Financial markets as disordered interacting systems: information, risk and illiquidity

Fabio Caccioli

Supervisor: Prof. Matteo Marsili

SISSA-International School for Advanced Studies

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

September 2010

Previously published work

1. F. Caccioli and M. Marsili
   Information efficiency and financial stability
   Economics **4**,2010-20 (2010).
2. F. Caccioli, M. Marsili and P. Vivo
   Eroding market stability by proliferation of financial instruments
   Eur. Phys. J. B **71**, 467 (2009).
3. F. Caccioli, S. Still, M. Marsili and I. Kondor
   Optimal liquidation strategies regularize portfolio selection
   submitted to EJF, online at `http://arxiv.org/abs/1004.4169` (2010)
4. S. Bradde, F. Caccioli, L. Dall'Asta and G. Bianconi
   Critical fluctuations in spatial complex networks
   Phys. Rev. Lett. **104**, 218701 (2010).
5. R. Potestio, F. Caccioli and P. Vivo
   Random Matrix approach to collective behavior and bulk universality in protein dynamics
   Phys. Rev. Lett. **103**, 268101 (2009).
6. F. Caccioli and L. Dall'Asta
   Non-equilibrium mean-field theories on scale-free networks
   J. Stat. Mech., P10004 (2009)
7. F. Caccioli, S. Franz and M. Marsili
   Ising model with memory: coarsening and persistence properties
   J. Stat. Mech., P07006 (2008).

Paper 1 covers the content of Chapter 2, paper 2 of Chapter 3 and paper 3 of Chapter 4. The remaining papers have not been included in the present thesis. Chapter 4 also includes results not yet published.

Contents

List of Figures

Chapter 1

Introduction

The last fifteen years have witnessed a growing interest for applications of statistical physics to economic driven problems [Bouchaud et al., 2008; Bouchaud and Potters, 2000; Challet et al., 2005; Mantegna and Stanley, 2000]. Notably, the huge amount of electronically stored financial data that has become accessible has led to the discovery of regularities in the statistical properties of economic systems. For instance, in the context of financial markets, it has been shown [Burda et al., 2003; Chakraborti et al., 2009; Mantegna and Stanley, 2000] that the distribution of price changes, company sizes, individual wealth etc are characterized by power law tails that are to a large extent universal. These findings have triggered the interest of physicists used to see the emergence of such collective properties in systems close to criticality [Cardy, 1996].
In addition to the empirical approach devoted to discover, characterize and verify such regularities, statical physicists have started to introduce models trying to explain the observed collective properties as emerging from the interactions between "elementary units" (agents) [Challet et al., 2005; Marsili and De Martino, 2006]. Certainly, economic systems are much more complex than physics systems, the interacting units being individuals that follow complex behavioral rules. Nevertheless, it may be reasonable to assume that a crowd of interacting individuals may present aspects of statistical regularities that can be captured by means stylized models. In this respect, on an abstract level, the nature of the problems addressed in economics are not so different from that considered in physics when, for instance, we try to figure out how a magnetic system can

exhibit spontaneous magnetization. In both cases the problem is that of finding how, out of individual interactions at the microscopic scale, collective properties may emerge. From this point of view statistical mechanics, allowing for a characterization in terms of phases and phase transitions, may represent a useful perspective to look at economic systems [Marsili and De Martino, 2006].

A very important aspect to consider when modeling an economic system, is the fact that each individual is different in the way he/she interacts with the environment. From the statistical mechanics point of view, this heterogeneity may be accounted for by considering systems with random couplings. In view of the considerable progress that has been achieved in the last decades by statistical physics of disordered systems, this way of modeling economic systems has revealed very fruitful since in many cases analytical solutions may be attained at least in the limit of very large systems [Challet et al., 2005].

In this thesis, we will try to pursue this line of thought by discussing three specific problems of economic interest through the prism of statistical mechanics. Interestingly, we will see how three problems which are in principle very different, in the statistical mechanics perspective may be shaped in such a way to represent different instances of the same problem. Notably, all the systems we will consider will be characterized by phase transitions which ultimately may be traced back to the same root.

The first part of this thesis will be devoted to the analysis of some hypothesis usually assumed in the modeling of financial markets. Mainstream economic theories, like those used by financial institutions to price derivatives or to determine optimal investment strategies [Bailey, 2005; Pliska, 1997], usually describe idealized markets where rational agents instantaneously correct any mis-pricing, so that prices correctly reflect the underlying reality and ensure optimal allocation of resources. In this framework, responsibilities for extreme events like market crashes are usually put on deviations of real markets from these ideal conditions [Mishkin, 1996; Mishkin and Herbertsson, 2006], that, if satisfied, would guarantee stable and properly functioning markets. A lot of discussion has been recently made concerning the role of "markets imperfections" in the recent economic crisis [Shiller, 2008; Shiller and Akerlof, 2009; Turnbull et al., 2008]. Here we wish however to tackle the problem of market stability from a different point of view, notably we want to understand whether the theories at the basis of financial engineering also bear some responsibilities [Bouchaud, 2008]. We will try to understand this point through the modeling of markets in terms of interacting heterogenous agents [Challet et al., 2005; Marsili and De Martino, 2006], trying to understand whether ideal markets are always synonymous of stable markets. Notably we will consider the following topics:

Information efficiency and market stability: Markets are said to be efficient if prices faithfully reflect the underlying reality, thus ensuring optimal allocation of resources. By means of a simple model of a market where agents trade on the basis of some private information structure and interact through the process of price formation, it is possible to show how the market can act as an information processing device that aggregates information scattered across different investors into prices [Berg et al., 2001]. In this context, market efficiency appears as an emergent property when the number of agents with different information structure is large enough. Upon introducing non-informed agents, we show that the latter start contributing significantly to the trading activity as the market becomes close to being information-efficient and we will argue that information efficiency might create the condition for bubble phenomena induced by the behavior of non-informed traders to set in [Caccioli and Marsili, 2010].

Proliferation of financial instrumens and market stability Arbitrage Pricing Theory is the theory that allows to compute prices of financial contracts and is therefore at the basis of financial innovation [Pliska, 1997]. The proliferation of financial instruments, introducing more ways for risk diversification, is usually expected to drive the system closer to the limit of efficient, arbitrage free complete market described by the APT [Merton and Bodie, 2005]. Despite this fact, we witnessed a tremendous crash in correspondence to the historical period of greater expansion in the repertoire of financial instruments. To shed some lights on this apparent paradox, we introduce a model of a market where derivatives on an underlying are traded and a feedback is introduced that accounts for the impact of trading derivatives on the underlying [Caccioli et al., 2009]. We show that, upon increasing the number of derivatives, the market converges as expected towards the limit of ideal market described by APT. At the same time, however, the same region of phase space is characterized by the presence of a phase transition, with large fluctuations and sharp discontinuities.

In both cases the bottom line will be that the conditions usually assumed to describe ideal markets may not be compatible with the stability of the market, in particular we will give support the idea that in fact *the more markets are close to ideal conditions the more they are prone to instabilities* [Brock et al., 2008; Caccioli and Marsili, 2010; Caccioli et al., 2009; Marsili, 2009].

While the first part of the thesis is aimed at a theoretical discussion of the concepts at the basis of financial engineering, the second part is related to a problem of more immediate application, namely that of portfolio optimization. This problem is related to the first question one usually asks when dealing with financial markets: "how do I need to invest my money?". This is a very interesting problem and the solution depends in general on investors' characteristics. On one hand one would like to maximize the expected profit of the investment, on the other hand one is also interested in minimizing the risk associated to it [Bailey, 2005; Markowitz, 1952, 1959]. The problem of finding the optimal investment strategy is then related to a compromise between these two aspects, so that usually the solution depends on investors' attitude toward risk. Despite this potential heterogeneity of investors' behavior, some general guidelines may nevertheless be found. For instance, it is intuitively clear that if one has to choose between portfolios with the same expected returns but different risk, one should go for the portfolio bearing the minimum risk. Conversely, in presence of portfolio bearing the same risk, one should choose that of greater expected returns. This is, in a nutshell, also the basis of the celebrated Capital Asset Pricing Model (CAPM) Bailey [2005] for the determination of the efficient portfolio frontier. We will focus here on the part of the problem concerning the minimization of risk, so that the relevant question we wish to address is "given a set of N assets, what is the optimal allocation of resources that minimizes a certain measure of risk?". In fact, we will face a problem more complex than this, since we will consider the realistic case where the estimation of risk is made on the basis of historical data. Since such data are usually noisy [Laloux et al., 1999], a problem of practical relevance is that of avoiding noise fitting. Statistical mechanics has already given important contributions to this field [Ciliberti et al., 2007; Kondor and Varga-Haszonits, 2008b; Pafka and Kondor, 2004], allowing for the determination of phase diagrams that discriminate in the phase space between a region where the optimization problem may be safely solved and a region where the presence of noise makes the optimization problem unfeasible. The main concern for practitioners, is that problems of real relevance sit at the boundary between the two regions, where large fluctuations set in. We will try to show that a simple solution for this problem may be naturally found by accounting for the impact of liquidation strategies when solving for the optimization problem. We will do this by introducing a feedback between traders' behavior and prices of securities [Eisler et al., 2009], once again explicitly accounting for interactions in financial markets. We will discuss in detail the problem of finding the optimal portfolio under Expected Shortfall (ES) in the case of linear and instantaneous market impact [Caccioli et al., 2010]. We will show that, once market impact is taken into account, a regularized version of the usual optimization problem naturally emerges. We characterize the typical behavior of the optimal liquidation strategies, in the limit of large portfolio sizes, and show how the market impact reduces the instability of ES in this context.

As we said the specific problems addressed in this thesis are in principle quite different, and indeed the concepts and the theories from which we step are usually treated separately in economics books (se for example Bailey [2005]). Nevertheless, we will show that statistical mechanics allows for a description of these systems which is ultimately very similar. Notably, once expressed in terms of systems with random couplings, all the problems under study will be related to the characterization of the minima of quadratic disordered Hamiltonians, and will be solved by means of replica calculations in the zero temperature limit. This approach will allow us to derive, for each system, the free energy in the limit of large markets. The link between the three specific problems under study will be further highlighted by the analysis of the free energies. Indeed we will show that, for each system, the free energy is characterized by a minimum that becomes more and more shallow upon increasing the degree of heterogeneity in the system. Eventually, the free energy will develop flat directions in the phase space and a corresponding degeneracy of minima. All the problems under study will then be characterized in terms of phase transitions that are triggered by the same mechanism, namely the emergence of flat directions in the phase space along which fluctuations may grow unbounded.

Chapter 2

Information efficiency and financial stability

According to Shiller [2008] "*the subprime crisis [...] is, at its core, the result of a speculative bubble in the housing market that begun to burst in the United States in 2006 and has now caused ruptures across many other countries in the form of financial failures and a global credit crunch*". A deep understanding of the reasons that lead to financial bubbles may then be of primary importance in order to develop measures for preventing market crashes. In the economic literature [Galbraith, 1954; Shiller, 2000, 2008; Shiller and Akerlof, 2009], speculative bubbles are often described as the result of irrational euphoria among investors, who believe that "*all will be better, that they are meant to be richer"* [Galbraith, 1954]. This excessive optimism about the future, makes people to invest their money into illiquid things like real estate until prices stop rising, the euphoria turns into panic and the bubble bursts. Although this picture provides an interesting description about the psychological patterns involved, no explanation is provided for what might have triggered the speculative euphoria. In this chapter, we address this point through the discussion, from a statistical mechanics point of view, of the Efficient Market Hypothesis (EMH). Market efficiency refers to the fact that prices of securities correctly reflect all the information available to investors. After a short introduction to the EMH, we will review in section 2.2 a simple model introduced by Berg et al. [2001]. This will allow us to set up a clear framework to understand how markets correctly aggregate information scattered across different investors into prices. In the subsequent section, we present a generalization of such model that accounts for the interplay between informed and non informed traders [Caccioli and Marsili, 2010]. In this context, we will show that non-informed traders take over as the market becomes efficient. When combined with the literature on heterogeneous agents models [Hommes, 2006; Lux and Marchesi, 1999], where it is shown that non-informed traders are responsible for speculative bubbles, our results suggests that market efficiency may indeed create the conditions for bubbles to be triggered.

2.1 Information Efficiency

The concept of market efficiency goes back to one of the most natural questions one can ask concerning financial markets, namely that of their predictability. A possible definition of market efficiency is be given in the following terms [Malkiel, 1992]: "*A market is efficient with respect to an information set if the public revelation of that information would not change the prices of the securities*". In brief, traders who have some information on the performance of an asset will buy or sell shares of the corresponding stock in order to make a profit. As a result, prices will move in order to incorporate this information, thus reducing the profitability of that piece of information. In equilibrium, when all informed traders are allowed to invest, prices must be such that no profit can be extracted from the market.
To give a proper definition of market efficiency, one should also specify the set of information he refers to. Different levels of efficiency, corresponding to different information sets taken into account, have been introduced in the literature according to the following classification (see for example [Bailey, 2005]):

- **weak form:** the set of information includes all current and past prices for the assets in the market.
- **semi-strong form:** the set of information includes all public information available to investors (e.g. news)
- **strong form:** the set of information comprises all information available to investors, including private information.

An important consequence of the Efficient Market Hypothesis is that the time evolution of asset prices should be unpredictable, essentially indistinguishable from a stochastic process like a random walk [Bailey, 2005; Mantegna and Stanley, 2000]. Indeed, if it were possible to predict future price changes an arbitrage opportunity, namely a safe way to make a profit, would appear. According to the common belief, such arbitrage opportunities are destroyed as soon as they start to be exploited, so that efficient arbitrage free markets should be a good approximation of real markets at least for sufficiently long time scales (i.e. much longer than the typical time scale needed for arbitrage opportunities to be eliminated). The validity of the Efficient Market Hypothesis has been object of a long debate since its formulation (for a review see [Bailey, 2005; Lo, 2007]), and no agreement has been achieved in the community so far. However, despite the fact that market can be inefficient, the efficient market hypothesis may always represent a useful benchmark, and indeed standard economics models like the Black-Scholes equation for option pricing are derived assuming market efficiency [Mantegna and Stanley, 2000].
In the following, we will refer to the strong form of information efficiency, for which is possible to introduce a clear and simple framework showing how markets act as information processing devices that aggregate information into prices. The point we are going to make is related to the relation between information efficiency and the behavior of investors. Investors can be classified in broad categories depending on the trading strategies they adopt, and their effect on the market can be very different [Bailey, 2005; Hommes, 2006]. In this respect, it is then quite important to understand under which conditions a specific kind of investors dominate the market. We will consider in the following two classes of investors:

- **fundamentalists**: traders who base their forecasts of future prices upon economics fundamentals (e.g. dividends, interest rates, price to earning ratio). Fundamentalists basically believe that markets may misprice a security in the short run, but that the fundamental price, i.e. the correct price that reflects the underlying reality, will eventually be reached. Profits can then be made by trading the mispriced security and waiting for the market to re-price the security towards its correct value. Fundamentalists are usually associated with a stabilizing effects. Let us suppose for instance that the current price of an asset is much lower than the fundamental one. Fundamentalists will then start buying shares of that stock in order to make profit, since they know that the correct value of the asset is higher than the current price. In doing so fundamentalists are increasing the demand for the asset, whose price will in turn increase becoming closer to the fundamental one. On the other hand, if the current price of the asset were higher than the fundamental one, fundamentalists would start selling causing the price to decrease. The idea is that fundamentalists will keep on buying/selling until the market price matches the fundamental one. Notice that if the market is efficient in the strong sense, fundamentalists cannot make any profit on the basis of their analysis, while in a weakly efficient market they can still make profit since prices at each time only reflect information about past prices.

- **chartists**: traders who base their forecasts upon extrapolation of patterns in past prices, trying to recognize and exploits trends. In some sense chartists believe past prices contain the relevant information they need to predict future trends. A very simple trend following rule may be the that of a linear extrapolation from past prices, so that one should buy if the price of an asset has raised and sell otherwise. At odds with fundamentalists, trend followers are usually associated with a destabilizing effect [Minsky, 1992]. Naively, in a market dominated by trend followers, a small up (down) trend in a stock will cause investors to buy (sell) shares of that stock in such a way to amplify the trend and eventually cause a bubble.

Research in Heterogeneous Agents Models [Hommes, 2006; Lux and Marchesi, 1999] has provided solid support to the thesis that when trading activity is dominated by non-informed traders (e.g. trend followers), bubbles and instabilities develop. In the following, we try to establish a connection between information efficiency and the trading activity of fundamentalists and trend followers, providing support to the idea that *non-informed traders dominate if and only if the market is sufficiently close to information efficiency.* In addition, as markets become informationally efficient, they develop a marked susceptibility to perturbations and instabilities.

2.2 Aggregation of Information in a Complex Market

Berg et al. [2001] introduced an elegant model of a market where agents with different bits of information trade trying to exploit their private information. Interaction between agents is provided through the process of price formation, which occurs according to the balance between demand and supply. This introduces a feedback between agents behavior and price movements, that allows information to be transfered by investors into prices. Under which conditions does the market become informationally efficient, so that no profit can be extracted by agents on the basis of their private information? This is the question we are going to answer in this section, where we introduce the model by Berg et al. [2001] and review its main features.

2.2.1 Definition of the model

We consider a market where a single asset is traded an infinite number of periods. The return of the asset has some degree of uncertainty, that we model in the following manner. We imagine that, at any time, the world can be in any of Ω states, and the return of the asset R^{ω_t} is determined at each period by the draw of a state of nature from the discrete set $\omega_t \in \{1 \dots \Omega\}$, according to a probability distribution π^ω which we take uniform across states in the following. Agents do not observe directly the state of nature, in fact they observe a signal associated to it. A signal is a function from the set of states $\{1 \dots \Omega\}$ to a signal space, that for simplicity is assumed to be $M = \{\pm 1\}$. We denote by k_i^ω the signal observed by agent i if state ω is realized. We focus on a random realization of this setup, so that each agent is endowed with a binary vector whose entries are drawn at random with uniform probability across states, such vector representing the information structure available to the agent. Returns are taken of the form $R^\omega = \overline{R} + \frac{\tilde{R}^\omega}{\sqrt{N}}$, where $\tilde{R}^\omega$ are gaussian variables with zero mean and variance

s^2. According to the signal $k_i^{\omega_t}$ they receive at time t, agents decide to invest a monetary amount $z_i^{k_i^{\omega_t}}$ on the asset. Once agents have invested their money, the price of the asset is computed balancing demand and supply. If we imagine that at each time there are N available units for the asset, the price at time t is determined as

$$Np^{\omega_t} = \sum_{i=1}^{N} \sum_{m=\pm 1} z_i^m \delta_{k_i^{\omega_t},m}. \tag{2.2.1}$$

It is clear at this level that agents interact through the process of price formation and that the interaction is of a mean field type, all agents interacting by means of the average quantity p^{ω_t}. At the end of each period, agents receive an amount R^{ω_t} for each unit of asset they have. If agent i has invested the amount $z_i^{k_i^{\omega_t}}$, she holds $z_i^{k_i^{\omega_t}}/p^{\omega_t}$ units of asset, so that her expected payoff is given by

$$u_i(z_i) = \frac{1}{\Omega} \sum_{\omega} \sum_{m} \delta_{k_i^{\omega_t},m} z_i^m \left(\frac{R^{\omega_t}}{p^{\omega_t}} - 1 \right). \tag{2.2.2}$$

Agents aim at maximizing such expected utility, so that the problem to solve is that of finding the optimal allocations $\{z_i^{\pm}\}$ $i = 1, \ldots, N$. Such problem can be solved with tools borrowed from the statistical mechanics of disordered systems in the limit of large markets, where the number of agents N as well as the number of states Ω is very large $N, \Omega \to \infty$, while their ratio $n = N/\Omega$ is kept fixed. In particular we are interested in understanding if the optimal allocations correspond, in some regime, to an efficient market, where information about returns in fully incorporated into prices, namely $p^{\omega} = R^{\omega}$ $\forall \omega$.

2.2.2 Information structures and information efficiency

Before proceeding to the solution of the optimal allocation problem, we make a little digression in order to discuss the relation between agents' information structure and information efficiency. This will be a way to clarify that we refer here to the strong form of market efficiency. From the information theoretic point of view, the information content of the signal k_i^{ω} can be quantified in one bit. Indeed, the entropy of the unconditional distribution over states, which is $\log \Omega$, is reduced to $\log(\Omega/2)$ by the knowledge of the signal k_i^{ω}. Hence, the information gain is $\log 2$, i.e. one bit. This information gain allows agent i to discriminate between two different conditional distributions of returns, whose means $\mathbb{E}_\pi[R^\omega | k_i^\omega = \pm 1]$ are separated by an amount of order $1/N$. Indeed, for any two states ω and ω', by assumption $R^\omega - R^{\omega'} \sim 1/\sqrt{N}$. Now take the average over the states ω and ω' such that $k_i^\omega = +1$ and $k_i^{\omega'} = -1$ respectively. Given that the expected value of R^ω and $R^{\omega'}$ are the same, one finds that the average of the difference is of the order of the standard deviation of R^ω, times the square root of the number $\Omega/2$ of samples, i.e.

$$|\mathbb{E}_\pi[R^\omega | k_i^\omega = +1] - \mathbb{E}_\pi[R^\omega | k_i^\omega = -1]| \sim s/\sqrt{N\Omega/2} \sim 1/N.$$

This difference is of the same order of the contribution of agents to the price p^ω, hence it allows to differentiate meaningfully their investments z_i^m, depending on the signal they receive.

Let us now discuss market efficiency. A market is efficient with respect to an information set if the public revelation of that information would not change the prices of the securities Fama [1970]. This means that the best prediction of future returns (or prices), conditional on the information set, are present prices. *Strong efficiency* refers to the case where the information set includes the information available to any of the participants in the market, including private information.

In our case, an agent who knew simultaneously the signals k_i^ω of all agents would be able to know the state ω, with probability one, for $\Omega \propto N$ and $N \to \infty$.

Indeed let $N_=$ be the number of pair of states ω and ω' which cannot be distinguished on the basis of the knowledge of all signals. For such pair of states, $k_i^\omega = k_i^{\omega'}$ must hold for all i, because otherwise there would be a signal $k_i^\omega \neq k_i^{\omega'}$ which allows to distinguish ω from ω'. The probability $P\{N_= > 0\}$ that there are at least two states ω and ω' with different returns $R^\omega \neq R^{\omega'}$, but which cannot be distinguished given the signals, is upper bounded by the expected value of $N_=$. The latter can be easily evaluated, since for each pair of states the probability of them not being distinguishable is $P\{k_i^\omega = k_i^{\omega'}, \forall i\} = 2^{-N}$. The number of pairs is $\Omega(\Omega - 1)/2$ so that

$$P\{N_= > 0\} \leq \mathbb{E}[N_=] = \Omega(\Omega - 1)2^{-(N+1)},$$

and this vanishes for $N \to \infty$ in the case $\Omega \propto N$ we consider here. We conclude that, if all signals were revealed, agents would be able to know which state ω has materialized. In this case, prices would not change only if $p^\omega = R^\omega$ for all states ω, which would be equivalent to the strong form of information efficiency Malkiel [1992].

2.2.3 Competitive equilibria

We now come back to the optimal allocation problem, namely the problem of finding the set of $\{z_i^\pm\}$ that maximizes investors' utility. The solution can be achieved both in a static and in a dynamic setting. First of all, it is important to define the type of equilibria we are looking for. We define a competitive equilibrium for the system described above as a set $\{z_i^\pm\}$ such that

- for every agent i and signal m

$$z_i^m \in \text{argmax}_{x \geq 0} \mathbb{E}_\pi \left[x \delta_{k_i^\omega, m} \left(\frac{R^\omega}{p^\omega} - 1 \right) \right] \tag{2.2.3}$$

- the market clears (i.e. aggregate supply matches aggregate demand) $Np^\omega = \sum_{i=1}^N \sum_{m=\pm 1} z_i^m \delta_{k_i^\omega, m}$.

A competitive equilibrium is then a solution of the optimal allocation problems, optimal in the sense that each agents, in correspondence to each signal, invests an amount of money that maximizes her utility, given the amount invested by other agents. Notice that in the competitive equilibrium setting, agents ignore the effect they have on the price, which they take as given. Notice also that if the expected return for a given signal is positive, then agents invest an infinite amount when receiving that signal. Conversely, they do not invest if the expected return is negative.

2.2.4 Learning to trade

An alternative way to solve the optimal allocation problem, is through a learning dynamics with boundedly rational agents. In particular, each agent $i > 0$ has a propensity to invest $U_i^m(t)$ for each of the signals $m = \pm 1$. Her investment

$z_i^m = \Phi(U_i^m)$ at time t is an increasing function of $U_i^m(t)$ ($\Phi : \mathbb{R} \to \mathbb{R}^+$) with $\Phi(x) \to 0$ if $x \to -\infty$ and $\Phi(x) \to \infty$ if $x \to \infty$. After each period agents update $U_i^m(t)$ according to the marginal success of the investment:

$$U_i^m(t+1) = U_i^m(t) + \Gamma \frac{\partial u_i^{\omega_t}}{\partial z_i^m}, \qquad i = 1, \dots, N \tag{2.2.4}$$

where $u_i^{\omega_t} = \sum_m \delta_{k_i^{\omega_t},m} z_i^m \left(\frac{R^{\omega_t}}{p^{\omega_t}} - 1 \right)$, ω_t is the state at time t and p^{ω_t} is the realized price at time t. The idea in Eq. (2.2.4) is that if for a given signal m agent i observes returns R^ω which are higher than prices, she will increase her propensity U_i^m to invest under that signal. Also in this case agents do not take into account their impact on the price, so that the partial derivative is taken at fixed p^{ω_t}.

2.2.5 A Hamiltonian for the system

The stationary properties of the system can be also characterize in terms of the minima of the following Hamiltonian function:

$$H = \frac{1}{2} \sum_{\omega=1}^{\Omega} (R^\omega - p^\omega)^2 , \tag{2.2.5}$$

where $p^\omega = \frac{1}{N} \sum_{i=1}^N \sum_{m=\pm 1} z_i^m \delta_{k_i^\omega,m}$. In order to show that the minima of this Hamiltonian correspond to the competitive equilibria of the market we can proceed in the following way. Let us call$\{z_i^{m*}\}$ the set of variables that minimizes (2.2.15).

- If $z_i^{m*} = 0$ then it must hold

$$\left. \frac{\partial H}{\partial z_i^m} \right|_{z_i^m = 0} = \Omega \mathbb{E}_\pi [(p^\omega - R^\omega) \delta_{k_i^\omega,m}] > 0. \tag{2.2.6}$$

  If on average $p^\omega > R^\omega$ when signal m is observed then it is not convenient for agents to buy shares of the stock, so z_i^m is set to zero.

- If $z_i^{m*} = s_i^*$ then it must hold

$$\left. \frac{\partial H}{\partial z_i^m} \right|_{z_i^m = s^*} = 0, \tag{2.2.7}$$

which is consistent with the fact that if $\mathbb{E}_\pi\left[R^\omega - p^\omega\right] = 0$ agents buy a finite amount of the stock.

The minima of the Hamiltonian correspond then to the competitive equilibria of the market, but what about the stationary solutions of the learning dynamics? The following argument may help in understanding that indeed the learning dynamics converges to the competitive equilibria. Taking the derivative of the Hamiltonian with respect to z_i^m, we see that

$$\frac{\partial H}{\partial z_i^m} = -\Omega \mathbb{E}_\pi \left[U_i^m(t+1) - U_i^m(t)\right]. \tag{2.2.8}$$

If we now consider a continuous time limit for equation (2.2.4), and compute the time derivative of the Hamiltonian we obtain that

$$\dot{H} = \sum_{i,m} \frac{\partial H}{\partial z_i^m} \dot{z}_i^m = -\Omega \sum_{i,m} \mathbb{E}_\omega \left[\left(\dot{U}_i^m\right)^2 \Phi'\left(U_i^m\right)\right] < 0, \tag{2.2.9}$$

given that $\Phi(U_i^m)$ is an increasing function of its argument. This is telling us that the Hamiltonian decreases along the trajectories generated by the learning dynamics, so that in the long time regime such dynamics converges to the minima of H, which in turn correspond to the equilibria of the competitive market.

