

An Agent-Based Stochastic Volatility Model

als Inaugural-Dissertation
zur Erlangung des akademischen Grades eines Doktors
der Wirtschaft- und Sozialwissenschaftlichen Fakultät
der Cristian-Albrechts-Universität zu Kiel

vorgelegt von

Dipl. Phys. Simone Alfarano,

Kiel, März 2006

An Agent-Based Stochastic Volatility Model

als Inaugural-Dissertation
zur Erlangung des akademischen Grades eines Doktors
der Wirtschaft- und Sozialwissenschaftlichen Fakultät
der Cristian-Albrechts-Universität zu Kiel

vorgelegt von

Dipl. Phys. Simone Alfarano,
aus Cagliari (Italien) geb. 21/06/1975

Kiel, März 2006

Gedruckt mit Genehmigung der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

Dekan: Prof. Dr. Andreas Drexl
Erstberichterstattender: Prof. Dr. Thomas Lux
Zweitberichterstattender: Prof. Dr. Friedrich Wagner

Tag der Abgabe der Arbeit: 20. Dezember 2005
Tag der mündlichen Prüfung: 9. Februar 2006

TO MY FAMILY

Contents

List of Figures	V
List of Tables	VII
1 Introduction: Heterogeneity and Interactions	1
2 Distributional Properties of Financial Returns	5
2.1 Leptokurtic Distributions and Fat Tails	7
2.2 Alternative Hypothesis to Stable Distributions	13
2.3 The Tail Index of Continuous Distributions and the Extreme Value Theory	14
2.3.1 The Central Limit Theorem	15
2.3.2 The Extreme Value Theory	18
2.3.3 Estimation of the tail index	21
2.3.4 Optimal tail size	24
2.3.5 Monte Carlo Simulations	26
2.4 Fat tails in financial data: The case of the Tokyo Stock Exchange	28
2.5 Conclusions	34
Appendix to Chapter 2	35
A 2.1 Description of the data	35
A 2.2 Derivation of the Hill Estimator	35
A 2.3 Optimal tail size for Student-t distribution	36
3 Serial Correlation Properties of Financial Time Series	39
3.1 Measuring Correlations	40
3.2 Serial Correlation in Raw Returns	42
3.3 The notion of volatility	47
3.4 Heuristic estimation methods for long memory processes	49
3.5 Conclusions	58
4 An analytical approach to an herding model	61
4.1 A simple herding model as a jump Markov process	61
4.1.1 Transition rates in the Kirman's model	62
4.1.2 Generalized transition rates	63
4.2 A theoretical finitary approach	65
4.3 The diffusion approximation	67

4.3.1	Derivation of the Fokker-Planck equation	68
4.4	Remarkable properties of the transition rates	70
4.4.1	Local and non-local interactions	70
4.4.2	Independence on the total number of individuals	73
4.4.3	Mean-reverting nature of the process	74
4.5	Equilibrium distribution	74
4.6	Integration of the Fokker-Planck equation	77
4.6.1	Expansion in Jacobi polynomials	78
4.6.2	Main result	79
4.6.3	Final comments	80
4.7	Autocorrelation	81
4.8	Special cases: the cosine of a random walk	84
4.9	Langevin equation and the micro-macro structure	85
4.9.1	Autocorrelation revisited	87
4.10	Different avenues for simulating the herding model	88
4.11	The Kirman model and the emergence of macroscopic laws	90
4.12	Conclusions	93
	Appendix to Chapter 4	94
	Appendix to Chapter 4	94
A 4.1	Backward equation	94
A 4.2	Boundary conditions	94
A 4.3	The Jacobi polynomials	96
5	An artificial financial market based on herding:	
	Analytical results	99
5.1	The artificial market	99
5.1.1	Implementation of the market	100
5.1.2	Aggregate excess demand	102
5.1.3	Computation of the returns	104
5.1.4	Stochastic volatility approximation	108
5.2	Statistical properties of the volatility process	112
5.2.1	Fokker-Planck equation for the volatility	112
5.2.2	Connection with the discrete model	114
5.2.3	Dynamical properties of the volatility process	114
5.2.4	General remarks	117
5.2.5	Equilibrium distribution of volatility	118
5.2.6	Conditional properties of the volatility: the autocorrelation	120
5.3	Asymptotic behavior of the volatility process:	
	the extreme asymmetric case $\varepsilon_1 \gg \varepsilon_2$	122
5.3.1	The limiting stochastic process	122
5.3.2	Conditional and unconditional properties of the process (5.54)	124
5.3.3	The connection with the discrete model	126
5.4	The returns process	127
5.4.1	Unconditional distribution of returns	127
5.4.2	Conditional properties of the returns	128
5.5	Conclusions	129

Appendix to Chapter 5	130
A 5.1 Analytical moments of the process (5.10)	130
A 5.2 Fokker-Plank equation for the volatility process (5.25)	130
A 5.3 Unconditional moments of the volatility (5.43)	130
A 5.4 Auto-covariance of the volatility, eq. (5.46)	131
A 5.5 Laguerre polynomials	133
A 5.6 Autocorrelation (5.64)	133
6 Estimation of the Parameters of an Agent-Based Model with Asymmetric Herding	135
6.1 Estimation of the parameters of the model	136
6.2 Discussion of the results	138
Appendix to Chapter 6	141
Conclusions	141
A 6.1 Unconditional distribution of returns: $\varepsilon_1\varepsilon_2$ -model with uniform noise	143
A 6.2 Unconditional distribution of returns: ε_2 -model with uniform noise	143
A 6.3 Unconditional distribution of returns with Gaussian noise	144
A 6.4 Discreteness of price records	144
7 General Conclusions	147
Acknowledgments	148
Bibliography	158

List of Figures

2.1	Deviation from gaussian behavior	8
2.2	Behavior of the sequential moments	12
2.3	Hill plot	23
2.4	OLS and Hill estimator.	30
3.1	Autocorrelation functions of returns	45
3.2	Autocorrelation functions of simple non-linear transformation of returns	48
3.3	Volatility clustering effect	50
3.4	DFA: comparison raw and absolute returns	55
3.5	GPH: comparison raw and absolute returns	56
4.1	Equilibrium discrete distributions	66
4.2	Equilibrium continuous distributions	76
4.3	Discreteness of the stationary distribution	77
4.4	Comparison continuous and finitary approach	78
4.5	Example of application of the FPE: uni-modal equilibrium distribution	82
4.6	Example of application of the FPE: bi-modal equilibrium distribution	83
4.7	N-independence in the micro-simulation approach	90
4.8	Emergence of the macroscopic equation	91
5.1	Stochastic volatility approximation	109
5.2	Typical behavior of the autocorrelation function from simulated data	112
5.3	Time series of volatility	115
5.4	Deterministic path of volatility	116
5.5	Equilibrium distribution for volatility	119
5.6	Autocorrelation function of volatility	121
5.7	Convergence of the equilibrium distribution	125
5.8	Autocorrelation function	126
5.9	Auto-correlation function of absolute returns	128
6.1	Comparison of the models	137
6.2	Comparison of the models	139
6.3	Fit empirical data	141
6.4	Fit empirical data	142

List of Tables

2.1	Jarque-Bera test	7
2.2	Alternative measures of leptokurtosis	9
2.3	Monte Carlo simulations for tail index estimation	27
2.4	Results of the application of several tail index estimation procedures	32
3.1	Sample autocorrelations	44
3.2	Ljung-Box test	46
3.3	BDS test for non-linearity of financial returns	46
3.4	R/S statistics and Hurst exponent	52
3.5	Detrended Fluctuation Analysis	54
3.6	GPH estimator	58
6.1	Estimation of the parameters	138
6.2	Hill estimates	140

CHAPTER 1

Introduction: Heterogeneity and Interactions

Financial markets are populated by heterogeneous traders. The traders have different beliefs about the future development of prices or volatilities, different budget constraints, different access to the market, different information.

A long debate in the economic literature, starting with Friedman [1953] and formalized by Fama [1965, 1970], exists on whether this plethora of different traders has a relevant impact on the price formation, and, if relevant, about the role that this heterogeneity plays in the market. The Efficient Market Hypothesis (EMH) states that the presence of non-rational traders can be largely neglected, since their idiosyncratic errors would be averaged out in the aggregate so that they could not significantly affect the market price. Eventually, they would progressively lose money in favor of rational investor, disappearing at the end from the market. The heterogeneity of traders is, then, drastically reduced to a homogeneous group of rational traders with ‘correct’ expectations¹, which ultimately ‘collapses’ into a single representative agent. Within this theoretical framework, the price formation is driven by absence of arbitrage opportunities, defined as the possibility of a riskless gain. The continuous arbitrage activity makes the market efficient, in the sense that the price incorporates all available information. The literature in favor of this framework, from both sides empirical and theoretical, is vast, hardly confineable in few articles or surveys. A short (and critical) review of the EMH is given by Shleifer [1999].

The faith in the market efficiency has eroded in the last decades, as theoretical and empirical evidence has been found against the ‘perfect’ informational efficiency of the markets. One of the main contributions of this literature recognizes that the heterogeneity of the traders matters, and that non-rational behavior can be a relevant determinant in the price formation. The research has concentrated on the analysis of the arbitrage mechanism and its close connection with the rationality of the investors as the ultimate engine to ensure market efficiency. From an individual perspective, several cognitive biases in the process of decision formation have been first detected as psychological phenomena. Successively, they have been acknowledged as being of economic relevance. Deviations from the Bayesian learning scheme [Kahneman and A., 1973], framing effects and misleading accounting for risk [Kahneman and Riepe, 1998] are just a few examples of revealed deviations from ‘pure’

¹It is relevant to highlight the difference between beliefs and expectations. The latter is a well-defined mathematical operator, while the former has more the character of subjective assessment.

rationality. The important aspect here is that those deviations are *not random*, but rather systematic errors in the decision making processes. From a collective perspective, social interactions amplify those biases, since individuals are typically influenced in their decisions by ‘the others’ (cf. Shiller [1984]). The idiosyncrasy of errors of non-rational investors is, therefore, a first important assumption of the EMH which might not be in conformity with the reality.

Taking into account this empirical evidence, a new view of financial investors has been developed. The field of Behavioral Finance aims to analyze the connections between the behavior of the traders and the process of price formation. In the models developed within this framework, the market is populated by (at least) two categories of traders: informed traders, called also sophisticated, smart, fundamental, arbitragers or rational traders, and non-informed investors, called also noise, chartistic, naive or liquidity traders, following the imagination of the various authors. Within the Behavioral Finance approach, the heterogeneity of the market participants is never neglected. The presence of non-rational traders and the correlation of their ‘errors’ rise the question whether the arbitragers have enough power to dominate the market and drive out the other participants. In other words, is the arbitrage mechanism able to generate the informationally efficient market also in the presence of non-rational interacting heterogeneous traders? If not, can the uninformed traders survive in this market?

De Long et al. [1990] proposed a model of financial market to shed light on this issue. They showed that, under limited arbitrage capability of the informed traders², a group of interacting traders with misleading knowledge of the underlying return distribution might have higher expected returns than the sophisticated investors. The presence of an endemic market risk created by the uninformed traders and the contemporaneous limited capability of the arbitragers are the key factors for the long run surviving possibility of the uninformed traders. Although this model does not give a definitive answer to the issue of the sustainability of heterogeneity in the market, it helps to understand the crucial role that **heterogeneity** of the traders and their **social interactions** plays in how financial markets work.

From an empirical perspective, the EMH predicts that price changes are due to incoming news or, put differently, stock prices do not react to non-information. As a consequence, within the EMH, the time series of returns must be the perfect images of the news hitting the market. Therefore, one might ask whether it is plausible that informationally efficient prices would give rise to the long list of extremely robust statistical findings, so called stylized facts, such as the conditional structure of the volatility—from the ARCH effect, to the multi-scaling of the level of fluctuations of returns—or the fatness of the tail of returns distribution (cf. Pagan [1996]). The presence of those complex empirical regularities embedded in the time series of prices may cast doubts on the simple one-to-one relationship between price changes

²The investors act within an overlapping generation framework. They are forced to sell their assets at the end of the period and, therefore, they have time-constraints in managing their financial positions.

and information as implied by the EMH. If we assume that the ‘relevant’ information is made up of a collection of non-correlated information —economical, political and even meteorological— it is hard to justify that such a composite ‘assortment’ of news possesses the complex temporal structure observed for volatility. Anyhow, a strict empirical validation of such a relationship is practically impossible, since the information process is not directly observable.³

These empirical findings might alternately be viewed as the imprint of an endogenous dynamics of the market not related to fundamental factors. This paradigm has been called ‘Interacting Agents Hypothesis’ (IAH) by Lux and Marchesi [1999], ‘Adaptive Beliefs System’ by Brock and Hommes [1997] and ‘The Adaptive Markets Hypothesis’ by Lo [2004]. Within this approach, several authors have attempted to model financial markets as a system of heterogeneous interacting agents, whose activities might be responsible for this intrinsic force.⁴ Proposed dynamic market models typically differ in the degree of heterogeneity of traders (for instance in the number of different types of investors) or in the way they interact (for instance, following the herd or exchanging information via genetic algorithms etc.). Despite all these differences, many of them can successfully replicate the key stylized facts and explain their universality as an emergent property of the interactions among traders.

Since many different models can reproduce the empirical regularities, one might ask whether it is possible to precisely quantify their ability to describe the data and to compare the different models. However, one of the main drawbacks of the agent-based approach is related to the complexity of the interactions, which typically prevents an analytical solution, leaving only the possibility for Monte Carlo simulations based on a rough calibration of the underlying parameters (see e.g. LeBaron [2000]). It is, then, very difficult to directly compare different models, or to assess their goodness-of-fit. Moreover, the complexity of the underlying structure of those models does not permit to clearly identify the fundamental factors that drive their dynamics.

A direct estimation of the parameters of agent-based models is, therefore, largely missing in the pertinent literature. As far as we know the first attempt is a recent contribution by Gilli and Winker [2003], who estimate some of the parameters of Kirman’s seminal herding model⁵, via an *indirect* simulated method of moments approach. Very recently, the literature on estimation of agent-based models counts

³Some authors have attempted to analyze this relationship. Roll [1984] has examined the future market of orange juice, where he claimed that the information process is mainly influenced by news about weather; he found that, although weather news help to determine future prices, they account for a small component of their movements. Cutler et al. [1991] analyzed the 50 largest one-day stock price movements in the US stock market, finding that the majority of them happened after days of no major announcements.

⁴A long, however partial, list of contributions in this vein ranges from the (very) early papers of Baumol [1957] and Zeeman [1974], to recent research on noise traders, fundamentalist/chartist interaction and ‘artificial’ financial markets (Arthur et al. [1997], De Long et al. [1990], Kirman [1993], Beja and Goldman [1980] being some prominent examples).

⁵Kirman [1991, 1993].

a number of new papers: Westerhoff and Reitz [2003], Boswijk et al. [2005].

Taking stock of the mentioned limitations of the agent-based approach, we will develop an analytical model of heterogeneous interacting agents to explain the key stylized facts of financial data along the line of the IAH. Those empirical regularities are described in the first two chapters: The first is devoted to the distributional properties of the returns, while in the second the temporal properties are analyzed. This part of the thesis focuses mainly on the robustness of these empirical regularities, by means of an extended analysis of a large pool of real time series.

The second part of the thesis is devoted to the development of an artificial financial market model with heterogeneous interacting agents. The heterogeneity of the traders is given by the different investment strategies they can adopt. More precisely, they are divided into two groups: fundamentalists and technical traders, following the well-established chartists-fundamentalists framework introduced by Beja and Goldman [1980]. Moreover, our agents might change attitude according to mutual social interactions. The interaction mechanism is closely related to the herding model proposed by Kirman [1993]. It is based on the tendency of the single individual to follow the ‘suggestions’ of his environment or to change his opinion because of idiosyncratic shocks. The main ingredient is related to the use of a probabilistic approach in order to describe the microscopic behavior of individuals. In comparison to the representative agent paradigm, which is based on the optimization calculus of independent agents, two advantages emerge out of the implementation of a probabilistic framework. On the one hand, we can introduce heterogeneity in the model going beyond the representative agent paradigm. On the other hand, we can analytically deal with the aggregation of this underlying heterogeneity by means of statistical methods, given the stochastic nature of the agents’ behavior. In chapter 4, we will, then, explain the mathematical apparatus necessary for the analytical description of the aggregation of the agents’ behavior. The following chapter is devoted to the implementation of the herding mechanism in an agent-based financial market and to the illustration of the connection with the stylized facts. The last chapter is devoted to a ‘rough’ estimations of some of the underlying parameters of the model, given the analytical solution of the unconditional distribution of returns.

Our main contribution to the agent-based literature is the development of an analytical model, that is able not only to reproduce the key stylized facts of financial markets, but also to precisely identify their origin in terms of the agents’ interactions. Thus, the analytical solution enables us to understand the links between the aggregation process of the heterogeneous micro-economic world and the emergence of macroscopic statistical regularities.

CHAPTER 2

Distributional Properties of Financial Returns

The statistical approach in modeling financial data started one century ago when Louis Bachelier wrote his Ph.D. Thesis “*Théorie de la Spéculation*” [Bachelier, 1900], proposing the Normal distribution for the price variations. His fundamental contribution comes from modeling the apparent unpredictability of price changes assuming explicitly an underlying random motion of financial time series. Formally, if we denote with $p(t)$ the price of the asset at time t , the hypothesis of Normal distributed variations is a natural consequence of the Central Limit Theorem (CLT in the following), considering daily increments $\Delta p(t, \Delta t)$ over an interval Δt , as a sum of a large number of random shocks δp_i at the intra-daily level, drawn independently from a common distribution $\rho(\delta p_i)$ with mean zero and finite variance. The Bachelier’s formalization is given by:

$$\Delta p(t, \Delta t) := p(t + \Delta t) - p(t) = \sum_{i=1}^N \delta p_i, \quad (2.1)$$

where

$$\delta p_i = p(t + i \delta t) - p(t + (i - 1)\delta t), \quad i = 1, \dots, N, \quad N\delta t = \Delta t. \quad (2.2)$$

Following Bachelier, the price dynamics is, consequentially, governed by a **random walk process** using the terminology of the modern theory of stochastic processes—see the book of Feller [1971]—defined as:

$$p(t + \Delta t) = p(t) + \varepsilon(t + \Delta t), \quad \varepsilon(t) \sim N(0, \Delta t \sigma^2), \quad (2.3)$$

where the summation in eq. (2.1) is now replaced by a random variable $\varepsilon(t)$, drawn from a Normal distribution with mean zero and variance that grows linearly with the considered time interval Δt . The CLT assures, in fact, that, the sum of a large number of δp_i , independently on the functional form of $\rho(\delta p_i)$, is well approximated by a Normal distribution (see the paragraph on the CLT in the following). From the stability-under-addition property of the Gaussian, it follows that the shape of the distribution of price increments on a certain time resolution (daily, weekly or monthly for instance) is again Gaussian, given the suitable scaling of the variables. More precisely, the value of the variance on a daily level, for instance, together with its linear scaling with time interval is enough to fully characterized the distribution of price increments and the price dynamics of the considered asset for higher levels

of time aggregation. The Bachelier's approach in modeling the price dynamics is based mainly on the interplay between the assumed pure randomness of the process at small time scale, and the emergence of non-random regularities at aggregated level.

The price increments were afterward replaced by relative price changes or logarithmic increments, called returns, as key variables for a statistical description of financial data, introducing then the geometrical Brownian motion as more appropriate model for the price dynamics. The Bachelier's assumption, in fact, considers implicitly that the variance of the price increments is independent of the price level. However, the intuition suggests that the degree of uncertainty of the expected return should be constant regardless of the price level. Therefore, the standard deviation of the increments should have an extra linear dependence on the price, and eq. (2.3) becomes:

$$\ln \left(\frac{p(t, \Delta t)}{p(t)} \right) = \varepsilon(t), \quad \varepsilon(t + \Delta t) \sim N(0, \Delta t \sigma^2). \quad (2.4)$$

The stationary variable will now be the log-difference of prices, and not the price increment itself. Anyway, the Gaussian assumption for price variation is simply translated into Normally distributed returns. The geometric Brownian motion is considered nowadays the prototype model for the dynamics of the price of a financial asset. In this chapter and in the following one, we will see that real financial time series exhibit important deviations from the 'pure' Gaussian geometric random-walk model in both their unconditional and conditional properties.

In this chapter we will focus on the statistical characterization of the unconditional distribution of returns, neglecting any information on time development and serial correlation among data. It means that every return is considered as an independent realization from a common distribution (IID-ness approximation). The continuous geometric random walk benchmark implies that such a distribution is Gaussian, as required by the Central Limit Theorem. However, it has been repeatedly observed in various market data that the unconditional distribution of returns substantially deviates from Normality: small and very large returns are, in fact, much more frequent than under the Gaussian hypothesis, rendering to the empirical distribution its peculiar leptokurtic shape. Many parametric models have been employed to characterize this ubiquitous finding, starting from the Lévy stable distribution family proposed by Mandelbrot [1963]. However, the parametric approach constrains the empirical analysis, that might be strongly "distorted" by the parametric model itself¹. Recent applications of non-parametric techniques, which by construction make very few ad-hoc hypothesis on the data, reveal new qualitative properties of financial time series. In that respect, the issue of modeling the unconditional return distribution has gained a considerable number of new insights from this changed viewpoint.

Through the development of this chapter, we will describe how these new techniques can be applied for characterizing the unconditional properties of financial returns, in

¹For example, the non-Gaussian stable model limits the value of the tail index to the interval (0,2).

particular to the analysis of extreme events. In the first section 2.1 we will introduce the peculiar property of leptokurtosis of empirical distributions, together with a number of models to describe this empirical finding, detailed in section 2.2. The key results of the Extreme Value Theory are presented in section 2.3 in connection with the Central Limit Theorem. Its recent applications to financial data and the development of new non-parametric methods based on it will be also presented. Finally, an empirical investigation of the Tokio stock market, based on these non-parametric techniques, is detailed in section 2.4.

2.1 Leptokurtic Distributions and Fat Tails

Let us start with a descriptive statistic of a representative pool of financial series, summarized in Table 2.1. A first glance to its entries shows that the means of the considered time series are close to zero, as compared to the standard deviations, and the absolute values of the skewness are significantly smaller than one. At first approximation, we can conclude that the empirical distributions are *well behaved, uni-modal and symmetric distributions around zero*.

	Mean	Std	Skewness	Kurtosis	J-B norm.	p-value
Gold	$1 \cdot 10^{-4}$	$1 \cdot 10^{-2}$	-0.02	11.39	5.12	0.0
Dax	$3 \cdot 10^{-4}$	$1 \cdot 10^{-2}$	-0.32	10.45	4.48	0.0
DB	$2 \cdot 10^{-4}$	$1 \cdot 10^{-2}$	-0.29	8.10	2.55	0.0
Siemens	$3 \cdot 10^{-4}$	$1 \cdot 10^{-2}$	-0.11	7.40	2.04	0.0
USD/DEM	$-8 \cdot 10^{-5}$	$0.7 \cdot 10^{-2}$	-0.16	4.14	0.70	0.0
Dax-Hf	$-1 \cdot 10^{-6}$	$2 \cdot 10^{-4}$	-0.76	73.61	225	0.0

Table 2.1: Results of Jarque-Bera test, with the descriptive statistics. The vanishing p-value for all cases underlines the limited power of the Normal distribution in fitting the empirical data. The J-B statistics is divided by the sample size. All the sample estimates of the mean, standard deviations, skewness and kurtosis refer to the time series of log-returns.

Giving our theoretical benchmark, the Gaussian random walk, the first question that arises in dealing with the financial data is: “Are returns Normally distributed?” [Pagan, 1996].

Figure 2.1(a) shows the time development of the log-returns from the Deutsche Bank. A simple inspection of this time series casts serious doubts on the validity of the Gaussian hypothesis, since we observe much more events outside the $\pm 3\sigma$ band, than expected under this hypothesis². The non-Gaussian character of the marginal distribution of returns, almost immediately noticed in the earlier stage of research in empirical finance (cf. Mandelbrot [1963] and references therein), is nowadays a consolidated statement in the literature [de Vries, 1994, Pagan, 1996, Lux and Ausloos, 2002]. It is easy to show, using a simple Normality test, like

²Under Normality, 99.73% of the data points should be included in this band. We expect on average 18 events - the number of returns is 6771, while the realized value is 113, almost 7 times higher!

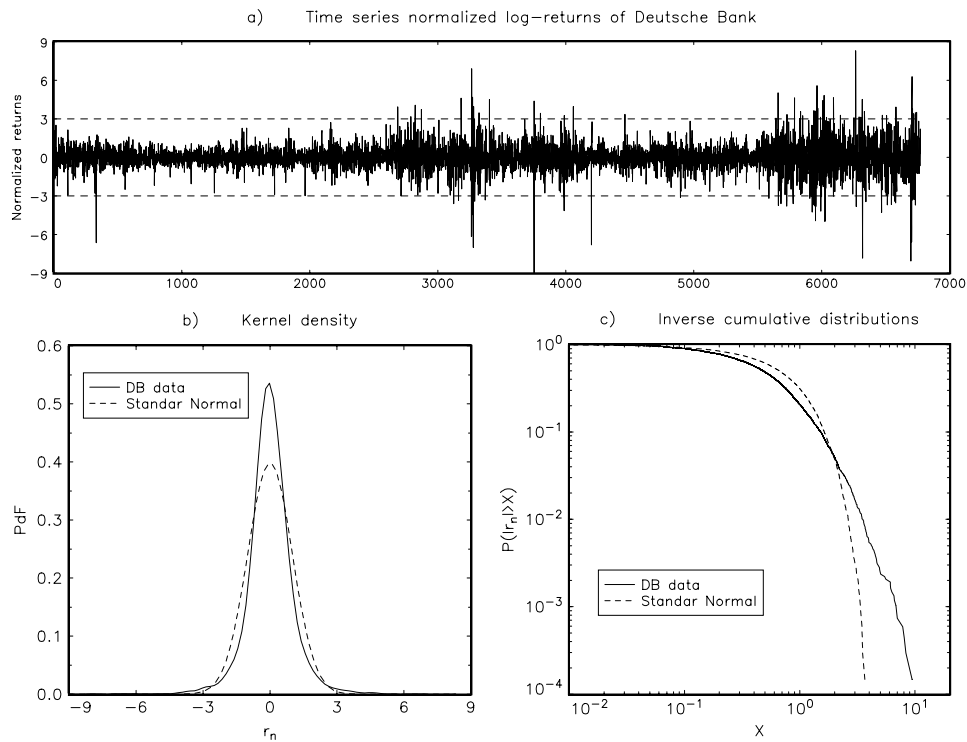


Figure 2.1: Panel (a) shows the time series of normalized daily log-returns for the Deutsche Bank. The dashed lines refer to the $\pm 3\sigma$ region. Sample probability density of normalized returns and a standard Normal are shown in panel (b). The leptokurtic shape of the empirical distribution is evident. Panel (c) shows the inverse cumulative distribution. Note that the tail of the empirical distribution considerably deviates from the behavior predicted by the Normal distribution.

the Jarque-Bera test [Jarque and Bera, 1980], that the null hypothesis of Normally distributed returns is rejected at a high significance level. This test is based on two measurements of deviation from the Gaussian behavior: the skewness, s , which accounts for the degree of asymmetry, and the kurtosis, κ , which quantifies the shape of the dispersion around the mean. The results of this test applied to the previous time series are given in Table 2.1. Interestingly, for our representative pool of data, the probability density functions exhibit a leptokurtic shape, i.e. a concentration of the probability mass in the center and in the tails of the distribution when compared to the Gaussian — as illustrated in Figure 2.1(b) for the Deutsche Bank. This finding is not limited to the analyzed time series, but it is observed in almost all financial data [de Vries, 1994]. It constitutes the first stylized fact: **leptokurtic empirical distributions**. In Figure 2.1(b) a normalized empirical distribution is compared with a standard Normal to illustrate its degree of leptokurtosis. Associated to the graphical representation of non-Normality, one defines a quantity to measure the deviation from the Gaussian behavior of a certain distribution $p(x)$: the kurtosis. It is defined as the normalized fourth cumulant³:

$$\kappa = \frac{E[(x - m)^4]}{\sigma^4} - 3, \quad (2.5)$$

where $m = E[x]$ and $\sigma^2 = E[(x - m)^2]$ are the mean and the variance of $p(x)$, respectively. Typically a symmetric bell-shape distribution can be classified as mesokurtic, leptokurtic or platykurtic, if $\kappa = 0$ (like-Gaussian behavior), $\kappa > 0$ (a narrow pick) or $\kappa < 0$ (a flat region around the mean), respectively. For “almost” symmetric normalized distributions, $p_n(x)$, — those with mean zero, variance one and negligible skewness — the first three cumulants coincide with the ones of the standard Normal. Hence the kurtosis can be directly interpreted as a measure of the “distance” between the distribution $p_n(x)$ and the Gaussian. All the time series listed in Table 2.1 exhibit a significative positive value of kurtosis that signals deviation from Normality.

	Gold	Dax	DB	Siemens	USD/DEM	Dax-Hf
$\hat{p}(0)$	0.73	0.47	0.53	0.51	0.54	1.18
\hat{F}	0.18	0.12	0.14	0.13	0.14	0.23

Table 2.2: \hat{F} is the fraction of standardized data in the interval $(-0.1257, +0.1257)$, and $\hat{p}(0)$ is the non-parametric kernel density at the origin.

However, it has been recently recognized that the kurtosis is an inadequate measure of deviation from the Gaussian behavior, since it depends on the convergence of the fourth moment, that is not *a priori* guaranteed in the empirical distributions; it depends, in fact, on the rate of decay of their tails. In order to overcome this problem, Pagan [1996] proposed two alternative quantities to account for deviations

³For the definition of the cumulants and their properties see the book of Feller [1971].

from Gaussian behavior of the standardized data: the first is the estimated fraction of realizations lying in the interval $\pm 0.1257, \hat{F}$. The second is the non-parametric density estimate at the origin, $\hat{p}(0)$. For a standard Normal random variable, they take the values 0.1 and 0.4, respectively. Table 2 shows the values of $\hat{p}(0)$ and \hat{F} for the considered time series, that are in line with those usually reported in the literature [Pagan, 1996], confirming the leptokurtic character of these distributions. These two alternative measures are easy to compute and, moreover, do not depend on the high moments, whose existence might be doubtful.

Pagan's suggestion of concentrating on the behavior of the peak around the mean omits the influence of the tail, which is a more informative region of the distribution (see the discussion on the Central Limit Theorem). Figure 2.1(c) illustrates, more in detail, the departure from the Gaussian behavior of the outer part of the empirical distribution, showing the presence of a larger portion of extreme events than expected under Normality. This finding goes under the name of *fat tail phenomenon*, that constitutes another ubiquitous regularity of financial data. It is important to highlight that, although strictly related, the two characterizations of the deviation from Normality, namely the leptokurtic shape and the fatness of the tails, focus attention on two different aspects. Whereas leptokurtosis refers to the behavior of the entire distribution, the tail characterization involves just a small fraction of extreme events belonging to its outer part. Far from being just an alternative graphical illustration to describe non-Normality, the analysis of the tail constitutes a new paradigm in the investigation of financial data (see section 2.3). Given this type of deviation from Gaussianity of the unconditional empirical distributions, it is necessary to introduce alternative measures, than the variance or kurtosis, in order to capture the dispersion of returns. A relative recent quantity is the *tail index*, defined as the highest finite absolute moment:

$$\alpha = \sup\{k > 0, \mu_k < +\infty\} , \quad (2.6)$$

where μ_k is the k-th central moment of absolute returns, given by the expression:

$$\mu_k = E[(|r - E[r]|)^k] . \quad (2.7)$$

The tail index, as a measure of the dispersion of returns, possesses several remarkable properties: i) it is a measure of the fatness of the tail; ii) it characterizes the behavior of extreme realizations (see the paragraph on extreme value theory); iii) it overcomes the problem of possible non-stationarity of the kurtosis. From empirical side, the estimates of α are concentrated on a very narrow range of variability between 2.5 and 5 (see section 2.4).

Following the historical development, in the early 1960s, Mandelbrot [1963] and Fama [1963] pointed out the insufficiency of the assumption of Gaussian distributed returns, showing up the heaviness of the tails of empirical distributions and their leptokurtic behavior, as well. Mandelbrot, in his famous article on the price of cotton, quoted several earlier papers dealing with the poor descriptive power of the Gaussian in modeling the return distribution — the first of them dated 1915

— where typically an anomalous amount of large entries non compatible with the Gaussian assumption was identified (see, for example Figure 2.1(a)). However none of these authors proposed an alternative hypothesis, as Mandelbrot did, when he introduced the symmetric Lévy stable family. Conversely, it was a common approach to eliminate the outliers and to maintain, then, the Gaussian assumption for the filtered data. Quoting Mandelbrot [1963]:

“One very common approach is to note that, *a posteriori*, large price-changes are usually traceable to well-determined ‘causes’ that should be eliminated before one attempts a stochastic model of the remainder. Such preliminary censorship obviously brings any distribution closer to the Gaussian. This is, for example, what happens when one restricts himself to the study of ‘quiet periods’ of price change. There need not to be any obvious discontinuity between ‘outliers’ and the rest of the distribution, however, and the above censorship is therefore usually undetermined.”

Skipping the hypothesis of Normally distributed returns of Bachelier, he preserved the stability-under-addition property and the central limit law in a *generalized* form, including distributions with infinite second moment. Analogously to the role of the Gaussian, the Lévy stable family is, in fact, an attractor in the space of distributions with infinite variance. In other words, a sum of independent random variables, drawn from a common distribution $p(x)$ with a maximum finite moment $\mu < 2$, converges asymptotically to a Lévy distribution with index⁴ μ .

Like in the case of Bachelier’s assumption, the essence of the Mandelbrot’s idea to model the empirical distribution of returns in terms of the Lévy stable family is based on arguments involving the invariance under time aggregation, in a generalized perspective. However, the lack of convergence of the variance of this kind of distributions prevents the use of many standard estimation procedures. In order to empirically justify the use of this parametric family, Mandelbrot analyzed the sequence of the recursive variances (sequential second moment), formed by adding on one observation at a time, which should converge to the unconditional value. If the second moment is not defined (like in the Lévy processes), then there is no asymptotic convergence, and one tends to observe large “jumps” in the sequential second

⁴The symmetric Lévy distributions are labeled by an index μ , identified with the maximum finite moment. From a mathematical viewpoint, the Lévy distribution can be conveniently described by its characteristic function (see Mantegna and Stanley [2000], Bouchaud and Potters [2000] and note 13), that, for the symmetric case, has a rather simple analytical expression

$$f_{\mu}(z) = \exp(-a_{\mu}\gamma|z|^{\mu}), \quad (2.8)$$

where a_{μ} is a constant dependent on μ , and γ is a scale parameter. It has been shown analytically that the tail behaves as a power law

$$p(|x|) \propto |x|^{-(1+\mu)}, \quad (2.9)$$

with a characteristic exponent $\mu < 2$, which prevents, in fact, the existence of the second and higher moments. As a result, the Lévy stable distribution has the desirable property of invariance under time aggregation, and, additionally, a leptokurtic shape and fat tails.

moment, as illustrated in Figure 2.2, as well as in the higher moments. Moreover,

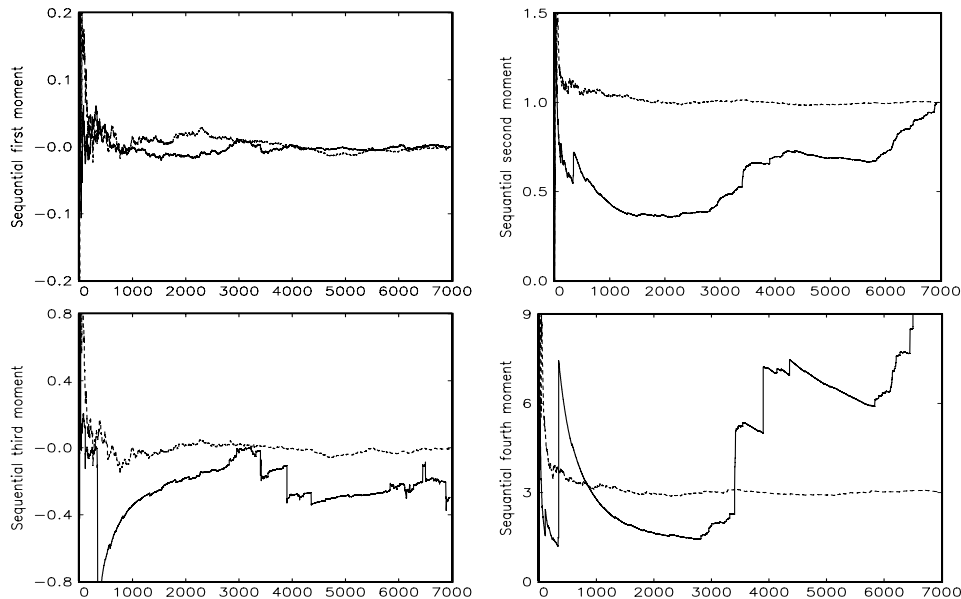


Figure 2.2: *The four panels show the temporal evolution of the sequential moments of normalized daily data from Deutsche Bank (solid line), and of a standard Normal random variable. Note that, except for the first moment, the sequential moments of real data exhibit much larger fluctuations than the correspondent moments of simulated Gaussian data, which might be the sign of non-stationarity. The presence of such large “jumps” in the sequential second moment has been employed by Mandelbrot as an evidence for the Lévy stable assumption.*

Mandelbrot has shown the apparent stability of the shape of the empirical distributions and the power law decay of the tail, by plotting the estimated probability density functions in a log-log scale for different levels of time aggregation. He found a tail index of $\mu = 1.6$ for the price of the cotton. A more extensive analysis of several stocks from New York Stock Exchange was later performed by Fama [1965], that seemed to confirm the stable Paretian hypothesis.

Mandelbrot’s assumption was very convincing from a theoretical viewpoint, based on the stability argument, and from several empirical evidences. Nevertheless, further studies based on cleverly designed tests for stability-under-addition, or on the speed of divergence of higher moments, pointed out the insufficiency of the Lévy stable assumption [Friedman and Vandersteel, 1982, Hall et al., 1989, Hsu et al., 1974, Lau et al., 1990, Upton and Shannon, 1979]. Recent non-parametric analysis of the tail behavior (see the following sections, specially 2.4) converge on the common conclusion of finiteness of the second moment, confirming, from a different viewpoint, the limited validity of the stable Paretian hypothesis. Even so, the main merit of Mandelbrot was to recognize that the occurrence of large events, non compatible with the Gaussian assumption, could not be associated to a simple presence of few outliers, but rather to a universal feature of the data, important enough to justify the introduction of a total new approach to describe them.

2.2 Alternative Hypothesis to Stable Distributions

The stability-under-addition property and the Central Limit Theorem give to the Bachelier's hypothesis and the Mandelbrot's generalization the necessary character of universality of a scientific theory. In fact, invoking the CLT, it is possible to abstract from the details of the particular distribution governing the high frequency returns. Moreover, the property of stability eliminates the dependence on the time resolution of the considered data, providing the suitable scaling. However, if the rejection of the stable Gaussian hypothesis for returns is a rather old finding, the insufficiency of the Lévy stable laws and its implications are relatively new results. It coincides, indeed, with the recent and easier availability of high frequency records (minute to minute or even transaction to transaction), which gives the opportunity to analyze a larger amount of data with non-parametric or semi-parametric techniques, and to investigate earlier stages of time aggregation. An excellent reference in this field is the book of Dacorogna et al. [2001], in which they describe in great detail the new techniques and recent insights concerning the analysis of high frequency data. In the next section some of these methods are introduced within the general framework of Extreme Value Theory (EVT hereafter).

Many other parametric distributions have been used to model the unconditional distribution of returns, going beyond the paradigm of stable distributions. They are driven more by a mathematical convenience to fit the empirical data, than guided by a strict statistical justification of the underlying assumptions. The common framework is the IID-ness of the random variables that exclude from the subsequent list stochastic processes with any kind of memory:

- *The Student-t distribution* with more than two degrees of freedom. The main feature of this parametric family is the power law behavior of the tail together with the leptokurtic shape, and the lack of stability under aggregation⁵. It has been shown that such a model can approximate with a good accuracy the empirical distributions for different time resolutions [Blattberg and Gonedes, 1974, Bouchaud and Potters, 2000].
- *The truncated Lévy flight* (TLF hereafter). The central part is described by a Lévy distribution, however all the moments of this distribution are finite, since it is truncated (i.e. it possesses an exponential decay of the tail or vanishing probability to observe values of the variable higher than a given threshold). The functional form of the re-scaled sum of a large number of variables distributed according to a TLF is still a TLF. Roughly speaking, this process is “almost” stable when the number N of elements of the sum is lower than a critical value N_c , since it is “close” to a genuine Lévy stable process, while it is close to a Gaussian behavior for $N \gg N_c$. Typically N_c is of the order of several hundreds⁶. This parametric family has been

⁵A mathematical proof of the convergence to a Gaussian together with the stability of the functional form of its tails is given in section (2.3.1).

⁶The critical value is $N_c \approx A(\alpha)l^\alpha$, where l is the cutoff length of range of the Lévy process with index α and $A(\alpha)$ is a function depending solely on the index α . The order of magnitude of N_c

successfully employed to describe the empirical distributions of high frequency data, that show an approximate Lévy behavior in the central part and a very slow convergence to the Gaussian under aggregation (see the book of Mantegna and Stanley [2000] and references therein).

- *Hyperbolic distribution*, which lies in between a Gaussian body and exponential tails. It is a leptokurtic distribution with a bounded excess of kurtosis (between zero and three, see Bouchaud and Potters [2000]). The distinguishing feature of this parametric family is that the logarithm of its density forms hyperbola. Note that the logarithm of a Gaussian is a parabola, which is the limit case of an hyperbola. Sørensen and Bibby [2003], among others, employ an hyperbolic distribution family to model the return distribution.

The zoology of contributions in the literature is richer than this very short review and it indicates the large degree of uncertainty on the underlying “true” distribution of returns; the existence of such a distribution is by itself another important issue. It is remarkable that, despite the vast heterogeneity of the models, the Gaussian random walk is *always nested* in all processes, that underlies once again the universal nature of this reference model.

In the following sections a new approach is presented, which avoids the more ambitious goal of modeling the entire distribution. The alternative viewpoint concentrates just on a small fraction of more informative extreme events.

2.3 The Tail Index of Continuous Distributions and the Extreme Value Theory

The demanding task of modeling the entire return distribution can be fruitfully simplified if one concentrates only on its tails. In the last decade, this approach has gained considerable attention, both from a practical and a theoretical viewpoint. The underlying philosophy can be summarized by the motto “let the data speak for themselves”. The idea is to use non-parametric techniques instead of estimating an optimal parameter of a potentially misspecified model. Its theoretical background is based on the Extreme Value Theory, which allows for a *one-parameter classification* of the continuous distributions based on the limiting behavior of their extreme realizations or, equivalently to some extent, their maxima and minima⁷. Moreover, it gives the possibility of abstracting from the details of the particular distribution, allowing, then, to use non-parametric or semi-parametric techniques (Hill estimator, for example) in order to estimate this key-parameter. On a more applied perspective, practitioners are mainly interested in the evaluation of the downside risk, i.e. to asses a precise likelihood for negative extreme events⁸, rather than evaluate the exact functional form of the minor fluctuations. The analysis of the tail constitutes,

of several hundreds refers to its estimation using high-frequency financial data (see [Mantegna and Stanley, 2000]).

⁷See theorems 1 and 2.

⁸The popularity of the recent VaR technique is a signal of this change in viewpoint.

nowadays, a crucial issue in risk management.

This section is devoted to illustrate the main results of the EVT, and its applications to the empirical analysis of financial data. In particular, we will focus on the characterization of the outer part of the return distribution. It is worth to highlight that the results are once again confined to the hypothesis of independent or weakly dependent data.

2.3.1 The Central Limit Theorem

The CLT states that a sum of N random variables, $x_N = \sum_{i=1}^N x_i$, independently drawn from a common distribution $p(x)$ with mean μ and finite variance σ^2 , converges to a Normal distribution with mean $N\mu$ and variance $N\sigma^2$, as N goes to infinity⁹. Formally:

$$\lim_{N \rightarrow \infty} P \left(u_1 \leq \frac{x_N - \mu N}{\sigma \sqrt{N}} \leq u_2 \right) = \frac{1}{\sqrt{2\pi}} \int_{u_1}^{u_2} e^{-\frac{1}{2}u^2} du \quad (2.10)$$

for all finite u_1 and u_2 . The form of the original distribution $p(x)$ does not affect the asymptotic behavior (that in any case is Gaussian), but only the speed of the convergence. The CLT only concerns the *central* part of the distribution of the sum X_N , but it does not provide information on the behavior of its extremes. In particular, the CLT does **not** imply that the laws governing the extreme realizations will be those governing the extremes of the limiting Normal distribution. Rather the tail behavior of the original distribution $p(x)$ might be preserved under certain conditions despite the overall attraction towards a Gaussian shape. This means that, in the aggregation process, the region around the mean is gradually better approximated by the Gaussian. Conversely, the outer part, independent of its functional form, might still follow a different type of extreme value distribution, although it is expelled towards large values of X_N , becoming consequently less and less ‘visible’. The next two soluble examples illustrate the dynamics of X_N towards the Gaussian regime, underlying the different behaviors of the region around the mean and the tails of the distribution.

Example (I) Let us consider, as a first example, a positive random variable that follows an exponential distribution given by:

$$p(x) = \begin{cases} 0 & x < 0 \\ \alpha e^{-\alpha x} & x \geq 0, \end{cases} \quad (2.11)$$

where $\alpha \neq 0$. The mean is $\mu = \alpha^{-1}$ and the variance is finite: $\sigma^2 = \alpha^{-2}$. The sum of N variables distributed according to (2.11) is then a Gamma distribution¹⁰:

$$p(x, N) = \begin{cases} 0 & x \leq 0 \\ \alpha^N \frac{x^{N-1} e^{-\alpha x}}{(N-1)!} & x > 0. \end{cases} \quad (2.12)$$

⁹See Feller [1971] for a generalization of the CLT to dependent and non-identical distributed variables, and to distributions with infinite variance.

¹⁰The general form of a Gamma distribution is $p(x) = a^b \frac{x^{b-1} e^{-ax}}{\Gamma(b)}$, where a and b are positive real number, and $\Gamma(b)$ is the Gamma function.

It is important to emphasize that the probability of negative realizations from such a distribution is zero, and, the outer part decays exponentially. These two elements seem to be in contradiction to the Gaussian shape, where the distribution has to converge to. Nevertheless, the Gaussian approximation holds in the central region, as we can show by expanding the distribution (2.12) around the most probable value:

$$\frac{d}{dx}p(x, N)|_{x^*} = 0 \quad x^* = (N - 1)\mu. \quad (2.13)$$

The Taylor expansion of $p(x, N)$ around its maximum x^* is given by

$$\ln[p(x, N)] = -\Omega(N - 1) - \ln(\mu) - \frac{\alpha^2(x - x^*)^2}{2(N - 1)} + \frac{\alpha^3(x - x^*)^3}{3(N - 1)^2} + O(x - x^*)^4. \quad (2.14)$$

The first order term in the expansion (2.14) is obviously zero, since it is done with respect to the maximum of the distribution. The function $\Omega(N)$ reads as:

$$\Omega(N) = \ln(N!) + N - N \ln N \approx \ln(\sqrt{2\pi N}), \quad (2.15)$$

where the last equality is written using the Stirling's approximation¹¹. If the third order term in the expansion is negligible, the distribution of x_N is well approximated by a Gaussian with mean $(N - 1)\mu$ and variance $(N - 1)\sigma^2$, as predicted by the CLT¹²:

$$\ln[p(x, N)] \approx -\frac{1}{2} \ln[2\pi(N - 1)\sigma^2] - \frac{1}{2} \left[\frac{x - (N - 1)\mu}{(N - 1)\sigma} \right]^2. \quad (2.16)$$

The condition of large N has to be satisfied together with the inequality:

$$|x - x^*| \ll mN^{2/3}, \quad (2.17)$$

which guarantees vanishing terms of order higher than the second in the expansion (2.14). Eq. (2.17) gives the order of magnitude of the ‘‘central’’ region in which the Gaussian approximation is satisfied. Eq. (2.17) is not a specific result limited to this example, but holds for asymmetric distributions with non-zero skewness. Note that the outer part of the distribution $p(x, N)$ decays exponentially like the original distribution $p(x)$, therefore much slower than a Gaussian. This result can be understood within the more general phenomenon of invariance of the tail under aggregation (see EVT). The Gaussian regime grows with N at the rate fixed by Eq. (2.17), and contemporary, the probability to observe the ‘true’ tail’s behavior goes to zero.

Example (II) The second example is more relevant for financial applications. Let us consider a Student-t distributed random variable with 3 degrees of freedom, $p(x)$ being its probability density function:

$$p(x) = \frac{2a^3}{\pi(x^2 + a^2)^2}. \quad (2.18)$$

¹¹The Stirling approximation holds for large factorials and is given by $N! \approx \sqrt{2\pi N} N^N e^{-N}$.

¹²To be precise, the CLT predicts a convergence to a Normal with mean and variance N times the mean and variance of the original distribution $p(x)$. However the relative difference between $(N - 1)$ and N vanishes for large N .

