Econophysics: An Experimental Course for Advanced Undergraduates in the Nanyang Technological University

IIM Kozhikode Society & Management Review
2(2) 79–99
© 2013 Indian Institute
of Management Kozhikode
SAGE Publications
Los Angeles, London,
New Delhi, Singapore,
Washington DC
DOI: 10.1177/2277975213507764
http://ksm.sagepub.com

Siew Ann Cheong

Abstract

In this article, I have described an experimental econophysics course for advanced undergraduates in the Nanyang Technological University (NTU) in Singapore. This course was offered for the first time in Semester II (January–April 2013) of the 2012/2013 academic year, and is a student-led participatory learning experiment aimed at developing an open-source textbook on econophysics. The course covered an introduction to the history of econophysics, a review of probability and statistics, statistical properties of single high-frequency financial time series, correlations within a high-frequency financial time series cross section, and agent-based models of financial markets. In spite of the heavier workload resulting from the experimental course format, feedbacks from the nine students taking the course were highly positive, and an improved version of the econophysics course will be offered again in the first half of 2014.

Keywords

Econophysics, advanced undergraduate course, Nanyang Technological University, open textbook

Introduction

Econophysics as a field of research is now 18 years old, if we count the Aug 1995 'Dynamics of Complex Systems' conference in Kolkata, India as the official start of the field (Stanley *et al.*, 1996). Following this conference, the inaugural econophysics conference was held in 1998 in Budapest, Hungary. In 2005, the first Econophysics Colloquium was held in Canberra, Australia, while the first Econophys-Kolkata meeting was held in Kolkata, India. Both have since grown into annual events, with the latest edition of the former held in Pohang, Korea, 29–31 July 2013, and the latest edition of the latter held in Kolkata, India 8–12 November 2012. With the growing interest in econophysics in Asia, the International Conference on Econophysics and the Asia-Pacific Econophysics Conference were also launched in June 2011 and September 2012 respectively. In 2014, we will see at least three econophysics meetings in Asia: Econophys-Kolkata in March, the International Conference on Econophysics in June, and the Econophysics Colloquium in November. Besides the proliferation of conferences, we can judge from the number of econophysics papers published every year that the field has matured. Looking especially at the recent papers (Moat *et al.*, 2013; Preis, Moat, & Stanley, 2013), it is also clear that many exciting research themes have emerged, waiting for the econophysics community to explore in detail.

In an opinion paper published in 2006, Gallegati *et al.* expressed worries about the state of econophysics at that time (Gallegati, Keen, Lux, & Omerod, 2006). Econophysicists were not paying enough attention to the economics literature, not sufficiently rigorous in their statistical analysis, overselling the notion of universality in human activities and organizations, and overconfident in statistical physics models based on energy conservation. In the years following, this situation has certainly changed. Even though econophysicists are still not publishing in the most prestigious economics journals, they are slowly penetrating finance journals (Bouchaud, Gefen, Potters, & Wyart, 2004). More importantly, there are increasing collaborations between economists and econophysicists, who are traditionally from nuclear physics, nonlinear physics, and statistical physics. This is a very encouraging trend.

Clearly, financial markets and economies are complex. More mathematical sophistication, not less, is needed to understand them. Nonlinear approaches, in addition to linear approaches, will be needed. After the Global Financial

Siew Ann Cheong is at the Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore. He is also a member of the Complexity Program in Nanyang Technological University.

Crisis of 2007–2009, part of the economics community realized these urgent needs, and called for more efforts devoted into developing the field of complexity economics (Arthur, 1999; Durlauf, 2005; Gallegati & Kirman, 2012). The Institute for New Economic Thinking (INET), created in late 2009 by a US$50 million seed funding by business magnate and philanthropist George Soros, now stands as the beacon for this movement within economics. However, there are also many economists, including very influential ones, who are reluctant to move on from equilibrium ideas about the economy. Financial institutions and regulators will obviously not wait for this academic debate to conclude. They need answers, and they need them now, however imperfect they may be, so long as they are more relevant to the problems they wrestle with day in day out.

This presents a fantastic opportunity to the econophysics community. Continuing the trend before the Subprime Crisis (*The Economist*, 2007), US undergraduates continue to pursue a major in economics (*The Brown Daily Herald*, 2013; University of Wisconsin-Madison, 2013). Although no systematic study has been done in Europe and Asia, we believe that interest in economics and finance is also growing. This is to be expected, when news media devote so much space to talking about the global financial crisis. At the bachelor's level, we are also inclined to believe an economics major training will not adequately prepare graduates to cope with economics and finance problems in the real world. Most of these graduates simply do not have the requisite mathematical sophistication. Econophysics graduates, on the other hand, trained in the mathematical methods of classical and quantum particles and fields, and also receive dedicated instruction to bridge the cultural divide between physical and social sciences, may be much better placed for growing employment opportunities in banks as well as government.

With these considerations at the back of our mind, we started an econophysics advanced undergraduate course, PH4410 Econophysics, in the Physics and Applied Physics degree programmes in the Nanyang Technological University, Singapore. In the spirit of encouraging undergraduates to take on a more interdisciplinary view of the real world, this course is also open to students from the Business Management (Banking and Finance), Computer Science, Economics, Mathematical Sciences, and Mathematics and Economics degree programmes, offered by the Nanyang Business School, the School of Computer Engineering, the School of Humanities and Social Sciences, and the School of Physical and Mathematical Sciences. With this course as the core, we also aim to build up an econophysics concentration consisting of two Physics and Applied Physics courses, and one course each from the Nanyang Business School (Banking and Finance) and the School of Humanities and Social Sciences (Economics). Eventually, when we are certain the demand is there, and the graduates are well received in the job market, we will propose an Econophysics degree programme, with an enrolment of 50 to 100 students every year.

In the sections to follow, we will give a brief introduction to the history of the Nanyang Technological University, its current academic makeup, and a brief description of its teaching and research. We will then describe the design of the econophysics course, giving an outline of the topics covered, its experimental participatory course structure, the learning objectives and how students are assessed. Finally, we will present outcomes from running this course the very first time, how we plan to fine tune it the second time round, before we conclude.

The Nanyang Technological University

The Nanyang Technological University (NTU) occupies a 200-hectare main campus in the west of Singapore. It is one of the four publicly funded universities in Singapore, the other three being the National University of Singapore, the Singapore Management University, and the Singapore University of Technology and Design (in collaboration with the Massachusetts Institute of Technology and Zhejiang University). The university started out as Nanyang University in 1955 with donations of money and land from Singaporean Chinese philanthropists. In this original incarnation, all courses were taught in Chinese. In 1980, the Singapore government ordered it to be merged with the Singapore University to form the National University of Singapore, because of the increasing difficulty for Chinese-educated university graduates to find employment.

At the same time, Singapore was rapidly growing its manufacturing industry, and needed a large number of engineers. The large engineering cohort received foundational instructions in the National University of Singapore for the first year, and in their second year, part of this cohort remained in the Faculty of Engineering in the National University of Singapore, while the rest continued their engineering education in the campus of the former Nanyang University, now renamed the Nanyang Technological Institute. Later, recognizing that a purely engineering campus with mostly young men posed a social problem that would be difficult to solve later on, Accounting and Business Management with mostly young women

undergraduates were moved from the National University of Singapore to the Nanyang Technological Institute in 1987. Then in 1991, the National Institute of Education merged with the Nanyang Technological Institute to form NTU. In the same year, the Wee Kim Wee School of Communications and Information was also created.

In 2002, the International Academic Advisory Panel commissioned by the Singapore Ministry of Education reviewed the higher education scene in Singapore, and among other things recommended to make NTU a comprehensive university. This led to the creation of a new School of Biological Sciences, a new School of Art, Design and Media, a new school of Physical and Mathematical Sciences, a new School of Humanities and Social Sciences, and the S. Rajaratnam School of International Studies. Starting last year, our university also became home to the Lee Kong Chian School of Medicine, which is a joint venture between NTU and Imperial College. At present, the various schools are organized into the College of Business, the College of Engineering, the College of Humanities, Arts and Social Sciences, the College of Science, National Institute of Education, S. Rajaratnam School of International Studies, and the Lee Kong Chian School of Medicine.

NTU started out with highly disciplinary research. However, research funding in Singapore is primarily driven by problems rather than scientific disciplines. As such, research in our university has evolved to become more interdisciplinary in nature. In particular, for energy and water research, it has organized its experts into the Energy Research Institute at Nanyang (ERI@N) and the Nanyang Environment and Water Research Institute (NEWRI), both of which are funded by the university and competitive research grants from various agencies. The university also hosts two highly interdisciplinary National Research Foundation (NRF) Research Centres of Excellence, the Earth Observatory of Singapore and the Singapore Centre of Environmental Life Science and Engineering. The university has also recently started a Complexity Programme modelled after the Santa Fe Institute, drawing experts from the various schools to look into interdisciplinary problems related to natural and man-made complex systems. On a whole, the university has risen steadily in terms of world ranking. As of 2012, it is ranked 86 in the Times Higher Education (THE) world university ranking, and 47 in the Quacquarelli Symonds (QS) world university ranking. It is one of the 10 fastest growing universities in the world.

The Nanyang Technological University offers 54 undergraduate degree programmes and 58 masters programmes, in addition to the PhD degree by research in the Interdisciplinary Graduate School as well as all schools and institutes. It has 23,500 undergraduate students, 10,000 graduate students, and 4,200 faculty and staff. Teaching remained highly disciplinary until recently, when the university started responding quickly to recent developments in industry, and regularly offers new degree programmes as well as updates to existing programmes at both the undergraduate and graduate levels. Over the last few years, various double degree and double major programmes such as Business and Computing, Mathematics and Economics, Biology and Psychology have seen strong interests from applicants. Besides encouraging schools to abolish traditional boundaries and mount joint courses and minors that will help our graduates develop an interdisciplinary edge in their approach to real-world problems, the university also has three elite programmes that are by design interdisciplinary. These are the C N Yang Scholars Program, the Renaissance Engineering Program, and the University Scholars Program. Within such a research and teaching environment, the development of the econophysics course, concentration, as well as degree programme is favoured at the highest level of the university administration.

PH4410 Econophysics Course Organization

As the course code suggests, my econophysics course is targeted at fourth-year advanced Physics and Applied Physics undergraduates, taught over a 13-week semester. The pre-requisites for this course are PH3201 Statistical Mechanics I and PH3502 Chaotic Dynamical Systems. These pre-requisites are more suggestive than restrictive, and they can be waived for Physics and Applied Physics students eager to take PH4410 Econophysics. For Computer Science, Business Management (Banking and Finance), Economics, Mathematical Sciences, Mathematics and Economics students wanting to take the course, these pre-requisites can similarly be waived provided they are judged to have adequate mathematical and statistical background.

Course Outline

The course is organized into three parts: (a) high-frequency financial time series (4 weeks); (b) high-frequency financial time series cross section (5 weeks); and (c) agent-based models of markets and economies (4 weeks). There is a

two-hour lecture every week, and another contact hour every week that is used every other week for tutorials, and the rest of the time for lectures. For this first offering of the course, I adopted an experimental course structure, based on the concept of an open textbook.

