POLITECNICO DI MILANO

FACOLTÁ DI INGEGNERIA DEI SISTEMI

Ingegneria Matematica

Orientamento in Finanza Quantitativa

Stochastic Orders in ARCH and GARCH Models

Relatore: Prof. Carlo Sgarra

Tesi di Laurea di:
Lorenzo Dall'Oca
Matr. 799508

Anno Accademico 2013 - 2014

Sommario

In questa tesi ho analizzato come nei modelli ARCH e GARCH con leverage gli ordinamenti stocastici si propagano dai parametri e dalle innovazioni ai valori dei logreturns, alle loro somme e ai prezzi delle opzioni scritte sullo stock modellizzato dal modello stesso.
Nel primo capitolo ho posto le basi teoriche per quanto riguarda gli ordinamenti stocastici, soffermandomi in particolare sull'ordinamento usuale, su quello convesso e su quello crescente convesso. Dopo averli definiti, ho fornito diverse interpretazioni sugli ordinamenti ed evidenziato condizioni necessarie o sufficienti affinché questi valgano. Infine ho riportato legami intercorrenti tra gli ordinamenti, mostrando se e quando uno implica l'altro.
Nel secondo capitolo, finalizzato all'introduzione del modello GARCH, ho riportato le definizioni piú importanti sull'analisi delle serie storiche, i primi modelli autoregressivi e a media mobile come quelli ARMA, ed infine ho introdotto i modelli ARCH e GARCH con le loro generalizzazioni tra cui quella del GARCH con leverage.
Nel terzo capitolo confluiscono i concetti principali dei due precedenti: infatti, passo dopo passo, mostro, grazie alla teoria del primo capitolo, come gli ordinamenti dei parametri e delle innovazioni si propaghino ai valori della volatilitá, da questi ai logreturns e, nel caso di innovazioni simmetriche, persino alle loro somme. Ho svolto le precedenti analisi inizialmente nel caso di un generico modello GARCH, prendendo spunto dall'articolo "Comparison Results for GARCH processes" di Bellini F., Pellerey F, Sgarra C. e Sekeh S.Y. [3]. Successivamente, seguendo la metodologia esposta nel sopracitato articolo, ho mostrato come gli stessi risultati si applichino anche al modello ARCH a piú passi. Infine ho provato a dimostrare la propagazione degli ordinamenti dei parametri in riferimento al GARCH con leverage. In questo ultimo caso, nonostante gli ordinamenti si diffondano ancora ai logreturns, non sono riuscito ad ottenere risultati significativi per quanto riguarda le somme dei logreturns. Mostro quindi nel dettaglio come l'introduzione di una non linearitá nel modello modifichi i risultati ottenuti nel caso generico.
Infine nel quarto capitolo verifico numericamente i risultati ottenuti in precedenza mostrando come l'ordinamento usuale stocastico su uno o tutti i parametri induca un ordinamento stocastico usuale sui moduli e sui quadrati dei logreturns, e un ordinamento convesso sui logreturns stessi e, in alcuni casi con innovazioni simmetriche, sulle loro somme.
Inoltre é possibile ottenere il valore di un possibile sottostante di opzioni finanziarie come il valore iniziale moltiplicato per l'esponenziale delle somme

dei logreturns. Chiaramente la funzione appena descritta é convessa e quindi anche i sottostanti stessi saranno ordinati in modo convesso. Ne segue che per payoff convessi otteniamo un ordinamento preciso sui prezzi delle opzioni legato all'ordinamento dei parametri del modello GARCH.
Questa tesi si colloca nel sempre piú ampio e approfondito campo delle applicazioni finanziarie degli ordinamenti stocastici. Nel mio caso specifico questi ultimi mi permettono di avere delle informazioni su determinate proprietá di ordinamento dei valori della serie simulata, sfruttando informazioni sui parametri iniziali.
Il caso che ho trattato é chiaramente discreto e non markoviano. Esiste parallelamente un'ampia letteratura su argomenti simili ma relativi a modelli continui e markoviani.
Ad esempio nell'articolo "Comparison of Option Prices in Semimartingale Models" di Bergenthum J. e Rüschendorf L. [4], vengono mostrate condizioni sufficienti sotto le quali sia possibile confrontare i prezzi di opzioni europee, su sottostanti differenti, calcolate rispetto a misure di martingale equivalenti.
Un altro ampio campo di applicazione, sempre nell'ambito dei processi markoviani, é quello relativo all'ordinamento di misure di martingala equivalente in semimartingale market models sotto le quali calcolare i prezzi delle opzioni.
Ad esempio nell'articolo "Bounds on Option Prices for Semimartingale Market Models" di Gushchin A.A. e Mordecki E. [10], si mostra come ricavare dei range sui prezzi delle opzioni con payoff convesso grazie ad un ordinamento indotto sulle misure coerenti. Chiaramente da questi range sará poi possibile individuare dei bounds sui prezzi al variare delle misure di martingala equivalenti.
Nell'articolo "Stochastic Orders and Risk Measures: Consistency and Bounds" di Baüerle N. e Müller A. [1], é invece sottolineato come, sotto alcune ipotesi, le misure di rischio siano consistenti con gli ordinamenti, in riferimento a quello usuale e a quello convesso. Da questa osservazione viene poi mostrato come ricavare dei bounds sulle misure di rischio relative ad un portafoglio.

Abstract

In this thesis i analyze how, in the ARCH and GARCH with leverage models, stochastic orders are propagated from parameters and innovations to logreturns, their sum and to the prices of options having as underlying the stock described respectively by ARCH and GARCH models.
In the first chapter i give some fundamental tools concerning stochastic order, giving particular attention to usual stochastic order, to the convex order and to the increasing convex order. In addiction to their definition, i give several interpretation of these orders and some necessary or sufficient condition to verify them. Finally i underline how these orders are related each other, showing if and when one imply another.
In the second chapter, aimed to introduce GARCH model, i report the most important definitions about time series analysis, then i describe the first "auto-regressive" and "moving average" models such ARMA models, and introduce ARCH and GARCH models, with their generalizations such as GARCH with leverage.
In the third chapter we find the main concepts of the two previous chapter: indeed, step by step, i show, thanks to the theory of the first chapter, how orders of parameters and innovations do transfer to volatility's values, from here to logreturns and, with symmetric innovations, to their sums. In a first place i did these analysis in the generic case of GARCH model taking the cue from the article "Comparison Results for GARCH processes" by Bellini F., Pellerey F, Sgarra C. and Sekeh S.Y. [3]. Subsequently, following the methodology explained in the above article, i showed how the same results can be applied in the ARCH(q) model, extending the results obtained in the article to multistep models. Finally i tried to prove the propagation of parameters' orders in the the GARCH with leverage model. In this last case, although the orders do propagate to the logreturns, i didn't reached significant results concerning the logreturs' sums. Indeed, i show in detail how the non linearity introduced in this model modifies the results obtained in the generic case.
Finally, in the fourth chapter, i give a numerical proof of the previous results showing how usual stochastic order transmits from one or from all parameters to the absolute value and to the square of logreturns and induces a convex order on logreturns and, in some cases with symmetric innovations, on their sums.
Moreover we can compute the value a possible options' underlying as the initial value multiplied by the exponential of logreturns' sums. Clearly the

just described function is convex, then the underlying itself will be convexly ordered. It then follow that for convex payoff we get a precise order for options's price related to GARCH parameters.
This thesis take place in the wide and deepen field of stochastic orders financial applications. In my specific case stochastic orders help in providing some orders's properties on the values of the simulated series, starting from information on model's parameters.
Clearly, the case i dealt with, is not Markovian. However exists an extensive literature about similar subjects but related to continue and Markovian models.
For example in the article "Comparison of Option Prices in Semimartingale Models" by Bergenthum J. and Rüschendorf L. [4], the authors face the problem of comparing prices of European options, written on different underlying, computed under martingale equivalent measure.
Another extensive scope, still concerning the Markov process, is the one related to the order of martingale equivalent measures in semimartingale market models. Under these measure is possible to compute option prices that should follow the same order of the measures. For example in the article "Bounds on Option Prices for Semimartingale Market Models" by Gushchin A.A. and Mordecki [10], is explained how to get a reliable range on option prices having a convex payoff function thanks to an order induced by the coherent measures. Clearly, from these ranges, it will be possible to identify bounds on prices in a certain set of martingale equivalent measures.
In the article "Stochastic Orders and Risk Measures: Consistency and Bounds" by Baüerle N. and Müller A. [1], is underlined how, under some hypothesis, the risk measures are coherent with orders, especially with the usual stochastic order and with the convex order. From this remark is then analyzed how to get bounds on risk measures in relation to a portfolio-choice problem.

Contents

List of Figures

Introduction

The importance of stochastic orders in finance and in the insurance market is testified by an extensive literature of their applications. First of all lets see a basic and intuitive definition of a stochastic order and how it result useful in the portfolio selection.
We say that $X \leq_{\mathcal{F}} Y$ if

$$\mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]$$

for all the function f belonging to some function class $\mathcal{F}$. Observe that this definition provides an order between stochastic variables defined in relation to a particular class of functions whenever the expected value exists.
Consider now two future prices of financial assets, for example the asset A and the asset B. Their outcomes are usually represented by random variables, respectively X_A and X_B. The situation in which an investor has to choose between this two assets may be modeled with a Von Neumann-Morgenstern utility function u, related to the risk aversion of the investor. According to the expected utility principle, the investor prefers asset B to A if and only if

$$\mathbb{E}[u(X_A)] \leq \mathbb{E}[u(X_B)].$$

If the inequality holds for all investors whose utility functions belongs to some function class, then this is exactly the definition of some stochastic order between X_A and X_B. This conclusion proves the importance of the stochastic orders in risk management: having sufficient or necessary condition on the random variables to establish if and how they are ordered may help to establish which one is more or less risky.
In option pricing we often take the expected value of payoff function, having as arguments the option's fixed parameters and the stochastic underlying asset. In simulating the underlying asset we usually use random variables. In this thesis the main goal is to establish which orders are propagated from the random variables to the asset's dynamic that they generate. This will allow us, with payoff functions belonging to particular function classes, to order the option prices consequently to the stochastic order between the random

variables that generated the underlying asset. We will investigate this kind of problem in the specific case of ARCH, GARCH and GARCH with leverage models.

Before focusing on the financial usage of stochastic order, we would like to underline their several applications in insurance market.

Risks are generally modeled by random variables or distribution functions. The diversity of insurance types, or simply of insured populations is reproduced by introducing individual or collective risk models with rare and extreme events and, on the other hand, models with moderate or even bounded risks. The possibility of order risks, in some way, may lead to the main theories for risk measures. It helps also in estimating the ruin probability (see "Asymptotic Ruin Probabilities for Risk Processes with dependent increments" by Müller A. and Pflug G. [15]) that is the evaluation of probabilities of rare and extreme events.

A simple example is related to the insurance deductible. Let say for example that a car insurance has a deductible, for a certain kind of accidents, fixed at 500 euros per year. It means that the insurer, will have to refund all the expense superior to that amount to the client. Let call X and Y the random variables describing the year expense of two clients due to accidents of the type considered. This variable may be related to the age of the insured, to the place where he live, to the number of accident he had in the last few years and to other similar factors. We clearly have that the sum payed to the policyholder is defined by $f(X) = (X - 500)^+$. This function clearly belongs to the increasing and convex function (see chapter 2) so we do have that, if $X \leq_{icx} Y$

$$\mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]$$

holds true. This means that the client represented by Y should have a higher premium than the one represented by X.

Stochastic orders have many important applications in risk management.

For example in "Stochastic Orders and Risk Measures: Consistency and Bounds" by Baüerle N. and Müller A. [1] the authors try to verify if and when a monotone risk measure ρ has the property that $X \leq_{st} Y$ implies $\rho(X) \leq \rho(Y)$ and that a convex risk measure has the property that $X \leq_{cx} Y$ implies $\rho(X) \leq \rho(Y)$. They found out that the crucial point is the probability space on which the risks are defined. These results are applied to bound the risk of a portfolio, that is a joint financial position $X_1 + .. + X_n$. Another important subject faced in the article is the portfolio optimization bounds consequently to the main results in the Coherent Risk Measure scope.

In "Stochastic orders and their applications in financial optimization" by Kijima and Ohnishi [12] the attention is focused on how the stochastic orders

are applied is the demand and shift effect problems in portfolio selection. In this article are considered important order such as the likelihood ratio order and the reversed hazard rate order that we are not going to consider in this thesis.
We briefly describe how the option pricing techniques have developed through the year and then we will focus on the subject dealt in this thesis.
First of all consider an European option price. The most famous model for calculating its price is the Black-Merton-Scholes formula. It shows an increasing dependence of the option price by the riskiness of the underlying asset that is completely driven by logreturns distribution variance. We may describe the uncertainty through the dispersion around the means and, consequently, the distribution function can be ordered according to their"peakdness" (see chapter 1 for more details). The conclusion is that, the larger is the dispersion, the higher the option prices. This conclusion leads once again to stochastic orders as an important instrument to order underlying assets in relation to their risk thanks to the peakdness order. Clearly in more complex models in turns out to be fundamental a rigorous approach to avoid wrong conclusions.
The Black-Merton-Scholes (BMS) has proven unable to handle and reproduce some empirical facts related to the logreturns.
First of all the so called"fat tails": the BMS model is based on a normal distribution that, often, underestimate some critical and extreme events, under-pricing options that are far out of the money. In fact the extreme events, that take an out of the money option to an at the money option are empirically much more frequent then in the normal distribution.
Second of all, "volatility clustering": this volatility's property underlines how often periods of high volatility are followed by periods of high volatility as periods of low volatility are followed by periods of low volatility. Clearly this property is completely ignored by BMS model in which the volatility is considered constant in the whole period.
Then we have "aggregation Gaussianity" and"leverage effect". The first one is the tendency of logreturns to variate their distribution depending from the time scale. As one increases the time scale over which returns are calculated, their distribution looks more and more like a normal distribution but this is generally false with different time scales. The second one, point out an empirical evidence of a negative correlation between the option prices and the volatility. This is another market feature ignored by the BMS model.
Finally we have"volatility smile". Fixed all the other parameters we may calculate the implicit BMS volatility, and draw its dependence from the strike. We found out that for equities is decreasing, higher for low strikes and lower for higher strikes, and for the commodities, on the other hand, we have the

opposite behavior. Even this fact is in contrast with the BMS assumption of constant volatility.
In order to provide a more detailed and precise description of the faced problems, several different models were introduced.
A first generalization was made adding jumps to logprices dynamics, leading to the Levy processes with finite or infinite variation and activities. This choice was made in order to increase the probability of extreme events and, possibly, to create an asymmetry between the higher and the lower tails of the logprice distribution.
A second and maybe more important step was made introducing the stochastic volatility models. There are several continuous models in this scope such as the Heston model, the Hull and White model and the Sabr model (used in particular for the interest rate derivatives). Moreover we also have important discrete models such as the Autoregressive Conditioned Heteroschedastic (ARCH) models introduced by Engle in "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the UK Inflation" [8] and their general extension (the GARCH model) proposed by Bollerslev in "General Autoregressive Conditional Heteroskedasticity" [6]. There is also a further generalization of the GARCH model, made up introducing a leverage effect (that is an other parameter) that keeps the volatility higher when the magnitude of the previous logreturn is high and lower when it is low. These models will be largely discussed in the following chapters.
The main goal of the stochastic volatility is to fix the shortcomings of BMS model such as volatility clustering, volatility smile and the leverage effect. The volatility clustering, for example, might be reproduced through mean reverting processes with slow velocity of mean reverting. The leverage effects is simply handled using for stochastic volatility's dynamics and for logprice's dynamics two different Brownian Motions negatively correlated. The same strategy is applied to create empirical justifiable smiles or skews for the strike-volatility graph.
Finally this two big categories were combined to exploit their different positive effects in modeling both the volatility, the logprice and their correlation. The most famous model is the Bates one where logprice dynamic has jumps and is negatively correlated with the stochastic volatility dynamic.
The literature on analysis concerning the propagation of stochastic orders in these pricing models is wide. However the most important articles focus on the continue case, especially in the Markovian case.
The first big class of problems is the one related to the comparison of same model under different probability measures. In fact, for every equivalent martingale measure we get a different option price. Dealing with a similar problem, El Karoui, Jeanblanc-Picqu and Shreve in "Robustness of the

Black and Scholes formula" [7] studied how relations between two different volatilities are propagated to the option's prices. Subsequently Henderson and Hobson in "Coupling and option price comparisons in a jump-diffusion model" [11] gave a simple criterion for ordering of option prices under various martingale measures. Starting from this article but in a different context, Møller, in "Stochastic orders in dynamic reinsurance markets" [17] determined optimal martingale measures such as the minimal martingale measure and the minimal entropy martingale measure, and some comparison results for prices under different martingale measures, leading to a simple stochastic ordering result for the optimal martingale measures.
Moreover it's clear how important would be to fix bounds to the possible price's values under different probabilities.
In a first place this subject was faced by Bellamy and Jeanblanc in "Incompleteness of markets driven by a mixed diffusion" [2] who concluded that in incomplete markets driven by mixed diffusion, the range of prices is too large. They pointed out as a lower bound the Black and Scholes function evaluated at the underlying asset's price while the upper bound was determined just under particular hypothesis. Consequently Gushchin A.A. and Mordecki have deepened this subject in "Bounds on Option Prices for Semimartingale Market Models" [10]. In this article is proposed a methodology to get the desired bounds in the case of general semimartingale market model. This goal is reached thanks to a partial ordering in the set of distributions of discounted stock prices at exercise time T, that is the law of a random variable $Z_t = \frac{S_T}{B_T}$ under an arbitrary equivalent martingale measure. This order allow to find extreme distributions and, correspondingly, upper and lower bounds for the range of option prices.
Moreover in "Comparison of Option Prices in Semimartingale Models" by Bergenthum J. and Rüschendorf L. [4], is faced the problem of comparing d-dimensional exponential semimartingales computed under different probabilities. The main result of this article gives sufficient conditions for the comparison for European options, with convex payoff function, with respect to martingale pricing measures. Sufficient conditions for these orderings are formulated in terms of the predictable characteristics of the stochastic logarithm of the stock price processes. These conditions are applied to the following models: exponential semimartingale, stochastic volatility models, Lévy processes and diffusion with jumps processes.
Another class of problems concerns the comparison of models under the same probability measure but with different parametric specification. This one is the kind of problem we are going to focus on.
Let now focus on the main models we are going to analyze in this thesis: the ARCH and GARCH processes.

The most general form of the ARCH(q) model is:

$$X_t = \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0 + \textstyle\sum_{i=1}^{q} \alpha_i X_{t-i}^2$$

with positive parameters. This model subsequently evolved in the GARCH(p,q) models where the dependence from the previous values of volatility is highlighted:

$$X_t = \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0 + \textstyle\sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

with positive parameters too. Observe that the GARCH(1,1) model is made up by three parameters assuming a numerical value (possibly stochastic) and the innovations, Z_t, which are random IID variables with assigned density function.
We will focus on the GARCH(1,1) with leverage model that is:

$$X_t = \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1(X_{t-1} + \delta|X_{t-1}|)^2 + \beta\sigma_{t-1}^2$$

In this model there is one more parameter, δ, modeling how the volatility react to positive or negative values of logreturns. We usual consider negative δ to underline that bad news usually have the greater effect.
The last formula may be written as:

$$\sigma_t^2 = \alpha_0 + (\tilde{\alpha}_1 + \tilde{\delta}\,\mathbb{I}_{\{X_{t-1}<0\}})X_{t-1}^2 + \beta\sigma_{t-1}^2$$

where $\mathbb{I}$ is the indicator function, $\tilde{\alpha}_1 = \alpha_1(1+\delta)^2$ and $\tilde{\delta} = -2\delta\alpha_1$. This one is a particular form of the Threshold GARCH (TGARCH) in which the parameters values depend on the value (in this case on the sign) of the previous logreturn (see "Threshold GARCH model: Theory and Applications" by Wu J. [21] for further details).
First of all lets see an example to justify the further investigations. We consider the density of the logreturn sums in GARCH(1,1) model as in (3.13). We take the first two parameters $\alpha_0 = 10^{-6}$ and $\alpha_1 = 0.08$ and we consider the third parameter β initially equal to 0.8, then 0.85 and finally 0.9. We consider iid standard Gaussian variables as innovations and an initial variance σ_0 fixed to 0.1. We simulate 100000 trajectories of length $N = 50$ and we show how the logreturns sum cumulative distribution functions develop in respect to the considered values of β.

Let then consider the sequence of "stock prices" defined by $P_{k+1} = P_0 e^{S_k}$, $k = 0, ...50$. We take as the initial value of the underlying $P_0 = 10$ euros and a zero interest rate. We then compute the Monte Carlo prices C of the call options with strike $K = 10$ euros given by

$$C = \mathbb{E}[max(P_{51} - K, 0)] = \mathbb{E}[max(P_0 exp(S_{50}) - 10, 0)]$$

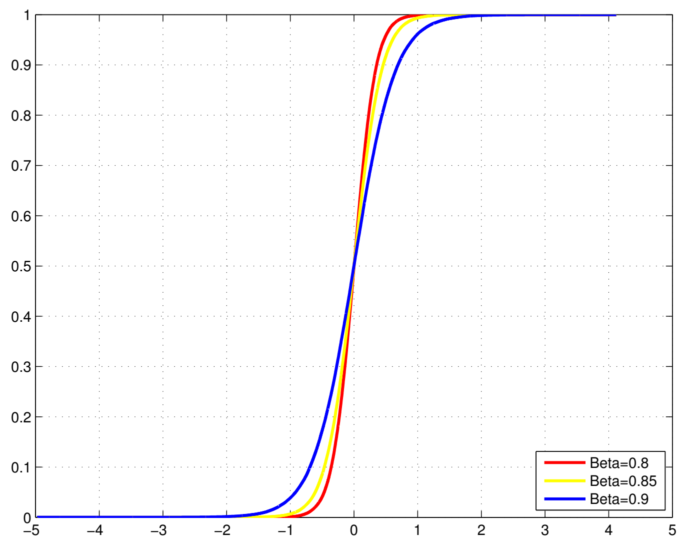


Figure 1: Distribution function of the logreturns sum with β =0.8 ,0.85, 0.9

where the expected values is taken for every value of the parameter β over the set of 1000000 simulated values of S_{50}.
We obtain the following prices:

a) $\beta = 0.8 \Rightarrow C = 1.3807$

b) $\beta = 0.85 \Rightarrow C = 1.8839$

c) $\beta = 0.9 \Rightarrow C = 3.4035$

Observe figure 1: as β increases, the probability of having lower and higher logreturns sum values increases. That means that the variance of the S_{50} variables increase. As a natural consequence the price of options with P_{51} as underlying will increase too.
It's then clear from the figure and from the option prices that the usual real number order relation implies some ordering on logreturn sums and option prices. Therefore it's natural to ask if this property holds true for other type of orders between the parameters and, possibly, between the innovations, particularly in the stochastic case.
We will then investigate which kind of stochastic orders between parameters and innovations are naturally propagated to the logreturns, to their sums, to the stock price and to options having the stock as underlying asset, through the ARCH, GARCH and GARCH with leverage models.

We start with Chapter 1 to describe the most famous stochastic orders: the so called usual stochastic order, the convex order, the increasing convex order and the peakdness order. We give necessary and sufficient conditions to verify that one or more that the above orders hold true. Moreover we describe the relations between these stochastic orders. To main sources for this chapter were “Stochastic Orders” by Shaked M. and Shanthikumar J. G. [20] and “Comparison Methods for Stochastic Models and Risks” by Müller and Stoyan.
In the second Chapter we recall some basic tools for Time Series Analysis and the natural development of the ARMA model to the ARCH and GARCH models. We also introduce the GARCH with leverage model and the TGARCH models. In this Chapter we took the cue from the book written by McNeil, Rüdiger, and Embrechts “Quantitative Risk Management” [13].
Finally in the third chapter we apply the stochastic orders results to the ARCH, GARCH and GARCH with leverage theory to conclude that with symmetric innovations, the convex order is propagated from the innovations to the logreturns, and to their sums just in the first two models. Besides the usual stochastic order for the parameters, imply the usual stochastic order on the absolute values of the logreturns and on their square, and the convex order on the logreturns in the three models. Moreover, in the first two models with symmetric innovations, the usual stochastic order of the parameters implies a convex order for the logreturn's sum. In order to write this part of the thesis the article “Comparison Results for GARCH processes” by Bellini F., Pellerey F, Sgarra C. and Sekeh S.Y. [3] was fundamental. In particular it really helped to understand how the proofs have to be build up.
In the last chapter we give some numerical proof of the Theorem of the previous chapter. We show how the stochastic orders are propagated in a first place taking just one stochastic parameter then considering all three parameters as stochastic. Finally we take the innovations ordered in the convex order to verify the propagation of convex order to logreturns and total logreturns.

Chapter 1

Stochastic Orders

1.1 The usual Stochastic Order

Consider two random variables X and Y such that:

$$P\{X > x\} \leq P\{Y > x\} \quad for\ all\ x \in (-\infty, \infty) \tag{1.1}$$

then X is said to be *smaller than* Y *in the usual stochastic order* (denoted by $X \leq_{st} Y$). Roughly speaking, (1.1) says that X is less likely than Y to take on large values, where "large" means any value greater than x, and that is the case for all x's. We may say that this order compares random variables according to their "*magintude*". Note that (1.1) is the same as

$$P\{X \leq x\} \geq P\{Y \leq x\} \quad for\ all\ x \in (-\infty, \infty) \tag{1.2}$$

It is easy to verify (by noting that every closed interval is an infinite intersection of open intervals) that $X \leq_{st} Y$ if, and only if

$$P\{X \geq x\} \leq P\{Y \geq x\} \quad for\ all\ x \in (-\infty, \infty) \tag{1.3}$$

In fact, we can write (1.1) and (1.3) in an equivalent way, as follows

$$P\{X \in U\} \leq P\{Y \in U\} \quad for\ all\ upper\ sets\ U \subseteq (-\infty, \infty) \tag{1.4}$$

Remember that in the univariate case, that is the real line, a set U is an upper set if, and only if, it is an open or a closed right half line. Moreover in (1.4) we can use the expected values instead of the probabilities to get:

$$\mathbb{E}[I_U(X)] \leq \mathbb{E}[I_U(Y)] \quad for\ all\ upper\ sets\ U \subseteq (-\infty, \infty) \tag{1.5}$$

where I_U denotes the indicator function of U. From (1.5) it then follows that if $X \leq_{st} Y$, then

$$\mathbb{E}[\sum_{i=1}^{m} a_i I_{U_i}(X)] - b \leq \mathbb{E}[\sum_{i=1}^{m} a_i I_{U_i}(Y)] - b \tag{1.6}$$

for all $a_i \geq 0$, $i = 1, 2, ..., m$, $b \in (-\infty, \infty)$, and $m \geq 0$.
Thanks to the following theorem we will give the most important characterization of the usual stochastic order.