The outer part of the inverse cumulative distribution decays following a cubic power law, therefore, the moments higher or equal to 3 are not defined. As outlined previously, the Student-t parametric family has been used to model the empirical return distribution. The distribution is symmetric, with mean zero and with a finite variance $\sigma^2 = a^2$. The sum of N distributed random variables X_N converges, hence, to a Gaussian. In order to show it, let us derive the characteristic function¹³ of $p(x)$

$$f(z) = (1 + a|z|)e^{-a|z|}, \quad (2.19)$$

that expanded around $z = 0$ leads to

$$f(z) \approx 1 - \frac{z^2 a^2}{2} + \frac{|z|^3 a^3}{3} + O(z^4). \quad (2.20)$$

The non-analytical term $|z|^3$ is not surprising, giving the divergence of all the moments higher or equal to 3. As a general rule, the characteristic function of a distribution with an asymptotic power law decay of the tail with index β is non-analytical around $z = 0$. The Taylor expansion around zero contains regular terms z^n , with $n < \beta$ and a non-analytical element $|z|^\beta$. In order to get information about the behavior of the sum of N variables distributed following (2.18), we compute the characteristic function of the N th convolution¹⁴:

$$f(z, N) = (1 + a|z|)^N e^{-aN|z|}. \quad (2.21)$$

The expansion around $z = 0$ leads to:

$$f(z, N) \approx 1 - \frac{z^2 N a^2}{2} + \frac{|z|^3 N a^3}{3} + O(z^4). \quad (2.22)$$

If one considers just the first two terms in the right hand side of Eq. (2.22), it is exactly the small z expansion of the characteristic function of a Gaussian¹⁵, with mean zero and variance Na^2 , as predicted by the CLT. Nevertheless, the non-analytical term $|z|^3$, that indicates the divergence of the third moment and, consequently, the cubic decay of the tail, does not disappear under convolution. It means that, although the sum converges to the Gaussian, the tail of $p(x, N)$ retains the same power law decay of the original distribution $p(x)$. It is possible to show that the cut-off

¹³ The characteristic function, denoted in the following with $f(z)$, is the Fourier transformation of the probability density function of a stochastic variable.

¹⁴The convolution between two distributions of independent random variables X_1 and X_2 , distributed according to $p_1(x_1)$ and $p_2(x_2)$ is given by:

$$p(x) = \int p_1(x')p_2(x - x')dx'$$

where $x = x_1 + x_2$ and $p(x)$ is its distribution. The convolution assumes a simple form using the Fourier transforms, for which it becomes simply a product. Therefore, the characteristic function for the sum of N IID random variables is the N power of the characteristic function of the single variable.

¹⁵The characteristic function of a Normal distributed variable with mean zero and variance σ^2 is $f(z) = \exp(-\frac{1}{2}\sigma^2 z^2)$. The expansion around $z = 0$ is, therefore, $f(z) \approx 1 - \frac{1}{2}\sigma^2 z^2$

value $x_0(N)$ between the Gaussian regime and the power law behavior grows with N as:

$$x_0(N) \sim a\sqrt{N \ln(N)}. \quad (2.23)$$

This exercise can be generalized to any tail exponent higher than 2 (since for smaller values the convergence is towards the Lévy family), without changes in the overall qualitative behavior¹⁶.

This long digression on the CLT and its limitations is very useful for a better understanding of the Extreme Value Theory, described in the next paragraph. The invariance of the tail behavior under aggregation in the two previous examples, rather than a mathematical curiosity, is a general phenomenon explained by the EVT, perfectly compatible with the convergence of the overall distribution to the Gaussian. Moreover, throughout these two exercises, it is possible to realize that extreme events, those events that belong to the tail, are particularly informative about the underlying generating process. In the two previous examples, for instance, the tails bring information on the functional form of the original distribution $p(x)$, distorted by the aggregation procedure, that flattens everything to the Gaussian world. Taking into account this information, one can, in fact, discriminate among the pool of hypothesis concerning the generating mechanisms at the early stage of aggregation, avoiding in this way the bias created by fitting the overall distribution.

2.3.2 The Extreme Value Theory

In modeling the extreme events of a IID random variable, the Extreme Value Theory¹⁷ is the counterpart of the Central Limit Theorem for predicting the feasible limiting stable distributions. However, while the CLT is concerned with “small” fluctuations around the mean resulting from an aggregation process, the EVT provides a classification of continuous distributions according to the asymptotic behavior of their extremes. The theory predicts three limiting stable distributions for the maximum values of a random variable¹⁸, called Generalized Extreme Value Distributions (GEV), and three associated Generalized Pareto Distributions (GPD). The main difference between GEV and GPD lies in the method to empirically identify the extreme events in real data: the *Block Maxima Method* and the *Peak Over Threshold Method*¹⁹.

GEV: Limiting Distribution for Extrema

Let us consider a stationary sequence of IID variables $\{x_i\}_{i=1}^N$ with a common distribution function P . By dividing the entire dataset in L non-overlapping sub-samples, and taking the maximum from every sub-sample, we will end up with a subset of maxima M_1, M_2, \dots, M_L , so-called *block maxima*. The limit law of this sequence $\{M_j\}_{j=1}^L \equiv M(L)$ is given by the following theorem:

¹⁶An excellent reference for a detailed description of this aspect of CLT, in a financial perspective, is the book of Bouchaud and Potters [2000], where these two examples are taken from.

¹⁷Part of the material presented here is published in [Alfarano and Lux, 2004].

¹⁸It is also possible to derive equivalent results for the minimum values.

¹⁹For a detailed but not too technical description of the techniques see Gilli *et al.* Gilli and Këllezli [2003].

Theorem 1 (Fisher and Tippet, Gnedenko [Beirlant et al., 1996], [Reiss and Thomas, 1997]): If there exist two normalizing constants $c_L > 0$ and $d_L \in \mathfrak{R}$, and a non-degenerate distribution H such that

$$\frac{M(L) - d_L}{c_L} \xrightarrow{d} H,$$

where the subscript d indicates convergence in distribution, then H belongs to one of the following extreme value distributions:

$$\text{Fréchet: } G_{1,\alpha}(x) = \begin{cases} 0 & x \leq 0 \\ e^{-x^{-\alpha}} & x > 0, \end{cases} \quad (2.24a)$$

$$\text{Weibull: } G_{2,\alpha}(x) = \begin{cases} e^{-(-x)^\alpha} & x \leq 0 \\ 1 & x > 0, \end{cases} \quad (2.24b)$$

$$\text{Gumbel: } G_3(x) = e^{-e^{-x}} \quad x \in \mathfrak{R}. \quad (2.24c)$$

Based on the previous theorem, the distributions can be classified into three categories: (i) *heavy-tailed* distributions, if the extremes follow the first type of law; (ii) *short-tailed* distributions with finite end-point, if the extremes obey the Weibull's type; (iii) *medium-tailed* distributions, if the extremes are governed by the third category. Note that in cases (i) and (ii) we have a one-parameter family of distributions, parametrized by the shape parameter α . Representative members of the three groups are respectively: the Student-t, the uniform distribution and the Normal. The von Mises representation of GEV provides a unified formula for the previous three limiting distributions (2.24a), (4.70) and (2.24c):

$$G_\gamma = \exp[-(1 + \gamma x)^{-\frac{1}{\gamma}}]. \quad (2.25)$$

For a positive γ we recover the Fréchet distribution, negative γ corresponds to the Weibull type, and the limit case $\gamma \rightarrow 0$ describes the Gumbel formula. The shape parameters of the two representations are related to each other by the formula $\alpha = \frac{1}{\gamma}$ for the distribution (2.24a), and $\alpha = -\frac{1}{\gamma}$ for the type (4.70). The Von Mises approach turns out to be very useful given that it nests all three types of limiting behavior in a unified framework and, via estimation of γ , allows inference on the relevant limit laws.

GPD: Limiting Distributions for the Tail

The second set of results focuses on the tails of the distributions instead of maxima; the selected events are, in this case, those events that exceed a given threshold u . Let us first introduce the Generalized Pareto Distributions (GPD in the following) using the so-called α -parameterization:

$$W_{1,\alpha} = 1 - (x)^{-\alpha}, \quad x \geq 1, \quad (2.26a)$$

$$W_{2,\alpha} = 1 - (-x)^\alpha, \quad -1 \leq x \leq 0, \quad (2.26b)$$

$$W_3 = 1 - \exp(-x), \quad x \geq 0. \quad (2.26c)$$

All three distributions (2.26a) to (2.26c) assume the value zero outside the pertinent intervals. For the GPD a similar one-parameter representation exists as with the extreme value distributions²⁰:

$$W_{\gamma,0,1} = 1 - (1 + \gamma x)^{-\frac{1}{\gamma}} \quad (2.27)$$

where for $\gamma > 0$, $\gamma < 0$ and $\gamma = 0$ we recover the first, second and third group, respectively. γ is called the shape parameter. The relations between α and γ are again $\alpha = \frac{1}{\gamma}$ for the first type, and $\alpha = -\frac{1}{\gamma}$ for the second type.

The basic result for the limiting behavior of the tail region of a distribution is:

Theorem 2 (Pickands, Balkema and de Haan Beirlant et al. [1996], Reiss and Thomas [1997]): Let us define the exceedance distribution function of a continuous distribution $F(x)$:

$$F^{[u]}(x) = Pr(X \leq x | X > u)$$

If $F^{[u]}(a_u x + b_u)$ has a continuous limiting distribution function as u goes to the right-end point²¹ $\omega(F)$ of F , then $F^{[u]}$ converges to one of the GPDs distributions:

$$|F^{[u]}(x) - W_{\gamma,u,\sigma_u}(x)| \rightarrow 0 \quad u \rightarrow \omega(F)$$

for shape, location and scale parameter γ , u and $\sigma_u > 0$, respectively. The theorem can also be formulated in terms of the α -parameterization.

The previous limiting theorem allows for a classification of distributions according to the behavior of their tails: hyperbolic decline, if the distribution converges to the first type of GPDs, distribution with finite end-point, if it converges to the second type, or exponential decline, if its limiting distribution belongs to the third type.

To summarize, through Theorems (1) and (2), the EVT allows to abstract from the specific distribution governing the fluctuations of the overall system (that can be a financial market or a geological structure, for instance) when investigating extremes, and to concentrate solely on the behavior of large observations. Moreover, the GPDs formalization is very flexible in describing the tail's behavior, although it depends on one parameter only, the *index* α . Additionally, the EVT provides the analytical expression for the limiting behavior of the tail, giving the possibility, via several *non-parametric* estimation methods of the key-parameter α , to discern the proper functional form of the empirical tail out of the three theoretical feasible tails behavior.

Focusing on financial data, the EVT turns out to be extremely useful to simplify the complex problem of modeling the returns distribution. The characterization of

²⁰If a random variable X has a distribution function F , then $\sigma X + \mu$ has a distribution function $F_{\mu,\sigma} = F[(X - \mu)/\sigma]$, where μ and σ are the location and scale parameter, respectively. Eq (2.27) represent the standard form of the GPDs in the γ parametrization, where the location parameter is 0 and the scale parameter is 1.

²¹The right-end point $\omega(F)$ of a distribution is defined as $\omega(F) \equiv \sup\{x : F(x) < 1\}$.

the tail in terms of one of the above three GPDs can greatly reduce the “zoology” of models present in the literature (several examples are listed in paragraph 2.2).

As a final remark, the EVT can shed some light on the results of the previous examples. The convergence of the exponential distribution and the Student-t distribution towards the Gaussian is explain by the CLT. Conversely, the EVT gives a rationale for the invariance of the tail of those two distributions under aggregation. The exponential distribution, in fact, belongs to the basin of attraction of the third type of GPD specified by (2.26c), while the tail of a Student-t converges to the first type of GPD given by eq. (2.26a).

2.3.3 Estimation of the tail index

The estimation of the index α is an important issue in empirical research dealing with extreme events, since it allows to precisely quantify the likelihood of big fluctuations, and to have, consequently, a better control of the related risk.

However, any attempt of estimation of the tail index has to cope with three key problems. The first one arises from the operative definition of what is an extreme realization, i.e. one has to decide which events, from the complete set of data points, belong to the subset relevant for the estimation of α . This problem is denoted as the *threshold selection problem*. Then, it is necessary to construct an estimator that provides a compromise between a potential bias, due to the only approximate validity of the power law, and the high variance due to too small a size of the selected subset, since extreme events are, by their definition, rare. Finally, the method should also provide appropriate confidence intervals for the estimates.

There are several simple heuristic regression approaches for the estimation of the tail index, that often use graphical procedures: mean excess function and mean log-excess function, for instance (see Beirlant et al. [1996], Reiss and Thomas [1997]). Among them, the most popular method employs the linear behavior of the cumulative distribution on a log-log scale (further details below). Beside the fact that these methods give only a first rough assessment on the value of the parameter α , they might be sometimes useful for discriminating among the three different types of tail behavior, namely power law, exponential decay or finite end point. A more rigorous procedure is the one proposed by Hill [1975], that nowadays is the standard tool to analyze power laws in economics and other related fields. In the following section a detailed description of the method is given.

Linear regression

If a random variable with a distribution $p(x)$ possesses a power law decay of the tail, the inverse cumulative distribution $P(x > X)$, in a log-log scale (see Figures 2.3 and 2.4(a)), exhibits a linear behavior in its outer part. One of the most popular graphical methods to asses the value of the tail index is to exploit this scaling property, using a simple linear regression that extends in the extreme part of $P(x > X)$. The simplicity of the estimator is obviously related to the properties of Ordinary

Least Squares procedure (OLS), and clearly it constitutes one of the main advantages of the method. Moreover it can be assigned confidence intervals and a goodness-of-fit measure to the estimate. However, one of the fundamental assumptions required for the application of OLS is not entirely fulfilled: the data are evidently strongly dependent by construction, which limits the applicability of the method. Moreover, the OLS procedure is not robust under the presence of outliers. A simple numerical exercise can illustrate this statement. Adding up a large value (100) to the Student-t distributed random variables, introduced in the subsequent Monte Carlo simulation, we can mimic the presence of an outlier. The mean of the estimates dramatically changes in the case of OLS (with the outliers the mean value is 2.31 and without is 2.80), while it is almost unchanged for the Hill estimator (from 2.73 to 2.88 with or without outliers, respectively). The tail size is fixed in both cases by the condition $|x_t| > 5$, where x_t are random realizations drawn from a Student-t distribution (see section 2.3.5). Furthermore, the range of the approximated linear behavior is left to the arbitrariness of the user, a further drawback shared with the Hill estimator.

Hill estimator

The Hill estimator is the conditional maximum likelihood estimator for heavy-tailed distributions. If we assume that the data points exceeding a given threshold u follow a Pareto distribution with index α , the distribution of realizations exceeding u reads:

$$P(x) = 1 - \left(\frac{u}{x}\right)^\alpha \quad x \geq u. \quad (2.28)$$

As has been shown by Hill, the (conditional) maximum likelihood estimator of the parameter α in eq. (2.28) assumes the particularly simple form:

$$\hat{\gamma}_{k,N} = (\hat{\alpha}_{k,N})^{-1} = \frac{1}{k} \sum_{i=1}^k [\ln x_{(N-i+1)} - \ln x_{(N-k)}], \quad (2.29)$$

with $x_{(i)}$ the order statistics of the series x , $x_{(N)} > x_{(N-1)} > \dots > x_{(1)}$, i.e. $x_{(N)}$ is the maximum of x , $x_{(N-1)}$ is the second largest value etc. We only consider values above the threshold u , $x_{(N-k)} = u > x_{(N-k-1)}$, where k is the number of selected points, out of the entire sample of N realizations. It has been shown that under some mild additional restrictions on the behavior of the underlying distributional function, $\hat{\gamma}_{k,N}$ is asymptotically Gaussian with mean γ (i.e. the inverse of the true index) and variance $(\gamma^2 k)^{-1}$. Given the asymptotic normality of $\hat{\gamma}$, the 95% confidence interval is computed as:

$$\hat{\gamma} \pm 1.96 \frac{\hat{\gamma}}{\sqrt{k}}. \quad (2.30)$$

It is important to emphasize that each different threshold value might lead to a different Hill estimator. The two subscripts N and k indicate dependence of the estimated value on the number of data points and on the chosen threshold, respectively. In the appendix (A 2.2) a mathematical proof of the formula (2.29) and a detailed description of the procedure are given. The easy implementation of eq. (2.29) together with the desirable asymptotic properties of consistency and Normality are the advantages of the Hill estimator over many other procedures.

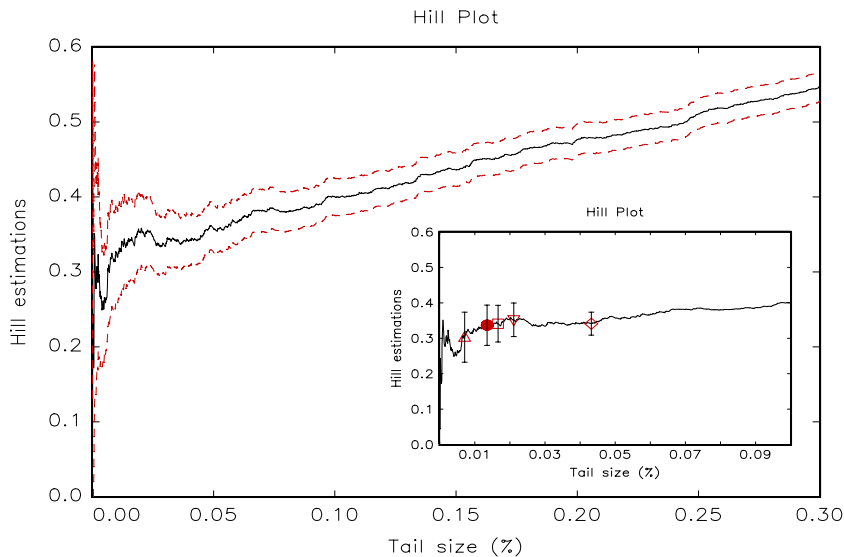


Figure 2.3: Hill plot for 10000 realizations from a random variable drawn from a Student-t distribution, with 3 degrees of freedom. The dashed lines represent the 95% confidence interval derived from eq. (2.30). The four points mark the estimates of γ chosen by the methods (1) through (4) of the next section: the square (1), the triangle (2), the inverted triangle (3) and the diamond (4). The estimated values are (0.007,0.30), (0.017,0.34), (0.021,0.35) and (0.043,0.34) respectively. The first number indicates the optimal tail size expressed as a fraction of the entire sample, and the second value refers to the corresponding Hill estimator. The full circle marks the estimate associated to the theoretical value of k_N^{opt} , from eq. (A 2.3) in appendix A 2.3.

Nevertheless, it is important to emphasize its potential sensitivity to the choice of the tail size. For better understanding of this problem, let us consider a simple example. In figure (2.3), a “Hill plot” is exhibited, i.e. the inverse tail index $\hat{\gamma}$ estimated using eq. (2.29) as a function of the tail size, for a Student-t random variable with 3 degrees of freedom. The theoretical value of the tail index for this distribution is exactly its degree of freedom [Feller, 1971]. Nevertheless, we can observe a monotonic decrease of the estimated value of γ , starting from $\hat{\gamma} = 0.55$ at 30% tail size, to $\hat{\gamma} = 0.32$ at the 1% tail size²². The underlying assumption of power law tails is the more accurate the further one goes to the outer part of the distribution. The large upward bias at, e.g., 30% tail sizes is, then, not too surprising, since the true power law behavior is strongly contaminated by the entries from more central parts of the distribution.

For this reason, it is not immediately obvious what the appropriate tail fraction

²²This phenomenon is not limited to simulated data, but it is also a frequent empirical fact observed, for example, in financial data, as shown by Lux [2000] among others.

should be that would give the “best” estimator for the ‘true’ parameter α . A possible practical approach could be an “eyeball” method, searching for a region in the Hill plot where the estimated values are approximately constant. However, there is no clear evidence of the existence of such a plateau, and, moreover, this approach has all the drawbacks of a subjective graphical data analysis.

A more rigorous and statistically meaningful approach is to use a quantitative criterion for an endogenous selection of the *optimal cut-off value*, based on the statistical properties of the Hill estimator.

2.3.4 Optimal tail size

We now turn to a data-driven criteria for an endogenous selection of the threshold value. A natural choice is the minimization of the mean squared error (MSE in the following) of the estimated $\hat{\gamma}(= \frac{1}{\alpha})$, since it implies a trade off between the bias, that increases with the tail size, and the variance, that decreases when extending the tail fraction. The optimal value is, then, defined as:

$$k_N^{opt} = \min_k E[(\hat{\gamma}_{k,N} - \gamma)^2] = \min_k \{Var[\hat{\gamma}_{k,N}] + Bias^2[\hat{\gamma}_{k,N}]\}. \quad (2.31)$$

It is relatively easy to evaluate the variance, given the asymptotic normality of the estimator (see [Hall, 1982], Goldie and Smith [1987]). In order to assess the contribution of the bias, it is necessary to introduce a second order expansion of the tail²³:

$$P(x|x > u) = 1 - ax^{-\alpha}[1 + bx^{-\beta} + o(x^{-\beta})], \quad (2.32)$$

called Hall condition (see [Beirlant et al., 1996], [Reiss and Thomas, 1997]). Note that the first and the second terms have the same functional form. This is a crucial assumption, since a slower decay term, such as a $\log x$, would prevent the asymptotic convergence towards the power law, while an exponential term would have a so rapid convergence that it would not affect the behavior of the tail. Anyhow, the expansion (2.32) applies to many text book distributions, such as the Student-t. For the Hall’s class of distributions, the first and the second moments of the Hill estimator (2.29) can be computed, which allows to derive the asymptotic MSE as a function of the underlying parameters of (2.32) (cf. Hall [1982]). Hall has shown that the optimal tail size is given by:

$$k_N^{opt} = \lambda(a, b; \alpha, \beta) N^{\frac{2\alpha}{2\alpha+\beta}}. \quad (2.33)$$

The major practical problem of using (2.33) is that it requires a preliminary estimate of α , as well as of the parameter β of the second-order term in the expansion (2.32).

Many methods for an endogenous selection of the optimal tail size are currently available in the literature. In general, we can classify them into two different categories, according to which a quantity is employed for the estimation of k_N^{opt} . To the first class belong those methods that derive estimators for MSE or asymptotic MSE, and then, apply a minimization algorithm to find the optimal value. The second

²³Other equivalent parameterizations exist in the literature, see for instance Beirlant et al. [1996], Reiss and Thomas [1997].

group relies on a direct estimation k_N^{opt} from eq. (2.33). In the following list, we give a brief description of the procedures, leaving the technical details, often quite involved, to the pertinent literature²⁴.

1. The first method introduced by Hall [1990] is based on a bootstrapping procedure, using sub-samples rather than the entire data set. The contribution of the bias is extracted via the minimization of the empirical MSE of the sub-sample, computed using an initial estimate of α from the overall sample. Hall assumes $\alpha = \beta$, which applies to some distributions, but, in general, turns out to be a quite restrictive assumption. Danielsson and de Vries [1997] generalize this approach including a moment estimator for β . Several parameters are left to the choice of the user: i) the starting value for $\gamma_n(k)$ ii) the sub-sample size $n_1 = fN$, expressed in terms of a fraction f of the entire sample of N data points, iii) the number of re-samples in the bootstrapping procedure.
2. A more general sub-sample bootstrap procedure is implemented in the method developed by Danielsson et al. [2001]. They show that the MSE of a suitable combination of first and of second moments of the Hill estimator leads to an equivalent and more convenient minimization problem than the MSE of the estimator itself. The auxiliary statistic is, indeed, a consistent estimator for α , and, moreover, no initial estimate is needed in order to compute the bootstrap MSE. The optimal tail is, then, a function of two estimates from sub-samples of different sizes, namely a given value n_1 and $n_2 = n_1^2/N$. The sub-sample size n_1 is an exogenous parameter, that can be expressed in terms of a fraction of the whole set of points.
3. Drees and Kaufmann [1998] construct a “stopping rule” for the Hill estimates to extract the bias. Their starting point is the observation that large deviations of the estimates from the expected order of magnitude, given by $\ln(\ln N)$, are attributable to the bias term. The optimal tail is, then, derived by using two different “stopping times”, after computing a consistent moment estimator for β from the pertinent Hill estimators. The stopping rule assumes the form $r_n = 2.5\gamma_N(k)N^\theta$, where $\gamma_N(k)$ is an initial estimate of γ , and θ is a number smaller than 0.5.
4. Beirlant *et al.* Beirlant et al. [1996] use as a starting point the relationship between the Hill estimator and Pareto quantile plots. The optimal value for the tail size is estimated via a weighted least squares regression derived from this relationship. Weights are, then, iteratively adjusted until convergence of both the weights and the tail fractions derived from them is obtained.
5. Finally, Beirlant *et al.* Beirlant et al. [1996] generalize the previous method by applying the same idea to the von Mises γ -parameterization. Their method, therefore, does not only allow for data-driven selection of the cut-off value k of data with a limiting Pareto distribution, but also nests the possibilities of exponential decline and finite endpoint.

²⁴For a concise description of the methods see Lux [1996a].

The estimates using the first four methods²⁵ are illustrated in the inset in Figure 2.3, together with the theoretical value k_N^{opt} derived from eq. (A 2.3). Despite the relative high variability of the estimated optimal tail sizes, the estimates of γ are all very close to the true value, since, on the basis of 95% confidence intervals, the value 0.33 can not be rejected.

2.3.5 Monte Carlo Simulations

In order to compare the different methods for an endogenous selection of the tail size, we have performed a Monte Carlo analysis using 100 samples of randomly generated variables from a Student-t distribution with 3 degrees of freedom, each composed by 10 000 independent realizations. The theoretical value of the tail index is 3. Moreover, for this distribution it is possible to derive the analytical expression of k_N^{opt} from eq. (2.33) for any given sample size N , since the expansion (2.32) holds exactly for this parametric family. Its value for a sample of 10 000 points is $k_N^{opt} = 135$ (details in the appendix). Table 2.3 summarizes the results of the numerical experiment.

The considered four procedures (numbered with 3, 4, 5 and 6 in Table 2.3) for the optimal tail size are characterized by several exogenous parameters which might eventually affect the estimation of the tail index and the optimal cut-off value -for instance the percent of order statistic over the total sample size in the method of Danielsson et al. [2001], or the stopping time parameter in the Drees and Kaufmann [1998] algorithm. In order to evaluate their robustness with respect to changes in the parameter sets, we perform our Monte Carlo simulations varying one or two key parameters (for the description see the previous paragraph). We observe that a progressive decrease of the number of considered extreme observations, due to a variation of the exogenous parameters, increases the standard deviation of the estimates, while contemporary reduces the bias, since the Pareto approximation becomes more accurate. However, a closer look to the entries of Table 2.3 reveals an impressive homogeneity of the RMSE (Root Mean Squared Error) within each method across different choices of parameters. Such regularity is not surprising, giving that the common goal of every procedure consists in minimizing the Mean Squared Error (the asymptotic or the empiric one). To be more precisely, we observe a slightly reduction of the RMSE for smaller mean values of the optimal tail sizes.

For a meaningful comparison of the methods, we consider the parameter sets that give the minimum RMSE, marked in Table 2.3 with a star (*). The mean values of the estimates exhibit a notable similarity; all the procedures, in fact, present mean values almost identical, but somewhat below the true value 3. The downward bias is related to the minimization of the MSE, the ultimately goal of all the methods, which produces, by its very definition, a bias for any finite sample size. The four procedures exhibit different degrees of variability of the estimates. Interesting, the behavior of the estimated indices mirrors in an equivalent behavior of the estimated optimal tail sizes. The medians²⁶, in fact, are very similar across the procedures

²⁵Conversely to the first four semi-parametric methods, the last approach does not rely on the Hill estimator.

²⁶The median is notoriously a more robust estimator than the mean against the presence of

	Mean	Med	Std	Max.	Min.	Bias	RMSE	Opt	Med	Std	Max.	Min.
1) Ordinary Least Squared												
$ r > 2$	2.54	2.54	0.09	2.77	2.32	0.46	0.47	1395	1395	33	1464	1313
$ r > 5$	2.80	2.82	0.31	3.56	1.93	0.20	0.37	155	154	14	201	123
$ r > 10$	2.79	2.69	0.74	4.92	1.06	0.21	0.77	21	22	5	34	10
2) Hill Estimator												
$ r > 2$	2.32	2.32	0.06	2.47	2.19	0.68	0.68	—	—	—	—	—
$ r > 5$	2.88	2.86	0.22	3.49	2.38	0.12	0.24	—	—	—	—	—
$ r > 10$	3.14	3.11	0.69	5.10	1.65	-0.14	0.70	—	—	—	—	—
$k_N^{opt} = 135$	2.90	2.89	0.25	3.55	2.27	0.10	0.26	—	—	—	—	—
3) Drees and Kaufmann [1998]												
$\theta = 0.25$	2.67	2.66	0.18	3.25	2.33	0.33	0.38	493	507	155	790	113
$\theta = 0.10$	2.77	2.73	0.27	3.58	2.19	0.23	0.35	304	295	201	865	21
$\theta = 0.05^*$	2.87	2.80	0.33	3.91	1.84	0.13	0.35	215	141	170	704	10
4) Danielsson et al. [2001]												
$f = 0.15$	2.83	2.73	0.47	5.66	2.13	0.17	0.49	309	254	292	1744	2
$f = 0.10$	2.86	2.80	0.53	6.32	1.95	0.14	0.55	277	188	286	1286	7
$f = 0.05^*$	2.88	2.81	0.38	4.32	2.15	0.12	0.41	281	181	294	1874	10
5) Danielsson and de Vries [1997]												
<i>(Starting value $\gamma_{0.05}$)</i>												
$f = 0.15$	2.81	2.84	0.20	3.26	2.43	0.19	0.28	260	248	86	522	129
$f = 0.10$	2.83	2.82	0.20	3.29	2.40	0.17	0.26	246	232	88	602	134
$f = 0.05^*$	2.87	2.82	0.22	3.39	2.45	0.13	0.26	195	181	71	496	84
<i>(Starting value $\gamma_{0.10}$)</i>												
$f = 0.15$	2.71	2.70	0.14	3.03	2.40	0.29	0.32	421	417	88	624	254
$f = 0.10$	2.74	2.73	0.14	3.15	2.44	0.26	0.30	355	345	75	682	201
$f = 0.05$	2.82	2.82	0.19	3.32	2.44	0.18	0.26	280	262	91	647	125
6) Beirlant et al. [1996]												
	2.92	2.91	0.65	7.35	2.34	0.08	0.65	421	100	250	1013	6

Table 2.3: Results of the Monte Carlo simulations. The rule of the parameters are described in section 2.3.4. The descriptive statistics for the tail size of the OLS and the Hill methods refers to the selected extreme observations, without any “optimal” criterium. The values for the Hill estimator are identical to those of the OLS procedure. The number of re-samples for the bootstrapping routines 4) and 5) is 50.

(except for method 6), while the standard deviations exhibit a much wider range. As a final point, the best performance is given by the method of Danielsson and de Vries [1997], which guarantees the lowest RMSE. On the contrary the fastest method in terms of computational time is the algorithm introduced by Drees and Kaufmann [1998]. Moreover, the method (5) gives average estimates very close to those computed using, as a tail size, its theoretical optimal value k_N^{opt} : the RMSE is identical, while the estimated mean for the optimal tail size is higher than k_N^{opt} .

Additionally, Table 2.3 provides the descriptive statistics for the OLS method applied to the same pool of time series, which might be of interest, since this method is very often applied in empirical work. It turns out that the OLS estimator produces always a higher RMSE than the more elaborated methods, on the other hand, given its simplicity and sufficient accuracy, the OLS methodology might be very useful for a first assessment of the index of the tail; however it should be accompanied with the more reliable methods.

In conclusion, the average values of the tail estimates and the average optimal cut-off values corresponding to the minimum RMSE among the different methods exhibit a notable homogeneity. The discrepancy across the four procedures emerges in the different degree of dispersion of the estimates and in the computational time.

2.4 Fat tails in financial data: The case of the Tokyo Stock Exchange

The EVT gives a rationale for a simple classification of the continuous distributions, based on the behavior of their tails. This theory, together with the empirical estimation methods derived from it, have been recently employed to analyze extreme fluctuations in financial data. Examples in this direction are the papers of Lux [2000], Loretan and Phillips [1994], Gilli and K ellezi [2003], among others. The conclusion of this large body of research is the almost universal consensus on Pareto tails of returns distributions, which means that empirical distributions belong to the Fr chet basin of attraction. These studies confirm, to some extent, the pioneering intuition by Mandelbrot [1963] of Pareto tails, but do not completely agree on the value of the exponent α . The question that arises is how fat are the empirical distributions, or, in other words, which is the value of the exponent α .

Mandelbrot [1963], by introducing the stable L vy, distribution implicitly set an upper bound to the tail index (α should be strictly lower than 2 to be compatible with the L vy stable assumption, cf. section 2.1). However, recent non-parametric analysis point to a tail index that varies within a narrow interval ranging between 2.5 and 5, which is outside the L vy stable regime. The surprising homogeneity of the estimates for different data (stocks or exchange rates) and frequencies (from ‘tick by tick’ to weekly data, cf. [Dacorogna et al., 2001]) has been considered by many authors as an imprint of an universal law, called by Gopikrishnan et al. [1998] *inverse cubic law for financial fluctuations*.

outliers. For this reason we also include in Table 2.3 the medians.

In this conclusive part of the chapter devoted to the distributional properties of financial returns, we consider again the problem of quantifying the tail index α . In order to do so, we performed an extended empirical investigation of the behavior of the tail of daily data from the Tokyo Stock Exchange (details of the data are presented in appendix), using the methods and techniques outlined in the previous section. We use a very large pool of stocks of about 1 200, comparing the different methods for an endogenous selection of the tail size (cf. Optimal tail size, section 2.3.3). Lux [2000] has compared the estimated values for α applying these techniques to the thirty largest German stock companies, a number that might be too small for a sharper conclusion concerning the universality of the tail index²⁷. The large sample employed here allows for a more comprehensive comparison of the performance of the different methods, and, moreover, for a more robust inference on the existence of this universal inverse cubic law. Plerou et al. [1999] have analyzed an amount of stocks of the same order of magnitude from the New York Stock Exchange, using the OLS estimation method and the Hill estimator with a fixed tail size. Their conclusion of an index $\alpha \approx 3$ might be influenced by a potential bias due to the rigidity in the choice of the tail size.

Figure 2.4(a) shows the complement of the cumulative distribution for normalized absolute returns in a log-log scale of a time series of a randomly chosen stock out of the entire pool²⁸. The outer part of the distribution is well approximated by a straight line, as shown in the inset of Figure 2.4(a), which signals a power law decay of the tail. This simple graphical representation can be used as an heuristic estimation method of the index α . We have already discussed the limitations of the OLS approach, however, given its simple implementation and its large popularity, we use this procedure as a starting point of our empirical investigation. The Monte Carlo analysis of the previous section shows that such procedure provides reasonably accurate estimates. Next, we employed the celebrated Hill estimator with an endogenous selection of the tail size. For the OLS method, we *arbitrarily* restrict the power-law approximation to normalized returns that are higher than 2 in absolute value, similarly to the analysis conducted by Plerou et al. [1999], which corresponds to 5% quantile of the Gaussian distribution. To emphasize the arbitrary nature of the results as a function of the chosen tail size, we perform the same analysis with an higher threshold value, namely $|r| > 4$. We set, then, the exogenous threshold u for the Hill estimator using two different criteria. In the first one, typically adopted in many studies, the tail is a given fraction, f , of the entire sample of absolute returns (indicated with N); therefore the cut-off value $u = r_{f*N}$ is a random variable depending on the sample realizations. In the second alternative, u is given and, therefore, the number of considered data points, that constitute the tail, is random.

²⁷However, the main idea of that paper was to investigate whether the data-driven methods for the optimal tail size confirm the stable Paretian hypothesis of Mandelbrot.

²⁸In order to enhance the statistics, we merged together the negative and positive returns, computed, then, the absolute values. The considered empirical distributions are almost symmetric, since the means are close to zero, when compared to the standard deviations. The skewness is smaller than one, except for few cases.

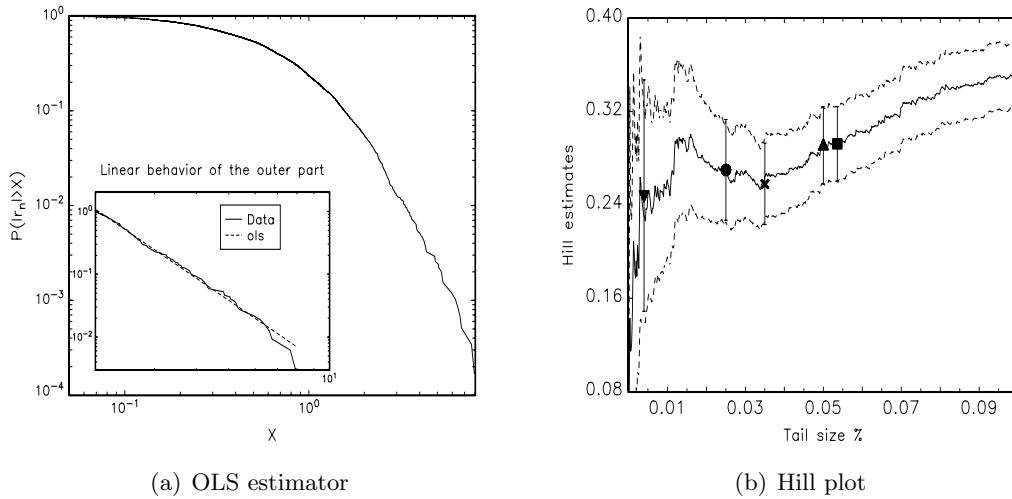


Figure 2.4: (a) Complement cumulative distribution for the normalized absolute returns of the pooled data of the randomly chosen time series. The inset shows the approximately linear behavior of the tail, that signals its power law decay. The fit is computed considering returns higher than two standard deviations ($|r_n| > 2$); the slope is $\alpha = 3.63 \pm 0.02$ with a $R^2 = 0.994$. (b) Hill estimator as a function of the tail size. The two points mark Hill estimates for two different criteria for the selection of the tail size: the triangle refers to the 5% tail size ($\gamma_{H_1} = 0.292 \pm 0.032$); the square labels the Hill estimate computed with normalized returns higher than 2 ($\gamma_{H_2} = 0.291 \pm 0.033$). The two estimates are not significantly different. The other three points (inverted triangle, circle and star) refer to the methods (4), (5) and (6) from Table 2.4, applied to this particular time series. The values are (0.248 ± 0.108) , (0.270 ± 0.043) and (0.258 ± 0.035) , respectively. The dashed lines represent the 95% confidence interval, computed using eq. (2.30).

Table 2.4 confirms the results observed in the literature: the estimates are peaked around a value slightly higher than three with a relative low dispersion, for both the Hill estimator, with a fixed threshold $|r| > 2$ or for 5% tail size, and the OLS procedure with $|r| > 2$. However, the alternative cut-off values, namely $|r| > 4$ or 1% tail size, for the first three methods give rise to different population means and standard deviations, both increasing for higher threshold values. The conclusion of Plerou et al. [1999] on inverse cubic law for financial returns, seems, then, strongly dependent on the chosen threshold value.

	Mean	Med	Std	Max.	Min.	Opt	Med	Std	Max.	Min.	# Fl.	# $\alpha < 2$
1) Ordinary Least Squares												
$ r > 2$	3.31	3.39	0.52	4.55	0.45	253	271	64	357	6	2	41
$ r > 4$	3.64	3.61	1.33	10.37	0.55	26	25	7	55	6	14	115
2) Hill fixed threshold												
$ r > 2$	3.12	3.19	0.36	4.11	0.46	—	—	—	—	—	—	12
$ r > 4$	4.09	4.01	1.19	11.49	0.88	—	—	—	—	—	—	43
3) Hill variable threshold												
5%	3.19	3.22	0.33	4.22	1.66	264	280	52	333	23	0	6
1%	3.84	3.79	0.74	7.18	1.39	52	56	10	66	7	1	18
4) Drees and Kaufmann [1998]												
	3.55	3.39	0.87	10.51	1.33	120	75	104	515	6	43	9
5) Danielsson et al. [2001]												
	3.84	3.59	1.26	16.78	0.98	136	59	227	2618	6	95	12
6) Danielsson and de Vries [1997]												
	3.50	3.50	0.47	5.51	1.70	139	134	46	631	18	0	6

Table 2.4: Results of the application of the data-driven selection methods for the tail size to the Japanese Stock Exchange, together with the Hill and OLS estimator. Note that we set an arbitrary minimum value to the tail size equal to 5. The procedure is not converging in the other case (the number of failures is labeled in the table with # Fl.). All the entries in the Table refer to observations for which the tail size is higher than this minimum value. For the omitted entries see the caption of Table 2.3.

It might be a sensible idea to compare the previous results, obtained with the Hill and the OLS methods, to the estimates computed from the methods for an endogenous selection of the tail size. We employ the first three procedure detailed in the previous section, excluding the algorithm of Beirlant et al. [1996], which is extremely slow and time consuming compare with the bootstrapping routines or the method of Drees and Kaufmann [1998]. Note that we set an arbitrary minimum value for the tail size (5 points), under which the routine is considered not to converge. The associated confidence interval would be, in fact, so large to prevent any sensible conclusion. The average estimates -computed among the cases for which we have convergence of the algorithms- of both the tail index and the optimal cut-off are very similar among the three methods. On the contrary, the dispersion around the mean shows significant disparities across the different procedures. Interesting, the three procedures, when applied to real data, give equivalent results as in the previous Monte Carlo numerical experiment in terms of the standard deviation of the estimates²⁹: The routine proposed by Danielsson and de Vries [1997], in fact, outperforms the others in terms of both a smaller standard deviation and a lower number of non-converging cases (see Table 2.4), while the procedure of Danielsson et al. [2001] turns out to be the less efficient.

One should be aware, however, that the assumption of independent observations does not hold for empirical data, for the well known phenomenon of volatility clustering (see next chapter). Kearns and Pagan [1997] have shown that the Hill estimator, when applied to IGARCH increments, possesses a standard deviation seven times higher than predicted by eq. (2.30). Moreover Rootzen et al. [1998] have shown that for dependent data, the Hill estimator remains only consistent but not asymptotically efficient. The dependency among data can, then, partially explain the higher values of the standard deviations of the estimates when compared to the standard errors from the simulated data.

The last column of Table 2.4 shows the number of estimates for which α is smaller than 2, therefore compatible with the stable Paretian hypothesis³⁰. The hypothesis of stable Paretian distribution is strongly rejected by the data, since we observe a number of cases that is approximately 1% of the entire pool of observation, which eventually limits the validity of such an assumption to a negligible fraction of the total pool of stocks.

In conclusion, our analysis confirms the existence of a power law behavior for financial fluctuations for the Japanese Stock Exchange, in line with other world financial markets, with an average index slightly higher than three³¹. What is really remark-

²⁹We refer to the relative order of magnitude of the standard errors across methods, rather than their absolute values, which are obviously different.

³⁰These events are selected among the estimates for which we observe a convergence of the algorithm.

³¹Note, however, that the empirical distributions of the tail indices show a non-negligible degree of skewness. Therefore a more robust estimator of their "typical values" might be the median, which is closer to three than the means

able is the regularity of that index across different markets, confirmed once again by our empirical investigation. On the other hand seems quite artificial to remark the existence of an “exact” inverse cubic law for financial fluctuations.

2.5 Conclusions

We have observed that the leptokurtic behavior of financial returns is a very pervasive empirical regularity. The characterization of this property in terms of the index of the tail provides a simple and effective method to quantify this ubiquitous finding, according to the theoretical foundation of Extreme Value Theory. The regularity of the tail estimates across different frequencies, markets and data constitutes a remarkable empirical property of financial data. Our analysis for the Japanese Stock Exchange confirms such an impressive homogeneity, despite large heterogeneity of the considered stocks (in terms of volumes and number of transactions, market capitalization or underlying economic activities) and economic phases that the Japanese market experienced (from the bubble of the ‘80, to the crash and the subsequent stagnation). Our novel contribution is the validation of the power law behavior of the tail using methods for an endogenous selection of the its size. The origin of such extraordinary regularity is one of the key issues of the present thesis.

Appendix to Chapter 2

A 2.1 Description of the data

In the following list the details of the time series used in this chapter and, later on, in the thesis are given:

- Gold: daily price of Gold from 01/78 to 12/98;
- DAX: index of the German Stock Exchange from 10/59 to 12/98;
- DB: daily price of the Deutsche Bank in the period 1974/2001
- Siemens is the time series of daily price of the correspondent German company quoted in Frankfurt in the period 1974/2001;
- daily USD/DM exchange rate US dollar against Deutsche Mark from 01/74 to 12/98;
- Dax/Hf: minute-to-minute value of the DAX index from 11/1988 to 12/1995;
- The data set consists of the daily prices of all the stocks traded in the Tokyo Stock Exchange in the period 01/1975 to 12/2001.

A 2.2 Derivation of the Hill Estimator

In this appendix we derive the maximum likelihood estimator of the index α of a Pareto law, developed by Hill [1975].

Assume we are interested in the tail behavior of a sequence of T IID variables x_i , which we assume are distributed approximatively according to a Pareto distribution³² with index α :

$$F(x) = 1 - \left(\frac{u}{x}\right)^\alpha \quad x \geq u, \quad (\text{A 2.1})$$

where u is a scaling parameter that regulates the amplitude of the tail. The corresponding density is given by:

$$f(x) = \alpha \frac{u^\alpha}{x^{\alpha+1}}.$$

The maximum of the log-likelihood function is the solution of the maximization problem:

$$L_T(\alpha) = \max_{\alpha} \sum_{i=1}^T \ln f(x_i) = \max_{\alpha} \sum_{i=1}^T [\ln \alpha + \alpha \ln u - (\alpha + 1) \ln x_i],$$

which yields

$$\frac{1}{\alpha} = \frac{1}{T} \sum_{i=1}^T [\ln x_i - \ln u].$$

³²The proof is based on the approximation given by eq. (A 2.1). However, the Hill's formula holds also for a more general class of distribution functions with regularly varying tails (see de Vries [1994]).

The implementation of this estimation procedure for a time series is straightforward. Let us assume have N observations of an IID random variable. The first step is the selection of a suitable subset of k sample elements that belong to the tail of the distribution. Then, taking descending order statistics, i.e. we arrange the data in the order $x_{(N)} > x_{(N-1)} \dots > x_{(N-k)}$, where now $u = \min(x_{(i)}) = x_{(N-k)}$. If we denote by γ the inverse of α , we end up with eq. (2.29) in the main text:

$$\hat{\gamma}_{k,N} = H_{k,N} = \frac{1}{k} \sum_{i=1}^k [\ln x_{(N-i+1)} - \ln x_{(N-k)}].$$

It is obvious that this procedure is optimal for data strictly following a Pareto distribution. Asymptotic consistency of the estimator for fat-tailed distributions has been demonstrated by Mason [1982], in the case of $k \rightarrow \infty$ and $k/N \rightarrow 0$; moreover, for IID sequences, asymptotic Normality has been demonstrated by Hall [1982] and Goldie and Smith [1987].

A 2.3 Optimal tail size for Student-t distribution

The probability density function of a Student-t random distributed variable with 3 degrees of freedom is:

$$p(x) = \frac{2}{\pi} \frac{a^3}{(a^2 + x^2)^2},$$

as in eq. (2.18). The associated distribution function is:

$$P(x) = \int_{-\infty}^x p(y) dy = \frac{1}{2} + \frac{1}{\pi} \frac{ax}{a^2 + x^2} + \frac{1}{\pi} \arctan\left(\frac{x}{a}\right). \quad (\text{A 2.2})$$

The asymptotic behavior of the tail of the distribution function, up to the second term, is given by the following expansion:

$$P_{\infty}(x) = 1 - \frac{2}{3\pi} \left(\frac{a}{x}\right)^3 + \frac{4}{5\pi} \left(\frac{a}{x}\right)^5 + O\left(\frac{1}{x^6}\right). \quad (\text{A 2.3})$$

To see this, we explicitly write the expansion at infinity of the two functions from eq. (A 2.2):

$$\begin{aligned} \frac{1}{\pi} \frac{ax}{a^2 + x^2} &= \frac{1}{\pi} \frac{a}{x} - \frac{1}{\pi} \left(\frac{a}{x}\right)^3 + \frac{1}{\pi} \left(\frac{a}{x}\right)^5 + O\left(\frac{a}{x}\right)^6, \\ \frac{1}{\pi} \arctan\left(\frac{x}{a}\right) &= \frac{1}{2} - \frac{1}{\pi} \frac{a}{x} + \frac{1}{3\pi} \left(\frac{a}{x}\right)^3 - \frac{1}{5\pi} \left(\frac{a}{x}\right)^5 + O\left(\frac{a}{x}\right)^6. \end{aligned}$$

Then it is straightforward to compute (A 2.3).

It has been shown that the value k_N^{opt} of tail size that minimize the asymptotic mean squared error for the Hill estimator (see Matthys and Beirlant [2000]) is given by:

$$k_N^{opt} = \left(\frac{C^{2\rho}(\rho+1)^2}{2D^2\rho^3} \right)^{\frac{1}{2\rho+1}} N^{\frac{2\rho}{2\rho+1}}, \quad (\text{A 2.4})$$

if the underlying distribution function belongs to the so-called Hall class, i.e. if holds:

$$1 - P(x) = Cx^{-\frac{1}{\beta}}[1 + Dx^{-\frac{\rho}{\beta}} + o(x^{-\frac{\rho}{\beta}})].$$

As can be see from eq. (A 2.3), the Student-t distribution belongs to the Hall class. We can, then, easily compute the optimal value k_N^{opt} , given the sample size N , using the following relations:

$$C = \frac{2}{3\pi}a^3, \quad D = -\frac{6}{5}a^5, \quad \beta = \frac{1}{3}, \quad \rho = \frac{2}{3}.$$

For $N = 10000$, the optimal value turns out to be $k_N \approx 135$.

CHAPTER 3

Serial Correlation Properties of Financial Time Series

In this chapter we consider the time structure of financial returns, focusing on some methods to detect and to quantify time correlations among time series. The random walk hypothesis, in its geometric or arithmetic version, implies independent prices increments. Following Pagan [1996], we address the question: “*Are financial series independently distributed over time?*”

In order to answer this question, we have to point out the crucial role of the time-ordering in the statistical analysis of time series. Starting with a series of returns $\{r_t\}_{t=1}^T$, the distributional properties of the variable r_t , such as the cumulative distribution analyzed in the previous chapter, are well-defined quantities invariant under permutations of the sequential order of the series itself. Conversely, when we introduce the concept of “time series”, implicitly, we assume that the sequential order matters, and it might bring useful information for the characterization of the underlying stochastic process. For instance, the existence of correlations among data implies that the past realizations of the stochastic process has some kind of influence on the future realizations. The first issue is, then, to define proper quantities to detect and to measure such correlations.

