. Specifically, Mandelbrot (1963) was the first to detect volatility clustering and empirical leptokurtosis. However, his interpretation of VC was not in terms of DCH. Instead, he argued that the proper probabilistic interpretations of the observed VC (and EL) should take the form of infinite unconditional variance in the theoretical distribution of stock returns. In other words, Mandelbrot intepreted these regularities not as evidence for non-linear temporal dependence in the returns generating process, but rather as evidence for the non-existence of the second-moments of the returns' unconditional distribution. In his seminal study of scientific progress, Kuhn (1962) argues that the way by which a scientist interprets a given body of evidence at a specific point in time depends on his own theoretical framework formed by the theoretical concepts prevailing at that time. At the beginning of 1960s, probability theory had already produced important limit theorems allowing for convergence of a sum of random varibales to non-Gaussian distributions with infinite variance (see Levy 1925). These results were at the heart of Mandelbrot's "infinite variance" interpretation of empirical leptokurtosis and volatility clustering. On the other hand, results on non-linear stochastic processes, characterised by the property of "asymptotic independence" (or mixing) were not available, or at least widely known, at that period.

As mentioned above, DCH (combined with FMR) motivated the birth of the univariate MD-GARCH model by Engle (1982) and Bollerslev (1986). One of the earliest theoretical result concerning this model is that it entails leptokurtosis of the unconditional distribution of R_t. Another result, which was also proved relatively early, was that $\{R_t\}$ is a second-order stationary process if $a_1 + a_2 < 1$, in which case the unconditional variance of R_t exists and is equal to $a_0/(a_1 + a_2)$. Under this parametric restriction, Diebold (1988) showed that MD-GARCH implies AG and AI. These results were based on the fact that $\{R_t\}$ being a covariance stationary and asymptotically independent (mixing) process, belongs to the domain of attraction of the normal law. This in turn implies that the classic Central Limit Theorem (CLT) apply which in turn entail AG.

The preceding discussion suggests that starting with the assumption that the high frequency returns process, $\{R_t\}$, is MD-GARCH, we are led to the conclusion that the low frequency returns process $\{R_\tau\}$ is approximately normal, since R_τ is the sum of the R_t's occured within τ. This property (AG) also has implications for the dependence properties of the sequence $\{R_\tau\}$. 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.

However, early estimates of a_1 and a_2 were found to be in the viscinity of the unit root area, that is $a_1 + a_2 \simeq 1$. These estimates gave rise to the so-called Integrated GARCH process (MD-IGARCH), that is a process described by (1) 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 raise the question of whether CLT applies in the presence of an infinite variance, which in turn generates doubts as to whether MD-IGARCH entails AG. To this end, 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 (1) with $a_1 + a_2 = 1$ the barely infinite variance hypothesis is stated as $E\,|R_t|^\delta < \infty$ for every $0 \leq \delta < 2$. If the sequence $\{R_t\}$ 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 $\{R_t\}$. 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.The difference between the "barely infinite variance" IGARCH process defined by (1) with $a_1 + a_2 = 1$ and the independent Stable Paretian process proposed by Mandelbrot (1963) is huge as far as their asymptotic properties are concerned. More specifically, in spite of having (barely) infinite variance, an MD-IGARCH process is in the domain of the attraction of the normal law. These results imply that both MD-GARCH and MD-IGARCH entail theoretical leptokurtosis, aggregational normality and aggregational independence. This means that DCH produces TL, AG, and AI even in the case in which DCH is quite strong (IGARCH). The theoretical relationships, $DCH \Longrightarrow \{TL, AG, AI\}$ may be translated in terms of empirical regularities as $VC \Longrightarrow \{EL, EAG, EAI\}$.

Moving to a multivariate setting, we may argue that the multivariate equivalent of VC is CC which in turn gave rise to M-MD-GARCH. To this end, Bauewens, Laurent and Rombouts (2006) claim that "the most obvious application of multivariate GARCH models is the study of the relations between volatilities and co-volatilities of several markets" (2006, pp. 79). Many alternative M-MD-GARCH models have been proposed in the literature stemming from alternative parsimonious specifications of the conditional covariance matrix (see, for example, Bollerslev 1990, Engle 2002). The question that we now face is whether any of these M-MD-GARCH specifications entail CFP. Such a result is obtained by Bollerslev, Engle and Wooldridge (1988) in the context of a multivariate GARCH-in-mean model, for the case in which the risk factor is the market portfolio. *(ADDITIONAL RESULTS????)*

To sum up: This section has surveyed theoretical results, showing that $CC \Longrightarrow \{CFP, EL, EAG, EAI\}$, where CFP is defined solely with respect to market portfolio. Note that CC by itself does not imply FMR, since the probabilistic property of dynamic heteroskedasticity" is independent to that of "martingale difference". Nonetheless, a multivariate GARCH model, in which the MD assumption has been imposed, is capable of functioning as a covering statistical law in D-S explanations for the full set of stock returns regularities, $\mathcal{S}_R$.

2 Deductive-Statistical Explanations of the SFM-AR model

Following the discussion in Introduction, we assume that $R_{i,t}$ is related to a single factor, M_t, via the following relationship:

$$R_{i,t} = a_i + (\beta_i + \beta_{i,t})M_t + u_{i,t}, \ i = 1, 2, ..., n \tag{2}$$

where a_i and β_i are real numbers, and $u_{i,t}$, $\beta_{i,t}$, are zero-mean sequences of random variables whose exact properties will be defined below. Equation (2) can be written in vector form as follows

$$\mathbf{R}_t = \mathbf{a} + M_t(\boldsymbol{\beta} + \boldsymbol{\beta}_t) + \mathbf{u}_t, \tag{3}$$

where $\mathbf{R}'_t = [R_{1,t}, R_{2,t}, \ldots, R_{n,t}]$, $\mathbf{a}' = [a_1, a_2, \ldots, a_n]$, $\boldsymbol{\beta}' = \left[\beta_{1,t}, \beta_{2,t}, \ldots, \beta_{n,t}\right]$ and $\mathbf{u}'_t = [u_{1,t}, u_{2,t}, \ldots u_{n,t}]$.

Assumption M: $\beta_{i,t}$ follows a zero-mean AR(1) process,

$$\beta_{i,t} = \varphi_i \beta_{i,t-1} + \varepsilon_{i,t}, \ \ |\varphi_i| < 1, \ 1 \leq i \leq n \tag{4}$$

and

$$\begin{bmatrix} \mathbf{u}_t \\ M_t \\ \boldsymbol{\varepsilon}_t \end{bmatrix} \sim NIID \left(\mathbf{0}, \begin{bmatrix} \Sigma_u & 0 & 0 \\ 0 & \sigma_m^2 & 0 \\ 0 & 0 & \Sigma_\varepsilon \end{bmatrix} \right) \tag{5}$$

where $\boldsymbol{\varepsilon}_t = [\varepsilon_{1,t}, \varepsilon_{2,t}, \ldots, \varepsilon_{n,t}]'$, $\Sigma_u = diag\left(\sigma^2_{u_1}, \sigma^2_{u_2}, \ldots, \sigma^2_{u_n}\right)$, $\Sigma_\varepsilon = (\sigma_{i,j})_{1 \leq i,j \leq n}$ and $E_{t-1}[M_t] = 0$, where by $E_{t-1}[\cdot]$ we denote the expectation conditional on the information set that is generated by all the random variables under consideration up to time $t-1$.

Let us denote $\sigma^2_{\varepsilon_i} = Var(\varepsilon_i) = \sigma_{i,i}$, $1 \leq i \leq n$,

$$\Sigma_\beta := Var(\boldsymbol{\beta}_t) = E\left[\boldsymbol{\beta}_t \boldsymbol{\beta}'_t\right] = \left(\frac{\sigma_{i,j}}{1 - \varphi_i \varphi_j} \right)_{1 \leq i,j \leq n}$$

and

$$\sigma^2_{\beta_i} = Var\left(\beta_i\right) = \frac{\sigma^2_{\varepsilon_i}}{1-\varphi_i^2} \ .$$

Note that under assumption **M**, equation (3) implies that $\mathbf{R}_t$ is a strictly stationary process with finite second moments. Equation (4) can be also written in vector form as

$$\boldsymbol{\beta}_t = \boldsymbol{\Phi}\boldsymbol{\beta}_{t-1} + \boldsymbol{\varepsilon}_t \ , \tag{6}$$

where $\boldsymbol{\Phi}$=$diag\left\{\varphi_1, \varphi_2, \ldots, \varphi_n\right\}$.

The most important result of this section, namely $SFM{-}AR \Longrightarrow \{CC, FMR, EL, EAG, EAI\}$ takes the form of the following two theorems.

Theorem 1

Equation (2) together with Assumption M, imply that

(i) The conditional covariance matrix of $\mathbf{R}_t$ is given by

$$Var_{t-1}\left(\mathbf{R}_t\right) = \sigma^2_m\Sigma_\varepsilon + \Sigma_u + \sigma^2_m\left(\boldsymbol{\beta} + \boldsymbol{\Phi}\boldsymbol{\beta}_{t-1}\right)\left(\boldsymbol{\beta} + \boldsymbol{\Phi}\boldsymbol{\beta}_{t-1}\right)'$$

(ii) The unconditional distribution of R_t is a mixture of normal distributions and is described by

$$\mathbf{R}_t \sim MN\left(\mathbf{a}, \Sigma_u + \sigma^2_m(\boldsymbol{\beta}+\boldsymbol{\beta}_t)(\boldsymbol{\beta}+\boldsymbol{\beta}_t)'\right) \ , \tag{7}$$

where MN stands for the mixed normal distribution.

(iii) The kurtosis coefficient of the unconditional distribution of $R_{i,t}$ is given by

$$Kurt(R_{i,t}) = \frac{E\left[(R_{i,t} - E[R_{i,t}])^4\right]}{\left(Var(R_{i,t})\right)^2} = 3 + \frac{12\beta^2\sigma^2_\beta\sigma^4_m}{\left(Var(R_{i,t})\right)^2}. \tag{8}$$

(iv) The process $\{\mathbf{R}_t - \mathbf{a}\}$ is a martingale difference.

Proof: see Appendix.

