The time-variation of volatility and the evolution of expectations

van der Weide, R.

Publication date
2012
Document Version
Final published version

The Time-Variation of Volatility and the Evolution of Expectations

Roy van der Weide

The Time-Variation of Volatility and the Evolution of Expectations

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties
ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op donderdag 8 maart 2012, te 12.00 uur

door

Roy van der Weide

geboren te Heemskerk

Promotiecommissie:

Promotores:	Prof. dr. C.H. Hommes
	Prof. dr. H.P. Boswijk
Overige leden:	Prof. dr. W.A. Brock
	Prof. dr. C.G.H. Diks
	Prof. dr. F.C.J.M. de Jong
	Prof. dr. A. Lucas
	Prof. dr. J. Tuinstra

Faculteit Economie en Bedrijfskunde

Aan mijn ouders, Anja en Willem

Acknowledgements

This thesis has been my labour of love for the past decade. I joined the PhD program in April 2000 following a four months backpacking adventure in South-East Asia with Niels Janssen (who had asked if I was interested in traveling to Asia with him as I was wrapping up my Masters dissertation a few weeks before our departure). I got close to finishing my PhD thesis three years later. Many people, including myself, have asked why I did not finish it then. After all, the introduction was the only chapter missing, and another chapter (a joint research project with Remco Peters) only needed some minor revision. So what took me so long?

Feeling optimistic, I left for Washington DC in June 2003 to work on a World Bank project that was supposed to take only a few months of my time. In the event I could not finish my thesis in the weekends and evenings, I would surely submit my thesis on my return to the Netherlands later that year. Fast forward to June 2011, I had now spent about five years of my life in Washington DC and three years in Lao PDR (South-East Asia) working for the World Bank. With time, I became increasingly immersed in my work at the World Bank which started taking precedence over finishing my thesis. As it turned out, I had become part of a large group of hopefuls that had joined the Word Bank and would submit their PhD thesis in the weekends that would follow. I will join a more exclusive group on March 8, 2012, when after a delay of eight years I will still make good on my promise of completing my PhD.

I owe my supervisors Cars Hommes and Peter Boswijk an enormous amount of gratitude for their patience and for not giving up on me over all those years. Cars Hommes has been a great support both academically and personally, both during my three years as a PhD student in Amsterdam and the years thereafter. Peter Boswijk joined as a supervisor some years later when it became apparent that time series econometrics would be the theme of the second half of my PhD research (which ended up as the first chapters of my thesis). Chapters two and three are the outcome of my joint research with Peter. Both of my supervisors have made invaluable contributions to all other parts of my thesis and have been an inspiration to my research more generally.

I am also very grateful to my commission members Buz Brock, Cees Diks, Jan Tuinstra, Andre Lucas and Frank de Jong, who after a long wait put in the effort to review my thesis and take part in my defense. The contribution of Cees Diks to this thesis goes well beyond his role as a commission member. The last two chapters of my thesis are the outcome of a research project Cees and I started in 2000 and continue to pursue today. I am also very grateful to Cees for his input and comments on the other chapters of my thesis. A special mention goes also to Jan Tuinstra for injecting a healthy dose of humour over the years.

Having been with CeNDEF from the time it was conceived, first as a research assistant and then as a PhD student, the research group has played an invaluable role in nurturing me as an academic. I have experienced my time at CeNDEF as a member of a closely knit family. Let me hereby extend my gratitude to all members of the family. I am looking forward to continuing being part of CeNDEF in one capacity or another.

Undoubtedly, there must have been a moment in the minds of everyone that I hold dear where he/she thought it would never come to this moment where I could share the defense of my PhD thesis with them. I owe it largely to all of you that we finally did it! Papa, mama, mijn lieve broer en zus, Jordanka, Nicole, Thierry, Hubert, Michelle, Tijana, Fabian, Remco, Ilse, Alex, Dennis, Soonhwa, Nico, Marieke, Himmat, Peter, Matthijs, Niels, Ria en Kees, it is with great relief, gratitude and affection that I am writing this to you.

A special thanks goes to Nicole Veninga for helping me with the Summary in Dutch, which although substantial still only denotes a small example of her love and support over the years. Another special thanks goes to Remco Peters for investing a great number of years in what appears to be evolving into a life-long research project (of which chapter four provides a snapshot) while juggling parenthood with the nurturing of a co-founded financial investment firm. Also Dennis Botman deserves a special mention. Were it not for Dennis we may well not be celebrating my defense coming March 8, 2012. Dennis introduced me to Washington DC. Had Dennis never gone to Washington, and thus never introduced Washington to me, chances are we would now be celebrating the eight year anniversary of my defense.

Finally, I want to express my love and gratitude to my parents, brother, sister and my grand parents, and my Jordanka. To my regret I have never been very good in telling them how much I love them. Surely, without their love and support throughout my life, I would

not be where I am today. Please know that I love you!

You will now understand why March 8, 2012, the day I will defend my PhD thesis, will be that much more special. Remember that Rome was not built in a day, and heerlijk duurt het langst!

February 16, 2012 Washington DC, Sofia, Copenhagen

Contents

Chapter 1

Introduction and thesis outline

Equity prices, house prices, oil prices, electricity prices, food prices, the price of a currency, they all fluctuate over time. The majority of these different products are traded on active and well-regulated markets where their prices are determined by demand and supply. While demand will largely be driven by consumption, all markets owe part of their liquidity to speculative behavior. The price series are seen to share a number of important empirical features, also known as stylized facts, even though market characteristics may be quite different.

Price bubbles, or boom and bust cycles, denote a prominent example of a stylized fact. During a ‘boom’ prices are seen to grow for an extended period of time, also known as a ‘bull market’. Market participants may believe that the price has risen well above their perception of the fundamental value, but choose not to opt out. Inevitably, prices will come down again, that is, a ‘bear market’ will occur. When they do, prices often come down by as much as they had gained earlier, but considerably faster. This is known as ‘asymmetric volatility’: “Volatility tends to be higher in bear markets” (Engle, 2004).

Another key stylized fact is that the volatility of price returns is universally found to vary over time. Volatility is defined as the standard deviation of the distribution from which the price returns are drawn. While this distribution is not directly observable, such that volatility is strictly speaking invisible, larger returns (in absolute value) are likely to be drawn from distributions with larger standard deviations. Some markets will be more volatile than others, yet most markets see phases of high volatility alternate with phases of low volatility.

This introductory chapter will give a general overview of the themes that are addressed in this thesis, which can be divided into two parts. Both parts are concerned with capturing the key stylized facts of empirical prices. The first part builds an "econometric model" for the time-variation of multivariate volatility. The objective here is to obtain a good fit of the volatility dynamics so to improve forecasts. In the second part we build "structural economic models" with the objective to gain a better understanding of what economic and behavioral mechanisms are responsible for the observed price and volatility dynamics. To make this opening chapter a coherent story that can be read on its own it embeds the subjects covered by this thesis into a larger review of the literature. The closing section of this introduction provides a focused summary of each of the thesis chapters.

1.1 Price bubbles: Boom and bust cycles

Arguably the first documented example of a price bubble, or the most famous, is the Dutch tulipmania that occured between 1634 and 1637. In an early example of a futures market, an organized gathering of traders in numerous taverns whose linkage to the physical market is debated, the price of rare tulip bulbs was seen to rise to extraordinarily high levels. At the height of the speculation, early February 1637, prized single bulbs were sold at an equivalent of an average person's annual income. When prices collapsed, these bulbs reportedly lost around 90 percent of their value in a short period of time. A careful analysis of the tulipmania can be found in Garber (1989, 1990). For a recent discussion offering different historical perspectives on price bubbles, see O'Hara (2008).

The Wall Street crash of 1929 is arguably the most profound stock market crash in the history of the United States. The Dow Jones Industrial Average (DJIA) lost around half of its value in a space of two months. It reached a peak value of 381.17 in September 1929 followed by a low of 198.6 in November 1929. During the boom hundreds of thousands of Americans had entered the stock market, which were in all likelihood drawn by the steady gains the market had accumulated in the years leading up to the crisis. Much of the funds that fueled the speculative bubble was borrowed money. The collapse marked the beginning of the Great Depression, which lasted for more than a decade until the second World War. In 1933 the United States put policies in place that would separate commercial banks from investment banks (the Glass-Steagall Act), designed to prevent history from repeating itself. (Interestingly, governments are considering similar policies today in a response to the global financial crisis of 2008-2011.) See e.g. Shiller (2005) for a discussion of the speculative bubble.

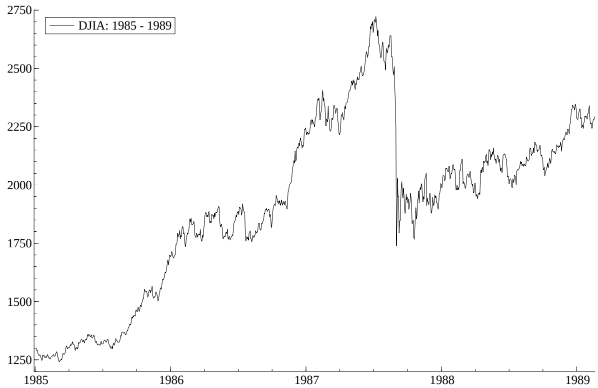

Figure 1.1: *Dow Jones Industrial Average (DJIA): 1985-1989*

Another famous stock market crash occured about half a century later, the crash of October 1987. Figure 1.1 shows the empirical time-series of the Dow Jones index (a mix of US equity prices) for the period 1985 to 1989. One immediately recognizes the bursting of the bubble on Black Monday, October 19, 1987, after an extended period of steady growth. The index gradually gained about 1000 points over the two year period prior to October 1987, and then lost it all in a matter of days.

A current example, and no less dramatic, is the 2008-2011 global crisis that includes a boom and bust in oil prices, a world food price crisis, and the real estate crisis that fueled a global financial crisis. The latter crises are all inter-connected. Figure 1.2 plots four different empirical price series for the period 2006 to 2009: (1) the Dow Jones Industrial Average, (2) the price of crude oil, (3) the price of Thai export rice, and (4) the Case-Shiller house price index. The time period covers both the boom and the bust. One can clearly see how the global financial crisis inter-links with the crisis in housing, oil and food prices. Note that the decline in house prices preceded, and possibly initiated, the global economic crisis. The price of oil came down after the bubble had burst in financial markets accompanied by economic recession, as this meant a collapse in the demand for oil.

1.2 Price volatility and uncertainty

Figures 1.3 shows the last 30 years of daily (log) price returns for the DJIA (the period 1980 to 2010). It can be seen how volatility varies over time. Periods characterized by large

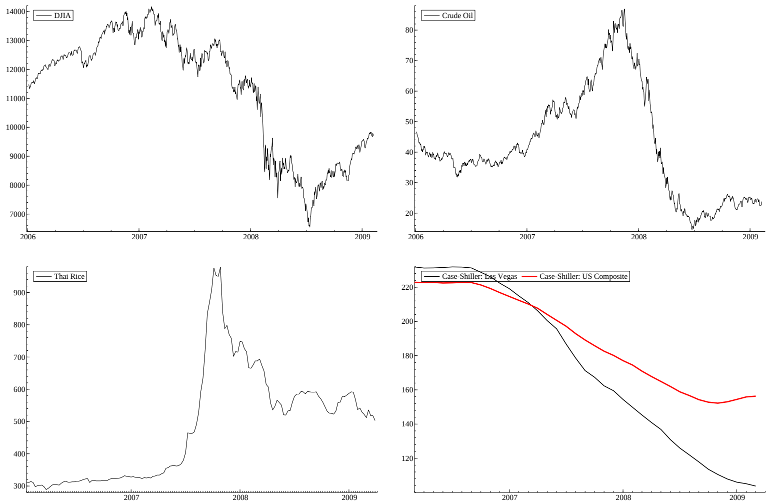

Figure 1.2: *Price of DJIA, Crude Oil, Thai Rice, and the Case-Shiller housing Price Index: 2006-2009*

(small) price changes may be associated with periods of high (low) volatility. Notice that all return series show a similar pattern: Large (small) price fluctuations today are likely to be followed by large (small) price fluctuations tomorrow, which is an indication that volatility is persistent (today's volatility would be a good predictor for tomorrow's volatility). Phases of high (low) volatility are eventually followed by phases of low (high) volatility, and the cycle repeats itself. This stylized fact is known as 'volatility clustering'. The empirical properties of stock return volatility have been studied by e.g. Anderson (1996), Hamilton and Lin (1996), and Anderson *et al.* (2001). For modeling and forecasting of volatility, see e.g. Anderson *et al.* (2003) and Poon and Granger (2003), and the references therein.

Volatility is associated with uncertainty and risk. An increase in the volatility of an asset's value implies an increase in uncertainty for those who have invested in the asset. A higher volatility means that the investor is at risk of making a higher loss (as well as making a larger profit). The financial sector has made this uncertainty and risk their business by developing increasingly sophisticated ways of measuring and managing risk. It has created a market for derivatives, which denote the primary instruments for mitigating financial risk, also known as 'hedging risk'. Popular examples of derivatives are 'Put' and 'Call' options. The Put option locks-in a minimum value for an asset at which it may be sold in case the actual value falls short of that. The Call option gives an investor looking

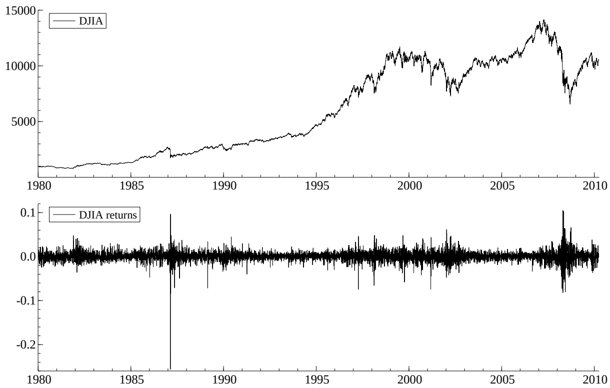

Figure 1.3: *DJIA prices and log returns: 1980-2010*

to buy an asset the guarantee that he/she will not have to pay more than is described in the contract. The prices of these option contracts may be viewed as insurance premiums as it insures the buyer of these contracts against unforeseen price movements.

The uncertainty in future prices and exchange rates brings challenges to many segments of the economy, from large economic agents to small entrepeneurs. Investment banks deal with the uncertain values of financial assets. Real-estate developers look out for changes in land and house prices. A government's monetary policy is in part shaped by expectations about future inflation rates. The profit of import/export companies may be exposed to exchange rate risk.

1.2.1 Macro: Energy price uncertainty

Energy being a primary commodity, large movements in the price of oil and sustained uncertainty in future prices will have significant impacts on the global economy. Bernanke (1983) predicts that optimizing firms will postpone investment decisions when faced with uncertainty about energy prices which will ultimately depress aggregate output. This is confirmed empirically by e.g. Cameron and Schnusenberg (2009), Elder and Serletis (2009, 2010), and Yoon and Ratti (2010). "[V]olatility in oil prices has had a negative and statistically significant effect on several measures of investment, durables consumption and aggregate output" (Elder and Serletis, 2010). Similarly, oil price shocks are found to depress cash flows and lower investments, see e.g. Dotsey and Reid (1992) and Jones and

Kaul (1996). Rising oil prices and heightened uncertainty have also been linked to food price increases (see e.g. Alghalith, 2010). Hamilton (1983) observed that rising oil prices are often followed by major downturns in economic activity.

The large price swings and high volatility in the global market for oil in recent years has renewed the interest into studying the determinants of oil price dynamics. Conjectures concerning the possible determinants include: speculative trading, unexpected contraction in supply (for example when OPEC decides to withhold oil supplies from the market), and global business cycles. The latter would for example attribute the recent surge in prices to the unexpected strong growth in the global economy driven largely by emerging Asia. In a recent empirical study, Kilian and Murphy (2010) find support for the latter, namely that "the sustained run-up in the real price of oil between 2003 and mid-2008 was caused primarely by shifts in the global flow demand for oil".

Whether speculative behavior also had a part in the rise and fall of oil and food prices of late has recently become a subject of popular debate. The opinions are divided. While it is a tempting conjecture, it is argued that speculators would need to stockpile the commodities for an extended period of time to sustain the rise in prices. The fall in prices would indicate when their stocks are being returned to the market. There is little evidence, however, on the existence of these stockpiles. By the same token, there is tentative evidence that a segment of the oil market in part relies on extrapolating price trends to determine demand, see e.g. ter Ellen and Zwinkels (2010).

It has long been recognized that commodity prices are extremely volatile, see for example Deaton and Laroque (1992): "For some commodities, there have been swings from trough to peak in just a few months. For countries whose export earnings and GNP are dependent on these commodities, such volatility poses major problems both of macroeconomic and microeconomic policy. An understanding of the stochastic processes governing these price movements is essential for macroeconomic management, for national consumption and saving policies, for agricultural pricing policies, and for the design of risk-sharing mechanisms between farmers, resource holders, and government".

1.2.2 Micro: Firms and farmers

A high price for oil is bad for businesses as it raises costs. It yields high production costs as well as high transport costs. Companies could incur the costs themselves, which would lower their margins and thus curb their profits. Alternatively, companies may opt to recover the extra costs by increasing consumer prices at the risk of lowering the demand for their produce. For selected products manufacturers will plausibly still see a decline for their produce even if they do not touch the price. Automobiles are a good example, particularly those characterized by excessive fuel consumption. Cameron and Schusenberg (2009) find that when oil prices are high manufacturers of large cars that are not fuel-efficient (e.g. SUVs) are more likely to see their stock prices decline than manufacturers of smaller passenger cars that have a more favourable fuel consumption.

High levels of uncertainty concerning future oil prices arguably also puts pressure on the present value of airline companies. Higher oil prices raises the costs of travel, which airlines cannot always pass on to their customers each time the price of oil goes up, so that it will cut into their profits. For one, they have to compete with other airlines, some of which might be less sensitive to changes in the price of oil, either because they are more fuel efficient or because they were more successful in hedging against the increase in the price of oil. Boswijk and van der Weide (2009) examine the time-varying correlations between changes in the price of oil and changes in the stock prices of selected airlines and transport companies. Upon comparing the airlines, they find that South-West is least sensitive to movements in oil prices, which may be attributed to South-West's active program to hedge fuel prices: "they buy fuel options when the price of oil is believed to be low to hedge against potentially high fuel prices years later. These hedges have for example helped South-West through the Iraq war and hurricane Katrina when oil prices indeed increased significantly. In the third quarter of 2008 ... South-West recorded its first loss in many years, partly because the then noticeable drop in oil prices had rendered the fuel hedging strategy of lesser value" (Boswijk and van der Weide, 2009).

Correlations are found to peak in periods with increased uncertainty in future oil prices; in 2002 with the build-up to the Iraq war, and in 2008 when oil prices reached their peak against the backdrop of a looming financial crisis. "At the outset of the financial crisis, oil prices were still climbing ... Thus when the recession in the United States was already looming, the price of oil was still high ... there was considerable uncertainty at the time whether the recession would bring down oil prices. Much of the (new) demand for oil has come from emerging economies such as China, as well as from Europe. It was uncertain if and how the recession in the United States would carry over to the emerging markets and Europe. In case of a decoupling the demand for oil would stay strong despite the recession originating in the United States. In this time of uncertainty, correlations increased in absolute value. When it then became clear that the recession would be global also oil prices came down, together with stock prices" (Boswijk and van der Weide, 2009).

A farmer would be an example of a smaller entrepreneur for which uncertainty in future prices brings a challenge. “Farmers the world over, in dealing with costs, returns, and risks, are calculating economic agents. Within their small individual, allocative domain they are fine-tuning entrepreneurs” (Nerlove, 1979). For the planning of the crop production, i.e. when deciding on how much to cultivate of each crop, the farmer will be relying on his/her estimates of post-harvest prices. Where farmers will inevitably make errors in the forecasting of agricultural prices, it may be expected that rising uncertainty concerning future prices will put pressure on their profits. “In so far as farmers formulate production plans on the basis of their price expectations, the failure of expected prices to correspond to realized prices will prevent ex-post profits from being maximized. Equally important, errors in price expectations may impair the efficiency of agriculture” (Heady and Kaldor, 1954).

In developing countries, where both infrastructure and markets are not yet fully developed, building expectations of future prices is often hampered by poor access to information. When a market for derivatives and insurance is also missing, farmers will find it difficult to hedge their risks against unforeseen price movements, whether that is due to demand shocks or supply shocks. Yet, also farmers in the developed world, where access to information is more efficient and where derivative markets are available to hedge against price risks, the price fluctuations in the commodity markets still represent a significant difficulty.

In April 2008, the New York times published an article entitled ‘Price Volatility Adds to Worry on U.S. Farms’, for which they interviewed a farmer from Illinois, Fred Grieder, who has been farming on 1500 acres for 30 years. A typical days work may range between 12 and 20 hours, which means long days of plowing, planting, fertilizing, and looking after the crops. “But Mr. Grieders days on the farm ... are getting even longer. He now has to keep a closer eye on the derivatives markets in Chicago, trying to hedge his risks so that he knows how much he will be paid in the future for crops he is planting now. And the financial tools he uses to make such bets are getting more expensive and less reliable ... todays crop prices are not just much higher, they also are much more volatile ... The price swing expected in March for soy beans was three times its monthly average, and the expected volatility in corn prices was twice its monthly average ... there is no question that the grain markets are now experiencing levels of volatility that are running well above the average levels over the last quarter-century ... Mr. Grieders crop insurance premiums rise with the volatility. So does the cost of trading in options, which is the financial tool he has used to hedge against falling prices ... In any case, at current levels of volatility, options trading becomes riskier, and therefore more expensive, too expensive for many farmers like Mr. Grieder, who now has to hedge with the recently less reliable futures contracts. That exposes him to the risk of having to put up more cash, to maintain his price protection, whenever a weather threat, shipping disruption or a fresh surge of money from Wall Street suddenly pushes up grain prices ... 'If you have got 50,000 bushels hedged and the market moves up 20 cents, that would be a USD10,000 day,' he said. 'If you only had USD10,000 in your margin account, you would have to sit down and write a check ... On an unusual day, he said, he might get four phone calls a day from his broker seeking additional margin ... he sometimes has to rely on his bank to advance him the margin he needs to keep those hedges in place, a worrisome requirement even for a successful farmer in an economy already struggling with a credit squeeze ... Farmers used to leave the market-watching to traders who work for big grain elevator companies. But with some of those companies now refusing to buy crops in advance because hedging has become so expensive and uncertain, farmers have to follow and trade in those markets themselves" (The New York Times, 2008).

1.3 The determinants of price- and volatility dynamics

What makes prices fluctuate to the extent that they do? Can the enormous swings in prices that define a boom and bust cycle be justified by changes in fundamentals, or are there more forces at work? Naturally, markets will re-evaluate the price of an asset as news comes in that concerns the intrinsic value, think of earning anouncements, newly found demand, technological discoveries, changes in the price of energy, changes in the interest rate, plans for a merger or acquisition etc. For example, Garber (1989, 1990) hypothesizes that during the boom and bust cycle of the tulipmania (1634-1637) the price of rare bulbs were still in touch with fundamentals. The nature of the process that ultimately transforms rare bulbs into commercially viable tulips, which essentially gives the owner the rights to the production of the new species of tulips, offers a rationale for the dramatic price movements. In fact, similar price swings can still be observed today allbeit to a lesser extent due to technological advancements over the years.

There is evidence however that suggests that these news events are not fully able to explain the magnitude of observed price changes in general, i.e. price changes are believed to be more volatile than changes in the fundamental value. Shiller (1990) refers to this discrepancy as 'excess volatility'. Also volatility is found to fluctuate, see e.g. Schwert (1989) who addresses the question as to "Why does stock market volatility change over time?". What makes prices and volatility fluctuate can arguably be traced back to both economic and behavioral factors.

Among the prominent economic determinants of market volatility are the volatility of macroeconomic factors such as GDP growth, industrial production, inflation, short term interest rate, and the exchange rate (see e.g. Engle and Rangel, 2008; and Engle *et al.*, 2009). For some macroeconomic factors levels too are found to play a role, "there is evidence that high inflation and slow growth of output are also positive determinants" (Engle *et al.*, 2009). For argicultural commodities, other economic factors such as the cost of storage may have an impact on price volatility.

Behavioral factors too may shape the dynamics of prices, think of economic agents updating their expectations or adjusting their degree of risk aversion (likely triggered by news events and/or observed price fluctuations). Economic agents determine how much they buy or sell of an asset based on their expectations concerning its future value in an effort to seek a profit or avoid a loss. In a world where prices fluctuate, at a minimum driven by changes in the fundamental value, agents will want to anticipate price movements. This leads them to form expectations about future prices, which they will update as new information (that includes newly observed prices) becomes available. A 'feedback loop' emerges where today's prices shape the expectations that will determine tomorrow's prices. It can be shown that this expectation formation process will generally create fluctuations of its own. Moreover, it is conceivable that the 'endogenous dynamics' will also act to amplify the significance of exogenous shocks to prices. The ways in which expectations shape the dynamics of prices can take different forms. See e.g. Brock and Hommes (1997, 1998), Lux and Marchesi (1999), Evans and Honkapohja (2003), Boswijk *et al.* (2007), Gaunersdorfer *et al.* (2008), Heemeijer *et al.* (2009), Burnside *et al.* (2011), and the references therein.

1.3.1 Rational expectations

The economist will need to make assumptions about the economy, think of market structure, production, depreciations and costs, as well as about the behavioral underpinnnings, think of preferences and expectations. Here we focus on the latter, in particular on the expectation formation process. The standard assumption is that agents have 'rational expectations'. This is often assumed to imply that agents either have perfect foresight (in deterministic models) or make forecasting errors that are zero on average (in stochastic models). In the original theory of rational expectations put forward by Muth (1961), it is not so much assumed that individuals (firms in Muth's case) were rational, but that the market as a whole, the aggregate, is rational. Individual forecasting errors are assumed to cancel out on average, so that the average individual has perfect foresight. This is weaker than assuming rationality at the individual level. Lucas (1972) denotes an early and well-known example that adopts the rational expectations framework.

In general, however, we have that a collective of agents with heterogeneous expectations behaves differently than a single 'representative agent' holding their average expectations. In other words, Muth (1961) relies heavily on assumptions of linearity. That does not mean that the hypothesis of Muth (1961), that aggregate expectations are rational, is unreasonable. It finds supports in early survey data on expectations, see e.g. Heady and Kaldor (1954): "For example, the average forecast for corn in 1949 differed from the realized price by about 1 per cent. Yet, only 52 per cent of the individual forecasts fell within plus or minus 10 per cent of the realized price".

Over time, rational expectations prevailed as a popular assumption, also when the linearity assumption is not satisfied. This means that in general all agents are assumed to be rational. It largely owes its success to its analytical tractability, and that the alternative is less well-defined, i.e. if expectations are not rational they could be almost anything.

Note that rational expectations are not necessarily inconsistent with price bubbles. It is harder however to rationalize boom and bust cycles, which is what we observe in practice. In a standard rational expectations asset pricing model, prices can either follow the fundamental or grow indefinetely, away from the fundamental (see e.g. Flood and Garber, 1980). For boom and bust cycles to emerge we need prices to switch between the fundamental and the bubble solution. This would arguably require a non-rational intervention; think of a fraction of the market questioning the bubble, fearing it has to stop some day.

In the end rational expectations is a rather strong and thereby an unrealistic assumption. Perfect foresight implies perfect knowledge of the underlying model, which often requires knowledge of the expectations of other market participants. Also, if all agents were rational, we would arguably not see as much trade as we do, see the no trade theorems of e.g. Milgrom and Stokey (1982) and Tirole (1982). Rational expectation models are arguably not well equipped to rationalize the observed data.

1.3.2 Boundedly rational expectations

As the evidence mounted against rational expectations the concept of bounded rationality emerged as a more realistic alternative assumption, see for example the survey by Sargent (1993).[1] "Economics and finance are witnessing an important paradigm shift, from a representative, rational agent approach towards a behavioral, agent-based approach in which markets are populated with boundedly rational, heterogenous agents using rule of thumb strategies [...] A boundedly rational agent forms expectations based upon observable quantities and adapts his forecasting rule as additional observations become available" (Hommes, 2006). The forecasting rules can be as basic as extrapolating trends or as sophisticated as econometric modeling where unknown parameters are identified from the data. Learning can also be accommodated, which may entail that the econometrician updates his/her model parameters as new data becomes available, or he/she might switch to a new model all together if that fits the data better.

Different agents may work with different models, update their models using different criteria and operate at different time horizons (think of long-term versus short-term investors), which introduces natural levels of heterogeneity. Models of expectation formation that go by these rules are deemed an accurate and realistic account of human behavior.

In theory, boundedly rational expectations may in time converge to rational expectations, if the underlying model does not vary over time (or very gradually) and if the model is not too complex, so that a Bayesian econometrician has a fair chance of learning the underlying mechanisms (see e.g. Cyert and DeGroot, 1974). By the same token, this may realistically never happen. Also heterogeneity might steadily decline over time until all agents agree with each other on future prices (in which case trade may cease to exist), but this probably too will not happen.[2] Realistically, any given market will have heterogeneity in abundance.

Consider the following three different mechanisms by which the expectation formation process may introduce endogenous dynamics in prices: (i) rational herding where agents will follow recent trends in prices even if that means a deviation from the perceived fundamental value; (ii) performance based switching between cost-efficient naive estimates and costly sophisticated estimates; and (iii) changing weights given to prior beliefs relative to observed data. Let us briefly elaborate on each of these mechanisms.

[1] Early ideas of bounded rationality can be traced back to Simon (1957), before Muth (1961) launched his concept of rational expectations.

[2] For evidence that heterogeneity is likely to prevail, see e.g. Lahiri and Sheng (2008).

Rational herding (see e.g. Scharfstein and Stein, 1990; Froot *et al.*, 1992): With this we have in mind a world where agents will follow each other, which may temporarily drive prices away from the fundamental value. One could think of different reasons why such herding behavior may be optimal. It is conceivable that following the crowd may at times lead to more accurate price forecasts. In periods where speculative forces are stronger than those of the 'underlying economic system', prices may be expected to diverge from the fundamental. It would then be rational to follow the herd rather than coordinate on fundamentals (see e.g. Froot *et al.*, 1992). It can be shown that a world where agents extrapolate price trends to predict future prices is prone to endogenous price fluctuations (see e.g. Hommes (2006) and the references therein).

Switching between naive and rational expectations (Brock and Hommes, 1997): It does not always pay to invest in costly estimates of future price movements. It only pays when the benefits exceed the costs. With a stylized model Brock and Hommes (1997) illustrate how a calculated switching between cheaply available 'naive expectations' and costly 'rational expectations' will introduce price dynamics where prices will fluctuate close to equilibrium for an extended period of time, but eventually move away from equilibrium. It is assumed that rational expectations have a stabilizing effect, i.e. when the entire market acts rationally, prices will tend to the equilibrium value. Similarly, naive expectations are assumed to destabalize the market. Consider an initial state where prices fluctuate close to equilibrium. In this case it is not hard to predict tomorrow's price. Naive expectations will then be just as good as the more sophisticated expectations, but cheaper. As naive expectations are now the cost-efficient option, more and more agents will shy away from investing in expensive estimates. Until almost the entire market is adopting naive expectations, which is when the price dynamics becomes unstable. Prices will start to diverge from equilibrium. As price fluctuations become larger, they also become more unpredictable. At some point the benefits no longer outweigh the costs of acquiring rational expectations, and hence rational expectations will start to attract more agents. Until the market as a whole becomes rational, which is when prices will return to equilibrium, and we are back to where we started.

Choosing between prior beliefs and observed data (Diks and van der Weide, 2011): In this world, agents act as Bayesian statisticians, they learn about the underlying model from observed price data, and combine this with their prior beliefs to obtain estimates of future prices. Analogous to Brock and Hommes (1997) it is assumed that the market is unstable when all agents entirely discard observed prices when building price expectations.

It is also assumed that agents believe in a changing world, i.e. that the model underlying the price dynamics from before the 1987 stock market crash (in a world without internet) is believed to be different from the model underlying today's price dynamics. This means that agents will give more weight to recent price data relative to older price data, and will eventually forget about distant observations. The mechanism works as follows. Consider again the initial state where prices are close to equilibrium. When prices show very little fluctuations, they offer little information. In time, the last period of substantial price variation will gradually be considered less relevant as the underlying model may have changed over time. If observed prices remain uninformative, in the form of little price variation, agents will start to attach more weight to their prior beliefs, and in effect less and less to observed data. Until prevailing market expectations are no longer shaped by observed data, which is when the market becomes unstable, and prices will start to deviate from equilibrium. Prices now fluctuate, making observed price data informative. Soon agents will be trading their priors for data. Then, as the underlying model is becoming known to the learning agents, the market stabilizes, and prices tend to equilibrium.

1.3.3 What do real-life expectations look like?

Models that accommodate a more realistic expectation formation process are more likely to be consistent with empirical data. What makes expectations more realistic? To answer that question, different ways of collecting data on expectations have been explored, think of: (i) surveys that directly ask for expectations held by economic agents such as farmers or financial analysts, (ii) laboratory experiments that allow the researcher to monitor the expectations and actions of human subjects as they interact in an artificial market, and (iii) inferring market expectations from observed price data. (The latter is only a realistic option for selective markets, such as the markets for futures, forward exchange rates, and options.)

In the mid-twentieth century, Heady and Kaldor (1954) conducted a three-year study of expectations held by farmers from the ten southern counties of Iowa in the United States. While farmers and producers in the United States have obviously developed over time, some of the distinctive features of expectation formation may be expected to apply today still. Moreover, some of the insights obtained by studying farmers from the developed world in the 1940s (shortly after the second World War) may well carry over to today's farmers in the developing world. From their interviews they learned that the farmers commonly "start the process of devising expected prices from current prices. The current price was then adjusted for the expected effects of important supply-and-demand forces. Where farmers possessed little information about these forces, there was a tendency to project either the current or the recent price trend". When it comes to forming expectations, financial specialists are not that different from farmers. Evidence on exchange rate expectations held by financial specialists, collected by different surveys over the years, is summarized in Hommes (2006): "A consistent finding from survey data is that at short horizons investors tend to use extrapolative chartists' trading rules, whereas at longer horizons investors tend to use mean reverting fundamentalists' trading rules".[3]

Laboratory experiments confirm that humans, when faced with simple decision problems under uncertainty, such as planning production or trading on a market, are more likely to employ simple heuristics than to act as rational agents. For example, the asset pricing experiments by Smith *et al.* (1988) show how price bubbles can be seen to emerge in artifial markets where all market participants have sufficient information to compute the fundamental value of the asset that is traded. This is confirmed by Hommes *et al.* (2005) whose experiments show that persistent price deviations from the fundamental are reinforced by trend following strategies. Subsequent experminents by Dufwenberg *et al.* (2005) and Haruvy *et al.* (2007) however find that these price bubbles tend to shrink or are elimenated all together when experienced traders are included (even if they are outnumbered by inexperienced traders). This tells us that laboratory experiments can be learned which then stabilizes the experimental market. Real markets however are considerably more complex. Waves of new technologies and exogenous shocks mean that real-life conditions are ever changing making them considerably harder to learn. In a recent experiment, Hussam *et al.* (2008) try to mimic these conditions by introducing increased dividend and liquidity uncertainty in which case bubbles are seen to re-emerge even with experienced traders. "[I]ncreased experience in the same environment is the only condition that has reliably elimenated price bubbles" (Hussam *et al.*, 2008). Inexperienced traders plausibly still have a part in nurturing price bubbles (with experienced traders unable to elimenate them). This is confirmed in a recent empirical study by Greenwood and Nagel (2009). Using data on actual mutual fund managers they find that inexperienced managers are more likely to adopt trend-chasing behavior, tend to extrapolate from the limited data they have observed in their career, and are more prone to the type of optimism that fuels bubbles. Consistent with these observations is that at the peak of the technology price bubble, it was the less experienced fund managers that were more heavily invested in the

[3]See e.g. Frankel and Froot (1987, 1990), Allen and Taylor (1990), Ito (1990), and Taylor and Allen (1992).

riskier technology stocks.

Observed price data too may hold information about expectations. Market prices are made from different ingredients, of which market expectations is one. While expectations are seen to go into market prices, it is not given that one can also get them out. In some cases you can. From option price data for example one can infer market expectations about future financial volatility: "Because option value depends critically on expected future volatility, the volatility expectation of market participants can be recovered by inverting the option pricing formula" (Dumas *et al.*, 1998).[4]

Jackwerth and Rubinstein (1996) and Bates (2000) are among the studies that exploited option price data to learn about the subjective distribution of volatility. They find that the subjective distributions of volatility from before and after the stock market crash of 1987 look very different. There is no evidence, however, that the objective distribution of volatility experienced a similar change. That suggests that market expectations have changed. The way the market now models the volatility process is consistent with the objective distribution of volatility, while volatility expectations from before the crash were clearly misspecified. This shows that market participants are not rational agents, but that they are learning. At times, learning may be triggered by dramatic events.

There is also evidence that economic agents are different, perceive information differently, and form different expectations, as might be expected. In a recent paper, Lahiri and Sheng (2008) note in their opening paragraph that "various survey data on expectations from many countries over [the] last fifty years have produced mounting evidence on substantial interpersonal heterogeneity in how people perceive the current and form inference about the future economic conditions". Heady and Kaldor (1954), in their study of U.S. farmers in the late 1940s, find that farmers use a variety of models involving various degrees of complexity: "No single procedure was employed by all farmers. Moreover, the same farmer often used more than one procudere, depending upon the amount of information possessed and upon the degree of confidence attached to it". This shows that economic agents do not only adopt different models, over time they also make calculated switches between models.

[4]Their statement holds true because the option value as a function of volatility is indeed invertible.