I have tentatively planned for the open textbook to have seven chapters, in line with the three parts of the course. These chapters are as follows:

1. History of Econophysics
2. Preliminaries on Probability and Statistics
3. Properties of High-frequency Financial Time Series
4. Memory in Financial Markets
5. Financial Time Series Cross Section and Random Matrix Theory
6. Cross Correlation Filtering and Time Series Clustering
7. Agent-based Models of Financial Markets

In Chapter 1, 'History of Econophysics', I aim to provide the historical background to the birth of econophysics, starting from the beginnings of classical physics and classical economics. I explained how the mathematization of classical mechanics started in 1687 with Sir Isaac Newton's formulation of the laws of motion and the development of calculus, was brought to fruition by Lagrange and Laplace in the form of analytical mechanics. With this, a complete mathematical description of the dynamics of point particles was in place, and physicists went on to obtain the mathematical description of classical fields before quantum physics was discovered. Classical economics started later with Adam Smith in 1776, it became an independent discipline through the works of Thomas Malthus and John Stuart Mill, among others, and started its mathematization programme in 1874 with Leon Walras, in a development that parallels physics. However, the mathematical trajectories of the two disciplines soon parted ways. While physics continued using more and more sophisticated mathematical languages and methods to describe quantum indeterminism and chaos in nonlinear systems, economics continued working with linear mathematics and equilibrium.

These two disciplines remained apart for several decades, before stock exchanges began computerizing their operations. This gave rise not only to faster trades (and hence a quicker redistribution of wealth), but also to huge volumes of transaction data. Physicists like Eugene H. Stanley and his collaborators then started looking at these data, drew quantitative conclusions about the character of stock market returns, and ushered in the era of econophysics. In this sense, physics and economics both grew out of their philosophical roots when the small amount of data that became available drove the mathematization of the two disciplines, to build quantitative theories capable of explaining such data. The birth of econophysics, on the other hand, was driven first by the sudden availability of large amounts of data, and then later by the availability of computing resources. The large data volume pointed to inadequacies behind the standard random walk paradigm, whereas computing power led to the development of exploratory agent-based models and simulations.

In the second chapter, I very quickly reviewed the preliminaries on probability and statistics that are needed to appreciate developments in econophysics. Formally going through these mathematical preliminaries is important, because the probability and statistics preparations of different degree programmes can vary a lot. I started out introducing the concept of a random variable, the notion of a probability distribution, and gave examples of common discrete and continuous distribution functions before going on to explain how expectation values are computed. I also reviewed the concepts behind parameter estimation, using linear regression as an example, before reminding students what hypothesis testing is all about. Finally, I talked about the relation behind information and entropy, and how to bring in concepts from equilibrium statistical mechanics.

The third and fourth chapters are on individual financial time series, and I started off talking about Louis Bachelier's PhD thesis based on the assumption that stock returns are generated by a Brownian process, and how this assumption lead to the testable hypothesis that stock returns should follow a Gaussian distribution. I then explained to students that this is indeed observed for returns computed on time horizons much longer than a day, but show serious discrepancies at shorter time horizons. In particular, returns of real stocks have fat tails that fall off asymptotically as power laws. I illustrated this using the daily stock prices of Apple Inc downloaded from Yahoo! Finance. This illustration is made into a series of three YouTube videos (http://www.youtube.com/watch?v=cRPYF0E47hc, http://www.youtube.com/watch?v=b-P8NSzhCUc, and http://www.youtube.com/watch?v=QdHXhg3RC_M).

In the fourth chapter, I talked about the other assumption behind the Brownian process model, that is, returns at different time lags should be uncorrelated. In actual fact, they show qualitatively different autocorrelation at different time lags. I defined the autocorrelation, and described the general finding that returns have strongly negatively autocorrelations at the time scale of minutes, and weak

positive autocorrelation at the time scale of days and weeks (Cont, 2001; Lewellen, 2002). Such autocorrelation manifests itself as momentum or mean reversion in the financial market. I also talked about the Hurst exponent, another measure of long-range autocorrelation, or memory in the market (Alvarez-Ramirez, Alvarez, Rodriguez, & Fernandez-Anaya, 2008; Cajueiro & Tabak, 2004; Carbone, Castelli, & Stanley, 2004; Di Matteo, Aste, & Dacorogna, 2003, 2005; Eom, Choi, Oh, Jung, 2008; Feigenbaum & Freund, 1996; Grech & Pamula, 2008; Lux, 1996; Peters, 1989; Matos, Gama, Ruskin, Sharkasi, & Crane, 2008). Finally, I talked about nonstationarity in financial time series, and how segmentation methods can be useful in understanding the distribution and meaning of this nonstationarity (Cheong *et al.*, 2012; Tóth, Lillo, & Farmer, 2010; Vaglica, Lillo, Moro, & Mantegna, 2008; Wong, Lian, & Cheong, 2009; Zhang *et al.*, 2011).

The fifth and sixth chapters of the planned open textbook are on financial time series cross section. In the fifth chapter, I introduce various cross correlation measures between time series, before moving on to the Random Matrix Theory (RMT), as a means for discriminating between pure noise and real signal in the time series cross section (Laloux, Cizeau, Bouchaud, & Potters, 1999; Plerou, 2002). In the sixth chapter, we follow up by talking about correlation filtering, from naïve threshold filtering (Kim & Jeong, 2005), to RMT filtering (Laloux, Cizeau, Potters, & Bouchaud, 2000), to filtering through the use of a minimal spanning tree (MST) (Mantegna, 1999) or a planar maximally filtered graph (PMFG) (Tumminello, Aste, Di Matteo, & Mantegna, 2005). I then introduced the concept of time series clustering (Liao, 2005), to discover the number of independent collective variables that can be used to simplify the description of the financial market. Various clustering methods were introduced.

Finally, in the last and seventh chapter, I talked about agent-based models of financial markets. We started first on the history of agent-based models, from both the economists' and physicists' perspective. Economists' agent-based models tend to be top-down, and involve many assumptions on the behaviour of agents. Although these look more realistic, their simulation results are also hard to analyze. Here I used the Santa Fe artificial market model (Palmer, Arthur, Holland, LeBaron, & Tayler, 1994) as the main example. In contrast, physicists' agent-based models tend to be bottom-up, that is, toy models. As a specific example, I talked about the Ising model, how to simulate it, and how its critical behaviour mirrors the tendency of markets to crash.

Course Structure

In this course, students are assessed based on four grading components. In the first grading component, students attend lectures, and are also divided into groups of three to five to read up the primary literature (original research articles, reviews, monographs) and help me develop chapters in the open textbook. I explained to students at the start of the course that they are all authors of the open textbook, and as instructor I merely play the role of an editor. I also explained to them that by reading the primary literature and summarizing it into a pedagogical form, they would truly learn the knowledge and wisdom contained in the econophysics literature. Since this would still be very challenging, I explained to the students that my lectures serve as road maps on what to read, and what level of details to read down to. I also dedicated one tutorial hour to explain how they can search for papers and reviews using Google Scholar and Web of Science using different combinations of key words. I told my students to limit their searches to highly cited papers. This will introduce biases, but I explained that we would fix these biases next year when the course runs again, because the students will have to do this again. Finally, I explained to them how they could synthesize different readings to produce coherent summaries.

In the second grading component, students develop end-of-chapter exercises and their solutions. I explained to them that they should develop a few of each category of questions. The categories are: (a) conceptual, (b) methodological, and (c) applications to real data. Again, this is a hard task for undergraduates, even if they work in groups. Therefore, I devoted one tutorial hour to go through a sample exercise shown in Appendix 1. Further, since these end-of-the-chapter exercises are not examination questions, each exercise can be self-contained, and not have too many parts.

In the third grading component, students develop MATLAB programmes that illustrate some of the methods discussed in class. Each group is only required to develop one MATLAB programme over the whole semester, pertaining to the open textbook chapter they are working on. Again, I devoted one tutorial hour to talk about this. The example used is shown in Appendix 2.

The final grading component is the final examination. This is a closed book examination, and students are told that there will be one question from each major part of the course, and a fourth question integrating more probability and statistics calculations. The examination is shown in Appendix 3.

As its structure suggests, this course emphasizes participatory learning by the student, in that they are primarily responsible for shaping the materials (textbook, problems, MATLAB demos) used in the course, with some guidance from me (in the form of roadmap lectures, and dedicated workshop-like tutorials). This is not the first time I have used such a student-driven course structure: I have made students develop lecture notes and exercises and MATLAB demos in the two times that I had taught a computational physics course previously. However, this is the most ambitious implementation of student-led learning that I have attempted, to eventually come up with an open source textbook on econophysics.

Outcomes

A total of nine students registered for PH4410 Econo-physics. Seven were fourth-year Physics and Applied Physics majors, one was third-year Physics and Applied Physics major, while the last student was a third-year Economics student on study exchange in NTU. Because of the small class, the students were organized into two groups for their participatory activities.

Open Textbook

In my original vision for the course, after students submitted their first drafts of the chapters in the middle of the semester, the open textbook would go live on a WordPress or equivalent platform. However, this goal proved to be unrealistic, as the students needed more than half a semester just to get their readings done. Moreover, the students chose to develop Chapters 6 and 7, and hence it would look weird launching an open textbook with the first five chapters missing. In the end, the students went through only one revision of the chapters. When this course runs again next year, I will load these two chapters onto a private website for the next batch of students, and get them to develop the remaining chapters. Should the progress be good, I will launch the open textbook. Otherwise, I will have one more batch of students work on the open textbook before it is made publicly available. In the long run, besides acting as the main reference for students taking this course, I will continue to involve students in updating the textbook, and perhaps developing special topics that are of current interest in econophysics research.

For the open textbook chapters, students were told to submit a reading list for approval one week after the tutorial workshop on literature search. The guideline I gave the students was to read only the most highly cited papers that are relevant to the two chapters. Further, the students were told to read at least 10 papers each. The approved reading lists are shown in Appendix 4. After reading these monographs, reviews, and papers, students summarized them using a pedagogical style into first drafts of the two respective chapters. I then commented on the first drafts, before the students revised the two chapters. Snippets of Chapters 6 and 7 are shown in Appendix 5.

End-of-Chapter Exercises

As mentioned in the 'Course Structure' section, after a tutorial workshop on how to develop problems that would clarify and reinforce the econophysics concepts and methods covered, students also developed end-of-chapter exercises for the chapters they are responsible for. Examples of these are shown in Appendix 6.

MATLAB Demos

After a tutorial workshop on how to develop MATLAB programmes that can help illustrate important econophysics concepts, students spent two to three weeks developing MATLAB demos for the open textbook chapters they are working on. These demos are listed in Appendix 7.

Conclusions

To conclude, I taught an experimental econophysics course in Semester II (January–April 2013) for the 2012/2013 academic year in the NTU's Physics and Applied Physics degree programme. NTU values interdisciplinarity in teaching, and thus PH4410 Econophysics is open to Business Management (Banking and Finance), Computer Science, Economics, Mathematical Sciences, Mathematics and Economics majors, in addition to Physics and Applied Physics majors. This course forms the core of the Econophysics concentration that we are currently offering, and in future may seed several courses for an Econophysics degree programme.

Nine students completed this course, and gave very positive feedbacks. In this first offering of the course, students were exposed to a participatory learning format, where they are guided by lectures to read up the original research papers in the field, to produce pedagogical materials for an

open textbook that we plan to eventually create. They also developed end-of-chapter exercises for this open textbook, as well as MATLAB demos to better illustrate some of the methods discussed in class. Unfortunately, because only the last two chapters (out of seven planned) were developed, the open textbook project could not go live. I am now planning a second offering of PH4410 Econophysics in Semester II of the 2013/2014 academic year, where I will again engage students in developing the remaining chapters of the open textbook. Should enough materials be developed, I hope to launch the open textbook at the end of April 2014.

Appendix 1. Sample Exercise Shown in Tutorial

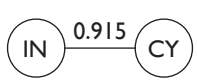
The Pearson cross correlations between the 10 Dow Jones US Industry Sector indices during a particular period are given below.

