DISS. ETH NO. ____________

Disentangling Financial Markets and Social Networks: Models and Empirical Tests

A dissertation submitted to

ETH ZURICH

for the degree of
Doctor of Sciences

presented by

Qunzhi Zhang
(张群智)

Msc. ETH in Management,
Technology and Economics, ETH Zurich

born on Sept. 2, 1974

citizen of China

accepted on the recommendation of

Prof. Dr. Didier Sornette, examiner
Prof. Dr. Thorsten Hens, co-examiner

2013

Acknowledgements

Doing the Ph.D. is a wonderful and challenging experience. I have received a lot of help from many people in the last three and a half years. First and foremost, I would like to express my sincerest gratitude to my supervisor Prof. Dr. Didier Sornette. He taught me not only knowledge but also the way of scientific thinking. He guided my studies with my Ph.D. projects, giving me new ideas and insights. By learning and doing after him, I successfully converted myself from a software engineer to a scientist with a multi-discipline portfolio. In addition to science, he also taught us ways to make our work more efficient and make us stronger both mentally and physically, such as good diet habits, physical practice, etc. I benefited a lot from his guidance.

Moreover, I am deeply grateful to Prof. Dr. Jeffrey Satinover, who guided me in the reverse-engineering project, and Prof. Dr. Thorsten Hens, who accepted to be my co-examiner and gave me a lot of guidance in my Ph.D. thesis. I also want to thank Prof. Dr. Theo Wehner for being the assessor of my doctoral examination.

My Ph.D. study would have been mission impossible without the support of my wife Qi Zhu. In the last years, she did her best to take care of me, my elder son, and our new son who was born in 2011. Especially after our second son was born, she has been undertaking most of the sleepless nights alone taking care of the baby. She should be credited with at least half of what I have achieved in my work.

I also want to thank Heidi Demuth for giving me a lot of help during my Ph.D. study. I am grateful to Dr. Ryan Woodard for always helping me like a big brother. I thank Dr. Thomas Maillart for all the helpful discussions and great ideas, and the exciting projects we have been working on. And I am very grateful that Hyun-U Sohn helped me translate the abstract into German carefully and promptly, and also gave me a bunch of useful suggestions to polish the English in my thesis.

Last but not least, I thank the members (current and former) of the Chair of Entrepreneurial Risks for their help: Dr. Mika Kastenholz, Dr. Peter Cauwels, Dr. Dorsa Sanadgol, Dr. Georges Harras, Dr. Andreas Hüsler, Prof. Dr. Alexander Saichev, Prof. Dr. Yannick Malevergne, Prof. Dr. Vyacheslav Yukalov, Shengsui Hu, Dr. Wanfeng Yan, Dr. Moritz Hetzer, Dr. Vladimir Filimonov, Dr. Riley Crane, Dr. Hilary Woodard, Maroussia Favre, Ryohei Hisano, Susanne von der Becke, Zalàn Forrò, Tatyana Kovalenko, Yura

Yoshihiro, Dr. Monika Gisler, Ya-ming Wang, Hao Meng, Prof. Dr. Rong Lan, Dr. Ching-Yi Tsai, Prof. Dr. Yong-hong Jin, Meng Li, Dr. Zhi-Qiang Jiang, Dr. Li Lin, Dr. Xiao-Hui Ni, Dr. Maxim Fedorovsky, Stanislav Gora, Amine Issami, et cetera.

Abstract

In this thesis we include papers using an interdisciplinary approach to study the complex financial systems and social networks. The papers are linked by an identical goal: to better understand the financial systems and social networks, and develop and test news models to predict statistically the future asset prices in the financial systems and predict the emergent phenomena in the social networks.

In the "Can media moods predict stock prices during and after the 2008 financial crisis?" paper we extract media moods from the Reuters US news using computational linguistics methods and study the relationships between these media moods and the S&P500 returns from January 1, 2007 to June 6, 2012. We report three major findings. First, negative moods Granger cause S&P500 returns with a negative coefficient. Second, the S&P500 returns Granger cause the negative moods also with a negative coefficient, showing the existence of a positive feedback loop between them. Third, we find that trading strategies based on media moods can generate both statistically significant and economically significant returns, and the extra returns cannot be explained by the Fama-French factors. The corresponding extracted α's (excess risk-adjusted returns unexplained by the Fama-French factors) are impressively high, in the range $0.1 - 0.2\%$ per day and thus dominate typical transaction costs and implementation slippage. This suggests that financial markets are not informationally efficient over the studied time period and that this results from the existence of the mutually reinforcing feedbacks between negative moods and negative S&P500 returns, which have been and are still present over this time period characterized by a very serious financial crisis and its ongoing development.

In the "Reverse engineering stock markets with mixed games and alpha generation" paper we construct virtual financial markets populated by artificial agents, who make decisions according four classes of backward-looking decision functions, with the goal of testing the weak form of the efficient market hypothesis (EMH). Our agent-based models (ABM) are populated by agents with bounded rationality and heterogeneous beliefs, which can be represented by the decision functions defining respectively the minority game, the majority game, the $-game and the delayed minority game. We extend a previous methodology and provide the main structural parameters, the specific trading strategies used by the agents, as well as the fractions of agents playing the four different games. This genuine reverse-engineered reconstruction of the real financial markets is applied to the 10-year time series of the S&P500, Dow Jones Industrial Average and Nasdaq 100 indexes from 1982 to 2012 in 700 experiments associated with different time windows. Our empirical results provide evidence that our ABM's can describe the behavior of a large proportion of investors in a real market. This is supported by (i) our finding that 654 out of 700 reverse engineering experiments with on three main U.S indexes predict the future return signs with statistically significant success rates, (ii) trades based on these predictions can generate statistically and economically significant returns, and (iii) there are statistically significant relations between market regimes and the corresponding parameters of reverse-engineered ABM's.

In the "Empirical test of the origin of Zipf's law in growing social networks" paper we report a detailed analysis of a burgeoning network of social groups, in which all ingredients needed for Zipf's law to apply are verifiable and verified. A recently developed theory predicts that Zipf's law corresponds to systems that are growing according to a maximally sustainable path in the presence of random proportional growth, stochastic birth and death processes. We estimate empirically the average growth r and its standard deviation σ as well as the death rate h and predict without adjustable parameters the exponent μ of the power law distribution $P(s)$ of the group sizes s. Using numerical simulations of the underlying growth model, we demonstrate that the empirical stability of Zipf's law over the whole lifetime of the social network can be attributed to the interplay between a finite lifetime effect and a large σ value. Our analysis and the corresponding results demonstrate that Zipf's law can be observed with a good precision even when the balanced growth condition is not realized, if the random proportional growth has a strong stochastic component and is acting on young systems under development.

All these results show that our approach is able to disentangle the financial systems and social networks from the complexity in terms of both understanding the underlying mechanisms of the systems and predicting them. We shall carry on this approach in the future with more theoretical and practical problems.

Zusammenfassung

Diese Arbeit umfasst wissenschaftliche Artikel, die einen interdisziplinären Ansatz zum Studium von komplexen Finanzsystemen und sozialen Netzwerken verwenden. Ein roter Faden durchzieht und verbindet alle Artikel miteinander: Es werden Modelle zur Medienanalyse entwickelt und getestet, sowohl um die Entwicklung von Anlagepreisen statistisch vorherzusagen als auch um emergente Phänomene in sozialen Netzwerken zu antizipieren.

In dem Artikel „Can media moods predict stock prices during and after the 2008 financial crisis?“ extrahieren wir die Stimmung in den Medien, repräsentiert durch Reuters US News, mithilfe von Methoden der rechnergestützen Linguistik und untersuchen die Beziehung zwischen der Stimmung in den Medien und den S&P 500 Renditen zwischen 1. Januar 2007 und 6. Juni 2012. Wir stellen dabei drei wesentliche Resultate fest. Erstens zeigen Tests, dass eine signifikante Granger-Kausalität zwischen negativen Stimmungen und S&P 500 Renditen besteht, mit einem negativen Koeffizienten. Zweitens zeigen Tests, dass eine signifikante Granger-Kausalität zwischen S&P 500 Renditen und negativen Meinungen besteht, ebenfalls mit einem negativen Koeffizienten. Zusammengenommen deutet dies auf die Existenz einer positiven Rückkopplung zwischen beiden Grössen hin. Drittens können Handelsstrategien, die auf den festgehaltenen Medienstimmungen basieren, statistisch wie ökonomisch signifikante Renditen generieren, die dabei nicht auf die Fama-French-Faktoren zurückzuführen sind. Die Alphas (Überschussrenditen, die nicht durch die Fama-French-Faktoren erklärt werden) sind dabei mit 0.1%–0.2% pro Tag beeindruckend hoch und dominieren typische Transaktionskosten und Implementierungsschlupf. Wir schliessen daraus, dass die Finanzmärkte über die betrachtete Zeitperiode, die von einer ernsten Finanzkrise gezeichnet war, nicht informationseffizient waren und dass der Grund hierfür in Rückkopplungsmechanismen zwischen negativer Stimmung und negativen S&P 500 Renditen zu suchen ist.

In dem Artikel „Reverse engineering stock markets with mixed games and alpha generation“ konstruieren wir virtuelle Finanzmärkte, die von künstlichen Agenten bevölkert sind, welche ihre Anlage- und Handelsentscheidungen anhand von vier Klassen rückwärtsblickender Entscheidungsfunktionen treffen, mit dem Ziel, die schwache Form der Hypothese effizienter Märkte zu testen. Unsere agentenbasierten Modelle (ABMs) sind bevölkert von Agenten beschränkter Rationalität und mit heterogenen Meinungen; ihre Zielfunktionen bilden respektive das Minority Game, das Majority Game, das $-Game und das verzögerte Minority Game ab. Wir erweitern eine frühere Methodologie und berechnen die wesentlichen strukturellen Parameter, die spezifischen Handelsstrategien, die von den Agenten benutzt werden, sowie den Anteil der Agenten, die die vier Spiele spielen. Diese rückentwickelte (reverse-engineering) Rekonstruktion realer Finanzmärkte wird auf 10-Jahres-Zeitreihen vom S&P500, Dow Jones Industrial Average and Nasdaq 100 zwischen 1982 und 2012 in 700 Experimenten mit verschieden positionierten Zeitfenstern angewendet. Die Ergebnisse zeigen, dass unsere ABMs das Verhalten eines grossen Anteils von Anlegern in einem echten Markt beschreiben können. Wir zeigen, dass (1) 654 von 700 Rückentwicklungsexperimenten auf den drei grossen US Indizes zukünftige Renditen mit statistisch signifikantem Erfolg vorhersagen können, (2) Transaktionen basierend auf diesen Prognosen statistisch wie ökonomisch signifikante Renditen generieren und (3) statistisch signifikante Beziehungen zwischen Marktregimen und den zugehörigen Parameterwerten der rückentwickelten ABMs bestehen.

In dem Artikel „Empirical test of the origin of Zipf's law in growing social networks“ berichten wir über eine detaillierte Analyse eines wachsenden Netzwerks sozialer Gruppen, in dem alle Komponenten enthalten sind, die es braucht, um das Zipf-Gesetz zum Tragen zu bringen. Eine kürzlich entwickelte Theorie sagt voraus, dass das Zipf-Gesetz Systeme kennzeichnet, die angesichts stochastischer Wachstums-, Geburten- und Sterberaten auf dem maximal nachhaltigen Pfad wachsen. Wir schätzen empirisch die durchschnittliche Wachstumsrate r und deren Standardabweichung σ sowie die Sterberate h und sagen ohne verstellbare Parameter den Exponenten μ des Potenzgesetzes $P(s)$ über den Gruppengrössen s voraus. Mithilfe numerischer Simulationen des zugrundeliegenden Wachstumsmodells demonstrieren wir, dass die empirische Stabilität des Zipf-Gesetzes über die gesamte Lebenszeit des sozialen Netzwerks auf das Zusammenspiel zwischen einem Effekt endlicher Lebenszeit und einem grossen σ-Wert zurückzuführen ist. Unsere Analyse und die zugehörigen Resultate demonstrieren, dass das Zipf-Gesetz selbst dann beobachtet werden kann, wenn die Bedingung balancierten Wachstums nicht gegeben ist, solange die stochastischen Wachstumsraten eine starke Zufallskomponente haben und auf ein junges System, das sich noch in Entwicklung befindet, Anwendung finden.

All diese Ergebnisse zeigen, dass unser Ansatz es vermag, die Komplexität von Finanzsystemen und sozialen Netzwerken herunterzubrechen, sowohl was das Verständnis des jeweils zugrundeliegenden Mechanismus angeht sowie was dessen Vorhersage betrifft. Wir werden diesen Ansatz in der Zukunft an weiteren Problemstellungen aus der Theorie wie Praxis weiterverfolgen.

Contents

Acknowledgements **i**

Abstract **iii**

Zusammenfassung **v**

1 Introduction **1**
1.1 Goals and Methodology 3
1.2 Abstracts of scientific papers 6
1.3 Contributions of the Ph.D. candidate to the papers 8

2 Can media moods predict stock prices during and after the 2008 financial crisis? **9**
2.1 Introduction 9
2.2 Previous works 12
2.3 Data 15
2.4 Data analysis methods and empirical results 18
2.5 Conclusions 29

3 Reverse engineering stock markets with mixed games and alpha generation **37**
3.1 Introduction 37
3.2 The ABM's 39
3.3 Calibrating ABM's and empirical results 45
3.4 Conclusions 60

4 Empirical test of the origin of Zipf's law in growing social networks **63**
4.1 Introduction 63
4.2 Theory and data 64
4.3 Data analysis and results 67
4.4 Conclusions 73

5 Conclusions **77**

List of Tables **79**

List of Figures **85**

Bibliography 87

1

Introduction

We have just witnessed the devastating power of the global financial crisis in 2008 (Shiller, 2008; Kolb, 2011) whose aftershocks are still shaking the world economy; the European sovereign debt crisis (Wonders, 2010), possibly triggered by it, followed only about one year later and as of this writing (Jan. 2013), European countries are still stuck in it. Preceding these two most recent events, there have been a number of financial and economic crises in the last 100 years (Kobrak and Wilkins, 2012; Galbraith, 1994; Kindleberger and Aliber, 2011), including the Great Depression (1929-1939) (Bernanke, 2004; Eichengreen and Temin, 2000; Rothbard, 2000; Robbins, 2009). And not only in the financial and economic domains. Crises also happened in the social domain: there were the great world wars in the last century, and there are still wars going on right now as well as political and geopolitical tensions in many areas of the world. A recent example of a critical transition in society is the "Arab Spring" revolution, a big surprise triggered by a rather small event (Anderson, 2011; Gelvin, 2012).

Crises generally surprise people as they usually strike without explicit warning signs. For instance, during the early phase of the global financial crisis, forecasters were predicting only a mild recession (Mishkin, 2011). Another example is that "stock market crashes are often unforeseen for most people, especially economists" (Sornette, 2003). Indeed, forecasting financial and economic crises is so difficult that economists are rethinking their theories and models in sometimes fundamental ways. Rogoff (2010) provides a good perspective on the development of economic theory. The difficulty of forecasting crises comes from the fact that both financial systems and social networks are complex: They are comprised of many individuals that interact, and due to this interaction the aggregate cannot be treated as simply the sum of its components – "more is different" (Anderson, 1972). For instance, symmetry of behaviour of individual investors who trade a stock may lead to a random walk of the stock's returns, but the symmetry could be broken by investors mimicking each other, leading to new phenomena such as bubbles (if the majority of the investors buy) and crises (if the majority of the investors sell) would emerge – phenomena which are unlikely to happen under the random walk assumption that the stock market is informationally efficient. In complex systems, there are many "new" features, such as positive feedback loops, non-linearity, power laws, critical states, phase transitions, and so on, which cannot be treated in a linear way. We need new theories and models to forecast crises in financial, economic, and social systems, monitor systemic risks, and possibly even prevent future crises from happening. This motivates us to apply an interdisciplinary approach with concepts and tools developed in diverse fields, including financial economics, statistical physics, and computer science.

In this thesis we focus on several concrete and important problems, so it includes papers on different but closely linked subjects. In the paper "Can media moods predict stock prices during and after the 2008 financial crisis?" we study if the stock market is predictable in terms of publicly available information. We find that in the period we study, negative news and negative returns reinforce each other, and that one can make profits by trading on news information. The results thus challenge the semi-strong form of the Efficient Market Hypothesis (EMH) (Fama, 1970). In the paper "Reverse engineering stock markets with mixed games and alpha generation", we use agent-based models (ABMs) to study the underlying mechanism of stock markets and use the resulting configuration to predict future stock prices. Tests show that our reverse-engineered ABMs can predict the sign of future returns with statistically significant success rates, and that one can trade profitably with these ABM-based strategies. We also find a relationship between the parameters of our reverse-engineered ABMs and historical regimes of the U.S. stock market. The results challenge the weak form of EMH and show that stock markets are predictable based on historical price information. In these two papers we study the informational efficiency of the stock markets and develop prediction tools from different points of view. But the strategy we apply are the same. We incorporate interdisciplinary concepts, methods and tools into our studies to disentangle systems from their complexity, by understanding the underlying mechanisms and predicting them. In the third paper, "Empirical test of the origin of Zipf's law in growing social networks", we apply the same strategy to complex social networks. As the model (Malevergne et al., 2010) predicts, Zipf's law of group sizes on a website emerges from gradients including i) the proportional growth of groups and ii) the birth of new groups as well as the death of existing ones.

The results thus show that we can predict phenomena in complex social networks if we understand the underlying interactions and have enough data.

To sum up, the papers in this thesis are linked by one identical goal: to develop and test new models and methods in order to understand complex financial and social systems and predict future asset prices and emergent phenomena. This thesis is organized as follows. The first chapter is the concise introduction of different papers, including their goals and methodology. The second chapter is the paper “Can media moods predict stock prices during and after the 2008 financial crisis?”. The third chapter is the paper “Reverse engineering stock markets with mixed games and alpha generation”. The fourth chapter is the paper “Empirical test of the origin of Zipf’s law in growing social networks”. The last chapter concludes.

1.1 Goals and Methodology

1.1.1 Can media moods predict stock prices during and after the 2008 financial crisis?

The efficient market hypothesis (EMH) (Fama, 1970) in finance theory asserts that financial markets are informationally efficient. In the spirit of the EMH, an “efficient markets model” commonly used by economists and market analysts to value stocks states that “real stock prices equal the present value of rationally expected or optimally forecasted future real dividends discounted by a constant real discount rate” or by a variable but stable real discount rate (Shiller, 1981). For stock valuation, therefore, it is very important to test the EMH.

There exists a rich literature using event study methods to test the EMH (Eckbo, 2007). The event study methods test whether stock returns change significantly when there are exogenous shocks in the form of news. Most researchers use daily stock prices and check abnormal returns only in 2 or 3 days around some peculiar class of news impacts, and their results tend to support the EMH (Antweiler and Frank, 2006). Nevertheless, increasing evidence in the behavioural finance literatures challenges the EMH. Subrahmanyam (2008) provides a good review and synthesis of this literature.

Our goal in this study is to test if a piece of new information about stock markets will be quickly incorporated into prices so that no one can profit from it, because the traditional opinion is that any arbitrage opportunity embedded in the information will disappear immediately if many investors try to exploit it. Our hypothesis, however, is that the arbitrage opportunity may not always disappear, because there could be a positive feedback loop in the relationships between news and stock prices: when bad news pushes stock prices down, the markets may not go to a new equilibrium state as the traditional point of view predicts, but further deviate from the equilibrium state as the sell actions cause more bad news in turn. We would especially like to test this hypothesis during a financial crisis. Furthermore, we would like to test if we can predict future stock prices based on news.

We use Reuters daily news downloaded from the Reuters US website Reuters.com (2012) between 2007 and 2012 to do our analysis. There are thousands of news stories everyday, mainly business and financial news, as well as breaking US and international news. It can be seen as a complete library of important events happening in the US and the world, and thus it is an ideal resource to study relationships between news and US stock prices. We use natural language processing (NLP) methods to perform a sentiment analysis in a coarse grained manner: we classify daily news articles as positive, neutral, or negative, and calculate the respective fractions of positive, neutral, and negative news for any given trading day. We then use econometric methods to study the linear relations between sentiments and US stock index returns, and we also construct (non-linear) trading strategies based on news sentiments to test the predictability of sentiments on US stock indexes returns. For stock prices data, we use daily stock prices, as downloaded from Yahoo Finance Yahoo.com (2012).

1.1.2 Reverse engineering stock markets with mixed games and alpha generation

The goal of reverse engineering stock markets is to understand the mechanism underlying them and predict future price changes based on it. To this end, we apply ABMs, which are well suited for describing the interactions between bounded rational agents. This bottom-up approach is based on two ideas. The first is to use ABMs to capture the decision-making processes of investors at the micro-level and aggregate the collective behaviour to price series at the macro level; the second is to search for configurations of ABMs that generate the best fit to real price series and then use these ABMs to predict future price changes.

In other words, we calibrate virtual stock markets by reverse-engineering historical stock price movements. A virtual stock market is comprised of N agents, each trading in the virtual stock market in a finite time span. During each period, normally a trading day, the agents decide to buy or sell. The agents make decisions based on historical information. Each agent has a limited memory length m and a limited number s of trading strategies to predict future price changes from an m-day price history. The agents assess the success rate of their trading strategies in the past T days. If the success rate of an agent's best trading strategy (the trading strategy with the highest success rate) is lower than a threshold τ, the agent will not be confident enough to trade. Here, the success rate means the percentage of correct price change directions a trading strategy would predict.

The agents interact with each other by playing four kinds of games: the minority game (Challet et al., 2005), the majority game (Challet et al., 2005), the \$-game (Andersen and Sornette, 2003), and the delayed minority game (Wiesinger et al., 2012). On each trading day, an agent checks her best trading strategy and decides to buy, sell, or do nothing if she is not confident enough; the collective buy and sell actions then change the stock price on that day; at last, the agents update the performance of their trading strategies by comparing the predicted direction of the price movement with the realized one. In this manner, the agents will generate an artificial price time series.

To calibrate our artificial stock markets, we use a genetic algorithm (GA) to let the parameters of our artificial stock market evolve and find the configuration which generates the best match between the artificial and the real time series. We thus get a calibrated ABM with optimal parameters in terms of reproducing the real time series. This configuration will then be used to predict future stock price movements and market regimes.

We use two methods to test the predictive power of the ABMs. First, we use them to predict the signs of future returns, and we prove that the success rates of these predictions are statistically significant, compared to random strategies. Second, we use ABM-based strategies to trade, and show that they can generate significantly positive abnormal returns using various statistical tests.

1.1.3 Empirical test of the origin of Zipf's law in growing social networks

Power law distributions (equation (4.1)) are ubiquitous characteristics of many natural and social systems. The function $p(s)$ is the density associated with the probability $P(s) = \Pr\{S > s\}$ that the value S of some stochastic variable, usually a size or frequency, is greater than s. Among power law distributions, Zipf's law states that $\mu = 1$, i.e., $P(s) \sim s^{-1}$ for large s. Zipf's law has been reported for many systems (Saichev et al., 2009), including word frequencies (Zipf, 1949), firm sizes (Axtell, 2001), city sizes (Gabaix, 1999), connections between Web pages (Kong et al., 2008) and between open source software packages (Maillart et al., 2008), Internet traffic characteristics (Adamic and Huberman, 2000), abundance of expressed genes in yeast, nematodes and human tissues (Furusawa and Kaneko, 2003) and so on. The apparent ubiquity and universality of Zipf's law has triggered numerous efforts to explain its validity. It is also essential to understand the origin(s) of Zipf's law.

We use a database from Amazee.com, which is a Web-based platform of collaboration. Using Amazee's Web-platform, anyone with an idea for a collaborative project can sign in and use the website to gather followers, who will together help the project owner to accomplish the project. An Amazee project can be of any type of activities, such as arts and culture, environment and nature, politics and beliefs, science and innovation, social and philanthropic, sports and leisure, and so on. Most of the projects are public, for instance, "build a strong community of Internet entrepreneurs in Switzerland to exchange information and have fun" (Web Monday Zurich), "connect all women working in the Swiss ICT industry" (Tech Girls Switzerland), "to provide fresh running water to each home in the small African village of Dixie" (Water for Dixie), and so on. Amazee.com provides a set of features covering the entire lifetime of a typical project, such as project planning, participants recruiting, fund raising, events and meetings hosting, communication, files archiving, and so on. Users join Amazee.com by either creating a new project, or participating in projects created by others. The Amazee data we analyze contains the complete recording in time of the activities of all users creating and joining all the projects in existence between February 2008 and April 2011.

With the Amazee data we empirically estimate the power law distributions of project sizes, as well as parameters such as the average project growth rate, the standard deviation of growth rates, and the hazard rate of projects existing from the website, and use a newly developed model (Malevergne et al., 2010) to predict the power law exponent with these estimated parameters, thus testing empirically the theory about the origin of the Zipf's law.

1.2 Abstracts of scientific papers

1.2.1 Can media moods predict stock prices during and after the 2008 financial crisis?

We extract media moods from the Reuters US news using computational linguistics methods and study the relationships between these media moods and the S&P500 returns from January 1, 2007 to June 6, 2012. We report three major findings. First, negative moods Granger cause S&P500 returns with a negative coefficient. Second, the S&P500 returns Granger cause the negative moods also with a negative coefficient, showing the existence of a positive feedback loop between them. Third, we find that trading strategies based on media moods can generate both statistically significant and economically significant returns, and the extra returns cannot be explained by the Fama-French factors. The corresponding extracted α's (excess risk-adjusted returns unexplained by the Fama-French factors) are impressively high, in the range $0.1 - 0.2\%$ per day and thus dominate typical transaction costs and implementation slippage. This suggests that financial markets are not informationally efficient over the studied time period and that this results from the existence of the mutually reinforcing feedbacks between negative moods and negative S&P500 returns, which have been and are still present over this time period characterized by a very serious financial crisis and its on-going development.