Parametric, semi-parametric or non-parametric methods are available in order to describe the intertemporal dependence among data. In the following we exclusively focus attention on non-parametric and semi-parametric methods, leaving out the vast literature of the parametric models to the pertinent literature. Like in the previous chapter related to the unconditional distribution, our choice is mainly dictated by the necessity to characterize properties of financial returns by making very general assumptions. We will, then, apply these methods to the same large set of financial time series of the Tokyo Stock Exchange, detailed in chapter 2.

3.1 Measuring Correlations

Correlation function

Given a stationary¹ time series $\{r_t\}_{t=-\infty}^{\infty}$, a quantity usually employed to measure correlations is the temporal two-point correlation function or auto-covariance, defined as:

$$Cov(k) = E[(r_{t-k} - \mu)(r_t - \mu)] , \quad (3.1)$$

where $E[\cdot]$ is the expectation operator and $\mu = E[r_t]$. The above definition is not sensitive to the expectation value of the process. The auto-correlation function (ACF hereafter) is conveniently normalized to avoid the influence of the scale of the fluctuations present in eq. (3.1), and defined as:

$$C_r(k) = corr(r_t, r_{t-k}) = \frac{Cov(k)}{Cov(0)} . \quad (3.2)$$

Given the choice of the normalization, the ACF is a quantity that ranges in the interval $[-1, +1]$. It is extremely important to highlight that the ACF detects whether **linear correlations** are present in the time series between the past observations and the actual realization. In other words, if a time series does not exhibit any significant autocorrelation, it does not imply absence of correlations among the data, but just lack of linear dependence between the adjacent realizations.

The sample counterpart of the ACF is the *correlogram*, given by:

$$\hat{\rho}(k) = \frac{\sum_{t=k+1}^T [(r_t - \hat{\mu})(r_{t-k} - \hat{\mu})]}{\sum_{t=1}^T [(r_t - \hat{\mu})^2]} , \quad (3.3)$$

where the expectation operator $E[\cdot]$ is replaced by its empirical counterpart, and $\hat{\mu}$ is the population mean. Eq. (3.3) introduces a bias in the evaluation of eq. (3.2) –see Bartlett [1946]– which is negligible for a large number of observations and relatively small number of considered lags.

Which information can we extract from the knowledge of the ACF? An interesting quantity is the *characteristic time-scale* τ_c , that, to some extent, gives an idea on the order of magnitude of the memory of the generating stochastic process. If we assume that the autocorrelation function is monotonically decreasing toward zero, the characteristic time-scale can be defined as the time that the ACF needs to reach a given fraction, usually identified with 0.5 or $1/e$, of the value at the first lag:

$$\frac{\hat{\rho}(\tau_c)}{\hat{\rho}(1)} \approx 0.5 \quad \text{or} \quad \frac{\hat{\rho}(\tau_c)}{\hat{\rho}(1)} \approx \frac{1}{e} . \quad (3.4)$$

Although operatively always computable in the case of monotonicity of the ACF, the characteristic time does not provide a direct information on the decay rate of the correlation. Using the characteristic time-scale, in fact, we cannot distinguish

¹A wide-sense stationary process is defined by the following conditions: i) the existence of the mean $E[r_t] < \infty$; ii) $R(t_1, t_2) = E[r_{t_1}r_{t_2}]$ is only a function of the difference $\tau = t_2 - t_1$, and is finite for every τ .

between an exponential or a power law decay of the correlation function. The decay rate is a more informative property of the stochastic process —see the connection between the behavior of the variance and the correlation decay described below. In order to obtain a more informative characterization of the time scale of the stochastic process, we introduce the following quantity, called *typical time-scale*, $\bar{\tau}$, of the process:

$$\bar{\tau} = \int_0^{\infty} C(\tau) d\tau . \quad (3.5)$$

The integral (3.5) might converge or not depending on the decay rate of the ACF. In the case of an **exponential decay** of the ACF, which is very common in many examples of natural processes, the integral (3.5) converges and we have the following identity:

$$\bar{\tau} = \int_0^{\infty} \exp \left[-\frac{\tau}{\bar{\tau}} \right] d\tau . \quad (3.6)$$

Eq. (3.6) tells us that the decay rate of the exponential function coincides with the typical time-scale of the stochastic process, in the case of an exponentially decay of the ACF. In general, many stochastic processes are characterized by a finite integral (3.5), and they are defined as *short-range* processes. This property gives the possibility for a clear separation of the time-development of the stochastic process in two regimes: a regime dominated by correlations and a regime characterized by pairwise independence between the realizations.

An ACF with an asymptotic **power law decay**, $C(\tau) \sim \tau^{\alpha-1}$ is characterized by a divergent integral (3.5), when the exponent is $0 < \alpha < 1$. In this case it is not possible to define any typical time-scale, and the process cannot be separated in two different regimes, namely correlated and pairwise independent observations, since all the time-scales are “mixed” together. Stochastic processes that exhibit this behavior are called long-memory or long-range correlated processes. Given the importance of this typology of processes for the financial econometrics, we provide a more precise mathematical definition of the long memory processes in terms of the ‘degree of persistence’ of their memory.

Definition 1: Let X_t be a stationary stochastic process, with finite variance σ^2 and, without loss of generality, mean zero. The long memory process is characterized by an autocorrelation $\rho(k)$ with an asymptotic hyperbolic decay. Formally:

$$\lim_{k \rightarrow \infty} \frac{\rho(k)}{c_{\rho} k^{-\alpha}} = 1 , \quad (3.7)$$

where $c_{\rho} > 0$ and $\alpha \in (0, 1)$. Consequently, the correlation coefficients are no longer summable, preventing the definition of a typical time $\bar{\tau}$ over which the data can be considered pairwise independent, as in the short memory processes.

The previous definition refers exclusively to asymptotic properties of the correlations. Def.(1), in fact, specifies only the ultimate rate of decay of the autocorrelation, without fixing the absolute size of any finite lag: each individual value can be arbitrarily small. For a better illustration of this aspect, let us consider a stationary

stochastic process² with N realizations $\{X_t\}_{t=1}^N$. The variance of the sum S_N of N variables X_t can be written as:

$$\text{Var}(S_N) = \sum_{i,j=1}^N C_{i,j} = N\sigma^2 + 2N \sum_{l=1}^N \left(1 - \frac{l}{N}\right) C(l), \quad (3.8)$$

where $C(0) = \sigma^2$ and $l = |i - j|$. If $C(l)$ decays faster than $1/l$, the sum converges to a finite value as N goes to infinity. Therefore the variance of the sum grows linearly for values of N large enough, as predicted by the CLT. On the contrary, if $C(l) \sim l^{-\alpha}$ has a power law decay for large l , with an exponent $0 < \alpha < 1$, then the sum does not converge to any finite value, even if the coefficients $C(l)$ are arbitrarily small. The joint effect of the long-range dependence is to enhance the size of the fluctuations, since the variance of the sum grows faster than linearly.

Spectral density

In addition to the characterization of the memory in terms of the autocorrelation, the long memory property can be equivalently defined in the frequency domain. The spectral density $S(\lambda)$ of a stationary stochastic process X_t is the Fourier transformation of the autocorrelation function:

$$S(\lambda) = \frac{\sigma^2}{2\pi} \sum_{k=-\infty}^{+\infty} \rho(k) e^{-ik\lambda}, \quad (3.9)$$

where λ s are the Fourier frequencies. Definition 1 can be rewritten in the frequency domain by the following mathematically equivalent definition (see theorem 2.1 in Beran [1994]):

Definition 2: Let X_t be a stationary process, for which exists a real number $\beta \in (0, 1)$ and a constant $c_f > 0$ such that:

$$\lim_{\lambda \rightarrow 0^+} \frac{S(\lambda)}{c_f |\lambda|^{-\beta}} = 1. \quad (3.10)$$

Then the process is called a stationary process with long memory. The spectral density of this type of stochastic process has, then, a pole at the origin.

3.2 Serial Correlation in Raw Returns

We can now apply the previous concepts to quantify time correlations in financial data. According to the Gaussian random walk benchmark (see chapter 2), the ACF of returns has to be identically zero for every time-lag, and the estimated quantities $\hat{\rho}(t)$ have to be asymptotically Normally distributed and independent random variables. Table 3.1 shows estimated values of ACF at different time lags;

²Again, we assume finite variance σ^2 and mean zero.

the associated confidence intervals are computed assuming Normally distributed estimates, with variance given by³:

$$\text{Var}(\hat{\rho}(t)) = \frac{1}{N} . \quad (3.11)$$

Figure 3.1 shows three examples of correlograms for raw returns. If we focus on daily data, the visual impression, confirmed by the entries in Table 3.1, is for a negligible correlation among returns. On the contrary, the minute-to-minute records of DAX exhibits a significant level of correlation, rapidly decaying with a characteristic time-scale of a few minutes. A positive fast decaying correlation function is observed in all high-frequency financial series [Mantegna and Stanley, 2000], with a decay-time ranging from few minutes for very liquid markets, to somewhat longer time for less traded stocks. The economic intuition behind those findings is the absence of continuing arbitrage opportunity in efficient markets. The statistically significant correlation in high-frequency data is anyway hardly exploitable if one takes into account transaction costs.

For a more quantitative validation of the null hypothesis of uncorrelated returns than simple graphical inspection of the autocorrelation, we use the Ljung-Box test Ljung and Box [1978]. Table 3.3 shows the results of this test when applied to the time series considered in Table 3.1. We can observe a high rejection probability of the null hypothesis, except for the exchange rate USD/DM. The very large values for minute-to-minute DAX data is related to the exponential decay of the autocorrelation. Even though the test rejects the hypothesis of non-correlated returns for the original time series, numerical experiments show higher p-values -often above the 5%- for smaller sample sizes (i.e. for 2000 data points). This dependence of the p-values on the size of the time series casts some doubts on the asymptotic Normality of the correlogram coefficients $\hat{\rho}(t)$. As recently shown by Davis and Mikosch [2000], the sample autocorrelation function (SACF hereafter) of heavy-tailed non-linear time series can have non-standard statistical properties. They have derived several theoretical properties of the SACF for a number of econometric models (e.g. GARCH and stochastic volatility models), which share the power-law decay in the tail of their unconditional distributions. In particular, they show that for stochastic processes with infinite fourth moment -which might be the case for financial returns, as detailed in chapter 2- the SACF remains a consistent estimator of the ACF, however with a rate of convergence slower than \sqrt{N} , expected under Normality. Consequentially, asymptotic confidence intervals of the SACFs are much wider than the ones predicted by Bartlett's formula. One should, therefore, be careful when drawing quantitative conclusions from many econometric testing procedures for detecting correlations, such as the Ljung-Box test [Ljung and Box, 1978].

Another interesting approach is to test the condition of independence by focusing on the estimated marginal and joint densities. If we have two independent stochastic

³Eq. (3.11) is, at least for a small number of lags, a good approximation of the more complicated Bartlett's formula [Bartlett, 1946].

	1	2	3	8	12	16
Gold	-0.050 (-0.078, -0.022)	-0.008 (-0.036, 0.020)	0.017 (-0.011, 0.045)	0.012 (-0.016, 0.040)	0.033 (0.005, 0.061)	0.014 (-0.013, 0.042)
Dax	0.070 (0.050, 0.090)	-0.056 (-0.076, -0.036)	-0.017 (-0.037, 0.003)	0.008 (-0.012, 0.028)	-0.018 (-0.038, 0.002)	0.009 (-0.011, 0.029)
DB	0.077 (0.053, 0.100)	-0.054 (-0.077, -0.030)	-0.026 (-0.050, -0.002)	0.020 (-0.003, 0.044)	-0.015 (-0.039, 0.009)	0.000 (-0.024, 0.024)
Siemens	0.060 (0.036, 0.084)	-0.029 (-0.053, -0.005)	-0.005 (-0.029, 0.019)	0.040 (0.016, 0.064)	0.006 (-0.018, 0.030)	0.034 (0.010, 0.058)
USD/DEM	-0.034 (-0.060, -0.009)	0.006 (-0.019, 0.032)	0.019 (-0.007, 0.044)	0.015 (-0.010, 0.041)	0.003 (-0.023, 0.028)	-0.014 (-0.040, 0.011)
Dax-Hf	0.430 (0.426, 0.433)	0.348 (0.344, 0.352)	0.265 (0.261, 0.269)	0.036 (0.032, 0.040)	0.002 (-0.002, 0.006)	-0.005 (-0.009, -0.001)

Table 3.1: Estimated values of the correlogram coefficients for different data sets. The numbers in brackets refer to the 95% confidence interval computed using the formula $2/\sqrt{N}$, where N is the sample size.

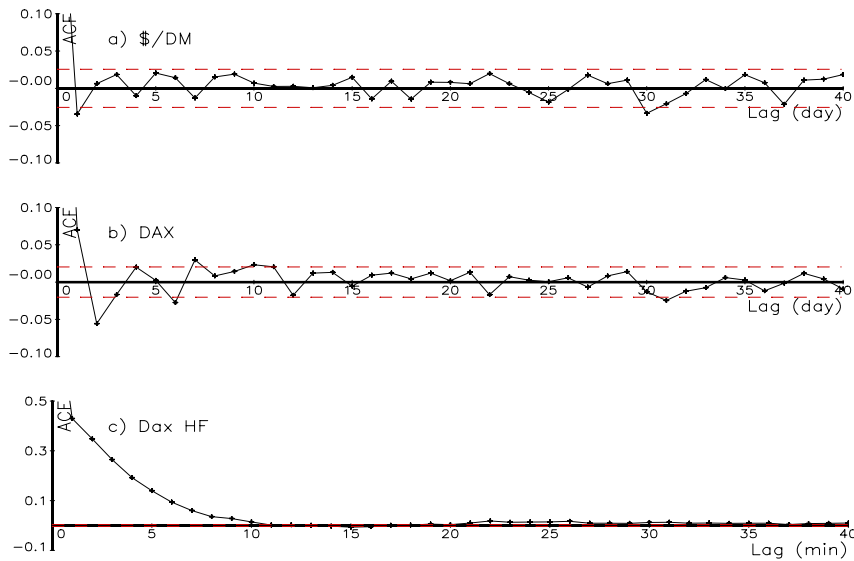


Figure 3.1: Panels (a), (b) and (c) show the correlograms of raw returns for three different time series: DAX, Siemens and DAXHf (see appendix A 2.1). The dashed lines in the first two panels refer to the 95% confidence intervals computed using the formula $2/\sqrt{N}$. Given the large number of observations, the confidence intervals are not visible in the case of DAXHf.

variables X and Y , their independence implies:

$$E[f_x(X)f_y(Y)] = E[f_{xy}(X, Y)] . \quad (3.12)$$

where with $f_{x,y}$ is the marginal density and f_{xy} is the joint density. The BDS test proposed by Brock et al. [1987] is based on the relation 3.12.⁴ Its null hypothesis is the iid-ness of returns, while the alternative hypothesis relies on general type of dependence. For this reason the BDS test is considered a test for non-linearity of the data.

Table 3.3 shows the results of the BDS tests for different time series, which are in line with other contributions from the pertinent literature (for a review see Barnett and Serletis [2000]). The pertinent test statistic is:

$$V_{\varepsilon,m} = \frac{\sqrt{n}(C_{\varepsilon,m} - C_{\varepsilon,1}^m)}{\sigma_{\varepsilon,m}} , \quad (3.13)$$

where $m = 2, 3, \dots$ is the correlation dimension, ε is a positive constant and $C_{\varepsilon,m}$ is the so-called correlation function —for more details see Barnett and Serletis [2000]. $V_{\varepsilon,m}$ has a limiting standard Normal distribution under the null hypothesis of IID

⁴The proof of the connection with the original BDS test and the formulation in terms of the relation 3.12 is given in Pagan [1996] pp. 27-29.

	4	8	12
Gold	14.30 (0.01)	19.49 (0.01)	49.12 (0.00)
Dax	83.12 (0.00)	103.88 (0.00)	111.93 (0.00)
DB	0.077 (0.00)	-0.054 (0.00)	-0.026 (0.00)
Siemens	33.23 (0.00)	50.43 (0.00)	56.33 (0.00)
USD/DEM	8.46 (0.08)	13.41 (0.10)	17.76 (0.12)
Dax-Hf	113 639 (0.00)	123 639 (0.00)	123 078 (0.00)

Table 3.2: Results of the Ljung-Box test applied to different time series at different lags. The absence of correlation is rejected at high significant level, for all time series, except for the exchange rate USD/DM. Note the larger values of the statistics for the high-frequency data than for the other daily series, which is related to the approximatively exponential decline observed in Figure 3.1.

	2	3	4	5	Result
Gold	22.6	27.2	31.0	35.1	Reject
Dax	22.1	29.4	35.6	40.6	Reject
DB	22.5	27.8	32.0	36.3	Reject
Siemens	24.1	31.6	37.6	43.5	Reject
USD/DEM	12.0	17.9	22.4	27.5	Reject
Dax-Hf	201	228	244	259	Reject

Table 3.3: Results of the BDS test. For the chosen time series the null hypothesis of *iid*-ness of returns is strongly rejected for all the considered embedding dimensions. In all these tests, we set ε equal to one times the standard deviation of the correspondent sample. The entries of the Table are given by the eq. 3.13.

data. We observe a strong rejection of the null hypothesis of independent and identically distributed returns. However, this rejection is consistent with several types of dependence in the data, which could result from a linear stochastic system, a non-linear stochastic system or a non-linear deterministic system.

All in all, if we focus on daily financial data⁵, the martingale hypothesis seems a quite reasonable assumption. It essentially describes the very high degree of (linear) unpredictability of financial returns. However, from a closer inspection of the time series with more refined statistical tests, one can conclude that the random walk model and its independent increments implication are just a first approximation for a meaningful description of the return dynamics⁶. These departure from iid-ness of the data, which are the rule rather than the exception, seems to be caused by temporal correlations in the higher moments (see next section) and are only very ‘dilute’ ones in the first moment.

3.3 The notion of volatility

The random walk benchmark implies not just uncorrelated returns, but *independent* increments under any transformation. Independence is a broader notion, “referring to the ability to express a joint density with marginals”, quoted by Pagan [1996]. Independence in the first moment of the joint two-points distribution $p_\tau(r_{t+\tau}, r_t)$ can be formalized in the formula:

$$E[r_t r_{t-k}] = E[r_t] E[r_{t-k}]. \quad (3.14)$$

Independence under general transformations implies that this is true for all the moments of $p_\tau(r_{t+\tau}, r_t)$. More precisely, if $h(\cdot)$ and $g(\cdot)$ are two measurable functions, the broadest definition requires the following condition:

$$E[g(r_t)h(r_{t-k})] = E[g(r_t)] E[h(r_{t-k})]. \quad (3.15)$$

Although it is practically impossible to test the relation (3.15) for the entire class of measurable functions, a feasible non-parametric approach is to replace the trial functions $g(\cdot)$ and $h(\cdot)$ with their polynomial expansions, and test the null hypothesis that all the pairwise terms $cov(r_t^m r_{t-k}^l)$ are zero. However this method, described for instance in the paper of Cameron and Trivedi [1993], has been rarely used in the literature. More attention has been devoted to some specific low values of m and l ; in particular the combinations $l = m = 1, 2$ for absolute returns⁷ have been extensively investigated in the last decade [Pagan, 1996, Ding et al., 1993]. Figure 3.2 shows a

⁵We have seen that high-frequency data exhibit very clearly the presence of time correlation (see Figure 3.1). However, these correlations can be easily traced back to “small” imperfections of the markets, such as the presence of transaction costs.

⁶It should be stressed that the random walk and the martingale hypothesis are not equivalent concepts. The first implies independent and identically distributed increments, while the latter only poses some constraints on the first moment of the conditional distribution.

⁷Another measure of the non-linear dependence in returns is the correlation of returns with subsequent squared or absolute values:

$$L(\tau) = corr(|r_{t+\tau}|, r_t) \quad \text{or} \quad L(\tau) = corr(r_{t+\tau}^2, r_t), \quad (3.16)$$

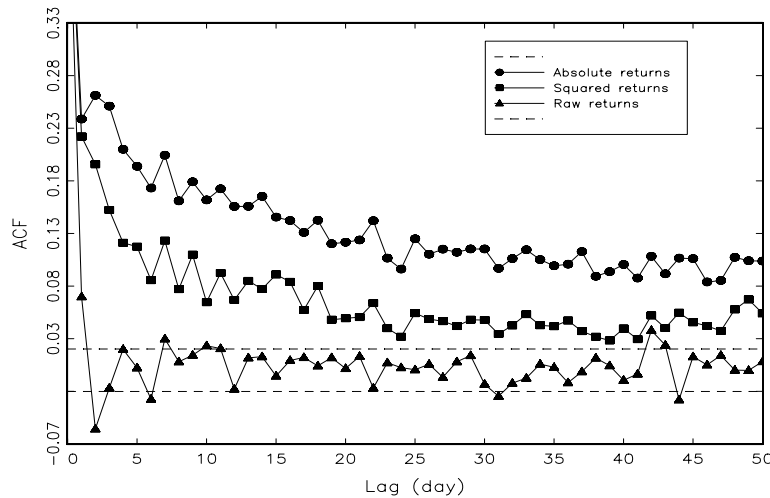


Figure 3.2: Panels (a), (b) and (c) show the correlograms of raw returns for raw, squared and absolute returns for the time series of DAX. The dashed lines refer to the 95% confidence interval.

comparison of the ACF of absolute, squared and raw returns, which reveals a positive and very persistent time-correlation for these simple non-linear transformations.

The existence of such high order return correlations, as opposed to the (almost) absence of autocorrelation of the returns themselves, is usually interpreted in terms of an auxiliary quantity called **volatility**, defined as a measure of the *conditional* variability of the return generating process as a function of past realizations. More formally, the above empirical evidences motivate a decomposition of the returns as a product of two quantities:

$$r(t, \Delta t) = \sigma(t, \Delta t)\varepsilon(t) , \quad (3.17)$$

where $\varepsilon(t)$ is a white noise process and $\sigma(t, \Delta t) > 0$ is the conditional volatility factor, whose dynamics is specified to match the empirical correlations. The decomposition (3.17) is the basic mathematical structure for a whole class of models⁸ that attempt to describe financial data, which has several theoretical implications:

called ‘leverage effect’. It means that a large negative return is more likely followed by large absolute price variations. Although the magnitude of this effect is negligible, when compared to the correlations in the level of fluctuations, it exhibits a very persistent behavior over an extended time horizon (see for instance the book of Bouchaud and Potters [2003] (pp. 137-141) and references therein).

⁸It has to be mentioned that the above decomposition is not the only mathematical structure that can explain the empirical facts on time-correlations of financial returns. One can imagine, in fact, a mixture of negative short-term correlations and positive long term dependence, which offset each other and lead to an apparent insignificant linear correlation. The assumptions underlined by eq. (3.17) are, at most, the simplest and the more intuitive, however not necessarily the more appropriated ones.

- The volatility process cannot be directly observed from the raw data, but has rather to be estimated using parametric or non-parametric models. In many empirical papers, the absolute or squared value of returns has been used as a proxy of volatility. However, as we can deduce from eq. (3.17), their values depend not just on $\sigma(t, \Delta t)$, but also on $\varepsilon(t)$.
- The auxiliary quantity σ is a latent variable, which is responsible for the majority of non-linearities detected in the data⁹, given that $\varepsilon(t)$ has no time correlation.
- The stochastic variable $\varepsilon(t)$ is itself an unobservable quantity, and its whiteness is a crucial ingredient to model the lack of (linear) predictability of returns and the resulting vanishing ACF.
- The formula (3.17) points out the fundamental distinction between the rule of the conditional variance $\sigma^2(t, \Delta t)$ and the unconditional variance $E[r^2(t, \Delta t)]$ when one attempts to model financial risk.
- The conditional volatility framework might be considered a generalization of the random walk model. The time-dependent standard deviation takes into account dependent increments.

In almost all financial data an alternation of periods of large and small market movements has been observed, the so-called **volatility clustering** phenomenon, which is the visual consequence of the empirically identified persistence in the level of the returns fluctuations. Figure 3.3 illustrates several examples of financial time series, comparing them with a time series of simulated independent Gaussian increments. Volatility clustering is another very robust feature of financial data, that has to be included in the list of stylized facts of financial data.

3.4 Heuristic estimation methods for long memory processes

The slow decay of the ACF for absolute and squared returns, as shown in Figure 3.2, rises the question whether these time series exhibit long-range dependence, or, more generally, whether the hidden volatility process is characterized by long memory. Starting with the seminal contribution by Ding et al. [1993], findings of long memory in simple non-linear transformations of stock returns, exchange rates or other financial assets are very numerous. Moreover, a parallel and active area of research has been developed on the detection of long-range dependence in raw returns, which, on the contrary, does not converge to such an unanimous consensus. This last issue is of primary importance in the debate on the efficiency of the market; the presence of long-memory in returns could be exploited by appropriate trading strategies and would, consequently, violate the efficient market hypothesis.

⁹A simple idea to incorporate the leverage effect in this theoretical framework is to introduce an appropriate negative correlation between the noise $\varepsilon(t)$ and the volatility process. This approach has been extensively used in the financial literature.

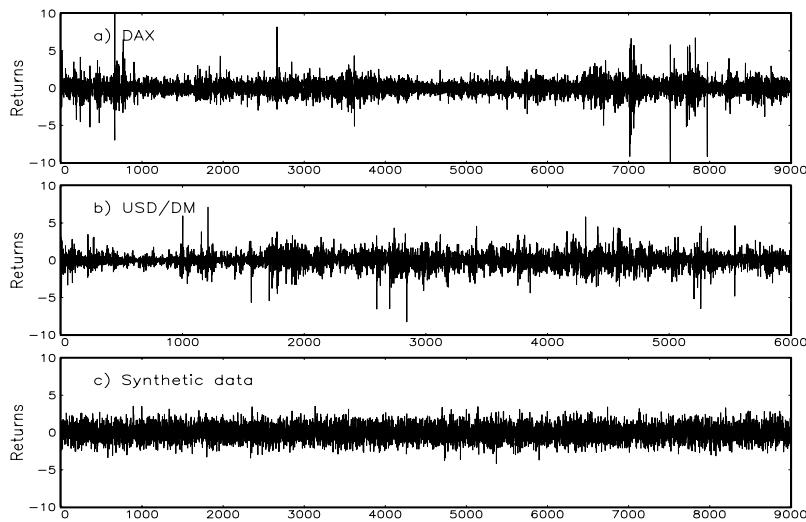


Figure 3.3: *The three panels show different return time series: (a) DAX daily log-returns; (b) USD/DM daily log returns; (c) random variables from a standard Normal. Note the presence of extreme events in the real data and the clusters of volatility, as compared to the ‘more regular’ behavior of the synthetic data.*

Turning to more practical problems, in order to detect and quantify the persistence of the level of the fluctuations, a number of different techniques have been developed in the literature, exploiting the scaling properties of long memory processes. It is interesting to highlight the broad spectrum of journals where these methods are published, ranging from hydrological engineering to physics or economics, which signals the ubiquitous role of the long-range dependence in natural phenomena. A common aspect of all the contributions is that the statistical inference for stationary long-memory processes is typically performed via semi-parametric techniques. They avoid, in fact, the detailed description of the short-run behavior of the system –which turns out to be essential if one attempts to estimate a parametric model– giving to the considered method the required flexibility to be applied to all these different contexts. In the following, a list of several methods is given:

- The rescaled adjusted range or R/S statistics, introduced originally by Hurst [1951] in the contest of hydrologic problems, and refined later by Mandelbrot [1973] and coauthors.
- The variance plot, which exploits the deviation from the linear behavior of the variance as a function of the sample size.
- The correlogram plot, which is based on the asymptotic property of the autocorrelation from the definition 1 in section 3.1.
- The power spectrum, which is related directly to the scaling behavior of the low-frequency part of the spectrum, as stated in definition 2 in section 3.1.

- Finally, the detrending fluctuation analysis (DFA hereafter), which is based on an alternative measure of the dispersion of the fluctuations rather than the rescaled range.

Simple transformations of these quantities (typically a logarithmic transformation) exhibit a linear behavior with a slope proportional to the Hurst exponent (H) or to the parameter of fractional differencing (d). A linear regression is, then, sufficient for an heuristic estimation of the long-memory parameter. Besides simplicity, these methods give only a rough assessment of the parameter H or d . However, they might be sometimes useful for discriminating among short- versus long-memory processes. In the following, a detailed description of three methods together with an illustrative application to financial data is given, leaving the explanation of the other techniques to the pertinent literature (for a review see the book of Beran [1994]).

R/S method

The *rescaled adjusted range* or *R/S* statistic is a semi-parametric method to estimate the Hurst exponent. It has been introduced by Hurst himself, when investigating the design of an optimal reservoir for the regulation of a river flow. His definition of the range of a stochastic process, in fact, reflects the underlying hydrological problem that he was trying to solve (for an interesting description along this line see footnote 12 of Lo [1991]). Going beyond the practical applications in hydrology, Mandelbrot and others have refined the methodology, supporting their research with a number of theoretical results on the asymptotic properties of the estimator. Moreover, they demonstrate the superiority of the *R/S* technique with respect to other methods based on alternative measures of the dispersion (variance analysis) or relying on the scaling of the autocorrelation. In particular, they have shown the robustness of the *R/S* estimator against leptokurtosis and fat-tailed marginal distribution of the underlying stochastic process.

Let us turn to the operational definition of the method. The *R/S* statistics is the range of the partial sums of deviations (positive or negative) of a time series from its mean, rescaled by its standard deviation (for more details see the book of Beran [1994]). It is, then, possible to demonstrate that the following asymptotic scaling law holds:

$$(R/S)_k \sim ak^H, \quad (3.18)$$

where H is the Hurst exponent, and k is the size of the intervals that constitute a partition of the original time series. The previous equation can be exploited in order to estimate the parameter H , via an ordinary least squares fit after a suitable log-linearization of eq. (3.18). Although the *R/S* technique is superior to other heuristic graphical methodologies, the method suffers from a number of important limitations: i) no asymptotic distribution theory has been derived for H ; ii) the estimator is strongly biased in the case of iid time series [Weron, 2002]; iii) the estimator is very sensitive to the departure from stationarity (e.g. the presence of slowly decaying trends) or short term-dependence.

Despite these limitations, quite a number of contributions from empirical finance and econometrics have employed the R/S methodology in order to investigate the presence of long term dependence in returns as well as in the volatility [Lo, 1991, Lux, 1996b]. The outcome of this research speaks clearly in favor of long memory in volatility, while ambiguous results can be found for returns. This controversial conclusions seem to be due to the bad performance of this estimator. It has been established, in fact, that the estimator possesses a positive bias for small and moderate samples, as shown by Weron [2002] among others, which can easily give the impression of a certain degree of long-memory, even if the considered stochastic process does not possess any correlation. Therefore, given the well-known limitations of the R/S methodology, we do not proceed further in the application of the estimator to investigate the nature of the dependence in our large sample of stocks from the Japanese market. Conversely, we confine our analysis only to some illustrative cases. Table 3.4 shows some examples of the estimation of the Hurst exponent using the R/S methodology, applied to different financial assets.

	H_r	$H_{ r }$	H_{r^2}
Gold	0.44	0.84	0.70
Dax	0.47	0.91	0.93
DB	0.48	0.80	0.68
Siemens	0.40	0.86	0.58
USD/DM	0.50	0.86	0.53

Table 3.4: Results of the estimation of the Hurst exponent using the rescaled range methodology for raw, squared and absolute returns. The results are indicative of absence of long-range dependence for returns, while strong indication for long-memory is observed for absolute returns. The estimated values of H for squared returns shows a somewhat ambiguous behavior (the value for USD/DM is very close to 0.5). The rescaled range varies from 50 points to the entire sample of the considered time series; an OLS estimation of H is computed on the linearized eq. (3.18). The procedure to compute the R/S statistic follows the methodology introduced by [Lo, 1991]. The truncation parameter in the valuation of the standard deviation using the Newey-West heteroscedasticity and autocorrelation consistent estimator is 5 for all the series.

DFA estimator

The detrended fluctuation analysis is the second method which we refer for the estimation of the Hurst exponent. It has been proposed by Peng et al. [1994], and applied by these authors for the investigation of correlations in the DNA sequences. The basic idea of the technique relies on a different way of measuring the dispersion of the fluctuations, compared to the rescaled range. Formally, given a sample of data $\{x_i\}_{i=1}^N$, one starts dividing the entire time series into d subseries of length n .

Next, for each subseries $m = 1, 2, \dots, d$, the user has to:

- create a cumulative time series $y_{i,m} = \sum_{j=1}^i x_{j,m}$ for $i = 1, 2, \dots, n$, where $x_{j,m}$ indicates the realization j in the sub-sample m ;
- fit a least squared line $\bar{y} = a_m i + b_m$ to each subseries $\{y_{i,m}\}_{i=1}^n$;
- compute the mean squared fluctuation of the integrated and detrended series:

$$f(m) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{i,m} - a_m i - b_m)^2}, \quad (3.19)$$

- finally, compute the mean value of $f(m)$ for all d subseries of length n :

$$F(n) = \frac{1}{d} \sum_{m=1}^d f(m). \quad (3.20)$$

The dependence of the variance on the box size follows a scaling law, given by:

$$F(n) \sim C n^{\tilde{H}}, \quad (3.21)$$

where C is a constant and \tilde{H} is related to the Hurst exponent H via the following transformation:

$$\tilde{H} = 2H - 0.5. \quad (3.22)$$

The DFA technique is more robust than the R/S statistics against spurious effects generated by non-linear trends in the data, which can give the impression of the presence of long-memory. Unfortunately, no asymptotic distribution theory has been derived for the DFA statistic so far. We can, however, rely on some Monte Carlo-based simulation studies for a numerical assessment of the confidence intervals (see Weron [2002] among others). Note that the DFA method can be generalized in order to eliminate higher order trends [Bunde et al., 2002], taking into account parabolic, cubic or quartic fits of the data, instead of limiting the filtering solely on a linear basis.

Table 3.5 exhibits the results of the application of the DFA to the 1219 assets of the Japanese stock exchange (see appendix A 2.1 for a detailed description of the data). As can be observed, the average value of the parameter \tilde{H} for the raw returns is very close to 0.5, which is the typical value of an iid stochastic process. Given the lack of an asymptotic theory for \tilde{H} , the judgment ‘very close’, however, remains a matter of subjective assessment, and it is not supported by a quantitative criterium. A different behavior is observed for absolute and squared returns, which seem to possess long-term dependence, since the average value of \tilde{H} is well above 0.5. Furthermore, in the majority of the cases (972 out of 1219), the value of \tilde{H} for absolute returns is higher than the corresponding value for squared returns. This is in line with the empirical finding that long memory is more pronounced for the first power of returns than for higher powers (see for instance Ding et al. [1993]). Figure

	Mean	Median	Std	Max	Min
\tilde{H}_r	0.487	0.481	0.041	0.662	0.376
$\tilde{H}_{ r }$	0.802	0.800	0.059	0.976	0.595
$\tilde{H}_{r,2}$	0.787	0.787	0.055	0.965	0.611

Table 3.5: Results of the estimation of the exponent \tilde{H} using the DFA methodology for raw, squared and absolute returns from the Japanese market. The length of the subintervals ranges from a minimum of 15 to a maximum of $0.85N$, where N is the total number of the data points. This range is, then divided in 50 equally spaced intervals, in the log-scale (see Figure 3.4).

3.4 shows an illustrative example of the application of the DFA to the time series of raw and absolute returns for a stock of the Japanese stock exchange. The scaling of $\langle F(t) \rangle$ as a function of the time resolution, in a log-log scale, is well fitted by a straight line, as predicted by eq. (3.21).

The GPH estimator

The method based on the scaling property of the spectral density, as detailed in section 3.1, was introduced by Geweke and Porter-Hudak [1983], and therefore known in the literature as **GPH estimator**. The authors demonstrate the equivalence of the spectral density of an ARFIMA model with the parameter of fractional differencing d , with a fractional Gaussian noise characterized by the Hurst exponent $H = d + 0.5$, making, then, possible the application of their estimator to a very general class of long-memory stochastic processes. Their main assumption is a semi-parametric approximation of the spectral density $S(\lambda)$ at the origin¹⁰:

$$S(\lambda) \sim c_f \lambda^{-2d} \quad \lambda \rightarrow 0^+, \quad (3.24)$$

where c_f is a positive constant and d is the parameter of fractional differentiation. The stationarity of the time series is guaranteed by the condition¹¹ $d \in (-0.5, 0.5)$. Eq. (3.24) can be translated into a linear relationship between $\ln[S(\lambda)]$ and $\ln(\lambda)$. Formally:

$$\ln[S(\lambda)] \sim \ln c_f - 2d \ln(\lambda). \quad (3.25)$$

¹⁰For an ARFIMA model the spectral density can be written in a closed-form:

$$S(\lambda) = \frac{\sigma^2}{2\pi} \{4 \sin^2(\lambda)\}^{-d} f_u(\lambda), \quad (3.23)$$

where d is the fractional differencing parameter and $f_u(\lambda)$ is the spectral density of the stationary linear component. If the condition $|\lambda| \ll 1$ holds, then eq. (3.23) can be written as eq. (3.24)

¹¹Velasco [1999], among others, have demonstrated the validity of the method for an enlarged interval. More precisely, he shows that in the interval $[0.5, 0.75)$, where the time series is non-stationary, asymptotic Normality and consistency is preserved as in the original interval $(-0.5, 0.5)$, while for values of d in the interval $[0.75, 1)$ the estimator is still consistent.

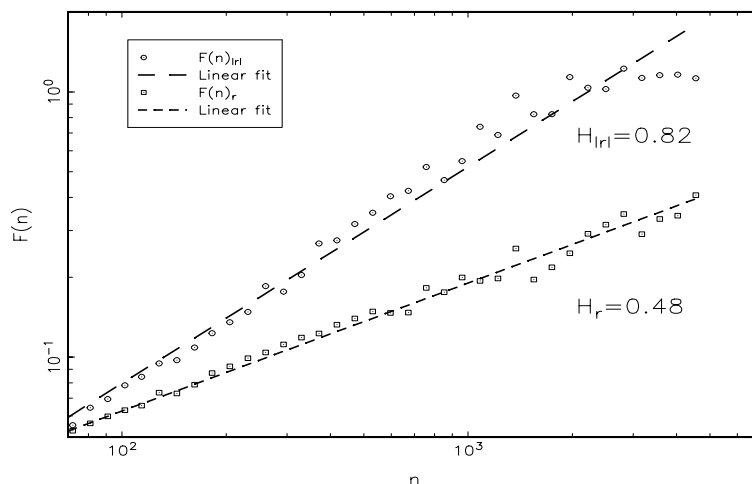


Figure 3.4: It is shown a typical behavior of a linear fit of the $F(n)$ as a function of the logarithm of the time window, according to the detrended fluctuation analysis. The two regressions refer to the raw and absolute returns.

The previous relation, which involves continuous quantities, can be rewritten in terms of the periodogram $I(\lambda_k)$, calculated at the *discrete* Fourier frequencies $\lambda_k = \frac{2\pi}{N}k$ for $k = 0, 1, \dots, (N-1)/2$, defined as:

$$I(\lambda_k) = \frac{1}{2\pi N} \left| \sum_{t=1}^N X_t \exp\left(\frac{-2\pi itk}{N}\right) \right|^2. \quad (3.26)$$

For short memory processes and, under some restrictions, even for strongly dependent data (see theorem 3.7 in the book of Beran [1994]), the periodogram is a consistent and unbiased estimator of $S(\lambda)$; the periodogram ordinates $I(\lambda_k)$ are, in fact, approximatively independent and exponentially distributed random variables with mean $S(\lambda_k)$. We can express eq. (3.25) in a more suitable form:

$$\ln[I(\lambda_k)] \sim \ln c_f - 2d \ln(\lambda_k) + \ln \xi_k, \quad (3.27)$$

where $\xi_k = \frac{I(\lambda_k)}{S(\lambda_k)}$ are independent random variables, drawn from the standard Exponential distribution. Introducing the following notations:

$$\begin{aligned} y_k &= \ln I(\lambda_k), \\ x_k &= \ln \lambda_k, \\ \beta_0 &= \ln c_f - c \quad -c = E[\ln \xi] = -0.577, \\ \beta_1 &= -2d, \end{aligned} \quad (3.28)$$

eq. (3.27) becomes a regression equation with iid errors, $e_k = c + \ln \xi_k$, with mean zero:

$$y_k = \beta_0 + \beta_1 x_k + e_k. \quad (3.29)$$

It is, then, straightforward to estimate the value of d by mean of a least squared regression. Besides simplicity, the non-parametric nature of the method constitutes its main advantage. Consistency is, then, obtained without any necessary characterization of the generating process, while, for asymptotic Normality, more restrictive assumptions are needed (see Geweke and Porter-Hudak [1983], Velasco [1999], Robinson [1995]). Interestingly, the knowledge of the asymptotic distribution of the estimator allows to explicitly test for the null-hypothesis of short *vs* long-term dependence.

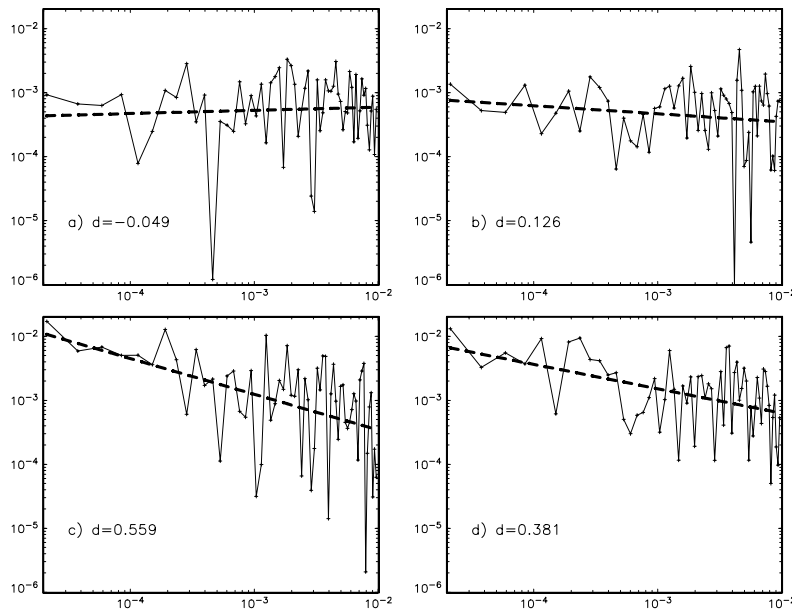


Figure 3.5: The four panels show the linear fit of the periodogram regression according to the GPH estimator [Geweke and Porter-Hudak, 1983]. Panels (a) and (c), which refer to the raw and absolute returns respectively, show the OLS on the raw power spectrum, while panels (b) and (d) exhibit the linear fit on the tapered spectrum. We applied the GPH to the same time series used in Figure 3.4. For the raw returns, the null hypothesis of absence of long-memory cannot be rejected at any conventional significance level, while it is rejected in the case of absolute returns, given the low p -values, virtually zero, for both non-tapered and tapered versions. The OLS has been performed considering the lowest $64 \approx \sqrt{N}$ frequencies, and neglecting the first $2 \approx N^{0.1}$ frequencies, as suggested by Robinson [1995]. Note that in all the panels we have used, scale in the y axes is the same which facilitates comparison.

Although the estimator presents some advantages, several problems arise with the approximations involved in the method: eq. (3.24) is a good approximation in the region around the origin, while it poorly captures the entire spectrum. To overcome

this problem, a ‘small’ interval of lower frequencies should be used in the regression, at a cost of a lower precision. Typically the number of frequencies is $N^{0.5}$, as suggested originally by Geweke and Porter-Hudak, or $N^{0.8}$, which turns out to be MSE-optimal (see Hurvich et al. [1998]). In the literature, a number of data-driven criteria are available, which avoid an *ad hoc* choice of the frequency cut-off. Moreover, for very low frequencies, the unbiasedness and independence of the periodogram ordinates become poor approximations, consequently the assumptions of the least squares regression do not hold any longer. A solution, proposed by Robinson [1995], is to leave out also the lowest l frequencies, creating a lower cut-off in the spectrum. Additionally, the disturbances in the regression (3.29) are highly skewed, reducing the efficiency of the GPH estimator, which is based on the underlying assumption of symmetric disturbances. In order to solve the problem, however, one should leave the OLS method and use a parametric approach. Some interesting improvements of the efficiency of the GPH estimator can be reached by applying weighted least squared with appropriate weights, as suggested by Robinson [1995].

Another extremely important issue is the robustness of the test against the presence of structural breaks or slowly decaying deterministic trends, that can be misspecified as long memory processes. Heyde and Dai [1996] have shown that the GPH is robust against fast decaying trends, while Sibbertsen [2003] has demonstrated that it is strongly biased for slowly decaying trends. In order to avoid or reduce this bias, the GPH-regression is modified using a tapered periodogram, defined as:

$$I_T(\lambda_k) = \frac{1}{2\pi A} \left| \sum_{t=1}^N w_t X_t \exp\left(\frac{-2\pi i t k}{N}\right) \right|^2, \quad A = \sum_{t=1}^N w_t^2. \quad (3.30)$$

In the case of the cosine-bell tapered, the weights w_t are given by:

$$w_t = \frac{1}{2} \left[1 - \cos\left(\frac{2\pi(t + 0.5)}{N}\right) \right]. \quad (3.31)$$

The role of the tapered is to reduce the influence of the smallest frequencies in the spectrum, which are connected to the presence of the trends. Velasco [1999] has shown the consistency and asymptotic normality of the tapered estimator. Sibbertsen [2003] performed Monte Carlo simulations, confirming the theoretical results. The main drawback of this technique results in an increased variance of the estimator, since its major effect is, roughly speaking, to reduce the effective length of the data. Figure 3.5 shows an example of the application of the GPH estimator to the same time series used in Figure 3.4, and a direct comparison between the tapered and raw version of the estimator.

Table 3.6 shows the descriptive statistic of the GPH estimates for the entire pool of stocks of the Japanese market. The results are pretty much in line with related contributions that refer to other markets, [Lux, 2001, Lo, 1991]: absence of long-range dependence in returns and strong evidence for long memory in absolute and squared returns. The dispersion around the mean, measured by the standard deviation, is very similar for the original GPH method and for the tapered alternative, although

	Mean	Median	Std	Max	Min
Range $N^{0.1} - N^{0.5}$					
d_r	-0.021	-0.022	0.112	0.347	-0.447
d_r^T	-0.017	-0.017	0.142	0.491	-0.573
$d_{ r }$	0.352	0.356	0.120	0.760	-0.085
$d_{ r }^T$	0.309	0.312	0.131	0.731	-0.115
d_{r^2}	0.301	0.305	0.128	0.834	-0.223
$d_{r^2}^T$	0.260	0.262	0.131	0.855	-0.235
Range $N^{0.1} - N^{0.8}$					
d_r	-0.042	-0.041	0.042	0.232	-0.263
d_r^T	-0.023	-0.021	0.051	0.187	-0.269
$d_{ r }$	0.205	0.204	0.037	0.405	0.036
$d_{ r }^T$	0.213	0.213	0.044	0.362	0.036
d_{r^2}	0.184	0.183	0.043	0.550	0.052
$d_{r^2}^T$	0.193	0.192	0.048	0.387	0.039

Table 3.6: Descriptive statistics of the estimates of fractional differencing parameter using the GPH methodology for raw, squared and absolute returns. The superscript T refers to the tapered version. The two parts refer to two different upper cut-off values for the frequency region.

we observe slightly higher values for the tapered version, as expected by its intrinsic properties. The enlarged frequency region, the second part of Table 3.6, exhibits a much smaller standard deviation of the estimates. However, they might be strongly biased due to the influence of the short-run behavior of the underlying process. A deeper look at the results concerning absolute and squared returns shows that the average estimates for the tapered version are slightly lower than the original estimates using the smaller frequency range, and systematically higher in the second range. Given the superior robustness of the tapered estimator against potential non-stationarity, this might indicate that a somewhat ‘correct’ average value of the fractional differencing parameter could lie within the interval bounded by the means of the tapered version in the two ranges. Finally, we can compare the estimates of d using the GPH method with those of H performed via the DFA (note that the two quantities are connected by the simple relation $H = d + 0.5$). The overall picture of correlations in returns and volatility is qualitatively equivalent. However, the linear relation between H and d is not exactly respected.

3.5 Conclusions

In this chapter, we have analyzed the time-correlation properties of financial returns and their simple transformations. The near absence of dependence in raw returns and the persistence correlation, for instance, in their absolute values has been reconciled in the framework of dynamic volatility processes for financial data. We have,

in fact, introduced the useful concept of volatility in order to describe the empirical regularities of financial time series. The precise characterization of those empirical regularities has been extensively analyzed by means of non-parametric statistical techniques using real data from the Tokyo Stock Exchange. The detailed empirical results presented in this chapter will be successively compared with the analogous properties of the agent-based model proposed in the next chapter.

An analytical approach to an herding model

4.1 A simple herding model as a jump Markov process

The following model is inspired by Kirman's seminal paper [Kirman, 1993] on information transmission in an ant colony. He provides a simple stochastic formalization of information transmission introduced to explain the macroscopic patterns emerging from entomologic experiments with ant colonies. The underlying scenario is one of foraging ants who have two identical sources of food at their disposal in the vicinity of their nest. Experimental settings show that at any point in time, a majority of the ant population concentrates on exploiting one particular food source, but over time switches may occur of the preferred source. Thus, averaging over time, a symmetric bimodal distribution of the frequency of ants visiting one or the other manger would result. In order to explain the alternation of the majority of ants in the exploitation of the two sources of food, Kirman introduces an interaction mechanism based on an exchange of information in pair-wise random meetings combined with an autonomous behavior due to stochastic search. In other words, the ants, when exploring a given searching area, S , might act as independent entities or might randomly meet a companion and, by exchanging pheromons, signal the presence of food in the area. The transmitted information is not the exact spatial position of the source of food, but it is simply limited to a signal of the type 'follow me', which might be perceived by the other ant with a certain probability. The recruitment mechanism is interpreted as a herding-based interaction, in the sense that one ant might give up its own private information following, then, the companion's 'suggestion'. The random meeting is limited to pair-wise interaction of two ants, which implicitly means that the density of ants in the area S is low in order to exclude multiple encounters. Additionally, *neither* the probability of following another ant *nor* the success in recruiting companions depend on the outcome of previous meetings. Kirman assumes that the ants are, therefore, memoryless.