	BM	CY	EN	FN	HC	IN	NC	TC	TL	UT
BM		0.872	0.825	0.898	0.726	0.900	0.789	0.818	0.709	0.832
CY			0.753	0.898	0.826	0.915	0.876	0.856	0.768	0.835
EN				0.750	0.663	0.776	0.720	0.745	0.607	0.759
FN					0.771	0.889	0.845	0.827	0.741	0.814
HC						0.827	0.913	0.804	0.772	0.770
IN							0.861	0.877	0.808	0.842
NC								0.852	0.783	0.819
TC									0.769	0.783
TL										0.729
UT										

Draw the minimal spanning tree (MST) of these cross correlations.

Solution. Since we get the same MST whether we work with the ultrametric distances $D_{ij} = \sqrt{2(1-C_{ij})}$ and start adding links from the one with the minimum distance, or we work directly with the cross correlations C_{ij} and start adding links from the one with the maximum correlation, let us work with C_{ij}.

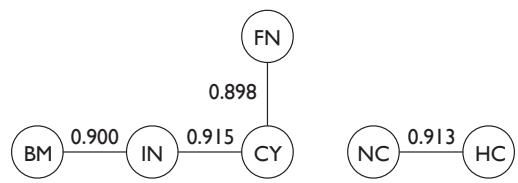
From the table above, we see that the maximum correlation is $C_{ij} = 0.915$, between IN and CY. We therefore connect these two industry sectors, as shown below.



From the table above, we see that the next largest correlation is $C_{ij} = 0.913$, between HC and NC. Since these two industry sectors are new to the MST, no cycles are possible, and we therefore connect these two industry sectors, as shown below.

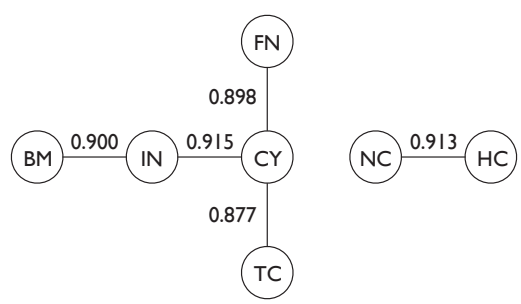
At this point, the four sectors are not all connected yet. From the table above, we see that the next largest correlation is $C_{ij} = 0.900$, between IN and BM. This link connects an existing industry sector, IN, to a new industry sector, BM, and hence there can be no cycles. We therefore draw a link between these two industry sectors, as shown below.

Next, we see from the table above that the next largest correlation is $C_{ij} = 0.898$. However, there are two pairs of industry sectors with this cross correlation. The first is FN and BM, and the second is FN and CY. In both pairs, only FN is new to the MST. Both BM and CY have already been added to the MST in previous iterations. We see also that we can add the FN-BM link or the FN-CY link, but not both, as doing so would introduce a cycle to the MST. In rare cases like this, we can choose to link FN and BM, or FN and CY. At the level of the MST, this choice is not particularly important, and we shall see elsewhere in the course that both links would be included if we decide to draw the planar maximally filtered graph (PMFG) of the cross correlations. With this in mind, we find the still incomplete MST shown below.



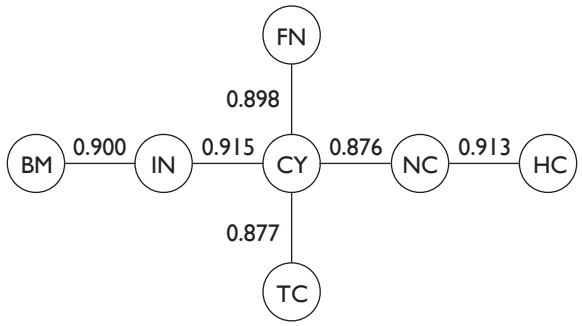
We are now more than halfway through the MST construction, and only four more industry sectors (EN, TC, TL, UT) need to be added. From the table above, we see that the next largest cross correlation is $C_{ij} = 0.889$, having already ignore the cross correlation $C_{ij} = 0.898$ between FN and BM because a cycle will be formed. The next largest cross correlation is between IN and FN, but we cannot draw a link between them, because a cycle will again be formed if we do.

Going further down, we find $C_{ij} = 0.877$, between TC and IN. IN is already in the incomplete MST, but TC is new, and thus no close cycle will be formed by drawing a link between these two industry sectors. We therefore accept the link, to get the updated MST shown below.



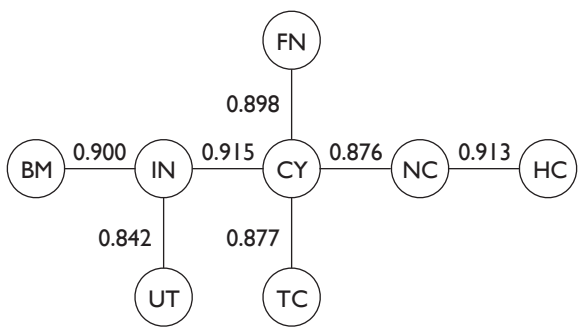
Continuing, we find the next largest correlation to be $C_{ij} = 0.876$, between CY and NC. Both industry sectors are not new to the MST, but they are in disconnected subgraphs, and hence no cycle will be formed if we draw a link between them.

Accepting this CY-NC link, we now have the incomplete MST shown below.

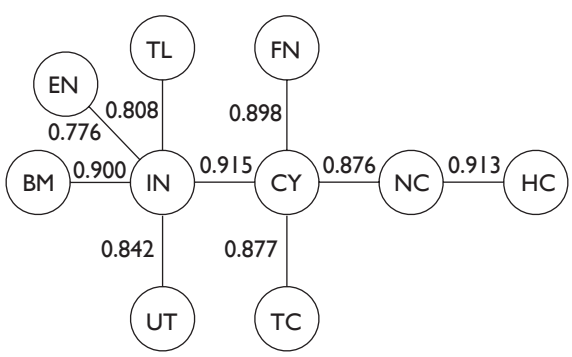


The incomplete MST is now a fully connected graph. However, we are still missing three more industry sectors, EN, TL, and UT. We continue descending the cross correlations, but the next five correlations (C_{ij} = 0.872 between BM and CY, C_{ij} = 0.861 between IN and NC, C_{ij} = 0.856 between CY and TC, C_{ij} = 0.852 between NC and TC, and C_{ij} = 0.845 between FN and NC) lead to cycles if we accept the links. Thus, we have to reject these links.

From the table above, we see that the next largest correlation after these is C_{ij} = 0.842 between IN and UT. Since UT is new to the MST, adding a link between it and IN will not result in a cycle. We therefore accept the link to get the incomplete MST shown below.



Since we are now missing only EN and TL, we no longer need to systematically go through all cross correlations. Instead, we can just focus on the cross correlations involving EN and TL. EN is most strongly correlated with IN, with C_{ij} = 0.776, while TL is most strongly correlated with IN, with C_{ij} = 0.808. Adding these two links, the completed MST looks like that shown below.



Appendix 2. Sample MATLAB Demo Shown in Tutorial

Write a MATLAB programme to simulate a one-dimensional Ising market model

$$H = -\sum_{i=1}^{N} J s_i s_{i+1} + h s_i$$

on a periodic chain with N spins $s_i = \pm 1$, where J is the coupling between neighbouring spins (how strongly a trader is influenced by traders he/she knows), and h is the external magnetic field (how strongly past successes influence present actions) using the Metropolis algorithm.

Solution. Before we start coding, we should understand that within the Metropolis algorithm, in each Monte Carlo step, we pick a random spin s_i and flip it. This trial move is accepted with unit probability if $\Delta E = E_f - E_i < 0$, and accepted with probability $\exp(-\beta \Delta E)$ if $\Delta E = E_f - E_i > 0$. Here, $\beta = 1/T$ is the inverse temperature of the simulation. The lower the temperature, the more rational traders are in the market.

Since only s_i is flipped, we have $\Delta E = +2Js_i(s_{i-1} + s_{i+1}) - 2hs_i$.

The MATLAB programme that we can use to simulate this at $\beta = 1$ is shown below.

```
% number of traders
N = 1000;
% duration of simulation
T = 1000000;
% simulation parameters
J = 1.0;
h = 0.5;
beta = 1.0;
% initial state
s = 2*(rand(1, N) < 0.5) – 1;
% Metropolis simulation
for t = 1:T
    % pick random trader
    i = floor(N*rand()) + 1;
    % calculate energy change
    if i == 1
        dE = 2*J*s(i)*(s(N) + s(i + 1)) – 2*h*s(i);
    else
        if i == N
            dE = 2*J*s(i)*(s(i – 1) + s(1)) – 2*h*s(i);
        else
            dE = 2*J*s(i)*(s(i – 1) + s(i + 1)) – 2*h*s(i);
        end
    end
    % accept flip?
    if dE < 0
        s(i) = –s(i);
    else
        if rand() < exp(–beta*dE)
            s(i) = –s(i);
        end
    end
end
```

Appendix 3. Final Examination for PH4410 Econophysics in Semester II, AY 2012/2013

The final examination for PH4410 Econophysics consists of four questions to be answered within two and a half hours. The questions are:

1. A pedestrian guide to random matrix theory.

Random matrix theory, first developed in nuclear physics, has led to exciting developments in wave optics, condensed matter physics, as well as number theory and statistics. There are several universal families of random matrices. A random symmetric 2 × 2 matrix

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}$$

is a member of one such family, the Gaussian orthogonal ensemble (GOE), if it satisfies three conditions:

1. for any 2 × 2 orthogonal transformation $\mathbf{OO}^T = \mathbf{1} = \mathbf{O}^T\mathbf{O}$, $\mathbf{A}' = \mathbf{OAO}^T$ is also a member of the GOE;
2. the matrix elements a_{11}, a_{12}, and a_{22} are statistically independent; and
3. the probability density $P(\mathbf{A})\, d\mathbf{A}$, where $d\mathbf{A} = da_{11}\, da_{12}\, da_{22}$ is given by

$$P(\mathbf{A})\, d\mathbf{A} \propto \exp(-a\, \mathrm{Tr}\mathbf{A}^2 + b\, \mathrm{Tr}\mathbf{A} + c\,)\, d\mathbf{A}, \qquad (1)$$

where $a > 0$, b, and c are real numbers.

a. Compare $P(\mathbf{A})$ in (1) with the normal distribution

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],$$

and explain the significance of a, b, and c.

(3 marks)

b. Random matrix theory predicts that, while a_{11}, a_{12}, and a_{22} are statistically independent, the eigenvalues λ_1 and λ_2 are strongly correlated. To see this, let us write $P(\mathbf{A})\, da_{11}\, da_{12}\, da_{22} = Q(\lambda_1, \lambda_2, \theta) d\lambda_1\, d\lambda_2\, d\theta$ in terms of the eigenvalues λ_1 and λ_2. θ is the angle the normalized eignvector u_1 makes with the x axis (see Figure 1).

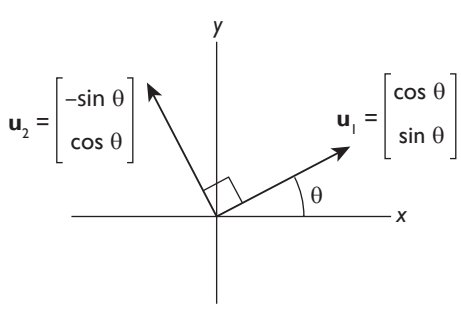


Figure 1.

(i) Show that the independent matrix elements of $\mathbf{A}$ can be written as

$$a_{11} = \lambda_1 \cos^2\theta + \lambda_2 \sin^2\theta,$$
$$a_{12} = (\lambda_1 - \lambda_2)\cos\theta \sin\theta,$$
$$a_{22} = \lambda_1 \sin^2\theta + \lambda_2 \cos^2\theta$$

in terms of λ_1, λ_2, and θ.