1.2.2 Reverse engineering stock markets with mixed games and alpha generation

We construct virtual financial markets populated by artificial agents, who make decisions according four classes of backward-looking decision functions, with the goal of testing the weak form of the efficient market hypothesis (EMH). Our agent-based models (ABM) are populated by agents with bounded rationality and heterogeneous beliefs, which can be represented by the decision functions defining respectively the minority game, the majority game, the $-game and the delayed minority game. We extend a previous methodology and provide the main structural parameters, the specific trading strategies used by the agents, as well as the fractions of agents playing the four different games. This genuine reverse-engineered reconstruction of the real financial markets is applied to the 10-year time series of the S&P500, Dow Jones Industrial Average and Nasdaq 100 indexes from 1982 to 2012 in 700 experiments associated with different time windows. Our empirical results provide evidence that our ABM's can describe the behavior of a large proportion of investors in a real market. This is supported by (i) our finding that 654 out of 700 reverse engineering experiments with on three main U.S indexes predict the future return signs with statistically significant success rates, (ii) trades based on these predictions can generate statistically and economically significant returns, and (iii) there are statistically significant relations between market regimes and the corresponding parameters of reverse-engineered ABM's.

1.2.3 Empirical test of the origin of Zipf's law in growing social networks

Zipf's power law is a general empirical regularity found in many systems. We report a detailed analysis of a burgeoning network of social groups, in which all ingredients needed for Zipf's law to apply are verifiable and verified. A recently developed theory predicts that Zipf's law corresponds to systems that are growing according to a maximally sustainable path in the presence of random proportional growth, stochastic birth and death processes. We estimate empirically the average growth r and its standard deviation σ as well as the death rate h and predict without adjustable parameters the exponent μ of the power law distribution $P(s)$ of the group sizes s. Using numerical simulations of the underlying growth model, we demonstrate that the empirical stability of Zipf's law over the whole lifetime of the social network can be attributed to the interplay between a finite lifetime effect and a large σ value. Our analysis and the corresponding results demonstrate that Zipf's law can be observed with a good precision even when the balanced growth condition is not realized, if the random proportional growth has a strong stochastic component and is acting on young systems under development.

1.3 Contributions of the Ph.D. candidate to the papers

1. Can media moods predict stock prices during and after the 2008 financial crisis? Co-authored by Qunzhi Zhang and Didier Sornette. Qunzhi Zhang and Didier Sornette design the research. Qunzhi Zhang collects data and analyzes data. Qunzhi Zhang and Didier Sornette write the paper.

2. Reverse engineering stock markets with mixed games and alpha generation. Co-authored by Qunzhi Zhang, Didier Sornette and Jeffrey Satinover. Didier Sornette, Jeffrey Satinover and Qunzhi Zhang design the research. Qunzhi Zhang writes the reverse engineering software, and does the simulations and data analysis. Qunzhi Zhang and Didier Sornette write the paper.

3. Empirical test of the origin of Zipf's law in growing social networks. Co-authored by Qunzhi Zhang and Didier Sornette. Qunzhi Zhang and Didier Sornette design the research. Qunzhi Zhang analyzes the data. Qunzhi Zhang and Didier Sornette write the paper.

2

Can media moods predict stock prices during and after the 2008 financial crisis?

2.1 Introduction

Our main contribution is to present a novel methodology to test for market informational inefficiency, when using massive news data that are publicly available. Building investment strategies using mood indicators constructed on the news data allows us to demonstrate the presence of highly statistically significant excess risk-adjusted returns. Our results suggest the existence of positive feedback loops in the investors' mood dynamics produced by the flow of news, the investment decisions themselves influencing news and moods via their impact of prices through a procyclical process. This type of positive feedback loops is totally different from the standard negative feedbacks associated with the exploitation of anomalies, which make them to be arbitraged away, as described by the Efficient Market Hypothesis (EMH). Our empirical results thus bring new insights into the EMH, a cornerstone of financial economics, and provide more precise information on the formation mechanism of financial bubbles and crises.

On 20 September, 2012, the news title "Apple klaut Bahnhofs-Uhr der SBB" (Apple steals the Swiss Federal Railways station clock) by Rotzinger and Benkö (2012) was published in the German language on the website of Blick, a local newspaper in Zurich, Switzerland. In the early morning of 21 September, 2012, a similar title (posted as lydia_emyeu (2012)) appeared in Chinese on a website in China called "Weiphone". Following this posting, this same piece of news has been cited by many main news channels in China on the same day. Meanwhile, hundreds of news articles telling the same story spread on the Internet in English, published by different news media.

This example illustrates how quickly language and geographical barriers are overcome nowadays for the propagation of news, be they economic, financial, business or political news. For financial investors, this suggests that public information spreads so fast on the web and in the media that it should not be possible to gain by investing on the basis of news that everyone can have access to so quickly. In other words, stock markets have become even more informationally efficient, since never before have news stories spread so rapidly around the World following the emergence of the Internet.

The importance of news and their relationship with financial markets have a long history. Indeed, financial markets can be considered essentially as the engines that transform information into prices and provide both funding channels for firms and investment opportunities for all. According to the "efficient market hypothesis" (EMH), which was introduced by **?** and Samuelson (1965, 1973), price movements are almost perfect instantaneous reactions to the information flow. The emphasis on the claim that stock prices fully reflect all publicly available information is called the semi-strong form of the EMH (**?**). Accordingly, whatever the internal structure of financial markets, according to the semi-strong form of the EMH, price changes just reflect exogenous news. Being of all possible types (geopolitical, environmental, social, financial, economic and so on), these news lead investors to continuously reassess their expectations of the cash flows that firms and investment projects could generate in the future. These reassessments are translated into readjusted demand/supply functions, which then push prices up or down as a function of their impact in the order books. As a consequence, observed prices are considered the best embodiments of present values. The EMH is based on arbitrage arguments, according to which any possible arbitrage opportunity contained in a piece of news will disappear fast as more and more investors start to exploit it, leading to its incorporation in the asset price. Therefore, in such informationally efficient stock markets, it is almost impossible to gain extra returns by exploiting public information, and in particular as provided by news.

There is a lot of empirical support for the EMH, embodied by the statements that the dynamics of price is well described by the random walk model and that returns are very poorly predicted by available academic models (Rubinstein, 2001; Welch and Goyal, 2008). However, few works have tested how the media moods, which is publicly available to all investors, affect stock prices. We hypothesize that media moods about the stock markets, either positive or negative, may act as a global external influence that affects the general behaviors of investors. If the amplitude of this external influence is strong enough, the stochastic actions of the investors may be unified or coordinated to some extent by the common mood, similarly to spins in a magnet aligning towards an external magnetic field. This could happen especially strongly during the development of a bubble and its subsequent crash phase (**?**Harras and Sornette, 2011). Therefore, the investors could be driven solely by the media moods relating to the stock markets, even in the absence of any new information. Contrary to the arbitrage opportunities contained in news information that tend to be removed by the collective actions of profit maximizing investors, we hypothesize that the effects of media moods may not be eliminated but could in fact be enhanced by the collective reaction of the investors to them. For instance, negative media moods may drive the investors to sell, while more sells depress market prices and lead to even more negative moods, leading to a self-reinforcing downward spiral. Unlike the negative feedbacks associated with the exploitation of arbitrage opportunities contained in a news piece, the positive feedback loops produced by the media moods could be responsible for a kind of market informational inefficiency.

The development of computational linguistics methods and the availability of huge amount of financial data, including both stock prices and news information, make it possible to test the following hypotheses against each other:

H_0: The market is always semi-strongly efficient.

H_1: The market is not semi-strongly efficient in the sense that there are arbitrage opportunities when using the media moods extracted from large information feeds, such as Reuters.

Thanks to the wide use of the Internet, news articles are publicly accessible, such as those on the Reuters website that covers the period from 1st January 2007 to present, which are rich in content, with all kinds of information on a daily basis, including economic news, financial news, business news, politics news, and so on. We will use this source to extract media moods, with the help of text analysis algorithms. Our methodology consists in finding trading strategies based on media moods extracted from the Reuters news that provide abnormal risk-adjusted returns. If we can find such a trading strategy that generates statistically significantly excess positive returns based on the media moods, we shall reject H_0 and accept H_1; otherwise, we shall reject H_1 and accept H_0.

Our results show clearly that it is possible to construct such a trading strategy and gain statistically significantly positive excess returns based on the media moods, so that our main conclusion is that H_0 is rejected in favor of H_1, which is accepted.

It is however not clear if this kind of market informational inefficiency occurs only temporarily during or shortly after the serious financial crisis, due to the fact that our testing period (January 1, 2007 to June 6, 2012) overlaps strongly with the duration of the crisis. In the most restricted understanding, one would consider the financial crisis as just lasting over the official duration of the recession in the U.S., which began in December 2007 and ended in June 2009, according to the U.S. National Bureau of Economic Research. But, other measures support a more extensive duration, starting with the revelations of really significant problems that could affect stock markets, which occurred in the summer of 2007. This was followed by marked global economic declines in December 2007, which accelerated particularly sharply downward in September 2008 with Lehman Brothers' bankruptcy and AIG's bailout. The crisis may be argued to have continued and to be still ongoing at the time of writing, as gauged for instance by the drastic remedies in place, such as the open-ended Quantitative Easing policies decided by the Federal Reserve in September 2012. Thus, we cannot exclude the fact that the failure of H_0 could be due to the abnormal regime of the economy and financial markets, and thus could be transitory in nature.

Even if this is the case, our findings still have a significant impact in providing a novel metric for market inefficiencies, whether they are intrinsic or associated with a special era with strong central bank and political interventions. To our knowledge, there have been no similar studies reporting this kind of market informational inefficiency that uses massive news data.

The rest of the paper is organized as follows. In the second section, we briefly review previous works studying the relationships between news and asset prices. The third section introduces the news data extraction method we have employed in this study and present the stock prices data used for our tests. The fourth section explains our data analysis method and presents our main results. The fifth section discusses the results and concludes.

2.2 Previous works

There is a rich literature examining the relations between price changes and news. A first important observation is that prices move much too much compared with what would be expected from the EMH (the so-called "excess volatility puzzle") (**?**LeRoy and Porter, 1981; LeRoy, 2008). This suggests that there is more to price dynamics than just the direct impact of exogenous news (e.g. the dynamics of dividends). There have also been many attempts to relate price changes to news, using the "event study" approach, from long time scales (Cutler et al., 1987; McQueen and Roley, 1993) to high frequency trading time scales (Fleming and Remolona, 1997; Fair, 2002; Joulin et al., 2008; Erdogan and Yezegel, 2009). The "event study" method consists in testing if individual news events lead to abnormal returns occurring at times around the release times of the news events. In an excellent introduction and summary of the event study method, Kotharj and Warner (2007) count more than 500 event studies papers appearing in the leading financial and business journals from 1974 to 2000 and the number continues to grow. Here, we cannot give credit to this enormous literature but only focus on the most relevant works to our own study.

As already mentioned, event studies can be roughly divided into two sets, those that are concerned with short time horizons and those investigating event impacts on long time scales. The latter generally applies to time windows of one year or more, while the former applies to much shorter event windows, down to the smallest tick time scale. For time scales of days to weeks ("short time" scales at this epoch), Fama (1991) concluded that, on average, stock prices adjust quickly to publicly available corporate information, such as investment decisions, dividend changes, changes in capital structure and corporate-control transactions. Moreover, Fama (1998) pointed out that market efficiency has been able to survive the challenge from the literature investigating abnormal returns generated by news-based strategies, by concluding that the so-called abnormal returns are essentially due to luck. In sum, most results of event studies have supported the EMH, which holds that markets react very fast to new information and any arbitrage opportunities disappear almost immediately after the information is available publicly.

The use of large news databases available electronically is now casting some doubts on this claim of the EMH. By classifying according to topic some 245,420 Wall Street Journal corporate news stories from 1973 to 2001 with the help of computational linguistics methods, Antweiler and Frank (2006) found that statistically significant abnormal returns can be observed for many days after the release of public information. They document that the pre-event and post-event abnormal returns have on average opposite signs, suggesting under-estimation before the event and over-shooting after the event. The results are found to be sensitive to the duration of the events and the average news impacts are stronger in a recession than in an expansion. New results have been recently obtained using computational linguistics methods that enable financial economists not only to classify news stories according to topic, but also to extract semantic meanings from financial documents. Tetlock et al. (2008) found that words carrying negative sentiment can be used to predict individual firm's accounting earnings and stock returns. Loughran and Mcdonald (2011) studied 10-K filings and developed a method to extract negative sentiments from financial documents. For this, they construct a word list sorted in terms of their positive, negative, uncertainty, litigious, strong modal and weak modal characters. The use of a term weighting scheme enabled them to lower the noise resulting from word misclassification. Statistical tests showed that the word lists are significantly related to announcement returns. These results suggest that textual analysis can contribute to the ability of financial economists to understand the impact of information on stock returns.

More than financial documents, Internet stock message boards and Twitter messages have also been studied in the financial literature. There are numerous reports on the existence of correlations between financial activity, proxied e.g. by transaction volume or price volatility, and news, sentiment indices, mood indicators, search volume and other measures of social activity. These studies are not directly relevant to the question that we revisit here in a novel form, namely of the existence of market price predictability and of arbitrage opportunities. More relevant is the study of Antweiler and Frank (2004) on how Internet stock message boards are related to stock market price moves. Using simple Bayesian methods to extract information from 1.5 million messages, Antweiler and Frank (2004) found that stock messages help predict market volatility, and the impact of the stock messages on stock returns is statistically significant, but economically small. Bollen et al. (2011) used a computational linguistics method to extract public mood from Twitter messages and found that Twitter mood helps predict stock market price moves. Da et al. (2011) proposed a new and direct measure of investor attention using search frequency in Google and showed that it has some predictability for stock price moves.

While these results are certainly enticing, they however suffer from a lack of precision on what is meant by predicting stock price moves, in particular in relation with the EMH and the possible existence of arbitrage. Indeed, the proponents of the EMH do not claim an absolute absence of predictability but only that any residual predictability cannot in general be exploited to enjoy statistically significant abnormal risk-adjusted gains. As an illustration, it is well known that financial returns exhibit auto-correlations over time scales than can go down to the tick-by-tick level, depending of the level of liquidity and transaction costs. With such auto-correlations, it would in principle be possible to develop very profitable trading strategies using the simple Wiener filter predictor, if only transaction costs were absent or much smaller. In other words, the level of auto-correlation at short time scales is just the one that is marginally too costly to arbitrage away by any reasonable trading strategy. This does not contradict in any way the EMH. Actually, Grossman and Stiglitz (1980) proved that a perfect informationally efficient market is impossible, since the informed investors would not be remunerated by trading and would not be motivated to translate their information into the prices. Therefore, financial markets need to fluctuate in some sense close to an efficient frontier but not exactly on it, in order to entice traders to reveal their information by their profitable investment activities. In this sense, much more than just predictability is needed to demonstrate an abnormal or puzzling behavior of financial markets. It is our goal in the present work to present a rigorous and systematic investigation of the possibility to use news to develop abnormal risk-adjusted returns. In particular, we demonstrate that financial markets are not anymore efficient in the semi-strong form, notably due to the presence of the growing number of electronic news outlets. It seems that investors as a whole have not had time to "digest" the technology to remove significant arbitrage opportunities. This is probably due to the arm race between the on-going explosive growth of electronic based news, the active development of ever more sophisticated machine learning algorithms and the technical challenges to develop operational implementations for arbitraging. This implies that the message of Grossman and Stiglitz (1980) is now more relevant than ever in our changing accelerating electronic world. Based on our results, we conjecture that, paradoxically, financial markets may be further away from perfect equilibrium than before the dominance of the recent era of electronic news and algorithm trading.

The study closest to ours is by Matthias W. Uhl (2011), who used sentiment classified news rated by Thomson Reuters at a very coarse level and at the monthly scale. Matthias W. Uhl (2011) showed that the Reuters sentiment index explains the Dow Jones Industrial Average returns better than macroeconomic news.

2.3 Data

2.3.1 Extracting media moods from Reuters news stories

Reuters makes all its US news stories from January 1, 2007 to present available online for readers and researchers. For instance, Reuters' 2007 news archives can be found at Reuters.com (2012). From January 1, 2007 to June 6, 2012, there are 5,255,784 news stories, covering most of the important events in the U.S. and in the World, in the fields of economics, business, finance, politics, technology, entertainment, and so on.

In principle, one could get a precise understanding about the evolution of the global media sentiment about the U.S. stock market, either positive or negative, by reading all the news stories. This is, however, almost impossible for human researchers, as there are on average more than 2500 news stories each day, which implies that one would need more than five years to read all the texts generated in the period from January 1, 2007 to June 2012, even if one could survey two stories per minute.

Fortunately, computational linguistic methods can help us quantify the massive text data and extract media moods more efficiently than in a manual way. Hopkins and King (2010) introduced an excellent method for extracting media moods from massive text documents, and showed that this automatic way can be additionally more reliable than the manual way of human beings. Unlike human beings, who need many years of training to understand and appreciate the subtleties about whether a news story reflects a positive or negative view with respect to the stock market, computer programs using statistical language models to categorize news stories can be trained and learn over a very short time. While human beings get tired after long hours of work and make mistakes, computers do not have these limitations.

The key idea is to categorize daily news stories into three categories, positive, neutral and negative, according to the sentiments related to the stock market. The media moods are thus represented by the fractions of positive and negative sentiments. For instance, if most of the news stories in one day express a negative (respectively positive) view of the stock market, the feeling of a reader of the news stories would be pessimistic (respectively optimistic). The investment strategy that we propose is based on the hypothesis is that one would buy stocks when the investor feels optimistic media moods, or would sell when she feels pessimistic media moods. In this way, the investor's decisions drive the market even more optimistically or pessimistically, through a self-reinforcing loop, leading to a new kind of informational market inefficiency, which has been neglected in financial economics.

The categorizing algorithm introduced in Hopkins and King (2010) works as follows. Let us denote the positive, neutral and negative news categories as $D_l (l = 1, 2, 3)$, respectively. A news story is nothing else than a list of words. Thus, for a combination of all possible word stems $W = \{w_1, w_2, \ldots, w_K\}$, we can summarize a news story i as a word stem profile $S_i = \{S_{i1}, S_{i2}, \ldots, S_{iK}\}$, with $S_{ik} = 1 (k = 1, 2, \ldots, K)$ if the word stem w_k has been used in the news story, or $S_{ik} = 0$ if not. The total number of words in the considered dictionary is K. Then, the law of total probability reads

$$P(S_i) = \sum_{l=1}^{3} P(S_i|D_l)P(D_l) \tag{2.1}$$

where $P(S_i)$ is the probability of the word stem profile i occurring within the news population, $P(S_i|D_l)$ is the probability of the word stem profile i occurring within the news in the category l, for instance, the negative news, and $P(D_l)$ is the probability of a specific news category l, namely the fraction of news in category l, for instance, the fraction of daily negative news. Let us denote $\mathbf{S}$ as the set $\{S_i, i = 1, 2, \ldots, 2^K\}$, and D as $\{D_1, D_2, D_3\}$, then equation 2.1 can be rewritten as an equivalent matrix expression:

$$\underset{2^K\times 1}{P(\mathbf{S})} = \underset{2^K\times 3}{P(\mathbf{S}|D)}\,\underset{3\times 1}{P(D)} \tag{2.2}$$

To estimate $P(D)$ with the equation 2.2, we must first know $P(\mathbf{S})$ and $P(\mathbf{S}|D)$. The former can be directly obtained from the daily news, while the latter has to be estimated from the training set, which is a collection of hand coded news. According to Hopkins and King (2010), the hand coded news can be either randomly sampled or not. For the training set, we can write down an equation similar to equation 2.2:

$$\underset{2^K\times 1}{\tilde{P}(\mathbf{S})} = \underset{2^K\times 3}{\tilde{P}(\mathbf{S}|D)}\,\underset{3\times 1}{\tilde{P}(D)} \tag{2.3}$$

where $\tilde{P}(\mathbf{S})$ is the probabilities of word stem profiles $\mathbf{S}$ occurring within the training set, $\tilde{P}(\mathbf{S}|D)$ is the probabilities of word stem profiles $\mathbf{S}$ occurring within the news in categories D in the training set, and $\tilde{P}(D)$ is the fractions of news in categories D in the training set. Since the training set is hand coded, $\tilde{P}(D)$ is already known. Moreover, $\tilde{P}(\mathbf{S})$ can be estimated by counting the occurrence frequency of each $S_i, (i = 1, 2, \ldots, 2^K)$ in $\mathbf{S}$. Thus, with equation 2.3 we can eventually obtain $\tilde{P}(\mathbf{S}|D)$.

The underlying hypothesis of Hopkins and King (2010) is the following equation:

$$P(\mathbf{S}|D) = \tilde{P}(\mathbf{S}|D) \tag{2.4}$$

which indicates that the probability of every word stem profile $S_i, (i = 1, 2, \ldots, 2^K)$ occurring within any news category $D_l (l = 1, 2, 3)$ in the test set is the same as in the training set. This hypothesis is reasonable as long as the training set is thought to be representative of the test set.

Using equation (2.4) in equation (2.2) leads to

$$\underset{2^K\times 1}{P(\mathbf{S})} = \underset{2^K\times 3}{\tilde{P}(\mathbf{S}|D)}\,\underset{3\times 1}{P(D)} \tag{2.5}$$

As we have mentioned, $P(\mathbf{S})$ can be estimated directly from the daily news by counting the occurrence frequency of each word stem profile $S_i, (i = 1, 2, \ldots, 2^K)$. $P(\mathbf{S})$, along with $\tilde{P}(\mathbf{S}|D)$ obtained from the training set, make equation 2.5 equivalent to a linear regression equation. Thus, linear methods can be used to estimate $P(D)$, namely the fractions of positive, neutral, and negative news, which is our ultimate target.

If the number K of words in our dictionary is too big, 2^K becomes too large for any standard computer to handle. Therefore, only subsets of **S** will be used to estimate $P(D)$.

As said above, we must estimate $\tilde{P}(\mathbf{S}|D)$ from a training set. The size of the training set recommended by Hopkins and King (2010) is 500. We follow this advise and our training set size is taken exactly equal to 500.

The 500 news stories are chosen randomly from the Reuters website Reuters.com (2012), and sample different years, months, and days. We use the following main criteria for collecting the training set. The first one is **length**. If a story is too short, containing only one or two sentences, it will not be chosen. The second one is **purity**. If a story contains complex sentiments, for instance, both positive and negative, it will not be chosen. The third one is **language**: only news in English will be chosen. Once a news story has been chosen, we copy and save its content into a text file, and code it immediately with one of the values 1, 0, -1, denoting positive, neutral, and negative, respectively, based on how the news story is related to the stock market. The coded results are also saved into a text file called the control file.

Hopkins and King (2010) have implemented the above algorithm in their software "Readme" (Hopkins et al., 2012). In this study, we apply the "Readme" software on the Reuters daily news by feeding it with the daily news and the control file. Its final output is the fractions of positive, neutral and negative news on each day.

2.3.2 Stock prices data

Our media moods data aggregates all publicly available information. To test how it is related to the stock market, a good proxy is the S&P500 index, because it aggregates and represents all the prices information of 500 top publicly traded American companies. Using an index including more firms provides a better representation of the US market than the other indices containing a smaller number of companies. The daily high, low, open and close prices data of the S&P500 index from January 1, 2007 to June 6, 2012 are collected from the Yahoo! Finance website(Yahoo.com (2012)).

2.4 Data analysis methods and empirical results

2.4.1 Descriptive statistics of data and relationships between positive and negative media moods

The "Readme" software applied to the Reuters daily news stories provides the fractions of positive, neutral and negative news for every day from

January 1, 2007 to June 6, 2012. In the following, we refer to the fraction of positive news as the "positive moods" and to the fraction of negative news as "negative moods", putting the neutral news aside. The dynamics of the media moods, as well as the S&P500 index, are shown in figure 1. Table 1 lists the descriptive statistics of the positive and negative moods and of the daily returns of the S&P500 index.

One can observe in figure 1 three main regimes characterizing the relations between the positive moods and negative moods over the period from January 1, 2007 to June 1, 2012.

1. Before September 2007, the positive moods level is higher than that of the negative moods. This reflects the still buoyant U.S. stock market that peaked in October 2007.
2. On September 2007, there is a transition with a large drop in positive moods and a steady increase of negative moods, so that the level of negative moods quickly exceeds that of positive moods. This transition roughly coincides with the first serious news released during the summer of 2007 about serious valuation and redemption problems facing majors funds trading collaterized debt obligations and other securities associated with credits on the U.S. real estate market. After September 2007, the level of the positive moods has never recovered in the sense of overpassing that of negative moods. Two hills in the positive moods level occur in early 2009 and in the spring of 2010, probably related to the policies of the Federal Reserve and U.S. Treasury and the so-called "quantitative easing" actions. But these hills are rather short-lived and insufficient to overcome the negative moods.
3. In mid-2010, there is a second transition towards a third regime characterized by an even larger gap between the dominant negative moods and the positive moods.

The Elliott, Rothenberg and Stock unit root test (Elliott et al., 1996) applied to the positive and to the negative media moods show that they have no unit roots and can thus be considered stationary. Visual inspection of Figure 1 suggests that the media moods are trend stationary: the positive moods have a downward trend, while the negative moods have an upward trend. We regress both the positive and negative moods with respect to time, and find that the positive moods have a statistically significant daily trend of -0.000078 per day while the negative moods have a statistically significant daily trend of $+0.000053$, consistent with the visual observation. We thus de-trend the media moods time series by subtracting their temporal trends, and all the analyses hereinafter are performed on the de-trended media moods time series.