Some comments are needed in order to clarify how we compute the minima of H. The sought result can be achieved through the computation of the zero temperature limit of the free energy of the system:

$$\lim_{\beta\to\infty} \lim_{N\to\infty} F_N(\beta) = \lim_{\beta\to\infty} \lim_{N\to\infty} \frac{1}{N\beta} \log Z_N(\beta), \tag{2.2.10}$$

where β is the inverse temperature and Z the partition function. However, given the presence of randomness in the system (returns R^ω and information vectors $\{k_i^\omega\}$ are quenched random variables), in principle each observable may depend on the specific realizations of disorder, including the free energy of the system: $F_N(\beta) \rightarrow F_N(\beta|\{R^\omega k_i^\omega\})$, where we have denoted by $F_N(\beta|\{R^\omega k_i^\omega\})$ the free energy for a given realization of the disorder. In principle the above computation should then be done given a realization of the disorder. Nevertheless, in the thermodynamic limit, we expect that the value of quantities which are extensive like the energy or the free energy do not depend on the specific realization of the quenched variables: [1]

$$\lim_{N\to\infty} F_N(\beta|\{R^\omega k_i^\omega\}) = F_\infty(\beta). \tag{2.2.11}$$

Quantities for which this is true are called self-averaging. It is now clear that the average over disorder of a self-averaging quantity will be equivalent to its disorder-independent value [Castellani and Cavagna, 2005; Mezard et al., 1987]:

$$\langle F(\beta|\{R^\omega k_i^\omega\})\rangle = F_\infty(\beta), \tag{2.2.12}$$

where we denoted by $\langle\cdots\rangle$ averages over the disorder. In order to compute the typical properties of the system one needs then to compute

$$\lim_{\beta\to\infty}\lim_{N\to\infty}\frac{1}{N\beta}\langle\log Z_N(\beta|\{R^\omega k_i^\omega\})\rangle. \tag{2.2.13}$$

This can be done resorting to the so called replica trick, namely exploiting the identity

$$\log Z = \lim_{m\to 0}\frac{Z^m - 1}{m} \tag{2.2.14}$$

and computing the partition function for a system made of m replicas of the original system. A section in the appendix will be devoted to a detailed calculation for the system under study. In the following we just focus on the results.

2.2.6 Transition to efficient market

As we have seen, both the study of competitive equilibria and the asymptotic properties of the learning dynamics turn into the characterization of the minima of the Hamiltonian function

$$H = \frac{1}{2}\sum_{\omega=1}^{\Omega}(R^\omega - p^\omega)^2\,, \qquad p^\omega = \frac{1}{N}\sum_{i=1}^{N}\sum_{m=\pm1} z_i^m \delta_{k_i^\omega,m}. \tag{2.2.15}$$

[1] Indeed in the large N limit we can imagine to divide the system into a large number of sub-systems much smaller than the original systems but still macroscopic. At this point we can express the total energy of the system as the sum of the energies of the sub-systems. A law of large numbers can then be invoked to show that the average properties of the system do not depend on the realizations of disorder.

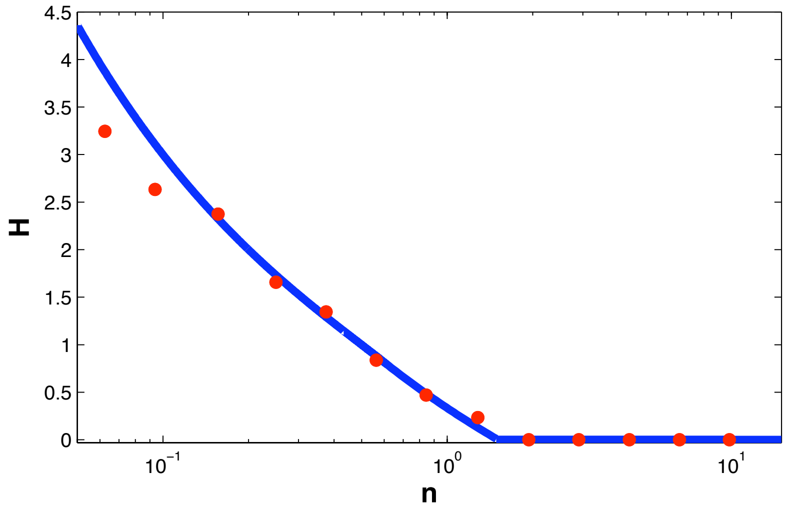

Figure 2.1: *Distance* $|p - R| = \sqrt{\sum_{\omega=1}^{\Omega} \left(R^{\omega} - p^{\omega,k_0}\right)^2}$ *of prices from returns in competitive equilibrium in absence of trend followers and zero information cost.*

The function H represents the squared distance of prices from returns, and is therefore a measure of market efficiency. The main result by Berg et al. [2001] is that, as more and more different types of informed agents enter the market, prices approach returns. In particular they show that there is a critical value n_c of informed traders beyond which $H = 0$, which implies that prices equal returns ($p^{\omega} = R^{\omega}$) for each state $\omega = 1, \ldots, \Omega$. Figure 2.1 shows the transition towards an efficient market upon increasing the number of agents with different information. The solid line refers to the analytical computation of the competitive equilibria achieved through the minimization of the disordered Hamiltonian (3.2.15), while dots refer to simulations made with the learning scheme. In passing, we can also see from the figure that the two approaches indeed lead to the same solutions. The region $H = 0$ is also characterized by a divergent *susceptibility*, which means that allocations $\{z_i^m\}$ have a marked dependence on structural parameters. The susceptibility χ relates a small uncertainty in a structural parameter, such as e.g. R^{ω}, to the uncertainty in allocations $\delta z_i^m \simeq \chi \delta R^{\omega}$. A divergent susceptibility $\chi \to \infty$ signals the fact that equilibria with different allocations are possible even for the same structural parameters, i.e. that the minimum of Eq. (2.2.15) is not unique.

The model by Berg et al. [2001] is quite important, because it shows in a clear framework how markets manage to aggregate information through the mechanism of price formation. Moreover, it shows at same time that the attempt by traders to exploit their information ends up in destroing the profitability of that information. Indeed if we look at the expected payoff of agents

$$u_i(z_i) = \frac{1}{\Omega}\sum_{\omega}\sum_{m}\delta_{k_i^\omega,m} z_i^m \left(\frac{R^\omega}{p^\omega} - 1\right)$$

we see that agents' profits are reduced as market become efficient, i.e. as $p^\omega \to R^\omega$. It is important to stress that information efficiency appears here as an emergent property, namely as the result of the interaction between agents, rather than being postulated from the beginning as often done in economics modeling.

The setting outlined above is our starting point for the next section, where we try to elucidate the relation between information efficiency and different trading strategies based on private or public information. This is indeed a very interesting topic, as it was realized long ago by Grossman and Stiglitz [1980]. Paradoxically, when markets are really informationally efficient, traders have no incentive to gather private information, because prices already convey all possible information. Hence traders' behavior doesn't transfer anymore information into prices, implying that efficient markets cannot be achieved in the long run. In the next section, we basically provide a statistical mechanics framework for the Grossman-Stiglitz paradox.

2.3 Market Efficiency, Informed and Non-Informed Traders

In this section we are going to consider a generalization of the model by Berg et al. [2001] presented above. In particular we will introduce non-informed traders who invest according to public available information and we will introduce an information cost for fundamentalists: traders that invest according to a private information structure have to pay a cost in order to gather their information. Non-informed traders, on the other hand, will mimic in our framework the behavior of chartists, e.g. trend followers.

2.3.1 The model

As before we consider a market where a single asset is being traded at discrete times $t = 0, 1, 2, \ldots$. Let there be N informed traders (fundamentalists) and N' uninformed traders (chartists) operating in the market. For simplicity, we assume that all uninformed traders adopt the same trading strategy, so that the description of non-informed traders can be given in terms of a single representative agent. We thus set $N' = 1$ and we shall refer to the chartist representative agent as agent $i = 0$[1]. In the market there are N units of asset available at each time and at the end of each period the asset pays a return. The return depends only on the state of nature in that period, $\omega = 1, \ldots, \Omega$, and is denoted by R^ω. The state of nature is determined, in each period, independently according to the uniform distribution on the integers $1, \ldots, \Omega$. Traders do not observe the state directly. Informed traders ($i = 1, \ldots, N$) receive a signal on the state according to some fixed private information structure, which is determined at the initial time and remains fixed. As before, the signal space is assumed to be $M = \{-, +\}$ and we denote by $k_i^\omega \in M$ the signal observed by trader i in state ω. Trader $i = 0$, instead, does not receive any signal on the state ω, but she observes a public variable $k_0 \in \{-1, +1\}$, which is drawn independently at random in each period[2].

Like in the previous section, we focus on a random realization of this setup, where the value of the return R^ω in state ω is drawn at random before the first period, and does not change afterwards. Returns thus only change because the state of nature changes. We then take R^ω Gaussian with mean $\bar{R}$ and variance s^2/N. Likewise, the information structure is determined by setting $k_i^\omega = +1$ or -1 with equal probability, independently across traders i and states ω. At the

[1]Similarly, each one of the informed agents labeled as $i = 1, \ldots, N$ is in fact a representative agent for all the agents acting according to a specific information structure.

[2]The case where k_0 depends on past market data (e.g. price differences) introduces no qualitative difference.

beginning of each period, a state ω and a public information k_0 are drawn, and private information k_i^ω is revealed to informed agents ($i > 0$). All traders decide to invest a monetary amount z_i^m in the asset: here $m \in M$, for $i > 0$, takes the value of the signals k_i^ω which agent $i > 0$ receives, whereas it equals k_0 for $i = 0$. The price of the asset p^{ω,k_0} is then derived from the market clearing condition

$$Np^{\omega,k_0} = \sum_{i=1}^{N} \sum_{m=\pm 1} z_i^m \delta_{k_i^\omega,m} + \sum_{m=\pm 1} z_0^m \delta_{k_0,m}, \tag{2.3.16}$$

which is the generalization of equation (2.2.1). Agents do not know the price at which they will buy the asset when they decide their investment z_i^m. The price depends on the state ω and on k_0 because the amount invested by each agent depends on the signal they receive, which depends on ω Pliska [1997]; Shapley and Shubik [1977]. The expected payoff of agents is

$$u_i^{k_0}(z_i) = \frac{1}{\Omega} \sum_{\omega} \sum_{m} \delta_{k_i^\omega,m} z_i^m \left(\frac{R^\omega}{p^{\omega,k_0}} - 1 \right). \tag{2.3.17}$$

How will agents choose their investments? As before, one can consider either competitive equilibria or take a dynamical approach where agents are assumed to learn over time how to invest optimally.

Competitive equilibria: agents aim at maximizing their expected utility, but they consider the price of the stock as given, thus neglecting their impact in the market. A competitive equilibrium for this model is given by a set of variables $\{z_i^\pm\}$ such that

- for every agent i and signal m

$$z_i^m \in \text{argmax}_{x \geq 0} \mathbb{E}_\pi \left[x \left(\delta_{k_i^\omega,m} \left(\frac{R^\omega}{p^\omega} - 1 \right) - \epsilon \frac{1 - \delta_{i,0}}{N} \right) \right]. \tag{2.3.18}$$

At odds with Berg et al. [2001], we also take into account ,through the term proportional to ϵ, that fundamentalists face a cost for the private information they gather. More precisely, investment is considered attractive only if the returns under signal m exceed prices by more than ϵ/N. No cost is instead charged on agent $i = 0$, since her trading only relies on public available information.

- the market clearing equation (2.3.16) holds.

Learning dynamics: As before, each informed agent $i > 0$ has a propensity to invest $U_i^m(t)$ for each of the signals $m = \pm 1$ and her investment $z_i^m = \Phi(U_i^m)$ at time t is an increasing function of $U_i^m(t)$ satisfying $\Phi(x) \to 0$ if $x \to -\infty$ and $\Phi(x) \to \infty$ if $x \to \infty$. At the end of each period, agents update $U_i^m(t)$ as

$$U_i^m(t+1) = U_i^m(t) + \left(R^{\omega_t} - p_t^{\omega_t, k_0}\right) \delta_{k_i^{\omega_t}, m} - \frac{\epsilon}{N} \delta_{k_i^{\omega_t}, m}, \qquad i = 1, \ldots, N. \tag{2.3.19}$$

Similarly, the non-informed agent updates her propensity to trade according to

$$U_0^m(t+1) = U_0^m(t) + \left(R^{\omega_t} - p_t^{\omega_t, k_0}\right) \delta_{k_0, m} \tag{2.3.20}$$

and invests an amount $z_0^m = \Phi(U_0^m)$, depending on the value $m = k_0$ of public information at time t.

Hamiltonian: These two different choices are going to bring to the same equilibria, that are again given by the minimization of a function, which takes the form

$$H_\epsilon = \frac{1}{4\Omega} \sum_{\omega=1}^{\Omega} \sum_{k_0 = \pm 1} \left[R^\omega - p^{\omega, k_0}\right]^2 + \frac{\epsilon}{2N^2} \sum_{i=1}^{N} \sum_{m=\pm 1} z_i^m, \tag{2.3.21}$$

where p^{ω, k_0} is given in Eq. (2.3.16) in terms of z_i^m, $i = 0, \ldots, N$, $m = \pm 1$. Exactly as before, it is possible to prove on one side that the condition for the minimum of the Hamiltonian is equivalent to the condition that gives competitive equilibria. On the other side, it is possible to show that the learning dynamics converges in the stationary regime towards the minima of H_ϵ.

2.3.2 Efficiency and Stability

The model we just described is a generalization of the one introduced by Berg et al. [2001], which can be recovered in the limit where no trend followers are allowed to trade and $\epsilon = 0$. What happens when we introduce chartists ($z_0 > 0$)

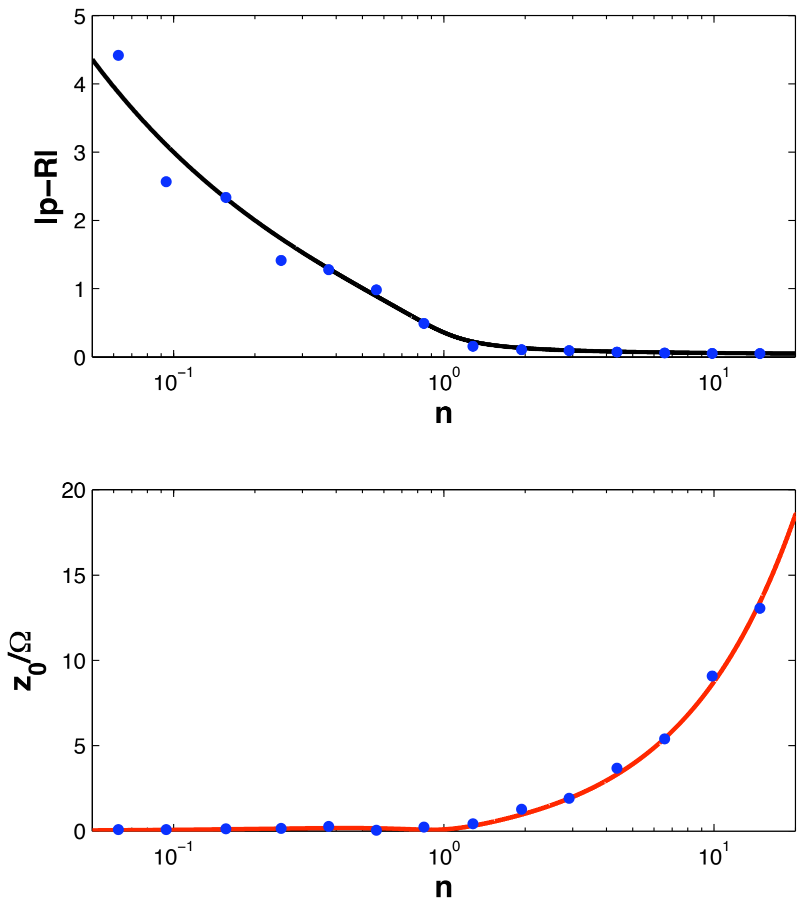

Figure 2.2: *Top panel: distance* $|p-R| = \sqrt{\sum_{\omega=1}^{\Omega} E_{k_0}\left[R^{\omega} - p^{\omega,k_0}\right]^2}$ *of prices from returns in competitive equilibrium. The full line represents the analytical solution for the case* $s = \overline{R} = 1$ *and* $\epsilon = 0.1$*, points refer to numerical simulations of systems with* $\Omega = 32$*,* $s = \overline{R} = 1$ *and* $\epsilon = 0.1$*. Bottom panel: monetary amount invested by the trend follower* z_0 *for the same values of the parameters.*

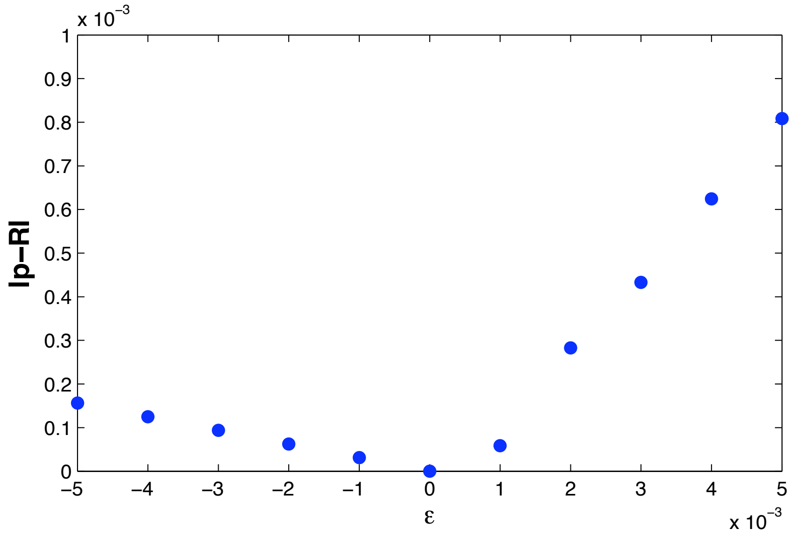

Figure 2.3: *Measure of market efficiency as a function of information cost.*

and information costs ($\epsilon > 0$)? Some simple heuristic arguments can be useful in order to understand the basic behaviour of the system. Let us consider the Hamiltonian (2.3.21). In the case of small ϵ and small $n = N/\Omega$, the first term in Eq. (2.3.21) dominates the second and the minimum is expected to be close to that without chartists. When n increases, however, the value of H (i.e. the first term in equation (2.3.21)) decreases making the two terms comparable. When this happens, i.e. when $n \approx n_c$ and $H \approx 0$, then it starts to become possible to achieve a small value of H_ϵ by decreasing the size of the second term increasing, at the same time, z_0^m in order to keep average prices of the same order of average returns. Hence we expect z_0^m to be large and of order N when the market becomes close to being information-efficient. The results of numerical simulations as well as the analytical solution for competitive market equilibrium (see Appendix A.1 for more details), shown in Fig. 2.2, confirm this picture. Upon increasing the number of informed agents, the system approaches the limit of efficient market. Correspondingly, the share of trades due to uninformed agent starts raising only once information has been aggregated by informed traders. It has to be noticed that the introduction of the information cost ϵ makes sure that a perfect efficiency of the market is recovered only at $\epsilon = 0$ (see Figure 2.3). It is then instructive

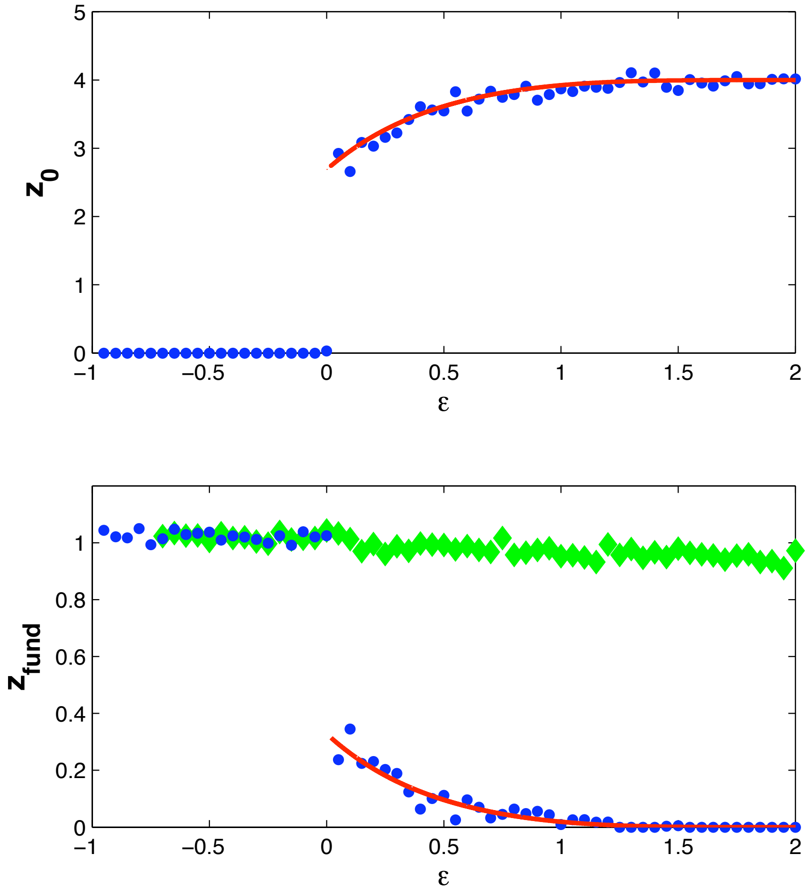

Figure 2.4: *Top panel: monetary amount invested by the trend follower. Bottom panel: monetary amount invested by a fundamentalist in presence (blue points) or absence (green diamonds) of the trend follower. Points refer to simulations of systems with* $\Omega = 32$*,* $n = 4$ *and* $s = \overline{R} = 1$*. Full lines represent the corresponding analytical solution.*

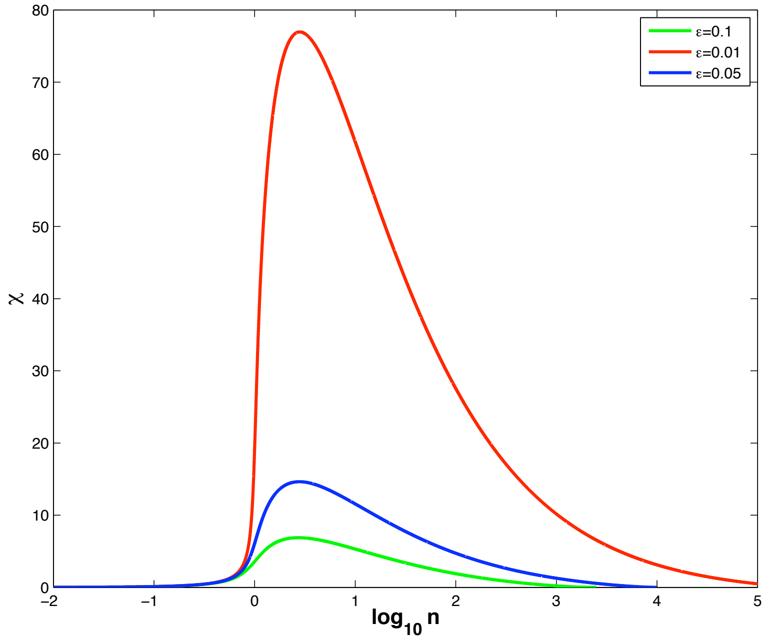

Figure 2.5: *Susceptibility as a function of market complexity for different information costs.*

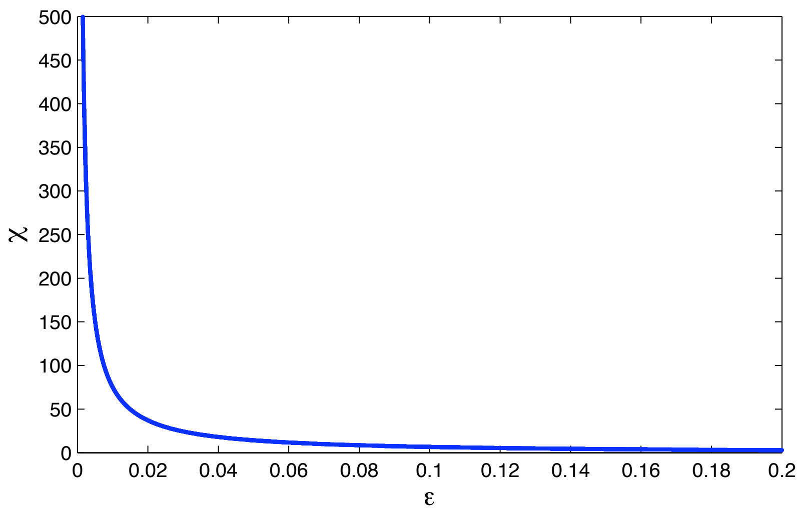

Figure 2.6: *Susceptibility as a function of information cost.*

to look at the behaviour of chartists as a function of ϵ[1]. Figure 2.4 shows signatures of a phase transition occurring at $\epsilon = 0$. Indeed, for $\epsilon < 0$ chartists barely operate in the market, while they start trading as soon as $\epsilon > 0$. As we mentioned in section (2.1), a market dominated by trend followers is a market where bubble phenomena may occur. Our results, once combined with the insights coming from Heterogeneous Agents Models [Hommes, 2006], suggest that market efficiency, usually considered a necessary condition for an ideal market, may in fact be a necessary condition for bubble phenomena to be triggered. In order to further explore the relation between information efficiency and market stability, it is instructive to look at the behavior of the susceptibility of informed traders, that quantifies their response to small changes in the constitutive parameters of the model. As shown in Figure 2.5, the susceptibility as a function of n is characterised by a peak in correspondence to the crossover towards an (almost) efficient market. The information cost ϵ plays here a nontrivial role, as evidenced by the behaviour of the susceptibility as the information cost ϵ is decreased. In this situation the peak in the susceptibility becomes more pronounced. When information costs vanish ($\epsilon \rightarrow 0$) the susceptibility diverges in the whole efficient phase, i.e. for n large enough ($n \geq n_c$). In Figure 2.6 we show the behavior of the susceptibility as a function of the information cost in the efficient phase. When there are no information costs ($\epsilon = 0$) the susceptibility is strictly infinite in all the region of phase space where the market is efficient ($n \geq n_c$). It is interesting to make some considerations, in this respect, about possible extensions of this framework where information costs are endogenous. It may be reasonable to assume that information should be cheaper and cheaper the more the market is close to information efficiency. Indeed, the more prices reflect private information, the less incentives agents have to gather information. This suggests that, with endogenous information costs, the market should approach the critical line $\epsilon = 0^+$ with $n \geq n_c$, where the behavior of informed traders is characterized by infinite susceptibility. In other words, the path towards an efficient market may be a path towards an unstable market.

[1] Notice that the interpretation of ϵ as information cost is meaningful only for non negative values of ϵ. However, since 2.3.21 is well defined for all real values of ϵ, we allow the parameter to take negative values to better characterize the properties of the Hamiltonian.

2.4 Summary and Perspectives

We have shown that, in a simple asset market model, non-informed traders contribute a non-negligible fraction of the trading activity only when the market becomes informationally efficient. In the simple setting studied here, non-informed traders do not have a destabilizing effect on the market as in the models of Hommes [2006]. At the same time, when non-informed traders dominate, their activity does not spoil information efficiency. This is due to the fact that the market discussed here is basically the repetition over time of the same single period framework. Nevertheless, we can see from our analysis that information efficiency is associated with a phase transition in the statistical mechanics sense, characterized by strong fluctuations and sharp discontinuities in the optimal allocations. This suggests that market efficiency carries in fact some seeds of instability. Moreover, when combined with the insights of the literature on Heterogeneous Agent Models [Hommes, 2006], the very fact that non-informed traders start trading massively when market efficiency is approached suggests in fact that information efficiency can trigger the occurrence of bubbles and instabilities. This issue has also been recently addressed by Goldbaum [2006], however the analysis was limited to a single type of fundamentalists ($N = 1$ in our case). A stronger case would require first to extend the framework of Hommes [2006] to the case of fundamentalists with many different types of private information, recovering a picture for information efficiency similar to that provided by Berg et al. [2001]. Then one should investigate the effect of introducing non-informed traders, i.e. genuine trend-followers. Besides understanding whether information efficiency is also in that case a necessary condition for non-informed traders to dominate, one could also address the interesting question of the effect of chartists on information efficiency. Ultimately, the discussion of these results suggests that excessive insistence on information efficiency in market regulation policies , as e.g. in the debate on the Tobin tax [Haq et al., 1996], could have the unintended consequence of propelling financial bubbles, such as those which have plagued international financial markets in the recent decades.

Chapter 3

Proliferation of derivatives and market stability

One of the most debated questions in relation to the recent economic crisis has certainly been that of the role of derivatives. A derivative is a financial contract whose value is linked to the future price movements of an underlying asset. Examples of derivatives are futures contracts, options, and mortgage-backed securities. Financial markets have increased tremendously in size and complexity in the last decade [Fund, 2008a,b] and the volume of exchanged derivatives has been often much larger that the corresponding underlying market. Notwithstanding the worrying warning by Warren Buffet "*In my view, derivatives are financial weapons of mass destruction, carrying dangers that, while now latent, are potentially lethal*" [Buffet, 2002], the increase in the repertoire of available financial instruments that we witnessed in the past years has been welcomed by most practitioners. Indeed, the growth in financial complexity as well as the unfettered access to trading were expected [Merton and Bodie, 2005] to drive the market closer to the theoretical limit of arbitrage free, complete market described by the Arbitrage Pricing Theory (APT) [Pliska, 1997], the theory at the basis of financial engineering.
APT makes it possible to give a present monetary value to future risks, and hence to price complex derivative contracts. In order to do this, APT relies on idealizations which allow one to neglect completely the feedback of trading on market's dynamics. These concepts are very powerful, and APT has been quite successful in stable market conditions. In addition, the proliferation of financial instruments

provides even further instrument for hedging and pricing other derivatives. So the proliferation of financial instruments produces precisely that arbitrage-free, complete market which is described by APT. Paradoxically, the recent economic crisis occurred in the period of greater expansion of financial markets, when real markets were closer as never before to the perfect world of APT. In the following, we will try to solve this apparent paradox by means of a description of financial markets in terms of an interacting system.
After a short introduction to APT, we introduce a model of a simple market where derivatives on an underlying market are traded. We will account in particular for a feedback between the derivative market and the underlying one. Indeed, the trading in derivatives generates demand in the underlying by banks trying to hedge risk. We will show that uncontrolled proliferation of derivatives drives the market towards the limit of ideal market considered in APT. We will also show, however, that as the market becomes complete it also approaches a critical line where a phase transition in the statistical mechanics sense occurs, characterized by large fluctuations and instabilities.