Theorem 1.1. *Given an increasing function ϕ, it is possible, for each m, to define a sequence of U_i's, a sequence of a_i's, and a b (all of which may depend on m), such that as $m \to \infty$ then (1.6) converges to*

$$\mathbb{E}[\sum_{i=1}^{m} a_i I_{U_i}(X) - b] \to \mathbb{E}[\phi(X)] \tag{1.7}$$

provided the expectations exist.

It then follows that:

Definition 1.1. $X \leq_{st} Y$ if, and only if, (1.7) hold for all increasing function ϕ for which the expectation exist.

We give another important characterization of the usual stochastic order in the following Theorem (here $=_{st}$ denotes equality in law).

Theorem 1.2. *Two random variables X and Y satisfy $X \leq_{st} Y$ if, and only if, there exist two random variables $\hat{X}$ and $\hat{Y}$, defined on the same probability space , such that*

$$\hat{X} =_{st} X \tag{1.8}$$

$$\hat{Y} =_{st} Y \tag{1.9}$$

and

$$P[\hat{X} \leq \hat{Y}] = 1. \tag{1.10}$$

Proof. Obviously (1.8), (1.9) and (1.10) imply that $X \leq_{st} Y$. In order to prove the necessity part of this Theorem, let F and G be, respectively, the distribution functions of X and Y, and let F^{-1} and G^{-1} be the corresponding right continuous inverses. Define $\hat{X} = F^{-1}(U)$ and $\hat{Y} = G^{-1}(U)$ where U is a uniform [0,1] random variable. Then it is seen that $\hat{X}$ and $\hat{Y}$ satisfy (1.8) and (1.9). Remark that $X \leq_{st} Y \Leftrightarrow P[X \leq x] \geq P[Y \leq x]$ for all $x \in (-\infty, \infty)$. It follows that (1.10) also holds. □

Afterwards the following notation will be used: for any random variable Z and an event, A, let $[Z|A]$ denote any random variable that has as its distribution the conditional distribution of Z given A.

Theorem 1.3. *a) If $X \leq_{st} Y$ and g is an increasing [decreasing] function, then $g(X) \leq_{st} [\geq_{st}] g(Y)$.*

b) Let $X_1, X_2, ..., X_m$ be a set of independent random variables and let $Y_1, Y_2, ..., Y_m$ be another set of independent random variables. If $X_i \leq_{st} Y_i$ for $i = 1, 2, ..., m$, then, for any increasing function $\psi : \mathbb{R}^m \to \mathbb{R}$, one has

$$\psi(X_1, X_2, ..., X_m) \leq_{st} \psi(Y_1, Y_2, ..., Y_m) \tag{1.11}$$

in particular,

$$\textstyle\sum_{j=1}^{M} X_j \leq \sum_{j=1}^{M} Y_j$$

That is, the usual stochastic order is closed under convolutions.

c) Let $\{X_j \; j = 1, 2, ...\}$ and $\{Y_j \; j = 1, 2, ...\}$ be two sequences of random variables such that $X_j \to_{st} X$ and $Y_j \to_{st} Y$ as $j \to \infty$, where $\to_{st}$ denotes convergence in distribution. If $X_j \leq_{st} Y_j$, $j = 1, 2...$ then $X \leq_{st} Y$.

d) Let X, Y and Θ be random variables such that $[X|\Theta = \theta] \leq_{st} [Y|\Theta = \theta]$ for all θ in the support of Θ. Then $X \leq_{st} Y$. That is, the usual stochastic order is closed under mixtures.

Proof. To prove the statement (a) recall that the combination of two increasing function is an increasing function. Then defined $\psi = \phi \circ g$ for every increasing function ϕ hold:

$$\mathbb{E}[\phi(g(X))] = \mathbb{E}[\psi(X)] \leq \mathbb{E}[\psi(Y)] = \mathbb{E}[\phi(g(Y))]$$

where the inequality follows from $X \leq_{st} Y$. We get the thesis using Definition 1. Let now consider (b). Without loss of generality we can assume that all the $2m$ random variables are independent because such an assumption does not affect the distributions of $\psi(X_1, X_2, ..., X_m)$ and $\psi(Y_1, Y_2, ..., Y_m)$. The proof is by induction on m. We start considering $m = 1$ and using (a) we get $g(X) \leq_{st} g(Y)$. Assume that (1.11) holds true for vectors of size $m - 1$. Let g and ϕ be increasing functions. Then

$$\mathbb{E}[\phi(\psi(X_1, X_2, ..., X_m))|X_1 = x] = \mathbb{E}[\phi(\psi(x, X_2, ..., X_m))] \leq \\ \mathbb{E}[\phi(\psi(x, Y_2, ..., Y_m))] = \mathbb{E}[\phi(\psi(X_1, Y_2, ..., Y_m))|X_1 = x]$$

where the equalities above follow from the independence assumption and the inequality follows from the induction hypothesis. Taking expectations with respect to X_1, we obtain

$$\mathbb{E}[\phi(\psi(X_1, X_2, ..., X_m))] \leq \mathbb{E}[\phi(\psi(X_1, Y_2, ..., Y_m))]$$

Repeating the argument, but now considering on $Y_2, ..., Y_m$ and using (1.11) with $m = 1$, we see that

$$\mathbb{E}[\phi(\psi(X_1, Y_2, ..., Y_m))] \leq \mathbb{E}[\phi(\psi(Y_1, Y_2, ..., Y_m))]$$

and this proves the result. The point (c) is a consequence of the convergence of the density functions of the X_n and Y_n to the distribution function of X and Y. In fact for every n we have that for every increasing function ϕ:

$$X_n \leq_{st} Y_n \Leftrightarrow \int_{-\infty}^{\infty} \phi(x) f_{X_n}(x)dx \leq \int_{-\infty}^{\infty} \phi(x) f_{Y_n}(x)$$

that, thanks to the convergence of the density functions, leads to

$$\int_{-\infty}^{\infty} \phi(x) f_X(x)dx \leq \int_{-\infty}^{\infty} \phi(x) f_Y(x)$$

that is equivalent to Definition (1.1).
To end the proof of the Theorem observe that for any increasing function ψ:

$$\mathbb{E}[\psi(X)] = \mathbb{E}_\Theta \mathbb{E}[\psi(X)|\Theta = \theta] \leq \mathbb{E}_\Theta \mathbb{E}[\psi(Y)|\Theta = \theta] = \mathbb{E}[\psi(Y)]$$

that, with Definition 1, is (d). □

Clearly, if $X \leq_{st} Y$ then $\mathbb{E}[X] \leq \mathbb{E}[Y]$. It's easy to find counterexamples which show that the converse is false.

Example 1.1. Let consider two discrete random variables X and Y such that:

$$P(X = 1) = 1/3, \; P(X = 2) = 1/3, \; P(X = 9) = 1/3 \Rightarrow \mathbb{E}[X] = 4$$

$$P(Y = 4.5) = 1 \Rightarrow \mathbb{E}[Y] = 4.5$$

then we have $\mathbb{E}[X] < \mathbb{E}[Y]$, but considering the increasing function $\psi(x) = x^3$ we get $\mathbb{E}[X^3] = 246$ $\mathbb{E}[Y^3] = 91.125$.

However, as the following result shows, if two random variables are ordered in the usual stochastic order and have the same expected values, they must have the same distribution.

Theorem 1.4. *If $X \leq_{st} Y$ and if $\mathbb{E}[h(X)] = \mathbb{E}[h(Y)]$ for some strictly increasing function h, then $X =_{st} Y$.*

Proof. First we prove the result when $h(x) = x$. Let $\hat{X}$ and $\hat{Y}$ be as Theorem 1.2. If $P(\hat{X} < \hat{Y}) > 0$, then $\mathbb{E}[X] = \mathbb{E}[\hat{X}] < \mathbb{E}[\hat{Y}] = \mathbb{E}[Y]$, a contradiction to the assumption $\mathbb{E}[X] = \mathbb{E}[Y]$. Therefore $X =_{st}= \hat{X} = \hat{Y} =_{st} Y$. Now let h be some strictly increasing function. Observe that if $X \leq_{st} Y$, then $h(X) \leq_{st} h(Y)$ and therefor from the above result we have that $h(X) =_{st} h(Y)$. The strict monotonicity of h yields $X =_{st} Y$. $\square$

Moreover observe that $X \leq_{st} Y$ implies all the inequality related to the odd moments (for example, $\mathbb{E}[X^3] \leq \mathbb{E}[Y^3]$).

A simple sufficient condition which implies the usual Stochastic Order is described next. The following notation will be used. Let $a(x)$ be defined on I, where I is a subset of the real line. The number of sign changes of a in I is defined by:

$$S^-(a) = supS^-[a(x_1), a(x_2), ..., a(x_m)], \tag{1.12}$$

where $S^-(y_1, y_2, ..., y_m)$ is the number of sign changes of the indicated sequence, zero terms being discarded, and the supremum in (1.12) is extended over all sets $x_1 < x_2 < ... < x_m$ such that $x_i \in I$ and $m < \infty$. We can then state the following Theorem:

Theorem 1.5. *Let X and Y be two random variables with (discrete or continuous) density functions of the form f and g, respectively. If*

$$S^-(g - f) = 1 \quad \textit{and the sign sequence is } -, +$$

then $X \leq_{st} Y$.

Proof. Fix $\overline{x}$ such that $f(\overline{x}) = g(\overline{x})$. We then have by the hypothesis

$$g(x) < f(x) \quad \forall x < \overline{x} \quad \textit{and} \quad g(x) > f(x) \quad \forall x > \overline{x}$$

We then have for $x < \overline{x}$:

$$P(X \geq x) = 1 - \int_{-\infty}^{x} f(x)dx < 1 - \int_{-\infty}^{x} g(x)dx = P(Y \geq x)$$

and for $x > \overline{x}$:

$$P(X \geq x) = \int_{x}^{\infty} f(x)dx < \int_{x}^{\infty} g(x)dx = P(Y \geq x)$$

That correspond to the characterization of the usual stochastic orders in (1.3). $\square$

Obviously we can state a similar sufficient condition based on the distribution functions.

Corollary 1.1. Let X and Y be two random variables with (discrete or continuous) distribution functions of the form F and G, respectively. If

$$S^-((G - F)') = 1 \quad \textit{and the sign sequence is } -, +$$

then $X \leq_{st} Y$.

1.2 The Convex Order

Let X and Y be two random variables such that

$$\mathbb{E}[\phi(X)] \leq \mathbb{E}[\phi(Y)] \text{ for convex [concave] functions } \phi : \mathbb{R} \to \mathbb{R} \tag{1.13}$$

provided the expectations exist. Then X is said to be *smaller than* Y *in the convex order*, denoted as $X \leq_{cx} Y$. Roughly speaking, convex functions are functions that take their (relatively) larger values over regions of the form $(-\infty, a) \cup (b, \infty)$ for $a \leq b$. Therefore if (1.13) holds, then Y is more likely to take"extreme" values than X. That is, Y is "more variable" than X. It should be mentioned here that in (1.13) it is sufficient to consider only functions ϕ that are convex on the union of the supports of X and Y rather than over the whole real line.
One can also define a *concave order* by requiring (1.13) to hold for all concave functions ϕ (denoted as $X \leq_{cv} Y$). However, $X \leq_{cv} Y$ if, and only if, $Y \leq_{cx} X$. Therefore, it is not necessary to have a separate discussion for the concave order.
Note that the functions $\phi_1(x) = x$ and $\phi_2(x) = -x$ are both convex. Therefore, from (1.13) it easily follows that:

$$X \leq_{cx} Y \Rightarrow \mathbb{E}[X] = \mathbb{E}[Y], \tag{1.14}$$

provided the expectations exist. Later it will be helpful to observe that if $\mathbb{E}[X] = \mathbb{E}[Y]$, then

$$\int_{-\infty}^{\infty} [F(u) - G(u)]du = \int_{-\infty}^{\infty} [\overline{F}(u) - \overline{G}(u)]du = 0 \tag{1.15}$$

provided the integrals exist, where $\overline{F}[F]$ and $\overline{G}[G]$ are the survival [distribution] functions of X and Y, respectively. The function $\phi(x) = x^n$, with $n \geq 2$ even, is convex, therefore from (1.13):

$$X \leq_{cx} Y \Rightarrow \mathbb{E}[X^n] \leq \mathbb{E}[Y^n] \quad \textit{with } n \textit{ even}$$

that is, all the even moments of X are lower then the Y's ones. In particular, using the equality between the expected values of the stochastically convex ordered variables:

$$X \leq_{cx} Y \Rightarrow Var[X] \leq Var[Y] \tag{1.16}$$

whenever $\text{Var}[Y] < \infty$.
For a fixed a, the function $\phi_a(x) = (x - a)_+$, and the function $\varphi_a(x) = (a - x)_+$, are both convex. Therefore, if $X \leq_{cx} Y$, then

$$\mathbb{E}[(X - a)_+] \leq \mathbb{E}[(Y - a)_+] \quad \textit{for all } a \tag{1.17}$$

and

$$\mathbb{E}[(a-X)_+] \leq \mathbb{E}[(a-Y)_+] \quad \textit{for all } a \tag{1.18}$$

provided the expectations exist. Alternatively, using a simple integration by parts:

$$\begin{gathered}\mathbb{E}[(a-X)_+] = \int_{\mathbb{R}} (a-u)_+ f(u)\,du = aF(a) - \int_{-\infty}^{a} u f(u)\,du = \\ aF(a) - (aF(a) - \int_{-\infty}^{a} F(u)\,du) = \int_{-\infty}^{a} F(u)\,du\end{gathered}$$

similarly

$$\mathbb{E}[(a-Y)_+] = \int_{-\infty}^{a} G(u)\,du$$

$$\mathbb{E}[(X-a)_+] = \int_{a}^{\infty} \overline{F}(u)\,du$$

$$\mathbb{E}[(Y-a)_+] = \int_{a}^{\infty} \overline{G}(u)\,du$$

It than follow that (1.17) and (1.18) can be rewritten as

$$\int_x^\infty \overline{F}(u)du \leq \int_x^\infty \overline{G}(u)du \tag{1.19}$$

and

$$\int_{-\infty}^x F(u)du \leq \int_{-\infty}^x G(u)du \tag{1.20}$$

provided the integrals exist. In fact, when $\mathbb{E}[X] = \mathbb{E}[Y]$, (1.19) is equivalent to $X \leq_{cx} Y$. To see this equivalence, note that every convex function can be approximated by (that is, is a limit of) positive linear combinations of the functions ϕ_a's, for various choices of a's, and of the function $\psi(x) = -x$. By (1.19), $\mathbb{E}[\phi_a(X)] \leq \mathbb{E}[\phi_a(Y)]$ for all a's, and this fact, together with the equality of the means of X and Y ($\Rightarrow \mathbb{E}[\psi(X)] \leq \mathbb{E}[\psi(Y)]$), implies (1.13). We thus have proved the first part of the following result. The other part is proven similarly.

Theorem 1.6. *Let X and Y be two random variables such that $\mathbb{E}[X] = \mathbb{E}[Y]$. Then*

(a) $X \leq_{cx} Y$ if, and only if, (1.19) holds.

(b) $X \leq_{cx} Y$ if, and only if, (1.20) holds.

Corollary 1.2. Let X and Y be two random variables such that $X \leq_{cx} Y$. Then $aX \leq_{cx} aY$ for every $a \in \mathbb{R}$.

Proof. Using the inequality in the definition of the convex order with $\psi(x) = x$ and $\phi(x) = -x$ we get $\mathbb{E}[X] = \mathbb{E}[Y]$ and consequently $\mathbb{E}[aX] = \mathbb{E}[aY]$ for every $a \in \mathbb{R}$. Using $F_{aX}(u) = F_X(u/a)$ and $G_{aY}(u) = G_Y(u/a)$, 1.20 and a change of variables:

$$\int_{-\infty}^{x} F_{aX}(u)\, du = a \int_{-\infty}^{x/a} F_X(v)\, dv \leq a \int_{-\infty}^{x/a} G_Y(v)\, dv = \int_{-\infty}^{x} G_{aY}(u)\, du$$

that implies the thesis. □

Moreover, when both the random variables X and Y of the previous Corollary have expected values zero, we have a stronger result that is proved using the same arguments of Theorem 1.6:

Corollary 1.3. Let X and Y be two random variables such that $X \leq_{cx} Y$ and $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. Then $aX \leq_{cx} bY$ for every $a \leq b \in \mathbb{R}_+$.

Proof. We have to prove that for every $\phi_\alpha(x) = (x - \alpha)_+$ and for the function $\psi(x) = -x$

$$\mathbb{E}[\phi_\alpha(aX)] \leq \mathbb{E}[\phi_\alpha(bX)] \quad \textit{and} \quad \mathbb{E}[\psi(aX)] \leq \mathbb{E}[\psi(bX)]$$

do hold. In fact it would then hold for every convex, function that is $aX \leq_{cx} bY$. Let start with ϕ_α: notice that $f(x) = (cx - a)_+$ is convex for every $c, a \in \mathbb{R}$ and that for $d \geq c > 0$ we have $\mathbb{E}[(cx - a)_+] \leq \mathbb{E}[(dx - a)_+]$. To prove this last statement consider separately the case of $a \geq 0$ and $a < 0$.

- $a \geq 0$ In this case we have that $(cx - a)_+ \leq (dx - a)_+ \forall x$ and the same inequality hold for the expected values.

- $a < 0$ In this case $(cx - a)_+ \leq (dx - a)_+$ holds just for $x \geq 0$. We then have to prove $\mathbb{E}[(cx - a)_+] \leq \mathbb{E}[(dx - a)_+]$ directly. From now on we will indicate as f_x the density function of the variable x. First of all notice that the thesis is equivalent to:

$$c \int_{\frac{a}{c}}^{\infty} x f_x(x) dx - a \int_{\frac{a}{c}}^{\infty} f_x(x) dx \leq d \int_{\frac{a}{d}}^{\infty} x f_x(x) dx - a \int_{\frac{a}{d}}^{\infty} f_x(x) dx$$

Moreover we have

$$- \int_{\frac{a}{c}}^{\infty} f_x(x) dx \leq - \int_{\frac{a}{d}}^{\infty} f_x(x) dx \Leftrightarrow \int_{\frac{a}{c}}^{\infty} f_x(x) dx \geq \int_{\frac{a}{d}}^{\infty} f_x(x) dx$$

that is true as we are integrating the same positive function on a bigger interval on the left side.
Then we just have to prove

$$c \int_{\frac{a}{c}}^{\infty} x f_x(x)dx \leq d \int_{\frac{a}{d}}^{\infty} x f_x(x)dx$$

We have that $\int_{\frac{a}{c}}^{\frac{a}{d}} x f_x(x)dx \leq 0$ as the function is negative in the considered interval. Than is trivial to prove:

$$c \int_{\frac{a}{c}}^{\infty} x f_x(x)dx \leq c \int_{\frac{a}{c}}^{\infty} x f_x(x)dx - c \int_{\frac{a}{c}}^{\frac{a}{d}} x f_x(x)dx = c \int_{\frac{a}{d}}^{\infty} x f_x(x)dx \leq d \int_{\frac{a}{d}}^{\infty} x f_x(x)dx$$

that is the desired statement.

We then have

$$\mathbb{E}[\psi_\alpha(aX)] = \mathbb{E}[(aX - \alpha)_+] \leq \mathbb{E}[(aY - \alpha)_+] \leq \mathbb{E}[(bY - \alpha)_+] = \mathbb{E}[\psi_\alpha(bY)]$$

Consider now $\psi(x)$:

$$\mathbb{E}[-aX] = -a\mathbb{E}[X] = 0 \quad and \quad \mathbb{E}[-bY] = -b\mathbb{E}[Y] = 0$$

that is

$$\mathbb{E}[\psi(aX)] \leq \mathbb{E}[\psi(bY)]$$

that ends the proof. □

Another way to describe the convex order can be deduced by adding a to both sides of the inequality in (1.17), it is seen that (1.17) can be rewritten as

$$\mathbb{E}[max\,\{X, a\}] \leq \mathbb{E}[max\,\{Y, a\}] \quad for\ all\ a \tag{1.21}$$

Thus, when $\mathbb{E}[X] = \mathbb{E}[Y]$, then (1.21) is equivalent to $X \leq_{cx} Y$. In a similar manner (1.18) can be rewritten.
The following Theorem provides another characterization of the convex order.

Theorem 1.7. *Let X and Y be two random variables such that $\mathbb{E}[X] = \mathbb{E}[Y]$. Then $X \leq_{cx} Y$ if, and only if,*

$$\mathbb{E}[|X - a|] \leq \mathbb{E}[|Y - a|] \quad for\ all\ a \in \mathbb{R} \tag{1.22}$$

Proof. Clearly, if $X \leq_{cx} Y$, then (1.22) holds. So suppose that (1.22) holds. Without loss of generality it can be assumed that $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. A straightforward computation gives:

$$\mathbb{E}[|X - a|] = a + 2\int_a^\infty \overline{F}(u)du = -a + 2\int_{-\infty}^a F(u)du \tag{1.23}$$

The result now follows from (1.19) or (1.20). □

An immediate consequence of (1.17) is shown next. Denote the supports of X and Y by $supp(X)$ and $supp(Y)$. Let $l_X = inf\{x : x \in supp(X)\}$ and $u_X = sup\{x : x \in supp(X)\}$. Define l_Y and u_Y similarly. Then we have that if $X \leq_{cx} Y$, then $l_Y \leq l_X$ and $u_Y \geq u_X$. As proof, suppose, for example, that $u_Y < u_X$. Let a be such that $u_Y < a < u_X$. Then $\mathbb{E}[(Y - a)_+] = 0 < \mathbb{E}[(X - a)_+]$, in contradiction to (1.17). Therefore we must have $u_Y \geq u_X$. Similarly using (1.18), it can be shown that $l_Y \leq l_X$. As a consequence we have that if X and Y are random variables whose supports are intervals, then:

$$X \leq_{cx} Y \Rightarrow supp(X) \subseteq supp(Y) \tag{1.24}$$

An important characterization of the convex order by construction on the same probability space is stated in the next Theorem.

Theorem 1.8. *Two random variables X and Y satisfy $X \leq_{cx} Y$ if, and only if, there exist two random variables $\hat{X}$ and $\hat{Y}$, defined on the same probability space , such that*

$$\hat{X} =_{st} X \tag{1.25}$$

$$\hat{Y} =_{st} Y \tag{1.26}$$

and $\hat{X}, \hat{Y}$ is a martingale, that is,

$$\mathbb{E}[\hat{Y}|\hat{X}] = \hat{X} \quad a.s. \tag{1.27}$$

Furthermore, the random variables $\hat{X}$ and $\hat{Y}$ can be selected such that $[\hat{Y}|\hat{X} = x]$ is increasing function in x in the usual stochastic order $\leq_{st}$.

Proof. It is not easy to proof the constructive part of Theorem 1.8. However, it easy to prove that if random variables $\hat{X}$ and $\hat{Y}$ as described in the Theorem exist, the $X \leq_{cx} Y$. Just note that if ϕ is a convex function, then by Jensen's Inequality,

$$\mathbb{E}[\phi(X)] = \mathbb{E}[\phi(\hat{X})] = \mathbb{E}[\phi(\mathbb{E}[\hat{Y}|\hat{X}])] \leq \mathbb{E}[\mathbb{E}[\phi(\hat{Y})|\hat{X}]] = \mathbb{E}[\phi(\hat{Y})] = \mathbb{E}[\phi(Y)]$$

which is (1.13). □

Let consider now the properties of the convex order. Unfortunately it turns out that this order is not close with respect to weak convergence, as the following example shows.

Example 1.2. Let

$$P(X_n = 1) = 1 \quad for\ all\ n \in \mathbb{N}^+$$

and

$$P(Y_n = n) = 1/n = 1 - P(Y_n = 0) \quad for\ all\ n \in \mathbb{N}^+.$$

Then the sequences (X_n) and (Y_n) converge in distribution to X and Y respectively where

$$P(X = 1) = P(Y = 0) = 1$$

Moreover, notice that $X_n \leq_{cx} Y_n$ in fact to prove it, is sufficient to prove that (1.13) holds for $\phi_a(x) = (x - a)_+$ for every a and for $\psi(x) = -x$. In fact every convex function can be approximated by positive linear combination of this functions. We have:

$$\mathbb{E}[\psi(X_n)] = -1 \leq -1 = \mathbb{E}[\psi(Y_n)]$$

$$\mathbb{E}[\psi(X_n)] = 1 - a \leq 1 - a = \mathbb{E}[\psi(Y_n)] \quad for\ a \leq 0$$

$$\mathbb{E}[\psi(X_n)] = 1 - a \leq 1 - \tfrac{a}{n} = \mathbb{E}[\psi(Y_n)] \quad for\ 0 \leq a \leq 1$$

$$\mathbb{E}[\psi(X_n)] = 0 \leq \tfrac{1}{n}(n - a)_+ = \mathbb{E}[\psi(Y_n)] \quad for\ a \geq 1$$

thus $X_n \leq_{cx} Y_n$ for all n. However, taking the increasing and convex function $f(x) = x$ we have

$$\mathbb{E}[f(X)] = \mathbb{E}[X] = 1 > 0 = \mathbb{E}[Y] = \mathbb{E}[f(Y)]$$

that contradicts $X \leq_{cx} Y$.

Therefore a stronger definition of convergence is needed to obtain a positive result. Let see it with other important properties.

Theorem 1.9. (*a*) *Let X and Y be two random variables. Then*

$$X \leq_{cx} Y \Leftrightarrow -X \leq_{cx} -Y \tag{1.28}$$

(*b*) *Let X, Y, and Θ be random variables such that $[X|\Theta = \theta] \leq_{cx} [Y|\Theta = \theta]$ for all θ in the support of Θ. Then $X \leq_{cx} Y$. That is, the convex order is closed under mixtures.*

(c) *Let $\{X_j, j = 1, 2, ...\}$ and $\{Y_j, j = 1, 2, ...\}$ be two sequences of random variables such that $X_j \to_{st} X$ and $Y_j \to_{st} Y$ as $j \to \infty$. Assume that*

$$\mathbb{E}[|X_j|] \to \mathbb{E}[|X|] \quad and \quad \mathbb{E}[|Y_j|] \to \mathbb{E}[|Y|] \quad as\ j \to \infty \tag{1.29}$$

If $X_j \leq_{cx} Y_j$, j =1,2,..., then $X \leq_{cx} Y$.

(d) *Let $X_1, X_2, ..., X_m$ be a set of independent random variables and let $Y_1, Y_2, ..., Y_m$ be another set of independent random variables. If $X_i \leq_{cx} Y_i$, for i = 1,2,...m then*

$$\sum_{j=1}^{m} X_j \leq_{cx} \sum_{j=1}^{m} Y_j. \tag{1.30}$$

That is, the convex order is closed under convolutions.