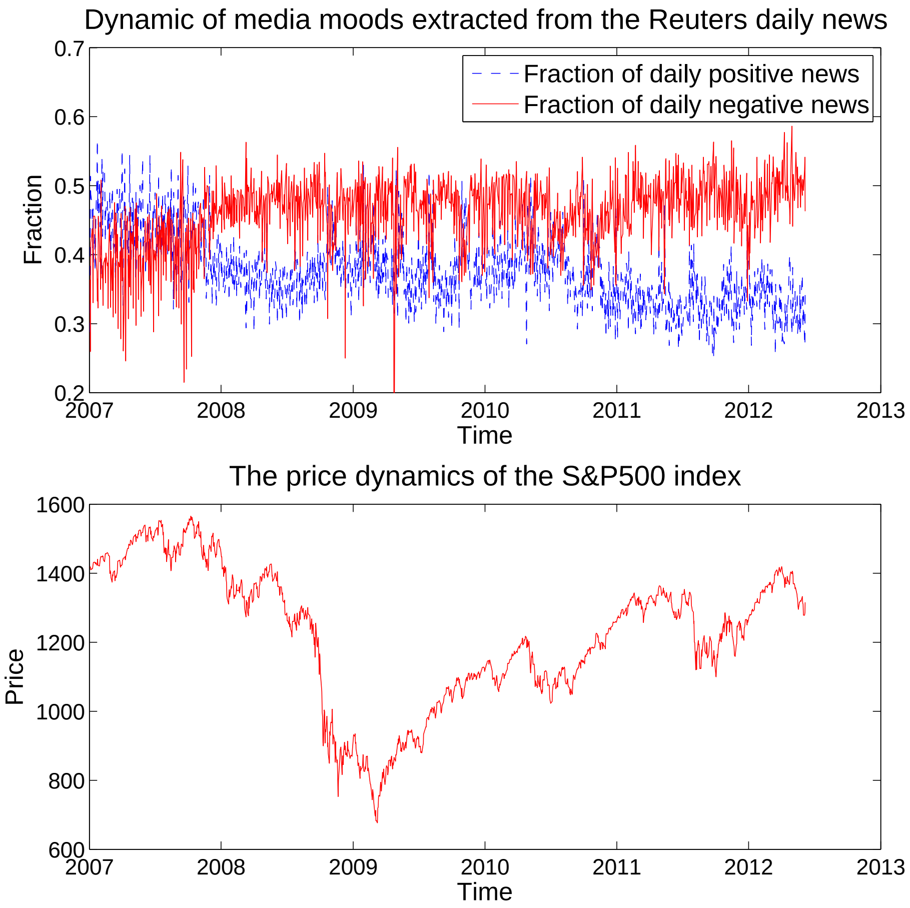

Fig. 1: The upper plot shows the dynamics of both positive (blue dashed line) and negative moods (red continuous line) extracted from the Reuters daily news from January 1, 2007 to June 6, 2012. The positive (respectively negative) mood for a given day is defined as the fraction of positive (respectively negative) news among all news articles provided by Reuters on that day. The lower plot shows the dynamics of the S&P500 index from January 1, 2007 to June 6, 2012.

Tab. 1: Descriptive statistics of the media moods and of the returns of the S&P500 index shown in figure 1 from January 1, 2007 to June 6, 2012.

	Positive moods	Negative moods	S&P500 returns
Number of Samples	1360	1360	1359
Mean value	0.3763	0.4586	-0.0000547
Minimum value	0.2485	0.0581	-0.0947
Maximum value	0.5671	0.5863	0.110
Median value	0.3722	0.4657	0.000817
Standard deviation	0.0533	0.0513	0.0163

The positive and negative moods are found to exhibit a significant negative correlation coefficient of -0.61. Moreover, by applying a VAR(p) model (Vector autoregression model with lag number p, see detailed description e.g. in (Hamilton, 1994)), we find that both the positive and the negative moods are not only autocorrelated but each time series Granger causes the other one, namely each time series can be used to improve the prediction of the other one. The best value p of the VAR(p) model is found equal to 10 for both the positive and negative moods, as detected by using the AIC (Akaike information criterion) (Akaike, 1974). The VAR(10) model reads

$$\mathbf{m_t} = \beta_0 + \sum_{i=1}^{10} \beta_\mathbf{i} \mathbf{m_{t-i}} + \epsilon_\mathbf{t} \,, \tag{2.6}$$

where $\mathbf{m_t} = (\mathbf{p_t}, \mathbf{n_t})'$ is the media moods vector containing both the positive moods $\mathbf{p_t}$ and the negative mood $\mathbf{n_t}$ at time t. The coefficient vectors $\beta_0, \beta_1, \ldots, \beta_{10}$ are the parameters to estimate and $\epsilon_\mathbf{t}$ denotes the white noise vector at time t.

Table 2 reports the results of the estimation of model (2.6), which shows that both the positive and negative moods are autocorrelated up to lag 10 days, and the lagged positive moods can alleviate the negative moods. Granger causality tests confirm that the positive moods Granger causes the negative moods, and the negative moods also Granger causes the positive moods. The orthogonal impulse response function plots shown in figure 2 constructed from model (2.6) present the relationships more intuitively: both the positive moods and the negative moods are self-enforcing, and the lagged positive moods alleviate the negative moods, but the lagged negative moods have little effect on the positive moods.

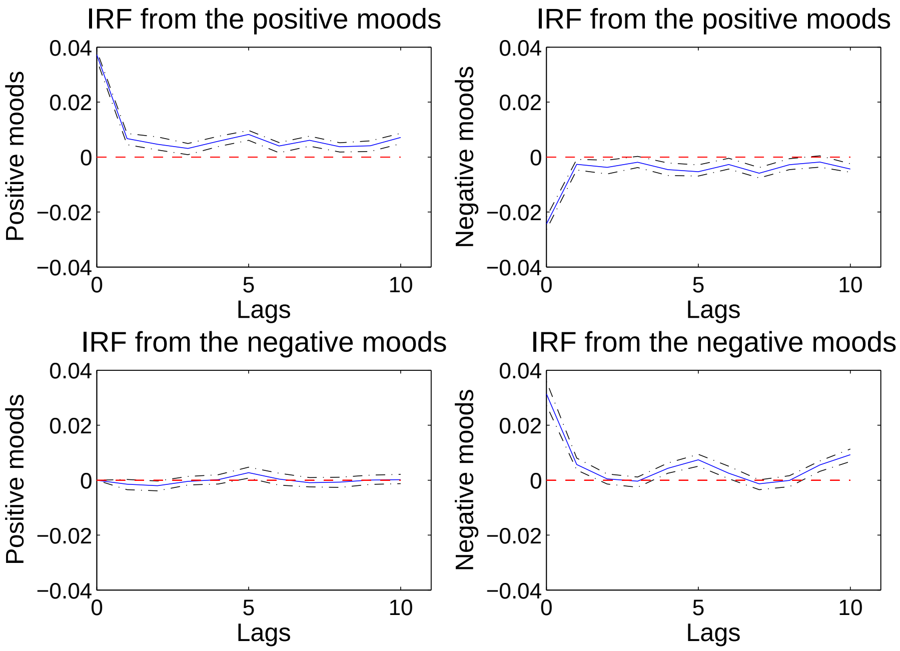

Fig. 2: The four plots display the orthogonal impulse response functions (IRF) of the media moods. The upper-left plot shows the response function of the positive moods receiving impulses from the positive moods; the upper-right plot shows the response function of the negative moods receiving impulses from the positive moods; the lower-left plot shows the response function of the positive moods receiving impulses from the negative moods; and the lower-right plot shows the response function of the negative moods receiving impulses from the negative moods. The solid blue lines in the plots are the orthogonal impulse response functions, the dash-dotted black lines are the 95% confidence intervals of the impulse response functions, and the red dashed lines indicate the 0 levels.

2.4.2 Detection of linear relationships between the media moods and the stock prices, as well as positive feedback loops

Since the positive moods and the negative moods are correlated, we cannot use the VAR model to study the relationships between the S&P500 returns and both positive and negative moods all together, as it would bias the parameter estimation. Therefore, we study the relationships between the returns and the positive moods, and the relationships between the returns and the negative moods, separately. For the positive moods, however, we find that it is not statistically significantly related to the returns so that, in this section, we shall discuss only the results concerning the relationships between returns and negative moods.

To study the linear relationships between returns and negative moods, we use again the VAR model. Based on the AIC and considering the simplicity of the model, we apply a VAR(3) model at this time, which reads

$$\mathbf{x_t} = \beta_0 + \beta_1 \mathbf{x_{t-1}} + \beta_2 \mathbf{x_{t-2}} + \beta_3 \mathbf{x_{t-3}} + \epsilon_\mathbf{t} \, , \qquad (2.7)$$

where $\mathbf{x_t} = (\mathbf{r_t}, \mathbf{n_t})'$, $\mathbf{r_t}$ denotes the S&P500 return on day t, and $\mathbf{p_t}$ denotes the negative mood on day t. The parameters β_0, β_1, β_2, and β_3 are the coefficient vectors of the independent variables at lags 1, 2, and 3, respectively, and $\epsilon_\mathbf{t}$ is the white noise vector.

Table 3 reports the estimation results of model (2.7), and the plots in figure 3 constructed from model (2.7) show the orthogonal impulse response functions of the S&P500 returns and the negative moods. Table 3 shows that the negative moods at lag 3 statistically significantly decreases the S&P500 returns at lag 0. And we have verified that the negative moods at lag 3 Granger causes the S&P500 returns at lag 0, namely that the negative moods at lag 3 help predict the S&P500 returns at lag 0. The results are statistically significant, though not economically significant, because the adjusted R^2 is only 0.021. Correspondingly, the lower-left plot in figure 3 shows how the S&P500 returns respond to changes in the negative moods. One can observe that the negative moods at lag 3 pushes down the returns at lag 0. The effects at lag 1 and 2 seem to be of the opposite (positive) sign, but the effects are not statistically significant, as shown in table 3.

We also use returns of the Dow Jones Index and the Nasdaq Index to estimate model (2.7), and get similar results. The relationships between the media moods and the index returns disappear, however, when we use the Korea Composite Stock Price Index (aka KOSPI) or the All-Ordinaries Stock Index. The phenomena is reasonable that both the Down Jones and the Nasdaq are U.S. indices, while the KOSPI and the All-Ordinaries are relatively independent of the U.S. stock market. All these results are shown in table 4. The results prove that the relationships between the media moods and the U.S. stock markets are robust. As we have mentioned before, since the S&P500 Index has 500 component stocks of leading companies publicly traded in the U.S stock market, much more than the other indices, we take it as the best representation of the U.S. stock market. In this paper we thus present only results with the S&P500 Index.

The above results state that increasing negative moods can predict that returns will be decreasing three days later. In turn, we find a feedback of the decreasing returns that increase the level of the negative moods, as shown from both table 3 and figure 3. Thus, there is a positive (or mutually reinforcing) feedback loop existing between the S&P500 returns and the negative moods. Along with the self-enforcing behavior of the negative moods, this raises the question: Is this positive feedback loop revealing of investors' irrational decisions, in the sense that investors follow too faithfully the media moods and thus create market information inefficiencies? In the next section, we shall study and answer this question.

The linear relationships between the daily S&P500 returns and the negative moods found here can be used in principle to predict future S&P500 returns. But this may not be feasible in practice because of the small obtained adjusted R^2. However, the results are significantly improved by increasing the time scale from daily to monthly returns, as shown in Table 5. The estimated parameters of a VAR(1) model are shown in Table 5, which reveals statistically significant linear relationships between the monthly average S&P500 returns and the monthly average negative moods. While the linear relationship is stronger, we find that the positive feedback loop disappears at the monthly scale.

Because we strive to get as much power as possible in order to test hypotheses H_0 versus H_1, we will use daily returns in the following in order to have much more statistics. The cost for this is more noise, as shown above. We address this issue by turning to nonlinear models.

2.4.3 Constructing trading strategies with a non-linear model of media moods

In order to test hypotheses H_0 versus H_1 as formulated above, we need to construct a trading strategy based on the media moods extracted from the Reuters news stories that trades on the S&P500 index and apply standard tests to decide whether it generates excess returns that are statistically significant on a risk-adjusted basis.

The linear models in the previous section provide statistically significant relationships between the S&P500 returns and the media moods but, as already mentioned, the results are not economically significant. To test if

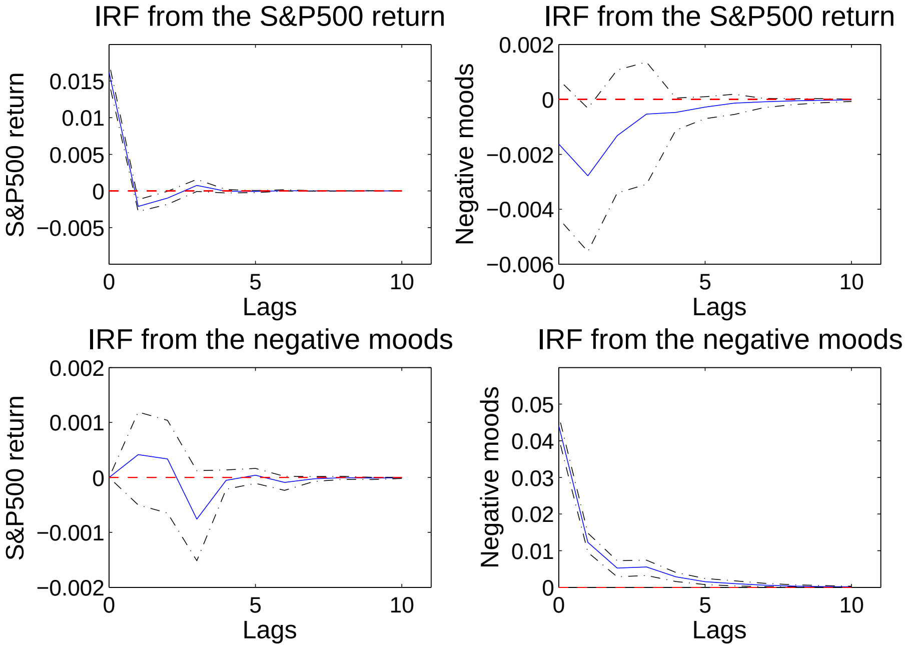

Fig. 3: The group of plots display the orthogonal impulse response functions of the S&P500 returns and the negative moods. The upper-left plot shows the response function of the S&P500 returns receiving impulses from the S&P500 returns, the upper-right plot shows the response function of the negative moods receiving impulses from the S&P500 returns, the lower-left plot shows the response function of the S&P500 returns receiving impulses from the negative moods, and the lower-right plot shows the response function of the negative moods receiving impulses from the negative moods. The solid blue lines in the plots are the orthogonal impulse response functions, the dash-dotted black lines are the 95% confidence intervals of the impulse response functions, and the red dashed lines indicate the 0 levels.

the positive feedback loop reflect the fact that investors make irrational decisions and thus create market informational inefficiency, we turn to non-linear models, namely, we use the Radial-Basis-Function (RBF) network to predict the S&P500 returns with the media moods. The RBF network is a kind of artificial neural network, and it has been proved (Park and Sandberg, 1991) that RBF networks having one hidden layer are capable of universal approximations. The RBF networks are thus strong potential tools to predict time series.

While we did not find statistically significant linear relationships between the positive moods and the S&P500 returns, we include the positive moods in the non-linear model, because they can affect the negative moods, and thus may affect the S&P500 returns non-linearly and indirectly through the negative moods. For simplicity and following the results of the linear models, we use lags of up to 3 days for the positive and negative moods to predict daily S&P500 returns.

The trading strategy works in the following way.

First, we train RBF networks with in-sample data. On each day in the in-sample period, we input into the RBF networks the positive and negative moods of the previous 3 days, and also the return on that day. Here, returns are calculated differently than in the previous section, where a return on a given day was defined as the difference of the log close prices of the day and its previous day. In the present section, a return on a given day is the difference of the log close price and the log open price. The justification for this definition is to align the target return (open to close) with the flow of media news, also occurring from open to close.

Second, we use the trained RBF networks to predict the out-of-sample returns. In the out-of-sample periods, if a predicted return is positive, we buy at the market open and sell at the market close; otherwise we short sell at the market open and buy back at the market close. We do not consider transaction costs in our trading strategy.

Third and last, the whole S&P500 index time series from January 1, 2007 to June 6, 2012 is split into blocks. The first block is an in-sample block, and the remain part is split into many out-of-sample blocks with the same length. Each out-of-sample block will be predicted by a RBF network trained with the in-sample data right before them. The length of the in-sample data and the length of the out-of-sample data are both fixed, but they can either be the same or different. The in-sample data window is moving, always followed by the out-of-sample block. Therefore, we can predict the return and trade everyday between January 1, 2007 and June 6, 2012, except in the first in-sample window starting from January 1, 2007.

Denote the length of the in-sample data by T_{in}, and the length of the out-sample data by T_{out}, we look for combinations of (T_{in}, T_{out}) such that the above trading strategy can generate statistically significantly positive risk-adjusted returns that are not due to chance. An important criterion to minimize data-snooping is that there should be many such winning combinations. This turns out to be the case. We shall report the results obtained for ($T_{in} = 180, T_{out} = 40$), which are representative of many other combinations. In particular, results remain robust when changing T_{in} and T_{out} in a neighborhood of these two values ($T_{in} = 180, T_{out} = 40$).

Figure 4 shows the cumulative Profit and Loss generated by the above strategy from September 20, 2007 to June 6, 2012, without considering the transaction costs. This corresponds to the entire time period between January 1, 2007 and June 6, 2012, minus the first $T_{in} = 180$ in-sample days. Comparing with the buy-and-hold as well as with the short-and-hold strategies, the strategy based on media moods generates a quite impressive cumulative returns.

2.4.4 Statistical tests of the trading strategy based on media moods

In order to assess the real value of our investment strategy, it is very useful to compare it with strategies that keep all its characteristics except the timing skills (Daniel et al., 2009). We thus refer to these benchmarks as "random strategies". The advantage in using them is that they are applied to the same data set, under the exactly same conditions, so that explicit or implicit factors that may conjure to promote abnormal returns in our initial strategy are also present for the random strategies. If present, these factors will translate into abnormal returns also visible in the random strategies. A given random strategy thus makes one and only one trade per day, either buying at the open and selling at the close or short selling at the open and buying back at close. The difference between random strategies and our's is that the trading decisions of the former are made randomly, not based on the media moods information.

Because there are two kinds of actions, long and short for each day, we consider an exhaustive set of random strategies classified according to their "long ratio" r_ℓ, defined as the fraction of long to short positions that the random strategy takes. In other words, each day, the random strategy chooses randomly to go long at the open (and sell at the close) with a probability equal r_ℓ and vice-versa with the complementary probability $1 - r_\ell$. We test all possible values for r_ℓ from 0 to 1 with a discrete step of 0.01. The limit value $r_\ell = 1$ corresponds to the buy-and-strategy, except for the overnight exit. The other limit value $r_\ell = 0$ corresponds to the sell-and-hold strategy, again except for the overnight exit. Each of the remaining other 99 classes of random strategies indexed by their r_ℓ value $\{0.01, 0.02, 0.03, \ldots, 0.97, 0.98, 0.99\}$ is sampled by generating 1000 random strategies with that specific r_ℓ. This allows us to obtain the percentile rank of the total cumulative return (denoted by τ_r) and the percentile rank of the Sharpe ratio (denoted by τ_{sr}) of our trading strategy based on the media moods within the population of the 1000 random trading results sampled for a given long ratio r_ℓ. Here, the ranks are ordered from the worst to the best performers, i.e., rank 1 (respectively 1000) corresponds to the worst (respectively best) random strategy in terms of the corresponding variable (cumulative return or Sharpe ratio). With the 99 possible values of the long ratios (excluding 0 and 1), we obtain 99 such τ_r's and τ_{sr}'s. We then pick the smallest τ_r and τ_{sr} among the values generated from the 99 testing long ratios. This corresponds to matching our strategy against the best possible random strategies according to their long ratio r_ℓ, which in a sense already gives some skills ex-post to the so-called random strategies. We can thus consider this procedure as disadvantageous for our strategy. This implies that a good performance of our strategy in such context where we bias the dice, so to speak, in favor of the random strategies, should be considered as really meaningful.

The results of these horse races in different time windows $[T_s, T_e]$ are reported in table 6, which shows that the percentile ranks of our strategy are higher than 97.5% for four periods among the total five periods. Only when considering the early period from September 20, 2007 to December 31, 2008, do we find a slightly smaller percentile rank of 92.5%, probably due to the very turbulent dynamics associated with the developing financial crisis, Lehman Brothers default and so on. But notwithstanding such turbulence and associated wild uncertainties in this time period, our strategy performs very well, when compared with the random strategies as well as with the buy-and-hold strategy or the S%P500 index. Thus, we conclude that our trading strategy based on the media moods has in general a probability less that 2.5% of being due to chance, suggesting strong support to accept hypothesis H_1 and reject H_0.

We perform another standard test, namely we try to remove the effects of the Fama-French factors (Fama and French, 1993) from our trading results based on the media moods. Regressing our time series of returns onto the Fama-French 3 factors model, we test if our trading strategy based on the media moods can generate statistically significantly positive α, i.e., if there is some abnormal positive return that cannot be explained by the Fama-French factors and that could thus be attributed to a new factor associate with the media moods information. We use the daily Fama-French factors data from the French (2012) website. Table 7 reports all the α's for the periods presented in table 6. We find that all α's are statistically significantly positive, with excess risk-adjusted daily returns between 0.109% and 0.203%. Here, "excess" means that these additional returns cannot be explained by the Fama-French factors. As a side remark, given the average daily values of the excess returns, it is clear that the reported effects remain valid in the presence of transaction costs and of implementation slippage, given the fact that our strategy trades only once per day (enter-exit) and transaction costs are a small fraction of the reported returns when strategies are implemented by large institutions, hedge-funds or through ETFs.

These results suggest that the media moods can generate both statistically significant and economically significant excess returns, which cannot be explained by the Fama-French factors. Therefore, there exists some market informational inefficiency related to the media moods. It seems that some market informational inefficiency does exist, at least during or shortly after a serious financial crisis, and this inefficiency is related to the positive feedback loop we have documented in the relationships between the stock prices and the media moods.

2.5 Conclusions

We have extracted daily positive and negative media moods from the Reuters US news archives from January 1, 2007 to June 6, 2012 by applying computational linguistics methods. We found that the positive and negative moods are negatively correlated. Both the positive and negative moods are autocorrelated, and more specifically self-enforcing. Moreover, the positive moods can alleviate the level of the next day negative moods, while the negative moods have no statistically significant effects on the next day positive moods.

We have analyzed the relationships between the S&P500 index and the media moods by first using linear models and have found that the negative moods with three days lag in the past predict a decrease of the S&P500 daily returns at lag 0, and the S&P500 daily returns at lag 1 in the past are negatively related to the level of the negative moods at lag 0. Thus, there exists a positive feedback loop in the relationships between the S&P500 daily returns and the negative moods. When using the monthly average returns of the S&P500 index and the monthly average negative moods, the positive feedback loop disappears, but the monthly average negative moods at lag 1 Granger cause the monthly average returns at lag 0 with a negative coefficient. The relationships between the positive moods and the S&P500 returns are however unclear, given the empirical data we extracted from the Reuters US news.

Although the linear relationships between the negative moods and the S&P500 daily returns are statistically significant, they are economically small. To test it is possible to use the media moods to predict the S&P500 daily returns, we constructed a trading strategy with Radial-Basis-Function (RBF) networks, which are a type of neural networks. Trained with in-

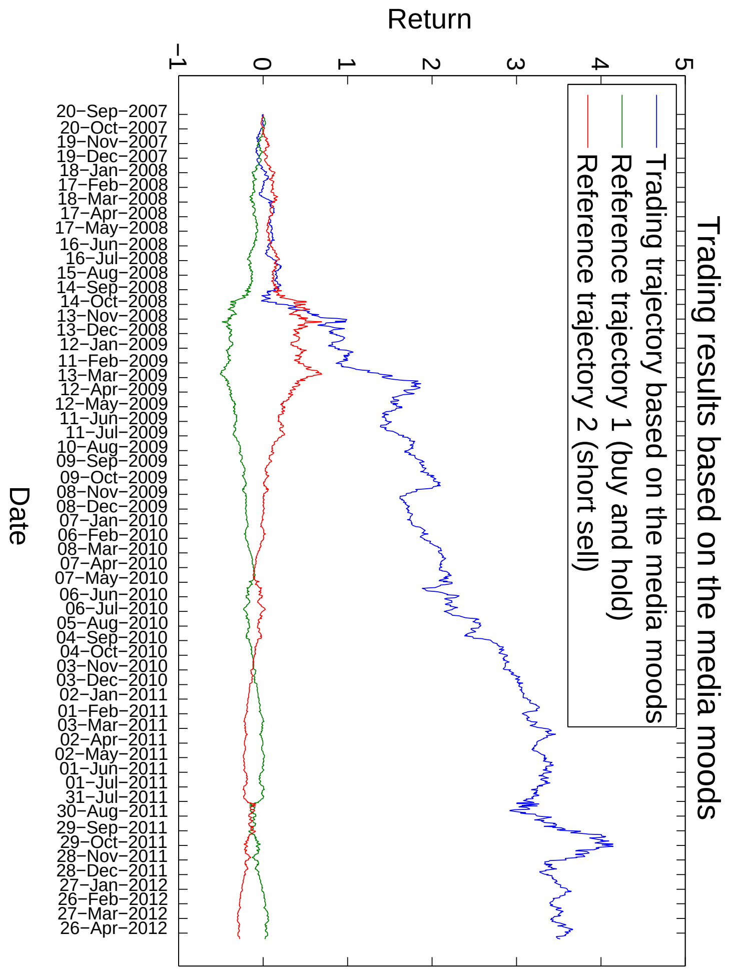

Fig. 4: P&L (cumulative return) without considering the transaction costs obtained by the strategy described in the text, which is based on the media moods (blue thick line), over the entire time period between January 1, 2007 and June 6, 2012, minus the first $T_{in} = 180$ in-sample days. The buy and hold strategy is represented by the thin green dashed line. The sell and hold (and buy back) strategy, consisting in shorting the S&P500 index at the beginning and holding until the end when the buy back occurs, is shown as the thin red dotted-dashed line.

sample data, the RBF networks have allowed us to predict the returns of out-of-sample data. We found that the trading results based on the predictions of the RBF networks are both statistically and economically significant. The chance for a trader to generate the same results as our trading strategy by using random trading strategies that we devised in a way to bias them favorably is statistically significantly very small. Moreover, by applying the Fama-French three factors model, we find that our trading strategy based on the media moods generates statistically significantly positive excess returns, which cannot be explained by the Fama-French factors. The corresponding extracted α's are impressively high, in the range $0.1 - 0.2\%$ per day and thus dominate typical transaction costs and implementation slippage. Since the RBF networks have been trained with a three day history of media moods of each present trading time, this suggests that there exists some market informational inefficiency, at least during or shortly after a serious financial crisis.