1.4 The benefits of having a good model

What are the benefits of modelling price dynamics? That largely depends on whether one is building an economic or an econometric model. Risk measurement, portfolio allocation, derivative pricing all benefit from having a statistical model that accurately descibes the time-variation of asset prices including its volatility process. A theoretical model of the economic and behavioral mechanisms underlying the observed price dynamics offers different benefits. It may offer explanations for observed regularities or anomalies, such as the recent price swings in the oil and food markets. By tweeking the models they may be used to study the effects of changes to the economy, think of innovations that alter the cost functions of producers. Similarly, economic models may be adopted to study the implications of policy interventions, think of price subsidies, regulations that alter market trading rules etc.

For example, the hope is that a better understanding of what is underlying the dramatic fluctuations in oil prices will help policy makers to decide between policies designed to curb the fluctuations in the oil market. Under the assumption that speculative trading plays an important role, "there has been considerable political pressure recently to impose regulatory limits on trading in oil futures markets" (Kilian and Murphy, 2010). Alternatively, if "surges in the global business cycle are the chief cause of high oil prices, then efforts aimed at reviving the global economy after the financial crisis are likely to cause the real price of oil to recover as well, creating a policy dilemma" (Kilian and Murphy, 2010).

What does a good model look like? Also this depends on whether one is looking at an economic or an econometric model. For an econometric model all that matters is that it fits the data well and provides good out-of-sample forecasts, so that risks will be accurately measured and derivates will be accurately priced. An economic model too is judged on how well it is able to describe the observed data, but importantly, the model is also judged on whether it manages to do so by relying on sensible economic and behavioral assumptions.

One could conceivably construct different economic models that are capable of replicating the same price dynamics. When tweeked, however, they may lead to different results. Thus, the relevance of the model as a tool for studying policy interventions relies crucially on whether the model has captured the appropriate underlying mechanisms. The assumptions the model is built on play an important role here. "Evidence from behavioral finance helps us to understand for example that the ... worldwide stock market boom, and then crash after 2000, had its origins in human foibles and arbitrary feedback relations ... The challenge for economists is to make this reality a better part of their models" (Shiller, 2003).

1.4.1 Building economic models

Having many agents with different expectations interacting with each other, however uncomplicated each individual rule, will likely lead to complicated dynamics. There is obviously a trade-off between realism and complexity. A strand of literature that wishes not to compromise on the complexity inherent to more realistic models of expectation formation relies on computers to trace the large numbers of interacting agents. Using computer simulations one can study aggregate outcomes, such as market prices, profits, trade volumes, and levels of heterogeneity without having to impose constraints on the number of trading rules and learning mechanisms (for an overview, see LeBaron, 2000). This is how the Santa Fe artificial stock market came into existence (see e.g. LeBaron *et al.*, 1999). "In the Santa Fe market, the actions of agents, which are based upon their expectations, are explicitly modelled and traced for each individual agent. With computers becoming cheaply available and faster, these artificial markets allow for more and more detailed modelling" (Diks and van der Weide, 2005). The approach also has disadvantages, one is that the mechanisms underlying the aggregate outcomes essentially become a black box (just as with observed data), i.e. "it is not always clear what exactly causes an observed simulation outcome" (Hommes, 2002). While one can of course still tweek the model, and then examine the changes in outcomes, the approach is not accommodating to gaining an understanding of how these changes come about.

The economic dynamics literature takes a different route. Here, the objective is to build agent-based models that make it possible to study the joint dynamics of expectations and aggregate outcomes analytically (for an overview, see e.g. Hommes (2006) and Hommes and Wagener (2009)). To achieve this analytical tractability, the realistic but complex expectation formation process is stripped to its bare essentials. The models "may be viewed as simple, stylized versions of the more complicated 'artificial markets' and computationally oriented agent-based simulation models" (Hommes, 2006). They aim to accommodate the key distinctive features of realistic expectations like econometric models aim to accommodate the key stylized facts of empirical data. To ensure that the model has every chance of providing an accurate account of the mechanisms underlying empirical data, one may impose the condition that the expectations held by agents in the model are consistent with the price data generated by the model (i.e. that obvious tests for misspecification will not reject the models), see e.g. Hommes and Sorger (1998), Hommes and Rosser (2001) and Sogner and Mitlohner (2002). "Economists once thought that behavior was either rational or impossible to formalize. We now know that models of bounded rationality are both possible and also much more accurate descriptions of behavior than purely rational models" (Barberis and Thaler, 2002).

1.4.2 Building econometric models

Here the objective is to best fit the data with a model that is as simple as possible, and with parameters that can be identified from the data. The econometric model cares not about the underlying economic or behavioral mechanisms, as long as the model is capable of describing the observed data. This means that the model and its parameters will typically not have an economic or behavioral interpretation attached. Stylized facts often function as the blue-print on which econometric models are build. If a new stylized fact is discovered, and deemed important, the model will be extended in a way that requires minimum changes yet provides a good fit.

For multivariare volatility models the number of parameters can easily rise very rapidly. The 'holy grail' is to find the model that is general, does not overfit the data, and yet is feasible in terms of estimation.

1.5 Overview of chapters

The thesis consists of five chapters in addition to this introductory chapter. The first part is concerned with fitting empirical data from financial markets, while the second part is concerned with rationalizing the empirical regularities of this data. Chapters 2 and 3 put forward an econometric model for (multivariate) financial volatility. Chapter 4 functions as a bridge between the two parts. It infers market expectations about financial volatility from empirical option price data. The last two chapters build a structural economic model. Chapter 5 develops an analytic framework designed to model the co-evolution of market expectations and prices. Chapter 6 considers some special cases by zooming in on a number of behavioral features.

Chapter 2 puts forward a new multivariate volatility model, which we will refer to as Generalized Orthogonal GARCH (GO-GARCH). It is a member of the ARCH-family, where ARCH stands for autoregressive conditional heteroskedasticity. The ARCH model arguably denotes the most popular specification to date for modelling the time-variation of financial volatility. It was introduced by Engle (1982) and generalized into GARCH by Bollerslev (1986). The multivariate extension, multivariate GARCH, provides a model specification for the time-variation of the covariance matrix (which has time-varying variances on- and covariances off the diagonal). Estimation of multivariate GARCH models can be problematic as the number of unknown parameters involved tends to rise rapidly with the dimension.

The first general multivariate GARCH models put forward in the literature adopt a relatively large number of parameters which leads to convergence difficulties of estimation algorithms (see e.g. Bauwens *et al.*, 2006). New specifications are often determined by means of practical considerations. The challenge is to find a parameterization of the covariance matrix that is feasible in terms of estimation at a minimum loss of generality.

The GO-GARCH model denotes a natural generalization of the O-GARCH model, and is nested as a special case in the more general BEKK model. It accomodates the key stylized facts of multivariate volatility while maintaining feasibility. Both artificial and empirical examples are included to illustrate the new choice of model. For the published version of the chapter, see van der Weide (2002).

In Chapter 3 we propose a new estimation method for the factor loading matrix in GO-GARCH models. The method is based on eigenvectors of suitably defined sample autocorrelation matrices of squares and cross-products of returns. The method is numerically more attractive than likelihood-based estimation. Furthermore, the new method does not require strict assumptions on the volatility models of the factors, and therefore is less sensitive to model misspecification. We provide conditions for consistency of the estimator, and study its efficiency relative to maximum likelihood estimation using Monte Carlo simulations. The method is applied to European sector returns. For the published version see Boswijk and van der Weide (2011).

Embedded in option prices are market expectations regarding future volatility. While the assumption of rational expectations has been a popular paradigm, it is difficult to ignore the subjective nature of expectations. The objective of Chapter 4 is to make market expectations visible as they evolve over time, and to price options in line with prevailing expectations, be they rational or non-rational. We put forward an analytically convenient option pricing framework that accommodates both stochastic volatility and asymmetric volatility. Daily estimates of the implied pdf of volatility are obtained by estimating the option pricing model one day at a time. We do not impose too much structure on how expectations are updated over time, but allow market expectations to take their course. See Peters and van der Weide (2011) for the working paper version.

With Chapter 5 we move to the modelling of the expectation formation process. We propose a new analytic framework for studying the joint time-variation of expectations and prices. Beliefs distributions are defined on a beliefs space representing a continuum of possible strategies agents can choose from. Agents base their choices on past performances and re-evaluate strategies as new information becomes available. By considering individual choices as random variables, which is natural in a random utility framework, heterogeneity in beliefs can be seen to act as a 'natural source of randomness'. Our framework gives rise to a random dynamical system (RDS), the stochastic properties of which are directly related to the time-varying beliefs distribution. We consider some asset pricing examples and discuss several conditions, that involve dependence among agents and unequal market impact, under which the randomness persists even as the number of agents tends to infinity. The chapter has been published as a working paper, see Diks and van der Weide (2003).

Chapter 6 considers a simple example of the framework introduced in Chapter four with the objective to investigate the effects on price dynamics of several behavioral assumptions: (i) herd behaviour; (ii) a-synchronous updating of beliefs; and (iii) heterogeneity in time horizons (memory) among agents. The benchmark model with many traders yields a random walk driven by news. Introducing herding is shown to modify the random walk to an ARIMA$(0, 1, 1)$ process, which is observationally equivalent to a reduction of the number of market participants. In terms of returns the model predicts MA(1) structure with a negative coeffient. Asynchronous updating leads to an MA(1) model for returns with GARCH$(1, 1)$ innovations, and predicts a relation between the ARCH and GARCH coefficients. Heterogeneity in memory leads to long-range dependence in returns. In the empirical section we perform a modest 'reality check' concerning the predicted sign of the MA coefficient and the relation between the ARCH and GARCH coefficients for exchange rate data. For the published version, see Diks and van der Weide (2005).

Chapter 2

GO-GARCH: A new multivariate volatility model

2.1 Introduction

The 'holy grail' in multivariate GARCH modeling is without any doubt a parameterization of the covariance matrix that is feasible in terms of estimation at a minimum loss of generality. The general multivariate GARCH models available parameterize the covariance matrix by a very large number of parameters that are hard to estimate, which often leads to convergence difficulties of estimation algorithms. Therefore, the choice of the multivariate model is often determined by means of practical considerations i.e. the ease of estimation. The strong restrictions are often not believed to reflect the 'truth', but they are imposed to guarantee feasibility.

Some of the best known multi-variate GARCH models available include the VECH model of Bollerslev *et al.* (1988), the constant correlation model of Bollerslev (1990), the factor ARCH model of Engle *et al.* (1990), and the BEKK model studied by Engle and Kroner (1995). For an overview of the multivariate GARCH models, as well as tests for misspecification, see the paper by Kroner and Ng (1998). An extensive survey of empirical applications of time-varying covariance models in finance can be found in Bollerslev *et al.* (1992). In particular the model of Bollerslev (1990) has been a popular choice for modelling high-variate time series. A test for its assumption of a constant correlation is introduced in a recent paper by Tse (2000). Shortly after, both Engle (2002) and Tse and Tsui (2002) generalized the model to allow for time-varying correlations.

A somewhat different approach is the Orthogonal GARCH (O-GARCH) or principal components GARCH method. The principal components approach has first been applied in a GARCH-type context by Ding (1994). Shortly after, Alexander and Chibumba (1996) introduced the strongly related O-GARCH model. Thereafter, O-GARCH has been a popular choice to model the conditional covariances of financial data (see e.g. Klaassen, 1999), mainly because the model remains feasible for large covariances matrices (see e.g. Alexander, 2002). Recently, the model has been elaborated along with applications by Alexander (1998, 2001).

The O-GARCH model implicitly assumes that the observed data can be linearly transformed into a set of uncorrelated components by means of an orthogonal matrix. These unobserved components can be interpreted as a set of uncorrelated factors that drive the particular economy or market, similar to that in the Factor (G)ARCH approach of Engle *et al.* (1990). The orthogonality assumption, however, appears to be very restrictive. Indeed, if a linkage with a set of uncorrelated economic components exists, why should the associated matrix be orthogonal? The O-GARCH model is also known to suffer from identification problems, mainly because estimation of the matrix is entirely based on unconditional information (the sample covariance matrix). For example, when the data exhibits weak correlation, the model has substantial difficulties to identify a matrix that is truly orthogonal (see e.g. Alexander, 2001).

The multivariate GARCH model proposed in this chapter can best be seen as a natural generalization of the O-GARCH model. Clearly, orthogonal matrices are very special, and they only reflect a very small subset of all possible invertible linear maps. The generalized O-GARCH model (GO-GARCH) allows the linkage to be given by any possible invertible matrix. Estimation of the matrix requires the use of conditional information, which in turn solves possible identification problems[1]. The parameters are relatively easy to estimate, so that a substantial increase in the degrees of freedom is obtained at a very affordable price.

The next section will introduce the generalized Orthogonal GARCH model (GO-GARCH). Estimation is discussed in Section 2.3. Sections 2.4 and 2.5 present some simulation results and an empirical example, respectively. Section 2.6 concludes.

[1]For example, the data is not required to exhibit strong dependence for the method to work.

2.2 Generalized Orthogonal GARCH

2.2.1 Notation

In a multivariate GARCH setting, the conditional covariance matrix of the m-dimensional zero mean random variable depends on elements of the information set up to time $t-1$, denoted by $\Im_{t-1}$. Assume that x_t is normally distributed and that its conditional covariance matrix V_t is measurable with respect to $\Im_{t-1}$, the multivariate GARCH model is then described by:

$$x_t \,|\Im_{t-1} \sim N\left(0, V_t\right), \tag{2.1}$$

where we have assumed that x_t is second order stationary so that $V = E(V_t)$ exists. The information set $\Im_t$ contains both lagged values of the squares and cross-products of x_t and elements of the conditional covariance matrices up to time t, i.e. lagged values of V_t. The challenge in multivariate GARCH modeling is to find a parameterization of V_t as a function of $\Im_{t-1}$ that is fairly general while feasible in terms of estimation.

In the following we will frequently use the terms conditional information and unconditional information. We specify unconditional information as information that can be extracted from the unconditional covariance matrix. By conditional information we mean the information set $\Im_t$ as introduced above.

2.2.2 Representation

The key assumption of the GO-GARCH model is the following:

Assumption 2.1 *The observed economic process $\{x_t\}$ is governed by a linear combination of uncorrelated economic components*[2] *$\{y_t\}$:*

$$x_t = Zy_t. \tag{2.2}$$

[2]Note that there might be more components than the number of variables observed, so that exposing a set of reliable components could be troublesome. However, as the components are assumed to be described by independent GARCH-type models, a new set of uncorrelated components can be constructed by aggregating the 'original' components. Under certain conditions, the (extracted) aggregated components are also described by GARCH-type processes, see for example Drost and Nijman (1993) in which temporal aggregation of GARCH processes is considered. However, it is known that the GARCH-type 'features' typically become weaker under aggregation. As a consequence, the accuracy with which the components are described by GARCH-type models increases as more components can be extracted, which will result in better fits.

The linear map Z that links the unobserved components with the observed variables is assumed to be constant over time, and invertible.

Without loss of generality[3], we normalize the unobserved components to have unit variance, so that:

$$V = Ex_t x_t^T = ZZ^T. \tag{2.3}$$

An explicit example, which we will denote the GO-GARCH(1,1) model, would be:

$$x_t = Zy_t \quad y_t \sim N(0, H_t), \tag{2.4}$$

where each component is described by a GARCH(1,1) process:

$$H_t = diag(h_{1,t}, ..., h_{m,t}) \tag{2.5}$$
$$h_{i,t} = (1 - \alpha_i - \beta_i) + \alpha_i y_{i,t-1}^2 + \beta_i h_{i,t-1} \quad i = 1, ..., m, \tag{2.6}$$

where $H_0 = I$ equals the unconditional covariance matrix of the components[4]. The conditional covariances of $\{x_t\}$ are given by:

$$V_t = ZH_tZ^T. \tag{2.7}$$

2.2.3 Identification

Let P and Λ denote the matrices with, respectively, the orthonormal eigenvectors and the eigenvalues of the unconditional covariance matrix $V = ZZ^T$.

Let us assume that an orthogonal linear linkage Z indeed exists, so that $x_t = Zy_t$. The unconditional covariance matrix V is then given by: $V = ZHZ^T$, where H is diagonal. Then the orthogonal matrix P, the O-GARCH estimator for Z, is only guaranteed to coincide with Z, when the diagonal elements of H are all distinct. Identification problems thus arise when some of the uncorrelated components have similar unconditional variance.

[3]Note that the unconditional variances of the components and the matrix Z are directly related. Let $\{y_t\}$ denote the components with original scaling, and let the normalized set of components be denoted by $\{\tilde{y}_t\}$, so that $\{\tilde{y}_t\} = \{Dy_t\}$, where D represents the diagonal normalization matrix. The observed process is then given by $\{x_t\} = \{Zy_t\} = \left\{\tilde{Z}\tilde{y}_t\right\}$, where $\tilde{Z} = ZD^{-1}$.

[4]Ling and McAleer (2003) provide a method for treating the initial value when it comes to asymptotic theory for multivariate GARCH.

To see this, suppose that all components have unit variance, so that $V = ZIZ^T = I$. Clearly, the matrix Z is no longer identified by the eigenvector matrix of V, as for every orthogonal matrix Q, we have $(ZQ)(ZQ)^T = I$. Note that the eigenvalues of V reflect the variances of the components when the model is well identified. The estimations should therefore be interpreted with caution when some of the eigenvalues are almost identical. Problems of this type are known to occur when, for example, the data exhibits weak dependence[5]. The next lemma states that the linkage Z is well identified when conditional information is taken into account.

Lemma 2.1 *Let Z be the map that links the uncorrelated components $\{y_t\}$ with the observed process $\{x_t\}$. Then there exists an orthogonal matrix U_0 such that:*

$$P\Lambda^{\frac{1}{2}}U_0 = Z. \tag{2.8}$$

Proof. The result follows directly from Singular Value Decomposition, see e.g. Horn and Johnson (1999). ■

Let the estimator for U_0 be denoted by U. Without loss of generality, we restrict the determinant of U to be 1[6].

It can be verified that the orthogonal matrices P and Λ have $\frac{m(m-1)}{2}$ and m degrees of freedom, respectively. Together with the $\frac{m(m-1)}{2}$ degrees of freedom for U, we have $m + m\,(m-1) = m^2$ degrees of freedom for the invertible matrix Z. The matrices P and Λ will be estimated by means of unconditional information, as they will be extracted from the sample covariance matrix V. Conditional information is required to estimate U_0.

Note that there is a continuum of matrices Q for which a set of linearly independent components $u_t = Qx_t$ can be obtained. For every choice of orthogonal matrix U, the linear transformation $Q = U^T\Lambda^{-\frac{1}{2}}P^T$ induces an uncorrelated series with unit variance: $Eu_tu_t^T = QVQ^T = U^TU = I$. Clearly, these components often still exhibit a form of nonlinear correlation. Therefore, linear independence can be very deceiving, as it might give the impression that the linkage between the observed variables and the uncorrelated components is uncovered, when more often it is not. The original components can only[7] be restored by means of the inverse of Z.

[5] Given that the observed data is normalized to have unit variance, which is common practice.

[6] More precisely, U is considered an element of $SO(m)$, which denotes the set of all m-dimensional orthogonal matrices with positive determinant.

[7] Equivalent matrices, in the sense that they only exchange variables for example, are included.

According to Lemma 2.1, the model is well identified as there exists a U_0 that is associated with the original Z. Indeed, the additional $\frac{m(m-1)}{2}$ degrees of freedom induced by the extra term U extends the representation to full generality, in the sense that any invertible linkage Z can in principle be estimated from the data, instead of orthogonal matrices only.

One way to parameterize the estimator for the orthogonal matrix U_0 would be by means of rotation matrices:

Lemma 2.2 *Every m-dimensional orthogonal matrix U with $\det(U) = 1$ can be represented as a product of $\binom{m}{2} = \frac{m(m-1)}{2}$ rotation matrices:*

$$U = \prod_{i<j} R_{ij}(\theta_{ij}) \quad -\pi \leq \theta_{ij} \leq \pi, \tag{2.9}$$

where $R_{ij}(\theta_{ij})$ performs a rotation in the plane spanned by e_i and e_j over an angle θ_{ij}.

Proof. See Vilenkin (1968). ■

The rotation angles[8] $\{\theta_{ij}\}$ are commonly referred to as the Euler angles, which can be estimated by means of maximum likelihood.

We have noted earlier already that the O-GARCH model suffers from identification problems, for example when the data exhibits weak dependence. These problems should not arise when conditional information is exploited, as proposed in the GO-GARCH model. For example, when the independent components appear to be observed directly, we expect the estimator for U_0 to be close to P^T, since $\hat{Z} = P\Lambda^{\frac{1}{2}}P^T = V^{\frac{1}{2}}$ is approximately diagonal when the data is virtually independent.

2.2.4 Time-varying correlations

The implied conditional correlations $\{R_t\}$ of the observed process $\{x_t\}$ can be computed as:

$$R_t = D_t^{-1} V_t D_t^{-1}, \quad D_t = (V_t \circ I)^{\frac{1}{2}}, \tag{2.10}$$

where $\{V_t\} = \{Z H_t Z^T\}$ denotes the conditional covariances of $\{x_t\} = \{Z y_t\}$, and where $\circ$ denotes the Hadamard product.

[8] Note that the values for the angles will depend on the ordering of the rotation matrices. The ordering should not affect the estimation results.

This theoretical example illustrates how possible lower and upper bounds for the correlation depend on the type of linear map Z_θ. Let Z_θ be the following two dimensional map:

$$Z_\theta = \begin{pmatrix} 1 & 0 \\ \cos\theta & \sin\theta \end{pmatrix}, \tag{2.11}$$

where θ measures the extent to which the uncorrelated components are mapped in the same direction. For $\theta = 0$ the map is not invertible yielding perfect correlation between the observed variables, whereas for $\theta = \frac{1}{2}\pi$ we have the identity map, so that the observed variables are completely uncorrelated. Let the conditional variances of the uncorrelated components be denoted by (h_{1t}, h_{2t}). It can be verified that the conditional correlation between the observed variables, denoted by ρ_t, is given by:

$$\rho_t = \frac{h_{1t}\cos\theta}{\sqrt{h_{1t}}\sqrt{h_{1t}\cos^2\theta + h_{2t}\sin^2\theta}}. \tag{2.12}$$

If we assume that $h_{it} > 0$, we can define $z_t = \frac{h_{2t}}{h_{1t}}$, so that ρ_t can be expressed as:

$$\rho_t = \frac{1}{\sqrt{1 + z_t\tan^2\theta}}. \tag{2.13}$$

For finite samples, the variable z_t will have finite lower and upper bounds. As a consequence, the conditional correlation ρ_t is also bounded.

Note that a constant linkage Z_θ gives rise to time-varying correlations between the observed variables. These correlations rise on average when the components are mapped more in the same direction. We can not exclude the possibility that the 'economic mechanism' Z evolves over time. If so, endogenizing Z and make it time-varying, might improve the fit of the time-varying correlations. Extending the GO-GARCH model to allow for a non-constant Z, however, is left for further research. A first step would be to test for a constant linkage, for example by means of test on structural change such as the Chow test.

2.3 Estimation

The parameters that need to be estimated by means of conditional information, include the vector θ of rotation coefficients that will identify the invertible matrix Z (see lemma 3), and the parameters (α, β) for the m uni-variate GARCH(1,1) specifications. The log likelihood $L_{\theta,\alpha,\beta}$ for the GO-GARCH model can be represented as:

$$L_{\theta,\alpha,\beta} = -\frac{1}{2}\sum_t m\log(2\pi) + \log|V_t| + x_t^T V_t^{-1} x_t \tag{2.14}$$

$$= -\frac{1}{2}\sum_t m\log(2\pi) + \log\left|Z_\theta H_t Z_\theta^T\right| + y_t^T Z_\theta^T (Z_\theta H_t Z_\theta^T)^{-1} Z_\theta y_t \tag{2.15}$$

$$= -\frac{1}{2}\sum_t m\log(2\pi) + \log\left|Z_\theta Z_\theta^T\right| + \log|H_t| + y_t^T H_t^{-1} y_t, \tag{2.16}$$

where $Z_\theta Z_\theta^T = P\Lambda P^T$ is independent of θ. For the initial value of H_t we take the identity matrix, which equals the implied unconditional covariance of $\{y_t\}$. Even in high-variate cases, when the covariance matrices are very large, it should not be a problem to maximize the log likelihood over the $\frac{m(m-1)}{2}+2m$ parameters. Note that in order to avoid convergence difficulties of estimation algorithms, we propose a kind two-step estimation. We exploit unconditional information first, so that the number of parameters for Z that are estimated through maximum likelihood is $\frac{m(m-1)}{2}$ instead of m^2 (see lemma 2).

Conditions for strong consistency of the maximum likelihood estimator for general multivariate GARCH are derived by Jeantheau (1998). These conditions are verified by Comte and Lieberman (2003) for the general BEKK model, in which a result of Boussama (1998), concerning the existence of a stationary and ergodic solution to the multivariate GARCH(p,q) process, is used.

It can be verified that the more general BEKK model has the GO-GARCH model nested as a special case. Strong consistency of the quasi MLE for GO-GARCH can therefore be established by appealing to Jeantheau's conditions, following Comte and Lieberman. To keep it simple, we focus on the GO-GARCH(1,1) model as in (2.4), but it can be verified that the results also hold for the more general GO-GARCH(p,q) model. To apply the results of Jeantheau (1998), we assume that starting value of the process is drawn from its stationary distribution P_{θ_0}, although Comte and Lieberman (2003) indicate that consistency holds for an arbitrary starting value. We refer to Ling and McAleer (2003) for a more extensive discussion of the treatment of the initial value and its implication for asymptotic properties.

Proposition 2.1 *Consider the GO-GARCH(1,1) model, where α_i and β_i correspond to the GARCH(1,1) parameters of the independent components. Assume that B0-B2 holds,* *and that the components are stationary, i.e.*

$$\alpha_i + \beta_i < 1 \quad \textit{for i=1,...,m.} \tag{2.17}$$

Then the MLE is consistent.

Proof. The result follows directly from the derivation of Comte and Lieberman (2003). ■

How to conduct inference is beyond the scope of this chapter. However, as Comte and Lieberman (2003) have proven asymptotic normality of the quasi-MLE for the BEKK model, having GO-GARCH nested as a special case, we conjecture that this property is also inherited by GO-GARCH. Some caution will be in place though, since we proposed a kind of two-step estimation which will affect the distribution of the estimator. For example, the standard errors might be underestimated by the Fisher Information matrix. We leave a precise study of the asymptotic distribution for further work. For tests on possible misspecification of the multivariate GARCH model see Kroner and Ng (1998), and the more recent paper by Tse (2002).

2.4 Simulation results

This section aims to illustrate the behavior of the GO-GARCH model by experimenting with artificial data.

2.4.1 Orthogonal linkage

We constructed the independent components by generating from 4 univariate GARCH(1,1) models to build a 4-variate time series. The conditional variance of each component is described by:

$$h_{i,t} = c_i + \alpha_i y_{i,t-1}^2 + \beta_i h_{i,t-1} \quad i = 1, ...4. \tag{2.18}$$

The values that are assigned to the parameters (c, α, β) are summarized in Table 2.1.

The parameters are chosen so that variances are nearly integrated, which is commonly observed in financial data. Also note that the parameters are chosen in such a way that

component	c	α	β
1	0.08	0.10	0.88
2	0.03	0.08	0.90
3	0.05	0.15	0.80
4	0.10	0.20	0.70

Table 2.1: *GARCH parameters*

some of the unconditional variances are identical. The length of the artificial data set is 3000 observations, which is equivalent to approximately 12 years of daily data.

The first orthogonal matrix considered is the identity matrix. As it preserves independence, the components will be observed directly. In the second part, we simulate with an orthogonal matrix that induces dependence among the observed variables.

Independent multivariate GARCH

In this part, we test whether the models are able to detect the independent nature of the observed data. It is known that the O-GARCH model can not deal properly with virtually independent data. In contrast, GO-GARCH should be able to estimate a linear representation that induces weak dependence or even independence. The results are presented in Table 2.2.

As expected, O-GARCH was not able to detect the independence of the process. The estimated matrix is far from being diagonal, so that conditional dependence is 'brought into' the residuals. The substantial errors in the GARCH parameters estimates also indicate that O-GARCH did not extract the independent components, but some dependent variables instead. The GO-GARCH model, however, performs very well in this example. The estimated linkage correctly reflects the independent nature of the data. Also the GARCH parameters are estimated properly[9].

Dependent multivariate GARCH

The independent components are described by exactly the same process as in the first part. The key difference is that in this example they are not observed directly. The observed process will be an orthogonal representation of the components that exhibits strong dependence. In principle, the O-GARCH model could also perform well in this example, as the observed variables are no longer independent, while the associated matrix is orthog-

[9]Note that some components might have been switched.

onal. However, note that some of the components have a similar scaling (unconditional variance). As a consequence, O-GARCH might still suffer from identification problems.

	O-GARCH				GO-GARCH			
$\widehat{Z^{-1}}$	0.39	−0.25	−0.64	−0.58	1.00	0.02	0.01	0.01
	−0.27	0.83	−0.06	−0.47	−0.01	1.00	0.01	0.01
	−0.88	−0.36	−0.32	−0.08	0.00	0.02	−1.00	0.00
	−0.06	−0.34	0.70	−0.66	0.00	0.00	0.04	−1.00
c	0.09	0.04	0.03	0.09	0.03	0.02	0.05	0.10
α	0.09	0.06	0.08	0.10	0.11	0.06	0.15	0.20
β	0.81	0.90	0.89	0.80	0.86	0.92	0.80	0.70

Table 2.2: *Estimates for the linkage and GARCH parameters*

The orthogonal matrix, denoted by Z, is constructed as a product of four rotation matrices, and is shown in Table 2.3. Table 2.4 summarizes the results.

map	Z
matrix	$\begin{pmatrix} \frac{\sqrt{3}}{2\sqrt{2}} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2\sqrt{2}} \\ \frac{\sqrt{3}}{2\sqrt{2}} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2\sqrt{2}} \\ -\frac{\sqrt{3}}{4} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{3}{4} \\ \frac{1}{4} & \frac{\sqrt{3}}{2\sqrt{2}} & \frac{\sqrt{3}}{2\sqrt{2}} & \frac{\sqrt{3}}{4} \end{pmatrix}$

Table 2.3: *Orthogonal linkage*

	O-GARCH				GO-GARCH			
c	0.05	0.08	0.05	0.07	0.10	0.02	0.05	0.03
α	0.06	0.07	0.07	0.06	0.20	0.06	0.15	0.11
β	0.89	0.84	0.89	0.87	0.70	0.92	0.80	0.86

Table 2.4: *Estimates for the GARCH parameters*

In the case of O-GARCH, the estimates for the GARCH parameters are clearly different from the true parameters suggesting that the model was not able to identify the independent components. In the previous subsection we have seen that the trivial orthogonal matrix, namely identity, could also not be identified by O-GARCH. Thus even when the linkage is truly orthogonal, there is no guarantee that O-GARCH is able to identify it. The model additionally requires that all the components have a different scaling, which might often not be the case.

When we look at the estimates of the GO-GARCH model, we find that the GARCH parameters of the components are estimated with reasonable accuracy[10]. From this we conclude that the linkage estimated by GO-GARCH can not be far from the 'truth' as we build it.

2.4.2 Non-orthogonal invertible linkage

In this subsection, non-orthogonal invertible matrices are chosen to link the independent components with the observed process. This will be an important example, as we generalized the O-GARCH model to be able to expose linkages that are not orthogonal. It follows that lower bounds for the conditional correlations can be observed when the linkage matrices approach singularity.

component	c	α	β
1	0.05	0.15	0.80
2	0.05	0.25	0.70

Table 2.5: *The GARCH parameters*

map	Z_1	Z_2	Z_3	Z_4
matrix	$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} \frac{1}{2} & 1 \\ 0 & 2 \end{pmatrix}$	$\begin{pmatrix} 2 & -2 \\ 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$

Table 2.6: *The invertible linkages*

map	Z_1	Z_2	Z_3	Z_4
V_i	$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$	$\begin{pmatrix} \frac{5}{4} & 2 \\ 2 & 4 \end{pmatrix}$	$\begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}$	$\begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$

Table 2.7: *The unconditional covariances*

map	Z_1	Z_2	Z_3	Z_4
W_i	$\begin{pmatrix} 1.41 & -1 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 2.24 & -2 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.47 & 0.75 \\ -0.94 & 0.75 \end{pmatrix}$	$\begin{pmatrix} -0.75 & 1.49 \\ 1.49 & -0.75 \end{pmatrix}$

Table 2.8: *The true linear representations*

In order to have a more controlled experiment, we confine ourselves to the 2-dimensional case. Similar to the first examples, we construct two independent components in order

[10]Note that some components have been switched.

to build up a bivariate time series. The conditional variances of both components are specified by means of the same GARCH model, as in (2.18). Also the sample size is chosen to be identical, namely 3000 observations. Table 2.5 lists the values at which the GARCH parameters were set to simulate the independent components. It can easily be verified that both components have unit unconditional variance, so that their unconditional covariance matrix equals the identity matrix.

We will consider four different invertible linear maps for the linkage. The associated matrices, denoted by Z_1 till Z_4, are shown in Table 2.6. The unconditional covariance matrix of the observed process is simply given by: $V_i = Z_i Z_i^T$. The covariances V_1 till V_4 are listed in Table 2.7.

The observed data is commonly normalized to have unit variance by a diagonal matrix D, so that the covariance matrices of the normalized series $\tilde{V}_i = D_i Z_i Z_i^T D_i^T$ has 1's along the diagonal. In our example, the diagonal elements of D_1till D_4 are easily seen to be $\left\{\frac{1}{\sqrt{2}}, 1\right\}$, $\left\{\frac{2}{\sqrt{5}}, \frac{1}{2}\right\}$, $\left\{\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{5}}\right\}$, and $\left\{\frac{1}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right\}$, respectively. It follows that the true matrix that links the normalized observed variables with its independent components, is given by $(D_i Z_i)^{-1}$. These matrices, denoted by W_i, are shown in Table 2.8.

Note that in all cases the orthonormal eigenvectors of the unconditional covariance matrix are given by $P = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$, so that O-GARCH is expected to estimate scaled versions of P for the linkages.

O-GARCH

map	Z_1		Z_2		Z_3		Z_4	
$\hat{W}_i$	0.54	0.54	0.51	0.51	−0.61	−0.61	0.53	0.53
	1.31	−1.31	2.17	−2.17	0.86	−0.86	1.59	−1.59
$\hat{c}_i$	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.06
$\hat{\alpha}_i$	0.18	0.12	0.21	0.13	0.14	0.23	0.11	0.12
$\hat{\beta}_i$	0.76	0.83	0.73	0.82	0.82	0.72	0.84	0.81

Table 2.9: *The estimates for the linkages and the GARCH parameters*

GO-GARCH

map	Z_1		Z_2		Z_3		Z_4	
$\hat{W}_i$	0.01	0.99	0.02	0.98	−0.48	−0.73	1.49	−0.74
	1.41	−1.01	2.23	−2.00	0.94	−0.76	0.77	−1.51
$\hat{c}_i$	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
$\hat{\alpha}_i$	0.25	0.14	0.25	0.14	0.14	0.25	0.25	0.14
$\hat{\beta}_i$	0.71	0.81	0.71	0.81	0.81	0.71	0.71	0.81

Table 2.10: *The estimates for the linkages and the GARCH parameters*

The results for both the O-GARCH and the GO-GARCH model are presented in Table 2.9 and 2.10, respectively. This example illustrates that the GO-GARCH model is able to deal with decompositions that are not of the orthogonal type. The estimated linkages are in all cases very close[11] to the 'truth', the matrices from Table 2.8. The estimates for the GARCH parameters are also accurate.

A priori we know that the O-GARCH model can not uncover the non-orthogonal linkages, as it restricts the matrix to be orthogonal. As a consequence, it extracts components that are not independent, which is reflected by the biased estimates for the GARCH parameters. Particularly in example 4, the O-GARCH estimates for the GARCH parameters show substantial error. The difference between the estimated correlations is therefore most notable in example 4, which can be seen in Figure 2.1.

In example 4, the GO-GARCH estimates for the correlations never fall below 0.8, say, whereas the correlations estimated by O-GARCH show much stronger declines and sometimes even drop till below 0.4. The reason for this effect is that the matrix from example 4 shows the strongest 'deviation' from an orthogonal matrix. The linkage from example 4 maps both independent components in almost the same direction which induces a strong correlation between the observed variables. Exactly the same feature is observed in the empirical example described in the next section. Indeed, it seems reasonable that observed variables that are strongly related exhibit high correlation at all times. As the linkage with the components that induces the high correlation is assumed constant over time, it will be surprising to observe periods in time where the variables suddenly appear almost uncorrelated. This feature is illustrated by a theoretical bivariate example, where the lower bound and upper bounds of the correlation are derived as a function of a characteristic parameter θ of the linkage Z_θ.

2.5 Empirical example

We include an example from real life, as an attempt to gain insight in the relation between observed economic and financial variables and the uncorrelated factors that are assumed to drive the market. Our example considers the Dow Jones Industrial Average (DJIA) versus the NASDAQ composite. The sample contains more than ten years of daily observations, starting at the first of January in 1990 and ending in October 2001. First, we estimate a

[11] Neglect signs, as they do not yield a different representation. Also note that some components might have been switched.