Before we present Theorem 2, concerning EAG and EAI, we need to introduce some further notation. Let us, first, define the k-period return $p_t - p_{t-k}$, where p_t is the logarithm of the stock price at time t. Since we study non-overlapping returns, the series of k-period returns under consideration will be of the form $\{\ldots, p_{t-k} - p_{t-2k}, p_t - p_{t-k}, p_{t+k} - p_t, \ldots\}$. For this reason, we introduce a new index, denoted by τ, which represents the k-period interval, in terms of t. More specifically, if t and τ correspond to the same moment in time, then $\tau + 1$ will coincide with $t + k$. In other words, one unit in terms of τ corresponds to k units in terms of t. **By dropping the subscript, i, for notational economy,** this change of index allows us to denote the k-period returns by

$$R_\tau(k) = p_t - p_{t-k} = \sum_{i=1}^{k} R_{t-k+i} \ .$$

Respectively, for the k-period return at lag 1, we use the notation

$$R_{\tau-1}(k) = p_{t-k} - p_{t-2k} = \sum_{i=1}^{k} R_{t-2k+i}$$

and so forth. In the subsequent paragraphs we will make use of the notation "$\tau - l$" and "$\tau + l$", instead of "$t - lk$" and "$t + lk$", where $l \geq 0$. We directly observe that since $R_\tau(k) - ka$ is a martingale difference process,

$$Var(R_\tau(k)) = kVar(R_t) = k\left(\left(\beta^2 + \sigma_\beta^2\right)\sigma_m^2 + \sigma_u^2\right) \ . \tag{9}$$

The next theorem proves that the sequence of weighted sums of returns, as described by SFM-AR, satisfies an invariance principle. To this end, let us fix some $t_0 \in \mathbb{Z}$, set

$$S_{t_0,k} = \sum_{i=1}^{k} \left(R_{t_0+i} - E\left[R_{t_0+i}\right]\right)$$

and for $0 \leq r \leq 1$, define

$$W_k(r) = \sum_{i=1}^{[rk]} \left(R_{t_0+i} - E\left[R_{t_0+i}\right]\right),$$

where for $r < 1/k$, $W_k(r) := 0$. Now we are ready to present Theorem 2.

Theorem 2

Equation (2) together with Assumption M, imply that

$$\frac{1}{\sqrt{Var(R_\tau(k))}} W_k \xrightarrow{D} W, \text{ as } k \to \infty,$$

where W is a standard Brownian motion and "$\xrightarrow{D}$" denotes the usual weak convergence on the real line.

Remark:

Note that $R_\tau(k)$ does not have a well defined limit as $k \to \infty$. This fact, does not allow us to obtain any conclusion with respect to the independence between $R_\tau(k)$ and $R_{\tau-1}(k)$ as $k \to \infty$, since the definition of asymptotically independent random sequences requires that they are stochastically bounded. On the other hand, Theorem 2 implies that

$$K_\tau(k) := (R_\tau(k) - E[R_\tau(k)]) / \sqrt{Var(R_\tau(k))} \xrightarrow{d} N(0,1) \text{ as } k \to \infty, \tag{10}$$

where by $N(0,1)$ we denote the standard Gaussian distribution. By virtue of (9), we can re-write (10) as follows:

$$(R_\tau(k) - E[R_\tau(k)]) / \sqrt{k} \xrightarrow{d} N\left(0, \left(\beta^2 + \sigma_\beta^2\right) \sigma_m^2 + \sigma_u^2\right) \tag{11}$$

as $k \to \infty$. The left hand sides in (10) and (11) provide us with sequences (of k) with well defined limits. Theorem 2 implies that for every τ, $K_{\tau-1}(k)$ and $K_\tau(k)$ are asymptotically independent as $k \to \infty$. In other words, this proves the asymptotic independence between the de-meaned and properly standardized long-horizon returns $K_{\tau-l}(k)$ and $K_\tau(k)$, for every $l \neq 0$, as the return horizon, k, tends to infinity.

3 Causal Mechanistic Explanation: M-MD-GARCH versus SFM-AR

The previous sections have shown that both M-MD-GARCH and SFM-AR entail all the empirical regularities of stock returns that belong to $\mathcal{S}_R$. Is this all that we require in order to claim that the regularities in $\mathcal{S}_R$ have actually been explained by M-MD-GARCH and SFM-AR? Or do we need to trace the causal process that led to the emergence of M-MD-GARCH and SFM-AR? In Salmon's terms: is explanation achieved merely by showing "*that* phenomena fit into a *nomic* nexus"? Or explanation requires the achievement of the far more ambitious task of demonstrating "*how* phenomena fit into a *causal* nexus"? (1984, pp. 20). In the context of explanations of stock returns regularities, the last question may be translated into the following one: "where do M-MD-GARCH and SFM-AR come from?". Do we need to answer this question in order to have an adequate explanation of $\mathcal{S}_R$? Or instead, is the subsumption of $\mathcal{S}_R$ under M-MD-GARCH or SFM-AR all that matters for the explantion of $\mathcal{S}_R$? In the context of D-S model, the answer to the last question is "yes". According to Hempel, the obvious question "what is the explanation of M-MD-GARCH or SFM-AR?", or "where does M-MD-GARCH or SFM-AR come from?" is an entirely different question, which does not have to be answered before an explantion of $\mathcal{S}_R$ is achieved.

More specifically, one of the distinctive features of the D-S type of explanations analyzed above is that once an empirical regularity S_j is derived from S_i, that is once the result $S_i \Longrightarrow S_j$ is established, a D-S explanation of S_j (by means of the broader regularity S_i) is achieved. Moreover, this explanation is complete in the sense that there is no need to inquire into the origins of the "covering" regularity S_i. The latter may have purely empirical grounds, that is, it may have been inductively inferred from the available data.

Next, assume that the aforementioned covering regularity S_i is explained by an even broader empirical regularity S_k. The fact $S_k \Longrightarrow S_i$ is considered to be another D-S explanation, (of S_i in terms of S_k) the presence of which does not add to the value of the first explantion, $S_i \Longrightarrow S_j$. Salmon (1984) comments on this issue as follows: " Such explanations (D-S) are complete. If one wants an explanation of a law that entered into the first explanation (S_i, in our case) it can be supplied by deriving that law from more general laws or theories (S_k, in our case). The result is another explanation. The fact that a second explanation of this sort can be given does nothing to impugn the credentials of the first explanation" (1984, pp. 156).

As already mentioned, Railton (1978, 1981) disagrees with this view by arguing that the principle aim of explanation is to enhance our understanding of "how the world works." His D-N-P model, which belongs to the class of C-M models of explanation, asserts that the mere subsumption of a narrow regularity, S_j, under the broader regularity, S_i does not constitute an explanation of S_j unless S_i is backed up with "an account of the mechanism(s) at work." More specifically, S_j is explained by placing it within a web of "inter-connected series of law-based accounts of all the nodes and links in the causal network culminating in the explanandum, complete with a fully detailed description of the *causal mechanisms* involved in the theoretical derivations of all covering laws involved" (1981, pp. 174, emphasis added). This means that in the case under study, neither M-MD-GARCH nor SFM-AR in themselves can form the sole basis for a satisfactory explanation of $\mathcal{S}_R$, unless either M-MD-GARCH or SFM-AR is derivable from a theory concerning the causal mechanism at work. Railton's view is best expressed in the following paragraph "If one inspects the best-developed explanations in physics or chemistry textbooks and monographs, one will observe that these accounts typically include not only derivations of lower-level laws and generalizations from higher-level theory and facts but also attempts to *elucidate the mechanisms* at work (1981, pp. 242, italics in original). This means that a D-N-P explanation of S_j is achieved only when the covering law regularity S_i - or more precisely the model that was originated by it - is itself deduced from the theoretical account of the chance mechanism at work.

The explanatory web mentioned above is what Railton defines as "an ideal explanatory text." The full derivation of M-MD-GARCH or SFM-AR from their elementary parts constitutes the full ideal text relevant to M-MD-GARCH or SFM-AR. In other words, full understanding of M-MD-GARCH or SFM-AR requires the full ideal D-N-P text.

The preceding discussion must have clarified the central role that the notion of "mechanism", in general, and "chance mechanism", in particular, play in the context of D-N-P explanations. This is why D-N-P explanation is usually thought to be a special case of the so-called "causal mechanistic explanations". But what exactly a "mechanism" is? Railton does not offer an exact definition of this concept, but rather he describes it indirectly by arguing that an "account of the mechanism(s)" is "a more or less complete filling-in of the links in the *causal chains*" (1978, pp. 748, emphasis added). Glennan (2002) gives the following definition "A mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations." (2002, pp. 344). Based on this definition, he then proceeds to define the so-called mechanistic explanation of empirical regularities: "To mechanistically explain a regularity, one describes a mechanism whose behavior is characterized by that regularity" (2002, pp. 346). Similarly to Railton's view of explanation, Glennan's account of explanation of an empirical regularity requires a description of the internal workings of the mechanism that produces this regularity. This view gives rise to the so-called mechanical model: "A mechanical model is a description of a mechanism, including (i) a description of the mechanism's behavior; and (ii) a description of the mechanism which accounts for that behavior." (2002, pp. 347).

Cohnitz *(????)* argues that the causal-mechanistic approach to explanation supports the following claims: "(I) If some explanatory mechanism M can be reduced to some more fundamental mechanism M', any explanation E' referring to M' contains more information about the ideal explanatory text than any explanation E referring to M". and "(II) If some explanation E' contains more information about the ideal explanatory text than some rival explanation E, E' has a higher explanatory value than E" *(????, pp. 23).*

The preceding discussion suggests that according to the mechanistic view of explanation, an adequate explanation of the set $\mathcal{S}_R$ of stock returns regularities is not achieved unless either M-MD-GARCH or SFM-AR are deduced from the chance mechanism at work in the stock market. In other words, these models should have been derived from some kind of theory about the mechanism that produces returns, "without appeal to particular facts". Only then, an "understanding of the process" by which returns are generated would have been achieved. Moreover, as will be discussed below, the extra-statistical condition, referred to as C1, implied by the phrase "without appeal to particular facts" is particularly binding for both M-MD-GARCH and SFM-AR. It must be emphasized at the outset that C1 imposes severe restrictions on the allowable origins of these models, which are not usually met in econometric practice. Wold (1969) remarks that "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).

The extent to which M-MD-GARCH and SFM-AR satisfy the conditions of D-N-P model of explanation, that is the extent to which these models have purely theoretical origins, is analyzed below.