The contribution of our paper is to identify a novel example of market informational inefficiency. The mainstream view is that the market is informationally efficient because an arbitrage opportunity contained in a piece of new information will disappear almost immediately as soon as sufficiently many investors start to exploit it. However, we find that there are cases when the stock market is not informationally efficient because the exogenous forces such as the media moods can generate positive feedback loops: the negative moods make the returns go down, and the decreasing returns further increase the negative moods. In those cases, one could exploit the arbitrage opportunities by short selling. However, the short sells will likely aggravate the negative returns and thus the negative moods - so the arbitrage opportunities will not disappear but be self-enforcing.

With the available empirical data, we have been unable to detect similar positive feedback loop in the relationships between the positive moods and the stock returns. The question is thus still open as to whether there are positive feedback loops in the relationships between the media moods, both positive and negative, and the stock returns, during non-crisis time. If such positive feedback loops could still be observed, this would suggest that there is fundamentally informational inefficiency in the stock markets.

Tab. 2: Results of the estimation of the VAR(10) model 2.6 on the positive and negative moods. The rows labeled by $\beta_0, \beta_1, \ldots, \beta_{10}$ list the estimated parameters, where the values in the parentheses are the standard errors of the corresponding parameters, the "Adj. R^2" row lists the adjusted R^2 of the two linear models embedded in model 2.6, and the "F-stat" row lists the F statistics of the two linear models, where the values in the parentheses are the corresponding p-values. The p_t (respectively n_t) column list the results of the linear model whose dependent variable is the positive (respectively negative) moods at lag 0. Moreover, "*" indicates that the marked parameter is statistically significant with a p-value less than 0.1, "**" indicates a p-value less than 0.05, "***" indicates a p-value less than 0.01, and "****" indicates a p-value less than 0.001.

		$\mathbf{p_t}$		$\mathbf{n_t}$	
β_0		0.107^{**}	(0.046)	0.143^{***}	(0.049)
β_1	$\mathbf{p_{t-1}}$	0.150^{****}	(0.036)	0.047	(0.037)
	$\mathbf{n_{t-1}}$	-0.047	(0.032)	0.182^{****}	(0.034)
β_2	$\mathbf{p_{t-2}}$	0.063^{*}	(0.035)	-0.107^{***}	(0.037)
	$\mathbf{n_{t-2}}$	-0.050	(0.032)	-0.016	(0.034)
β_3	$\mathbf{p_{t-3}}$	0.052	(0.035)	-0.029	(0.038)
	$\mathbf{n_{t-3}}$	0.008	(0.032)	-0.015	(0.035)
β_4	$\mathbf{p_{t-4}}$	0.127^{****}	(0.035)	-0.013	(0.038)
	$\mathbf{n_{t-4}}$	0.013	(0.032)	0.135^{***}	(0.034)
β_5	$\mathbf{p_{t-5}}$	0.223^{***}	(0.035)	0.013	(0.038)
	$\mathbf{n_{t-5}}$	0.099	(0.033)	0.018^{****}	(0.035)
β_6	$\mathbf{p_{t-6}}$	0.012^{***}	(0.035)	-0.014	(0.038)
	$\mathbf{n_{t-6}}$	0.017	(0.033)	0.001	(0.035)
β_7	$\mathbf{p_{t-7}}$	0.080^{**}	(0.035)	-0.128^{****}	(0.037)
	$\mathbf{n_{t-7}}$	-0.010	(0.032)	-0.043	(0.035)
β_8	$\mathbf{p_{t-8}}$	-0.009	(0.035)	0.009	(0.037)
	$\mathbf{n_{t-8}}$	-0.015	(0.032)	0.000	(0.034)
β_9	$\mathbf{p_{t-9}}$	-0.003	(0.035)	0.104^{***}	(0.037)
	$\mathbf{n_{t-9}}$	-0.019	(0.032)	0.112^{***}	(0.034)
β_{10}	$\mathbf{p_{t-10}}$	0.083^{**}	(0.034)	0.055	(0.037)
	$\mathbf{n_{t-10}}$	-0.026	(0.032)	0.187^{****}	(0.034)
Adj. R^2		0.271		0.274	
F-stat.		26.1	(0.000)	26.47	(0.000)

Tab. 3: Results of the estimation of the VAR(3) model (2.7) using the S&P500 returns. The rows labeled $\beta_0, \beta_1, \ldots, \beta_{10}$ list the estimated parameters, where the values in parentheses are the standard errors of the corresponding parameters. The "Adj. R^2" row lists the adjusted R^2 of the two linear models embedded in model (2.7). The "F-stat" row lists the F statistics of the two linear models, where the values in the parentheses are the corresponding p-values. The r_t (respectively n_t) column lists the results of the linear model whose dependent variable is the returns (respectively negative moods) at lag 0. Moreover, "*" indicates that the marked parameter is statistically significant with a p-value less than 0.1, "**" indicates a p-value less than 0.05, "***" indicates a p-value less than 0.01, and "****" indicates a p-value less than 0.001.

		$\mathbf{r_t}$		$\mathbf{n_t}$	
β_0		0.001	(0.006)	0.252^{****}	(0.015)
β_1	$\mathbf{r_{t-1}}$	-0.130^{****}	(0.036)	-0.144^{*}	(0.074)
	$\mathbf{n_{t-1}}$	0.009	(0.010)	0.277^{****}	(0.027)
β_2	$\mathbf{r_{t-2}}$	-0.074^{***}	(0.027)	-0.048	(0.075)
	$\mathbf{n_{t-2}}$	0.006	(0.010)	0.043	(0.028)
β_3	$\mathbf{r_{t-3}}$	0.028	(0.027)	-0.010	(0.074)
	$\mathbf{n_{t-3}}$	-0.018^{*}	(0.010)	0.083^{***}	(0.027)
Adj. R^2		0.021		0.102	
F-stat.		5.862^{****}	(0.000)	26.58^{****}	(0.000)

Tab. 4: Results of the estimation of the VAR(3) model (2.7) using different indices other than the S&P500. The rows labeled $\beta_0, \beta_1, \ldots, \beta_{10}$ list the estimated parameters, where the values in parentheses are the standard errors of the corresponding parameters. The "Adj. R^2" rows lists the adjusted R^2 of the two linear models embedded in model (2.7). The "F-stat" rows lists the F statistics of the two linear models, where the values in the parentheses are the corresponding p-values. The r_t (respectively n_t) columns list the results of the linear model whose dependent variable is the returns of the corresponding index (respectively negative moods) at lag 0. Moreover, "*" indicates that the marked parameter is statistically significant with a p-value less than 0.1, "**" indicates a p-value less than 0.05, "***" indicates a p-value less than 0.01, and "****" indicates a p-value less than 0.001.

Indices		Dow Jones				Nasdaq			
		$\mathbf{r_t}$		$\mathbf{n_t}$		$\mathbf{r_t}$		$\mathbf{n_t}$	
β_0		0.015	(0.013)	0.438^{****}	(0.039)	0.018	(0.015)	0.438^{****}	(0.039)
β_1	$\mathbf{r_{t-1}}$	-0.129^{****}	(0.027)	-0.139^{*}	(0.080)	-0.107^{****}	(0.027)	-0.123^{*}	(0.070)
	$\mathbf{n_{t-1}}$	0.007	(0.011)	0.294^{****}	(0.033)	0.005	(0.013)	0.293^{****}	(0.033)
β_2	$\mathbf{r_{t-2}}$	-0.075^{***}	(0.027)	-0.092	(0.081)	-0.064^{**}	(0.027)	-0.084	(0.070)
	$\mathbf{n_{t-2}}$	-0.001	(0.012)	-0.082^{**}	(0.034)	0.000	(0.013)	-0.083^{**}	(0.034)
β_3	$\mathbf{r_{t-3}}$	0.041	(0.027)	-0.030	(0.080)	0.036	(0.027)	0.007	(0.070)
	$\mathbf{n_{t-3}}$	-0.025^{**}	(0.011)	0.023	(0.033)	-0.028^{**}	(0.013)	0.024	(0.033)
Adj. R^2		0.021		0.139		0.014		0.140	
F-stat.		3.938^{****}	(0.000)	22.96^{****}	(0.000)	2.957^{***}	(0.001)	23.00^{****}	(0.000)

Indices		KOSPI				All-Ordinaries			
		$\mathbf{r_t}$		$\mathbf{n_t}$		$\mathbf{r_t}$		$\mathbf{n_t}$	
β_0		0.016	(0.015)	0.439^{****}	(0.039)	0.021^{*}	(0.012)	0.436^{****}	(0.038)
β_1	$\mathbf{r_{t-1}}$	0.015	(0.028)	-0.094	(0.073)	-0.017	(0.027)	-0.075	(0.088)
	$\mathbf{n_{t-1}}$	-0.014	(0.013)	0.250^{****}	(0.034)	-0.026^{**}	(0.010)	0.303^{****}	(0.033)
β_2	$\mathbf{r_{t-2}}$	0.001	(0.028)	-0.018	(0.073)	0.008	(0.027)	-0.160^{*}	(0.088)
	$\mathbf{n_{t-2}}$	0.010	(0.013)	-0.058^{*}	(0.035)	0.014	(0.011)	-0.105^{***}	(0.034)
β_3	$\mathbf{r_{t-3}}$	-0.006	(0.028)	-0.031	(0.073)	-0.071^{***}	(0.027)	-0.006	(0.088)
	$\mathbf{n_{t-3}}$	-0.016	(0.013)	0.031	(0.034)	-0.015	(0.010)	0.030	(0.033)
Adj. R^2		-0.002		0.119		0.004		0.139	
F-stat.		0.693	(0.732)	18.95^{****}	(0.000)	1.534	(0.121)	22.94^{****}	(0.000)

Tab. 5: Linear relationships between the monthly average S&P500 returns and the monthly average negative moods modeled by a VAR(1) model. The "Const." row lists the values of the constant parameter in the linear models. The $\mathbf{\bar{r}_{t-1}}$ row lists the coefficients of the monthly average returns at lag 1. The $\mathbf{\bar{n}_{t-1}}$ row lists the coefficients of the monthly average negative moods at lag 1. In the above rows, the values in the parentheses are the standard errors of the corresponding coefficients. The "Adj. R^2" row lists the adjusted R^2 of the linear models. The F-stat. row lists the F statistics of the linear models, where the parentheses give the p-values. The $\mathbf{\bar{r}_t}$ column contains the estimation results of the linear model whose dependent variable is the monthly average returns at lag 0. The $\mathbf{\bar{n}_t}$ column contains the estimation results of the linear model whose dependent variable is the monthly average negative moods at lag 0. Moreover, "*" indicates that the marked parameter is statistically significant with a p-value less than 0.1, "**" indicates a p-value less than 0.05, "***" indicates a p-value less than 0.01, and "****" indicates a p-value less than 0.001.

	$\mathbf{\bar{r}_t}$		$\mathbf{\bar{n}_t}$	
Const.	0.011^{**}	(0.005)	0.200^{****}	(0.044)
$\mathbf{\bar{r}_{t-1}}$	0.095	(0.122)	-0.866	(0.992)
$\mathbf{\bar{n}_{t-1}}$	-0.027^{**}	(0.013)	0.528^{****}	(0.105)
Adj. R^2	0.053		0.286	
F-stat.	2.819	(0.067)	14.01	(0.000)

Tab. 6: Performance of our strategy based on the media moods and comparison with random strategies. We consider five different time periods starting with different starting times T_s and ending times T_e. All dates are given in the "year-month-day" format. The r_t column lists the total returns of the trades over the corresponding time interval $[T_s, T_e]$. The $\bar{r}$ column lists the average annual returns of our strategy for each time window and the sr column lists the annualized Sharpe ratios (using zero risk-free interest rate). The τ_r column lists the percentile rank of the total returns of our strategy compared with random strategies, as explained in the text. The τ_{sr} column lists the percentile rank of the Sharpe ratios of our strategy compared with random strategies, as explained in the text.

No.	T_s	T_e	r_t	$\bar{r}$	sr	τ_r	τ_{sr}
1	2007-01-01	2008-12-31	95.04%	58.38%	1.659	92.5%	92.5%
2	2007-01-01	2009-12-31	171.1%	49.50%	1.570	97.5%	97.5%
3	2007-01-01	2010-12-31	304.1%	47.50%	1.708	97.5%	97.5%
4	2007-01-01	2012-06-06	351.8%	35.77%	1.398	97.5%	97.5%
5	2008-04-16	2012-06-06	137.6%	27.89%	1.351	95.0%	95.0%

Tab. 7: List of the α's of our trading strategy based on the media moods obtained for the five periods shown in table 6. The α's are obtained as the intercepts of the regression of the time series of returns of our strategy in each time window as a function of the three Fama-French factors model. The parentheses give the standard errors of the corresponding α's. "*" indicates that the marked α is statistically significant with a p-value less than 0.1, "**" indicates a p-value less than 0.05, and "***" indicates a p-value less than 0.01.

No.	α	
1	0.203^{*}	(0.120)
2	0.187^{**}	(0.084)
3	0.185^{***}	(0.062)
4	0.140^{***}	(0.047)
5	0.109^{**}	(0.045)

3

Reverse engineering stock markets with mixed games and alpha generation

3.1 Introduction

The efficient market hypothesis (EMH) (Fama, 1970) is the basis of the neoclassic financial economic theory. Its weak form states that all the information on the history of asset prices has been incorporated into their current prices, so one cannot profit from using them in excess of the expected risk-adjusted return based on the market risk factor. The semi-strong and strong forms of EMH extend this statement to all publicly available and to private information, respectively. A rich behavioral finance literature has provided numerous pieces of evidence that challenge the semi-strong form and the strong form of EMH (see (Subrahmanyam, 2008) for a detailed review of this literature). The weak form of EMH, however, is generally considered to be more robust (Fama, 1991, 1998). Nevertheless, the occurrence of the recent financial crisis of 2007 - 2008, ensuing "great recession" and the on-going European sovereign debt crisis, accompanied by strong bullish markets suggest the existence of significant anomalies occurring in financial markets. In particular, asset returns may exhibit transient dependence structures that are incompatible with the no-arbitrage principle. This motivates us to develop a new set of tools to probe the anomalies that can develop in financial markets, and devise better predicting methods for assets prices. For this, we develop an interdisciplinary approach with concepts and tools developed in diverse fields, including financial economics, statistical physics and computer science.

Our main tool is agent-based modeling (aka ABM's), also known as agent-based computational economics (ACE). ABM provides a possible alternative to equilibrium models, because it relaxes some of their restrictive assumptions by adopting a bounded rationality framework (Simon, 1955b; Rubinstein, 1997). By taking into account the heterogeneity of the preferences and skills exhibited by different agents, by allowing deviations from equilibrium and by embracing a fundamental out-of-equilibrium dynamical view of the world, ABM offers the possibility to account for financial bubbles, market instabilities and crises as well as regime shifts modeled with an endogenous approach. By their structure, ABM is an ideal tool to study complex interactions between agents as occurs in stock markets. We refer to (Hommes, 2006; Hommes and Wagener, 2009; Chiarella et al., 2009; Evstigneev et al., 2009) for authoritative reviews on agent-based models from different perspectives. The method that we develop and present below is inspired by works of statistical physicists and financial economists, including Arthur (1994); Challet and Zhang (1997); Challet et al. (2000); Jefferies et al. (2001); Andersen and Sornette (2003); Wiesinger et al. (2012).

We consider four classes of ABM that have been studied separately in the literature and mix them in what we will refer to as mixed games. These four types of ABM are respectively the minority game, the delayed minority game, the majority game and the \$-game. In the minority and delayed minority games Challet and Zhang (1997); Challet et al. (2000); Jefferies et al. (2001), agents are rewards by playing strategies whose outcome follow the minority choice. This class of rewards emphasize situations encountered in entry situations, when an investor needs to get the best price when buying or selling ahead of the crowd. Minority games capture some of the mechanisms associated with changes of regimes. The majority and the \$ games (Andersen and Sornette, 2003; Wiesinger et al., 2012) describe agents who are prone to herding by the mechanism of rewards given to follow the majority. This creates positive feedback loops, thus providing the possibility for bubbles and crashes to develop. Our accompanying paper (Zhang et al., 2013) shows that the interplay between these four ABM in our mixed games reproduces the most important stylized facts of stock returns.

The main contribution of our paper is to extend the calibration of our mixed games to real financial time series, along the lines of Andersen and Sornette (2005); Wiesinger et al. (2012), using a more flexible and powerful methodology, with much larger time series. We call this calibration process "reverse engineering", because we not only determine the parameters of the models (such as number of agents, fraction of agents in each game, memory length, number of strategies per agent, and so on), but also determine the set of specific strategies used by each of our virtual agent so that their aggregate behavior best match the realized time series of financial returns. In this way, our ABM's disclose some of the microscopic mechanisms at work in stock markets and use this knowledge to predict their aggregate macroscopic behaviors. To address the criticism that "Prediction is a thorn in the side of ABM" (Elsenbroich, 2011), we use our ABM's to predict the future return signs, and construct trading strategies based on the predictions of future returns. We also investigate the relations between market regimes and the parameters of our calibrated ABM's. We find that our ABM's have statistically significant prediction power and generate robust positive alpha's, as tested with both the three factor Fama-French model and the four factor Carhart-Fama-French model. Moreover, the relations between market regimes and the parameters of ABM's provides novel ABM-based diagnostics of market regimes and their switching phases. Taken together, these results challenge the weak form of the EMH, since the predictions made by our ABM's are based only on the price history.

The remaining parts of this paper are organized as follows. In the next section 3.2, we present the agent-based models and the methodology. In section 3.3, we introduce the ABM calibrating method, and report the main results obtained on predicting future return signs, trading based on the predictions, and studying the relations between market regimes and the calibrated parameters of the ABM's. The last section concludes.

3.2 The ABM's

3.2.1 General definitions

We model stock markets within a bottom-up approach in which interactions between investors at the microscopic level are aggregated to generate the macroscopic dynamics of stock prices. The agents are abstractions of humain investors, living in the virtual digital world and interacting with each other by buying or selling shares of virtual stocks. The collective actions of the agents form the dynamics of the stock prices in the virtual stock markets. Specifically, we build virtual stock markets containing only one asset and a fixed number N agents who trade over Z discrete time periods. At any given time $t \in \{1, \dots, Z\}$ in a virtual stock market, each agent makes her decision to buy one share of the asset, to sell one share of the stock or to do nothing (short selling is allowed). When all the agents have made their actions, the imbalance between the amounts of shares to buy or sell by the agents determines the return r_t of the virtual asset at time t. Thus, from $t = 1$ to $t = Z$ the virtual market generates a return time series $r := \{r_1, \ldots, r_Z\}$ of length Z.

Next, we discuss how agents in our virtual stock markets form the stock prices, and show how the pricing mechanism in our virtual stock markets is linked to the real stock markets.

3.2.2 Agent decision making rules

Agents in our ABM's are boundedly rational, they "satisfice" rather than "optimize", as Simon (1956) argued. Following the concept of bounded rationality, we assume that agents have both limited knowledge about the virtual asset prices and limited computation capacity, so that they are not able to maximize their utility functions even if they had full knowledge of it. Our bounded rational agents therefore make trading decisions based on history prices data of the virtual asset. In contrast with fully informed perfectly rational agents, the beliefs formed by our agents could turn out to be wrong. Moreover, our agents need only to know qualitative characteristics of their utility functions in order to calculate their preferences concerning their trading decisions.

To be specific on the characteristics of our boundedly rational agents, we borrow from the literatures on the minority game, the \$-game and the majority game (Challet and Zhang, 1997; Challet et al., 2000; Jefferies et al., 2001; Andersen and Sornette, 2003; Wiesinger et al., 2012). In this approach, the behaviors of agents are sufficiently simplified to allow for extensive simulations on modern CPU cores, though the computations remain intensive, especially for the reverse-engineering of the ABM on long real financial time series.

We now describe the decision making process of a typical agent. This agent records the history of the asset price change, namely the positive and negative signs of the returns over only m time steps back from present. In other words, the memory of the agent is of length m, and is the same for all agents. The information that an agent uses to form her decision can thus be represented as a binary vector of length m, in which 1 indicates a positive return and 0 a negative return.

We assume that all agents have sufficient wealth so that, even if they underperform, they can still continue to play the investment game until the end of the Z time period. The decision of each agent is chosen in the triplet $\{+1, 0, -1\}$, where $+1$ corresponds to buying one share of the virtual asset, 0 means that the agent does nothing and -1 represents that she sells one share of the virtual asset. To reduce the complexity in our ABM's, we let the agents buy or sell exactly one share every time she decides to trade.

Agents make decisions using their trading strategies, which are based on the history of the asset price and embody their beliefs. We denote μ_t

the history information that each agent holds at time t, before she makes her decision. Thus, μ_t is a binary vector of length m, containing the stock price change directions from $t - m$ to $t - 1$. The agents use naive trading strategies to map the input information to their actions, denoted by $f(\mu_t) : \{0, 1\}^m \rightarrow \{+1, -1\}$. The size of the trading strategy space, denoted by $\mathbf{F} := \{f | \forall f : \{0, 1\}^m \rightarrow \{+1, -1\}\}$, is thus 2^{2^m}. Due to the limited capacity of the agents, we assume that they cannot remember and track the performance of all these 2^{2^m} trading strategies, but only a few of them. We denote by s the number of trading strategies of each agent, with $s << 2^{2^m}$. This number s is the same for all agents, while the specific used strategies are different from agent to agent, with possible overlaps. The trading strategy set of a given i agent is denoted by $\mathrm{F}_i := \{f_i^1, f_i^2, \ldots, f_i^s\}$, and obviously $\mathrm{F}_i \subset \mathbf{F}$.

The decision problem of an agent is now reduced to that of choosing the best trading strategy at time t, 'best' in some sense to be specified. To satisfice herself, we posit that the agent always chooses the trading strategy that is most probably generating positive returns among the s trading strategies from her trading strategy set. She does not know ex-ante what trading results a given trading strategy can produce at time t. The best she can do is to learn from the history performance of all her s trading strategies. The four games used here assume that the agent chooses the trading strategy with the highest performance recorded in the last m time steps. This is a backward-looking and myopic approach, embodying the behaviors documented in real financial traders and investors, who tend to look at past recent returns as a predictors of future ones (Hommes, 2006; Hommes and Wagener, 2009; Chiarella et al., 2009; Evstigneev et al., 2009). Note that this class of strategies is diametrically different from the forward-looking strategies of fully rational agents with forward expectations (Blanchard, 1979). In our ABM's, we define a simple performance assessment function for each trading strategy of a given agent i, which is the sum of the history payoffs of the trading strategy:

$$U(f_i^j, t) = \sum_{\zeta=1}^{t-1} \pi(f_i^j(\mu_\zeta)) \,, \tag{3.1}$$

where $U(f_i^j, t)$ denotes the performance assessment function of trading strategy j, $\forall j = 1, \ldots, s$, of the agent i, $\forall i = 1, \ldots, N$, at time t, and π is the payoff function of the agents' trading strategies, which returns a positive value if the trading strategy f_i^j has predicted an action consistent with the realized market behavior, or a negative value if the prediction of the trading strategy f_i^j is not correct as expected. For instance, if a trading strategy of a trend follower returns a "buy" action at time t, and most agents really buy at that time and thus push the price up, the payoff will be positive because the trading strategy has made a successful prediction; otherwise the payoff will be negative. This performance assessment function thus models the learning process of the agent from the history data, based on her belief. The best trading strategy of the agent at time t is defined as

$$f_i^* = \mathrm{argmax}_{f_i^j} U(f_i^j, t)\,, \tag{3.2}$$

indicating the preference of the agent over her trading strategies at any time t.

A trading strategy always predicts a buy or a sell action. But an agent can choose to do nothing when she is not confident enough with the success rate of the best trading strategy. This idea was introduced by Jefferies et al. (2001), and this ingredient can help model the liquidity in the real stock markets. Without this ingredient, there is always enough liquidity in our ABM's, while this is not always the case in for real financial markets. This observation is captured in virtual stock markets equipped with this ingredient, which can lead at some times to a drying up of liquidity when a large fraction of agents are uncertain on the performance of their s strategies and do not trade.