3.1 Asset Pricing

In this section we try to introduce the basic concepts of Arbitrage Pricing Theory by means of a simple example. Let us consider a financial market where a risk-less asset and a risky asset are issued. We focus on a single period framework, so that there are two relevant times $t = 0$ (today) and $t = 1$ (tomorrow). In order to model the uncertainty about the future, we say that the market at $t = 1$ can be in one of two states $\omega \in 1, 2$, according to a probability distribution π^ω. The risk-less asset cost 1 today and pays $B(\omega) = 1 + r$ tomorrow , while the risky asset costs 1 today and pays $S(1) = 1 + u$ if $\omega = 1$ or $S(2) = 1 - d$ if $\omega = 2$. Imagine now that a contract C is introduced that pays C^ω at $t = 1$. The question we want to answer is that of determining the correct price of this contract at time 0. In order to do this, we can try to replicate the payoff of C by means of a linear combination of the two assets B and S, namely we write

$$C = \alpha_B B + \alpha_S S \tag{3.1.1}$$

and we look for coefficients that satisfy

$$C(1) = \alpha_B(1+r) + \alpha_S(1+u) = C^1 \tag{3.1.2}$$
$$C(2) = \alpha_B(1+r) + \alpha_S(1-d) = C^2. \tag{3.1.3}$$

The solution of these equations leads to

$$\alpha_B = \frac{C^2(1+u) - C^1(1-d)}{(1+r)(u+d)} \tag{3.1.4}$$
$$\alpha_S = \frac{C^1 - C^2}{u+d}. \tag{3.1.5}$$

Now that we have a portfolio that reproduces the payoffs of assets S at $t = 1$, it is reasonable to set the price of the contract C at $t = 0$ equal to the one of the replicating portfolio, namely

$$p_C = \alpha_B + \alpha_S = \frac{1}{1+r}\left[\frac{r+d}{u+d}C^1 + \frac{u-r}{u+d}C^2\right]. \tag{3.1.6}$$

Notice that this can be written as

$$p_C = \frac{1}{1+r}\mathbb{E}_q[C^\omega], \tag{3.1.7}$$

where we have defined the probability distribution q^ω as $q^1 = \frac{r+d}{u+d}$, $q^2 = \frac{u-r}{u+d}$. The above result tells us that the price of the asset C is the present discounted value of its expected future payoff, and is one of the fundamentals results of arbitrage pricing theory. Notice, in passing, that the computed price does not depend on the probability over states π^ω.

The probability measure q introduced above is usually called **equivalent martingale measure** or **risk neutral measure**. What is the meaning of such a measure? In a market where both assets S and B are traded, agents are typically indifferent with respect to the two assets. This means that the expected payoff of the two should be the same. The equivalent martingale measure is the probability distribution which implements this statement: $\mathbb{E}_q[B] = \mathbb{E}_q[S] = 1 + r$. The risk neutral measure is then the probability distribution used by investors in order to model the uncertainty of the market. Notice that for the above probability to be meaningful some conditions must be satisfied, notably $r > -d$ and $u > r$ [1]. Let us comment on the meaning of such conditions. Let us suppose for instance that $u > r$, but also $-d > r$: in that case asset S would outperform asset B for both states $\omega = 1, 2$. In an analogous (but opposite) way, if $-d < r$ but $u < r$ S would over-perform B. In such situations, an investor who goes short[2] on the dominated asset and buys shares of the dominating one would make a profit without bearing any risks, exploiting what is known as an **arbitrage opportunity**. Notice that the presence of arbitrage opportunities affects equation (3.1.6) through the presence of negative weights, so that it is no more possible to define an equivalent martingale measure in presence of arbitrage opportunities. This is basically the content of the arbitrage pricing theorem, which states that, in a market where there are no arbitrage opportunities, there exists an equivalent martingale measure.

Notice that an essential element that allowed us to compute a unique price for the contract C was that the number of assets was equal to the number of states. Let us try to generalize the above ideas to the case with two assets (B and S) but three states ($\omega \in \{1, 2, 3\}$). Let us look for an equivalent martingale measure q^ω, i.e. for a risk measure such that $\mathbb{E}_q[B] = \mathbb{E}_q[S] = 1 + r$. In this case, one has then to satisfy a system of two equations by fixing the value of the three variables q^1, q^2 and q^3. Different risk neutral measures can thus be defined, so that investors may not agree in assigning a price to the contract. In this situation, market are usually referred to as incomplete markets, in contrast to the case of **complete markets**, where the number of independent assets is equal to the number of states.

[1] Notice that these conditions also imply $u + d > 0$, since $u > r > -d \rightarrow u + d > r + d > 0$. An alternative set of conditions would be $r < -d$ and $u < r$, that would in turn imply $u + d < 0$, preserving the meaning of probabilities for the coefficients of equation (3.1.6).

[2] short selling is the practice of selling assets that have been borrowed from a third party with the intention of buying identical assets back at a later date to return to the lender. The short seller hopes to profit from a decline in the price of the assets between the sale and the repurchase, as the seller will pay less to buy the assets than the seller received on selling them.

3.1.1 The world of asset pricing

The considerations of the previous section can be generalized, in a more systematic way, in the so called Arbitrage Pricing Theory [Pliska, 1997]. A more general framework can be described as follows. There are only two times $t = 0$ (today) and 1 (tomorrow). The world at $t = 1$ can be in any of Ω states and π^ω is the probability that state $\omega = 1, \ldots, \Omega$ occurs. We consider a market where one risk-less asset (bond) and K risky assets are traded. The risk-less asset costs one today and pays one tomorrow, in all states[1]. The price for the k-th risky assets is one at $t = 0$ and is $1 + r_k^\omega$ at $t = 1$ if state ω materializes, where $k = 1, \ldots, K$. Prices of assets are assumed given at the outset. Portfolios of assets can be built in order to transfer wealth from one state to the other. A portfolio $\vec{\theta}$ is a linear combination with weights θ_k, $k = 0, \ldots, K$ on the riskless and risky assets. The value of the portfolio at $t = 0$ is

$$V_\theta(t = 0) = \sum_{k=0}^{K} \theta_k,$$

which is the price the investor has to pay to buy $\vec{\theta}$ at $t = 0$. The return of the portfolio, i.e. the difference between its value at $t = 1$ and at $t = 0$, is given by

$$r_\theta^\omega \equiv \sum_{k=1}^{K} \theta_k r_k^\omega.$$

The content of the theory relies on the following steps

No-arbitrage It is assumed that returns r_k^ω are such that there is no portfolio $\vec{\theta}$ whose return $r_\theta^\omega \geq 0$ is non-negative for all ω and strictly positive on at least one state ω [2].

Equivalent martingale measure The absence of arbitrages implies the existence of an Equivalent Martingale Measure (EMM) q^ω which satisfies

$$E_q[r_k] \equiv \sum_{\omega} q^\omega r_k^\omega = 0, \qquad \forall k = 1, \ldots, K.$$

[1]This is equivalent to considering, for the sake of simplicity, discounted prices, namely $r = 0$ in the language of the previous section, right from the beginning).

[2]Namely there is no way to make a profit without bearing any risks.

Let us try to give some intuition in order to understand the important relation between the absence of arbitrage opportunities and the existence of risk neutral measures.. Let us consider the simplest case of a market where there is just one risky asset. An arbitrage opportunity in this case would appear if, for instance, $r_1^\omega \geq= 0\ \forall\omega > 1$ and $r_1^1 > 0$. It is clear that, in this situation, it is not possible to find any set of positive values for the Ω variables q^ω satisfying $\sum_\omega q^\omega r_1^\omega = 0$, so that no equivalent martingale measure exists if arbitrage opportunities arise.

Valuation of contingent claims A contingent claim f is a contract between a buyer and a seller where the seller commits to pay an amount f^ω to the buyer dependent on the state ω (the contract C that we considered in the previous section was a contingent claim). If the seller can build a portfolio θ of securities such that $f^\omega = f_0 + r_\theta^\omega$, i.e. which replicates f, then the seller can buy the portfolio and ensure that she can meet her commitment. The value of the replicating portfolio provides then a price for the contract f, which can be expressed in the form

$$V_f = E_q[f] = \sum_\omega q^\omega f^\omega$$

Claims f for which this construction is possible are called marketable. If this is not possible, the parties may differ in their valuation of f because of their different perception of risk. Put differently, there may be many different EMM's consistent with the absence of arbitrage, each giving a different valuation of the contract f.

Complete markets If there are at least Ω linear independent vectors among the vectors $(r_k^1, \ldots, r_k^\Omega)$, $k = 1, \ldots, K$ and $(1, \ldots, 1)$, then any possible claim is marketable, which means that it can be *priced*. In such a case the market is called *complete* and the risk neutral measure is unique.

Summarizing, the logic of financial engineering is: assuming that markets are arbitrage free, the price of any contingent claim, no matter how exotic, can be computed. This involves some consideration of risk, as long as markets are incomplete. But if one can assumes that markets are complete, then prices can be computed in a manner which is completely independent of risk. Note indeed that the probability distribution over π^ω plays no role at all in the above construction. It is also worth to remark that asset returns r_k^ω do not depend on the type of portfolios which are traded in the market. A complete, arbitrage free market is the best of all possible worlds and one is tempted to argue that this is indeed a good approximation of real financial markets when these markets expand in both complexity and volumes.

3.2 A picture of the market as an interacting system

Why should markets be arbitrage free? According to the standard folklore, this is because otherwise "everybody would 'jump in' [...] affecting the prices of the security" Pliska [1997]. If an arbitrage opportunity appears, many investors will indeed try to exploit it in order to make profit. As a result, prices of securities will move in such a way as to eliminate that arbitrage opportunity. As we already said, within the context of APT prices of risky assets are fixed at the outset. However, the very fact that those assets are traded implies that some information has been aggregated into their prices [Bouchaud et al., 2008]. Moreover, trading in derivative contracts implies a trading activity in the underlying market by issuers of such contracts, who need to meet their commitments. When the number of traders and the volume of traded assets are very large, usually such impact on the price is neglected. In the following we want however to account for this (microscopic) feedback between underlying and derivative markets, in order to see if such interactions may become relevant in some regimes. Let us then assume that prices are affected not only when "everybody jumps in" but anytime someone trades, even though for individual trades the effect is very small. We will consider a market where some assets are traded, and we will account for the presence of a creative financial industry that develops derivative contracts based on the underlying assets. In this picture, derivatives will simply be contracts which deliver a particular amount of assets in a given state. In the simplified picture of the market we shall discuss below, prices depend on the balance between demand

and supply: if demand is higher than supply return is positive, otherwise it is negative. Demand comes from either an exogenous state contingent process or is generated from the derivative market in order to match contracts. Even if markets are severely incomplete, financial institutions (which we shall call banks for short in what follows) will match the demand for a particular derivative contract if that turns out to be profitable, i.e. if the revenue they extract from it exceeds a risk premium. We will be interested in characterizing the behavior of the market as the number of derivatives (financial complexity) increases. Indeed one should expect that, when financial complexity is large enough, the market becomes complete because each claim can be replicated by a portfolio of bond and derivatives. So, the proliferation of financial instruments leads finally to the efficient, arbitrage-free, complete market described by APT. However, we will show that the road to such ideal markets can be plagued by singularities which arise upon increasing financial complexity.

3.2.1 A stylized model

We consider the one period framework described above, and we denote by π^ω be the probability that state $\omega = 1, \ldots, \Omega$ materializes. For simplicity, we imagine there is a single risky asset[1] and a risk-less asset. We shall implicitly take discounted prices, so we set the return of the risk-less security to zero. The price of the risky asset is 1 at $t = 0$ and $1 + r^\omega$, in state ω. However, rather than defining at the outset the return of the asset in each state, we assume it is fixed by the law of demand and supply. In the market we consider, banks develop and issue financial instruments (derivative contracts). In this simplified world, a financial instrument is a pair (c, a^ω) where c is the cost payed to the bank at $t = 0$ by the investor and a^ω is the amount of risky asset it delivers in state ω at $t = 1$. We imagine there is a demand s_0 for each of N possible financial instruments (c_i, a_i^ω), $i = 1, \ldots, N$. The return of the asset at $t = 1$ in each state ω is given by

$$r^\omega = d_0^\omega + \sum_{i=1}^{N} s_i a_i^\omega \qquad (3.2.8)$$

[1] Generalization to more assets is straightforward.

where d_0^ω is assumed to arise from investors' excess demand whereas the second term is generated by banks hedging derivative contracts, s_i being the supply of instrument i provided by banks. Since we wish to reproduce the typical features of APT, we assume the existence of a risk neutral measure q^ω such that

$$\mathbb{E}_q\left[r^\omega\right] = 0 \tag{3.2.9}$$

In order to enforce this condition in the following we will take $\mathbb{E}_q\left[d^\omega\right] = 0$ as well as $\mathbb{E}_q\left[a_i^\omega\right] = 0 \ \ \forall i$. Notice that the condition for the exogenous demand is required because (3.2.9) should hold also in absence of traded derivatives. The condition on the derivative contracts is instead more restrictive than required by (3.2.9), so we take it as a simplifying assumption. Moreover, we assume that the risk neutral measure is given exogenously.

The supply s_i of instrument i is fixed by banks in order to maximize their expected profit, which for a unit supply ($s_i = 1$) is given by:

$$u_i = \left[c_i - \sum_{\omega=1}^{\Omega} \pi^\omega a_i^\omega (1 + r^\omega)\right]. \tag{3.2.10}$$

Banks will sell derivatives if the expected profit is large enough, and will not sell it otherwise. Specifically, we assume that banks match the demand of investors (i.e. $s_i = s_0$) if $u_i > \bar{u}_i$, whereas $s_i = 0$ if $u_i < \bar{u}_i$. When $u_i = \bar{u}_i$ the supply takes instead a finite value $s_i \in (0, s_0)$. Here $\bar{u}_i$ can be interpreted as a risk premium related to the risk aversion of banks.

3.2.2 A typical large complex market

Financial markets are quite complex, with all sorts of complicated financial instruments. This situation is reproduced in our framework by assuming that demand d_0^ω and derivatives a_i^ω are drawn independently at random. Furthermore we take the limit $\Omega \to \infty$. This situation clearly defies analytical approaches for a specific realization. However, it is possible to characterize the statistical behavior of typical realizations of such large complex markets. In order to do that, some comments on the scaling of different quantities, with the number of states Ω, is in order. The interesting region is the one where the number of derivatives N is of the same order of the number of states Ω ($N \sim \Omega$). In this regime we expect the market to approach completeness. For this reason we introduce the variable $n = N/\Omega$. Assuming that r^ω is a finite random variable in the limit $\Omega, N \to \infty$, requires d_0^ω to be a finite random variable – e.g. normal with mean $\bar{d}$ and variance Δ. Likewise, we shall take a_i^ω as random variables with zero average and variance $1/\Omega$. Indeed, the second term of Eq. (3.2.8) with $a_i^\omega \sim 1/\sqrt{\Omega}$ is of the order of $\sqrt{N/\Omega}$, which is finite in the limit we are considering[1].

It is convenient, in the following discussion, to introduce the parameters

$$\frac{\epsilon_i}{\Omega} \equiv \bar{u}_i - \left[c_i - c_i^{(0)}\right]. \tag{3.2.11}$$

where $c_i^{(0)} = \sum_\omega \pi^\omega a_i^\omega$ is the expected price of instrument i, at $t = 0$. The dependence on Ω in the equation above is motivated by the fact that the variance of $c_i^{(0)}$ and of the second term in Eq. (3.2.10), is of order $\pi^{\omega 2} \sim 1/\Omega^2$. This implies that the relevant scale for the r.h.s. of Eq. (3.2.11) is of order $1/\Omega$. The parameter ϵ_i encodes the risk premium which banks require for selling derivative i. We first take the simplifying assumption that $\epsilon_i = \epsilon$ does not depend on i, in order to illustrate the generic behavior of the model. In the following, we shall discuss the general case where ϵ_i depends on i. As in the previous chapter, we will focus on the case of competitive equilibria, that can be achieved by means of an adaptive dynamics of banks. As before, the statistical properties corresponding to the competitive equilibrium will be related to those of the minima of a global function (Hamiltonian).

Competitive equilibria: A competitive equilibrium for the market is given by banks choosing the supply s_i so as to maximize their profit (3.2.10), considering returns r^ω as given. When $\epsilon_i = \epsilon$ a competitive equilibrium is then defined as a set of variables $\{s_i^*\}$ such that

$$s_i^* = \text{argmax}_{x \in [0, s_0]} \left[-x \left(\sum_\omega \pi^\omega a_i^\omega r^\omega - \epsilon \right) \right], \tag{3.2.12}$$

where returns are given by $r^\omega = d_0^\omega + \sum_{i=1}^N s_i a_i^\omega$. In the above expression, the term proportional to ϵ accounts for the earning obtained by selling

[1] Equivalently, one could assume a_i^ω to be of order one, but introduce a coefficient $\lambda = 1/\sqrt{\Omega}$ in Eq. (3.2.8) as a finite market depth.

derivatives as well as for the risk premium required by banks. The term proportional to $\sum_\omega \pi^\omega a_i^\omega r^\omega$ refers instead to the fact that banks need to trade on the underlying to match derivative contracts.

Learning dynamics: Let us assume that the context outlined above is repeated for many periods (e.g. days), indexed by $t = 1, \ldots$. Let $\omega(t)$ be the state of the market at time t and assume this is drawn independently from the distribution π^ω in each period. Accordingly, the returns $r^{\omega(t)}$ are still determined by Eq. (3.2.8), with $\omega = \omega(t)$ and $s_i = s_i(t)$, the supply of instruments of type i in period t. In order to determine the latter, banks estimate the profitability of instrument i on historical data. They assign a score (or attraction) $U_i(t)$ to each instrument i, which they update in the following manner:

$$U_i(t+1) = U_i(t) + u_i(t) - \bar{u}_i = U_i(t) - a_i^{\omega(t)} r^{\omega(t)} - \frac{\epsilon}{\Omega}. \tag{3.2.13}$$

Notice that, by Eq. (3.2.10) and (3.2.11), $U_i(t)$ increases (decreases) if $u_i > \bar{u}_i$ ($u_i < \bar{u}_i$). Banks supply instrument i according to the simple rule

$$s_i(t) = \begin{cases} 0 & \text{if } \; U_i(t) \leq 0 \\ U_i(t)/\Omega & \text{if } \; 0 \leq U_i(t) \leq s_0 \\ s_0 & \text{if } \; U_i(t) > s_0 \end{cases} \tag{3.2.14}$$

Therefore, if instrument i provides an expected utility larger than the margin $\bar{u}_i$, its score will increase and the bank will sell it more likely. Conversely, an instrument with $u_i(t) < \bar{u}_i$, on average, has a decreasing score and it will not be offered by banks.

Hamiltonian: The competitive market equilibrium, as well as the stationary regime of the learning dynamics, is given by the solution of the minimization of the function:

$$H = \frac{1}{2} \sum_{\omega=1}^{\Omega} \pi^\omega \left(r^\omega\right)^2 + \frac{\epsilon}{\Omega} \sum_{i=1}^{N} s_i \tag{3.2.15}$$

over the variables $0 \leq s_i \leq s_0$, where r^ω is given in terms of s_i by Eq. (3.2.8). A proof of the statement above follows by direct inspection of the first order conditions and the observation that $\frac{\partial H}{\partial s_i} = -(u_i - \bar{u}_i)$. Imagine $\{s_i^*\}$ to be the minimum of H. If $s_i^* = 0$ it must be that

$$\left.\frac{\partial H}{\partial s_i}\right|_{s_i=0} = \bar{u}_i - u_i > 0. \tag{3.2.16}$$

If $u_i < \bar{u}_i$, it is not convenient for banks to sell derivative i, which is consistent with zero supply ($s_i^* = 0$). Likewise, if $s_i^* = s_0$ the first order condition yields $u_i > \bar{u}_i$, which is the condition under which banks should sell as many derivatives as possible. If instead $0 < s_i^* < s_0$, then $u_i = \bar{u}_i$ which is consistent with perfect competition among banks.
On the other hand:

$$\left.\frac{\partial H}{\partial s_i}\right|_{s_i=0} = -\mathbb{E}_\pi[U_i(t+1) - U_i(t)], \tag{3.2.17}$$

so it is possible to show that, along the trajectories generated by the adaptive dynamics, $\dot{H} < 0$ and the learning scheme converges towards the minima of H.
Notice that, as far as competitive equilibria are considered, if ϵ is negligible a consequence of banks maximizing their utility, is that return's volatility (the first term of Eq.(3.2.15)) is reduced[1]. Actually, with $\epsilon > 0$, only those derivatives which decrease volatility are traded ($s_i > 0$). This is an "unintended consequence" of banks' profit seeking behavior, i.e. a feature which emerges without agents aiming at it.

3.2.3 Market completeness and market stability

From a purely mathematical point of view, we notice that H is a quadratic form which depends on N variables $\{s_i\}$, through the Ω linear combinations given by the returns r^ω. It is intuitively clear that, as N increases, the minimum of H becomes more and more shallow and, for large N, it is likely that there will be directions in the space of $\{s_i\}$ (i.e. linear combinations of the s_i) along which

[1] In the dynamical setting, the situation is actually more complex. Indeed the variables s_i that appear in the Hamiltonian correspond to time averages in the stationary state of the variables $s_i(t)$ updated according to (3.2.14), so that the Hamiltonian accounts only for the part of volatility due to the presence of different states $\omega = 1\ldots,\Omega$.

H is almost flat. The location of the equilibrium is likely to be very sensitive to perturbation along these directions. These qualitative conclusions can be put on firmer grounds by the theoretical approach which we discuss next. A full charac-

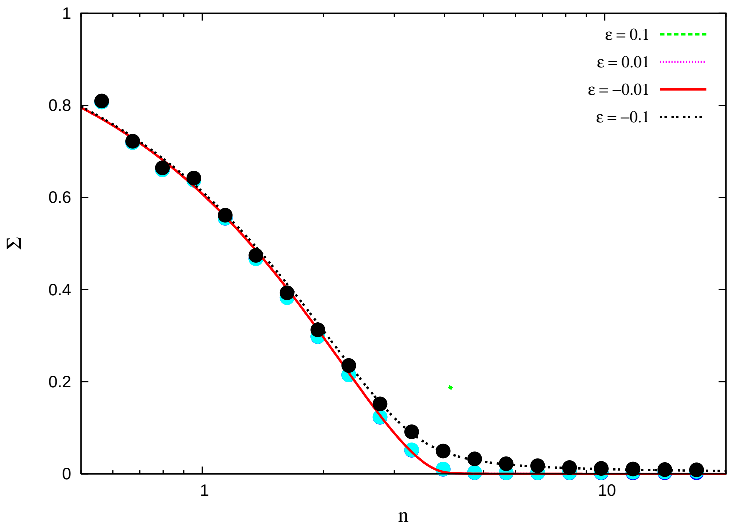

Figure 3.1: *Volatility* $\Sigma = E_\pi\left[(r - E_\pi[r])^2\right]$ *in competitive equilibria (full lines), as a function of* $n = N/\Omega$*, for different values of* ϵ*. Points refer to the variance of* $\bar{r}^\omega$ *computed in numerical simulations of a system with* $\Omega = 64$ *(*$s_0 = 1$*).*

terization of the solution of the minimization of H, in the limit $\Omega, N \to \infty$ with $N/\Omega = n$ fixed, can be derived with tools of statistical mechanics of disordered systems (see appendix B.1), following similar lines of those of chapter 2. Indeed, mathematically the model is quite similar to the Grand Canonical Minority Game studied by Challet and Marsili [2003]. The solution can be summarized in the following "representative" derivative problem: given a normal random variable z with mean zero and unit variance, the supply of the "representative" derivative is given by

$$s_z = \max\left\{s_0, \min\left\{0, \sqrt{g+\Delta}z - \epsilon(1+\chi)\right\}\right\} \tag{3.2.18}$$

where

$$g = nE_z[s_z^2] \tag{3.2.19}$$

$$\chi = \frac{nE_z[s_z z]}{\sqrt{g+\Delta} - nE_z[s_z z]} \tag{3.2.20}$$

are determined self-consistently in terms of expected values $E_z[\ldots]$ over functions of the normal variable z. Loosely speaking, z embodies the interaction through the market of the "representative" derivative with all other derivatives. Any quantity, such as the average supply of derivatives

$$\bar{s} \equiv \lim_{\Omega\to\infty} \frac{1}{N}\sum_{i=1}^{N} s_i^* = E_z[s_z] \tag{3.2.21}$$

can be computed from the solution. Figure 3.1 plots the volatility $\Sigma = E_\pi\left[(r - E_\pi[r])^2\right]$ as a function of n for small values of ϵ for the case $s_0 = 1$. As it can be seen, as the market grows in financial complexity, fluctuations in returns of the risky asset decrease and, beyond a value $n^* \simeq 4.14542\ldots$, they become very small (of order ϵ^2, for small ϵ). The expected return $E_\pi[r] = \bar{d}/(1+\chi)$ also decreases, keeping a bounded Sharpe ratio $E_\pi[r]/\sqrt{\Sigma} = \bar{d}/\sqrt{g+\Delta}$.

The situation in the region $n > n^*$ and $\epsilon \ll 1$ closely resembles the picture of an efficient, arbitrage free complete market. Unfortunately this is also the locus of a sharp discontinuity, as shown in Fig. 3.2. This plots $\bar{s}$ as a function of n for different values of ϵ. As it can be seen, as the market grows in financial complexity, it passes from a regime ($n < n^*$) where its behavior is continuous and smooth with ϵ to one ($n > n^*$) which features a sharp discontinuity at $\epsilon = 0$[1]. In particular, the discontinuity manifests clearly in the limit $s_0 \to \infty$ when the demand of investors is unbounded. Then while for $\epsilon > 0$ the average supply $\bar{s}$ is finite, for $\epsilon < 0$ and $n > 2$ the supply $\bar{s} \to \infty$ diverges as $s_0 \to \infty$. In other words, while in one region ($\epsilon > 0$ and $\epsilon < 0$ for $n < 2$) the equilibrium is controlled by the supply side, in the other region ($\epsilon < 0$, $n > 2$) the equilibrium is dominated by the demand side. This distinction, as we shall see in the next section, applies to a more general model where ϵ_i depends on i.

The consequences of the singularity at $n > n^*$ and $\epsilon = 0$ manifest strongly in response functions: the market behavior close to the singularity is quite sensitive to small perturbations. For example a small change in the risk perception of banks (i.e. in ϵ) can provoke a dramatic change in the volume of trading in the derivative

[1] It should be noted that N is not the actual value of derivatives traded, but the number of derivatives for which there is a demand. The number of derivatives with $s_i > 0$, is well approximated by $N\text{Prob}\{s_z > 0\}$, which is less than Ω, for $\epsilon > 0$.

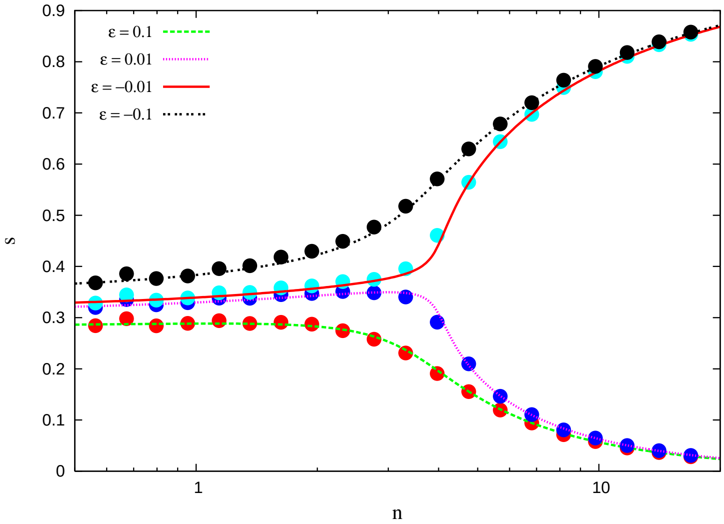

Figure 3.2: *Average supply $\bar{s}$ in competitive equilibria (full lines) for different values of ϵ. Points refer to the variance of $\bar{r}^{\omega}$ computed in numerical simulations of a system with $\Omega = 64$ ($s_0 = 1$).*

market. The effects of this increased susceptibility appears dramatically in the case where the interaction is repeated over time and banks learn and adapt to investors' demand. This setting not only allows us to understand under what conditions will banks learn to converge to the competitive equilibrium, but it also sheds light on the emergent fluctuation phenomena.