3.1 Theoretical Origins of M-MD-GARCH

Since M-MD-GARCH is a direct descendent of the univariate MD-GARCH model, the discussion will be focused on the origins of the latter. The theoretical origins of MD-GARCH are poor if non-existent. This model was born out of purely empirical considerations of the behaviour of stock returns. The success of this model does not stem from its theoretical underpinnings, but rather from its ability to generate forecasts for the volatility of stock returns. It is noteworhty that, despite the widespread adoption of the MD-GARCH models in the empirical finance literature, these models originated in the context of empirical macroeconomics. Engle (2003) describes the genesis of this model as an attempt to "...get variances into macroeconomic models, because some people thought it was actually not the expected value of economic variables but rather their variability that was relevant for business cycle analysis" (2003, pp. 1176). Indeed, first this model was applied to the UK inflation rate (see Engle 1982). Moreover, the MD-GRCH models were not developed as a direct attempt to capture the volatility patterns that were observed in macroeconomic time series but instead with the aim of obtaining a powerful test for detecting bilinearity. As Engle remarks "...I discover the model from the test, rather than the other way round" (2003, pp. 1177).

The purely empirical origins of MD-GARCH are also reflected on the plethora of alternative conditional variance specifications that have been suggested in the literature during the last thirty years or so. Each time that a "new regularity" was detected (or was thought to be detected) MD-GARCH was enjoyed (or suffered) ad hoc modifications in an attempt to account for this regularity. As a result, the econometrics literature (with the present paper being no exception) is flooded with acronyms whose common subset is the word ARCH *(EXAMPLES????)*.

The preceding discussion suggests that MD-GARCH and M-MD-GARCH, having no theoretical origins, do not satisfy the conditions of the D-N-P model, thus failing to explain the regularities in $\mathcal{S}_R$. As a result, the usual claim that "conditional heteroskedasticity explains leptokurtosis" is considered to be false under the prism of the D-N-P model.

3.2 Theoretical Origins of SFM-AR

Since SFM-AR is an extention of the constant-beta SFM, the discussion will naturally begin from the theoretical origins of the latter. SFM is defined as follows:

$$R_{i,t} = a_i + \beta_i M_t + u_{i,t},\ i = 1, 2, ..., n \tag{12}$$

where M_t is the unanticipated changes of the risk factor, X_t, that is $M_t = X_t - E(X_t \mid \Phi_{t-1})$. Concerning, the non-systematic term, $u_{i,t}$, a minimal assumption is that it is a martingale-difference with respect to the publicly available information set Φ_{t-1}. For reasons that will become clear in the sequel, let us introduce the multiple factor model, MFM,

$$R_{i,t} = a_i + \beta_{1,i} M_{1,t} + \beta_{2,i} M_{2,t} + ... + \beta_{k,i} M_{k,t} + u_{i,t},\ i = 1, 2, ..., n, \tag{13}$$

where $M_{l,t} = X_{l,t} - E(X_{l,t} \mid \Phi_{t-1})$, $l = 1, 2, ..., k$. The first thing to note is that according to the inventors of this model, SFM (or MFM) violates the D-N-P condition C1, which states that the thoretical derivation of the model should be made "without appeal to particular facts". Roll and Ross (1980) state explicitly this fact: "...the APT (Arbitrage Pricing Theory) is based on a linear return generating process *as a first principle...*" (1980, pp. 1074). They motivate the birth of this model by appealing not to a theory but to observable facts, namely to "the single most widely-acknowledged empirical regularity of asset returns, their common variability" (1980, pp1073).

However, this statement does not do justice to the theoretical origins of SFM (or MFM). Even if all that motivated SFM was to account for the common variability of returns, the specific way by which SFM does so is causal. More specifically, SFM accounts for the common variabilty of stock returns by adopting the "common cause principle", (CC). This principle states that if two events A and B are correlated in the sense that

$$P(A \cap B) > P(A)P(B) \tag{14}$$

then either A probabilistically causes B, or B probabilistically causes A, or A and B are independent but there is a common cause, C, causing both A and B. In the latter case the three events form a "conjuctive fork", ACB, that is, they satisfy the following relationships:

$$
\begin{aligned}
P(A \cap B \mid C) &= P(A \mid C)P(B \mid C) \\
P(A \cap B \mid \overline{C}) &= P(A \mid \overline{C})P(B \mid \overline{C}) \\
P(A \mid C) &> P(A \mid \overline{C}) \\
P(B \mid C) &> P(B \mid \overline{C})
\end{aligned} \tag{15}
$$

Moreover, as Reichenbach *(1946????)* showed, relationships (15) imply (14). These relationships imply that: (i) conditioning on a common cause and conditioning on the absence of a common cause, renders the effects independent, and (ii) the existence of a common cause raises the probability of both effects individually.

The CC interpretation of SFM assumes that M_t is the *only* common cause of $R_{i,t}$, $i = 1, 2, ..., n$. This in turn implies that the observed correlations among the $R_{i,t}$'s stem solely from their common causal relationship to M_t. This causal interpretation of SFM imposes a restriction on the error terms $u_{i,t}$, $i = 1, 2, ..., n$ (which is sometimes ignored in the literature), namely that their correlation matrix, Σ_u, is diagonal. The diagonality of Σ_u may be thought of as a "theoretical" restriction reflecting the view that the common risk factor M_t "screens off" the correlations among the $R_{i,t}$'s.

Further analysis of the CC interpretation of SFM reveals that this model is likely to have deeper theoretical origins. More specifically, what is the "causal process" by which a change in M_t brings about a change in $R_{i,t}$, $i = 1, 2, ..., n$? Economic theory answers this question as follows: Assume that investors in the market are risk averse so that they require a (time-invariant) risk premium, ρ_i, in order to hold the risky asset i. This premium is the expected returns promised by the asset i over and above the return of a risk-free asset, that is

$$E(p_{i,t} \mid \Phi_{t-1}) - p_{i,t-1} = \rho_i + r \tag{16}$$

where $p_{i,t}$ is the logarithm of the price of asset i at period t. The above relationship shows that the demand for the risky asset i at $t-1$, and hence the current price $p_{i,t-1}$, depend on investors' expectation, $E(p_{i,t} \mid \Phi_{t-1})$, for next period's asset price. This expectation in turn depends on the expectation, $E(X_t \mid \Phi_{t-1})$ that investors form about the future level of the variable X_t. This is because investors themselves have chosen this particular variable to "drive" their decisions. This reasoning may be represented by the following causal chain:

$$E(X_t \mid \Phi_{t-1}) \longrightarrow E(p_{i,t} \mid \Phi_{t-1}) \longrightarrow p_{i,t-1}. \tag{17}$$

How does the price of the asset i change between periods $t-1$ and t, that is how the returns $R_{i,t} = p_{i,t} - p_{i,t-1}$ are generated? At period t, the actual value of the factor X_t is realized. Investors compare the actual value X_t with their expectations $E(X_t \mid \Phi_{t-1})$ for this value that formed at period $t-1$. If $X_t - E(X_t \mid \Phi_{t-1}) > 0$, that is when a positive surprise occurs, investors update their expectations $E(X_{t+1} \mid \Phi_t)$ upwards, which in turn, via the causal chain (17), leads to the formation of a price $p_{i,t}$ higher than $p_{i,t-1}$. As a result, positive (negative) surprises concerning the factor X lead to positive (negative) returns for asset i. Moreover, the response of asset price to a factor surprise is not the same for all the n assets. This makes investors to require a greater risk premium for holding asset i than that for holding asset j if the response (beta) of asset i is greater than that (beta) of asset j.

It is important to note that in spite of the fact that all the quantities in (17) appear to be simultaneously determined at period t-1, in reality (that is, in actual rather than in model's time) the underlying events occur in a well-defined temporal order. First, an event from the information set Φ_{t-1} materializes, second the expectation $E(X_t \mid \Phi_{t-1})$ is formed, third, the expectation $E(p_{i,t} \mid \Phi_{t-1})$ is generated and finally the price $p_{i,t-1}$ is determined. This temporal order characterising the sequence of relevant events leading to the formation of price is consistent with the philosophical principle that "causes precede their effects in time". Of course, the time intervals between the occurence of the aforementioned events may be quite small. The Efficent Market Hypothesis (EMH) asserts that the speed at which the information Φ_{t-1} is processed and reflected to price $p_{i,t-1}$ is so high that the events may be thought of as "simultaneous". However, the "instantaneous" process of information, usually met in the definitions of EMH, should not be interpreted literally but instead should be taken to imply that this information processing is completed in very short time. We shall return to the issue of temporal priority of causes over their effects in the next Section, in which the role of the exact time of occurence of the relevant events in the identification of risk factors will be analyzed in detail.

EMH is usually identified with the rational expectations hypothesis (REH). REH asserts that investors' subjective expectations, $\mathcal{E}(X_t \mid \Phi_{t-1})$, coincide with the objective mathematical expectations, that is $\mathcal{E}(X_t \mid \Phi_{t-1}) = E(X_t \mid \Phi_{t-1})$. REH may be thought of as referring to properties (expectations) of the constituent parts (investors) of the stock market mechanism. Moroever, SFM in conjuction with REH implies that $\{R_{i,t}\}$ (or more accurately $\{R_{i,t} - a_i\}$) is a martingale difference sequence. This can be easily seen by operating on both sides of (12) with the objective operator $E(\cdot \mid \Phi_{t-1})$. On the contrary, if $\mathcal{E}(X_t \mid \Phi_{t-1}) \neq E(X_t \mid \Phi_{t-1})$, then it can be shown that

$$E(R_{i,t} \mid \Phi_{t-1}) = a_i + b_i \mathcal{E}(X_t \mid \Phi_{t-1})$$

with $\mathcal{E}(X_t \mid \Phi_{t-1}) \in \Phi_{t-1}$. In such a case, $\{R_{i,t} - a_i\}$ is not MD with respect to Φ_{t-1}.

Remark

The MD property of asset returns in the context of SFM is "derived", rather than imposed. As a result the corresponding empirical regularity FMR is derived from the (more fundamental) regularity CF. On the contrary, in the context of the MD-GARCH model, MD has to be imposed as an independent assumption, since it is not deducible from DCH.

The preceding paragraphs have presented a brief outline of the argument underlying the causal interpretation of (12) or (13). This argument is theoretical in the sense that it could have been advanced out of purely theoretical considerations of the "mechanism" under study, without any appeal to empirical evidence.