The entomologic experiment can be formalized within the mathematical apparatus introduced in the previous section. It possesses, in fact, all the necessary ingredients: a large group of individuals (ants) characterized by a finite number of categories (different information about the *two* sources of food), that interact with each other exchanging information on their environment. At the micro level, the interaction among ants can be modeled as a Markov chain, given the randomness and memoryless character of the encounters. Note that the system is permeated by randomness. The encounters, the behavior of a single ant searching for food and

also the outcome of the recruitment mechanism possess random ingredients. The underlying idea is that if we do not model the precise details of the various aspects of the system (the search behavior and the exchange of pheromons, for instance), randomness seems a suitable choice in order to keep simplicity and to avoid arbitrary assumptions. On the other hand, the large number of constituents allows to apply powerful mathematical tools in order to aggregate over this random elements and to derive meaningful macroscopic laws. The ultimate goal of the formalism, in fact, is to provide a dynamical characterization of macroscopic quantities, such as the evolution of the number of ants near one source of food, or the average time to observe a switch of majority, and not to focus on the detailed reasons or incentives for the ants to follow this or that searching path. Moreover, each constituent ‘looses’ his identity, since we somehow aggregate over all the system, and, therefore, the individuals are indistinguishable (see [Aoki, 1996]).

4.1.1 Transition rates in the Kirman’s model

The core of the model consists in the mathematical formalization of the information transmission mechanism in terms of transition probabilities. To set the stage, let us first restate the mechanics of the recruitment-herding model proposed by Kirman [1993]. The system (the ant colony) is populated by a fixed number of individuals N , each of them being either in state 1 or 2 (states stand for sources of food). The number of agents in the first state will be denoted by n , so that $n \in \{0, 1, \dots, N\}$ defines the state of the system (in the opposite state the number of agents will obviously be $N - n$). Its stochastic evolution is governed by (i) random meetings of two agents, after which one of them might follow the companion, therefore changing state, and (ii) autonomous switches, which model the possibility for a random change of state. The stochastic population dynamics evolves according to the probabilities of changing from state n at time t to some state n' at time $t + \Delta t$. The memoryless of the agents is the crucial assumption to formalize the model as a Markov process. Let these conditional probabilities be denoted by $\rho(n', t + \Delta t | n, t)$ which are related to the transition rates per unit time, $\pi(n \rightarrow n')$ by $\rho(n', t + \Delta t | n, t) = \pi(n \rightarrow n') \Delta t$, for small time increments Δt . Since in the limit of continuous time $\Delta t \rightarrow 0$, multiple switches during one incremental time unit become increasingly unlikely, one can confine the analysis to $n' = n \pm 1$ with transition probabilities:

$$\begin{aligned} \rho(n + 1, t + \Delta t | n, t) &= (N - n) \left(a + \frac{1 - \delta}{N} n \right) \Delta t, \\ \rho(n - 1, t + \Delta t | n, t) &= n \left(a + \frac{1 - \delta}{N} (N - n) \right) \Delta t, \\ \rho(n, t + \Delta t | n, t) &= 1 - \rho(n + 1, t + \Delta t | n, t) - \rho(n - 1, t + \Delta t | n, t), \end{aligned} \tag{4.1}$$

where $1 - \delta$ is the probability of a successful recruitment. An obvious restriction to eq. (4.1) is that the probability $\rho(n, t + \Delta t | n, t)$ should be positive, which fixes an upper bound to the elementary time-step Δt :

$$\Delta t \leq \frac{2}{N(1 - \delta + 2a)}. \tag{4.2}$$

To restate the features of the herding ant model sketched in the previous section, one should consider the following elements. First, the term proportional to the number

of ants in the state, via the parameter a , models their independent random switches, while the quadratic term $\frac{1-\delta}{N}(N-n)n$ governs the recruitment mechanism acting in the market via the dependence on the fraction of individuals in the other state. The probability to meet an agent in the other state is proportional to the overall fraction of agents in the alternative state. Once two agents meet, a switch of one agent to the alternative state happens with probability $1 - \delta$. Note that the symmetric quadratic term with respect of the two states in eq. (4.1) models the assumed perfect symmetry in the recruitment mechanism. It means that there is not any preferential tendency to switch in a specific state due to herding behavior. The Markovian nature of the transition probabilities reflects, then, the lack of memory of the agents.

More technically, the above transition probabilities define a finite Markov chain, or, in particular, a birth-and-death process¹ with the following properties²:

- the transition probabilities are *homogeneous*, i.e. time independent;
- the chain is *irreducible*, i.e. every state can be reached from every other state in a finite number of steps;
- the chain is *aperiodic*, since the probability $p(n, t + \Delta t | n, t) \neq 0$ for all the states;
- the chain is ergodic, since it is aperiodic and irreducible;
- the equilibrium distribution $P_{e,N}(n)$ of the chain exists and is unique; moreover, it will be reached for any starting distribution $P_{0,N}(n)$.³

The underlying Markovian nature of the chain, and also of the entire formalism, is a crucial ingredient, in order to exploit a series of theoretical results like the existence and uniqueness of the equilibrium distribution. It should be stressed that the Markovian assumption is a very restrictive constrain on the transition probabilities of the model. Nevertheless, we will show that this very simple model can generate a multiplicity of interesting scenarios. We might, therefore, argue that, despite the limitation of the memory of the stochastic processes, this formalism possesses enough descriptive power to be fruitfully applied in many different situations.

4.1.2 Generalized transition rates

The transition probabilities given by (4.1) can be generalized in order to obtain an enhanced flexibility in the model. A deeper look at eq. (4.1) shows an underlying symmetry. They are, in fact, invariant under the transformation

$$n \rightarrow N - n, \quad (4.3)$$

which amounts to an interchange of the states. This invariancy depends on the fact that the recruitment mechanism is independent of the type of agents, and the

¹A birth-and-death process is limited to transitions between neighboring states.

²For all the details on the Markov chain properties and the related theorems we refer to the books of Feller [1971] and Kelly [1979].

³The subscript N indicates the finite number of agents.

random switching component a is equal in both states. The system, on average, does not exhibit any preferential direction between the two states. Intuitively, in the entomologic experiment, this invariance could be explained in terms of perfect identity of the two sources of food and perfect identity of the ants⁴. It should be pointed out here that the original intension of Kirman was to account for an endogenous emergence of temporary asymmetric feeding of the ants, given an overall symmetry of the sources of food. In order to adapt the model to other contexts than the ant colony, nothing prevents one from generalizing the original setting to account for exogenous asymmetries. In this respect, an alternative setting could be based on asymmetric transition probabilities, i.e. a new mathematical formalization which breaks the invariance of the system under the transformation (4.3). A quite intuitive modification might be to replace the autonomous switching parameter a by two different parameters a_1 and a_2 , which could be interpreted as an intrinsic asymmetry between the two states. A combination $a_1 > a_2$, for instance, can generate a biased movement of the agents towards the state 2, therefore breaking the perfect symmetry implied in eq. (4.1). Pretty much in the same way, we can introduce a differentiation in the herding probability δ , with two new values δ_1 and δ_2 , which breaks the symmetric character of the recruitment mechanism. We can show later that the symmetric interaction in the herding mechanism is a crucial ingredient in the model, in order to obtain non trivial dynamics. In the following, when not explicitly mentioned, we will refer to the symmetric herding mechanism, implied by identical herding coefficients.

An additional interesting modification is to go beyond the pair-wise character of the recruitment mechanism. If we replace a N -dependent herding parameter - the factor $(1 - \delta)/N$ in the transitions (4.1) - by a constant parameter b independent on the number of agents, we introduce a *global coupling* among all the individuals, instead of an indirect interaction via the random pair-wise meetings. It is not immediately obvious that this tiny algebraic modification in the transition rates can drastically change the nature of the interactions within the population. A more detailed explanation of the arguments will be given later on in this chapter.

We introduce now the new transition rates, which are a generalization of the Kirman's formalization⁵:

$$\begin{aligned}\rho(n+1, t + \Delta t \mid n, t) &= (N - n)(a_1 + bn) \Delta t, \\ \rho(n-1, t + \Delta t \mid n, t) &= n(a_2 + b(N - n)) \Delta t, \\ \rho(n, t + \Delta t \mid n, t) &= 1 - \rho(n+1, t + \Delta t \mid n, t) - \rho(n-1, t + \Delta t \mid n, t).\end{aligned}\tag{4.4}$$

The above transition probabilities define a one-step process with the same math-

⁴Kirman [1993] pointed out the self reinforcing character of one of the setting of the experiment. In order to maintain the same level of food and, therefore, to guarantee for the perfect symmetry of the two sources, "the experimenter has to supply more food", generating an asymmetry in its flow between the two sources. To avoid this artificial situation, an alternative setting has been proposed where two symmetric bridges lead to just one source of food. Even for this setting, the ants behavior shows an asymmetric exploitation, in this case of the bridges rather than the sources of food.

⁵We can recover the Kirman's formalization if we posit $a = a_1 = a_2$ and $b = \frac{(1-\delta)}{N}$.

ematical properties of the Markov chain given by eq. (4.1), namely homogeneity, ergodicity, existence and uniqueness of the equilibrium distribution. However, the underlying symmetry under the transformation (4.3) does not hold anymore, which mirrors in an asymmetric equilibrium distribution (see section 4.2). The upper limit for the elementary time step, in order to have positive quantities in eq. 4.4), is:

$$\Delta t \leq \frac{2}{bN \left(N + \frac{a_1+a_2}{b} + \left(\frac{a_2-a_1}{2b} \right)^2 \right)}, \quad (4.5)$$

which scales approximatively as $\sim \frac{1}{N^2}$.

4.2 A theoretical finitary approach

Given the transition probabilities (4.1) and (4.4), we can explicitly compute the equilibrium distribution, denoted by $P_{e,N}(n)$, where the subscript N indicates the *finite* and arbitrary number of agents. We know already that $P_{e,N}(n)$ exists and is unique, therefore we have to find a strategy to operatively compute it. In order to do so, let us start from the Chapman-Kolmogorov equation for the chain:

$$P(n'_{t+\Delta t}) - P(n'_t) = \sum_{n' \neq n} \left(\rho(n'|n)P(n_t) - \rho(n|n')P(n'_t) \right), \quad (4.6)$$

where n' belong to the set of first neighbors, i.e. $n' \in \{n-1, n, n+1\}$. With $P(n_t)$, we denote the probability that at time t there are n agents in state 1. We adopt the notation $\rho(n|n')$ for the conditional probabilities instead of the more lengthy expressions in eqs. (4.1) and (4.4). Let us assume that for any couple $n' \neq n$ we have:

$$\rho(n'|n)P(n_t) = \rho(n|n')P(n'_t). \quad (4.7)$$

If we plug eq. (4.7) into eq. (4.6), then we obtain:

$$P(n_{t+\Delta t}) = P(n_t) \equiv P_{e,N}(n). \quad (4.8)$$

The probability satisfying eq. (4.8) is dynamically invariant and, by the uniqueness theorem, is equal to the equilibrium distribution $P_{e,N}(n)$ (see Garibaldi et al. [2003]). Equation (4.7) is called detailed balance condition, which states that, at the equilibrium, the probability flux from an arbitrary state n of the chain to another state n' should equal the reverse flux from n' to n . The detailed balance condition is directly connected to an important property of the chain. It is possible to show that for a stationary Markov chain the detailed balance condition holds if and only if the chain is *reversible* (see Kelly [1979] pp.5-7). Time-reversibility of the chain implies that, at the equilibrium, the probability of going from a state n_0 to any other state of the chain n depends solely on n and n_0 , but not on the particular path. In other words, the equilibrium properties of the chain are invariant under a time reversal transformation (see ch. 1 in Kelly [1979]).

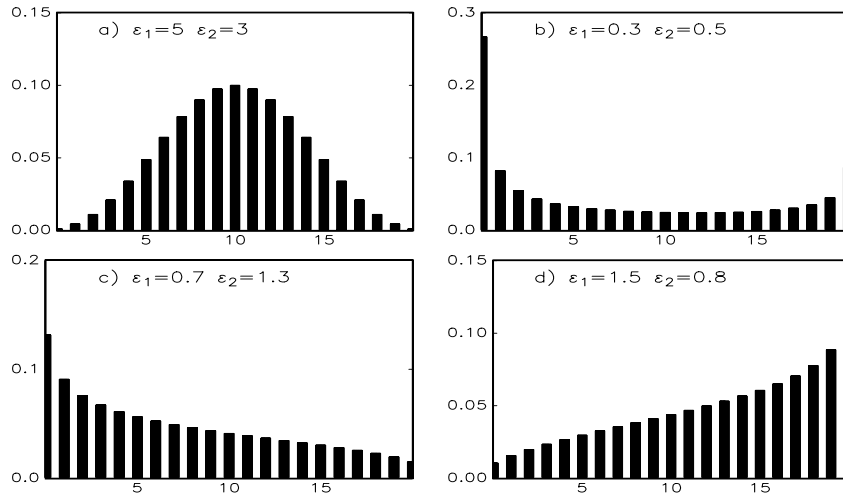


Figure 4.1: The four panels show different equilibrium probability distributions derived from eq. (4.12) for several choices of the parameters ε_1 and ε_2 . In the panel (a) a symmetric uni-modal distribution, in (b) a bi-modal distribution and in panels (c) and (d) two cases of asymmetric monotonic distributions are shown. Note that the probability distribution is drawn as an histogram to highlight its discrete character. The underlying number of agents is $N = 20$.

Turning now to the derivation of the equilibrium distribution, eq. (4.7) provides us with a simple recursive equation for the equilibrium probabilities:

$$\frac{P_{e,N}(n+1)}{P_{e,N}(n)} = \frac{(N-n)(a_1 + bn)}{(n+1)(a_2 + b(N-n-1))}. \quad (4.9)$$

If we assume that $b \neq 0$, we can define two new parameters:

$$\varepsilon_1 = \frac{a_1}{b} \quad \text{and} \quad \varepsilon_2 = \frac{a_2}{b}. \quad (4.10)$$

We can rewrite eq. (4.9) as follows:

$$\frac{P_{e,N}(n+1)}{P_{e,N}(n)} = \frac{(N-n)(\varepsilon_1 + n)}{(n+1)(\varepsilon_2 + N - n - 1)}. \quad (4.11)$$

By induction (see Mansour and de Palma [1984]), we arrive at the final solution⁶:

$$P_{e,N}(n) = \frac{\Gamma(N+1)}{\Gamma(\varepsilon_1 + \varepsilon_2 + N)} \frac{\Gamma(\varepsilon_1 + \varepsilon_2)}{\Gamma(\varepsilon_1)\Gamma(\varepsilon_2)} \frac{\Gamma(\varepsilon_1 + n)}{\Gamma(n+1)} \frac{\Gamma(\varepsilon_2 + N - n)}{\Gamma(N - n + 1)}. \quad (4.12)$$

We can alternatively express eq. (4.12) as a Polya distribution (see Garibaldi et al. [2003]):

$$\text{Polya}(n; N, \varepsilon_1, \varepsilon_2) = \frac{N!}{(\varepsilon_1 + \varepsilon_2)^{[N]}} \frac{\varepsilon_1^{[n]}}{n!} \frac{\varepsilon_2^{[N-n]}}{(N-n)!}, \quad (4.13)$$

⁶Note that if x is an integer, $\Gamma(x+1) = x!$.

where $x^{[N]} \equiv x(x+1)\dots(x+N-1)$ is called ascending factorial [Aoki, 2002]. Figure 4.1 shows some examples of possible distributions for several choices of the two parameters.

The finitary approach for the dynamics of a Markov chain is in general complicated, and in very few cases can lead to close form solutions. In our case, for instance, it is not easy to deal analytically with the resulting Polya distribution, given the presence of various factorials, although we have been able to derive a closed-form expression for the equilibrium distribution. The finitary approach can be overcome introducing a continuous approximation for the stochastic process (4.4) in the limit of a large collection of agents. We will show in the following sections two different approximation schemas for the jump Markov process (4.4), the Fokker-Plank equation and the Langevin equation.

4.3 The diffusion approximation

In this section we will derive a continuous approximation for the stochastic process (4.4). It means that the discrete stochastic process governing our population dynamics can be approximated by a continuous diffusion process in a new variable:

$$z = \frac{n}{N}. \quad (4.14)$$

This approximation leads to the so called Fokker-Planck equation, which is a elliptic partial differential equation governing the time evolution of the probability density $p(z, t)$, given the starting distribution $p_0(z_0)$ and the boundary conditions [Van Kampen, 1992]. In the following, we focus attention on the generalized form of the transition probabilities (4.4). The analysis of the original Kirman's model within our formalism will be described in details in section 4.11, in relation to its local character and the consequences for the macroscopic behavior in the limit of large system size.

The remaining part of the chapter will be rather technical, since it provides the application of the stochastic formalism for the analysis of the herding model. It is not possible, in fact, to avoid the presentation of many proofs and derivations. Despite the fact that some of the results are already known in the literature, some other constitute a novelty when interpreted within an economic setting which focuses on different questions than the natural sciences, where this stochastic formalism is mainly applied. For example, the global/local framework related to some formal differences between eq. (4.1) and eq. (4.4) constitutes a novel viewpoint. Moreover, many of the results of the calculations of the following part of the thesis are necessary in order to derive economically significant consequences in the last two chapters. Finally, the mathematical elegance of the formalism which describes the herding model should be appreciated. It embeds in a unitary framework several subclasses of classical stochastic processes that exhibit closed-form solutions.

4.3.1 Derivation of the Fokker-Plank equation

Let us start deriving the Fokker-Plank equation as a second-order Taylor approximation to the continuum of our population dynamics given by eq. (4.4). In order to do so, we have to express the transition probabilities (4.4) as a function of the variable z , that, if the total number of agents N is large enough, can be treated as a continuous quantity. The relation between the transition rates (4.4), expressed in terms of the variables n or z , is given by the following formula:

$$\pi_n^\pm = N^2 \pi_z^\pm . \quad (4.15)$$

The transition rates (4.4) are now functions⁷ of z :

$$\pi(z \rightarrow z + 1/N) = \pi_z^+ = (1 - z) \left(\frac{a_1}{N} + b z \right) , \quad (4.16)$$

$$\pi(z \rightarrow z - 1/N) = \pi_z^- = z \left(\frac{a_2}{N} + b (1 - z) \right) . \quad (4.17)$$

Let us introduce the “step” operators⁸ \mathbf{E} and \mathbf{E}^{-1} , using the notation of Van Kampen [1992]. Their effect on an arbitrary function $f(n)$ are respectively to add to or to drop off one unit to their integer argument n . Formally:

$$\mathbf{E}[f(n)] = f(n + 1) \quad \text{and} \quad \mathbf{E}^{-1}[f(n)] = f(n - 1) . \quad (4.18)$$

With the aid of these operators $\mathbf{E}[\cdot]$ and $\mathbf{E}^{-1}[\cdot]$, the Chapman-Kolmogorov equation (4.6) for the one-step process can be conveniently rewritten:

$$\frac{\partial}{\partial t} P_{n,t} = (\mathbf{E} - 1)[\pi_n^- P_{n,t}] + (\mathbf{E}^{-1} - 1)[\pi_n^+ P_{n,t}] , \quad (4.19)$$

where $P_{n,t}$ is the probability to have n agents at time t in state 1, and π^\pm are the transition rates (4.4). The time derivative in the left-hand side of eq. (4.19) stands for the continuous time approximation for the evolution of the probability $P_{n,t}$. This equivalent form of the Chapman-Kolmogorov equation (4.6) is called, in the pertinent literature, *Master equation*, which more clearly, can be interpreted as a “gain-loss equation for the probabilities of the separate states n ” (cf. [Van Kampen, 1992] pp. 97).

We can now approximate the Master equation (4.19) with the Fokker-Plank equation. With the new variable z and leaving out the obvious time dependence of $P_{z,t}$, Eq. (4.19) assumes the form:

$$\frac{\partial}{\partial t} P_z = (\mathbf{E} - 1)[\pi_z^- N^2 P_z] + (\mathbf{E}^{-1} - 1)[\pi_z^+ N^2 P_z] , \quad (4.20)$$

where the transition rates are given by (4.16) and (4.17). Note that the probabilities are invariant under this transformation, therefore $P_n = P_z$. This transformation

⁷The transition rates are the transition probabilities per time unit. The new transitions read as $\pi_z^\pm = \pi(z \rightarrow z \pm 1/N)$.

⁸Note that the step operator is denoted in bold letter to distinguish it from the expectation operator, which is denoted in normal character.

amounts to label the variable with a different name, from n to z , without changing its cardinality. The probability density of the variable z is defined according to the following limit:

$$p_{z,t} = \lim_{N \rightarrow \infty} \frac{P_{z,t}}{\Delta z} = \lim_{N \rightarrow \infty} N P_{z,t} . \quad (4.21)$$

where p_z is a continuous function of z . Eq. (4.20) can be rewritten as:

$$\frac{\partial}{\partial t} p_z = N^2 \{ (\mathbf{E} - 1) [\pi_z^- p_z] + (\mathbf{E}^{-1} - 1) [\pi_z^+ p_z] \} . \quad (4.22)$$

Now, since the step operator acts just on continuous functions, we may use its Taylor expansion up to the second order:

$$\mathbf{E}[f(z)] = f(z + \Delta z) = f(z) + \Delta z \frac{df}{dz}(z) + \frac{1}{2} \Delta z^2 \frac{d^2 f}{dz^2}(z) + o(\Delta z^2) , \quad (4.23)$$

where $\Delta z = 1/N$. Formally, we can substitute the step operator with the following expansion:

$$\mathbf{E} = 1 + \Delta z \frac{\partial}{\partial z} + \frac{1}{2} \Delta z^2 \frac{\partial^2}{\partial z^2} + o(\Delta z^2) . \quad (4.24)$$

Note that the derivatives in eq. (4.23) are replaced in eq. (4.24) by the partial derivatives, since the function $f(\cdot)$, which the operator is applied to, might have more than one variable. The expansion for \mathbf{E}^{-1} is simply obtained from the previous formula replacing Δz with $-\Delta z$. Using the expansions (4.24) for \mathbf{E} and \mathbf{E}^{-1} until the second order, we end up with:

$$\frac{\partial}{\partial t} p_z = N^2 \left\{ -\Delta z \frac{\partial}{\partial z} [(\pi_z^+ - \pi_z^-) p_z] + \frac{1}{2} \Delta z^2 \frac{\partial^2}{\partial z^2} [(\pi_z^+ + \pi_z^-) p_z] \right\} . \quad (4.25)$$

The N^2 factor in front of the equation disappears⁹, and we arrive finally to the Fokker-Planck equation:

$$\frac{\partial}{\partial t} p_z = -\frac{\partial}{\partial z} A_z p_z + \frac{1}{2} \frac{\partial^2}{\partial z^2} D_z p_z , \quad (4.26)$$

where the drift and diffusion functions are respectively given by:

$$A_z = N(\pi_z^+ - \pi_z^-) = (a_1 - (a_1 + a_2)z) , \quad (4.27)$$

and

$$D_z = \pi_z^+ + \pi_z^- = \frac{1}{N} (a_1 - (a_1 - a_2)z) + 2b z(1 - z) . \quad (4.28)$$

Some comments are in order here. The drift function can be rewritten in terms of the expected value $E[z] = \frac{a_1}{a_1 + a_2}$ of the process (see paragraph 4.5):

$$A(z) = b (\varepsilon_1 + \varepsilon_2)(E[z] - z) . \quad (4.29)$$

⁹The independence of the overall dynamics on the number of agents N is a remarkable property of the system, which will be analyzed in details in section 4.4.

The process is characterized by a linear mean reversion towards its unconditional mean.

The diffusion function, that governs the properties of the fluctuations of the system, is composed of two terms: the first term is a linear N -dependent function of z , that becomes negligible for large values of N ; the second term is a symmetric and quadratic function of z , independent on N . The two terms characterize *two different regulating mechanisms* for the emergence of endogenous fluctuations. The linear term governs the fluctuations due to its intrinsic granularity, that obviously vanishes in the continuous approximation. The second term is directly related to the **non-local interactions** among its constituents, and, consequently, does not disappear when enlarging the system size (see paragraph 4.4 for further comments on this point). If we are not too close to the boundaries, the herding mechanism is dominating, and the fluctuations are governed by the quadratic term, while the linear N -dependent term can be neglected. However, near the boundaries, the two effects are comparable in magnitude, therefore the linear term should be taken into account. In the next paragraph, we will analyze its measurable effect on the stationary distribution.

Finally, it should be emphasized that the Fokker-Planck equation (4.26) is an approximation to the continuum of the intrinsically discrete stochastic process (4.4). This equation holds if $\frac{\Delta z}{1-z} \ll 1$ and $\frac{\Delta z}{z} \ll 1$, i.e. not too close to the boundaries, where the discreteness of the variable z cannot be any longer neglected. More precisely, the discreteness of the process can be neglected if the $z(1-z) \gg \frac{1}{2N}(\varepsilon_1 - (\varepsilon_1 - \varepsilon_2)z)$, i.e. when the herding term is dominating with respect to the granular term in the diffusion. The previous inequality is equivalent to constrain the interval of variability of z into the interval

$$\frac{\varepsilon_1}{N} \ll z \ll 1 - \frac{\varepsilon_2}{N}. \quad (4.30)$$

However, given the inverse dependence of these constrains on the size of the system N , the ‘critical’ region might be tuned to any arbitrarily small interval.

4.4 Remarkable properties of the transition rates

The simple herding mechanism, based on the transition rates (4.4), possesses some remarkable properties deeply rooted in the mathematical formalization of the transition probabilities. Before characterizing in great details the statistical properties of our system in the following paragraphs, we want to shed some light on the connection between these statistical properties and the mathematical ingredients necessary to generate them, with a special focus on their robustness with respect to changes in the formalization of eq. (4.4).

4.4.1 Local and non-local interactions

In many cases, the inherent fluctuations of a discrete stochastic system attenuate as the number of its constituents is progressively increased. This property has been

extensively applied in physics and chemistry to justify the use of the deterministic approach based on differential equations for the explanation of many natural phenomena. A more refined approach relies on the application of the theory of stochastic processes, which is a natural candidate for the description of discrete environments. The relation between these two approaches has been exhaustively studied in the recent past (see for instance the theoretical work of Van Kampen [1992] and the illuminating work of Nicolis and Prigogine [1977]). The two approaches might converge to an equivalent description for large systems (i.e. composed of a large number of constituents), in the sense that the probability distribution becomes sharply peaked around the macroscopic trajectory. In order to end up with this remarkable equivalence, the transition rates $W(Y|X)$ of the underlying Markovian process, that characterized changes among the states of the system, have to be extensive¹⁰, i.e.

$$W(Y|X) = Nw(y|x) , \quad x = X/N , \quad y = Y/N . \quad (4.31)$$

The capital letters indicate quantities enumerated in terms of constituents, and the small letters denote concentrations.

Extensive transition rates describe systems whose evolution is driven by local mechanisms. In order to justify this assertion, namely the relation between the extensive transition probabilities with the local character of the interactions, we will provide some qualitative arguments. Let us start considering a system composed by a large number of independent constituents. To describe the asymptotic behavior of such a system, we may advocate the CLT, expecting Gaussian fluctuations around the mean of some aggregate quantities. The ratio of the amplitude of the fluctuations relative to their mean, then, will converge to zero; more precisely, we expect a dependence of the type $O\left(\frac{1}{\sqrt{N}}\right)$. The convergence to a well defined mean with vanishing fluctuations is the basic determinant for the equivalence between the stochastic and deterministic approach for the description of models composed by independent ‘particles’.

This scenario can be generalized to include some sort of interactions among the constituents, which typically generate correlations among the different particles¹¹. If the correlations are ‘weak’ enough, we can still apply the CLT and arrive to the Gaussian regime for the fluctuations like for independent particles, with well defined mean for aggregate quantities, that can still be describe by deterministic laws. Here the term ‘weak’ stands for a limited number of constituents involved in the interaction mechanism. The term ‘weak’ can be interpreted in this context as ‘local’, since the range of the interactions is limited to few neighbor constituents. On the contrary, the mechanism has a non-local or, in extreme cases, global character if the range of the correlations among the elements extends over macroscopic distances, i.e. involves a macroscopic fraction of elements. The long-range correlations prevent the

¹⁰This point has been made clear by Mansour and de Palma [1984] by means of the stochastic process described in this chapter.

¹¹For particular strongly interacting systems, a simple change of coordinates can decouple the elements into a new equivalent system composed of non-interacting components -for instance the phonons gas in a cold solid [Chandler, 1987]

use of the CLT and, therefore, we cannot expect the asymptotic Gaussian regime for the fluctuations. Spacial or temporal correlations in the level of fluctuations might be observed for any system size. It should be stressed here that extensive transition rates guarantee convergence to the Gaussian regime, while non-extensive systems *might* or *might not* exhibit convergence to the Gaussian regime. Finally, note that the mathematical formalization of the transition rates is not by itself a clear indication of the local or non-local nature of the interactions. This classification is based on the interpretation of the different asymptotic behavior of the system for a large number of constituents.

Turning now to the transition rates (4.4), it is straightforward to check that the extensivity property is not satisfied. It would be correct just in the case of an inverse dependence of b on the system size, as $b \sim O(1/N)$. This particular dependence is embedded in the Kirman's version of the herding model (see eq. (4.1), where $b = (1 - \delta)/N$). In his version, the interaction among ants is governed by pairwise, and, by its very definition, local meeting of two ants in their random searching for food, together with a stochastic self-conversion recruitment mechanism, based on exchange of pherormons. It is easy to show that, given the inverse dependence of b on the total number of ants N , the diffusion function is proportional to $1/N$ —see eq. (4.28)— which leads to vanishing fluctuations in the limit of large system size with the predicted inverse dependence $O\left(\frac{1}{\sqrt{N}}\right)$ (thermodynamic limit).¹²

Conversely, in our setting we have assumed that all the parameters remain bounded for increasing values of N , i.e. $b, a_1, a_2 \sim O(1)$. The non-extensivity of the transition rates (4.4) can be correlated to the presence of non-local interactions among the agents. What we observe is, in fact, non-vanishing fluctuations when we enlarge the system size (see next paragraph), which is in contradiction to the assumption of local interactions among the particles. A natural interpretation of the independence of the fluctuations on N is to assume a *global coupling* among all the agents, which explains the persistence of the fluctuations when enlarging the system size. It means that every agent in one group somehow *feels* the influence of *all* agents in the other group. In our setting, therefore, the interactions are not limited to a random meeting of two individuals, but, on the contrary, a global coupling of the agents is considered. The nature of this coupling is not modeled in detail in the present version of the model, but it is simply assumed by default in the formalization of the transition probabilities (4.4). This leaves open the question whether this global interaction exists and which is the underlying mechanism responsible for this non-local coupling. Once again, we interpret the non-trivial behavior in the thermodynamic limit as due to a pervasive interaction mechanism which involves all the constituents independent of their number. Once again, this interpretation is not directly derived from eq. (4.4) or a detailed interaction mechanism, but rather from the asymptotic behavior given by the Fokker-Plank equation (4.26).

¹²The thermodynamic limit simply refers to the limit for very large number of constituents of the system. Note that in physics “large” means of the order of the Avogadro number (10^{23}), which obviously is not a feasible number if we confine ourselves within a sociological environment.

Finally, it is very interesting to connect the presence of local interactions, with the formal requirement of extensive transition probabilities and the consequence of vanishing fluctuations for large system size. In many papers in the literature of economic agent-based models (cf. [Egenter et al., 1999, Lux and Schornstein, 2005, Challet and Marsili, 2004]), it has been observed that the ‘interesting’ properties of the fluctuations (namely non-Gaussianity and long range dependence) progressively disappear when the system size is enlarged. Typically, in such models is not straightforward to identify the extensive or non-extensive character of the transition probabilities among the different categories of agents. The line of our arguments might interpret this outcome as a natural consequence of ultimately underlying local interactions among the agents.

4.4.2 Independence on the total number of individuals

The derivation of the FPE, as detailed in the previous paragraph, might help to comprehend the origin of the anomalous behavior of the process (4.4) in the thermodynamic limit. Note that the diffusion function is independent of the number N of individuals, if we neglect the first term in eq. (4.28).¹³ This very important feature of the model is a consequence of two crucial assumptions: the symmetric and parabolic functional form of the herding components.

A deeper look at the transition rates from eqs. (4.16) and (4.17) shows that the quadratic terms, which encapsulate the herding tendency, are identical, therefore they cancel out when we compute the drift —see eq. (4.27). It means that, at first order, the net flow of probability due to the herding component is zero. The constant in front of the second term in the transition rates must be, then, identical. A slight difference of the values of b in the two terms —e.g. let us define b_1 and b_2 with $b_1 \neq b_2$ — would change the behavior of the system dramatically. In this case, all the individuals will migrate to one of the two states, depending on the sign of the difference $b_1 - b_2$, without any further conversion to the other state. The assumption $b_1 = b_2$ is equivalent to assuming that the global coupling among the individuals has identical strength, independently of the state the individuals belong to. This assumption can be justified assuming that the hidden mechanism responsible for the interaction among the constituents, not specified here, is homogeneous over the individuals.

Moreover, the diffusion function is proportional to the sum of the two rates, as we can notice from eq. (4.28). It corresponds to the second order term in the expansion of the Master equation, and, therefore brings a dependence on $1/N^2$. However, this factor cancels out *exactly* with the dependence of the herding component on N^2 , due to overall quadratic herding term, which is ultimately related to the linear assumption of the influence of the raw number of individuals in the other state. The diffusion term, then, remains always finite, and the fluctuations persist even for a very large system size. The collective nature of the interactions, mathematically formalized in non-extensive transition rates, its homogeneity among agents, given

¹³As already outlined, this requirement gives the further conditions $\frac{\epsilon_1}{N} \ll z \ll 1 - \frac{\epsilon_2}{N}$.

by the condition $b_1 = b_2$, and the quadratic functional form of the herding term are ultimately responsible for non-vanishing fluctuations in the thermodynamic limit.

Which is the consequence of this important property of the system? Since the Fokker-Plank equation is virtually independent on the system size, the only requirement is that N is large enough: *N turns out to be a free parameter of the model.* We, therefore, might describe systems with very different sizes within our framework.

4.4.3 Mean-reverting nature of the process

The symmetric structure of the herding term has a further implication on the dynamics of the process (4.4). It turns out to be a linear-mean reverting process because the interacting herding components in the computation of the drift cancel out exactly. The drift term is, then, dominated by the independent and autonomous switches between the two strategies, which tends to make the system converging to its unconditional mean. The difference in the values of a_1 and a_2 is, then, equivalent to introducing an external influence in the decision of the agents. In other words, if we are in the case $a_1 > a_2$, agents prefer to switch from the state 1 to the state 2. However, we do not provide any theoretical explanation, in terms of preferences or incentives, in order to justify the underlying difference among the two states. We rather assume it as given.

4.5 Equilibrium distribution

In the present paragraph we focus on the derivation of the equilibrium distribution $p_e(z)$ (called also unconditional or stationary distribution) of the process, define as the asymptotic probability density of z for $t \rightarrow \infty$. We have already derived the equilibrium distribution $P_{e,N}(n)$ in the finitary approach in paragraph 4.2. Here we will compute the equilibrium distribution for the continuous variable z in the framework of FPE.

Let us firstly define the probability current associated with the eq. (4.26):

$$J(z, t) = A(z)p(z, t) - \frac{1}{2} \frac{\partial}{\partial z} D(z)p(z, t). \quad (4.32)$$

which allow us to express the FPE as a continuity equation (see Van Kampen [1992] pp. 193) for the flow of probability:

$$\frac{\partial}{\partial t} p(z, t) + \frac{\partial}{\partial z} J(z, t) = 0. \quad (4.33)$$

To compute the equilibrium distribution $p_e(z, N)$, we use the standard formula:

$$p_e(z, N) = \frac{K_N}{D_N(z)} \exp \left(\int^z \frac{2A(y)}{D_N(y)} dy \right), \quad (4.34)$$

obtained with the condition $\frac{\partial}{\partial t} p(z, t) = 0$ and vanishing current at the boundaries,

$J(0) = J(1) = 0$. We refer to this constraint as *reflecting boundary conditions*.¹⁴ The dependence on N of the equilibrium distribution is brought about by the term proportional to $1/N$ in the diffusion function (4.28), that we denoted as $D_N(z)$, which brings information on the granularity of the system. Writing $A(y)$ as

$$A(y) = b \left[\varepsilon_1 \left(1 - y + \frac{\varepsilon_2}{2N} \right) - \varepsilon_2 \left(y + \frac{\varepsilon_1}{2N} \right) \right],$$

the integral in the argument of the exponential is straightforward:

$$\int^z \frac{2A(y)}{D_N(y)} dy = \varepsilon_1 \ln \left(z + \frac{\varepsilon_1}{2N} \right) + \varepsilon_2 \ln \left(1 - z + \frac{\varepsilon_2}{2N} \right). \quad (4.35)$$

Inserting this integral into eq. (4.34), we arrive at the following formula for $p_e(z, N)$:

$$p_e(z, N) = K_N(\varepsilon_1, \varepsilon_2) \left(z + \frac{\varepsilon_1}{2N} \right)^{\varepsilon_1 - 1} \cdot \left(1 - z + \frac{\varepsilon_2}{2N} \right)^{\varepsilon_2 - 1}. \quad (4.36)$$

The normalization constant $K_N(\cdot, \cdot)$ follows from $\int p_e(z, N) dz = 1$. If we neglect the $1/N$ term in D_N , we can express K_N as the Beta function¹⁵:

$$B(\varepsilon_1, \varepsilon_2) = \int_0^1 z^{\varepsilon_1 - 1} (1 - z)^{\varepsilon_2 - 1} dz = \frac{\Gamma(\varepsilon_1)\Gamma(\varepsilon_2)}{\Gamma(\varepsilon_1 + \varepsilon_2)}, \quad (4.40)$$

which leads to the equilibrium distribution:

$$p_e(z) = \frac{1}{B(\varepsilon_1, \varepsilon_2)} z^{\varepsilon_1 - 1} \cdot (1 - z)^{\varepsilon_2 - 1}. \quad (4.41)$$

The distribution (4.41) is known in the probability literature as the Beta distribution, which is one of the most flexible distributions in a bounded domain. Figure 4.2 shows different types of equilibrium distribution functions embedded in the formula (4.41). It is a remarkable property of the model that it possesses such extreme

¹⁴The FPE is a partial differential equation of order one in the time variable t and of order two in the 'spatial' variable z . In order to integrate eq. (4.26) we need one initial condition for the time variable, given by the initial probability distribution $p_0(z)$, and two further conditions for z . The reflecting boundary conditions are a possibility to provide these constraints. Moreover, they guarantee the conservation of probability over time (see Van Kampen [1992]). Other possible conditions are absorbing or periodic boundaries, that, however, are not analyzed in the thesis. Absorbing boundaries, for instance, do not conserve the overall probability, which, in fact, is 'absorbed' over time until vanishing.

¹⁵The Euler Beta function is defined for $a, b > 0$ as:

$$\text{Beta}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 t^{a-1} (1-t)^{b-1} dt. \quad (4.37)$$

The Gamma function $\Gamma(a)$ is defined as:

$$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt, \quad (4.38)$$

where $a > 0$. From this definition, one derives the following recursive formula:

$$\Gamma(a+1) = a\Gamma(a). \quad (4.39)$$

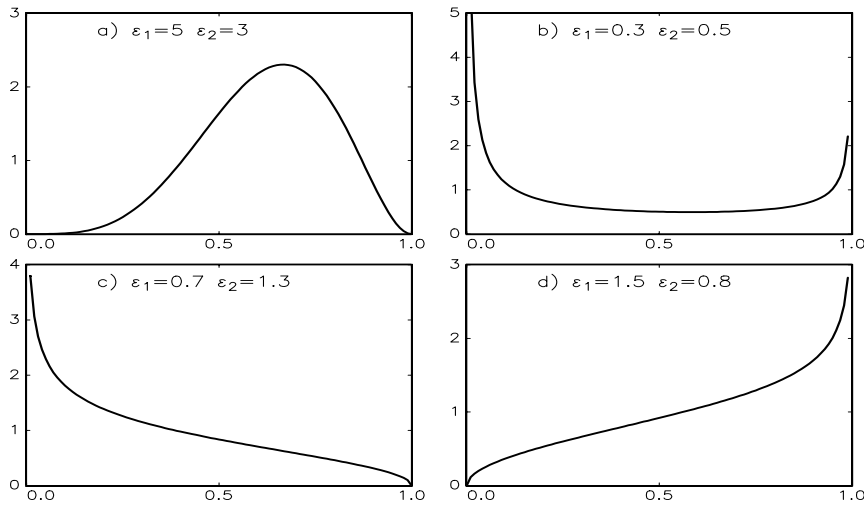


Figure 4.2: The four panels show different equilibrium probability densities derived from eq. (4.41) for several choices of the parameters ε_1 and ε_2 . In the panel (a) a uni-modal distribution, in (b) a bi-modal distribution and in panels (c) and (d) two cases of asymmetric monotonic distributions are shown.

flexibility of the resulting equilibrium distribution, despite the very few underlying parameters. Interestingly, the parameter a_1 , a_2 and b do not enter directly into the expressions (4.36) and (4.41), but only via the ratios $\varepsilon_1 = a_1/b$ and $\varepsilon_2 = a_2/b$. This is easy to understand if one thinks about the very definition of the equilibrium distribution; it does not involve any consideration on time development or dynamics, therefore the characterizing parameters and defining quantities have to be time-dimensionless.

In Figure 4.3, the equilibrium distribution (4.36) is exhibited for increasing values of N . Note their almost perfect agreement with the continuous approximation (4.41), as shown in the inset: all the curves collapse on $p_e(z)$. However, in the main panel, note that deviation among the curves $p_e(z, N)$ and the theoretical distribution at $N \rightarrow \infty$, $p_e(z)$, in the region close to the boundaries $z = 0$ and $z = 1$. This small discrepancy is due to the breakdown of the continuous approximation at these extremes. In Figure 4.4 we observe the convergence of eq. (4.12), based on a finite number of agents N , to its continuous approximation, given by eq. (4.36), for an increasing number of agents. Note that the pronounced deviation arises as soon as the continuous approximation does not hold i.e. in the region $\frac{\varepsilon_1}{N} \ll z$.

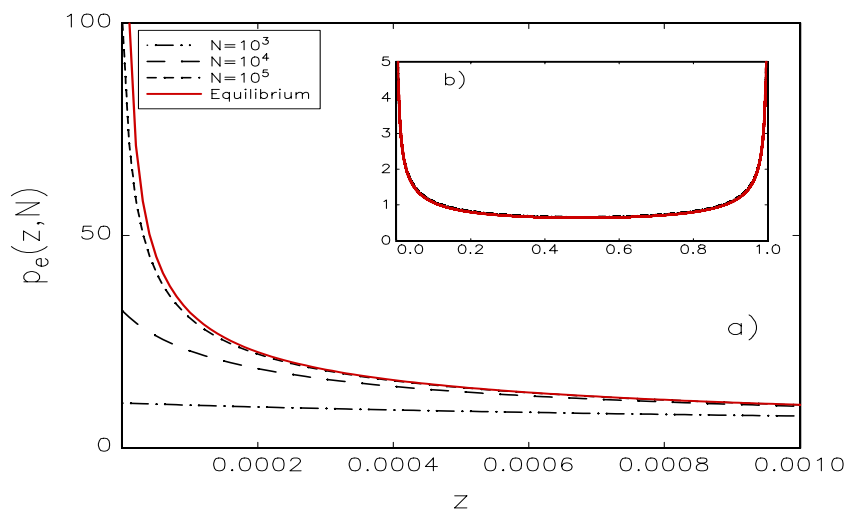


Figure 4.3: Panel (b) shows the almost perfect agreement between the equilibrium distribution from eq. (4.41) and the equilibrium distributions $p_e(z, N)$, in the case of different values of N . All the curves collapse on $p_e(z)$. Panel (a) shows a magnification of the entire distribution functions in the region close to the left boundary, $z = 0$. Note that for an increasing value of N , the distribution $p_e(z, N)$ gets closer to the curve $p_e(z)$. For any finite N , the densities $p_e(z, N)$ at zero are always finite, while $p_e(z)$ is unbounded. The chosen parameters are $\varepsilon_1 = \varepsilon_2 = 0.5$.

Moments of the equilibrium distribution

Given eqs. (4.39), (4.40) and (4.41), it is straightforward to compute the mean value of the distribution $p_e(z)$:

$$E[z] = \frac{\varepsilon_1}{\varepsilon_1 + \varepsilon_2}. \quad (4.42)$$

For higher moments, the following general formula holds:

$$E[z^k] = \frac{\Gamma(\varepsilon_1 + \varepsilon_2)}{\Gamma(\varepsilon_1 + \varepsilon_2 + k)} \frac{\Gamma(\varepsilon_1 + k)}{\Gamma(\varepsilon_1)}, \quad (4.43)$$

which can be derived using the definition of Beta function (4.40). It allows, then, to compute the variance of the distribution:

$$\text{Var}[z] = E[z^2] - E[z]^2 = \frac{\varepsilon_1 \varepsilon_2}{(\varepsilon_1 + \varepsilon_2 + 1)(\varepsilon_1 + \varepsilon_2)^2}. \quad (4.44)$$

4.6 Integration of the Fokker-Planck equation

In this paragraph we aim to integrate the FPE given by eq. (4.26), which possesses the unique equilibrium distribution $p_e(z)$ given by eq. (4.41), under reflecting boundary conditions. Let us denote by $p(z, t|p_0)$ the distribution of the variable z at time t , with an initial distribution $p_0(z, 0)$ at a time 0.

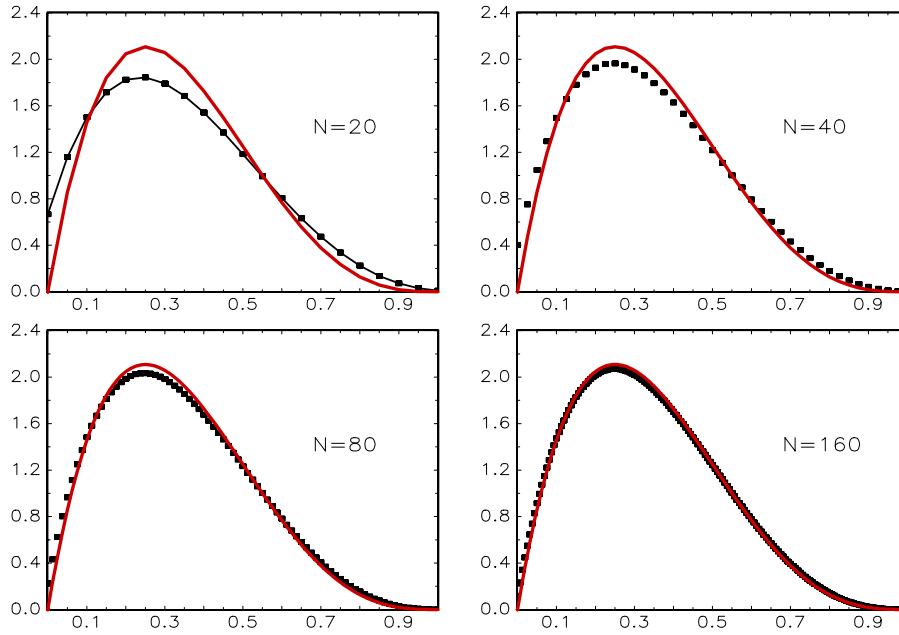


Figure 4.4: The four panels show the evolution of the equilibrium probability distribution $N \cdot p_{e,N}$ and the probability density $p_e(z)$ for increasing values of the number of agents N . Note that for $N = 160$, the agreement between the two curves seems to become satisfactory. The conditions for applicability of the continuous approximation, $\frac{\varepsilon_1}{N} \ll z \ll 1 - \frac{\varepsilon_2}{N}$, are, in fact, fulfilled by increasing the number of agents. The chosen parameters are $\varepsilon_1 = 2$ and $\varepsilon_2 = 4$.

4.6.1 Expansion in Jacobi polynomials

Given the stationarity of the underlying FPE, it is worthwhile to write the probability density $p(z, t)$ as an explicit function of the equilibrium distribution $p_e(z)$, as following:

$$p(z, t) = p_e(z)Q(z, t), \quad (4.45)$$

where $Q(z, t)$ is a function that accounts for the time dependence of the $p(z, t)$. In appendix A 4.1 it is shown that the function $Q(z, t)$ satisfies the *backward* or *adjoint* equation¹⁶:

$$\partial_t Q(z, t) = A(z)\partial_z Q(z, t) + \frac{1}{2}D(z)\partial_z^2 Q(z, t). \quad (4.46)$$

We posit $Q(z, t) = T(t)q(z)$ to try the separation of the variables t and z . Eq. (4.46) separates into:

$$\frac{d_t T(t)}{T(t)} = \frac{A(z)d_z q(z) + \frac{1}{2}D(z)d_z^2 q(z)}{q(z)} = -\lambda. \quad (4.47)$$

¹⁶In the following, for notational convenience, the derivative $\frac{d}{dz}$ is replaced by the symbol d_z and the symbol ∂_z reads as the partial derivative $\frac{\partial}{\partial z}$.