At odds with the competitive equilibrium setting, in the present case the volatility of returns also acquires a contribution from the fluctuations of the variables $s_i(t)$, which is induced by the random choice of the state $\omega(t)$. Hence, we can distinguish two sources of fluctuations in returns

$$r(t) = \bar{r}^{\omega(t)} + \delta r(t) \tag{3.2.22}$$

the first depending on the stochastic realization of the state $\omega(t)$, the second from fluctuations in the learning dynamics. Here $\bar{r}^{\omega}$ is the average return of the asset in state ω, and it can be shown that it converges to its competitive equilibrium value, $\forall \omega$. Indeed the contribution of $\bar{r}^{\omega}$ to the volatility $\Sigma = E_{\pi}\left[(\bar{r} - E_{\pi}[\bar{r}])^2\right]$ closely follows the theoretical curve in Fig. (3.1).

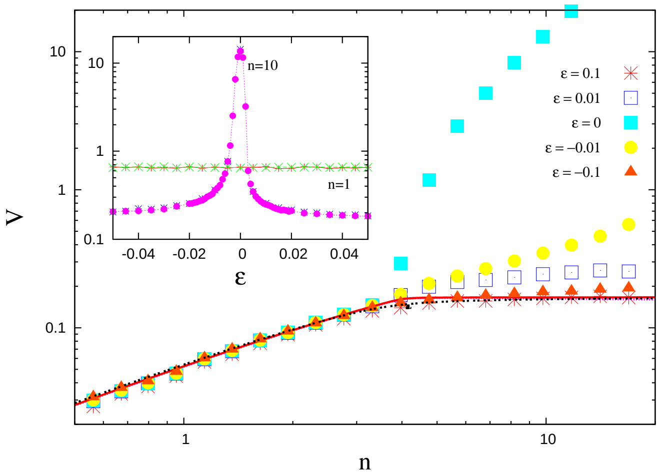

Figure 3.3: *Dynamical contribution to the volatility* $V = E_\pi[\delta r^2]$ *in numerical simulations of a system with* $\Omega = 64$ *for different values of* ϵ *(points). Lines refer to the theoretical prediction in the approximation of independent variables* $s_i(t)$. *Inset: Total volatility* $\Sigma + V$ *in numerical simulations for* $n = 1$ *(*$+$ *for* $\Omega = 128$ *and* $\times$ *for* 256*) and* $n = 10$ *(*$*$ *for* $\Omega = 32$ *and* $\bullet$ *for* $\Omega = 64$*) as a function of* ϵ.

The dynamical contribution to the volatility $V = E_\pi[\delta r^2]$ instead shows a singular behavior which reflects the discontinuity of $\bar{s}$ across the line $\epsilon = 0$ for $n > n^*$. Our theoretical approach also allows us to estimate this contribution to the fluctuations under the assumption that the variables $s_i(t)$ are statistically independent. As Fig. 3.3 shows, this theory reproduces remarkably well the results of numerical simulations for $n < n^*$ but it fails dramatically for $n > n^*$ in the region close to $\epsilon = 0$. The same effect arises in Minority Games Haq et al. [1996]; Marsili and De Martino [2006], where its origin has been traced back to the assumption of independence of the variables $s_i(t)$. This suggests that in this region, the supplies of different derivatives are strongly correlated. This effect has a purely dynamical origin and it is reminiscent of the emergence of persistent correlations arising from trading in the single asset model of Bouchaud and Wyart [2007].

We have considered ϵ as a fixed parameter up to now. Actually, in a more refined model, ϵ_i should depend on i and it should be fixed endogenously in terms of the incentives of investors and banks. At the level of the discussion given thus far, it is sufficient to say that the scale of incentives of banks and investors, and hence of ϵ, is fixed by the average return $\bar{r} = \bar{d}/(1+\chi)$ or by the volatility $\sqrt{\Sigma + V}$. Both these quantities decrease as n increases, so it is reasonable to conclude that ϵ should be a decreasing function of n, in any model where it is fixed endogenously. In other words, as the financial complexity (n) increases, the market should follow a trajectory in the parameter space (n, ϵ), which approaches the critical line $n > n^*$, $\epsilon = 0$. In the next section, we extend our analysis to the case where ϵ_i depends on i.

3.2.4 Asset dependent risk premia

We now generalize the model presented in the previous section to the case of a asset dependent risk premia, in order to see how heterogeneity in the risk perception affects the market in relation to its stability. We consider the case where the $\{\epsilon_i\}$ are taken as quenched asset-dependent random variables drawn from a gaussian distribution with mean $\bar{\epsilon}$ and variance σ_ϵ^2. In analogy with the previous treatment, the competitive equilibrium is equivalent to finding the ground state of the Hamiltonian

$$H = \frac{1}{2}\sum_{\omega} \pi^{\omega}(r^{\omega})^2 + \sum_{i} \frac{\epsilon_i}{\Omega} s_i. \tag{3.2.23}$$

As before, the solution can be summarized in a representative derivative problem. Given a normal random variable z with 0 mean and unit variance, the supply of the representative derivative is given by

$$s_z = \min\left\{ s_0, \max\left\{ 0, \frac{\sqrt{\frac{g+\Delta}{(1+\chi)^2} + \sigma_\epsilon^2}\, z - \bar{\epsilon}}{\nu} \right\}\right\}, \tag{3.2.24}$$

where

$$\nu = \frac{1}{1+\chi} \tag{3.2.25}$$

and

$$g = n\mathrm{E_z}[\mathrm{s_z^2}], \tag{3.2.26}$$

$$\chi = \frac{n\mathrm{E_z}[\mathrm{s_z z}](1+\chi)}{\sqrt{g + \Delta + \sigma_\epsilon^2(1+\chi)^2}} \tag{3.2.27}$$

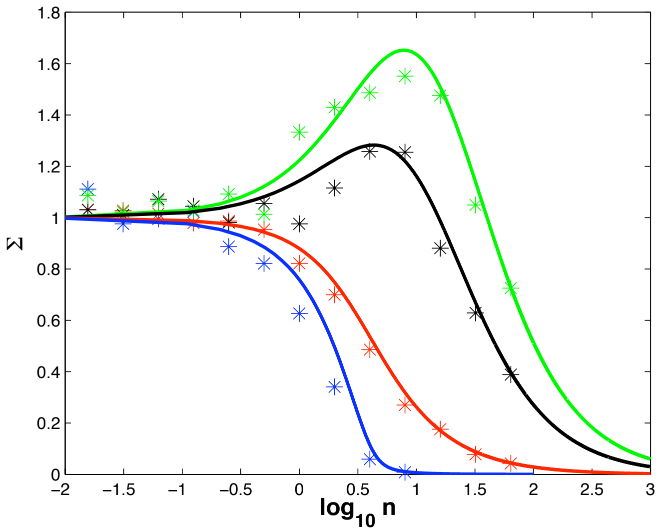

Figure 3.4: *Volatility as a function of* n *for different values of* σ_ϵ^2 *and* $\bar{\epsilon} = 0.1$ *(*$s_0 = 1$*). From top to bottom* $\sigma_\epsilon^2 = 20, 10, 1$ *and* 0.01.

are determined self-consistently.

The main features of this generalized problem are as follows:

1. As for the homogeneous case, the fluctuations of returns eventually become very small as the market complexity n increases (see figure (3.4) where we plot the volatility of returns $\Sigma = \mathbb{E}_\pi[r - \mathbb{E}_\pi[r]^2]$). However, we can observe that the value n^* beyond which the volatility approaches zero depends on the width of the risk premium distribution, and it increases with σ_ϵ^2.

2. The sharp transition previously observed in the behavior of the average supply for large n becomes smooth as soon as the risk premium distribution has a finite width (see Figure (3.5)).

The sharpness of the crossover in the behavior of s as a function of $\bar{\epsilon}$ is enhanced as the demand s_0 for derivatives increases. This signals a transition from a supply limited equilibrium, where the main determinant of the supply s_i of derivatives is banks' profits, to a demand limited equilibrium, where s_i is mostly limited by the finiteness of the investors' demand. In order to make this point explicit, it is

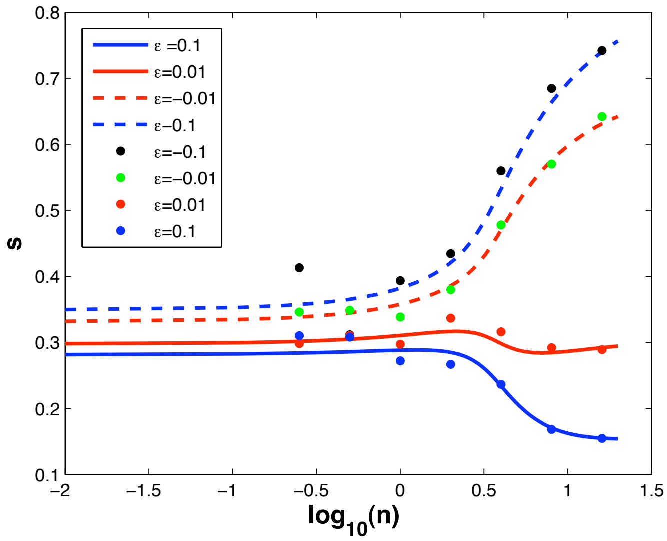

Figure 3.5: *Supply as a function of* n *for different values of* $\bar{\epsilon}$, *for* $\sigma_\epsilon^2 = 0.01$ *and* $s_0 = 1$.

instructive to investigate the case of unbounded supply ($s_0 \to \infty$), because the transition becomes sharp in this limit. The representative derivative, in this case, is described by

$$s_z = \max\left\{0, \frac{\sqrt{\frac{g+\Delta}{(1+\chi)^2} + \sigma_\epsilon^2} z - \bar{\epsilon}}{\nu}\right\}, \tag{3.2.28}$$

and, as before,

$$g = n\mathrm{E_z}[\mathrm{s_z^2}], \tag{3.2.29}$$

$$\chi = \frac{n\mathrm{E_z}[\mathrm{s_z z}](1+\chi)}{\sqrt{g+\Delta+\sigma_\epsilon^2(1+\chi)^2}}. \tag{3.2.30}$$

Similarly to the case $\sigma_\epsilon = 0$, the system displays a phase separation in the $(\bar{\epsilon}, n)$ plane between a region in which the average supply remains finite and a region in which the volume of the traded assets diverges (see left panel of Figure 3.6).

Introducing a gaussian distribution of risk premia with variance σ_{ϵ^2} has then very different effects depending on whether the supply is bounded or not. In

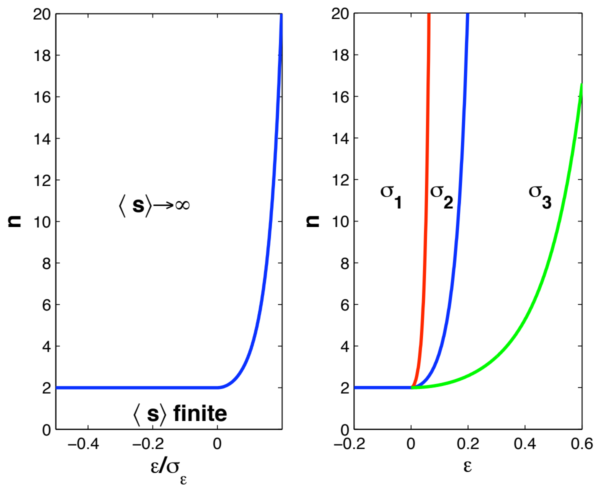

Figure 3.6: *Phase diagram for the case of unbounded supply. Left panel: the plane is divided into two region. Above the blue line the supply diverges, while bleow it remains finite. Right panel: critical line for different values of* σ_ϵ *(*$\sigma_3 > \sigma_2 > \sigma_1$*).*

the first case σ_ϵ acts as a regularizer preventing the occurrence of a sharp phase transition. On the other hand, as $s_0 \to \infty$, $\sigma_\epsilon \neq 0$ entails a deformation of the critical line that tends to flatten along the $n = 2$ line as σ_ϵ grows (see right panel of Figure 3.6). In this case, heterogeneity in the risk perception by banks makes the system more unstable, since for each value of $\bar{\epsilon}$ there exists a critical value of n where a phase transition occurs. This is clearly shown in Figure 3.7, where we plot both the volume of traded derivatives and the volatility of the underlying market as a function of financial complexity. The growth in market complexity leads here to an increase in the volume of traded derivatives. Correspondingly, however, we observe a peak in the volatility of the underlying asset as the volume of trading starts raising. This signals that an instability of the underlying market occurs when the market becomes complete, suggesting that the path towards a complete market may be a path towards an unstable market. Notice, in this respect, that the introduction of a Tobin tax, namely a cost proportional to the volume of traded assets, would have the effect of shifting the average risk premium: $\bar{\epsilon} \to \tilde{\epsilon} = \bar{\epsilon} + \epsilon_t$, where we have indicated with $\epsilon_t > 0$ the cost that banks have to pay for each unit of traded asset. This would in turn result in a

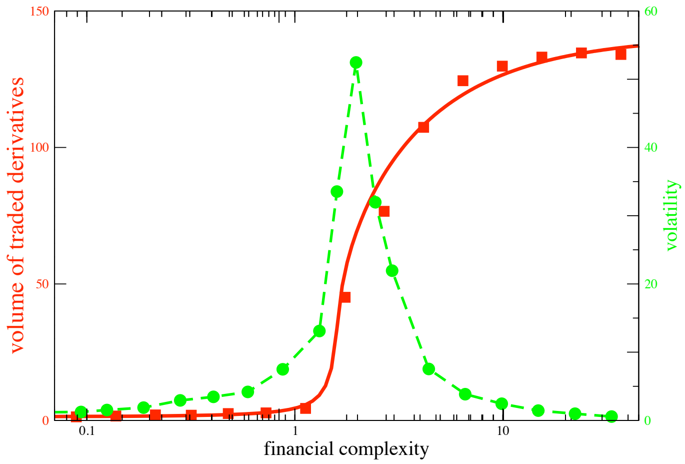

Figure 3.7: *Red curve: average supply of derivatives (left axes) as a function of financial complexity (i.e. n). Green curve: volatility of the underlying (right axis) as a function of financial complexity.*

shift of the critical value of the financial complexity: $n_c \rightarrow \tilde{n_c} > n_c$. the effect of such a tax would then be that of retarding, but not of preventing, the occurrence of the phase transition. For the system to attain complete stability, a super-linear tax would be in this case needed. This would then be quite different with respect to the case of minority games, where it was shown that the introduction of a Tobin tax drives the system far from the critical line [Bianconi et al., 2009] Further investigation is certainly needed concerning policy related issues, since at the level of the present modeling these are just interesting speculations.

3.3 Summary and perspectives

We presented a very stylized description of an arbitrage free market where we accounted for a feedback between derivative and underlying markets. As the number of financial instruments grows, the system was shown to approach the limit of complete, efficient, arbitrage free market described by APT. Interestingly, the uncontrolled proliferation of financial instruments was shown to drive

the market through a transition from a supply limited to a demand limited equilibrium, where large fluctuations and instabilities may arise.
It has to be said that, in order for these kind of models to be more realistic, some improvements certainly are needed. For instance, the demand for derivatives can be derived from a state dependent utility function for consumption at $t = 1$. This could also be a way to obtain an endogenous risk neutral measure. We believe that, while accounting for these effects can make the model more appealing, the collective behavior of the model will not significantly change. Indeed, the qualitative behavior discussed above is typical of a broad class of models [Marsili and De Martino, 2006] and mainly depends on the proliferation of degeneracies in the equilibria of the model.

In summary, the considerations of these two chapters suggest that the ideal view of financial markets on which financial engineering is based may not be compatible with market stability. The proliferation of financial instruments makes the market look more and more similar to an ideal arbitrage-free, efficient and complete market, as well as the increase in the heterogeneity of investors who have access to the market makes the market more informationally efficient. In both cases, however, this occurs at the expense of market stability.
In contrast with the axiomatic equilibrium picture on which financial engineering is based, the models we discussed provide a coherent, though stylized, picture of financial markets as systems made of interacting units. In this picture, concepts such as no-arbitrage, perfect competition, market efficiency or completeness arise as emergent properties of the aggregate behavior, rather than being postulated from the outset. We believe that such an approach can potentially shed light on the causes of and conditions under which liquidity crises, arbitrages and market crashes occur.

Chapter 4

Optimal liquidation strategies regularize portfolio selection

In the first part of this thesis we described financial markets as systems of interacting agents. Notably, accounting for an interaction between agents' behaviors and price processes, we argued that the concept of ideal financial market may be linked to that of market instability. In this chapter, we are going instead to consider a different problem, that of portfolio selection, which in real practice is plagued by instabilities of risk measures and we will show that, in this context, accounting for interactions may be a way to tame fluctuations [Caccioli et al., 2010; Still and Kondor, 2009].

Portfolio selection refers to the problem of finding an optimal investment policy, optimal in the sense of bearing the minimum risk. Different measures for quantifying risk have been introduced so far, all of them being characterized by instabilities [Kondor and Varga-Haszonits, 2008b; Kondor et al., 2007; Varga-Haszonits and Kondor, 2008]. Indeed, when looking for an optimal investment policy, one need to estimate risk starting from historical data concerning the performances of the assets in the market. Such estimation may be strongly biased when time series are too short with respect to the number of assets involved. In this case deviations of the estimated optimal portfolio with respect to the true optimal one may grow unbounded and degeneracies in the space of solutions may arise Ciliberti et al. [2007]; Kondor and Varga-Haszonits [2008a]. Finding a way to reduce fluctuations in this region may be of crucial importance for practitioners,

since optimization of large portfolios in real life operates precisely in the regime where instabilities occurs.
After a brief introduction to the problem and a short review of the main results obtained so far exploiting techniques borrowed from statistical mechanics, we will show that the sought reduction of fluctuations may be achieved accounting for liquidation strategies in portfolio selection. Indeed, the value of a portfolio should always be estimated once the strategy for liquidating the portfolio has been considered [Acerbi and Scandolo, 2007], since prices will move according to it. We will consider in this chapter the specific case of Expected Shortfall in order to illustrate the effect of market impact in portfolio selection. We will show how regularized optimization problems may be derived when market impact is taken into account. Notably, we will discuss in some detail the case of linear as well as instantaneous market impact, showing explicitly how the accessible region of phase space is increased with respect to the standard procedure. We will end with some analysis concerning real data in order to see whether the strength of market impact may be quantified by means of portfolio optimization techniques.

4.1 The Markovitz problem

Let us consider a market where N assets are traded and let us assume we want to invest in the market, namely we want to buy a portfolio of assets that we represent through a vectors of weights $\vec{w} = \{w_1 \dots w_N\}$. The weight w_i represents the relative fraction of the total investment invested on asset i, so that a budget constraint of the kind $\sum_{i=1}^{N} w_i = 1$ must be satisfied. The obvious problem we have to face now is that of choosing how to invest across different assets, i.e. that of selecting the optimal set of weights. Optimal with respect to what? There are two objectives one wish to achieve through an investment policy:

- maximizing the expected return of the portfolio,
- minimizing the risk associated to the investment.

The problem of finding the optimal portfolio was formulated by Markowitz [Markowitz, 1952, 1959] in terms of a tradeoff between reward and risk. The idea is that in presence of portfolios with equal risks one should prefer that of greater expected return, while if expected returns are equal one should choose the portfolio bearing smaller risk. In the following we are going to focus on the part of the problem related to risk, so that the problem we consider is that of choosing the set of weights that minimizes a certain measure of risk. The simplest way to quantify the risk associated to a portfolio is that of measuring its variance, i.e. its volatility [Elton and Gruber, 1995]. If we introduce the matrix $\sigma_{i,j}$ measuring the covariance between assets i and j, the Markowitz optimization problem reduces to finding the minimum of

$$\sigma_P^2 = \sum_{i,j=0}^{N} w_i \sigma_{i,j} w_j, \tag{4.1.1}$$

under the budget constraint $\sum_{i=1}^{N} w_i = 1$[1]. The optimal set of weights is thus given in terms of the inverse of the covariance matrix as

$$w_i^* = \frac{\sum_j \left(\sigma^{-1}\right)_{i,j}}{\sum_{j,k} \left(\sigma^{-1}\right)_{j,k}}. \tag{4.1.2}$$

In practice, the covariance matrix is unknown and has to be estimated from measurements on the market. If we indicate by $x_{i.t}$ the return of asset i at time t, an estimate of the covariance matrix can be given from historical data as

$$\sigma_{i,j} = \frac{1}{T} \sum_t x_{i,t} x_{j,t}, \tag{4.1.3}$$

where T is the number of time records we have access to. For a portfolio of N assets and time series of length T, we need to estimate $\mathcal{O}(N^2)$ elements out of data sets including $\mathcal{O}(NT)$ returns. In order for the estimate to be precise, we would then need $N << T$. However, such a situation rarely holds in practice, so that noisy estimates may arise [Kondor et al., 2007]. It is therefore of crucial importance to understand how the typical features of the optimal solution computed on the basis of historical data change as a function of the ratio N/T.

[1]Notice that w_i can take any real value, so that no ban on short-selling is considered. Although unrealistic, this setup is in fact the most suitable to highlight the problems related to the instability of risk measures.

As a measure of the estimation error, the following quantity was introduced by Pafka and Kondor [2002]

$$q_0 = \frac{\sum_{i,j} w_i^* \sigma_{i,j}^{(0)} w_j^*}{\sum_{i,j} w_i^{*(0)} \sigma_{i,j} w_j^{*(0)}}, \tag{4.1.4}$$

where $\sigma_{i,j}$ denotes the empirical covariance matrix, $\sigma_{i,j}^{(0)}$ the true covariance matrix, w_i^* the optimal weights as computed from the empirical matrix and $w_i^{*(0)}$ the true optimal weights[1]. In the simple case of independent gaussian returns, corresponding to the identity covariance matrix, Pafka and Kondor [2002] were able to compute q_0 in the limit $N, T \to \infty$ with fixed N/T, obtaining

$$q_0 = \frac{1}{1 - \frac{N}{T}}. \tag{4.1.5}$$

The divergence of the estimation error for $N = T$ has a clear meaning. The rank of the empirical covariance matrix is in fact the minimum between N and T, so that such matrix develops zero eigenvalues as soon as $T < N$ and the optimization problem becomes meaningless, fact that is signaled by the divergence of q_0 as $T \to N$. Figure 4.1 shows a comparison between the result of equation (4.1.5) and numerical simulations. The simulations where carried on by generating artificial time series of length T representing assets returns. Returns where drawn at random independently across assets and times from gaussian distributions of zero mean and standard deviation $1/\sqrt{N}$. The empirical covariance matrix was then computed from the time series, as well as the optimal weights were computed according to (4.1.2) from the empirical covariance matrix. The estimation error has then be computed according to (4.1.4).

Another interesting quantity to look at is the behavior of the intensity of fluctuations in the optimal weights. Notably, Kondor et al. [2007] showed that such fluctuations become wilder as T approaches N. Figure 4.2 shows the optimal weights computed from covariance matrices estimated using time series with different length together with a comparison with the true optimal weights. It is clear from the picture that fluctuations increase as T becomes smaller. The

[1]Notice that the above quantity is denoted as q_0^2 in [Pafka and Kondor, 2002]. We choose a slightly different definition for later convenience.

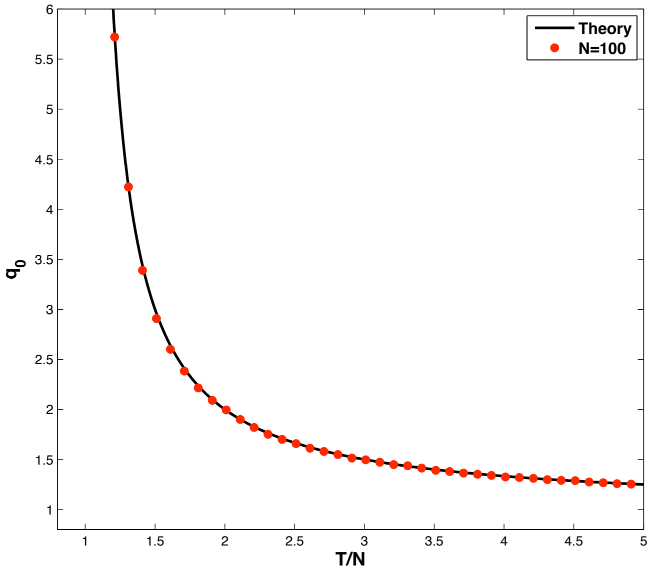

Figure 4.1: *Behavior of the estimation error as a function of T/N. The solid line represents the result of equation (4.1.5). Red dots refer to simulations with $N = 100$.*

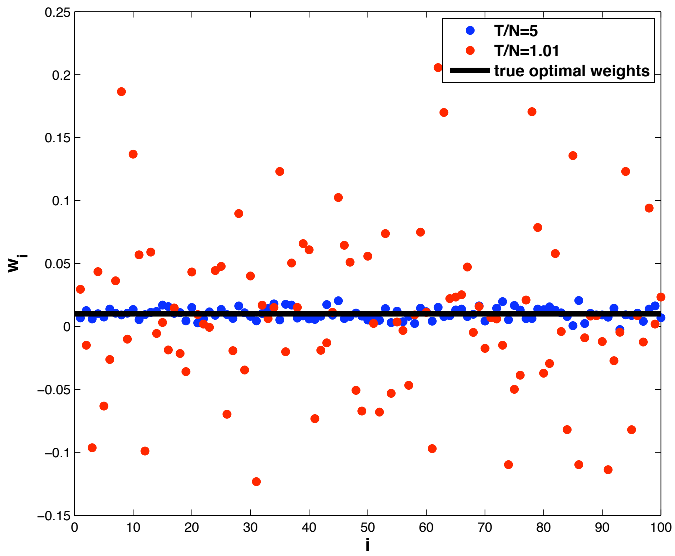

Figure 4.2: *Optimal weights computed from empirical covariance matrices for* $N = 100$*,* $T = 101$ *(red dots) and* $T = 500$ *(blue dots). The solid line refers to the optimal weights as computed from the true covariance matrix.*

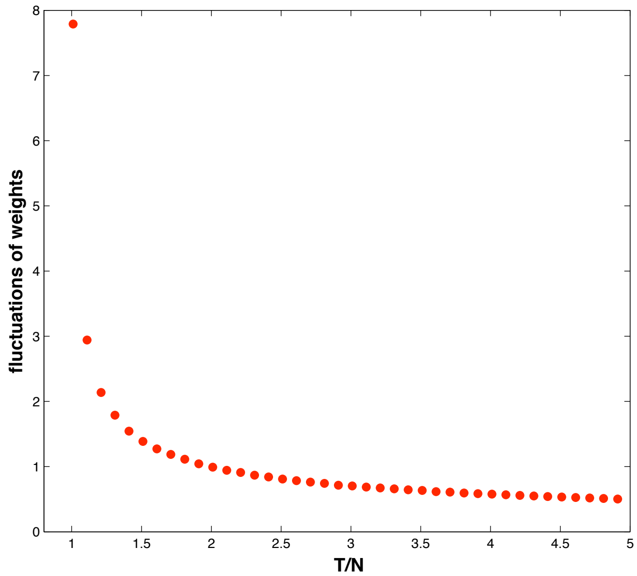

Figure 4.3: *Behavior of the fluctuations of the optimal weights as a function of* T/N.

behavior of fluctuations as a function of T/N is explicitly shown in Figure 4.3. The divergence of q_0, as well as the increase of fluctuations as T approaches N, suggests the presence of a phase transition at $T = N$. We will further elaborate on this in the next section, where we will introduce a risk measure that is more important for applications than the variance we have considered so far. In this context, we will discuss the emergence of a phase transition in the statistical mechanics sense.

4.2 Instability of risk measures

In the previous section we have introduced the problem of portfolio optimization using as a risk measure the variance of portfolio returns. This is certainly a good measure to characterize portfolio fluctuations in the case of normally distributed returns. Real portfolio, however, often display long tailed distributions that call for a different quantification of risk [Bouchaud and Potters, 2000]. In fact alternative risk measures abound in the literature as well as in the practice [Kondor et al., 2007]. We will focus in the following on the particular case of the Expected Shortfall.