One important caveat of SFM that has serious implications for its causal interpretation concerns the identity of the risk factor X_t (or, more generally, the identity of $X_{1,t}, X_{2,t}, ..., X_{k,t}$ in (13)). To this end, we may distinguish two cases: (i) The risk factors are not measurable, that is they are not expressed in terms of observable random variables. The employment of unobservable theoretical concepts in the construction of a scientific theory is justified so long as these concepts appear only in higher-level hypotheses being eventually eliminated as the theory passes to lower-level hypotheses or empirical generalizations (see, for example, Braithwaite 1964). However, in the case under study, the unobservable entities, namely the risk factors appear to be present not in the higher-level assumptions (or the most primitive parts of the mechanism at work) but in the lower-level empirical hypotheses, namely SFM itself. In such a case, the only viable alternative is to approximate the true unobservable factors with some observable proxies. In fact, this is what is usually done in practice. However, as will be shown below this practice has very serious implications for the explanatory status of SFM. (ii) The risk factors are observable. In this case, the theoretical account of the mechanism at work ought to reveal the identity of these factors, if the ideal explanatory text is to be achieved. As in the previous case, if the risk factors that are eventually identified as such are not the true causal (and observable) factors, then the explanatory value of SFM is seriously weakened.

On this point, Roll and Ross themselves raise the question "What are the common or systematic factors?" (1980, pp. 1077). In searching the identity of these factors, Roll and Ross argue as follows: "If there are only a few systematic components of risk, one would expect these to be related to fundamental economic aggregates, such as GNP, or to interest rates or weather (*although no causality is implied by such relations*)" (1980, pp. 1077, emphasis added). Roll and Ross seem to suggest that the systematic components of risk (the real causal factors) are likely to be non-identifiable. However, they suggest that the true factors are likely to be related to observable macroeconomic variables such as GNP.

As a result, in empirical tests of APT, these macroeconomic variables can approximate the true causal risk factors. However, one important question that has already been raised above is the following: What are the implications of replacing the true causal factors with the approximating ones for the causal interpretation of (12) or (13)? The answer to this question is "devastating". Salmon has forcefully argued that if we wish to put "cause" back to "because", then we should be able to identify the true causal factors for the phenomenon of interest (see Salmon 1971, 1984, 1989). If we fail to do so, then the purported explanation is flawed, because irrelevancies are harmless to predictions but fatal to explanations. To draw an analogy, "explaining" the returns $R_{i,t}$ by means of the symptomatic factor Y_t rather than the true causal factor X_t is equivalent to explaining the level of rainfall in a particular region by means of the readings of a well-functioning barometer rather than the atmospheric conditions in that region.

The failure of SFM (or MFM) to identify the identity of the causal factors, provided that these factors are indeed identifiable (observable) is surely a major drawback of this model as far as its explanatory status is concerned. Nevertheless, the analysis of the origins of SFM presented above seems to suggest that some parts of the ideal explanatory text have been produced. More specifically, the aforementioned theory may be thought of as accounting for the "structure", but not for the "identity of the constituent parts", of the chance mechanism at work. In fact, this theory already delivers more than was intended to do, because SFM (or MFM) was not meant to be explained by some higher-level theory, but instead it was put forward as the starting point for developing APT. Roll and Ross state explicitly this fact: "...the APT is based on a linear return generating process *as a first principle*..." (1980, pp. 1074). As already mentioned, they motivate this model by appealing not to a theory but to observable facts, namely to the common variability of asset returns. As Roll and Ross admit "We do consider the basic underlying causes of the generating process of returns to be potentially important area of research, but we think it is an area that can be investigated seperately from testing asset pricing theories" (1980, pp. 1077).

The question which naturally arises at this point is the following: If we think of the stock market as a chance mechanism that generates the observed stock returns regularities, then is there any detailed theory that trace the causal chains leading up to SFM (or MFM), with the identity of the factor(s) being identified by the theory itself? The answer to this question is partially yes. As already mentioned, a special version of SFM corresponds to the case in which X_t represents the returns of the "market portfolio". A well-known theory that leads to this result is offered by the Capital Asset Pricing Model (CAPM). This case is analyzed below.

3.2.1 SFM and CAPM

It is well known that the history of asset pricing theory starts with CAPM, proposed by Sharpe (1964) and Litner (1965). One feature of CAPM that makes it quite attractive as a theory that describes the internal workings of the stock market mechanism is that it starts by postulating fundamental properties of the constituent parts of the mechanism. More specifically, CAPM is based on some (rather strong) assumptions concerning investors' preferences and behaviours. Indeed, Roll and Ross admit that "elegant derivations of the CAPM equation have been concocted beginning from the first principles of utility theory" (1980, pp 1074). Since CAPM is very well known, we shall focus only to the points which are relevant for the intended discussion.

An intermediate result in the context of CAPM is that the expected utility $\mathcal{E}(U_j)$ of investor j is a function of only two parameters, namely the objective mean $\mu(R_p)$ and the objective standard deviation $\sigma(R_p)$ of the returns, R_p, of the portfolio p of risky assets. This result, hereafter referred to as Expected Utility Property (EUP), consists of two parts. The first part is that each agent j evaluates alternative portfolios p in terms of only the first two subjective moments $\mu_j^e(R_p)$ and $\sigma_j^e(R_p)$ of R_p. The second part is that all the agents have homogeneous beliefs in the sense

$$\begin{aligned} \mu_j^e(R_p) &= E(R_p) \equiv \mu(R_p),\ \forall j \\ \sigma_j^e(R_p) &= \sigma(R_p),\ \forall j. \end{aligned} \tag{18}$$

Since EUP is fundamental in deriving the CAPM-based SFM equation, the question which naturally arises at this point is how EUP itself is derived. Two alternative types of sufficient conditions for EUP have been suggested in the literature. The first type imposes restrictions on the distribution of returns, whereas the second imposes restrictions on the form of investors utility functions. As will be shown below, these two sets have different implications for the explanatory status of the CAPM-based SFM.

The first type of conditions, referred to as CAPM-D, assumes that the joint distribution of the returns R_i of all the existing assets, $i = 1, 2, ..., n$ is Gaussian (or more generally, elliptical see Owen and Rabinovitch 1983). A further assumption within CAPM-D is that each investor's subjective joint probability distribution $F^e_{j,[R_1,R_2,...,R_n]}(y_1, y_2, ..., y_n)$ of returns coincides with the corresponding objective one $F_{[R_1,R_2,...,R_n]}(y_1, y_2, ..., y_n)$, that is

$$F^e_{j,[R_1,R_2,...,R_n]}(y_1, y_2, ..., y_n) = F_{[R_1,R_2,...,R_n]}(y_1, y_2, ..., y_n),\ \forall (y_1, y_2, ..., y_n) \in \mathbb{R}^n. \tag{19}$$

with the latter, $F_{[R_1,R_2,...,R_n]}(y_1, y_2, ..., y_n)$, being Gaussian (or more generally, elliptical). The rationality condition (19) implies that all investors share the same ex ante views $\boldsymbol{\mu}_j^e$ and Σ_j^e of the mean vector $\boldsymbol{\mu}$ and the covariance matrix Σ of the random variables $R_1, R_2, ..., R_n$ respectively, and also that these views turn out to be "correct". Specifically,

$$\begin{aligned} \boldsymbol{\mu}_j^e &= \boldsymbol{\mu},\ \forall j \\ \Sigma_j^e &= \Sigma,\ \forall j \end{aligned} \tag{20}$$

As a consequence of (20), and since any R_p is a linear function of the jointly normal random variables $R_1, R_2, ..., R_n$, it follows that all investors agree that the distribution of R_p is normal with mean and standard deviation equal to $E(R_p)$ and $\sigma(R_p)$ respectively. Hence, equalities (18) obtain.

The relations (18) state that the agents not only share the same ex-ante views of the mean and standard deviation of the returns R_p of each possible portfolio p, but these views coincide with the "true" or "objective" mean and standard deviation of R_p. Moreover, the joint normality assumption implies that the distribution of any possible R_p is not only perceived by investors to be normal but it is actually normal. The rationality assumption establishes a link between ex ante beliefs and ex post realizations and is based on the thesis that investors do not make systematic mistakes. More specifically, each investor tests the "correctness" of his own beliefs $\boldsymbol{\mu}_j^e$ and Σ_j^e by comparing these beliefs with the objective moments $\boldsymbol{\mu}$ and Σ, respectively. If he observes systematic deviations between the subjective and objective moments then he tends to revise his beliefs in the direction of the objective moments. This procedure, referred to as *statistical learning* ensures that each investor will eventually equate his subjective beliefs with the corresponding objective ones, and by this mechanism the homogeneity of beliefs is achieved. Ross (1978) comments on this as follows: " ... it is natural to assume that investors do not err in a systematic fashion in their a priori beliefs. It follows that the ex post distribution from which returns are drawn will be the ex ante one on which investors based their actions" (pp. 889). From these assumptions, it follows immediately that each investor's expected utility $\mathcal{E}(U_j)$ is now defined with respect to the true mathematical operator $E(\cdot)$, that is $\mathcal{E}(U_j) = E(U_j)$ and also that $E(U_j)$ is a function of $E(R_p)$ and $\sigma(R_p)$, thus ensuring EUP.

The preceding analysis reveals a fundamental inconsistency in the causal chain of events leading to the CAPM-based SFM. More specifically, in order for the investors to act in the manner implied by EUP, the result of their actions, namely the joint distribution of returns, must be known. Put it differently, before the cause is determined the effect must have occured. This circularity is usually overshadowed by the fact that CAPM-D is intended to describe investors' behaviour in a state of equilibrium. As such, however, CAPM-D does not describe the path to equilibrium, or put differently, the causal mechanism by which this equilibrium is achieved. This may be thought of as an instance of what Hoover (1993) refers to as "a great historical divide in economics between analyses based on process and analyses based on equilibrium" (1993, pp. 695). For these reasons, we can conclude that CAPM-D does not satisfy the conditions for an adequate D-N-P explanation of SFM.