Eq. (4.47) is satisfied for every t and z only in the case that both, the left and right hand sides, are equal to an arbitrary constant, denoted as $-\lambda$. The time-dependent part reads now as:

$$T(t) = e^{-\lambda bt} . \quad (4.48)$$

It follows from eq. (4.47) that the function $q(z)$, satisfied the following equation:

$$z(1-z)\partial_z^2 q(z) + [\varepsilon_1 - (\varepsilon_1 + \varepsilon_2)z]\partial_z q(z) + \lambda q(z) = 0 , \quad (4.49)$$

called *hypergeometric differential equation* in the mathematical literature.¹⁷ As detailed in the appendix A 4.2, the reflecting boundary conditions of our original problem constrain λ to assume specific and discrete values, given by:

$$\lambda_n = n(n + \varepsilon_1 + \varepsilon_2 - 1) , \quad (4.50)$$

where n is an integer number. For this choice of λ_n , the general differential equation (4.49) becomes the so-called Jacobi differential equation:

$$z(1-z)\partial_z^2 q(z) + [\varepsilon_1 - (\varepsilon_1 + \varepsilon_2)z]\partial_z q(z) + \lambda_n q(z) = 0 , \quad (4.51)$$

whose general solutions are Jacobi polynomials, as detailed in appendix (A 4.2). They form a *complete set* of basis functions in the compact domain $[0, 1]$, which allows to express *every* other function¹⁸, in the same domain as a linear combination of them — see appendix A 4.3.

Finally, the solution of the FPE (4.26) with the reflecting boundary conditions assumes the form of a linear combination of Jacobi polynomials with time-dependent coefficients:

$$p(z, t|p_0) = p_e(z) \sum_{n=0}^{\infty} \frac{a_n}{h_n} P_n^{(\varepsilon_1, \varepsilon_2)}(z) \exp[-\lambda_n bt] , \quad (4.52)$$

where a_n are fixed by the initial distribution $p_0(z, 0)$. The constants h_n are defined in the appendix A 4.3.

4.6.2 Main result

To fully solve the integration problem of the FPE, we have to specify the coefficients a_n in the expansion (4.52) as a function of a given starting distribution $p_0(z, 0)$. We concentrate on two cases that are useful for the further development of the thesis.

The first chosen starting distribution is a Dirac's delta peaked at an interior value $z_0 \in (0, 1)$:

$$p_0(z, 0) = \delta(z - z_0) . \quad (4.53)$$

¹⁷Note that in the pertinent literature (Arfken [1985]) this type of equation might be parametrized differently than in eq. (4.49).

¹⁸This is true for continuous and squared integrable functions with respect to the metric $p_e(z)$ in the interval $(0, 1)$.

Using the relation (A 4.29), it is easy to derive that $a_n = P_n^{(\varepsilon_1, \varepsilon_2)}(z_0)$. The complete solution of the FPE with a peaked starting distribution in z_0 is then:

$$p(z, t|z_0, 0) = p_e(z) \sum_{n=0}^{\infty} \frac{1}{h_n} P_n^{(\varepsilon_1, \varepsilon_2)}(z_0) P_n^{(\varepsilon_1, \varepsilon_2)}(z) \exp[-\lambda_n bt]. \quad (4.54)$$

As a second relevant example, we assume that the initial value z_0 occurs with the equilibrium distribution $p_e(z_0)$. The complete solution is then given by the previous equation multiplied by an extra factor $p_e(z_0)$:

$$p(z, t|p_e(z_0)) = p_e(z) p_e(z_0) \sum_{n=0}^{\infty} \frac{1}{h_n} P_n^{(\varepsilon_1, \varepsilon_2)}(z_0) P_n^{(\varepsilon_1, \varepsilon_2)}(z) \exp[-\lambda_n bt]. \quad (4.55)$$

In Figures 4.5 and 4.6 the evolution of the probability density $p(z, t)$ according to the FPE (4.26) is shown, for two different sets of parameters. These two examples illustrate the great informational content in the FPE, which, in fact, fully characterizes the dynamics of the system.

4.6.3 Final comments

It should be emphasized that the FPE, and the associate dynamics of the probability density, is **deterministic**, although the underlying state of the system z follows a stochastic process (4.4). We renounce to give a fully deterministic description of the evolution of the variable z , as it should be if the underlying stochastic variable would follow a deterministic process governed by a differential equation. However, via the FPE, we can still formulate a deterministic law of motion for the probability density $p(z, t)$, given the starting distribution $p_0(z, 0)$.

Taking into account the stochastic nature of the dynamic of z , can we somehow introduce a meaningful quantity to characterize in a deterministic way the stochastic motion of z itself? A quite intuitive approach might be to follow the evolution of the mean of $p(z, t)$, denoted by \bar{z} .¹⁹ It means that we define a macroscopic quantity, the mean of the process, and we consider its evolution as a representative dynamics of the system. It is well known (see Gardiner [2003], Van Kampen [1992]) that the following equation holds for stochastic processes with linear drift functions:

$$\frac{d}{dt} \bar{z} = A(\bar{z}), \quad (4.56)$$

where $A(\cdot)$ is given by eq. (4.27). Equation (4.56) simply states that the mean of $p(z, t)$ follows a deterministic differential equation of first order, whose solution is given by:

$$\bar{z}_t = E[z] - (\bar{z}_0 - E[z]) e^{-(\varepsilon_1 + \varepsilon_2)bt}. \quad (4.57)$$

\bar{z}_0 is the mean value of the starting distribution $p_0(z, 0)$ and $E[z]$ is the unconditional mean of the process. As a result, we obtain that \bar{z} converges to its unconditional

¹⁹Alternative definitions might rely on the mode of the distribution, or other measures of the typical value

value $E[z]$ with an exponential rate. If we take into account the example of Figure 4.6, the macroscopic equation predicts no evolution for \bar{z} , since the starting value \bar{z}_0 coincides with the unconditional value of the process. On the contrary, we observe a non trivial dynamics of $p(z, t)$. The contrast between the macroscopic approximation given by eq. (4.56) and the more complete description in terms of the FPE is evident. The endogenous fluctuations of the system can be neglected only at a cost of a meaningless description of its time-evolution. This example gives a clear indication of the ineffectiveness of the ‘small noise’ approach (Gardiner [2003]) for our model. It consists in the separation of the dynamics of the system in a deterministic evolution of the mean and a ‘small’ *superimposed* noise perturbation. We have seen that, for non-extensive systems, this approach might lead to inconsistent results (see paragraph 4.4).

4.7 Autocorrelation

In addition to the unconditional moments, the knowledge of the FPE allows to compute all the conditional moments such as the autocorrelation function.

Let us start with the definition of the covariance:

$$E[z(t)z(0)] = \int_0^1 \int_0^1 dz dz_0 z p(z, t|z_0) z_0 p_0(z_0, 0), \quad (4.58)$$

where the starting value z_0 is averaged over the equilibrium distribution $p_e(z_0)$ and $p(z, t|z_0)$ is given by eq. (4.53). By means of the expansion in Jacobi polynomials, we can express the covariance in a very compact and elegant form (see [Gardiner, 2003] pp. 131):

$$E[z(t)z(0)] = \sum_{n=0}^{\infty} \frac{1}{h_n} e^{-\lambda_n bt} \left[\int_0^1 dz z P_n(z) p_e(z) \right]^2. \quad (4.59)$$

Using the relations (A 4.26) and (A 4.27), we express the variable z as a function of the first two Jacobi polynomials:

$$z = \frac{1}{\varepsilon_1 + \varepsilon_2} P_1(z) + E[z] P_0(z). \quad (4.60)$$

If we plug eq. (4.60) into eq. (4.59) and applying the orthogonality property, it is straightforward to express the previous formula as:

$$E[z(t)z(0)] = E[z]^2 + \text{Var}[z] e^{-(\varepsilon_1 + \varepsilon_2)bt}, \quad (4.61)$$

where $\text{Var}[z]$ is given by eq. (4.44). The autocorrelation is defined as:

$$C_z(t) = \frac{E[z(t)z(0)] - E[z]^2}{E[z^2] - E[z]^2}, \quad (4.62)$$

which turns out to be a pure exponential:

$$C_z(t) = e^{-b(\varepsilon_1 + \varepsilon_2)t}. \quad (4.63)$$

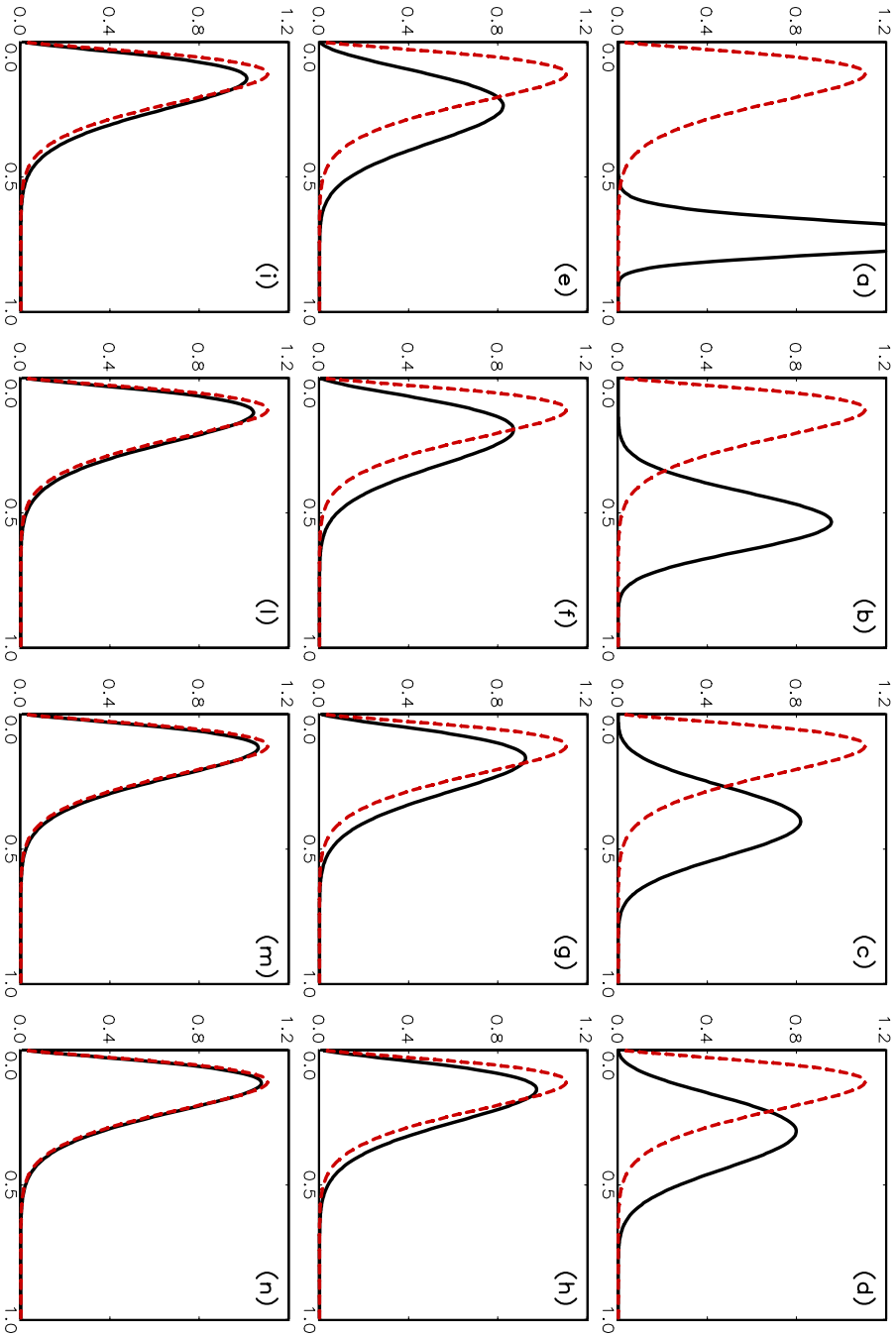


Figure 4.5: The sequence of figures shows the evolution of the probability density $p(z, t)$ (continuous line) governed by the FPE form eq. (4.26). The starting distribution is peaked around the starting value 0.7, shown in panel (a). Note the convergence of $p(z, t)$ towards the equilibrium distribution (dashed line), e.g. in panel (n). The underlying parameters are $\varepsilon_1 = 2.5$, $\varepsilon_2 = 12$ and $b = 0.01$. The expansion in Jacobi polynomials counts 50 terms.

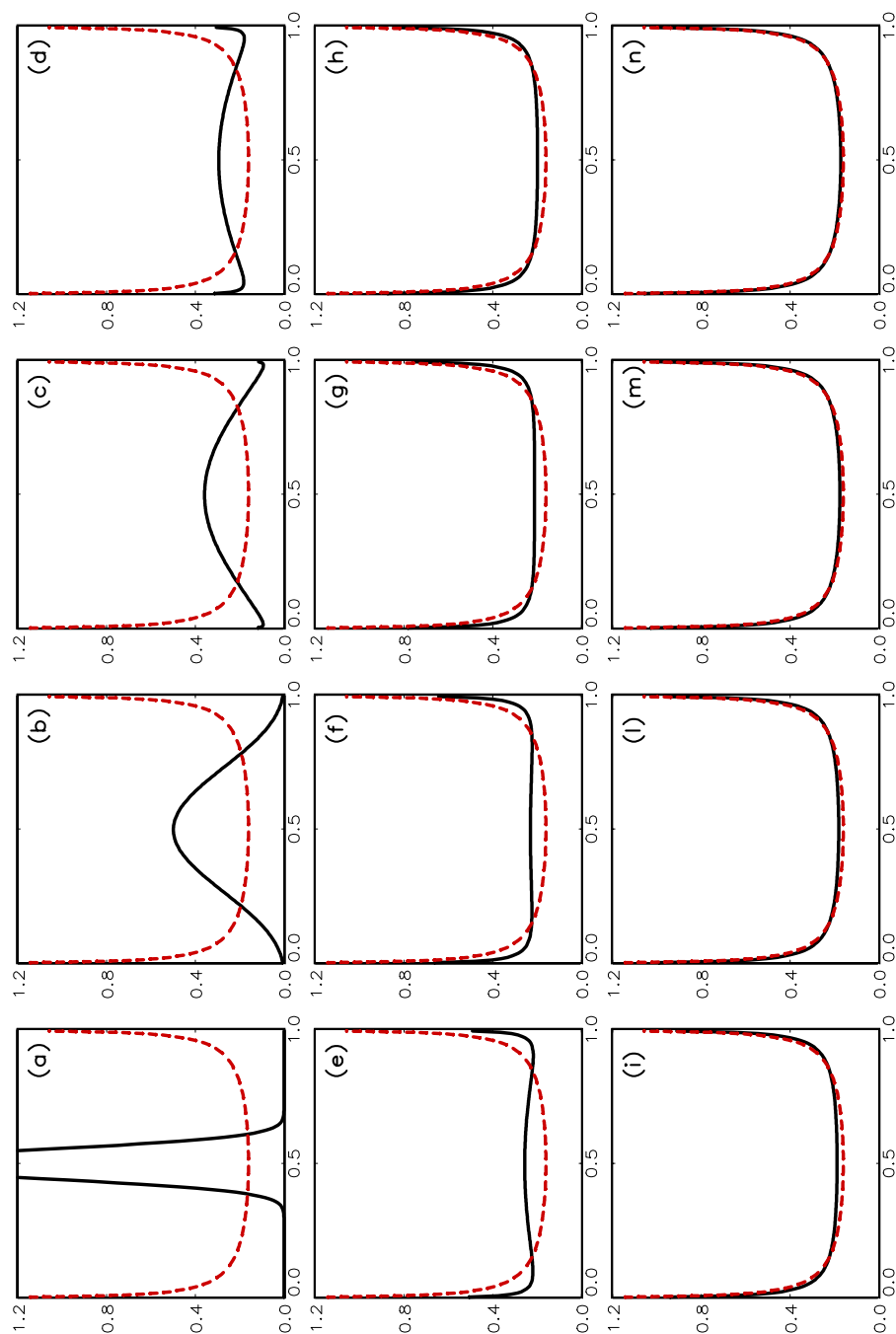


Figure 4.6: In this second example, the convergence of $p(z, t)$ (continuous line) is towards a bi-modal equilibrium distribution (dashed line). The underlying parameters are $\varepsilon_1 = \varepsilon_2 = 0.5$ and $b = 0.01$. The expansion in Jacobi polynomials counts 50 terms.

The decay rate is given by the first eigenvalue λ_1 multiplied by the herding parameter b . This result is not confined solely to our model. On the contrary, it is a general property of linear Markovian processes, i.e. those processes whose mean follows a deterministic linear equation, like eq. (4.56) [Gardiner, 2003].

The auto-covariance of a function $f(z)$ of the original stochastic variable z is given by the more general expression:

$$E[f(z_t)f(z_0)] = \sum_{n=0}^{\infty} \frac{1}{h_n} e^{-\lambda_n b t} \left[\int_0^1 dz f(z) P_n(z) p_e(z) \right]^2. \quad (4.64)$$

From the previous formula, we can compute in close form the autocorrelation function if $f(\cdot)$ is a polynomial with finite degrees of freedom. In general, it is possible to express in closed form all the functions that are expressed as linear combination of a finite number of Jacobi polynomials. If $f(\cdot)$ does not exhibit this property, we might not find a close form solution for the autocorrelations. We have then to rely on numerical computation of a truncated version of the series (4.64).

Two points are worth mentioning here. The asymptotic form of the autocorrelation is exponential. In fact, if the integral in eq. (4.64) increases at most as a polynomial on n , for t larger than $1/(a_1 + a_2)$, the first exponential dominates, and the asymptotic form of the auto-covariance is:

$$E[f(z_T)f(z_0)] \sim e^{-b(\varepsilon_1 + \varepsilon_2)t} \quad \text{for } T > \frac{1}{(a_1 + a_2)}. \quad (4.65)$$

This asymptotic form does not come as a surprise since we know that Markovian processes cannot be long memory processes *strictu sensu*. The time $1/(a_1 + a_2)$ plays the important role of a natural time scale of the system, over which asymptotic theory can be applied. However, if we consider time lags smaller than $1/(a_1 + a_2)$, we might find deviations from the asymptotic exponential decay, i.e. a more pronounced curvature than the exponential decay. The reason is that many time scales from eq. (4.64) are mixed together, which might even give the impression of a hyperbolic decay.

4.8 Special cases: the cosine of a random walk

It is well-known that the Jacobi polynomials are a very general class of orthogonal polynomials. Many classical polynomials can be obtained as special cases of them for suitable choices of the underlying parameters. In our framework, which are these sets of parameters ?

The first restriction to the underlying parameters is to posit $\varepsilon_1 = \varepsilon_2 = \varepsilon$. The resulting polynomials are the *Gegenbauer* polynomials. We call this case the symmetric set, to distinguish it from the general asymmetric conditions for $\varepsilon_1 \neq \varepsilon_2$.

In the symmetric case, among all the possible values of ε , we distinguish two particular values: $\varepsilon = 1$ and $\varepsilon = \frac{1}{2}$. In the first case we obtain the *Legendre* polynomials

and the resulting unconditional distribution is uniform – see eq. (4.41).

The second case is particularly elegant. We recover the *Chebyshev* polynomials with a resulting symmetric bimodal equilibrium distribution. Interestingly, for $\varepsilon = \frac{1}{2}$ the integration of the FPE is greatly simplified. The FPE reads now as:

$$\frac{\partial}{\partial t} p_z = -\frac{\partial}{\partial z} \left(\frac{1}{2} - z \right) p_z + \frac{1}{2} \frac{\partial^2}{\partial z^2} 2b z(1-z) p_z . \quad (4.66)$$

With the new variable $z = \frac{1}{2}(1 - \cos \theta)$, the previous FPE takes the simple form:

$$\frac{\partial}{\partial t} p_\theta = b \frac{\partial^2}{\partial \theta^2} p_\theta . \quad (4.67)$$

The associated current is:

$$j(\theta, t) = \frac{\partial}{\partial \theta} p(\theta, t) , \quad (4.68)$$

which has to vanish at the borders to fulfil the reflecting boundary conditions. The resulting dynamics of θ is a pure diffusion (no drift and constant diffusion function) in the bounded interval $[0, \pi]$. Basically for $\varepsilon = 0.5$, the dynamics of z is given by the projection on the diameter of a point that moves as a random walk back and forth in the positive semi-circle.

4.9 Langevin equation and the micro-macro structure

We introduced in paragraph 4.3 the Fokker-Planck equation as the approximation to the continuum of the discrete stochastic process (4.4). The expansion in series of Jacobi polynomials is an exact solution of the integration problem of the FPE. In the case of a very small value of the parameter b , which will be the case for application to financial data, the truncation of this series, however, turns out to be a poor approximation if it involves only few elements (two or three), and it obviously becomes non-manageable if the number of terms is increased²⁰. In this paragraph, we intend to introduce an alternative approximation scheme of the stochastic process (4.4), which comes up with another important type of equation called Langevin equation in the pertinent literature. The basic idea of this simplified approach is to find an appropriate time interval, that we call *macro-interval*, for which the conditional distribution of the discrete variable z is well approximated by a Gaussian. In the following we sketch a proof of the Langevin equation to approximate the discrete process (4.4) based on some heuristic arguments. Let us start with the identity:

$$z_{t+\Delta t} = z_t + \sum_{i=1}^M \eta_{t+i\Delta t_0} , \quad (4.69)$$

where $\Delta t = M\Delta t_0$ and $\eta(t + i\Delta t_0)$ might take the values in the finite set $S \equiv \{0, 1/N, -1/N\}$; Δt_0 is the elementary time interval for a possible change of strategy

²⁰There are numerical problems when the terms of the series are increased, since the series does not converge uniformly to the function $p(z, t)$. A small number of terms are not sufficient for an accurate description of $p(z, t)$ when $b \cdot t \leq 1$.

of one agent. The previous decomposition is based on a separation of the dynamics of z into two different time scales: a micro time scale Δt_0 and a macro time scale Δt . During the micro interval Δt_0 , the variation of z is constrained to the numerable set S . If we concentrate on aggregate variation Δz during the macro-time interval Δt , we sum up over many of these elementary increments, loosing the information on the fine structure of the dynamics of z , i.e. the information on the movements of every single individual. Our aim is, however, to give up this too detailed perspective for a simpler and, at the same time, meaningful description of the system. In order to do so, we have to find a proper time scale, long enough to aggregate the tiny details, and not too long to lose relevant information. This perspective is called in the pertinent literature mesoscopic approach, a sort of middle way between the micro and macroscopic approach. We have, then to find the proper value for the aggregation level M . The random variable $\eta(t + (i + 1)\Delta t_0)$ can take just three values with the following probabilities:

$$\eta_{i+1} = \begin{cases} +\frac{1}{N} & p_i \Delta t_0 = (1 - z_i) \left(\frac{\varepsilon_1}{N} + z_i \right) b N^2 \Delta t_0, \\ 0 & 1 - (p_i + q_i) \Delta t_0, \\ -\frac{1}{N} & q_i \Delta t_0 = z_i \left(\frac{\varepsilon_2}{N} + 1 - z_i \right) b N^2 \Delta t_0, \end{cases} \quad (4.70)$$

where we explicitly take into account the dependence on the previous value of z (here $z_i = z(t + i\Delta t_0)$ and $\eta_{i+1} = \eta(t + (i + 1)\Delta t_0)$). For notational convenience, we label the time step with i . The mean μ_{i+1} and variance σ_{i+1}^2 of η_{i+1} are given respectively by:

$$\mu_{i+1} = \frac{\Delta t_0}{N} (p_i - q_i), \quad (4.71)$$

$$\sigma_{i+1}^2 = E[\eta_{i+1}^2] - \mu_{i+1}^2 = \frac{\Delta t_0}{N^2} (p_i + q_i) + o\left(\frac{\Delta t_0}{N^2}\right). \quad (4.72)$$

Following the Central Limit Theorem (CLT), a sum of M independent random variables, drawn from a common distribution, with mean μ and finite variance σ^2 , converges to a Normal with mean $M\mu$ and variance $M\sigma^2$. We aim to approximate the sum in (4.69) by a normally distributed random variable. The requirement of finite variance is surely fulfilled by the distribution of the variable η , however, the iid-ness assumption does not strictly hold. The probabilities $p\Delta t_0$ and $q\Delta t_0$ depend on z , which dynamically changes with η itself. We have to impose a further condition on z : it should not vary ‘too much’ during the time interval Δt , in such a way that it can be treated as a constant. Under this approximation, we recover the condition of identically distributed variables. Finally, M and, consequentially, Δt should be large enough to assure the convergence towards the Gaussian, but not too large to prevent the approximation of constant z .

We approximate the sum in eq. (4.69) with a Gaussian distribution with mean $M\mu$ and standard deviation $\sqrt{M}\sigma$. Therefore we end up with:

$$z_{t+\Delta t} = z_t + (\varepsilon_1 - (\varepsilon_1 + \varepsilon_2) z_t) M b \Delta t_0 + \sqrt{2z_t(1 - z_t) M b \Delta t_0} \cdot \xi_t, \quad (4.73)$$

where ξ_t is a Normally distributed random variable. The inequality $M b \Delta t_0 \ll 1$ should be satisfied in order to guarantee small deviations of the variable z and,

therefore, to preserve the approximation of constant z during the interval Δt . Recalling the inequality (4.5), we come up with the following scaling relation between the number of micro steps M and the number of agents N :

$$M = k \frac{N^2}{2}, \quad (4.74)$$

with an arbitrary small number k . Finally, the Langevin approximation for the stochastic process eq. (4.69) is given by:

$$\Delta z = (\varepsilon_1 - (\varepsilon_1 + \varepsilon_2)z_t)b\Delta t + \sqrt{2b\Delta t z_t(1-z_t)} \cdot \xi_t, \quad (4.75)$$

where $\Delta t = M\Delta t_0$. The limit of such an approximation are related to the assumption of the CLT, therefore to the goodness of the Gaussian approximation for the sum in eq. (4.69). Near the edges, $z = 0$ or $z = 1$, the variable z can not be considered constant, and the assumption of identically distributed variables does not hold any longer²¹. Additionally, eq. (4.75) does not incorporate neither the conditions of boundedness of z to the compact interval $[0, 1]$, nor the natural boundary conditions of the eqs. (4.1) and (4.4). The boundary conditions, then, should be put in the equation (4.75) ‘by hand’. They are given by:

$$\text{if } z(t) > 1 \quad \text{then} \quad \frac{z(t + \Delta t) + z(t)}{2} = 1, \quad (4.76)$$

$$\text{if } z(t) < 0 \quad \text{then} \quad \frac{z(t + \Delta t) + z(t)}{2} = 0. \quad (4.77)$$

A glance to eqs. (4.76) and (4.77) shows that they are equivalent to a reflection around the edges of the domain of z , $z = 1$ and $z = 0$, respectively.²²

4.9.1 Autocorrelation revisited

The ACF of z can be calculated by means of the LE using a recursive method. Let us start with the auto-covariance:

$$E[z_{t+\Delta t}z_0] = E[z_tz_0] + E[A(z_t)z_0], \quad (4.80)$$

where Δt is a macro-time interval. It is not surprising that the diffusion term vanishes since we know that the ACF is governed solely by the drift term. Introducing the notation $F_t = E[z_tz_0]$, eq. (4.80) reads:

$$F_{t+\Delta t} = (1 - b\Delta t(\varepsilon_1 + \varepsilon_2))F_t + \varepsilon_1E[z_0]. \quad (4.81)$$

which leads to the final equation²³:

$$E[z_tz_0] = (E[z^2] - E[z]^2) \exp(-bt(\varepsilon_1 + \varepsilon_2)) + E[z]^2. \quad (4.82)$$

²¹Note that also the FPE is not a valid description of the system near the borders.

²²Alternative choices might be possible. For example, the boundary conditions can also be implemented as following:

$$\text{if } z(t) > 1 \quad \text{then} \quad z(t + \Delta t) = 1, \quad (4.78)$$

$$\text{if } z(t) < 0 \quad \text{then} \quad z(t + \Delta t) = 0. \quad (4.79)$$

²³The exponential decay appears under the condition $b\Delta t \ll 1$.

Note that eq. (4.82) is identical to eq. (4.61). The two equations, namely the FPE and the LE, lead to the same ACF for z . The recursive method based on the Langevin equation is, however, not feasible for non-linear functions, except for a few cases (for instance for z^2). For arbitrary functions $f(z)$ we should rely on eq. (4.61).

4.10 Different avenues for simulating the herding model

The theoretical analysis of the Markov chain governed by eq. (4.4) can help to shed some light on the issue of simulating the herding model. We will detail here three different approaches, commonly used in the agent-based model literature.

Following multi-particle simulations in statistical physics, different avenues exist for simulating the herding model (4.4). The first, obvious choice would be a true microscopic simulation keeping track of the state of each individual agent and determining its switches over time by random number draws. Of course, the continuous-time framework would have to be simulated in discretised form. We should be careful about the restriction imposed on the time unit Δt_0 by the normalization of the probabilities, given by eq. (4.5). The maximum admissible time increment obviously decreases hyperbolically with the population size, $\Delta t_0 \propto N^{-2}$. Microscopic simulations, therefore, become increasingly more time consuming with larger population size.

As an alternative, we could resort to simulating the aggregated outcome of the stochastic dynamics in terms of the population configuration which is summarized by the variable n , without taking into account the information on the history of every single individual. It is convenient to choose Δt_0 in such a way that it allows the highest “efficiency” of the macroscopic simulations, i.e. such that it minimizes the probability to observe no change in n . This is equivalent to using (4.5) as an equality. Of course, only the smallest possible change in n can be observed during the micro-step Δt_0 of the simulation. In a similar way as for the Langevin equation (see section 4.9), we may introduce a distinction between micro-time steps Δt_0 and macro time increments Δt in which many increments of n may be observed. To illustrate the dynamics and to provide a justification of a “useful” macro time scale, consider the following scenario. For $n = 0$, the system can evolve like follows: it may remain unchanged with probability:

$$\frac{N + \varepsilon_2 + 0.25(\varepsilon_2 - \varepsilon_1)^2}{N + \varepsilon_1 + \varepsilon_2 + 0.25(\varepsilon_2 - \varepsilon_1)^2}, \quad (4.83)$$

which is close to 1, or it may change by one unit with the small probability:

$$\frac{\varepsilon_1}{N + \varepsilon_1 + \varepsilon_2 + 0.25(\varepsilon_2 - \varepsilon_1)^2}. \quad (4.84)$$

Therefore, the average number of iterations needed to observe a change of one agent is approximately equal to N . To observe larger increments of n (as changes by one unit are negligible, in particular for the case of large N), we need to multiply for

an additional factor kN , where k is an arbitrary number smaller than 1. From eq. (4.5), a sensible choice appears to be:

$$\Delta t = \frac{b}{2} N^2 \Delta t_0, \quad (4.85)$$

In the simulations, one then iterates the process $\frac{b}{2} N^2$ times until one stores the current value of n as one realization at the macroscopic time scale Δt . The most interesting aspect of this approach is that it guarantees *invariance* of the dynamics of the macroscopic variable n with respect to the number of agents due to the flexibility of the chosen macro time scale. This does not come as a surprise since we have already noticed that the drift and diffusion functions in the FPE (4.26) are independent on N . Figure 4.7 provides an illustration of this feature in which we indeed observe no qualitative difference in the behavior of time series for different sizes of the population.

All in all, in order to have a meaningful micro-simulation algorithm for the simulation of the Markov chain governed by eq. (4.4) in the case of various population sizes, we have to introduce two different time scales. Note that the scaling of the macro-time with the number of individuals is non-linear. Again, far to be mathematical curiosity, the non-linearity of the macro time-scaling reflects the non-extensivity of the transition probabilities (4.4).

One might, as an alternative, simulate the model using a Langevin equation providing a Gaussian approximation to the stochastic dynamics over $\Delta t / \Delta t_0$ micro time steps per time unit Δt . Although for small step sizes Δt , the Langevin equation indeed provides a close approximation to the underlying agent-based model, it has the drawback that it might violate the built-in boundaries $y \in [0, 1]$ of the population dynamics. Despite these drawback, (4.75) has the important advantage that it facilitates numerical simulations. In Figure 4.7 we can observe the qualitatively good agreement between the micro-simulation based on eq. (4.4) and the Langevin approach.

Remark. It is interesting to notice that if one starts with the FPE and the LE, the underlying discrete stochastic process (4.4) cannot be completely identified. The approximations given by (4.26) and (4.75) hold for a ‘large’ number of agents, N , without further specification on its order of magnitude. It implies that, in order to recover the discrete process (4.4), one might exogenously impose an arbitrary number N of agents, with the only requirement to be large, with the further condition $\frac{\varepsilon_1}{N} \ll z \ll 1 - \frac{\varepsilon_2}{N}$. In other words, the number of agents is a free parameter of the model. Anyway, if we concentrate on the continuous fraction z , the absolute number of agents is not a relevant quantity. To illustrate the implication of the vanishing N-dependence in the LE, Figure 4.7 shows the dynamics of the fraction z for different numbers N , ranging from 50 to $5 \cdot 10^4$, computed via the microsimulation approach (4.4) as compared to the simulation performed via the Langevin equation (4.75). The qualitative behavior of the dynamics is similar for all the six panels. The micro/macro algorithm for the simulation of the discrete stochastic process (4.4)

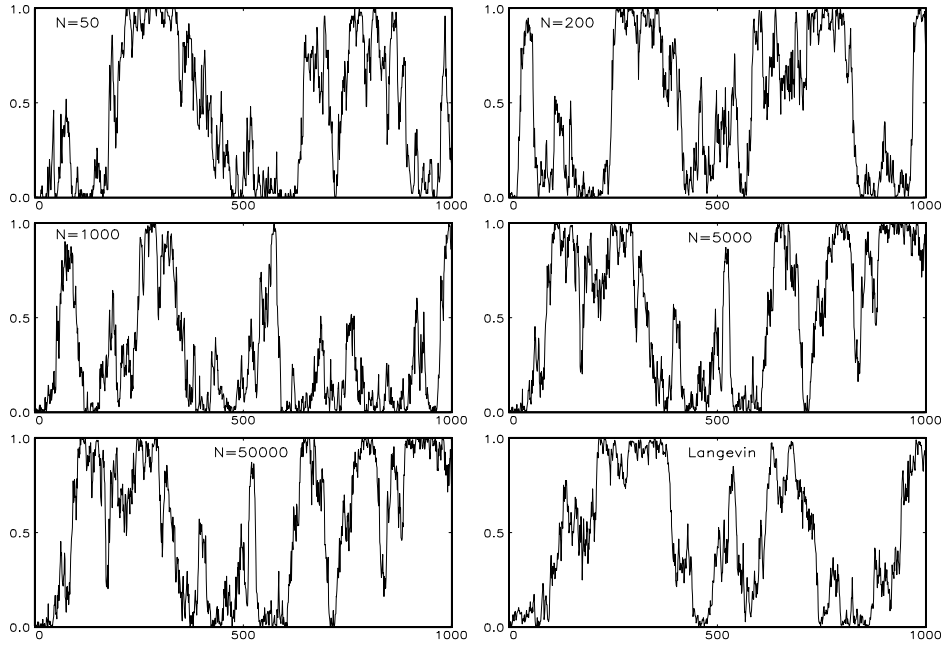


Figure 4.7: *The six panels show snapshots of the dynamics of the population index for increasing values of the number of agents N . The right bottom panel shows a trajectory computed via the Langevin equation.*

guarantees the invariance of the dynamics with respect to the parameter N . The LE, then, constitute an alternative approach to simulate the process, which is extremely more efficient in terms of computational-time.

4.11 The Kirman model and the emergence of macroscopic laws

Until now, we have analyzed the stochastic process (4.4) and its possible approximations. In this section, we want to turn back to the original process introduced by Kirman and governed by eq. (4.1) within our framework. The transitions (4.1) differ from the transitions (4.4) in their scaling with the size of the system. They are, in fact, extensive, according to the definition (4.31). This feature is responsible for the very different behavior observed in the limit $N \rightarrow \infty$. In order to show it, let us start with the following choice of parameters: the constant governing the autonomous switching being $a = \frac{k}{N_0}$, where N_0 is the starting reference population size and $k = 0.5$ and the probability of a conversion being equal to $1 - \delta$. A glance at the transition rates in eq. (4.1) shows that the effective herding parameter (the constant in front of the quadratic term) is $\frac{1-\delta}{N}$, for a market populated by N traders. Without loss of generality, we further assume that $\delta = 0$, which implies a probability of conversion equal to 1. Given the extensivity property, we can conveniently move to the intensive formalism in terms of the variable z . The ratio between the

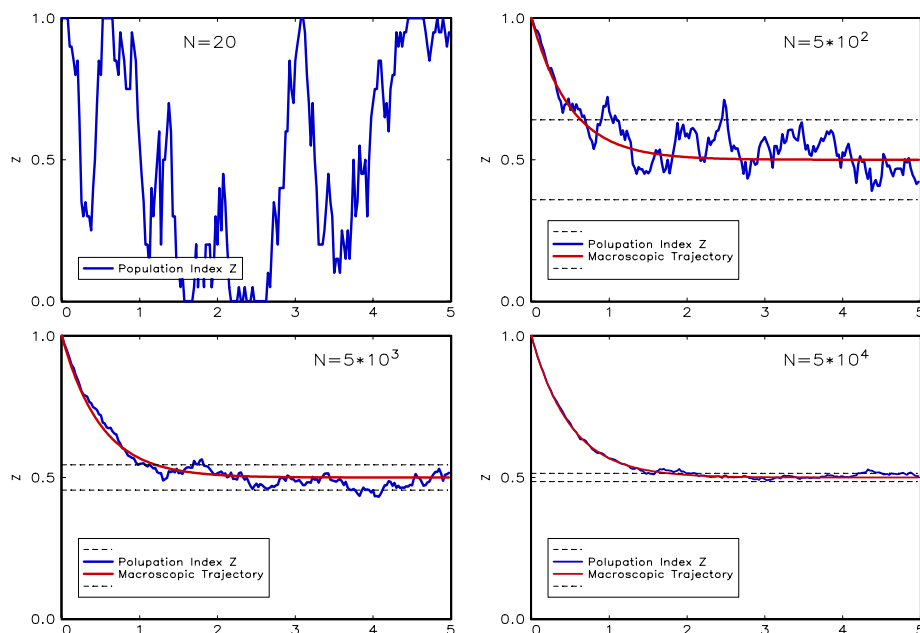


Figure 4.8: The four panels show the emergence of the macroscopic skeleton of the system when governed by the extensive transition probabilities (4.1), for an increasing number of agents. Note the clear exponential decay of the trajectory of z for large population size, well described by the deterministic eq. (4.57). The dashed lines represent the standard deviation of equilibrium distribution which the variable z converges to. The underlying parameters are $N = 20$, $\delta = 0$ and $a = 0.5/N_0$. The time is measured in natural unit of a .

autonomous term and the herding parameter is, then, defined by:

$$\varepsilon_N = \frac{N}{2N_0}. \quad (4.86)$$

The N -dependence in ε_N is the major difference with respect to the non-extensive formalization of the process. When the number of agents trading in the market N coincides with the reference size N_0 , the market falls in the herding dominating regime, with a bimodal probability density function of z -see eq. (4.41). If one now increases the number of agents over the starting level N_0 and, at the same time, leaves unchanged the other parameters, the equilibrium distribution will undergo a transition from a bimodal shape to a narrowly peaked distribution around the average value $E[z] = 0.5$. The linear dependence of ε on the number of agents N brings about a convergence of the equilibrium distribution $p_e(z)$ to a Gaussian.

In order to show it, let us express the equilibrium distribution (4.41) as a function of the new variable $x = 2z - 1$, which has the advantage to be symmetric around 0.

Let us further assume that $\varepsilon_N \gg 1$, which implies that $N \gg N_0$. Eq. (4.41) can be written in terms of the variable x as follows:

$$p_e(x) = \frac{\Gamma(2\varepsilon_N)}{\Gamma(\varepsilon_N)^2} \frac{1}{2^{2\varepsilon_N-1}} (1-x^2)^{\varepsilon_N-1}. \quad (4.87)$$

With the aid of the Stirling approximation²⁴, the ratio of the two Gamma functions transforms into:

$$\frac{\Gamma(2\varepsilon_N)}{\Gamma(\varepsilon_N)^2} \sim_{\varepsilon_N \gg 1} \sqrt{\frac{\varepsilon_N}{4\pi}} 2^{2\varepsilon_N}. \quad (4.89)$$

The x -dependence factor in eq. (4.87) can be written as:

$$(1-x^2)^{\varepsilon_N-1} = \exp((\varepsilon_N-1)\ln(1-x^2)) \sim \exp(-\varepsilon_N x^2), \quad (4.90)$$

where the last step follows from the approximation $\ln(1-x^2) \approx -x^2$, for $|x| \ll 1$. Plugging the approximations (4.89) and (4.90) into (4.87), we obtain:

$$p_e(x, \varepsilon_N \gg 1) = \frac{1}{\sqrt{2\pi} \sigma} \exp\left(-0.5 \left(\frac{x}{\sigma}\right)^2\right), \quad (4.91)$$

which is a Gaussian with variance:

$$\sigma^2 = \frac{1}{2\varepsilon_N}. \quad (4.92)$$

The transition from a bimodal to a unimodal peaked distribution implies the existence of a critical number of agents N_C , which we might conveniently define by the relation²⁵:

$$\varepsilon = \frac{N_C}{N_0} k = 1 \quad ; \quad N_C = \frac{N_0}{k}.$$

The critical size of the market, N_C , might be employed to distinguish between two regimes: An interacting regime, where the random meetings among agents have a relevant impact on the population dynamics, with a resulting bimodal distribution, and an ‘individualistic’ regime, where the market can be considered as a collection of independent and non-interacting agents. The resulting Gaussian probability distribution for the case $N \gg N_C$ is, in fact, the asymptotic distribution for a collection of independent agents. The approach to the equilibrium can be well described by eq. (4.57). In the case of extensive transition rates and a large number of components, the system can be described by a deterministic equation for the mean and a ‘small’ *superimposed* noise perturbation (see Gardiner [2003] and especially Aoki [1996, 2002] for several applications of this formalism to economics).

The Kirman’s model based on local interactions among the individuals collapse into a trivial aggregate behavior if the number of individuals is increased over a critical threshold. The extensivity of the transitions probability plays the crucial role in his formalization. Figure 4.8 shows graphically the convergence to the Gaussian and the emergence of the macroscopic trajectory.

²⁴The Stirling approximation states that:

$$\Gamma(y+1) = \sqrt{2\pi y} y^y e^{-y}, \quad (4.88)$$

if $y \gg 1$.

²⁵Note that in the case $\varepsilon_N = 1$ the equilibrium distribution of x or z is uniform.

4.12 Conclusions

In this chapter we have introduced a generalization of the herding model developed by Kirman [1993]. The symmetric transition rates of his model, have been replaced by asymmetric ones, which makes the system more flexible in describing various scenarios: namely unimodal or bimodal distributions, as in the original Kirman's model, and monotonic increasing or decreasing distributions. A second important modification is allowing for a non-extensive character of the transition rates, which leads to independence of the fluctuations of the system size.

The dynamics of the system is described in terms of the Fokker-Planck equation, which provides a very powerful analytical tool in order to characterize the aggregate behavior of the group of agents. The FPE has been integrated in terms of an infinite expansion in Jacobi polynomials. This decomposition allows the determination of several statistical properties, such as the autocorrelation function of the raw variable describing the system, namely z , and its simple non-linear transformations.

Appendix to Chapter 4

A 4.1 Backward equation

Let us start with the general FPE, with stationary drift and diffusion functions $A(z)$ and $D(z)$, respectively:

$$\partial_t p(z, t) = -\partial_z[A(z)p(z, t)] + \frac{1}{2}\partial_z^2[D(z)p(z, t)]. \quad (\text{A 4.1})$$

Denote by $p_e(z)$ the equilibrium distribution, which, by definition, satisfied:

$$-\partial_z[A(z)p_e(z)] + \frac{1}{2}\partial_z^2[D(z)p_e(z)] = 0. \quad (\text{A 4.2})$$

The reflecting boundary conditions $J(0) = J(1) = 0$ give, then, the further relation:

$$-A(z)p_e(z) + \frac{1}{2}\partial_z[D(z)p_e(z)] = 0. \quad (\text{A 4.3})$$

According to the definition (4.45), we can express the probability density $p(z, t)$ as the product of the stationary distribution $p_e(z)$ and the function $Q(z, t)$. By direct substitution of eq. (4.45) in the FPE (A 4.1), we find:

$$p_e \partial_t Q = \left(\frac{1}{2}\partial_z^2[Dp_e] - \partial_z[Ap_e] \right) Q + (\partial[Dp_e] - Ap_e) \partial_z Q + \frac{1}{2}Dp_e \partial_z^2 Q, \quad (\text{A 4.4})$$

where we omitted the dependence on z and t of all the functions. Using the relations (A 4.2) and (A 4.3), we end up with the *backward equation* for $Q(z, t)$:

$$\partial_t Q(z, t) = A(z)\partial_z Q(z, t) + \frac{1}{2}D(z)\partial_z^2 Q(z, t), \quad (\text{A 4.5})$$

which proof eq. (4.46).

A 4.2 Boundary conditions

The function $q(z)$ has to satisfied the hypergeometric differential equation (4.49), that is an homogeneous differential equation of second order. The standard form is given by:

$$\frac{d^2}{dz^2}q(z) + L(z)\frac{d}{dz}q(z) + M(z)q(z) = 0, \quad (\text{A 4.6})$$

where the function $L(z)$ is:

$$L(z) = \frac{\varepsilon_1 - (\varepsilon_1 + \varepsilon_2)z}{z(1-z)}, \quad (\text{A 4.7})$$

and $M(z)$ reads as:

$$M(z) = \frac{\lambda}{z(1-z)}. \quad (\text{A 4.8})$$

The properties of the solutions of this equation, denoted by $q(z; \varepsilon_1, \varepsilon_2, \lambda)$, depend on the behavior of the functions $M(z)$ and $L(z)$ in their domain. If they are regular in an arbitrary point z_0 , the solution will also be regular in z_0 . Moreover, singularities

of $L(z)$ and $M(z)$ are also singularities for $q(z; \varepsilon_1, \varepsilon_2, \lambda)$. Without entering into the details of the theory of differential equations, the singular points of the functions $L(z)$ and $M(z)$ are called Fuchsian singularities, denoted by \tilde{z} , if the two following limits are finite:

$$\lim_{z \rightarrow \tilde{z}} (z - \tilde{z})L(z) = l(\tilde{z}), \quad (\text{A 4.9})$$

$$\lim_{z \rightarrow \tilde{z}} (z - \tilde{z})^2 M(z) = m(\tilde{z}). \quad (\text{A 4.10})$$

Loosely speaking, the two conditions (A 4.9) and (A 4.10) state that if the functions $L(z)$ and $M(z)$ diverge ‘regularly enough’ around the singularities, then the solution is, to some extent, well-behaved. Focusing on our problem, equation (A 4.6) possesses two Fuchsian singularities at $z = 0$ and $z = 1$, since the functions $L(z)$ and $M(z)$ are unbounded; by elementary calculations, it turns out that $l(0) = \varepsilon_1$, $l(1) = \varepsilon_2$ and $m(0) = m(1) = 0$.

The general solutions of (A 4.6) are *hypergeometric functions* $q(z; \varepsilon_1, \varepsilon_2, \lambda)$ (see Arfken [1985]). If we plug in the Taylor series expansion of the solution $q = \sum_{k=0}^{\infty} a_k (z - z_0)^k$ around a regular point z_0 , in eq. (A 4.6), we obtain the recursion formula:

$$a_{k+1} = a_k \frac{k(k + \varepsilon_1 + \varepsilon_2 - 1) - \lambda}{(k + 1)(k + \varepsilon_1)}, \quad (\text{A 4.11})$$

which provides the values of the coefficients a_k , given the initial value a_0 . The Fuchsian points should be treated separately. The expansion of the solution around a singular point assumes the the form:

$$q_{\tilde{z}}(z; \varepsilon_1, \varepsilon_2, \lambda) = (z - \tilde{z})^{\rho(\tilde{z})} \sum_{k=0}^{\infty} a_k (z - \tilde{z})^k, \quad (\text{A 4.12})$$

where $\rho(\tilde{z})$ is a solution of the *characteristic equation*²⁶:

$$\rho^2 + [l(\tilde{z}) - 1]\rho + m(\tilde{z}) = 0. \quad (\text{A 4.13})$$

In our case, we end up with the following expansions around the two singular points, $\tilde{z}_1 = 0$ and $\tilde{z}_2 = 1$, respectively:

$$q_0(z; \varepsilon_1, \varepsilon_2, \lambda) = z^{1-\varepsilon_1} \sum_{k=0}^{\infty} a_k z^k \quad (\text{A 4.14})$$

$$q_1(1 - z; \varepsilon_1, \varepsilon_2, \lambda) = (1 - z)^{1-\varepsilon_2} \sum_{k=0}^{\infty} a_k (1 - z)^k. \quad (\text{A 4.15})$$

More relevant for our calculations is the behavior of the derivative of $q(\cdot)$ around the two singularities, given by:

$$q(\cdot)'_0(z; \varepsilon_1, \varepsilon_2, \lambda) \approx \text{const } z^{-\varepsilon_1}, \quad (\text{A 4.16})$$

²⁶In the expansion (A 4.12) around the singularities is compatible with the definition of Fuchsian points in eqs. (A 4.9) and (A 4.10). Plugging those expansions in the main equation (A 4.6) we end up with a second order equation for the coefficient $\rho(\tilde{z})$ for every Fuchsian point.

$$q(\cdot)'_1(1-z; \varepsilon_1, \varepsilon_2, \lambda) \approx \text{const} (1-z)^{-\varepsilon_2} . \quad (\text{A 4.17})$$

Turning now to the solution the differential equation (4.49), its initial conditions are provided by the reflecting boundary condition imposed by the conservation of the probability, i.e. $J(0) = J(1) = 0$. Given the relation (A 4.2) and the decomposition of the probability density $p(z, t) = p_e(z)q(z)e^{-\lambda bt}$, the general expression for the current (4.32) transforms into:

$$J(z, t) = e^{-\lambda bt} D(z) p_e(z) \frac{d}{dz} q(z) \sim z^{\varepsilon_1} (1-z)^{\varepsilon_2} \frac{d}{dz} q(z) . \quad (\text{A 4.18})$$

Taking into account the expansions (A 4.16) and (A 4.17), the current is always finite and different from zero at the boundaries, unless λ assumes the discrete values:

$$\lambda_n = n(n + \varepsilon_1 + \varepsilon_2 - 1) , \quad (\text{A 4.19})$$

which leads to solutions of the differential equation (A 4.6) that are polynomials. The recursive equation (A 4.11), in fact, gives rise to n-th non-vanishing coefficients, while, for $k > n$ all the coefficients are identically zero. In other words, the reflecting boundary conditions force the parameter λ , in eq. (A 4.6), to assume solely discrete values fixed by (A 4.19). Moreover, the solutions of the differential equation are polynomials, in fact Jacobi polynomials, that satisfy the eigenvalues equation:

$$z(1-z) \frac{d^2}{dz^2} q(z) + [\varepsilon_1 - (\varepsilon_1 + \varepsilon_2)z] \frac{d}{dz} q(z) = -\lambda_n q(z) . \quad (\text{A 4.20})$$

A 4.3 The Jacobi polynomials

The Jacobi polynomials²⁷ are eigenvectors of the Jacobi differential equation (A 4.20), where the parameter λ_n assumes the discrete values given by (A 4.19). They form a *complete orthogonal system* of polynomials in the compact domain $[0, 1]$, with respect to the weighting function $p_e(z)$. It means that:

$$\int_0^1 P_n(z) P_m(z) p_e(z) dz = h_n \delta_{n,m} , \quad (\text{A 4.21})$$

where h_n is a normalization function given by:

$$h_n = \frac{1}{n!} \frac{\Gamma(\varepsilon_1 + n)}{\Gamma(\varepsilon_1)} \frac{\Gamma(\varepsilon_2 + n)}{\Gamma(\varepsilon_2)} \frac{\Gamma(\varepsilon_1 + \varepsilon_2)}{\Gamma(\varepsilon_1 + \varepsilon_2 + n - 1)} \frac{1}{\varepsilon_1 + \varepsilon_2 + 2n - 1} . \quad (\text{A 4.22})$$

The normalization is based on the condition:

$$P_n^{(\varepsilon_1, \varepsilon_2)}(1) = \binom{n + \varepsilon_2 - 1}{n} . \quad (\text{A 4.23})$$

²⁷The properties of the Jacobi polynomials listed in this section are standard results in the literature of classical polynomials (see Abramowitz and Stegun [1972], ch. 22), therefore we will not give the detailed calculation. However, the reader should be careful when comparing with other sources, since the normalization can be different. Moreover, in some text-books the polynomials are expressed as a function of the variable $x = 2z - 1$, without obviously changing the content of the theory.