Proof. To prove the statement (a) observe that

$$\psi(x) \text{ is convex} \Leftrightarrow \psi(-x) \text{ is convex}$$

It follows that:

$$X \leq_{cx} Y \Leftrightarrow \mathbb{E}[\psi(X)] \leq \mathbb{E}[\psi(Y)]\ \ \forall\psi\ convex \Leftrightarrow \mathbb{E}[\psi(-X)] \leq \mathbb{E}[\psi(-Y)]\ \ \forall\psi\ convex \Leftrightarrow -X \leq_{cx} -Y$$

Moreover under the given assumption for any convex function ψ:

$$\mathbb{E}[\psi(X)] = \mathbb{E}_{\Theta}\mathbb{E}[\psi(X)|\Theta = \theta] \leq \mathbb{E}_{\Theta}\mathbb{E}[\psi(Y)|\Theta = \theta] = \mathbb{E}[\psi(Y)]$$

that proves point (b). In order to prove part (c) of the Theorem we will use the characterization of the convex order given in Theorem 1.7. Without loss of generality it can be assumed that $\mathbb{E}[X_j] = \mathbb{E}[Y_j] = \mathbb{E}[X] = \mathbb{E}[Y] = 0$ for all j. From (1.23) we have that

$$\mathbb{E}[|X_j - a|] = -a + 2\int_{-\infty}^{a} F_j(u)du \quad for\ all\ a$$

where F_j denotes the distribution function of X_j. In particular, when $a = 0$, it is seen that

$$\mathbb{E}[|X_j|] = 2\int_{-\infty}^{0} F_j(u)du.$$

Therefore

$$\mathbb{E}[|X_j - a|] = \mathbb{E}[|X_j|] - a + 2\int_{0}^{a} F_j(u)du.$$

Using (1.29) it is seen that as $j \to \infty$, the latter expression converges to

$$\mathbb{E}[|X - a|] = \mathbb{E}[|X|] - a + 2\int_{0}^{a} F(u)du.$$

where F is the distribution function of X. That is, for all a,

$$\mathbb{E}[|X_j - a|] \to \mathbb{E}[|X - a|] \quad as\ j \to \infty$$

Similarly

$$\mathbb{E}[|Y_j - a|] \to \mathbb{E}[|Y - a|] \quad as\ j \to \infty$$

The result now follows from Theorem 1.7. To prove part (d) of the Theorem note that part (b) can be rephrased as follows: let Z_1, Z_2, and Θ be independent random variables and let g be a bivariate function such that

$$g(Z_1, \Theta) \leq_{cx} g(Z_2, \Theta) \quad for\ all\ \theta\ in\ the\ support\ of\ \Theta \tag{1.31}$$

Then

$$g(Z_1, \Theta) \leq_{cx} g(Z_2, \Theta)$$

If Z_1 and Z_2 satisfy $Z_1 \leq_{cx} Z_2$, then the function g defined by $g(z, \theta) = z + \theta$ satisfies (1.31), since the order $\leq_{cx}$ is closed under shifts. Thus we have shown that if $Z_1 \leq_{cx} Z_2$ and Θ is any random variable independent of Z_1 and Z_2, then

$$Z_1 + \Theta \leq_{cx} Z_2 + \Theta \tag{1.32}$$

Repeated applications of (1.32) yield part (d) of Theorem 1.9. □

We will need the following Corollary:

Corollary 1.4. Let X, Y and Z be three random variables such that $X \leq_{cx} Y$, with Z independent from both X and Y. Then $XZ \leq_{cx} YZ$.

Proof. For every realization of the variable $Z = z$ we get $Xz \leq_{cx} Yz$ by Corollary 1.1 (recall that the variables X and Y are not influenced from the values of Z). The thesis follows then easily from point (b) of the previous Theorem. □

It should be pointed out, in contrast to part (a) of Theorem 1.9, that if X and Y are such $X \leq_{cx} Y$, it is not necessarily true that $X \leq_{cx} -Y$ also, even when $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. This can be seen easily from (1.24).
Without condition (1.29) the conclusion of part (c) of Theorem 1.9, as shown in the next example, need not to be true.

Example 1.3. Let the X_j's be all uniformly distributed on [0.5,1.5]. And let the Y_j's be such that

$$P\{Y_j = 0\} = (j-1)/j \quad P\{Y_j = j\} = 1/j, \quad j \geq 2$$

Note that the distributions of the Y_j's converge to a distribution that is degenerate at 0. Here $X_j \leq_{cx} Y_j$, $j = 2, 3, ...$, but it is not true that $X \leq_{cx} Y$.

Observe that, given the two random variables and their distributions it is sometimes not clear how to verify that $X \leq_{cx} Y$. We now point out several simple conditions that imply the convex order. Recall the notation $S^{-}(a)$ (defined in (1.12)) for the number of sign change of the function a.

Theorem 1.10. *Let X and Y be two random variables with equal means, density function f and g, distribution functions F and G, and survival functions $\overline{F}$ and $\overline{G}$, respectively. Then $X \leq_{cx} Y$ if any of the following conditions hold:*

$$S^{-}(g-f) = 2 \quad \text{and the sign sequences is } +, -, + \tag{1.33}$$

$$S^{-}(\overline{F}-\overline{G}) = 1 \quad \text{and the sign sequences is } +, - \tag{1.34}$$

$$S^{-}(G-F) = 1 \quad \text{and the sign sequences is } +, - \tag{1.35}$$

Proof. We will prove the result for the continuous case; the proof in the discrete case is similar. Suppose that $S^{-}(g-f) = 2$ and that the sign sequence is $+, -, +$. Let a and b $(a < b)$ be two of the crossing points. Denote $I_1 = (-\infty, a]$, $I_2 = (a, b]$, and $I_3 = (b, \infty)$. Then $g(x) - f(x) \geq 0$ on I_1, $g(x) - f(x) \leq 0$ on I_2 and $g(x) - f(x) \geq 0$ on I_3. Therefore

$$G(x) - F(x) = \textstyle\int_{-\infty}^{x} [g(u) - f(u)]du$$

is increasing on I_1, decreasing on I_2 and increasing on I_3. It is also clear that

$$\lim_{x \to -\infty} [G(x) - F(x)] = \lim_{x \to \infty} [G(x) - F(x)] = 0.$$

Combining all these observation shows that $S^{-}(G-F) = 1$ and that the sign sequence is $+, -$. Now suppose that $S^{-}(G-F) = 1$ and that the sign sequence is $+, -$. Let c be a crossing point. Denote $J_1 = (-\infty, c]$ and $J_2 = (c, \infty)$. Then $G(x) - F(x) \geq 0$ on J_1 and $G(x) - F(x) \leq 0$ on J_2. Clearly:

$$\lim_{x \to -\infty} \textstyle\int_{-\infty}^{x} [G(u) - F(u)]du = 0$$

and from the equality of the means (see (1.15)) it follows that

$$\lim_{x \to \infty} \textstyle\int_{-\infty}^{x} [G(u) - F(u)]du = 0$$

Combining this observations shows that (1.20) holds. This proves that (1.33) and (1.35) imply $X \leq_{cx} Y$. Note that $S^{-}(\overline{F}-\overline{G}) = S^{-}(G-F)$ with the same sign sequence. This observation, together with (1.35), shows that (1.34) implies $X \leq_{cx} Y$. □

1.3 The Increasing Convex Order

The order we are going to define has the purpose to compare random variables according both to their "location" (or "magnitude") and their "spread". Let X and Y be two random variables such that

$$\mathbb{E}[\phi(X)] \leq \mathbb{E}[\phi(Y)] \qquad \textit{for all increasing convex [concave] functions } \phi : \mathbb{R} \to \mathbb{R} \tag{1.36}$$

provided the expectations exist. Then X is said to be *smaller than Y in the increasing convex [concave] order* (denoted by $X \leq_{icx} Y[X \leq_{icv} Y]$). Roughly speaking, if $X \leq_{icx} Y$, then X is both "smaller" and "less variable" than Y in some stochastic sense.
One can also define a *decreasing convex [concave] order* (denoted by requiring (1.36) to hold for all decreasing convex [concave] functions ϕ (denoted by $\leq_{dcx}$ $[\leq_{dcv}]$). The terms "decreasing convex" and "decreasing concave" are counter intuitive in the sense that if X is smaller than Y in the sense of either of these two orders, then X is "larger" than Y in some stochastic sense. These orders can be easily characterized using the orders $\leq_{icx}$ and $\leq_{icv}$. Therefore, it is not necessary to have a separate discussion for these orders.
In analogy with Theorem 1.9 (a), the orders $\leq_{icx}$ and $\leq_{icv}$ are related to each other as follows.

Theorem 1.11. *Let X and Y be two random variables. Then*

$$X \leq_{icx} [\leq_{icv}] Y \Leftrightarrow -X \geq_{icv} [\geq_{icx}] - Y \tag{1.37}$$

Proof. Observe that a function ϕ satisfies $\phi(x)$ is increasing and convex in x if, and only if, $\psi(x) = -\phi(-x)$ is increasing and concave in x. That is, using (1.28)

$$\mathbb{E}[\phi(X)] \leq \mathbb{E}[\phi(Y)] \Leftrightarrow \mathbb{E}[\phi(-X)] \leq \mathbb{E}[\phi(-Y)] \Leftrightarrow -\mathbb{E}[\phi(-X)] \geq \\ -\mathbb{E}[\phi(-Y)] \Leftrightarrow \mathbb{E}[\psi(X)] \geq \mathbb{E}[\psi(Y)]$$

for every increasing and convex ϕ and every increasing and concave ψ. □

Note that the function ϕ, defined by $\phi(x) = x$, is increasing and is both convex and concave. Therefore, from (1.36) it follows that

$$X \leq_{icx} Y \Rightarrow \mathbb{E}[X] \leq \mathbb{E}[Y]$$

and that

$$X \leq_{icv} Y \Rightarrow \mathbb{E}[X] \leq \mathbb{E}[Y]$$

provided the expectations exist.
Let $\overline{F}$ [F] and $\overline{G}$ [G] be the survival [distribution] functions of X and Y, respectively. For a fixed a, the function ϕ_a, defined by $\phi(x) = (x-a)_+$, is increasing and convex. Therefore if $X \leq_{icx} Y$, then

$$\mathbb{E}[(X-a)_+] \leq \mathbb{E}[(Y-a)_+] \quad for\ all\ a, \tag{1.38}$$

provided the expectations exist. Using a simple integration by part we can rewrite (1.38) as (1.19), provided the integrals exist.
We may apply an analogous argument to the $\leq_{icv}$. For any real number a let a_- denote the negative part of a, that is, $a_- = a$ if $a \leq 0$ and $a_- = 0$ if $a > 0$. For a fixed a, the function ζ_a defined by $\zeta_a(x) = (x-a)_-$, is increasing and concave. Therefore, if $X \leq_{icv} Y$, then

$$\mathbb{E}[(X-a)_-] \leq \mathbb{E}[(Y-a)_-] \tag{1.39}$$

for all a, provided the expectation exist. Alternatively, again using a simple integration by parts, it is seen that (1.39) can be rewritten as

$$\int_{-\infty}^{x} F(u)du \geq \int_{-\infty}^{x} G(u)du \tag{1.40}$$

for all x, provided the integrals exist. Recall now that every increasing convex [concave] function can be approximated (that is, is a limit of) positive linear combinations of the functions ϕ_a's [ζ_a'2], for various choice of a's. By (1.19), $\mathbb{E}[\phi_a(X)] \leq \mathbb{E}[\phi_a(Y)]$ for all a, and this fact implies (1.36) in the convex case. Similarly, by (1.20), $\mathbb{E}[\zeta_a(X)] \leq \mathbb{E}[\zeta_a(Y)]$ for all a, and this fact implies (1.36) in the concave case. We thus proved the following result.

Theorem 1.12. *Let X and Y be two random variables. Then $X \leq_{icx} Y$ [$X \leq_{icv} Y$] if, and only if, (*1.19*) [(*1.40*)] holds.*

An important characterization of the increasing convex and the increasing concave orders by construction on the same probability space is stated next.

Theorem 1.13. *Two random variables X and Y satisfy $X \leq_{icx} Y$ [$X \leq_{icv} Y$] if, and only if, there exist two random variables $\hat{X}$ and $\hat{Y}$, defined on the same probability space , such that*

$$\hat{X} =_{st} X \tag{1.41}$$

$$\hat{Y} =_{st} Y \tag{1.42}$$

and $\hat{X}, \hat{Y}$ is a submartingale [$\hat{Y}, \hat{X}$ is a supermartingale], that is,

$$\mathbb{E}[\hat{Y}|\hat{X}] \geq \hat{X} \ \mathbb{E}[\hat{X}|\hat{Y}] \leq \hat{Y} \quad a.s. \tag{1.43}$$

Furthermore, the random variables $\hat{X}$ and $\hat{Y}$ can be selected such that $[\hat{Y}|\hat{X} = x]$ [$[\hat{X}|\hat{Y} = x]$] is increasing function in x in the usual stochastic order $\leq_{st}$.

Proof. It is not easy to proof the constructive part of Theorem 1.13. However, it easy to prove that if random variables $\hat{X}$ and $\hat{Y}$ as described in the Theorem exist, the $X \leq_{icx} Y$. For example, if the first inequality in (1.43) holds and if ϕ is a increasing convex function, then by Jensen's Inequality,

$$\mathbb{E}[\phi(X)] = \mathbb{E}[\phi(\hat{X})] \leq \mathbb{E}[\phi(\mathbb{E}[\hat{Y}|\hat{X}])] \leq \mathbb{E}[\mathbb{E}[\phi(\hat{Y})|\hat{X}]] = \mathbb{E}[\phi(\hat{Y})] = \mathbb{E}[\phi(Y)]$$

which is (1.36). □

Also in the $\leq_{icx}$ order we have a property similar to (1.11) that is

Theorem 1.14. *Let $X_1, X_2, ..., X_m$ be a set of independent random variables and let $Y_1, Y_2, ..., Y_m$ be another set of independent random variables. If $X_i \leq_{icx} Y_i$ for $i = 1, 2, ..., m$ then*

$$g(X_1, X_2, ..., X_m) \leq_{icx} g(Y_1, Y_2, ..., Y_m) \tag{1.44}$$

for every increasing and componentwise convex function g.

Proof. Without loss of generality we can assume that all the $2m$ random variables are independent because such an assumption does not affect the distributions of $g(X_1, X_2, ..., X_m)$ and $g(Y_1, Y_2, ..., Y_m)$. The proof is by induction on m. We start considering $m = 1$ and observing that the composition of two increasing and convex functions is still increasing and convex. It follows from the definition of $\leq_{icx}$ order that if g is increasing and convex then $\mathbb{E}[\phi(g(X))] \leq \mathbb{E}[\phi(g(Y))]$ for every increasing and convex function ϕ. That means $g(X) \leq_{icx} g(Y)$. Assume that 1.44 holds true for vectors of size $m-1$. Let g and ϕ be increasing and componentwise convex functions. Then

$$\begin{gathered}\mathbb{E}[\phi(g(X_1, X_2, ..., X_m))|X_1 = x] = \mathbb{E}[\phi(g(x, X_2, ..., X_m))] \leq \\ \mathbb{E}[\phi(g(x, Y_2, ..., Y_m))] = \mathbb{E}[\phi(g(X_1, Y_2, ..., Y_m))|X_1 = x]\end{gathered}$$

where the equalities above follow from the independence assumption and the inequality follows from the induction hypothesis. Taking expectations with respect to X_1, we obtain

$$\mathbb{E}[\phi(g(X_1, X_2, ..., X_m))] \leq \mathbb{E}[\phi(g(X_1, Y_2, ..., Y_m))]$$

Repeating the argument, but now considering on $Y_2, ..., Y_m$ and using 1.44 with $m = 1$, we see that

$$\mathbb{E}[\phi(g(X_1, Y_2, ..., Y_m))] \leq \mathbb{E}[\phi(g(Y_1, Y_2, ..., Y_m))]$$

and this proves the result. □

1.4 Peakedness Order

In this section we discuss a variability order that applies to random variables with symmetric distribution functions. It stochastically compares random variables according to their distance from their center of symmetry.
In the article written by Birnbaum "On random variable with comparable peakedness" [5] we may find more detailed properties related to variables with ordered peakedness. Let X be a random variable with a distribution function that is symmetric about μ, and let Y be another random variable with a distribution function that is symmetric about ν. Suppose that

$$|X - \mu| \leq_{st} |Y - \nu|$$

Then X is said to be *smaller than Y in the peakedness order* (denoted by $X \leq_{peak} Y$).
In the following result we state a characterization for this order when the two variables have the same mean:

Theorem 1.15. *Let X and Y be two random variables with different distribution functions, but with the same mean. Suppose that the distribution functions F and G, of X and Y, respectively, are symmetric about the common mean. Then $X \leq_{peak} Y$ if, and only if,*

$$S^{-}(G - F) = 1 \quad \textit{and the sign sequence is } +, -$$

*where S^{-} is defined as (*1.12*)*

1.5 Relations between the different orders

The increasing and convex function's set is a subset of both the increasing functions and the convex functions. It than follows that

$$X \leq_{st} Y \Rightarrow X \leq_{icx} Y \tag{1.45}$$

and

$$X \leq_{cx} Y \Rightarrow X \leq_{icx} Y \tag{1.46}$$

Generally the opposite implication is wrong. The next Theorem state an equivalence between the $\leq_{icx}$ and the $\leq_{cx}$ in the particular case of $\mathbb{E}[X] = \mathbb{E}[Y]$.

Theorem 1.16. *The following statement are equivalent:*

(i) $X \leq_{cx} Y$

(ii) $X \leq_{icx} Y$ *and* $\mathbb{E}[X] = \mathbb{E}[Y]$.

Proof. Assume (i) holds. (ii) easily follows from (1.46) and from (1.14). Let then (ii) hold. Let f be an arbitrary convex function. Assume for the moment that there is some finite α such that

$$x \mapsto f(x) + \alpha x \quad \textit{is increasing} \tag{1.47}$$

Then $\mathbb{E}[f(X)] + \alpha\mathbb{E}[X] \leq \mathbb{E}[f(Y)] + \alpha\mathbb{E}[Y]$ and hence $\mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]$, as $\mathbb{E}[X] = \mathbb{E}[Y]$. If such an α does not exist, then approximate f monotonically by

$$f_n(x) = \begin{cases} f(x) & \text{for } x \geq -n \\ f(-n) + f'_+(-n)(x+n) & \text{otherwise} \end{cases}$$

where f'_+ denotes the right derivative that always exists because of convexity. All these functions are convex and fulfill (1.47) with $\alpha = -f'_+(-n)$. Hence the assertion follows from the monotone convergence Theorem. □

Moreover a weaker implication from $\leq_{icx}$ order to $\leq_{cx}$ is stated in the following Theorem:

Theorem 1.17. *Let X and Y be two random variables such that $X \leq_{icx} Y$. Then we have that $|X| \leq_{cx} |Y|$.*

Proof. This Theorem is based on the fact that

$$\phi(x) \textit{ increasing and convex} \Rightarrow \phi(|x|) \textit{ convex}$$

□

Let now see a further, useful relation between the three stochastic orders we introduced.

Theorem 1.18. (*a*) *Two random variables X and Y satisfy $X \leq_{icx} Y$ if, and only if, there exists a random variable Z such that*

$$X \leq_{st} Z \leq_{cx} Y$$

(*b*) *Two random variables X and Y satisfy $X \leq_{icx} Y$ if, and only if, there exists a random variable Z such that*

$$X \leq_{cx} Z \leq_{st} Y$$

(*c*) *Two random variables X and Y satisfy $X \leq_{icv} Y$ if, and only if, there exists a random variable Z such that*

$$X \leq_{cv} Z \leq_{st} Y$$

(*d*) *Two random variables X and Y satisfy $X \leq_{icv} Y$ if, and only if, there exists a random variable Z such that*

$$X \leq_{st} Z \leq_{cv} Y$$

Proof. First we prove the part (a). The convex order and the usual stochastic order both imply the increasing convex order. It than follows from $X \leq_{st} Z \leq_{cx} Y$ that $X \leq_{icx} Y$. So suppose that $X \leq_{icx} Y$. Let $\hat{X}$ and $\hat{Y}$ be defined on the same probability space, as in Theorem 1.13. Define $\hat{Z} = \mathbb{E}[\hat{Y}|\hat{X}]$. It is seen that $\mathbb{E}[\hat{Y}|\hat{Z}] = \mathbb{E}[\hat{Y}|\hat{X}] = \hat{Z}$. Thus by Theorem 1.8, $\hat{Z} \leq_{cx} \hat{Y}$. Also, by Theorem 1.13, $\hat{X} \leq \hat{Z}$, and therefor by Theorem 1.2, $\hat{X} \leq_{st} \hat{Z}$. Letting $\hat{Z}$ have the same distribution as $\hat{Z}$, we obtain the stated result.
Now prove part (b). Again it is obvious that $X \leq_{cx} Z \leq_{st} Y \Rightarrow X \leq_{icx} Y$. So suppose that $X \leq_{icx} Y$. Let $\hat{X}$ and $\hat{Y}$ be defined on the same probability space, as in Theorem 1.13. Define $\hat{Z} = \hat{Y} + \hat{X} - \mathbb{E}[\hat{Y}|\hat{X}]$. Then, by Theorem 1.13, $\hat{Z} \leq \hat{Y}$, and therefore, by Theorem 1.2, $\hat{Z} \leq_{st} \hat{Y}$. Also, $\mathbb{E}[\hat{Z}|\hat{X}] = \hat{X}$, and thus, by Theorem 1.8, $\hat{X} \leq_{cx} \hat{Z}$. Letting Z have the same distribution as $\hat{Z}$, we obtain the stated result.
Part (c) and (d) can be proven similarly. Alternatively, using Theorem 1.11, part (c) can be obtained form part (a), and part (d) can be obtained form part (b). □

Chapter 2

ARCH and GARCH Models for Changing Volatility

2.1 Fundamentals of Time Series Analysis

This section provides a short summary of the essentials of classical univariate time series analysis with focus on that which is relevant for modeling risk-factor return series. The first three section are developed especially from the book written by McNeil, Rüdiger, and Embrechts "Quantitative Risk Management" [13] while in the GARCH section i considered also the Nelson's article "Stationarity and persistence in GARCH(1,1) models" [18].
A time series model for a single risk factor is a stochastic process $(X_t)_{t\in\mathbb{Z}}$ i.e. a family of random variables, indexed by the integers and defined on some probability space $(\Omega, \mathcal{F}, P)$.
Assuming they exist, we define the *mean function* $\mu(t)$ and *autocovariance function* $\gamma(t,s)$ of $(X_t)_{t\in\mathbb{Z}}$ by

$$\mu(t) = \mathbb{E}[X_t] \quad t \in \mathbb{Z}$$

$$\gamma(t,s) = \mathbb{E}[(X_t - \mu(t))(X_s - \mu(s))] \quad t,s \in \mathbb{Z}$$

It follows that the autocovariance function satisfies $\gamma(s,t) = \gamma(t,s)$ for all t, s, and $\gamma(t,t) = var(X_t)$.
Generally the processes we consider will be stationary in one or both of the following two senses.

Definition 2.1 (strict stationarity)**.** The time series $(X_t)_{t\in\mathbb{Z}}$ is strictly stationary if

$(X_{t_1}, ..., X_{t_n}) \approx (X_{t_1+k}, ..., X_{t_n+k})$, (i.e. the two vectors has the same distribution) for all $t_1, ...t_n$, $k \in \mathbb{Z}$ and for all $n \in \mathbb{N}$.

Definition 2.2 (covariance stationarity)**.** The time series $(X_t)_{t\in\mathbb{Z}}$ is *covariance stationary* (or weakly or second-order stationary) if the first two moments exist and satisfy

$$\mu(t) = \mu \quad t \in \mathbb{Z}$$

$$\gamma(t, s) = \gamma(t + k, s + k) \quad t, s, k \in \mathbb{Z}$$

The aim of these two definitions is to underline the similarities of the time series in any epoch in which we might observe it. Systematic changes in mean, variance or the covariance between equally observations are inconsistent with stationarity.
It may be easily verified that a strictly stationary time series with finite variance is covariance stationary, but it is important to note that we may define infinite-variance processes (including certain ARCH or GARCH processes) which are strictly stationary but not covariance stationary.
From Definition (2.2) we have that for all s, t we have

$$\gamma(t - s, 0) = \gamma(t, s) = \gamma(s, t) = \gamma(s - t, 0)$$

so that the covariance between X_t and X_s only depends on their temporal separation $|s-t|$, which is known as the *lag*. Thus, for a covariance-stationary process we write the autocovariance function of one variable:

$$\gamma(h) := \gamma(h, 0) \quad for\ all\ h \in \mathbb{Z}$$

Noting that $\gamma(0) = var(X_t)$, for all t, we can now define the autocorrelation function of a covariance-stationary process.

Definition 2.3 (autocorrelation function)**.** The autocorrelation function (ACF) $\rho(h)$ of a covariance-stationary process $(X_t)_{t\in\mathbb{Z}}$ is

$$\rho(h) = \rho(X_h, X_0) = \gamma(h)/\gamma(0) \quad for\ all\ h \in \mathbb{Z}$$

We speak of the autocorrelation or *serial correlation* $\rho(h)$ at lag h. In classical time series analysis the set of serial correlations and their empirical analogues estimated from data are the objects of principal interest. The study of autocorrelations is known as *analysis in the time domain*.
The basic building blocks for creating useful time series models are stationary processes without serial correlation, known as *white noise* processes and defined as follows.

Definition 2.4 (white noise). $(X_t)_{t\in\mathbb{Z}}$ is a white noise process if it is covariance stationary with autocorrelation function

$$\rho(h) = \begin{cases} 1, & \text{h} = 0 \\ 0, & \text{h} \neq 0 \end{cases}$$

A white noise process centered to have mean zero with variance σ^2=var(X_t) will be denoted WN(0,σ^2). A simple example of a white noise process is a series of iid random variables with finite variance, and this is known as a *strict white noise* process.

Definition 2.5 (strict white noise). $(X_t)_{t\in\mathbb{Z}}$ is a strict white noise process if it is a series of iid, finite-variance random variables.