The second type of sufficient conditions for EUP, hereafter refered to as CAPM-U, are restrictions on the form of investrors' utility functions, which in turn imply restrictions on investors's preferences. Specifically, utility functions are assumed to be quadratic. Under quadratic utility functions, each investor expected utility $\mathcal{E}(U_j)$ is a function of his own subjective beliefs on the first two moments of R_p, namely $\mu_j^e(R_p)$ and $\sigma_j^e(R_p)$. However, this fact alone does not ensure EUP, since this property requires homogeneity of beliefs in the sense $\mu_j^e(R_p) = \mu^e(R_p)$ and $\sigma_j^e(R_p) = \sigma^e(R_p)$, $\forall j$. Put differently, the homogeneity of beliefs is not ensured in any way by the assumption of quadratic utilities. This assumption must be established independently of the assumption of quadratic utilities. Therefore, the question which arises at this point is how this homogeneity of beliefs is established in the absence of any reference to something "objective" such as rationality condition (19). In the case of CAPM-D, we showed that the homogeneity of beliefs is achieved by the mechanism of statistical learning. In the case of CAPM-U, the homogeneity of beliefs has to be achieved by another mechanism. We refer to this mechanism as *Bayesian conditionalization* and we analyze its basic structure below.

First of all, each investor has his own measures of beliefs (subjective probabilities) of events of the type $(R_1 \in [a, b], R_2 \in [c, d], ..., R_m \in [k, l])$. If CAPM-U is construed as a normative theory of decision under uncertainty, then these subjective probabilities have to satisfy the constraints of probability calculus (see Kyburg 1978). The usual justification for this rationality of subjective beliefs is that no investor is susceptible to a Dutch Book, with the latter being a sequence of bets which the investor may accept individually, but which collectively guarantee investor's loss with absolute certainty *(see ????)*. This new type of rationality ensures that each investor computes $\mu_j^e(R_p)$ and $\sigma_j^e(R_p)$ from his own subjective probabilities on $R_1, R_2, ..., R_m$ according to the rules of probability calculus. Second, assume that initially $\mu_j^e(R_p)$ and $\sigma_j^e(R_p)$ are initially different among the investors. However, the ranking of alternative portfolios by a specific investor, s, in terms of his own $\mu_s^e(R_p)$ and $\sigma_s^e(R_p)$ may turn out to be systematically more profitable than the rankings of the other investors. In the light of this evidence, the other investors are likely to update their beliefs bringing them closer to those of investor s. If this updating by conditionalization continues for some periods, then the beliefs of all the investors will converge towards the beliefs of the most profitable investor, s. Sooner or later, all the investors, through the process of Bayesian conditionalization will end up having adopted the beliefs of the investor s, and homogeneity of beliefs will have been achieved. Note that this process does not involve any reference to the objective probability distribution of returns. However, such a concept can still be defined in terms of hypothetical limiting frequencies, or (after some periods at which the CAPM-U procedure is repeated) actual relative frequencies. The question is how do these objective probabilities relate to the subjective ones. In the case that there exists an objective reality whose probabilistic structure is independent of agents beliefs, or probabilities whose behavior is in the words of Fitelson, Hajek and Hall (2006) "independent of anyone's mental state", then standard "convergence theorems" may be used to claim that subjective beliefs will converge to the corresponding objective probabilities. To this end Fitelson, Hajek and Hall (2006) argue: "Bayesians reply that various convergence theorems show roughly that in the long run, agents who do not give probability 0 to genuine possibilities, and whose stream of evidence is sufficiently rich, will eventually be arbitrarily close to certain regarding the truth about the world in which they live" *(????, pp. ???)*. However, in the case under study, there is no "reality" in the form of a physical system, existing independently of the investors' subjective beliefs. The only reality that is relevant are the observed stock returns themselves, whose emergence and hence relative frequencies depend heavily on investors subjective beliefs. Since there is no clear dichotomy between subjective beliefs and objective world, the convergence theorems mentioned above might not apply. However, another type of convergence may be contemplated: If all the investors end up having homogeneous subjective probabilities (produced by Bayesian conditionalization), then it is tempting to assume that the objective probabilities which will eventually emerge will be equal to these subjective probabilities. In a sense the equation (19) re-emerges although from a different route. In CAPM-D, it is the objective probabilities that come first and dictate the way at which the subjective probabilities are formed. In CAPM-U, it is the other way round; the subjective probabilities come first, dictate investos' actions and generate returns, whose relative frequencies are the objective probabilities.

The preceding discussion seems to suggest that CAPM-U is closer to achieving the theoretical objectives set by the causal mechanistic view of explanation. However, the assumption of quadratic utilities impose a strong and rather unrealistic restrictions on investors preferences. Moreover, as analyzed above, the theory is still in need of the assumption that all investors agree on the means and standard deviations of all the available candidate portfolios. The way by which such homogeneity is achieved may be the process of learning by experience discussed above. However, many authors believe that the assumption of homogeneity of beliefs is actually imposed without any realistic justification. As Ross (1978) puts it "Given that such homogeneity is going to be imposed eventually, it would seem natural to begin the CAPM story with restrictions on distributions, rather than preferences" (1978, pp. 888). In other words, it seems "more natural" to obtain the key assumption of CAPM by imposing restrictions on the returns generating process itself rather than by imposing restrictions on investors' preferences. As Ross puts it: "A theory that obtains strong implications for equilibrium asset prices from restrictions on perceived distributions and permits heterogeneity in preferences is surely to be preferred to one which obtains similar market implications, but imposes restrictions on preferences along with strong similarity of beliefs" (1978, pp. 888). In view of the above analysis, we conclude that we cannot have, in the context of CAPM, a fully satisfactory account of the causal mechanism that leads up to the CAPM-based SFM.

The inability of CAPM to provide a well articulated account of the chance mechanism at work, leading up to the CAPM-based SFM equation, is also reflected on the fact that this equation cannot derive a causal interpretation. This shortcoming is due to the violation of the temporal asymmetry between cause and effect implied by the nature of the variable identified by CAPM as the risk factor. More specifically, the unique risk factor in the CAPM-based SFM is the (excess) returns, $R_{M,t}$, of the "market" portfolio. This factor $R_{M,t}$, being defined as the weighted average of the returns of all the assets in the economy, includes the returns, $R_{i,t}$, of the specific asset i. This in turn deprives the CAPM-based SFM from a causal interpretation consistent with the principle that causes preceed their effects in time. Indeed, $R_{M,t}$ is not determined until all the returns in the economy, including $R_{i,t}$, are determined. In Wold's words the equation (12) is "not realizable in the sense of computer simulation". More specifically, the right-hand variable, $R_{M,t}$, in (12) "makes input in the computer simulation, and the left-hand side variables ($R_{i,t}$, in the present case) are the output of the simulation; it is not realizable to require that the output of a simulation process is part of the input." (1969, pp. 465). Therefore, $R_{M,t}$ cannot be thought of, even in principle, as a causal determinant of $R_{i,t}$.

4 Empirical Identification of Risk Factors

The analysis of the previous sections provides us with the necessary background to address the following question: To what extent is it currently possible to derive the ideal explanatory text concerning the empirical regularities of stock returns? Below we list the conclusions we have reached with respect to the extent to which a derivation of SFM (or MFM) from a theoretical account of the chance mechanism at work, is currently possible :

(i) A complete version of such an account, describing the causal nexus in full detail is clearly missing. Our best theories, based on the common cause principle, EMH and forward looking behaviour, can offer at best an outline of the structure of the chance mechanism. Returns are generated by investors' efficient reactions to (rationally) unanticipated changes in common risk factors.

(ii) Our best theories have failed, so far, to identify the identity of the true causal risk factors. This failure is more sever if these factors are indeed identifiable (observable). Those that have attempted to do so (CAPM) pay a rather high "identification cost", namely they are forced to sacrifice the fundamental causal principle, stipulating that causes precede their effects in time. Moreover, the single factor identified by CAPM, namely the returns on the market portfolio, is unobservable, which results in some proxies for the market portfolio being employed in empirical applications (see Roll's (1977) critique of CAPM) . This, however, raises doubts on the empirical validity of the CFP regularity itself. Is CFP a true or a spurious empirical regularity? This point is emphasized in (iii) below.

(iii) Failure to identify the true causal risk factors makes the empirical truth of the CFP regularity suspect. Indeed, if M_t in (12) ($M_{1,t}$, $M_{2,t}, ..., M_{k,t}$ in (13)) is (are) not the true causal factor(s), then the empirical observation that β_i in (12) ($\beta_{1,i}$, $\beta_{2,i}, ..., \beta_{k,i}$ in (13)) is (are) time-varying and persistent, may simply be wrong. Put it differently, the time variation in betas may not be a genuine empirical regularity, but rather an accidental one, resulting from model misspecification. In such a case, the D-S explanations of $\mathcal{S}_R$, based on SFM-AR, obtained in Section 3, are *potential* rather than *actual* explanations.

(iv) Related to (iii), any future attempts towards unravelling the identity of risk factors must ensure that the purported factors are genuine, rather than spurious common causes. This is intimately related to the issue of temporal priority of causes over their effects.

The following analysis assumes that the true causal factors are indeed observable and examines alternative, non-theoretical, routes for identifying them. More specifically, in the absence of a well-developed causal theory that reveals the identity of the risk factors in MFM, (under the assumption that these factors are identifiable), the question which naturally arises is whether these factors can be identified empirically. Put differently, if theoretical reasoning alone has failed to achieve this task, then is it possible to appeal to empirical procedures instead? However, before we turn our attention to this question, we must first answer the following: Why is the identification of the true causal risk factors so important for the explanation of the regularities in $\mathcal{S}_R$? After all, the theoretical results of Section 3 have shown that SFM-AR offers D-S explanations of all the regularities in $\mathcal{S}_R$, without putting any conditions either to the identity or even to the probabilistic properties of the risk factors. Indeed, SFM-AR explains all the regularities in $\mathcal{S}_R$ mainly by means of the autoregressive nature of beta, regardless of which factor is associated with this beta. Therefore, one is justified to ask, why is the issue of factor identification deemed as fundamental for the adequacy of either D-S or C-M explanations of $\mathcal{S}_R$?

The answer to this question consists of the following two parts:

(i) As mentioned above, the presence of autoregressive betas *alone* can indeed offer D-S explanations for all the elements of $\mathcal{S}_R$, conditional on the fact that betas are indeed autoregressive. However, establishing the autoregressive nature of betas, to begin with, requires the employment of the true causal risk factors in regression models such as (12) or (13). As already mentioned, β_i in (12) may be autoregressive with respect to a spurious non-causal factor M_t, but time-invariant with respect to the true causal factor M_t'.