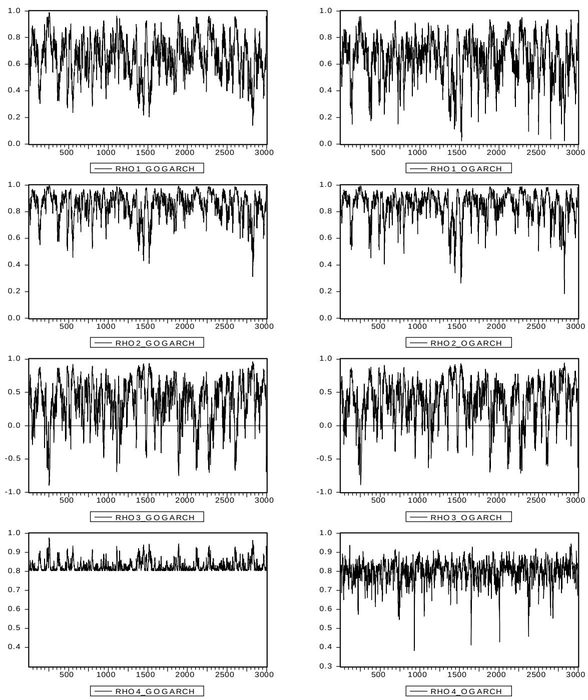

Figure 2.1: *The estimated time evolutions of the (conditional) correlations for all four examples. The figures on the left correspond with the GO-GARCH model, and the figures on the right with O-GARCH.*

first order[12] vector autoregressive (VAR) model to account for the linear structure present in the data. Subsequently, we use the residuals to estimate the conditional covariances from which the (conditional) correlations between the DJIA and the NASDAQ can be computed. Questions that arise naturally include: (i) Are non-orthogonal linkages common in real life examples? (ii) Will allowing for a more general linkage (all invertible matrices) typically induce a better description of the time-varying correlations between economic and financial variables?

We estimate both the O-GARCH and GO-GARCH model, and compare their results. The estimates are summarized in Table 2.11.

To address the question whether non-orthogonal linkages can be found in financial data,

[12]Higher order specifications do not significantly contribute to a better linear fit.

we first verify whether the estimated unrestricted matrix is approximately orthogonal. Let $\widehat{Z_{go}^{-1}}$ denote the unrestricted representation, then $(\widehat{Z_{go}^{-1}}(\widehat{Z_{go}^{-1}})^T = \begin{pmatrix} 1.49 & -1.09 \\ -1.09 & 1.95 \end{pmatrix}$ should be close to the identity matrix, which it is not. This confirms our conjecture that the orthogonality assumption of O-GARCH is probably too restrictive, in that it might exclude many of the linkages 'observed' in financial markets. In order to quantify the impact of the restrictions imposed by O-GARCH, we compute the Likelihood Ratio statistic (LR) to test the O-GARCH specification against GO-GARCH for several lengths of the time series. The results are listed in Table 2.12. For all lengths of the time-series considered, the hypothesis of an orthogonal linkage is rejected at a 1% level.

model	O-GARCH		GO-GARCH	
$\widehat{Z^{-1}}$	−0.55	−0.55	−1.18	0.32
	1.19	−1.19	0.58	−1.27
c	0.009	0.003	0.009	0.003
α	0.070	0.039	0.054	0.079
β	0.922	0.957	0.939	0.915

Table 2.11: *The estimates for the linkages and GARCH parameters*

Length	250	500	1000	3082
LR	23.7	13.8	42.6	731.4

The critical value of χ_1^2 at a 1% level is 6.63.

Table 2.12: *Likelihood-Ratio Test of O-GARCH against GO-GARCH*

Even though the GO-GARCH model provides a better fit when compared with O-GARCH, it might be that our more general model is still seriously misspecified. For this reason, we include a simple test for misspecification. Since we are interested in heteroskedasticity in particular, we estimate a VAR model on the squares and the product of the two standardized residuals to verify whether the conditional covariance has been modelled correctly. The results, which are shown in Table 2.13, suggest that GO-GARCH is not seriously misspecified. The remaining structure found in the standardized GO-GARCH residuals is fairly weak. In contrast, the residuals from the O-GARCH model still seem to exhibit significant persistence in volatility.

To examine to what extent the restrictions on the linkage affect the (conditional) correlations, we compare the implied correlations of both models. The time evolution of the correlations is shown in Figure 2.2, which reveals several interesting features. Perhaps

model	O-GARCH			GO-GARCH		
variable	$\varepsilon_{1,t}^2$	$\varepsilon_{2,t}^2$	$\varepsilon_{1,t}\varepsilon_{2,t}$	$\varepsilon_{1,t}^2$	$\varepsilon_{2,t}^2$	$\varepsilon_{1,t}\varepsilon_{2,t}$
c	4.11	4.29	4.22	3.55	4.39	3.89
$\varepsilon_{1,t-1}^2$	1.40^{**}	0.08	-0.27^{**}	1.09^{*}	0.52	0.70
$\varepsilon_{2,t-1}^2$	1.06^{**}	0.36	-0.09	0.92	0.90	0.81
$\varepsilon_{1,t}\varepsilon_{2,t}$	-2.44^{**}	-0.44	0.36	-1.99^{*}	-1.46	-1.52
$adj.R^2$	0.14	0.00	0.01	0.00	0.00	0.00

*,** significant at the 5% and 1% level.

Table 2.13: *Misspecification test: VAR model on the squares and the product of the two residuals*

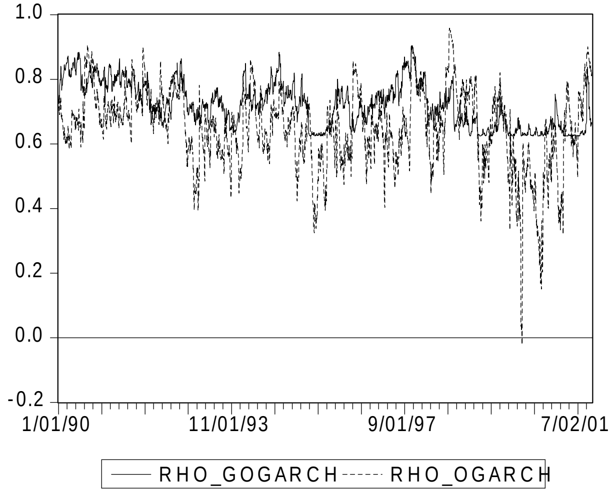

Figure 2.2: *The estimated correlations between the DJIA and the NASDAQ*

most striking is that the correlations estimated by GO-GARCH are much less volatile. Furthermore, the GO-GARCH correlations never seem to fall below 0.6, say, whereas the more volatile correlations estimated by O-GARCH show several substantial drops. In the beginning of 2000, the correlation according to O-GARCH is even just below zero where the GO-GARCH correlations show no decline at all. Since the matrix estimated by GO-GARCH is explicitly not orthogonal, we have reason to believe that O-GARCH often underestimates the correlations. Indeed, it seems plausible that the DJIA and the NASDAQ, which are strongly related, exhibit high correlation at all times. Assuming that the linkage does not change over time, it will be surprising to observe periods where the DJIA and the NASDAQ appear to be almost unrelated. The differences observed for the year 2000, however, are kind of extreme. This might suggest that the linkage is not constant over time, or stronger, that it was subject to a structural change. Indeed, at the beginning of 2000 we experienced a technological boom which could explain our findings. To test the assumption of a constant linkage will be left for further research.

2.6 Concluding remarks

A new type of multivariate GARCH model is proposed that can best be seen as a generalization of the O-GARCH model. It supports the assumption that the observed variables are driven by some unobserved uncorrelated components, linked by a linear map. In order to identify these components, we only need invertibility of the associated matrix. Under the null of O-GARCH, however, the matrix is assumed orthogonal which only covers a very small subset of all possible invertible matrices. Moreover, even when the matrix is truly orthogonal, the estimator proposed by O-GARCH is not always able to identify it. The GO-GARCH model considers every invertible matrix as a possible linkage, which will be parameterized in such a way that it is not expected to complicate estimation while excluding any identification problems.

The model is tested on both artificial and financial data. The simulation results show that the model correctly estimates both orthogonal and non-orthogonal invertible linkages. The results are not affected by the scaling of the uncorrelated factors or a possible weak dependence among the observed variables. The latter is known to be responsible for the identification problems of O-GARCH, which is confirmed by our experimental results. The nature of the linkage, for example whether it is orthogonal or not, is strongly related with the implied correlations between the observed variables. This relation is made explicit by a theoretical example, and illustrated by some of the simulation results. The effect of the linkage on the correlations is also observed in the empirical example, the Dow Jones Industrial Average versus the NASDAQ. A likelihood ratio test rejects the hypothesis that the associated matrix is orthogonal. In addition a simple test for misspecification suggests that the GO-GARCH model provides a better description of the process, and the conditional correlations in particular. We argue that by restricting the matrix to be orthogonal, O-GARCH will often underestimate the correlations. The differences are kind of extreme during the year 2000, which coincides with the technology boom that initiated early that year. It could be that the linkage is not constant over time and that it experienced a structural change during the technological bust in 2000. A test for such a structural break, and perhaps even extending the model to allow for a time-varying linkage, is left for further research. Probably the most important question that remains, is to what extent is GO-GARCH able to improve the modelling of very large covariance matrices. Indeed, it would be very interesting to compare the model with other recently developed multivariate GARCH models, such as the Dynamic Conditional Correlation model of Engle (2002). The misspecification tests recently proposed by Tse (2002) can be used to measure performance.

Chapter 3

Method of moments estimation of GO-GARCH models

3.1 Introduction

The GO-GARCH model was proposed by van der Weide (2002), as a generalization of the orthogonal GARCH model of Ding (1994) and Alexander (2001). The starting point of the model is that an observed vector of returns can be expressed as a non-singular linear transformation of latent factors (either independent or conditionally uncorrelated) that have a GARCH-type conditional variance specification.

A restricted version of the model where only a subset of the latent factors has a time-varying conditional variance has been analyzed recently by Lanne and Saikkonen (2007). This shows that a parsimonious version of the factor GARCH model by e.g. Diebold and Nerlove (1989) and Engle *et al.* (1990) is nested as a special case (where the variance matrix of the idiosyncratic error term will not be of full rank). The closely related model proposed by Vrontos *et al.* (2003) is also nested as a special case by imposing structure on the linear transformation. Recently, Fan *et al.* (2008) studied a general version of the model by relaxing the assumption of independent factors to conditionally uncorrelated factors.

Note that one has considerable flexibility in specifying models for the factors. One could in principle also consider stochastic volatility as an alternative to the GARCH-type models (see e.g. Doz and Renault, 2006). For surveys on multivariate volatility models we refer to e.g. Bauwens *et al.* (2006), and Silvennoinen and Teräsvirta (2009). For an overview on common features, see Urga (2007), and for a glossary to volatility models, see Bollerslev (2008).

The model is designed to balance generality against ease of estimation. Time-varying variances, time-varying correlations and (asymmetric) volatility spillovers are accommodated, which denote the key stylized facts of multivariate financial data. Moreover, the model is closed under linear transformations, i.e., a vector of portfolio returns satisfies the same model, albeit with other parameter values, as the original vector of asset returns. The model is also closed under temporal aggregation, see Hafner (2008).

Van der Weide (2002) proposed a two-step estimation method that requires joint maximum likelihood (ML) estimation of parameters that feature both in the linear transformation (between factors and observed data) and in the univariate GARCH specifications for the individual factors. Not all parameters of the linear transformation need to be estimated by ML, more than half can be identified from the spectral decomposition of the unconditional variance matrix. While the method works well, and numerical optimization of the likelihood function often converges without difficulties for dimensions up to fifteen, maximum likelihood can become problematic when the dimension is particularly large and/or when the model used to specify the likelihood function is considerably misspecified.

This chapter puts forward a three-step estimation method that is easy to implement and is numerically attractive. The first two steps define a method-of-moments (MM) estimator for the linear transformation that does not require any Newton-type optimization of an objective function, but instead only involves iterated matrix rotations, so that the method is free of numerical convergence problems regardless of the dimension. We identify the linear transformation by using the fact that the latent factors are heteroskedastic. All that is assumed is that the factors exhibit persistence in variance and have finite fourth moments. Note that the idea of identification through heteroskedasticity in simultaneous equation models is not new; see e.g. Sentana and Fiorentini (2001) and Rigobon (2003). The third and final step involves estimation of the univariate GARCH-type models for each of the factors.

An obvious application of multivariate GARCH models involves forecasting the conditional variance matrix for the purpose of optimal portfolio selection, hedging and risk management, and option pricing. Naturally, the models may also be used to examine patterns in conditional correlations and volatilities over time. Does volatility in one market spill over to other markets? Are correlations increasing or decreasing, i.e., are markets moving closer together over time, and do correlations jump up in periods of extreme volatility (such as in a financial crisis)? In a modest empirical example in this chapter we examine the correlation over time between returns on European sector indices, and find that the degree of comovement between volatility and correlation depends on the general state of the economy.

The outline of the chapter is as follows. In Section 3.2, we formulate the GO-GARCH model, and discuss currently available estimation methods. Section 3.3 introduces our method-of-moments estimator, and discusses how information from autocorrelation matrices at different lags may be efficiently combined. In Section 3.4 we use Monte Carlo simulations to study the efficiency of our estimator relative to (quasi-) maximum likelihood in low-dimensional systems. Section 3.5 contains an empirical application to a vector of ten European industry returns. Section 3.6 concludes.

3.2 The GO-GARCH model

3.2.1 Model and assumptions

Consider an m-vector time series $\{x_t\}_{t\geq 1}$, representing a vector of (daily) returns on m different assets. Letting $\{\mathcal{F}_t\}_{t\geq 0}$ denote the filtration generated by $\{x_t\}_{t\geq 1}$, we assume that any possibly non-zero conditional mean has been subtracted from x_t, so that $E(x_t|\mathcal{F}_{t-1}) = 0$. The GO-GARCH model imposes a structure on the conditional variance matrix $\Sigma_t = \text{var}(x_t|\mathcal{F}_{t-1}) = E(x_t x_t'|\mathcal{F}_{t-1})$, implied by:

Assumption 3.1 *The process $\{x_t\}_{t\geq 1}$ satisfies the representation*

$$x_t = Z y_t = Z H_t^{1/2} \varepsilon_t, \tag{3.1}$$

$$H_t = \text{diag}(h_{1t}, \ldots, h_{mt}), \tag{3.2}$$

where Z is an $m \times m$ non-singular matrix, where $\{\{h_{it}\}_{t\geq 1}, i = 1, \ldots, m\}$ are positive, $\{\mathcal{F}_{t-1}\}$-adapted processes with $E(h_{it}) = 1$, and where $\{\varepsilon_t\}_{t\geq 1}$ is a vector martingale difference sequence, with $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ and $\text{var}(\varepsilon_t|\mathcal{F}_{t-1}) = I_m$.

The model implies that the observed vector of returns x_t can be written as a non-singular transformation of a latent vector process y_t (of the same dimension m), the components

y_{it} of which satisfy

$$E(y_{it}|\mathcal{F}_{t-1}) = 0, \quad \mathrm{var}(y_{it}|\mathcal{F}_{t-1}) = h_{it}, \quad \mathrm{cov}(y_{it}, y_{jt}|\mathcal{F}_{t-1}) = 0, \qquad i \neq j = 1, \ldots, m,$$

i.e., the components of y_t are conditionally uncorrelated. The original formulation of the GO-GARCH model involved the stronger assumption of independence of the components of y_t, but for the methods presented in the present chapter, the conditional uncorrelatedness assumption (proposed by Fan *et al.*, 2008) suffices. The assumptions also imply that y_t is a covariance-stationary process with mean 0 and unconditional variance $E(H_t) = I_m$. This in turn implies that x_t is covariance-stationary with (conditional) mean zero, conditional variance

$$\Sigma_t = \mathrm{var}(x_t|\mathcal{F}_{t-1}) = ZH_tZ', \tag{3.3}$$

and unconditional variance

$$\Sigma = \mathrm{var}(x_t) = ZZ'. \tag{3.4}$$

The conditional variances h_{it} are assumed to follow a GARCH-type structure. One possibility, as considered by van der Weide (2002), is to assume separate univariate GARCH(1,1) specifications

$$h_{it} = (1 - \alpha_i - \beta_i) + \alpha_i y_{i,t-1}^2 + \beta_i h_{i,t-1}, \qquad \alpha_i, \beta_i \geq 0, \quad \alpha_i + \beta_i < 1, \tag{3.5}$$

which, under a suitable starting value assumption on h_{i0}, implies independence of the components y_{it}. Fan *et al.* (2008) propose a more flexible structure, where h_{it} may depend on $y_{j,t-k}, j \neq i, k \geq 1$. A simple extension of (3.5) is their extended GARCH(1,1) specification:

$$h_{it} = \left(1 - \sum_{j=1}^{m} \alpha_{ij} - \beta_i\right) + \sum_{j=1}^{m} \alpha_{ij} y_{j,t-1}^2 + \beta_i h_{i,t-1}, \qquad \alpha_{ij}, \beta_i \geq 0, \quad \sum_{j=1}^{m} \alpha_{ij} + \beta_i < 1. \tag{3.6}$$

Intermediate versions, where some of the $\alpha_{ij}, j \neq i$ are restricted to zero, can also be considered. It should be emphasized that Assumption 3.1, as well as the estimation methods proposed in this chapter, also allow for various other specifications of the conditional variance process, including leverage effects and models formulated in terms of log-volatilities. The assumption also allows for stochastic volatility, as long as the projection h_{it} of the latent stochastic volatility process of y_{it} on the observed price history $\mathcal{F}_{t-1}$ is correlated with lagged squares $y_{i,t-k}^2$.

Consider the polar decomposition of Z:

$$Z = SU, \tag{3.7}$$

where S is a positive definite symmetric matrix, and U is an orthogonal matrix. Using (3.4), it follows that $\Sigma = S^2$, so that S is the symmetric square root of Σ, i.e., $S = PL^{1/2}P'$, where PLP' is the spectral decomposition of Σ. This implies that part of the matrix Z may be identified from the unconditional information $\Sigma = \text{var}(x_t)$, and the problem of estimating Z may be reduced to the problem of identifying the orthogonal matrix U from the conditional information. In other words, defining the (unconditionally) standardized returns

$$s_t = \Sigma^{-1/2} x_t = S^{-1} x_t,$$

we find that s_t follows a GO-GARCH specification $s_t = Uy_t$ with an orthogonal link matrix U.

Note that van der Weide (2002) and Boswijk and van der Weide (2006) consider, instead of (3.7), the singular value decomposition $Z = PL^{1/2}U^*$, where $U^* = P'U$ is another orthogonal matrix. This leads to analyzing the standardized principal components $s_t^* = L^{-1/2}P'x_t$, satisfying $s_t^* = U^* y_t$. Here we follow Lanne and Saikkonen (2007) in using the polar decomposition, which circumvents identification problems that arise when Σ has eigenvalues with a multiplicity. As an extreme example, if $\Sigma = I_m$, then $S = I_m$ and $L = I_m$, but P may be an arbitrary orthogonal matrix; in such cases the principal components s_t^* would form an arbitrary orthogonal transformation of the observation vector x_t, whereas $s_t = x_t$. Note also that the O-GARCH model of Alexander (2001) assumes that the standardized principal components s_t^* are independent GARCH processes, which corresponds to a special case of our model with $U^* = I_m$ (hence $s_t^* = y_t$), which in the parametrization considered here requires $U = P$.

3.2.2 Reduced factor models

Lanne and Saikkonen (2007) analyze a special case of the GO-GARCH model with independent components, in which only a subset of the components of y_t have a time-varying conditional variance. The motivation for this is that if the number of assets m is large, then it may be reasonable to expect that the conditional variance matrix Σ_t can be described by a number $r < m$ of heteroskedastic factors. Indeed, the model then reduces to a parsimoniously parametrized version of the factor ARCH model of Engle *et al.* (1990)

and Diebold and Nerlove (1989).

The variance matrix H_t can in this case be expressed as

$$H_t = \begin{bmatrix} H_{1t} & 0 \\ 0 & I_{m-r} \end{bmatrix}, \qquad H_{1t} = \mathrm{diag}(h_{1t}, \ldots, h_{rt}).$$

Partitioning $Z = [Z_1 : Z_2]$ and $U = [U_1 : U_2]$ conformably with H_t, the model implies

$$\begin{aligned} \Sigma_t &= Z_1 H_{1t} Z_1' + Z_2 Z_2' = \Sigma + Z_1 (H_{1t} - I_r) Z_1', \\ \mathrm{var}(s_t | \mathcal{F}_{t-1}) &= U_1 H_{1t} U_1' + U_2 U_2' = I_m + U_1 (H_{1t} - I_r) U_1'. \end{aligned}$$

These representations imply that in the reduced-factor model, the matrix U_2 is only identified by the properties $U_1' U_2 = 0$ and $U_2' U_2 = I_{m-r}$. In other words, U_2 and hence $Z_2 = SU_2$ are only identified up to orthogonal transformations of their columns.

In a companion paper (Boswijk and van der Weide, 2008), we propose a testing procedure for the hypothesis of a reduced-factor model based on the same sample autocorrelation matrix that will be used in the next section to estimate U . Unless indicated otherwise, in the present chapter we assume a full-factor GO-GARCH model.

3.2.3 Currently available estimation methods

In this subsection we briefly review the currently available methods for estimating the model implied by Assumption 3.1, or specific versions thereof. Although the GO-GARCH model can be considerably more parsimonious than alternative multivariate GARCH models, for larger m it will become harder to maximize its likelihood function over the entire parameter space, which has motivated the development of two-step approximations of maximum likelihood, or alternative methods that are easier to apply in larger dimensions.

Gaussian maximum likelihood estimation of the model with independent GARCH factors was analyzed by van der Weide (2002). He considered the standardized returns s_t as observable time series, which leads to a log-likelihood function of the form

$$\ell(\theta) = -\frac{1}{2} \sum_{t=1}^{n} \left\{ m \log(2\pi) + \log |H_t(\theta)| + s_t' U(\theta_1) H_t(\theta)^{-1} U(\theta_1)' s_t \right\}, \tag{3.8}$$

where $\theta = (\theta_1', \theta_2')$, with θ_1 a vector of dimension $\frac{1}{2}m(m-1)$ characterizing the $m \times m$ orthogonal matrix $U = U(\theta_1)$, and θ_2 a $2m$-dimensional parameter vector of GARCH parameters. A convenient parametrization of $U(\theta_1)$ as the product of $\frac{1}{2}m(m-1)$ rotation matrices, each characterized by one parameter, is discussed in van der Weide (2002). Note that H_t depends on U via $y_{t-1} = U's_t$, so that $H_t = H_t(\theta)$ is characterized by the full parameter vector θ. By applying the general asymptotic results of Comte and Lieberman (2003) for BEKK models (in which the GO-GARCH model is nested), conditions for consistency and asymptotic normality of the maximum likelihood estimator are obtained.

In practice s_t is not observed, and will have to be estimated by $\hat{s}_t = \hat{\Sigma}^{-1/2}x_t$, with $\hat{\Sigma}$ the sample variance matrix $n^{-1}\sum_{t=1}^{n} x_t x_t'$. (If m is large relative to n, one could also consider shrinkage-type estimators of Σ, see e.g. Ledoit and Wolf, 2004). Therefore, in practice the procedure of van der Weide (2002) is a two-step approximation of maximum likelihood. If we let $\theta_0 = \text{vech}(\Sigma)$, then full maximum likelihood would entail maximizing (3.8), with s_t replaced by $\Sigma(\theta_0)^{-1/2}x_t$, over $(\theta_0', \theta_1', \theta_2')'$. Lanne and Saikkonen (2007) derive asymptotic properties of such a full maximum likelihood procedure for the reduced-factor model considered in Section 3.2.2.

Boswijk and van der Weide (2006) proposed a non-linear least-squares estimator of U, based on the autocorrelation properties of the matrix-valued process $S_t = s_t s_t' - I_m$. Let $\hat{B}$ be the minimizer, over all symmetric matrices, of the least-squares criterion

$$Q(B) = \frac{1}{n}\sum_{t=1}^{n} \text{tr}(S_t - BS_{t-1}B)^2.$$

Using the fact that $S_t = UY_tU'$, with $Y_t = y_t y_t' - I_m$, it follows that

$$Q(B) = n^{-1}\sum_{t=1}^{n} \text{tr}(Y_t - AY_{t-1}A)^2, \qquad A = U'BU.$$

Boswijk and van der Weide (2006) derive conditions under which the probability limit of $\hat{A} = U'\hat{B}U$ is a diagonal matrix, which in turn implies that the eigenvector matrix $\hat{U}$ of $\hat{B}$ will be a consistent estimator of U. This estimator can be embedded in a three-step procedure: first estimate Σ to construct $\hat{s}_t$, then estimate U based on $\hat{s}_t$, and finally estimate the GARCH parameters based on $\hat{y}_t = \hat{U}'\hat{s}_t$.

An alternative estimator of U was proposed by Fan *et al.* (2008). The starting point of their analysis is that the conditionally uncorrelated restriction $E(y_{it}y_{jt}|\mathcal{F}_{t-1}) = 0$ is equivalent to $E[y_{it}y_{jt}I(B)] = 0$ for all $B \in \mathcal{F}_{t-1}$, where $I(.)$ is the indicator function. Let u_i denote the i-th vector of U, so that $y_{it} = u_i's_t$, let $\mathcal{B}$ be a collection of subsets of $\mathbb{R}^m$, and let p be an arbitrary integer. Then the columns of U should satisfy the (population) criterion

$$\Psi(U) = \sum_{1\le i<j\le m} \sum_{B\in\mathcal{B}} \sum_{k=1}^{p} |u_i' E[s_t s_t' I(s_{t-k} \in B)]u_j| = 0. \tag{3.9}$$

Fan *et al.* (2008) propose to estimate U by minimizing a sample analog of $\Psi(U)$, and provide a bootstrap inference procedure for this estimator, and for a test of the conditionally uncorrelatedness hypothesis. Again, this estimator of U should be preceded by the estimation of Σ and s_t, and followed by the estimation of the (extended) GARCH models for $\hat{y}_t = \hat{U}'\hat{s}_t$.

All methods considered in this subsection require numerical maximization of a criterion function over a high-dimensional parameter space. Therefore, as m increases, each of these methods is likely to run into numerical problems, such as failure of a Newton-type optimization procedure to converge, or the possibility of ending up in a local maximum. The estimator proposed in the next section, on the other hand, only requires the calculation of common eigenvectors of a sequence of sample moment matrices, and therefore can be applied to arbitrary dimensions m.

3.3 Method of moments estimation

3.3.1 The estimator based on a single lag

The starting point of our method-of-moments estimator is the same as in Boswijk and van der Weide (2006), i.e., the autocorrelation properties of the (mean-zero) matrix-valued processes $S_t = s_t s_t' - I_m$ and $Y_t = y_t y_t' - I_m$. For the autocorrelation matrices of these processes to be well-defined (and consistently estimated by their sample analogs) and to be able to identify U from these, we make the following assumption.

Assumption 3.2 *The process $\{y_t\}_{t\ge1}$ is strictly stationary and ergodic, and has finite fourth moments $\kappa_i = E(y_{it}^4) < \infty, i = 1,\dots,m$. Furthermore, the autocorrelations $\rho_{ik} = \operatorname{corr}(y_{it}^2, y_{i,t-k}^2)$ and cross-covariances $\tau_{ijk} = \operatorname{cov}(y_{it}^2, y_{i,t-k}y_{j,t-k})$ satisfy, for some integer p,*

$$\min_{1\le i\le m} \max_{1\le k\le p} |\rho_{ik}| > 0, \qquad \max_{1\le k\le p, 1\le i<j\le m} |\tau_{ijk}| = 0.$$

The stationarity assumption, as well as the assumptions on the moments, would be implied by independent GARCH processes for y_{it}, under suitable parameter restrictions to guarantee a finite kurtosis, see He and Teräsvirta (1999). Because estimated GARCH parameters in practice do not always satisfy the finite kurtosis restrictions, this assumption is not without loss of generality. In the next section, we investigate the sensitivity of our method to deviations from this assumption through Monte Carlo simulations. The non-zero autocorrelation assumption allows us to identify U from the first p autocorrelation coefficients of y_{it}^2. It would be hard to think of processes that do display volatility clustering but violate this assumption (i.e., with $\text{corr}(y_{it}^2, y_{i,t-k}^2) = 0$ for all $k = 1, \dots, p$). Finally, the zero cross-covariances τ_{ijk} exclude dependence in h_{it} on whether $y_{i,t-k}$ and $y_{j,t-k}$ have the same sign. Although this may exclude particular asymmetries in volatility, note that the assumption does allow for the extended GARCH model (3.6), possibly augmented with $y_{i,t-1}$ and $y_{j,t-1}$ (but not their product) to allow for leverage effects.

Define the autocovariance matrices

$$\Gamma_k(y) = E(Y_t Y_{t-k}), \qquad k = 1, 2, \dots \tag{3.10}$$

Note that $\Gamma_k(y)$ does not contain all separate k-th order (cross-) autocovariances of squares and cross-products of y_t (which would require vectorizing Y_t), but is an $m \times m$ matrix with elements

$$\Gamma_k(y)_{ij} = \sum_{\ell=1}^{m} \text{cov}(y_{it} y_{\ell t}, y_{\ell,t-k} y_{j,t-k}).$$

Therefore, Assumptions 3.1 and 3.2 imply, using $\text{var}(y_{it}^2) = E(y_{it}^4) - E(y_{it}^2)^2 = \kappa_i - 1$,

$$\Gamma_k(y)_{ij} = \text{cov}(y_{it}^2, y_{i,t-k} y_{j,t-k}) = \begin{cases} (\kappa_i - 1)\rho_{ik}, & j = i, \\ \tau_{ijk} = 0, & j \neq i, \end{cases}$$

or in other words

$$\Gamma_k(y) = \text{diag}((\kappa_1 - 1)\rho_{1k}, \dots, (\kappa_m - 1)\rho_{mk}).$$

For the corresponding autocorrelation matrix, we thus find

$$\Phi_k(y) = \Gamma_0(y)^{-1/2} \Gamma_k(y) \Gamma_0(y)^{-1/2} = \text{diag}(\rho_{1k}, \dots, \rho_{mk}).$$

For the process $s_t = U y_t$, the corresponding autocovariance and autocorrelation matrices satisfy

$$\Gamma_k(s) = E(S_t S_{t-k}) = E(U Y_t U' U Y_{t-k} S') = U \Gamma_k(y) U',$$

and hence

$$\Phi_k(s) = \Gamma_0(s)^{-1/2} \Gamma_k(s) \Gamma_0(s)^{-1/2} = U \Phi_k(y) U'.$$

Because $\Gamma_k(y)$ and $\Phi_k(y)$ are diagonal matrices and U is an orthogonal matrix, we find that under Assumptions 3.1 and 3.2, U may be identified by the eigenvectors of either $\Gamma_k(s)$ or $\Phi_k(s)$.

Consider the sample analogs of $\Gamma_k(s)$ or $\Phi_k(s)$:

$$\hat{\Gamma}_k(s) = \frac{1}{n}\sum_{t=k+1}^{n} S_t S_{t-k} = \frac{1}{n}\sum_{t=k+1}^{n}(s_t s_t' - I_m)(s_{t-k}s_{t-k}' - I_m), \tag{3.11}$$

$$\hat{\Phi}_k(s) = \hat{\Gamma}_0(s)^{-1/2}\hat{\Gamma}_k(s)\hat{\Gamma}_0(s)^{-1/2}, \tag{3.12}$$

where $\hat{\Gamma}_0(y)^{-1/2}$ is the symmetric square root of $\hat{\Gamma}_0(y)^{-1}$. We define our estimator $\hat{U}_k$ as the matrix of eigenvectors of the symmetrized version $\tilde{\Phi}_k(s) = \frac{1}{2}(\hat{\Phi}_k(s) + \hat{\Phi}_k(s)')$ of $\hat{\Phi}_k(s)$. Although in principle one could also take the eigenvectors of the corresponding symmetric version of $\hat{\Gamma}_k(s)$ as estimator of U, preliminary Monte Carlo experiments have indicated that the standardization used to construct $\hat{\Phi}_k(s)$ leads to a more efficient estimator.

3.3.2 Combining information from different lags

Although one could in principle use the estimator $\hat{U}_k$ proposed in the previous subsection for one particular choice of the lag length k, we may obtain a more efficient estimator by combining information from different lags. This is relevant in particular for daily financial data, where the autocorrelation function of the squares typically is small but slowly decaying. This implies that the eigenvalues $\{\rho_{ik}\}_{i=1}^{m}$ of $\Phi_k(s)$ will be close to zero (and and hence close to each other), yielding weakly identified eigenvectors for fixed k. Provided that the autocorrelation functions $\{\rho_{ik}\}_{k=1}^{\infty}$ are sufficiently different, pooling the information from different $\Phi_k(s)$ matrices will then increase the efficiency of the estimator.

Let p denote the maximal lag length, and let $\tilde{\Phi}_k = \tilde{\Phi}_k(s)$. The property that each of the population matrices $\Phi_k = \Phi_k(s)$ have the same matrix of eigenvectors is shared with the so-called common principal components (CPC) model, see Flury (1984), where $\{\Phi_k\}_{k=1}^{p}$ represent covariance matrices of p different random vectors. Under the assumption that these are Gaussian, and that we have p independent i.i.d. samples, Flury (1984) derives the maximum likelihood estimator of the common eigenvector matrix U. Because these assumptions are clearly violated, we consider a closely related least-squares estimator, discussed by Beaghen (1997). This estimator minimizes the criterion function

$$S(U) = \sum_{k=1}^{p} \mathrm{tr}(U'\tilde{\Phi}_k U - \mathrm{diag}(U'\tilde{\Phi}_k U))'(U'\tilde{\Phi}_k U - \mathrm{diag}(U'\tilde{\Phi}_k U)) \qquad (3.13)$$

over all orthogonal matrices U. Although this minimization problem does not have a closed-form solution, Beaghen (1997) shows that the so-called F-G algorithm of Flury and Gautschi (1986) can be easily adapted to this least-squares estimator. This algorithm involves an iteration of rotations until the first-order conditions are satisfied, which can be rewritten as

$$\begin{aligned} u_i' \left(\sum_{k=1}^{p} (\lambda_{ki} - \lambda_{kj}) \tilde{\Phi}_k \right) u_j &= 0, \qquad i \neq j = 1, \ldots, m, && (3.14) \\ \lambda_{ki} &= u_i' \tilde{\Phi}_k u_i, \qquad i = 1, \ldots, m; \quad k = 1, \ldots, p, && (3.15) \end{aligned}$$

with u_i the columns of U. This shows that the solution diagonalizes a weighted average of the matrices $\tilde{\Phi}_k$, with weights proportional to the difference in eigenvalues. Matrices $\tilde{\Phi}_k$ that are less informative about the matrix U, because of the near multiplicity of their eigenvalues, are therefore given less weight.

The estimator we propose can be summarized as follows.

Summary 3.1 *Starting from an m-vector of daily returns $\{x_t\}_{t=1}^n$, possibly corrected (by least-squares) for a constant mean and serial correlation, the model is estimated in the following steps:*

1. *estimate the unconditional variance matrix $\hat{\Sigma} = n^{-1} \sum_{t=1}^n x_t x_t'$, its spectral decomposition $\hat{\Sigma} = PLP'$, and hence its symmetric square root $S = PL^{1/2}P'$ and the standardized returns $s_t = S^{-1}x_t = PL^{-1/2}P'x_t$;*

2. *calculate the matrix-valued series $S_t = s_t s_t' - I_m$, its sample autocovariance matrices $\hat{\Gamma}_k(s) = n^{-1} \sum_{t=1}^n S_t S_{t-k}$, $k = 0, \ldots, p$, and its sample autocorrelation matrices $\hat{\Phi}_k(s), k = 1, \ldots, p$, from (3.12);*

3. *using the symmetrized autocorrelation matrices $\tilde{\Phi}_k = \frac{1}{2}(\hat{\Phi}_k(s) + \hat{\Phi}_k(s)')$, estimate U by minimizing the criterion function $S(U)$ given in (3.13), based on the adapted F-G algorithm;*

4. *estimate the conditionally uncorrelated components y_t by $\hat{y}_t = \hat{U}'s_t$, and estimate separate GARCH-type models for the components of y_{it} by quasi-maximum likelihood.*

3.3.3 Consistency

In this subsection we prove consistency of the estimator $\hat{U}$ defined in the previous section. We use the square root $d(\cdot,\cdot)$ of a symmetric version of the distance measure $D(\cdot,\cdot)$ for orthogonal matrices introduced by Fan *et al.* (2008):

$$\begin{aligned} d(U,\hat{U}) &= \sqrt{\tfrac{1}{2}\left[D(U,\hat{U}) + D(\hat{U},U)\right]}, \\ D(\hat{U},U) &= 1 - \frac{1}{m}\sum_{i=1}^{m} \max_{1\le j\le m} \left|u_i'\hat{u}_j\right|. \end{aligned}$$

The motivation for $D(\cdot,\cdot)$ is that in the model $s_t = Uy_t$, the columns of the matrix U may be reordered and multiplied by -1, by changing rows of y_t in the same way. In other words, U is invariant under permutation and sign change of its columns. The modification $d(\cdot,\cdot)$ is a distance function that satisfies the properties of a metric (symmetry, triangle inequality[1]), provided that an orthogonal matrix is identified by its equivalence class.

An identification assumption needed for consistency of $\hat{U}$ is the following:

Assumption 3.3 *In the model defined by Assumptions 3.1 and 3.2,*

$$\max_{1\le k\le p}\ \min_{1\le i<j\le m} \left|\rho_{ik} - \rho_{jk}\right| > 0.$$

The assumption excludes the possibility that two squared components y_{it}^2 and y_{jt}^2 have the same autocorrelation function for $k = 1,\dots,p$. The reason for this assumption is that the autocorrelations are the eigenvalues of the matrix $\Phi_k(s)$, and if this matrix has eigenvalues with a multiplicity, then the corresponding submatrix of eigenvectors is only identified up to orthogonal transformations. Because such transformations will typically destroy the property of the true matrix U, that $U's_t = y_t$ is a vector of conditionally uncorrelated components, this would result in an inconsistent estimator $\hat{U}$.