4.2.1 Expected Shortfall

Given a portfolio of weights $\vec{w} = \{w_1, \ldots, w_N\}$ and N assets with return $\vec{x} = \{x_1, \ldots, x_N\}$ drawn from the probability distribution $p(\vec{x})$, we define the loss $l(\vec{w}|\vec{x}) = -\sum_i w_i x_i$. The probability for such loss to be smaller than a threshold α is

$$P_<(\vec{w}, \alpha) = \int \prod_i dx_i p(\vec{x}) \theta(\alpha - l(\vec{w}, \vec{x})), \tag{4.2.6}$$

with $\theta(x) = 1$ if $x > 0$ and $\theta(x) = 0$ otherwise. The associated βVaR, which represents the possible minimal loss assuming a confidence level β, is defined as

$$\beta\mathrm{VaR}(\vec{w}) = \min\{\alpha : P_<(\vec{w}, \alpha) \geq \beta\}, \tag{4.2.7}$$

while the Expected Shortfall $\mathrm{ES}(\vec{w})$ is given by

$$\mathrm{ES}(\vec{w}) = \frac{1}{1-\beta} \int \prod_i dx_i p(\vec{x}) l(\vec{w}|\vec{x}) \theta\big(l(\vec{w}|\vec{x}) - \beta\mathrm{VaR}(\vec{w})\big). \tag{4.2.8}$$

Expected Shortfall (ES) is then the mean loss above a high quantile.[1] It gives a more faithful representation of large losses than Value at Risk (VaR) [Jorion, 2000] that can be identified by the quantile itself. In addition, ES can be computed by fast linear programming algorithms [Rockafellar and Uryasev, 2000] and, most significantly, it was shown [Acerbi, 2002, 2004; Acerbi and Tasche, 2002] to belong to the set of coherent risk measures [Artzner et al., 1999]. The calculation of the ES can be obtained through the minimization of the function [Rockafellar and Uryasev, 2000]

$$F_\beta(\vec{w}, \epsilon) = \epsilon + \frac{1}{1-\beta} \int \prod_i dx_i p(\vec{x}) [l(\vec{w}|\vec{x}) - \epsilon]^+ \tag{4.2.9}$$

as

$$\mathrm{ES}(\vec{w}) = \min_\epsilon F_\beta(\vec{w}, \epsilon), \tag{4.2.10}$$

[1] Note the sign convention: in the context of risk measures losses are counted positive.

with $[x]^+ = (x + |x|)/2$. Approximating the integral in (4.2.9) by sampling the probability distributions of returns

$$\int \prod_i dx_i p(\vec{x}) f(\vec{x}) \to \frac{1}{T} \sum_\tau f(\vec{x}_\tau), \tag{4.2.11}$$

the problem can be reduced to the calculation of the minimum of the cost function

$$E[v, \{u_\tau\}] = (1 - \beta)T\epsilon + \sum_{\tau=1}^{T} u_\tau \tag{4.2.12}$$

under the constraints

$$u_\tau \geq 0 \quad \forall \tau,$$
$$u_\tau + \epsilon + \sum_{i=1}^{N} x_{i,\tau} w_i \geq 0 \quad \forall \tau$$

and

$$\sum_i w_i = N,$$ [1]

where we have defined $u_\tau = [-\epsilon - \sum_i x_{i,\tau} w_i]^+$.
Given that the cost function is monotonous in u_τ, the first two constraints enforce the fact that $u_\tau = -\epsilon - \sum_i x_{i,\tau} w_i$ if $-\epsilon - \sum_i x_{i,\tau} w_i > 0$ and $u_\tau = 0$ otherwise.

4.2.2 Instability of Expected Shortfall

We now focus on a limit case of ES and discuss a very simple example useful to easily understand the reason for instabilities that may arise when looking for the optimal portfolio [Kondor et al., 2007]. Let us consider the special case of ES corresponding to $\beta = 1$. The problem reduces in this case to that of finding the weights that minimize the maximal loss (ML)

$$ML(\vec{w}) = \max_{t=1,\ldots,T} \left[-\sum_i w_i x_{i,t} \right], \tag{4.2.13}$$

so that we need to find

$$\vec{w}^* = \mathrm{argmin}_{\vec{w}} \left[ML(\vec{w}) \right]. \tag{4.2.14}$$

[1] From now on, we choose to normalize the sum to N for later convenience, while usually the normalization is set to unity.

Let us consider the very simple case of two assets and two times: $N = T = 2$. Since the weights sum to 1 we take $w_1 = w$ and $w_2 = 1-w$. The losses associated to the two periods $t = 1, 2$ are then

$$l_1 = -wx_{11} - (1-w)x_{21} = w(x_{21} - x_{11}) - x_{21} \tag{4.2.15}$$

$$l_2 = -wx_{12} - (1-w)x_{22} = w(x_{22} - x_{1,2}) - x_{22}. \tag{4.2.16}$$

It is clear at this stage that, depending on the slopes of these two straight lines, the maximal loss $\max\{l_1, l_2\}$ may or may not be bounded from below. Notably, if the two slopes have opposite sign, the maximal loss is bounded from below and the optimization problem has a solution. In contrast, if the slopes have the same sign, the maximal loss is not bounded from below and the minimum runs away to infinity, as well as the value of the objective function (4.2.13) goes to minus infinity. What is the meaning of two slopes with the same sign?
This situation can happen when

- $x_{21} > x_{11}$ and $x_{22} > x_{12}$,
- or $x_{21} < x_{11}$ and $x_{22} < x_{12}$.

In both cases we have an asset that dominates the other for all $t = 1, 2$. Obviously, if the return of an asset is always greater than the return of an other asset, going infinitely short[1] in the dominated asset and infinitely long[2] in the dominating one will produce profit without bearing any risk.

More generally, being a conditional average, ES is not bounded from below: if a portfolio produces a large gain, rather than a loss, then ES takes a large negative value. Now, on a finite sample it may happen that one of the items, or a combination of items, *dominates* the others, i.e. produces a larger return at each time point than the rest. When such an apparent arbitrage occurs,

[1]short selling is the practice of selling assets that have been borrowed from a third party with the intention of buying identical assets back at a later date to return to the lender. The short seller hopes to profit from a decline in the price of the assets between the sale and the repurchase, as the seller will pay less to buy the assets than the seller received on selling them.

[2]A long position in a security means the holder of the position owns the stock and will profit if the price of the security will increase.

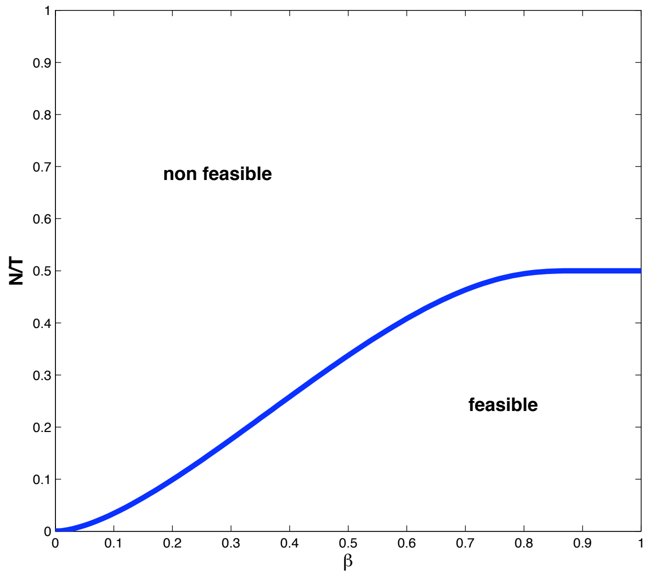

Figure 4.4: *Phase diagram for the optimization problem under Expected Shortfall.*

the optimization of ES suggests to go as long as possible in the dominating asset and correspondingly short in the dominated ones. In particular, if there are no other constraints except for the fixed budget, this leads to a runaway solution and to a seemingly infinite return. Therefore, by considering an ensemble where returns are generated from a given distribution, for finite N and T the optimization of ES will have a finite solution with a probability always less than one. This probability quickly approaches one as N/T goes to zero, and quickly approaches zero as N/T goes to infinity. Moreover, it was shown [Ciliberti et al., 2007; Kondor et al., 2007] that the transition between the two limits becomes sharper and sharper as N and T go to infinity such that their ratio is fixed, which is the realistic limit to consider for large institutional portfolios. In this limit there will be a critical value of the ratio N/T where a sharp transition occurs between the region where the optimization of ES leads to a finite solution and the one where it does not. This instability of ES was pointed out, for the case of independent gaussian distributed returns, by Kondor et al. [2007], where the phase diagram was determined numerically, while Ciliberti et al. [2007] derived an analytic expression for the critical line exploiting tools borrowed from the

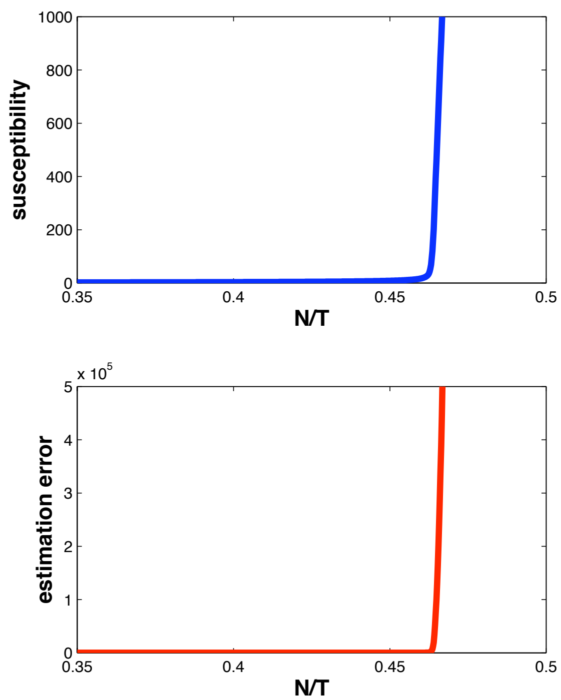

Figure 4.5: *Divergence of susceptibility and estimation error for the ES problem in the case* $\beta = 0.7$.

statistical physics of random systems. Figure 4.4 reproduces the phase diagram computed by Ciliberti et al. [2007]. The region labeled as feasible is the region where a finite solution may be identified, while in the unfeasible region weights run to infinity and large fluctuations set in. This is highlighted by the divergence of the susceptibility as well as of the estimation error in the whole unfeasible region (see Figure 4.5). We do not reproduce here the calculation carried on in Ciliberti et al. [2007], since we are going to provide the solution for a more general problem in the following sections.

4.2.3 Coherent risk measures

The generality of the argument presented above suggests that the presence of accidental statistical arbitrages causes instabilities in portfolio selection not only under ES, but also for all the risk measures that are unbounded from below. It was indeed demonstrated [Kondor and Varga-Haszonits, 2008b] that the coherence axioms [Artzner et al., 1999] imply the appearance of a similar instability. Coherent risk measures have been introduced [Artzner et al., 1999] as an attempt to characterize risk in axiomatic way, giving solid theoretical foundations to the way of quantifying risk. Let us imagine we have a risk measure $\rho(\vec{w}, \mathbf{X})$, where $\vec{w}$ represent a portfolio and $\mathbf{X}$ is an $N \times T$ matrix whose entry $x_{i,t}$ reprents the return of asset i at time t. In our case, $\rho(\vec{w}, \mathbf{X})$ represents a measure of the loss probability associated to the portfolio $\vec{w}$ when the risk is estimated on the basis of the historical data encoded in the matrix $\mathbf{X}$. The risk measure ρ is said to be coherent with respect to the sample $\mathbf{X}$ if it satisfies the following requirements:

- *monotonicity* $\vec{u}$ is such that $\sum_i u_i x_{i,t} > 0 \ \forall t \Rightarrow \rho(\vec{u}|\mathbf{X}) \leq 0$ (if a portfolio has only positive returns, its risk should be negative);
- *sub-additivity* $\rho(\vec{u} + \vec{v}|X) \leq \rho(\vec{u}|\mathbf{X}) + \rho(\vec{v}|\mathbf{X})$ (the risk of two portfolios together cannot get any worse that adding the two risks separately);
- *positive homogeneity* $a > 0 \Rightarrow \rho(a\vec{u}|\mathbf{X}) = a\rho(\vec{u}|\mathbf{X})$ (if you increase the size of your portfolio by a factor a, you also increase the risk by the same factor);
- *translational invariance* $\rho(\vec{u}|\mathbf{X}+a) = \rho(\vec{u}|\mathbf{X}) - a$ (adding cash to a portfolio reduces the risk of the investment),

where $\vec{u}$ and $\vec{v}$ represent two portfolios and $\mathbf{X}+\mathbf{a}$ denotes a matrix obtained adding the real number a to each element of the matrix $\mathbf{X}$. The problem to solve for the selection of the optimal portfolio reads

$$\vec{w}^* = \mathrm{argmin}_{\vec{w}}\left[\rho(\vec{w}|\mathbf{X})\right], \tag{4.2.17}$$

$$\sum_i w_i = N. \tag{4.2.18}$$

Let us introduce the definition of dominant portfolio. Let us consider two portfolios $\vec{u}$ and $\vec{v}$. We say that $\vec{u}$ dominates $\vec{v}$ over the sample $\mathbf{X}$ if $\sum_i u_i x_{i,t} \geq \sum_i v_i x_{i,t}$ $\forall t = \{1,\dots T\}$. In presence of strict inequalities we say that $\vec{u}$ strictly dominates $\vec{v}$. A general result about the feasibility of the optimization problem (4.2.17) was proven by Kondor and Varga-Haszonits [2008b]:

Theorem: *if there are two portfolios $\vec{u}$ ans $\vec{v}$ such that $\vec{u}$ strictly dominates $\vec{v}$ the optimization problem $\min_{\vec{w}}\rho(\vec{w}|\mathbf{X})$, $\sum_i w_i = N$ has no solution.*

This results generalizes the considerations of the previous section concerning the role of apparent arbitrages arising when we consider finite time series. The presence of such historical arbitrages constitutes a sufficient condition for the optimization problem to be unfeasible. [1]

While keeping in mind the generality of these results, in the following we will keep on considering the specific case of Expected Shortfall.

4.3 Regularized portfolio selection

Problems of portfolio selection in real life usually lie in the range $N \sim T$, where a transition from a feasible to an unfeasible region may occur and large fluctuations might set in. As we said, the origin of such fluctuations may be traced back to over-fitting of noisy samples. It is therefore of primary interest the development of techniques allowing for a reduction of noise fitting in the estimation of risk. In this respect, within the context of ES, it was recently proposed to consider a regularized problem where a further constraint to the optimization problem is added, so to reduce the available space for optimal solutions. Notably, Still and

[1] This condition was proven to be also necessary in the case of the Maximal Loss. Kondor and Varga-Haszonits [2008b].

Kondor [2009] proposed to solve in place of (4.2.12) the following optimization problem

$$\min_{\vec{w},\vec{u},\epsilon}\left[(1-\beta)T\epsilon+\sum_{\tau=1}^{T}u_\tau+\frac{1}{2C}\sum_i w_i^2\right] \tag{4.3.19}$$

under the constraints

$$u_\tau \geq 0 \quad \forall\tau,$$

$$u_\tau+\epsilon+\sum_{i=1}^{N}x_{i,\tau}w_i \geq 0 \quad \forall\tau,$$

and

$$\sum_i w_i = N,$$

where C is a positive constant that was shown to be related to the capacity of a certain class of learning machines [Still and Kondor, 2009]. The addition of a term proportional to the L_2 norm of the vector of weights, was justified in view of a mapping of the optimization problem onto a modified support vector regression problem [Vapnik, 1995]. Moreover, from the financial point of view the introduction of such a term can be justified in terms of an enhanced portfolio diversification.

An in-depth analysis of the relation between the portfolio optimization problem and learning machine theory is presented by Still and Kondor [2009]. In the following we show that the regularized problem may be derived by accounting for the impact of liquidation strategies when searching for the optimal portfolio. After providing a derivation of the optimization problem (4.3.19) from this point of view, we will characterize the typical properties of the solutions of such problem and we will prove that the introduction of a regularizer is indeed useful to tame fluctuations and enhance the fitting performances on finite data samples.

4.3.1 Regularization from market illiquidity

To generate cash, an investor has to liquidate (part of) his portfolio. The set up of the portfolio optimization problem discussed so far ignores the fact that this liquidation may have an impact on asset prices. Consider the case that an investor has a portfolio of N assets that we represent through a vector $\vec{q}_t$ =

$(q_{1,t}, \ldots, q_{N,t})$, where $q_{i,t}$ is the quantity of the i-th stock that the investor has on day $t = 0, 1, 2, \ldots$. Let us assume that the liquidation of a part of the portfolio with weights $\vec{w}_t$ on day t affects prices of securities in a linear way

$$\vec{p}_{t+1} = \vec{p}_t + \vec{x}_t - \eta\vec{w}_t. \tag{4.3.20}$$

Here $\vec{x}_t$ is the vector of returns, and η is an impact parameter. Notice that investment is taken to move prices in the direction opposite to trading: selling $(w_{i,t} > 0)$ will cause prices to fall and buying $(w_{i,t} < 0)$ will push prices up. The cash flow generated on day t is then given by

$$c_t = \vec{w}_t \cdot \vec{p}_{t+1} = \vec{w}_t \cdot \vec{p}_t + \vec{w}_t \cdot (\vec{x}_t - \eta\vec{w}_t) \tag{4.3.21}$$

The first part $\vec{w}_t \cdot \vec{p}_t$ is known at time t, so risk only enters in the second part, $\vec{w} \cdot \vec{x} - \eta||\vec{w}||^2$, where we have dropped the subscript t to simplify the notation. Similarly to what is done in classical portfolio theory [Markowitz, 1952], we consider the problem of finding the portfolio of minimal risk, for a given present value $\sum_i w_i p_{i,t} = WN$ of the realized cash flow. The parameter W plays the role of a normalization, and is customarily set to one, because risk is usually linear in the size of the portfolio. Here, however, the size of the portfolio matters as the impact of liquidation strategies on prices depends on the size. We therefore keep W as an independent parameter. In order to further simplify the notation, we consider $p_i = 1$, $\forall i$, so that we have the constraint

$$\sum_i w_i = WN. \tag{4.3.22}$$

As before, we take the expected shortfall as a risk measure. The loss is now given by $l(\vec{w}|\vec{x}) = -\vec{w} \cdot \vec{x} + \eta||\vec{w}||^2$, and we then have to find the minimum of the cost function

$$E_\eta[v, \{u_\tau\}] = (1-\beta)Tv + \sum_{\tau=1}^{T} u_\tau \tag{4.3.23}$$

under the constraints

$$\begin{aligned} u_\tau &\geq 0 \ \forall\tau, & (4.3.24)\\ u_\tau + v + \sum_{i=1}^{N} w_i x_{i,\tau} - \eta||\vec{w}||^2 &\geq 0 \ \forall\tau, & (4.3.25)\\ \sum_i w_i &= WN. & (4.3.26) \end{aligned}$$

All of the T inequality constraints contain a term that is independent of τ ($\tau = 1, \dots, T$), given by

$$\epsilon = v - \eta \|\vec{w}\|^2. \tag{4.3.27}$$

Substitution of $\epsilon + \eta\|\vec{w}\|^2$ for v in the cost function, Eq. (4.3.23), and multiplication by $\frac{1}{2(1-\beta)T\eta}$ leads us to the regularized expected shortfall problem (4.3.19):

$$\min_{\vec{w},\vec{u},\epsilon} \left[\frac{1}{2}\|\vec{w}\|^2 + C \left(\frac{1}{T} \sum_{\tau=1}^{T} u_\tau + (1-\beta)\,\epsilon \right) \right] \tag{4.3.28}$$

$$\text{s.t.} \quad \vec{w} \cdot \vec{x}_\tau + \epsilon + u_\tau \geq 0; \quad u_\tau \geq 0; \quad \forall \tau, \tag{4.3.29}$$

$$\sum_i w_i = WN. \tag{4.3.30}$$

with

$$C = \frac{1}{2(1-\beta)\eta}. \tag{4.3.31}$$

We recognize that the term proportional to η in Eq. (4.3.21) acts as a regularizer.

4.3.2 Stability of regularized Expected Shortfall

In order to develop some intuition about the role of the L_2 regularizer introduced in the optimization problem under ES, we consider as before the limit case of Maximal Loss in the simple situation where the optimization involves two assets $i = 1, 2$ and two periods $t = 1, 2$. The maximal loss is defined in this case as[1]

$$ML(\vec{w}) = \max_{t=1,\dots,T} \left[-\sum_i w_i x_{i,t} + \frac{\eta}{2} \sum_i w_i^2 \right], \tag{4.3.32}$$

and we need to find

$$\vec{w}^* = \text{argmin}_{\vec{w}} \left[ML(\vec{w}) \right]. \tag{4.3.33}$$

[1] The $\beta \to 1$ limit is a bit tricky to recover from the ES optimization problem. A section in the appendix is devoted to the derivation of the correct limit.

We thus compute the losses associated to times $t = 1, 2$:

$$l_1 = -wx_{11} - (1-w)x_{21} + \frac{\eta}{2}w^2 + \frac{\eta}{2}(1-w)^2 \quad (4.3.34)$$

$$= \eta w^2 + (x_{21} - x_{11} - \eta)w - x_{21} - \frac{\eta}{2} \quad (4.3.35)$$

$$l_2 = -wx_{12} - (1-w)x_{22} + \frac{\eta}{2}w^2 + \frac{\eta}{2}(1-w)^2 \quad (4.3.36)$$

$$= \eta w^2 + (x_{22} - x_{12} - \eta)w - x_{22} - \frac{\eta}{2}, \quad (4.3.37)$$

where we have implemented the budget constraint taking $w_1 = w$ and $w_2 = 1-w$. It is clear at this stage that, given the fact that the losses are convex quadratic functions, there exists always a finite and unique minimum for the maximal loss as soon as $\eta > 0$.

More in general, imagine that there are two portfolios $\vec{w}^+$ and $\vec{w}^-$, each properly normalized (i.e. $\sum_i w_i^\pm = WN$), with $\vec{w}^+\vec{x}_\tau \geq \vec{w}^-\vec{x}_\tau$ for all $\tau = 1, \ldots, t$ and $\vec{w}^+\vec{x}_\tau > \vec{w}^-\vec{x}_\tau$ for at least one τ. Then, when $\eta = 0$, minimal Expected Shortfall would be realized by selling K units of $\vec{w}^-$ and buying $K + 1$ units of $\vec{w}^+$, with $K \to \infty$. This, as we said before, is the origin of the instability in coherent risk measures. Such infinite returns cannot be realized, however, by liquidating a real portfolio because prices will adjust. In the linear approximation discussed here, when $\eta > 0$, the investment behavior discussed above is going to modify future returns, because $x_{i,t+1} \to x_{i,t+1} - \eta w_i$, thereby eliminating the apparent arbitrage. This effect eliminates the apparent arbitrage and reflects precisely the logic behind the no-arbitrage hypothesis.

4.3.3 Behavior of large random minimal risk portfolios under L_2 regularized Expected Shortfall

We have argued that the instability of risk measures can be alleviated by regularization discussing a very simple example. We present now a more general discussion generalizing the calculation by Ciliberti et al. [2007] to see how the regularizer takes care of the instability. The calculation is reported in some details in appendix C.1, here we just quote the result and discuss its consequences, namely the removal of the singularity of the risk measure. We also refer the interested reader to the related literature within statistical learning theory, such as [Dietrich et al., 1999; Malzahn and Opper, 2005; Opper, 1995; Opper and Haussler, 1995] and references therein.

Given a realization for the history of returns $\{x_{i,\tau}\}$ drawn from a Gaussian distribution[1], the calculation proceeds by considering the partition function

$$Z_\gamma(\{x_{i,\tau}\}) = \int_{V(\{x_{i,\tau}\})} d\vec{Y} e^{-\gamma E[\vec{Y}]}, \tag{4.3.38}$$

where γ is the inverse temperature, we have used the notation $\vec{Y}$ to indicate the set of variables, and $V(\{x_{i,\tau}\})$ represents the portion of phase space where all constraints are satisfied. The minimum cost can then be computed as

$$\lim_{N\to\infty} \lim_{\gamma\to\infty} -\frac{\log Z_\gamma(\{x_{i,\tau}\})}{N\gamma}. \tag{4.3.39}$$

In order to compute typical properties of the ensemble, we average over the probability distribution of returns, that is we compute the average of $\log Z_\gamma(\{x_{i,\tau}\})$. This can be achieved through the replica trick exploiting the identity

$$\langle \log Z \rangle = \lim_{n\to 0} \left\langle \frac{\partial Z^n}{\partial n} \right\rangle, \tag{4.3.40}$$

where we have denoted by $\langle \cdots \rangle$ averages over $\vec{Y}$. We showed in the previous sections that both statistical and financial considerations lead us to a cost function of the form

$$E[v, \{u_\tau\}] = (1-\beta)T\epsilon + \sum_{\tau=1}^{T} u_\tau + \tilde{\eta}\|\vec{w}\|^2,$$

where $\tilde{\eta}$ can be expressed in terms of C or η. Starting from this cost function, after some manipulations, it is possible to express the energy in terms of three order parameters as in Ciliberti et al. [2007]

$$E(\tilde{\epsilon}, \tilde{q}_0, \Delta) = \Delta \left[t(1-\beta)\tilde{\epsilon} - \frac{\tilde{q}_0}{2} + \frac{t}{2\sqrt{\pi}} \int_{-\infty}^{\infty} ds e^{-s^2} g(\tilde{\epsilon} + s\sqrt{2\tilde{q}_0}) \right] \tag{4.3.41}$$
$$+ \frac{W^2}{2\Delta} + \tilde{\eta}\tilde{q}_0\Delta^2, \tag{4.3.42}$$

[1] Here $x_{i,t}$ are taken as i.i.d. Gaussian variables with zero mean and variance $1/\sqrt{N}$. The latter ensures a meaningful limit $N, T \to \infty$ with $N/T = n$ constant, and is also realistic for typical cases where $N \sim 10^3 - 10^4$.

where Δ is the susceptibility, $q_0 = \tilde{q}_0\Delta^2 = \sum w_i^2/N$, $\epsilon = \tilde{\epsilon}\Delta$, $t = T/N$ and

$$g(x) = \begin{cases} 0, & x \geq 0 \\ x^2, & -1 \leq x \leq 0 \\ -2x-1, & x < -1 \end{cases} . \tag{4.3.43}$$

The difference with respect to Ciliberti et al. [2007] is that now we have an additional term proportional to q_0. Indeed the extra term $\tilde{\eta}\tilde{q}_0\Delta^2$ precisely maps into the term $\tilde{\eta}\|\vec{w}\|^2$ added to the objective function, if one considers the definition of $\tilde{q}_0$.

Let us now discuss how the term proportional to $\|\vec{w}\|^2$ in the cost function takes care of the instability in the portfolio optimization problem. The saddle point equations corresponding to (4.3.42) read

$$-1 + \frac{t}{\sqrt{2\pi\tilde{q}_0}} \int dse^{-s^2} sg'(\tilde{\epsilon} + s\sqrt{2\tilde{q}_0}) + 2\tilde{\eta}\Delta = 0, \tag{4.3.44}$$

$$1 - \beta + \frac{1}{2\sqrt{\pi}} \int dse^{-s^2} g'(\tilde{\epsilon} + s\sqrt{2\tilde{q}_0}) = 0, \tag{4.3.45}$$

$$-\frac{w^2}{2\Delta^2} + t(1-\beta)\tilde{\epsilon} - \frac{\tilde{q}_0}{2} + \frac{t}{2\sqrt{\pi}} \int dse^{-s^2} g(\tilde{\epsilon} + s\sqrt{2\tilde{q}_0}) + 2\tilde{\eta}\Delta\tilde{q}_0 = 0. \tag{4.3.46}$$

Notice that the variables $\tilde{\epsilon}$ and $\tilde{q}_0$ are finite [1], since they have already been rescaled with respect to the original variables as $\tilde{\epsilon} = \epsilon/\Delta$ and $\tilde{q}_0 = q_0/\Delta^2$. This should be sufficient to conclude that all integrals over the variable s are finite. In order to see if a solution with divergent susceptibility exists, let us now impose the condition $\Delta \to \infty$ on the saddle point equations. We first note that, in order for (4.3.46) to be satisfied, $\tilde{q}_0\Delta$ has to be finite as $\Delta \to \infty$, i.e. $\tilde{q}_0 = \alpha/\Delta$ with α finite. This is in contrast with a similar constraint we can deduce from equation (4.3.44), where we find that a solution exists only if $\Delta\sqrt{\tilde{q}_0}$ is finite. Indeed if we multiply all terms of (4.3.44) for $\sqrt{\tilde{q}_0}$ and impose $\tilde{q}_0 = \alpha/\Delta$ we see that all the terms are bounded except the last one which diverges as $\sqrt{\Delta}$. We thus conclude that no solution with divergent susceptibility can be found as long as $\tilde{\eta} > 0$.

[1] The divergence of $\tilde{\epsilon}$ and $\tilde{q}_0$ is prevented by equation (4.3.45), that admits solutions as long as the two variables are kept finite.

The numerical solution of the saddle point equations confirms this prediction. In figures 4.6 and 4.7 we show the behavior of $q_0 = \frac{1}{N}\sum_i w_i^2$ and of the susceptibility. We can clearly observe that the divergence, which is present for $\tilde{\eta} = 0$ disappears as soon as $\tilde{\eta} > 0$. This is further confirmed by figure 4.8, where we show that, in the unfeasible region of the original problem, the susceptibility diverges at $\tilde{\eta} = 0$. Finally, the case of independent gaussian variables is a quite useful benchmark since it allows for a direct comparison with the (known) correct solution, namely that with $w_i = 1\ \forall i$. Figure 4.9 shows the optimal weights computed with and without regularization for $t = 2.25$ compared with the true solution. The reduction in weights fluctuations due to the introduction of the regularizer is quite clear in the picture.