(7 marks)

(ii) Write down the Jacobian $\|J\|$ associated with the change of integration variables from (a_{11}, a_{12}, a_{22}) to $(\lambda_1, \lambda_2, \theta)$, and show, without explicitly evaluating the determinant, that $\|J\| \propto |\lambda_1 - \lambda_2|$.

(10 marks)

(iii) Hence, or otherwise, show that the joint probability distribution function for λ_1 and λ_1 is

$$R(\lambda_1, \lambda_2)\, d\lambda_1\, d\lambda_2 = \int_0^{2\pi} Q(\lambda_1, \lambda_2, \theta) d\lambda_1\, d\lambda_2\, d\theta$$
$$\propto |\lambda_1 - \lambda_2| \exp[-a(\lambda_1^2 + \lambda_2^2) + b(\lambda_1 + \lambda_2)]\, d\lambda_1\, d\lambda_2.$$

(5 marks)

2. Minimal spanning tree and minimal spanning forest.

The minimal spanning tree (MST) is a very convenient tool to visualize the cross correlations between different financial instruments.

a. Describe how you would construct the MST of N stocks, starting from their Pearson cross correlations $-1 \le C_{ij} \le 1$.

(10 marks)

b. Figure 2 shows the MST of the 36 Nikkei 500 industries constructed from their indices in the second half of 2007. The thickness of the link between two industries i and j is proportional to the cross correlation C_{ij} between them. The number beside each link tells us the order in which the links were added to the MST.

Explain the significance of the hubs NELI and NMAC, in relation to peripheral nodes like NFIN and NAIR.

(5 marks)

c. Sketch the minimal spanning forest of the 36 Nikkei 500 industries.

(10 marks)

3. Price formation through a double auction market.

Consider a simple model of price formation through a double auction market with two groups of traders: one which puts in *bid* orders uniformly distributed between the limits p_1 and p_3, and another which puts in *ask* orders uniformly distributed between the limits p_2 and p_4, with $p_1 < p_2 < p_3 < p_4$.

(a) If we simulate such a model to generate a price time series $p(t)$, what is the average price $< p >$ that we should expect?

(5 marks)

(b) What is the variance we expect to find in the price time series?

(5 marks)

(c) What is the average bid-ask spread we expect to find?

(5 marks)

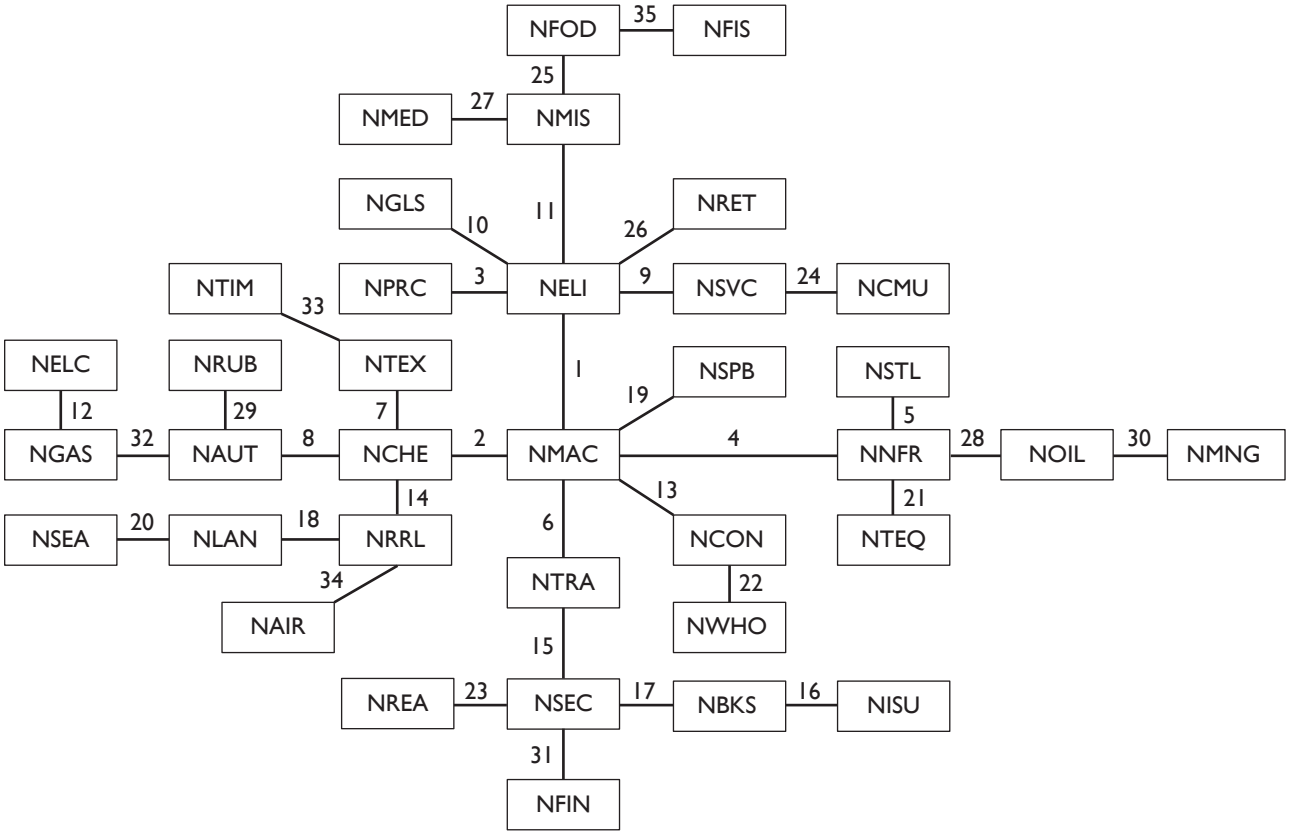


Figure 2.

(d) What is the variance we expect to find for the bid-ask spread?

(5 marks)

(e) Will this model produce volatility clustering in the price time series? Explain.

(5 marks)

4. Agents with bounded rationality.

The economist Herbert Simon proposed that the rationality of individuals is limited to the information they have, the cognitive resources they can muster, and the time available to make decisions. Because bounded rationality is still a mode of decision making, most economists focused on building models of the decision process. Such an approach makes the development of agent-based models (ABM) extremely challenging.

In contrast, econophysicists consider simple agents whose rationality is bounded by the amount of information it can integrate from its neighbours. A toy model of bounded rationality is the Ising model

$$H = -\sum_{(i,j)} J_{ij} s_i s_j + \sum_i h_i s_i$$

on a square lattice shown in Figure 3. In this model, an agent can buy ($s_i = +1$) or sell ($s_i = -1$), and it arrives at this decision by considering what its neighbours are doing ($-\sum_{(i,j)} J_{ij} s_i s_j$) and also its own decision ($h_i s_i$).

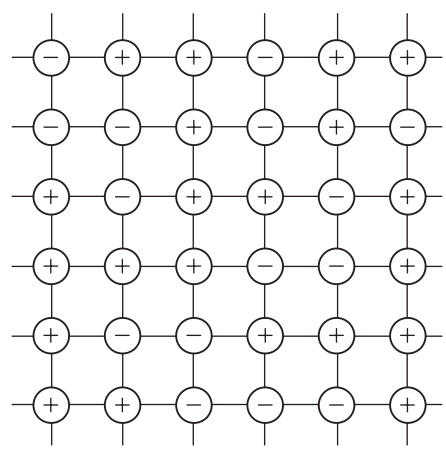


Figure 3.

a. Describe how you would simulate such a model at inverse temperature $\beta = T^{-1}$, using the Metropolis algorithm.

(10 marks)

b. Elaborate on how this toy model provides a stylized explanation of booms and busts in the financial markets.

(10 marks)

c. In real markets, traders do not just buy and sell. They sometimes decide to hold. We can incorporate this

behaviour into the model by going from $s_i = \pm 1$ to $s_i = -1, 0, +1$, keeping the Hamiltonian H the same.

Describe the most important differences between the dynamics of this modified Ising model and that of the original Ising model.

(5 marks)

Appendix 4. Approved Reading Lists for Open Textbook Chapters

Chapter 6. Cross Correlation Filtering and Time Series Clustering

1. Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society, 7*(1), 48–50.
2. Harris, L. (1991). Stock price clustering and discreteness. *Review of Financial Studies, 4*(3), 389–415.
3. Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B, 11*(1), 193–197.
4. Hill, R. J. (1999). Comparing price levels across countries using minimum spanning trees. *Review of Economics and Statistics, 81*(1), 135–142.
5. Gavrilov, M., Anguelov, D., Indyk, P., & Motwani, R. (2000). Mining the stock market: which measure is best?. *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* pp. 487–496, ACM.
6. Kullmann, L., Kertesz, J., & Mantegna, R. N. (2000). Identification of clusters of companies in stock indices via Potts super-paramagnetic transitions. *Physica A, 287*, 412–419.
7. Groenen, P. J. F., & Franses, P. H. (2000). Visualizing time-varying correlations across stock markets. *Journal of Empirical Finance, 7*(2), 155–172.
8. Bonanno, G., Lillo, F., & Mantegna, R. N. (2001). Levels of complexity in financial markets. *Physica A, 299*(1), 16–27.
9. Bonanno, G., Lillo, F., & Mantegna, R. N. (2001). High-frequency cross-correlation in a set of stocks. *Quantitative Finance, 1*(1), 96–104.
10. Longin, F. & Solnik, B. (2001). Extreme correlation of international equity markets. *The Journal of Finance, 56*(2), 649–676.
11. Onnela, J.-P., Chakraborti, A., Kaski, K., & Kertész, J. (2002). Dynamic asset trees and portfolio analysis. *The European Physical Journal B, 30*(3), 285–288.
12. Kullmann, L., Kertesz, J., & Kaski, K. (2002). Time-dependent cross-correlations between different stock returns: A directed network of influence. *Physical Review E, 66*(2), 026125.
13. Onnela, J.-P., Chakraborti, A., Kaski, K., Kertész, J., & Kanto, A. (2003). Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E, 68*(5), 056110.
14. Micciché, S., Bonanno, G., Lillo, F., & Mantegna, R. N. (2003). Degree stability of a minimum spanning tree of price return and volatility. *Physica A, 324*(1–2), 66–73.
15. Onnela, J.-P., Chakraborti, A., Kaski, K., & Kertesz, J. (2003). Dynamic asset trees and Black Monday. *Physica A, 324*, 247–252.
16. Bonanno, G., Calderelli, G., Lillo, F., Micciché, S., Vandewalle, N., & Mantegna, R. N. (2004). Networks of equities in financial markets. *The European Physical Journal B, 38*(2), 363–371.
17. Pattarin, F., Paterlini, S., & Minerva, T. (2004). Clustering financial time series: An application to mutual funds style analysis. *Computational Statistics & Data Analysis, 47*(2), 353–372.
18. Tumminello, M., Di Matteo, T., Aste, T., & Mantegna, R. N. (2005). A tool for filtering information in complex systems. *Proceeding of the National Academy of Sciences of the United States of America, 102*(30), 10421–10426.
19. Jung, W.-S., Kwon, O., Kaizoji, T., Chae, S., & Moon, H.-T. (2005). Grouping in the stock markets of Japan and Korea. arXiv:physics/0511224 (27 November).
20. Aste, T., Di Matteo, T., Tumminello, M., & Mantegna, R. N. (2005). Correlation filtering in financial time series. In *Proceedings of the SPIE Third International Symposium on Fluctuations and Noise, 5848*, 100–109.
21. Coronnello, C., Tumminello, M., Lillo, F., Micciche, S., & Mantegna, R. N. (2005). Sector identification in a set of stock return time series traded at the London Stock Exchange. *Acta Physica Polonica B, 36*(9), 2653–2679.
22. Keogh, E., & Lin, J. (2005). Clustering of time-series subsequences is meaningless: Implications for previous and future research. *Knowledge and Information Systems, 8*(2), 154–177.
23. Chen, J. R. (2005). Making subsequence time series clustering meaningful. *Proceedings of the Fifth IEEE International Conference on Data Mining*, IEEE.
24. Liao, T. W. (2005). Clustering of time series data—a survey. *Pattern Recognition, 38*(11), 1857–1874.
25. Kim, D.-H., & Jeong, H. (2005). Systematic analysis of group identification in stock markets. *Physical Review E, 72*(4), 046133.
26. Toth, B., & Kertesz, J. (2006). Increasing market efficiency: Evolution of cross-correlations of stock returns. *Physica A, 360*(2), 505–515.
27. Aste, T., & Di Matteo, T. (2006). Dynamical networks from correlations. *Physica A, 370*(1), 156–161.
28. Tóth, B., & Kertész, J. (2006). Increasing market efficiency: Evolution of cross-correlations of stock returns. *Physica A, 360*(2), 505–515.
29. Heimo, T., Saramäki, J., Onnela, J.-P., & Kaski, K. (2007). Spectral and network methods in the analysis of correlation matrices of stock returns. *Physica A, 383*(1), 147–151.
30. Toth, B., & Kertesz, J. (2007). On the origin of the Epps effect. *Physica A, 383*(1), 54–58.
31. Tumminello, M., Coronnello, C., Lillo, F., Micciché, S., & Mantegna, R. N. (2007). Spanning trees and bootstrap reliability estimation in correlation-based networks.