The completeness property is expressed by:

$$\sum_{n=0}^{\infty} \frac{1}{h_n} p_e(z) P_n^{(\varepsilon_1, \varepsilon_2)}(z) P_n^{(\varepsilon_1, \varepsilon_2)}(z_0) = \delta(z - z_0). \quad (\text{A } 4.24)$$

The polynomial of degree n is given by the following sum:

$$P_n^{(\varepsilon_1, \varepsilon_2)}(z) = \frac{1}{n!} \sum_{i=0}^n (-1)^i \binom{n}{i} \frac{\Gamma(\varepsilon_2 + n)}{\Gamma(\varepsilon_2 + i)} \frac{\Gamma(\varepsilon_1 + n)}{\Gamma(\varepsilon_1 + n - i)} z^{n-i} (1-z)^i. \quad (\text{A } 4.25)$$

The first two polynomials are:

$$P_0^{(\varepsilon_1, \varepsilon_2)}(z) = 1, \quad h_0 = 1, \quad (\text{A } 4.26)$$

$$P_1^{(\varepsilon_1, \varepsilon_2)}(z) = (\varepsilon_1 + \varepsilon_2)z - \varepsilon_1, \quad h_1 = \frac{\varepsilon_1 \varepsilon_2}{\varepsilon_1 + \varepsilon_2 + 1}. \quad (\text{A } 4.27)$$

Given the completeness of the Jacobi set, every function $f(z)$ can, then, be expanded in series of Jacobi polynomials, as following:

$$f(z) = \sum_{n=0}^{\infty} a_n p_e(z) P_n^{(\varepsilon_1, \varepsilon_2)}(z), \quad (\text{A } 4.28)$$

The coefficients a_n are given by:

$$a_n = \frac{1}{h_n} \int_0^1 f(y) P_n^{(\varepsilon_1, \varepsilon_2)}(y) p_e(y) dy, \quad (\text{A } 4.29)$$

which is derived using the property of orthogonality of the polynomials.

CHAPTER 5

An artificial financial market based on herding: Analytical results

In this chapter we employ the herding mechanism detailed in the previous part of the thesis as the main ingredient for an artificial financial market populated by heterogeneous interacting agents¹. In the first part of this chapter, we describe the structure of the market. The two states of the herding model are identified with fundamentalists and technical traders, following the well-established dichotomy of financial investors introduced by Beja and Goldman [1980] and nowadays widespread in financial agent-based literature. Using, then, two simple behavioral rules for the formalization of the excess demand of the groups and a Walrasian market clearing mechanism, the market equilibrium price is expressed as a function of the external and internal elements which influence the market, namely the fundamental information and the dynamics of the traders. The last step regards the analytical approximation of the underlying stochastic process which governs the returns. It is shown that this process can be expressed as a stochastic volatility decomposition, which is the typical framework for the description of the financial time series —see chapter 3.

This remarkable equivalence allows for an analytical treatment of the outcome of the agent-based financial market model proposed here. The last part of the chapter is devoted to a detailed analysis of the statistical properties of the stochastic process proposed to model the returns dynamics. The connection between the microscopic features of the investors and the macroscopic aggregate quantities characterizing the returns process is described along the whole chapter. Given the analytical characterization of the entire model, special emphasis is put on the origin of the fluctuations, and on the connection with the stylized facts.

5.1 The artificial market

In this paragraph we will present the implementation of the herding mechanism, detailed in the previous chapter, in an artificial financial market. The description of the characteristics of the market players is presented in the first part. The second part is devoted to the elaboration of an analytically tractable approximation of the returns dynamics.

¹Part of the material of this and of the following chapter has been published in Alfarano et al. [2005].

5.1.1 Implementation of the market

As we have stressed in the introduction to this chapter, we believe that real traders are heterogeneous — e.g. with respect to expectations, available information, market power etc. — and interact not only via a centralized institution, the market, but also in a decentralized way², gathering information for instance through channels like interpersonal communication or mass-media.

In order to simplify this complex view of real markets, we divide the fixed number of traders N into two categories or types:

- N_F *fundamentalists*, who buy or sell according to the deviation between the spot price p and the fundamental value p_F ;
- N_C *technical traders* who rely on chartistic techniques rather than fundamental information.

The fundamental value p_F is constant over time and exogenously given. The two initially arbitrary states of the previous chapter are now employed to characterize two different attitudes of traders. Precisely, if an agent belongs to the former state 1 we will call him fundamentalist, while the second state stands for technical traders.

We assume that financial agents exchange information about their own beliefs, e.g. by giving suggestions or being interested in the particular attitude of the other investors. In this way they mutually influence their strategies. The literature on financial investments as a socio-economical activity is vast. The paper of Shiller [1984] can be considered as a seminal contribution initiating this branch of the literature. Quoting from this paper:

“Investing in speculative assets is a social activity. Investors spend a substantial part of their leisure time discussing investment, reading about investments, or gossiping about others’ successes or failures in investing. It is thus plausible that investors’ behavior (and hence prices of speculative assets) would be influenced by social movements.”

The recent explosion of the activities of on-line forums devoted to financial investments constitutes an example for this type of social interaction.

If we argue that interpersonal communication plays a *crucial* role in the behavior of financial agents, the herding model detailed in the previous chapter might be a suitable mechanism to mathematically formalize this type of interactions. This mechanism is based on the simple idea that a trader is more likely to change his investment attitude, let say from a strategy based on chartistic techniques to a strategy based on fundamental information, if in his environment a consistent number of

²The literature on financial market microstructure — see the book of O’Hara [1995] and references therein — is typically assuming a centralized market structure, so-called star market architecture, where the traders interact only with the market maker, while indirectly exchanging information via the market price.

agents adopts the later strategy.

Our market can be, therefore, casted in the broad category of *bipolar dynamical markets*. We define as *bipolar* a market where the heterogeneity of the traders is limited to two distinctive categories, in our case fundamentalists and technical traders. The bipolar structure of an artificial financial market is wide spread in the literature on agent-based models — see the papers of Beja and Goldman [1980], Frankel and Froot [1986], Kirman [1993] and Lux [1995]. We assume here that the fundamentalists just do the work of bringing prices towards fundamentals, while technical-traders might act to deviating from it. The sharp distinction between fundamentalists and technical traders seems rather artificial, since the separation among investors' attitudes may be blurred. Nonetheless, this clear-cut distinction simplifies the complex 'ecology' of real markets.

In our setting, the investors can change their attitude according to the transition rates, specified in eqs. (4.4). This leads to the *dynamical* characterization of our artificial market. We argue that the switching between both groups might model a certain degree of indeterminacy of the strategy which an individual is adopting. In other words, a 'real' agent trading in the market is not applying a pure fundamentalist or chartistic strategy. Typically, a mixture of technical trading and fundamental analysis is the basis of the adopted market strategy. We believe that the unrealistic sharp distinction between both groups of traders is attenuated by the possibility for an investor to change attitude.

Instead of focusing on the detailed reasons of each trader to follow one specific strategy, according to his or her risk attitudes or preferences, we adopt a probabilistic view as a mechanism responsible for changes of strategies. The probabilistic approach, that enters into the model via the *transition probabilities* (4.4), allows for an interesting simplification in dealing with a large pool of interacting heterogeneous agents. On the one hand, we can model in a meaningful way our lack of information of the risk preferences of every single trader, avoiding *ad hoc* assumptions on their preferences. On the other hand, we can nevertheless keep a certain degree of heterogeneity among the interacting agents. Moreover, the probabilistic approach together with the large number of agents gives rise to unexpected and interesting regularities at aggregated level. The aggregation procedure is not a mere sum of every single individual behavior, but it shows new features of the system, which are not present at the micro level³.

As described in the previous chapter, the switches of the agents between the two attitudes follow a pure stochastic process. Obviously, the pure randomness of the switching mechanism is a strong simplification of reality. In the transition rates (4.4), in fact, we do not encounter any dependence on the price dynamics. In many models a feedback dependence between the price dynamics and the traders' behavior has been introduced. Examples in the literature are the model of Lux and March-

³For a very interesting discussion on the emergence of aggregate macroscopic laws see the paper of Ramsey [1996].

esi [1999], which is related to our model, and the Santa Fe' artificial Stock Market [Arthur et al., 1997], just to quote some of them. It might be probably true that both social interactions and price signals influence the investment decisions of the traders. For analytical tractability, we focus solely on the social interactions as the key-factor influencing the transitions among the different attitudes.

Changes in the relative values of the three key-parameters, namely a_1 , a_2 and b , lead to different market scenarios. For example, $a_1 > a_2$ would mean that, on average, the propensity of autonomous conversion of a former fundamentalist is higher than the probability for a switch in the opposite direction. The consequence is a dominance of the technical trader attitude in this particular setting. It is important to stress that the difference in the two parameters a_1 and a_2 *exogenously* introduces a preferential attitude. This distinction is not justified in terms of better survival possibilities or other economical arguments, but it will be rather estimated through the empirical data.

All in all, our bipolar dynamical market is, on the one hand, a generalization of the homogeneous landscape of the EMH and, on the other hand, an important simplification of the complete heterogeneity that we think real markets are permeated with.

5.1.2 Aggregate excess demand

The trading attitudes of the agents translate into a market price via two behavioral rules for demand and supply. As it is typically done in the literature, we set a simple formula to describe the aggregate excess demand for each group, namely fundamentalists and technical traders.

Fundamentalists

Fundamentalists' excess demand is given by:

$$ED_F = N_F \ln \frac{p_F}{p} . \quad (5.1)$$

We assume that each fundamentalist is characterized by the same reaction to deviations from the fundamental value, buying (selling) whenever he perceives an undervaluation (overvaluation) of the stock price. The aggregate excess demand of this group is, then, the sum of the demand of a 'representative' fundamentalist times the number of fundamentalists, N_F . The peculiar reaction function $\ln \frac{p_F}{p}$ greatly simplified the final analytical form for returns —see eqs. (5.8) and (5.10). However, a different reaction function involving, for instance, the absolute difference $(p_F - p)$ instead of the relative one, does not drastically affect the general behavior of the model. Anyhow, the excess demand of the fundamentalists can be alternatively written as:

$$\ln \frac{p_F}{p} \simeq \frac{p_F - p}{p} , \quad (5.2)$$

which holds if the deviation from the fundamental value is small relative to the market price. We, then, recover the functional form employed in many other con-

tributions in the literature (in the model of Lux and Marchesi [1999], for instance).

Technical traders

We assume that the technical traders' aggregate excess demand takes the form:

$$ED_C = -r_0 \sum_{i=1}^{N_C} \theta_i(\vec{p}), \quad (5.3)$$

where $\theta_i(\vec{p})$ reflects the particular chartistic strategy followed by the technical trader i , which is based on the history of the price, denoted by \vec{p} . The constant r_0 is a scale factor that accounts for their impact on the price formation, and the expression is multiplied by -1 for notational convenience. Typically, in the literature of chartists/fundamentalists, the chartistic technique that every group of chartists is using is specified. Trend following or trend chasing, moving average techniques are the most common inference procedures to extract information from the price chart. However, any book on technical trading —Edwards et al. [2001]— shows a much larger spectrum of existing possibilities. Contrary to this large part of the literature, we model the aggregate excess demand of the technical traders' group without accounting for specific trading rules. The crucial element of the technical traders' excess demand is the aggregate impact of the overall pool of strategies. We can, in fact, rewrite eq. (5.3) in the following way:

$$ED_C = -r_0 N_C \frac{\sum_{i=1}^{N_C} \theta_i(\vec{p})}{N_C}. \quad (5.4)$$

We argue that a useful way to model the aggregate impact of many heterogeneous chartist techniques might be to describe it with as a stochastic process. We introduce, then, the 'mood' or 'sentiment' of the technical traders as following:

$$\xi = \frac{\sum_{i=1}^{N_C} \theta_i(\vec{p})}{N_C}. \quad (5.5)$$

The quantity ξ can be related to the vast literature of fads or fashion in financial markets — see the book of Shleifer [1999] and the paper of Shiller [1984], among others.

The excess demand of the technical traders is now:

$$ED_C = -r_0 N_C \xi. \quad (5.6)$$

Interestingly, the independence of ξ on the number of technical traders N_C implies the presence of a high degree or eventually a perfect correlation among them, following the implementation of De Long et al. [1990]. An aggregate demand of a collection of independent individuals with random, possibly standard normally distributed, orders would bring about a term of the form $\sqrt{N_C}$, as implied by the central limit theorem. The assumption of aggregate linear dependence implies a deviation from the CLT, and consequentially some form of correlation among agents. The nature of this correlation and the precise statistical features of the stochastic process ξ will be clarified later in correspondence of the properties of the resulting time series of returns.

Price

We compute the equilibrium price within a Walrasian scenario by simply setting the total excess demand equal to zero:

$$ED_F + ED_C = 0. \quad (5.7)$$

We, then, end up with the following formula for the market price:

$$p = p_F \cdot \exp\left(r_0 \frac{z}{1-z} \xi\right), \quad (5.8)$$

where z and $1 - z$ are the fractions of the technical traders and fundamentalists among the total number of agents, respectively. The functional form of the equilibrium price p of eq. (5.8) is composed by two factors, which possess a different nature. The first factor depends on the fundamental price, or, in other words, on the exogenous information hitting the market. The second component is clearly related to the interactions among the traders.

The use of a Walrasian updating of the price implies that the market is always at the equilibrium and, therefore, the market price is the equilibrium price. This might be a reasonable assumption for daily data. It should not be surprising, then, if the model would not conform to the statistical properties of high frequency data. Moreover, it should be noticed that the use of the Langevin equation for the description of the dynamics of the agents already implies a time aggregation process (see section 4.9). The use of high frequency data (minute-to-minute for instance) could not be compatible with this underlying aggregation of agents (and their orders).

5.1.3 Computation of the returns

We define, now, the log-returns, computed over a time horizon Δt , as follows:

$$r(t, \Delta t) = \ln(p_{t+\Delta t}/p_t). \quad (5.9)$$

Note that the time-unit Δt of the return process is different from the elementary time-unit of the population changes $\Delta \tau$; we, therefore, refer to the former as micro-time and the latter as macro-time. Essentially, during a macro-time Δt , z is aggregated over movements of many agents between the two states. This aggregation procedure plays a crucial role in the theoretical model. The introduction of two time scales, namely micro and macro, allows for a clear-cut separation between the dynamics of the equilibrium price, which involves the longer time unit, and the change of the attitude of a single individual, which occurs in a much shorter time-interval. The large number of agents, then, permits a statistical approach for the dynamic of the price, approximated by a diffusion process in continuous time.

Turning now to the computation of returns, they are given by:

$$r(t, \Delta t) = r_0 \left[\frac{z(t + \Delta t)}{1 - z(t + \Delta t)} \xi(t + \Delta t) - \frac{z(t)}{1 - z(t)} \xi(t) \right], \quad (5.10)$$

where Δt is the time interval over which they are computed. Introducing the new variable:

$$\sigma(t) = r_0 \frac{z(t)}{1 - z(t)}, \quad (5.11)$$

eq. (5.10) can be more compactly rewritten as:

$$r(t, \Delta t) = \sigma(t + \Delta t) \xi(t + \Delta t) - \sigma(t) \xi(t) = \Delta[\sigma(t) \cdot \xi(t)], \quad (5.12)$$

where $\Delta[\cdot]$ is the differential operator. Note that the dependence of the fundamental value p_F disappears from the computation of returns, since we have assumed that it is constant over time. Alternatively, if we assume that the fundamental value p_F follows a random walk with constant drift μ and a constant diffusion D , the previous formula can be generalized as:

$$r(t, \Delta t) = \mu \Delta t + \chi(t) + [\sigma(t + \Delta t) \xi(t + \Delta t) - \sigma(t) \xi(t)]. \quad (5.13)$$

where $\chi(t)$ is a Normal random variable with zero mean and standard deviation D .

A complete analytical characterization of eq. (5.10), or equivalently (5.12), turns out to be cumbersome, considering the positive correlation of the variable z over time and the presence of two sources of randomness, namely z and ξ , which are not separable. However, we can provide some partial quantitative results. First, it is easy to check that the *expected value is identically zero*, $E[r(t, \Delta t)] = 0$, under the quite general assumption that both the unconditional mean values of $\sigma(t)$ and $\xi(t)$ are finite. This is certainly true for $\sigma(t)$ given the stationarity of the process $z(t)$ (see below for more details). For $\xi(t)$ it depends on the parametric choice of its underlying stochastic process.

The variance of $r(t)$ depends crucially on which would be the underlying stochastic process governing the evolution of $\xi(t)$. Following the original model of De Long et al. [1990] on the impact of a group of noise traders in the market populated by informed traders, the stochastic variable ξ should be iid. However two important drawbacks have to be taken into account. Firstly, their model is based on an overlapping generation framework, which implies too long time-units compared to our original intention to model daily data. The De Long's formalization, in fact, is realistically confined to a time period of at least several years, if we interpret literally the overlapping generation framework as a live cycle of an individual. Additionally, the mere application of the De Long iid assumption generates, in our setting, abrupt variations of the market price —see for instance eq. (5.8)— which is not in harmony with the empirical data.

Alternatively, if we assume that the variable $\xi(t)$ follows a random walk, the returns process (5.10) will be non-stationary, since its variance increases linearly over time. It easy to show that:

$$E[r(t, \Delta t)^2] = 2E[\xi^2](E[\sigma_t^2] - E[\sigma_{t+\Delta t}\sigma_t]), \quad (5.14)$$

which, given the stationarity of the variable $\sigma(t)$, proves the overall non-stationarity of the time series of returns. However, along with the random walk choice for ξ

comes also the non-stationarity of the time series of prices, as we can see from eq. (5.8).

In order to induce stationarity to the returns time series and avoiding the unrealistic price behavior implied by the iid assumption, we might assume that ξ follows an AR(1) process⁴:

$$\xi_t = a\xi_{t-\Delta t} + \eta_t , \quad (5.15)$$

with the autoregressive coefficient a very close to 1 and the disturbances iid distributed with $E[\eta_t] = 0$ and $Var[\eta_t] = \sigma_0^2$. It means that the mood, sentiment or momentum of the technical traders might considerably deviate from its expected value ($E[\xi] = 0$) for quite some time, however, reverting eventually to its unconditional mean. The economic intuition behind the AR parametric choice can be qualitatively justified by possible behavioral assumptions on the dynamics of the aggregate technical traders' mood. We assume that there is a certain degree of coordination among their expectations about the future development of the price, which mirrors in a persistent but temporary positive (or negative) aggregated value of their excess demand for the asset. If we refer to the model of De Long et al. [1990], the noise traders' expectations are correlated (in fact fully correlated) which create an endemic market risk that the arbitragers have to face, rather than a sum of idiosyncratic risks formed by independent noise traders.

It should be emphasized that the stationarity of the stochastic variable ξ carries over the time series of the price, as we can see from eq. (5.8) and Figure 5.1. Nonetheless, the 'almost' unit root characterization of our AR process for ξ guaranties a behavior of the price that might look non-stationary for finite sample sizes. One may ask whether the price of a financial asset follows a 'pure' random walk or, alternatively, an autoregressive process with an AR coefficient very close, however lower, to the unit root, superimposed to a stochastic or deterministic exogenous growth. The non-stationarity of financial data, then, comes along trends generated, for instance, by an exogenous growth of the overall economy, rather than by the unit root implied by the random walk⁵. To model this type of non-stationarity, we can easily generalize eq. (5.10) along the line of eq. (5.13), and assume a stationary stochastic process for ξ .

An alternative possibility to induce stationarity in the returns time series might be to adopt for the time development of ξ a random walk process, with a bounded domain and with reflecting boundaries⁶:

$$\xi_t = \xi_{t-\Delta t} + \eta_t . \quad (5.16)$$

It means that the 'sentiment' of the technical traders cannot anymore indefinitely

⁴Obviously a more complex ARMA process might be also employed, however the further complexity does not add more explanatory power to the model. In our case, simplicity is the fundamental ingredient that we require.

⁵See the paper of Shiller [1981] and its analysis of stock prices.

⁶Without loss of generality we can assume a symmetric interval centered in zero.

grow, but it is constrained within a specific interval.

Remark One can complain that the dynamics of the technical traders ‘mood’ or ‘sentiment’ possesses a rather intangible nature. Additionally, quite a number of assumptions have been introduced in the literature to model their behavior. Obviously, since we cannot directly observe the dynamics of the opinion formation, we have to rely on the indirect consequences of our parametric implementations. The ultimate aim of our work is to set the stage for an analytical approach to an agent-based artificial market. Therefore, we *must* precisely formalize also those aspects of the model that are not directly observable in reality. Our hope is that the limitations given by many parametric assumptions involved in the setting of the model, might be overcome by the interesting final results.

5.1.4 Stochastic volatility approximation

The stochastic process governed by eq. (5.10) provides us with a limited pool of analytical results, constraining our investigation to rely on Monte Carlo simulations (see Figure 5.1). In order to support our analysis with more analytical insights in the equilibrium price dynamics, we should simplify the underlying stochastic process (5.10).

Let us first focus on other more quantitative results regarding the process (5.10). In order to do so, we specify the underlying dynamics of the aggregate opinion of the technical traders, ξ , as an AR(1) process. We can, then, compute some specific moments of (5.10): $E[r_{t,\Delta t}^2]$, $E[r_{t,\Delta t}^4]$ and first lag of the auto-covariance $E[r_{t+\Delta t} \cdot r_t]$. These moments are given by the following expressions (see appendix A 5.1)⁷:

$$E[r^2] \approx E[\sigma^2] \cdot \sigma_0^2, \quad (5.17)$$

$$E[r^4] \approx E[\sigma^4] \cdot E[\eta^4], \quad (5.18)$$

$$E[r_{t+\Delta t} \cdot r_t] \approx -2E[\sigma^2] \cdot \sigma_0^2 \cdot \frac{1-a}{1+a}. \quad (5.19)$$

Using eqs. (5.17) and (5.19), the first lag of the auto-correlation can easily be derived, leading to the expression:

$$C(1) = -2 \frac{1-a}{1+a} \approx 0. \quad (5.20)$$

The last result has a very intuitive economic appeal. The (linear) correlation among adjacent returns is close to zero. The small negative value is related to the mean reverting nature of the AR process⁸. The process (5.10) squares well with the empirical evidence of absence of linear correlation among returns. Here the importance of the assumption $a \approx 1$ should be stressed. A value of the AR-coefficient higher than one, on one hand, would induce a non-stationarity of the time series of returns. On the other hand, if its value would be lower but not close to one, it would have the consequence of creating a significant dependence in the time series of returns, which is not compatible with the behavior of real data.

Eqs. (5.17) and (5.18) allow for the computation of the kurtosis:

$$\kappa[r] = \frac{E[r^4]}{E[r^2]^2} = \kappa[\sigma] \cdot \kappa[\eta]. \quad (5.21)$$

If we assume a standard Normal distribution for η , the leptokurtic nature of the return distribution depends directly on the distribution of the volatility, and ultimately on the values of the parameters ε_1 and ε_2 . Given the power-law decay of

⁷The dependence of the variables on time t and on the time-interval Δt is omitted when not necessary.

⁸In principle, the small negative correlation is at odds with the informational efficiency and martingale nature of financial prices. However, the small mean-reverting tendency of the ξ and, consequently, the returns can easily be blurred by the noise level generated by the finiteness of the data sample and would not be detected if a were not too far from 1 (typical values in our simulations are $a = 0.99$).

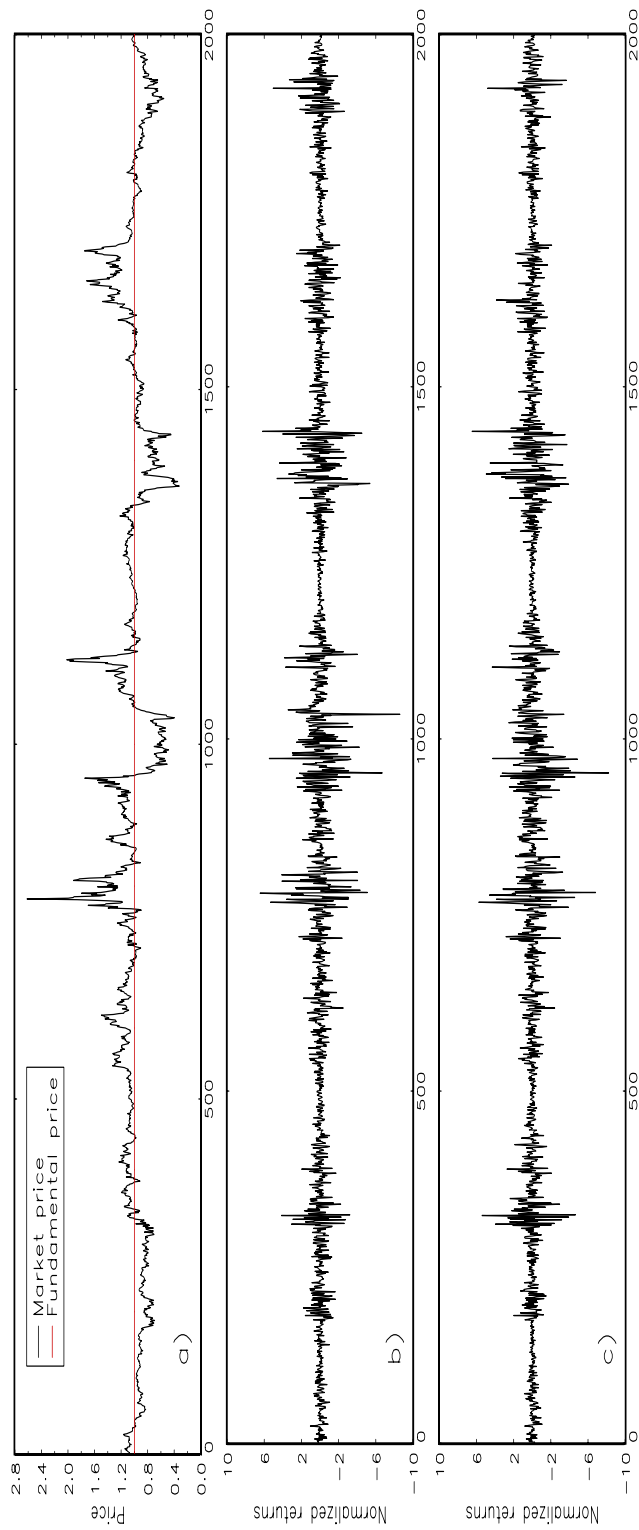


Figure 5.1: Panel (a) shows the time evolution of the price as computed from eq. (5.8). Panel (b) exhibits the time series of log-returns derived from eq.(5.12). The lower panel (c) is computed using the stochastic approximation introduced eq. 5.25. Note that the two time series of returns, namely the ‘correct’ and the approximated one, are in good agreement with each other. The parameters of the volatility process are: $\varepsilon_1 = 3.2$, $\varepsilon_2 = 3.9$ and $b = 0.0025$. The volatility time series has been computed using the Euler approximation —see eq. 5.36. For the dynamics of the sentiment of the technical traders an AR(1) process has been used. The values of its parameters are: $a = 0.99$ and $\sigma_0 = 0.2$ and its increments are normally distributed.

the distribution of σ (see next paragraph), our process can generate the leptokurtic shape of returns distribution found in the empirical literature (see chapter 2).

At this point, we can proceed to analyze the process (5.10) by means of Monte Carlo simulations, then to investigate several relevant statistical features of the syntectic time series (Hill estimator, Hurst exponent, unit root, just to mention some of them), following the seminal methodology of Lux and Marchesi [1999, 2000], which is applied nowadays to a long list of agent-based models. This approach has been used to provide a sort of goodness-of-fit for different agent-based models; more precisely, the goodness of the model under investigation is established on which extent the above econometric techniques, when applied to syntectic data, are in qualitatively agreement with the empirical measurements. However, this approach has several important drawbacks, when applied to these class of models:

- First, and most important, *which values of the parameters* have to be chosen?
- Are the *numerical results robust* under changes of some or all the parameters?
- Which is the *precise role* that a single parameter has in the generating mechanism and in the properties of the time series of returns?
- Can we evaluate the *goodness-of-fit* of the model in describing empirical data?

It is very difficult to answer to the previous list of questions for the majority of the agent-based models in the pertinent literature given their lack of analytical results. The complexity of these models, in fact, constrain their analytical tractability and one has to rely on computer simulations. In principle, this is not a severe methodological problem, but it constitutes an important limitation if one wants to answer to the previous questions.

More interesting for our goal is the structure of eqs. (5.17), (5.18) and (5.19). The same results might be obtained if the generating return process would be of the form:

$$r_{t,\Delta t} = \sigma_t \cdot \eta_{t+\Delta t} . \tag{5.22}$$

where $\eta_{t+\Delta t}$ are the increments of the AR(1) process governing the dynamics of ξ . Motivated by this simple consideration, we can approximate eq. (5.10) by assuming a ‘faster’ dynamics for ξ compared to that of the variable $z/(1-z)$, which can be considered to be constant during a small time interval Δt . This approximation amounts to separate the time scales governing the switching process among attitudes and the underlying dynamics of the ‘mood’ of the technical traders. The mathematical structure of eq. (5.22) turns out to be much more convenient for a full analytical characterization of the statistical properties of the returns, leading to a *closed-form solution* for several interesting quantities: all the unconditional moments and, for particular cases, even the unconditional return distribution. The autocorrelation function of returns and their simple transformations can be computed as well.

Under the assumption given by eq. (5.22), eq. (5.10) can be approximated by:

$$r(t, \Delta t) = r_0 \frac{z(t)}{1 - z(t)} \eta(t, \Delta t) \quad (5.23)$$

where we redefine:

$$\eta(t, \Delta t) \equiv \xi(t + \Delta t) - \xi(t) . \quad (5.24)$$

Eq. (5.23) can be accordingly rewritten as:

$$r(t, \Delta t) = \sigma(t) \eta(t, \Delta t) , \quad (5.25)$$

where we assume that $\eta(t, \Delta t)$ is iid with a given distribution $p(\eta)$. The level of accuracy of the approximation (5.25) depends on the choice of the underlying parameters, namely a_1 , a_2 and b , and the value that the variable z assumes at time t . For given values of $a_{1,2}$ the accuracy between eqs. (5.10) and (5.23) is a decreasing function of b , which, in fact, governs the diffusion part of the size of the increments Δz (see Figure 5.1).

Equation (5.25) is the key result of the present chapter. Starting from the discrete stochastic process governing the behavior of a pool of interacting agents, we end up with an equation that describes the aggregate dynamics of the artificial financial market. It exhibits a so-called stochastic volatility structure, i.e. is given by the product of a white noise, η , and a conditional volatility factor, σ , which incorporates the dependence on the past observations –see the volatility decomposition of eq. (3.17). The iid-ness of the noise term η guarantees the absence of linear correlation of returns, as empirically observed (see chapter 3). The positive correlations of non-linear transformations of returns, squared or absolute values for instance, are then governed by the correlations in the volatility $\sigma(t)$, which originate from and are related to the dynamical properties of $z(t)$.

Figure 5.1 shows a typical price pattern from a Monte Carlo simulation of eq. (5.8). The market price fluctuates around the fundamental value $p_F = 1$, with both periods of positive and negative deviations from it, that we can interpret as bubbles, and subsequent realignments, interpreted as crashes. The corresponding time series of returns exhibits volatility clusters, which arise in close correspondence to deviations from the fundamental value -see panel (b) in Figure 5.1. This intermittent behavior of the returns is related to the change in the market attitude of the traders. Periods of high volatility correspond to a large fraction of technical traders acting in the market; *vice versa*, only minor fluctuations occurs when the market is dominated by fundamentalists. The market as a whole exhibits excess volatility. In our simulation, all the fluctuations of returns are, in fact, generated by the speculative activities of traders, and are disconnected from the fundamental price, here assumed to be constant. The social interaction among the traders, formalized by the herding mechanism described in chapter 4, then, provides the ultimate “engine” for the market dynamics. The behavior of the autocorrelation of raw returns and their simple non-linear transformations reflect this peculiar intermittent dynamics;

absence of linear correlation in returns and positive significant correlation in absolute and squared returns -as measure of volatility— are robust features of the model, as illustrated in the bottom panel of Figure 5.2.

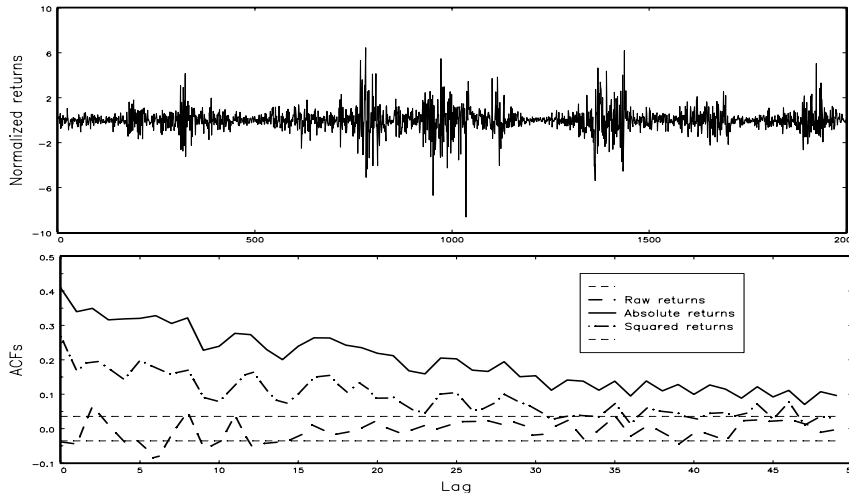


Figure 5.2: The bottom panel shows the behavior of the autocorrelation of raw, absolute and squared returns of the time series in panel (c) of Figure 5.1. The dashed lines represent the 95% confidence interval.

5.2 Statistical properties of the volatility process

The second part of the chapter is devoted to the detailed analysis of the statistical properties of the return process as given by eq. (5.25).

In order to derive the properties of the return process (5.25), we firstly focus on the stochastic properties of the volatility $\sigma(t)$, given by

$$\sigma(t) = \sigma_0 \frac{z(t)}{1 - z(t)}. \quad (5.26)$$

The previous equation can be viewed as a non-linear transformation of the variable z , whose dynamical evolution is governed by the Fokker-Planck equation (4.26), introduced in chapter 4. All the stochastic properties of the variable σ can be derived from analogous properties of the variable z , taking into account the transformation (5.26).

5.2.1 Fokker-Planck equation for the volatility

From eq. (5.26), we can transform the FPE (4.26), expressed in terms of the variable z , into the FPE that governs the evolution of the conditional probability density $p(\sigma|\sigma_i)$ of the volatility σ , given its starting distribution $p_i(\sigma)$. Instead of a mere

algebraic application of the text book formula for the change of variable of the FPE, we detail briefly few steps of its derivation, which provides very interesting insights on the nature of the drift and diffusion terms of the volatility process.

We start from known drift and diffusion functions⁹ $A(x)$ and $D(x)$ respectively, both expressed in terms of a variable x . Let y be a new variable and $y(x)$ a one-to-one time-independent mapping of x into y . Let us, then, denote the Jacobean as $J = \frac{dx}{dy}$. In order to derive the new F-P equation, we can apply the following transformations for the diffusion and drift functions, respectively:

$$\tilde{D} = \frac{D}{J^2}, \quad (5.27)$$

and

$$\tilde{A} = \frac{A}{|J|} - \frac{1}{2} \frac{J'}{J} \tilde{D}, \quad (5.28)$$

where \tilde{A} and \tilde{D} are the diffusion and the drift functions expressed in terms of the variable y . The new diffusion function is proportional to the original diffusion function. The transformed drift is composed by two terms: the first term is the (scaled) original drift rate; the second is given by a contribution of the diffusion term. Interestingly, for linear transformations of the two variables, x and y , this contribution is identically zero, and the resulting drift and diffusion functions are proportional to the original functions.

We can now apply the transformations (5.27) and (5.28) to our particular case (see appendix A 5.2). We know from chapter 4 that, for the stochastic equation governing the dynamics of z , the diffusion function is directly related to the herding interactions among agents, while the drift term depends solely on their autonomous behavior. This clear separation gets lost if we concentrate on the stochastic process of the volatility, as given by eq. (5.26). Via the transformations (5.27), the new diffusion function is given by:

$$\tilde{D}(\sigma) = 2b \frac{\sigma}{\sigma_0} \left(1 + \frac{\sigma}{\sigma_0}\right)^2. \quad (5.29)$$

We have seen in the previous chapter that the fluctuations of the stochastic process governing the dynamics of the variable z is generated by the changes of state of the agents due to the herding interaction. The diffusion function of the volatility process is directly proportional to the correspondent diffusion function of the variable z ; therefore the fluctuations of the volatility process are directly correlated to the agents' herding interactions.

The drift function is given by:

$$\tilde{A}(\sigma) = \underbrace{b(\sigma_0 + \sigma)(\varepsilon_1 - \varepsilon_2\sigma)}_{\text{Autonomous behavior}} + \underbrace{2b\sigma(\sigma_0 + \sigma)}_{\text{Interactions}}, \quad (5.30)$$

⁹Here we assume that the drift and diffusion function do not depend explicitly on time.

Given the non linearity of the transformation (5.26), the new drift function is composed of two terms:

- the first one related to the autonomous behavior of the agents, which arises from the first term in eq. (5.28);
- the second one is proportional to the diffusion function and, therefore, depends on the herding interaction among the agents.

Note that the drift is not anymore a linear function of the state variable σ .

The knowledge of the drift and diffusion function provides us with all the information to characterize the stochastic process for volatility. The unconditional distribution and the autocorrelation function of the variable σ can be easily derived as simple transformations of the the analog properties of z .

5.2.2 Connection with the discrete model

We must always have in mind that the diffusion process (4.26) is an approximation to the dynamics of the stochastic *discrete* variable n for a large number of agents, N . In the previous chapter, we have discussed the limitations implied by this approximation to the stochastic variable z . In this paragraph we discuss the connection between the discrete model (4.4) and the dynamics of the volatility σ .

In order do so, let us start from the diffusion function (4.28), which also takes into account the contribution to the fluctuations given by the granularity of the system. We, then, impose that the contribution of the interactions is much larger than the granularity term. We have then the following inequality:

$$2b \frac{\sigma}{(\sigma_0 + \sigma)^2} \gg \frac{a_1}{N} + \frac{a_2 - a_1}{N} \frac{\sigma}{\sigma_0 + \sigma}, \quad (5.31)$$

which is eq. (4.30) expressed in terms of σ . Solving the correspondent inequalities, we end up with:

$$\frac{(\varepsilon_1 - \varepsilon_2)^2}{8N\varepsilon_2} \ll \frac{\sigma}{\sigma_0} \ll \frac{2N}{\varepsilon_2}. \quad (5.32)$$

As we can see from eq. (5.32), for any finite values of ε_1 and ε_2 , the interval of variability of the diffusion approximation can be set arbitrarily close to $(0, \infty)$, increasing the number of agents N .

5.2.3 Dynamical properties of the volatility process

As introduced in the chapter 4, we can express the dynamics of the variable σ as a stochastic diffusion approximation. In order to derive it, we can follow two different approaches: the first one relies on the transformation, via the Ito's lemma, of the stochastic equation governing the diffusion approximation of the variable z , given in eq. (4.26). The second one invokes some kind of equivalence of a stochastic differential equation to the FPE equation (see the books of Gardiner [2003] and Van Kampen [1992]), the former defined by the drift and diffusion functions in

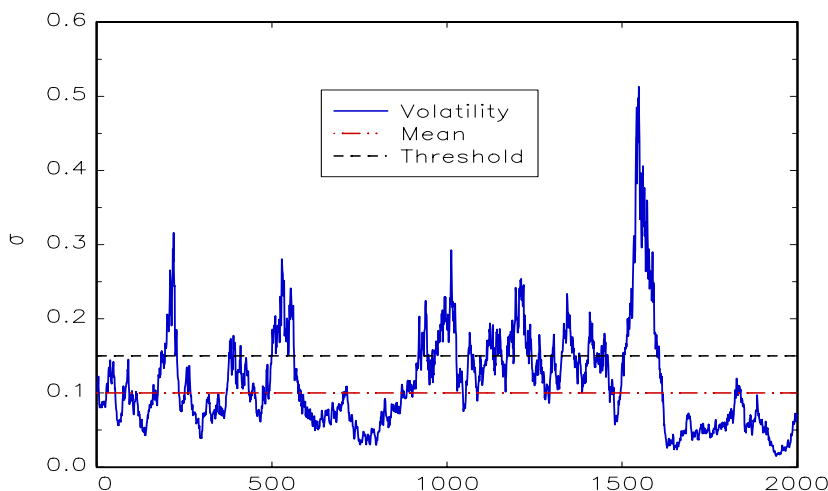


Figure 5.3: A typical time series of the volatility process is shown. The two dashed lines represent the unconditional mean $E[\sigma]$ and threshold value σ_T of the process (5.33). The parameter values are: $\varepsilon_1 = 3$, $\varepsilon_2 = 4$, $b = 0.001$ and $\sigma_0 = 1$. The time series is generated by means of the Euler approximation given by eq. (5.36).

eqs. (5.30) and (5.29). Both approaches lead to the following stochastic differential equation:

$$d\sigma = \sigma_0 \left(1 + \frac{\sigma}{\sigma_0}\right) \left(\varepsilon_1 + (2 - \varepsilon_2) \frac{\sigma}{\sigma_0}\right) b dt + \left(1 + \frac{\sigma}{\sigma_0}\right) \sqrt{2b \frac{\sigma}{\sigma_0}} dw, \quad (5.33)$$

where dw is a Wiener increment.

As we have noticed elsewhere, the drift function is non-linear in σ . Its parabolic shape creates an asymmetry between the values of σ below or above the threshold value σ_T , for which the drift function becomes zero, given by:

$$\sigma_T = \sigma_0 \frac{\varepsilon_1}{\varepsilon_2 - 2}. \quad (5.34)$$

Differently from processes characterized by linear drift functions, where $\sigma_T \equiv E[\sigma]$, we observe that, for the process governed by eq. (5.33), $\sigma_T > E[\sigma]$ ¹⁰, as illustrated in Figure 5.3. Given the non linear behavior of the drift function, it takes more time to converge to σ_T from lower values than from higher values of σ . Therefore, the frequency of lower levels of volatility is greater, which shift the unconditional mean, $E[\sigma]$, below σ_T . In order to show this asymmetry, we study the deterministic skeleton of eq. (5.33). We ‘switch off’ the noise term in eq. (5.33), obtaining:

$$\frac{d\sigma}{dt} = b(\varepsilon_2 - 2) \left(1 + \frac{\sigma}{\sigma_0}\right) (\sigma_T - \sigma). \quad (5.35)$$

¹⁰Note that $E[\sigma] = \sigma_0 \frac{\varepsilon_1}{\varepsilon_2 - 1}$, as detailed in the next paragraph.

In Figure 5.4 the dynamics of $\sigma(t)$ is shown, governed by the deterministic process (5.35), for different starting values σ_i , which are above and below the reference value σ_T . This asymmetric behavior generates very high and sharp peaks of the volatility and long periods of low volatility. Interestingly, the contribution to the drift due to agents' interaction is always positive –see eq. (5.30)– which increases the average value of the volatility. The autonomous component is the responsible of the mean-reverting nature of the volatility process. Once again, we can precisely identify the contributions of the relevant components of the agents' behavior to the aggregate market dynamics.

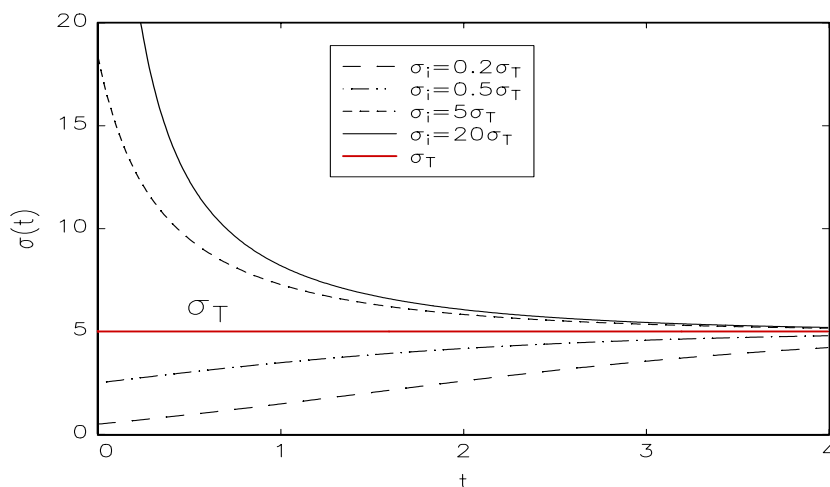


Figure 5.4: The behavior of the solution of the differential equation 5.35 for different starting values. The parameter values are: $\varepsilon_1 = 3$, $\varepsilon_2 = 4$, $b = 0.001$ and $\sigma_0 = 1$.

The discretization of the continuous time equation (5.33), called in the pertinent literature Euler approximation in analogy with the Euler method for solving differential equation, is given by:

$$\Delta\sigma = \sigma_0 \left(1 + \frac{\sigma}{\sigma_0}\right) \left(\varepsilon_1 + (2 - \varepsilon_2)\frac{\sigma}{\sigma_0}\right) b\Delta t + \left(1 + \frac{\sigma}{\sigma_0}\right) \sqrt{2b\Delta t \frac{\sigma}{\sigma_0}} \xi(t), \quad (5.36)$$

where $\xi(t)$ is a standard Normal random variable and Δt is the discrete time-unit. The discrete-time equation (5.36) can easily be used to simulate the process (5.33) in a computer code. Moreover, we have to impose boundaries conditions to prevent the process to go in the region of negative values. In order to do so, we impose the further boundary conditions:

$$\text{if } \sigma(t) < 0 \quad \text{then} \quad \frac{\sigma(t + \Delta t) + \sigma(t)}{2} = 0. \quad (5.37)$$

Note that the conditions (5.37) are given as *additional* external equations. They are

essential in the discrete approximation (5.36), whereas for the continuous equation (5.33) they are built in.

5.2.4 General remarks

Before describing the mathematical properties of the stochastic process (5.33) and its limiting cases, some general remarks should be mentioned. Several stochastic volatility models have been proposed to generalize the Black and Scholes approach for option pricing. They are based on a trade-off between mathematical convenience to have closed-form solutions (or at least some tractable series expansions) and the need to deviate from constant volatility assumption of the geometric Brownian motion for the dynamics of the underlying asset. Typically, the stochastic process that governs the dynamics of the returns capture the GARCH effect and the leptokurtosis of the return distribution. However, not all of the existing models exhibit a Pareto law for the probability of the extreme events (an example is the model proposed by Heston [1993]). More important, all the existing models are based on phenomenological assumptions, without specifying the source or detailing the mechanism that generates the fluctuations of returns. The stochastic process of the underlying asset is, in fact, exogenously given. This characteristic has relevant implications from both fundamental and practical viewpoints. The fundamental aspect of the problem lies in the identification of the source of the fluctuations. Is the source of the fluctuations exogenous (e.g. news) or endogenous (for instance the behavior of the agents)? By which factor is it influenced? From more practical purposes, any hedging strategy has to take into account not only the statistical properties of the source of uncertainty, but also its nature. The volatility models that exist in the econometrics literature can describe reasonably well the statistical properties of financial fluctuations, however, they do not provide an explanation on their origins.