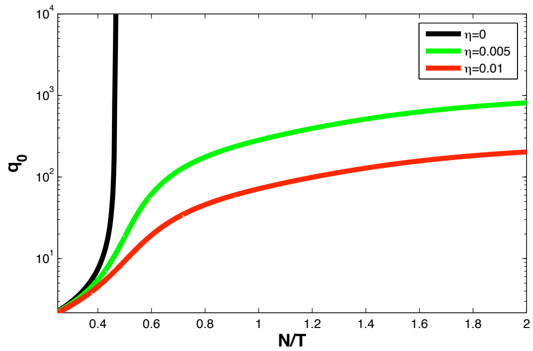

Figure 4.6: *q_0 as a function of N/T for different values of $\tilde{\eta}$ and $\beta = 0.7$.*

Let us now comment on the generality of the result. Concerning the linear assumption in Eq. (4.3.20), we observe that the estimate of market impact functions is a matter of active current research Eisler et al. [2009]. In double auction markets, if one restricts attention to the instantaneous impact of market orders, the effect on the price depends on the shape of the order book. In order to discuss this case in some more detail, let $\rho_i(p,t)$ be the density of limit orders for asset i at time t, and consider the situation where a market order for a quantity w_i arrives at time t. If $p_{i,t-1}$ is the current price and $p_{i,t-1} + x_{i,t}$ is the price (of the

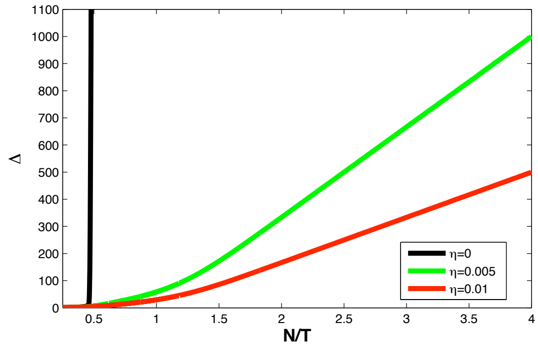

Figure 4.7: *Susceptibility as a function of N/T for different values of $\tilde{\eta}$ and $\beta = 0.7$.*

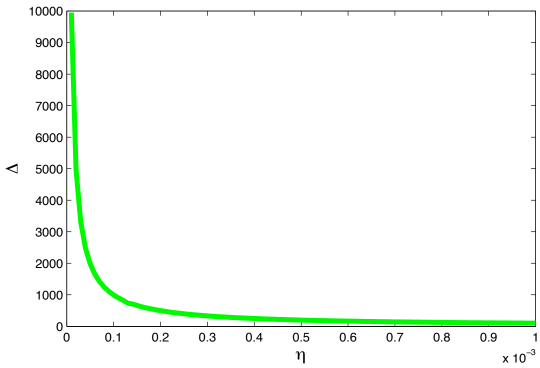

Figure 4.8: *Susceptibility as a function of $\tilde{\eta}$ for the case $t = 1.5$ and $\beta = 0.7$.*

transaction which occurred) just before the order arrives, then the price $p_{i,t}$ at which the transaction will take place is given by

$$w_i = \int_{p_{i,t-1}+x_{i,t}}^{p_{i,t}} dp \rho_i(p,t). \tag{4.3.47}$$

A linear impact, as the one assumed in Eq. (4.3.20), then corresponds to an order book with a constant density of limit orders. Hence, a measure of η is given by the density of the order book close to the best bid/ask. Since the density of the order book fluctuates and liquidity varies across assets, $\eta_{i,t}$ could also be taken as an asset dependent stochastic quantity.

Then the computation of the ES can still be performed in terms of the cost

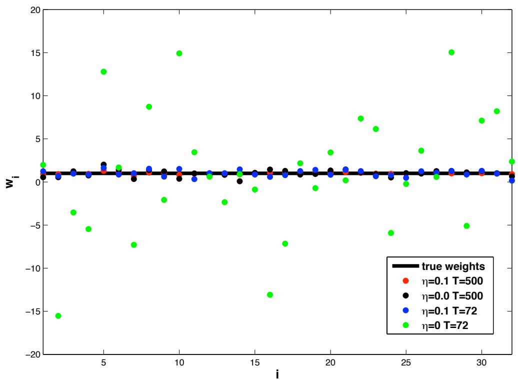

Figure 4.9: *Optimal weights for $N = 32$. Green dots: $T = 72$ and $\eta = 0$. Blue dots: $T = 72$ and $\eta = 0.1$. Black dots: $T = 500$ and $\eta = 0$. Red dots: $T = 500$ and $\eta = 0.1$. The solid line refers to the target optimal weights.*

function (4.3.23), but now with the T constraints

$$u_\tau + v + \sum_{i=1}^{N} w_i x_{i,\tau} - \sum_i \eta_{i,\tau} w_i^2 \geq 0 \tag{4.3.48}$$

in place of Eq. (4.3.25). In this case the mapping to a simple L_2 regularizer that we have laid out in section 4.3.1, is then complicated by the fact that the impact term depends on τ and cannot be absorbed into a τ independent constant. Nevertheless we do not expect the essential features of the problem to change with respect to the case of a constant η.

Note furthermore that different assumptions for the market impact function lead to different regularizers. For example, considering the instantaneous impact and Eq. (4.3.47) in the presence of a bid-ask spread, we expect the price to bounce from the bid to the ask, depending on the direction of trading (i.e. on the sign of w_i). This suggests a term proportional to the sign of w_i in the equation for the price, which, in turn, would then introduce an $L1$ regularizer. We address in the next subsection the effect of considering an L_1 regularizer in the portfolio optimization problem, in order to see that also in this case the regularized problem is more stable than the standard one.

4.3.4 Behavior of large random minimal risk portfolios under L_1 regularized expected shortfall

In order to develop some intuition about the problem, we focus again on the case of the Maximal Loss, that in presence of the L_1 regularizer we define as

$$ML(\vec{w}) = \max_{t=1,\dots,T} \left[-\sum_i w_i x_{i,t} + \eta ||\vec{w}|| \right], \tag{4.3.49}$$

where $||\vec{w}||$ is the L_1 norm form the vector of weights. Let us consider, as we already did for the L_2 case, the simple case of two asset and two times. The loss associated to each time can be expressed as

$$l_1 = -wx_{11} - (1-w)x_{21} + \eta|w| + \eta|1-w| \tag{4.3.50}$$
$$l_2 = -wx_{12} - (1-w)x_{22} + \eta|w| + \eta|1-w|, \tag{4.3.51}$$

where we have taken into account the budget constraint by taking $w_1 = w$ and $w_2 = 1 - w$. We distinguish now four different cases:

- if $0 < w < 1$, then $l_t = w(x_{2t} - x_{1t}) + \eta - x_{2t}$
- if $w > 1$, then $l_t = w(x_{2t} - x_{1t} + 2\eta) - x_{2t} - \eta$
- if $-1 < w < 0$, then $l_t = w(x_{2t} - x_{1t} - 2\eta) + \eta - x_{2t}$
- if $w < -1$, then $l_t = w(x_{2t} - x_{1t}) - \eta - x_{2t}$.

The instability may arise in the following situations:

- if $w < -1$, $x_{21} - x_{11} > 0$ and $x_{22} - x_{21} > 0$ (the two straights lines have both positive slope).
- if $w > 1$, $x_{21} - x_{11} < -2$ and $x_{22} - x_{21} < -2\eta$ (the two straights lines have both negative slope).

If we assume that $x_{i,t}$ are independent normal distributed random variables, the instability is then present with probability $P(\eta) = 1/4(1 + erfc^2(\eta))$. This has to be compared with the result in absence of regularizer, where there is an instability with probability 0.5. Since $P(\eta) \leq 0.5$, we conclude that the stability of the system is augmented by the presence of the regularizer, even though the instability is not removed [1].

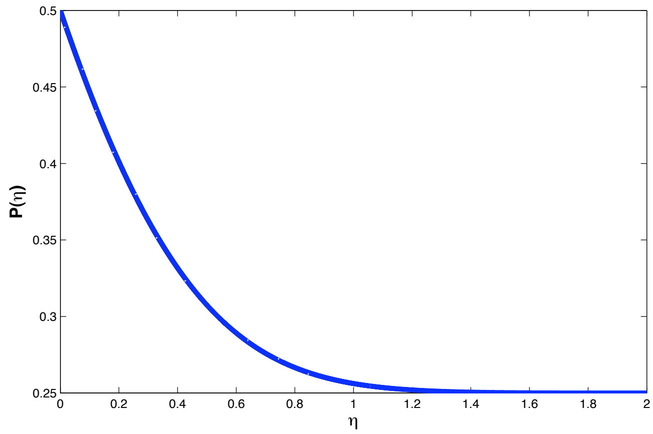

Figure 4.10: *Probability that the system is unstable as a function of η*

From the intuition that we gain from this very simple case we may expect that a shift of the transition towards the unfeasible region may be obtained introducing the L_1 regularizer in the optimization problem under ES. This generalized problem reads

$$\min_{\vec{w},\vec{u},\epsilon}\left[(1-\beta)T\epsilon+\sum_{\tau=1}^{T}u_\tau+\eta|\vec{w}|\right], \tag{4.3.52}$$

$$\text{s.t.}\quad \vec{w}\cdot\vec{x}_\tau+\epsilon+u_\tau\geq 0;\quad u_\tau\geq 0;\quad \forall\tau, \tag{4.3.53}$$

$$\sum_i w_i = WN. \tag{4.3.54}$$

Again, we can easily extend the calculation of Ciliberti et al. [2007] valid for independent gaussian returns. After some effort, the saddle point free energy

[1]Notice that with the same argument it is possible to understand that for any L^p norm with (integer) $p>1$ the instability disappears.

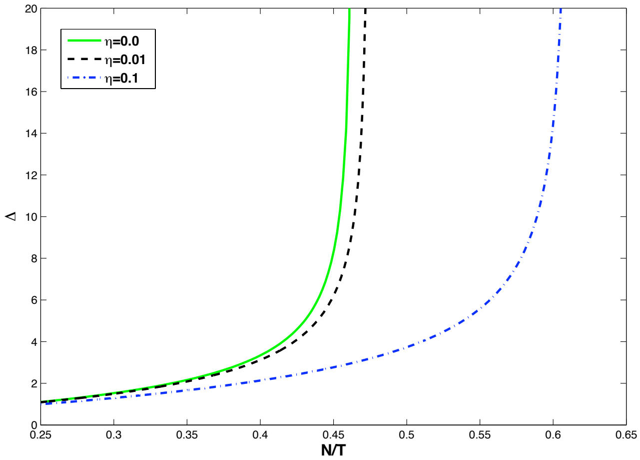

Figure 4.11: *Susceptibility as a function of T/N for different values of η for the L_1 regularize ES.*

may be written in terms of six order parameters (details about the calculation are reported in appendix C.3):

$$F(\tilde{\lambda}, \tilde{\epsilon}, \tilde{q}_0, \Delta, \tilde{\hat{q}}_0, \hat{\Delta}) = -\tilde{\lambda}W - t(1-\beta)\epsilon + \Delta\tilde{\hat{q}}_0 + \Delta^2\hat{\Delta}\tilde{q}_0 + \frac{1}{\gamma}\left\langle \log \int_{-\infty}^{\infty} dw e^{-\gamma V(w,z)} \right\rangle_z + \frac{t\Delta}{2\sqrt{\pi}} \int_{-\infty}^{\infty} dse^{-s^2} g(\tilde{\epsilon} + s\sqrt{2\tilde{q}_0}), \tag{4.3.55}$$

where $V(w,z) = \hat{\Delta}w^2 + \eta|w| - \tilde{\lambda}w - zw\sqrt{-2\tilde{\hat{q}}_0}$ and $\langle \cdots \rangle_z$ represent an average over the normal variable z.

The solution of the saddle point equations that can be derived from the above expression, confirms the expectation that the L_1 regularization produces a shift towards higher values of N/T of the feasible-unfeasible transition. Even tough the L_1 regularizer is not enough to prevent the occurrence of such transition, nevertheless the accessible region of phase space is greater with respect to the standard situation. The shift in the critical point due to the regularization is clearly shown in Figures 4.11 and 4.12, where we plot the behavior of the susceptibility and of the estimation error.

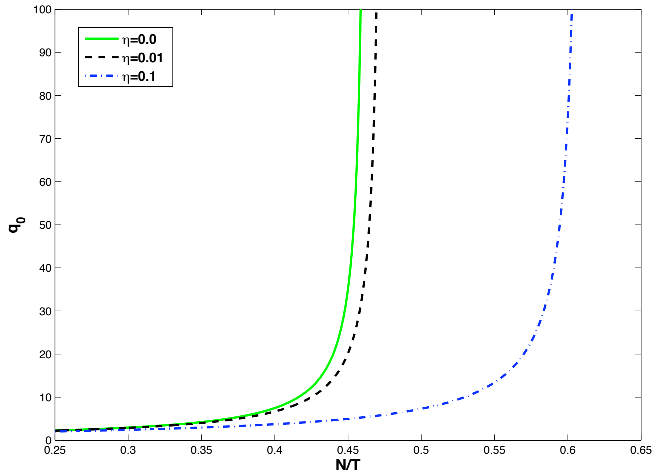

Figure 4.12: *Estimation error as a function of T/N for different values of η for the L_1 regularize ES.*

In summary, we have considered a generalized problem of portfolio selection where we accounted for the impact of liquidation strategies. We have explicitly shown in the case of Expected Shortfall that, once market impact has been taken into account, one has to solve a regularized problem where the norm of the vector of weights enters into the cost function. We have considered the cases of linear and instantaneous market impact, showing in both cases that the accessible region of phase space is augmented by the presence of the regularizer. Since market impact is a real feature of financial markets, one may at this point wonder whether the effect of finite liquidity may be detected in real data by means of an augmented stability of optimal portfolio solutions with respect to the case of purely random returns. Figure 4.13 refers to an analysis carried on using data from the New York Stock Exchange. Notably, we considered daily returns of the 41 more representative assets from ranging from 1970 to 2007. We can then rely on time series of length $T_{max} = 6908$. In order to see whether some differences with respect to the case of random data may arise when solving the optimal portfolio problem, we compared the results of the linear programming problem (4.2.12) obtained from the real time series and random time series. In order to make averages we proceeded in the following way. For a fixed value of T, we divided the real time

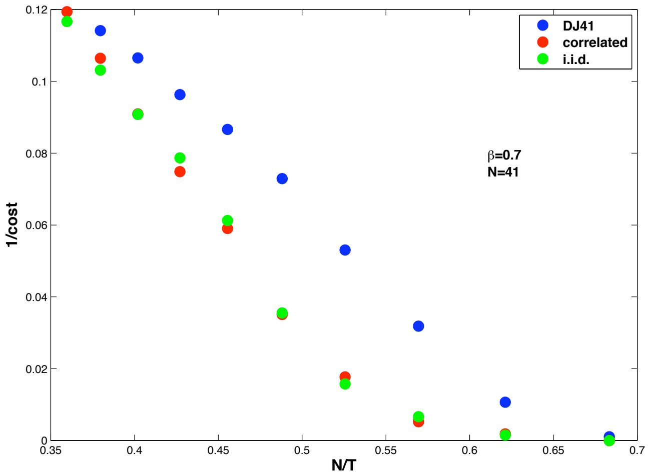

Figure 4.13: *Inverse of the minimum of the cost function for real data taken from the Dow Jones (blue dots), i.i.d. gaussian variables (green dots) and random variables with the same correlation of real data (red dots) .*

series into chunks of length T, we computed the solution of the linear programming problem and we averaged over chunks. Artificial data have been treated in the same way starting from random time series of length T_{max}. Figure 4.13 shows the inverse of the cost function computed for the optimal portfolio in three cases: real returns, independent gaussian returns and gaussian returns generated with the same correlation of real ones. Keeping in mind that in the unstable region of the optimization problem the cost function diverges, we can distinguish from the figure that a difference between solutions computed using real returns and solutions computed from random data exists. In particular the unstable region is shifted towards greater values of the ratio N/T. This may be due to the presence of a market impact in real data.

4.4 Summary and perspectives

We considered in this chapter the topic of portfolio optimization. Given a set of N assets, the problem we considered was that of finding the portfolio of minimal risk. As a measure of risk, we considered the so called Expected Shortfall, which belongs to the class of coherent risk measures. Coherent risk measures are characterized by instabilities when time series are too short with respect to the size of portfolio. We argued that a possible remedy for such instabilities may be the introduction of regularizers in the optimization problem. Notably, we showed that regularized problems naturally arise when accounting for the impact of portfolio liquidation. We explicitly showed that the introduction of an L_2 regularizer in the optimization problem under ES completely removes the instability of the risk measure, as well as the presence of an L_1 regularizer is enough to shift the instability, thus increasing the accessible volume of phase space. This is of particular interest since practical cases of portfolio selection operate in the regime $N \sim T$ where risk measures become unstable, so that taming fluctuations in this regime may be a good way to select a portfolio closer to the real optimal one. We also presented some preliminary results concerning the analysis of real data, that support the idea that market impact should be accounted for when searching for optimal investment strategies.

Chapter 5

Conclusions

In this thesis we have discussed problems of economic relevance through the prism of statistical mechanics of disordered systems. In all the problems we have considered, the aim was that of understanding how the presence of interactions between units at the "microscopic" scale reflects at the macroscopic level in non trivial emergent collective properties. The very complex nature of economic systems, built of individuals that interact according to heterogeneous behavioral rules, has been modeled by considering systems with random couplings. In this respect, statistical mechanics of disordered systems provides tools that allows for the analytical characterization of systems of heterogeneous agents and for the determination of phase diagrams to describe the typical properties of these systems in terms of few relevant parameters.

The first part of the thesis was devoted to discuss the consequences of some idealizations usually assumed in the modeling of financial markets. Notably, we first considered the topic of information efficiency. In the context of a simple model, we showed that as market become efficient they start being dominated by trend followers, which are usually associated with a destabilizing effect on the market. Moreover, the region of phase space where the market is perfectly efficient turns out to be a critical line where a phase transition of the second order occurs. In the context of the model we have introduced, we claimed that information efficiency may play a non-trivial role in triggering the occurrence of bubble events. A natural development of the present work is that of proving this

claim. A first step in this direction may be that of extending the framework by Hommes [2006] to the case of fundamentalists with different bits of information and to recover in this context the picture of information efficiency we have discussed in chapter 2. Another direction for future investigations may as well be that of introducing a dynamics for information costs, in such a way to see whether the market follows a self-organized path towards the critical line.
As a second point, we addressed the problem of the effect on the underlying market of the proliferation of financial instruments. By accounting for a feedback between derivative market and underlying one, which is due to financial institutions trading on the underlying for hedging derivatives, we showed that the proliferation of financial instruments drives the market towards a state that closely resemble that of the efficient, arbitrage free complete market described by APT. Also in this case, however, the same region of phase space is the locus of a sharp phase transition. In this respect we argued that the path towards ideal markets may be a path towards unstable markets, a point that should be taken into account when deciding for regulatory policies. Possible directions for future research include *i)* a deep discussion about possible stabilizing effects of taxes on trading activities, such as the Tobin tax; *ii)* the introduction of an utility function for consumption from which demand for derivatives may be derived; *iii)* extension of the present model to the case of an endogenous generated risk neutral measure, as that considered by Marsili [2009].

Although the problems we discussed may be of relevance in the discussion concerning policy issues, the tone of our discussion was definitely on the theoretical side in the first part of this thesis. The second part of the thesis was instead devoted to a problem of immediate impact on real practice, namely that of portfolio optimization. The problem is that of finding the portfolio that minimizes a certain measure of risk. In practice, the real risk is unknown and the optimization problem is done on the basis of estimations of risk made from historical data. Real data are however noisy and this may cause troubles related over-fitting, i.e. the optimal portfolio computed on the basis of historical data may be dramatically different from the true optimal one. Statistical mechanics has revealed very useful also in this context, allowing for the determination of precise phase diagrams where a sharp phase transition discriminates a region where the optimization problem has solutions and one where the optimization problem cannot be solved. Notably, the latter is characterized by large fluctuations and divergent estimation errors. In this context, we showed that accounting for the impact of liquidation strategies on the market leads to a regularized optimization problem where such problems are drastically reduced. Natural extensions of the present work concern *i)* the derivation of general results for the class of coherent risk measures; *ii)* the study of a generalized ES problem with a mixed L_1 and L_2 regularization, *iii)* a deeper analysis of real data, in order to show whether regularized optimization problem may improve the real practice.

In summary, we tried to show that concepts and tools borrowed from statistical mechanics may give an useful perspective in the discussion of problems related to financial markets. In the three cases we have discussed the major contribution of statistical mechanics has been that of allowing for a proper understanding of the connection between interactions at the "microscopic" scale and collective properties that appear at the "macroscopic" scale. This allowed also to put in the same framework problems in principle very different that, once seen through the lenses of statistical mechanics, have been recognized as different expressions of the same phenomena.

Appendix A

Information efficiency and financial stability

A.1 The statistical mechanics analysis

The competitive equilibrium solution of our problem can be obtained through the minimisation of the following Hamiltonian function

$$H_\epsilon = \frac{N^2}{4\Omega}\sum_{\omega,k_0}(R^\omega - p^{\omega,k_0})^2 + \frac{\epsilon}{2}\sum_{i,m} z_i^m, \tag{A.1.1}$$

with $p^{\omega,k_0} = \frac{1}{N}\sum_{i,m} z_i^m \delta_{k_i^\omega,m} + \sum_{k_0} \frac{z_0^{k_0}}{N}$. In order to compute the minima of H we introduce the partition function

$$Z(\beta) = \int_0^\infty dz_0^+ \int_0^\infty dz_0^- \cdots \int_0^\infty dz_N^+ \int_0^\infty dz_N^- e^{-\beta H_\epsilon\{z_i^m\}}. \tag{A.1.2}$$

In the limit $\beta \to \infty$ integrals are dominated by those configurations $\{z_i^m\}$ that minimise the Hamiltonian. The central quantity to compute is the free energy $f_\beta = -\beta^{-1}\log Z(\beta)$, which has to be averaged over the realisations of the disorder, namely $\{k_i^\omega, R^\omega\}$. In the following we are going to consider $k_i^\omega = \pm 1$ with equal probability $\forall\ i,\ \omega$, and we take $R^\omega = R + \frac{\tilde{R}}{\sqrt{N}}$, where $\tilde{R}$ are Gaussian variables with zero mean and variance equal to s^2. In order to compute the average over the disorder $\langle f_\beta \rangle$ we can resort to the so called replica trick through the identity $\log Z = \lim_{M\to 0}(Z^M - 1)/M$. The problem reduces then to that of computing the average over the disorder of the partition function of M non interacting replicas of the system:

$$\begin{aligned}\langle Z^M \rangle &= \Big\langle \mathrm{Tr}_{\{z\}} \prod_a \delta\left(N\overline{R} - \sum_i \overline{z}_{i,a} - \overline{z}_{0,a}\right) \times \\ &\quad e^{-\beta\left[\sum_{a,\omega,k0}(NR^\omega - \sum_{i,m} z_i^m \delta_{k_i^\omega,m} - z_0^{k0})^2 + \epsilon \sum_{i,a} \frac{z_{i,a}^+ + z_{i,a}^-}{2}\right]} \Big\rangle,\end{aligned}$$

with $a \in \{1,\ldots,M\}$, $i \in \{1,\ldots,N\}$, $\omega \in \{1,\ldots,\Omega\}$, $m \in \{-1,1\}$ and $k0 \in \{-1,1\}$ and $\overline{z}_{i.a} = (z_{i,a}^+ + z_{i,a}^-)/2$. We verified through numerical simulations that, for the specific public signal k_0 that we considered in this paper, $\langle z_0^+ \rangle = \langle z_0^- \rangle$ so, in order to simplify the calculation, we make the assumption $z_0^+ = z_0^- = z_0$. After performing a Hubbard-Stratonovich transformation in order to linearize the quadratic term of the Hamiltonian, taking the average over the quenched variables introduces an effective interaction between replicas:

$$\begin{aligned}\langle Z^n \rangle &= \Big\langle \int \{dQ_{a,b}\}\{d\hat{Q}_{a,b}\}\{d\hat{R}\}\mathrm{Tr}_{\{z\}} \\ &\quad e^{-\sum_{a,b}\hat{Q}_{a,b}\left(NQ_{a,b} - \sum_i \Delta_i^a \Delta_i^b\right) - \sum_a \hat{R}_a\left(N\overline{R} - \sum_i \overline{z}_{i,a} - z_{0,a}\right)} \times \\ &\quad e^{-\beta N/\Omega \sum_{a,b,\omega}(\tilde{R}^\omega)^2\left(\frac{\beta Q_{a,b}}{\alpha} + \delta_{a,b}\right)^{-1} - \beta\epsilon\sum_{i,a}\overline{z}_{i,a}} \times \\ &\quad e^{-\frac{\Omega}{2}\mathrm{Tr}\log\left(\frac{\beta Q_{a,b}}{\alpha} + \delta_{a,b}\right)} \Big\rangle,\end{aligned}$$

where we have introduced the overlap matrix $Q_{a,b}$ and the variables $\Delta_{i.a} = (z_{i,a}^+ - z_{i,a}^-)/2$, while $\hat{Q}_{a,b}$ and $\hat{R}_a$ are conjugated variables that come from integral representations of δ functions:

$$\delta(X - X_0) \propto \int d\hat{X} e^{-\hat{X}(X - X_0)}. \tag{A.1.3}$$

In order to make further progress we consider the replica symmetric ansatz, namely we take

$$Q_{a,b} = q_0 + \alpha\frac{\chi}{\beta}\delta_{a,b} \tag{A.1.4}$$

$$\hat{Q}_{a,b} = -\frac{\beta^2 \hat{q}_0}{\alpha^2} + \frac{\beta^2\hat{q}_0/\alpha^2 + \beta w/\alpha}{2}\delta_{a,b} \tag{A.1.5}$$

The resulting expression is handled in such a way to be able to use saddle point methods in the limit $N, \beta \to \infty$ (see Caccioli et al. [2009] for more details on a similar calculation). The final result is given in terms of the free energy

$$f(q_0, \chi, \hat{q}_0, w, \hat{R}, z_0) = \frac{s^2 + q_0}{1+\chi} + 2\frac{\hat{R}\overline{R}}{\alpha} - 2\frac{\hat{R}z_0}{\alpha} + \frac{\chi\hat{q}_0}{\alpha} - \frac{wq_0}{\alpha} + \frac{2}{\alpha}\Big\langle \min_{z\geq 0}\{V(z)\}\Big\rangle_t, \tag{A.1.6}$$

with the potential $V(z)$ given by

$$V(z) = \frac{w}{2}\Delta^2 - \sqrt{\hat{q}_0}t\Delta - \hat{R}\overline{z} + \epsilon\overline{z} \tag{A.1.7}$$

and where we used $\langle \cdots \rangle_t$ to denote averages over the normal variable t. The corresponding saddle point equations are

$$\begin{aligned} w &= \frac{\alpha}{1+\chi} && \text{(A.1.8)}\\ \hat{q}_0 &= \frac{\alpha(s^2+q_0)}{(1+\chi)^2} && \text{(A.1.9)}\\ \overline{R} &= z_0 + \langle \Delta^* \rangle_t && \text{(A.1.10)}\\ q_0 &= \langle \Delta^{*2} \rangle_t && \text{(A.1.11)}\\ \chi &= \frac{\langle t\Delta^* \rangle_t}{\sqrt{\hat{q}_0}} && \text{(A.1.12)}\\ \hat{R} &= 0, && \text{(A.1.13)} \end{aligned}$$

where

$$\begin{aligned} \Delta^*(t) &= \theta(t-\tau)\frac{\sqrt{\hat{q}_0}}{w}(t-\tau) + \theta(-t-\tau)\frac{\sqrt{\hat{q}_0}}{w}(-t-\tau) && \text{(A.1.14)}\\ \tau &= \frac{\epsilon}{\sqrt{\hat{q}_0}}. && \text{(A.1.15)} \end{aligned}$$

Using these equations it is possible to compute $\langle H_\epsilon \rangle = \frac{q_0+s^2}{(1+\chi)^2}$. It is useful to define the three functions

$$\begin{aligned} \psi_r(\tau) &= 2\int_\tau^\infty dte^{-t^2/2}(t-\tau) = \sqrt{\frac{2}{\pi}}e^{-\tau^2/2} - \tau\mathrm{erfc}\left(\frac{\tau}{\sqrt{2}}\right) && \text{(A.1.16)}\\ \psi_q(\tau) &= 2\int_\tau^\infty dte^{-t^2/2}(t-\tau)^2 = (1+\tau^2)\mathrm{erfc}\left(\frac{\tau}{\sqrt{2}}\right) - \sqrt{\frac{2}{\pi}}\tau e^{-\tau^2/2} && \text{(A.1.17)}\\ \psi_\chi(\tau) &= 2\int_\tau^\infty dte^{-t^2/2}t(t-\tau) = \mathrm{erfc}\left(\frac{\tau}{\sqrt{2}}\right) && \text{(A.1.18)} \end{aligned}$$

It is now possible to express equations (A.1.10), (C.1.6) and (B.1.23) in terms of these non-linear functions.We can now look for a parametric solution in terms of τ, and consider α as an independent variable. From the definition of τ we have $\hat{q}_0 = \epsilon^2/\tau^2$. Inserting equation (A.1.8) into equation (B.1.23) we find

$$\alpha = \frac{1+\chi}{\chi}\psi_\chi(\tau), \tag{A.1.19}$$

while from equation (C.1.6) we get

$$q_0 = \frac{\epsilon^2}{\tau^2}\frac{\psi_q(\tau)}{\psi_\chi^2(\tau)}\chi^2. \tag{A.1.20}$$

Inserting these expressions into equation (A.1.9) we obtain

$$\frac{\epsilon^2}{\tau^2} = \frac{s^2\psi_\chi(\tau)}{\chi(1+\chi)} + \frac{\epsilon^2}{\tau^2}\frac{\psi_q(\tau)\chi}{\psi_\chi(\tau)(1+\chi)}, \tag{A.1.21}$$

from which

$$\chi_\pm = \frac{-1 \pm \sqrt{1 + 4\psi_\chi(\tau)s^2\frac{\tau^2}{\epsilon^2}\left(1 - \frac{\psi_q(\tau)}{\psi_\chi(\tau)}\right)}}{2(1-\psi_q(\tau)/\psi_\chi(\tau))}. \tag{A.1.22}$$

Since χ has the meaning of a distance between replicas the only physical solution is $\chi = \chi_+$. Inserting this expression for χ in the previous equations makes possible to express all order parameters and α in terms of the functions ψ_r, ψ_q, ψ_χ and of the free parameters ϵ and τ.