(ii) Apart from D-S explanations, the risk-factor identification is a necessary condition for explanatory adequacy for C-M explanations as well. The necessity of factor identification is a manifestation of the philosophical thesis that "Explanation consists in identifying causal relations" (Kitcher 1985, pp. 638). Let us analyze this thesis in the context of the following case: Imagine for a moment, that the variables $R_{i,t}$ and M_t in (12) represented the amount of rainfall in a particular region over a specific period of time, t, and the readings (just before t) of a well-functioning barometer, respectively. Moreover, assume that we could use this statistical model to deduce some probabilistic properties of $R_{i,t}$ that correspond to some empirical regularities, $\mathcal{S}_R$, of $R_{i,t}$. According to the D-S model, we have obtained perfectly adequate D-S explanations of $\mathcal{S}_R$. Moroever, apart from regularities, this model could be used for explaining or predicting the occurence of a singular event, for example the amount of rainfall occured in a specific point in time t_0 (see, Hempel 1965, for the so-called Inductive-Statistical explanations of single events and the Explanation-Prediction symmetry thesis). According to the C-M model however, the empirical law represented by (12) cannot form the basis for an adequate explanation of the empirical regularities in $\mathcal{S}_R$. This is due to the fact that these regularities are explained in terms of a *symptomatic* factor, such as the barometer readings, instead of a true causal factor, such as, for example, the atmospheric conditions in that region.To repeat Salmon's aphorism, causal irrelevancies might be harmless for predictions but fatal for explanations.

The preceding discussion must have made clear why the identification of the true causal risk factors is essential for the adequacy of both D-S and C-M explanations of the empirical regularities of stock returns. It is now time to turn our attention to examining the possibility of identifying the causal risk factors empirically. It must be noted at the outset that this possibility, even if successfully implemented, signifies a departure from the strict ideal standards set by the D-N-P model of explanation. However, knowing the causes of a specific effect is explanatory useful even if we are unable to provide all the causal links and interactions leading up to that effect. For example, it is explanatory relevant to have established that "smoking" probabilistically causes "lung cancer", even if we cannot describe in full detail all the biochemical reactions in which this process is completed. Kitcher (1985) remarks on the possibility of having diffferent levels of explanation as follows: "There are degrees of scientific understanding, and, equally, degrees of *un*understanding. An explanation is ideally complete if it eliminates all our ununderstanding. An account may advance our understanding without being ideally complete. We may have to settle for less." (1985, pp. 633) Indeed, the possibility that, quite often, we have to settle for less was already ackowledged by Railton, who admits that obtaining an ideal explaantory text is often impossible. He asks this question: "Is it preposterous to suggest that any such ideal could exist for scientific explanation and understanding?

Has anyone ever attempted or even wanted to construct an ideal causal or probabilistic text?" (1981, pp 246-247). Railton answers his own question as follows: "It is not preposterous if we recognize that the actual ideal is not to *produce* such texts, but to have the ability (in principle) to produce arbitrary parts of them." (1981, pp 246-247). Put differently, the absence of the full text does not imply complete lack of understanding of the observed regularities. As Psillos (2002) puts it, the ideal D-N-P text "is more of a regulative ideal than what, in practice, we need and should strive for. In practice, what we (or the scientists) need and should strive for is "explanatory information" relevant to the explanandum. Such information, if indeed it is information relevant to the explanandum, will be part of the ideal D-N-P text. By producing such parts, no matter how underdeveloped and incomplete they may be, scientists understand why a certain explanandum happens. Finding more and more bits of the ideal texts, we move closer to the ideal of a full understanding" (2002, pp. 260). Railton (1981) himself refers to the ideal D-N-P text as "a yardstick for proffered explanations of chance phenomena" and also allows for these proffered explantions to take various forms and "still be successful in virtue of communicating information about the relevant ideal text" (1981, pp. 246-247).

In the case under study, MFM which involves only the true causal risk factors with the latter being empirically identified, may be thought of as an intermediate case between a statistical model of purely inductive origins and one which is produced from a theoretical account of the chance mechanism at work. As such, the explanatory status of this model may be thought of as lying between the simple D-S and the ideal D-N-P ones.

Given the discussion above, can we identify the true causal risk factors empirically? The answer to this question depends to a large extent on whether true causal relationships can be reduced to empirical correlations. One of the first attempts towards this direction is Salmon's (1971, 1984) Statistical Relevance (S-R) model of explanation, which is based on the "screening-off" relations and conjuctive forks, introduced in the previous Section. In the context of S-R an explanation of an event is no longer an argument (as in the I-S case) but rather "an assemblage of factors relevant to the occurrence or nonoccurence of the event to be explained, along with the associated probability values" *(Causal and theoretical explanation blue book pp.108).* This in turn implies that an S-R explantion of an event E has been achieved when the probability of E conditional on all the relevant factors for E has been obtained, regardless of whether this probability is high or low. This probability is interpreted as the relative frequency of E within the "homogeneous" reference class, that is the reference class which is generated by partitioning the initial class solely by relevant conditions. Partitioning by irrelevant conditions/factors, that is factors that are screened-off by relevant ones, is prohibited in the context of S-R model. As Salmon *(???- athird dogma...pp.98)* puts it: "On this model, high probability is not the desideratum; rather the amount of relevant information is what counts. According to the S-R model, *a statistical explanation consists of a probability distribution over a homogeneous partition of an initial reference class.* A homogeneous partition is one that does not admit of further relevant subdivision. The subclasses in the partition must also be maximal - that is, the partition must not involve any irrelevant subdivisions" (emphasis added). This is because although irrelevant partitions are "harmless to arguments" such as I-S they are "fatal to explanations" such as S-R. With the term "maximal" Salmon means that the reference class is the broadest possible one in the sense that the initial class has not been erroneously narrowed by irrelevant factors.

The preceding discussion focuses on explanations of single events or *local* explanations. In the context of SFM, which is stated in terms of random variables, explanation refers to types of events (those represented by the corresponding random variables), which means that explanation is *global.* Niiniluoto (1982) draws a distinction between these two modes of explanation as follows: "Explanation may be local or global with respect to the explanandum depending upon whether the explanandum is a particular statement or a whole class of statements" (1982, pp. 171). An S-R global explanation of empirical regularities in $\mathcal{S}_R$ requires the identification of the risk factors (and only those) which are statistically relevant for $R_{i,t}$. Initially, Salmon (1971) held the view that the set of all the statistically relevant factors coincides with the set of all the genuinely causal factors. Later he changed his view arguing that statistical relevance relations alone, are not sufficient to unravel causal relevance relations. Instead, they merely constitute evidence of causal relations. In his 1984 book Salmon admits: "It seemed obvious at the time that statistical relevance relations had some sort of explanatory power in themselves. As I have said repeatedly throughout this book, that view now appears to be utterly mistaken...Their fundamental import lies in the fact ...that they constitute evidence for causal relations." (1984, pp. 191-192). One of the main reasons for the inability of statistical relations to identify causal ones is that there is a certain type of factors, which function as screeners-off without being causal. This case will be analyzed in detail in the next subsection.

The preceding analysis suggests that the reduction of causal relations to statistical relevance relations is extremely difficult (if not impossible) to be achieved without any extra-statistical information. The latter may come from accepting the aforementioned principle of "temporal priority" asserting that "causes precede their effects in time". In view of the significance of the issue of temporal priority in the empirical identification of risk factors for asset returns, we analyze it in more detail below.

4.1 The Issue of Temporal Priority of Causes over their Effects

In his analysis of probabilistic causality, Suppes (1970) assumes explicitly that "a cause precedes its effect in time" (1970, pp. 11). As a result, the timing of occurrence of the various events "are included in the formal characterization of the probability space" (1970, pp. 12). Suppes defines the event $B_{t'}$ to be a *prima facie cause* (or a prima facie *positive* cause) of the event A_t if and only if

$$\begin{aligned} (i)\ t' &< t \\ (ii)\ P(B_{t'}) &> 0 \\ (iii)\ P(A_t &\mid B_{t'}) > P(A_t) \end{aligned} \tag{21}$$

An event which satisfies the conditions (21) that is an event which is a prima facie cause of A_t can still prove to be a spurious cause of A_t. More specifically, $B_{t'}$ is a spurious cause of A_t if and only if there is a t'' such that

$$t'' < t' < t \tag{22}$$

and an event $C_{t''}$ such that $P(B_{t'} \cap C_{t''}) > 0$ and

$$P(A_t \mid B_{t'} \cap C_{t''}) = P(A_t \mid C_{t''}) \tag{23}$$

In Salmon's terminology, Suppes's condition (23) states that $B_{t'}$ is a spurious cause of A_t if it is screened-off by an earlier event $C_{t''}$. In Suppes's own terms, a prima facie cause becomes spurious "if there exists an earlier event that eliminates the effectiveness of the cause when that event occurs" (1970, pp. 25).

Condition (23) can be employed to characterise an event $B_{t'}$ as a non-spurious or genuine cause of A_t : $B_{t'}$ is a (prima facie genuine) cause of A_t if (21) holds and there is no further event $C_{t''}$ with $t'' \leq t'$such that (23) holds. Another useful definition is that of "direct cause" (see Definition 5, pp. 28). More specifically, if there is no event $C_{t''}$ with

$$t' < t'' < t \tag{24}$$

satisfying the probabilistic relationship (23), then $B_{t'}$ is called a direct cause of A_t. In other words, $B_{t'}$ is a direct cause of A_t if there is no causal chain of the form $B_{t'} \longrightarrow C_{t''} \longrightarrow A_t$ satisfying the causal Markov property.

The preceding analysis implies that the probabilistic relationship (23) may have quite different implications for the causal status of $B_{t'}$ for A_t depending on the time of occurence of the third event $C_{t''}$. In particular, if $C_{t''}$ occured at a time earlier than $B_{t'}$ then $B_{t'}$ is a spurious cause of A_t. If however, $C_{t''}$ occured later than $B_{t'}$ then (23) is consistent with the case in which $B_{t'}$ is an indirect cause of A_t, that is it causes A_t through its direct causing of $C_{t''}$. In this case, the screening-off of $B_{t'}$ from the intermediate event $C_{t''}$ is the result of the Markov property according to which the information of the intermediate event $C_{t''}$ renders information about the earlier event $B_{t'}$ irrelevant to the probability of the later event A_t. The moral from this analysis is that (unfortunately for the probabilistic theories of causality) there are certain types of events, referred to as "intermediate causes" which exhibit screening-off properties without being genuine (common) causes.