The present work is devoted to bridge the two fields that, from different perspectives, try to describe the stylized facts of financial data. On one hand the financial econometrics and, on the other hand, the agent-based approach. In particular, a direct connection between the stochastic volatility models of Ahn and Gao [1999] and our generalization of the model introduced by Kirman [1993], detailed in the previous chapter, is an extraordinary and surprising result —see section 5.3. The big advantage of our model with respect to the models proposed in the financial econometrics is the ability to explain the origin of the randomness present in the market. It is in fact, very clear how the interactions based on herding among agents play the crucial role in the emergence of the market fluctuations. We can, in fact, precisely identify the source of the aggregate regularities of the returns in terms of the agents behavioral assumptions. Additionally, the stochastic volatility structure together with all the battery of theoretical results allows for a direct estimations of the underlying parameters, which might provide new insights into the data and, indirectly, into the behavior of the traders.

5.2.5 Equilibrium distribution of volatility

To compute the equilibrium distribution of the volatility $\sigma(t)$, denoted by $\rho_e(\sigma)$, we can adopt two different methods: the first is applying eq. (4.34) to the drift and diffusion functions, as given by (5.30) and (5.29); the alternative approach relies on the general formula:

$$\rho_e(\sigma) = \frac{dz}{d\sigma} \cdot p_e(z), \quad (5.38)$$

which gives a direct relation between the equilibrium probability density functions of the two variables, in our case z and σ . In the following, we concentrate on the second approach.

Using the definition of σ as given by eq. (5.26), we can express the variable z as a function of σ , as follows:

$$z = \sigma / (\sigma_0 + \sigma). \quad (5.39)$$

Plugging eq. (5.39) and the expression of the equilibrium distribution $p_e(z)$ from eq. (4.41) into eq. (5.38), and considering that

$$\frac{dz}{d\sigma} = \frac{\sigma_0}{(\sigma_0 + \sigma)^2},$$

we can obtain the equilibrium distribution for the variable σ :

$$\rho_e(\sigma) = \frac{1}{\sigma_0} \frac{1}{B(\varepsilon_1, \varepsilon_2)} \left(\frac{\sigma}{\sigma_0} \right)^{\varepsilon_1 - 1} \left(\frac{\sigma_0}{\sigma + \sigma_0} \right)^{\varepsilon_1 + \varepsilon_2}. \quad (5.40)$$

It can be more conveniently rewritten as:

$$\rho_e(\sigma) = \frac{1}{\sigma_0} \frac{1}{B(\varepsilon_1, \varepsilon_2)} \frac{\sigma_0}{\sigma} \left(\frac{\sigma_0}{\sigma} + 1 \right)^{-\varepsilon_1} \cdot \left(\frac{\sigma}{\sigma_0} + 1 \right)^{-\varepsilon_2}. \quad (5.41)$$

As we can see from the eq. (5.41), the two parameters ε_1 and ε_2 governs different regions of the distribution. The first dominates the behavior of the distribution near the origin, while the second governs the decay of the tail. The factor σ_0 is simply a scale parameter. Not surprisingly, the great flexibility of the Beta distribution, from which the probability density (5.40) is derived, is also preserved. Depending on the choice of the two parameters, in fact, we can have quite different shapes of the equilibrium distribution $\rho_e(\sigma)$:

- monotonic decreasing distribution with a pole at the origin, for $\varepsilon_1 < 1$ and $\varepsilon_2 > 0$;
- monotonic decreasing distribution with a finite value at the origin, for $\varepsilon_1 = 1$ and $\varepsilon_2 > 0$;
- a unimodal distribution, for $\varepsilon_1 > 1$ and $\varepsilon_2 > 0$.

Figure 5.5 compares the three cases.

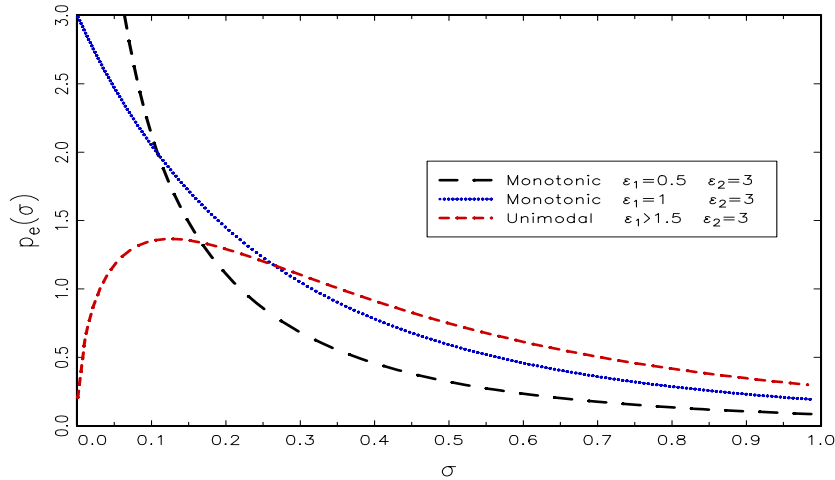


Figure 5.5: The three curves are illustrative examples of different behaviors of the distribution given by eq. (5.40) for different choices of the parameters. In all cases we set $\sigma_0 = 1$.

The region near zero behaves as a power law with an exponent ε_1 . Interestingly, the right tail of the distribution (5.40) also exhibits a power law decay for $\sigma \gg \sigma_0$,

$$\rho_e(\sigma) \sim \left(\frac{1}{\sigma}\right)^{\varepsilon_2+1}. \quad (5.42)$$

As detailed in paragraph 5.4, the power law decay of the the tail, under fairly general conditions, carries over the equilibrium distribution of returns. This property is compatible with the empirically identified power-law tails described in the chapter 2. Moreover, the property of fat tailed return distribution is an essential consequence of our framework¹¹.

Remarks. As it is obvious from eq. (5.42), the exponent of the tail is related to the parameters characterizing the behavior of the fundamentalists, namely the ratio between the tendency of autonomous switches from fundamentalist to technical trader attitude, governed by a_2 , and the herding parameter b . The direct connection between a macroscopic quantity, the index of the tail of the return distribution, and the parameters governing the agents' behavior, is a remarkable result of the thesis.

¹¹In principle it is the volatility process that exhibits the power law decay of the tail. However, under fairly general conditions, this property carries over the distribution of returns. For more details see section 5.4.1

Unconditional moments

As detailed in appendix A 5.3, the q -moment of the equilibrium distribution of the volatility is given by:

$$m_q = \sigma_0^q \frac{B(\varepsilon_1 + q, \varepsilon_2 - q)}{B(\varepsilon_1, \varepsilon_2)} = \frac{\Gamma(\varepsilon_1 + q)}{\Gamma(\varepsilon_1)} \frac{\Gamma(\varepsilon_2 - q)}{\Gamma(\varepsilon_2)}. \quad (5.43)$$

Not surprisingly, the moments diverge¹² for $q \geq \varepsilon_2$, because of the power law decay of the tail (recall the definition of the tail index as the highest finite moment of the distribution). The first two integer moments are:

$$E[\sigma] = \sigma_0 \frac{\varepsilon_1}{\varepsilon_2 - 1}, \quad E[\sigma^2] = \sigma_0^2 \frac{\varepsilon_1(\varepsilon_1 + 1)}{(\varepsilon_2 - 1)(\varepsilon_2 - 2)}. \quad (5.44)$$

To have a finite first and second moment, we have to restrict the value of the parameter ε_2 to the interval $(2, \infty)$. If $\varepsilon_1 > 1$, the distribution will be unimodal, with the mode σ^* located at:

$$\sigma^* = \sigma_0 \frac{\varepsilon_1 - 1}{\varepsilon_2 + 1}. \quad (5.45)$$

5.2.6 Conditional properties of the volatility: the autocorrelation

The knowledge of the FPE of the stochastic variable σ gives us a complete description of the time evolution of the process. In principle, we can compute the conditional distribution of the variable σ , that we denote by $\rho(\sigma, t|\sigma_i)$, given its initial probability density function $p_i(\sigma)$. The line of reasoning would be pretty much the same as for the unconditional equilibrium distribution of σ , as derived from eq. (5.38). We will not proceed in this direction, since the resulting expansion in Jacobi polynomials for $\rho(\sigma, t|\sigma_0)$ cannot be easily handled for an estimation procedure of the underlying parameters of the model, which is our ultimately aim of the present work.

More interesting and useful for the estimation of the underlying parameters is the determination of the auto-covariance of σ , which helps to compute the autocorrelation of returns and their transformations. The computation of this quantity is in principle straightforward using the result derived in chapter 4. Eq. (4.64) gives, in fact, the auto-covariance of an arbitrary variable that is a function of the original stochastic variable z . Unfortunately, the non-linear function (5.26) cannot be expressed as a linear combination of a *finite* number of Jacobi polynomials, which would allow for a closed-form solution for the autocorrelation. Therefore, we have to rely on numerical evaluation of the infinite series (4.64) –the details of the calculation are presented in appendix (A 5.4). We end up with the following expression:

$$E[\sigma_t \cdot \sigma_i] = \sigma_0^2 \sum_{n=0}^{\infty} e^{-b\lambda_n t} F_n(\varepsilon_1, \varepsilon_2), \quad (5.46)$$

where $F_n(\varepsilon_1, \varepsilon_2)$, introduced in the appendix (A 5.4), is an analytical function of the parameters ε_1 and ε_2 . The previous equation helps us to derive the auto-correlation

¹²The gamma function diverges in fact at the origin.

of volatility:

$$C_\sigma(t) = \frac{\sigma_0^2 \sum_{n=0}^{\infty} e^{-b\lambda_n t} F_n(\varepsilon_1, \varepsilon_2) - E[\sigma]^2}{E[\sigma^2] - E[\sigma]^2}. \quad (5.47)$$

Figure 5.6 shows the behavior of the autocorrelation of volatility, computed with 30 terms in the expansion (5.47), as compared with its exponential asymptotic behavior, which includes just the first two terms of the sum in eq. (5.47). For large enough time lags, the autocorrelation is dominated by the first non-zero eigenvalue, which correspond to an exponential decay of the autocorrelation function. In this case, the contribution to the sum of the higher order eigenvalues can be neglected. Note, in fact, that the asymptotic behavior is a good approximation of the autocorrelation for time intervals higher than 1 (measured in natural time, see the caption in Figure 5.6). In the region $t < 1$, we observe a pronounced deviation from the exponential decay.

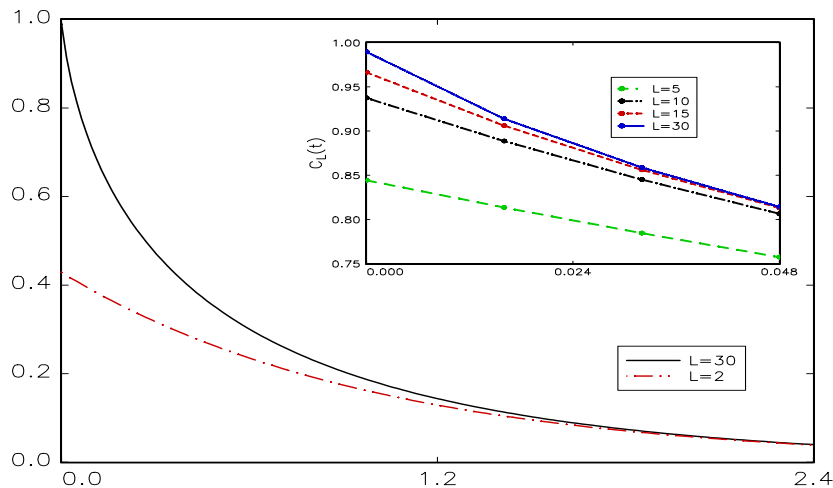


Figure 5.6: The main panel shows the behavior of the autocorrelation (continuous line) as compared to its asymptotic behavior (dashed line). Note that the time unit is expressed in terms of characteristic time τ_c , defined in eq. (5.48). The number of terms in the sum (5.46) is 30, while the asymptotic curve is computed using the first two terms. The inset shows the autocorrelation computed for an increasing number of terms in eq. (5.47). The accuracy reaches a satisfactory level for $L > 15$. The underlying parameters are: $\varepsilon_1 = 5$, $\varepsilon_2 = 3$ and $b = 0.002$ –see appendix (A 5.4) for a more detailed description of the numerical evaluation of eq. (5.47).

Some comments are in order here. The leading exponential decay of the variable σ does not come as a surprise giving its underlying Markovian nature. The inverse of the asymptotic decay rate defines a characteristic relaxation time of the process, τ_c ,

given by:

$$\tau_c = \frac{1}{b(\varepsilon_1 + \varepsilon_2)}. \quad (5.48)$$

Note that, fixing the parameter ε_1 and ε_2 , τ_c can be arbitrarily set to any given positive value, tuning the parameter b . Deviations from the exponential decay arise from the contributions of higher eigenvalues in eq. (5.46), which cannot be neglected for time intervals smaller than τ_c . For smaller time intervals, several time scales appear due to the non-linear transformation (5.26). In this respect, note the difference to the exact exponential decay for all time intervals of the autocorrelation of the variable z , given in eq. (4.63), which, in fact, possesses a linear drift term.

The positive autocorrelation of the volatility σ is an immediate consequence of some degree of persistence of its fluctuations. The volatility process, governed by the drift and diffusion eqs. (5.30) and (5.29), gives rise to the volatility clustering phenomenon (see Figure 5.3), which is a further key-stylized fact of financial data that the model can reproduce. The dependence on the past realizations of the volatility is a direct consequence of the Markovian assumption of the transition rates of the discrete process (4.4). The GARCH effect at the aggregate level, however, is not exogenously imposed to the dynamics of the volatility, but it is an endogenous consequence of the agent-based underlying model. We ‘break’ then the common adage “to get GARCH you need to begin with GARCH”, which is a feature of other contributions in the literature, Pagan [1996].

5.3 Asymptotic behavior of the volatility process: the extreme asymmetric case $\varepsilon_1 \gg \varepsilon_2$

An interesting limiting case for the stochastic process (5.33) originates under the condition

$$\varepsilon_1 \gg \varepsilon_2, \quad (5.49)$$

together with a specific normalization, which fixes the value of σ_0 . Let us denote with ε_1 - ε_2 -model the unrestricted model given by eq. (5.33) and with ε_2 -model the alternative limiting case. We will show in the next chapter that the values of the parameters of the unrestricted model estimated using real data from different markets fall in the case ε_2 -model. Later on we will discuss the connection of the limiting model with the discrete model (4.4), pointing out the implications for the agents’ behavior of the condition $\varepsilon_1 \gg \varepsilon_2$.

5.3.1 The limiting stochastic process

Let us start imposing the following normalization condition to the process (5.33):

$$E[\sigma] = \sigma_0 \frac{\varepsilon_1}{\varepsilon_2 - 1} = 1, \quad (5.50)$$

which is equivalent to fix the scale of the fluctuations of the stochastic variable σ . Under this condition, we can express the scale parameter σ_0 as a function of the other two parameters ε_1 and ε_2 , therefore reducing the total number of parameters.

Under the normalization condition (5.50), the volatility process (5.33) can be rewritten as:

$$d\sigma = b((\varepsilon_2 - 1) + \sigma(2 - \varepsilon_2))dt + \sigma \frac{b \varepsilon_1}{\varepsilon_2 - 1} ((\varepsilon_2 - 1) + \sigma(2 - \varepsilon_2))dt + \sqrt{2\sigma} \sqrt{\frac{b(\varepsilon_2 - 1)}{\varepsilon_1}} \cdot dw + \sigma \sqrt{2\sigma} \sqrt{\frac{b \varepsilon_1}{\varepsilon_2 - 1}} \cdot dw . \quad (5.51)$$

If we assume the following condition:

$$\sigma \frac{\varepsilon_1}{\varepsilon_2 - 1} \gg 1 , \quad (5.52)$$

the first term in the drift and in the diffusion can be neglected. It means that the variable σ should be higher than the ratio $\frac{\varepsilon_2 - 1}{\varepsilon_1}$, which is small given the original assumption $\varepsilon_1 \gg \varepsilon_2$. When ε_1 becomes very large, the characteristic time τ_c , defined in eq. (5.48), becomes very small. A further assumption is that τ_c remains constant and finite when ε_1 increases. For this purpose, we introduce a new parameter which remains finite for very large ε_1 :

$$a = b \cdot \varepsilon_1 . \quad (5.53)$$

The previous condition is equivalent to impose a scaling to the b parameter. Now, τ_c assumes the finite value $\tau_c = 1/a$, independent of ε_1 . We end up with a new stochastic process for the volatility¹³:

$$dv = \frac{\sigma a}{\varepsilon_2 - 1} ((\varepsilon_2 - 1) + v(2 - \varepsilon_2))dt + v \sqrt{\frac{v a}{\varepsilon_2 - 1}} dw . \quad (5.54)$$

In the continuous financial literature, the process (5.54) is called 3/2 process, because of the power of v in the diffusion term. It has been successfully employed by Ahn and Gao [1999] for modeling interest rate dynamics and as a benchmark model for option pricing in a stochastic volatility environment (see the book of Lewis [2000]).

Interestingly, the stochastic process (5.54) can be obtained as a transformation of the following stochastic differential equation¹⁴:

$$du = (\varepsilon_2 - u)a dt + \sqrt{2au} dw' , \quad (5.55)$$

under the simple non-linear transformation of the variable u :

$$v = \frac{\varepsilon_2 - 1}{u} . \quad (5.56)$$

This remarkable equivalence can be easily proved applying the Ito's lemma. The unconditional and conditional properties of the process (5.55) are well known and can be employed to derive the analog properties of the process (5.54).

¹³We denote with v the stochastic variable governed by the ε_2 -model to distinguish it from the variable σ , which follows the ε_1 - ε_2 -model. Nevertheless, both variables govern the volatility process in our artificial financial market model.

¹⁴The stochastic process (5.55) has been introduced by Heston [1993] for option pricing in a stochastic volatility environment. The big advantage of this model is that it has essentially a closed form solution for the option evaluation which is easy to implement (cf. Lewis [2000]).

5.3.2 Conditional and unconditional properties of the process (5.54)

The statistical properties of the process (5.54) can be derived from the analog properties of the process (5.55). Let us start with the analysis of the former. It is known in the financial literature as squared-root process (given the power in the diffusion term, cf. Lewis [2000]), Cox-Ingersoll-Ross or CIR process [Cox et al., 1985], which are the authors that have originally introduced this stochastic differential equation to model interest rate dynamics, or Heston model [Heston, 1993], which has derived a closed-form solution for an European call option using (5.55) as the volatility process of the underlying asset. The statistical properties of the process are well known and described in almost all textbooks on stochastic processes applied to finance. In the following, we briefly list those properties leaving the mathematical details to the pertinent literature (see the book of Lewis [2000] and references therein):

- The conditional distribution $p(u(t)|u(s))$ is a non-central Chi-squared distribution (cf. Ahn and Gao [1999]). It can equivalently be written as a linear combination of Laguerre polynomials (see appendix A 5.5):

$$p(u(t)|u(s)) = \sum_{n=0}^{\infty} a_n e^{-na(t-s)} L_n^{\varepsilon_2}(u) . \quad (5.57)$$

- The unconditional distribution is given by:

$$p_e(u) = \frac{1}{\Gamma(\varepsilon_2)} u^{\varepsilon_2-1} e^{-u} , \quad (5.58)$$

which actually is the Gamma distribution. Note that the distribution of the volatility exhibits an exponential decay of the tail.

- The autocorrelation function is given by:

$$C_u(t) = e^{-at} . \quad (5.59)$$

Given the linear drift function in (5.55), we recover the ‘pure’ exponential autocorrelation function.

Unconditional distribution

We can derive the unconditional distribution for the process (5.54) in several ways: Directly from the stochastic differential equation (5.54) using eq. (4.34) or using the probability transformation of the distribution (5.58) under the change of variable (5.56). Let us derive it as a limiting distribution of eq. (5.40) for ε_1 large. We can rewrite eq. (5.40) as follows:

$$\rho_e(\sigma) = \frac{\Gamma(\varepsilon_1 + \varepsilon_2)}{\Gamma(\varepsilon_1) \varepsilon_1^{\varepsilon_2}} \cdot \frac{(\varepsilon_2 - 1)^{\varepsilon_2}}{\Gamma(\varepsilon_2)} \cdot \frac{1}{\left(1 + \frac{\varepsilon_2 - 1}{\sigma} \frac{1}{\varepsilon_1}\right)^{\varepsilon_1}} \cdot \frac{1}{\left(\sigma + \frac{\varepsilon_2 - 1}{\varepsilon_1}\right)^{\varepsilon_2}} \cdot \frac{1}{\sigma} , \quad (5.60)$$

under the normalization condition (5.52). The first factor converges to 1 for $\varepsilon_1 \gg \varepsilon_2$; the third factor converges to $e^{-\frac{\varepsilon_2+1}{\sigma}}$; the fourth factor becomes $v^{-(\varepsilon_2+1)}$, if $\sigma \gg \frac{\varepsilon_2-1}{\varepsilon_1}$

which is the condition (5.52). Finally, we end up with the following formula:

$$p_e(v) = \frac{(\varepsilon_2 - 1)^{\varepsilon_2}}{\Gamma(\varepsilon_2)} \cdot \frac{1}{v^{\varepsilon_2-1}} e^{-\frac{\varepsilon_2+1}{v}}. \quad (5.61)$$

which is the inverse Gamma distribution. Note that the right tail decays as a power law, with the same exponent as the distribution (5.40), as it should be expected. Figure 5.7 shows the convergence of the distribution (5.40) to its limiting distribution (5.61) for increasing values of ε_1 .

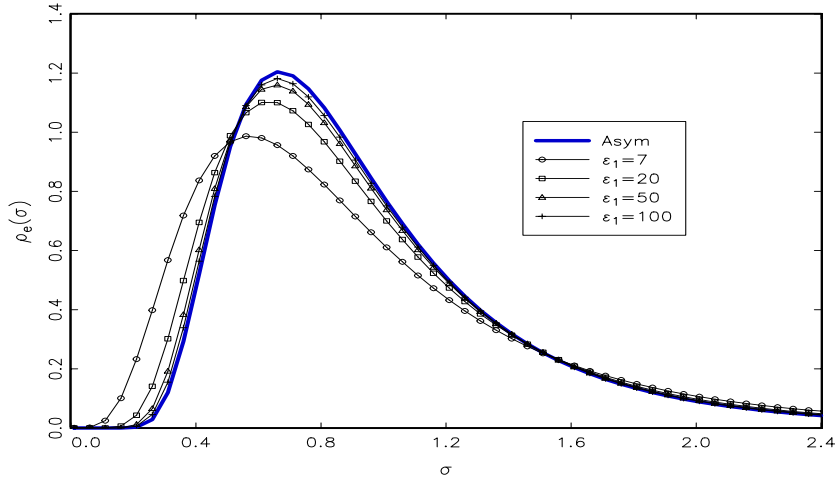


Figure 5.7: Convergence of the unconditional distribution of the ε_1 - ε_2 -model to its asymptotic functional form, for increasing values of ε_1 . Parameter value: $\varepsilon_2 = 5$.

Unconditional moments

The unconditional moments are given by:

$$m_q = (\varepsilon_2 - 1)^q \frac{\Gamma(\varepsilon_2 - q)}{\Gamma(\varepsilon_2)}, \quad (5.62)$$

which can be derived as the asymptotic behavior of eq. (5.43) for ε_1 very large, or directly from the probability distribution (5.61). The first and the second moments are given by:

$$m_1 = 1 \quad \text{and} \quad m_2 = 1 + \frac{1}{\varepsilon_2 - 2}. \quad (5.63)$$

Note that the mean is 1 since the normalization condition (5.50). The variance diverges for $\varepsilon_2 = 2$.

Autocorrelation function

The autocorrelation function of the process (5.54) can be calculated using the general formula (4.64) and the properties of the stochastic process (5.55) –see appendix A 5.6. The autocorrelation is given by:

$$C(t) = (\varepsilon_2 - 2) ({}_2F_1(1, 1; \varepsilon_2; e^{-at}) - 1) , \quad (5.64)$$

where the hypergeometric function is given by the following series representation

$${}_2F_1(1, 1; \varepsilon_2; z) = \sum_{n=0}^{\infty} n! \frac{\Gamma(\varepsilon_2)}{\Gamma(\varepsilon_2 + n)} z^n . \quad (5.65)$$

The previous formula holds under the condition $\varepsilon_2 > 2$, since for lower values the variance is not defined. Figure 5.8 shows the autocorrelation function as given by eq. (5.64) compared to its asymptotic behavior, which includes just the first two terms in the series (5.65). Note the deviation from the exponential decay for time interval $t < \tau_c = 1/a$, where the contribution of the higher terms in the expansion (5.65) is relevant.

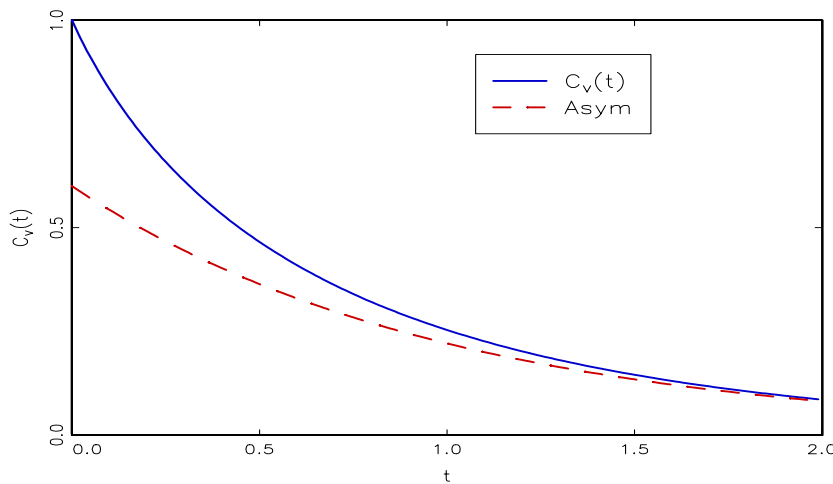


Figure 5.8: *Autocorrelation function of the ε_2 -model compared to its exponential asymptotic behavior (natural time unit). The series (5.65) is computed using the first 100 terms. The parameter values are: $\varepsilon_2 = 5$ and $a = 0.01$.*

5.3.3 The connection with the discrete model

In this paragraph we want to analysis the connection between the extreme asymmetric model in connection with the discrete model (4.4). We have already discussed the connection of the ε_1 - ε_2 -model with its agent-based counterpart in section 4.1.2, providing the interval of validity for the diffusion approximation as a function of the

underlying parameters ε_1 , ε_2 and the number of agents N . The asymptotic case exhibits an additional layer of complexity with respect to the previous case, since both ε_1 and N have to be very large'. In order to preserve the diffusion approximation we have to set a relation between ε_1 and N . Under the condition $\varepsilon_1 \gg \varepsilon_2$, the mean value of the variable z will be close to the right edge of the interval $[0, 1]$. However, it cannot enter in the region $1 - \varepsilon_2/N$, which is the limit of validity of the diffusion approximation. Therefore, the following inequality must be satisfied:

$$\frac{1 - E[z]}{1 - (1 - \frac{\varepsilon_2}{N})} \gg 1, \quad (5.66)$$

which implies that $N \gg \varepsilon_1$.

Under the $\varepsilon_1 \gg \varepsilon_2$ condition, the large majority of traders will belong to the technical group. We denote, in fact, this scenario as an extremely asymmetric case.

5.4 The returns process

In the previous paragraphs, we have analyzed in details the volatility process (5.25) and the asymptotic case $\varepsilon_1 \gg \varepsilon_2$. In order to be able to compare the model with real data, we have to specify the unconditional distribution of the noise term in eq. (5.25). Since the noise term is not an observable, we might choose various parametric distributions.

5.4.1 Unconditional distribution of returns

Given the unconditional distribution of the volatility from eq. (5.40), we can compute the unconditional returns distribution $p_e(r)$. In order to do so, we can apply the following formula:

$$p_e(r) = \int_0^\infty p(r|\sigma)\rho_e(\sigma)d\sigma, \quad (5.67)$$

where $p(r|\sigma)$ is the conditional distribution of r given the value of the volatility, which basically is the unconditional distribution of η . Obviously, the existence of a closed-form solution for the integral (5.67) depends crucially on the parametric functional form of the noise term in eq. (5.25). Eventually it is possible to compute numerically the value of the integral. However, we can derive some general properties of the distribution $p_e(r)$ that are to a large extent independent of the particular functional form of $p(\eta)$. If we assume that $p(\eta)$ is symmetric around its median 0, this *symmetry* carries over the distribution $p_e(r)$. If we, then, assume *continuity* of $p(\eta)$, the same holds for the returns distribution.

Interestingly, taking into account the independence of the two variables σ and η in eq. (5.25) and the underlying symmetry of the distribution $p(\eta)$, the moments of the absolute returns distribution are given by:

$$m_q(|r|) = m_q(\sigma) \cdot m_q(|\eta|). \quad (5.68)$$

Therefore, assuming that all moments of $p(\eta)$ exist, the finite highest moment for the returns distribution is ε_2 . It means that the power law character of the volatility

carries over the distribution of returns.¹⁵ We have already computed the kurtosis of the returns distribution in eq. (5.21) under the hypothesis of Normally distribute noise term η , showing its leptokurtic character.

All in all, for a whole class of noise term distributions $p(\eta)$, $p_e(r)$ is a *continuous, symmetric and leptokurtic distribution, with a power law decay of the tail with an exponent ε_2* .

Our parametric choice for the distribution $p(\eta)$ is constrained by the possibility to estimate the underlying parameters of the model, which is one of the main goals of the present research.

5.4.2 Conditional properties of the returns

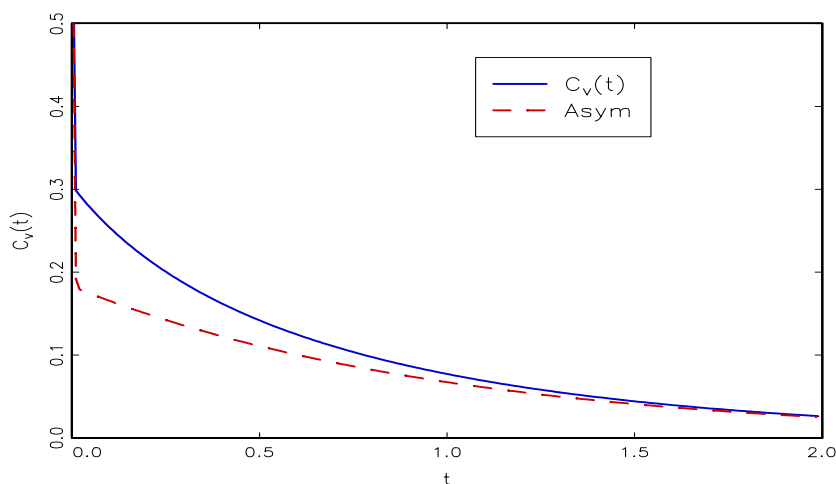


Figure 5.9: *The autocorrelation function of absolute returns with a Gaussian noise term η for the ε_2 -model, compared to its asymptotic behavior. The parameter values are: $\varepsilon_2 = 5$ and $a = 0.01$. The theoretical drop factor is $D_{|r|} = \frac{1}{2\pi-3} = 0.305$.*

In the previous section we have focused attention on the unconditional properties of returns distribution. In this section we concentrate on the autocorrelation function of returns and their simple transformations.

The autocorrelation of returns is identically zero for every time interval:

$$C_r(r_{t+\tau}r_t) = 0 \quad \text{for every } \tau > 0 . \tag{5.69}$$

¹⁵In order to have a power law decay with an index ε_2 for the return distribution, the restrictive condition of the existence of all moments for the distribution $p(\eta)$ can be replaced by the more general condition of existence of moments of order higher than ε_2 .

This property is derived from the iid-ness of the noise term in eq. (5.25) and its independence on the volatility process. Eq. (5.69) squares well with the behavior of real returns, at least for lower frequencies (daily data for instance). From an economic perspective, the vanishing autocorrelation is directly related to the (linear) unpredictability of financial returns. Note that the property (5.69) is essentially independent of the functional form of the distribution $p(\eta)$.

We can compute the autocorrelation function for the absolute returns, which we will use to estimate the parameters of the model. The autocorrelation of absolute returns is given by:

$$C_{|r|}(t) = \frac{\left(E[\sigma_t \sigma_0] - E[\sigma]^2\right) E[|\eta|^2]}{E[\sigma^2] E[\eta^2] - E[\sigma]^2 E[|\eta|^2]}, \quad (5.70)$$

where we have used the independence of the two variables σ and η . It can be rewritten as:

$$C_{|r|}(t) = \frac{\left(\frac{E[\sigma_t \sigma_0]}{E[\sigma]^2} - 1\right)}{\frac{E[\sigma^2]}{E[\sigma]^2} \frac{E[\eta^2]}{E[|\eta|^2]} - 1}, \quad (5.71)$$

which depends on the distribution of η via the ratio of its two particular moments $\frac{E[\eta^2]}{E[|\eta|^2]}$.

In Figure 5.9 we show a representative example of the autocorrelation of returns with the volatility process governed by the ε_2 -model in the case of Gaussian noise η . Note that the decay rate of the autocorrelation of absolute returns is governed solely by the volatility process. The abrupt jump at the first time-lag is related to the presence of the iid-noise term, whose amplitude is given by:

$$C_{|r|}(1) \equiv D_{|r|} = \frac{\gamma_\sigma - 1}{\gamma_\sigma \gamma_\eta - 1}, \quad (5.72)$$

where

$$\gamma_\sigma = \frac{m_2(\sigma)}{m_1^2(\sigma)} \quad \text{and} \quad \gamma_\eta = \frac{m_2(\eta)}{m_1^2(|\eta|)}. \quad (5.73)$$

5.5 Conclusions

In this chapter we have shown how the herding model, detailed in the previous chapter, can be integrated into an artificial financial market. We have illustrated by analytical techniques that our model can reproduce the key stylized facts of financial time series. The main result of our contribution is not only in the good qualitative agreement of the model with real data but, more important, the direct connection of the aggregate agents' dynamics with the stochastic volatility framework typically employed in financial econometrics. The important advantage of our model in comparison with the econometric approach is the precise behavioral identification of the determinants for the emergence of the market fluctuations. We have shown that the social interactions among agents are responsible for the fluctuations, which are, therefore, an endemic feature of financial markets.

Appendix to Chapter 5

A 5.1 Analytical moments of the process (5.10)

Let us recall the equation of the motion of the returns as given by eq. (5.10):

$$r(t, \Delta t) = \sigma(t + \Delta t) \xi(t + \Delta t) - \sigma(t) \xi(t), \quad (\text{A } 5.1)$$

where ξ_t is given by an AR(1) stochastic process:

$$\xi(t + \Delta t) = a\xi(t) + \eta(t). \quad (\text{A } 5.2)$$

Recall that $a \approx 1$, $E[\eta] = 0$ and $E[\eta^2] = \sigma_0^2$. Plugging eq. (A 5.2) into eq. (A 5.1), it is easy to arrive at the following expression:

$$E[r^2] = E[\sigma^2] \sigma_0^2 + E[\xi^2] E[\sigma^2] (a - 1)^2, \quad (\text{A } 5.3)$$

where we have used the approximation $E[\sigma_{t+\Delta t} \cdot \sigma_t] \approx E[\sigma^2]$. Considering that $E[\xi^2] = \frac{\sigma_0^2}{1-a^2}$ for an AR(1) process, we arrive to eq. (5.17) in the main text. In a similar way it is possible to derive eq. (5.18).

Eq. (5.19) can be derived taking into account the following expressions:

$$E[\xi_{t+2\Delta t} \cdot \xi_t] = a^2 E[\xi^2], \quad (\text{A } 5.4)$$

$$E[\xi_{t+2\Delta t} \cdot \xi_{t+\Delta t}] = a E[\xi^2], \quad (\text{A } 5.5)$$

$$E[\sigma_{t+2\Delta t} \cdot \sigma_{t+\Delta t}] \approx E[\sigma_{t+\Delta t} \cdot \sigma_t] \approx E[\sigma_t^2]. \quad (\text{A } 5.6)$$

A 5.2 Fokker-Planck equation for the volatility process (5.25)

In order to derive the Fokker-Planck equation for the new variable σ , we apply the transformations (5.28) and (5.27). Using the following results:

$$z = \frac{\sigma}{\sigma_0 + \sigma}, \quad 1 - z = \frac{\sigma_0}{\sigma_0 + \sigma} \quad (\text{A } 5.7)$$

$$J \equiv \frac{dz}{d\sigma} = \frac{\sigma_0}{(\sigma_0 + \sigma)^2}, \quad \left| \frac{J'}{J^3} \right| = \frac{2(\sigma_0 + \sigma)^3}{\sigma_0^2}, \quad (\text{A } 5.8)$$

$$A(\sigma) = b \left(\varepsilon_1 + \varepsilon_2 \frac{\sigma}{\sigma_0} \right) (\sigma_0 + \sigma), \quad D(\sigma) = b \frac{\sigma_0 \sigma}{(\sigma_0 + \sigma)^2}, \quad (\text{A } 5.9)$$

we arrive at eqs. (5.30) and (5.29) for $\tilde{A}(\sigma)$ and $\tilde{D}(\sigma)$.

A 5.3 Unconditional moments of the volatility (5.43)

Let us start from the definition of the q -th moment of the distribution (5.40):

$$m_q = \frac{1}{\sigma_0} \frac{1}{B(\varepsilon_1, \varepsilon_2)} \int_0^\infty \sigma^q \left(\frac{\sigma}{\sigma_0} \right)^{\varepsilon_1 - 1} \left(\frac{\sigma_0}{\sigma + \sigma_0} \right)^{\varepsilon_1 + \varepsilon_2} d\sigma. \quad (\text{A } 5.10)$$

Under a change of variable $\sigma = \sigma_0 \frac{z}{1-z}$, the integral (A 5.10) becomes:

$$m_q = \sigma_0^q \frac{1}{B(\varepsilon_1, \varepsilon_2)} \int_0^1 z^{\varepsilon_1 - 1 + q} (1 - z)^{\varepsilon_2 - 1 - q} dz. \quad (\text{A } 5.11)$$

Recalling the definition of the Beta function from eq. (4.37), we end up with eq. (5.43) in the main text.

A 5.4 Auto-covariance of the volatility, eq. (5.46)

In order to derive the auto-covariance, we need to compute the following integral:

$$I_n(\varepsilon_1, \varepsilon_2) = \int_0^1 f(z) P_n(z) p_e(z) dz, \quad (\text{A 5.12})$$

where $f(z) = \sigma_0 \frac{z}{1-z}$ (without loss of generality, we posit $\sigma_0 = 1$). Plugging the formula of the Jacobi polynomial of order n from eq. (A 4.25), the equilibrium distribution from eq. (5.40) and using the definition of the Beta function in eq. (4.37), the integral (A 5.12) takes the form:

$$I_n(\varepsilon_1, \varepsilon_2) = \frac{\Gamma(\varepsilon_1 + \varepsilon_2)}{\Gamma(\varepsilon_1)\Gamma(\varepsilon_2)} \cdot \frac{\Gamma(\varepsilon_1 + n)\Gamma(\varepsilon_2 + n)}{n! \Gamma(\varepsilon_1 + \varepsilon_2 + n)} T_n(\varepsilon_1, \varepsilon_2), \quad (\text{A 5.13})$$

where $T_n(\varepsilon_1, \varepsilon_2)$ is given by:

$$T_n(\varepsilon_1, \varepsilon_2) = \sum_{i=0}^n (-1)^i \binom{n}{i} \frac{n + \varepsilon_1 - i}{\varepsilon_2 + i - 1}. \quad (\text{A 5.14})$$

The previous sum belongs to the large class of *binomial sums*¹⁶, for which closed-form solutions exist in only a few cases. The sum (A 5.14) can be simplified, arriving at the alternative formulation:

$$T_n = (\varepsilon_1 + \varepsilon_2 + n - 1) S_n(\varepsilon_2) - \delta_{n,0}, \quad (\text{A 5.15})$$

where $S_n(\varepsilon_2)$ is given by:

$$S_n(\varepsilon_2) = \sum_{i=0}^n (-1)^i \binom{n}{i} \frac{1}{\varepsilon_2 + i - 1}. \quad (\text{A 5.16})$$

The previous sum can be expressed in a closed form. Let us start imposing the following normalization:

$$S_n(\varepsilon_2) = \frac{1}{\varepsilon_2 - 1} \sum_{i=0}^n (-1)^i \binom{n}{i} \frac{\varepsilon_2 - 1}{\varepsilon_2 + i - 1}, \quad (\text{A 5.17})$$

in such a way that the first term of the sum is 1. If we label by a_i the i -th term of the sum (A 5.17), the ratio of two consecutive terms is given by:

$$\frac{a_{i+1}}{a_i} = \frac{(i + \varepsilon_2 - 1)(i - n)}{(i + \varepsilon_2)(i + 1)}, \quad (\text{A 5.18})$$

which is a ratio of two polynomials. Therefore, the sum (A 5.17) can be expressed as a hypergeometric function¹⁷:

$$S_n(\varepsilon_2) = \frac{1}{\varepsilon_2 - 1} {}_2F_1(\varepsilon_2 - 1, -n; \varepsilon_2; 1). \quad (\text{A 5.20})$$

¹⁶Binomial sums are defined as $\sum_{i=0}^n \binom{n}{i} a_i$.

¹⁷The hypergeometric functions are defined as follows (see the book of Arfken [1985]):

$${}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\Gamma(a+n)}{\Gamma(a)} \frac{\Gamma(b+n)}{\Gamma(b)} \frac{\Gamma(c)}{\Gamma(c+n)} z^n. \quad (\text{A 5.19})$$

Applying the Gauss's hypergeometric theorem¹⁸, we end up with:

$$S_n(\varepsilon_2) = \frac{\Gamma(\varepsilon_2 - 1) n!}{\Gamma(\varepsilon_2 + n)}. \quad (\text{A } 5.22)$$

Proceeding in the derivation of the auto-covariance, we have to evaluate the factor:

$$F_n(\varepsilon_1, \varepsilon_2) = \frac{1}{h_n} I_n(\varepsilon_1, \varepsilon_2)^2, \quad (\text{A } 5.23)$$

where h_n is given by eq. (5.50). For $n > 0$, we obtain:

$$F_n(\varepsilon_1, \varepsilon_2) = \frac{\Gamma(\varepsilon_2 - 1)^2 \Gamma(\varepsilon_1 + \varepsilon_2) \Gamma(\varepsilon_1 + n) n!}{\Gamma(\varepsilon_1) \Gamma(\varepsilon_2) \Gamma(\varepsilon_2 + n) \Gamma(\varepsilon_1 + \varepsilon_2 + n - 1)} (\varepsilon_1 + \varepsilon_2 + 2n - 1). \quad (\text{A } 5.24)$$

For $n = 0$, we have:

$$F_0(\varepsilon_1, \varepsilon_2) = (\varepsilon_1 + \varepsilon_2 - 1)^2. \quad (\text{A } 5.25)$$

From the definition of the auto-covariance function, we obtain:

$$E[\sigma^2] = \sum_{n=0}^{\infty} F_n(\varepsilon_1, \varepsilon_2). \quad (\text{A } 5.26)$$

We know that the second moment converges to a finite value if $\varepsilon_2 > 2$, and obviously the series in eq. (A 5.26). For $\varepsilon_2 \leq 2$, the series does not converge and the second moment is not defined. In order to show this, let us evaluate the asymptotic behavior of the term of the series (A 5.26), which is straightforward using Stirling's approximation. We end up with:

$$F_n(\varepsilon_1, \varepsilon_2) \sim \frac{1}{n^{2\varepsilon_2 - 3}}. \quad (\text{A } 5.27)$$

For $\varepsilon_2 \leq 2$, the asymptotic term decays as $1/n^\alpha$, with $\alpha \leq 1$. For $\varepsilon_2 > 2$, the asymptotic term decays as $1/n^\alpha$, with $\alpha > 1$. The term $E[\sigma_t \sigma_0] = \sum_{n=0}^{\infty} F_n e^{-\lambda_n b t}$ converges for $t > 0$, $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$.

The numerical evaluation of the auto-covariance does not present major problems. The ratio of Gamma functions in eq. (A 5.24) can be easily computed recursively. Obviously, the series (5.46) has to be truncated to include the first N_0 terms, without virtually any sensitive constraint (see Figure 5.6). Finally, we can compute the moments $E[\sigma_t^q \cdot \sigma_0^q]$ in terms of hypergeometric functions. However, the expression will be much more complicated to handle numerically. The dependence on the parameter b is anyway captured by the auto-covariance, which is sufficient for our estimation purposes.

¹⁸The Gauss's hypergeometric theorem states that:

$${}_2F_1(a, b; c; 1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}. \quad (\text{A } 5.21)$$

A 5.5 Laguerre polynomials

The Laguerre polynomials¹⁹, denoted as $L_n^\alpha(x)$, are orthogonal polynomials with respect to the weight function:

$$w(x) = x^\alpha e^{-x}, \quad (\text{A 5.28})$$

for $\alpha > -1$. They satisfy the orthogonality condition:

$$\int_0^\infty w(x) L_m^\alpha(x) L_n^\alpha(x) dx = \frac{\Gamma(n + \alpha + 1)}{n!} \delta_{n,m}. \quad (\text{A 5.29})$$

The Laguerre polynomials arise when solving the stochastic differential equation (5.54). Starting from eq. (5.54), we can write the associated Fokker-Plank equation and solve it pretty much in the same way as we did in chapter 4 for the equation (4.26) (namely, separation of the time variable from the ‘space’ variable, imposing the reflecting boundaries condition which generate discrete eigenvalues, solving the correspondent differential equation, which turns out to be the Laguerre differential equation). Without entering into the details of the mathematical derivation, we must modified the previous relation to adapt it to our notation. The Laguerre polynomial of order n has the following explicit representation:

$$L_n^{\varepsilon_2}(u) = \sum_{i=0}^n (-1)^i \binom{n + \varepsilon_2 - 2}{n - i} \frac{1}{i!} u^i, \quad (\text{A 5.30})$$

where we explicitly take into account the dependence²⁰ on ε_2 . The orthogonality condition is, then modified as:

$$\int_0^\infty p_e(u) L_m^{\varepsilon_2}(u) L_n^{\varepsilon_2}(u) du = k_n \delta_{n,m}, \quad (\text{A 5.31})$$

where k_n is given by:

$$k_n = \frac{\Gamma(n + \varepsilon_2)}{\Gamma(\varepsilon_2) n!}. \quad (\text{A 5.32})$$

The extra factor $\Gamma(\varepsilon_2)$ comes from the equilibrium distribution $p_e(u)$, which is different from the classical weight function $w(x)$ defined previously.

A 5.6 Autocorrelation (5.64)

To compute the autocorrelation of the stochastic process (5.55), we can use the general formula (4.64) replacing the Jacobi with the Laguerre polynomials, inserting the correspondent equilibrium distribution.

¹⁹The following orthogonal polynomials are denoted in the pertinent literature generalized Laguerre polynomials, to distinguish them from the Laguerre polynomials, which refer to an exponential weight function, $w(x) = e^{-x}$. We use the notation Laguerre polynomials to denote the generalized polynomials since there will be no ambiguity in the text.

²⁰The relation between the parameter ε_2 and the coefficient α , typically used in the literature, is $\alpha = \varepsilon_2 - 1$.

We have first to evaluate the integral:

$$F_n(\varepsilon_2) = \int_0^\infty \frac{1}{u} p_e(u) L_n^{\varepsilon_2}(u) = \frac{\Gamma(\varepsilon_2 - 1)}{\Gamma(\varepsilon_2)}, \quad (\text{A } 5.33)$$

which can be computed using the explicit representation of the Laguerre polynomials (A 5.30) and some elementary algebra. Interestingly, the integral is constant and independent on n . The auto-covariance is given by the general formula:

$$E[v_t \cdot v_0] \equiv \sum_{n=0}^{\infty} \frac{F_n^2}{k_n} e^{-n \cdot at}, \quad (\text{A } 5.34)$$

which can be rewritten as:

$$E[v_t \cdot v_0] = \left(\frac{\Gamma(\varepsilon_2 - 1)}{\Gamma(\varepsilon_2)} \right)^2 \sum_{n=0}^{\infty} \frac{n! n! \Gamma(\varepsilon_2)}{n! \Gamma(n + \varepsilon_2)} = \left(\frac{\Gamma(\varepsilon_2 - 1)}{\Gamma(\varepsilon_2)} \right)^2 {}_2F_1(1, 1, \varepsilon_2, z), \quad (\text{A } 5.35)$$

where $z = e^{-at}$ and ${}_2F_1(\cdot)$ is the hypergeometric function. Inserting the values of the first and second moment into the definition of the autocorrelation function, we easily obtain eq. (5.64). The inequality $\varepsilon_2 > 2$ arises as a condition for the existence of the second moment.

CHAPTER 6

Estimation of the Parameters of an Agent-Based Model with Asymmetric Herding

In order to estimate the underlying parameters of the model, we can apply various methodologies, with different degrees of efficiency and complexity. In principle, the expansion of the conditional distribution $p(\sigma|\sigma_0)$ in Jacobi polynomials can be employed for the Maximum Likelihood method, which notoriously is the most efficient method, at least asymptotically. However, the evaluation of the infinite series of eq. (4.55) leads to very complex numerical problems. The literature on estimation of the stochastic volatility models, which our model belongs to, counts a large number of contributions and it is still rapidly growing. The accessibility of cheap computer power makes available a whole set of numerical procedures, which could not be affordable a few years ago. The non-feasible ML method is typically replaced by methodologies which rely on simulations of the non-analytically feasible likelihood. The Markov Chain Monte Carlo methods (cf. Eraker [2001]), particle filters (cf. Doucet et al. [2001]) and indirect estimation methods based on ‘latent’ stochastic processes (cf. Gallant and Tauchen [1996]) are some examples of new and complex estimation procedures, which are computationally very intensive.

Therefore, rather than implementing those very sophisticated techniques for the estimation of the parameters of the model, our goal is mainly concerned to find a trade off between simplicity of the considered method and ‘reliability’ of the results. At this stage of the research, our estimation exercise does not aim to quantify in the most efficient way the values of all parameters, but rather to assess or ‘guestimate’ their range of variability, and to understand to which extent the model describes the empirical data.