International Journal of Bifurcation and Chaos, 17(7), 2319–2329.
32. Brida, J. G., & Risso, W. A. (2008). Multidimensional minimal spanning tree: The Dow Jones case. *Physica A, 387*(21), 5205–5210.
33. Eom, C., Oh, G., Jung, W.-S., Jeong, H., & Kim, S. (2009). Topological properties of stock network based on minimal spanning tree and random matrix theory in financial time series. *Physica A, 388*(6), 900–906.
34. Toth, B., & Kertesz, J. (2009). The Epps effect revisited. *Quantitative Finance, 9*(7), 793–802.
35. Nanda, S. R., Mahanty, B., & Tiwari, M. K. (2010). Clustering Indian stock market data for portfolio management. *Expert Systems with Applications, 37*(12), 8793–8798.
36. Tse, C. K., Liu, J., & Lau, F. (2010). A network perspective of the stock market. *Journal of Empirical Finance, 17*(4), 659–667.
37. Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666.
38. Silva, B., & Marques, N. C. (2011). Clustering stock market values with a self-organized feature map. *Proceedings of the 15th Portuguese Conference on Artificial Intelligence, Multi-Agent Systems: Theory and Applications*.

Chapter 7. Agent-based Models of Financial Markets

1. Shelling, T. C. (1969). Models of segregation. *The American Economic Review, 59*(2), 488–493.
2. Bak, P., Tang, C., & Wiesenfeld, K. (1988). Self-organized criticality. *Physical Review A, 38*(1), 365–374.
3. Holland, J., & Miller, J. (1991). Artificial adaptive agents in economic theory. *The American Economic Review, 81*(2), 365–370.
4. Levy, M., Levy, H., & Solomon, S. (1994). A microscopic model of the stock market: cycles, booms, and crashes. *Economics Letters, 45*(1), 103–111.
5. Palmer, R. G., Arthur, W. B., Holland, J. H., LeBaron, B., & Tayler, P. (1994). Artificial economic life: A simple model of a stock market. *Physica D, 75*(1–3), 264–274.
6. Arthur, W. B. (1995). Complexity in economic and financial markets: Behind the physical institutions and technologies of the marketplace lie the beliefs and expectations of real human beings. *Complexity, 1*(1), 20–25.
7. Bak, P., & Paczuski, M. (1995). Complexity, contingency, and criticality. *Proceeding of the National Academy of Sciences of the United States of America, 92*(15), 6689–6696.
8. Levy, M., Solomon, S., & Ram, G. (1996). Dynamical explanation for the emergence of power law in a stock market model. *International Journal of Modern Physics C, 7*(1), 65–72.
9. Bak, P., Paczuski, M., & Shubik, M. (1997). Price variations in a stock market with many agents. *Physica A, 246*(3–4), 430–453.
10. Vandewalle, N., Boveroux, Ph., Minguet, A., & Ausloos, M. (1998). The crash of October 1987 seen as a phase transition: Amplitude and universality. *Physica A, 255* (1–2), 201–210.
11. Parunak, H. V. D., Savit, R., & Riolo, R. L. (1998). Agent-based modeling vs. equation-based modeling: A case study and users' guide. *Lecture Notes in Computer Science, 1534*, 10–25.
12. Farmer, J. D. (1999). Physicists attempt to scale the ivory towers of finance. *Computing in Science & Engineering*, *1*(6), 26–39.
13. Farmer, J. D., & Lo, A. W. (1999). Frontiers of finance: Evolution and efficient markets. *Proceedings of the National Academy of Sciences of the United States of America, 96*(18), 9991–9992.
14. LeBaron, B., Arthur, W. B., & Palmer, R. (1999). Time series properties of an artificial stock market. *Journal of Economic Dynamics & Control, 23*(9–10), 1487–1516.
15. Savit, R., Manuca, R., & Riolo, R. (1999). Adaptive competition, market efficiency, and phase transitions. *Physical Review Letters, 82*(10), 2203–2206.
16. Batten, D. F. (2000). *Discovering Artificial Economics: How Agents Learn and Economies Evolve*. Boulder, CO: Westview Press.
17. Gilbert, N., & Terna, P. (2000). How to build and use agent-based models in social science. *Mind & Society, 1*(1), 57–72.
18. LeBaron, B. (2000). Agent-based computational finance: Suggested readings and early research. *Journal of Economic Dynamics and Control, 24*(5–7), 679–702.
19. Kaizoji, T. (2000). Speculative bubbles and crashes in stock markets: An interacting-agent model of speculative activity. *Physica A, 287*(3–4), 493–506.
20. Lux, T., & Marchesi, M. (2000). Volatility clustering in financial markets: A microsimulation of interacting agents. *International Journal of Theoretical and Applied Finance, 3*(4), 675.
21. Sznajd-Weron, K., & Sznajd, J. (2000). Opinion evolution in closed community. *International Journal of Modern Physics C, 11*(6), 1157–1165.
22. Challet, D., Marsili, M., & Zhang, Y.-C. (2000). Modeling market mechanism with minority game. *Physica A, 276* (1–2), 284–315.
23. Bornholdt, S. (2001). Expectation bubbles in a spin model of markets: Intermittency from frustration across scales. *International Journal of Modern Physics C, 12*(5), 667–674.
24. Chen, S.-H. & Yeh, C.-H. (2001). Evolving traders and the business school with genetic programming: A new architecture of the agent-based artificial stock market. *Journal of Economic Dynamics and Control, 25*(3–4), 363–393.
25. Jefferies, P., (2001). Hart, M. L., Hui, P. M., & Johnson, N.F. (2001). From market games to real-world markets. *The European Physical Journal B, 20*(4), 493–501.
26. LeBaron, B. (2001). A builder's guide to agent-based financial markets. *Quantitative Finance, 1*(2), 254–261.
27. LeBaron, B. (2001). Empirical regularities from interacting long- and short-memory investors in an agent-based stock market. *IEEE Transactions on Evolutionary Computation, 5*(5), 442–455.

28. LeBaron, B. (2001). Evolution and time horizons in an agent-based stock market. *Macroeconomic Dynamics*, *5*(2), 225–254.
29. Raberto, M., Cincotti, S., Focardi, S. M., & Marchesi, M. (2001). Agent-based simulation of a financial market. *Physica A, 299*(1–2), 319–327.
30. Berry, B. J. L., Kiel, L. D., & Elliott, E. (2002). Adaptive agents, intelligence, and emergent human organization: Capturing complexity through agent-based modeling, *Proceeding of the National Academy of Sciences of the United States of America, 99*(3), 7187–7188.
31. Bouchaud, J.-P. (2002). An introduction to statistical finance. *Physica A, 313*(1–2), 238–251.
32. Iori G. (2002). A microsimulation of traders activity in the stock market: The role of heterogeneity, agents' interactions and trade frictions. *Journal of Economic Behavior & Organization, 49*(2), 269–285.
33. Macy, M. W., & Flache, A. (2002). Learning dynamics in social dilemmas. *Proceedings of the National Academy of Sciences of the United States of America, 99*(3), 7229–7236.
34. Macy, M. W., & Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology, 28*, 143–166.
35. Sznadjd-Weron, K., & Weron, R. (2002). A simple model of price formation. *International Journal of Modern Physics C, 13*(1), 115.
36. Tesfatsion, L. (2003). Agent-based computational economics: Modeling economies as complex adaptive systems. *Information Sciences, 149*(4), 263–269.
37. Borshchev, A., & Filippov, A. (2004). From system dynamics and discrete event to practical agent based modeling: Reasons, techniques, tools. *Proceedings of the 22nd International Conference of the System Dynamics Society,* no. 22.
38. Gilbert, N. (2004). Agent-based social simulation: Dealing with complexity. *The Complex Systems Network of Excellence, 9*(25), 1–14.
39. Gatti, D. D., Guilmi, C. D., Gaffeo, E., Guilioni, G., Gallegati, M., & Palestrini, A. (2005). A new approach to business fluctuations: Heterogeneous interacting agents, scaling laws and financial fragility. *Journal of Economic Behavior and Organization, 56*(4), 489–512.
40. Alfarano, S., Lux, T., & Wagner, F. (2006). Estimation of a simple agent-based model of financial markets: An application to Australian stock and foreign exchange data. *Physica A, 370*(1), 38–42.
41. LeBaron, B. (2006). Chapter 24: Agent-based computational finance. *Handbook of Computational Economics, 2*, 1187–1233.
42. Janssen, M. A. & Ostrom, E. (2006). Empirically based agent-based models. *Ecology and Society, 11*(2), 37.
43. Railsback, S. F., Lytinen, S. L., & Jackson, S. K. (2006). Agent-based simulation platforms: Review and development recommendations. *Simulation, 82*(9), 609–623.
44. Tesfatsion, L. (2006). Agent-based computational economics: Growing economies from the bottom up. *Artificial Life, 8*(1), 55–82,
45. Tesfatsion, L. (2006). Chapter 16: Agent-based computational economics: A constructive approach to economic theory. *Handbook of Computational Economics, 2*, 831–880.
46. Fagiolo, G., Moneta, A., & Windrum, P. (2007). A critical guide to empirical validation of agent-based models in economics: Methodologies, procedures, and open problems. *Computational Economics, 30*(3), 195–226.
47. Matthews, R. B., Gilbert, N. G., Roach, A., Polhill, G., & Gotts, N. M. (2007). Agent-based land-use models: A review of applications. *Landscape Ecology, 22*(10), 1447–1459.
48. Samanidou, E., Zschischang, E., Stauffer, D., & Lux, T. (2007). Agent-based models of financial markets. *Reports on Progress in Physics, 70*(3), 409–450.
49. Windrum, P., Fagiolo, G., & Moneta, A. (2007). Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation, 10*(2), Article 8.
50. Zhou, W.-X. & Sornette, D. (2007). Self-organizing Ising model of financial markets. *The European Physical Journal B, 55*(2), 175–181.
51. Weidlich, A., & Veit, D. (2008). A critical survey of agent-based wholesale electricity market models. *Energy Economics, 30*(4), 1728–1759.
52. Yang, J.-S., Kwak, W., Kaizoji, T., & Kim, I.-M. (2008). Increasing market efficiency in the stock markets. *The European Physical Journal B, 61*(2), 241–246.
53. Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics, 7*(2), 174–196.
54. Macal, C. M. & North, M. J. (2010). Tutorial on agent-based modelling and simulation. *Journal of Simulation, 4*, 151–162.
55. Chakraborti, A., Toke, I. M., Patriarca, M., & Abergei, F. (2011). Econophysics review: II. Agent-based models. *Quantitative Finance, 11*(7), 1013–1041.
56. Chen, S.-H., Chang, C.-L., & Du, Y.-R. (2012). Agent-based economic models and econometrics. *The Knowledge Engineering Review, 27*(2), 187–219.
57. Chen, S.-H. (2012). Varieties of agents in agent-based computational economics: A historical and an interdisciplinary perspective. *Journal of Economic Dynamics and Control, 36*(1), 1–25.
58. Zhao, K., Yen, J., Ngamassi, L.-M., Maitland, C., & Tapia, A. H. (2012). Simulating inter-organizational collaboration network: A multi-relational and event-based approach. *Simulation, 88*(5), 617–633.