Let us summarise the aforementioned probabilistic causal relationships as follows:

(a) The conditions (22), (21) and (23) imply the following causal structure, C1:

$$\begin{array}{ccc} & & A_t \\ & \nearrow & \\ C_{t''} & & \\ & \searrow & \\ & & B_{t'} \end{array}$$

(b) The conditions (24), (21) and (23) imply the following causal structure, C2:

$$B_{t'} \longrightarrow C_{t''} \longrightarrow A_t.$$

The preceding analysis suggests that any theory that aims at revealing the identity of the genuine, causal, common factors of returns must explicitly take into account the times of occurence of events, represented by these factors.

4.2 Identification of Risk Factors in Practice

As mentioned above, the task of empirically identifying the exact set of true causal risk factors in MFM faces serious difficulties, stemming from the problem of reducing causal relationships to statistical correlations. How can we be certain that the employed risk factors $X_{1,t}, X_{2,t}, ..., X_{k,t}$ are truly causal? Put differently how can we be certain that some or all of these factors are not symptomatic? One source of difficulty is the fact that the true causal risk factors may be unobservable. Various solutions to the problem of "factor specification" have been suggested in the literature including solutions of purely statistical nature such as principal component analysis (see Chamberlain and Rothschild 1983, Connor and Korajczyk 1985, 1986). Another popular solution amounts to approximationg the risk factors by the so-called "mimicking portfolios". A well-known example in this direction is the approach of Fama and French (1993) in which characteristics such as the firm size or the book-to-market ratio are supposed to capture the effects of some unobservable risk factors (see also, Rosenberg, Reid and Lanstein 1984, Chan, Hamao and Lakonishok 1991). An immediate implication of such an approach is that the mimicking portfolios are by definition symptomatic factors in the sense that if the true causal factors had been included, the mimicking portfolios would become statistically irrelevant. In other words, the true factors would screen off the mimicking portfolios. This in turn raises questions on the explanatory power, in terms of statistical relevance of the factor models that employ mimicking portfolios as factors.

To overcome this problem, other approaches assume that the set of the causal factors in question is a subset of the set of all possible observable macroeconomic variables. Chen Roll and Ross (1986), follow this approach and identify the following risk factors for stock returns: unanticipated changes in inflation, unanticipated changes in GDP, unanticipated changes in the default premium of corporate bonds, and unanticipated shifts in the yield curve. A necessary condition for accepting these four factors as the causal ones is that any other possible macroeconomic variable is screened-off by these four factors. The screening-off procedure is implemented in two steps: In the first step m time series regressions are run with $R_{i,t}$ $i = 1, 2, ..., n$ being the dependent variable and a set of potential factors $\mathcal{D}$ being the independent variables. Assume that this step is finished when a subset $\mathcal{D}^s$ of the initial set of candidate variables is found to be statistically significant for all the m time series regression under study. In the second step, we decide which of the variables in $\mathcal{D}^s$ explain the cross-sectional variation in returns, that is which variables are actually priced by the market. Assume we end up with $\mathcal{D}^c$ which contain only the priced variables. The variables in $\mathcal{D}^s - \mathcal{D}^c$ are screened-off in the second step of the procedure. The final set

$\mathcal{D}^c$ of the surviving variables in both steps of the screening-off procedure are the causally relevant factors. A rather surprising result emerging from the study of Chen Roll and Ross (1986) is that "well established" factors such as the value-weighted New York Stock Exchange Index turn out to be symptomatic since they are screened off in the second step by apparently, more relevant macroeconomic factors.

However, the aforementioned procedure is problematic because it does not account for the exact time of occurence of the relevant events. Let us analyze this in more detail. In practical applications, the estimation of (13) is usually carried out by employing data observed in one of the usual frequencies (for example monthly). As already mentioned, the common time subscript in (13) means that the factors $X_{j,t}$, $j = 1, 2, ..., k$ and the dependent variable $R_{i,t}$ are simultaneously observed which in turn implies that the corresponding events represented by these variables occur simultaneously. However, in reality this is not the case. Some of these factors may be observed early in the month, some others in the mid of the month and the rest towards the end of the month. In view of the discussion in the previous section, this differentiation in the actual time of observation raises some important issues about the causal status of each of the factor for $R_{i,t}$. For example assume that the values $x_{1,t}$ of $X_{1,t}$ are usually observed on 5 January whereas the values $x_{2,t}$ of $X_{2,t}$ are usually observed on 20 January. The values y_t of $R_{i,t}$ are observed on 31 January. In the econometrician's dataset, however, these differences are wiped out and $x_{1,t}$, $x_{2,t}$ and y_t are treated as the January observations of $R_{i,t}$, $X_{1,t}$ and $X_{2,t}$, respectively. This in turn implies that the event $B_{1t} = \{X_{1,t} = x_{1t}\}$ occurs at a time earlier than $B_{2t} = \{X_{2t} = x_{2t}\}$ and that both B_{1t} and B_{2t} occur earlier than $A_t = \{R_{i,t} = y_t\}$. Assume that the estimation results from a linear regression of $R_{i,t}$ on $X_{1,t}$ and $X_{2,t}$ suggest that $X_{1,t}$ is insignificant at the usual significance levels, whereas $X_{2,t}$ is not. At the same time, a regression of $R_{i,t}$ on $X_{1,t}$ alone, suggests that $X_{1,t}$ is a significant "explanatory" factor for $R_{i,t}$. How does the combination of the results from the two regression ought to be interpreted? If the aforementioned differences in the actual time of observations are ignored, and $X_{1,t}$ and $X_{2,t}$ are assumed to be simultaneously observed, then the results imply that $X_{1,t}$ is a spurious causal factor of $R_{i,t}$ being screened-off by $X_{2,t}$. On the other hand, if the fact that $X_{1,t}$ is observed prior to $X_{2,t}$ is taken into account, then the insignificance of $X_{1,t}$ does not imply that $X_{1,t}$ is a spurious causal factor of $R_{i,t}$. On the contrary, $X_{1,t}$ may well be an indirect causal factor of $R_{i,t}$, causing $R_{i,t}$ through its effect on $X_{2,t}$. For this to be the case, a Markov condition in the chain $B_{1,t} \longrightarrow B_{2,t} \longrightarrow A_t$ must hold for all possible events $B_{1,t}$, $B_{2,t}$, A_t that can be defined in terms of the random variables $X_{1,t}$, $X_{2,t}$ and $R_{i,t}$, respectively. In this case, we end up having erroroneously identified $X_{2,t}$, instead of $X_{1,t}$ as a causal factor of $R_{i,t}$. If a necessary condition for the explanation of a regularity is to identify all the genuine causal factors of it and only those, then the aforementioned statistical procedure has clearly failed to satisfy this condition.

4.3 From SFM to SFM-AR

The preceding analysis has shown that SFM, as opposed to MD-GARCH, enjoys some degree of theoretical justification, in terms of theoretical properties of the mechanism at work such as "operation of a common cause", "efficient use of available information" and "forward looking behaviour". Moreover, SFM gives a broad outline of the functioning of this mechanism in the state of equilibrium without touching upon the issue of how this equilibrium state is achieved. As already mentioned, the transition from SFM to SFM-AR was motivated by purely empirical considerations. This in turn generates the question of whether there is any "causal" theory, complementary to the basic theory underlying SFM, that accounts for the time-variation and persistence in betas. It is important to note at the outser that the time variation of betas does not (necessarily) alter the "equilibrium" status of the factor models. Indeed, as was shown in Section 3, under some parameter restrictions, SFM-AR represents a second-order stationary returns process.

To this end, Berk, Green and Naik (1999) suggest a theoretical model which implies that a firm's systematic risk and expected returns change through time in a predictable way as a result of temporal variations in firm's growth and investment opportunities. More specifically, this model illustrates how the stochastic behaviour of systematic risk is driven by firm's value maximizing choices, with the latter exhibiting some degree of persistence. In a similar vein, Avramov and Chordia (2006) attribute some of the well-known anomalies of the empirical literature, such as the size and book-to-market effects to the persistent behavior of betas (see also Petkova and Zhang, 2005, Ang and Chen, 2007, and Zhang, 2005). Although, these studies do not satisfy the (rather impossible) task of providing the full account of the causal mechanism at work that produces $SFM - AR$, nonetheless they enhance our understanding of the possible origins of the persistent variation in sytematic risk. In doing this, they convey relevant information for the explanandum.

5 Conclusions

This paper examined the sense in which statistical models of stock returns explain empirical regularities. It was argued that there are alternative definitions of "statistical explanation", developed in the philosophy of science literature, according to which a statistical model can be classified as "explanatory" rather than"descriptive". It was shown that a statistical model may be called "explanatory" even if its birth was motivated purely by empirical considerations. Moreover, a statistical model can be explanatory in more than one ways, or it may be explanatory in one sense, but non-explanatory in another. This paper focused on two distinct models of probabilistic explanation, namely D-S and D-N-P, with the latter being a prominent member of the class of C-M models of explanation.

In the context of D-S, an explanation of an empirical regularity, S_i, by a statistical model M_1 is achieved, when S_i is deduced from M_1. D-S poses no restrictions on the origins of M_1 other than M_1 is not born out of the probabilistic interpretation of S_i itself (in which case, M_1 explains S_i trivially). On the other hand, D-N-P sets the much stricter condition of M_1 being derived from a theoretical account of the chance mechanism at work.

The analysis in this paper showed that the M-MD-GARCH model satisfies the D-S criteria, since it entails the empirical regularities CFP, EL, EAG, and EAI. SFM-AR also satisfies the D-S criteria with respect to CC, FMR, EL, EAG and EAI. Note that M-MD-GARCH and SFM-AR imply (CC, FMR) and CFP, respectively in a trivial way, since (CC, FMR) and CFP were the "generating regularities" of M-MD-GARCH and SFM-AR, respectively.

The explanatory status of SFM-AR and particularly M-MD-GARCH changes drastically, with respect to the C-M criteria set by the D-N-P model. There is no doubt that the origins of M-MD-GARCH are purely empirical. This model is a direct descendent of the univariate MD-GARCH model which, according to the testimony of its initiator, emerged from a well-designed mispecification test for dynamic heteroskedasticity. There was no theory of how the internal workings of the stock market mechanism produce volatility clustring, which is the motivating regularity of this model.