6.1 Estimation of the parameters of the model

For the comparison of the model with the empirical data, we have to estimate the whole set of its underlying parameters, namely the scale parameter σ_0 , the ratios $\varepsilon_{1,2}$ and the scale of the two constants a_1 and a_2 or equivalently the parameter b . The major complication in the estimation procedure is related to the occurrence of two sources of randomness in the dynamics of returns in eq. (5.25), namely the stochastic change of σ and the contemporaneous presence of the noise term η , whereas only one measurement (the value of the return) is available at each time step.

Among the large spectrum of possible techniques, we have chosen the ‘equilibrium’ maximum likelihood (EML hereafter). We can use, in fact, the observed sample of returns to estimate the parameters σ_0 and $\varepsilon_{1,2}$ by maximum likelihood fit to the analytically known (or numerically computed) unconditional distribution of the absolute returns. The EML, however, is only an approximation of the “true” likelihood, since we are not considering any information on the time evolution of the volatility process (5.26). We pretend, in fact, that the realizations of the Markovian process (5.26) are independent and identically distributed, according to the unconditional distributions given by eq. (5.41) or eq. (5.61). The advantage of using this approximation is the simplicity of its implementation and the reduced computational burden with respect to the mentioned more elaborated methods. On the other hand, only parameters from the unconditional distribution can be estimated. We will not estimate the parameter b , which, in fact, does not appear in the expression of the unconditional distribution. The method gives asymptotically consistent estimates as the sample size $T \rightarrow \infty$, if the sampling frequency $\Delta = \Delta_T \rightarrow 0$ and the increase of T is faster than the decrease of Δ_T , i.e. $T\Delta_T \rightarrow \infty$. The estimates are asymptotically normally distributed, under the additional condition $T\Delta_T^2 \rightarrow 0$ (cf. Genon-Catalot et al. [1999]). For an application of this approximation to stochastic volatility models see Genon-Catalot et al. [1999].

Turning now to the estimation of the parameters, out of the three parameters that enter in the unconditional distribution, namely ε_1 , ε_2 and σ_0 , the later can be expressed in terms of the other two parameters by imposing the following normalization on the empirical data¹:

$$E[|r|] = 1. \tag{6.1}$$

The normalized absolute returns will be denoted by the variable V :

$$V = \frac{|r|}{E[|r|]}. \tag{6.2}$$

Due to the relation $E[\sigma] = \varepsilon_1/(\varepsilon_2 - 1)$, equation (6.1) implies the following value of σ_0 , which defines the scale of the fluctuations:

$$\sigma_0 = \frac{\varepsilon_2 - 1}{\varepsilon_1} \cdot \frac{1}{E[|\eta|]}. \tag{6.3}$$

¹The normalization is based on the implicit assumption of existence of the population mean. Therefore eq. (6.3) holds under the condition $\varepsilon_2 > 1$, which guarantees the existence of the mean of the process (5.26) (note that ε_2 governs the asymptotic behavior of the tail and, consequently, the existence of the moments).

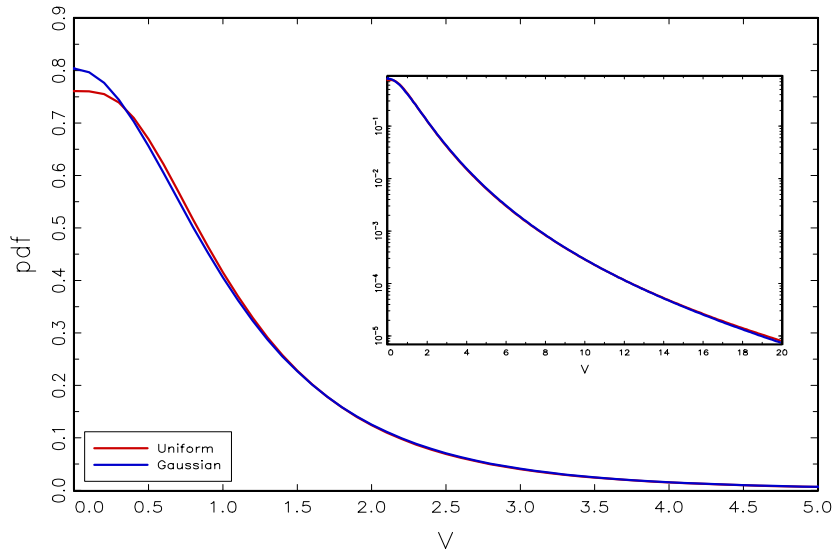


Figure 6.1: The figure compares the probability density functions of the $\varepsilon_1\varepsilon_2$ -model with the uniform and Gaussian noise term. The parameters for the uniform-noise-model are $\varepsilon_1 = 5.6$ and $\varepsilon_2 = 5$, for the Gaussian-noise-model $\varepsilon_1 = 100$ and $\varepsilon_2 = 5$. For this choice of ε_1 both densities have the same first and second moment. The inset shows the pdf in log-linear scale for a better visualization of the tail region. Note the almost perfect correspondence between the two curves.

In order to fully identify the stochastic process that governs the dynamics of the returns from eq. (5.25), we have to precisely specify the distribution of the noise term η . In the following we concentrate on two choices: the uniform and the Gaussian distributions. The uniform distribution has the advantage to allow for a closed-form solution for the return distribution for both the $\varepsilon_1\varepsilon_2$ -model and its asymptotic variant.² The alternative choice, namely a Gaussian distributed noise term, has been extensively used in the literature of stochastic volatility models (for instance the original version of GARCH has conditional Normally distributed increments). The two main reasons for its wide use are related to the property of invariance under ag-

²The uniform distribution is not the only functional form of the noise term that guarantees a closed-form solution for the unconditional distribution of returns. A probability density of the form:

$$p(\eta) = K \sum_{q=0}^n a_q \eta^q, \quad (6.4)$$

leads to a closed form solution as well. Essentially, it is a linear combination of a finite number of power of the variable η . The functional form of the solution, however, exhibits an increasing complexity as soon as the terms of the series (6.4) are added. The uniform distribution leads to the simplest functional form for the distribution of returns. It should also be mentioned that the series (6.4) might be not the most general expression for a closed-form solution of $p(V)$.

gregation of the Gaussian distribution and to its connection with the Central Limit Theorem. In our model, the noise η represents the increment of the stochastic process which governs the dynamics of the sentiment of the technical traders' group. The Gaussian distribution of the increments is, then, the natural choice if we assume that the 'sentiment' of the group is composed by many independent shocks, whose aggregate value converges to normally distributed variable. Following this line of reasoning, the uniform distribution would imply some sort of correlation among the technical traders, which is not formalized in the present version of the model.

The computation of the unconditional distribution for absolute returns in the case of the uniformly distributed random term η is given in the appendix of this chapter, for both $\varepsilon_1\varepsilon_2$ -model and for the ε_2 -model. The unconditional distribution for the case of conditionally Gaussian increments is computed by numerical integration, performed with the method of Newton with constant step size. Note that the EML for all the models has been performed under the restriction $\varepsilon_1 \leq 200$.

In Figure (6.1) we compare the probability density functions for the $\varepsilon_1\varepsilon_2$ -model in the case of Gaussian and uniform noise term. Figure (6.2) shows an analogous comparison of the ε_2 -model with the two alternative noise terms. To eliminate the influence of the scale, all the figures are drawn under the condition $E[V] = 1$. For the $\varepsilon_1\varepsilon_2$ -model the value of ε_1 has been chosen such that the second moment is the same in both cases, namely uniform and Gaussian. The behavior for $V > E[v]$ is dominated by the asymptotic eq. (5.42) and therefore very similar for the two different noises. The distributions mainly differ for small V , i.e. in the region $V < E[v]$.

6.2 Discussion of the results

	Data Set	$\hat{\varepsilon}_1$	$\hat{\varepsilon}_2$	$-\ln L_{\varepsilon_1, \varepsilon_2}$	$\hat{\varepsilon}_2$	$-\ln L_{\varepsilon_2}$	p-value
Unif.	Gold	3.2 ± 0.4	3.9 ± 0.4	4942.3	2.41 ± 0.06	4979.0	0.00
	DB	5.9 ± 1.0	4.4 ± 0.4	6688.8	3.13 ± 0.08	6704.8	0.00
	Siemens	6.4 ± 1.2	4.0 ± 0.3	6588.0	2.97 ± 0.07	6601.9	0.00
	DAX	16 ± 5	4.9 ± 0.3	9533.4	4.2 ± 0.1	9537.5	0.00
Gauss.	Gold	6.8 ± 2.2	4.1 ± 0.5	4942.2	3.8 ± 0.09	4960.0	0.00
	DB	160 ± 28	4.6 ± 0.2	6686.2	4.5 ± 0.1	6686.3	–
	Siemens	181 ± 22	4.5 ± 0.2	6578.9	4.3 ± 0.1	6579.8	–
	DAX	200	7.3 ± 0.4	9561.3	7.2 ± 0.3	9558.7	–

Table 6.1: *Estimated parameters for the $\varepsilon_1\varepsilon_2$ -model and the ε_2 -model for the case of uniformly and Normally distributed noise. The last three columns refer to the ε_2 -model. The missing entrances in the p-values of the likelihood-ratio test are due to the errors of the numerical integration.*

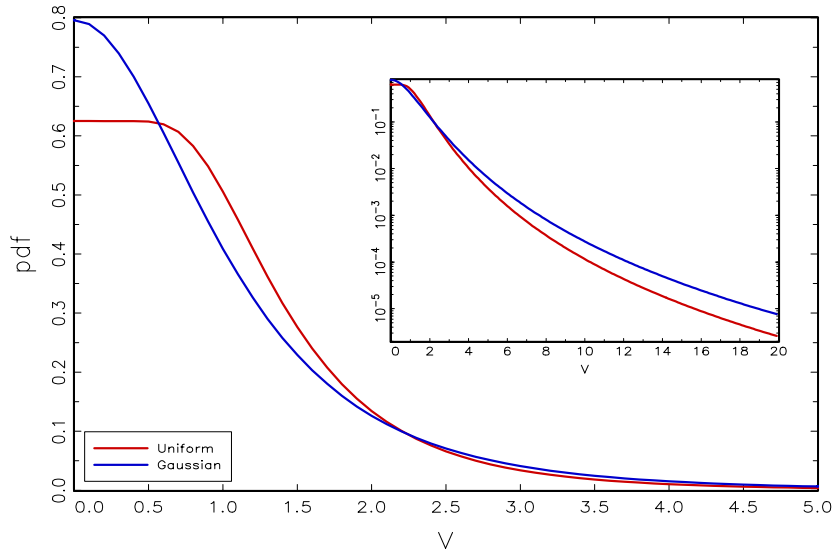


Figure 6.2: The figure compares the probability density functions of the $\varepsilon_1\varepsilon_2$ -model with uniform and Gaussian noise term. The parameters for the uniform-noise-model is $\varepsilon_2 = 5$, for the Gaussian-noise-model $\varepsilon_2 = 5$. The inset shows the pdf in log-linear scale for a better visualization of the tail region.

Table 6.1 summarizes the results of our estimation exercise for a pool of representative financial time series. Several interesting considerations can be drawn from the results of the estimation procedure. A common pattern emerges from the comparison of the values of the parameters across the different data sets and the different functional forms of the noise term. We observe that the case $\varepsilon_1 > \varepsilon_2$ is the most common outcome of the estimation procedure, which holds for Gaussian as well as for uniform noise —the only exception is Gold with the uniform noise term³. An important property of our model is that the parameters of the distribution can be related to the behavior of the agents —see chapter 6. For example, $\varepsilon_1 > \varepsilon_2$ indicates that, *on average*, the market is dominated by a technical trader attitude (i.e. $E[z] > 0.5$). The revealed asymmetry of the parameters reflects an analogous asymmetric behavior of the traders. The condition $\varepsilon_1 > \varepsilon_2$ implies, in fact, a **dominance of the technical traders' attitude** on the market. While the overall tendency is preserved for the two cases, namely uniform and Gaussian noise, the magnitude of this dominance is much stronger using the Gaussian noise.

³In order to assess if the difference between ε_1 and ε_2 is significant, we performed a likelihood-ratio test under the restriction $\varepsilon_1 = \varepsilon_2$. The test, independently done for every series, does not reject the hypothesis that the parameters are equal for the time series of Gold, DB and Siemens, in the case of uniform noise term. However, the overall behavior of the parameters indicates the persistent pattern $\varepsilon_1 > \varepsilon_2$.

Interestingly, while the behavior of the parameter ε_1 changes considerably across series and noises, the value of the parameter ε_2 is fairly stable —excluding the case of DAX. Recalling that the parameter ε_2 should coincide with the so-called tail index of the unconditional distribution, we compare the pertinent results to those obtained with the standard conditional ML estimator of the tail index (Hill estimator). As can be seen from Table 6.1 and (6.2), in all cases, the estimated values of $\hat{\varepsilon}_2$ are somewhat above the 95% interval for the semi-parametric Hill estimator for the tail index.⁴ The slight tendency of the parametric estimation towards higher values, when compared with the semi-parametric estimator, may be explained by the influence of the center of the distribution on the estimated values of ε_2 . Nevertheless, we observe a relatively small interval of variability of $\hat{\varepsilon}_2$ in harmony with the remarkable homogeneity of the tail index for empirical data.

Data Set	$\hat{\alpha}_H$	95% interval
Gold	2.9	(2.4, 3.4)
DB	3.4	(2.9, 4.0)
Siemens	3.7	(3.2, 4.3)
DAX	3.1	(2.9, 3.6)

Table 6.2: *The Hill estimates for the four series in Table 6.1 with the correspondent 95% confidence interval —see chapter 2.*

Comparing the values of the likelihood, it seems that the Gaussian noise is providing a better fit to the data than the uniform noise. However, it is hard to draw a definitive conclusion, since the two models are not nested and the considered pool of series is small —e.g. in the case of the German index DAX, the uniform noise has a higher likelihood. On the other hand, the values of the likelihood differ in most of the cases for one or two units. This difference falls into the error of the numerical procedure to compute the integrals (A 6.13) and (A 6.14).

Table 6.1 shows the results of the likelihood ratio test under the restriction $\varepsilon_1 \rightarrow \infty$, which is the condition for the emergence of the ε_2 -model —see chapter 6. The extreme asymmetric model is always rejected in the case of uniform noise. For the Gaussian noise, we can reject at a high significance level the asymptotic version for the Gold time series. It is not possible a reliable application of the likelihood-ratio test for the other time series due to the presence of the errors of the numerical integration.

Figures (6.3) and (6.4) show the fit of the data for those cases which are not in line with the general pattern of the estimation exercise: i) the $\varepsilon_1\varepsilon_2$ -model with its

⁴One might remark, however, the well known fact that the asymptotic distribution underestimates the finite-sample variability of estimates for processes with volatility clustering (cf Kearns and Pagan [1997]).

asymptotic counterpart with the Gaussian noise; ii) the case of the DAX with the $\varepsilon_1\varepsilon_2$ -model with uniform noise compared with the ε_2 -model both with Gaussian noise. This graphical comparison clarifies the origin of the discrepancy of the values of the likelihood for the chosen cases.

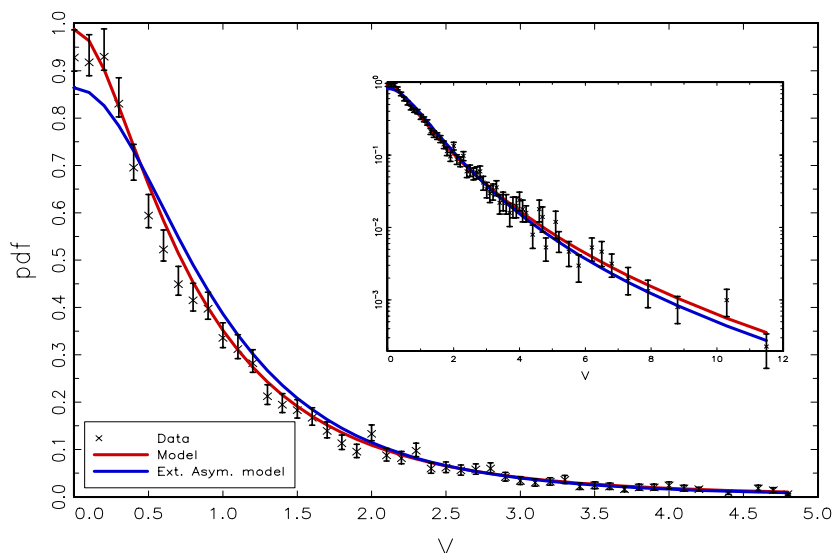


Figure 6.3: The main panel shows the probability density of the time series of Gold as a function of the normalized volatility V together with the distribution given by eqs. (A 6.13) and (A 6.14) with estimated parameters given in Table 6.1. The inset shows for both cases the pdf in a log-linear plot for a better visualization of the tail region. Although the likelihood ratio test rejects the asymptotic version of the $\varepsilon_1\varepsilon_2$ -model, the graphical comparison shows a quite similar behavior. The rejection is probably due to somewhat better fit of the $\varepsilon_1\varepsilon_2$ -model in the region around zero, which counts for a large fraction of the overall events —the number of normalized absolute returns smaller than 0.5 is 2165, which constitutes 43% of the entire sample. The graph also shows intervals of \pm one standard deviation, which are computed assuming a Normal distribution for the entries in every bin of the histogram. The same procedure is applied also in Figure (6.4).

Conclusions

Our estimation exercise reveals the power of the model in describing the behavior of real data. The unconditional distribution is well described by the $\varepsilon_1\varepsilon_2$ -model irrespective of the chosen functional form of the noise —at least for the two distributions considered so far. The values of the parameter ε_2 is not exactly in line with the semi-parametric Hill estimator of the tail index, however its small interval of variability is remarkable. Moreover, a striking result is the evidence for an asymmetric

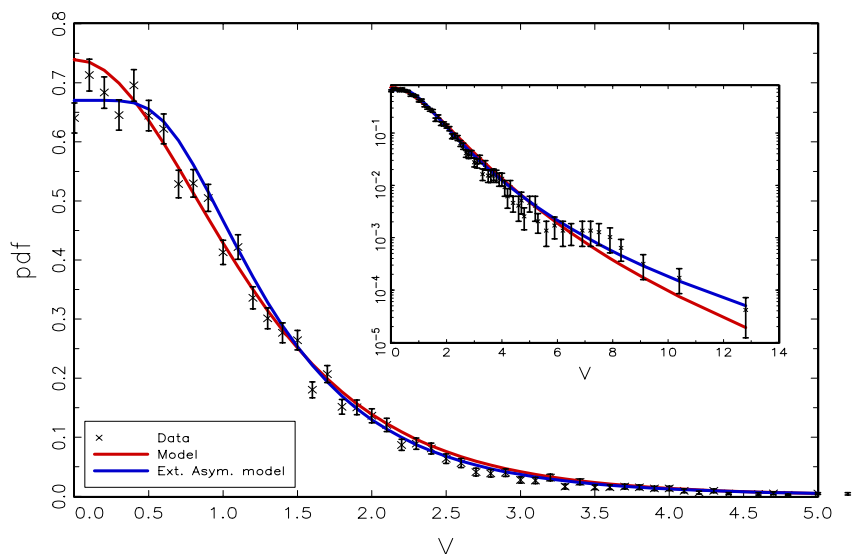


Figure 6.4: The main panel shows the probability density of the time series of DAX as a function of the normalized volatility V together with the distribution given by eqs. (A 6.4) and (A 6.14) with estimated parameters given in Table 6.1. The inset shows for both cases the pdf in a log-linear plot for a better visualization of the tail region. The significant difference in the values of the likelihood might be explain in the different fit of the region around zero. However, a visual inspection shows that both models describe the data reasonable well.

behavior of the traders with respect to both chartistic and fundamentalistic investment strategy. Considering a Gaussian noise term, the extreme asymmetric model turns out to be more appropriate in describing real data than the correspondent $\varepsilon_1\varepsilon_2$ -model.

So far, the estimation procedure has been applied only to those parameters that enter in the unconditional distribution. The information on the time evolution of the model is a crucial aspect that has to be investigated in future research. On one hand, the inclusion in the estimation procedure of the time correlation should improve the quality of the estimation itself. On the other hand, we can analyze whether the model can reproduce quantitatively the time correlations exhibited by real data, and described in detail in chapter 3. In principle, the inclusion of the time-correlation properties of the process (5.26) in the estimation procedure would allow to estimate the volatility σ , which implies the possibility to forecast the market volatility. Rather than being just an econometric exercise, the connection of the variable σ with the agent-based model would also allow to estimate the actual fraction of the fundamentalists and chartists in real markets.

Appendix to Chapter 6

A 6.1 Unconditional distribution of returns: $\varepsilon_1\varepsilon_2$ -model with uniform noise

The returns are the product of the noise variable η , distributed according to $p(\eta)$, and the volatility variable σ , distributed according to the equilibrium distribution $\rho_e(\sigma)$ from equation (5.41). If we assume that $p(\eta)$ is uniformly distributed in the interval $[-1, 1]$, the conditional distribution of $|r| = V$ for a given value of σ , denoted by $p(V|\sigma)$, takes the following expression:

$$p(V|\sigma) = \frac{1}{\sigma\sigma_0} . \tag{A 6.1}$$

The unconditional distribution of V is, then, obtained by integrating $p(V|\sigma)$ over σ with respect to its equilibrium distribution $\rho_e(\sigma)$:

$$p_u(V) = \int p(V|\sigma) \rho_e(\sigma) d\sigma . \tag{A 6.2}$$

The subscript indicates that the noise term is uniformly distributed. Taking into account that the domain of the variable $|\eta|$ is bounded, the integral (A 6.2) has the following integration limits:

$$p_u(V) = \int_V^\infty \frac{1}{\sigma\sigma_0} \rho_e(\sigma) d\sigma . \tag{A 6.3}$$

Inserting $\rho_e(\sigma)$ from eq. (5.41), we find

$$p_u(V) = \frac{1}{\sigma_0^2 B(\varepsilon_1, \varepsilon_2)} \int_V^\infty \left(\frac{\sigma}{\sigma_0}\right)^{\varepsilon_1-2} \left(\frac{\sigma_0}{\sigma + \sigma_0}\right)^{\varepsilon_1+\varepsilon_2} d\sigma . \tag{A 6.4}$$

If we make the substitution $u = \sigma_0 t / (1 - t)$, the integral (A 6.4) can be reduced to the incomplete beta function $\beta(\cdot; \cdot, \cdot)$, defined by:

$$\beta(x; a, b) = \frac{1}{B(a, b)} \int_0^x t^{a-1} (1-t)^{b-1} dt , \tag{A 6.5}$$

with $B(a, b)$ being the Beta function given in eq. (4.37). Finally we obtain:

$$p_u(V) = \frac{1}{\sigma_0} \frac{\varepsilon_2}{\varepsilon_1 - 1} \left[1 - \beta\left(\frac{V}{V + \sigma_0}; \varepsilon_1 - 1, \varepsilon_2 + 1\right) \right] . \tag{A 6.6}$$

A 6.2 Unconditional distribution of returns: ε_2 -model with uniform noise

The unconditional distribution of absolute returns for the ε_2 -model, denoted as $p_u^\infty(V)$, can be easily computed using the results of the case of $\varepsilon_1\varepsilon_2$ -model. Starting from eq. (A 6.3) and inserting the expression for the probability distribution of the volatility⁵ v as given by eq. (5.61), we end up with the following integral:

$$p_u^\infty(V) = 2 \frac{(\varepsilon_2 - 1)^{\varepsilon_2}}{\Gamma(\varepsilon_2)} \int_V^\infty \frac{1}{v^{\varepsilon_2+2}} e^{-\frac{\varepsilon_2-1}{v}} dv . \tag{A 6.7}$$

⁵Note that we now denote with v the volatility to distinguish the cases $\varepsilon_1\varepsilon_2$ -model and $\varepsilon - 1$ -model.

With the substitution $v = (\varepsilon_2 - 1)/u$, we arrive at the expression:

$$p_u^\infty(V) = \frac{2}{(\varepsilon_2 - 1)\Gamma(\varepsilon_2)} \int_0^{\frac{2(\varepsilon_2 - 1)}{V}} u^{\varepsilon_2} e^{-u} du, \quad (\text{A 6.8})$$

which can be rewritten using the lower incomplete gamma function⁶:

$$p_u^\infty(V) = \frac{2}{(\varepsilon_2 - 1)\Gamma(\varepsilon_2)} \gamma\left(\varepsilon_2 + 1, \frac{2(\varepsilon_2 - 1)}{V}\right). \quad (\text{A 6.10})$$

A 6.3 Unconditional distribution of returns with Gaussian noise

We can use eq. (A 6.2) to compute the unconditional distribution of the absolute returns for the Gaussian noise term, denoted with $p_g(V)$. The integral (A 6.2) can be expressed as follows:

$$p_g(V) = \frac{\sqrt{2}}{\sigma_0\sqrt{\pi}} \int_0^\infty \frac{1}{\sigma} \exp\left(-\frac{1}{2}\left(\frac{V}{\sigma_0\sigma}\right)^2\right) \rho_e(\sigma) d\sigma. \quad (\text{A 6.11})$$

The previous integral can be rewritten in a boded domain with the substitution $\sigma = \frac{z}{1-z}$. With the following relation:

$$\rho_e(\sigma)d\sigma = p_e(z)dz, \quad (\text{A 6.12})$$

we end up with the expression:

$$p_g(V) = \frac{\sqrt{2}}{B(\varepsilon_1, \varepsilon_2)\sigma_0\sqrt{\pi}} \int_0^1 \exp\left(-\frac{1}{2}\left(\frac{V(1-z)}{\sigma_0 z}\right)^2\right) z^{\varepsilon_1 - 2}(1-z)^{\varepsilon_2} dz, \quad (\text{A 6.13})$$

which we, then, have to integrate numerically.

The integral for the unconditional distribution of returns in the case of the ε_2 -model can be analogously computed, which leads to the following expression:

$$p_g(V) = \frac{\sqrt{2}}{(\varepsilon_2 - 1)\Gamma(\varepsilon_2)\sqrt{\pi}} \int_0^\infty \exp\left(-\left(\frac{V u}{\pi(\varepsilon_2 - 1)}\right)^2 - u\right) u^{\varepsilon_2} du. \quad (\text{A 6.14})$$

A 6.4 Discreteness of price records

In the estimation procedure, an important issue arises from the discrete nature of the records of prices. Close inspection of our empirical time series shows a variability of the precision of the entries that changes over time. The DAX, for instance, exhibits a decimal precision for the first 7064 data points, which is increased to the second digit afterwards; the price of Gold is recorded, for a long period, with a precision

⁶The lower incomplete gamma function is defined as:

$$\gamma(a, x) = \int_0^x t^{a-1} e^{-t} dt. \quad (\text{A 6.9})$$

of a quarter of dollar, but changes to smaller increments later on. The discreteness of prices is reflected in an artificial large variability of small returns ($|r| < E[|r|]$), where certain values are more frequent than others, due to some sort of threshold effect. Consequently, the empirical proxy for volatility $|r|$ is poorly approximated by a continuous variable in the region close to zero, while the key variable in our model is continuous over its entire range. In order to induce continuity, one could add a small white Gaussian noise with mean zero to the time series of prices, to avoid the discreteness of small returns. If one takes a standard deviation of the noise of 0.05, this should be sufficiently small to just marginally affect the original price level, but large enough to avoid the discreteness in the last decimal⁷.

⁷In the case of Gold price the standard deviation is 0.1, since the minimum increment is 1/4.

CHAPTER 7

General Conclusions

We have presented an extremely simple model of an agent-based artificial financial market, which nevertheless is able to reproduce the key stylized facts of financial time series. Due to its very simple structure, we were able to obtain a full analytical solution for i) the unconditional distribution of returns, which possesses a power law decay of the tail, as empirically identified in the financial econometrics literature; ii) the autocorrelation of absolute returns, which exhibits a slow decay, resembling the long range dependence of real data iii) the autocorrelation of returns which by construction exhibits an ‘almost’ vanishing degree of correlation, indicating unpredictability of the returns in close agreement with real data.

We have shown that the stochastic process governing the returns dynamics of the artificial market turns out to be equivalent to a stochastic volatility decomposition. In contrast to previous research in financial econometrics, which is assuming a particular stochastic process as exogenously given, our work identifies the process as an outcome of an aggregation of the behavior of interacting heterogeneous agents. Thus, we bridge the gap between the econometric approach to financial economics and the agent-based simulation approach. Deriving the stochastic differential equation governing the returns dynamics from the theoretical model has the crucial advantage of enabling us to clearly identify the nature of the randomness present in the artificial market. Moreover, we can link the statistical properties of the fluctuations (for instance the volatility clustering or the fatness of the distribution of returns) to the features of the microscopic interactions among traders (Markovian nature of the herding model or non extensivity property of the transition rates).

The connection of the model with an econometric framework allows for a direct estimation of some parameters of the model using real data. What emerges is an asymmetric behavior of the traders with respect to the available investment strategies, namely fundamentalists and chartistic techniques.

Acknowledgments

I am finally able to write the last words of the thesis. It was not just a scientific challenge, but also an incredible life experience. It has been really amazing how, while sitting in the same office, I could be in contact with persons from almost all the continents. This gave to me the opportunity to open my narrower perspective on the multi-cultural world we are living in.

I would like to thank Professor Lux for both the academic and personal support during these years. More important, I would like to thank him for giving me the opportunity to start (and to continue) my adventure in the academic world.

I would like to thank Professor Wagner for his great effort to patiently teach me the secrets of statistical physics. He has been (and he continues to be) my guide in the field of science that gives me always inspiration: Physics.

Many other persons gave to me a precious help in these years to accomplish my work. Thanks to all the colleges for the small as well the big suggestions.

I am particularly in debt with Eleni, Emanuel and Lorenzo not only for the daily help in writing the thesis... but for the the most precious friendship I could find around me.

Finally I want to thank my family that, although far, has been always close to me and to whom is dedicated this work.

L'ultima parola e' per Eva: *grazie*.

Bibliography

- M. Abramowitz and I. A. Stegun. Orthogonal polynomials. In M. Abramowitz and I. A. Stegun, editors, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. U.S. Department of Commerce, New York, 1972.
- D. Ahn and B. Gao. A parametric nonlinear model of term structure dynamics. *Review of Financial Studies*, 12:721–762, 1999.
- S. Alfarano and T. Lux. Extreme value theory as a theoretical background for power law behavior. In Claudio Cioffi-Revilla, editor, *Power Laws in the Social Sciences: Discovering Complexity and Non-Equilibrium Dynamics in the Social Universe*. 2004.
- S. Alfarano, T. Lux, and F. Wagner. Estimation of agent-based models: the case of an asymmetric herding model. *Computational Economics*, 26:19–49, 2005.
- M. Aoki. *New Approaches to Macroeconomic Modeling: Evolutionary Stochastic Dynamics, Multiple Equilibria, and Externalities as Field Effects*. University Press, Cambridge, 1996.
- M. Aoki. *Modeling Aggregate Behavior and Fluctuations in Economics*. University Press, Cambridge, 2002.
- G. Arfken. *Mathematical Methods for Physicists*. Academic Press, Orlando, FL, 1985. 3rd ed.
- W. B. Arthur, J. H. Holland, B. LeBaron, R. Palmer, and P. Tayler. Asset pricing under endogenous expectations in an artificial stock market. *Economic Notes*, 26:297–330, 1997.
- L. Bachelier. Théorie de la spéculation. *Ann. Sci. Ecole Norm. Sup.*, 17:21–86, 1900. English Translation in Cootner, P. H. (Ed.) *The Random Character of Stock Market Prices*, MIT Press, Cambridge, 1964, 17-78.
- W. Barnett and A. Serletis. Martingales, nonlinearity and chaos. *Journal of Economic Dynamics and Control*, 24:703–729, 2000.
- M. S. Bartlett. On the theoretical specification and sampling properties of autocorrelated time-series. *J. Roy. Statist. Soc. Supplement*, 8:27–41, 1946.
- W. J. Baumol. Speculation, profitability and stability. *Review of Economics and Statistics*, 39:263–271, 1957.

- J. Beirlant, P. Vynckier, and J.L. Teugels. *Practical Analysis of Extreme Values*. University Press, Leuven, 1996.
- A. Beja and M. B. Goldman. On the dynamic behavior of prices in disequilibrium. *Journal of Finance*, 35:235–248, 1980.
- J. Beran. *Statistics for Long-Memory Processes*. Chapman and Hall, London, 1994.
- R. C. Blattberg and N. J. Gonedes. A comparison of stable and student distributions as statistical model for stock prices. *Journal of Business*, 47:244–280, 1974.
- H.P. Boswijk, C.H. Hommes, and S. Manzan. Behavioral heterogeneity in stock prices, 2005. CeNDEF Working paper 05-12, University of Amsterdam.
- J-P Bouchaud and M. Potters. *Theory of Financial Risks. From Statistical Physics to Risk Management*. Univerity press, Cambrige, 2000.
- J-P Bouchaud and M. Potters. *Theory of Financial Risks. From Statistical Physics to Risk Management*. Univerity press, Cambrige, 2003. Second edition.
- W. A. Brock, W. D. Dechert, and J. A. Scheinkman. A test for independence based on the correlation dimension. Department of Economics, University of Wisconsin, 1987. Manuscript.
- W. A. Brock and C. H. Hommes. A rational route to randomness. *Econometrica*, 65:1059–1095, 1997.
- A. Bunde, S. Havlin, E. Koscielny-Bunde, and H. J. Schellnhuber. Atmospheric persistence analysis: novel approaches and application. In A. Bunde, J. Kropp, and H. J. Schellnhuber, editors, *Theories of Disaster - Scaling Laws Governing Weather, Body, and Stock Market Dynamics*, pages 171–191. Springer, Berlin Heidelberg, 2002.
- A.C. Cameron and P.K. Trivedi. Tests of independence in parametric models: with applications and illustrations. *Journal of Business and Economic Statistics*, 45: 1–28, 1993.
- D. Challet and M. Marsili. From minority game to the real markets. *Physical review E*, 68:168–176, 2004.
- D. Chandler. *Introduction to modern statistical mechanics*. Oxford university press, New York, 1987.
- J. Cox, J. Ingersoll, and S. Ross. A theory of term structure of interest rates. *Econometrica*, 53:385–407, 1985.
- D. Cutler, J. Poterba, and L. Summer. Speculative dynamics. *Review of Economic Studies*, 56:529–546, 1991.
- M. M. Dacorogna, R. Gençay, U. Müller, and O. V. Pictet. *An Introduction to High-Frequency Finance*. Accademic press, 2001.

- J. Danielsson, L. de Hann, Peng L., and C.G. de Vries. Using a bootstrap method to choose the optimal sample fraction in tail index estimation. *Journal of Multivariate Analysis*, 76:226–248, 2001.
- J. Danielsson and C.G. de Vries. Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance*, 4:241–257, 1997.
- R.A. Davis and T. Mikosch. The sample autocorrelations of financial time series models. In W.J. Smith R.L. Fitzgerald, A.T. Walden, and P. Young, editors, *Nonlinear and Nonstationary Signal Processing*, pages 247–274. Cambridge University Press, Cambridge, England, 2000.
- B. J. De Long, A. Shleifer, L. H. Summers, and R. J. Waldmann. Noise trade risk in financial markets. *Journal of Political Economy*, 98:703738, 1990.
- C. G. de Vries. Stylized facts of nominal exchange rate returns. In F. van der Ploeg, editor, *The Handbook of International Macroeconomics*, pages 348–389. Blackwell, Oxford, 1994.
- Z. Ding, R. Engle, and C. Granger. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1:83–106, 1993.
- A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York, 2001.
- H. Drees and E. Kaufmann. Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and Their Applications*, 75:149–172, 1998.
- R. D. Edwards, J. Magee, and W. H. C. Bassetti. *Technical Analysis of Stock Trends*. St. Lucie Press, London, 2001.
- E. Egenter, T. Lux, and D. Stauffer. Finite-size effects in Monte Carlo simulations of two stock market models. *Physica A*, 268:250–256, 1999.
- B. Eraker. MCMC Analysis of Diffusion Models With Application to Finance. *Journal of Business and Economic Statistics*, 19:177–191, 2001.
- E. F. Fama. Mandelbrot and the stable paretian hypothesis. *Journal of Business*, 35:420–429, 1963.
- E. F. Fama. The behavior of stock market prices. *Journal of Business*, 35:420–429, 1965.
- E. F. Fama. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 35:420–429, 1970.
- W. Feller. *An Introduction to Probability Theory and its Applications*. John Wiley and sons, New York, 1971.

- J. Frankel and K. A. Froot. Understanding the U.S. dollar in the eighties: The expectations of chartists and fundamentalists. *Economic Record*, Special Issue, December:24–38, 1986.
- D. Friedman and S. Vandersteel. Short-run fluctuations in foreign exchange rates. *Journal of International Economics*, 13:171, 1982.
- M. Friedman. The case for flexible exchange rates. In *Essays in Positive Economics*. University of Chicago Press, Chicago, 1953.
- A. Gallant and G. Tauchen. Which Moment to Match? *Econometric Theory*, 12: 657–681, 1996.
- C. W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Springer, 2003. Third edition.
- U. Garibaldi, M.A. Penco, and P. Viarengo. An exact physical approach for market participation models. In R. Cowan and Jonard N., editors, *Heterogeneous agents, Interactions and Economic Performances, Lecture Notes in Economics and Mathematical Systems*. Springer, Berlin, 2003.
- V. Genon-Catalot, T. Jeantheau, and C. Laredo. Parameter estimation for discretely observed stochastic volatility models. *Bernoulli*, 5:855–872, 1999.
- J. Geweke and S. Porter-Hudak. The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4:221–238, 1983.
- M. Gilli and E. K ellezi. An application of extreme value theory for measuring risk. <http://www.unige.ch/ses/metri/gilli/evtrm/CSDA-08-02-2003.pdf>, 2003. Manuscript.
- M. Gilli and P. Winker. A global optimization heuristic for estimating agent based models. *Computational Statistics and Data Analysis*, 42:299–312, 2003.
- C.M. Goldie and R.L. Smith. Slow variation with remainder: Theory and applications. *Quarterly Journal of Mathematics*, 38:45–47, 1987.
- P. Gopikrishnan, M. Meyer, L.A.N. Amaral, and H.E. Stanley. Inverse cubic law for the distribution of stock price variations. *Eur. Phys. J. B*, 3:139–140, 1998.
- J.A. Hall, B.W. Brorsen, and S.H. Irwin. The distribution of futures prices: A test of the stable paretian and mixture of normal hypotheses. *Journal of Financial and Quantitative Anaysis*, 24:105–116, 1989.
- P. Hall. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society, Series B*, 44:37–42, 1982.
- P. Hall. Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*, 32:177–203, 1990.

- S.L. Heston. A closed-form solution for option with stochastic volatility with application to bonds and currency options. *The Review of Financial Studies International Journal of Modern Physics C*, 6:327–343, 1993.
- C. Heyde and W. Dai. On the robustness to small trends of estimation based on the smoothed periodogram. *Journal of Time Series Analysis*, 17:141–150, 1996.
- B. M. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3:1163–1173, 1975.
- D.-A. Hsu, R.B. Miller, and D.W. Wichern. On the stable paretian behavior of stock-market prices. *Journal of American Statistic Association*, 69:108–113, 1974.
- H. E. Hurst. Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civil Engineers*, 116:770–799, 1951.
- C. Hurvich, R. Deo, and J. Brodsky. The mean squared error of Geweke and Porter-Hudaks estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis*, 19:1946, 1998.
- C.M. Jarque and A.K. Bera. Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals. *Economic Letters*, 6:255–259, 1980.
- D. Kahneman and Tversky A. On the Psychology of Prediction. *Psychological review*, 80:237–251, 1973.
- D. Kahneman and Riepe. Aspect of investor psychology. *Journal of Portfolio Management*, 24:52–65, 1998.
- P. Kearns and A.R. Pagan. Estimating the density tail index for financial time series. *Review of Economics and Statistics*, 79:171–175, 1997.
- F. Kelly. *Reversibility and Stochastic Networks*. Wiley & Sons, New York, 1979.
- A. Kirman. Epidemics of opinion and speculative bubbles in financial markets. In M. P. Taylor, editor, *Money and Financial Markets*, pages 354–368. Blackwell, Cambridge, 1991.
- A. Kirman. Ants, rationality, and recruitment. *Quarterly Journal of Economics*, 108:137–156, 1993.
- A.H.-L. Lau, H.-S. Lau, and J.R. Wingender. The distribution of stock returns: new evidence against the stable model. *Jornal of Business Economics and Statistics*, 8(2):217–224, 1990.
- B. LeBaron. Agent based computational finance: Suggested readings and early research. *Journal of Economic Dynamics and Control*, 24:679–702, 2000.
- A. L. Lewis. *Option Valuation under Stochastic Volatility*. Finance Press, California, 2000.

- G. Ljung and G. Box. On a measure of lack of fit in time series models. *Biometrika*, 65:297–303, 1978.
- A. W. Lo. Long-term memory in stock market prices. *Econometrica*, 59:1279–1313, 1991.
- A. W. Lo. The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management*, 30:15–29, 2004.
- M. Loretan and P. C. B. Phillips. Testing the covariance stationarity of heavily-tailed time series. *Journal of Empirical Finance*, 1:211–248, 1994.
- T. Lux. Herd behaviour, bubbles and crashes. *Economic Journal*, 105:881–896, 1995.
- T. Lux. The limiting extremal behavior of speculative returns: an analysis of intradaily data from the Frankfurt stock exchange. *Applied Financial Economics*, 11: 299–315, 1996a.
- T. Lux. Long-term stochastic dependence in financial prices: evidence from the German stock market. *Applied Economics Letters*, 3:701–706, 1996b.
- T. Lux. On moment condition failure in German stock returns: an application of recent advances in extreme value statistics. *Empirical Economics*, 20:641–652, 2000.
- T. Lux. The stable Paretian hypothesis and the frequency of large returns: An examination of major German stocks. *Applied Financial Economics*, 6:463–475, 2001.
- T. Lux and M. Ausloos. Market fluctuations I: Scaling, multiscaling and their possible origins. In A. Bunde, J. Kropp, and H. J. Schellnhuber, editors, *Theories of Disaster - Scaling Laws Governing Weather, Body, and Stock Market Dynamics*, pages 373–409. Springer, Berlin Heidelberg, 2002.
- T. Lux and M. Marchesi. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397:498–500, 1999.
- T. Lux and M. Marchesi. Volatility clustering in financial markets: A micro-simulation of interacting agents. *International Journal of Theoretical and Applied Finance*, 3:67–702, 2000.
- T. Lux and S. Schornstein. Genetic learning as an explanation of stylized facts of foreign exchange markets. *Journal of Mathematical Economics*, 41:169–196, 2005.
- B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 35: 394–419, 1963.
- B. Mandelbrot. Statistical methodology for nonperiodic cycles: from the covariance to R/S analysis. *Rev. Econ. Soc. Meas.*, 10:259–290, 1973.

- M. M. Mansour and A. de Palma. On the stochastic modelling of systems with non-local interactions. *Physica A*, 128:377–382, 1984.
- R.N. Mantegna and H.E. Stanley. *An Introduction to Econophysics. Correlation and Complexity in Finance*. Univerisy Press, Cambrige, 2000.
- D.M. Mason. Laws of large numbers for sums of extreme values. *Annals of Probability*, 10:754–764, 1982.
- G. Matthys and J. Beirlant. Adaptive threshold selection in tail index estimation. In P. Embrechts, editor, *Extremes and Integrated Risk Management*, pages 37–49. Risk Books, London, 2000.
- G. Nicolis and I. Prigogine. *Self-organization in non-equilibrium systems*. Wiley, New York, 1977.
- M. O’Hara. *Market Microstructure Theory*. Blackwell, Cambridge, 1995.
- A. Pagan. The econometrics of financial markets. *Journal of Empirical Finance*, 3: 15–102, 1996.
- C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, Stanley H. E., and A. L. Goldberger. Mosaic organization of DNA nucleotidies. *Physical Review E*, 49:1685–1689, 1994.
- V. Plerou, P. Gopikrishnan, L. Amaral, , M. Meyer, and H.E. Stanley. Scaling of the distribution of price fluctuations of individual companies. *Physical Review E*, 60:6519–6529, 1999.
- J. B. Ramsey. On the existence of macro variables and of macro relationships. *Journal of Economic Behavior and Organisation*, 30:275–299, 1996.
- R.-D. Reiss and M. Thomas. *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhäuser, Basle, 1997.
- P. M. Robinson. Log-periodogram regression of time series with long range dependence. *Annals of Statistics*, 23:10481072, 1995.
- R. Roll. Orange juice and weather. *American Economic Review*, 74:861–880, 1984.
- H. Rootzen, M.R. Leadbettr, and L. de Hann. On the distribution of tail array sums for strongly mixing stationary sequences. *Annals of Applied Probability*, 8: 868,885, 1998.
- R. Shiller. Do prices move toomuch to be justified by subsequent changes in dividends. *American Economic Review*, 71:421–436, 1981.
- R. Shiller. Stock prices and social dynamics. *Brookings Papers on Economic Activities*, 2:457–498, 1984.
- A. Shleifer. *Inefficient Markets. An introduction to behavioral finance*. Oxford University Press, New York, 1999.

- P. Sibbertsen. Log-periodogram estimation of the memory parameter of a long-memory process under trend. *Statistics and Probability Letters*, 61:261–268, 2003.
- M. Sørensen and B.M. Bibby. Hyperbolic processes in finance. In S. Rachev, editor, *Heavy Tailed Distributions in Finance*, pages 211 – 248. Elsevier Science, 2003.
- D.E. Upton and D.S. Shannon. The stable paretian distribution, subordinated stochastic processes, and asymptotic lognormality: an empirical investigation. *Journal of Finance*, 34(4):1031–1039, 1979.
- N. G. Van Kampen. *Stochastic processes in Physics and Chemistry*. North Holland, Amsterdam, 1992. Revised edition.
- C. Velasco. Non-stationary log-periodogram regression. *Journal of Econometrics*, 91:325–371, 1999.
- R. Weron. Estimating long-range dependence: finite sample properties and confidence intervals. *Physica A*, 312:285–299, 2002.
- F.H. Westerhoff and S. Reitz. Nonlinearities and cyclical behavior: The role of chartists and fundamentalists. *Studies in Nonlinear Dynamics and Econometrics*, 7, 2003.
- E. C. Zeeman. On the unstable behaviour of stock exchanges. *Journal of Mathematical Economics*, 1:39–49, 1974.

CURRICULUM VITAE

PERSONAL DATA

Name: **Simone**
Surname: **Alfarano**
Date of Birth: **21/06/1975**
Place of Birth: **Cagliari (Italy)**
Nationality: **Italian**
E-mail: **alfarano[at]bwl.uni-kiel.de**

EDUCATION

M. Sc. in Physics
at University of Cagliari, Italy (*Dec. 1999*)

Ph. D. in Quantitative Economics (*Feb. 2006*)
at the Christian-Albrechts University of Kiel, Germany (*Feb. 2006*)

ACTUAL POSITION

Assistant Professor

Department of Economics
Christian-Albrechts University of Kiel, Germany

AWARDS AND FELLOWSHIPS

Post-Lauream Scholarship for Young Researchers
[Regione Autonoma della Sardegna, Italy 2000-2003]

**European Union Fellowship for Young Researchers
New Conference**
[University of Salerno, Italy, 13 - 15 September 2001]

SNDE Fellowship
(Society of Nonlinear Dynamics and Econometrics)
[Federal Reserve Bank of Atlanta, USA, 14 - 15 March 2002]

DAAD Scholarship for Research Exchange Program
(Deutscher Akademischer Austausch Dienst)
[Ph.D. Program in Quantitative Economics, University of Kiel,
Germany, February - March 2003]

2005

Wehia 2005

Workshop on Economics with Heterogeneous Interacting Agents
[Essex university, Colchester, UK] *Talk*

2004

Wehia 2004

Workshop on Economics with Heterogeneous Interacting Agents
[Kyodai Kaikan Kyoto University, Kyoto, Japan] *Talk*

10th International SCE Conference

Society for Computational Economics
[University of Amsterdam, Amsterdam, The Netherlands] *Talk*

First Bonzenfreies Colloquium

[University of Alessandria, Alessandria, Italy]

Invited Speaker

First Workshop on Computational Finance

[University of Kiel, Kiel, Germany,] *Talk*

2003

11th Annual Symposium of the Society of Non-linear Dynamics and Econometrics

[Villa La Pietra, Firenze, Italy] *Talk*

8th Spring Meeting of Young Economists

[Catholic University of Leuven, Belgium] *Talk*

Wehia 2003

Workshop on Economics with Heterogeneous Interacting Agents
[Institute of World Economics, Kiel, Germany] *Talk*

27th AMASES Meeting

Applied Mathematics to Social Sciences and Economics
[University of Cagliari, Italy] *Talk*

2002

10th Annual Symposium of the Society of Nonlinear Dynamics and Econometrics

[Federal Reserve Bank of Atlanta, USA] *Talk*

Workshop on Concepts for Complex Adaptive Systems

[Hanse Wissenschaftskolleg, Delmenhorst, Germany] *Poster*

Wehia 2002

Workshop on Economics with Heterogeneous Interacting Agents
[Abdus Salam International Center, Trieste, Italy] *Talk*

8th International SCE Conference

Society for Computational Economics

[University of Marseille, Aix en Provence, France] *Talk*

2001

Conference on Dynamical Networks in Complex Systems

[University of Kiel, Germany] *Attendance*

New Conference: New Methods for the New Millennium

[University of Salerno, Italy] *Talk*

SEMINARS

Colloquium on Financial Markets

Institute of World Economics of Kiel

[Kiel, Germany, October 2001]

Department of Physics, University of Kiel

[Kiel, Germany, November 2001]

Department of Econometrics, University of Geneva

[Geneva, Switzerland, March 2003]

Department of Physics, University of Cagliari

[Cagliari, Italy, May 2003]

REFEREE FOR:

Studies in Nonlinear Dynamics and Econometrics

European Physics Journal

Journal of Economic Dynamics and Control

Applied Mathematical Finance

Hiermit versichere ich, die vorliegende Dissertation selbständig angefertigt und keine anderen als die angegebenen Hilfsmittel verwendet zu haben.

Kiel März 2006