Appendix 5. Snippets from Chapters 6 and 7 of the Open Textbook

Chapter 6

After our discussion in part A about correlation filtering and MST, we can now start using them to make stocks clusters. There

are different methods on doing time series clustering for stocks. In this chapter we will only discuss some of them; hierarchical clustering, k-means clustering, spectral clustering, and many others. Next, we will try to discuss more in depth each of them.

Hierarchical Clustering

Hierarchical structures exist in different complex systems, and the financial market is no exception. A hierarchical structure is defined as the organization of the elements into clusters each consisting of subclusters and so on up to a certain level (Simon, 1962). We can observe this hierarchy by mapping the hierarchical tree from the correlation matrix that we studied in the previous part. The hierarchical tree is based on the correlation between stocks, therefore a correlation matrix, an MST, or a PMFG would suffice to make a hierarchical tree. Different correlation based network can be associated with the same hierarchical tree.

Now let us consider the triangular version of the 4 × 4 correlation matrix that we have in the previous section.

$$C = \begin{pmatrix} 1.00 & 0.13 & 0.90 & 0.81 \\ & 1.00 & 0.57 & 0.34 \\ & & 1.00 & 0.71 \\ & & & 1.00 \end{pmatrix}$$

Suppose we name each stock based on its index such as stock **1**, **2**, **3**, and **4**. From this correlation matrix, we can proceed to make a hierarchical tree with two methods, average linkage cluster analysis (ALCA) and single linkage cluster analysis (SLCA). The example below is about making SLCA. This algorithm is based on Mantegna's research on 2008.

1. Set a matrix $\boldsymbol{B} = \boldsymbol{C} = \begin{pmatrix} 1.00 & 0.13 & 0.90 & 0.81 \\ & 1.00 & 0.57 & 0.34 \\ & & 1.00 & 0.71 \\ & & & 1.00 \end{pmatrix}$
2. Choose the maximum correlation b_{hk} in the correlation matrix $\boldsymbol{B}$. In this case we choose h and k to be a simple cluster of one element each while they can actually be larger clusters. In this case, the maximum correlation b_{hk} = 0.90 where $h = 1$ and $k = 3$. For all $i \in h$ and $j \in k$, sets the elements $p_{ij}^{<}$ of the matrix $\boldsymbol{C}_{SLCA}^{<}$ as $p_{ij}^{<} = p_{ji}^{<} = b_{hk}$.
3. Merge cluster h and k into a single cluster q. The merging operation identifies a node in a rooted tree connecting h and k at the correlation b_{hk}. So the merged index will be the stock corresponding to the indices 3 and 4. Note that after merger the matrix size will be $(M-1) \times (M-1)$ instead of $M \times M$.
4. Redefine the matrix $\boldsymbol{B}$:

$$\begin{cases} b_{qj} = Max[b_{hj}, b_{kj}],\ j \notin h \text{ and } j \notin k \\ b_{ij} = b_{ji},\ otherwise \end{cases}$$

After the first implementation of algorithm $\boldsymbol{B}$ changed as shown as below:

$$\boldsymbol{B} = \begin{pmatrix} 1.00 & 0.57 & 0.81 \\ & 1.00 & 0.34 \\ & & 1.00 \end{pmatrix}$$

The above matrix $\boldsymbol{B}$ corresponds to:

	1/3	2	4
1/3	1.00	0.57	0.81
2	0.57	1.00	0.34
4	0.81	0.34	1.00

1/3 represents the cluster made of stock **1** and **3**. Following the algorithm, we get each element of the matrix to be the highest correlation of the element with the new cluster. Note that stock **1** and stock **3** are connected in the hierarchical tree at correlation 0.90 which was the largest correlation between stock **1** and stock **3**.

If we continue the procedure, we will get the corresponding matrix:

$$\boldsymbol{B} = \begin{pmatrix} 1.00 & 0.57 \\ 0.57 & 1.00 \end{pmatrix}$$

Where now the cluster **1/3** is merged with stock **4** at correlation 0.81. Applying the procedure one more time, we get the cluster **1/3/4** to be connected with stock **2** at correlation 0.57. At this point the matrix size is **1**, and thus we can make the corresponding hierarchical tree shown below:

Whereas the corresponding matrix associated with SLCA is

$$\boldsymbol{C}_{SLCA}^{<} = \begin{pmatrix} 1.00 & 0.57 & 0.90 & 0.81 \\ & 1.00 & 0.57 & 0.57 \\ & & 1.00 & 0.81 \\ & & & 1.00 \end{pmatrix}$$

ALCA will later be encountered in the exercise problem.

This hierarchical structure might be useful to extract meaningful economic taxonomy. In our dummy correlation matrix, this tree might not be very useful. However, if we are analyzing considerably more stocks we might arrive to some interesting conclusive. For example, Mantegna detected that ores companies are affected differently than aluminium and copper companies which traditionally classified as raw materials companies according to

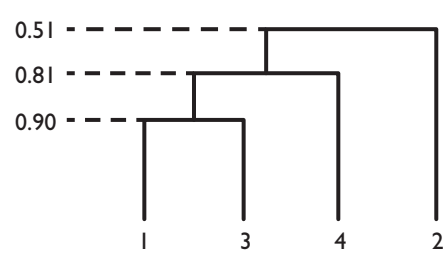


Figure 1. Hierarchical tree representing the correlation matrix, not drawn to scale.

Forbes. In his research he found that ore companies form different cluster than the two other kinds of companies based on the hierarchical tree.

Hierarchical clustering can thus be visualized better with using colour map. Consider colour map of S&P 100 in no particular order.

By using the algorithm presented by the Matlab demo, we can apply SLCA to obtain the corresponding hierarchical tree. Sorting the stocks according to their hierarchy will result on the colour map below.

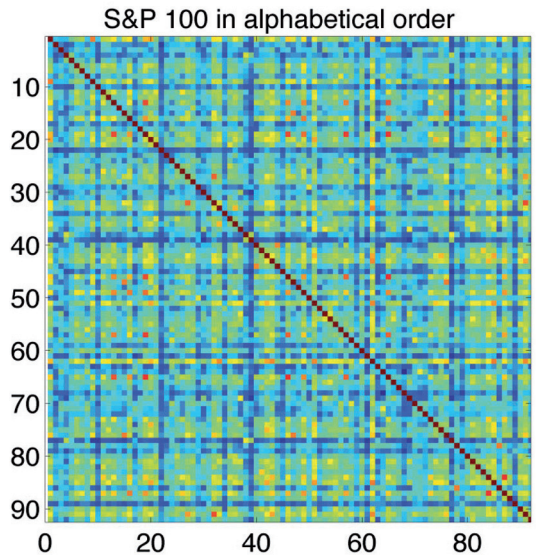


Figure 2. The colours represent correlations between stocks where lighter colours have more correlation than darker colours.

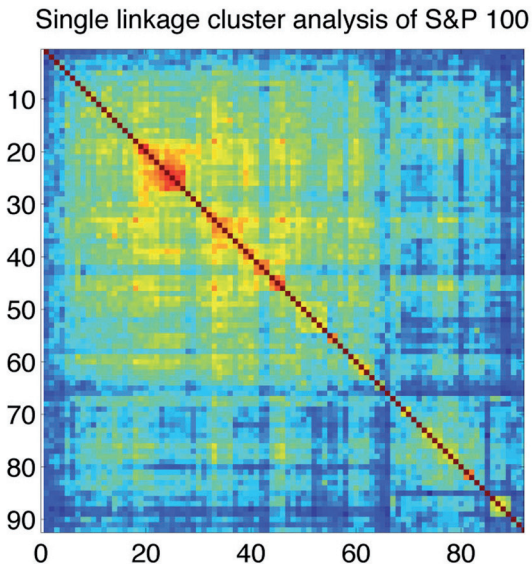


Figure 3. Ordered colour map with SLCA.

After the stocks are ordered according to the SLCA, some yellowish squares started to appear. Those squares signify clusters since it means that highly correlated stocks are placed nearly located to each other in the hierarchical tree. The arrow shows where the cluster is. The more stocks we include in the calculation, the more obvious the clusters can be seen.

Chapter 7

Genetic Algorithm

The technique employed for the agents to adapt and evolve is called the Genetic Algorithm (GA). This was applied to the SFI Market model. GA is a stochastic search algorithm based on the mechanics of natural selection (Darwin, 1897) and population genetics (Mettler *et al.*, 1988). It is modelled after natural genetic operators that enable biological populations to effectively and robustly adapt to their environment and to changes in their environment.

Since the algorithm is based on biological evolution, it makes sense to first describe the process from the biological point of view, before we take on the financial perspective. Evolution occurs in organic entities capable of encoding the structure of living organisms. These entities are known as chromosomes. Chromosomes transform due to mutations and crossovers. The objective of mutation is to introduce diversity into the population, whereas a crossover enhances the schemata[13]. The diagram below illustrates a reproduction process of the cell (agent) after the genetic operator modifies the offspring's string, or introduces new strings into the system.

Now that you have some sense of how a GA works, we are going to introduce the rules of GA as they were conceived by John Holland[13]. There are four basic elements:

1. Chromosome: different types of decisions, resources, etc.
2. Creator: generalize by an operator that starts by generating an initial pool of chromosomes.
3. Evaluator: rates each chromosome by giving the highest rating to those that are able to solve their task most of the time.
4. Generator: uses mutate and crossover operations to randomly reconfigure the solutions specified by the chromosomes.

The elements are an example of how an organism evolves into an effective system, and also a process of natural selection of the fittest chromosome. This selection process consists of a few modules, namely, population, interaction, evaluation, reproduction

Figure 4. Cross-over operator.

and modification[13]. The sequence of a genetic algorithm is as follows:

1. We first initialize a number of traders
2. Each trader has a string that acts as their strategies during the interactions.
3. During the 'Market Period', the Traders interact with the system:
   (a) The algorithm evaluates the fitness of the population. e.g., wealth of the agent.
   (b) System removes agents based on the fitness of the agents. Traders that did not make as much profit as the rest would be remove from the model.
   (c) The agents reproduce with the probability that is proportional to their fitness criterion. With this, the subsequent generation will contain strategies that will make them 'fitter' through the trading process.
   (d) Perform modification to the 'newly created traders' using genetic operators.
4. Outputs the end result for the 'Market cycle' and loop the process

In general, we are trying to evolve the population, or rather teach the agents such that they are best adapted to the environment. However, with only this process, the agents do not evolve over time and the agents left behind after going through GA will be one of the agents we created at the start.

Classifier System

The classifier system (CS) is an adaptive rule-based system which learns syntactically simple string rules, called classifiers, as introduced by Holland and Reitman (1978). It has the ability to learn to classify messages from the environment into general sets. This is similar to the feedback process in control system in many respects. The mechanism of the CS works with rules as condition/action rules—each rule is a simple message as the message act as conditions to the other rules. When more than one rule is triggered by a condition, the one with the highest strength will win.