The success rate of a trading strategy during a time period of length T is defined as the ratio of the number of times that the trading strategy generates positive payoffs to the number of period length$T$:

$$\mathrm{sr}(f_i^j) = \frac{1}{T} \sum_{\zeta=t-T}^{t-1} \mathbf{1}_{R^+}(\pi(f_i^j(\mu_\zeta)))\,, \tag{3.3}$$

where $\mathbf{1}_{R^+}$ is an indicator function, which is equal to 1 when the payoff is positive and 0 otherwise. The agent chooses to do nothing when $\mathrm{sr}(f_i^*) < \tau$, where τ is a threshold for the agent to take actions according to the predictions of the best trading strategy. Thus, the agent's choice $a_i^t(\mu_t)$ at t can be summarized by

$$a_i^t(\mu_t) = \begin{cases} f_i^*(\mu_t) & \text{if } \mathrm{sr}(f_i^*) \geq \tau \\ 0 & \text{if } \mathrm{sr}(f_i^*) < \tau \end{cases} \quad \forall t \in \{1, \ldots, Z\} \tag{3.4}$$

To sum up, our ABMs use the following ingredients.

1. N bounded rational agents trade a single virtual asset.
2. The agents are endowed with sufficient initial wealth so that they will not go bankrupt and trade until the end of the game, i.e., over Z discrete time periods.
3. The agents make trading decisions based on the history of price

change directions.

4. The agents have a limited memory of length m, which is the same for all agents.

5. The agents use trading strategies to make decisions. Each agent has the same number s of trading strategies, but different agents have in general different trading strategies.

6. The agents make decisions by learning from the historical performance of their trading strategies. They assess the performance of their trading strategies by using equation (3.1), and choose the best trading strategies according to equation (3.2).

7. The agents further check the success rates of the best trading strategies, and decide to trade or not based on equation (3.4).

The parameters N, m, s, τ, T, as well as the trading strategy sets of the agents are thus the key parameters of our ABM's that we will be aiming at estimating.

3.2.3 Price formation from the collective actions of agents

The aggregate actions over all agents in the virtual market is defined as

$$A_t = \sum_{i=1}^{N} a_i^t(\mu_t) \,. \tag{3.5}$$

If A_t is positive, there are more buyers than sellers and the price will go up and the opposite when A_t is negative. We use a linear response function of the return as a function of the aggregate variable A_t:

$$r_t = \frac{A_t}{\lambda} \,, \tag{3.6}$$

where λ is a normalization factor called liquidity. There are some theoretical justification for such a linear relation (Kyle, 1985; Farmer, 2002) together with an on going debate on its precise validity versus the existence of nonlinear impact functions (Lillo et al., 2003; Almgren, 2003; Bouchaud et al., 2006; Farmer et al., 2012). For our purpose, we stay out of this debate and stick to the simple linear impact function (3.6).

3.2.4 Beliefs of agents: mixing of the 4 games

We use a mixture of four games to model the heterogeneous beliefs among the agents: the minority game, the majority game, the \$-game, and the delayed minority game. We assume each agent holds an invariant belief, though the agents can have different beliefs. The fraction of agents holding one of the four beliefs, i.e. making their choices based on one of the four payoff functions, is assumed to be fixed during the whole lifetime of a given virtual stock market.

The four games are characterized by four different payoffs.

1. *Minority game*: the payoff function of trading strategies for agents obeying the minority game rule is proportional to minus the product of their actions and of the aggregate actions of all agents:

   $$\pi^{\mathrm{mg}}(f_i^j(\mu_t)) = -\kappa f_i^j(\mu_t) A_t \,, \tag{3.7}$$

   where $\kappa > 0$ is a normalization constant. Equation (3.7) means that if the j-th trading strategy of the agent i at time t makes a trading decision that is different from the collective actions of all agents, the payoff of that trading strategy is positive; otherwise, the payoff is negative. Hereinafter, we use the suffix mg to denote variables related to the minority game.

2. *Delayed minority game*: It is a one-time step delayed minority game. Minority game players can be considered as moderate fundamentalists, while the delayed minority game players are more radical: a delayed minority game player expects that the choice of the majority at time $t+1$ will be opposite to her choice at time t, i.e. she believes that the majority will push the stock price away from its fundamental value at time $t+1$. The corresponding scoring function of trading strategies of a delayed minority game playing agent is

   $$\pi^{\mathrm{dmg}}(f_i^j(\mu_t)) = -\kappa f_i^j(\mu_t) A_{t+1} \,. \tag{3.8}$$

   Hereafter, we use the suffix dmg to denote variables related to the delayed minority game.

3. *The majority game*: An agent playing the majority game believes that she can profit from an upward or a downward trend in stock prices. She mimics the behavior of the majority. The corresponding payoff function of trading strategies of an agent i playing the majority game is

   $$\pi^{\mathrm{majg}}(f_i^j(\mu_t)) = \kappa f_i^j(\mu_t) A_t \,, \tag{3.9}$$

   which is the exact opposite to the payoff (3.7) for the minority game. Hereafter, we use the suffix majg to denote variables and expressions related to the majority game.

4. *The \$-game*: it is a one time-step delayed majority game. The \$-game players are smarter trend followers than the majority game players, as the latter mimic decisions of the majority blindly. Instead of expecting they are among the majority at time t, the \$-game playing agents anticipate that at time $t + 1$ the majority will make the same decision as they have made at time t, i.e. they want to predict trends. The corresponding scoring function for a \$-game playing agent is

$$\pi^{\mathrm{dg}}(f_i^j(\mu_t)) = \kappa f_i^j(\mu_t) A_{t+1} \,. \tag{3.10}$$

In the following, we use the suffix dg to denote variables related to the \$-game.

3.3 Calibrating ABM's and empirical results

3.3.1 Calibrating ABM's

In an accompanying paper (Zhang et al., 2013), we show that the mixed-game virtual stock markets constructed by combining the four ABM are able to reproduce the main stylized facts of real financial markets (Cont, 2001; Chakraborti et al., 2011). This is encouraging because not many models are able to account for not just one but many distinct empirical traits of financial returns. Indeed, explaining empirical observations is a first requirement for any model aiming at providing insights on the inner mechanisms of financial markets. Most models stop at this level however, content to provide some story on the reported anomalies or paradoxes (Hommes, 2002; Sornette and Zhou, 2006; Parisi et al., 2013). However, as explained in an ambitious research carried out in Los Alamos National Laboratories aimed at validating the US nuclear stewardship program (Sornette et al., 2007, 2008), a truly convincing validation approach needs to include a systematic and continuous process of developing new tests and new predictions on phenomena as different as possible from the stylized facts that have been tested in a first phase. The present paper follows this strategy by using the mixed-game virtual stock markets to reverse-engineer financial systems and, in this way, diagnose and predict both different regimes and financial returns.

Our approach to calibrate our ABM's with real time series extends that developed by Wiesinger et al. (2012), which itself extended (Andersen and Sornette, 2005). We improve on these previous works in particular by including in the estimation procedure the four fractions of players that are active in each of the four constitutive games, while Wiesinger et al. (2012) postulated fixed values. Indeed, there is a lot of evidence supporting the hypothesis that financial markets are characterized by regime shifts, with changing types of investment styles as a function of the central bank monetary policy (increasing or decreasing interest rates) and macroeconomic conditions. For ABM, we refer in particular to Lux and Marchesi (1999) who stressed first the importance of including time varying fractions of investing styles in order to account for the stylized facts of financial returns. Another important improvement of our present study compared with previous ones is to work with much larger data sets, so that our results are much more statistically robust.

To reverse-engineer financial markets with our mixed-game virtual stock markets, we proceed as follows. First, the input is chosen as a time series of returns, here the daily returns of several well known financial indexes. The descriptive statistics of the time series we reverse engineer are presented in table 8. For a given financial time series, we select an arbitrary window of length W_{is} that we refer to as the in-sample window, and calibrate the mixed-game virtual stock market model to the returns in that window. The calibration is performed by solving the following optimization problem:

$$\text{minimize: } \sum_{t=0}^{W_{is}} (r_t^{\mathrm{r}} - r_t^{\mathrm{abm}})^2 \,, \tag{3.11}$$

where the minimization is performed over the set of five parameters (number of agents $N \in \{3, \ldots, 103\}$, memory length $m \in \{2, \ldots, 8\}$, number of strategies per agent $s \in \{1, \ldots, 16\}$, threshold for action $\tau \in [0, 1]$ and duration of scoring counter $T \in \{1, \ldots, 25\}$) and over the set of all strategies defined within the four games (minority, delayed minority, majority and dollar games). In expression (3.11), r_t^{r} is the return of the input real financial index at time t and r_t^{abm} is the return of the mixed ABM at the same time t. The optimization (3.11) amounts to find the mixed game model that best replicates the returns of the real stock markets, given the same price information history within the in-sample window.

The technical procedure to solve the optimization problem (3.11) is explained in (Wiesinger et al., 2012). Since the problem is highly underconstrained and a priori ill-conditioned, such a fitting exercise will always return good fits, but with no insurance that the calibrated parameters and strategies provide any real value or insight. Therefore, it is essential to complement the optimization step (3.11) in the in-sample window with out-of-sample tests. Specifically, we use the parameters and strategies determined in the in-sample window to run the mixed game and predict the returns in the out-of-sample window just following the in-sample window, which we choose of length $W_{os} = 16$ days. The results reported below do not change significantly when changing the duration W_{os} of the out-of-sample window, as long as it is not too large. This value $W_{os} = 16$

days is a compromise between (i) having sufficient daily returns in the out-of-sample window to get statistically significant results and (ii) be not too far away from the in-sample window so that the parameters and strategies are still relevant. This second requirement is dictated by the fact that there is no hope that our mixed-game model could be the genuine generating process of the real financial time series. It can only be an approximation or imperfect representation, like a local tangent projective approximation of the complex unknown generating process. Such local tangent projective representation requires a periodic re-calibration of the model, in the same way that the tangent to a nonlinear curve evolves with the position on the curve. The tangent provides a useful informative representation in the neighborhood of the point of estimation (here a trend) but fails to extrapolate non-locally. Our mixed-game model is arguably better than just a linear model, accounting for the highly nonlinear strategies used by agents, but the general argument still holds that even the best nonlinear model has to be calibrated again beyond a certain horizon of validity.

Finally, we apply the above procedure to many in-sample windows and their associated out-of-sample windows within the chosen time series. For each time series in table 8, we perform 100 reverse engineering experiments. The in-sample window lengths W_{is} are changing from 40 to 400 days, with a step of 20 days, while the out-of-sample window length W_{os} is fixed to 16 days. This defines 700 experiments in total, in which every specific in-sample data window length appears twice. The value of the whole calibration process of our mixed-game models is quantified by how well are predicted the returns in the out-of-sample windows. We present several statistical tests to ascertain the value of this procedure, which are presented in the sequel.

3.3.2 Success rates in predicting the direction of future price changes

Our first statistical test compares the signs of the predicted returns with the signs of the realized returns in the out-of-sample windows. For this, we define the 'success rate' as the ratio of the correctly predicted return signs to the total number of realized returns (which is the same as the duration measured in days) in all out-of-sample windows. Notwithstanding the slight boundary effect introduced by the first in-sample window, the 'success rate' is essentially the fraction of return signs successfully predicted by our mixed game.

In order to interpret how good are our 'success rates', we compare them with those obtained by 1000 random strategies for each out-of-sample window. A given random strategy is defined as follows. It predicts a positive return with probability f_+ and a negative return with probability

Tab. 8: Descriptive statistics of the time series used in our reverse engineering process. The "Start time" and "End time" use the "Year-Month-Day" format, the "Mean" column gives the average returns, the "Sd" column lists the standard deviations of returns, the "Median" column shows the median values of the returns, the "Min" column gives the minimum returns, and the "Max" column lists the maximum returns.

Index	Start time	End time	Mean	Sd	Median	Min	Max
S&P500	1992-01-02	2001-12-31	0.0004	0.0100	0.0004	-0.0711	0.0499
S&P500	2002-01-02	2011-12-30	0.0001	0.0140	0.0008	-0.0947	0.1096
Nasdaq	1992-01-02	2001-12-31	0.0004	0.0255	0.0015	-0.7001	0.1720
Nasdaq	2002-01-02	2011-12-30	0.0004	0.0151	0.0011	-0.1111	0.1185
Dow Jones	1982-01-04	1991-12-31	0.0005	0.0117	0.0004	-0.2563	0.0967
Dow Jones	1992-01-02	2001-12-31	0.0004	0.0099	0.0006	-0.0745	0.0486
Dow Jones	2002-01-02	2011-12-30	0.0001	0.0130	0.0005	-0.0820	0.1051

$1 - f_+$, where f_+ is the observed fraction of days with positive returns in the corresponding real financial time series. Note that, while the random strategies toss random coins, they are actually endowed with the hindsight of using the realized value of f_+ for the whole time series, an information that is not available in the real-time forecasting set-up implemented in our truly out-of-sample causal prediction scheme. Therefore, the result presented below that our mixed-game predictions are significantly better than most random strategies can be considered as conservative. In other words, the reported p-values (fractions of random strategies performing better than our mixed game) can be considered as upper bounds.

We have run 700 different experiments, 100 for each of the seven 10 year long financial time series shown in table 9. Each experiment is defined by its in-sample window and corresponding 16-days out-of-sample window where return predictions are compared with realized returns. Out of the 700 experiments, we find that only 46 of them (6.6%) have p-values larger than 0.1. In addition, except for the daily returns of the Dow Jones index from 1982 to 1991 and from 1992 to 2001, the other five indices have no more than 5% of their experiments underperforming the random strategies at the 90% confidence level. Table 9 shows the number ("Count") of experiments out of 100 of them for each financial time series with insignificant prediction power at the 90% confidence level. Note that half of the experiments with insignificant prediction power occur for the Dow Jones index from 1982 to 1991, for reasons unknown to us.

Another result is that, in 109 out of the 700 experiments (15.6% of all the experiments), the success rate is higher than the ratio f_+ of positive

Tab. 9: Numbers of experiments ("Count") with insignificant prediction power at the 90% confidence level. For each of the seven indices, 100 experiments have been performed and their corresponding 'success ratios' have been compared with those of 1000 random strategies. See main text for explanations.

Index	Start time	End time	Count
S&P500	1992-01-02	2001-12-31	4
S&P500	2002-01-02	2011-12-30	0
Nasdaq	1992-01-02	2001-12-31	3
Nasdaq	2002-01-02	2011-12-30	5
Dow Jones	1982-01-04	1991-12-31	23
Dow Jones	1992-01-02	2001-12-31	7
Dow Jones	2002-01-02	2011-12-30	4

returns in the real time series used to implement the random strategies. This success rate may be considered as a natural benchmark, since it is obtained by construction by the static buy-and-hold strategy. Since the buy or sell decision of the random strategies for any given day is independent of the sign of the return on that day, the theoretical average success rate of random strategies is equal to f_+^2, which is of the order of 0.30 for the typical values $f_+ \approx 0.55$ observed in the studied financial time series, as seen in table 10. No random strategy in our sample of thousands of trials can achieve a success rate of f_+ or larger, as found in 15.6% of our experiments.

There are many other tests that we have performed that confirm the significance of the predictive power of the mixed games. Table 10 presents a subset of the results by showing the results of 25 experiments.

In summary, these different results suggest that our reverse-engineered mixed-games have true predictive power, much beyond what can be attributed to chance.

3.3.3 Trading strategies based on the ABM's

Strategies can achieve statistically abnormally large success rates, while still not being in contradiction with the no-arbitrage principle and the efficient market hypothesis, if these strategies fail to provide a positive abnormal risk-adjusted return. We thus investigate the profits-and-losses (P&L) properties of our strategy based on buying (respectively selling) at the open of a given day when our method predicts a positive (respectively negative) return for the day, and closing the position at the end of the day. We compare the P&L of our strategy with that of random strategies that buy (respectively sell) at the opening of each day randomly with a probability b (respectively with a probability $1 - b$), and close their positions at the end of the day. We scan b uniformly between 0 and 1. For each time window, we generate 100 different b's ($b_1, \ldots, b_{100}$) and generate 3000 random strategies for each b, leading to a total of 300,000 random strategies. This generalizes the random strategies considered in the previous subsection on the success rates, by including both the buy-and-hold ($b = 1$) and the sell-and-hold ($b = 0$) strategies, as well as a continuum of intermediate strategies. Transaction costs are not included.

We compare our strategy constructed from the 700 experiments with the random strategies in terms of two indicators of performance: the total P&L and the Sharpe ratio (with zero risk-free interest rate). We find that more than 15% of the 700 experiments can generate good total P&L, in the sense that none of these P&L values can be achieved by 90% of the random strategies. Similarly, 15% of the experiments can create significant Sharpe ratios, which cannot be achieved by 90% of the random strategies. In the same way, if we set the significance level to 95%, we find 6.78% of the 700 experiments outperform the random strategies, and when we set the significance level to 99%, we find 1.6% of the 700 experiments outperform the random strategies. These fractions are higher than can be achieved using random strategies. For instance, at the significance level 90%, we should find only 10% random strategies that are better than our ABM based strategies compare with the fraction 15% found among our 700 experiments.

The question is then whether the fractions 15% versus 10%, or 6.78% versus 5% or 1.6% versus 1% are due to statistical fluctuations or not. To answer this question, we construct a statistical test based on the null hypothesis that the 700 experiments perform on average identically to the random strategies. The corresponding alternative hypothesis is that the ABM based strategies perform better. If the null hypothesis is true, for a given significance level $1-\alpha_+$ the number of the ABM based strategies that outperform the random strategies is a binomial distribution $B(700, \alpha_+)$. Let $\alpha_+ = 0.1$. The p-value associated with the observation that 15% of the experiments outperform the random strategies is 0. For $\alpha_+ = 0.05$, the p-value of getting 6.78% better strategies is 0.023, and the p-value of getting 1.6% better strategies for $\alpha_+ = 0.01$ is 0.066. These p-values suggest that we can reject the null hypothesis, and we thus accept the alternative hypothesis that the 700 experiments on average perform better than the random strategies.

To ascertain further the statistical significance of our results, we perform other tests. We regress the time series of returns generated by our strategy on the three Fama-French factor model (Fama and French, 1993), to test for the existence of significant abnormal positive return α (defined as the intercept). We find that 21 out of the 700 experiments have significant positive α. This proportion seems to be very small but actually hides a genuine skill as we now explain. The returns of the indices that the strategies trade (going "long" or "short" on them) are well explained by the three factor model taken from (Fama and French, 1993):

$$R(t) - RF(t) = a + b[RM(t) - RF(t)] + sSMB(t) + hHML(t) + e(t) \tag{3.12}$$

In this equation (3.12), the intercept a is the α quantifying, if positive. the excess performance. Our ABM based strategies are usually long or short the indices roughly 50% of the time. Calling f_l the fraction of time where one of our strategy is long an index and $1 - f_l$ the fraction of time where that strategy is short the index, the average return of that strategy, if no skill is present, should be given by

$$\begin{aligned} E(R^{\mathrm{ABM}}(t) - RF(t)) &= (2f_l - 1)a + (2f_l - 2)E(RF(t)) \\ &+ (2f_l - 1)bE[RM(t) - RF(t)] \\ &+ (2f_l - 1)sE(SMB(t)) + (2f_l - 1)hE(HML(t))\,. \end{aligned} \tag{3.13}$$

Because $f_l \leq 1$ and actually close to 0.5 in general, $2f_i - 1$ is quite smaller than 1, which would thus even tend to reduce significantly the strategy's α below that of the indices. Actually, the story becomes even more interesting when one realizes that S&P500 and Dow Jones Indices have significant *negative* α's (ranging from about -0.009% to -0.02% for daily return in the data we use for this paper). Launching 20,000 random strategies for each of our 700 experiments, we find that only less than 0.1% of them can achieve the same or slightly better results than 3% of our ABM bases strategies (the 21 out of of the 700). Thus, while small, the finding that 21 out of the 700 experiments have significant positive α is highly significant. We note that this line of reasoning is similar to the ensemble approaches of Romano and Wolf (2005) and Barras et al. (2010).

Table 11 summarizes our results for representative trading strategies, each based on trading signals from a single reverse engineering experiment and without considering the transaction costs. These representative trading strategies are among the 15% that outperform the random strategies at the significance level 90%.

Table 11 shows that the strategies can generate not only statistically significant but economically significant returns. The excess fractions of such strategies in the 700 experiments thus give us a way to arbitrage. Since the strategies we test are obtained by a reverse-engineering of mixed games using only financial market returns, i.e., history price information, our results suggest a rejection of the weak form of the efficient market hypothesis.

3.3.4 Structures of ABM's and market regimes

Our mixed-game virtual stock market dynamics are generated by the interplay between four types of agents, obeying respectively minority, delayed minority, majority and dollar games. As the relative proportion of agents of different types changes, the nature of the resulting price dynamics will also exhibit distinct properties. For instance, if most agents in a mixed-game virtual stock market play the majority game or the $-game, the chance that the mixed-game virtual stock market generates a bubble or a crash will be high, because majority trading captures the tendency for traders to herd. Changing the threshold τ for trading has also a significant impact on the properties of the generated price dynamics.

If, as suggested by the previous results, our mixed-game virtual stock markets constructed by combining the four ABM are good models of real financial market returns, there should be a relationship between the parameters characterizing the mixed-game models and the realized financial returns. In other words, the calibration of our mixed-game virtual stock market models to the real data offers the possibility of identifying the existence of distinct market regimes, the switching times between them, and their characteristics in terms of the key parameters defining the mixed-games. In addition, deterministic relations between the real returns and the parameters of the reverse-engineered virtual stock markets would provide further evidence supporting the relevance of the ABM and shed light on the main mechanisms at work during different market regimes.

Specifically, following exactly the same procedure of reverse engineering calibration of our ABM on real financial time series as explained above, we regress the real returns in the out-of-sample data window on the calibrated parameters of the mixed-game virtual stock market. The linear regression models read

$$\hat{r} = a_0 + a_1 \hat{fr}_{\text{act}} + \epsilon\,, \tag{3.14}$$

$$\dots\,,$$

$$\hat{r} = a_0 + a_6 \hat{fr}^{\text{dmg}} + \epsilon\,, \tag{3.15}$$

$$\hat{r} = a_0 + a_1 \hat{fr}_{\text{act}} + a_2 \hat{\tau} + \epsilon\,, \tag{3.16}$$

$$\hat{r} = a_0 + a_1 \hat{fr}_{\text{act}} + a_3 \hat{fr}^{\text{mg}} + \epsilon\,, \tag{3.17}$$

$$\dots\,,$$

$$\hat{r} = a_0 + a_1 \hat{fr}_{\text{act}} + a_6 \hat{fr}^{\text{dmg}} + \epsilon\,, \tag{3.18}$$

$$\dots\,,$$

$$\dots\,,$$

$$\hat{r} = a_0 + a_1 \hat{fr}_{\text{act}} + a_2 \hat{\tau} + a_3 \hat{fr}^{\text{mg}} + a_4 \hat{fr}^{\text{majg}} + a_5 \hat{fr}^{\text{dg}} + \epsilon\,, \tag{3.19}$$

$$\dots\,,$$

$$\hat{r} = a_0 + a_1 \hat{fr}_{\text{act}} + a_2 \hat{\tau} + a_4 \hat{fr}^{\text{majg}} + a_5 \hat{fr}^{\text{dg}} + a_6 \hat{fr}^{\text{dmg}} + \epsilon\,, \tag{3.20}$$

where $\hat{r}$ is the average return in an out-of-sample window or multiple out-of-sample windows, $\hat{fr}_{\text{act}}$ is the average fractions of active agents[1], $\hat{\tau}$ is the average τ, and $\hat{fr}^{\text{mg}}$, $\hat{fr}^{\text{majg}}$, $\hat{fr}^{\text{dg}}$ and $\hat{fr}^{\text{dmg}}$ are the average fractions of agents who play the minority game, the majority game and the $-game and the minority game, respectively. The parameters to estimate are the coefficients a_i's, and ϵ is a white noise. Because the fractions $\hat{fr}^{\text{mg}}$, $\hat{fr}^{\text{majg}}$, $\hat{fr}^{\text{dg}}$ and $\hat{fr}^{\text{dmg}}$ are linearly correlated, we actually use a variant of the regressions (3.16-3.20) with residuals of the mutual regressions between the fraction variables $\hat{fr}^{\text{mg}}$, $\hat{fr}^{\text{majg}}$, $\hat{fr}^{\text{dg}}$ and $\hat{fr}^{\text{dmg}}$ themselves used as the independent variables. The use of an average τ and of average agent fractions is associated with the fact that, in many cases, we regress also the returns averaged over multiple out-of-sample windows. The regressions (3.14-3.20) use the variables $\hat{fr}_{\text{act}}$, $\hat{\tau}$, $\hat{fr}^{\text{mg}}$, $\hat{fr}^{\text{majg}}$, $\hat{fr}^{\text{dg}}$ and $\hat{fr}^{\text{dmg}}$ because they are most relevant and also because they are found to be stationary within each in-sample window. Because we do not know ex-ante which independent variables can better explain the average returns, we include all combinations of variables in regressions (3.14-3.20) and choose the best ones according to the estimation results.

We use the AIC criteria Akaike (1974) to determine the best model among regressions (3.14-3.20). Estimating the parameters of these regressions for the 700 reverse engineering experiments, we find a lot of statistically significant results. Since it is impossible to report all the results here and because it is difficult to distinguish clear trends in such a massive set of regressions, we apply a coarse-grained method by using averages of multiple out-of-sample windows and stack all 100 experiments performed on the same index together. Moreover, we divide the time series of the indexes from 1982 to 2012 into 6 different regimes, and report the estimation results of the best models among regressions (3.14-3.20) in the coarse-grained representation for each of the market regime separately. Specifically, we classify six major regimes from 1982 to 2012 [2] within which we analyze the performance of the ABM based strategies:

---

[1] with a positive τ, not all agents will trade, so that the reverse engineering approach allows determines the fraction of active agents.