A parametric solution can be found also for the case of α fixed and ϵ variable. From equation (B.1.23) we find

$$\chi = \frac{\psi_\chi(\tau)}{\alpha - \psi_\chi(\tau)}. \tag{A.1.23}$$

From equation (C.1.6)

$$q_0 = \frac{\epsilon^2}{\tau^2}\frac{(1+\chi)^2}{\alpha^2}\psi_q(\tau). \tag{A.1.24}$$

Finally, inserting this expression in equation (A.1.9) we can now express ϵ as:

$$\frac{\epsilon^2}{\tau^2} = \frac{\alpha s^2}{(1+\chi)^2}\frac{1}{1 - \frac{\psi_q(\tau)}{\alpha}}. \tag{A.1.25}$$

As before, using this expression, is now possible to write the order parameters in terms of ψ_r, ψ_q, ψ_χ and of the free parameters α and τ.

Appendix B

Proliferation of derivatives and market stability

B.1 The statistical mechanics analysis

The problem is the one of finding the ground state of the Hamiltonian

$$H = \frac{1}{2}\sum_{\omega} \pi^{\omega}(r^{\omega})^2 + \sum_i \frac{\epsilon_i}{\Omega} s_i = \frac{1}{2\Omega}\sum_{\omega=1}^{\Omega}(r^{\omega})^2 + \sum_{i=1}^{N} f(s_i). \tag{B.1.1}$$

Let's write down the partition function

$$\begin{aligned}
Z_{q,a} &= \mathrm{Tr}_{r,s} e^{-\beta H} \delta\left(\sum_{\omega} q^{\omega} r^{\omega}\right) \prod_{\omega=1}^{\Omega} \delta\left(r^{\omega} - d^{\omega} - \sum_i s_i a_i^{\omega}\right) && \text{(B.1.2)}\\
&= \int \frac{du}{2\pi} \mathrm{Tr}_{r,s} e^{-\beta \sum_i f(s_i)} \prod_{\omega=1}^{\Omega} e^{iur^{\omega}q^{\omega} - \frac{\beta}{2} r^{\omega 2}} \delta\left(r^{\omega} - d^{\omega} - \sum_i s_i a_i^{\omega}\right) && \text{(B.1.3)}\\
&= \int \frac{du}{2\pi} \mathrm{Tr}_{r,s,\xi} e^{-\beta \sum_i f(s_i)} \prod_{\omega=1}^{\Omega} e^{iur^{\omega}q^{\omega} - \frac{\beta}{2} r^{\omega 2} + ir^{\omega}\xi^{\omega} - id^{\omega}\xi^{\omega} - i\xi^{\omega}\sum_i s_i a_i^{\omega}} && \text{(B.1.4)}
\end{aligned}$$

where we have used the shorthand Tr to indicate integrals on the variables in the index. Here it is understood that all variables s_i are integrated from 0 to s_0 and all variables r^ω and ξ^ω are integrated over all the real axis, with a factor $1/(2\pi)$ for each ω. The next step is to replicate this, i.e. to write

$$\begin{aligned} Z^m_{q,a} &= \int \frac{d\vec{u}}{2\pi} \mathrm{Tr}_{\vec{r},\vec{s},\vec{\xi}} e^{-\beta\sum_{i,a} f(s_{i,a})} \prod_{\omega=1}^{\Omega} e^{i\sum_a [u_a r^\omega_a q^\omega - \frac{\beta}{2} {r^\omega}^2_a + i r^\omega_a \xi^\omega_a - i d^\omega \xi^\omega_a] - i\sum_i a^\omega_i \sum_a s_{i,a}\xi^\omega_a} \\ &= \int \frac{d\vec{u}}{2\pi} \mathrm{Tr}_{\vec{s},\vec{\xi}} e^{-\beta\sum_{i,a} f(s_{i,a})} \prod_{\omega=1}^{\Omega} e^{-\frac{1}{2\beta}\sum_a (u_a q^\omega + \xi^\omega_a)^2 - i d^\omega \sum_a \xi^\omega_a - i\sum_i a^\omega_i \sum_a s_{i,a}\xi^\omega_a} \end{aligned}$$

where the sums on a runs over the m replicas. In the second equation above, we have performed the integrals over r^ω_a. We can now perform the average over the random variables a^ω_i which will be assumed to be normal with zero average and variance $1/\Omega$. This yields

$$\begin{aligned} \langle Z^m_q \rangle &= \int d\mathbf{G} \int \frac{d\vec{u}}{2\pi} \mathrm{Tr}_{\vec{\xi}} \prod_{\omega=1}^{\Omega} e^{-\frac{1}{2\beta}\sum_a (u_a q^\omega + \xi^\omega_a)^2 - i d^\omega \sum_a \xi^\omega_a - \frac{1}{2}\sum_{a,b} \xi^\omega_a \xi^\omega_b G_{a,b}} \langle I_s(\mathbf{G}) \rangle_\epsilon \\ \langle I_s(\mathbf{G}) \rangle_\epsilon &= \left\langle \mathrm{Tr}_{\vec{s}} e^{-\beta\sum_{i,a} f(s_{i,a})} \prod_{a\le b} \delta\left(\Omega G_{a,b} - \sum_i s_{i,a}s_{i,b} \right) \right\rangle_\epsilon, \end{aligned}$$

where the average $\langle \ldots \rangle_\epsilon$ is taken over gaussian variables with mean $\bar{\epsilon}$ and variance σ^2_ϵ and the symbol $d\mathbf{G}$ stands for integration over all the independent entries of the matrix $\mathbf{G}$. In order to evaluate the latter we use a standard delta function identity bringing into play the matrix $\mathbf{R} = R_{a,b}$ conjugated to the overlap matrix $\mathbf{G} = G_{a,b}$. The ϵ-average is evaluated as follows

$$\begin{aligned} \langle I_s(\mathbf{G}) \rangle_\epsilon &= \int d\mathbf{R} \mathrm{Tr}_{\vec{s}} e^{\sum_{a\le b} R_{ab}[\Omega G_{ab} - \sum_i s_{i,a}s_{i,b}]} \int_{-\infty}^{+\infty} \prod_i \left[d\epsilon_i e^{-\frac{(\epsilon_i - \bar{\epsilon})^2}{2\sigma^2_\epsilon} - \beta\sum_a f(s_{i,a})} \right] \\ &= \int d\mathbf{R} \mathrm{Tr}_{\vec{s}} e^{\sum_{a\le b} R_{ab}[\Omega G_{ab} - \sum_i s_{i,a}s_{i,b}] - \beta\sum_{i,a} \bar{\epsilon} s^a_i + \frac{\Omega\beta^2\sigma^2_\epsilon}{2}\sum_{a,b} G_{a,b}}, \end{aligned}$$

where the quadratic term $(\sum_a s^a_i)^2$ arising from the gaussian integration has been replaced by the overlap matrix. Taking the replica symmetric (RS) ansatz for $\mathbf{G}$ and $\mathbf{R}$

$$G_{ab} = g + \frac{\chi}{\beta}\delta_{a,b}, \qquad R_{ab} = -\beta^2 r^2 + \frac{\beta^2 r^2 + \beta\nu}{2}\delta_{a,b} \tag{B.1.5}$$

the $\vec{s}$-independent part of the exponent in the integrand takes the form

$$\begin{aligned}
&\Omega \sum_{a\leq b} R_{a,b}G_{a,b} + \tfrac{\Omega\beta^2\sigma_\epsilon^2}{2}\sum_{a,b} G_{a,b} \\
&= \Omega m\left(g+\tfrac{\chi}{\beta}\right)\left(\tfrac{\beta\nu}{2}-\tfrac{\beta^2 r^2}{2}\right) - \tfrac{\Omega m(m-1)}{2} g\beta^2 r^2 + \tfrac{\Omega\sigma_\epsilon^2}{2}(m^2\beta^2 g + m\beta\chi) \\
&\simeq \tfrac{\Omega m\beta}{2}\left(\nu g - r^2\chi + \sigma_\epsilon^2\chi + \nu\chi/\beta\right)
\end{aligned}$$

where we have neglected terms of order m^2 in view of the $m \to 0$ limit. This yields

$$I_s(\mathbf{G}) = \int d\mathbf{R} e^{\frac{\Omega m\beta}{2}[\nu g - r^2\chi+\sigma_\epsilon^2\chi+\nu\chi/\beta]} W_{\bar{\epsilon}}[\nu,r]^N, \tag{B.1.6}$$

where we have defined

$$W_{\bar{\epsilon}}[\nu,r] = \left[\mathrm{Tr}_{\vec{s}} e^{-\beta\sum_a\left[\bar{\epsilon}s_a+\frac{\nu}{2}s_a^2\right]+\frac{1}{2}\left(\beta r\sum_a s_a\right)^2}\right]^N. \tag{B.1.7}$$

In the limit $m \to 0$ this quantity can be evaluated as follows

$$\begin{aligned}
W_{\bar{\epsilon}}[\nu,r] &= e^{N\log\left[\mathrm{Tr}_{\vec{s}} e^{-\beta\sum_a\left[\bar{\epsilon}s_a+\frac{\nu}{2}s_a^2\right]+\frac{1}{2}(\beta r\sum_a s_a)^2}\right]} \\
&= \exp\left\{N\log\left[\mathrm{Tr}_{\vec{s}} e^{-\beta\sum_a\left[\bar{\epsilon}s_a+\frac{\nu}{2}s_a^2\right]}\left\langle e^{z\beta r\sum_a s_a}\right\rangle_z\right]\right\}
\end{aligned}$$

where we have performed a Hubbard-Stratonovich transformation in order to decouple the $\{s^a\}$ variables introducing the average $\langle\ldots\rangle_z$ over the gaussian variable z. Clearly:

$$W_{\bar{\epsilon}}[\nu,r] = \exp\left\{N\log\left\langle\left(\int_0^{s_0} ds e^{-\beta[s^2+s(\bar{\epsilon}-zr)]}\right)^m\right\rangle_z\right\}. \tag{B.1.8}$$

Exploiting the usual identity

$$\log\langle X^m\rangle \simeq m\langle\log X\rangle, \tag{B.1.9}$$

valid for $m \to 0$, we finally obtain

$$W_{\bar{\epsilon}}[\nu,r] = \exp\left\{Nm\left\langle\log\int_0^{s_0} ds e^{-\beta[s^2+s(\bar{\epsilon}-zr)]}\right\rangle_z\right\}. \tag{B.1.10}$$

Inserting (B.1.10) into (B.1.6) we eventually get

$$I_s(\mathbf{G}) = \int d\mathbf{R} e^{\Omega m\beta\left\{\frac{1}{2}[\nu g - r^2\chi+\sigma\epsilon^2\chi+\nu\chi/\beta]+\frac{n}{\beta}\left\langle\log\int_0^\infty ds e^{-\beta[\bar{\epsilon}s+\nu s^2/2-rsz]}\right\rangle_z\right\}} \tag{B.1.11}$$

After inserting (B.1.11) into (B.1.5) we have to perform the $(m \cdot \Omega)$ integrals in ξ_a^ω. For each ω

$$J[\chi, g, u] \quad = \int d\vec{\xi} e^{-\frac{1}{2\beta}\sum_a (u_a q^\omega + \xi_a)^2 - id^\omega \sum_a \xi_a - \frac{1}{2}\sum_{a,b}\xi_a \xi_b G_{a,b}} \tag{B.1.12}$$

The matrix $\mathbf{G}$, in the RS ansatz has two distinct eigenvalues:

$$a_\parallel \quad = mg + \tfrac{\chi}{\beta} \qquad \text{multiplicity } 1 \tag{B.1.13}$$

$$a_\perp \quad = \tfrac{\chi}{\beta} \qquad \text{multiplicity } m-1. \tag{B.1.14}$$

Therefore the determinant of $\left(\mathbf{G} + \frac{\mathbf{I}}{\beta}\right)$ is clearly

$$\det\left(\mathbf{G} + \frac{\mathbf{I}}{\beta}\right) = e^{-\frac{1}{2}\left[m \log(\chi+1) + \log\left(1 + \frac{\beta m g}{1+\chi}\right)\right]}. \tag{B.1.15}$$

The expression (B.1.15) assists performing the gaussian integral in (B.1.12) as

$$J[\chi, g, u] = e^{-\frac{1}{2}\log\left(1 + \beta m \frac{g}{1+\chi}\right) - \frac{m}{2}\log(1+\chi) + \frac{\beta m}{2(1+\chi+m\beta g)}(uq^\omega + id^\omega)^2 - \frac{\beta m}{2}(uq^\omega)^2} \tag{B.1.16}$$

where we have rescaled u as $u \to \beta u$ and considered for u^a the form $u^a = u \; \forall a$. Taking the average over q^ω in the limit $m \to 0$ we use the fact that

$$\langle e^{mf(q)} \rangle_q \simeq 1 + m\langle f(q) \rangle_q \simeq e^{m\langle f(q)\rangle_q}$$

which in our case yields

$$\langle f(q) \rangle_q = -\frac{\beta\chi}{1+\chi}u^2 - \frac{1}{2}\frac{\beta\langle d^2\rangle}{1+\chi},$$

having taken for q an exponential distribution with average 1 . Hence

$$\langle Z_q^m \rangle \quad = \quad \int d\mathbf{G} \int d\mathbf{R} \int \frac{d\vec{u}}{2\pi} e^{m\Omega\beta F} \tag{B.1.17}$$

$$F \quad = \quad -\frac{u^2\chi}{1+\chi} - \frac{1}{2}\frac{g + \langle d^2\rangle}{1+\chi} - \frac{1}{2\beta}\log(1+\chi) + \frac{1}{2}\left[\nu g - r^2\chi + \sigma_\epsilon^2\chi + \frac{\nu\chi}{\beta}\right]$$
$$+ \frac{n}{\beta}\left\langle \log \int_0^{s_0} ds e^{-\beta[\bar{\epsilon}s + \nu s^2/2 - rsz]} \right\rangle_z . \tag{B.1.18}$$

We first observe that the saddle point equation for u yields $u = 0$. Then we take the limit $\beta \to \infty$ which gives

$$F = -\frac{1}{2}\frac{g + \langle d^2\rangle}{1+\chi} + \frac{1}{2}\left[\nu g - r^2\chi + \sigma_\epsilon^2\chi\right] - n\left\langle \min_{0\le s\le s_0}\left[\bar{\epsilon}s + \frac{\nu}{2}s^2 - rsz\right]\right\rangle_z \tag{B.1.19}$$

The saddle points equations, obtained differentiating (B.1.18) with respect to the order parameters and sending β to ∞, read

$$
\begin{aligned}
r^2 &= \frac{g+\langle d^2\rangle}{(1+\chi)^2}+\sigma_\epsilon^2 && \text{(B.1.20)}\\
\nu &= \frac{1}{1+\chi} && \text{(B.1.21)}\\
g &= n\langle s_z^2\rangle_z && \text{(B.1.22)}\\
r\chi &= n\langle s_z z\rangle_z && \text{(B.1.23)}\\
& && \text{(B.1.24)}
\end{aligned}
$$

where $s_z = \min\{s_0, \max\{0, (rz-\bar{\epsilon})/\nu\}\}$, since, in our case, the supply is limited to $0 \le s \le s_0$. The above set of equations can then be recasted in the form

$$
\begin{aligned}
s_z &= \min\left\{1, \max\left\{0, \left(z\sqrt{\tfrac{g+\langle d^2\rangle}{(1+\chi)^2}+\sigma_\epsilon^2}-\bar{\epsilon}\right)(1+\chi)\right\}\right\} && \text{(B.1.25)}\\
g &= n\mathbb{E}_z[s_z^2] && \text{(B.1.26)}\\
\chi &= \tfrac{n\mathbb{E}_z[s_z z](1+\chi)}{\sqrt{g+\Delta+\sigma_\epsilon^2(1+\chi)^2}}. && \text{(B.1.27)}
\end{aligned}
$$

The above calculation can also be performed, in order to probe the solution, introducing an auxiliary field h^ω coupled to the returns r^ω. This allows also to easily compute the average and the fluctuations of returns by deriving with respect to h the logarithm of the free energy and then setting $h=0$:

$$
\begin{aligned}
\bar{r} &= \sum_\omega \pi^\omega r^\omega = \frac{\bar{d}}{1+\chi} && \text{(B.1.28)}\\
\delta r^2 &= \sum_\omega \pi^\omega (r^\omega-\bar{r})^2 = \frac{g+\langle d^2\rangle-\langle d\rangle^2}{(1+\chi)^2} && \text{(B.1.29)}
\end{aligned}
$$

B.2 Computation of the critical line

We show here how it is possible to find the critical line in the case of unbounded supply. Let us consider the case $s \in [o,\infty)$, so that $s_z = \max\left\{0, \frac{r}{\nu}(z-z_0)\right\}$, with $z_0 = \frac{\bar{\epsilon}}{r}$. From the saddle pont equations we get

$$
g = \frac{n[\langle d^2\rangle + \sigma_\epsilon^2(1+\chi)^2]I_2(z_0)}{1-nI_2(z_0)} \qquad \text{(B.2.30)}
$$

where we have defined $I_2(z_0) = \int_0^\infty dz \frac{e^{-z^2/2}}{\sqrt{2\pi}}(z-z_0)^2$. Inserting now this expression for g into equation (B.1.20) and explointing $r = \bar{\epsilon}/z_0$ we obtain

$$\frac{\bar{\epsilon}^2}{z_0^2}(1 - nI_2(z_0)) = \sigma_\epsilon^2 + \frac{\langle d^2 \rangle}{(1+\chi)^2} \tag{B.2.31}$$

that using (B.1.23) can be writen as

$$\frac{\bar{\epsilon}^2}{z_0^2}(1 - nI_2(z_0)) = \sigma_\epsilon^2 + \langle d^2 \rangle (1 - nI_1(z_0))^2, \tag{B.2.32}$$

where $I_1 = \int_0^\infty dz \frac{e^{-z^2/2}}{\sqrt{2\pi}}(z-z_0)z$. If we now look for a solution in the case $\chi \to \infty$ the above equation reduces to

$$\frac{\bar{\epsilon}^2}{\sigma_\epsilon^2}\left(1 - \frac{I_2(z_0)}{I_1(z_0)}\right) = z_0^2 \tag{B.2.33}$$

and we also have that $n = 1/I_1(z_0)$. These equations define the critical line. In particular it is clear at this level that the dependence on the parameters of the risk premia distribution enters through the ratio $\bar{\epsilon}/\sigma_\epsilon$

Appendix C

Optimal liquidation strategies regularize portfolio selection

C.1 The replica calculation for the L_2 regularized Expected Shortfall

We present here the replica calculation used to solve the following optimization problem: find the minimum of the cost function

$$E[\epsilon, \{u_\tau\}] = (1-\beta)T\epsilon + \sum_{\tau=1}^{T} u_\tau + \tilde{\eta}\|w\|^2$$

under the constraints

$$u_\tau \geq 0,$$

$$u_\tau + \epsilon + \sum_{i=1}^{N} x_{i,\tau} w_i \geq 0$$

and

$$\sum_i w_i = WN.$$

The replicated partition function, corresponding to the partition function of n copies of the system can be computed as

$$
\begin{aligned}
Z_\gamma^n[x_{i,\tau}] &= \int_{-\infty}^{\infty} \prod_{a=1}^{n} d\epsilon^a \int_0^{\infty} \prod_{\tau=1}^{T} \prod_{a=1}^{n} du_\tau^a \int_{-\infty}^{\infty} \prod_{i=1}^{N} \prod_{a=1}^{n} dw_i^a \int_{-\infty}^{\infty} \prod_{a=1}^{n} d\lambda^a \\
&\times \int_0^{\infty} \prod_{\tau=1}^{T} \prod_{a=1}^{n} d\mu_\tau^a \int_{-\infty}^{\infty} \prod_{\tau=1}^{T} \prod_{a=1}^{n} d\hat{\mu}_\tau \exp\left\{ \sum_a \lambda^a (\sum_i w_i^a - WN) \right\} \\
&\times \prod_\tau \exp\left\{ \sum_a i\hat{\mu}_\tau^a \left(u_\tau^a + \epsilon^a + \sum_i x_{i,\tau} w_i^a - \mu_\tau^a \right) \right\} \\
&\times \exp\left\{ -\gamma \sum_a (1-\beta) T \epsilon^a - \gamma \sum_{a,\tau} u_\tau^a - \gamma \tilde{\eta} \sum_i w_i^{a2} \right\}.
\end{aligned}
$$

Averaging over the quenched variables $\{x_{i,\tau}\}$ and introducing the overlap matrix $Q_{a,b} = \frac{1}{N} \sum_i w_i^a w_i^b$ one obtains

$$
\begin{aligned}
Z_\gamma^n[x_{i,\tau}] &= \int [D\epsilon][Du][Dw][D\lambda][D\mu][D\hat{\mu}][DQ][D\hat{Q}] \exp\left\{ \sum_a \lambda^a (\sum_i w_i^a - WN) \right\} \\
&\times \prod_\tau \exp\left\{ -\frac{1}{2} \sum_{a,b} \hat{\mu}_\tau^a Q_{a,b} \hat{\mu}_\tau^b \right\} \exp\left\{ \sum_{a,b} \hat{Q}_{a,b} \left(NQ_{a,b} - \sum_i w_i^a w_i^b \right) \right\} \\
&\times \exp\left\{ -\gamma \sum_a (1-\beta) T \epsilon^a - \gamma \sum_{a,\tau} u_\tau^a - \gamma \tilde{\eta} \sum_i w_i^{a2} \right\} \\
&\times \prod_\tau \exp\left\{ i \sum_a \hat{\mu}_\tau^a \left(u_\tau^a + \epsilon^a - \mu_\tau^a \right) \right\}.
\end{aligned}
$$

We can now perform the Gaussian integral over the variables $\{\hat{\mu}_\tau^a\}$:

$$\begin{aligned}
Z_\gamma^n[x_{i,\tau}] &= \int [D\epsilon][Du][Dw][D\lambda][D\mu][DQ][D\hat{Q}] \exp\left\{\sum_a \lambda^a (\sum_i w_i^a - WN)\right\} \\
&\times \exp\left\{-\gamma \sum_a (1-\beta)T\epsilon^a - \gamma \sum_{a,\tau} u_\tau^a - \gamma\tilde{\eta} \sum_i w_i^{b2}\right\} \\
&\times \exp\left\{\sum_{a,b} \hat{Q}_{a,b} \left(NQ_{a,b} - \sum_i w_i^a w_i^b\right)\right\} \\
&\times \prod_\tau \exp\left\{-\frac{1}{2}\sum_{a,b} \left(u_\tau^a + \epsilon^a - \mu_\tau^a\right) Q_{a,b}^{-1} \left(u_\tau^b + \epsilon^b - \mu_\tau^b\right)\right\} \\
&\times \exp\left\{-\frac{T}{2} \operatorname{tr}\log Q\right\}.
\end{aligned}$$

We are now allowed to perform a Gaussian integration over the variables $\{w_i^a\}$, which is going to bring the inverse of the operator $\hat{Q}_{a,b} + \gamma\tilde{\eta}\delta_{a,b}$ into the game:

$$\begin{aligned}
Z_\gamma^n[x_{i,\tau}] &= \int [D\epsilon][Du][D\lambda][D\mu][DQ][D\hat{Q}] \exp\left\{\sum_a -w\lambda^a N\right\} \\
&\times \exp\left\{-\gamma \sum_a (1-\beta)T\epsilon^a - \gamma \sum_{a,\tau} u_\tau^a\right\} \\
&\times \prod_\tau \exp\left\{-\frac{1}{2}\sum_{a,b} \left(u_\tau^a + \epsilon^a - \mu_\tau^a\right) Q_{a,b}^{-1} \left(u_\tau^b + \epsilon^b - \mu_\tau^b\right)\right\} \\
&\times \exp\left\{\sum_{a,b} \hat{Q}_{a,b} N Q_{a,b}\right\} \exp\left\{-\frac{T}{2} \operatorname{tr}\log Q\right\} \exp\left\{-\frac{nN}{2}\log 2\right\} \\
&\times \exp\left\{-\frac{N}{2} \operatorname{tr}\log(\hat{Q} + \gamma\tilde{\eta}\delta_{a,b})\right\} \exp\left\{\frac{N}{4}\sum_{a,b} \lambda^a \left(\hat{Q}_{a,b} + \gamma\tilde{\eta}\delta_{a,b}\right)^{-1} \lambda^b\right\}.
\end{aligned}$$

Integrating now over the $\{\lambda^a\}$ we obtain

$$\begin{aligned} Z_\gamma^n[x_{i,\tau}] &= \int [D\epsilon[Du][D\mu][DQ][D\hat{Q}] \exp\left\{-\gamma\sum_a (1-\beta)T\epsilon^a - \gamma\sum_{a,\tau} u_\tau^a\right\} \\ &\times \prod_\tau \exp\left\{-\frac{1}{2}\sum_{a,b}\left(u_\tau^a + \epsilon^a - \mu_\tau^a\right) Q_{a,b}^{-1}\left(u_\tau^b + \epsilon^b - \mu_\tau^b\right)\right\} \\ &\times \exp\left\{N\sum_{a,b}\hat{Q}_{a,b}Q_{a,b}\right\}\exp\left\{-\frac{T}{2}\mathrm{tr}\log Q\right\}\exp\left\{-\frac{nN}{2}\log 2\right\} \\ &\times \exp\left\{-\frac{N}{2}\mathrm{tr}\log(\hat{Q}+\gamma\tilde{\eta}\delta_{a,b})\right\}\exp\left\{-Nw^2\sum_{a,b}(\hat{Q}_{a,b}+\gamma\tilde{\eta}\delta_{a,b})\right\}. \end{aligned}$$

Introducing the variables $y_\tau^a = \mu_\tau^a - u_\tau^b$ and $z_\tau^a = \mu_\tau^a + u_\tau^b$ and integrating over the $\{z_\tau^a\}$ one is left with

$$\begin{aligned} Z_\gamma^n[x_{i,\tau}] &= \int [D\epsilon][DQ][D\hat{Q}][\\ &\times \exp\left\{-Nw^2\sum_{a,b}(\hat{Q}_{a,b}+\gamma\tilde{\eta}\delta_{a,b}) - \gamma(1-\beta)\sum_a T\epsilon^a + N\sum_{a,b}\hat{Q}_{a,b}Q_{a,b}\right\} \\ &\times \exp\left\{-Tn\log\gamma - \frac{T}{2}\mathrm{tr}\log Q - \frac{N}{2}\mathrm{tr}\log(\hat{Q}+\gamma\tilde{\eta}\delta_{a,b}) - \frac{nN}{2}\log 2\right\} \\ &\times \exp\left\{T\log Z_\gamma(\{\epsilon^a, Q\})\right\}, \end{aligned}$$

where we have defined

$$\begin{aligned} Z_\gamma(\{\epsilon^a, Q\}) &= \int\prod_a dy^a \exp\left\{-\frac{1}{2}\sum_{a,b}(y^a-\epsilon^a)Q_{a,b}^{-1}(y^b-\epsilon^b)\right\} \\ &\times \exp\left\{\gamma\sum_a y^a\theta(-y^a)\right\}. \end{aligned}$$