A strict white noise (SWN) process centered to have mean zero and variance σ^2 will be denoted SWN(0,σ^2).
Although SWN is the easiest kind of noise process to understand, it is not the only noise that we will present. We will later see that covariance-stationary ARCH and GARCH process are in fact white noise processes.
Our further noise concept that we use, particularly when we come to discuss volatility and GARCH processes, is that of a *martingale difference sequence.* To discuss this concept we further assume that the time series $(X_t)_{t\in\mathbb{Z}}$ is adapted to some *filtration* $(\mathcal{F})_{t\in\mathbb{Z}}$ which represents the *accrual of information* over time. The sigma algebra $\mathcal{F}_t$ represents the available information at time t and typically this will be the information contained in past and present values of the time series $(X_s)_{s\leq t}$, which we refer to as the *history* up to time t and denote by $\mathcal{F}_t = \sigma(\{X_s : s \leq t\})$; the corresponding filtration is known as the *natural filtration.*
In a martingale-difference sequence the expectation of the next value, given current information, is always zero, and this property may be appropriate for financial return data. A martingale difference is often said to model our winnings in consecutive rounds of a *fair game.*

Definition 2.6 (martingale difference). The time series $(X_t)_{t\in\mathbb{Z}}$ is known as a martingale-difference sequence with respect to the filtration $(\mathcal{F})_{t\in\mathbb{Z}}$ if $\mathbb{E}[|X|] < \infty$, X_t is $\mathcal{F}_t$-measurable (*adapted*) and

$$\mathbb{E}[X_t|\mathcal{F}_{t-1}] = 0 \quad for\ all\ t \in \mathbb{Z}$$

Obviously the unconditional mean of such a process is also zero:

$$\mathbb{E}[X_t] = \mathbb{E}[\mathbb{E}[X_t|\mathcal{F}_{t-1}]] = 0 \quad for\ all\ t \in \mathbb{Z}$$

Moreover, if $\mathbb{E}[X_t^2] < \infty$ for all t, then autocovariance satisfy

$$\gamma(t,s) = \mathbb{E}[X_t X_s] = \begin{cases} \mathbb{E}[\mathbb{E}[X_t X_s|\mathcal{F}_{s-1}]] = \mathbb{E}[X_t \mathbb{E}[X_s|\mathcal{F}_{s-1}]] = 0, & \text{t} \leq \text{s} \\ \mathbb{E}[\mathbb{E}[X_t X_s|\mathcal{F}_{t-1}]] = \mathbb{E}[X_s \mathbb{E}[X_t|\mathcal{F}_{t-1}]] = 0, & \text{t} \geq \text{s} \end{cases}$$

Thus a finite-variance martingale-difference sequence has zero mean and zero covariance. If the variance is constant for all t, it is a white noise process.

2.2 ARMA process

The family of classical ARMA processes are widely used in many traditional applications of time series analysis. They are covariance-stationary processes that are constructed using white noise as a basic building block. As a general notation convention in this section we will denote noise by $(\varepsilon_t)_{t\in\mathbb{Z}}$.

Definition 2.7 (ARMA process)**.** Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be WN(0,σ_ε^2). The process $(X_t)_{t\in\mathbb{Z}}$ is a zero-mean ARMA(p,q) process if it is a covariance-stationary process satisfying difference equations of the form

$$X_t - \phi_1 X_{t-1} - ... - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon t - 1 + ... + \theta_q \varepsilon_{t-q} \quad \forall t \in \mathbb{Z} \tag{2.1}$$

(X_t) is an ARMA process with mean μ if the centered series $(X_t - \mu)_{t\in\mathbb{Z}}$ is a zero-mean ARMA(p,q) process.

Recalling Definition (2.2) we can observe that all ARMA processes are covariance stationary. Whether the process is strictly stationary or not will depend on the exact nature of the driving white noise, also known as the process of *innovations*. If the innovations are iid, or themselves form a strictly stationary process, then the ARMA process will also be strictly stationary. We now restrict our study to processes satisfying (2.1) and having a representation of the form:

$$X_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \tag{2.2}$$

where the ψ_i are coefficients which must satisfy

$$\sum_{i=0}^{\infty} |\psi_i| < \infty \tag{2.3}$$

We will call these kind of ARMA processes *causal* ARMA.

Observation 2.1. The so-called absolute summability condition (2.3) is a technical condition which ensures that $\mathbb{E}[|X_t|] < \infty$. This guarantees that the infinite sum in (2.2) converges absolutely, almost surely, meaning that both $\sum_{i=0}^{\infty} |\psi_i||\varepsilon_{t-i}|$ and $\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$ are finite with probability one.

We now verify by direct calculation that casual ARMA processes are indeed covariance stationary and calculate the form of their autocorrelation function before going on to look at some simple standard examples.

Theorem 2.1. *Any process satisfying (2.2) and (2.3) is covariance stationary with an autocorrelation function given by*

$$\rho(h) = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+|h|}}{\sum_{i=0}^{\infty} \psi_i^2} \tag{2.4}$$

Proof. Obviously, for all t we have $\mathbb{E}[X_t] = 0$ and $var(X_t) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i^2 < \infty$ due to (2.3). Moreover the autocovariances are given by

$$cov(X_t, X_{t+h}) = \mathbb{E}(X_t X_{t+h}) = \mathbb{E}[\textstyle\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \sum_{j=0}^{\infty} \psi_j \varepsilon_{t+h-j}]$$

Since (ε_t) is white noise, it follows that $\mathbb{E}[\varepsilon_{t-i}\varepsilon_{t+h-j} \neq 0 \Leftrightarrow j = i + h$, and hence that

$$\gamma(h) = cov(X_t, X_{t+h}) = \sigma_\varepsilon^2 \textstyle\sum_{i=0}^{\infty} \psi_i \psi_{i+|h|} \quad h \in \mathbb{Z}$$

which depends only on the lag h and not on t. The autocorrelation function follows easily from $\gamma(0) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i^2$. □

Example 2.1 (MA(q) process)**.** It is clear that a pure moving-average process

$$X_t = \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \tag{2.5}$$

forms a simple example of a causal process of the form (2.2). It is easily inferred from (2.3) that the autocorrelation function is given by

$$\rho(h) = \tfrac{\sum_{i=0}^{\infty} \theta_i \theta_{i+|h|}}{\sum_{i=0}^{\infty} \theta_i^2}$$

where $\theta_0 = 1$. For $|h| > q$ we have $\rho(h) = 0$ and the autocorrelation function is said to *cut off* at lag q. If this feature is observed in the estimated autocorrelations of empirical data, it is often taken as an indicator of moving-average behavior.

Example 2.2 (AR(1) process)**.** The first-order AR process satisfies the set of difference equations

$$X_t = \phi X_{t-1} + \varepsilon_t \quad \forall t \tag{2.6}$$

This process is casual if and only if $|\phi| < 1$, and this may be understood intuitively by iterating the equation (2.6) to get

$$X_t = \phi(\phi X_{t-2} + \varepsilon_{t-1}) + \varepsilon_{t-2} = \phi^{k+1} X_{t-k-1} + \textstyle\sum_{i=0}^{k} \phi^i \varepsilon_{t-i}$$

Using more careful probabilistic arguments it may be shown that the condition $|\phi| < 1$ ensures that the first term disappears as $k \to \infty$ and the second term converges. The process

$$X_t = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i} \tag{2.7}$$

turns out to be the unique solution of the defining equation (2.6). It may be easily verified that this is a process of the form (2.2) and that $\sum_{i=0}^{\infty} |\phi|^i = (1-|\phi|)^{-1}$ so that (2.3) is satisfied. Looking at the form of the solution (2.7), we see that the AR(1) process can be represented as an MA(∞) process: an infinite-order moving average process.
The autocovariance and autocorrelation functions of the process may be calculated from (2.4) and (2.2) to be

$$\gamma(h) = \tfrac{\phi^{|h|}\sigma_\varepsilon^2}{1-\phi^2}, \quad \rho(h) = \phi^{|h|}, \quad h \in \mathbb{Z}$$

Thus the ACF is exponentially decaying with possibly alternating sign.
In the case of general ARMA process, the issue of whether this process has a causal representation of the form (2.2) is resolved by the study of two polynomials in the complex plane, which are given in terms of the ARMA model parameters by

$$\tilde{\phi}(z) = 1 - \phi_1 z - ... - \phi_p z^p$$

$$\tilde{\theta}(z) = 1 + \theta_1 z + ... + \theta_q z^q$$

Provided that $\tilde{\phi}(z)$ and $\tilde{\theta}(z)$ have no common roots, then the ARMA process is a causal process satisfying (2.2) and (2.3) if and only if $\tilde{\phi}(z)$ has no roots in the unit circle $|z| \leq 1$. The coefficient ψ_i in the representation (2.2) are determined by the equation

$$\textstyle\sum_{i=0}^{\infty} \psi_i z^i = \frac{\tilde{\theta}(z)}{\tilde{\phi}(z)}$$

Example 2.3 (ARMA(1,1) process)**.** For the process given by

$$X_t - \phi X_{t-1} = \varepsilon_t + \theta\varepsilon_{t-1} \quad \forall t \in \mathbb{Z}$$

the complex polynomials are $\tilde{\phi}(z) = 1 - \phi z$ and $\tilde{\theta}(z) = 1 + \theta z$ and these have no common roots provided $\phi + \theta \neq 0$. The solution of $\tilde{\phi}(z) = 0$ is $z = \frac{1}{\phi}$ and this is outside the unit circle provided $|\phi| < 1$, so that this is the condition for causality (as in the AR(1) model of the previous example).
The representation (2.2) can be obtained by considering

$\sum_{i=0}^{\infty} \psi_i z^i = \frac{1+\theta z}{1-\phi z} = (1+\theta z)(1+\phi z+\phi^2 z^2+...), \quad |z|<1$

and is easily calculated to be

$$X_t = \varepsilon_t + (\phi+\theta)\sum_{i=0}^{\infty}\phi^{i-1}\varepsilon_{t-i} \tag{2.8}$$

Using (2.4) we may calculate that for $h \neq 0$ the ACF is

$$\rho(h) = \frac{\phi^{|h|-1}(\phi+\theta)(1+\phi\theta)}{1+\theta^2+2\phi\theta}$$

Equation (2.8) shows how the ARMA(1,1) process may be thought of as an MA(∞) process. In fact, if we impose the condition $|\theta| < 1$, we can also express (X_t) as the AR(∞) process given by

$$X_t = \varepsilon_t + (\phi+\theta)\sum_{i=0}^{\infty}(-\theta)^{i-1}X_{t-i} \tag{2.9}$$

If we rearrange this to be an equation for ε_t, then we see that we can, in a sense, reconstruct the latest innovation ε_t form the entire history of the process $(X_s)_{s\leq t}$.
The condition $|\theta| < 1$ is known as *invertibility* condition, and for the general ARMA(p,q) process the invertibility condition is that $\tilde{\theta}(z)$ should have no roots in the unit circle $|z| \leq 1$. In practice, the models we fit to real data will be both invertible and causal solution of the ARMA-defining equations.

Consider a general invertible ARMA model with non-zero mean. For what comes later it will be useful to observe that we can write such models as

$$X_t = \mu_t + \varepsilon_t \quad \mu_t = \mu + \sum_{i=1}^{p}\phi_i(X_{t-i}-\mu) + \sum_{j=1}^{q}\theta_j\varepsilon_{t-j} \tag{2.10}$$

Since we have assumed invertibility, the terms ε_{t-j}, and hence μ_t, can be written in terms of the infinite past of the process up to time $t-1$; μ_t is said to be *measurable* with respect to $\mathcal{F}_{t-1} = \sigma(\{X_s : s \leq t-1\})$.
If we make the assumption that the white noise $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a martingale-difference sequence with respect to $(\mathcal{F}_t)_{t\in\mathbb{Z}}$, then $\mathbb{E}[X_t|\mathcal{F}_{t-1}] = \mu_t$. In other words, such an ARMA process can be thought of as putting a particular structure on the conditional mean μ_t of the process. ARCH and GARCH processes will later be seen to put structure on the conditional variance $var(X_{t-1}|\mathcal{F}_t - 1)$.

2.3 ARCH processes

Definition 2.8. (ARCH process) Let $(Z_t)_{t\in\mathbb{Z}}$ be SWN(0,1). The process $(X_t)_{t\in\mathbb{Z}}$ is an ARCH(p) (Auto Regressive Conditionally Heteroscedastic) process if it is a strictly stationary and if it satisfies, for all $t \in \mathbb{Z}$ and some strictly positive-valued process $(\sigma_t)_{t\in\mathbb{Z}}$, the equations

$$X_t = \sigma_t Z_t \tag{2.11}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 \tag{2.12}$$

where $\alpha_0 > 0$ and $\alpha_i \geq 0$, $i = 1, ..., p$.

Let $\mathcal{F}_t = \sigma(\{X_s : s \leq t\})$ again denote the sigma algebra representing the history of the process up to time t so that $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ is the natural filtration Clearly, the construction (2.12) ensures that σ_t is *measurable* with respect to $\mathcal{F}_{t-1}$. This allows us to calculate that, provided $\mathbb{E}[|X_t|] < \infty$,

$$\mathbb{E}[X_t|\mathcal{F}_{t-1}] = \mathbb{E}[\sigma_t Z_t|\mathcal{F}_{t-1}] = \sigma_t \mathbb{E}[Z_t|\mathcal{F}_{t-1}] = \sigma_t \mathbb{E}[Z_t] = 0 \tag{2.13}$$

so the ARCH process has the martingale-difference property with respect to $(\mathcal{F}_t)_{t\in\mathbb{Z}}$. If the process is covariance stationary, it is simply a white noise, as discussed in the previous section.

Observation 2.2. The independence of Z_t and $\mathcal{F}_{t-1}$ that we have assumed above follows from the fact that an ARCH process must be casual, i.e. the equations (2.11) and (2.12) must have a solution of the form $X_t = f(Z_t, Z_{t-1}, ...)$ for some f so that Z_t is independent of previous values of the process.

If we simply assume that the process is a covariance-stationary white noise (for which we will give a condition in the next Theorem), then $\mathbb{E}[X_t^2] < \infty$ and

$$\text{var}(X_t|\mathcal{F}_{t-1}) = \mathbb{E}[\sigma_t^2 Z_t^2|\mathcal{F}_{t-1}] = \sigma_t^2 \text{var}(Z_t) = \sigma_t^2$$

Thus the model has the interesting property that its conditional standard deviation σ_t, or *volatility*, is a continually changing function of the previous squared values of the process. If one or more of $|X_{t-1}|, ..., |X_{t-p}|$ are particularly *large*, then X_t is effectively drawn from a distribution with large variance, and may itself be large; in this way the model generates volatility clusters.
The name ARCH refers to this structure: the model is autoregressive, since X_t clearly depends on the previous X_{t-i}, and conditionally heteroscedastic, since the conditional variance changes continually.

2.3.1 ARCH(1)

We now analyze some of the properties of the ARCH(1) model. These properties extend to the whole class of ARCH and GARCH (Generalized ARCH) models, but are easier to introduce in the simplest case.
Using $X_t^2 = \sigma_t^2 Z_t^2$ and (2.12) in the case of $p = 1$, we deduce that the squared ARCH(1) process satisfies

$$X_t^2 = \alpha_0 Z_t^2 + \alpha_1 Z_t^2 X_{t-1}^2 \tag{2.14}$$

A detailed mathematical analysis of the ARCH(1) model involves the study of equation (2.14), which is a *stochastic recurrence equation* (SRE). We would like to know when this equation has stationary solution expressed in terms of the infinite history of the innovations, i.e. solutions of the form $X_t^2 = f(Z_t, Z_{t-1}, ...)$.
For ARCH models we have to distinguish carefully between solutions that are covariance stationary and solutions that are only strictly stationary. It is possible to have ARCH(1) models with infinite variance, which obviously cannot be covariance stationary.
Equation (2.14) is a particular example of a class of recurrence equations of the form

$$Y_t = A_t Y_{t-1} + B_t \tag{2.15}$$

where $(A_t)_{t\in\mathbb{Z}}$ and $(B_t)_{t\in\mathbb{Z}}$ are sequences of iid rvs. Sufficient conditions for a solutions are

$$\mathbb{E}[max\,\{0, ln|B_t|\}] < \infty \quad and \quad \mathbb{E}[ln|A_t|] < 0 \tag{2.16}$$

where $ln^+x = max(0,\ln\, x)$. The unique solution is given by

$$Y_t = B_t + \sum_{i=1}^{\infty} B_{t-i} \prod_{j=0}^{i-1} A_{t-j} \tag{2.17}$$

where the sum converges absolutely, almost surely.
We can develop some intuition for the condition (2.16) and the form of the solution (2.17) by iterating equation (2.15) k times to obtain

$$Y_t = A_t(A_{t-1}Y_{t-2} + B_{t-1}) + B_t = B_t + \textstyle\sum_{i=1}^{k} B_{t-i} \prod_{j=0}^{i-1} A_{t-j} + Y_{t-k-1} \prod_{i=0}^{k} A_{t-i}$$

The conditions (2.16) ensure that the middle term on the right-hand side converges absolutely and the final term disappears. In particular note that:

$$\tfrac{1}{k+1} \textstyle\sum_{i=0}^{k} ln|A_{t-i}| \to \mathbb{E}[ln|A_t|] < 0 \quad a.s.$$

by the strong law of large numbers

$$\prod_{i=0}^{k} |A_{t-i}| = e^{(\sum_{i=0}^{k} ln|A_{t-i}|)} \to 0 \quad a.s.$$

which shows the importance of the $\mathbb{E}[ln|A_t|] < 0$ condition. The solution (2.17) to the standard recurrence equation is a strictly positive process (being a function of iid variables $(A_s, B_s)_{s\leq t}$), and the $\mathbb{E}[ln|A_t|] < 0$ condition turns out to be the key to the strict stationary of ARCH and GARCH models.
The squared ARCH(1) model (2.14) is a stochastic recurrence relation of the form (2.15) with $A_t = \alpha_1 Z_t^2$ and $B_t = \alpha_0 Z_t^2$. Thus the conditions (2.16) translate into the requirements that $\mathbb{E}[ln^+|\alpha_0 Z_t^2|] < \infty$, which is automatically true for the ARCH(1) process as we have defined it, and $\mathbb{E}[ln(\alpha_1 Z_t^2)] < 0$. This is the condition for a strictly stationary solution of the ARCH(1) equations and it can be shown that it is in fact a necessary and sufficient condition for strict stationary. From (2.17), the solution of the equation (2.14) takes the form

$$X_t^2 = \alpha_0 \sum_{i=0}^{\infty} \alpha_1^i \prod_{j=0}^{i} Z_{t-j}^2 \tag{2.18}$$

If the (Z_t) are standard normal innovations, then the condition for a strictly stationary solution is approximately $\alpha_1 < 3.562$; perhaps somewhat surprisingly, if the (Z_t) are scaled t innovations with fur degrees of freedom and variance 1, the condition is $\alpha_1 < 5.437$. Strict stationary depends on the distribution of the innovations but covariance stationary does not; the necessary and sufficient condition for covariance stationarity is always $\alpha_1 < 1$, as we now prove.

Theorem 2.2. *The ARCH(1) process is a covariance-stationary white noise process if and only if $\alpha_1 < 1$. The variance of the covariance-stationary process is given by $\frac{\alpha_0}{1-\alpha_1}$.*

Proof. Assuming covariance stationarity it follows from (2.14) and $\mathbb{E}[Z_t^2] = 1$ that

$$\sigma_x^2 = \mathbb{E}[X_t^2] = \alpha_0 + \alpha_1 \mathbb{E}[X_{t-1}^2] = \alpha_0 + \alpha_1 \sigma_x^2$$

Clearly, $\sigma_x = \frac{\alpha_0}{1-\alpha_1}$ and we must have $\alpha_1 < 1$.
Conversely, if $\alpha_1 < 1$, then by Jensen inequality,

$$\mathbb{E}[ln(\alpha_1 Z_t^2)] \leq ln\mathbb{E}[\alpha_1 Z_t^2)] = ln(\alpha_1) < 0$$

and we can use (2.18) to calculate that

$$\mathbb{E}[X_t^2] = \alpha_0 \sum_{i=0}^{\infty} \alpha_1^i = \frac{\alpha_0}{1-\alpha_1}$$

The process $(X_t)_{t\in\mathbb{Z}}$ is a martingale difference with a finite, non-time-dependent second moment. Hence it is white noise process. $\square$

It is clear from (2.18) that the distribution of the (X_t) in an ARCH(1) model bears a complicated relationship to the distribution of the innovations (Z_t). Even if the innovations are Gaussian, the stationary distribution of the time series is not Gaussian, but rather leptokurtic distribution with more slowly decaying tails.

Let know introduce some basic tool concerning the higher order moments of an ARCH(1) process. These conditions will be useful in the next section for the GARCH(1,1) model. More than the explicit formulas, we want to evidence how important are innovations' property in relation to the ARCH process itself.
We will see in the next chapter that there is a strict relationship between the $\leq_{icx}$ order and the kurtosis of the ordered variables.

Theorem 2.3. *For $m \geq 1$, the strictly stationary ARCH(1) process has finite moments of order $2m$ if and only if $\mathbb{E}[Z_t^{2m}] < \infty$ and $\alpha_1 < (\mathbb{E}[Z_t^{2m}])^{-1/m}$.*

Proof. We rewrite (2.18) in the form $X_t^2 = Z_t^2 \sum_{i=0}^{\infty} Y_{t,i}$ for positive random variables $Y_{t,i} = \alpha_0 \alpha_1^i \prod_{j=1}^{i} Z_{t-j}^2$, $i \geq 1$, and $Y_{t,0} = \alpha_0$. For $m \geq 1$ the following inequalities hold (the latter being Minkowsky's inequality):

$$\mathbb{E}[Y_{t,1}^m] + \mathbb{E}[Y_{t,2}^m] \leq \mathbb{E}[(Y_{t,1} + Y_{t,2})^m] \leq ((\mathbb{E}[Y_{t,1}^m])^{1/m} + (\mathbb{E}[Y_{t,2}^m])^{1/m})^m.$$

Since

$$\mathbb{E}[X_t^{2m}] = \mathbb{E}[Z_t^{2m}]\mathbb{E}[(\textstyle\sum_{i=0}^{\infty} Y_{t,i})^m]$$

it follows that

$$\mathbb{E}[Z_t^{2m}] \textstyle\sum_{i=0}^{\infty} \mathbb{E}[Y_{t,i}^m] \leq [X_t^{2m}] \leq \mathbb{E}[Z_t^{2m}](\sum_{i=0}^{\infty} (\mathbb{E}[Y_{t,i}^m])^{1/m})^m$$

Since $\mathbb{E}[Y_{t,i}^m] = \alpha_0^m \alpha_1^{im} (\mathbb{E}[]Z_t^{2m})^i$, it may be deduced that all three quantities are finite if and only if $\mathbb{E}[Z_t^{2m}] < \infty$ and $\alpha_1^m \mathbb{E}[Z_t^{2m}] < 1$ $\square$

For example for a finite fourth moment (m=2) we require $\alpha_1 < 1/\sqrt{3}$ in the case of Gaussian innovations and $\alpha_1 < 1/\sqrt{6}$ in the case of t innovations with six degrees of freedom; for t innovations with four degrees of freedom the fourth moment is undefined.
Assuming the existence of a finite fourth moment, it is easy to calculate its value, and also that of the kurtosis of the process. We square both sides of (2.14), take expectations of both sides and then solve for $\mathbb{E}[X_t^4]$ to obtain

$$\mathbb{E}[X_t^4] = \tfrac{\alpha_0^2 \mathbb{E}[Z_t^4](1-\alpha_1^2)}{(1-\alpha_1)^2(1-\alpha_1^2\mathbb{E}[Z_t^4])}$$

The kurtosis of the stationary distribution κ_X can then calculated to be

$$\kappa_X = \frac{\mathbb{E}[X_t^4]}{\mathbb{E}[X_t^2]^2} = \frac{\kappa_Z(1-\alpha_1^2)}{(1-\alpha_1^2\kappa_Z)}$$

where $\kappa_Z = \mathbb{E}[Z_t^4]$ denotes the kurtosis of the innovations. Clearly when $\kappa_Z > 1$, the kurtosis of the stationary distribution is inflated in comparison with that of the innovation distribution; for the Gaussian or t innovations $\kappa_X > 3$, so the stationary distribution is leptokurtic.

2.4 GARCH process

Definition 2.9 (GARCH(p,q))**.** Let $(Z_t)_{t\in\mathbb{Z}}$ be SWN(0,1). The process $(X_t)_{t\in\mathbb{Z}}$ is a GARCH(p,q) process if it is strictly stationary and if it satisfies, for all $t \in \mathbb{Z}$ and some strictly positive-valued process $(\sigma_t)_{t\in\mathbb{Z}}$, the equations

$$X_t = \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{2.19}$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, ..., p$, and $\beta_j \geq 0$, $j = 1, ...q$.

the GARCH process are *generalized* ARCH processes in the sense that the square volatility σ_t^2 is allowed to depend on previous squared volatilies, as well as previous squared values of the process.
In practice, low order GARCH models are most widely used and we will concentrate on the GARCH(1,1) model. In this model periods of high volatility tend to be *persistent*, since $|X_t|$ has a chance of being large if *either* $|X_{t-1}|$ is large *or* σ_{t-1} is large. It follows from (2.19) that for a GARCH(1,1) model we have

$$\sigma_t^2 = \alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta)\sigma_{t-1}^2 \tag{2.20}$$

which is again a stochastic recurrence relation of the form $Y_t = A_t Y_{t-1} + B_t$ as in (2.15). This time it is a stochastic recurrence relation for $Y_t = \sigma_t^2$ rather than X_t^2, but its analysis follows easily form the ARCH(1) case.
The conditions $\mathbb{E}[ln|A_t|] < 0$ for a strictly stationary solution of (2.15) translates to the condition $\mathbb{E}[ln(\alpha_1 Z_t^2 + \beta)] < 0$ for (2.20) and the general solution (2.17) becomes

$$\sigma_t^2 = \alpha_0 + \alpha_0 \sum_{i=1}^{\infty} \prod_{j=1}^{i} (\alpha_1 Z_{t-j}^2 + \beta) \tag{2.21}$$

If $(\sigma_t^2)_{t\in\mathbb{Z}}$ is a strictly stationary process, then so is $(X_t)_{t\in\mathbb{Z}}$, since $X_t = \sigma_t Z_t$ and $(Z_t)_{t\in\mathbb{Z}}$ is simply strict white noise. The solution of the GARCH(1,1) defining equations is then

$$X_t = Z_t \sqrt{\alpha_0 (1 + \sum_{i=1}^{\infty} \prod_{j=1}^{i} (\alpha_1 Z_{t-j}^2 + \beta))} \tag{2.22}$$

and we can use this to derive the condition for covariance stationarity.