Theorem 3.1 *Consider the MM-CPC estimator $\hat{U}$, minimizing (3.13). Then, under As-*

[1] Although we have not been able to prove the triangle inequality, numerically it appears that $d(\cdot,\cdot)$ satisfies this property; the original distance $D(\cdot,\cdot)$ violates this property.

sumptions 3.1–3.3, and as $n \to \infty$,

$$d(U, \hat{U}) \xrightarrow{P} 0.$$

Proof. From the law of large numbers for stationary ergodic Markov chains, see Jensen and Rahbek (2007), it follows that under Assumptions 3.1 and 3.2, and as $n \to \infty$,

$$\hat{\Gamma}_k(s) \xrightarrow{P} \Gamma_k(s), \qquad \hat{\Phi}_k(s) \xrightarrow{P} \Phi_k(s).$$

This implies that $\hat{U}$ converges in probability to a matrix satisfying (3.14)–(3.15), with $\tilde{\Phi}_k(s)$ replaced by $U \operatorname{diag}(\rho_{1k}, \ldots, \rho_{mk})U'$. If the eigenvalues ρ_{ik} are distinct for at least one $k = 1, \ldots, p$, as implied by Assumption 3.3, then these first-order conditions are only satisfied by a matrix that is in the same equivalence class as U (defined by permutation and sign change of the columns), so that $d(U, \hat{U}) \xrightarrow{P} 0$. ∎

A next step is to derive the asymptotic distribution of the estimator. A starting point of this would be to derive conditions under which a joint asymptotic normality applies to $(\tilde{\Phi}_1(s), \ldots, \tilde{\Phi}_p(s))$, which would lead to asymptotic normality of $\hat{U}$, analogously to the results obtained by Flury (1984) and Beaghen (1997). It is clear that such results would require, at the minimum, finite eighth moments of y_{it}, which is likely to be violated in practical applications, and is therefore not considered here. Moreover, asymptotic normality of the parameter estimators in $\hat{U}$ is in itself not a very useful result, unless it helps us to make inference on the parameters directly characterizing the volatility dynamics, or helps us to evaluate estimation uncertainty in volatility and correlation forecasts. We expect that such results are more easily obtained by bootstrap procedures; we leave this for future work.

3.4 Monte Carlo simulations

In this section we study the finite-sample performance of the estimator proposed in this chapter, in comparison with maximum likelihood. We focus on a trivariate system ($m = 3$), and we consider four different data-generating processes (DGPs) for the conditionally uncorrelated process $\{y_t\}_{t \geq 1}$. The (relative) efficiency of the two estimators is evaluated using the root mean square distance (RMSD), i.e., the square root of the average of $d(U, \hat{U})^2$, over 5000 Monte Carlo replications. For both the MM-CPC estimator and the ML estimator, the distance $d(U, \hat{U})$ is invariant to U, in the sense that if $\hat{U}_1$ and $\hat{U}_2$ are estimates based on data generated using U_1 and U_2, respectively, then $d(U_1, \hat{U}_1) = d(U_2, \hat{U}_2)$. Therefore, the choice of U in the DGP is irrelevant. Three sample sizes, with $n \in \{500, 1000, 2000\}$, are considered.

We first consider two DGPs in which the components of y_t are independent Gaussian GARCH(1,1) processes, so the maximum likelihood estimator is based on a well-specified model:

- DGP A: $(\alpha_1, \beta_1) = (0.055, 0.94)$; $(\alpha_2, \beta_2) = (0.16, 0.8)$; $(\alpha_3, \beta_3) = (0.25, 0.65)$;
- DGP B: $(\alpha_1, \beta_1) = (0.095, 0.9)$; $(\alpha_2, \beta_2) = (0.26, 0.7)$; $(\alpha_3, \beta_3) = (0.25, 0.65)$.

Under DGP A, all components of y_t have finite kurtosis, so that Assumption 3.2 is satisfied. The autocorrelation decay rate $\alpha + \beta$ varies from 0.995 to 0.9, reflecting empirically relevant persistence in the autocorrelation functions of y_{it}^2. The initial autocorrelations are such that the three autocorrelation functions cross each other, implying that the average autocorrelations (over p lags) for the three series can be close to each other (depending on p). Under DGP B, the first and second component have infinite kurtosis; this DGP is included to investigate how sensitive the estimator is to deviations from the finite-fourth-moment assumption.

The number of lags used in the MM-CPC estimator is fixed at $p = 100$. Preliminary simulations have indicated that for the type of persistence in autocorrelations considered here, the efficiency does not improve much beyond $p = 50$, so that $p = 100$ may be considered a conservative choice. It should be emphasized that the improvements over $p = 1$ are substantial: the RMSD of the MM estimator based on only the first lag is typically about 3 to 5 times larger than that of the MM-CPC estimator with $p = 100$. The results, for the first two DGPs are given in Figure 3.4, top panels.

We observe that for the first two DGPs, the MM-CPC estimator has a RMSD which is about two and a half times larger than that of the ML estimator. This should be seen as the price that we pay for using a more robust and computationally less intensive estimator than ML. The RMSD of both estimators appears to decrease with the sample size at approximatly the rate $n^{-1/2}$. (In a bivariate model where U is characterized by a single angle ϕ, it can be shown that the average $d(U, \hat{U})^2$ is approximately equal to half the mean squared error of $\hat{\phi}$.) Most striking is that the MM estimator in DGP B, which does not satisfy Assumption 3.2 because some of its components have infinite kurtosis, has the same qualitative behaviour as the estimator in DGP A, and in fact has a smaller RMSD.

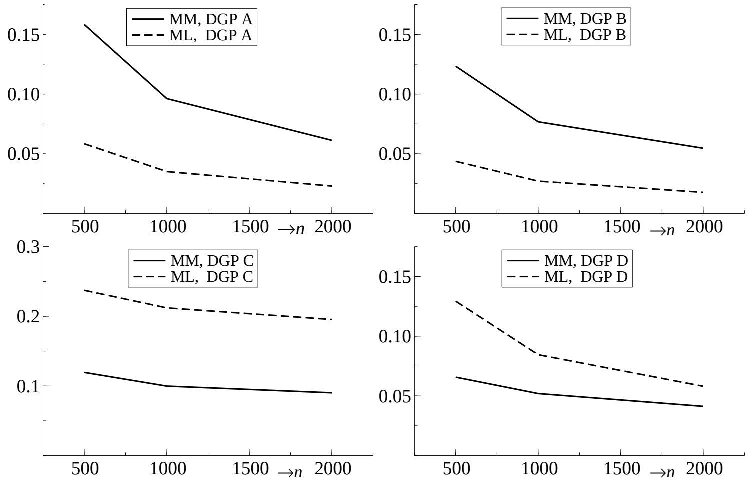

Figure 3.1: *Root mean square distance of MM and ML estimators.*

Therefore, although the definition of the moment matrices $\Gamma_k(s)$ and $\Phi_k(s)$ requires the existence of fourth moments, we observe that the estimator is not negatively affected by a departure from this assumption.

To investigate the effect of misspecification on the relative performance of the two methods, we now consider the following data-generating processes. DGP C is an extended GARCH specification, of the form

$$\begin{aligned} h_{1t} &= 0.005 + 0.035 y_{1,t-1}^2 + 0.02 y_{3,t-1}^2 + 0.94 h_{1,t-1}, \\ h_{2t} &= 0.04 + 0.08 y_{1,t-1}^2 + 0.08 y_{2,t-1}^2 + 0.8 h_{2,t-1}, \\ h_{3t} &= 0.1 + 0.25 y_{3,t-1}^2 + 0.65 h_{3,t-1}, \end{aligned}$$

where the standardized innovations ε_{it} now follow a standardized Student's $t(5)$ distribution (with ε_{it} independent of ε_{jt}, $i \neq j$). Extending the results of Hafner (2003), it can be shown that the components of y_t have finite kurtosis under this DGP. Clearly, the ML estimator is now based on a misspecified likelihood (assuming independent GARCH processes), whereas the MM estimator is still valid. Finally, DGP D has the same GARCH parameters as DGP A, but now the standardized innovations ε_{it} follow a standardized Student's $t(3)$ distribution (implying infinite kurtosis for all components of y_t). In this case the misspecification of the likelihood function only concerns the shape of the innovation density, not the volatility dynamics.

The results in the bottom panels of Figure 3.4 show that for these two DGPs and the sample sizes considered here, the MM estimator is in fact more efficient than the ML estimator. Most striking about the results for DGP C is that the RMSD appears to decrease slower as a function of the sample size than in the other cases. Comparison of the results for DGP D and DGP A (which differ only in the innovation distribution) shows that the performance of the MM estimator has improved by having fat-tailed innovations (despite the lack of finite fourth moments), whereas the ML estimator performs much worse.

The superiority of the ML estimator for the first two DGPs indicates that this is the preferred method of estimation when the dimension of the problem is not too large, and the latent processes are known to be independent Gaussian GARCH processes. On the other hand, the results for DGP C and D illustrate that maximum likelihood estimation can be sensitive to both volatility misspecification and innovation distribution misspecification, and that the MM-CPC method can be more robust in such cases. Furthermore, for larger dimensions, likelihood maximization will run into convergence problems, especially for misspecified models, which is not the case for the MM-CPC method.

3.5 Empirical application to European sector indices

In this section we analyse an empirical GO-GARCH model for Dow Jones STOXX 600 European stock market industry indices. From `www.stoxx.com`, we downloaded daily data, from the beginning of January 1992 through the end of January 2010, on the 10 industry indices, yielding $n = 4650$ daily log-returns.

Table 3.5 provides some descriptive statistics. The sample means are relatively small, which is due to the recent financial crisis: by the end of 2009 many stock prices were back at their 2003 levels. At the high end we find Basic Materials, at the low end we have Consumer Services. The latter is also among the least volatile industries, whereas the most volatile is the Technology industry. Note that while the first-order autocorrelation is statistically significant for some of the industries, the coefficients are all small in size. For the analysis all first-order autocorrelation is removed from the data.

Estimates not reported in the chapter show that all unconditional correlations range between 0.5 and 0.9, which suggests that the industries exhibit strong linkages. The in-

Industry	mean	std. dev.	ac(1)
Basic Materials	0.085	0.235	0.027*
Consumer Goods	0.060	0.169	−0.004
Consumer Services	0.028	0.187	0.013
Financials	0.036	0.239	0.057**
Industrials	0.049	0.196	0.058**
Health Care	0.072	0.184	0.024*
Oil and Gas	0.065	0.231	−0.003
Technology	0.037	0.314	0.040**
Telecom	0.050	0.251	0.026*
Utilities	0.065	0.176	−0.020

*,** significant at the 10% and 5% level.

Table 3.1: *Annualized means and standard deviations, and first-order autocorrelations of returns.*

dustries that show some of the highest correlations include: Industrials, Financials and Consumer Services. Industries that exhibit comparatively low correlations include: Oil and Gas, Technology and Health Care.

For our method-of-moments estimator, denoted by $\hat{U}_{MM}$, we used $p = 250$ lags. For the maximum likelihood estimator, denoted by $\hat{U}_{ML}$, we assume independent Gaussian GARCH(1,1) factors. The distance between the two estimates, $d(\hat{U}_{MM}, \hat{U}_{ML}) = 0.385$, indicates that the two methods provide different estimates for the link matrix U. How large these differences really are, and the consequences of this, can best be seen by comparing the estimates of the GARCH parameters, and ultimately the estimates of the volatilities and correlations.

The estimates for the GARCH parameters are shown in Table 3.5. For most factors, we observe that the MM and ML estimates are not that different from each other.

Figure 3.5 presents the estimated time-varying annualized volatilities of the 10 industry returns, obtained from the GO-GARCH model estimated by maximum likelihood or the method of moments. These are compared with the volatilities obtained from univariate GARCH(1,1) models for the returns, to check if the GO-GARCH model produces sensible estimates. We observe that for all returns, the three volatility estimates follow similar patterns, which are also remarkably similar accross industries. After a relatively stable period with low volatilities in the 1990s, volatilities tend to increase and display more variation around the time of the burst of the internet bubble in 2000. After a few years this again is followed by a stable period, which ends with the peak in the credit crisis in

Factor	$(\hat{\alpha}, \hat{\beta})_{ML}$	$(\hat{\alpha}, \hat{\beta})_{MM}$
1	$(0.055, 0.943)$	$(0.051, 0.948)$
2	$(0.074, 0.920)$	$(0.083, 0.910)$
3	$(0.088, 0.894)$	$(0.082, 0.901)$
4	$(0.105, 0.892)$	$(0.089, 0.908)$
5	$(0.088, 0.903)$	$(0.064, 0.931)$
6	$(0.027, 0.968)$	$(0.028, 0.967)$
7	$(0.033, 0.965)$	$(0.032, 0.966)$
8	$(0.075, 0.923)$	$(0.060, 0.938)$
9	$(0.024, 0.973)$	$(0.026, 0.971)$
10	$(0.053, 0.937)$	$(0.035, 0.961)$

Table 3.2: *Estimated GARCH parameters for the factors.*

September–October 2008, leading to a sharp increase in volatility, followed by a gradual decline. By late January 2010 (the end of our sample), volatilities are back to their 2007 levels.

Industries mainly differ in the impact of the internet bubble crash, which, as expected, clearly led to higher volatilities in the Technology and Telecom industries than in other industries. The main difference between the outcomes of the different estimation methods is the height of the volatility peaks. For example, for October 30, 2008, the GO-GARCH-based estimated volatilities of the Technology returns are 1.31 (ML) and 1.25 (MM), respectively, whereas the univariate GARCH-based volatility estimate is 0.67. Closer inspection reveals that on average, the ML-based volatilities are closer to the univariate GARCH estimates than the MM-based estimates, although the differences are very small; this is most evident in the results for the Technology and Telecom industries.

Note that the 10 industries yield a total of 45 different pairs. There would be little value added to report all 45 estimates of the conditional correlation processes. Pairs whose conditional correlation exhibits considerable variation over time would be most interesting, and would provide a test whether the different estimators identify the same trends and patterns in conditional correlation. The Oil and Gas industry is an example of such an industry. Correlations between Industrials and Oil and Gas for example are seen to fall as low as 0.25 in 2002, and then climb to levels as high as 0.85 in 2009. For many other pairs, especially those with high unconditional correlation, we observe relatively little variation in conditional correlation over time (which leaves little space for the two estimators to potentially disagree on). Figure 3.5 presents estimates for the conditional correlations

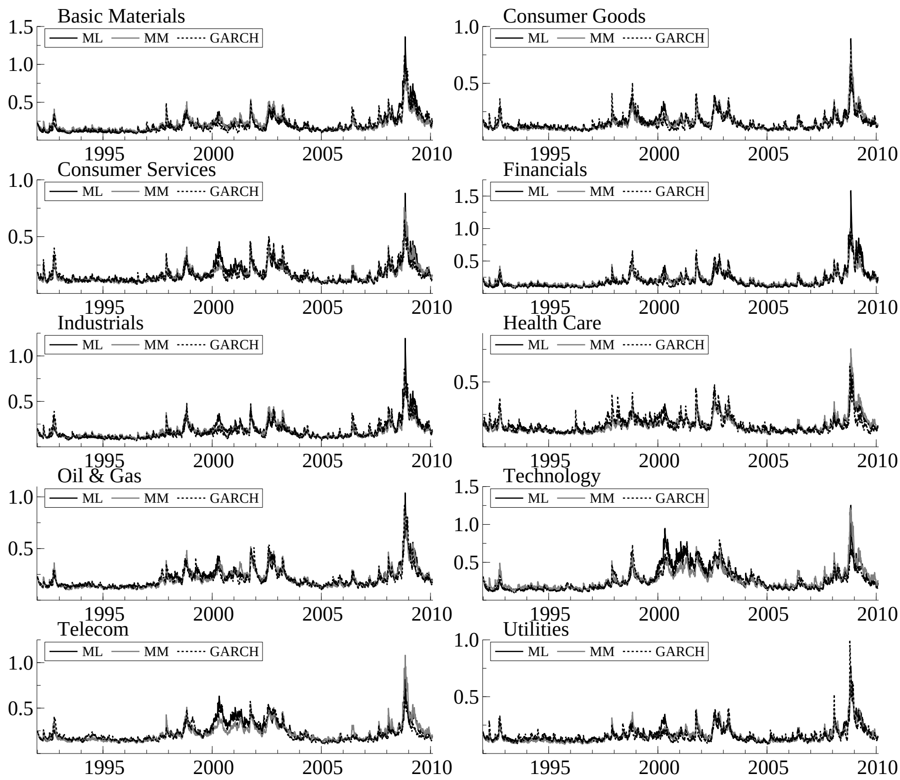

Figure 3.2: *Estimated annualized volatilities: MM, ML, and univariate GARCH.*

between the Oil and Gas industry and a choice of six other industries. We observe that, although the estimates for U show some difference, MM and ML largely agree on the conditional correlations.

What is apparent is that correlations are relatively low during the first period of higher volatility (between 1999 and 2003), while they climb from around 0.50 to 0.80 during the second period of high volatility in our sample period (between 2008 and 2009). This shows that correlations need not necessarily be higher in periods of higher volatility. What sets these two periods apart is that between 1999 and 2003 the rise in volatility is more gradual in a time when stock prices too are rising (which was the case for all industries except Technology and Telecom; not reported here). The rise in volatility in 2008 is more abrupt, in a time when stock prices started coming down as the financial crisis unfolded. This is

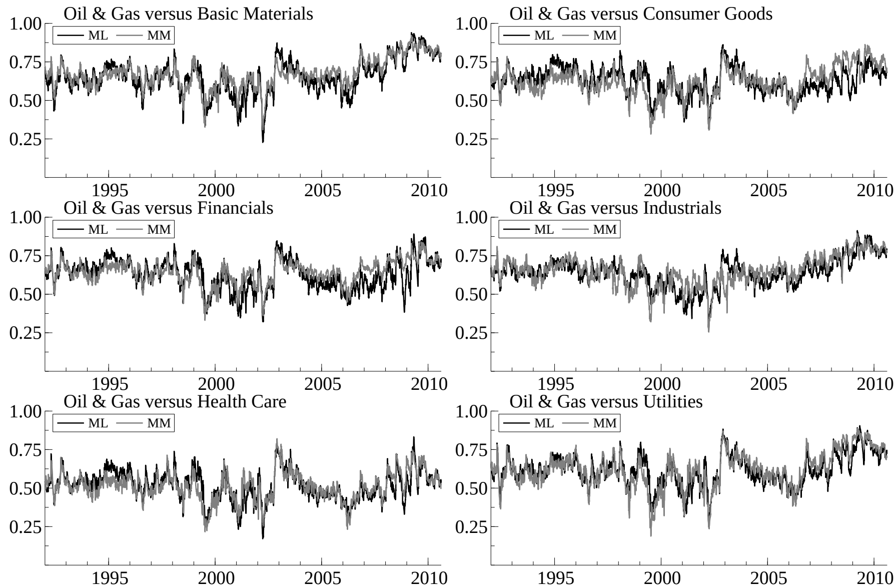

Figure 3.3: *Estimated correlations: Oil and Gas versus selected industries.*

consistent with earlier findings that correlations tend to increase in times of crisis, which are characterized by periods of extreme volatility combined with falling stock prices (see e.g. Longin and Solnik, 2001; Ang and Bekaert, 2002; Ang and Chen, 2002; Das and Uppal, 2004; and Bekaert *et al.*, 2005).

3.6 Conclusion

We have put forward a method-of-moments estimator for the factor loading matrix in GO-GARCH models. The method is based on the common eigenvectors of suitably defined sample autocorrelation matrices of squares and cross-products of the observed data. This means that estimation does not require any Newton-type optimization of an objective function, so that it is free of numerical convergence problems regardless of the dimension. The parameters from the univariate GARCH-type models can be estimated separately for each individual factor, given our estimate for the factor loading matrix, which makes the method particularly easy to implement.

Our method of estimation provides an alternative to the estimator originally proposed by van der Weide (2002), which jointly estimates parameters that feature both in the factor loading matrix and the univariate GARCH-type specifications by means of maximum likelihood. ML estimates of the factor loading matrix thus depend on the choice of GARCH-type models used to specify the likelihood function.

In a Monte Carlo experiment, the ML estimator is found to be more efficient than the MM estimator when the likelihood function is consistent with the data generating process. The loss in efficiency is the price the MM estimator pays for its numerical convenience (no optimization required), and the fact that no assumptions need to be made concerning the model specification for the individual factors. On the other hand, Monte Carlo simulations show that the MM estimator can be more efficient than ML when the latter is based on a misspecified likelihood function. Moreover, the MM estimator is a welcome alternative whenever ML experiences convergence difficulties. ML estimation can become problematic when the dimension is particularly large and/or when the model used to specify the likelihood function is considerably misspecified, while the MM estimator does not suffer from such problems.

Chapter 4

Volatility: Expectations and realizations

"Not to be absolutely certain is, I think, one of the essential things in rationality" (Bertrand Russell, 1947).

4.1 Introduction

At the heart of financial markets are market expectations. Embedded in option prices are market expectations regarding future volatility of the asset on which the option contract is written. While the assumption of rational expectations has been a popular (as well as convenient) paradigm, it is difficult to ignore the subjective nature of expectations as evidenced in recent empirical studies. Han (2008) and Constantinides *et al.* (2009) denote two such examples for the options market. "A growing literature shows that S&P 500 options are mispriced or not efficiently priced relative to a large class of rational option pricing models ... These lead to calls for research outside traditional rational option pricing" (Han, 2008; page 387). The proposed way forward is to extend the option pricing framework "to incorporate imperfect market and/or imperfect rationality" (Han, 2008; page 410), which will ultimately fit the empirical data better.[1]

[1] Han (2008) finds that investor sentiment, for example, plays an important role in determining option prices; "there is no guarantee that sentiment-induced mispricing will get corrected over a given horizon because of unpredictable investor sentiment in the future" (Han, 2008; page 390). As is to be expected, "the impact of sentiment is stronger when there are more limits to arbitrage" (Han, 2008; page 408).

The objective of this chapter is to make market expectations visible as they evolve over time, and to price options in line with prevailing expectations, be they rational or non-rational. Like financial volatility, market expectations are hidden. If not rational, market expectations can take on many different forms. Capturing expectations accurately is key for pricing options accurately. The expectation formation process denotes an integral part of the option pricing model.

Consider a European Call option contract that gives the owner the right to buy a specified quantity of an asset on which the contract is written (the underlying) at a specified future time (the expiration date) for a specified price (the strike price). Given the current price of the underlying asset, and the risk-free interest rate prevailing until expiration, the value of the option contract will be determined by the anticipated future volatility of the asset price. Agents trading option contracts are speculating on future volatility, in effect making the option market a market for financial volatility. "Because option value depends critically on expected future volatility, the volatility expectation of market participants can be recovered by inverting the option valuation formula" (Dumas, Fleming and Whaley, 1998).

Traditionally, implied volatility expectations are obtained by inverting the Black-Scholes and Merton (BSM) option pricing model that is built on the assumption that volatility fluctuates deterministically over time. The BSM model can be inverted for each strike price, but it predicts that all implied volatilities are the same. The well-documented smile and smirk pattern, that emerges when implied volatility is plotted against the strike price, contradicts the assumption of deterministic volatility.

We now know that volatility is best described by a stochastic process. Assuming that the market acknowledges the stochastic nature of volatility, implied volatility expectations will take on the form of a probability density function (pdf). Unlike traditional implied volatility, the implied pdf cannot be obtained from data on a single strike price. Instead, the option data for all strike prices jointly determine the implied pdf of volatility. So there can be no disagreement between different strike prices, there is only one implied pdf between them. The latter collapses to the traditional measure of implied volatility only in the event that the market believes volatility to be deterministic.

The 1987 stock market crash provides a great example where market expectations can be seen to change over time, arguably due to learning behaviour. Constantinides *et al.* (2009) note that "before the crash, option traders were using average historical volatility to price options and were not actively forecasting volatility changes". Although justifiable at the time perhaps, it does not sit well with rational expectations; "option traders were extensively using the BSM pricing model and the dictates of this model were imposed on the option prices even though these dictates were not necessarily consistent with the time series behavior of index prices" (Constantinides *et al.*, 2009; page 1271). Market participants like most humans are creatures of habit, until they are shaken out of their comfort zone. In their studies of this transition, Jackwerth and Rubinstein (1996) and Bates (2000), among others, compare the subjective distribution of asset prices obtained from option prices to the objective distribution. They find that through the eyes of the market, the volatility process before and after the crash of 1987 is fundamentally different. As the objective distributions show no sign of such change in the volatility process, this suggests that the market has changed instead.[2] Market expectations are more aligned with objective data after the crisis. Yet, recent findings by Constantinides *et al.* (2009) still "cast doubts on the hypothesis that the option market is becoming more rational over time, particularly after the crash" (Constantinides *et al.*, 2009; page 1271).

We put forward an analytically convenient option pricing framework that accommodates both stochastic volatility and asymmetric volatility. The latter, also known as the Fischer Black effect, captures an important stylized fact of asset prices, namely that volatility rises in times of negative price changes and drops in times of positive price changes. Option pricing models that do not accommodate this feature are found to be seriously misspecified (see e.g. Andersen *et al.*, 2002). The way we incorporate asymmetric volatility may be viewed as the simplest possible extension of the Hull and White (1987) stochastic volatility model. It offers an alternative to the Heston (1993) model. Both our model and Heston (1993) are nested as a special case in the generalized Black-Scholes model of Garcia *et al.* (2003a, 2010) (see also Garcia and Renault, 2001, and Garcia *et al.*, 2003b).

Daily estimates of the implied pdf of volatility are obtained by estimating the option pricing model one day at a time.[3] This means that we choose not to capture the dynamics of the volatility process, as perceived by the market, into a parametric model and then estimate the model parameters using the time series of option prices (from which one could re-construct the implied pdfs of volatility). Instead we keep the volatility model as simple as possible. Our sole interest is the implied distribution of volatility which

[2]Note that over time, agents may change their beliefs as well as their preferences. They may learn about the volatility process, and they may adjust their risk-aversion. If either of these changes happened abruptly, and were large enough, both might in principle be able to explain the findings linked to the crash of 1987. In this chapter, our interest is to inspect the subjective expectations held by the market as they evolve over time during every-day-trading-days, in a post-crash period, and assess market rationality.

[3]Estimation is based on the cross-section of option prices for different strike prices.

can be estimated directly, without having to first identify the model that underlies the distribution of volatility. Each new day's implied distribution of volatility is treated as a new (independent) parameter, whose estimate is not tied to estimates for other days in the sample. We worry not whether our simplistic model for volatility can rationalize the obtained time-series of implied pdfs. Market expectations are driven by many factors, including sentiment, therefore it is unlikely that any one model will be able to capture its time-variation. We explicitly do not want to impose too much structure a priori on how the market updates its expectations over time, but allow market expectations to take their course, and permit them to deviate freely from rational expectations.

Note that volatility here refers to the average variance over the remaining lifetime of the option. Even though the implied pdf of volatility is the natural measure to consider for volatility expectations in a world where volatility is stochastic, studies of the implied pdf of volatility and its time-variation are still uncommon. Upon replacing tradition implied volatility with the implied pdf of volatility, a popular empirical question that is worthwhile revisiting is whether volatility expectations contained in option prices yield better predictions of future volatility than predictors based on past (realized) volatility (see e.g. Busch *et al.* (2011), and the references therein).

Under the standard assumption that volatility risk is not priced, comparing the implied pdf to estimates of the objective pdf of average volatility opens the door for testing of market rationality. The objective pdf, however, is not easily estimated. For any given period in time we observe at best the level of average volatility that has been realized, but not the distribution from which it has been drawn. Yet, under suitable assumptions, daily obervations of realized volatility will lend itself for estimating moments of the objective pdf that can then be compared to moments of the subjective pdf. An example of such a moment is the degree of persistence in volatility, which determines how fast the volatility of average volatility declines when increasing the time to expiration. Similarly, estimates of realized volatility may be compared to the first moment of the implied pdf (although the former denotes a noisy estimate).

Our first empirical findings include: (i) the first moment of the implied pdf closely follows estimates of average realized variance, (ii) estimates of implied persistence in variance suggest that the market is fully aware of the fact that variance exhibits long-range dependence, and (iii) market expectations about future average variance appear to exhibit a degree of foresight.

In addition to fitting our option pricing model to the data, we include an analytic study of the implied variance function. The analytic expressions obtained show how the implied variance function is shaped by the model parameters. These results help with gauging the ability of the model to fit the empirical data, and may also be used in the estimation of the model parameters.

The remainder of the chapter is organized as follows. Section 4.2 presents our data. The theoretical framework is presented in Section 4.3 which includes the derivation of the option pricing model. Subsequently we study the implied variance function in Section 4.4. In Section 4.5 we discuss the method of estimation and provide a brief evaluation of model performance. Section 4.6 presents the empirical results. Finally, Section 4.7 concludes.

4.2 Data

4.2.1 Underlying value

For the underlying value we consider the German DAX stock index. The data used in our empirical illustration covers a period of one year: December 15, 2005 until December 15, 2006. We derive the price data from the most actively traded future on the DAX, which is the one with the shortest time to expiration. For the future's price we take the average between the bid and ask price recorded daily at 17:30 (the option prices too are recorded at 17:30).

The DAX stock index is corrected for dividends on the underlying stocks, which means that dividends are automatically reinvested. We view this as a convenient property. Alternatively, we would first need to derive the (expected) dividends until expiration ourselves, and correct the underlying prices accordingly.

Since we need not worry about dividends, the price of the underlying at time t denoted by S_t can be obtained by:

$$S_t = F_{t,T} e^{-r(T-t)}, \tag{4.1}$$

where $F_{t,T}$ denotes the future price at time t with expiration $T > t$, and where r denotes the instanteneous interest rate (which for the sake of simplicity is assumed constant). Our interest rate data consists of quotes on the 1-12 month Euribor rate. The time t anual interest rate for any $0 < t < T$ is obtained by means of linear interpolation.

4.2.2 Realized volatility

We construct a measure of realized variance for the underlying value by using tick-by-tick data on the DAX future(s) for the period December 15, 2005 until December 15, 2006. While the DAX future is traded from 9:00 until 20:00, the future is not traded very actively any more after the closing of stock trading at 17:30. For this reason we will focus on intraday future prices between 9:00 and 17:30.

Let $F_{t_{n,j}}$ denote the jth observation of the future price at day n corresponding to time t_n.[4] The length of the interval between subsequent obervations measures two minutes, which yields a total of 255 observations per day. The sum of the intraday squared returns will define our measure of quadratic variation rv_{t_n} at day n:

$$rv_{t_n} = \left(\log F_{t_{n,1}} - \log F_{t_{n-1,255}}\right)^2 + \sum_{j=2}^{255} \left(\log F_{t_{n,j}} - \log F_{t_{n,j-1}}\right)^2 .$$

The DAX future is among the most liquid futures on the European market which helps curb the effects of market microstructure, even at two minute intervals.

We use the daily realized variances to construct measures of average (anual) variance over periods with varying times to expiration $T-t$ (as t approaches T). A confirmation that realized variance provides accurate estimates of the stochastic variance process, such that standardized returns are indeed standard normal, can be found in e.g. Peters and de Vilder (2006). For a more elaborate discussion of realized variance we refer to Barndorff-Nielsen and Shephard (2002), and the references therein.

4.2.3 Option prices

Our option data consists of both puts and calls on the German DAX stock index, which are of the European type, with expiration date December 15, 2006 (the third Friday of the month). For each trading day between December 15, 2005 and December 15, 2006 (one year to expiration), we have both bid and ask prices recorded at 17:30 for a large series of strike prices. The number of strikes available varies by day. There tends to be no trade in options when the difference between their strike price and the value of the DAX has become too large. Also new strikes are introduced as the underlying value reaches new price levels over time.

[4] For notational convenience we supressed the indicator for expiration time.

Let us briefly summarize the construction of the option price data we use in the empirical illustration. For each strike price we have at most four daily observations:

$$(C_b, C_a, P_b, P_a) \tag{4.2}$$

where C_b, P_b (C_a, P_a) denote the bid (ask) price of the put and call option, respectively. For notational convenience we supressed the indicator for t. As there are no quoting obligations at the German option market, it is possible that the vector above is only partially observable. Provided there exists at least one bid price, and one ask price, the vector (C_b, C_a, P_b, P_a) can in principle be fully reconstructed using the put-call parity:

$$S + P^a - C^b = Ke^{-r(T-t)}, \tag{4.3}$$

where K denotes the strike price, and where r denotes the interest rate.

We estimate r by exploiting the put-call parity for strike prices for which we have both the call and the put price.[5] This yields an implied interest rate for each strike price. We take the average as our estimate for r. It follows that the estimated r closely matches the trend of the (interpolated) Euribor rate. We do observe a small bias. For much of the period, our estimate for r is slightly lower (2.7 versus 3 percent at the start of the period in December, 2005; 3 versus 3.5 half a year later). The two converge however toward expiration. Our estimate for r is used in the estimation of the option pricing model.

As puts and calls are directly related, we focus on one of them, which will be the call price. We either use the actual quote on the call option, or the price derived from the corresponding put using the put-call parity. If both are available, the one with the smaller bid-ask spread is used. We decided to drop option prices if they fall below 50 cents. Once the appropriate adjustments have been made, the price of the call we will work with is the mid-price:

$$C(t, T, K) = \frac{C_b(t, T, K) + C_a(t, T, K)}{2}. \tag{4.4}$$

Let N_t denote the number of option prices available at time t, and let $K_i(t, T)$ with $i = 1, \ldots, N_t$ denote the corresponding strike prices. As we will be estimating implied pdf's (of average volatility) for each day t separately, it is important that N_t is sufficiently large as it denotes the number of observations available to construct these daily estimates. Let m_t (M_t) measure the minimum (maximum) strike price as a fraction of the price of

[5]We exclude options whose price is less than 50 cents and options that are far out of the money (with a log moneyness below -0.10).

the underlying at day t:

$$m_t = \min\{K_1(t,T),\ldots,K_{N_{t,T}}(t,T)\}/S_t \quad (4.5)$$

$$M_t = \max\{K_1(t,T),\ldots,K_{N_{t,T}}(t,T)\}/S_t. \quad (4.6)$$

Expiry	# Days	$\bar{N}$	$\bar{m}$	$\bar{M}$
2006-12	257	26.2	0.852	1.184

Table 4.1: *Summary statistics of option price data set*

Table 4.1 presents the summary statistics for our option price data set; it shows the total number of trading days, the average number of observations per trading day, and the average of m_t and M_t. Note that averages are taken over the total number of trading days.

4.3 Option pricing framework

4.3.1 Model for underlying value

This section introduces our model for the stock price process. We formulate our model in discrete time, which provides an analytically convenient framework that is accommodating to asymmetric volatility.[6] The model is built on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $0 = t_0 < \ldots < t_N = T$ denote the timepoints at which we observe price quotes. For notational convenience we assume they are equidistant[7], $\Delta t = t_n - t_{n-1}$. At each timepoint t_n the information available to the market is denoted by $\mathcal{F}_n$, $n = 0, \ldots, N$. Formally, $\mathcal{F}_n, n = 0, \ldots, N$ is an increasing sequence of σ-algebras with $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_N = \mathcal{F}$.

Conditional on $\mathcal{F}_{n-1}$ the price of a stock S_n is determined by:

$$S_n = S_{n-1}\exp\left(\mu_n \Delta t - \frac{V_n}{2}\Delta t + (\beta V_n - \gamma_n)\Delta t + V_n^{\frac{1}{2}}\Delta t^{\frac{1}{2}} U_n\right) \quad (4.7)$$

$$V_n = h_n(W_n), \quad (4.8)$$

where $V_0, \ldots, V_N$ denotes the variance process, which is assumed bounded in L^2. The random variables U_n and W_n, $n = 1, \ldots, N$, are measurable with respect to $\mathcal{F}_n$. Conditional

[6]For arguments in favour of discrete time, other than analytical convenience, see e.g. Brennan (1979) and Rubinstein (1976a).

[7]It is straightforward to generalize this to allow for non-equidistant timepoints.

on $\mathcal{F}_{n-1}$ the random variable U_n is normally distributed with mean zero and unit variance. The variable W_n is independent of $\mathcal{F}_{n-1}$. Moreover, the sequences $(U_n)_{n=1}^N$ and $(W_n)_{n=1}^N$ are independent.

Assumption 4.1 *The processes $(\gamma_n)_{n=0}^N$ is deterministic and satisfies:*

$$\gamma_n = \Delta t^{-1} \log(E[\exp(\beta V_n \Delta t)|\mathcal{F}_{n-1}]), \tag{4.9}$$

which yields:

$$\mu_n = \Delta t^{-1} \log E[(S_n/S_{n-1})|\mathcal{F}_{n-1}]. \tag{4.10}$$

With Assumption 4.1 we rule out the possibility of arbitrage opportunities where persistently low (high) values of V_n would yield predictably higher (lower) returns. While Assumption 4.1 places strong restrictions on the stochastic volatility model, it is found to be effective in the sense that the framework accommodates asymmetric (stochastic) volatility, fits the empirical data well (as we will show later), and preserves analytical convenience. For a number of popular stochastic volatility models, for which γ_n would be stochastic, the option pricing formula may be viewed as an accurate approximation where it is no longer exact.[8]

β governs the degree of asymmetric volatility.[9] For non-zero β we have that stock returns are correlated with the variance process. Negative values are consistent with the Fischer Black effect, a key stylzed fact of the stock price process: Volatility is high in times of negative price changes, and volatility is low in times of positive price changes.