We now take the replica symmetric (RS) ansatz

$$Q_{a,b} = \begin{cases} q_1, & a=b \\ q_0, & a\neq b \end{cases} \tag{C.1.1}$$

$$\hat{Q}_{a,b} = \begin{cases} \hat{q}_1, & a=b \\ \hat{q}_0, & a\neq b. \end{cases} \tag{C.1.2}$$

and we define the susceptibility $\Delta q = q_1 - q_0$ as well as $\Delta\hat{q} = \hat{q}_1 - \hat{q}_0$. For Q^{-1} we then have

$$Q^{-1}_{a,b} = \begin{cases} (\Delta q - q_0)/(\Delta q)^2 + \mathcal{O}(n), & a = b \\ -q_0/(\Delta q)^2 + \mathcal{O}(n), & a \neq b \end{cases} \tag{C.1.3}$$

The effective partition function $Z_\gamma(\{\epsilon^a, Q\})$ reads

$$\begin{aligned} Z_\gamma(\{\epsilon^a, R^a, Q\}) &= \int \prod_a dx^a \exp\left\{-\frac{1}{2}\sum_{a,b} x^a Q^{-1}_{a,b} x^b\right\} \\ &\times \exp\left\{\gamma \sum_a (x^a + \epsilon^a)\theta(-x^a - \epsilon^a)\right\}, \end{aligned}$$

where we have defined $x^a = y^a - \epsilon^a$. By introducing a Gaussian variable s with measure $dP_{q_0}(s) = \frac{ds}{\sqrt{2\pi q_0}} e^{-s^2/2q_0}$ we obtain, in the limit $n \to 0$,

$$\frac{1}{n}\log(Z_\gamma(\{\epsilon^a, Q\}) = \frac{q_0}{2\Delta q} + \int dP_{q_0}(s) \log B_\gamma(s, \epsilon, \Delta q),$$

where

$$B_\gamma(s, \epsilon, \Delta q) = \int dx \exp\left\{-\frac{(x-s)^2}{2\Delta q} + \gamma(x+\epsilon)\theta(-x-\epsilon)\right\}.$$

If we also consider that

$$\operatorname{tr}\log Q = n(\log \Delta q + q0/\Delta q)$$

and

$$\operatorname{tr}\log(\hat{Q} + \gamma\tilde{\eta}\delta_{a,b}) = n(\log(\Delta\hat{q} + \gamma\tilde{\eta}) + \hat{q}_0/(\Delta\hat{q} + \gamma\tilde{\eta})),$$

we finally obtain the free energy

$$\begin{aligned} -\frac{\gamma F(q_0, \Delta q, \hat{q}_0, \Delta\hat{q}, \epsilon)}{nN} &= q_0\Delta\hat{q} + \hat{q}_0\Delta q + \Delta q\Delta\hat{q} - w^2\left(\Delta\hat{q} + \gamma\tilde{\eta}\right) - \gamma t(1-\beta)\epsilon \\ &- t\log\gamma + t\int dP_{q_0}(s)\log B_\gamma(\epsilon, s, R, \Delta q) - \frac{t}{2}\log\Delta q \\ &- \frac{\log 2}{2} - \frac{1}{2}\left(\log(\Delta\hat{q} + \gamma\tilde{\eta}) + \frac{\hat{q}_0}{\Delta\hat{q} + \gamma\tilde{\eta}}\right), \end{aligned}$$

where we have put $T = tN$. From the saddle point equations for $\hat{q}_0$ and $\Delta\hat{q}$ we get

$$\begin{aligned} \Delta\hat{q} + \gamma\tilde{\eta} &= \frac{1}{2\Delta q} \\ \hat{q}_0 &= \frac{w^2 - q_0}{2(\Delta q)^2}. \end{aligned}$$

Exploiting these relations the free energy becomes

$$\begin{aligned}-\gamma f(\epsilon, q_0, \Delta q) &= -\frac{\gamma F}{nN} = \frac{1}{2} - t\log\gamma - \gamma t(1-\beta)\epsilon \\ &+ t\int dP_{q_0}(s)\log B_\gamma(\epsilon, s, R, \Delta q) + \frac{1-t}{2}\log\Delta q + \frac{q_0 - w^2}{2\Delta q} - \gamma\tilde{\eta}q_0.\end{aligned}$$

Notice that

$$\Delta q = \frac{1}{2N}\sum_i (w_i^a - w_i^b)^2 \tag{C.1.4}$$

is the squared distance between two approximate solution of the optimization problem, drawn with a Gibbs measure with energy E. As $\gamma \to \infty$ the Gibbs measure gets more and more peaked on the optimal solution. If the latter is unique, we expect $\Delta q \to 0$. Indeed, given that the measure is nearly Gaussian, we expect $\Delta q \sim 1/\gamma$. Hence, in the large γ limit, it is natural to rescale $\Delta q = \Delta/\gamma$ keeping ϵ and q_0 independent of γ. In this limit we obtain the energy function

$$\begin{aligned}E &= t(1-\beta)\epsilon - \frac{q_0 - w^2}{2\Delta} - \int_{-\infty}^{-\Delta}\frac{dx}{\sqrt{2\pi q_0}}e^{-(x-\epsilon)^2/(2q_0)}\left(x + \frac{\Delta}{2}\right) \\ &+ \frac{t}{2\Delta}\int_{-\Delta}^{0}\frac{dx}{\sqrt{2\pi q_0}}e^{-(x-\epsilon)^2/(2q_0)}x^2.\end{aligned}$$

We now define $\tilde{x} = x/\Delta$, $\tilde{\epsilon} = \epsilon/\Delta$ and $\tilde{q}_0 = q_0/\Delta^2$. The reason for this change of variables is that we want to expose the singular behavior at the phase transition in terms of a single divergent quantity[1] Δ. Hence, we anticipate that $\tilde{\epsilon}$ and $\tilde{q}_0$ are going to attain finite values at the transition. In terms of the rescaled variables, we have

$$E(\tilde{\epsilon}, \tilde{q}_0, \Delta) = \frac{w^2}{2\Delta} + \Delta\left[t(1-\beta)\tilde{\epsilon} - \frac{\tilde{q}_0}{2} + \frac{t}{2\sqrt{\pi}}\int_{-\infty}^{\infty} ds e^{-s^2} g(\tilde{\epsilon} + s\sqrt{2\tilde{q}_0})\right] + \tilde{\eta}t\tilde{q}_0\Delta^2$$

where

$$g(x) = \begin{cases} 0, & x \geq 0 \\ x^2, & -1 \leq x \leq 0 \\ -2x - 1, & x < -1 \end{cases} \tag{C.1.5}$$

and $\tilde{q}_0$ and $\tilde{\epsilon}$ are the solutions of the saddle point equations

$$-1 + \frac{t}{\sqrt{2\pi\tilde{q}_0}}\int ds e^{-s^2} s g'(\tilde{\epsilon} + s\sqrt{2\tilde{q}_0}) + 2\tilde{\eta}\Delta = 0, \tag{C.1.6}$$

[1]It helps to note that Δ is the susceptibility.

$$1 - \beta + \frac{1}{2\sqrt{\pi}} \int dse^{-s^2} g'(\tilde{\epsilon} + s\sqrt{2\tilde{q}_0}) = 0, \tag{C.1.7}$$

$$-\frac{w^2}{2\Delta^2} + t(1-\beta)\tilde{\epsilon} - \frac{\tilde{q}_0}{2} + \frac{t}{2\sqrt{\pi}} \int dse^{-s^2} g(\tilde{\epsilon} + s\sqrt{2\tilde{q}_0}) + 2\tilde{\eta}\Delta\tilde{q}_0 = 0. \tag{C.1.8}$$

C.2 The Maximal Loss problem

We show here how to recover the correct $\beta \to 1$ limit leading to the Maximal Loss problem. The problem of finding the set of weights that minimizes the Maximal Loss (4.3.49) can be cast into that of finding the minimum of the cost function

$$E[u] = u \tag{C.2.9}$$

under the constraints

$$u + \sum_i w_i x_{i,t} \geq 0 \quad \forall t$$

and

$$\sum_i w_i = WN.$$

Let us show how this can be recovered starting from the general optimization problem under Expected Shortfall, where

$$\begin{aligned}
E[\epsilon, \{u_\tau\}] &= (1-\beta)T\epsilon + \sum_{\tau=1}^{T} u_\tau, && \text{(C.2.10)}\\
u_\tau &\geq 0 \ \ \forall \tau, && \text{(C.2.11)}\\
u_\tau + \epsilon + \sum_{i=1}^{N} x_{i,\tau} w_i &\geq 0 \ \ \forall \tau, && \text{(C.2.12)}\\
\sum_i w_i &= WN. && \text{(C.2.13)}
\end{aligned}$$

The first observation is that, for $\epsilon \geq ML$, (C.2.12) is satisfied for any set of $\{u_{i,\tau}\}$ satisfying (C.2.11). The minimum of the cost function can then be obtained by taking ϵ equal to the Maximal Loss and $u_{i,\tau} = 0 \ \forall i, \tau$. By comparing the resulting expression for (C.2.10) with (C.2.9), we can see that the two are equivalent if we keep $T(1-\beta) = 1$. If we now introduce the regularization, the cost function for the Maximal Loss problem reads

$$E[u, \{w_i\}] = u + \frac{T}{2C}\|\vec{w}\|^2.$$

As in section 4.3.1, an equivalent expression can be obtained by introducing the effect of the price impact. The two approaches are equivalent once we have taken

$$C = \frac{T}{2\eta}. \tag{C.2.14}$$

Notice that one can derive Eq. (C.2.14) also by taking $1-\beta = 1/T$ in (4.3.31), which is indeed the appropriate confidence level for maximal loss, because in a finite time window of T points, the worst possible outcome occurs with probability $1-\beta = 1/T$.

C.3 The replica calculation for the L_1 regularized Expected Shortfall

We present here the replica calculation for the L_1 regularized ES. We need to find the minimum of

$$E[\epsilon, \{u_\tau\}] = (1-\beta)T\epsilon + \sum_{\tau=1}^{T} u_\tau + \tilde{\eta}\|\vec{w}\|$$

under the constraints

$$u_\tau \geq 0,$$

$$u_\tau + \epsilon + \sum_{i=1}^{N} x_{i,\tau} w_i \geq 0$$

and

$$\sum_i w_i = WN.$$

The replicated partition function, corresponding to the partition function of n copies of the system can be computed as

$$\begin{aligned}
Z_\gamma^n[x_{i,\tau}] &= \int_{-\infty}^{\infty}\prod_{a=1}^{n} d\epsilon^a \int_0^\infty \prod_{\tau=1}^{T}\prod_{a=1}^{n} du_\tau^a \int_{-\infty}^{\infty}\prod_{i=1}^{N}\prod_{a=1}^{n} dw_i^a \int_{-\infty}^{\infty}\prod_{a=1}^{n} d\lambda^a \\
&\times \int_0^\infty \prod_{\tau=1}^{T}\prod_{a=1}^{n} d\mu_\tau^a \int_{-\infty}^{\infty}\prod_{\tau=1}^{T}\prod_{a=1}^{n} d\hat{\mu}_\tau \exp\left\{\sum_a \lambda^a(\sum_i w_i^a - WN)\right\} \\
&\times \prod_\tau \exp\left\{\sum_a i\hat{\mu}_\tau^a\left(u_\tau^a + \epsilon^a + \sum_i x_{i,\tau}w_i^a - \mu_\tau^a\right)\right\} \\
&\times \exp\left\{-\gamma\sum_a (1-\beta)T\epsilon^a - \gamma\sum_{a,\tau} u_\tau^a - \gamma\tilde{\eta}\sum_i |w_i^a|\right\}.
\end{aligned}$$

Averaging over the quenched variables $\{x_{i,\tau}\}$ and introducing the overlap matrix $Q_{a,b} = \frac{1}{N}\sum_i w_i^a w_i^b$ one obtains

$$\begin{aligned}
Z_\gamma^n[x_{i,\tau}] &= \int [D\epsilon][Du][Dw][D\lambda][D\mu][D\hat{\mu}][DQ][D\hat{Q}] \exp\left\{\sum_a \lambda^a(\sum_i w_i^a - WN)\right\} \\
&\times \prod_\tau \exp\left\{-\frac{1}{2}\sum_{a,b}\hat{\mu}_\tau^a Q_{a,b}\hat{\mu}_\tau^b\right\} \exp\left\{\sum_{a,b}\hat{Q}_{a,b}\left(NQ_{a,b} - \sum_i w_i^a w_i^b\right)\right\} \\
&\times \exp\left\{-\gamma\sum_a (1-\beta)T\epsilon^a - \gamma\sum_{a,\tau} u_\tau^a - \gamma\tilde{\eta}\sum_i |w_i^a|\right\} \\
&\times \prod_\tau \exp\left\{i\sum_a \hat{\mu}_\tau^a\left(u_\tau^a + \epsilon^a - \mu_\tau^a\right)\right\}.
\end{aligned}$$

We can now perform the Gaussian integral over the variables $\{\hat{\mu}_\tau^a\}$:

$$\begin{aligned} Z_\gamma^n[x_{i,\tau}] &= \int [D\epsilon][Du][Dw][D\lambda][D\mu][DQ][D\hat{Q}] \exp\left\{\sum_a \lambda^a (\sum_i w_i^a - WN)\right\} \\ &\times \exp\left\{-\gamma\sum_a (1-\beta)T\epsilon^a - \gamma\sum_{a,\tau} u_\tau^a - \gamma\tilde{\eta}\sum_i |w_i^a|\right\} \\ &\times \exp\left\{\sum_{a,b}\hat{Q}_{a,b}\left(NQ_{a,b} - \sum_i w_i^a w_i^b\right)\right\} \\ &\times \prod_\tau \exp\left\{-\frac{1}{2}\sum_{a,b}\left(u_\tau^a + \epsilon^a - \mu_\tau^a\right) Q_{a,b}^{-1}\left(u_\tau^b + \epsilon^b - \mu_\tau^b\right)\right\} \\ &\times \exp\left\{-\frac{T}{2}\mathrm{tr}\log Q\right\}. \end{aligned}$$

Introducing now the variables $y_\tau^a = \mu_\tau^a - u_\tau^b$ and $z_\tau^a = \mu_\tau^a + u_\tau^b$ and integrating over the $\{z_\tau^a\}$ we obtain

$$\begin{aligned} Z_\gamma^n[x_{i,\tau}] &= \int [D\epsilon][Du][Dw][D\lambda][DQ][D\hat{Q}] \exp\left\{\sum_a \lambda^a (\sum_i w_i^a - WN)\right\} \\ &\times \exp\left\{-\gamma\sum_a (1-\beta)T\epsilon^a - \gamma\sum_{a,\tau} u_\tau^a - \gamma\eta\sum_i |w_i^a|\right\} \\ &\times \exp\left\{\sum_{a,b}\hat{Q}_{a,b}\left(NQ_{a,b} - \sum_i w_i^a w_i^b\right)\right\} \\ &\times \prod_\tau \exp\left\{-\frac{1}{2}\sum_{a,b}\left(u_\tau^a + \epsilon^a - \mu_\tau^a\right) Q_{a,b}^{-1}\left(u_\tau^b + \epsilon^b - \mu_\tau^b\right)\right\} \\ &\times \exp\left\{-\frac{T}{2}\mathrm{tr}\log Q - TN\log\gamma + T\log Z_\gamma(\{\epsilon^a, Q\})\right\} \end{aligned}$$

where

$$\begin{aligned} Z_\gamma(\{\epsilon^a, Q\}) &= \int \prod_a dy^a \exp\left\{-\frac{1}{2}\sum_{a,b}(y^a - \epsilon^a) Q_{a,b}^{-1}(y^b - \epsilon^b)\right\} \\ &\times \exp\left\{\gamma\sum_a y^a \theta(-y^a)\right\}. \end{aligned}$$

In order to make further progress, let us consider the replica symmetric ansatz

$$Q_{a,b} = \begin{cases} q_1, & a = b \\ q_0, & a \neq b \end{cases} \tag{C.3.15}$$

$$\hat{Q}_{a,b} = \begin{cases} \hat{q}_1, & a = b \\ \hat{q}_0, & a \neq b. \end{cases} \tag{C.3.16}$$

and introduce the following rescaling relations

$$\begin{aligned} \Delta q &= q_1 - q_0 = \Delta/\gamma, && \text{(C.3.17)} \\ \hat{\Delta q} &= \hat{q}_1 - \hat{q}_0 = \hat{\Delta}\gamma, && \text{(C.3.18)} \\ \lambda^a &= \tilde{\lambda}^a/\gamma, && \text{(C.3.19)} \\ \hat{q}_0 &= \tilde{\hat{q}}_0/\gamma^2. && \text{(C.3.20)} \end{aligned}$$

The $\vec{w}$-dependent part of the partition function

$$\int [Dw] e^{-\gamma F_w} = \int [Dw] e^{\sum_{ia} \lambda^a w_i^a - \gamma\eta \sum_i |w_i^a| - \sum_{a,b} \hat{Q}_{a,b} \sum_i w_i^a w_i^b} \quad , \tag{C.3.21}$$

exploiting the identity $\log\langle X^n \rangle \simeq n \langle \log X \rangle$ valid for $n \to 0$, after some manipulations gives the following contribution to the free energy

$$F_w = \frac{nN}{\gamma} \Big\langle \log \int dw e^{-\gamma\left[\hat{\Delta} w^2 + \eta|w| - \tilde{\lambda} w - zw\sqrt{-2\tilde{\hat{q}}_0}\right]} \Big\rangle_z , \tag{C.3.22}$$

where we have denoted by $\langle \cdots \rangle_z$ averages over the normal variables z. Upon introducing also the new variables $\epsilon = \tilde{\epsilon}\Delta$ and $q_0 = \tilde{q}_0 \Delta^2$, after some algebra we eventually obtain the full free energy

$$\begin{aligned} \frac{F(\lambda, \tilde{\epsilon}, \tilde{q}_0, \Delta, \hat{q}_0, \hat{\Delta})}{Nn} &= -\lambda W - t(1-\beta)\epsilon + \Delta \hat{q}_0 + \Delta^2 \hat{\Delta} \tilde{q}_0 && \text{(C.3.23)} \\ &+ \frac{1}{\gamma} \Big\langle \log \int_{-\infty}^{\infty} dw e^{-\gamma V(w,z)} \Big\rangle_z + \frac{t\Delta}{2\sqrt{\pi}} \int_{-\infty}^{\infty} ds e^{-s^2} g(\tilde{\epsilon} + s\sqrt{2\tilde{q}_0}), \end{aligned}$$

where

$$V(w, z) = \hat{\Delta} w^2 + \eta|w| - \lambda w - zw\sqrt{-2\hat{q}_0} \tag{C.3.24}$$

and

$$g(x) = \begin{cases} 0, & x \geq 0 \\ x^2, & -1 \leq x \leq 0 \\ -2x - 1, & x < -1 \end{cases} . \tag{C.3.25}$$

References

C. Acerbi. Spectral Measures of Risk: a Coherent Representation of Subjective Risk Aversion. *Journal of Banking and Finance*, 26(7):1505–1518, 2002. 58

C. Acerbi. Coherent representations of subjective risk-aversion. In G. Szegö, editor, *Risk Measures for the 21st Century.* John Wiley and Sons., 2004. 58

C Acerbi and G Scandolo. Liquidity risk theory and coherent measures of risk, 2007. Available at SSRN: http://ssrn.com/abstract=1048322. 52

C. Acerbi and D. Tasche. On the Coherence of Expected Shortfall. *Journal of Banking and Finance*, 26(7):1487–1503, 2002. 58

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent Measures of Risk. *Mathematical Finance*, 9:203–228, 1999. 58, 63

R. R. Bailey. *The Economics of Financial Markets.* Cambridge University Press: Cambrdige, 2005. 2, 4, 5, 8, 9

J. Berg, M. Marsili, A. Rustichini, and R. Zecchina. Statistical mechanics of asset markets with private information. *Quantitative Finance*, 1(2):203–211, 2001. 3, 7, 11, 18, 19, 21, 22, 28

G. Bianconi, T. Galla, M. Marsili, and P. Pin. Effects of tobin taxes in minority game markets. *Journal of Economics Behavior and Organization*, 70:231–240, 2009. 49

J.-P. Bouchaud. Economics need a scientific revolution. *Nature*, 455:1181, 2008. 3

J.-P. Bouchaud and M. Wyart. Self-referential behaviour, overreaction and conventions in financial markets. *Journal of Economic Behavior and Organization*, 63:1–24, 2007. 44

J.-P. Bouchaud, J. D. Farmer, and Lillo F. How markets slowly digest changes in supply and demand, 2008. Available at arXiv:0809.0822. 1, 35

J.-Ph. Bouchaud and M. Potters. *Theory of Financial Risk - From Statistical Physics to Risk Management*. Cambridge University Press, Cambridge, UK, 2000. 1, 57

W. A. Brock, C.H. Hommes, and F.O.O. Wagener. More hedging instruments may destabilize markets, 2008. CeNDEF Working paper 08-04 University of Amsterdam. 4

E. E. Buffet. Berkshire hathaway annual report 2002, 2002. 29

Z. Burda, J. Jurkiewicz, and M. A. Nowak. Is econophysics a solid science? *Acta Physica Polonica*, B 34:87–132, 2003. 1

F. Caccioli and M. Marsili. Information efficiency and financial stability. *Economics: The Open-Access, Open-Assessment E-Journal*, 4:2010–20, 2010. 3, 4, 8

F. Caccioli, M. Marsili, and P. Vivo. Eroding market stability by proliferation of financial instruments. *European Physical Journal B*, 71(4):467–479, 2009. 4, 85

F. Caccioli, S. Still, M. Marsili, and I. Kondor. Optimal liquidation strategies regularize portfolio selection, 2010. Available on arXiv:1004.4169. 5, 51

L. Cardy. *Scaling and Renormalization in Statistical Physics*. Cambridge University Press, 1996. 1

T. Castellani and A. Cavagna. Spin glass theory for pedestrians. *Journal of Statistical Mechanics*, page P05012, 2005. 17

A. Chakraborti, I. M. Toke, M. Patriarca, and F. Abergel. Econophysics: Empirical facts and agent-based models, 2009. available at arXiv:0909.1974. 1

D. Challet and M. Marsili. Criticality and market efficiency in a simple realistic model of the stock market. *Physical Review E*, 68:036132, 2003. 41

D. Challet, M. Marsili, and Y.-C. Zang. *Minority games: interacting agents in financial markets*. Oxford University Press, 2005. 1, 2, 3

S. Ciliberti, I. Kondor, and M. Mezard. On the feasibility of portfolio optimization under expected shortfall. *Quantitative Finance*, 7:389–396, 2007. 5, 51, 61, 63, 68, 69, 70, 75

Rainer Dietrich, Manfred Opper, and Haim Sompolinsky. Statistical mechanics of support vector networks. *Phys. Rev. Lett.*, 82:2975–2978, 1999. 69

Z. Eisler, J.-P. Bouchaud, and J. Kockelkoren. The price impact of order book events: market orders, limit orders and cancellations. 2009. Available on arXiv: 0904.0900. 5, 71

E. J. Elton and M. J. Gruber. *Modern Portfolio Theory and Investment Analysis*. Wiley, New York, 1995. 53

E. F. Fama. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2):383–417, 1970. 13

Int. Monetary Fund. Global financial stability report (april 2008), 2008a. 29

Int. Monetary Fund. Global financial stability report (october 2008), 2008b. 29

J. K. Galbraith. *The Great Crash 1929*. Boston: Houghton Mifflin Company, 1954. 7

D. Goldbaum. Self-organization and the persistence of noise in financial markets. *Journal of Economic Dynamics and Control*, 30(9–10):1837–1855, 2006. 28

S. J. Grossman and J. E. Stiglitz. On the impossibility of informationally efficient markets. *American Economic Review*, 70(3):393–408, 1980. 19

M. U. Haq, I. Kaul, and I. Grunberg. *The Tobin Tax: Coping with Financial Volatility*. Oxford: Oxford University Press., 1996. 28, 44

C. H. Hommes. Heterogeneous agent models in economics and finance. In *Handbook of Computational Economics 2: Agent–Based Computational Economics*. Amsterdam: Elsevier, 2006. 8, 9, 10, 27, 28, 81

P. Jorion. *VaR: The New Benchmark for Managing Financial Risk*. McGraw-Hill, New York, 2000. 58

I. Kondor and I. Varga-Haszonits. Divergent estimation error in portfolio optimization and in linear regression. *Eur. Phys. J.*, B64:601–605, 2008a. 51

I. Kondor and I. Varga-Haszonits. Feasibility of portfolio optimization under coherent risk measures, 2008b. submitted to Quantitative Finance. 5, 51, 63, 64

I. Kondor, S. Pafka, and G. Nagy. Noise sensitivity of portfolio selection under various risk measures. *Journal of Banking and Finance*, 31:1545–1573, 2007. 51, 53, 54, 57, 59, 61

L. Laloux, P. Cizeau, J.-Ph. Bouchaud, and M. Potters. Noise Dressing of Financial Correlation Matrices. *Phys. Rev. Lett.*, 83:1467–1470, 1999. 5

A. W. Lo. Efficient markets hypothesis. In *The New Palgrave: a Dictionary of Economics*. New York: Palgrave MacMillan, 2007. 9

T. Lux and M. Marchesi. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 387:498–500, 1999. 8, 10

B. Malkiel. Efficient market hypothesis. In *New Palgrave Dictionary of Money and Finance*. London: MacMillan, 1992. 8, 14

Dörthe Malzahn and Manfred Opper. A statistical physics approach for the analysis of machine learning algorithms on real data. *Journal of Statistical Mechanics (JSTAT)*, 2005. 69

R. N. Mantegna and H. E. Stanley. *An introduction to Econophysics: correlations and complexity in finance*. Cambridge University Press., 2000. 1, 9

H. Markowitz. Portfolio selection. *Journal of Finance*, 7:77–91, 1952. 4, 52, 66

H. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. J. Wiley and Sons, New York, 1959. 4, 52

M. Marsili. Complexity and financial stability in a large random economy. SSRN Working Paper, Abdus Salam International Centre Theoretical Physics, Italy, 2009. 4, 81

M. Marsili and A. De Martino. Statistical mechanics of socio-economic systems with heterogeneous agents. *Journal of Physics A*, 39:R465–R540, 2006. 1, 2, 3, 44, 50

R. C. Merton and Z. Bodie. Design of financial systems: Towards a synthesis of function and structure. *Journal of Investment Management*, 3(1):1–23, 2005. 3, 29

M. Mezard, G. Parisi, and M. A. Virasoro. *Spin Glass Theory and Beyond*. Singapore: World Scientific., 1987. 17

H. P. Minsky. The financial instability hypothesis. The Jerome Levy Economics Institute Working Paper No. 74. Levy Economics Institute, New York, 1992. 10

F. S. Mishkin. Understanding financial crises: A developing country perspective. NBER Working Papers No. 5600. National Bureau of Economic Research, Cambridge, MA, 1996. 2

F. S. Mishkin and T. T. Herbertsson. Financial stability in iceland. Iceland Chamber of Commerce report, May 2006, Iceland Chamber of Commerce, Reykjavíc, 2006. 2

Manfred Opper. Statistical mechanics of learning: Generalization. In *The Handbook of Brain Theory and Neural Networks, (1st edition)*, pages 922–925. MIT Press, 1995. 69

Manfred Opper and David Haussler. Bounds for predictive errors in the statistical mechanics of supervised learning. *Phys. Rev. Lett.*, 75:3772–3775, 1995. 69

S. Pafka and I. Kondor. Noisy covariance matrices and portfolio optimization. *Eur. Phys. J.*, B 27:277–280, 2002. 54

S. Pafka and I. Kondor. Estimated correlation matrices and portfolio optimization. *Physica*, A 343:623–634, 2004. 5

S. R. Pliska. *Introduction to Mathematical Finance: Discrete Time Models*. Oxford: Blackwell., 1997. 2, 3, 21, 29, 33, 35

R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–41, 2000. 58

L. Shapley and M. Shubik. Trade using one commodity as a means of payment. *Journal of Political Economy*, 85(5):937–968, 1977. 21

R. J. Shiller. *Irrational Exuberance*. Princeton University Press, 2000. 7

R. J. Shiller. *The Subprime Solution: How Today's Global Financial Crisis Happened, and What to Do about It*. Princeton University Press, 2008. 2, 7

R. J. Shiller and G. A. Akerlof. *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism*. Princeton University Press, 2009. 2, 7

S. Still and I. Kondor. Regularizing portfolio optimization, 2009. To appear on New Journal of Physics. Available on arXiv: 0911.1694. 51, 64, 65

S. M. Turnbull, M. Crouhy, and R. A. Jarrow. The subprime credit crisis of 07. SSRN Working Paper, Samuel Curtis Johnson Graduate School of Management, Cornell University, 2008. 2

V. Vapnik. *The Nature of Statistical Learning Theory*. Springer Verlag, New York, 1995. 65

I. Varga-Haszonits and I. Kondor. The instability of downside risk measures. *J. Stat. Mech.*, P12007, 2008. 51