On the other hand, SFM-AR seems to have much stronger theoretical origins. Of course, the existing theory does not describe in detail the causal nexus of the market mechanism that brings about the observed behaviour of stock returns. Rather, it gives an outline of the general structure of this mechanism, operating in a state of equilibrium, which is based on the principles of "common cause", "efficient information processing" and "forward looking behaviour". The existing theory, however, is not capable of identifying the identity of the risk factors that appear in SFM-AR. This task is usually carried out empirically, using standard statistical methods. It was argued that the empirical identification of the true causal risk factors is at best questionable due to the difficulty of reducing probabilistic causal relationships to statistical correlations. However, even if the empirical identification of the causal factors were succesful, the explanation of the stock returns regularities in terms of these (true) factors would still not satisfy the D-N-P ideal. Indeed, in a case like this, we would have obtained an accurate description of the "outter" behaviour of the chance mechanism at work without, however, having gained any insight into the internal workings of the mechanism. To achieve the latter task, more theoretical work aiming at identifying the causal risk factors from first principles seems to be necessary if an "ideal" causal mechanistic explanation of the empirical regularities of stock returns is to be achieved.

1. Ellis 1956

2. Brewer and Lambert 200?

3. Box and unter 1965

4. Box and Draper 1987

5. Lehmann 1990

6. Hempel and Oppenheim 1948

7. Hempel 1962

8. Hempel 1965

9. Railton 1978

10. Railton 1981

11. Mandelbot 1963

12. Engel 1982, 2002, 2003

13. Bollerslev 1987

14. Diebold 1988

15. Drost and Nijman 1993

16. Meddahi and Renault 2004

17. Roll and Ross 1980

18. Christodoulakis 2001

19. Sharpe 1964

20. Blume 1971

21. Blume 1976

22. Fabozzi and Francis 1977

23. Fisher and Kamin 1985

24. Sunder 1980

25. Bos and Newbold 1984

26. Collins Ledolter and Rayburn 1987

27. Andersen et al. 2005

28. Jostova and Philipor 2005

29. Peligrad and Utev 2005
30. Bollerslev, Engle and Wooldridge 1988
31. Hempel 1966
32. Ross 1978
33. Kinoshite 1990
34. Friedman 1974
35. Kuhn 1962
36. Levy 1925
37. Engle 1982
38. Bollerslev 1986, 1990
39. Diebold 1988
40. Nelson 1990
41. Kourogenis and Pittis 2008
42. Bradley 1988
43. Peligrad 1990
44. Zhang and Lin 2012
45. Bauewens, Laurent & Rombouts 2006
46. Salmon 1971, 1984, 1989
47. Railton 1978, 1981
48. Glennan 2002
49. Cohnitz ???
50. Wold 1969
51. Roll and Ross 1980
52. Reichenbach 1946 ?
53. Braithwaite 1964
54. Litner 1965
55. Rose 1978
56. Hoover 1993

57. Kyburg 1978

58. Fitelson, Hajek and Hall 2006

59. Kitcher 1985

60. Psillos 2002

61. Railton 1981

62. Niiniluoto 1982

63. Suppes 1970

64. Chamberlain and Rothschild 1983

65. Connor and Korajczyk 1985, 1986

66. Fama and French 1993

67. Rosenberg, Reid and Lanstein 1984

68. Chan, Hamao and Lakonishok 1991

69. Chen, Roll and Ross 1986

70. Berk, Green and Naik 1999

71. Avramov and Chordia 2006

72. Petkova and Zhang 2005

73. Ang and Chen 2007

74. Zhang 2005

APPENDIX

Proof of Theorem 1:

(i) We have

$$\begin{aligned} Var_{t-1}(\mathbf{R}_t) &= E_{t-1}\left[(M_t(\boldsymbol{\beta}+\boldsymbol{\beta}_t)+\mathbf{u}_t)(M_t(\boldsymbol{\beta}+\boldsymbol{\beta}_t)+\mathbf{u}_t)'\right] \\ &= \sigma_m^2 E_{t-1}\left[(\boldsymbol{\beta}+\boldsymbol{\beta}_t)(\boldsymbol{\beta}+\boldsymbol{\beta}_t)'\right]+\Sigma_u \\ &= \Sigma_u+\sigma_m^2\left(\left(\boldsymbol{\beta}+\boldsymbol{\Phi}\boldsymbol{\beta}_{t-1}\right)\left(\boldsymbol{\beta}+\boldsymbol{\Phi}\boldsymbol{\beta}_{t-1}\right)'+\Sigma_\varepsilon\right) \\ &= \sigma_m^2\Sigma_\varepsilon+\Sigma_u+\sigma_m^2\left(\boldsymbol{\beta}+\boldsymbol{\Phi}\boldsymbol{\beta}_{t-1}\right)\left(\boldsymbol{\beta}+\boldsymbol{\Phi}\boldsymbol{\beta}_{t-1}\right)' \qquad (25) \end{aligned}$$

(ii) First note that from the independence between $\mathbf{u}_t$, w_t and $\boldsymbol{\varepsilon}_t$, postulated in assumption **M**, conditional on the realization of $\boldsymbol{\beta}_t$ and all the information that is generated up to time $t-1$, we have that $E_{\boldsymbol{\beta}_t}[\mathbf{R}_t] = \mathbf{a}$ and $Var_{\boldsymbol{\beta}_t}[\mathbf{R}_t] = \Sigma_u + \sigma_m^2(\boldsymbol{\beta}+\boldsymbol{\beta}_t)(\boldsymbol{\beta}+\boldsymbol{\beta}_t)'$. The result then follows from the normality of the random vector.

(iii) First note that

$$\begin{aligned}(Var(R_{i,t}))^2 &= \left(\left(\beta_i^2+\sigma_{\beta_i}^2\right)\sigma_m^2+\sigma_{u_i}^2\right)^2 \\ &= \sigma_{u_i}^4+\beta_i^4\sigma_m^4+\sigma_m^4\sigma_{\beta_i}^4+2\left(\beta_i^2\sigma_m^2\sigma_{u_i}^2+\beta_i^2\sigma_m^4\sigma_{\beta_i}^2+\sigma_m^2\sigma_{u_i}^2\sigma_{\beta_i}^2\right) \quad (26)\end{aligned}$$

Moreover, for the fourth central moment of $R_{i,t}$ we have

$$\begin{aligned}E\left[(R_{i,t}-E[R_{i,t}])^4\right] &= E\left[\left(\beta_i M_t+\beta_{i,t}M_t+u_{i,t}\right)^4\right] \\ &= E\left[u_{i,t}^4+M_t^4\beta_i^4+M_t^4\beta_{i,t}^4+6u_{i,t}^2M_t^2\beta_i^2+6u_{i,t}^2M_t^2\beta_{i,t}^2+6M_t^4\beta_i^2\beta_{i,t}^2\right] \\ &= 3\left[\left(\sigma_{u_i}^4+\beta_i^4\sigma_m^4+\sigma_{\beta_i}^4\sigma_m^4\right)+2\left(\beta_i^2\sigma_{u_i}^2\sigma_m^2+\sigma_{\beta_i}^2\sigma_{u_i}^2\sigma_m^2+3\beta_i^2\sigma_{\beta_i}^2\sigma_m^4\right)\right] \\ &= 3\left(Var(R_{i,t})\right)^2+12\beta_i^2\sigma_{\beta_i}^2\sigma_m^4\ , \quad (27)\end{aligned}$$

where we have used the fact that for the Gaussian distributions, the third moment is zero and fourth moment equals to three times the square of the second. Hence, the kurtosis coefficient of the unconditional distribution of stock returns is given by

$$Kurt(R_{i,t}) = \frac{E\left[(R_{i,t}-E[R_{i,t}])^4\right]}{(Var(R_{i,t}))^2} = 3+\frac{12\beta_i^2\sigma_{\beta_i}^2\sigma_m^4}{(Var(R_{i,t}))^2}. \quad (28)$$

(iv) By virtue of (3) and Assumption M, we have

$$E_{t-1}[\mathbf{R}_t-\mathbf{a}] = E_{t-1}[M_t(\boldsymbol{\beta}+\boldsymbol{\beta}_t)+\mathbf{u}_t] = E_{t-1}[M_t]\,E_{t-1}[\boldsymbol{\beta}+\boldsymbol{\beta}_t] = 0$$

Proof of Theorem 2:

For the proof we use the relatively recent invariance principle of Peligrad and Utev (2005), stated below:

Theorem PU (Invariance Principle of Peligrad and Utev (2005)): *Let* $\{X_i\}_{i\in\mathbb{Z}}$ *be a stationary sequence with* $E[X_0]=0$ *and* $E[X_0^2]<\infty$*. Assume that*

$$\sum_{n=1}^{\infty}\frac{\|E[S_n|\mathcal{F}_0]\|_2}{n^{3/2}}<\infty\ . \tag{29}$$

Then, $\left\{\max_{1\le k\le n} S_k^2/n\right\}_{n\ge 1}$ *is uniformly integrable and* $n^{-1/2}W_n \xrightarrow{D} \sqrt{\eta}W$*, where* η *is a non-negative random variable with finite mean* $E[\eta]=\sigma^2$ *and independent of* $\{W(t)\}_{t\ge 0}$*. Moreover,* η *is determined by the limit* $\lim_{n\to\infty}\left(E[S_n^2|\mathcal{I}]/n\right)=\eta$ in L_1 *, where* $\mathcal{I}$ *is the invariant sigma field. In particular,* $\lim_{n\to\infty}\left(E[S_n^2]/n\right)=\sigma^2$.

Assumption **M** implies that

$$E\left[R_{t_0+i}-E\left[R_{t_0+i}\right]|\mathcal{F}_{t_0}\right]=0 \text{ a.e.}$$

Therefore $\|E[S_{t_0,k}|\mathcal{F}_{t_0}]\|_2=0$ and condition (29) is trivially satisfied. By virtue of the existence of finite second moments for all random variables involved, we can apply Theorem PU. From the joint normality of β_t, M_t and u_t, we have that $\{R_t\}_{t\in\mathbb{Z}}$ is ergodic, hence the invariant σ-field is trivial. Applying, now, (9) we obtain

$$\eta \;=\; \lim_{n\to\infty}\frac{E[S_k^2]}{k}=\lim_{k\to\infty}\frac{Var\left(R_\tau(k)\right)}{k}=\beta^2\sigma_m^2+\sigma_u^2+\sigma_\beta^2\sigma_m^2\in\mathbb{R}.$$

QED