V_n measures the variance of the log asset price return when V_n is known. In what follows we will derive quadratic variation rv_{t_n} and treat this as realizations of V_n. Note that at any time prior to time t_n both V_n and rv_{t_n} are random variables. While we observe rv_{t_n} *ex post*, we do not observe realizations of V_n. Instead we estimate the distribution of V_n conditional on the information that is available at times prior to t_n, and then consider the first moment of this distribution as an estimate for V_n (and compare this to rv_{t_n}). Conditional on $\mathcal{F}_{n-1}$

[8] Examples of such stochastic volatility models are the autoregressive models and simple regime switching models for parameter values that correspond to a strong persistence in volatility.

[9] Asymmetric volatility can be captured in different ways. In our model the volatility shocks that correlate with the asset return are modeled as deviations from the 'local mean'. This choice is largely motivated by analytic convenience; it follows that the latent stochastic volatility process in this case only impacts option prices via its average over time to expiration. The stochastic volatility model of Heston (1993) denotes a well-known alternative where volatity shocks are modeled as incremental changes in volatility. While this imposes fewer restrictions on the volatility process, the pricing of options now require integration over the joint distribution of average volatility and the volatility at expiration.

we have that $var[\log{(S_n/S_{n-1})}|\mathcal{F}_{n-1}] = E[V_n|\mathcal{F}_{n-1}]\Delta t + (\beta - \frac{1}{2})^2 var[V_n|\mathcal{F}_{n-1}]\Delta t^2$. The second term however can be made arbitrarily small relative to the first term by choosing Δt arbitrarily small, so that $var[\log{(S_n/S_{n-1})}|\mathcal{F}_{n-1}] \approx E[V_n|\mathcal{F}_{n-1}]\Delta t$.

Average variance V is given by:

$$V = \sum_{n=1}^{N} V_n/N. \tag{4.11}$$

Let φ denote the probability density function of V. The first two moments of φ will be denoted by $\lambda = E[V]$ and $\nu = var[V]$. Let us define $\gamma = T^{-1}\log(E[\exp(\beta TV)])$. If the variance of V is positive then the distribution of stock price returns is leptokurtic. It follows that skewness largely depends on the parameter β.

Without loss of generality, we pick $t_0 = 0$ as our reference point at which options with expiration time T will be priced.

4.3.2 Pricing of options

When "trading takes place only at discrete intervals, it is in general not possible to construct a portfolio containing the contingent claim and the underlying asset in such proportions that the resulting portfolio return is non-stochastic" (Brennan, 1979).[10] But even in continuous time, when volatility is stochastic and the corresponding risks cannot be hedged, markets are incomplete, such that we cannot price derivatives via risk-free replicating portfolios, as can be done in e.g. Black and Scholes (1973) and Cox, Ross and Rubinstein (1979). In the more realistic setups where volatility is stochastic and trading is not continuous, it is however still possible to derive option pricing formulas. It follows that preferences and consumption will then play a role.

The Stochastic Discount Factor (SDF) provides a general framework for asset pricing. The fundamental equation states that the price P_n of an asset at time t_n with payoff G_{n+1} solves:

$$P_n = E\left[m_{n+1}G_{n+1}|\mathcal{F}_n\right], \tag{4.12}$$

where the (positive) random variable m_{n+1} denotes the SDF. Expectations are taken over the objective probability distribution. The SDF m_{n+1} is used to price all assets (including derivatives), with the bond price $B_{n,n+1}$ ($E[m_{n+1}] = B_{n,n+1}$) and the stock price S_n

[10] For conditions under which the Black-Scholes formula can be obtained within a discrete-time model see e.g. Merton (1973) and Rubinstein (1976b). They assume, however, that volatility is non-stochastic.

($E[m_{n+1}S_{n+1}] = S_n$) nested as obvious special cases. For a recent overview on option pricing see Garcia *et al.* (2003a, 2010), who adopt the SDF framework to derive a generalized Black-Scholes model.

In a multi-period setting the SDF is given by the concatenation of the one-period SDFs:

$$m_0^N = \prod_{n=1}^{N} m_n. \tag{4.13}$$

Accordingly, the price of a contingent claim P_0 at t_0 with payoff $G(S_N)$ is given by:

$$P_0 = E\left[m_0^N G(S_N)|\mathcal{F}_0\right]. \tag{4.14}$$

Consider the following two assumptions on the SDF which we borrow from Garcia *et al.* (2003a):

Assumption 4.2 *The distribution of* $(\log(m_n), \log(S_n/S_{n-1}))$ *conditionally on* $\mathcal{F}_{n-1}$ *and* W_n *is bivariate normal for* $n = 1, \ldots, N$.

Assumption 4.3 *The pair* $(\log(m_n), \log(S_n/S_{n-1}))$ *is independent of* $\mathcal{F}_{n-1}$ *given* $W_1, ...W_{n-1}$ *for* $n = 1, \ldots, N$.

It follows that when Assumptions 4.2 and 4.3 are satisfied, and $U_1, \ldots, U_N$ are i.i.d. normal, contingent claims can be priced via risk neutral valuation relationships (RNVR), which permits the following convenient representation (see Garcia *et al.*, 2003a):

$$P_0 = e^{-rT} E_{\mathbb{Q}}\left[G(S_N)|\mathcal{F}_0\right]. \tag{4.15}$$

Here $\mathbb{Q}$ denotes the risk-neutral probability measure. Under this new measure the drift $\exp(\sum_{n=0}^{N} \mu_n \Delta t)$ of the stock price process is shifted to $\exp(rT)$.

Albeit not explicitly, Assumptions 4.2 and 4.3 make assumptions about investors preferences. By definition the SDF is determined by:

$$m_{n+1} = \beta \frac{u'(C_{n+1})}{u'(C_n)}, \tag{4.16}$$

where C_n denotes aggregate consumption at time t_n, and u the investor's utility function. Equation (4.16) can be derived from optimizing utility. Note that Assumption 4.2 is satisfied when we assume a power utility function, and that returns on aggregate wealth $\log(C_n/C_{n-1})$ equal returns on the stock $\log(S_n/S_{n-1})$. Brennan (1979) was the first to show that these are necessary and sufficient for RNVR in a one-period setting[11]. Assumption 4.3 is included to extend these results to the multi-period case. Note that we implicitly make the standard assumption that volatility risk is noncompensated.

The following theorem derives the price of an European call option, given our model for the underlying value, as a function of β and φ, and the Black-Scholes option pricing formula.

Theorem 4.1 *Given our model for the underlying value, and Assumptions 4.1, 4.2 and 4.3, we have that the price of a call* $C_\varphi(K,S,r,T) = e^{-rT}E_{\mathbb{Q}}[(S_T-K)^+]$ *at time* $t_0 = 0$ *with expiration date* T *and strike price* K *solves:*

$$C_\varphi(K,S,r,T) = E_V[e^{(\beta V-\gamma)T}C_{BS}(K,S,r(V),T,V)] \tag{4.17}$$

$$= \int_0^\infty e^{(\beta v-\gamma)T}C_{BS}(K,S,r(v),T,v)\varphi(v)dv, \tag{4.18}$$

with $r(v) = r + (\beta v - \gamma)$, *and where* $C_{BS}(K,S,r(v),T,v)$ *denotes the Black-Scholes option pricing formula with volatility* $\sqrt{v}$*:*

$$C_{BS} = S\Phi(d_+) - Ke^{-r(v)T}\Phi(d_-) \tag{4.19}$$

$$d_\pm = (\ln(S/K) + (r(v) \pm \frac{v}{2})T)/\sqrt{vT}, \tag{4.20}$$

with Φ *the standard normal distribution function.*

Proof. The proof follows directly from Garcia *et al.* (2003a, 2010). It can be verified that our model is nested as a special case of their generalized Black-Scholes model (GBS); equations (4.7) and (4.8), and Assumptions 4.1 to 4.3, are seen to satisfy Assumptions 2.2 to 2.4 of Garcia *et al.* (2003a).[12] The GBS solves (see eq. (2.16) in section 2.6.2 of Garcia

[11]Assumption 4.2 thus implicitly assumes that preferences can be described by a power utility function, as Brennan (1979) does. Yet, Brennan (1979) assumes these preferences only to establish conditional joint lognormality of (m_{n+1}, S_{n+1}), which allows him to derive RNVR. For this reason Garcia *et al.* (2003a, 2010) directly assume Assumption 4.2 instead. Note that Assumption 4.3 is also borrowed from Garcia *et al.* (2003a, 2010).

[12]In Garcia *et al.* (2010), the same assumptions are implied by the model given in eq. (2.27) and (2.28) in section 2.6.

et al., 2003a):[13]

$$C_0 = E_0[\xi_{0,T} S_0 \Phi(d_+) - K e^{-rT}\Phi(d_-)], \tag{4.21}$$

where $\xi_{0,T}$ solves:

$$\begin{aligned}
\xi_{0,T} &= E\left[m_0^N \frac{S_N}{S_0}|W_1, ..., W_N\right] && (4.22)\\
&= e^{-rT} E_{\mathbb{Q}}\left[\frac{S_N}{S_0}|W_1, ..., W_N\right] && (4.23)\\
&= e^{(\beta V - \gamma)T}, && (4.24)
\end{aligned}$$

with $E[\xi_{0,T}] = 1$. The first $E[.]$ takes expectations over the joint distribution of the stochastic discount factor m_0^N and the log-return S_N/S_0. In the second step we make use of the RNVR that holds due to Assumptions 4.2 and 4.3 (see eq. (2.8) and (2.9) on page 12 of Garcia *et al.*, 2003a). Substituting this into equation (2.16) from Garcia *et al.* (2003a) then yields:

$$C_\varphi(K, S, r, T) = S E_V\left[e^{(\beta V - \gamma)T}\Phi(d_+) - K e^{-rT}\Phi(d_-)\right], \tag{4.25}$$

where expectations are taken over stochastic (average) variance V. After rearranging terms we obtain equation (4.17). ■

It can be verified that our model is nested as a special case of the generalized Black-Scholes model (see e.g. Garcia *et al.*, 2010). Note that for $\beta = 0$ we obtain the Hull and White (1987) option pricing model.[14]

4.4 Implied variance function

Let us define money-ness by $x = \log(\frac{K}{\exp(rT)S_0})$. It will also be convenient to define:

$$\begin{aligned}
c_{BS}(x, T, v) &= C_{BS}(K, S, r, T, v)/S\\
&= \Phi(d_+(x, T, v)) - e^x \Phi(d_-(x, T, v)),
\end{aligned}$$

[13] The original equation (2.16) features the ratio $B^*(0,T)/B(0,T)$ where $B^*(0,T)$ denotes the bond price conditional on the latent state variable W. For our model $B^*(0,T) = B(0,T)$.

[14] For an early overview on the empirical performance of alternative option pricing models, which include the popular Hull and White (1987) and Heston (1993) models, see e.g. Bakshi *et al.* (1997).

where $d_{\pm}(x,T,v) = -\frac{x}{\sqrt{vT}} \pm \frac{1}{2}\sqrt{vT}$. Similarly, we define c_φ by:

$$\begin{aligned}
c_\varphi(x,T) &= C_\varphi(K,S,r,T)/S \\
&= E_V[e^{(\beta V-\gamma)T} C_{BS}(K,S,r(V),T,V)/S] \\
&= E_V[e^{(\beta V-\gamma)T} c_{BS}(x_V,T,V)] \\
&= E_V[e^{(\beta V-\gamma)T} \left(\Phi(d_+(x_V,T,V)) - e^{x_V}\Phi(d_-(x_V,T,V))\right)] \\
&= E_V[e^{(\beta V-\gamma)T}\Phi(d_+(x_V,T,V)) - e^{x}\Phi(d_-(x_V,T,V))],
\end{aligned}$$

where $x_V = x - (\beta V - \gamma)T$. (Note that in obtaining the last equation we used that: $e^{(\beta V-\gamma)T}e^{x_V} = e^x$.)

Definition 4.1 *The implied variance function $I(x)$ for the option pricing model from Theorem 4.1 is defined as the solution to the following integral equation:*

$$\begin{aligned}
c_{BS}(x,T,I(x)) &= \int_0^\infty e^{(\beta v-\gamma)T} c_{BS}(x-(\beta v-\gamma)T,T,v)\varphi(v)dv \\
&= c_\varphi(x,T).
\end{aligned}$$

The function $I(x)$ is single-valued for each x, which follows from the invertibility of c_{BS} as a function of variance v.

Next we will derive selected properties of $I(x)$. As symmetric volatility ($\beta = 0$) versus asymmetric volatility ($\beta \neq 0$) denotes an important distinction, both analytically and empirically, we will discuss them in separate subsections.

4.4.1 Symmetric stochastic volatility

In this subsection we will assume $\beta = 0$.

Theorem 4.2 *The implied variance function $I(x)$ is quasi-convex:*

$$\begin{aligned}
&\partial I(x)/\partial x < 0 \qquad if \quad x < x^* \\
&\partial I(x)/\partial x > 0 \qquad if \quad x > x^*,
\end{aligned}$$

where x^* denotes the unique minimum that is attained at $x^* = 0$.

Proof. To proof follows from Theorem 4.2 of Renault and Touzi (1996), in which it is assumed that volatility risk is noncompensated (Assumption 2.3), as it is in our framework.

■

Theorem 4.3 *The implied variance function $I(x)$ is symmetric:*

$$I(x) = I(-x) \quad \forall x. \tag{4.26}$$

Proof. The proof follows directly from Proposition 3.1 of Renault and Touzi (1996). ∎

Theorem 4.2 and 4.3 show that $I(x)$ is both quasi-convex and symmetric for $\beta = 0$. Note that $I(x)$ need not be a convex function, as is illustrated by the following example.

Example 4.1 *Suppose that the pdf of V is given by:*

$$\varphi(v) = \frac{1}{v_1 - v_0} 1_{[v_1, v_0]}(v), \qquad v_1 > v_0 > 0. \tag{4.27}$$

Then $v_0 < I(x) < v_1$, and hence the function $I(x)$ cannot be convex.

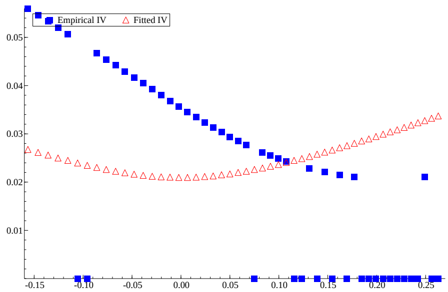

Figure 4.1: *Mismatch when imposing symmetric volatility (June 23, 2006)*

It follows that the symmetry and the location of the minimum of the implied variance function do not depend on φ. Provided that φ has positive dispersion (the market believes in stochastic volatility), we will observe a smile where the minimum is attained at $x^* = 0$. While the empirical implied variance function exhibits a smile pattern, the theoretical implied variance function in case of symmetric volatility ($\beta = 0$) typically does not fit the empirical data well. Figure 4.1 shows a typical mismatch. Two observations are apparant. First, and most notably, the horizontal alligment of the theoretical miminum ($x^* = 0$) does not match with the empirical minimum of the smile. Second, while the theoretical smile is always symmetric, the empirical smile is not.

4.4.2 Asymmetric stochastic volatility

Asymmetric volatility (obtained for $\beta \neq 0$) will introduce both asymmetry in the implied variance function and a shift in where it attains its minimum.

Conjecture 4.1 *The minimum of $I(x)$ satisfies:*

$$x^* \approx -2\beta T I(x^*). \tag{4.28}$$

If the relationship from Conjecture 4.1 is not exact, our simulation results suggest that it holds as a close approximation. Assuming that Conjecture 4.1 holds true in general, it confirms that for $\beta < 0$ the minimum of the implied variance function is attained at $x^* > 0$, which is consistent with the empirical data. Given that β and $I(x^*)$ are finite, the conjecture also predicts that the location of the minimum tends to zero ($x^* \to 0$) when time to expiration tends to zero ($T \to 0$).

Corollary 4.1 *Assuming Conjecture 4.1 holds true and $|\beta| < \infty$, we have:*

$$\lim_{T \to 0} x^*(T) = 0. \tag{4.29}$$

For the empirical variance function we indeed see that the minimum moves closer to $x^* = 0$ as time to expiration T becomes smaller. Where the empirical minimum does not exactly respect $x^* \to 0$ for $T \to 0$, it holds approximately. Finally, note that $\beta = 0$ implies $x^* = 0$ for any time to expiration T, such that the conjecture is also consistent with what we know for $\beta = 0$.

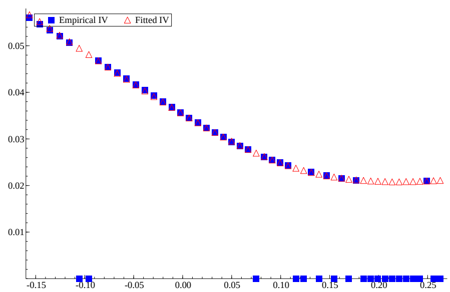

Figure 4.2: *Fit for asymmetric empirical smile (June 23, 2006)*

Figure 4.2 plots the empirical implied volatility function together with the fitted theoretical values where we allowed for $\beta < 0$. We used the same empirical data as in Figure 4.1. Now that asymmetric volatility is accounted for, we observe a nearly perfect match. The minimum of the implied variance function perfectly coincides with the empirical minimum (compare Figure 4.1 with Figure 4.2), and also the skewness introduced for $\beta < 0$ is in line with the empirical smile.

Conjecture 4.1, assuming it holds true, also suggests that we can estimate β as a function of the minimum $(x^*, I(x^*))$ of the empirical implied variance function: $\hat{\beta} = -\frac{1}{2}\frac{x^*}{TI(x^*)}$.

Next we will derive the asymptotic left and right slopes of the theoretical implied variance function. If both the left and right slope coefficients can be observed for the empirical implied variance function, they too can be used to construct an estimate for β.

Let α_R and α_L measure the slopes of linear "asymptotes" to implied variance:

$$\alpha_R(T) := \lim_{x\to\infty} \sup \frac{I(x,T)}{|x|/T} \tag{4.30}$$

$$\alpha_L(T) := \lim_{x\to-\infty} \sup \frac{I(x,T)}{|x|/T}. \tag{4.31}$$

Using this notation, the tail slopes of $I(x)$ (in absolute value) will be α_R/T and α_L/T. For any model for the underlying value S_T, the coefficients α_R and α_L belong to the interval $[0, 2]$ (see Lee, 2004), which confirms that $I(x)$ becomes flat when T tends to infinity. Given a model, the exact values are entirely determined by the following two moments of the distribution of S_T:

$$p := \sup\{p : E[S_T^{1+p}|F_0] < \infty\} \tag{4.32}$$

$$q := \sup\{q : E[S_T^{-q}|F_0] < \infty\}. \tag{4.33}$$

The relationship between (p, q) and (α_R, α_L) is given by (see Lee, 2004):

$$\alpha_R = 2 - 4\left(\sqrt{p^2 + p} - p\right) \tag{4.34}$$

$$\alpha_L = 2 - 4\left(\sqrt{q^2 + q} - q\right). \tag{4.35}$$

In the next theorem p and q are derived for our model.

Theorem 4.4 *Let $E[V] = \lambda$ and $var[V] = \nu$. In case average variance V is Gamma or Inverse Gaussian distributed, the moments p and q, as defined in eq. (4.32) and (4.33),* *are given by:*

$$p = (\sqrt{\xi} - \frac{1}{2}) - \beta \quad (4.36)$$

$$q = (\sqrt{\xi} - \frac{1}{2}) + \beta, \quad (4.37)$$

where

$$\xi = (\beta - \frac{1}{2})^2 + \frac{\lambda}{\nu T} c_\varphi. \quad (4.38)$$

with $c_\varphi = 1$ in case of Inverse Gaussian, and $c_\varphi = 2$ in case of Gamma.

Proof. The proof is given in the Appendix. ■

For $\beta < 0$ the result implies $\alpha_R > \alpha_L$, which is consistent with empirical data. Note that only for $\beta = 0$ we have $p = q$. In other words, $\alpha_R = \alpha_L$ if and only if $\beta = 0$, which denotes the symmetric case.

Corollary 4.2 *Let φ denote the Gamma or Inverse Gaussian pdf. Then, the degree of asymmetry β can be derived from:*

$$\beta = \frac{1}{2}(q - p), \quad (4.39)$$

where p and q are uniquely determined from α_R and α_L. Subsequently, λ/ν can be solved from the expression for q (or p) from Theorem 4.4:

$$\frac{\lambda}{\nu} = q^2 T c_\varphi^{-1} - 2q(\beta - \frac{1}{2}) T c_\varphi^{-1}. \quad (4.40)$$

Proof. The proof is straightforward, and therefore omitted. ■

It thus follows that the degree of asymmetry β can easily be derived from the empirical values of α_R and α_L, independent of the values of the other model parameters.

The next theorem provides an analytical approximation of $I(x)$ as a function of the moments of φ without making further assumptions about the functional form of φ.

Theorem 4.5 *Let $E[V] = \lambda$, $var[V] = \nu$, and $\gamma = T^{-1}\log(E[\exp(\beta T V)])$. Without further assumptions about the pdf φ, the implied variance function $I(x)$ can be approximated* *by:*

$$I(x) \approx \lambda + \frac{\kappa}{4v_0}\left(\frac{x^2}{v_0T} - 1 - \frac{v_0T}{4}\right) + \beta\kappa T\left(\beta + \frac{1}{2} + \frac{x}{v_0T} + (v_0T)^{\frac{1}{2}}(\beta - \frac{2\psi}{\kappa T})a(x)\right),$$

where

$$a(x) = \frac{\Phi(d_-(x))}{\phi(d_-(x))},$$

with $d_-(x) = \frac{1}{2}(v_0T)^{\frac{1}{2}} - x(v_0T)^{-\frac{1}{2}}$. *The functions* Φ *and* ϕ *denote the standard normal cumulative distribution and density function. Variance level* v_0 *is defined by:*[15]

$$v_0 = \begin{cases} \gamma/\beta & if \quad \beta \neq 0 \\ \lambda & if \quad \beta = 0 \end{cases}$$

The parameters (ψ, κ) *are given by:* $\psi = v_0 - \lambda$, *and* $\kappa = \nu + (v_0 - \lambda)^2$.

Proof. The proof is given in the Appendix. ■

In the symmetric volatility case ($\beta = 0$), the approximation of the implied variance function is indeed symmetric around $x = 0$ where it attains its minimum:[16]

$$I(x|\beta = 0) \approx \lambda + \frac{\nu}{4\lambda^2}\left(\frac{x^2}{T} - \lambda - \frac{\lambda^2}{4}T\right). \tag{4.41}$$

It follows that the curvature of the 'smile' is determined by $\nu/T\lambda^2$. The implied variance function becomes flat when $\nu/T \to 0$, while the smile will be most prominent close to expiration ($T \to 0$), or when ν is large. For $\nu = 0$, regardless of β, we know that $I(x) = \lambda$, which also holds for our approximation. ($\nu \to 0$ implies $v_0 \to \lambda$ and $\kappa \to 0$.)

The approximation for $I(x)$ is seen to share the important features of empirical implied variance. For $\beta \neq 0$ asymmetry is introduced by the asymmetric function $a(x)$. For $\beta < 0$ the minimum indeed shifts to the right. As may be expected, accuracy of the approximation is found to be highest for at- and around the money options, while divergence can be observed for far out of the money and far in the money options.

The analytical properties derived in this subsection show that the location and shape of the implied variance function is sensitive to the choice of φ. If we keep the first two moments of φ fixed but vary a third moment, the implied variance function will generally

[15] Note that $\beta \to 0$ implies $\gamma \to \beta\lambda \to 0$.

[16] This approximation for the symmetric case can also be found in Ball and Roma (1994).

alter its shape. That is not to say that the time-variation in φ can a priori not accurately be described by two degrees of freedom. Higher moments, which largely determine the type of distribution (e.g. Gamma, Inverse Gamma and Inverse Gaussian), may or may not be constant over time.

4.5 Estimation

This first subsection describes the estimation procedure adopted. The second subsection provides a brief evaluation of the goodness-of-fit of the option pricing model.

4.5.1 Loss-function for parametric estimator

For this chapter we consider a parametric estimator for the implied pdf φ. We assume a Generalized Inverse Gaussian (GIG) distribution for φ with three degrees of freedom, which has Gamma, Inverse Gamma and Inverse Gaussian nested as special cases.[17] β will be estimated jointly with φ.

Common loss functions measure either the errors in option prices or the errors in implied volatility.[18] Our experience is that minimizing the errors in option prices is numerically more attractive. Even so, we are keen to obtain an accurate fit of the implied volatility function. As our objective is to estimate the model parameters for each day separately, the number of observations will typically range between 20 and 40.

We choose to minimize the following loss function l:

$$l(\theta) = \sum_i \left(\log\left(\tau + C_{\varphi,i}(\theta)\right) - \log\left(\tau + C_i\right)\right)^2, \tag{4.42}$$

where θ denotes the parameter vector, $C_{\varphi,i}(\theta)$ denotes the model prediction, and where C_i denotes the observed option price at strike price K_i. $\tau > 0$ is included to curb the divergence of $\log(\tau + C)$ for small option prices. We have set $\tau = 5$ in our empirical example, but results are robust to the choice of τ provided it is not too close to zero. (To check whether the numerical optimization procedure converged to a global minimum, and not a local minimum, we repeat the procedure with different initial values.)

[17] Note that the GIG distribution is closed under aggregation, which denotes a convenient property. By averaging variance we will not leave the class of probability density functions.

[18] In view of consistency one may want to employ the same loss function for both estimation and evaluation (see e.g. Christoffersen and Jacobs, 2004).

Each of the model parameters is found to have a distinctive effect on the location and shape of the implied volatility function; the first moment of φ largely controls the vertical allignment. Curvature increases with the dispersion of φ, while β (the degree of asymmetry in volatility) largely governs the location of the minimum and the skewness of the implied volatility function. This should help with the identification of the parameters. It also helps that while the number of observations is limited, they are often found to provide an accurate description of the implied volatility function with little noise.

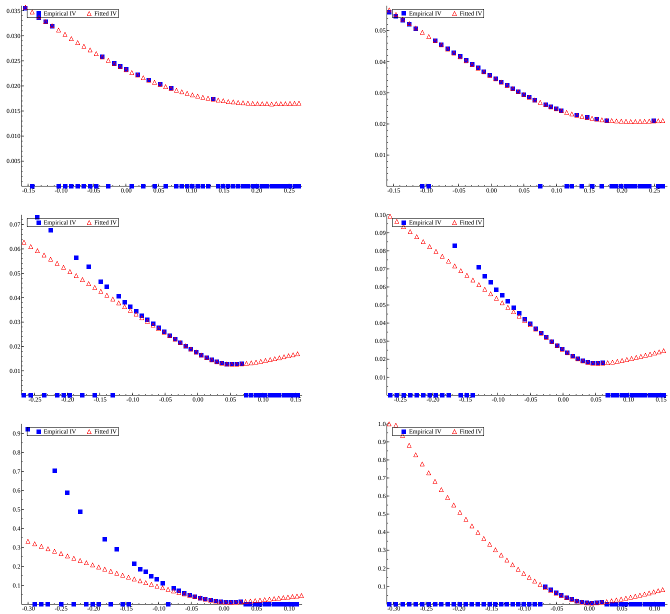

Figure 4.3: *Fitted empirical implied variance for: (a) 1 year (top-left); (b) half a year (top-right); (c) 6 weeks (middle-left); (d) 2 weeks (middle-right); (e) 3 days (bottom-left); and (f) 2 days (bottom-right) to expiration*

4.5.2 Model performance

This subsection provides a first evaluation of model performance. Figure 4.3 shows six fits for different, yet typical, implied variance functions. The time to expiration varies from two days in the top left figure to a year in the bottom right figure. It is stimulating to observe that our model, given its simplicity and analytical convenience, fits the data surprisingly well for all strike prices with longer maturities. For shorter maturities the best fits are observed for options that are not too far out of the money.[19]

When discrepancies between empirical and predicted values are observed, they are mostly at the far left or right end of the figure. By the same token the option price is not very sensitive to volatility for far in and out of the money options, such that the fit of the option price is still very good (which is not shown in the figure). Moreover, the right tail of the implied volatility function is often poorly represented in the empirical data. It refers to high values of the strike price where call option prices are small. Due to the lack of data it will generally be harder to fit this side of the implied volatility function. (Measurement error may also be more of an issue at this end since option prices are not quoted with infinite precision, which matters when prices are very small.) Overall, the model appears well equiped to handle the empirical smiles observed in practice.

4.6 Empirical results

This section shows by means of a modest empirical example how estimates of the implied pdf of volatility φ evolves over time. We assume a GIG distribution for φ with three degrees of freedom. The parameters are estimated jointly with β for each day of the sample. Our focus will be on the first two moments of the implied pdf φ: $E[V]$ and $sd[V]$. The estimates will comprise a time-series of market expectations concerning the distribution of future average variance. These will be compared with the time-series of realized variance.

The time-variation in (φ, β), as we move foward t_0, may reflect both the conditioning on new information over time and the variation in the model h_n (see eq. 4.8) that is adopted by the market. While these are different types of variation, they may well co-exist over time. We will not try to disentangle them, but focus on their joint outcome.

The first panel of Figure 4.4 shows the first moment ($E[V]$) of φ overlayed with realized variance (qv) as well as realized variance averaged over the time left to expiration ($avg[qv]$) for a period of one year with expiration at the end of the time-series. The second panel shows the dispersion of φ measured by the standard deviation ($sd[V]$).

Observation 4.1 *The first moment of the implied pdf of volatility closely matches average realized volatility.*

[19] We arguably need more degrees of freedom for the implied pdf of volatility to obtain a better fit of the implied variance function for shorter maturities.

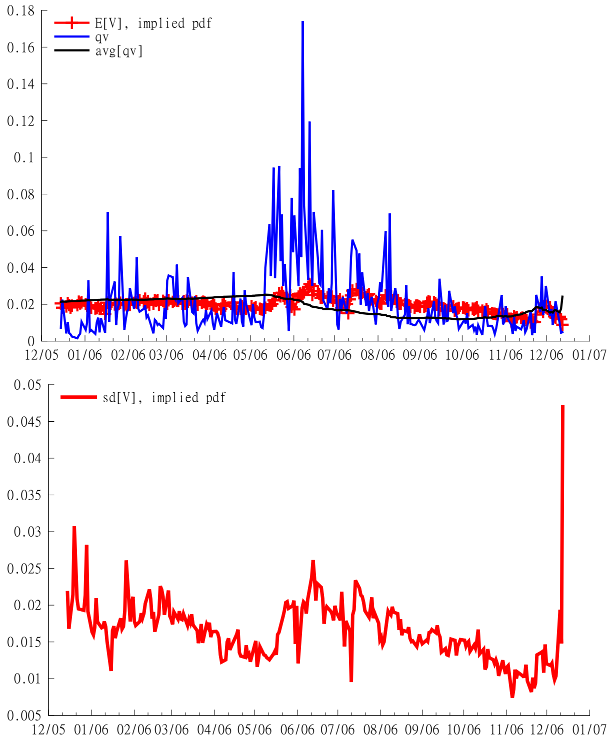

Figure 4.4: *Time-series of: (a) first moment of implied pdf E[V], realized variance qv, average realized variance avg[qv] (top panel), and (b) second moment of implied pdf sd[V] (bottom panel)*

It can be seen how $E[V]$ is particularly close to $avg[qv]$ during the first months of the year. Between six and seven months prior to expiration, realized variance jumps to a higher level and exhibits frequent peaks. $E[V]$ climbs during this period, and remains high for an extended period of time, noticably higher than $avg[qv]$. Expectations are nevetheless steadily reduced as realized variance continues to decline following the turbulent summer of 2006 (June to August). The fact that expectations do not immediately close the gap with average realized variance is consistent with a market contemplating the possibility that the remaining months leading up to expiration may bring a new surge in volatility. In November 2006, with 6 weeks left to expiration, expectations and realizations were fully aligned. Note that a modest rise in volatility emerged in December 2006 that was immediately picked up by the market.

Observation 4.2 *The implied volatility of average volatility (second moment of the implied pdf) does not tend to zero when the time to expiration is increased.*

This observation suggests that the market is fully aware of the fact the volatility exhibits strong persistence (long-range dependence). The variance of average variance is indeed expected not to show a rapid decline when we increase the time to expiration. Whether the subjective degree of volatility persistence matches objective estimates of persistence is an empirical question left for future research. In the bottom panel of Figure 4.4 it can be seen how $sd[V]$ moves up and down over the course of a year. While it increases during the weeks/days prior to expiration (as one would expect), it does not fade away as we move further away from expiration (where the period over which variance is averaged becomes larger).

Observation 4.3 *Market expectations about future average volatility appear to exhibit a degree of foresight.*

Figure 4.5 shows the one year period divided into three parts of about 4 months each. What is interesting is that expectations remain close to future average realized variance even in periods where present day realized variance is noticably lower (best seen in top panel) or noticably higher (best seen in middle panel). This may be an indication that the market is not fooled by current events as it correctly anticipates future levels of average variance.

Figure 4.6 zooms in on the last six weeks. We can see here how expectations respond faster to movements in realized variance closer to expiration. In the first three weeks of this period realized variance is seen to be on the decline, which is closely followed by market expectations, even though future average variance remains steady. Then, three weeks to expiration, realized variance rises to a higher level. This is again immediately picked up by the market which stays on top of movements in volatility until expiration. The bottom panel shows once more how the perceived volatility of average variance increases towards expiration, as one would expect, but that the trend is visibly not exponential.

Observation 4.4 *Estimates of β appear to diverge in the limit where time to expiration tends to zero.*

See Figure 4.7. Note that this occurs when the empirical minimum of the implied variance function does not tend to zero money-ness (i.e. β diverges if we do not have $x^* \to 0$ for

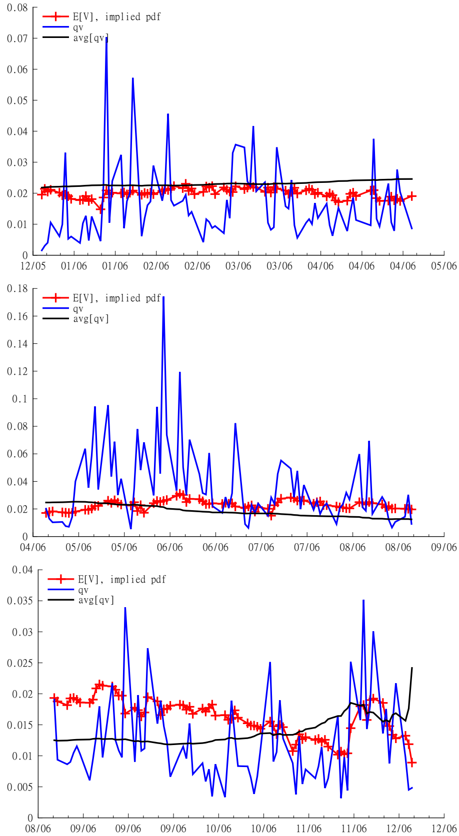

Figure 4.5: *Time-series of first moment of implied pdf E[V], realized variance qv, and average realized variance avg[qv] for: (a) end of Dec 2005 - beginning of May 2006 (top panel), (b) beginning of May 2006 - 3rd week of Aug 2006 (middle panel), and (c) 3rd week of Aug 2006 - 3rd week of Dec 2006 (bottom panel)*

$T \to 0$), see the section on the implied variance function. Whether this is particular for options on the DAX futures, or whether the minimum of the implied variance function does not tend to zero at expiration more generally remains an empirical question. In this case, our option pricing model appears better equipped to fit options with medium to long time to expiration. While our model still provides reasonable fits for options with short maturities, estimates of β in that case become unreasonably large.

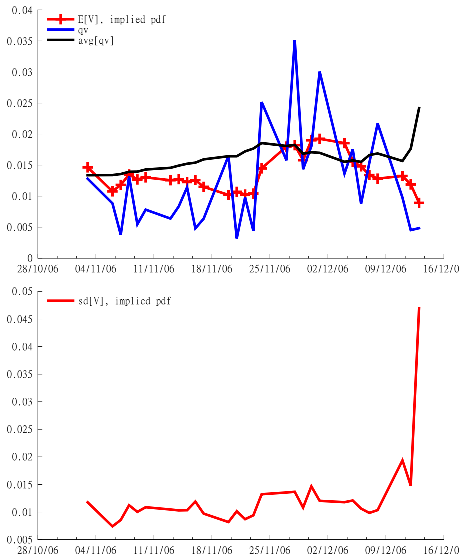

Figure 4.6: *Time-series of: (a) first moment of implied pdf E[V], realized variance qv, average realized variance avg[qv] (top panel), and (b) second moment of implied pdf sd[V] (bottom panel)*

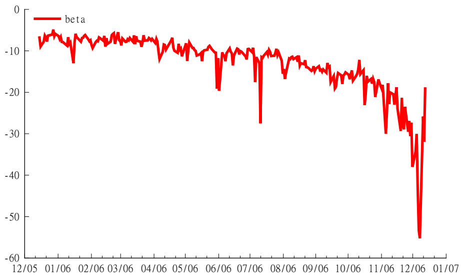

Figure 4.7: *Time-series of* β

4.7 Concluding remarks

The value of European options depends crucially on market expectations about the distribution of future average volatility. The options market may therefore be viewed as a market for volatility.

The objectives of this chapter are three-fold. First, we put forward an option pricing model that accommodates both stochastic and asymmetric volatility, that fits the data well, and preserves analytical convenience. Second, we advocate for the use of the implied pdf of volatility as an alternative to the traditional measure of implied volatility, which was introduced when volatility was still considered a deterministic process. The former is more meaningful in a world where volatility is stochastic. Third, we adopt the option pricing model to estimate the implied pdf of volatility for each day in the sample. The time-series of the implied pdf describes how market expectations about future volatility evolve over time. Data on the German stock index (DAX) is used for the empirical application.