Theorem 2.4. *The GARCH(1,1) process is a covariance-stationary white noise process if and only if $\alpha_1 + \beta < 1$. The variance of the covariance-stationary process is given by $\frac{\alpha_0}{(1-\alpha_1-\beta)}$*

Proof. We use a similar argument to Theorem (2.2) and make use of (2.22) □

Using a similar approach to Theorem (2.3) we can use (2.22) to derive conditions for the existence of higher moments of a covariance-stationary GARCH(1,1) process. For the existence of a fourth moment, a necessary and sufficient condition is that $\mathbb{E}[(\alpha_1 Z_t^2 + \beta)^2] < 1$, or alternatively that

$$(\alpha_1 + \beta) < 1 - (\kappa_Z - 1)\alpha_1^2$$

Assuming this to be true we calculate the fourth moment and kurtosis of X_t. We square both sides of (2.20) and take expectations to obtain

$$\mathbb{E}[\sigma_t^4] = \alpha_0^2 + (\alpha_1^2 \kappa_Z + \beta^2 + 2\alpha_1\beta)\mathbb{E}[\sigma_t^4] + 2\alpha_0(\alpha_1 + \beta)\mathbb{E}[\sigma_t^2]$$

Solving for $\mathbb{E}[\sigma_t^4]$, recalling that $\mathbb{E}[\sigma_t^2] = \mathbb{E}[X_t^2] = \frac{\alpha_0}{(1-\alpha_1-\beta)}$, and setting $\mathbb{E}[X_t^4] = \kappa_Z \mathbb{E}[\sigma_t^4]$, we obtain

$$\mathbb{E}[X_t^4] = \frac{\alpha_0^2 \kappa_Z (1-(\alpha_1+\beta)^2)}{(1-\alpha_1-\beta)^2(1-\alpha_1^2\kappa_Z-\beta^2-2\alpha_1\beta)}$$

from which it follows that

$$\kappa_X = \frac{\kappa_Z(1-(\alpha_1+\beta)^2)}{(1-(\alpha_1+\beta)^2-(\kappa_Z-1)\alpha_1^2)}$$

Again it is clear that the kurtosis of X_t is greater than that of Z_t, whenever $\kappa_Z > 1$, such as for Gaussian and scaled t innovations.

Higher-order ARCH and GARCH models have the same general behavior as ARCH(1) and GARCH(1,1), but their mathematical analysis becomes more tedious. The condition for a strictly stationary solution of the defining stochastic recurrence equation has been derived, but it is complicated. The necessary and sufficient condition that this solution is covariance stationary is $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$.

A squared GARCH(p,q) process has the structure

$$X_t^2 = \alpha_0 + \sum_{i=1}^{max(p,q)} (\alpha_i + \beta_i) X_{t-i}^2 - \sum_{j=1}^{q} \beta_j V_{t-j} + V_t$$

where $\alpha_i = 0$ for $i = p+1, ...q$ if $q > p$, or $\beta_j = 0$ for $j = q+1, ..., p$ if $q < p$. This resembles the ARMA(max(p,q),q) process and is formally such a process provided $\mathbb{E}[X_t^4] < \infty$.

2.4.1 GARCH with leverage

One of the main criticism of the standard ARCH and GARCH model is the rigidly symmetric way in which the volatility reacts to recent returns, regardless of their sign. Economic theory suggests that market information should have an asymmetric effect on volatility, whereby bad news leading to a fall in the equity value of a company tends to increase the volatility. This phenomenon has been called a *leverage effect*, because a fall in equity value causes an increase in the debt-to-equity ratio so-called leverage of company and should consequently make the stoke more volatile. At a less theoretical level it seems reasonable that falling stock values might lead to a higher level of investor nervousness than rises in value of the same magnitude.
One method of adding a leverage effect to a GARCH(1,1) model is by introducing an additional parameter into the volatility equation (2.14) to get

$$\sigma_t^2 = \alpha_0 + \alpha_1 (X_{t-1} + \delta |X_{t-1}|)^2 + \beta \sigma_{t-1}^2 \tag{2.23}$$

We assume that $\delta \in [-1, 1]$ and $\alpha_1 \geq 0$ as in the GARCH(1,1) model. Observe that (2.23) may be written as

$$\sigma_t^2 = \begin{cases} \alpha_0 + \alpha_1 (1+\delta)^2 X_{t-1}^2 + \beta \sigma_{t-1}^2 & X_{t-1} \geq 0 \\ \alpha_0 + \alpha_1 (1-\delta)^2 X_{t-1}^2 + \beta \sigma_{t-1}^2 & X_{t-1} < 0 \end{cases}$$

and hence that

$$\frac{\partial \sigma_t^2}{\partial X_{t-1}^2} = \begin{cases} \alpha_1 (1+\delta)^2 \sigma_{t-1}^2 & X_{t-1} \geq 0 \\ \alpha_1 (1-\delta)^2 \sigma_{t-1}^2 & X_{t-1} < 0 \end{cases}$$

The response of volatility to the magnitude of the most recent return depends on the sign of that return, and we generally expect $\delta < 0$, so bad news has the greater effect.

2.4.2 Threshold GARCH

Observe that (2.23) may easily be rewritten in the form

$$\sigma_t^2 = \alpha_0 + \tilde{\alpha}_1 X_{t-1}^2 + \tilde{\delta}\, \mathbb{I}_{\{X_{t-1} < 0\}} X_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{2.24}$$

where $\mathbb{I}$ is the indicator function, $\tilde{\alpha}_1 = \alpha_1(1+\delta)^2$ and $\tilde{\delta} = -4\delta\alpha_1$. Equation (2.24) gives the most common version of a threshold GARCH (or TGARCH) model. In effect, a threshold has been set at level zero, and at time t the dynamics depend on whether the previous value of the process X_{t-1} (or innovation Z_{t-1}) was below or above this threshold. However, it is also possible to set non-zero thresholds in TGARCH models, so this represents a more general class of model than GARCH with leverage.
In a less common version of threshold GARCH the coefficients of the GARCH effects depend on the signs of previous values of the process; this gives a first-order process of the form

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta\sigma_{t-1}^2 + \delta\, \mathbb{I}_{\{X_{t-1}<0\}}\sigma_{t-1}^2 \tag{2.25}$$

or alternatively

$$\begin{cases} \sigma_t^2 = \omega_0 + \alpha_0 X_{t-1}^2 + \beta_0\sigma_{t-1}^2, & if\ X_{t-1} > 0 \\ \sigma_t^2 = \omega_1 + \alpha_1 X_{t-1}^2 + \beta_1\sigma_{t-1}^2, & if\ X_{t-1} \le 0 \end{cases} \tag{2.26}$$

As in the standard GARCH(1,1) model we impose the non-negative constraints on all parameters to ensure the volatility to be non-negative. However, the conventional stationary conditions for GARCH model may not apply here. Since the volatility can fall into two different regimes, it is possible that conditional variance is not stationary in one regime but stationary in the other. See [21] for further details concerning the conditions on this model's parameters to guarantee stationarity and other properties. We will consider just GARCH with leverage with threshold set to zero that is easier to analyze then the general model belonging to this family of processes.

Chapter 3

Stochastic Comparison

Consider a stock price P_t at time t. We simulate its value at time $t+1$ with the formula:

$$P_{t+1} = P_t e^{X_t}$$

where X_t is the logreturn simulated with the ARCH and GARCH models. Observe that $P_{t+1} = P_0 e^{\sum_{i=0}^{t} X_i} = P_0 e^{S_t}$. We then have that the stock price is an increasing convex function of the total logreturns. Thanks to equation (1.46) and Theorem (1.14):

$$S_t \leq_{cx} \tilde{S}_t \Rightarrow S_t \leq_{icx} \tilde{S}_t \Rightarrow P_{t+1} \leq_{icx} \tilde{P}_{t+1}$$

We then have that for every option with increasing and convex payoff χ he have that

$$P_{t+1} \leq_{icx} \tilde{P}_{t+1} \Rightarrow C = \mathbb{E}[\chi(P_{t+1})] \leq \mathbb{E}[\chi(\tilde{P}_{t+1})] = \tilde{C}$$

where C is the option price.
It's then clear how important is to prove the propagation of convex order to the total logreturns as it implies the order between some option prices.
In the first four section we analyze the propagation of stochastic orders from the innovations to logreturns and total logreturns under some hypothesis. From the fifth section on we consider a model at a time and see if the previous results do apply and if there is a propagation of any order from the parameters to the logreturns and total logreturns.

3.1 General GARCH models

We want now to analyze if the stochastic orders of the innovations is propagated to the GARCH process itself and eventually how. We consider model of two different very general forms; the first model (M1) is:

$$\begin{cases} X_n = \sigma_n \varepsilon_n, \quad n = 0, 1, .. \\ \varepsilon_n \perp \sigma_n, \quad \mathbb{E}[\varepsilon_n] = 0 \\ \sigma_{n+1} = f^I(|\varepsilon_n|, \sigma_n) \end{cases} \tag{3.1}$$

with $f^I : \mathbb{R}_+^2 \to \mathbb{R}_+$, increasing and componentwise convex.
The second model (M2) is

$$\begin{cases} X_n = \sigma_n \varepsilon_n, \quad n = 0, 1, .. \\ \varepsilon_n \perp \sigma_n, \quad \mathbb{E}[\varepsilon_n] = 0 \\ \sigma_{n+1}^2 = f^{II}(\varepsilon_n^2, \sigma_n^2) \end{cases} \tag{3.2}$$

with $f^{II} : \mathbb{R}_+^2 \to \mathbb{R}_+$, increasing and componentwise convex. In both cases the innovations ε_n are independent and identically distributed. Observe that the recursive dynamics in (M1) is defined in terms of volatility σ_n, while in (M2) in terms of variance, σ_n^2 .
As we discussed in the previous chapter the usual GARCH(1,1) model is a particular case of both M1 and M2, and is defined as follows:

$$\begin{cases} X_n = \sigma_n \varepsilon_n, \quad n = 0, 1, .. \\ \sigma_{n+1}^2 = \alpha_0 + \alpha_1 X_n^2 + \beta \sigma_n^2 \end{cases} \tag{3.3}$$

With $\alpha_0, \alpha_1, \beta > 0$ and, possibly, $\alpha_1 + \beta < 1$, in order to guarantee covariance stationarity as we saw in Theorem (2.4). Both models start with a possibly random $\sigma_0 > 0$, by drawing a random ε_0.
In order to express the explicit solution of the recursive expressions of σ_n and σ_n^2, we introduce the following notations

$$\sigma_1 = g_1^I(|\varepsilon_0|, \sigma_0) = f^I(|\varepsilon_0|, \sigma_0)$$

$$\sigma_{n+1} = g_{n+1}^I(|\varepsilon_0|, |\varepsilon_1|, ..., |\varepsilon_n|, \sigma_0) = f^I(g_n^I(|\varepsilon_0|, |\varepsilon_1|, ..., |\varepsilon_{n-1}|, \sigma_0), |\varepsilon_n|) \tag{3.4}$$

for $n = 1, 2, 3, ...$, and

$$\sigma_1^2 = g_1^{II}(\varepsilon_0^2, \sigma_0^2) = f^{II}(\varepsilon_0^2, \sigma_0^2)$$

$$\sigma_{n+1}^2 = g_{n+1}^{II}(\varepsilon_0^2, \varepsilon_1^2, ..., \varepsilon_n^2, \sigma_0^2) = f^{II}(g_n^{II}(\varepsilon_0^2, \varepsilon_2, ..., \varepsilon_{n-1}^2, \sigma_0^2), \varepsilon_n^2) \tag{3.5}$$

for $n = 1, 2, 3, ...$
We now want conditions to describe monotonicity and convexity properties of σ_n or σ_n^2 in terms of the properties of f^I and f^{II}.

We then state the following general Theorem that we are going to apply in a particular case:

Theorem 3.1. *Consider the recurrence equation:*

$$X_{n+1} = f(Z_n, X_n)$$

Write $\mathbf{z}_n = (Z_0, Z_1, .., Z_n)$ *and*

$$X_0 = g_1(Z_0, X_0) = f(Z_0, X_0)$$

$$X_{n+1} = g_{n+1}(\mathbf{z}_n, X_0) = f(g_n(\mathbf{z}_{n-1}, X_0), Z_n) \quad for\ n = 1, 2, 3, ... \tag{3.6}$$

(a) If f is increasing in both variables, then all g_n are increasing in $(\mathbf{z_n}, X_0)$.

(b) If f is increasing in the second variable X and convex in the vector argument (z, X), then all g_n are convex in $(\mathbf{z}_n, x_0)$

(c) If f is increasing in the second variable X and componentwise convex in X and Z, then all g_n are componentwise convex in X_0 and all Z_i for $i = 0, ..., n$.

Proof. (a) g_0 is increasing in Z_0 by the monotonicity of f in Z. Proceeding with the induction step, since

$$g_{k+1}(Z_0, ..., Z_{k+1}, X_0) = f(g_k(Z_0, ...Z_k, X_0), Z_{k+1})$$

we get that g_{k+1} is increasing in all its argument because g_k is increasing in $(\mathbf{z_k}, X_0)$ for the induction hypothesis and f is increasing. We have then established the monotonicity of g_k for all k=0,1,2,...

(b) Again, g_0 is convex in (Z_0, X_0) by the convexity of f in (Z, X). For $0 < \lambda < 1$, assuming g_k is convex in $(\mathbf{z_k}, X_0)$, it is

$$\begin{aligned}
&g_{k+1}(\lambda \mathbf{z}'_{k+1} + (1-\lambda)\mathbf{z}''_{k+1} \ , \ \lambda X'_0 + (1-\lambda)X''_0) \\
&= f(g_k(\lambda \mathbf{z}'_k + (1-\lambda)\mathbf{z}''_k \ , \ \lambda X'_0 + (1-\lambda)X''_0) \ , \ \lambda Z'_{k+1} + (1-\lambda)Z''_{k+1}) \\
&\quad \le f(\lambda g_k(\mathbf{z}'_k, X'_0) + (1-\lambda)g_k(\mathbf{z}''_k, X''_0) \ , \ \lambda Z'_{k+1} + (1-\lambda)Z''_{k+1}) \\
&\qquad \le \lambda f(g_k(\mathbf{z}'_k, X'_0) \ , \ Z'_{k+1}) + (1-\lambda) f(g_k(\mathbf{z}''_k, X''_0) \ , Z''_{k+1}) \\
&\qquad\qquad = \lambda g_{k+1}(\mathbf{z}'_{k+1}, X'_0) + (1-\lambda) g_{k+1}(\mathbf{z}''_{k+1}, X''_0)
\end{aligned} \tag{3.7}$$

that is the convexity of g_{k+1} in $(\mathbf{z}_n, X_0)$.

(c) This point is proven similarly as (b).

□

From our hypothesis, using this Theorem we get the following:

Lemma 3.1. Let $g^I_{n+1}, g^{II}_{n+1} : \mathbb{R}^{n+2}_+ \to \mathbb{R}_+$ be defined as in (3.4) and (3.5). Then g^I_{n+1} and g^{II}_{n+1} are increasing and componentwise convex.

The Lemma's thesis follows easily from the points (a) and (c) of Theorem 3.1.
The previous Lemma is particularly useful in the case of iterative models or stochastic recurrence as it gives conditions that don't request different analysis for every iterative step.

3.2 Univariate Comparison of X_n

The aim of this section is to establish comparison results for X_n when the distributions of the innovations are changed from ε_k to $\tilde{\varepsilon}_k$. In order to get these results, the assumption that the innovations are identically distributed is not necessary, while the independence assumption is essential. In the following statement only the distribution of a single innovation ε_k will be changed, and the impact of this change on X_n will be investigated.
We will see that in the general context M1 and M2, the ordering that are naturally propagated from the innovations ε_k to X_n are the $\leq_{st}$ and the $\leq_{icx}$ ordering between absolute values or squared variables.
In order to better interpret these orders, in the next section we will see that the $\leq_{st}$ ordering between absolute values of the logreturns or squares can be seen as a *variability ordering*, while the $\leq_{icx}$ ordering between them can be interpreted as a *kurtosis* ordering.
In order to establish these results, we proceed in two steps: first we consider the volatilities σ_n and then the variables X_n. The first step is an immediate consequence of Lemma 3.1:

Theorem 3.2 (Comparison of σ_n and σ_n^2)**.**

a) Let σ_{n+1} be as in (3.1) and $|\varepsilon_k| \leq_{st} |\tilde{\varepsilon_k}|$; it follows that $\sigma_{n+1} \leq_{st} \tilde{\sigma}_{n+1}$.

b) Let σ_{n+1} be as in (3.1) and $|\varepsilon_k| \leq_{icx} |\tilde{\varepsilon_k}|$; it follows that $\sigma_{n+1} \leq_{icx} \tilde{\sigma}_{n+1}$.

c) Let σ^2_{n+1} be as in (3.2) and $\varepsilon_k^2 \leq_{st} \tilde{\varepsilon}_k^2$; it follows that $\sigma^2_{n+1} \leq_{st} \tilde{\sigma}^2_{n+1}$.

d) Let σ^2_{n+1} be as in (3.2) and $\varepsilon_k^2 \leq_{icx} \tilde{\varepsilon}_k^2$; it follows that $\sigma^2_{n+1} \leq_{icx} \tilde{\sigma}^2_{n+1}$.

Proof. Since from Lemma 3.1 in model M1 we do have (3.4) with g^I_{n+1} increasing and convex, item a) and b) follow respectively from point (b) of Theorem 1.3 and from Theorem 1.14. Similarly, since from Lemma 3.1 in model M1 we do have (3.5) with g^{II}_{n+1} increasing and convex, from the same Theorems we get c) and d). □

The comparison results for σ_n and σ_n^2 lead to the following comparisons of the variables X_n:

Theorem 3.3 (Comparison of X_n and X_n^2)**.**

a) Let X_n be as in (3.1) and $|\varepsilon_k| \leq_{st} |\tilde{\varepsilon_k}|$; it follows that $|X_n| \leq_{st} |\tilde{X}_n$.

b) Let X_n be as in (3.1) and $|\varepsilon_k| \leq_{icx} |\tilde{\varepsilon_k}|$; it follows that $|X_n| \leq_{icx} |\tilde{X}_n$.

c) Let X_n be as in (3.2) and $\varepsilon_k^2 \leq_{st} \tilde{\varepsilon}_k^2$; it follows that $X^2 \leq_{st} \tilde{X}_n^2$.

d) Let X_n be as in (3.2) and $\varepsilon_k^2 \leq_{icx} \tilde{\varepsilon}_k^2$; it follows that $X^2 \leq_{icx} \tilde{X}_n^2$.

Proof. Observe that $|X_n| = \sigma_n|\varepsilon_n|$ and $X_n^2 = \sigma_n^2\varepsilon_n^2$ with σ_n independent from ε_n. Then both $|X_n|$ and X_n^2 are increasing functions of their arguments. Using the hypothesis and the previous Theorem er have that the item a) and c) follow from Theorem 1.3. Similarly item b) and d) follow from Theorem 1.14. □

A natural question that arises at this point is if also the convex order is propagated that is if $\varepsilon_k \leq_{cx} \tilde{\varepsilon}_k \Rightarrow X_n \leq_{cx} \tilde{X}_n$. This is indeed the case of model M1 and M2 (we will prove this just for the M1 case). We start with a simple lemma:

Lemma 3.2. Let σ and $\tilde{\sigma}$ be nonnegative, with $\sigma \leq_{st} \tilde{\sigma}$. Let ε be independent from σ and $\tilde{\sigma}$, with $\mathbb{E}[\varepsilon] = 0$; then $\sigma\varepsilon \leq_{cx} \tilde{\sigma}\varepsilon$.

Proof. Using Theorem 1.2 we construct two identically distributed copies of σ and $\tilde{\sigma}$ on the same probability space $(\Omega, \mathcal{F}, P)$, such that $\sigma \leq \tilde{\sigma}$ almost surely. We then have, using Corollary 1.2 (with $a = \sigma$, $b = \tilde{\sigma}$ and $X = Y = \varepsilon$), that for every realization of σ

$$\mathbb{E}[\psi(\sigma\varepsilon)] \leq \mathbb{E}[\psi(\tilde{\sigma}\varepsilon)] \quad for\ all\ convex\ functions\ \psi$$

We can now conclude using Theorem 1.9, point (b). □

We can then state an important Theorem:

Theorem 3.4 (Propagation of Convex Order)**.** *Let X_n be as in (3.1) and $\varepsilon_k \leq_{cx} \tilde{\varepsilon}_k$; it follows that $X_n \leq_{cx} \tilde{X}_n$.*

Proof. First of all we remark that since $\varepsilon_k \leq_{cx} \tilde{\varepsilon}_k$, it follows using (1.46) that $\varepsilon_k \leq_{icx} \tilde{\varepsilon}_k$. We then have by Theorem 1.17 that $|\varepsilon_k| \leq_{cx} |\tilde{\varepsilon}_k|$ and, finally (using again (1.46)), $|\varepsilon_k| \leq_{icx} |\tilde{\varepsilon}_k|$. We now get from Theorem 3.2 (point (b)) that $\sigma_{n+1} \leq_{icx} \tilde{\sigma}_{n+1}$. From Theorem 1.18 (point (a)) there exist a random variable $\overline{\sigma}_{n+1}$ such that:

$$\sigma_{n+1} \leq_{st} \overline{\sigma}_{n+1} \leq_{cx} \tilde{\sigma}_{n+1}$$

By Lemma 2.1, we have that $\sigma_{n+1} \leq_{st} \overline{\sigma}_{n+1}$ implies $\sigma_{n+1}\varepsilon_{n+1} \leq_{cx} \overline{\sigma}_{n+1}\varepsilon_{n+1}$ (with $\varepsilon_{n+1} \perp \sigma_{n+1}$, $\varepsilon_{n+1} \perp \overline{\sigma}_{n+1}$ and $\mathbb{E}[\varepsilon_{n+1}] = 0$). On the other hand by Corollary 1.3 $\overline{\sigma}_{n+1} \leq_{cx} \tilde{\sigma}_{n+1}$ implies that $\overline{\sigma}_{n+1}\varepsilon_{n+1} \leq_{cx} \tilde{\sigma}_{n+1}\varepsilon_{n+1}$. By transitivity:

$$X_{n+1} = \sigma_{n+1}\varepsilon_{n+1} \leq_{cx} \tilde{\sigma}_{n+1}\varepsilon_{n+1} = \tilde{X}_{n+1}$$

□

3.3 The relevant orderings

In the preceding section the orderings defined by $|X| \leq_{st} |Y|$, $X^2 \leq_{st} Y^2$, $|X| \leq_{icx} |Y|$ and $X^2 \leq_{icx} Y^2$ have arisen naturally.
In order to better understand their meaning, especially concerning the innovations, in the following lemmas we identify some necessary and sufficient conditions in the continuous and symmetric case. We have the following:

Lemma 3.3. Let X and Y be symmetric with zero mean, continuous distributions F and G. The following conditions are equivalent:

(a) $X^2 \leq_{st} Y^2$;

(b) $|X| \leq_{st} |Y|$;

(c) $X \leq_{peak} Y$, where $\leq_{peak}$ is the order introduced in the first Chapter;

(d) $S^-(G-F) = 1$ with sign sequence +,-, where $S^-(G-F)$ is the number of intersections between G and F as defined in (1.12)

Proof. The equivalence of (a) and (b) is an immediate consequence of Theorem 1.3 (point (b)). The equivalence of (b) and (c) is the definition of the peakedness ordering, while the equivalence between (c) and (d) follows from Theorem 1.15 (recall that we have $\mathbb{E}[X] = \mathbb{E}[Y] = 0$). □

Lemma 3.4. Let X and Y be symmetric with continuous distribution F and G. The following conditions are equivalent:

(a) $X^2 \leq_{icx} Y^2$;

(b) $\int_x^\infty \overline{F}(u)udu \leq \int_x^\infty \overline{G}(u)udu$ for each $x \geq 0$, where $\overline{F}(u) = 1 - F(u)$ and $\overline{G}(u) = 1 - G(u)$;

(c) $\mathbb{E}[(X^2-k)^+] \leq \mathbb{E}[(Y^2-k)^+]$ for each $k \geq 0$.

Proof. $(a) \Leftrightarrow (b)$:
Observe that under our hypothesis:

$$F_{X^2}(t) = P(X^2 < t) = P(|X| < \sqrt{t}) = F(\sqrt{t}) - F(-\sqrt{t}) = \\ F(\sqrt{t}) - (1 - F(\sqrt{t}) = 2F(\sqrt{t}) - 1$$

and $\overline{F}_{X^2}(t) = 2 - 2F(\sqrt{t})$, for $t \geq 0$ (the same relations hold for G_{Y^2} and $\overline{G}$). To prove the equivalence of (a) and (b) we show the equivalence between (b) and (1.19):

$$\int_t^\infty \overline{F}_{X^2}(x)dx \leq \int_t^\infty \overline{G}_{Y^2}(x)dx \Leftrightarrow \int_t^\infty 2(1-F(\sqrt{x})dx \leq \int_t^\infty 2(1-G(\sqrt{x})dx \Leftrightarrow \\ \int_t^\infty \overline{F}(\sqrt{x})dx \leq \int_t^\infty \overline{G}(\sqrt{x})dx \Leftrightarrow \int_{\sqrt{t}}^\infty \overline{F}(u)\cdot u\cdot du \leq \int_{\sqrt{t}}^\infty \overline{G}(u)\cdot u\cdot du$$

where $u = \sqrt{x}$. We conclude using Theorem (1.12).
$(a) \Leftrightarrow (c)$:
The implication $(a) \Rightarrow (c)$ is trivial. Therefore assume that $\mathbb{E}[(X^2-k)_+] \leq \mathbb{E}[(Y^2-k)_+]$ for all $k \geq 0$, and let f be an arbitrary increasing convex function. We have to consider three cases.

I Assume for he moment that $\lim_{t\to-\infty} f(t) = 0$. It is well known that f then is the maximum of a countable set $\{l_1, l_2, ...\}$ of increasing linear functions; take, for example, the lines of support in all rational points. Now define

$$f_n(t) = max\{0, l_1(t), l_2(t), ..., l_n(t)\}.$$

Then f_n converges to f from below, and each f_n is piecewise linear with a finite number of kinks. Therefore f_n can be written as

$$f_n(x) = \sum_{i=1}^n a_{in}(x - b_{in})_+$$

for some constant $a_{in} \geq 0$ and $b_{in} \in \mathbb{R}$. Hence

$$\mathbb{E}[f_n(X^2)] = \sum_{i=1}^n a_{in}(X^2 - b_{in})_+ \leq \sum_{i=1}^n a_{in}(Y^2 - b_{in})_+ = \mathbb{E}[f_n(Y^2)].$$

Applying the monotone convergence Theorem implies $\mathbb{E}[f(X^2)] \leq \mathbb{E}[f(Y^2)]$.

II If $lim_{t\to-\infty} f(t) = \alpha \in \mathbb{R}$, then the problem can be reduced to case I by considering the function $f - \alpha$.

III Let $lim_{t\to-\infty} f(t) = -\infty$. Then $f_n(x) = max\{f(x), -n\}$ fulfills the assumptions of case 2 for all n, and f_n converges to f monotonically. Hence the assertion follows from the monotone convergence Theorem.