1. **1982 - Oct. 1987**. Overall decreasing Fed rates from a very high level to fight the inflation era of the 1970s to a low around 4.75% in 1993, punctuated by two spikes. The second of these spike was an attempt by the Fed to slow down the bubble developing from 1982 to Oct 1987, date at which a great worldwide crash occurred, with the US markets specifically crashing on black Monday 19 Oct 1987.

2. **Oct. 1987 - 1993**. Post bubble regime.

3. **1993 - 2000**. Flat or increasing Fed rate and the growth of the dot-com bubble.

4. **2000 - 2003**. Aggressive decrease of the Fed rate to fight the recession associated with the burst of the dot-com and biotech bubble.

5. **2003 - Oct. 2007**. Flat followed by slow increase of the Fed rate, jointly with the development of a set of co-inflating bubbles occurring in many different asset classes, including real-estate, oil, soft commodities, stock markets and financial derivatives. This global leverage bubble has prepared the economy for the “great recession" (Sornette and Woodard, 2010; Sornette and Cauwels, 2012).

6. **End 2007 - Present**. Lowered Fed rate and stabilization at almost zero value, occurrence of the great recession and successive phases of quantitative easing (QE1, QE2, Twist and QE-infinity).

Table 12 presents the results of the estimation of the best models among regressions (3.14-3.20). The dependent variable, namely the average returns, are calculated from 16 adjacent out-of-sample data windows, i.e. corresponds to averages over 256 data points. The independent variables are also calculated from 16 mixed-game virtual stock markets, which are calibrated in 16 in-sample data windows and used to predict the 16 out-of-sample windows. The independent and dependent variables of the 100 reverse engineering experiments performed for each given index time series are stacked together to estimate the regressions (3.14-3.20) for each given market regime, when the corresponding market regime indeed occurs within the considered experiment. Some cells of table 12 remain empty, which means that the corresponding variable is not included in the estimation. One can observe that variable $\hat{fr}^{\mathrm{dmg}}$ has never been selected as being significant by the AIC criterion.

---

[2] as a reference, figure 5 shows the prices dynamics of the indexes.

The very small p-values of the F tests presented in table 12 indicate the high statistical significance of the estimated coefficients of regression (3.14-3.20). For each market regime, we find highly significant linear relationships between the real returns and the calibrated ABM parameters. Moreover, the R^2's show that the parameters of ABM's can explain between approximately 10 to 30 percent of the variances of index returns. Concerning the threshold variable $\hat{\tau}$, we find that it is significant during the bubble regimes, while it is insignificant after crashes. Accordingly, the variable $\hat{fr}_{\mathrm{act}}$ is usually significant and positively related to index returns during a bubble regime, while it is usually insignificant or negatively related to indexes after crashes. However, we cannot verify the causal relationships between the dependent and independent variables from these results, because they are averages over multiple out-of-sample windows. Nevertheless, these two variables can still be diagnostic variables of market regimes.

Another interesting result is that the linear relations between index returns and the variables $\hat{fr}^{\mathrm{majg}}$ and $\hat{fr}^{\mathrm{dg}}$ are stronger than those between index returns and the variable $\hat{fr}^{\mathrm{mg}}$. This can be rationalized by the fact that the majority game and the \$-game capture very well the mechanism of herding, usually associated with bubbles and crashes. These two variables, however, cannot tell us what they will bring, bubbles or crashes. In future studies, it would be interesting to investigate more deeply the trading strategies of agents playing those two games and check the possibilities that they either create bubbles or crashes.

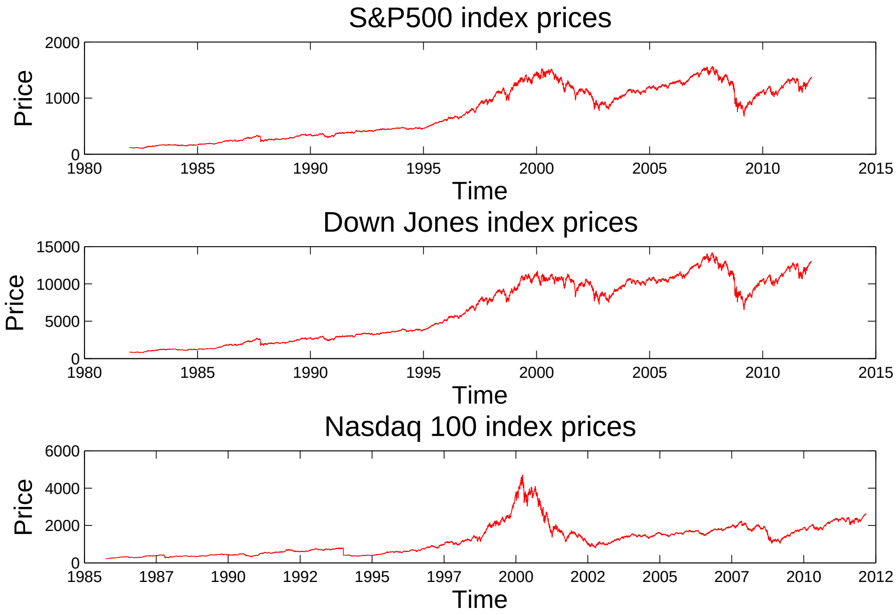

Fig. 5: Prices dynamics of the S&P500, Dow Jones Industrial Average, and Nasdaq 100 indexes from 1982 to 2012.

Tab. 10: Results of 25 experiments, using the reverse-engineered mixed games to predict the sign of future returns in out-of-sample windows of 16 trading days. **Index** lists the names of the indexes, **Start year** lists the start year of the corresponding time series, **End year** lists the end year of the time series, **Data points** list the numbers of data points predicted by the ABM's, **Positive ratio** lists the ratio of positive returns in the real time series, W_{is} lists the in-sample window sizes of the experiments, **Success rate** lists the total success rates of the ABM's, and **p-value** lists the p-values of the prediction power tests. In all experiments, the out-of-sample window sizes of the experiments is fixed at $W_{os} = 16$ days. For all experiments, the p-values are much smaller than 0.001.

Index	Start year	End year	Data Points	Positive ratio	W_{is}	Success rate
S&P500	2002	2011	2080	0.550	380	0.556
Nasdaq	1992	2001	2352	0.546	120	0.556
Nasdaq	1992	2001	2144	0.549	320	0.555
S&P500	2002	2011	2224	0.552	240	0.554
S&P500	2002	2011	2256	0.550	200	0.553
Nasdaq	1992	2001	2304	0.545	160	0.553
S&P500	2002	2011	2288	0.548	180	0.552
S&P500	2002	2011	2144	0.551	320	0.552
S&P500	2002	2011	2128	0.553	340	0.552
S&P500	2002	2011	2240	0.550	220	0.552
S&P500	2002	2011	2176	0.551	280	0.551
S&P500	2002	2011	2336	0.547	120	0.551
S&P500	2002	2011	2160	0.552	300	0.550
Nasdaq	1992	2001	2064	0.548	400	0.550
Nasdaq	2002	2011	2176	0.547	280	0.550
S&P500	2002	2011	2208	0.552	260	0.549
Nasdaq	2002	2011	2176	0.547	280	0.549
Nasdaq	1992	2001	2304	0.545	160	0.549
S&P500	2002	2011	2240	0.550	220	0.549
Nasdaq	2002	2011	2224	0.547	240	0.549
Nasdaq	1992	2001	2240	0.543	220	0.548
Dow Jones	2002	2011	2208	0.538	260	0.548
S&P500	2002	2011	2160	0.552	300	0.547
S&P500	2002	2011	2224	0.552	240	0.547
Nasdaq	2002	2011	2144	0.546	320	0.547

Tab. 11: Summary performance of trading strategies constructed on trading signals from a single reverse engineering experiment and without considering the transaction costs. **Index**'s are the names of the indexes, **Start year**'s are the start year of the corresponding time series, **End year**'s are the end year of the time series, **Data points**'s are the numbers of data points used in the ABM prediction process, W_{is}'s are the in-sample window sizes of the experiments, **Total returns**'s and **Sharpe ratios**'s are the total returns (P&L) and annualized Sharpe ratios of the ABM's based trading strategies respectively, PV_r's and PV_{shr}'s are the p-values of testing the returns and Sharpe ratios of the ABM's based trading strategies against those of the random trading strategies respectively, **Annual return**'s are the annual returns of the ABM's based trading strategies (for instance, 0.601 corresponds to an annual return of 60.1%) , and the α's are obtained as the intercepts of the regression of the time series of returns of the ABM's based trading strategies as a function of the three Fama-French factors model. For all experiments, the out-of-sample window size is fixed to $W_{os} = 16$ days. The parentheses give the standard errors of the corresponding α's. "*" indicates that the marked α is statistically significant with a p-value less than 0.1, "**" indicates a p-value less than 0.05, and "***" indicates a p-value less than 0.01.

Index	Start year	End year	Data points	W_{is}	Total return	Sharpe ratio	PV_r	PV_{shr}	Annual return	α
S&P500	1992	2001	2128	340	5.117	1.287	0.00	0.00	0.601	$0.045(0.018)^{**}$
S&P500	1992	2001	2112	360	4.869	1.254	0.00	0.00	0.576	$0.043(0.018)^{**}$
S&P500	1992	2001	2112	360	4.159	1.178	0.00	0.00	0.492	$0.047(0.021)^{**}$
S&P500	1992	2001	2192	280	4.284	1.151	0.00	0.01	0.489	$0.028(0.013)^{**}$
S&P500	1992	2001	2192	280	4.027	1.132	0.00	0.01	0.459	$0.040(0.017)^{**}$
Dow Jones	1982	1991	2416	60	5.726	1.064	0.00	0.02	0.592	$0.043(0.022)^{**}$
S&P500	1992	2001	2144	320	3.431	1.056	0.00	0.01	0.400	$0.023(0.014)^{*}$
Dow Jones	1982	1991	2384	80	5.227	1.027	0.00	0.02	0.548	$0.048(0.023)^{**}$
Nasdaq	1992	2001	2352	120	44.041	0.999	0.00	0.00	4.681	$0.137(0.051)^{***}$
Dow Jones	1982	1991	2256	220	4.192	0.998	0.00	0.02	0.465	$0.040(0.022)^{*}$
Nasdaq	1992	2001	2400	60	39.265	0.953	0.00	0.00	4.090	$0.126(0.051)^{**}$
Nasdaq	1992	2001	2128	340	23.586	0.884	0.00	0.01	2.771	$0.116(0.057)^{**}$
Nasdaq	2002	2011	2176	280	4.482	0.818	0.00	0.02	0.515	$0.054(0.027)^{**}$
S&P500	2002	2011	2160	300	3.491	0.813	0.00	0.01	0.404	$0.049(0.024)^{**}$
Nasdaq	1992	2001	2304	160	18.265	0.783	0.00	0.02	1.982	$0.103(0.053)^{*}$
S&P500	2002	2011	2288	180	3.212	0.739	0.00	0.02	0.351	$0.040(0.024)^{*}$
S&P500	2002	2011	2288	180	2.819	0.689	0.00	0.03	0.308	$0.038(0.022)^{*}$
Nasdaq	2002	2011	2416	40	4.226	0.664	0.00	0.04	0.437	$0.053(0.032)^{*}$

Tab. 12: Linear regression (3.14) between realized financial returns and some of the parameters of the ABM's during different market regimes. **Index** lists the names of the indexes. **Start time** contains the start times in "year-month" format of the corresponding regimes, and similarly **End time** gives the end time of the regimes. **Data points** lists the number of data points of the dependent and independent variables. **Intercept** presents the coefficient a_0, i.e. the intercept of regression (3.14). $\hat{fr}_{\text{act}}$, $\hat{\tau}$, $\hat{fr}^{\text{mg}}$, $\hat{fr}^{\text{majg}}$, $\hat{fr}^{\text{dg}}$ and $\hat{fr}^{\text{dmg}}$ list the estimated parameters for the corresponding variables. **F-test** shows the p-values of the corresponding F tests. R^2 contains the related R-squares. The parentheses give the standard errors of the corresponding parameters, and "*" indicates that the estimated parameter is statistically significant with a p-value less than 0.1, "**" indicates a p-value less than 0.05, and "***" indicates a p-value less than 0.01.

Index	Start time	End time	Data points	Intercept	$\hat{fr}_{\text{act}}$	$\hat{\tau}$	$\hat{fr}^{\text{mg}}$	$\hat{fr}^{\text{majg}}$	$\hat{fr}^{\text{dg}}$	$\hat{fr}^{\text{dmg}}$	F-test	R^2
Dow Jones	1982-01	1987-10	500	-0.00767*** (0.00185)	0.00102 (0.00116)	0.00304* (0.00166)	0.00090 (0.00196)	0.00860*** (0.00198)	0.01404*** (0.00185)		0.00000	0.18026
Dow Jones	1987-12	1993-01	245	-0.00142 (0.00175)	-0.00095 (0.00110)	-0.00091 (0.00157)	-0.00159 (0.00172)	0.00408** (0.00185)	0.00909*** (0.00156)		0.00000	0.21214
S&P500	1993-01	2000-01	600	-0.00242*** (0.00094)	0.00141** (0.00061)	0.00134 (0.00087)	-0.00131 (0.00107)	0.00274*** (0.00106)	0.00649*** (0.00097)		0.00000	0.15835
Dow Jones	1993-01	2000-01	600	-0.00345*** (0.00102)	0.00211*** (0.00071)	0.00311*** (0.00101)	-0.00238** (0.00106)	0.00330*** (0.00118)	0.00531*** (0.00098)		0.00000	0.15619
Nasdaq	1993-01	2000-01	600	-0.02115*** (0.00260)	0.01444*** (0.00144)	0.02091*** (0.00215)	-0.00267 (0.00306)	0.00425 (0.00323)	0.01311*** (0.00309)		0.00000	0.20023
Dow Jones	2000-01	2003-01	36	0.00055 (0.00047)	0.00068** (0.00029)			-0.00480* (0.00242)			0.06792	0.15040
Nasdaq	2000-01	2003-01	36	-0.00021 (0.00137)	-0.00346** (0.00167)						0.04612	0.11192
S&P500	2003-01	2007-10	400	-0.00317** (0.00140)	0.00279*** (0.00088)	0.00400*** (0.00121)	0.00012 (0.00114)	-0.00255* (0.00137)	0.00090 (0.00136)		0.00017	0.06019
Nasdaq	2003-01	2007-10	400	-0.00239** (0.00118)	0.00232*** (0.00076)	0.00274** (0.00113)	0.00095 (0.00111)	-0.00055 (0.00120)	-0.00000 (0.00117)		0.00880	0.03818
Dow Jones	2003-01	2007-10	400	-0.00549*** (0.00124)	0.00486*** (0.00080)	0.00563*** (0.00110)	0.00156 (0.00118)	-0.00195 (0.00129)	0.00177 (0.00127)		0.00000	0.11857
S&P500	2007-12	2012-01	236	-0.00253 (0.00201)	0.00020 (0.00125)	0.00291 (0.00182)	0.00047 (0.00232)	0.00400* (0.00233)	-0.00034 (0.00254)		0.00000	0.21695
Nasdaq	2007-12	2012-01	236	0.00137 (0.00275)	-0.00280 (0.00169)	-0.00198 (0.00244)	-0.00186 (0.00255)	0.00538* (0.00293)	0.00394 (0.00285)		0.00008	0.10771
Dow Jones	2007-12	2012-01	236	-0.00158 (0.00215)	-0.00239* (0.00143)	-0.00042 (0.00203)	0.00443** (0.00218)	0.00605** (0.00245)	0.00172 (0.00252)		0.00000	0.31871

3.4 Conclusions

We have constructed virtual financial markets populated by artificial agents, who make decisions according four classes of backward-looking decision functions, with the goal of testing the weak form of the efficient market hypothesis (EMH). Our agent-based models (ABM) are populated by agents with bounded rationality and heterogeneous beliefs, which can be represented by the decision functions defining respectively the minority game, the majority game, the \$-game and the delayed minority game. Zhang et al. (2013) have shown that the players using the reward functions of the minority and delayed minority games behave approximately as fundamentalists, as they tend to minimize the imbalance between their buy and sell actions. In contrast, the majority game and the \$-game players behave as trend followers. The mixed-game virtual stock markets that we have constructed combine agents playing all of the four games, leading to realistic price dynamics exhibiting the standard stylized facts, including the transient existence of bubbles and crashes (this is reported in the accompanying paper (Zhang et al., 2013)). Using the mixed-games, we have calibrated our ABM's to 10 year long real financial index returns. We have extended a previous methodology and provide the main structural parameters,N, m, s, τ, T, the specific trading strategies used by the N agents , as well as the fractions of agents playing the four different games. This gives a genuine reverse-engineered reconstruction of the real financial markets. Using these calibrated mixed games on the 10-year time series of the S&P500, Dow Jones Industrial Average and Nasdaq 100 indexes from 1982 to 2012 in 700 experiments, we have assessed the performance of their predictions on future daily returns in out-of-sample time windows of 16 days. We found that 654 out of the 700 reverse engineering experiments generate statistically significant success rates of predicting the future return signs. We then developed trading strategies implementing the predictions of future return signs and found that many such strategies are both statistically significant and economically significant. The probability that one get the same number of such strategies randomly is 0. Random strategies, even when using the information on the fraction of positive returns, fail to account for our performance. Regression 21 time series of the returns generated by our strategies on the three factor Fama-French model and on the four factor Carhart-Fama-French show statistically significantly positive abnormal risk-adjusted returns α's. Regressing the returns generated by our strategies on the calibrated parameters of the ABM shows that the threshold parameter τ is significantly and positively related to indexes returns during bubble regimes, while the relation is insignificant after crashes. The fractions of active agents are also related to index returns in a similar way. The fractions of the majority game and the \$-

game players are more obviously related to index returns than the fractions of the minority games. The relation between the delayed minority game and index returns is always insignificant. These results suggest that the calibrated parameters of our ABM's can help us diagnose market regimes.

In conclusion, our results challenge the weak form of the efficient market hypothesis. Transient deviations from efficiency are mostly due to the role of trend followers as captured by the majority game and the $-game players. The behavior of these trend followers create positive feedback loops that have been shown elsewhere to be the engine of bubbles and crashes (Sornette and Zhou, 2006; Sornette, 2003).

4

Empirical test of the origin of Zipf's law in growing social networks

4.1 Introduction

Power law distributions,

$$p(s) \sim 1/s^{1+\mu} , \tag{4.1}$$

are ubiquitous characteristics of many natural and social systems. The function $p(s)$ is the density associated with the probability $P(s) = \Pr\{S > s\}$ that the value S of some stochastic variable, usually a size or frequency, is greater than s. Among power law distributions, Zipf's law states that $\mu = 1$, i.e., $P(s) \sim s^{-1}$ for large s. Zipf's law has been reported for many systems (Saichev et al., 2009), including word frequencies (Zipf, 1949), firm sizes (Axtell, 2001), city sizes (Gabaix, 1999), connections between Web pages (Kong et al., 2008) and between open source software packages (Maillart et al., 2008), Internet traffic characteristics (Adamic and Huberman, 2000), abundance of expressed genes in yeast, nematodes and human tissues (Furusawa and Kaneko, 2003) and so on. The apparent ubiquity and universality of Zipf's law has triggered numerous efforts to explain its validity. It is also essential to understand the origin(s) of Zipf's law.

Since H. Simon's pioneering work (Simon, 1955a, 1960; Ijri and Simon, 1977), a crucial ingredient in the generating mechanism of Zipf's law is understood to be Gibrat's rule of proportional growth (Gibrat, 1931), more recently rediscovered under the name of "preferential attachment" in the context of networks (Barabasi and Albert, 1999). Expressed in continuous time in terms of the size $S(t)$ of a firm, a city or, more generally, a social group, Gibrat's rule corresponds to the geometric Brownian motion

$$dS(t) = S(t)\,(r\,dt + \sigma\,dW(t)) \ , \tag{4.2}$$

where the stochastic growth rate $r+\sigma dW/dt$ is decomposed into its average r and its fluctuation part $\sigma dW/dt$ with an amplitude determined by the standard deviation σ, while $W(t)$ is a standard Wiener process. Gibrat's rule alone cannot produce (4.1), since the solution of equation (4.2) has a (non-stationary) log-normal distribution. Simon and many other authors invoked an addition ingredient, corresponding to various modifications of the multiplicative process when $S(t)$ becomes small. Then, under very general conditions, the distribution of S becomes a power law, with an exponent μ that is a function of the distribution of the multiplicative factors (Kesten, 1973; Sornette, 1998).

The fact that the exponent μ is often found close to 1 requires another crucial ingredient. One particularly intriguing proposition is that Zipf's law corresponds to systems that are growing according to a maximally sustainable path (Gabaix, 1999; Malevergne et al., 2010). In other words, the set of stochastically growing entities $\{S_i(t), i = 1, 2, ..., n, ..\}$ is delicately poised at a dynamical critical growth point. Within a general framework in which (i) entities are born at random times, (ii) grow stochastically according to (4.2), and (iii) can disappear or die according to various stochastic processes with some hazard rate h, the explicit calculation of the exponent μ confirms the above optimal growth condition associated with Zipf's law ($\mu = 1$) (Malevergne et al., 2010).

Here, we present an empirical test of the optimal growth condition for Zipf's law by testing the formula for exponent μ (see below) on a unique database obtained from a Web platform of collaborative social projects (Amazee.com). In this dataset, we verify empirically that proportional growth holds, we measure the parameters r, σ and h independently, and determine the exponent μ of the power law distribution of project sizes. We show that the theory leading to the maximum sustainable growth principle explains remarkably well the empirical value, with no adjustable parameters.

4.2 Theory and data

4.2.1 Summary of theoretical predictions

The theory is based on the following assumptions (Gabaix, 1999; Saichev et al., 2009; Malevergne et al., 2010). Consider a population of social groups (firms, cities, projects, and so on), which can take different forms and can be applied in many different contexts.

1. There is a flow of group entries, i.e., a sequence of births of new groups. The times $\{t_1, t_2, ..., t_i < ...\}$ of entries of new groups follow a Poisson process with constant intensity (generalizations do not modify the key result (Saichev et al., 2009)).

2. At time t_i, $i \in \mathcal{N}$, the initial size of the new entrant group i is a random variable $s_{0,i}$. The sequence $\{s_{0,i}\}_{i\in\mathcal{N}}$ is the result of independent and identically distributed random draws from a common random variable $\tilde{s}_0$. All the draws are independent of the entry dates of the groups.

3. Gibrat's rule of proportional growth holds. This means that, in the continuous time limit, the size $S_i(t)$ of the i^{th} group at time $t \geq t_i$, conditional on its initial size s_0^i, is solution to the stochastic differential equation (4.2), where the drift r and the volatility σ are the same for all groups but the Wiener process $W_i(t)$ is specific to each project i.

4. Groups can exit (disappear) at random, with constant hazard rate $h \geq 0$, which is independent of the size and age of the group.

Under these conditions, the central result of Malevergne et al. (2010) reads as follows.

Proposition 1: *Defining*

$$\mu := \frac{1}{2}\left[\left(1 - 2\cdot\frac{r}{\sigma^2}\right) + \sqrt{\left(1 - 2\cdot\frac{r}{\sigma^2}\right)^2 + 8\cdot\frac{h}{\sigma^2}}\right], \tag{4.3}$$

provided that $\mathrm{E}\left[\tilde{s}_0^{\mu}\right] < \infty$, *and for times much larger than*

$$t_{\text{transient}} = \left[\left(r - \frac{\sigma^2}{2}\right)^2 + 2\sigma^2 h\right]^{-1/2}, \tag{4.4}$$

the average distribution of project's sizes follows an asymptotic power law with tail index μ *given by (4.3), in the following sense: the average number of projects with size larger than* s *is proportional to* $s^{-\mu}$ *as* $s \to \infty$.

The condition $\mathrm{E}\left[\tilde{s}_0^{\mu}\right] < \infty$ just means that the initial random sizes of entrant groups are drawn from a distribution with a tail thinner than a power law with exponent μ. It could be a power law with an exponent larger than μ or any distribution decaying faster than power laws for large $\tilde{s}_0$ values.

Following Proposition 1, we can state the following

Corollary 1: The exponent μ of the distribution of sizes takes the value 1 corresponding to Zipf's law, if and only if $r = h$.

In order to understand the meaning of Corollary 1, notice that $r - h$ represents the average growth rate of an incumbent group. Indeed, considering a group present at time t, during the next instant dt, it will either exit with probability $h \cdot dt$ (and therefore its size declines by a factor -100%) or grow at an average rate equal to $r \cdot dt$, with probability $(1 - h \cdot dt)$. The coefficient r is therefore the conditional growth rate of projects, conditioned on not having died yet. Then, the unconditional expected growth rate over the small time increment dt of an incumbent group is $(r - h) \cdot dt + O(dt^2)$. The statistically stationary regime, in the presence of a stationary population of group forming individuals, corresponds to condition $r = h$. Malevergne et al. (Malevergne et al., 2010) showed that this condition can be easily generalized to the case where the population of group forming individuals grows itself with some exponential rate, as is the minimal viable group size (Malevergne et al., 2010). Then, this condition translates into that for the maximum sustainable growth of the universe of groups, as mentioned above.

4.2.2 Strategy to test the theory and description of our data set

Our strategy is to find an empirical dataset in which (i) all ingredients of the theory can be verified explicitly, (ii) all parameters r, σ and h measured directly and (iii) the empirical distribution of group sizes can be compared with to prediction (4.1) with (4.3).