When we compare the first moment of the implied pdf of volatility to estimates of realized volatility, it appears that the market is able to gauge the level of future average volatility quite well. This can perhaps best be seen during periods of low volatility when future volatility is high, and vice versa. Estimates of the second moment of the implied pdf suggest that the market adopts a model that predicts long-range dependence in stochastic volatility. This too is consistent with the stylized facts of realized volatility.

Let us conclude with suggestions for future research. First on the agenda is to test the option pricing model for misspecification using a variety of option price data. Second, given estimates of market expectations over time, a logical next step would be to infer possible learning behaviour. How is the subjective probability distribution of future average volatility updated over time as new observations become available? A better understanding of the interaction between beliefs and observables would allow for more accurate estimation of the implied pdf's, and hence for more accurate option pricing, and hedging.

A third direction for future research is to explore the role of heterogeneity. "A further potential advantage of the discrete time approach is that it permits the introduction of heterogeneous probability assessments across investors and even individual uncertainty as to the parameters of the underlying probability distributions, thus removing the most restrictive feature of the continuous time model which is the assumption that the parameters of the underlying stochastic processes are known with certainty and agreed upon by all investors" (Brennan, 1979). Note, however, that introducing heterogeneity may call for agent based modelling where different demands for option contracts and the underlying stock, due to different expectations, are cleared by some market maker mechanism. For early studies both model uncertainty and heterogeneity of beliefs within an option pricing model see e.g. Buraschi and Jiltsov (2006) and Brock at al. (2009).

4.8 Appendix

Theorem 4.6 *Let $E[V] = \lambda$ and $var[V] = \nu$. In case average variance V is Gamma or Inverse Gaussian distributed, the moments p and q, as defined in eq. (4.32) and (4.33), are given by:*

$$p = (\sqrt{\xi} - \frac{1}{2}) - \beta \tag{4.43}$$

$$q = (\sqrt{\xi} - \frac{1}{2}) + \beta, \tag{4.44}$$

where

$$\xi = (\beta - \frac{1}{2})^2 + \frac{\lambda}{\nu T} c_\varphi. \tag{4.45}$$

with $c_\varphi = 1$ in case of Inverse Gaussian, and $c_\varphi = 2$ in case of Gamma.

Proof. It follows that $z^+ = \sup\{z : E[S_T^z|F_0] < \infty | z > 1\}$ and $z^- = \sup\{z : E[S_T^z|F_0] < \infty | z \leq 0\}$. We have $q = -z^-$ (for $z \leq 0$), and $p = z^+ - 1$ (for $z > 1$). Observe that S_T^z condtional on V (and on F_0) is log-normal distributed, with:

$$E[S_T^z|V] = c \exp\left((\beta - \frac{1}{2})VTz + \frac{1}{2}VTz^2\right) \tag{4.46}$$

$$= c \exp(\tau(z)V), \tag{4.47}$$

where $\tau(z) = (\beta - \frac{1}{2})Tz + \frac{1}{2}Tz^2$. The constant c depends on S_0, and on the model parameters, but not on z. $E[S_T^z] = E_\varphi[E[S_T^z|V]]$ may thus be derived as the expected value of the RHS of eq. (4.47), with variance V as the random variable with pdf φ. Given the functional form of $E[S_T^z|V]$, we have that $E[S_T^z]$ follows directly from the moment-generating-function of φ, which we shall denote $g(t) = E_\varphi[\exp(tV)]$ (with slight abuse of notation, as t is also used to indicate time).

For the Gamma distribution with two parameters θ and k, such that $E[V] = k\theta$ and $var[V] = k\theta^2$, it follows that $g(t) = (1 - \theta t)^{-k}$ for $t < \theta^{-1}$. This implies:

$$E[S_T^z] = c\left(1 - \theta((\beta - \frac{1}{2})Tz + \frac{1}{2}Tz^2)\right)^{-k}. \tag{4.48}$$

Given that $k > 0$, this will tend to infinity when $(\beta - \frac{1}{2})Tz + \frac{1}{2}Tz^2 \to 1/\theta$. Solving this for z yields:

$$z^{+,-} = (\frac{1}{2} - \beta) \pm \sqrt{(\beta - \frac{1}{2})^2 + 2\frac{\lambda}{\nu T}}, \tag{4.49}$$

where $1/\theta = \lambda/\nu$, which implies the followig values for p and q:

$$\begin{aligned} p &= -1 + z^{+} && (4.50)\\ &= (\sqrt{\xi_2} - \frac{1}{2}) - \beta && (4.51)\\ q &= -z^{-} && (4.52)\\ &= (\sqrt{\xi_2} - \frac{1}{2}) + \beta, && (4.53) \end{aligned}$$

where $\xi_2 = (\beta - \frac{1}{2})^2 + 2\frac{\lambda}{\nu T}$.

For the Inverse Gaussian distribution with two parameters μ and l, such that $E[V] = \mu$ and $var[V] = \mu^3/l$, the moment-generating-function is given by:

$$g(t) = \exp\left(\frac{l}{\mu}\left(1 - \sqrt{1 - 2\mu^2 t/l}\right)\right). \tag{4.54}$$

This implies that $E[S_T^z]$ will exist only when $1 - 2\mu^2\tau(z)/l \geq 0$. The values of z for which we border non-existance thus satisfy:

$$\tau(z) = \frac{1}{2}\frac{l}{\mu^2} = \frac{1}{2}\frac{\lambda}{\nu}. \tag{4.55}$$

Solving for z yields:

$$z^{+,-} = (\frac{1}{2} - \beta) \pm \sqrt{(\beta - \frac{1}{2})^2 + \frac{\lambda}{\nu T}}. \tag{4.56}$$

If we define $\xi_1 = (\beta - \frac{1}{2})^2 + \frac{\lambda}{\nu T}$, and derive p and q given the solution for z, we will obtain the expressions stated in the proposition. ■

Theorem 4.7 *Let $E[V] = \lambda$, $var[V] = \nu$, and $\gamma = T^{-1}\log(E[\exp(\beta T V)])$. Without further assumptions about the pdf φ, the implied variance function $I(x)$ can be approximated by:*

$$I(x) \approx \lambda + \frac{\kappa}{4v_0}\left(\frac{x^2}{v_0 T} - 1 - \frac{v_0 T}{4}\right) + \beta\kappa T\left(\beta + \frac{1}{2} + \frac{x}{v_0 T} + (v_0 T)^{\frac{1}{2}}(\beta - \frac{2\psi}{\kappa T})a(x)\right),$$

where

$$a(x) = \frac{\Phi(d_-(x))}{\phi(d_-(x))},$$

with $d_-(x) = \frac{1}{2}(v_0T)^{\frac{1}{2}} - x(v_0T)^{-\frac{1}{2}}$. *The functions* Φ *and* ϕ *denote the standard normal cumulative distribution and density function. Variance level* v_0 *is defined by:*[20]

$$v_0 = \begin{cases} \gamma/\beta & if \quad \beta \neq 0 \\ \lambda & if \quad \beta = 0 \end{cases}$$

The parameters (ψ, κ) *are given by:* $\psi = v_0 - \lambda$, *and* $\kappa = \nu + (v_0 - \lambda)^2$.

Proof. Let us begin with restating the expression for the option price:

$$C_\varphi = \int_0^\infty e^{(\beta v - \gamma)T} C_{BS}(r + (\beta v - \gamma), v)\varphi(v)dv, \tag{4.57}$$

where φ denotes the probability density function for average variance V. Expressed as an expected value, we have:

$$\begin{aligned} C_\varphi &= E[e^{(\beta V - \gamma)T} C_{BS}(r + (\beta V - \gamma), V)] && (4.58) \\ &= E[f(V)]. && (4.59) \end{aligned}$$

Consider the following second-order Taylor expansion for f at $V = v_0$:

$$\begin{aligned} f(V) &\approx f(v_0) + f^{'}(v_0)(V - v_0) + \frac{1}{2}f^{''}(v_0)(V - v_0)^2 \\ &= f(v_0) + f^{'}(v_0)((V - \lambda) - (v_0 - \lambda)) + \frac{1}{2}f^{''}(v_0)((V - \lambda) - (v_0 - \lambda))^2. \end{aligned}$$

If we take expectations over V, with $E[V] = \lambda$ and $var[V] = \nu$, we obtain:

$$E[f(V)] \approx f(v_0) - f^{'}(v_0) + \frac{1}{2}f^{''}(v_0)\left(\nu + (v_0 - \lambda)^2\right), \tag{4.60}$$

which provides a second-order approximation for our option pricing formula C_φ from eq. (4.59).

Let us take $v_0 = \gamma/\beta$ in case $\beta \neq 0$, and $v_0 = \lambda$ otherwise. Note that our price equation features the expectation of the Black-Scholes price $C_{BS}(r(V), V)$ where the interest rate depends on the variance: $r(V) = r + (\beta V - \gamma)$. For our choice of v_0, this reduces to the risk-free interest rate: $r(v_0) = r$. As a result, $f(v_0) = C_{BS}(r, v_0)$.

[20] Note that $\beta \to 0$ implies $\gamma \to \beta\lambda \to 0$.

A first order Taylor approximation of the Black-Scholes price $C_{BS}(I)$ yields:

$$C_\varphi = C_{BS}(I) \approx C_{BS}(r, v_0) + (I - v_0)\frac{\partial C_{BS}}{\partial V}(r, v_0). \tag{4.61}$$

For ease of notation we will supress the indication that the derivatives are evaluated at $V = v_0$ with $r(v_0) = r$. Note that the partial derivatives will be a function of x.

By combining the approximations from eq. (4.60) and eq. (4.61), we obtain:

$$I(x) \approx v_0 - (v_0 - \lambda)f'(v_0)\left(\frac{\partial C_{BS}}{\partial V}\right)^{-1} + \frac{1}{2}\left(\nu + (v_0 - \lambda)^2\right) f''(v_0)\left(\frac{\partial C_{BS}}{\partial V}\right)^{-1}.$$

Let us restate the definition for f: $f(V) = e^{(\beta V - \mu)T} C_{BS}(r + (\beta V - \mu), V)$. Differentiating this with respect to V yields:

$$\frac{\partial f}{\partial V} = e^{(\beta V - \mu)T}\left(\beta T C_{BS} + \frac{\partial C_{BS}}{\partial V} + \frac{\partial C_{BS}}{\partial r}\frac{\partial r}{\partial V}\right). \tag{4.62}$$

It follows that $\frac{\partial r}{\partial V} = \beta$. If we evaluate the derivative at $V = v_0$, we find: $f'(v_0) = \beta T C_{BS} + \frac{\partial C_{BS}}{\partial V} + \beta\frac{\partial C_{BS}}{\partial r}$. By differentiating once more with respect to V, and evaluating the derivative at $V = v_0$, we obtain:

$$\frac{\partial^2 f}{\partial V^2} = \beta^2 T^2 C_{BS} + 2\beta T\frac{\partial C_{BS}}{\partial V} + 2\beta^2 T\frac{\partial C_{BS}}{\partial r} + \frac{\partial^2 C_{BS}}{\partial V^2} + 2\beta\frac{\partial^2 C_{BS}}{\partial r \partial V} + \beta^2\frac{\partial^2 C_{BS}}{\partial r^2}.$$

This would be a good moment to review the 'Greeks', the partial derivatives of the Black-Scholes formula $C_{BS} = S_0\Phi(d_1(x)) - Ke^{-rT}\Phi(d_2(x))$:

$$\begin{aligned}
\frac{\partial C_{BS}}{\partial V} &= \frac{1}{2}T^{\frac{1}{2}}v_0^{-\frac{1}{2}}S_0\phi(d_1(x)) \\
\frac{\partial^2 C_{BS}}{\partial V^2} &= \frac{1}{4}T^{\frac{1}{2}}v_0^{-\frac{3}{2}}S_0\left[\left((v_0T)^{-\frac{1}{2}}x + \frac{1}{2}(v_0T)^{\frac{1}{2}}\right)\phi'(d_1(x)) - \phi(d_1(x))\right] \\
\frac{\partial C_{BS}}{\partial r} &= KTe^{-rT}\Phi(d_2(x)) \\
\frac{\partial^2 C_{BS}}{\partial r^2} &= Ke^{-rT}T^{\frac{1}{2}}v_0^{-\frac{1}{2}}\left(T\phi(d_2(x)) - T^{\frac{3}{2}}v_0^{\frac{1}{2}}\Phi(d_2(x))\right) \\
\frac{\partial^2 C_{BS}}{\partial r \partial V} &= \frac{1}{2}T^{\frac{1}{2}}v_0^{-\frac{1}{2}}\left(\frac{x}{v_0T} - \frac{1}{2}\right)KTe^{-rT}\phi(d_2(x)),
\end{aligned}$$

where $d_2(x) = d_1(x) - (v_0T)^{\frac{1}{2}}$, with $d_1(x) = \frac{1}{2}(v_0T)^{\frac{1}{2}} - x(v_0T)^{-\frac{1}{2}}$. Dividing these derivatives, as well as C_{BS} itself, by $\frac{\partial C_{BS}}{\partial V}$, gives us:

$$C_{BS}\left(\frac{\partial C_{BS}}{\partial V}\right)^{-1} = 2T^{-\frac{1}{2}}v_0^{\frac{1}{2}}\left(\frac{\Phi(d_1(x))}{\phi(d_1(x))} - \frac{\Phi(d_2(x))}{\phi(d_2(x))}\right) \tag{4.63}$$

$$\frac{\partial^2 C_{BS}}{\partial V^2}\left(\frac{\partial C_{BS}}{\partial V}\right)^{-1} = \frac{1}{2}v_0^{-2}\left(\frac{x^2}{T} - v_0 - \frac{1}{4}Tv_0^2\right) \tag{4.64}$$

$$\frac{\partial C_{BS}}{\partial r}\left(\frac{\partial C_{BS}}{\partial V}\right)^{-1} = 2(v_0T)^{\frac{1}{2}}\frac{\Phi(d_2(x))}{\phi(d_2(x))} \tag{4.65}$$

$$\frac{\partial^2 C_{BS}}{\partial r^2}\left(\frac{\partial C_{BS}}{\partial V}\right)^{-1} = 2T\left(1 - (v_0T)^{\frac{1}{2}}\frac{\Phi(d_2(x))}{\phi(d_2(x))}\right) \tag{4.66}$$

$$\frac{\partial^2 C_{BS}}{\partial r\partial V}\left(\frac{\partial C_{BS}}{\partial V}\right)^{-1} = \frac{x}{v_0} - \frac{1}{2}T. \tag{4.67}$$

In these derivations we used that $\frac{\phi(d_1(x))}{\phi(d_2(x))} = e^x$, and $\frac{\phi'(d_1(x))}{\phi(d_1(x))} = -d_1(x)$.

Now define: $\psi = (v_0 - \lambda)$, $\kappa = \nu + (v_0 - \lambda)^2$, and $a_i(x) = \frac{\Phi(d_i(x))}{\phi(d_i(x))}$ for $i = 1, 2$. If we substitute these results into our expression for the approximation of $I(x)$, we find:

$$\begin{aligned} I(x) \approx\ & \lambda + \beta(v_0T)^{\frac{1}{2}}(\beta\kappa T - 2\psi)(a_1(x) - a_2(x)) + 2\beta(v_0T)^{\frac{1}{2}}(\beta\kappa T - \psi)a_2(x) \\ +\ & \left(\beta\kappa T + \beta\kappa(\frac{x}{v_0} - \frac{1}{2}T) + \beta^2\kappa T\right) - \beta^2\kappa T(v_0T)^{\frac{1}{2}}a_2(x) \\ +\ & \frac{1}{4}\kappa v_0^{-1}\left(\frac{x^2}{v_0T} - 1 - \frac{1}{4}v_0T\right). \end{aligned}$$

It can be verified that all terms with $a_2(x)$ cancel each other out. Finaly, the result stated in the proposition can be obtained by rearranging the expression that remains. ■

Chapter 5

Heterogeneity as a natural source of randomness

5.1 Introduction

Though it takes little to realize that expectations about future prices are essential for investment decisions made by market participants, it is difficult to study the evolution of expectations explicitly. Expectations, let alone the beliefs on which they are based, are hardly ever directly observable in practice. For stock markets, for example, one typically observes realized prices, which, at best, reflect market aggregated expectations rather than individual expectations.

Recent results indicate that, to a large extent, the dispersion of beliefs among market participants bears important consequences to the behavior of the aggregates over time. In the recent literature on heterogeneity several theoretical models have been proposed and a number of relations between heterogeneity and stylized facts were derived and validated empirically. Shalen (1993) and Michaely and Vila (1996), for example, concluded that dispersion of beliefs enhances both trading volume and volatility. This supports the empirical evidence reported by Gallant *et al.* (1992) of a positive correlation between volume and volatility. Recent work also indicates that heterogeneity of expectations might be responsible for several other empirical observations. For example, Diether *et al.* (2002) presented evidence that the dispersion of opinions on future returns affects future earnings negatively, while Ziegler (2002) found that heterogeneity of beliefs can lead to the well-known "smile effect" in implied volatility derived from option prices.

Apparently, economic observables depend not only on the average belief, but, more generally, on the distribution of beliefs among market participants. Being both omnipresent and hidden, the dynamics of dispersed beliefs is of our key interest in this chapter. The objective is to model the evolution of beliefs explicitly. While doing so, we wish to keep track of the unpredictable nature of individual preferences and relate it to the unpredictable nature of observables. The latter is important for investigating the extent to which economic observables are affected by uncertainty of choice at the agent level. Together, these objectives motivate the introduction of the concept of a *continuous beliefs system* (CBS). By formulating economic dynamic models in terms of a CBS it is possible to address questions directly related to the beliefs distribution and the endogenous randomness in the dynamics, such as: (i) To what extent does the degree of heterogeneity interact with observables such as prices? (ii) Which aspects of the beliefs distribution matter most for the random dynamics? (iii) Do prices and beliefs evolve similarly in small and large markets? (iv) How do differences in market impact and dependence among agents' choices affect the joint dynamics of prices and beliefs?

Among the first to provide an analytical framework in which agents adapt their beliefs over time are Brock and LeBaron (1996) and Brock and Hommes (1997). The agents' incentives to switch beliefs are provided by the observed differences in past performance such as realized profits. Although, in this setup agents can choose from various beliefs, or expectation functions, the number of strategies available to the agents is finite. Typical applications are concerned with the simple case in which there are only two different belief types (see e.g. Brock and Hommes, 1998). Evidently a small number of belief types is insufficient to obtain a realistic beliefs distribution. Since each fraction of the population associated with a belief type becomes a state variable, using the Brock and Hommes (1997) framework to build a model with a large number of beliefs will typically lead to analytically intractable dynamics.

Recently, Brock *et al.* (2005) considered the price dynamics of a market with a continuum of traders in a so-called large type limit (LTL) where the number of strategies available to the agents tends to infinity. The belief types are drawn at random from a given distribution of possible types. Under fairly general conditions this (random) sequence of deterministic dynamical systems gives rise to unique deterministic limiting dynamics in the LTL. In this way, an LTL provides insights into the dynamics of a large market in the presence of a large number of strategies.

Simply because the limiting deterministic price dynamics was of their main interest, Brock *et al.* (2005) assigned a merely supporting role to the beliefs distribution. The evolution of the beliefs distribution remains more or less hidden because the state variables describing the beliefs distribution are eliminated in the derivation of the price dynamics. Also, by considering a continuum of agents from the start, their approach implicitly assumes that preferences of individual agents are negligible. As a result the concept of an LTL is not very convenient for our purpose: modeling the evolution of beliefs, while taking into account effects of the unpredictable nature of individual preferences. With this in mind, a CBS follows a different route. In contrast with an LTL, which is based on a continuum of agents who can choose from a finite (but increasing) number of strategies, a CBS is based on a continuum of strategies available to any number of agents. The CBS dynamics of the beliefs distribution and the economic observables can be explicitly formulated for any number of agents with arbitrary market weights. This permits several scenarios which are hard to study using an LTL. For example, since the CBS approach does not assume individual preferences to be negligible, it can be used to examine large market (many agent) limits in which the market is dominated by a small number of agents, or where the agents' expectations are strongly correlated.

A CBS contains the following ingredients. The basic concept underlying a CBS is a *beliefs space*, Ω, in which a class of possible point predictors is represented by a continuous (either scalar or vector valued) parameter θ. At each time t agents form beliefs regarding future economic variables by deciding on a single belief parameter θ in the beliefs space. They do so by evaluating the possible strategies, given the information available prior to time t, using a *performance measure*. This performance measure might be based, for example, on the history of past prediction errors, or on profits a strategy would have realized in the past. In each period individual agents choose the strategy which optimizes their expected subjective utility. Due to differences in tastes, they may differ in the strategies used. The central object in a CBS is a time dependent probability density function $\phi_t(\theta)$ on Ω, called the *beliefs distribution*. The belief, (or predictor, or strategy) $\theta_{i,t}$ actually employed by agent i at time t, from the modeler's point of view, is considered to be a random variable distributed according to the beliefs distribution. This can be thought of as representing the empirical fact that even when agents have identical information, they may still have differences of opinion regarding future revenues (see e.g. Frankel and Froot, 1990; Kandel and Pearson, 1995). The specific form of the beliefs distribution given the history of realized past observables, can be expressed in terms of past performances using a continuous choice model. Given the individual beliefs of agents just prior to time t, the individual net demand functions collectively determine the next observable (price) to be quoted publically. This involves some *aggregation mechanism*, such as a market maker setting the price at a zero aggregated net demand value. After the observable becomes public, the performances of strategies can be re-evaluated, a new beliefs distribution arises, etc. From a dynamical point of view, the ongoing evaluation of strategies by agents as new information becomes available results in the co-evolution of the distribution of beliefs and observable aggregates.

In a CBS, the beliefs distribution is a deterministic function of past observables. The strategies used by the agents are modeled as random variables, distributed according to the beliefs distribution, which provides a natural mechanism for endogenous randomness. We refer to this as a 'natural source of randomness', as it can be associated naturally with the uncertainty of choice. In fact this natural source of randomness can help explain part of the market fluctuations we observe in daily life. For example, the excess volatility as observed in financial markets (see Shiller, 1981) might to some extent be 'natural'. In some of the applications presented later it will become clear that most of the nontrivial structure arising from heterogeneous beliefs feedback loops resides in the noise rather than in the deterministic part of the dynamics.

One might expect the randomness to disappear from the dynamics as the number of agents tends to infinity. However, it is not hard to find mild conditions under which the choices of individual agents do not average out to give a deterministic law for the aggregate observables in the limit as the number of agents tends to infinity. For example, if the dependence among the agent's choices is sufficiently large, aggregates may remain random variables even for an infinity of agents. Also, for some combinations of the class of predictors and the performance measure, the conditions for the law of large numbers may not be satisfied, providing another source of randomness which does not vanish if the number of agents tends to infinity.

In general, a CBS gives rise to a Random Dynamical System (RDS) where the randomness carries a natural interpretation. For convenient, but often typical, choices of the performance measure and the class of predictors, the CBS reduces to a low dimensional RDS. The state variables of this RDS are directly related to prices and characteristics of the beliefs distribution, such as the degree of heterogeneity. Since the CBS approach allows stochastic price dynamics to be derived explicitly, it can be used as an analytic alternative to numerical studies based on computationally intensive agent models (for an overview, see LeBaron, 2000).

The remaining sections are organized as follows. In section 5.2 the concepts of a beliefs distribution and the continuous choice model are introduced. Section 5.3 describes a continuous beliefs system (CBS) in which the co-evolution of beliefs and public information such as market prices is taken into account. As an illustration the implied dynamics in a standard asset pricing context is examined by means of some stylized examples in section 5.4. In section 5.5 the mechanism by which endogenous noise arises from the continuous choice dynamics is described, and the role of the number of agents and their dependence considered. Section 5.6 summarizes and discusses the results.

5.2 Continuous beliefs distributions

In this section we describe how continuous beliefs distributions are obtained with the continuous choice model, which can be seen as a natural generalization of the well-known discrete choice model. The discrete version has among others been employed in an economic dynamic context by Brock and Hommes (1997). For clarity of exposition we will discuss both choice models in an agent based economic dynamic setting, starting with the more familiar discrete choice model.

Agent based models represent market participants as (a typically large number of) agents, who can select among a number of alternative strategies. If the strategies among which the agents can choose consists of a finite set of strategies, $s_1, \ldots, s_m$ say, then agents employing strategy s_i, $i = 1, \ldots, m$ are said to be of type i. McFadden (1973) derived an expression for the probability P_i that an individual will select strategy i, starting from the concept of random utility functions. It is assumed that the utility function of agent j can be written as

$$V_j(s) = U(s) + \epsilon_j(s)$$

where $U(s)$ is a non-stochastic "common" utility function representing the tastes of the population, and $\epsilon_j(s)$ is stochastic and reflects the idiosyncrasies of individuals in tastes. The individuals choose the alternative which optimizes their subjective expected utility. Under the assumption that the disturbances of the utility function follow an extreme value distribution, it can be shown that this leads to the *multinomial logit model*:

$$P_i = \frac{\mathrm{e}^{\beta U(s_i)}}{\sum_{l=1}^{m} \mathrm{e}^{\beta U(s_l)}},$$

where $U(s_l)$ is the utility associated with alternative l. The parameter β is referred to as the *intensity of choice*, and is related to the scale of the noise term $\epsilon_j(s)$. The larger the value of β, the smaller the noise, and the larger the probability that an agent chooses the option which actually optimizes $U(s)$. This is why $1/\beta$ is sometimes interpreted as the propensity of agents to err, presuming they actually all wish to optimize $U(s)$.

In the presence of a continuum of belief types it is convenient to introduce a finite dimensional measurable space Ω, containing all possible strategies that can be selected by the agents. We will refer to Ω as the *beliefs space*. Each possible choice, that is, each element θ in Ω uniquely represents a possible strategy agents can choose from. Note that the choice of the beliefs space is not unique, since any one-to-one transformation of the beliefs space Ω into another space, Ω', say, will again yield a suitable beliefs representation. To ensure that integrals over the beliefs space can be defined independently of the chosen representation, we explicitly denote the integration measure on Ω by ν.

In analogy with the discrete choice model, we wish to represent the diversity of belief types by a probability measure over the beliefs space. The distribution of strategies can be obtained from the generalization of the discrete choice model referred to as mixed discrete/continuous choice models. As in the discrete choice setting, it is convenient to adopt a random utility approach (Hanemann, 1984; Dagsvik, 1994; Resnick and Roy, 1994). The random part of the utility function of an agent affects the strategy a particular agent considers optimal. Therefore, the strategies employed by individual agents in a random utility framework are random variables.[1] The continuous choice analogue of the multinomial logit model is the *continuous logit model* (see e.g. Ben-Akiva and Watanatada 1981; Dagsvik, 1994). The probability that an agent selects a strategy in a subset A of Ω is given by

$$P(A) = Z^{-1} \int_A \mathrm{e}^{\beta U(\vartheta)} \, \nu(\mathrm{d}\vartheta) \tag{5.1}$$

with $Z = \int_\Omega \mathrm{e}^{\beta U(\vartheta)} \, \nu(\mathrm{d}\vartheta)$. As in the discrete choice setting, β represents the intensity of choice.

Whenever ν is continuous, a corresponding density $v(\theta)$ exists such that $\nu(\mathrm{d}\theta) = v(\theta)\,\mathrm{d}\theta$. In that case, assuming that $U(\theta)$ is also continuous, the beliefs distribution P has an associated pdf, denoted by $\phi(\theta)$, given by

$$\phi(\theta) = Z^{-1}\mathrm{e}^{\beta U(\theta)} v(\theta), \tag{5.2}$$

[1] Note that random utility does not imply that individual agents perceive their own utility functions to be random, only that they are random to the econometrician.

with $Z = \int_\Omega e^{\beta U(\vartheta)} v(\vartheta) d\vartheta$. For ease of presentation, in the subsequent sections we only consider cases where ν is continuous, and hence $v(\theta)$ exists.

The function $v(\theta)$ is nonnegative, and can be used to put different weights on different parts of the beliefs space. We refer to $v(\theta)$ as the *opportunity function*. In the case where a particular representation of the beliefs space has been fixed, the opportunity function can be thought of as reflecting the *a priori* faith of individuals in parameters within certain regions of the parameter space. Small values of $v(\theta)$ in a certain region then reflect the agents' tendency of avoiding parameter values in that region. For example, if in a particular representation it is reasonable to assume aversion against using extreme strategies, this can be represented by small values of $v(\theta)$ for these strategies. In that case, such "extreme" strategies need to outperform more common strategies to a large extent before agents are likely to use them.

Although derived differently, in the LTL framework (Brock *et al.*, 2005) extensive use is being made of a probability measure with the functional form given by equation (5.1), all market averages in the large type limit being defined with respect to this measure. For an LTL, the measure ν represents the probability measure on Ω from which an increasing set of strategies is chosen at random. However, as it turns out, in the CBS framework it is not necessary for ν to be normalizable. For example, if $\Omega = \mathbf{R}^m$, $v(\theta) = 1$ on Ω and $U(\theta)$ is quadratic in θ with a single maximum, then $\phi(\theta)$ is a multivariate normal probability density function. In general, the probability measure P given in equation (5.1) is a well-defined pdf if and only if $Z = \int_\Omega e^{\beta U(\vartheta)} \nu(d\vartheta)$ is positive and finite, i.e. if $e^{\beta U(\vartheta)}$ is ν-integrable. Clearly, when ν is non-normalizable it can no longer be interpreted as a probability measure. An alternative interpretation is the following. For two disjoint subsets A and B in Ω, with finite integration measures $\nu(A)$ and $\nu(B)$, the ratio $\nu(A)/(\nu(A)+\nu(B))$ can be interpreted as the conditional probability that an agent chooses a parameter value in A, in the absence of any information (U constant), conditional on this choice being either in A or in B.

5.3 Continuous beliefs systems

In this section we discuss the co-evolution of economic observables and the beliefs distribution in a CBS. Next to the beliefs space and the beliefs distribution, two additional ingredients are added. Firstly, we assume that agents evaluate strategies according to some *performance measure*, which might for example be based on past prediction errors, or profits a strategy would have realized in the past, given the information available now (ex post profits). Secondly, a market mechanism, or more generally, an *aggregation mechanism* is required which translates the individual beliefs of agents into publically available information such as prices. For example, in a cobweb framework the price equation might have the form

$$p_t = D^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} S(p_t^e(\theta_{i,t-1}))\right),$$

where $S(p_t^e(\theta_{i,t-1}))$ is firm i's supply and $\frac{1}{n}\sum_{i=1}^{n} S(p_t^e(\theta_{i,t-1}))$ denotes the average supply, and $D(p)$ demand at a given price p.

For clarity of exposition this section is presented against the background of a standard dynamic asset pricing framework. Today's price will be assumed to equal the market's expected present value of tomorrows pay-off. Rather than averaging over functions of beliefs, as in the cobweb framework, prices are conveniently obtained by averaging directly over the different beliefs. The modifications required for applications in more general dynamic settings are fairly straightforward. Throughout, time will be considered to be discrete. For a recently proposed generalization of discrete choice models to continuous time, we refer the interested reader to Dagsvik (2002).

The price of the risky asset at time t will be denoted by p_t. For simplicity we discuss the case where the agents are myopic. Prior to time t, they form expectations about the price of the asset at time $t+1$ (the next time they can sell the asset) including possible dividends payed in the period between time t and $t+1$. The information available to agents just before time t is denoted by $\mathcal{F}_t$. In simple cases the information set could consists of a historic record of past prices up to and including p_{t-1}, but in general it might include exogenous variables, such as the interest rate. The possible strategies from which the agent can choose to predict future prices are represented by a function f_θ of the observables in the information set, parameterized by θ. For a given set of information, we can then consider this predictor as a function of θ. In our examples, the available information on which predictions are conditioned is strictly the observed past price history:

$$p_{t+1}^e(\theta) = f_\theta(p_{t-1}, p_{t-2}, \ldots).$$

Thus, the prediction of p_{t+1} made by agent i, based on price information up to and including p_{t-1}, using strategy $\theta_{i,t-1}$, is denoted by $p_{t+1}^e(\theta_{i,t-1})$.

To give an example of a beliefs space, the class of d-th order linear predictors consists of all predictors of the form

$$p^e_{t+1}(\theta) = \theta_0 + \theta_1 p_{t-1} + ... + \theta_d p_{t-d}.$$

In this case the beliefs are represented in $\mathbf{R}^{d+1}$ and the beliefs distribution is a probability distribution on this space.

The pdf of the time dependent beliefs distribution conditional on $\mathcal{F}_t$ is denoted by $\phi_{t-1}(\theta)$. Schematically, the expectations feedback can then be represented as

$$\ldots \to p_{t-1} \to \phi_{t-1}(\theta) \to p_t \to \phi_t(\theta) \to \ldots.$$

Notice that, regardless of the details of the price formation mechanism, it is possible to investigate how a newly established price p_t affects the beliefs distribution. After p_t becomes part of the information set, agents can re-evaluate the available strategies, and at this point the continuous choice model can be invoked to obtain an expression for the new beliefs distribution:

$$\phi_t(\theta) = Z_t^{-1} v(\theta) \mathrm{e}^{\beta U_t(\theta)} \tag{5.3}$$

the time dependent analogue of equation (5.2).

Typically, we consider cases where the utility function $U_t(\theta)$ is based on performance or fitness measure of the strategies θ, such as last period's net ex post profits or squared prediction errors. We take into account dependence on further past performances by introducing memory in the model. The evolution of the fitness measure for strategy θ can for example be modeled as:

$$U_t(\theta) = \alpha U_{t-1}(\theta) + (1-\alpha)\pi_t(\theta), \tag{5.4}$$

where $\pi_t(\theta)$ is the performance in period t, and $\alpha \in [0, 1)$ is a memory parameter.[2] The utility function then becomes a geometrically weighted sum of ex post performances of strategy θ. An interpretation of the memory parameter α is the following; when analysts or traders consider new strategies it is common to perform back-testing, that is, to test the candidate strategies using historic data. Strategies that would have performed best in the past are more likely to be selected for future trading. The traders are likely to put more

[2]Alternatively one might define $U_t(\theta) = \alpha U_t(\theta) + \pi_t(\theta)$, but this is in fact equivalent, since for $\alpha \in [0, 1)$ this gives a utility differing by a factor $(1-\alpha)$ which can be absorbed by β in the beliefs distribution in equation (5.3).

weight on more recent observations, depending on the time scale on which they perceive their world to be approximately stationary.

By substituting equation (5.4) into equation (5.3) the effect of memory can be written as

$$\begin{aligned} \phi_t(\theta) &\propto v(\theta)\mathrm{e}^{\alpha\beta U_{t-1}(\theta)+(1-\alpha)\beta\pi_t(\theta)} \\ &\propto [v(\theta)]^{1-\alpha}[\phi_{t-1}(\theta)]^{\alpha}\mathrm{e}^{(1-\alpha)\beta\pi_t(\theta)}, \end{aligned} \tag{5.5}$$

which gives an update of the beliefs distribution in terms of the previous beliefs distribution and the last performance measure. Note that in general, since $\phi_t(\theta)$ is an infinite dimensional state variable, this may lead to a very complicated dynamical system. In general the models thus obtained need not be analytically tractable, and to solve the dynamic equations numerical methods may be required. However, in some cases the evolution of $\phi_t(\theta)$ can be completely described by a finite number of variables such as its first k moments, in which case it becomes finite dimensional. This class of models can be used to obtain insights into the interaction of observables and the beliefs distribution analytically. Therefore, in this chapter we will focus on models that are analytically tractable.

Next we consider the formation of the observables p_t, for a finite number, n, of agents, each of which is assigned a strategy according to the previous beliefs distribution $\phi_{t-1}(\theta)$. As mentioned before, some aggregation mechanism is required for determining the next observable from the individual beliefs of the agents. We assume that each predictor implies a unique associated excess demand function, and that given the beliefs of all agents, a unique equilibrium price p_t can be set, such that the market clears. Under the standard assumptions of mean-variance optimization today's market clearing price can be seen to equal the present value of aggregate beliefs concerning future prices and dividends:

$$(1+r)p_t = \frac{1}{n}\sum_{i=1}^{n} p^e_{t+1}(\theta_{i,t-1}) + \bar{y} \tag{5.6}$$

where n is the number of agents, $p^e(\theta_{i,t-1})$ denotes the expected price of individual i based on past observables available up to and including $t-1$, and $\bar{y}$ expected future dividends. Since our interest in the asset pricing model is strictly concerned with its illustrative use, we refer the reader to e.g. Brock and Hommes (1998) for details on the dynamic asset pricing framework.

In interpreting the price equation it is essential to realize that the $\theta_{i,t-1}$ are random variables, representing the individual strategies chosen by the agents. The dynamical system for a finite number of agents is thus stochastic. In the limit where the number of agents tends to infinity, the dynamics can, under certain conditions, become deterministic.

Consider the following assumption:

Assumption 5.1 *(Cross-sectional Independence) The strategies $\theta_{i,t-1}$ employed by agent i at time t, for each fixed time t are independent random variables, distributed according to the population distribution of beliefs prior to time t, $\phi_{t-1}(\theta)$.*

Assuming independence over agents might seem reasonable, since it is always possible to consider expectations of groups of correlated agents as expectations of a single agent representative of this group. The effect of dependence then is merely a reduction in the effective number of agents. A stronger assumption is made if one additionally assumes temporal independence of the idiosyncratic noise terms of each agent over time. This stronger assumption is reasonable if the time interval corresponding to one time step in the model is large compared to the time scale on which idiosyncratic preferences of single agents change over time.

The following theorem is concerned with the almost sure behavior of the model in the limit where the number of agents tends to infinity.