□

The first Lemma shows that for symmetric variables the orderings $|X| \leq_{st} |Y|$ and $X^2 \leq_{st} Y^2$ are variability comparisons equivalent to the peakedness ordering, that in this case boils down to item (d), that is the validity of a single cut condition between the distribution functions. In the typical econometric applications these orderings are however not very relevant since the innovations satisfy $\mathbb{E}[\varepsilon_k^2] = 1$, and hence $\mathbb{E}[\varepsilon_k^2] \leq_{st} \mathbb{E}[\tilde{\varepsilon}_k^2]$ would imply $\varepsilon_k^2 =_{st} \tilde{\varepsilon}_k^2$ (see Theorem 1.4).
In the normalized case the ordering $X^2 \leq_{icx} Y^2$ becomes equivalent to $X^2 \leq_{cx} Y^2$; as stated in Theorem (1.16).
We now give a sufficient and a necessary condition on the increasing and convex order:

Lemma 3.5. Let X and Y be symmetric with continuous distributions F and G and with $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 1$.

(a) If the densities of X and Y cross 4 times, with density of X being lower in the tails and in the center, and higher in the intermediate region, then $X^2 \leq_{icx} Y^2$.

(b) If $X^2 \leq_{icx} Y^2$ and X and Y have finite fourth moments, then $\beta_2(X) < \beta_2(Y)$, where β_2 is Pearson's kurtosis coefficient.

Proof.

(a) Under our hypothesis we get $f_{X^2}(t) = \frac{f(\sqrt{t})}{\sqrt{t}}$ for $t > 0$ deriving from the $F_{X^2}(t)$'s formulas obtained in the previous Lemma. Since X and Y are symmetrical, we have that the four intersection points between the densities f and g are symmetrical with respect to the origin. Hence the densities of X^2 and Y^2 cross in two points and since $\mathbb{E}[X^2] = \mathbb{E}[Y^2]$ from Theorem 1.10 we have that $X^2 \leq_{cx} Y^2$ (that implies the increasing convex order).

(b) In our case $\beta_2(X) = \mathbb{E}[X^4]$ and hence the thesis follows from the definition of the convex order.

□

This Lemma shows that the comparison $X^2 \leq_{icx} Y^2$ can be interpreted as a classical kurtosis ordering; the crossing condition is usually referred in kurtosis ordering literature as a Dyson-Finucan condition. You can see the article "A note on kurtosis" by Finucan [9] for further details.

3.4 Convex Comparison for Total Logreturns

In financial applications the variables X_n typically represent logreturns, that are additive quantities. The over-the-period total return is given by $S_n = \sum_{k=1}^{n} X_k$. It is therefore natural to ask if some of the comparison results obtained in the second section of this chapter. We now consider the case of convex order, that is, we wonder if $\varepsilon_k \leq_{cx} \tilde{\varepsilon}_k \Rightarrow S_n \leq_{cx} \tilde{S}_n$.
The problem is not trivial since S_n cannot be expressed as a sum of independent variables, so the standard results about convex ordering of sums cannot be applied; we are able to prove a positive result in the case of model M1 and for symmetric innovations. We start with a basic lemma:

Lemma 3.6. Let $\phi \in C^2(\mathbb{R})$ be convex and $g_i \in C^2(\mathbb{R})$ be *convex and nonnegative.* Let $a, b \in \mathbb{R}$ and $P_m := \{-1, 1\}^m$. It follows that

$$h(u) = \sum_{\underline{p}\in P_m} \phi(a + bu + \sum_{i=1}^{m} p_i g_i(u))$$

is convex.

Proof. We can compute:

$$h'(u) = \sum_{\underline{p}\in P_m} \phi'(a + bu + \sum_{i=1}^{m} p_i g_i(u)) \;\cdot\; (b + \sum_{i=1}^{m} p_i g_i'(u))$$

$$h''(u) = \sum_{\underline{p}\in P_m} \phi''(a + bu + \sum_{i=1}^{m} p_i g_i(u)) \;\cdot\; (b + \sum_{i=1}^{m} p_i g_i'(u))^2 \quad + \\ \sum_{\underline{p}\in P_m} \phi'(a + bu + \sum_{i=1}^{m} p_i g_i(u)) \;\cdot\; (\sum_{i=1}^{m} p_i g_i''(u))$$

Let start considering the term before the plus. It's surely positive in fact it is composed by the second derivative of ϕ that is positive by hypothesis and by a square. Let consider now the second term is given by

$$A_m = \sum_{\underline{p}\in P_m} \phi'(a + bu + \sum_{i=1}^{m} p_i g_i(u)) \;\cdot\; (\sum_{i=1}^{m} p_i g_i''(u))$$

Let us denote with $\underline{P}$ a random vector with a discrete uniform distribution on P_m; clearly $\mathbb{E}[\underline{P}] = \underline{0}$ and the components of $\underline{P}$ are independent. Notice that A_m is the sum of all possible values of the function $f(\underline{P}) = \phi'(a + bu + \underline{g}(u) \cdot \underline{P})(\underline{g}''(u) \cdot \underline{P})$ with $\underline{P} \in P_m$. We can then write it as the expected value of $f(\underline{P})$ if we add to every coefficient of the summation its probability i.e. $\frac{1}{2^m}$. We then get

$$A_m = 2^m \mathbb{E}[\phi'(a + bu + \underline{g}(u) \cdot \underline{P})(\underline{g}''(u) \cdot \underline{P})]$$

Since the functions $\phi'(a + bu + \underline{g}(u) \cdot \underline{P})$ and $(\underline{g}''(u) \cdot \underline{P})$ are componentwise increasing in $\underline{P}$ (that is their covariance is positive), from the covariance inequality $(0 \leq Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y])$ it follows that

$$\begin{aligned} A_m &= 2^m \mathbb{E}[\phi'(a + bu + \underline{g}(u) \cdot \underline{P})(\underline{g}''(u) \cdot \underline{P})] \\ &\geq 2^m \mathbb{E}[\phi'(a + bu + \underline{g}(u) \cdot \underline{P})]\ \mathbb{E}[(\underline{g}''(u) \cdot \underline{P})] = 0 \end{aligned}$$

since $\mathbb{E}[\underline{P}] = 0$. □

We remark that in this Lemma the smoothness requirements on ϕ and on the g_i can be dropped; we preferred this formulation in order to simplify the proof. Since in this section we consider only model M1, we define

$$g_n(\varepsilon_0, \varepsilon_1, ..., \varepsilon_{n-1}, \sigma_0) := g_n^I(|\varepsilon_0|, |\varepsilon_1|, ..., |\varepsilon_{n-1}|, \sigma_0)$$

from Lemma 3.1, it is clear that g_n is even and ccx. We have

$$\begin{aligned} S_n = X_0 + X_1 + ... + X_n = \sigma_0\varepsilon_0 + \sigma_1\varepsilon_1 + ... + \sigma_n\varepsilon_n = \\ = \sigma_0\varepsilon_0 + g_1(\varepsilon_0, \sigma_0)\varepsilon_1 + ... + g_n(\varepsilon_0, ..., \varepsilon_{n-1}, \sigma_0)\varepsilon_n = \\ = S_n(\varepsilon_0, \varepsilon_1, ..., \varepsilon_n, \sigma_0) \end{aligned} \tag{3.8}$$

The main problem is proving the propagation of convexity to the sums is that S_n in sot a componentwise convex function of the innovations ε_k; indeed, each g_k in (3.8) is multiplied by a possibly negative innovation ε_k. This prevents the applications of standard results and requires the development of a specific technique based on Lemma 3.6. The basic idea is that in the case of symmetric innovations it is possible to restore the convexity by averaging over all the possible sign changes, as in Lemma 3.6. This will be done in a recursive way; we start with the following:

Lemma 3.7. Let X_n and S_n be as in (3.1) and (3.8). Let ϕ be convex and ε_i be symmetric. Then the function

$$h(\varepsilon_0, ..., \varepsilon_k, \sigma_0) := \mathbb{E}_{\varepsilon_{k+1}, ..\varepsilon_n}[\phi(S_n(\varepsilon_0, ..., \varepsilon_n, \sigma_0))] \tag{3.9}$$

is convex in ε_k for each fixed value of $\varepsilon_0, ...\varepsilon_{k-1}, \sigma_0$.

Proof. To avoid notional burdening we drop the arguments of the functions g_i. Since the innovations are symmetric and g_i is even, we can write:

$$\begin{aligned} \mathbb{E}_{\varepsilon_{k+1}, ..., \varepsilon_n}[\phi(S_n(\varepsilon_0, ...\varepsilon_n, \sigma_0))] = \mathbb{E}_{\varepsilon_{k+1}, ..., \varepsilon_n}[\phi(\sigma_0\varepsilon_0 + ... + g_n\varepsilon_n)] = \\ \mathbb{E}_{\varepsilon_{k+1}, ..., \varepsilon_n}[\frac{1}{2^{n-k}} \sum_{\underline{p} \in P_{n-k}} \phi(\sigma_0\varepsilon_0 + ... + p_1 g_{k+1}\varepsilon_{k+1} + ... + p_{n-k} g_n \varepsilon_n) 1_{\varepsilon_{k+1} \geq 0, ..., \varepsilon_n \geq 0}] \end{aligned} \tag{3.10}$$

the second equality is the key of the all proof. In fact we have completely eliminated from the expected value the case of negative innovations. It might

not be clear how we used the hypothesis on g_i. Observe that if we take for example $p_j = -1$ (with $k+1 < j < n$), that correspond to consider $-\varepsilon_h$ (that is negative), we should have as argument of the g_k, $k > j$, $-\varepsilon_h$. Instead in the previous formula we have ε_h. This is allowed just in our case, i.e. with g_i even.
Denoting by

$$\overline{h}(\varepsilon_0, ..., \varepsilon_k, ..., \varepsilon_n, \sigma_0) = \frac{1}{2^{n-k}} \sum_{\underline{p} \in P_{n-k}} \phi(\sigma_0\varepsilon_0 + g_1\varepsilon_1 + ... + g_k\varepsilon_k + p_1 g_{k+1}\varepsilon_{k+1} + ... + p_{n-k} g_n \varepsilon_n)$$

we have that

$$h(\varepsilon_0, ..., \varepsilon_k, \sigma_0) = \mathbb{E}_{\varepsilon_{k+1},...,\varepsilon_n}[1_{\varepsilon_k+1 \geq 0, ..., \varepsilon_n \geq 0}\ \overline{h}(\varepsilon_0, ..., \varepsilon_k, ...\varepsilon_n, \sigma_0)]$$

and $\overline{h}$ is convex in ε_k from Lemma 3.6 ($u = \varepsilon_k$, $a = \sigma_0\varepsilon_0 + g_1\varepsilon_1...$, $b = g_k$ and with $g_i = g_{i+k}\varepsilon_{i+k}$). It follows that also $h(\varepsilon_0, ..\varepsilon_k, \sigma_0)$ is convex in ε_k for each value of $\sigma_0, \varepsilon_0, ..., \varepsilon_{k-1}$. □

We can finally state the result on the propagation of the convex order to S_n:

Theorem 3.5. *Let X_n and S_n be as in (3.1) and (3.8). Let ε_i be symmetric. If also $\tilde{\varepsilon}_k$ is symmetric and $\varepsilon_k \leq_{cx} \tilde{\varepsilon}_k$, then $\tilde{S}_n := S_n(\varepsilon_0, ..., \tilde{\varepsilon}_k, ..., \varepsilon_n, \sigma_0) \geq_{cx} S_n(\varepsilon_0, ..., \varepsilon_k, ..., \varepsilon_n, \sigma_0)$.*

Proof. Let ϕ be convex. From the independence of the ε_i we can write

$$\begin{aligned}\mathbb{E}[\phi(\tilde{S}_n)] &= \mathbb{E}_{\varepsilon_0,...,\varepsilon_{k-1}}\mathbb{E}_{\tilde{\varepsilon}_k}\mathbb{E}_{\varepsilon_{k+1},...,\varepsilon_n}[\phi(S_n(\varepsilon_0, ..., \tilde{\varepsilon}_k, ..., \varepsilon_n, \sigma_0))] = \\ &= \mathbb{E}_{\varepsilon_0,...,\varepsilon_{k-1}}\mathbb{E}_{\tilde{\varepsilon}_k}[h(\varepsilon_0, ..., \varepsilon_{k-1}, \tilde{\varepsilon}_k, \sigma_0)]\end{aligned}$$

where as in (3.9)

$$h(\varepsilon_0, ..., \tilde{\varepsilon}_k, \sigma_0) := \mathbb{E}_{\varepsilon_{k+1},..\varepsilon_n}[\phi(S_n(\varepsilon_0, ..., \varepsilon_{k-1}, \tilde{\varepsilon}_k, \varepsilon_{k+1}, ..., \varepsilon_n, \sigma_0))]$$

is a convex function of $\tilde{\varepsilon}_k$ for each value of $\sigma_0, \varepsilon_1, \varepsilon_{k-1}$ from Lemma 3.7.
Since $\tilde{\varepsilon}_k \geq_{cx} \varepsilon_k$ it follows that

$$\mathbb{E}_{\tilde{\varepsilon}_k}[h(\varepsilon_0, ..., \varepsilon_{k-1}, \tilde{\varepsilon}_k, \sigma_0)] \geq \mathbb{E}_{\varepsilon_k}[h(\varepsilon_0, ..., \varepsilon_{k-1}, \varepsilon_k, \sigma_0)]$$

that gives

$$\begin{aligned}&\mathbb{E}[\phi(\tilde{S}_n)] = \mathbb{E}_{\varepsilon_0,...,\varepsilon_{k-1}}\mathbb{E}_{\tilde{\varepsilon}_k}[h(\varepsilon_0, ..., \varepsilon_{k-1}, \tilde{\varepsilon}_k, \sigma_0)] \geq \\ &\geq \mathbb{E}_{\varepsilon_0,...,\varepsilon_{k-1}}\mathbb{E}_{\varepsilon_k}[h(\varepsilon_0, ..., \varepsilon_{k-1}, \varepsilon_k, \sigma_0)] = \mathbb{E}[\phi(S_n)]\end{aligned}$$

that is $\tilde{S}_n \geq_{cx} S_n$. □

3.5 ARCH(q) case

Before dealing with the more complicated GARCH and GARCH with leverage case, let see the case of the ARCH(q) model:

$$\begin{cases} X_n = \sigma_n \varepsilon_n, \quad n = 0, 1, ... \\ \sigma_n^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i X_{t-i}^2 \quad n = q, q+1, ... \end{cases} \tag{3.11}$$

The main difference from this model and the M1 and M2 ones is that the so called f function here don not depend directly just from the previous value of the innovation but also from the previous p values. Despite this little difference we can see this model as the M2 model but directly with the g function:

$$\begin{cases} X_n = \sigma_n \varepsilon_n, \quad n = 0, 1, .. \\ \sigma_n^2 = g_n^{ARCH}(\sigma_0, ..., \sigma_{p-1}, \varepsilon_p^2, ..., \varepsilon_{n-1}^2, \alpha_0, \alpha_1, ..., \alpha_p) \quad n = q, q+1, .. \end{cases} \tag{3.12}$$

The g^{ARCH} function has a complicated but explicit formulation and it is obviously increasing and componentwise convex.
We can then apply the same arguments of the previous sections to prove that Theorems (3.3), (3.4) and (3.5) do hold true also in the ARCH(q) case.
As we are going to do in the following section we now want to see if some ordering of the parameters is propagated to the logreturns and total logreturns. Considering stochastic parameters we may then state the following theorem:

Theorem 3.6. *Let X_n be as in (3.11). If we consider random parameters $\alpha_0 \leq_{st} \tilde{\alpha}_0$, $\alpha_i \leq_{st} \tilde{\alpha}_i$ for $i = 1...p$, then $|X_n| \leq_{st} |\tilde{X}_n|$, $X_n^2 \leq_{st} \tilde{X}_n^2$ and $X_n \leq_{cx} \tilde{X}_n$.*

Proof. Since σ_n and σ_n^2 are increasing functions of the parameters, if $\alpha_0 \leq_{st} \tilde{\alpha}_0$, $\alpha_i \leq_{st} \tilde{\alpha}_i$ for $i = 1...p$, it follows that $\sigma_n \leq_{st} \tilde{\sigma}_n$ and $\sigma_n^2 \leq_{st} \tilde{\sigma}_n^2$. As in the proof of Theorem 3.3 it follows that $|X_n| \leq_{st} |\tilde{X}_n|$, $X_n^2 \leq_{st} \tilde{X}_n^2$. From Lemma 3.2, $\sigma_n \leq_{st} \tilde{\sigma}_n$ implies that $X_n \leq_{cx} \tilde{X}_n$. □

We then want to see if some order is propagated to the logreturns sums. We do that just in the case of symmetric innovations. The proof is quite similar to the one we are going to apply in the GARCH case.

Theorem 3.7. *Let X_n be as in (3.11) and S_n be as in (3.8). Let ε_i be symmetric. If we consider random parameters $\alpha_0 \leq_{st} \tilde{\alpha}_0$, $\alpha_i \leq_{st} \tilde{\alpha}_i$ for $i = 1...p$, then $S_n \leq_{cx} \tilde{S}_n$.*

Proof. As before, we write

$$S_n = \sigma_0\varepsilon_0 + ... + \sigma_{q-1}\varepsilon_{q-1} + g_q^{ARCH}(\varepsilon_0, ..., \varepsilon_{q-1}, \sigma_0, ..., \sigma_{q-1}, \alpha_0, \alpha_1, ...\alpha_q)\varepsilon_q + \\ ... + g_n^{ARCH}(\varepsilon_0, ..., \varepsilon_{n-1}, \sigma_0, ..., \sigma_{q-1}, \alpha_0, \alpha_1, ...\alpha_q)\varepsilon_n$$

where the functions g_i are nondecreasing in the parameters α_0, α_i for $i = 1...q$, and even in respect to ε_i. Let ϕ be any convex function. We first want to prove that $\mathbb{E}[\phi(S_n)]$ is nondecreasing in the parameters. From the symmetry of the innovations ε_i we can write:

$$\mathbb{E}[\phi(S_n)] = \mathbb{E}_{\varepsilon_0,...,\varepsilon_n}[\phi(\sigma_0\varepsilon_0 + ... + g_n^{ARCH}\varepsilon_n)] = \\ = \mathbb{E}_{\varepsilon_0,...,\varepsilon_n}[\tfrac{1}{2^{n+1}}\textstyle\sum_{\underline{p}\in P_{n+1}} \phi(\sigma_0 p_0\varepsilon_0 + ... + p_n g_n^{ARCH}\varepsilon_n)1_{\varepsilon_0\geq 0,...,\varepsilon_n\geq 0}]$$

Denoting by

$$\overline{h}(\varepsilon_0, ...\varepsilon_n, \sigma_0, \alpha_0, \alpha_1, \beta) = \\ \tfrac{1}{2^{n+1}}\textstyle\sum_{\underline{p}\in P_{n+1}} \phi(\sigma_0 p_0\varepsilon_0 + ... + p_n g_n^{ARCH}\varepsilon_n)1_{\varepsilon_0\geq 0,...,\varepsilon_n\geq 0},$$

we see that $\overline{h}$ is nondecreasing in the parameters:

$$\tfrac{\partial\overline{h}}{\partial\alpha_0} = \tfrac{1}{2^{n+1}}\textstyle\sum_{\underline{p}\in P_{n+1}} \phi'(\sigma_0 p_0\varepsilon_0 + ... + p_n g_n^{ARCH}\varepsilon_n) \cdot (p_q\varepsilon_q g_q'^{ARCH} + ... + \\ p_n\varepsilon_n g_n'^{ARCH})\ 1_{\varepsilon_0\geq 0,...,\varepsilon_n\geq 0} \geq 0$$

from the multivariate covariance inequality, as in the proof of Lemma 3.6. In fact we have that

$$\tfrac{\partial\phi'(..)}{\partial p_i} = \phi''(...)(\sigma_i\varepsilon_i)\ 1_{\varepsilon_i\geq 0} \geq 0 \quad \text{for } i \leq q-1 \\ \tfrac{\partial\phi'(..)}{\partial p_i} = \phi''(...)(g_i^{ARCH}\varepsilon_i)\ 1_{\varepsilon_i\geq 0} \geq 0 \quad \text{for } i \geq q \\ \tfrac{\partial(p_1\varepsilon_1 g_1'^{ARCH} + ... + p_n\varepsilon_n g_n'^{ARCH})}{\partial p_i} = \varepsilon_i g_i'^{ARCH}\ 1_{\varepsilon_i\geq 0} \geq 0 \\ \mathbb{E}_{\underline{p}}[p_1\varepsilon_1 g_1'^{ARCH} + ... + p_n\varepsilon_n g_n'^{ARCH}] = 0$$

where $g'^{ARCH} = \frac{\partial g^{ARCH}}{\partial\alpha_0} \geq 0$.
The same reasoning shows that $\frac{\partial\overline{h}}{\partial\alpha_i} \geq 0$.
It follows that $\mathbb{E}[\phi(S_n)]$ is nondecreasing in α_0, α_i; but then if $\alpha_0 \leq_{st} \tilde{\alpha}_0$, $\alpha_i \leq_{st} \tilde{\alpha}_i$ for $i = 1...q$,

$$\mathbb{E}[\phi(S_n(\alpha_0, \alpha_1, ..., \alpha_q))] \leq \mathbb{E}[\phi(S_n(\tilde{\alpha}_0, \tilde{\alpha}_1, ..., \tilde{\alpha}_q))]$$

that is $S_n \leq_{cx} \tilde{S}_n$. □

3.6 The GARCH (1,1) case

We focus now on the GARCH (1,1) model specified by

$$\begin{cases} X_n = \sigma_n \varepsilon_n, \quad n = 0, 1, .. \\ \sigma_{n+1}^2 = \alpha_0 + \alpha_1 X_n^2 + \beta \sigma_n^2 \end{cases} \tag{3.13}$$

with $\alpha_0, \alpha_1, \beta > 0$. For this model the recursive dynamic of the volatility or of the variance (3.5) can easily be explicated as follows:

$$\sigma_{n+1}^2 = \sigma_0^2 \prod_{i=1}^{n+1} (\beta + \alpha_1 \varepsilon_{n-i+1}^2) + \alpha_0 [1 + \sum_{k=1}^{n} \prod_{i=1}^{k} (\beta + \alpha_1 \varepsilon_{n-i+1}^2)] \tag{3.14}$$

From this expression it is immediate that σ_{n+1}^2 and σ_{n+1} are nondecreasing functions of the parameters α_0, α_1 and β (for more detailed computations consider the next section with $\delta = 0$).
We already remarked that this model is a special case of both M1 and M2, so all the comparison result for varying innovations of the preceding section do hold. In this section we are interested in establishing comparison results for different parameters $\alpha_0, \alpha_1, \beta$. As mentioned in the introduction, it is natural that an increase in $\alpha_0, \alpha_1, \beta$ should correspond to an increase in the variability of X_n and S_n; in this section we prove it rigorously. Without any additional effort, we can consider stochastic parameters $\alpha_0, \alpha_1, \beta$:

Theorem 3.8. *Let X_n be as in (3.13). If we consider random parameters $\alpha_0 \leq_{st} \tilde{\alpha}_0$, $\alpha_1 \leq_{st} \tilde{\alpha}_1$ and $\beta \leq_{st} \tilde{\beta}$, then $|X_n| \leq_{st} |\tilde{X}_n|$, $X_n^2 \leq_{st} \tilde{X}_n^2$ and $X_n \leq_{cx} \tilde{X}_n$.*

Proof. Since σ_n and σ_n^2 are increasing functions of the parameters, if $\alpha_0 \leq_{st} \tilde{\alpha}_0$, $\alpha_1 \leq_{st} \tilde{\alpha}_1$ and $\beta \leq_{st} \tilde{\beta}$, it follows that $\sigma_n \leq_{st} \tilde{\sigma}_n$ and $\sigma_n^2 \leq_{st} \tilde{\sigma}_n^2$. As in the proof of Theorem (3.3) it follows that $|X_n| \leq_{st} |\tilde{X}_n|$, $X_n^2 \leq_{st} \tilde{X}_n^2$. From Lemma (3.2), $\sigma_n \leq_{st} \tilde{\sigma}_n$ implies that $X_n \leq_{cx} \tilde{X}_n$. □

The last point is to prove the convex comparison of the sums S_n; again, this is nontrivial since the X_n are not independent; we provide a proof in the case of symmetric innovations.