We have found such a database, with Amazee.com, which is a Web-based platform of collaboration. Using Amazee's Web-platform, anyone with an idea for a collaborative project can sign in and use the website to gather followers, who will together help the project owner to accomplish the project. An Amazee project can be of any type of activities, such as arts and culture, environment and nature, politics and beliefs, science and innovation, social and philanthropic, sports and leisure, and so on. Most of the projects are public, for instance, "build a strong community of Internet entrepreneurs in Switzerland to exchange information and have fun" (Web Monday Zurich), "connect all women working in the Swiss ICT industry" (Tech Girls Switzerland), "to provide fresh running water to each home in the small African village of Dixie" (Water for Dixie), and so on. Amazee.com provides a set of features covering the entire lifetime of a typical project, such as project planning, participants recruiting, fund raising, events and meetings hosting, communication, files archiving, and so on. Users join Amazee.com by either creating a new project, or participating in projects created by others. The Amazee data we analyze contains the complete recording in time of the activities of all users creating and joining all the projects in existence from February 2008 till April 2011.

Projects can be seen as proxies of many naturally occurring entities, such as social groups, firms, cities, investment vehicles, and so on, each driven by some goal, competition, and interaction within social networks. The detailed knowledge of the activity of the participants of all projects provides a remarkable opportunity to dissect and understand the dynamics of such systems. In the present study, we restrict our attention to the simplest measure of size, namely the number S_i of members of project i.

4.3 Data analysis and results

4.3.1 Empirical estimation of the power law distributions of project sizes

Amazee's platform started in February, 2008, which can be taken as the birthday of the ecology of projects. We analyze nine snapshots of the database, one for approximately every four months from October 2008 to April 2011. The first snapshot is eight months after the birth of the operations on Amazee.com. With the parameter values for r, σ and h determined below, formula (4.4) predicts a transient of 50-400 days. Therefore, we should observe a reasonable convergence to the expected power law distribution in each snapshot.

Table 13 and Fig 6 confirm that the distributions of project sizes obtained for these nine snapshots are power laws (4.1). Indeed, two kinds of statistical tests we use all validate these power laws. The two tests are (i) the parametric bootstrap based Kolmogorov-Smirnov (K-S) test and (ii) the uniformly most powerful unbiased (UMPU) test of Pareto against the lognormal distributions (Malevergne et al., 2011). Parametric bootstrap based K-S tests must be used in our case because the parameter μ in (4.1) is not known but is estimated in the calibration procedure. All the p-values for the K-S test applied to the nine snapshots are found larger than 0.05, indicating that one cannot reject the null hypothesis that the upper tails of the distributions of project sizes in all the nine snapshots are power laws. The second test, the UMPU test, compares the power law family as the null hypothesis with the lognormal distribution family as an alternative hypothesis. Lognormal distributions are often very hard to distinguish from power laws. Therefore, comparing their explanatory power relative to that of the power law family is natural. We find that all the p-values of the

Tab. 13: Descriptive statistics of the sizes of Amazee's projects at different times, showing that most projects have a size of just a few individuals while a few projects have hundreds to more than one thousand members. Dates are in format day/month/year.

Date	Projects Number	Mean Size	Minimum Size	Maximum Size	Median Size
01/10/2008	451	6	1	227	3
24/01/2009	864	11	1	1106	2
20/05/2009	1125	10	1	1114	2
13/09/2009	1275	9	1	1115	2
07/01/2010	1403	9	1	1117	2
03/05/2010	1579	9	1	1120	2
27/08/2010	1749	9	1	1121	2
21/12/2010	2033	9	1	1123	2
16/04/2011	2231	9	1	1126	2

UMPU tests applied to the nine snapshots are found larger than 0.05. One can thus accept the null hypothesis of power laws in the UMPU test and reject the alternative hypothesis of lognormal distributions.

Because the numbers of project members are integers, the exponents μ corresponding to the empirical distributions shown in Fig 6 are estimated using the maximum likelihood method (ML) with the normalized discrete version of (4.1), $p(s) = \frac{s^{-(1+\mu)}}{\zeta(1+\mu)}$, where $\zeta(x)$ is the Riemann zeta function: $\zeta(x) = \sum_{s=1}^{\infty} s^{-x}$. The exponents are found around 1.0, with confidence intervals always including 1.0. We check the robustness of this conclusion by estimating the exponents μ for the nine snapshots as a function of a lower threshold above which the MLE is performed. For all the snapshots, we find stable estimations, with the 95% confidence intervals including the value 1, as shown in Table 14. We can thus conclude that Zipf's law always holds in this dataset.

4.3.2 Empirical test of Gibrat's law of proportional growth

We now test formula (4.3) and its underlying model. For this, we test if model (4.2) holds and proceed to estimate the parameters r, σ and h. The proportional growth model posits that, for sufficiently small time intervals Δt, the mean $\mathrm{E}[\Delta S]$ and the standard deviation $\sigma_{\Delta S}$ of the increment of

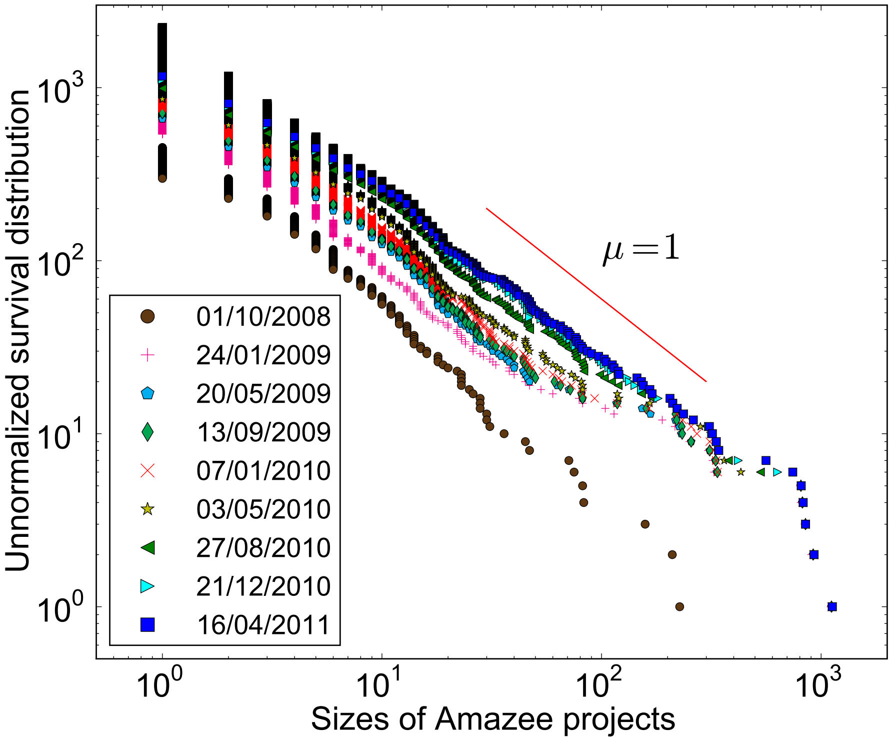

Fig. 6: Symbols in different colors and shapes: unnormalized survival distributions of Amazee project sizes measured for nine snapshots on different days shown as the legend. The short red line as a reference shows Zipf's law - the power law exponent μ equals to 1.

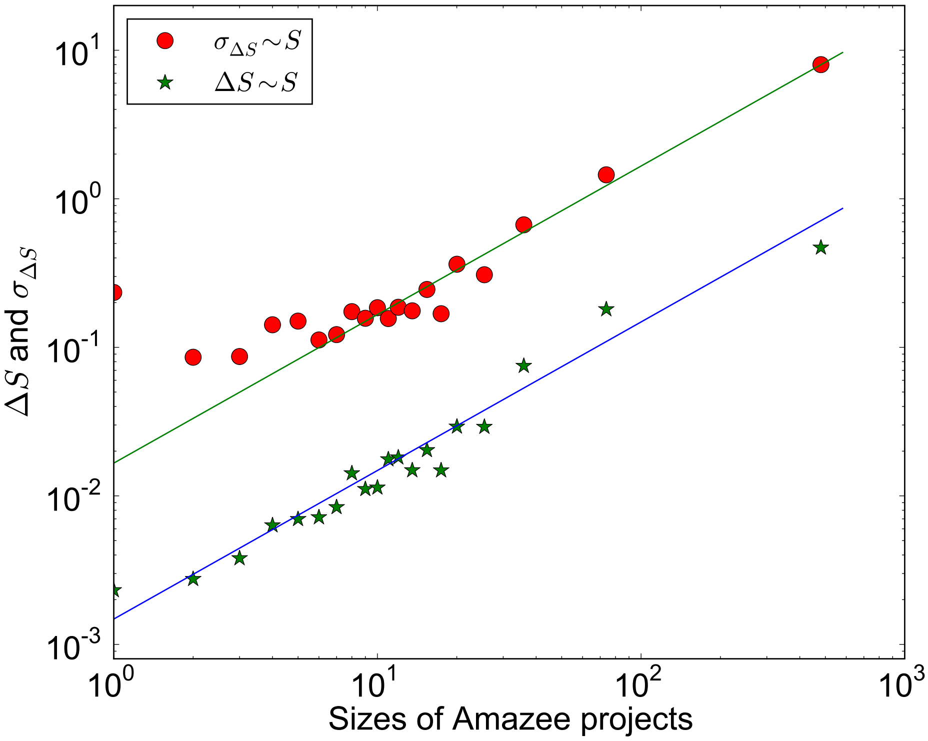

Fig. 7: Test of Gibrat's law for the proportional growth of Amazee project sizes until 16 April 2011. The slopes of the fitted straight lines are exactly 1.

the size S of a given project should both be proportional to S. To test this proposition, we extract all the $(S_i(t), \Delta S_i(t))$ pairs, where $i = 1, 2, ..., N$ denotes all the projects in the dataset that has totally N projects, and $t = 1, 2, ..., T$ denotes the time from the first day 1 to the last day T in the dataset, and all these data pairs are pooled together in 100 size intervals over all nine snapshots. For each of the 100 size intervals, Figure 7 plots the average daily increase of project sizes ($\mathrm{E}[\Delta S]$) and its standard deviation $\sigma_{\Delta S}$ as a function of S. Linear regressions give very high R^2's larger than 0.995, confirming that Gibrat's law holds. Note that $\sigma_{\Delta S}$ is much larger than $\mathrm{E}[\Delta S]$, i.e., the stochastic component of the proportional growth clearly dominates (an essential condition for a power law to emerge in the model (Saichev et al., 2009)). After verifying the Gibrat's law, we then estimate r and σ using the Maximum likelihood method, and the results are in Table 14.

4.3.3 Empirical analysis of the birth and death processes of amazee.com projects

Next, we find that the rate of birth of new projects on amazee.com is approximately described by a Poisson process, such that the probability that n projects are born in a given day is given by

$$Pr\{n\} = \frac{\lambda^n}{n!} e^{-\lambda} \, , \tag{4.5}$$

where $\lambda \approx 2.4$ is the mean number of new born projects per day. Numerical analysis shows that the deviation of the real birth process from a Poisson process brings very small errors into the exponents predicted by (4.3). The sensitivity of (4.3) to the distribution of either the birth process or the death process, as well as the convergence of the power law exponents discussed below, explored by numerical analysis and simulation, has been investigated by Saichev et al. (Saichev et al., 2009).

Many projects eventually stop growing, when they have reached their goals or in the presence of operational problems, and thus are inaccessible from the Amazee website. To the users, this means that the projects do not exist any more. We find that the distribution of project lifetimes ℓ (the lifetime of a project is the number of days between its birthday and the day it disappears from the Amazee website) is very well approximated by the exponential law

$$Pr\{\ell \geq T\} = e^{-hT} \, , \tag{4.6}$$

where h is the death hazard rate, whose maximum likelihood estimations are reported in Table 14 for the nine snapshots of Amazee's database.

4.3.4 Comparison between theoretical and empirical values of the power law exponents

Using the empirically determined values of r,σ and h, we are now in position to test the theoretical prediction (4.3) for the exponents μ of the proportional growth model in the presence of stochastic birth and death process.

As reported above, the empirically determined values of the power law exponent μ are, within statistical fluctuation, close to 1, the value corresponding to Zipf's law. However, as shown in Table 14, r is always larger than h. Therefore, the presence of Zipf's law we have observed in the nine snapshots cannot be explained by Corollary 1, which requires $r = h$. We therefore consider the following two possible mechanisms, which can also explains the presence of Zipf's law in the empirical distributions.

1. In a growing social system where $r > h$, the theory predicts that one should observe a power law with $\mu < 1$. However, it usually takes a long time for the exponent to converge to the stationary value predicted by expression (4.3). Since a system with a finite lifetime tends to underpopulate large groups, the effective power law exponent tends to be larger in the transient establishment of the population as it slowly evolves to its stationarity distribution. An approximate Zipf's law can thus emerge as a compensation between this transient effect of a cross-over from a short-term transient thin-tailed distribution to a very heavy distribution with exponent smaller than 1. This suggests that Zipf's law can be observed in the early stages of the social system, although the measured power law exponent will converge to the value predicted by (4.3) in the long run.
2. If σ^2 is large enough such that $\sigma^2 \gg h$ and $\sigma^2 \gg r$, Zipf's law will hold approximately according to the prediction of expression (4.3). This case could happen in an "old" system, in which σ^2 grows to very large values.

In order to understand how an approximate Zipf's law could be obtained as observed empirically, we have performed numerical simulations of artificial world of projects that follow the laws of random Poisson birth-proportional growth-random deaths, which are the ingredients of the theoretical predictions as explained before. Our simulations confirm that it takes a very long time for the power law exponents to converge to the predicted values listed in the μ (TH) column of Table 14 (this is all the more true when these asymptotic values are smaller than 1). In order to understand quantitatively the impact of a finite lifetime of the ecology of Amazee projects, we simulate the model underlying (4.3), using the empirical parameters λ, r, σ, h estimated empirically. We then record the distributions of simulated project sizes at finite times corresponding to the nine snapshots used for the empirical analysis. We then analyze the distributions for these nine snapshots in our synthetic universe following exactly the same procedure as done for the empirical data. The corresponding power law exponents are reported under the name μ (SIM) in Table 14. One can observe the excellent agreement between the empirical exponents μ (MLE) and the theoretical values μ (SIM) that take into account the finite lifetime of the Amazee system and for the values of the parameters found empirically. Specifically, almost all empirical exponents lie in the 95% confidence interval of the simulated exponents, with no adjustable parameters!

Our simulations thus make clear that the the deviations between empirical exponents and asymptotic values (4.3) may be in significant part explained by a finite lifetime effect. However, the second mechanism of a large σ^2 also contributed, especially at later times after August 2010, when the value of σ grew tremendously, so that σ^2 is much larger than both r and h.

In summary, at early times, the value of the empirical power law exponent close to 1 is mainly due to a finite lifetime effect. At later times, the exponent converges better and better to its asymptotic theoretical value, but the later tends to grow towards 1 as σ^2 grows correspondingly. The interplay between these two mechanisms explains well the observed stability of the empirical distributions, which are very close to Zipf's law over much of the history of Amazee.

4.4 Conclusions

The detailed empirical analysis of the burgeoning social networks on Amazee has provided a unique set-up to test the origin of Zipf's law in a system in which all ingredients needed for Zipf's law to apply are verifiable and verified. Indeed, the Amazee system underwent different regimes, from a relatively small standard deviation σ of the relative project size growth rates to large values, although all the nine snapshots show the same approximate Zipf's law. Using numerical simulations of the underlying growth model, we have demonstrated that the empirical stability of Zipf's law over the whole lifetime of the Amazee world can be attributed to a quite subtle interplay between a finite lifetime effect and a large σ value. Our analysis and the corresponding results demonstrate that Zipf's law can be observed with a good precision even when the balanced growth condition ($r = h$) is not realized, if the random proportional growth has a strong stochastic component and is acting on young systems under development.

Our analysis provides a novel validation of the underlying model of (4.3). In spite of the complex dynamics in the network of Amazee projects, the model predicts correctly Zipf's law at different times. More than the ubiquitous presence of Zipf's law or of power laws, our analysis reveals more information relevant to the evolution the system. These ingredients (random Poisson birth-proportional growth-random deaths) provide the basis for possible predictions of the future evolution of the system.

There are situations where Zipf's law holds for some sets and not for other sets with seemingly similar population and for some measures of sizes and not for others. For instance, Podobnik et al. (Podobnik et al., 2010) has documented Zipf's law for stock equity and market capitalization of NASDAQ companies. However, they have found that Zipf's law does not hold for NYSE firms. In addition, alternative measures of sizes, such as debt and assets, have been found to be generally power-law distributed but not necessarily with the exponent corresponding to Zipf's law. This implies that the mechanism of proportional growth has to be refined to take into account the multidimensional nature of firms and their interactions in their complex networks.

Tab. 14: For each of the nine snapshots of the amazee database, we report the parameters r, σ and h as explained in the text. Reporting these parameters in expression (4.3) yields the predicted exponents μ (TH), and simulating numerically the model in Malevergne et al. (2010) with these parameters we estimate the μ (SIM), which are compared with the empirical exponents μ estimated by maximum likelihood (MLE). Except μ (TH), all the parameters are estimated by MLE, and we report also the 95% confidence interval of each estimated parameter. For each cell value in the "μ (TH)" column, the 95% confidence interval is obtained by parametric bootstrapping over 100'000 samples. Dates are in format day/month/year.

Date	r	σ	h	μ (MLE)	μ (TH)	μ (SIM)
01/10/2008	0.0127 [0.0110, 0.0142]	0.164 [0.162, 0.165]	0.00373 [0.00314, 0.00432]	1.15 [0.979, 1.31]	0.555 [0.482, 0.631]	1.17 [0.926, 1.41]
24/01/2009	0.00721 [0.00658, 0.00785]	0.113 [0.113, 0.114]	0.00185 [0.00161, 0.00209]	0.918 [0.840, 0.997]	0.479 [0.424, 0.536]	0.937 [0.834, 1.04]
20/05/2009	0.00487 [0.00447, 0.00527]	0.0993 [0.0990, 0.0996]	0.00112 [0.000982, 0.00125]	0.934 [0.847, 1.02]	0.482 [0.433, 0.532]	1.15 [0.963, 1.33]
13/09/2009	0.00337 [0.00310, 0.00364]	0.0848 [0.0846, 0.0850]	0.000739 [0.000652, 0.000826]	0.989 [0.889, 1.09]	0.485 [0.438, 0.534]	1.03 [0.937, 1.12]
07/01/2010	0.00255 [0.00235, 0.00276]	0.0755 [0.0753, 0.0756]	0.000530 [0.000468, 0.000591]	0.982 [0.887, 1.08]	0.486 [0.439, 0.534]	1.04 [0.953, 1.12]
03/05/2010	0.00226 [0.00210, 0.00243]	0.0714 [0.0712, 0.0715]	0.000422 [0.000374, 0.000470]	0.922 [0.722, 1.12]	0.467 [0.424, 0.511]	1.18 [1.04, 1.31]
27/08/2010	0.00237 [0.00201, 0.00273]	0.173 [0.172, 0.173]	0.000437 [0.000394, 0.000481]	1.02 [0.904, 1.14]	0.875 [0.852, 0.898]	0.974 [0.753, 1.19]
21/12/2010	0.00216 [0.00184, 0.00247]	0.169 [0.169, 0.170]	0.000361 [0.000326, 0.000396]	1.03 [0.914, 1.14]	0.878 [0.857, 0.900]	0.957 [0.791, 1.12]
16/04/2011	0.00184 [0.00158, 0.00210]	0.154 [0.153, 0.154]	0.000325 [0.000294, 0.000355]	1.01 [0.902, 1.13]	0.875 [0.854, 0.897]	0.967 [0.842, 1.09]

5

Conclusions

The papers in this thesis deliver two main messages: i) complex financial and social systems cannot be treated as simply the sum of their individual components, as the non-linearity of the systems can lead to surprises; ii) nevertheless, we can disentangle these systems from the complexity by understanding the underlying interactions and coarse-graining.

The first two papers challenge both the weak form and the semi-strong form of EMH. In "Can media moods predict stock prices during and after the 2008 financial crisis?", we identify a positive feedback loop in the relationships between negative news and negative returns by using coarse-grained news data. This strongly hints in the direction of the likely formation mechanism of financial bubbles and crashes. Moreover, our empirical results show that one can generate both statistically and economically significant returns based on newsflows and that the extra returns are not explained by the Fama-French factors. In "Reverse engineering stock markets with mixed games and alpha generation", we find that ABMs have statistically significant success rates of predicting the sign of future returns. Using ABM-based strategies, one can generate statistically significant positive returns. In addition to those returns, we also find linear relationships between the ABMs parameters and historical stock market regimes. The former paper uses coarse-grained data, and the latter paper models the interaction between agents. Both papers apply economic insights and methods yet at the same time incorporate ideas from complexity theory in statistical physics. Combining these ideas with tools from data collection, big data analysis, and scientific computing, we are able to disentangle the systems from the complexity and show that asset prices are to some extent statistically predictable. Following the same approach, in "Empirical test of the origin of Zipf's law in growing social networks" we predict the exponents of the power law size distributions of groups on a website by calculating the growth rates and standard deviations of the groups, as well as their birth and hazard rates, at different times. The ubiquitous Zipf's law, an emergent phenomenon, is thus successfully explained by the microscopic behaviour of the groups. The results again evidence the predictive power of our approach.

The contributions of these papers are thus important. The first paper discloses that negative news is not digested by investors immediately but affects the investors for a longer time than predicted by EMH because of the positive feedback loop between negative news and negative returns. The second paper models how investors trade based on historical price information. Because of the existence of trend-followers, modelled by majority game and $-game players, stock markets are not informationally efficient. Both papers present methods to predict stock prices. In the future, we can combine the methods in these two papers to better model the regime switching of the stock markets. The third paper verifies Zipf's law empirically. The methodology applied therein can be applied to many other systems.

Our interdisciplinary approach has made successful predictions in both financial systems and social networks. It reveals the importance of combining insights from financial economics, concepts from statistical physics, and tools from computer science. In the future, we shall apply this approach to more problems, both theoretical and practical.

List of Tables

1 Descriptive statistics of the media moods and of the returns of the S&P500 index shown in figure 1 from January 1, 2007 to June 6, 2012. 21

2 Results of the estimation of the VAR(10) model 2.6 on the positive and negative moods. The rows labeled by $\beta_0, \beta_1, \ldots, \beta_{10}$ list the estimated parameters, where the values in the parentheses are the standard errors of the corresponding parameters, the "Adj. R^2" row lists the adjusted R^2 of the two linear models embedded in model 2.6, and the "F-stat" row lists the F statistics of the two linear models, where the values in the parentheses are the corresponding p-values. The p_t (respectively n_t) column list the results of the linear model whose dependent variable is the positive (respectively negative) moods at lag 0. Moreover, "*" indicates that the marked parameter is statistically significant with a p-value less than 0.1, "**" indicates a p-value less than 0.05, "***" indicates a p-value less than 0.01, and "****" indicates a p-value less than 0.001. 32

3 Results of the estimation of the VAR(3) model (2.7) using the S&P500 returns. The rows labeled $\beta_0, \beta_1, \ldots, \beta_{10}$ list the estimated parameters, where the values in parentheses are the standard errors of the corresponding parameters. The "Adj. R^2" row lists the adjusted R^2 of the two linear models embedded in model (2.7). The "F-stat" row lists the F statistics of the two linear models, where the values in the parentheses are the corresponding p-values. The r_t (respectively n_t) column lists the results of the linear model whose dependent variable is the returns (respectively negative moods) at lag 0. Moreover, "*" indicates that the marked parameter is statistically significant with a p-value less than 0.1, "**" indicates a p-value less than 0.05, "***" indicates a p-value less than 0.01, and "****" indicates a p-value less than 0.001. 33

4 Results of the estimation of the VAR(3) model (2.7) using different indices other than the S&P500. The rows labeled $\beta_0, \beta_1, \ldots, \beta_{10}$ list the estimated parameters, where the values in parentheses are the standard errors of the corresponding parameters. The "Adj. R^2" rows lists the adjusted R^2 of the two linear models embedded in model (2.7). The "F-stat" rows lists the F statistics of the two linear models, where the values in the parentheses are the corresponding p-values. The r_t (respectively n_t) columns list the results of the linear model whose dependent variable is the returns of the corresponding index (respectively negative moods) at lag 0. Moreover, "*" indicates that the marked parameter is statistically significant with a p-value less than 0.1, "**" indicates a p-value less than 0.05, "***" indicates a p-value less than 0.01, and "****" indicates a p-value less than 0.001. . 34

5 Linear relationships between the monthly average S&P500 returns and the monthly average negative moods modeled by a VAR(1) model. The "Const." row lists the values of the constant parameter in the linear models. The $\bar{\mathbf{r}}_{\mathbf{t-1}}$ row lists the coefficients of the monthly average returns at lag 1. The $\bar{\mathbf{n}}_{\mathbf{t-1}}$ row lists the coefficients of the monthly average negative moods at lag 1. In the above rows, the values in the parentheses are the standard errors of the corresponding coefficients. The "Adj. R^2" row lists the adjusted R^2 of the linear models. The F-stat. row lists the F statistics of the linear models, where the parentheses give the p-values. The $\bar{\mathbf{r}}_{\mathbf{t}}$ column contains the estimation results of the linear model whose dependent variable is the monthly average returns at lag 0. The $\bar{\mathbf{n}}_{\mathbf{t}}$ column contains the estimation results of the linear model whose dependent variable is the monthly average negative moods at lag 0. Moreover, "*" indicates that the marked parameter is statistically significant with a p-value less than 0.1, "**" indicates a p-value less than 0.05, "***" indicates a p-value less than 0.01, and "****" indicates a p-value less than 0.001. 35