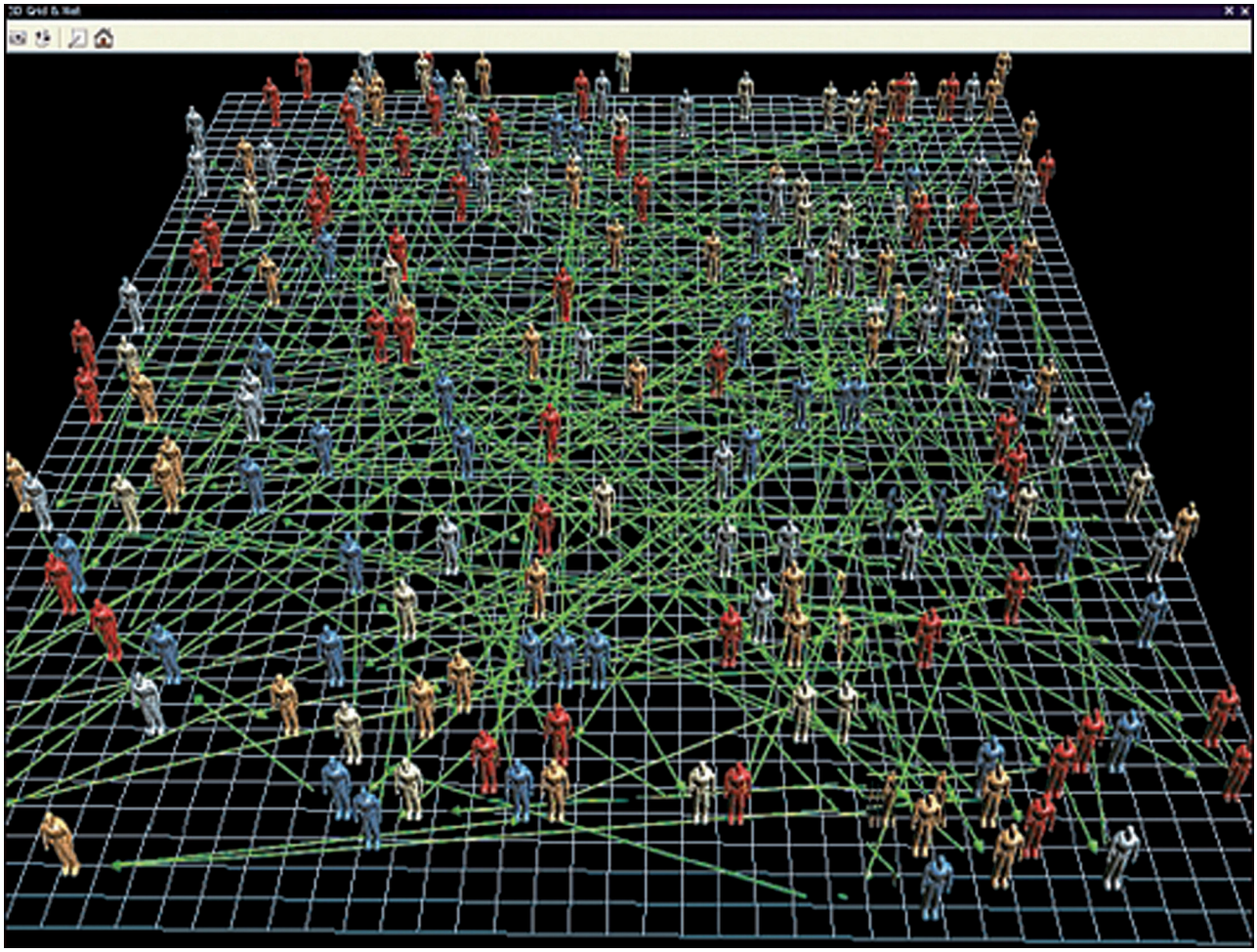


Figure 5. An illustration of agents interacting in on a lattice.

Source: http://www.scidacreview.org/0802/html/abms.html

Therefore the mechanism requires a certain form of strength to be calculated. This is done with the Bucket-Brigade Algorithm. When a rule wins, it pays the condition money in terms of strength and when it triggers other rules, the winner pays him. This covers as a useful measure to the rules contribution to the system.

By coupling the CS and the GA, the GA will act to revise the rules and reproduces by the strength of the rule. The complex system simulated by these 2 mechanism causes clusters and hierarchies of rules to emerge.

Appendix 6. End-of-Chapter Exercises

Chapter 6

Ex. 6.1. (Ian Beng Hau Tan)

In this exercise, you will be guided through the process of generating the correlation matrix in MATLAB. You will then be guided through methods to generate the MST using either MATLAB or Gephi, an open source visualization software for networks.

1. Use the *Historical Stock Data* utility written by Josiah Renfree on the MATLAB File Exchange website, which automatically downloads stock prices to the MATLAB array structure price (i). For this part of the exercise, read the documentation on the *Historical Stock Data* utility and download the daily returns for the Dow Jones 30 over a period of two years.

   We will use the Pearson correlation coefficient, given by

$$\rho_{ij} = \frac{\langle Y_i Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle}{\sqrt{\left(\langle Y_i^2 \rangle - \langle Y_i \rangle^2\right)\left(\langle Y_j^2 \rangle - \langle Y_j \rangle^2\right)}}$$

   where Y is the log return of a stock, $Y_i = \ln P_i(t) - \ln P_i(t - \Delta t)$. For this exercise, use P, the closing price of a stock i at the end of an interval t. We denote Δt as the *time horizon*.
2. For each stock i in the Dow Jones 30, generate the time series P_i in MATLAB using a time horizon of 1 trading day.
3. Generate the log return series Y_i.
4. Generate the correlation matrix, using the command z = corr(y).
5. Generate the MST:

   With MATLAB. First, create a linkage tree by using the linkage function (requires MATLAB Statistics Toolbox), or generate your own matrix of unique links and convert it into a tree structure. Next, use this linkage tree as an input into graphminspantree (requires MATLAB Bioinformatics Toolbox). Those who do not have access to either toolbox may try the free MatlabBGL toolbox on the MATLAB File Exchange.

   Without MATLAB. Use the open source software Gephi, available for Mac, Windows and Linux platforms. Generate a matrix of unique links, as shown in Section 6.1.3. The matrix of unique links can be saved in UCINET DL or CSV formats with a slight modification (see the online documentation in the Gephi web site). Import this DL or CSV file into Gephi and the software will do the rest.

   Bonus: generating the PMFG
   The steps are similar to that for the MST. The easiest way to generate the PMFG in MATLAB is to use the (free) PMFG routine developed by Tomaso Aste in MATLAB File Exchange. This routine requires the (also free) MatlabBGL toolbox, also available on MATLAB File Exchange. Alternatively, generate the matrix of unique links and input this into Gephi.

Chapter 7

Ex. 7.4. (Henry Pui Loong Lee & Boon Kin Teh)

In this exercise, we will demonstrate how to write a basic programme that simulates the agent-based Ising model using the Metropolis algorithm, which is given below.

Agent based Ising model (Metropolis algorithm)

(a) A trader is randomly drawn from the square lattice say trader i, and try to flip the decision made.
$S_i \to -S_i$

(b) Calculate the energy difference flipping, $H(t)$ and after flipping, $H(t_+)$ as

$$\Delta H = H(t_+) - H(t) = 2a \sum_{j=NN} S_i(t) S_j(t) + 2be\, S_i(t).$$

(c) Then decide the flip is accepted or not

$$\begin{cases} \Delta H < 0 & \textit{always accept} \\ \Delta H > 0 & \text{Accept with probabily } \exp(-k\Delta H) \end{cases}$$

(d) Iterate steps (i) to (iii).

1. First, we need to have a $N \times N$ lattice, in which the spins point up or down randomly. Write a programme that will generate a $N \times N$ square lattice with random decision, which are either buy or sell.
2. The Metropolis algorithm is a stochastic process, as it picks a trader randomly from the lattice in step (b). Write a programme that does this.
3. In the real world, there are a large number of traders in the financial market. However, to simulate such a large system we require a lot of computational power. We therefore impose periodic boundary conditions on the system. In periodic boundary condition, the boundaries of each end are connected to the opposite end. You can imagine the square lattice lying on the surface of a torus.

   After picking a trader, we have to decide whether the trial flip is accepted, according to the detail balance rule stated in step (c).

Write a programme that calculates the energy difference as in steps c) in the algorithm, ensuring periodic boundary conditions.

4. Now you have all the pieces to simulate the agent based-Ising model, write an iterative programme that combines all the code.

Appendix 7. MATLAB Demos

Chapter 6

```
% File:        hiertree.m
% Authors:     Tan Beng Hau
%
% Description: A dendrogram is constructed based on a given
%              clustering method. The dendrogram sorts accord-
%              ing to the ultrametric distance. This sorting is
%              used to create a colour map of the clustering.
%              For more distinct clusters, use more stocks in the
%              portfolio.
% Requires:    Matlab Statistics Toolbox
%
% Acknowledgements:
% hist_stock_data.m downloader from Matlab File Exchange
% Kenneth Lee for advice

% Download the stock prices for the stocks from a text file
labels = importdata('TICKERS100.txt'); % S&P 100
price = hist_stock_data('01012010','31122011',
        'TICKERS100.txt');

for i = 1:numel(labels)
    price2(:,i) = [price(i).AdjClose(end:-1:1)];
    % other ways may result in structtodouble error.
end

y = diff(log(price2)); %log returns

% Initialise the correlation matrix
z = corr(y);

D = pdist(z,'correlation');
% Sornette's metric distance
% Another approach is to define your own distance.
ds = sqrt(2*D);

% Linkage options given by the Statistics Toolbox are:
% single, average, ward, complete, weighted, median, centroid,
  ward
% If the linkage option is not defined, carries out SLCA by
  default.
% I believe centroid clustering is k-means, where k can be
  defined in
% cluster(tree,k).

% Metric information:
% 'correlation' calculates 1 – r_ij, where r_ij is the Pearson
% correlation coefficient.
tree = linkage(z,'single','correlation');
% Scale tree weights to Sornette's metric distance
for i=1:size(tree,1)
    tree(i,3) = sqrt(2*tree(i,3));
end

% I didn't implement this but it should be easy to.
% t_cluster = cluster(tree, 'maxclust',8);

% Perform the leaf ordering based on the Sornette metric
leafOrder = optimalleaforder(tree,ds,'Transformation','inverse');
[H,T1,outperm] =
dendrogram(tree,100,'Reorder',leafOrder,'labels',labels,'Color
Threshold','default','Orientation','left');
xlabel('Ultrametric distance'); ylabel('Stocks');
title('Dendrogram of the S&P 100');

% Construct the 'random' colour map

figure(2)
colormap(hot); % To reverse gray colormap, use colormap
(flipud(gray)).
cmin = 0; cmax = 1; % Define the color mapping space
imagesc(z)
caxis([cmin cmax]);
title('S&P 100 in no particular order');

% Reconstruct the matrix from the SLCA ordering
CM_reconst = z(outperm,outperm);

% Construct the SLCA-ordered colour map

figure(3)
colormap(hot); % To reverse gray colormap, use colormap
(flipud(gray)).
cmin = 0; cmax = 1; % Define the color mapping space
imagesc(CM_reconst)
caxis([cmin cmax]);
title('Single linkage cluster analysis of S&P 100');
```