Theorem 5.1 *(Law of Large Numbers) Given $\mathcal{F}_t$, under Assumption 5.1, if n tends to infinity, the aggregate*

$$\bar{p}_t^e = \frac{1}{n}\sum_{i=1}^{n} p_{t+1}^e(\theta_{i,t-1})$$

converges a.s. to

$$E_t\left[p_{t+1}^e(\theta_{i,t-1})\right],$$

if and only if $E_t\left[\left|p_{t+1}^e(\theta_{i,t-1})\right|\right] < \infty$.

Proof: By Assumption 5.1, the strategies $\theta_{i,t-1}$ conditionally on $\mathcal{F}_t$ are IID random variables in Ω, which implies that the corresponding predictions $p_{t+1}^e(\theta_{i,t-1})$ of agents are also IID. The result is immediate from Kolmogorov's strong law of large numbers for the IID random variables $p_{t+1}^e(\theta_{i,t-1})$, given $\mathcal{F}_t$ (see e.g. Resnick, 1998, p. 220). □

Note that the mean expected price, whenever it is finite, can be expressed as

$$E_t\left[p_{t+1}^e(\theta_{i,t-1})\right] = \int_\Omega p_{t+1}^e(\vartheta)\phi_{t-1}(\vartheta)\mathrm{d}\vartheta. \tag{5.7}$$

Theorem 5.1 states that, given the information public at time t, a necessary and sufficient condition for the aggregate expectation about p_{t+1} to converge a.s. to the mean expectation over the beliefs distribution, is the existence of this mean. This result has the following corollary concerning the dynamics.

Corollary 5.1 *(Deterministic Dynamics) Under Assumption 5.1 the observable p_t in the limit $n \to \infty$ almost surely tends to a deterministic function of the information $\mathcal{F}_t$ available at time t if and only if $E_t[p^e_{t+1}(\theta_{i,t-1})] < \infty$.*

This deterministic limiting dynamics can be interpreted as a first order approximation to the price dynamics for a system with a large but finite number of agents. Later it will become clear that the random deviations around the deterministic limit in general can not be represented by a random variable with constant (state-independent) distributional properties.

Notice that discrete choice models can be considered as special cases of a CBS (when allowing for discrete opportunity distributions) in which agents can only choose among a finite number of alternative strategies θ_l, $l = 1, \ldots, m$. Provided that the expected future prices of each of those strategies are finite, the average expected price is well-defined, so that the strong law of large numbers applies, and the dynamics converges to a deterministic dynamical system with probability one as the number of traders tends to infinity. A discrete choice setting with a continuum of agents, as considered by Brock and Hommes (1997), should thus always lead to functional determinism in the absence of exogenous noise provided that $p^e_{t+1}(\theta_l)$ is finite for each $l = 1, \ldots, m$.

At this point let us briefly compare equation (5.7) with it's LTL analogue. The population mean of the predictor function $p^e_{t+1}(\theta)$ can be written as

$$E_t[p^e_{t+1}(\theta)] = \int_\Omega p^e_{t+1}(\vartheta)\phi_{t-1}(\vartheta)\mathrm{d}\vartheta = \frac{\int_\Omega \mathrm{e}^{\beta U_{t-1}(\theta)} p^e_{t+1}(\vartheta)v(\vartheta)\mathrm{d}\vartheta}{\int_\Omega \mathrm{e}^{\beta U_{t-1}(\theta)} v(\vartheta)\mathrm{d}\vartheta}. \tag{5.8}$$

This expression can be seen to be of the same analytic form as the population average given by Brock *et al.* (2005). An important difference, however, is that we do not require $v(\theta)$ to integrate to one. That is, $v(\theta)$ itself is not required to be normalizable, and instead we only require $\phi_{t-1}(\theta)$ to be be a well-defined pdf (i.e. $\mathrm{e}^{\beta U_{t-1}(\theta)}v(\theta)$ should be integrable). In this sense, equation (5.8) provides a generalization of the analogous expression obtained with the LTL approach. This indicates that an LTL has an associated CBS, but that the converse does not always hold. In fact some of the example CBSes that will be discussed later use a non-normalizable function $v(\theta)$, and hence do not have an LTL representation. Furthermore, even if the beliefs distribution is well-defined, the mean expected price given in equation (5.7) need not exist. In that case the dynamics does not allow for a deterministic LTL, but as will be shown later, the associated CBS might still be meaningful, with the price being a random variable. While an LTL always gives rise to a deterministic dynamical system, a CBS gives rise to an RDS, containing deterministic dynamics as a special case.

Before moving to a concrete example, note that we only consider point predictors in this chapter. However, some price formation mechanisms can not be formulated in terms of point predictors only. For example, Guesnerie (2002) considers agents who have subset predictors (e.g. an interval) in mind rather than point predictors. Under certain conditions, the agents, by a rationality assumption and a common knowledge argument, can find agreement on a unique trading price. Such a mechanism could be incorporated in a CBS framework, provided that the more general predictors can be represented by a finite number of parameters, and the price agents eventually agree to trade on can be expressed explicitly in terms of the agents' individual beliefs parameters.

5.4 Example: Asset pricing with first order linear beliefs

We consider an example of a CBS in which agents believe that first order linear rules provide adequate predictions. The role of the performance measure is investigated by focusing on two cases: squared prediction errors and squared logarithmic prediction errors, leading to normal and log-normal beliefs distributions respectively.

For illustrational purposes we focus on a simple class of linear predictors. Agents choose a strategy θ that represents their perceived future growth rate. The expected price at time $t+1$ associated with predictor θ is then

$$p^e_{t+1}(\theta) = \theta p_{t-1}. \tag{5.9}$$

The utility function associated with belief type θ is given by equation (5.4). The price equation, (5.6), in this particular case becomes, in the limit of infinitely many agents:

$$\begin{aligned} (1+r)p_t &= \int \vartheta p_{t-1}\phi_{t-1}(\vartheta)\,\mathrm{d}\vartheta + \bar{y} \\ &= \mu_{t-1}p_{t-1} + \bar{y}, \end{aligned} \tag{5.10}$$

where μ_{t-1} represents the average belief parameter at time $t-1$. In the next subsections we consider the updating of the beliefs distribution for two different performance measures.

Throughout we assume that the parameters satisfy the following conditions: $\alpha \in [0,1)$, $r \in [0,\infty)$ and $\beta \in [0,\infty)$.

5.4.1 Updating according to squared prediction errors

First we examine the case where $\Omega = \mathbf{R}$, and the performance measure is minus the squared prediction error:

$$\pi_t(\theta) = -\left(p_t^e(\theta) - p_t\right)^2 = -\left(\theta p_{t-2} - p_t\right)^2. \tag{5.11}$$

After the price p_t has been realized and observed, the distribution of beliefs is updated according to the continuous choice model. The new distribution describing the dispersion of belief types is then given by equation (5.5). Upon substitution of the expression for $\pi_t(\theta)$ into equation (5.5), we obtain:

$$\phi_t(\theta) \propto \left[\phi_{t-1}(\theta)\right]^\alpha \exp\left[-\beta(1-\alpha)\left(\theta p_{t-2} - p_t\right)^2\right]. \tag{5.12}$$

Note that we have chosen a constant opportunity function, i.e. $v(\theta) = 1$, representing the simplest case where agents hold no aversion against extreme parameter values.[3]

Since the exponent contains only up to second order forms in θ with a negative coefficient for the quadratic term in θ, the distribution of beliefs in each period can be described by a normal distribution. Here we have implicitly assumed that $\phi_{t-1}(\theta)$ is also normal, which can be justified by assuming that the dynamics has been running since the infinite past.

If we denote the mean and variance of $\phi_t(\theta)$ by μ_t and σ_t^2 respectively, we can write

$$\phi_t(\theta) = \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left[-\frac{1}{2\sigma_t^2}\left(\theta - \mu_t\right)^2\right]. \tag{5.13}$$

Together with equation (5.12) this gives

$$\phi_t(\theta) \propto \exp\left[-\frac{\alpha}{2\sigma_{t-1}^2}(\theta - \mu_{t-1})^2 - \beta(1-\alpha)(\theta p_{t-2} - p_t)^2\right]. \tag{5.14}$$

By comparing the coefficients of θ^2 and θ in the exponents in equations (5.13) and (5.14),

[3] In fact, since $v(\theta) = 1$ is not normalizable, this is an example of a CBS which can not be obtained directly using the LTL approach. However, using the LTL approach, one might arrive at the CBS dynamics by taking $v(\theta)$ to be the pdf of a $N(0, s^2)$ distributed random variable, and taking the limit where s^2 tends to infinity.

the mean belief parameter μ_t, and the variance of the belief parameter, σ_t^2, can be seen to evolve according to

$$\begin{aligned}\frac{\mu_t}{\sigma_t^2} &= \alpha\frac{\mu_{t-1}}{\sigma_{t-1}^2} + 2\beta(1-\alpha)p_t p_{t-2} \\ \frac{1}{\sigma_t^2} &= \frac{\alpha}{\sigma_t^2} + 2\beta(1-\alpha)p_{t-2}^2.\end{aligned}$$

The remaining factor, which is independent of θ, is absorbed into the normalization factor Z_t. Notice that the average belief parameter μ_t and belief variance σ_t^2 completely describe the evolution of beliefs and are fully determined by past prices.

The full deterministic CBS, which also includes the price equation given in equation (5.10), becomes:

$$\begin{aligned}(1+r)p_t &= \mu_{t-1}p_{t-1} + \bar{y} \\ \frac{\mu_t}{\sigma_t^2} &= \alpha\frac{\mu_{t-1}}{\sigma_{t-1}^2} + 2\beta(1-\alpha)p_t p_{t-2} \\ \frac{1}{\sigma_t^2} &= \alpha\frac{1}{\sigma_{t-1}^2} + 2\beta(1-\alpha)p_{t-2}^2.\end{aligned} \tag{5.15}$$

The fixed point, which is the solution of $(p_t, \mu_t, \sigma_t^2) = (p^*, \mu^*, \sigma^{2*})$, is given by

$$(p^*, \mu^*, \sigma^{2*}) = \left(\frac{\bar{y}}{r}, 1, \frac{1}{2\beta}\right).$$

In section 5.4.3 the local stability properties of the dynamics near the fixed point are examined.

5.4.2 Updating according to squared logarithmic prediction errors

Next we consider the case with $\Omega = \mathbf{R}^+$, with a performance measure which taken to be minus the squared logarithmic prediction error:

$$\pi_t(\theta) = -\left(\ln p_t^e(\theta) - \ln p_t\right)^2 = -\left(\ln\theta - \ln p_t + \ln p_{t-2}\right)^2. \tag{5.16}$$

An argument for using logarithmic prediction errors rather than just mean squared prediction errors is that this error measure is independent of the price level. This leads to dynamics which scales with the price level and is equivalent before and after possible stock splits.

Using a uniform opportunity function again, i.e. $v(\theta) = 1$, substitution of the expression for $\pi_t(\theta)$ as in equation (5.16) into equation (5.5) gives:

$$\phi_t(\theta) = Z_t'^{-1}[\phi_{t-1}(\theta)]^\alpha \exp\left[-\beta(1-\alpha)\left(\ln\theta - \ln p_t + \ln p_{t-2}\right)^2\right], \tag{5.17}$$

where again Z_t' is a normalization factor independent of θ, but not necessarily equal to Z_t in equation (5.3).

In this case the exponent contains up to second order forms in $\ln\theta$ with a negative coefficient for the quadratic term in $\ln\theta$, which indicates that the distribution of beliefs in each period can be described by a log-normal distribution of the form

$$\phi_t(\theta) = \frac{1}{\sqrt{2\pi}\sigma_t\theta}\mathrm{e}^{-\frac{(\ln\theta-\mu_t)^2}{2\sigma_t^2}} = \frac{1}{\sqrt{2\pi}\sigma_t}\mathrm{e}^{-\frac{(\ln\theta-\mu_t)^2}{2\sigma_t^2}-\ln\theta}. \tag{5.18}$$

The evolution of the beliefs distribution follows from combining equations (5.17) and (5.18), which results in

$$\phi_t(\theta) \propto \exp\left[-\alpha\ln\theta - \frac{\alpha}{2\sigma_{t-1}^2}(\ln\theta-\mu_{t-1})^2 - \beta(1-\alpha)(\ln\theta - \ln p_t + \ln p_{t-2})^2\right]. \tag{5.19}$$

A comparison of the coefficients of $\ln\theta$ and $(\ln\theta)^2$ in the exponents in equations (5.18) and (5.19), gives the evolution rules for μ_t and σ_t^2, and the following deterministic CBS is obtained:

$$\begin{aligned}(1+r)p_t &= \exp\left[\mu_{t-1} + \frac{1}{2}\sigma_{t-1}^2\right]p_{t-1} + \bar{y} \\ \frac{\mu_t}{\sigma_t^2} &= \alpha\frac{\mu_{t-1}}{\sigma_{t-1}^2} + (1-\alpha)\left(2\beta(\ln p_t - \ln p_{t-2}) + 1\right) \\ \frac{1}{\sigma_t^2} &= \frac{\alpha}{\sigma_{t-1}^2} + 2\beta(1-\alpha).\end{aligned} \tag{5.20}$$

Note that σ_t^2 does not interact with the other variables, and simply tends to its steady state value $1/(2\beta)$ at an exponential rate. Also note that the state-variables $\{\mu_t, \sigma_t^2\}$ no longer denote the mean and variance of the beliefs distribution. In the limit of an infinite number of agents, the average belief becomes

$$\int \vartheta\phi_{t-1}(\vartheta)\,\mathrm{d}\vartheta = \exp\left[\mu_{t-1} + \frac{1}{2}\sigma_{t-1}^2\right],$$

while the degree of heterogeneity as measured by the variance, is given by:

$$\mathrm{Var}_t[\theta] = \exp\left[2\mu_{t-1} + \sigma_{t-1}^2\right]\left(\exp\left[\sigma_{t-1}^2\right] - 1\right). \tag{5.21}$$

It can be verified that the fixed point solution of the system is given by:

$$(p^*, \mu^*, \sigma^{2*}) = \left(\frac{\bar{y}}{1 + r - \exp\left(\frac{3}{4\beta}\right)}, \frac{1}{2\beta}, \frac{1}{2\beta}\right).$$

Note that for all finite positive values of β the fixed point price p^* is larger than the fundamental value $\bar{y}/r$. Only when β tends to infinity (i.e. the propensity to err tends to zero) the fixed point price tends to the fundamental price.

Remark: The examples discussed here and in the previous subsection are analytically tractable due to the quadratic terms in the exponents. A closed form analytic derivation of the dynamics might become cumbersome or even impossible if: (i) the beliefs distributions can not be represented within a finite parameter class, closed under updating according to equation (5.5) (it is closed under updating e.g. if $\pi_t(\theta)$ is a finite order polynomial in θ), or (ii) agent's decisions are based on more complicated optimization procedures such as those typical in a dynamic programming contexts. In those cases one might proceed using appropriate generalization of methods known in discrete choice simulation and estimation (see e.g. Keane and Wolpin, 1994). These generalizations are beyond the scope of this chapter and left for future research.

As noted above, we consider models that allow for an analytic derivation of the dynamics, although perhaps exceptional, important from a theoretical point of view, since they may help providing insights into the interaction between prices and the beliefs distribution. In the models derived above one can explicitly see how prices are formed given the beliefs distribution, and how realized prices in turn affect both the census and the dispersion of beliefs.

5.4.3 Local stability

The deterministic dynamical systems derived in the previous subsections, provide analytical descriptions of the co-evolution of both prices and the distribution of beliefs. Although our main motivation for considering these stylized examples is to illustrate how a CBS can be used to obtain insights into the interaction of prices and the beliefs distribution, these simple models also enable one to investigate how the dynamics is affected by the model parameters. For example, how does the steady state and its stability depend on behavioral parameters such as the intensity of choice β and the degree of memory α?

We present a brief bifurcation analysis of the deterministic skeletons just derived. However, it should be realized that the value of such an analysis strictly speaking is limited to the deterministic case only. Since we wish to illustrate the CBS methodology conceptually without putting too much stress on particular cases we choose to limit ourselves to a discussion on the local stability around the fixed points. The price equation, (5.10), suggests that increasing the interest rate has a stabilizing effect on the dynamics. Intuitively, also the memory parameter is important for stability, since it has a smoothing effect. The local bifurcation analysis is carried out mainly to check whether this intuition is justified.

The next proposition, which is proved in the Appendix, gives the local stability conditions:

Proposition 5.1 *For $\alpha \in [0,1)$ and $\beta \in [0,\infty)$ and $r \in [0,\infty)$ the CBSes given in equations (5.15) and (5.20) are locally stable around the fixed point if $g(\alpha,a) > 0$, where*

$$g(\alpha,a) = 1 - 3a^2 - 2a\alpha + 5a^2\alpha + a\alpha^2 - 2a^2\alpha^2,$$

with $a = (1+r)^{-1}$ for the squared prediction error case (equation 5.15) and $a = (1+r)^{-1}\exp\left(\frac{3}{4\beta}\right)$ for the squared logarithmic prediction error case (equation 5.20).

Figure 5.1 shows the line $g(\alpha,a) = 0$ where the bifurcation occurs. The dynamics is locally stable around the fixed point for parameter values below (right) of this bifurcation curve, and unstable for parameter values above (left) of the bifurcation curve. The memory parameter can be seen to have a stabilizing effect, provided that a is not too large. For any value of α in $[0,1)$ the dynamics can be made locally stable by increasing the interest rate (decreasing a) sufficiently. In contrast with the squared prediction error case, with squared logarithmic prediction errors the intensity of choice, β, also affects stability, since stability is determined, apart from α, by $a = (1+r)^{-1}\exp\left(\frac{3}{4\beta}\right)$ rather than r only. In the limit where the memory parameter tends to one, the equilibrium will lose its stability once the interest rate falls below $r = \exp\left(\frac{3}{4\beta}\right) - 1$. Thus in the squared logarithmic prediction error case, the interest rate can be very low while sustaining stability, as long as the intensity of choice is correspondingly large.

It is possible to show that the points on the interior of the primary bifurcation curve (i.e. for $\alpha \in (0,1)$) for both models correspond to Neimark-Sacker bifurcations. On the bifurcation curve at least one eigenvalue crosses the unit circle. Assuming that one eigenvalue

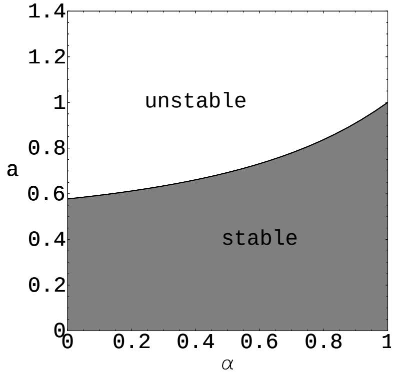

Figure 5.1: *Bifurcation curve in the (α, a)-plane. The dynamics of the CBSes described by equations (5.15) and (5.20) are locally stable for parameter values below (right) of this curve, and unstable for parameter values above (left) of it. The parameter a is related to the model parameters through $a = (1 + r)^{-1}$ for the squared prediction error case (equation 5.15) and $a = (1+r)^{-1} \exp\left(\frac{3}{4\beta}\right)$ for the squared logarithmic prediction error case (equation 5.20).*

crosses the unit circle at a real value (i.e. ± 1) leads to $\alpha = 0$ or $\alpha = 1$, i.e. the boundaries of the allowed values for α, since we required $\alpha \in [0, 1)$. It follows that on points in the interior of the bifurcation line the dynamics become unstable due to a complex conjugated pair of eigenvalues crossing the unit circle.

As a final remark on the bifurcations of the two models discussed above, it should be mentioned that numerical evidence indicates that squared prediction errors give rise to a supercritical bifurcation while squared logarithmic prediction errors lead to a subcritical bifurcation. Using squared prediction errors the stable fixed point breaks up into a quasiperiodic motion on a (topological) circle, which increases in size if the memory parameter is lowered, and eventually breaks up giving way to globally unstable dynamics. In the squared logarithmic prediction error case case, when one crosses from the stable to the unstable region in the parameter space, there is no stable quasiperiodic solution close to the fixed point. In fact, numerically we have only seen cross-overs from locally stable to globally unstable dynamics in the squared logarithmic prediction error case.

5.5 Natural sources of randomness

Traditionally, randomness in economic models is associated with exogenous shocks, for example due to news affecting a company's future earnings, or fluctuations of the interest rate. The CBS framework provides several possible natural sources of randomness, which enter via the market aggregate and can be associated with the uncertainty of choice.

In the previous section we have shown how the CBS approach can be used to derive deterministic dynamical systems in the presence of an infinity of agents. So far, two conceptual differences with the LTL approach (Brock *et al.*, 2005) became apparent: i) the CBS dynamics are expressed not only in terms of prices, but also in terms of variables which fully describe the distribution of beliefs among agents, and ii) the measure ν in the beliefs space need not be normalizable for a CBS. In fact, both examples in the previous section are based on non-normalizable measures ν, and hence cannot be obtained directly as an LTL.

In this section we exploit another important difference between the LTL and CBS approaches. Recall that the LTL approach starts with a continuum of agents, and examines the limiting dynamics as the number of possible strategies tends to infinity. In the CBS approach the starting point is a continuum of strategies. There the behavior of large markets can be studied by examining limits where the number of agents tends to infinity. Because dependence among agents' choices and heterogeneous market impact are not excluded *a priori*, because we avoided the assumption of a continuum of agents, large market limits under various conditions can be considered. As it turns out, for those scenarios the randomness associated with individual choices typically do not average out in the large market limit.

Several possible natural sources of randomness can be identified, which will be considered in the sequel. Firstly, a finite number of traders gives rise to stochasticity, because traders are assigned a belief $\theta_{i,t-1}$ at random from the beliefs distribution $\phi_{t-1}(\theta)$. Secondly, so far, we have implicitly assumed that no agents have a dominant market impact. This need be the case, for example, if wealth is not evenly distributed among agents. In that case the market impact of the wealthiest agents might still be important as the number of traders tends to infinity. Thirdly, dependence among agents can prevent the dynamics from becoming deterministic when the number of agents tends to infinity. If, for example, agents coordinate (at least partly) on a random variable, such as an exogenous variable, or on predictors announced by some 'leading' agents who quote their predictors publically in an early stage, then the stochastic properties of that variable show up in the price dynamics. In the extreme case where all agents have identical idiosyncrasies, they would all use the same (random) strategy. Finally, for certain combinations of the utility function and the predictor function $p^e_{t+1}(\theta_{i,t-1})$, the law of large numbers may not apply because the mean of the aggregate need not exist. In those situations, the price dynamics might still be defined as a stochastic dynamical system, in the limit where the number of agents tends to infinity.

5.5.1 Finite number of agents

In general, for a finite number of agents, the CBS approach leads to a random dynamical system, the only exceptions being cases where all $p^e_{t+1}(\theta_{t,i-1})$ are identical with probability one. In general it might not be straightforward to derive the distributional properties of the aggregate. However, in case the mean and variance of the aggregate are finite, one can appeal to the central limit theorem which states that the distribution of the aggregate is asymptotically normally distributed. The following theorem is a direct application of the central limit theorem.

Theorem 5.2 *(Central Limit Theorem) Given $\mathcal{F}_t$, under Assumption 5.1, if n tends to infinity, the random variable*

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} p^e_{t+1}(\theta_{i,t-1}) - E_t[p^e_{t+1}(\theta_{i,t-1})]\right)$$

converges in distribution to

$$W \sim N(0, \sigma_t^2)$$

with

$$\sigma_t^2 = \mathrm{Var}_t[p^e_{t+k}(\theta_{i,t-1})]$$

if and only if $E_t[|p^e_{t+1}(\theta_{i,t-1})|] < \infty$ *and* $\mathrm{Var}_t[p^e_{t+1}(\theta_{i,t-1})] < \infty$.

Example: Logarithmic prediction errors

In this example we consider the asset pricing example discussed in section 5.4.2, with a performance measure based on logarithmic squared prediction errors. We assume that agents make decisions regarding their strategy independently of the other agents (Assumption 1). The predictors $p^e_{t+1}(\theta)$ are linear in the parameter θ, and the price p_t follows from equations (5.6) and (5.9):

$$(1+r)p_t = \frac{1}{n}\sum_{i=1}^{n} p^e_{t+1}(\theta_{i,t-1}) + \bar{y}.$$

The conditional variance of $\theta_{i,t-1}$, given $\mathcal{F}_t$, is equal to (see equation 5.21):

$$S_t^2 = \mathrm{Var}_t[\theta_{i,t-1}] = \mathrm{e}^{2\mu_{t-1}+\sigma^2_{t-1}}(\mathrm{e}^{\sigma^2_{t-1}} - 1),$$

so that the conditional variance of $(1+r)p_t = \frac{1}{n}p^e_{t+1}(\theta_{i,t-1}) + \bar{y}$ equals $\frac{1}{n}S_t^2 p_{t-1}^2$. With the asymptotic normal approximation given by Theorem 2, the price equation for a finite number, n, of agents becomes

$$(1+r)p_t \simeq \left(\exp\left[\mu_{t-1} + \frac{1}{2}\sigma^2_{t-1}\right] + \frac{S_t}{\sqrt{n}}\epsilon_t\right) p_{t-1} + \bar{y},$$

with $\epsilon_t \sim N(0,1)$. If the number of agents, n, tends to infinity, the noise variance tends to zero and the price equation becomes deterministic. The full dynamics in the large number of agents tends to the 'deterministic skeleton' described by equation (5.15). Notice that the analogous asymptotic normal approximation is exact for the squared prediction error case, even for a finite number of agents n, since there the $\theta_{i,t}$ are normally distributed (see section 5.4.1). Also note the emergence of conditional heteroscedasticity. The time-varying variance $\frac{1}{n}S_t^2 p_{t-1}^2$ of the noise term is proportional to the evolving degree of heterogeneity, S_t^2.

5.5.2 Market impact

Differences in market impact, which may be driven by wealth differences, among agents can lead to the persistence of randomness, even for infinitely many agents. For simplicity we only consider the effect of an unequal market impact without considering the endogenous evolution of wealth, since the latter is beyond the scope of this chapter. Endogenous wealth effects have for example been studied by Cabrales and Hoshi (1996) in a discrete choice framework with two types of agents.

Let us denote the market weight of agent i by w_i, normalized such that $\sum_i w_i = 1$. The mean expected future price becomes

$$\bar{p}^e_{t+1} = \sum_{i=1}^{n} w_i p^e_{i,t+1}$$

which has a conditional mean equal to

$$E_t\left[\sum_{i=1}^{n} w_i p^e_{i,t+1}\right] = E_t[p^e_{i,t+1}]$$

and, assuming conditional independence of $\theta_{i,t-1}$ given $\mathcal{F}_t$, conditional variance

$$\mathrm{Var}_t\left[\sum_{i=1}^{n} w_i p^e_{i,t+1}\right] = \left(\sum_{i=1}^{n} w_i^2\right) \mathrm{Var}_t[p^2_{i,t+1}].$$

The term between brackets on the right hand side of this equation, is known as the Herfindahl index of concentration. If all agents have equal weights, $w_i = \frac{1}{n}$, the conditional variance is proportional to one over n. For all other weight distributions the conditional variance is larger. A large concentration (large Herfindahl index) thus implies a small effective number of market participants and vice versa. The conditional variance suggests defining an effective number of agents, n_{eff}, say, in terms of the inverse of the Herfindahl index:

$$n_{\text{eff}} = \left(\sum_{i=1}^{n} w_i^2\right)^{-1}.$$

The effective number of agents thus defined represents the number of agents that would give rise to the derived conditional variance in case their market impacts would be equal.

5.5.3 Interaction and dependence among agents

As a next source of randomness that can persist if the number of agents tends to infinity, we consider dependence among agents. Several types of dependence can be incorporated in the utility function by adding an interaction term, as in Brock and LeBaron (1996) and Brock and Durlauf (2001). In this approach individual utility depends not only on the past performance of strategies, but also on the choices made by other agents. The joint pdf of the beliefs parameters of n agents is written as

$$\phi_t(\theta_1, \ldots, \theta_n) \sim \exp\left(\sum_{i=1}^{n} \beta U_t(\theta_i) - \frac{J}{2}\sum_{i,j} h(\theta_i, \theta_j)\right),$$

where $u_t(\theta)$ denotes the performance measure, and $\frac{J}{2}h(\theta_i, \theta_j)$ is a function which captures the additional utility derived from interactions. The degree of dependence is measured by $J > 0$. Although formally the θ_i depend on time, the subscript is dropped here for simplicity.

Following Brock and LeBaron (1996) we consider a utility function which depends on squared prediction errors, and an interaction term that specifies that each agent prefers parameter values close to the consensus. One obtains

$$\phi_t(\theta_1, \ldots, \theta_n) \sim \exp\left(-\sum_{i=1}^{n} \frac{(\theta_i - \mu_t)^2}{2\sigma_t^2} - \frac{J}{2}\sum_{i=1}^{n}(\theta_i - \bar{\theta})^2\right), \qquad (5.22)$$

where $\bar{\theta} = \frac{1}{n}\sum_{i=1}^{n}\theta_i$. Equation (5.22) contains linear and quadratic terms in the parameters θ_i and hence represents a multivariate gaussian pdf. Upon comparing this with that of a multivariate normal

$$\frac{1}{(2\pi)^{n/2}\|\Sigma\|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \Sigma^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right),$$

where Σ is the variance covariance matrix of the θ_i, the statistical properties of the market average $\bar{\theta} = \frac{1}{n}\sum_{i=1}^{n}\theta_i$ can be derived. As in the case without interaction the expected value of the market average is given by $E_t[\bar{\theta}] = \theta_0 = \mu_t$. Due to permutation symmetry of the agents all diagonal elements of Σ are identical, as well as the off-diagonal elements. The same holds for Σ^{-1}. This implies that both Σ and Σ^{-1} have $\mathbf{1}$ as an eigenvector. We denote the corresponding eigenvalue of Σ by λ, so that the corresponding eigenvalue of Σ^{-1} is $1/\lambda$. For the elements of Σ^{-1} we find

$$(\Sigma^{-1})_{i,i} = \frac{1}{\sigma_t^2} + 2\frac{n-1}{n}J,$$

and

$$(\Sigma^{-1})_{i,j} = \frac{J}{n} \qquad \text{for } i \neq j.$$

The eigenvalue λ can now be determined from:

$$\Sigma^{-1}\mathbf{1} = \frac{1}{\lambda}\mathbf{1} = \left(\frac{1}{\sigma_t^2} + \frac{(n-1)}{n}J\right)\mathbf{1},$$

giving $\lambda = \frac{1}{\sigma_t^2} + \frac{(n-1)}{n}J$. For the variance of $\bar{\theta}$ one then finds

$$\text{Var}[\bar{\theta}] = \frac{1}{n^2}\text{Var}\left[\sum_{i=1}^{n}\theta_i\right] = \frac{1}{n^2}\mathbf{1}^T\Sigma\mathbf{1} = \frac{\lambda}{n} = \frac{1}{\frac{n}{\sigma_t^2} + (n-1)J},$$

which indicates that the interaction reduces the variance of the market average $\bar{\theta}$. Thus, when agents coordinate on the population mean, meaning that no agent prefers to be far away from the consensus, the variance of the population mean $\bar{\theta}$ reduces. Notice that the mean market expectation $E_t[\bar{\theta}] = \mu_t$ is identical with or without interaction.

The effective number of agents can be obtained by relating the conditional variance of the aggregate expectation $\bar{\theta}$ to the conditional variance in the absence of interaction. Upon defining n_{eff} through $\sigma_t^2/n_{\text{eff}} = \text{Var}[\bar{\theta}]$, one directly obtains

$$n_{\text{eff}} = n + (n-1)J\sigma_t^2.$$

In the example just considered the interaction resulted in a reduction of the variance of the mean expectation $\bar{\theta}$. An increase in the variance of the market average $\bar{\theta}$ can also occur. A different and perhaps more realistic scenario might be one in which agents coordinate partly on some source of information, being, for example a public exogenous variable. Such a source of information might, but need not necessarily, be related to economic fundamentals. In case the 'signal' on which agents coordinate is an exogenous noise source, this can increase the correlation of the agent's predictors, and hence lead to an increase of the variance of the market average. This scenario can be represented by the following joint pdf:

$$\phi_t(\theta_1, \ldots, \theta_n) \sim \exp\left(-\sum_{i=1}^{n}\frac{(\theta_i - \mu_t)^2}{2\sigma_t^2} - \frac{J}{2}\sum_{i=1}^{n}(\theta_i - X_t)^2\right),$$

where X_t represents any random variable on which the agents coordinate. In case X_t is taken to be equal to the parameter θ_i of one of the agents, agent 1, say, so that $X_t = \theta_1$, this covers the situation where one agent announces his strategy publically at an early stage, after which others coordinate on this (noisy) signal. Clearly, the randomness that enters the joint pdf has a similar effect on all individual choices. If X_t is positive (negative), the expected value of $\bar{\theta}$ increases (decreases). Moreover, this effect does not average out when the number of agents is increased. Coordination on a (possibly endogenous) random variable thus provides an additional source of randomness in the price dynamics.

5.5.4 Inherent randomness

As a final source of randomness we consider an example of the dynamics in a case where $p^e_{t+1}(\theta_{i,t-1})$ does not have a finite conditional mean, as a result of which the law of large numbers does not apply. The aggregate expectation $\frac{1}{n}\sum_{i=1}^n p^e_{t+k}(\theta_{i,t-1})$ does not tend to a constant, but tends in distribution to a well defined random variable when n tends to infinity. We consider an asset pricing model with agents choosing among constant predictors: $p^e_{t+1}(\theta) = \theta$. For simplicity we put $\alpha = 0$ (no memory, only the last observed price is taken into account). A situation in which the mean does not exist arises if for the performance measure one takes

$$\pi_t(\theta) = -\log(1 + (\theta - p_t)^2),$$

The beliefs distribution can now become fat-tailed. For the beliefs distribution one obtains

$$\phi_t(\theta) = \frac{\Gamma(\beta)}{\Gamma(\frac{1}{2})\Gamma(\beta - \frac{1}{2})}(1 + (\theta - p_t)^2)^{-\beta},$$

from which it follows that

$$\sqrt{2\beta - 1}\,(\theta_{i,t-1} - p_{t-1}) \sim t(2\beta - 1),$$

a t-distribution with $2\beta - 1$ degrees of freedom. Thus, the $\theta_{i,t-1}$ are distributed symmetrically around p_{t-1}, and for $\beta < \frac{3}{2}$ the mean does not exist.

For $\beta = 1$, $\theta_{i,t} - p_t$ given $\mathcal{F}_t$ is Cauchy(0,1) distributed. Since this distribution is closed under averaging, this gives $\bar{p}^e_{n,t+1} = \frac{1}{n}\sum_{i=1}^n \theta_{i,t-1} - p_{t-1} \sim$ Cauchy(0,1). The price equation becomes

$$(1 + r)p_t = p_{t-1} + \bar{y} + \eta_t,$$

where η_t is a Cauchy(0,1) distributed random variable. The resulting price process is a first order autoregressive process with a fat-tailed noise term. The price dynamics is stochastic, and the distribution of the noise term is independent of the number of agents.

We conclude that the remark made by Muth (1961) that "allowing for cross-sectional differences in expectations is a simple matter because their aggregate effect is negligible as long as the forecasts are not strongly correlated" only applies if the distribution of predictors is sufficiently well-behaved.

Notice that the non-vanishing randomness of the type described here can only occur with a continuum of strategies. If the number of strategies is finite, provided that each of them predicts a finite price, the mean expectation always exists, in which case, by the law of large numbers, the dynamics become deterministic when the number of traders tends to infinity.

5.6 Concluding remarks

Many will agree that most economic observables, such as realized prices, are to a large extent determined by the expectations of economic agents. In turn, individual expectations are typically shaped by the observed past, as agents will try to incorporate the underlying economic laws at force in their predictions for tomorrow. New observations provide agents with an incentive to update their beliefs whenever their belief appears inconsistent with these observations. One would therefore expect individual beliefs to co-evolve over time with realized prices. Being evidently important but not directly observable, the evolution of the distribution of beliefs among market participants is at the heart of our concept.

To set up a framework for studying the evolution of beliefs distributions, we defined probability density functions on a space of possible strategies, which is called the beliefs space. The continuous choice model is employed to update the beliefs distribution, while the incentives for the agents to switch are provided by the past performances of strategies. Co-evolution of a distribution of beliefs with observed prices thus emerges due to the ongoing evaluation of predictors. This approach to modeling the evolution of the beliefs distribution over time provides several insights into the nature of its feedback with economic observables. The results and implications of our concept are at least two-fold.

Firstly, while being applicable in a wide range of theoretical models where expectations feedback plays a key role, the concept allows one to model the joint dynamics of the beliefs distribution and observables (prices, say) explicitly in terms of a *random dynamical system*. The beliefs distribution becomes a state variable which is endogenously shaped by the assumptions about the functional form of predictors, and the type of performance measure (squared prediction error or profits etc.) employed. For illustrative purposes, we have included some stylized examples that lead to a normal- and a log-normal distribution for individual beliefs. In these cases, the distribution of beliefs is fully described by the average belief and the dispersion of beliefs, the latter of which can be related to the degree of heterogeneity. The CBS then prescribes explicitly how the dispersion of beliefs affects prices, and in turn how prices affect the dispersion of beliefs.