6 Performance of our strategy based on the media moods and comparison with random strategies. We consider five different time periods starting with different starting times T_s and ending times T_e. All dates are given in the "year-month-day" format. The r_t column lists the total returns of the trades over the corresponding time interval $[T_s, T_e]$. The $\bar{r}$ column lists the average annual returns of our strategy for each time window and the sr column lists the annualized Sharpe ratios (using zero risk-free interest rate). The τ_r column lists the percentile rank of the total returns of our strategy compared with random strategies, as explained in the text. The τ_{sr} column lists the percentile rank of the Sharpe ratios of our strategy compared with random strategies, as explained in the text. 35

7 List of the α's of our trading strategy based on the media moods obtained for the five periods shown in table 6. The α's are obtained as the intercepts of the regression of the time series of returns of our strategy in each time window as a function of the three Fama-French factors model. The parentheses give the standard errors of the corresponding α's. "*" indicates that the marked α is statistically significant with a p-value less than 0.1, "**" indicates a p-value less than 0.05, and "***" indicates a p-value less than 0.01. 36

8 Descriptive statistics of the time series used in our reverse engineering process. The "Start time" and "End time" use the "Year-Month-Day" format, the "Mean" column gives the average returns, the "Sd" column lists the standard deviations of returns, the "Median" column shows the median values of the returns, the "Min" column gives the minimum returns, and the "Max" column lists the maximum returns. 48

9 Numbers of experiments ("Count") with insignificant prediction power at the 90% confidence level. For each of the seven indices, 100 experiments have been performed and their corresponding 'success ratios' have been compared with those of 1000 random strategies. See main text for explanations. 49

10 Results of 25 experiments, using the reverse-engineered mixed games to predict the sign of future returns in out-of-sample windows of 16 trading days. **Index** lists the names of the indexes, **Start year** lists the start year of the corresponding time series, **End year** lists the end year of the time series, **Data points** list the numbers of data points predicted by the ABM's, **Positive ratio** lists the ratio of positive returns in the real time series, W_{is} lists the in-sample window sizes of the experiments, **Success rate** lists the total success rates of the ABM's, and **p-value** lists the p-values of the prediction power tests. In all experiments, the out-of-sample window sizes of the experiments is fixed at $W_{os} = 16$ days. For all experiments, the p-values are much smaller than 0.001. 57

11 Summary performance of trading strategies constructed on trading signals from a single reverse engineering experiment and without considering the transaction costs. **Index**'s are the names of the indexes, **Start year**'s are the start year of the corresponding time series, **End year**'s are the end year of the time series, **Data points**'s are the numbers of data points used in the ABM prediction process, W_{is}'s are the in-sample window sizes of the experiments, **Total returns**'s and **Sharpe ratios**'s are the total returns (P&L) and annualized Sharpe ratios of the ABM's based trading strategies respectively, PV_r's and PV_{shr}'s are the p-values of testing the returns and Sharpe ratios of the ABM's based trading strategies against those of the random trading strategies respectively, **Annual return**'s are the annual returns of the ABM's based trading strategies (for instance, 0.601 corresponds to an annual return of 60.1%) , and the α's are obtained as the intercepts of the regression of the time series of returns of the ABM's based trading strategies as a function of the three Fama-French factors model. For all experiments, the out-of-sample window size is fixed to $W_{os} = 16$ days. The parentheses give the standard errors of the corresponding α's. "*" indicates that the marked α is statistically significant with a p-value less than 0.1, "**" indicates a p-value less than 0.05, and "***" indicates a p-value less than 0.01. 58

12 Linear regression (3.14) between realized financial returns and some of the parameters of the ABM's during different market regimes. **Index** lists the names of the indexes. **Start time** contains the start times in "year-month" format of the corresponding regimes, and similarly **End time** gives the end time of the regimes. **Data points** lists the number of data points of the dependent and independent variables. **Intercept** presents the coefficient a_0, i.e. the intercept of regression (3.14). $\hat{fr}_{\mathrm{act}}$, $\hat{\tau}$, $\hat{fr}^{\mathrm{mg}}$, $\hat{fr}^{\mathrm{majg}}$, $\hat{fr}^{\mathrm{dg}}$ and $\hat{fr}^{\mathrm{dmg}}$ list the estimated parameters for the corresponding variables. **F-test** shows the p-values of the corresponding F tests. R^2 contains the related R-squares. The parentheses give the standard errors of the corresponding parameters, and "*" indicates that the estimated parameter is statistically significant with a p-value less than 0.1, "**" indicates a p-value less than 0.05, and "***" indicates a p-value less than 0.01. 59

13 Descriptive statistics of the sizes of Amazee's projects at different times, showing that most projects have a size of just a few individuals while a few projects have hundreds to more than one thousand members. Dates are in format day/month/year. 68

14 For each of the nine snapshots of the amazee database, we report the parameters r, σ and h as explained in the text. Reporting these parameters in expression (4.3) yields the predicted exponents μ (TH), and simulating numerically the model in Malevergne et al. (2010) with these parameters we estimate the μ (SIM), which are compared with the empirical exponents μ estimated by maximum likelihood (MLE). Except μ (TH), all the parameters are estimated by MLE, and we report also the 95% confidence interval of each estimated parameter. For each cell value in the "μ (TH)" column, the 95% confidence interval is obtained by parametric bootstrapping over 100'000 samples. Dates are in format day/month/year. 75

List of Figures

1 The upper plot shows the dynamics of both positive (blue dashed line) and negative moods (red continuous line) extracted from the Reuters daily news from January 1, 2007 to June 6, 2012. The positive (respectively negative) mood for a given day is defined as the fraction of positive (respectively negative) news among all news articles provided by Reuters on that day. The lower plot shows the dynamics of the S&P500 index from January 1, 2007 to June 6, 2012. 20

2 The four plots display the orthogonal impulse response functions (IRF) of the media moods. The upper-left plot shows the response function of the positive moods receiving impulses from the positive moods; the upper-right plot shows the response function of the negative moods receiving impulses from the positive moods; the lower-left plot shows the response function of the positive moods receiving impulses from the negative moods; and the lower-right plot shows the response function of the negative moods receiving impulses from the negative moods. The solid blue lines in the plots are the orthogonal impulse response functions, the dash-dotted black lines are the 95% confidence intervals of the impulse response functions, and the red dashed lines indicate the 0 levels. 22

3 The group of plots display the orthogonal impulse response functions of the S&P500 returns and the negative moods. The upper-left plot shows the response function of the S&P500 returns receiving impulses from the S&P500 returns, the upper-right plot shows the response function of the negative moods receiving impulses from the S&P500 returns, the lower-left plot shows the response function of the S&P500 returns receiving impulses from the negative moods, and the lower-right plot shows the response function of the negative moods receiving impulses from the negative moods. The solid blue lines in the plots are the orthogonal impulse response functions, the dash-dotted black lines are the 95% confidence intervals of the impulse response functions, and the red dashed lines indicate the 0 levels. 25

4 P&L (cumulative return) without considering the transaction costs obtained by the strategy described in the text, which is based on the media moods (blue thick line), over the entire time period between January 1, 2007 and June 6, 2012, minus the first $T_{in} = 180$ in-sample days. The buy and hold strategy is represented by the thin green dashed line. The sell and hold (and buy back) strategy, consisting in shorting the S&P500 index at the beginning and holding until the end when the buy back occurs, is shown as the thin red dotted-dashed line. 30

5 Prices dynamics of the S&P500, Dow Jones Industrial Average, and Nasdaq 100 indexes from 1982 to 2012. 56

6 Symbols in different colors and shapes: unnormalized survival distributions of Amazee project sizes measured for nine snapshots on different days shown as the legend. The short red line as a reference shows Zipf's law - the power law exponent μ equals to 1. 69

7 Test of Gibrat's law for the proportional growth of Amazee project sizes until 16 April 2011. The slopes of the fitted straight lines are exactly 1. 70

Bibliography

Adamic, L., Huberman, B., 2000. The nature of markets in the world wide web. Quarterly Journal of Electronic Commerce 1, 5–12.

Akaike, H., Dec 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6), 716–723.

Almgren, R., 2003. Optimal execution with nonlinear impact functions and trading-enhanced risk. Applied Mathematical Finance 10 (1), 1–18.

Andersen, J., Sornette, D., Jan 2003. The $-game. European Physical Journal B 31 (1), 141–145.

Andersen, J., Sornette, D., 2005. A mechanism for pockets of predictability in complex adaptive systems. Europhys. Lett. 70 (5), 697–703.

Anderson, L., Jun 2011. Demystifying the arab spring. Foreign Affairs 90 (3), 2–7.

Anderson, P. W., 1972. More is different. Science 177 (4047), 393 –396.

Antweiler, W., Frank, M. Z., 2004. Is all that talk just noise? the information content of internet stock message boards. The Journal of Finance 59 (3), 1259–1294.

Antweiler, W., Frank, M. Z., Aug 2006. Do us stock markets typically overreact to corporate news stories? SSRN eLibrary.
URL `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=878091`

Arthur, W. B., May 1994. Inductive reasoning and bounded rationality. The American Economic Review 84 (2), 406–411.

Axtell, R., 2001. Zipf distribution of u.s. firm sizes. Science 293, 1818–1820.

Barabasi, A.-L., Albert, R., 1999. Statistical mechanics of complex networks. Science 286, 509–512.

Barras, L., Scaillet, O., Wermers, R., February 2010. False discoveries in mutual fund performance: Measuring luck in estimated alphas. The Journal of Finance 65 (1), 179–216.

Bernanke, B. S., Jan 2004. Essays on the Great Depression. Princeton University Press.

Blanchard, O., May 1979. Backward and forward solutions for economies with rational expectations. The American Economic Review 69 (2), 114–118.

Bollen, J., Mao, H., Zeng, X., Mar 2011. Twitter mood predicts the stock market. Journal of Computational Science 2 (1), 1–8.

Bouchaud, J.-P., Kockelkoren, J., Potters, M., 2006. Random walks, liquidity molasses and critical response in financial markets. Quantitative Finance 6 (2), 115–123.

Chakraborti, A., Muni Toke, I., Patriarca, M., Abergel, F., 2011. Econophysics review: I. empirical facts. Quantitative Finance 11 (7), 991–1012.

Challet, D., Marsili, M., Zhang, Y.-C., Feb 2000. Modeling market mechanism with minority game 276, 284–315.

Challet, D., Marsili, M., Zhang, Y.-C., Jan 2005. Minority Games: Interacting Agents in Financial Markets. Oxford University Press, USA.

Challet, D., Zhang, Y.-C., Dec 1997. Emergence of cooperation and organization in an evolutionary game. Physica A: Statistical Mechanics and its Applications 246 (3), 407–418.

Chiarella, C., Dieci, R., He, X.-Z., 2009. Heterogeneity, market mechanisms, and asset price dynamics. Handbook of Financial Markets: Dynamics and Evolution (North-Holland, Elsevier) chapter 5, 277–344.

Cont, R., 2001. Empirical properties of asset returns: stylized facts and statistical issues. Quantitative Finance 1 (2), 223–236.

Cutler, D., Poterba, J., , Summers, L., 1987. What moves stock prices? Journal of Portfolio Management 15 (3), 4–12.

Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. The Journal of Finance 66 (5), 1461–1499.

Daniel, G., Sornette, D., Woehrmann, P., 2009. Look-ahead benchmark bias in portfolio performance evaluation. Journal of Economic Behavior and Organization 36 (1), 121–130.

Eckbo, B. E. (Ed.), Jun 2007. Handbook of Corporate Finance, Volume 1: Empirical Corporate Finance, 1st Edition. North Holland.

Eichengreen, B., Temin, P., 2000. The gold standard and the great depression. Contemporary European History 9 (02), 183–207.

Elliott, G., Rothenberg, T. J., Stock, J. H., Jul 1996. Efficient tests for an autoregressive unit root. Econometrica 64 (4), 813–836.

Elsenbroich, C., 2011. Explanation in agent-based modelling: Functions, causality or mechanisms? Journal of Artificial Societies and Social Simulation 15 (3), 1.

Erdogan, O., Yezegel, A., 2009. The news of no news in stock markets. Quantitative Finance 9 (8), 897–909.

Evstigneev, I., Hens, T., Schenk-Hoppé, K., 2009. Emergence of cooperation and organization in an evolutionary game. Handbook of Financial Markets: Dynamics and Evolution (North-Holland, Elsevier) chapter 9, 507–566.

Fair, R., 2002. Events that shook the market. Journal of Business 75 (4), 713–731.

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

Fama, E. F., 1991. Efficient capital markets: Ii. The Journal of Finance 46 (5), 1575–1617.

Fama, E. F., Sep 1998. Market efficiency, long-term returns, and behavioral finance. Journal of Financial Economics 49 (3), 283–306.

Fama, E. F., French, K. R., Feb 1993. Common risk factors in the returns on stocks and bonds 33, 3–56.

Farmer, J. D., Nov 2002. Market force, ecology and evolution. Industrial and Corporate Change 11 (5), 895–953.

Farmer, J. D., Gerig, A., Lillo, F., Waelbroeck, H., 2012. How efficiency shapes market impact. Quantitative Finance, (http://arxiv.org/abs/1102.5457).

Fleming, M., Remolona, E., 1997. What moves the bond market? FRBNY Economic Policy Review 3 (4), 31–50.

French, K. R., 2012. Kenneth r. french - data library.
URL `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

Furusawa, C., Kaneko, K., 2003. Zipfs law in gene expression. Phys. Rev. Lett. 90, 088102.

Gabaix, X., 1999. Zipf's law for cities: An explanation. Quarterly Journal of Economics 114, 739–767.

Galbraith, J. K., Jul 1994. A Short History of Financial Euphoria, reprint Edition. Penguin Books.

Gelvin, J. L., Mar 2012. The Arab Uprisings: What Everyone Needs to Know. Oxford University Press, USA.

Gibrat, R., 1931. Les inégalités economiques. Librarie du Recueil Sirey, Paris.

Grossman, S., Stiglitz, J., 1980. On the impossibility of informationally efficient markets. The American Economic Review 70 (3), 293–408.

Hamilton, J. D., Jan 1994. Time Series Analysis, 1st Edition. Princeton University Press.

Harras, G., Sornette, D., 2011. How to grow a bubble: A model of myopic adapting agents. Journal of Economic Behavior and Organization 80 (1), 137–152.

Hommes, C., 2002. Modeling the stylized facts in finance through simple nonlinear adaptive systems. Proc. Nat. Acad. Sci. USA 99 (Suppl. 3), 7221–7228.

Hommes, C., 2006. Heterogeneous agent models in economics and finance. Handbook of Computational Economics (Elsevier B.V.), Edited by Leigh Tesfatsion and Kenneth L. Judd vol. 2, chapter 23, 1109–1186.

Hommes, C., Wagener, F., 2009. Complex evolutionary systems in behavioral finance. Handbook of Financial Markets: Dynamics and Evolution (North-Holland, Elsevier) chapter 4, 217–276.

Hopkins, D., King, G., Knowles, M., Melendez, S., 2012. Readme: Software for automated content analysis.
URL `http://gking.harvard.edu/readme`

Hopkins, D. J., King, G., 2010. A method of automated nonparametric content analysis for social science. American Journal of Political Science 54 (1), 229–247.

Ijri, Y., Simon, H. A., 1977. Distributions and the sizes of business firms, north-holland, new york.

Jefferies, P., Hart, M., Hui, P., Johnson, N., 2001. From market games to real-world markets. The European Physical Journal B - Condensed Matter and Complex Systems 20 (4), 493–501.

Joulin, A., Lefevre, A., Grunberg, D., Bouchaud, J.-P., 2008. Stock price jumps: news and volume play a minor role. Wilmott Magazine Sept/Oct, 1–7.

Kesten, H., 1973. Random difference equations and renewal theory for products of random matrices. Acta Math. 131, 207–248.

Kindleberger, C. P., Aliber, R. Z., Sep 2011. Manias, Panics and Crashes: A History of Financial Crises, Sixth Edition, sixth edition, revised edition Edition. Palgrave Macmillan.

Kobrak, C., Wilkins, M. (Eds.), Nov 2012. History and Financial Crisis: Lessons from the 20th century. Routledge.

Kolb, R. W., Jan 2011. The Financial Crisis of Our Time. Oxford University Press, USA.

Kong, J., Sarshar, N., Roychowdhury, V., 2008. Experience versus talent shapes the structure of the web. Proc. Natl. Acad. Sci. USA 105 (37), 13724–13729.

Kotharj, S., Warner, J. B., Jun 2007. Handbook of Corporate Finance, Volume 1: Empirical Corporate Finance, Ch. ECONOMETRICS OF EVENT STUDIES. In: Eckbo (2007).

Kyle, A., May 1985. Continuous auctions and insider trading. Econometrica 53 (6), 1315–1336.

LeRoy, S. F., 2008. Excess volatility tests. The New Palgrave Dictionary of Economics 2nd ed.

LeRoy, S. F., Porter, R. D., 1981. The present-value relation: Tests based on implied variance bounds. Econometrica 49, 555–574.

Lillo, F., Farmer, D., Mantegna, R., 2003. Master curve for price impact function. Nature 421, 129–130.

Loughran, T., Mcdonald, B., 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. The Journal of Finance 66 (1), 35–65.

Lux, T., Marchesi, M., 1999. Scaling and criticality in a stochastic multi-agent model of a financial market. Nature 397, 498–500.

lydia_emyeu, 2012. 苹果时钟应用图标被指照抄瑞士铁路时钟.
URL `http://www.weiphone.com/iPad/news/2012-09-21/The_Apple_the_clock_application_icon_is_alleged_copied_the_Swiss_railway_clock_487226.shtml`

Maillart, T., Sornette, D., Spaeth, S., von Krogh, G., 2008. Empirical tests of zipf's law mechanism in open source linux distribution. Physical Review Letters 101 (21), 218701(1)–218701(4).

Malevergne, Y., Pisarenko, V., Sornette, D., Mar. 2011. Testing the pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. Physical Review E 83 (3), 036111.
URL `http://link.aps.org/doi/10.1103/PhysRevE.83.036111`

Malevergne, Y., Saichev, A., Sornette, D., 2010. Zipf's law and maximum sustainable growth. Submitted to American Economics Review.
URL `http://ssrn.com/abstract=1083962`

Matthias W. Uhl, 2011. Reuters Sentiment and Stock Returns. KOF Working Papers 288.

McQueen, G., Roley, V., 1993. Stock prices, news, and business conditions. Review of Fin. Studies 6 (3), 683–707.

Mishkin, F. S., Feb 2011. Over the cliff: From the subprime to the global financial crisis. Journal of Economic Perspectives 25 (1), 49–70.

Parisi, D., Sornette, D., Helbing, D., 2013. Financial price dynamics vs. pedestrian counterflows: A comparison of statistical stylized facts. Physical Review E 87 (2), 012804.

Park, J., Sandberg, I. W., Jun 1991. Universal approximation using radial-basis-function networks. Neural Computation 3 (2), 246–257.

Podobnik, B., Horvatic, D., Petersen, A. M., Urosevic, B., Stanley, H. E., Oct 2010. Bankruptcy risk model and empirical tests. Proceedings of the National Academy of Sciences 107 (43), 18325 –18330.

Reuters.com, 2012. Business & Financial News, Breaking US & International News.
URL `http://www.reuters.com`

Robbins, L., Mar 2009. The Great Depression. Transaction Publishers.

Rogoff, K., Sep 2010. Three challenges facing modern macroeconomics. SSRN eLibrary.
URL `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1889366`

Romano, J., Wolf, M., July 2005. Stepwise multiple testing as formalized data snooping. Econometrica 73 (4), 1237–1282.

Rothbard, M. N., Jun 2000. America's Great Depression, 5th Edition. Ludwig Von Mises Institute.

Rotzinger, U., Benkö, T., 2012. Apple klaut bahnhofs-uhr der sbb. URL `http://www.blick.ch/news/wirtschaft/apple-klaut-bahnhofs-uhr-der-sbb-id2040821.html`

Rubinstein, A., Dec 1997. Modeling Bounded Rationality. The MIT Press.

Rubinstein, M. E., 2001. Rational markets: Yes or no? the affirmative case. Financial Analysts Journal 57 (3), 15–29.

Saichev, A., Malevergne, Y., Sornette, D., 2009. Theory of zipf's law and beyond. Lecture Notes in Economics and Mathematical Systems 632.

Samuelson, P. A., 1965. Proof that properly anticipated prices fluctuate randomly. Industrial Management Review 6, 41–49.

Samuelson, P. A., 1973. Proof that properly discounted present values of assets vibrate randomly. The RAND Journal of Economics 4, 369–374.

Shiller, R. J., Jun 1981. Do stock prices move too much to be justified by subsequent changes in dividends? The American Economic Review 71 (3), 421–436.

Shiller, R. J., Aug 2008. The Subprime Solution: How Today's Global Financial Crisis Happened, and What to Do about It, first edition Edition. Princeton University Press.

Simon, H., 1955a. On a class of skew distribution functions. Biometrika 52, 425–440.

Simon, H., 1960. Some further notes on a class of skew distribution functions. Information and Control 3, 80–88.

Simon, H. A., Feb 1955b. A behavioral model of rational choice. The Quarterly Journal of Economics 69 (1), 99–118.

Simon, H. A., Mar 1956. Rational choice and the structure of the environment. Psychological Review 63 (2), 129–138.

Sornette, D., 1998. Multiplicative processes and power laws. Phys. Rev. E 57 (N4), 4811–4813.

Sornette, D., 2003. Why Stock Markets Crash (Critical Events in Complex Financial Systems). Princeton University Press.

Sornette, D., Cauwels, P., Dec. 2012. The illusion of the perpetual money machine. Notenstein Academy White Paper Series (`http://www.notenstein.ch/en/news/publications/whitepapers` and `http://ssrn.com/abstract=2191509`).

Sornette, D., Davis, A. B., Ide, K., Vixie, K. R., Pisarenko, V., Kamm, J. R., 2007. Algorithm for model validation: Theory and applications. Proc. Nat. Acad. Sci. USA 104 (16), 6562–6567.

Sornette, D., Davis, A. B., Kamm, J. R., Ide, K., 2008. A general strategy for physics-based model validation illustrated with earthquake phenomenology, atmospheric radiative transfer, and computational fluid dynamics. Lecture Notes in Computational Science and Engineering, Book Series: Computational Methods in Transport: Verification and Validation F. Graziani and D. Swesty (Eds.), Springer-Verlag, New York (NY) vol. 62, 19–73 (http://arxiv.org/abs/0710.0317).

Sornette, D., Woodard, R., 2010. Financial Bubbles, Real Estate bubbles, Derivative Bubbles, and the Financial and Economic Crisis. Proceedings of APFA7 (Applications of Physics in Financial Analysis),"New Approaches to the Analysis of Large-Scale Business and Economic Data," Misako Takayasu, Tsutomu Watanabe and Hideki Takayasu, eds., Springer (`http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1407608`).

Sornette, D., Zhou, W.-X., 2006. Importance of positive feedbacks and overconfidence in a self-fulfilling ising model of financial markets. Physica A 370, 704–726.

Subrahmanyam, A., 2008. Behavioural finance: A review and synthesis. European Financial Management 14 (1), 12–29.

Tetlock, P. C., Saar-Tsechansky, M., Macskassy, S., 2008. More than words: Quantifying language to measure firms' fundamentals. The Journal of Finance 63 (3), 1437–1467.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. The Review of Financial Studies 21 (4), 1455–1508.

Wiesinger, J., Sornette, D., Satinover, J., Jan 2012. Reverse engineering financial markets with majority and minority games using genetic algorithms. Computational Economics DOI 10.1007/s10614-011-9312-9 (http://ssrn.com/abstract=1553821).

Wonders, G., May 2010. The Imminent Crisis: Greek Debt and the Collapse of the European Monetary Union. CreateSpace Independent Publishing Platform.

Yahoo.com, 2012. Yahoo! finance - business finance, stock market, quotes, news.
URL `http://finance.yahoo.com`

Zhang, Q., Sornette, D., Satinover, J., 2013. Mixed-game virtual stock markets combining minority, delayed minority, majority and dollar agent-based models. working paper ETH Zurich.

Zipf, G., 1949. Human behavior and the principle of least effort. Addison-Wesley Press, Cambridge, Mass., USA.

Curriculum Vitae

Qunzhi Zhang
Gender: male, married, two sons
Phone: +41 44 632 92 87
Email: qzhang@ethz.ch

EXPERIENCE

2008-present: research on the relations between financial markets and social networks using event-based studies, agent-based models, machine learning and econometric/statistical methods

- Study of the relationships between news and stock prices
- Reverse engineering of stock markets with agent-based models
- Investigation of the origin of the ubiquitous Zipf's law in social networks
- Collection of data from social networks by developing a mediawiki based website with a prediction market
- Development of an earthquake data collection, presentation and analysis website for young students in Switzerland

1996-2007: worked mainly as a software product manager and software engineer, developing large-scale websites, server-side and desktop software products, and online games.

EDUCATION

ETH Zürich, Ph.D. candidate at the Chair of Entrepreneurial Risks, 2013
ETH Zürich, Msc. ETH in Management, Technology and Economics, 2009
East China University of Science and Technology, Shanghai, Bachelor in Chemical Engineering, 1996

SKILLS

- Language: Bilingual skills in Chinese (native) and English (proficient)
- Scientific skills:
  - ■ Build and analyze agent-based models to reverse engineer stock markets
  - ■ Analyze big economic, financial and social data with statistics and econometrics methods, and apply empirical tests with models in economics, finance and econophysics.

  - Scientific programming with c/c++, python, R, matlab, as well as parallel scientific computing
- Management skills:
  - Software product management
  - Systematically studied theories about general management, strategic management, marketing management, financial management, risk management, information system management, etc
- Computer skills:
  - Software engineering, including the analysis, design, implementation, testing, and deployment processes, with different kinds of software products, such as business products, online games, websites, etc
  - Programming with many popular programming languages such as c/c++, java, python, scala, pascal, php, javascript, etc
  - Database management and programming
  - Big data collection from Internet