Chapter 7

```
% function [M, E] = Metropolis_Stock(N,J,B,steps)
% to watch the movie, use this code
% Metropolis(250,1,0.001,1000)
% Metropolis_Stock(250,–1,0.001,1000)  %  market converges to
%                                         equilibrium
% Metropolis_Stock(128,2/3,0,1000)     %  noisy normal market
%=========================================
% [M, E] = Metropolis(N,J,B,steps)
%          Updates a single site at a time.
%          Choose sites at random to ensure no bias. The
%          probabilistic part of the algorithm is done using a
%          random number generator
%
% sigma      -  Ising Matrix 'space'
%               Every i,j represents ONE Agent
% N          -  number of rows and cols
% J          -  Coupling Constant
```

```
% B          - Magnetic field (1 = all black -1 = all white)
%
% steps      - number of iterations
% dE         - Change in Energy
% beta(1/kT) - inverse temperature times interaction
               strength
%              (beta >= 0)
%              (critical value = 0.8813736)
%              beta has no meaning if <=0
%              Beta determines how "much" agents interact
%              with each other
% E          - Higher E → "closer" to equlibrum
%==========================================

figure(1)
set(gcf,'Position',[200 50 500 500],'Color','White')

N=250;  %  128;%250; % no of traders
J=1;    %  coupling constant
B=0.01; %  0.5 dissappear %0;%.1; %–0.5:0.5
steps=1000;
M=zeros(steps);
E=zeros(steps);
randTol = 0.1;
%  The tolerance, dampens the spin flip process rand('twister');
%  seed rand()
sigma = (–1).^(round(rand(N))); % initial market state
%  First we generate a random initial configuration→ 'Ising Space'

%  beta=2*log(1+sqrt(2)); % 2*log(1+sqrt(2)) bring to
   equilibrium
%  state or o.w market cool down –0.5;
%  Beta will be determine by a GA
RandNum = rand(steps,1);
%  get rand first before the for loop. Func called once

market_mood=2;   %randi(4);
%  first mood
old_market_mood=market_mood;
%  for first mood
market_cycle=50;
%  how often the market mood changes in steps(time)
behaviour_of_market=randi(4,(steps/market_cycle)+1,1);
%  preallocate number of 'moods' needed
current_cycle=1;
%  count no of cycles

if ~isinteger(int8(steps/market_cycle))
   disp(['You have not entered a market cycle divisible
by',num2str(steps)])
   market_cycle=100;
end

for i=1:steps
      neighbours = circshift(sigma, [0 1]) + ...
                   % up
                   circshift(sigma, [0 –1]) + ...
                   % down
                   circshift(sigma, [1 0]) + ...
                   % right
                   circshift(sigma, [–1 0]);
                   % left

   dE = (J*(sigma .* neighbours) + B*sigma);
   % Calculates the change in energy of flipping a spin
   % dE = –2 * neighbours + 4 + 2 * B * sigma;
   % if dE <= 0 we make the proposed Change,
   % if dE > 0 we make the change with probability e^(dE/kT)
   % here kT =1

   % beta=stock_market1(neighbours,sigma,N);

   [beta
   market_mood]=stock_market2(neighbours,N,
   market_mood,old_market_mood,RandNum(i,1)); %super
   stock market!
   old_market_mood=market_mood;
   % stock_market returns a NON ZERO market_mood, we
   can keep it

   current_cycle=((i/market_cycle ~=
   current_cycle)*current_cycle)+(i/market_cycle ==
   current_cycle)*(current_cycle+1);
   % if steps NOT reach use old mood, else change mood of
   market

   market_mood=(i/market_cycle == current_cycle-1)*
   behaviour_of_market(current_cycle);
   %  market mood changes every 500 steps←set by user
   %  current_cycle was increase since the first
   %  step so i need to -1 to the current_cycle

   %  prob = exp(dE);
   %  probability whether or not to flip a spin
      prob = exp(-beta.*dE); % change beta to a NxN matrix!
   %  the negative is with the dE term
   transitions = (rand(N) < prob ).*(rand(N) < randTol) * –2 +
   1; % The 'new' space

   sigma = sigma .* transitions;
   M(i) = sum(sum(sigma));
   % E(i) = –sum(sum(dE))/2;
   % Divide by two because of double counting
   E(i) = sum(sum(dE))/2;
   % Divide by two because of double counting
   %=====================================
   %==========Output phase==================
   image((sigma+1)*128);
   plot_title = sprintf('%dx%d Stock Market variant of
   Metropolis Ising model',N,N);     %  converts to string
   title(plot_title);
   xlabel(sprintf('J = %0.2f, M = %0.2f, E = %0.2f, i= %d',
   J, M(i)/N^2, E(i)/N^2 , i));
   %set(gca,'YTickLabel',[],'XTickLabel',[]);
   set(gca,'YTickLabel',[],'XTickLabel',[],'XTick',[],
   'YTick',[]);
   axis square; colormap copper; drawnow;
%==========================================
end %end of for loop!
```

```
figure(2)
set(gcf,'Position',[200 50 1800 700],'Color','White')
%subplot(1,2,1)
hold on
plot(1:steps,E(:,1),'–')
xlim([0 steps])
%ylim([0 2*10^5]) %????????????????????????????
title(['Price']);

%subplot(1,2,2)
%hold on
%plot(1:steps,beta,'–')
%xlim([0 steps])
%ylim([–1 1]) %????????????????????????????
%title(['beta']);
```

Acknowledgements

I would like to thank my division head, Anh Tuan Phan, for giving me the opportunity to develop this econophysics course. I would especially like to thank the nine students, Boon Kin Teh, Ian Beng Hau Tan, Wei Liang Quek, Muhammad Al Atiqi, Henry Pui Loong Lee, Chee Wooi Tan, and Muhammad Uzair Arshed for putting in enormous efforts to develop materials for the open textbook, and also for lively discussions during lectures that made this course so enjoyable to teach.

References

Alvarez-Ramirez, J., Alvarez, J., Rodriguez, E., & Fernandez-Anaya, G. (2008). Time-varying Hurst exponent for US stock markets. *Physica A, 387*(24), 6159–6169.

Arthur, W. B. (1999). Complexity and the economy. *Science, 284*(5411), 107–109.

Bouchaud, J.-P., Gefen, Y., Potters, M., & Wyart, M. (2004). Fluctuations and response in financial markets: The subtle nature of 'random' price changes. *Quantitative Finance, 4*(2), 176–190.

The Brown Daily Herald (2013, 5 April). Economics see growing pains as students look for 'marketable skills'. Retrieved from http://www.browndailyherald.com/2013/04/05/economics-sees-growing-pains-as-students-look-for-marketable-skills/

Cajueiro, D. O., & Tabak, B. M. (2004). The Hurst exponent over time: Testing the assertion that emerging markets are becoming more efficient. *Physica A, 336*(3–4), 521–537.

Carbone, A., Castelli, G., & Stanley, H. E. (2004). Time-dependent Hurst exponent in financial time series. *Physica A, 344*(1–2), 267–271.

Cheong, S. A., Fornia, R. P., Lee, G. H. T., Kok, J. L., Yim, W. S., Xu, D. Y., & Zhang, Y. (2012). The Japanese economy in crises: A time series segmentation study. *Economics E-Journal, 6*, article 5.

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

Di Matteo, T., Aste, T., & Dacorogna, M. M. (2003). Scaling behaviors in differently developed markets. *Physica A, 324*(1–2), 183–188.

Di Matteo, T., Aste, T., & Dacorogna, M. M. (2005). Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *Journal of Banking & Finance, 29*(4), 827–851.

Durlauf, S. N. (2005, June). Complexity and empirical economics. *The Economic Journal, 115*(504), F225–F243.

The Economist (2007, 14 January). Economics, the gut major. Retrieved from http://www.economist.com/blogs/freeexchange/2007/01/economics_the_gut_major

Eom, C., Choi, S., Oh, G., Jung, W.-S. (2008). Hurst exponent and prediction based on weak-form efficient market hypothesis of stock markets. *Physica A, 387*(18), 4630–4636.

Feigenbaum, J. A. & Freund, P. G. O. (1996). Discrete scale invariance in stock markets before crashes. *International Journal of Modern Physics B, 10*(27), 3737–3745.

Gallegati, M., Keen, S., Lux, T., & Omerod, P. (2006). Worry trends in econophysics. *Physica A, 370*(1), 1–6.

Gallegati, M., & Kirman, A. (2012). Reconstructing economics: Agent based models and complexity. *Complexity Economics, 1*(1), 5–31.

Grech, D., & Pamula, G. (2008). The local Hurst exponent of the financial time series in the vicinity of crashes on the Polish stock exchange market. *Physica A, 387*(16–17), 4299–4308.

Kim, D.-H., & Jeong, H. (2005). Systematic analysis of group identification in stock markets. *Physical Review E, 72*, 046133.

Laloux, L., Cizeau, P., Bouchaud, J.-P., & Potters, M. (1999). Noise dressing of financial correlation matrices. *Physical Review Letters, 83*(7), 1467–1470.

Laloux, L., Cizeau, P., Potters, M., & Bouchaud, J.-P. (2000). Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance, 3*(3), 391–397.

Lewellen, J. (2002). Momentum and autocorrelation in stock returns. *The Review of Financial Studies, 15*(2), 533–563.

Liao, T. W. (2005). Clustering of time series data—A survey. *Pattern Recognition, 38*(11), 1857–1874.

Lux, T. (1996). Long-term stochastic dependence in financial prices: Evidence from the German stock market. *Applied Economics Letters, 3*(11), 701–706.

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B, 11*(1), 193–197.

Matos, J. A. O., Gama, S. M. A., Ruskin, H. J., Sharkasi, A. A., & Crane, M. (2008). Time and scale Hurst exponent analysis for financial markets. *Physica A, 387*(15), 3910–3915.

Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E., & Preis, T. (2013, 8 May). Quantifying *Wikipedia* usage patterns before stock market moves. *Scientific Reports, 3*, 1801.

Palmer, R. G., Arthur, W. B., Holland, J. H., LeBaron, B., & Tayler, P. (1994). Artificial economic life: A simple model of a stock market. *Physica D, 75*(1–3), 264–274.

Peters, E. E. (1989). Fractal structure in the capital markets. *Financial Analysts Journal, 45*(4), 32–37.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N., Guhr, T., & Stanley, H. E. (2002). Random matrix approach to cross correlations in financial data. *Physical Review E, 65*(6), 066126.

Preis, T., Moat, H. S., & Stanley, H. E. (2013, 25 April). Quantifying trading behavior in financial markets using *Google Trends*. *Scientific Reports, 3*, 1684.

Stanley, H. E., Afanasyev, V., Amaral, L. A. N, Buldyrev, S. V., Goldberger, A. L., Havlin, S., *et al.* (1996). Anomalous fluctuations in the dynamics of complex systems: From DNA and physiology to econophysics. *Physica A, 224*(1–2), 302–321.

Tóth, B., Lillo, F., & Farmer, J. D. (2010). Segmentation algorithm for non-stationary compound Poisson processes. *The European Physical Journal B, 78*(2), 235–243.

Tumminello, M., Aste, T., Di Matteo, T., & Mantegna, R. N. (2005). A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences of the United States of America, 102*(30), 10421–10426.

University of Wisconsin-Madison, College of Letters & Science News (2013, 6 February). Demand grows for Economics major. Retrieved from http://news.ls.wisc.edu/feature/demand-grows-for-economics-major/

Vaglica, G., Lillo, F., Moro, E., & Mantegna, R. N. (2008). Scaling laws of strategic behavior and size heterogeneity in agent dynamics. *Physical Review E, 77*(3), 036110.

Wong, J. C., Lian, H., & Cheong, S. A. (2009). Detecting macroeconomic phases in the Dow Jones Industrial Average time series. *Physica A, 388*(21), 4635–4645.

Zhang, Y., Lee, G. H. T., Wong, J. C., Kok, J. L., Prusty, M., & Cheong, S. A. (2011). Will the US economy recover in 2010? A minimal spanning tree study. *Physica A, 390*(11), 2020–2050.