Theorem 3.9. *Let X_n be as in (3.13) and S_n be as in (3.8). Let ε_i be symmetric. If we consider random parameters $\alpha_0 \leq_{st} \tilde{\alpha}_0$, $\alpha_1 \leq_{st} \tilde{\alpha}_1$ and $\beta \leq_{st} \tilde{\beta}$, then $S_n \leq_{cx} \tilde{S}_n$.*

Proof. As before, we write

$$S_n = \sigma_0\varepsilon_0 + g_1(\varepsilon_0, \sigma_0, \alpha_0, \alpha_1, \beta)\varepsilon_1 + ... + g_n(\varepsilon_0, ..., \varepsilon_{n-1}, \alpha_0, \alpha_1, \beta)\varepsilon_n$$

where the functions g_i are nondecreasing in the parameters $\alpha_0, \alpha_1, \beta$ and even in respect to ε_i. Let ϕ be any convex function. We first want to prove that $\mathbb{E}[\phi(S_n)]$ is nondecreasing in the parameters $\alpha_0, \alpha_1, \beta$. From the symmetry of the innovations ε_i we can write:

$$\mathbb{E}[\phi(S_n)] = \mathbb{E}_{\varepsilon_0,...,\varepsilon_n}[\phi(\sigma_0\varepsilon_0 + ... + g_n\varepsilon_n)] =$$
$$= \mathbb{E}_{\varepsilon_0,...,\varepsilon_n}[\tfrac{1}{2^{n+1}} \textstyle\sum_{\underline{p}\in P_{n+1}} \phi(\sigma_0 p_0\varepsilon_0 + ... + p_n g_n\varepsilon_n) 1_{\varepsilon_0\geq 0,...,\varepsilon_n\geq 0}]$$

Denoting by

$$\overline{h}(\varepsilon_0, ...\varepsilon_n, \sigma_0, \alpha_0, \alpha_1, \beta) = \tfrac{1}{2^{n+1}} \textstyle\sum_{\underline{p}\in P_{n+1}} \phi(\sigma_0 p_0\varepsilon_0 + ... + p_n g_n\varepsilon_n)\ 1_{\varepsilon_0\geq 0,...,\varepsilon_n\geq 0}$$

we see that $\overline{h}$ is nondecreasing in $\alpha_0, \alpha_1, \beta$; indeed we can compute:

$$\tfrac{\partial\overline{h}}{\partial\alpha_0} =$$
$$\tfrac{1}{2^{n+1}} \textstyle\sum_{\underline{p}\in P_{n+1}} \phi'(\sigma_0 p_0\varepsilon_0 + ... + p_n g_n\varepsilon_n) \cdot (p_1\varepsilon_1 g_1' + ... + p_n\varepsilon_n g_n')\ 1_{\varepsilon_0\geq 0,...,\varepsilon_n\geq 0} \geq 0$$

from the multivariate covariance inequality, as in the proof of Lemma 3.6. In fact we have that

$$\tfrac{\partial\phi'(..)}{\partial p_i} = \phi''(...)(g_i\varepsilon_i)\ 1_{\varepsilon_i\geq 0} \geq 0$$
$$\tfrac{\partial(p_1\varepsilon_1 g_1' + ... + p_n\varepsilon_n g_n')}{\partial p_i} = \varepsilon_i g_i'\ 1_{\varepsilon_i\geq 0} \geq 0$$
$$\mathbb{E}_{\underline{p}}[p_1\varepsilon_1 g_1' + ... + p_n\varepsilon_n g_n'] = 0$$

where $g' = \frac{\partial g}{\partial\alpha_0} \geq 0$.
The same reasoning shows that $\frac{\partial\overline{h}}{\partial\alpha_1} \geq 0$ and $\frac{\partial\overline{h}}{\partial\beta} \geq 0$.
It follows that $\mathbb{E}[\phi(S_{\underline{n}})]$ is nondecreasing in $\alpha_0, \alpha_1, \beta$; but then if $\alpha_0 \leq_{st} \tilde{\alpha}_0$, $\alpha_1 \leq_{st} \tilde{\alpha}_1$ and $\beta \leq_{st} \tilde{\beta}$,

$$\mathbb{E}[\phi(S_n(\alpha_0, \alpha_1, \beta))] \leq \mathbb{E}[\phi(S_n(\tilde{\alpha}_0, \tilde{\alpha}_1, \tilde{\beta}))]$$

that is $S_n \leq_{cx} \tilde{S}_n$. □

3.7 The GARCH with leverage case

As we saw in the previous chapter we may consider the leverage effect in the GARCH model with a simple change in the formula (3.13):

$$\begin{cases} X_n = \sigma_n\varepsilon_n, \quad n = 0, 1, .. \\ \sigma_t^2 = \alpha_0 + \alpha_1(X_{t-1} + \delta|X_{t-1}|)^2 + \beta\sigma_{t-1}^2 \end{cases} \tag{3.15}$$

We now rewrite this formula and the usual GARCH one, to better see their common aspects

$$\begin{aligned}\sigma_t^2 = \alpha_0 + \alpha_1(X_{t-1} + \delta|X_{t-1}|)^2 + \beta\sigma_{t-1}^2 = \\ = \alpha_0 + \alpha_1(X_{t-1} + \delta\sigma_{t-1}|\varepsilon_{t-1}|)^2 + \beta\sigma_{t-1}^2 = \\ = \alpha_0 + \alpha_1 X_{t-1}^2 + \alpha_1\delta^2\sigma_{t-1}^2\varepsilon_{t-1}^2 + 2\alpha_1 X_{t-1}\delta\sigma_{t-1}|\varepsilon_{t-1}| + \beta\sigma_{t-1}^2 = \\ = \alpha_0 + \alpha_1 X_{t-1}^2 + \alpha_1\delta^2\sigma_{t-1}^2\varepsilon_{t-1}^2 + 2\alpha_1\varepsilon_{t-1}\delta\sigma_{t-1}^2|\varepsilon_{t-1}| + \beta\sigma_{t-1}^2 = \\ = \alpha_0 + \alpha_1 X_{t-1}^2 + \sigma_{t-1}^2(\alpha_1\delta^2\varepsilon_{t-1}^2 + 2\alpha_1\varepsilon_{t-1}\delta|\varepsilon_{t-1}| + \beta) = \\ = \alpha_0 + \sigma_{t-1}^2(\alpha_1\varepsilon_{t-1}^2 + \alpha_1\delta^2\varepsilon_{t-1}^2 + 2\alpha_1\varepsilon_{t-1}\delta|\varepsilon_{t-1}| + \beta)\end{aligned} \quad (3.16)$$

Let now see the usual GARCH

$$\begin{aligned}\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta\sigma_{t-1}^2 = \\ = \alpha_0 + \alpha_1\sigma_{t-1}^2\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 = \\ = \alpha_0 + \sigma_{t-1}^2(\alpha_1\varepsilon_{t-1}^2 + \beta)\end{aligned} \quad (3.17)$$

It's now easier to see the analogies with (3.14). The leverage GARCH closed formula is then:

$$\begin{aligned}\sigma_{n+1}^2 = \sigma_0^2\prod_{i=1}^{n+1}(\alpha_1\varepsilon_{n-i+1}^2 + \alpha_1\delta^2\varepsilon_{n-i+1}^2 + 2\alpha_1\varepsilon_{n-i+1}\delta|\varepsilon_{n-i+1}| + \beta) + \\ + \alpha_0[1 + \sum_{k=1}^{n}\prod_{i=1}^{k}(\alpha_1\varepsilon_{n-i+1}^2 + \alpha_1\delta^2\varepsilon_{n-i+1}^2 + 2\alpha_1\varepsilon_{n-i+1}\delta|\varepsilon_{n-i+1}| + \beta)]\end{aligned} \quad (3.18)$$

We now have to prove that σ_{n+1}^2 is nondecreasing in $\alpha_0, \alpha_1, \beta$.

(α_0) We just have to prove that

$$\frac{\partial\sigma_{n+1}^2}{\partial\alpha_0} = 1+\sum_{k=1}^{n}\prod_{i=1}^{k}(\alpha_1\varepsilon_{n-i+1}^2+\alpha_1\delta^2\varepsilon_{n-i+1}^2+2\alpha_1\varepsilon_{n-i+1}\delta|\varepsilon_{n-i+1}|+\beta) \geq 0 \quad (3.19)$$

We have $\beta \geq 0$ by hypothesis so the previous inequality is equivalent to

$$\alpha_1\varepsilon_i^2 + \alpha_1\delta^2\varepsilon_i^2 + 2\alpha_1\varepsilon_i\delta|\varepsilon_i| \geq 0 \quad for\ i = 0, ...n$$

that is

$$\begin{cases}\alpha_1\varepsilon_i^2(\delta-1)^2, & if\ \varepsilon_i < 0 \\ \alpha_1\varepsilon_i^2(\delta+1)^2, & if\ \varepsilon_i \geq 0\end{cases} \geq 0$$

with $\alpha_1 \geq 0$ and $\delta \in [-1, 1]$.

(α_1) No simplify the notation let define:

$$c_i = \alpha_1 \varepsilon_{n-i+1}^2 + \alpha_1 \delta^2 \varepsilon_{n-i+1}^2 + 2\alpha_1 \varepsilon_{n-i+1} \delta |\varepsilon_{n-i+1}| + \beta$$

Then we can write

$$\frac{\partial \sigma_{n+1}^2}{\partial \alpha_1} = \sigma_0^2 \sum_{i=1}^{n+1} \frac{\partial c_i}{\partial \alpha_1} \prod_{k=1,\ k\neq i}^{n+1} c_i + \sum_{k=1}^{n} \sum_{i=1}^{k} \frac{\partial c_i}{\partial \alpha_1} \prod_{j=1,\ j\neq i}^{k} c_i \tag{3.20}$$

We already proved that $c_i \geq 0$. We only need to verify that $\frac{\partial c_i}{\partial \alpha_1} \geq 0$. That is:

$$\frac{\partial c_i}{\partial \alpha_1} = \begin{cases} \varepsilon_i^2 (\delta - 1)^2, & if\ \varepsilon_i < 0 \\ \varepsilon_i^2 (\delta + 1)^2, & if\ \varepsilon_i \geq 0 \end{cases} \qquad \geq 0$$

(β) Clearly $\frac{\partial c_i}{\partial \beta} = 1$. Then

$$\frac{\partial \sigma_{n+1}^2}{\partial \beta} = \sigma_0^2 \sum_{i=1}^{n+1} \prod_{k=1,\ k\neq i}^{n+1} c_i + \sum_{k=1}^{n} \sum_{i=1}^{k} \prod_{j=1,\ j\neq i}^{k} c_i \geq 0 \tag{3.21}$$

once again from $c_i \geq 0$.

We now have proved that σ_{n+1} is non decreasing in $\alpha_0, \alpha_1, \beta$ also with the leverage correction. It is then possible to apply the same arguments of Theorems 3.8 to prove that with $\alpha_0 \leq_{st} \tilde{\alpha}_0$, $\alpha_1 \leq_{st} \tilde{\alpha}_1$ and $\beta \leq_{st} \tilde{\beta}$ we have $|X_n^{lev}| \leq_{st} |\tilde{X}_n^{lev}|$, $X_n^{2\ lev} \leq_{st} \tilde{X}_n^{2\ lev}$ and $X_n^{lev} \leq_{cx} \tilde{X}_n^{lev}$.

Theorem 3.10. *Let X_n be as in (3.18). If we consider random parameters $\alpha_0 \leq_{st} \tilde{\alpha}_0$, $\alpha_1 \leq_{st} \tilde{\alpha}_1$ and $\beta_1 \leq_{st} \tilde{\beta}_1$, then $|X_n| \leq_{st} |\tilde{X}_n|$, $X_n^2 \leq_{st} \tilde{X}_n^2$ and $X_n \leq_{cx} \tilde{X}_n$.*

To prove the convex comparison of the S_n^{lev} we tried to repeat the proof of Theorem 3.9. The big difference from that proof is that here we have g functions depending from the previous signs of innovations. We show in what follows that this prevents us from proving the desired theorem.
Let X_n be as in (3.15) and S_n be as in (3.8). Let ε_i be symmetric. If we consider random parameters $\alpha_0 \leq_{st} \tilde{\alpha}_0$, $\alpha_1 \leq_{st} \tilde{\alpha}_1$ and $\beta \leq_{st} \tilde{\beta}$, we then would like to have $S_n^{lev} \leq_{cx} \tilde{S}_n^{lev}$.
As before, we write

$$S_n^{lev} = \sigma_0\varepsilon_0 + g_1^{lev}(\varepsilon_0, \sigma_0, \alpha_0, \alpha_1, \beta)\varepsilon_1 + ... + g_n^{lev}(\varepsilon_0, ..., \varepsilon_{n-1}, \alpha_0, \alpha_1, \beta)\varepsilon_n$$

where the functions g_i^{lev} are nondecreasing in the parameters $\alpha_0, \alpha_1, \beta$. Let ϕ be any convex function. The first thing we need to prove, is that $\mathbb{E}[\phi(S_n^{lev})]$ is nondecreasing in the parameters $\alpha_0, \alpha_1, \beta$. But here is where the problems arise.
To simplify the expressions define $g_i^{lev}(\underline{p}) = g^{lev}(\varepsilon_0 p_0, ..., \varepsilon_{i-1}p_{i-1}, \sigma_0, \alpha_0, \alpha_1, \beta)$. From the symmetry of the innovations ε_i we can write:

$$\mathbb{E}[\phi(S_n^{lev})] = \mathbb{E}_{\varepsilon_0,...,\varepsilon_n}[\phi(\sigma_0\varepsilon_0 + ... + g_n^{lev}\varepsilon_n)] =$$
$$= \mathbb{E}_{\varepsilon_0,...,\varepsilon_n}[\tfrac{1}{2^{n+1}} \textstyle\sum_{\underline{p}\in P_{n+1}} \phi(\sigma_0 p_0\varepsilon_0 + ... + p_n g_n^{lev}(\underline{p})\varepsilon_n)1_{\varepsilon_0\geq 0,...,\varepsilon_n\geq 0}]$$

Denoting by

$$\overline{h}(\varepsilon_0, ...\varepsilon_n, \sigma_0, \alpha_0, \alpha_1, \beta) =$$
$$\tfrac{1}{2^{n+1}} \textstyle\sum_{\underline{p}\in P_{n+1}} \phi(\sigma_0 p_0\varepsilon_0 + ... + p_n g_n^{lev}(\underline{p})\varepsilon_n)\ 1_{\varepsilon_0\geq 0,...,\varepsilon_n\geq 0},$$

we need to prove that $\overline{h}$ is nondecreasing in α_0, α_1 and β. To do that we use the same argument of Lemma 3.6. We show why this is not necessary true just in the case of α_0, because for the other variables is very similar.
Let then compute:

$$\tfrac{\partial\overline{h}}{\partial\alpha_0} = \tfrac{1}{2^{n+1}} \textstyle\sum_{\underline{p}\in P_{n+1}} \phi'(\sigma_0 p_0\varepsilon_0 + ... + p_n g_n^{lev}(\underline{p})\varepsilon_n)\ \cdot\ (p_1\varepsilon_1 g_1'^{lev}(\underline{p}) + ... +$$
$$p_n\varepsilon_n g_n'^{lev}(\underline{p}))\ 1_{\varepsilon_0\geq 0,...,\varepsilon_n\geq 0} =$$
$$\mathbb{E}_{\underline{p}}[\phi'(\sigma_0 p_0\varepsilon_0 + ... + p_n g_n^{lev}(\underline{p})\varepsilon_n)\ \cdot\ (p_1\varepsilon_1 g_1'^{lev}(\underline{p}) + ... + p_n\varepsilon_n g_n'^{lev}(\underline{p}))\ 1_{\varepsilon_0\geq 0,...,\varepsilon_n\geq 0}]$$

We need to show three statement to use the covariance inequality an conclude $\frac{\partial\overline{h}}{\partial\alpha_0} \geq 0$:

(a) $X(\underline{p}) = \phi'(\sigma_0 p_0\varepsilon_0 + ... + p_n g_n^{lev}(\underline{p})\varepsilon_n)\ 1_{\varepsilon_0\geq 0,...,\varepsilon_n\geq 0}$ is componentwise increasing in $\underline{p}$.

(b) $Y(\underline{p}) = p_1\varepsilon_1 g_1'^{lev}(\underline{p}) + ... + p_n\varepsilon_n g_n'^{lev}(\underline{p})\ 1_{\varepsilon_0\geq 0,...,\varepsilon_n\geq 0}$ is componentwise increasing in $\underline{p}$, where with g' we denote the partial derivative of g with respect to α_0.

(c) The expected value of $Y(\underline{p})$ with respect to $\underline{p}$ is zero.

In what follows we use (3.18), (3.20) and

$$\frac{\partial c_i}{\partial p_j} = \begin{cases} 0 & if \quad i \neq j \\ 2\alpha_1\varepsilon_i^2 p_i(1+\delta^2+2\delta sign(p_i)) & if \quad i = j \end{cases} \tag{3.22}$$

that is positive with p_i positive and negative with p_i negative.
However, if we consider strictly the domain of the p_i variables, we see that:

$$\delta \geq 0 \Rightarrow c_i(p_i = -1) \leq c_i(p_i = 1)$$
$$\delta < 0 \Rightarrow c_i(p_i = -1) > c_i(p_i = 1)$$

that means, with a fixed value of δ, we have a strictly monotone dependence of c_i from p_i.
Let now show why the above statements are generally false:

(a) We have that

$$\frac{\partial X(\underline{p})}{\partial p_i} = \phi''(...) \cdot (g_i^{lev}(\underline{p})\varepsilon_i + p_{i+1}\frac{\partial g_{i+1}^{lev}}{\partial p_i}(\underline{p})\varepsilon_{i+1} + ... + p_n \frac{\partial g_n^{lev}}{\partial p_i}(\underline{p})\varepsilon_n)$$
$$1_{\varepsilon_0 \geq 0, \ldots, \varepsilon_n \geq 0} \quad (3.23)$$

and with $j \geq l$ and $m = j - l + 1$

$$\frac{\partial g_{j+1}^{lev}}{\partial p_l} = \sigma_0^2 \frac{\partial c_m}{\partial p_l} \prod_{i=1\ i \neq m}^{j+1} c_i + \sum_{k=m}^{j} \frac{\partial c_m}{\partial p_l} \prod_{i=1\ i \neq m}^{k} c_i \quad (3.24)$$

The dependence of $\frac{\partial X(\underline{p})}{\partial p_i}$ from others p_k's means that the point (a) is not true.

(b) We have

$$\frac{\partial Y(\underline{p})}{\partial p_i} = g_i'^{lev}(\underline{p})\varepsilon_i + p_{i+1}\frac{\partial g_{i+1}'^{lev}}{\partial p_i}(\underline{p})\varepsilon_{i+1} + ... + p_n \frac{\partial g_n'^{lev}}{\partial p_i}(\underline{p})\varepsilon_n$$
$$1_{\varepsilon_0 \geq 0, \ldots, \varepsilon_n \geq 0} \quad (3.25)$$

and with $j \geq l$ and $m = j - l + 1$

$$\frac{\partial g_{j+1}'^{lev}}{\partial p_l} = \sum_{k=m}^{j} \frac{\partial c_m}{\partial p_l} \prod_{i=1\ i \neq m}^{k} c_i \quad (3.26)$$

The dependence of $\frac{\partial Y(\underline{p})}{\partial p_i}$ from others p_k's means that the point (b) is not true.

(c) Using the linearity of the expected value and the independence of p_i from g_i^{lev} we have the thesis.

It's then clear that the GARCH with leverage case is not solvable in the same way of the previous cases. Although we can't prove it's true, we can't even prove it's false. Besides the numerical simulations show that with the given hypothesis, $S_n^{lev} \leq_{cx} \tilde{S}_n^{lev}$ do hold.

Chapter 4

Numerical simulations

In order to avoid repetitions, we will focus just on the ARCH(3) and GARCH with leverage models. In fact we have shown that in the ARCH case we may get the same conclusion as in the classic GARCH model. Essentially, there are three kind of stochastic order's propagation we dealt with in this thesis :

(I) from the innovations to logreturns and to total logreturns

(II) from the parameters to logreturns, their absolute values and square

(III) from the parameters to the total logreturns

As we proved that (I) and (III) hold in the cases of ARCH(q) and GARCH(1,1) models while (II) holds in each of the three models examined, we now want to give a numerical proof of it. On the other side we are interested in showing how, in the GARCH with leverage model, (I) and (III) seem to hold true. In this Chapter we will use the notation F_X to indicate the distribution function of the variable X. Consider the model in (3.11) with $q = 3$:

$$\begin{cases} X_n^A = \sigma_n \varepsilon_n, \quad n = 0, 1, ... \\ \sigma_n^{2\ A} = \gamma_0 + \sum_{i=1}^{3} \gamma_i X_{t-i}^{2\ A} \quad n = q, q+1, ... \end{cases}$$

and the model in (3.18):

$$\begin{cases} X_n = \sigma_n \varepsilon_n, \quad n = 0, 1, .. \\ \sigma_{n+1}^2 = \alpha_0 + \alpha_1 (X_n + \delta |X_n|)^2 + \beta \sigma_n^2 \end{cases}$$

The choices to be done in the following numerical simulations, besides the values of the parameters, are:

- the initial values, i.e. σ_0 in the GARCH with leverage case and σ_0^A, σ_1^A and σ_2^A in the ARCH case
- which parameters has to be ordered in $\leq_{st}$ order
- which probabilistic distribution assign to each stochastic parameter
- how many step of the simulation consider
- the distribution of the innovations
- the number N of simulations on which the distribution functions are computed

We always choose N=100000 and a Gaussian(0,1) distribution for the innovations ε_n.
As the initial values we consider $\sigma_0 = 0.002$ for the GARCH with leverage case and σ_0^A, σ_1^A and σ_2^A equal to 0.001 for the ARCH case.
For the parameters distribution we pick out the absolute value of a Normal variable. In fact we need positive parameters to guarantee the positivity of σ_n^2.
For every simulation we show just the result related to 100 steps simulations because there's no significant difference in the graphs with the number of steps varying. Finally, in order to deepen the roles of the three different parameters, we first analyze the case of just one stochastic parameter and, at the end, we show that the same results hold true also taking three stochastic parameters.
The main goal is to give a numerical proof of Theorems (3.6), (3.7) and of (3.10).

4.1 GARCH(1,1) with leverage: stochastic α_1

We first want to analyze the role of each parameter. α_0 can be interpreted as the constant part of the volatility's evolution. It's then reasonable that with higher values of this parameter we obtain higher values of the volatility and consequently of the logreturns.
The α_1 parameter represent the adjustment to past shocks. In fact the higher is α_1, the higher is the contribute to σ_{n+1}'s value due to X_n's value.
Differently from the previous α_0 case, here is not that evident that an order between two different α_1 should imply an order between the respective X_n.
Consider an high α_1: if on one side it keeps the logreturns on high values when

the previous logreturn is high, on the other side it prevents low logreturns to come back to higher values.
Finally we have β: it obviously describes how the precedent value of the volatility influences the following one. The numerical results we are going to show on the case of stochastic α_1 hold true also taking, as stochastic, one of the other two parameters.
Consider the values:

- $\alpha_0 = 10^{-7}$
- $\beta = 0.8$
- $\delta = -0.2$

We then consider $\alpha_1 \sim |X|$ with $X := N(0.25, 0.25)$ and $\tilde{\alpha}_1 \sim |Y|$ with $Y := N(0.15, 0.1)$.

We see from Figure 4.1 that the condition requested from Corollary (1.1) is satisfied by the distribution functions of $\tilde{\alpha}_1$ and α_1 but with sign sequence +,-. It then follows that $\tilde{\alpha}_1 \leq_{st} \alpha_1$.
We then compute the 100 steps 100000 simulations of the logreturns series with the given parameters. Hence we calculate the difference between the distribution function respectively of X_n and of $\tilde{X}_n$, $|X_n|$ and $|\tilde{X}_n|$, X_n^2 and $\tilde{X}_n^2$.
From the graphs (b) and (c) of Figure 4.2 it's clear (using Corollary (1.1)) that $|\tilde{X}_{100}| \leq_{st} |X_{100}|$ and $\tilde{X}_{100}^2 \leq_{st} X_{100}^2$ hold. Moreover from graph (a) we can observe that $F_{\tilde{X}_{100}} - F_{X_{100}}$ changes its sign once with sign sequence -,+, that means, using Theorem (1.10), $\tilde{X}_{100} \leq_{cx} X_{100}$.

We take the total logreturns S_{100} and $\tilde{S}_{100}$ to see if there's some stochastic order between them. You can see the result in Figure 4.2 (d), where it's clear that the condition of Theorem (1.10) is satisfied. Therefore, according to this numerical example, with symmetric innovations we have that a stochastic order on α_1 implies a convex order between the total logreturns.

4.2 GARCH(1,1) with leverage: stochastic α_0, α_1 and β

Consider as parameters:

- $\alpha_0 \sim |X|$ with $X := N(2 \cdot 10^{-7}, 10 \cdot -9)$ and $\tilde{\alpha}_0 \sim |Y|$ with $Y := N(10^{-7}, 10^{-8})$

- $\alpha_1 \sim |X|$ with $X := N(0.25, 0.25)$ and $\tilde{\alpha}_1 \sim |Y|$ with $Y := N(0.15, 0.1)$
- $\beta \sim |X|$ with $X := N(0.8, 0.6)$ and $\tilde{\beta}_1 \sim |Y|$ with $Y := N(0.6, 0.4)$

We first show the graphs related to the distribution of the parameters to verify their orders in Figure 4.3. It's clear, using as in the previous section Corollary (1.1), that $\tilde{\alpha}_0 \leq_{st} \alpha_0$, $\tilde{\alpha}_1 \leq_{st} \alpha_1$ and $\tilde{\beta} \leq_{st} \beta$. Then, as in the previous section, we show the graphs with the distribution of the innovations, of their absolute values, of their square and of their sums in Figure 4.4. Using once again Corollary (1.1), all the desired results hold true. Moreover, as in the previous section, the total logreturns are ordered in respect to the convex order thanks to Theorem (1.10).

4.3 ARCH(3): stochastic γ_0, γ_1, γ_2 and γ_3

The role of each parameter is clearer in this model: as in the previous case γ_0 is the constant part of the volatility's dynamic. The other parameters describe the adjustment to the previous three shocks.
We consider as parameters:

- $\gamma_0 \sim |X|$ with $X := N(2 \cdot 10^{-7}, 2 \cdot 10^{-6})$ and $\tilde{\gamma}_0 \sim |Y|$ with $Y := N(10^{-7}, 10^{-6})$
- $\gamma_1 \sim |X|$ with $X := N(0.65, 0.65)$ and $\tilde{\gamma}_1 \sim |Y|$ with $Y := N(0.5, 0.6)$
- $\gamma_2 \sim |X|$ with $X := N(0.12, 0.6)$ and $\tilde{\gamma}_2 \sim |Y|$ with $Y := N(0.18, 0.2)$
- $\gamma_3 \sim |X|$ with $X := N(0.05, 0.15)$ and $\tilde{\gamma}_3 \sim |Y|$ with $Y := N(0.07, 0.05)$

From Figure 4.6 it's clear, using as in the previous section Corollary (1.1), that $\tilde{\gamma}_0 \leq_{st} \gamma_0$, $\tilde{\gamma}_1 \leq_{st} \gamma_1$, $\tilde{\gamma}_2 \leq_{st} \gamma_2$ and $\tilde{\gamma}_3 \leq_{st} \gamma_3$. Then we show the graphs with the distribution of the innovations, of their absolute values, of their square and of their sums in Figure 4.4. Using once again Corollary (1.1), all the desired results hold true. In Figure 4.4 you can see that the total logreturns are ordered in respect to the convex order thanks to Theorem (1.10). We have then showed that Theorems (3.6) and (3.7) hold true.

4.4 Innovations $\leq_{cx}$ ordered

We now want to show a numerical proof of the results stated in Theorems (3.4) and (3.5) for the ARCH case and see if, in this case, the same results hold true also for the GARCH with leverage model.
Let start with the GARCH with leverage model:

- $\alpha_0 = 10^{-7}$
- $\alpha_1 = 0.25$
- $\beta = 0.8$
- $\delta = -0.2$

and consider a 50 step GARCH with leverage series. As innovations we take iid standard Gaussian. Then we simulate another GARCH series with the same exact parameters of the first one but taking as tenth innovation ($\tilde{\varepsilon}_1 0$), a Gaussian variable with mean 0 and variance 3. Observe that all the innovations considered are symmetric. We first show, using the same argument as before, that $\tilde{\varepsilon}_{10} \geq_{cx} \varepsilon_{10}$ using Figure 4.7.

We then compute the logreturns and the total logreturns and their distribution function. In Figure 4.8 it's clear, thanks the usual Theorem concerning the sign change of the difference of the distribution functions, that $X_{50} \leq_{cx} \tilde{X}_{50}$ and $S_{50} \leq_{cx} \tilde{S}_{50}$.

Let now consider an ARCH model with the following parameters:

- $\gamma_0 = 2 \cdot 10^{-7}$
- $\gamma_1 = 0.65$
- $\gamma_2 = 0.12$
- $\gamma_3 = 0.05$

the innovations ε_{10} and $\tilde{\varepsilon}_{10}$ with the same values of the previous GARCH case. We consider 50 step simulations. In Figure 4.9 you can see that the hypothesis of Theorem (1.10) are respected, that is $X_{50}^{ARCH} \leq_{cx} \tilde{X}_{50}^{ARCH}$ and $S_{50}^{ARCH} \leq_{cx} \tilde{S}_{50}^{ARCH}$.
We have then shown that Theorems (3.4) and (3.5) do hold in the ARCH case.

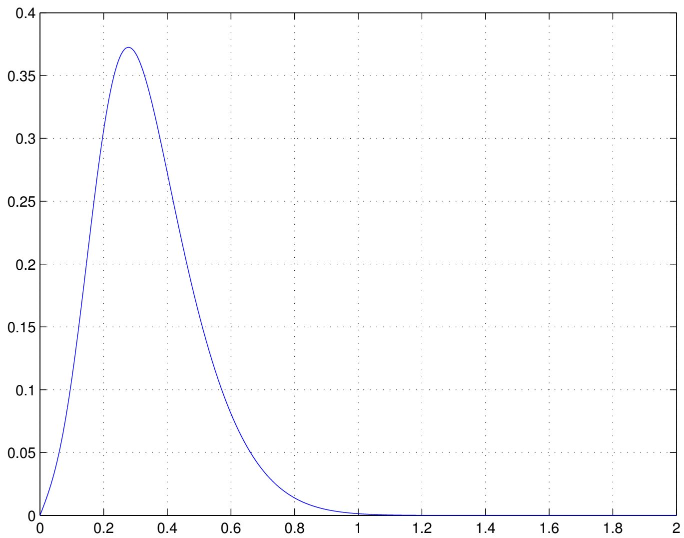
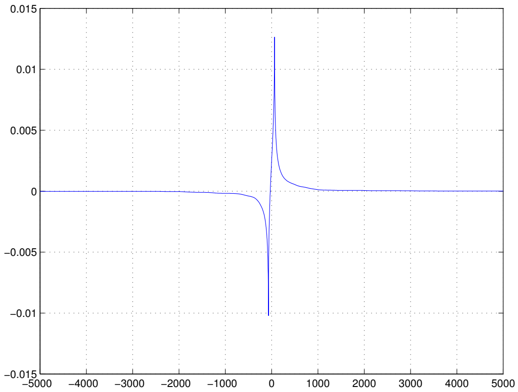
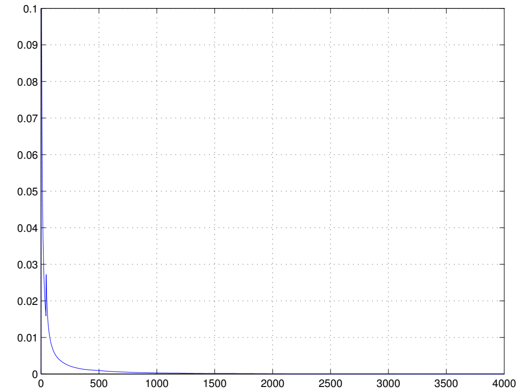


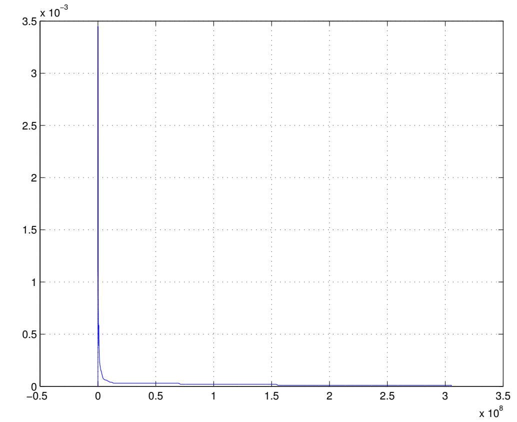
Figure 4.1: Difference of the distribution functions of $\tilde{\alpha}_1$ and α_1: $F_{\tilde{\alpha}_1} - F_{\alpha_1}$



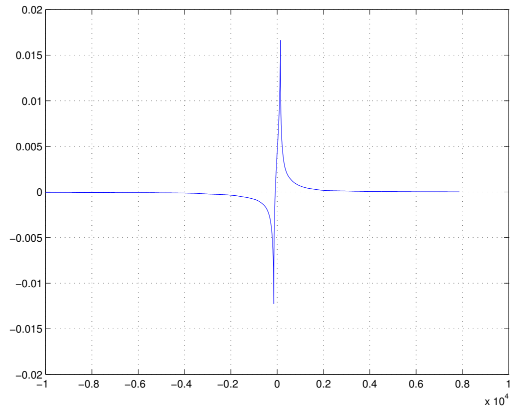
(a) $F_{\tilde{X}_{100}} - F_{X_{100}}$



(b) $F_{|\tilde{X}_{100}|} - F_{|X_{100}|}$

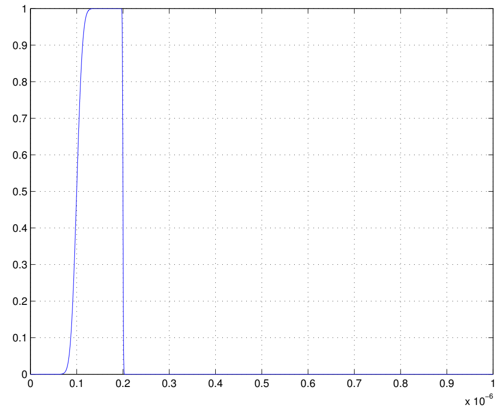


(c) $F_{\tilde{X}^2_{100}} - F_{X^2_{100}}$

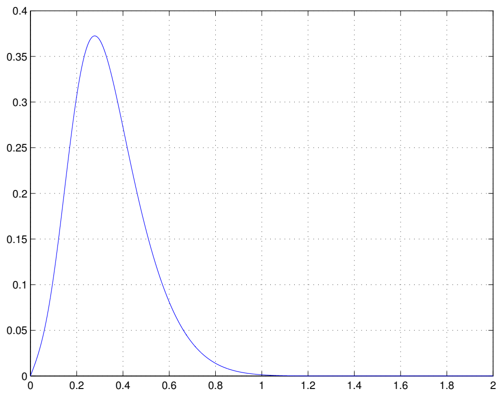


(d) $F_{\tilde{S}_{100}} - F_{S_{100}}$

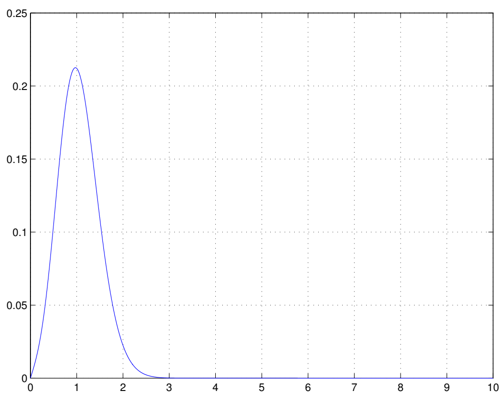
Figure 4.2: Comparison on 100 steps GARCH with leverage time series



(a) $F_{\tilde{\alpha_0}} - F_{\alpha_0}$

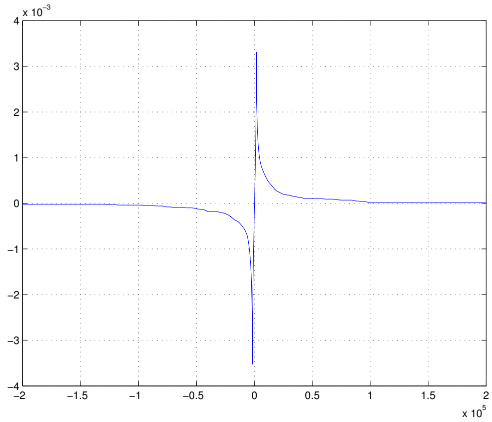


(b) $F_{\tilde{\alpha_1}} - F_{\alpha_1}$

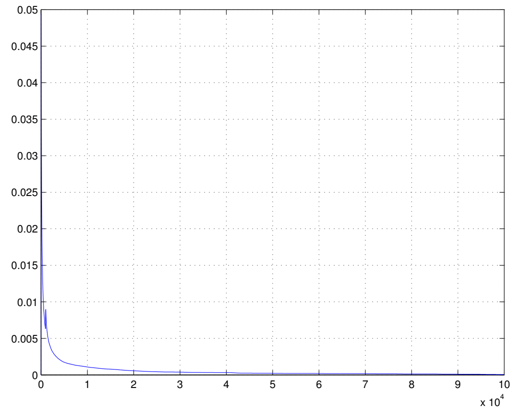


(c) $F_{\tilde{\beta}} - F_{\beta}$

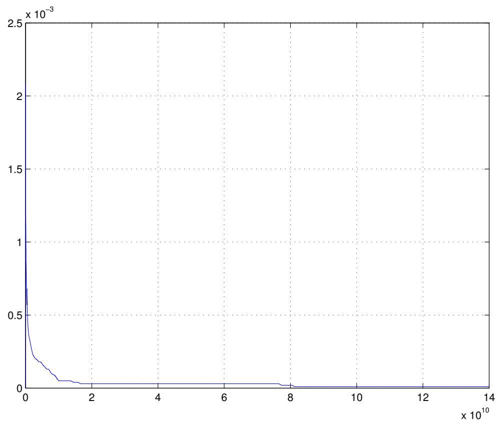
Figure 4.3: Difference of the distribution functions of the GARCH with leverage parameters



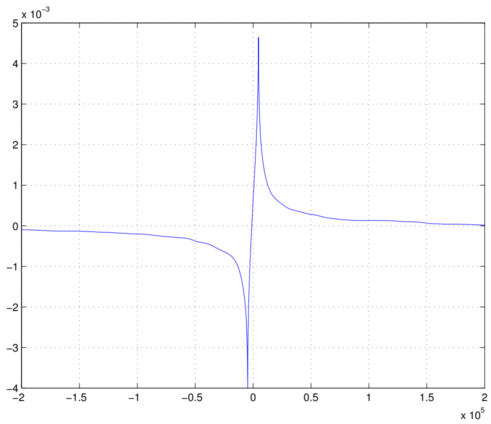
(a) $F_{\tilde{X}_{100}} - F_{X_{100}}$



(b) $F_{|\tilde{X}_{100}|} - F_{|X_{100}|}$

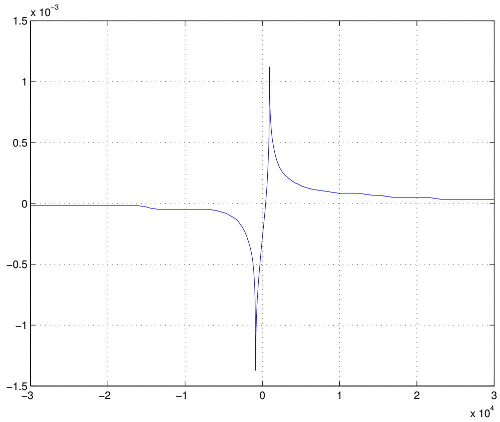


(c) $F_{\tilde{X}^2_{100}} - F_{X^2_{100}}$

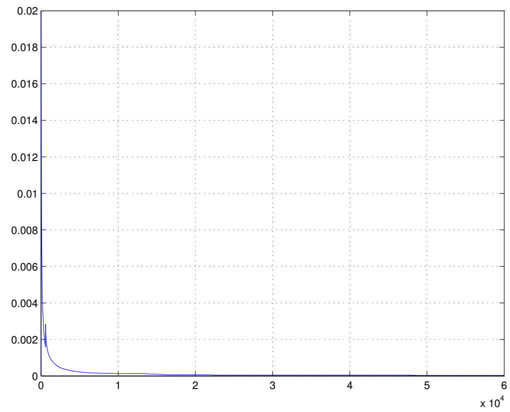


(d) $F_{\tilde{S}_{100}} - F_{S_{100}}$

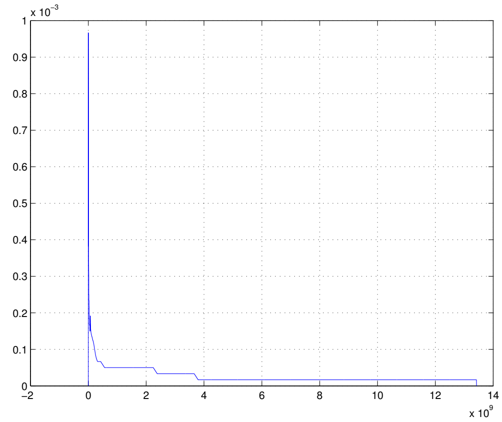
Figure 4.4: Comparison results on 100 steps GARCH with leverage time series



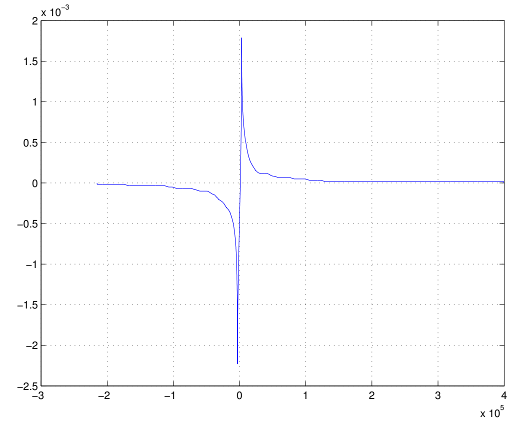
(a) $F_{\tilde{X}_{100}^{ARCH}} - F_{X_{100}^{ARCH}}$



(b) $F_{|\tilde{X}_{100}^{ARCH}|} - F_{|X_{100}^{ARCH}|}$

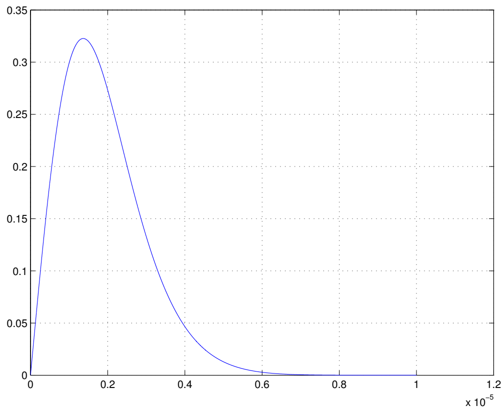


(c) $F_{\tilde{X}_{100}^{2\ ARCH}} - F_{X_{100}^{2\ ARCH}}$

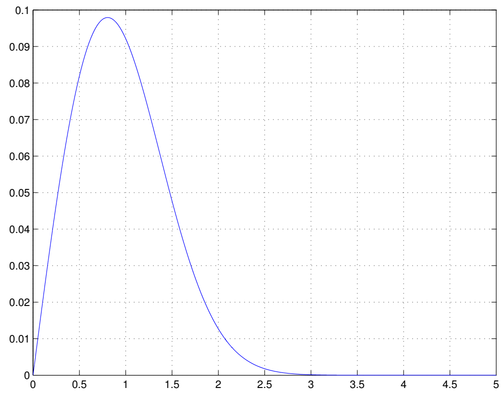


(d) $F_{\tilde{S}_{100}^{ARCH}} - F_{S_{100}^{ARCH}}$

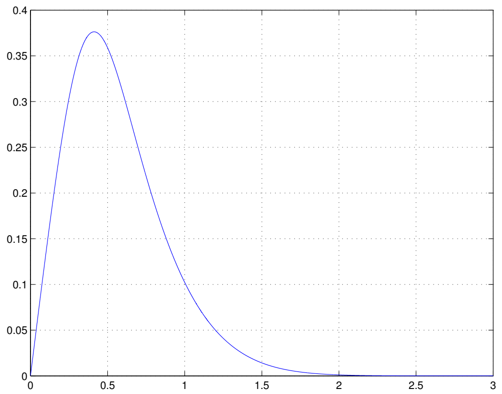
Figure 4.5: Comparison results on 100 steps ARCH time series



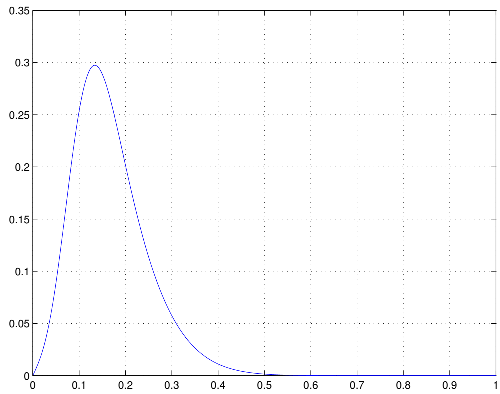
(a) $F_{\tilde{\gamma}_0} - F_{\gamma_0}$



(b) $F_{\tilde{\gamma}_1} - F_{\gamma_1}$



(c) $F_{\tilde{\gamma}_2} - F_{\gamma_2}$



(d) $F_{\tilde{\gamma}_3} - F_{\gamma_3}$

Figure 4.6: Difference of the distribution functions of the ARCH parameters

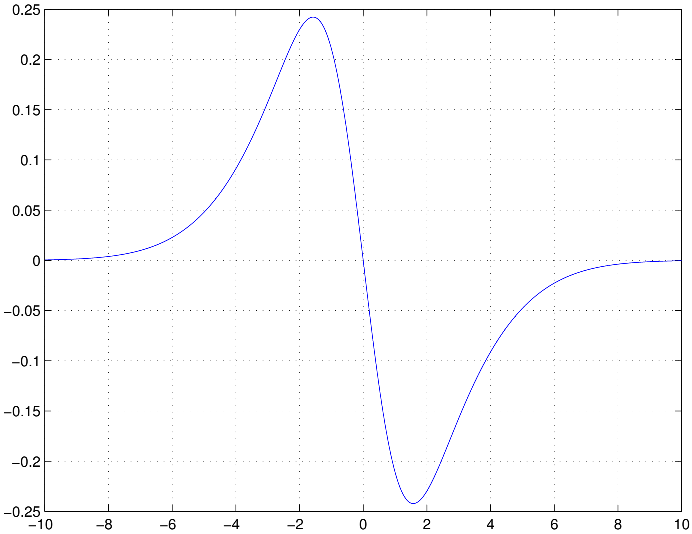
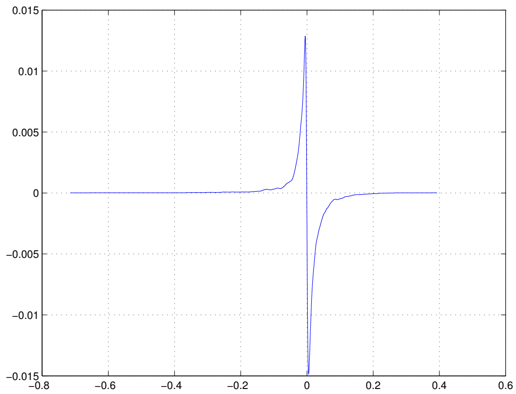
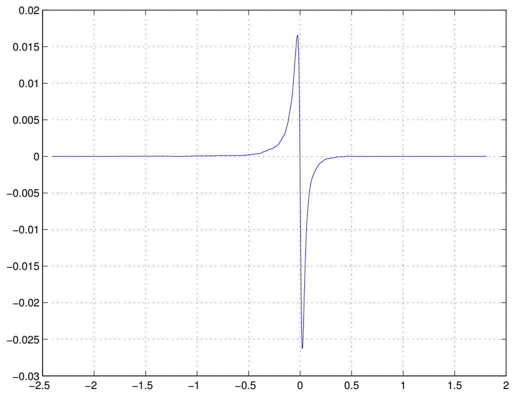


Figure 4.7: Difference of the distribution function of $\tilde{\varepsilon}_{10}$ and ε_{10}: $F_{\tilde{\varepsilon}_{10}} - F_{\varepsilon_{10}}$

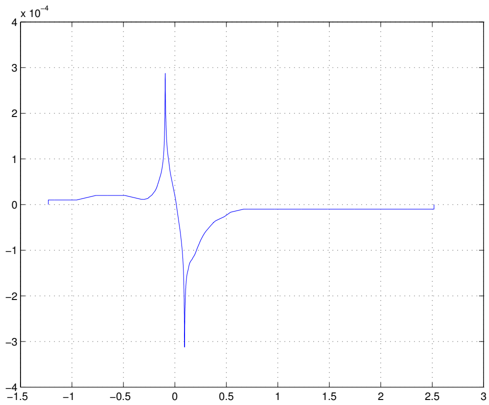


(a) $F_{\tilde{X}_{50}} - F_{X_{50}}$

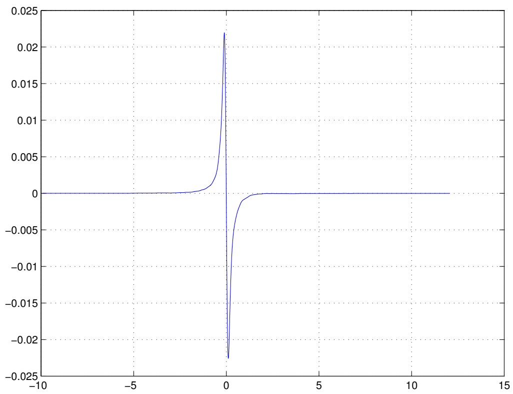


(b) $F_{\tilde{S}_{50}} - F_{S_{50}}$

Figure 4.8: Comparison results on 50 steps GARCH with leverage time series



(a) $F_{\tilde{X}_{50}^{ARCH}} - F_{X_{50}^{ARCH}}$



(b) $F_{\tilde{S}_{50}^{ARCH}} - F_{S_{50}^{ARCH}}$

Figure 4.9: Comparison results on 50 steps ARCH time series

Conclusions

As underlined at the beginning of the last chapter, there are three kind of stochastic order's propagation we dealt with in this thesis :

(I) from the innovations to logreturns and to total logreturns

(II) from the parameters to logreturns, their absolute values and square

(III) from the parameters to the total logreturns

At the same time we considered three specific models: the ARCH, the GARCH and the GARCH with leverage. Taking the cue from the arguments developed in the article "Comparison Results for GARCH processes" by Bellini, Pellerey , Sgarra and Sekeh [3] concerning the GARCH(1,1) model, we tried to expand those results to the other models.
We can obtain the ARCH model from the GARCH one taking all the β_i parameters equal to zero. So it was obvious that the same results should also apply in the ARCH(1) case. I then considered the general ARCH(q) model and showed that the same proof of the above article is worthy also in this case. Using the same arguments we may get the same results adding more steps to the GARCH(1,1) that is considering a GARCH(p,q) model.
Finally i considered the GARCH with leverage case. The non linearity with respect to the innovations introduced by this model is reflected in the no more increasing dependence of the volatility from the innovations themselves. The argument used in Theorems (3.3), (3.2) and (3.4) is therefore no more valid and the point (I) can't be proved. Moreover the dependence of the so called g_i functions from the previous signs of the innovations prevents from using the argument of Theorem (3.9) and prove the point (III).
However the partial derivative of the volatilities computed by the GARCH with leverage model is positive in respect to the parameters. This allows us to state the Theorem (3.10) to prove point (II).
The numerical simulations of the last chapter show that the propagation (I) and (III) for the GARCH with leverage model do hold , at least sometimes. It might be subject of future studies the research of conditions maybe on the

parameters or maybe on the distribution of the innovations that guarantee the propagations (I) and (III).
At the same time the hypothesis of symmetrical innovations considered in the ARCH and GARCH cases that is fundamental for proving the theorem might be dropped. In fact there's no numerical evidence concerning the importance of this hypothesis.
Another subject that has to be deepen is the one related to multivariate comparisons of logreturns. I did focused on scalar variables X_n and S_n but there is a wide literature concerning the multivariate approach to orders' propagation.
The basic tool for a such kind of studies is the definition of some multivariate stochastic orders. You find a detailed description of these orders in "Stochastic Orders" by Shaked M. and Shanthikumar J. G. [20].
Some theory related to the multivariate stochastic orders can be found in "Stochastic convexity on general space" by Meester and Shanthikumar [14], especially the convex one. They also give examples and applications from queuing theory, coverage processes, reliability and branching processes.

References

[1] Baüerle, N., Müller, A. (2006). Stochastic Orders and Risk Measures: Consistency and Bounds. *Insurance: Mathematics and Economics*, **38** (1), 132-148.

[2] Bellamy, N., Jeanblanc, M. (2000). Incompleteness of markets driven by a mixed diffusion. *Finance and Stochastic*, **4**, 209-222.

[3] Bellini, F., Pellerey, F., Sgarra, C., Sekeh, S.Y. (2012). Comparison Results for Garch Processes. **Preprint**.

[4] Bergenthum, J., Rüschendorf, L. (2006) Comparison of option prices in semimartingale models. *Finance and Stochastic*, **10** (2) 222-249.

[5] Birnbaum, Z. W. (1948). On random variables with comparable peakedness. *Ann. Math. Stat.*, **19**, 76-81.

[6] Bollerslev T., (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, **31**, 307-327.

[7] El Karoui, N., Jeanblanc-Picqu, M., Shreve, S.E. (1998). Robustness of the Black and Scholes formula. *Math. Fin.*, **8**, 93-126.

[8] Engle, R. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the UK Inflation. *Econometria*, **50**, 987-1008.

[9] Finucan, H.M. (1964). A note on kurtosis. *Journal of the Royal Statistical Society Series B (Methodological)*, **26** (1), 111-112.

[10] Gushchin, A.A., Mordecki, E. (2002). Bounds on option prices for semimartingale market models. *Proc. Steklov Inst. Math.*, **237**, 73-113.

[11] Henderson, V., Hobson, D. (2003). Coupling and Option Price Comparisons in a Jump-Diffusion Model. *Stochastics and Stochastics Reports*, **75** (3) 79-101.

[12] Kijima, M., Ohnishi, M. (1999). Stochastic orders and their applications in financial optimization. *Mathematical Methods of Operations Research*, **50** (2), 351-372.

[13] McNeil, A.J., Rüdiger, F., Embrechts, P. (2005). *Quantitative Risk Management* Princeton University Press.

[14] Meester, L.E., Shanthikumar, J.G. (1999). Stochastic convexity on general space. *Math. Oper. Res.*, **24**, 472-494.

[15] Müller, A., Pflug, G. (2001). Asymptotic ruin probabilities for risk processes with dependent increments. *Insurance: Mathematics and Economics*, **28** (3), 381-392.

[16] Müller, A., Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*, J. Wiley & Sons.

[17] Møller, T., (2004). Stochastic orders in dynamic reinsurance market. *Finance and Stochastic*, **8**, 479-499.

[18] Nelson, D.B., (1990). Stationarity and persistence in GARCH(1,1) models. *Econometric Theory*, **6**, 318-334.

[19] Rüschendorf, L. . Ordering of insurance Risk. In *Encyclopedia of Quantitative Risk Analysis and Assessment*, VOL **3**, J. Wiley & Sons.

[20] Shaked, M., Shanthikumar, J.G., (2006). *Stochastic Orders*, 2nd edn., Springer.

[21] Wu, J., (2011). Threshold GARCH Model: Theory and Application.