Secondly, our concept provides what we refer to as 'natural sources of randomness'. In many economic models, noise, which is required when matching stylized facts is the objective, is often included ad-hoc as additive exogenous shocks. Popular justifications include: exogenous news shocks, model approximation error, and noise induced by trading. Our natural source of randomness introduces a type of endogenous uncertainty that can not be associated with exogenous shocks but rather with the uncertainty of choice deriving from unpredictable aspects of the preferences of economic agents. Moreover, the endogenous noise term is shown to inherit its statistical properties directly from the beliefs distribution. To understand where the stochasticity derives from, note that in a CBS the beliefs distribution describes the likelihood, through the eyes of the econometrician, according to which individual agents select their beliefs. When the number of agents is finite, the aggregate belief by definition is a random variable. From simulation studies (e.g. Lux and Marchesi, 1999) it is known that the stochastic nature of the dynamics in simulated markets vanishes when the number of agents becomes large. This is typically what the CBS framework would predict if agents have comparable market weights and their predictions are well-behaved (have finite mean and variance) and are not strongly dependent across agents. However, the CBS approach shows that there are several scenarios for which endogenous noise remains significant in a large market. This can occur if, for example: (i) wealth is disproportionately distributed (ii) agents' choices exhibit dependence; and (iii) when the combination of the performance measure and the functional form of beliefs induces nonexistence of the mean (with respect to the beliefs distribution) of the predictor function.

A number of extensions and generalizations might prove worthwhile. For example connecting the concepts of a CBS with that of "rational beliefs", the latter being recently developed by Kurz (2001). In the work of Kurz, a beliefs distribution describes the individual belief of an agent, and reflects the uncertainty individuals hold regarding future economic variables. This concept is designed to generalize rational expectations theory where all agents have exact knowledge about the future. Although Kurz's approach is different from ours, in particular since the beliefs distribution represents a different object, Kurz's definition of 'Endogenous Uncertainty' as 'that component of the volatility of quantities and prices in the economy which is generated by the distributions of beliefs' has similar implications. Both approaches are able to relate their notion of endogenous uncertainty to common financial market "anomalies" such as excess volatility, and ARCH-type structure. When comparing our approach with that by Kurz, a logical extension of the CBS would be to endow agents with notions of uncertainty about future values rather than having point predictors. Then each agent's belief is described by a pdf, so that agents will generally have more dimensions to disagree upon.

One area for future research is the analysis of the random dynamical systems obtained within the CBS framework, using stochastic bifurcation theory. Stochastic bifurcation theory is closely connected to the theory of random dynamical systems, a new field which currently is under rapid development (see e.g. Arnold (1998)). In stochastic bifurcation theory, progress has been relatively fast on continuous time random dynamical systems when compared to discrete time random dynamical systems. The bifurcation theory of discrete time random dynamical systems, which is still in its infancy, would clearly be useful for characterizing the dynamical systems of the type considered here.

5.7 Appendix

Below a proof of Proposition 5.1 is given. The proof uses the fact that the characteristic equations for the local dynamics of the CBS with squared prediction errors (section 5.4.1) and squared logarithmic prediction errors (section 5.4.2) are closely connected. In fact it will be shown that the stability conditions are identical in terms of a if one defines

$$a = (1+r)^{-1}$$

for the squared prediction error case, and

$$a = (1+r)^{-1} e^{\frac{3}{4\beta}}$$

for the squared logarithmic prediction error case.

Proof of Proposition 5.1

For the squared prediction error case (section 5.4.1), upon defining $w_t = \frac{1}{\sigma_t^2}$, the dynamics can be casted in the form

$$\begin{aligned}
p_t &= \left(\mu_{t-1} p_{t-1} + \bar{y}\right)/(1+r) \\
q_t &= p_{t-1} \\
\mu_t &= \left(\alpha \mu_{t-1} + 2\beta(1-\alpha) p_t q_{t-1}\right) \tfrac{w_{t-1}}{w_t} \\
w_t &= \alpha w_{t-1} + 2\beta(1-\alpha) q_{t-1}^2,
\end{aligned}$$

where p_t and w_t on the right hand side of the equation for μ_t are functions of the state variables at time $t-1$, as specified by the first and last equation. The Jacobian matrix evaluated at the equilibrium evaluates to

$$J_{\mathrm{SPE}} = \begin{pmatrix} a & 0 & ap^* & 0 \\ 1 & 0 & 0 & 0 \\ (1-\alpha)\frac{a}{p^*} & -\frac{1-\alpha}{p^*} & \alpha + (1-\alpha)a & 0 \\ 0 & 4\beta p^* & 0 & \alpha \end{pmatrix},$$

with $a = (1+r)^{-1}$.

Similarly, the CBS with squared logarithmic prediction errors (section 5.4.2) can be expressed as

$$\begin{aligned} p_t &= \left(\exp\left(\mu_{t-1} + \tfrac{1}{2w_{t-1}}\right) p_{t-1} + \bar{y}\right)/(1+r) \\ q_t &= p_{t-1} \\ \mu_t &= \left(\alpha\mu_{t-1}w_{t-1} + (1-\alpha)\left(2\beta(\ln p_t - \ln q_{t-1}) + 1\right)\right)/w_t \\ w_t &= \alpha w_{t-1} + 2\beta(1-\alpha), \end{aligned}$$

the Jacobian at the equilibrium becomes

$$J_{\mathrm{SLPE}} = \begin{pmatrix} a & 0 & ap^* & -\frac{ap^*}{8\beta^2} \\ 1 & 0 & 0 & -\frac{(1-\alpha)a}{8\beta^2} \\ (1-\alpha)\frac{a}{p^*} & -\frac{1-\alpha}{p^*} & \alpha + (1-\alpha)a & 0 \\ 0 & 0 & 0 & \alpha \end{pmatrix},$$

where $a = \frac{e^{\frac{3}{4\beta}}}{1+r}$.

Although the two Jacobian matrices differ, their characteristic equations coincide due to the fact that the upper left 3×3 sub-matrices are identical, and because of the three off-diagonal zeros in the last column/row respectively. In both cases the corresponding characteristic equation can be written as

$$\det\left[\lambda I - J\right] = \left[\lambda^3 - (2a + \alpha(1-a))\lambda^2 + \alpha\lambda a + (1-\alpha)a\right](\lambda - \alpha) = 0.$$

Since α is an eigenvalue, a necessary condition for stability is $\mid \alpha \mid < 1$. This condition is automatically satisfied under our assumption that $\alpha \in [0, 1)$. It can be readily verified that $(0, 0, 0, 1)'$ is the corresponding eigenvalue in both cases, so that the stability question reduces to that of the upper left 3×3 sub-matrix

$$\tilde{J} = \begin{pmatrix} a & 0 & ap^* \\ 1 & 0 & 0 \\ (1-\alpha)\frac{a}{p^*} & -\frac{1-\alpha}{p^*} & \alpha + (1-\alpha)a \end{pmatrix}.$$

Application of the conditions for stability derived in Jury (1974) to the characteristic equation

$$\det\left[\lambda I - \tilde{J}\right] = \lambda^3 - (2a + \alpha(1-a))\lambda^2 + \alpha\lambda a + (1-\alpha)a = 0$$

of the remaining eigenvalues, leads to several conditions on the parameters which, for

$\alpha \in [0, 1)$, can be summarized as

$$1 - 3a^2 - 2a\alpha + 5a^2\alpha + a\alpha^2 - 2a^2\alpha^2 > 0.$$

□

Chapter 6

Herding, A-synchronous updating and heterogeneity in memory

6.1 Introduction

The current literature on agent based economic models can be roughly divided into two types: (i) computational finance models and (ii) economic dynamic models. Both approaches aim to discover what is behind the stylized facts commonly exhibited by empirical financial time series.

In computational finance, artificial stock markets are used as a tool for simulating time series of aggregate variables, such as prices, volumes, etcetera (for an overview, see LeBaron, 2000). To improve the understanding of markets with a large number of interacting heterogeneous agents, LeBaron *et al.* (1999) for example developed the Santa Fe artificial stock market. In the Santa Fe market, the actions of agents, which are based upon their expectations, are explicitly modelled and traced for each individual agent. With computers becoming cheaply available and faster, these artificial markets allow for more and more detailed modelling. Although detailed models may possibly generate more realistic price dynamics, their increasing complexity also makes simulation less attractive as a tool for the analytic study of the mechanisms linking the heterogeneity of beliefs with the stylized facts of prices.

The economic dynamics literature follows a somewhat different route. There agent based modelling explicitly aims at an analysis of the joint dynamics of beliefs and prices (see e.g. Brock and Hommes, 1997, 1998). To achieve the necessary analytical tractability, several simplifying assumptions are usually made, such as a continuum of agents who at each time can choose from a finite number of belief types. Typically a low dimensional deterministic dynamical system is derived, allowing the modeller to perform a detailed bifurcation analysis, providing insights in the long-run behaviour of the dynamics for different economic and behavioral parameters. The natural trade-off of this approach is that it at best provides an approximation to the *deterministic* aspects of the price process of interest. Indeed, the 'holy grail' in economic dynamics is beyond any doubt *the* heterogeneous agent model that is just as realistic in terms of its price dynamics as it is tractable analytically.

In previous work (Diks and van der Weide, 2003) we introduced the concept of a *continuous beliefs system* (CBS). This framework is built around a continuous *beliefs space* representing the possible point predictors agents can choose from. On this space a time dependent *beliefs distribution* is defined, which is updated according to a continuous choice model. As new market prices become available, the beliefs distribution is updated depending on past performances of strategies. Coevolution of the beliefs distribution with price dynamics thus arises as newly realized prices feed the ongoing evaluation of the strategies by the agents. As is natural in a continuous choice setting, individual choices are considered to be random variables, distributed according to the beliefs distribution. The preferences of agents provide an endogenous noise source which affects price dynamics. Because the randomness in prices can be associated naturally with the diversity of beliefs, the latter has been labelled a 'natural source of randomness'.

Diks and van der Weide (2003) discuss several scenarios under which the natural randomness in prices does not average out in the limit as the number of agents tends to infinity. These include: (i) a disproportionate distribution of market impacts; (ii) particular combinations of utility functions and predictors, and (iii) dependence among agents, which may arise for example when agents partly coordinate on a random variable, not necessarily related to economic fundamentals ('sunspots'). All these examples are based on a violation of at least one of the conditions of the law of large numbers. In the last case (dependence) it is often impossible to distinghuish between an infinite number of dependent agents, or a finite number of independent agents. For convenience we therefore introduced the notion of an effective number of agents, n_{eff}, defined as the number of *independent* agents required to obtain stochastic price dynamics with similar stochastic dynamics. For a finite effective number of agents the dynamics is random by construction. In general, a CBS gives rise to a random dynamical system, allowing for the stochastic price dynamics to be described explicitly. It therefore offers an analytic alternative to the simulation based studies commonly performed in computational finance.

In sum, a CBS shares several advantages of both computational finance models and economic dynamic models. Rather than obtaining the stochastic macroscopic dynamics through laborious simulation, the macroscopic behaviour of a CBS with agents who can choose among a continuum of alternative strategies, can often be derived analytically. The aggregate behaviour can typically be captured in a few simple dynamic equations, for an arbitrary number of agents. The analytic tractability can be used to study features such as the conditional mean and variance of prices analytically, and relate these to behavioural characteristics.

The objective of this chapter is to examine the effects on price dynamics of a number of behavioral assumptions. We introduce a simple benchmark CBS and expand it in the following directions: (i) herd behavior; (ii) a-synchronous updating of beliefs; and (iii) heterogeneity in the memory of agents. The behavioral models obtained in this way give rise to aggregate dynamics which are represented by well-known classes of econometric models. At the center is the random walk obtained for our benchmark CBS. We will show that that: (i) herd behaviour introduces moving average (MA) structure yielding an ARIMA model; (ii) an ARIMA model with GARCH disturbances is obtained if beliefs are not updated synchronously; and (iii) a fractionally integrated (ARFIMA) process with long-range dependence is obtained if the memory of agents exhibits heterogeneity. As an immediate result the parameters of these standard econometric models can be related directly to the parameters of the behavioral model.

Near the end of the chapter we will only carefully scratch the surface of the possible empirical applications. This is done by estimating some of the stylized models using daily exchange rate data. We wish to emphasize honestly at the outset, however, that while interpreting the empirical results in terms of behavioral parameters, one should keep in mind that the models are stylized and, for clarity of exposition, deliberately stripped from a range of features that may become important for a realistic summary of the data. For example, we are aware of the fact that our natural source of randomness is not the only relevant source of randomness, and that more factors than those included in our behavioral model may be responsible for MA and/or GARCH structure in the empirical data. Although omitting additional relevant model aspects will surely bias our parameter estimates, we consider comparing the signs of the estimated parameters with those expected from our theory a sensible 'reality check'. If successful, such an empirical validation may hopefully open doors to more detailed empirical assessments in the future.

Finally note that inherent in our approach is the assumption that the agents believe in structure in prices that can be profitably exploited by 'technical trading rules' summarized in terms of point predictors, which we will refer to interchangeably as strategies or beliefs. Empirical evidence supporting this belief in predictability in the foreign exchange markets is provided by e.g. Sweeney (1986), Taylor and Allen (1992), Levich and Thomas (1993) and LeBaron (1999). Furthermore, our approach is entirely built around the heterogeneity of beliefs. Some arguments for expectations being heterogeneous have been provided by, among others, Frankel and Froot (1990). In a theoretical exchange rate model, Frankel and Froot (1988) particularly employ the heterogeneity of expectations to explain some of their empirical findings.

In the following section we briefly review the basic concepts of a CBS. The introduction of the benchmark CBS and its extensions are described in section 6.3. In section 6.4 the models will be subjected to a basic empirical validation. Section 6.5 concludes.

6.2 Continuous beliefs systems

In this section we briefly review the concept of a continuous beliefs system. The main ingredients for describing the co-evolution of prices and the beliefs distribution in a CBS are the following. As is common in economic dynamics models, agents are assumed to predict future prices by using a point predictor which is a function of the information available to them (we typically have in mind past realized prices). The point predictors agents can choose from are represented in a beliefs space Ω, parameterized by a continuous parameter θ, called the *belief parameter*. Each value of $\theta \in \Omega$ represents a predictor $x^e_{t+1}(\theta)$ of the next price in terms of the information $\mathcal{F}_t$ available to the agents. For example if agents believe in a linear world and expect tomorrow's price x_{t+1} to be a linear function $x^e_{t+1}(\theta) = \theta_0 + \theta_1 x_t + \ldots + \theta_d x_{t-d+1}$ of the last d prices,then the beliefs space consists of the $d+1$-dimensional Euclidean space $\mathbf{R}^{d+1}$.

Given the individual beliefs $\{\theta_{i,t-1}\}$, a realization of the price x_t is obtained via some market clearing mechanism. Agents' utility functions and predictors imply individual demand functions, and the market price is defined to be the unique price x_t for which the market clears.[1] Of course the appropriate market clearing price equation depends on the

[1]Uniqueness of market clearing prices is not self-evident. However, if demands are strictly decreasing

context, and we will have to provide some additional arguments to justify our particular choice made for the benchmark model later.

As new prices become available, agents update their evaluations of parameter values by confronting them with the newly arrived information. As agents have differences in tastes and interpret identical information differently they will typically choose different predictors. This is captured by a continuous choice model, which defines a probability distribution on the beliefs space Ω. The probability density function of this time dependent *beliefs distribution* is denoted by $\phi_t(\theta)$. The actual belief parameters $\theta_{i,t}$ individual agents use to base their next prediction on are modelled as random variables distributed according to $\phi_t(\theta)$. In this way endogenous randomness enters the price dynamics in a natural way. In the simplest case the $\theta_{i,t}$ are assumed to be independent conditional on information available up to and including time t.

To see how the beliefs distribution evolves, let us consider the time when x_t has just been quoted. After x_t becomes public, agents re-evaluate the strategies using a *performance measure*, $\pi_t(\theta)$. The performance depends on past realized prices and might for example be based on *ex post* prediction errors, or the *ex post* profits realized by a strategy. A typical choice is to take $\pi_t(\theta)$ to consist of minus the *ex post* squared prediction error:

$$\pi_t(\theta) = -(x_t^e(\theta) - x_t)^2. \tag{6.1}$$

This choice is less restrictive as it may seem, since for $\phi_t(\theta)$ only *differences* in performance play a role. Using this property it can be shown that minus squared prediction errors are equivalent to risk adjusted profits when agents hold identical beliefs about conditional variances (see e.g. Hommes, 2001). In practice, decisions to change strategies are not based on yesterday's performances only, but rather on the success over a larger history. Indeed, financial analysts typically test their candidates on a large sample of past price history, also known as back-testing. Strategies that have performed best over some history of the sample are more likely to be selected for future trading. The simplest way to implement, or mimic, this back-testing is to consider a geometrically down-weighted average of past performances rather than only the last performance, so that the weighted average utility becomes:

$$U_t(\theta) = (1-\alpha)\sum_{i=0}^{\infty} \alpha^i \pi_{t-i}(\theta),$$

in x_t, as will be the case in our benchmark CBS, uniqueness of clearing prices is guaranteed.

which can be recast into the more practical form

$$U_t(\theta) = \alpha U_{t-1}(\theta) + (1-\alpha)\pi_t(\theta). \tag{6.2}$$

Given the average past utility the continuous logit model provides the following pdf for the beliefs distribution just after time t:[2]

$$\phi_t(\theta) = \frac{\mathrm{e}^{\beta U_t(\theta)}}{Z_t}, \tag{6.3}$$

where Z_t is a normalization constant independent of θ. The parameter β is referred to as the *intensity of choice*. The inverse $1/\beta$ is also known as the *propensity to err*. If agents are sensitive to small differences in utility, this will be reflected by a large value of β. Notice that Eq. (6.3) implies that only differences in performance play a role. Modifications of $\pi_t(\theta)$ that are independent of θ will be absorbed by the normalization factor Z_t, leaving $\phi_t(\theta)$ unchanged. With this in mind it is not surprising that risk adjusted profits and minus squared prediction errors turn out to be equivalent under certain conditions.

Substitution of Eq. (6.2) into Eq. (6.3) gives

$$\phi_t(\theta) = \frac{\mathrm{e}^{\alpha\beta U_{t-1}(\theta)+(1-\alpha)\beta\pi_t(\theta)}}{Z_t} = \frac{[\phi_{t-1}(\theta)]^{\alpha}\mathrm{e}^{(1-\alpha)\beta\pi_t(\theta)}}{Z'_t}, \tag{6.4}$$

which expresses how a new beliefs distribution can be obtained from the previous beliefs distribution and the last performance measure. The variable Z'_t is again a normalization constant. Although $\phi_t(\theta)$ is an infinite dimensional state variable, the time evolution of $\phi_t(\theta)$ can often be fully specified by a finite number of variables such as its first k moments, in which case the dynamics becomes finite dimensional. In summary, the expectations feedback can be represented schematically as

$$\ldots x_{t-1} \to \phi_{t-1}(\theta) \rightsquigarrow x_t \to \phi_t(\theta) \rightsquigarrow x_{t+1} \to \ldots,$$

where "$\to$" indicates a deterministic step and "$\rightsquigarrow$" a step involving endogenous noise.

It is important to note that the beliefs distribution $\phi_t(\theta)$ is determined only by the performance measure and past realised prices, and otherwise evolves independently of the individual choices of agents. The beliefs distribution is a convenient concept introduced

[2]In Diks and van der Weide (2003) the right-hand-side was multiplied by a so-called opportunity function $\varphi(\theta)$ which we take equal to unity here.

in the choice literature to model the fact that agents have idiosyncrasies in tastes, and that their choices are not identical given identical information. The fact that identical information is interpreted differently by agents is confirmed empirically by Kandel and Pearson (1995). The strategies $\theta_{i,t}$ used by the agents, from the point of view of the econometrician, are random variables, representing the individual choices of agents, so that x_{t+1} is a stochastic variable, even in the absence of exogenous shocks.

6.3 A simple benchmark CBS and its extensions

In this section we introduce a benchmark CBS and examine the effects on the dynamics of extending the model in the following behavioral directions: (i) herding, (ii) asynchronous updating, and (iii) heterogeneity in memory. The CBS formulation enables one to obtain insights into how these behavioral assumptions affect the interaction between the stylized facts and the unobserved beliefs distribution. The benchmark model is an extremely simple CBS deliberately stripped from any features that are not of direct interest. Although taking such a highly stylized model as a reference point may seem overly restricted, it clarifies the exposition and helps to isolate the effects of each of the extensions.

We wish our benchmark CBS to be as simple as possible while preserving economic relevance and bringing out the essentials of the CBS concept. Typically a CBS will be used for modelling the joint dynamics of beliefs and prices of well-specified economic units, such as shares, which are traded sequentially in a large market. Prices in such markets are determined by the expectations held by agents regarding future prices and cash flows, as well as their preferences concerning returns and risk. Agents interact by submitting demand functions depending on expected utility, while their realized utility is determined by realized market clearing prices. Regardless of preferences, it is clear that better abilities to predict future prices given the information available will help agents in realizing a higher utility. In this sense the setting resembles an ever ongoing version of Keynes' beauty contest.

Motivated by these arguments we consider a very stylized market mechanism which assumes prices to equal the average prediction over all agents in the market, i.e. the *market expectation*:

$$x_t = \frac{1}{n}\sum_{i=1}^{n} x^e_{t+1}(\theta_{i,t-1}), \tag{6.5}$$

where n denotes the number of agents. Price equations used in economic dynamic models often involve similar market expectations. Depending on the type of market, also the risk free interest rate and expectations concerning dividend payments may enter the price equation. We have in mind agents who are myopic in the sense that they are concerned only with maximizing their one-step-ahead utility. At each time then, agents act as if they are entering the market today while knowing that they will leave the market tomorrow. As a result, only their expectations concerning the next price and additional cash flows (such as dividends) play a role. Typically, dividend payments (if any) are rare and announced well in advance, which means that expectations regarding short run profits are driven by expectations concerning future prices, rather than dividends. If, moreover, time intervals are sufficiently short (e.g. one to five days, see Lehmann, 1991) the interest rates over a single interval are sufficiently small for agents to consider their predicted future price to be a 'fair price' today as well. Under those conditions, if we assume agents to be mean-variance optimizers the market clearing price is given by the right hand side of Eq. (6.5).[3]

The simplest continuous class of point predictors consists of a continuous range of different constants. We incorporate this by saying that agents with belief θ predict tomorrow's price to be equal to θ:

$$x^e_{t+1}(\theta) = \theta.$$

Heterogeneity will arise if agents employ different beliefs in that they have in mind different parameter values $\theta_{i,t}$. The conditional variance $\mathrm{Var}_t[\theta_{i,t}]$ can then be considered to be a measure for the degree of heterogeneity. Forecast surveys of foreign exchange market participants, for example, indicate that such a dispersion of beliefs indeed exists, as found by Frankel and Froot (1990). An argument they provide in favour of heterogeneity is the high trading volume observed in the foreign exchange market, since it takes differences of opinion to trade.

For the performance measure we take minus squared prediction errors (Eq. 6.1), and memory is included in the utility function through Eq. (6.2). Having observed and compared overall performance, all agents subsequently adapt their beliefs. The new distribution of beliefs is determined by means of the continuous choice model, and is given by Eq. (6.3). Because the utility function is a quadratic function in the belief parameter θ, it follows

[3] A mean-variance optimizing agent with belief θ expects a risk-adjusted short-term profit of $d_t[x^e_{t+1}(\theta) - x_t] - \frac{1}{2}cd_t^2$, where d_t is his demand and c is a risk aversion parameter times the perceived conditional variance of x_{t+1}. Maximizing this with respect to demand d_t, gives the demand function $d_t = (x^e_{t+1}(\theta) - x_t)/c$. If c is homogeneous among agents, the unique equilibrium price is given by Eq. (6.5). Heterogeneity in risk aversion and or perceived risk would lead to a weighted average $x_t = \sum w_i x^e_{t+1}(\theta_{i,t+1}) / \sum w_i$ with weights $w_i = c_i^{-1}$, which in a large market is equivalent to Eq. (6.5).

that, for all t, the distribution of beliefs is described by a normal distribution:

$$\phi_t(\theta) = \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left[-\frac{1}{2}\left(\frac{\theta - \mu_t}{\sigma_t}\right)^2\right]. \tag{6.6}$$

Normality implies that the beliefs distribution is fully specified by its mean μ_t and variance σ_t^2. Using Eq. (6.4) it can be seen that the dynamics of μ_t and σ_t is governed by:

$$\begin{aligned} \mu_t &= \alpha\left(\frac{\sigma_t^2}{\sigma_{t-1}^2}\right)\mu_{t-1} + 2(1-\alpha)\beta\sigma_t^2 x_t \\ \frac{1}{\sigma_t^2} &= \frac{\alpha}{\sigma_{t-1}^2} + 2(1-\alpha)\beta. \end{aligned} \tag{6.7}$$

In the long run σ_t^2 will converge to $\sigma^2 = \frac{1}{2\beta}$, and the dispersion of beliefs thus tends to a constant proportional to the propensity to err. If the propensity to err is high, agents are less sensitive to differences in performances of strategies. As a consequence, a large propensity to err implies a larger dispersion of beliefs. If we substitute $\sigma^2 = \frac{1}{2\beta}$ into the dynamics for the average belief μ_t, we obtain:

$$\mu_t = \alpha\mu_{t-1} + (1-\alpha)x_t. \tag{6.8}$$

Having derived the dynamics for beliefs, we now shift focus to its interaction with the price equation Eq. (6.5) which in the constant predictor case becomes:

$$x_t = \frac{1}{n}\sum_{i=1}^{n} x_t^e(\theta_{i,t-1}) = \frac{1}{n}\sum_{i=1}^{n}\theta_{i,t-1},$$

For simplicity, we consider the system in the limit as the number of agents tends to infinity. In the absence of dependence among agents, the law of large numbers applies, and it follows that the observed price quoted as the average prediction will converge to μ_{t-1}:

$$x_t = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\theta_{i,t-1} = E[\theta_{i,t-1}] = \mu_{t-1}. \tag{6.9}$$

Together with the 'dynamics' of the mean μ_t described by Eq. (6.8) this gives:

$$x_t = \alpha x_{t-1} + (1-\alpha)x_{t-1} = x_{t-1}.$$

Thus in the absence of uncertainty, the price will be constant at all times.

Exogenous shocks, such as news, are assumed to affect the price equation only through the expectations of agents. This assumption appears natural, since the only way news can possibly affect prices is indeed through expectations of individuals. The dynamics become

$$\begin{aligned} x_t &= \mu_{t-1}, \\ \mu_t &= \alpha\mu_{t-1} + (1-\alpha)x_t + \xi_{t+1}, \end{aligned} \tag{6.10}$$

where ξ_{t+1} is a noise term representing news shocks arriving between time t and $t+1$. Upon substituting the first of those equations in the second, it can be seen that μ_t follows a random walk driven by ξ_t. The first equation states that x_t simply equals μ_{t-1}, the average belief under agents given all observables up to and including time $t-1$, and the news that has become available before time t.[4] In terms of increments we thus have, in the presence of exogenous shocks $\{\xi_t\}$:

$$\Delta x_t \equiv x_t - x_{t-1} = \xi_t.$$

The random walk model is consistent with the efficient market hypothesis (EMH) and has been traditionally viewed as *the* model for financial variables, such as asset prices and exchange rates. Indeed, the random walk has long been considered unbeatable, a view epitomized by Malkiel (1990). Only recently it seems that it has to give in to a growing list of statistical evidence against the EMH, some of which is collected in Lo and MacKinlay (1997). As a consequence we see that more sophisticated models have begun to make their appearances. Kilian and Taylor (2001), for example, introduced an exchange rate model which is close to a unit root process near the fundamental, but in which the exchange rate is driven back to the fundamental when their difference becomes too large. They report empirical evidence supporting predictability of exchange rates at long horizons, but at short horizons exchange rates are hard to distinguish from a random walk.

6.3.1 Herd behaviour

If the conditions for the law of large numbers (LLN) are not satisfied, the price dynamics may be stochastic in a large (many trader) market even in the absence of exogenous

[4] The random walk price dynamics can alternatively be derived with agents forming expectations of the type $x_t^e(\theta) = R_t + \theta$, where R_t is a publicly known estimate of the fundamental exchange rate with increments $R_t - R_{t-1} = \xi_t$. In that case μ_t is stationary, while the exchange rate follows a random walk, $x_t = R_t + \text{cnst}$.

uncertainty. In the introduction we briefly mentioned a number of such 'sources of randomness' that might cause the LLN to fail. These are addressed in more detail in Diks and van der Weide (2003). Here we will focus on the dependence among agents arising from herd behaviour, as a source of 'natural randomness'. It turns out that sufficient correlation between the agents will indeed violate the conditions for the LLN to apply.

In the literature many different definitions of herding have been described. The type of herding we wish to address is that where conditional on public information, the strategies used by the agents exhibit dependence. In particular we have in mind situations where agents herd to publically announced opinions of others. During each period some 'early' agents reveal their expectations, allowing other agents to adapt their expectations towards those of the 'leaders' before submitting their market orders. This type of herding is 'concurrent' in that it involves agents coordinating simultaneously on common signals. Concurrent herding is different from studies that associate herding with for example informational cascades or learning (see e.g. Bikhchandani *et al.*, 1992; Avery and Zemsky, 1998 and Nelson, 2002). In informational cascades uninformed agents try to mimic, or learn from, the more informed by extracting information from realized prices. This type of herding therefore has a 'sequential' nature. As one would expect, having in mind 'concurrent' or 'sequential' herding makes a difference. For example, Nelson (2002) finds that 'high correlation among agents' actions does not necessarily imply herding'. As in our model dependence among actions must originate from dependence in strategies, we would refer to such correlation among actions as herding by definition.

Possible motivations for (concurrent) herding are numerous. First note that according to Kandel and Pearson (1995), even though each individual typically has his own interpretation of public signals only few behave completely independent. A motivation is that being too different from the rest can be risky and might jeopardize career perspectives or reputation (see e.g. Trueman, 1994 and Scharfstein and Stein, 1990). Hong *et al.* (2000) have recently strengthened these arguments by providing some empirical evidence in favour of them. When delving into experience, they also find that younger analysts forecast closer to the average forecast, and that they are more likely to be terminated when they deviate from the consensus.

Even when accurate (short term) forecasting is the only objective agents have, herding might still be rational. Scharfstein and Stein (1990), as an oblique remark not that: "...investing is also driven by group psychology, which weakens the link between information and market outcomes." Indeed, note that in our benchmark CBS prices are entirely driven by speculation (group psychology).[5] Froot *et al.* (1992) have given the thought more attention, and concluded the following. In the short term, speculative forces may be stronger than those of the 'underlying economic system'. As a consequence, price series generally show fluctuations around the fundamental, where one can always find short periods in time where prices exhibit divergence. In such periods it is thus rational to *coordinate* on speculative behaviour (herding) rather than on fundamentals.

Kirman (1993) describes herding as arising from a process, which involves agents that make choices first, 'recruit' other agents to make the same decision. There are several possible forms in which this recruiting takes place, which we will not delve into here. What is important for our model is that there may be mechanisms that introduce dependence among the strategies selected by the agents. The 'recruiting' scenario is an example mechanism which gives just this type of dependence in our model.

Herding towards a leading agent can be implemented easily in a CBS by introducing interaction terms into the utility function.[6] For simplicity we assume that there is only a single leading agent. Let $\theta_{0,t}$ denote the parameter value announced publically by the leading agent, which we assume to be distributed according to the original beliefs distribution given by Eq. (6.6). Upon taking a new utility function $\tilde{U}_t(\theta) = U_t(\theta) - J(\theta - \theta_{0,t})^2$, the beliefs distribution of the other agents becomes

$$\phi_t(\theta) = Z_t^{-1} \exp\left[\beta U_t(\theta) - \beta J(\theta - \theta_{0,t})^2\right]. \tag{6.11}$$

The interaction strength $J > 0$ can be considered as a measure for the desire of agents to select a parameter value close to that of the leader. Straightforward comparison of coefficients of powers of θ in the exponent with those in Eq. (6.6) gives the following values for the mean $\tilde{\mu}_t$ and variance $\tilde{\sigma}_t^2$ of the beliefs distribution under herding:

$$\tilde{\mu}_t = \mu_t + \frac{J(\mu_t - \theta_{0,t})}{J+1} \equiv \mu_t,$$

and

$$\tilde{\sigma}^2 = \frac{\sigma^2}{1 + \sigma^2 J}.$$

[5]Lux (1995) is another paper that considers the possibility where agents pay more attention to each other than to the fundamentals. He assumes a market in which initially only optimists and pessimists can be distinguished. The evolution of the number of optimists is completely determined by the numbers of optimists and pessimists that surround individuals. At a later stage he assumes the presence of a fundamentalist.

[6]See Diks and van der Weide (2003), where social interaction is incorporated in a CBS, in analogy with Brock and Durlauf (2001) who consider social interaction in a dynamic discrete choice setting.

In the latter relation the time indices have been dropped as σ_t^2 was already known to be constant over time (see Eq. (6.7) and the discussion thereafter).

Under the assumption that market weights of individual agents (including the leading agent's) are negligible, realized prices are given by the mean of the pdf in Eq. (6.11):

$$x_t = \tilde{\mu}_{t-1} = \mu_{t-1} + \sigma\lambda u_t, \tag{6.12}$$

where $\{u_t\} \equiv \left\{(\mu_{t-1} - \theta_{0,t-1})/\sigma\right\}$ is a sequence of independent standard normal random variables.

If we incorporate the dynamics for the average belief μ_t, which is given in Eq. (6.8), it follows that the market's price dynamics is described by the following ARIMA model:

$$x_t = x_{t-1} + \frac{J}{J+1}\sigma\left[u_t - \alpha u_{t-1}\right].$$

The amount of endogenous uncertainty entering the price dynamics is proportional to $J/(J+1)$. Since this function is increasing in J and takes values between 0 and 1, it is convenient to define a measure of herding λ, via

$$\lambda \equiv \frac{J}{J+1},$$

so that the price dynamics in the absence of news can be concisely stated as

$$\Delta x_t = \sigma\lambda\left(u_t - \alpha u_{t-1}\right), \tag{6.13}$$

where $\Delta x_t \equiv x_t - x_{t-1}$. As herd behaviour acts to affect the *effective* number of agents, it should come as no surprise that a change in the extent of herding λ can not be distinguished from a reduction of the number of agents in the CBS. Only their joint effect can be identified.[7]

At this stage all stochasticity in the model can be associated with endogenous uncertainty. That is, the randomness observed by the econometrician reflects the random nature with which individual agents select beliefs from the distribution of beliefs. If in addition

[7] Diks and van der Weide (2003) related the amount of endogenous noise to both the market impact of agents, and the correlation between agent's decisions. The effective number of agents (which is inversely proportional to the variance of endogenous noise) is given by $n_{\text{eff}} = \left(\sum_{i=1}^n w_i^2 + \rho\sum_i\sum_{j\neq i} w_i w_j\right)$ where n is the total number of agents, $\{w_i\}$ represent their market impacts, and where ρ is the correlation between agent's beliefs. Taking equal weights, $w_i = \frac{1}{n}$, and letting n tend to infinity as we do here, only leaves $n_{\text{eff}} = \rho^{-1}$.

we include independent, identically distributed (IID) exogenous shocks ξ_t to the system, we find:

$$\begin{aligned} \Delta x_t &= \sigma\lambda\left(u_t - \alpha u_{t-1}\right) + \rho\xi_t \\ &\sim \eta\left(u'_t - \psi u'_{t-1}\right), \end{aligned} \tag{6.14}$$

where $\{u'_t\}$ is a sequence of IID standard normal random variables, η a scaling parameter, and $-\psi$ a moving average coefficient. The relation between our original parameters $(\sigma, \lambda, \alpha, \rho)$ and the new parameters (η, ψ) is the following:

$$\begin{aligned} \eta &= \left(\sigma^2\lambda^2 + \rho^2\right)^{\frac{1}{2}} \\ \psi &= \frac{\alpha\sigma\lambda}{\left(\sigma^2\lambda^2 + \rho^2\right)^{\frac{1}{2}}}. \end{aligned} \tag{6.15}$$

The equivalence of the two representations in Eq. (6.14) follows directly from comparing the auto-covariance functions of both representations. While both representations describe the same ARIMA(0,1,1) model, the reduction from four to two parameters shows that the first representation was over-parameterized. The original (behavioral) model parameters can therefore not be identified all independently. For an introduction to ARIMA and related models we refer to Mills (1999).

It is worth noting that we have established a link between behavioral finance and econometric modelling. As a direct result, one can associate the parameters from the econometric model with economically relevant and/or behavioral quantities. In this particular example, the scale of the endogenous noise appears in the econometric model as the product $\sigma\lambda$ of the dispersion σ of the beliefs distribution and the herding parameter λ, while ρ measures the size of the exogenous noise.

6.3.2 A-synchronous updating

So far, as is also common in economic dynamic models based on discrete choice, we have assumed implicitly that the time scale on which agents re-evaluate strategies is short with respect to the trading time. It seems reasonable to assume that in the absence of costs for evaluating and updating beliefs, every agent will indeed decide to update every period. However, if updating is either time consuming or costly, agents are likely to update less frequently. Hong *et al.* (2000) found empirical evidence supporting this view. In their empirical study they concluded that inexperienced analysts revise their forecasts more frequently than experienced analysts.

In general, if the time scales on which agents review strategies is longer in practice than the clock rate at which prices/exchange rates are quoted, the dynamics of both prices and the distribution of beliefs will be affected. Indeed, Brock and LeBaron (1996) find that time series with more realistic properties, such as strong persistence of volatility, are obtained in the case of less frequent updating. The effects of the time scale for evaluations can be incorporated in the model in several ways. One might take the approach of Brock and LeBaron (1996) where all agents update their beliefs all at once with a certain fixed probability per time unit.

An alternative approach, which we will follow here, is to assign a probability δ that individual agents do not update their beliefs, and a probability $1-\delta$ that individual agents do update their belief, at each time step. The latter is generally more difficult to implement, since it implies that the beliefs distribution becomes a mixture with weight δ on the old utility function, and $1-\delta$ on the new, that is,

$$\begin{aligned} \phi_t(\theta) &= \delta\phi_{t-1}(\theta) + (1-\delta)\tilde{\phi}_t(\theta) \\ \tilde{\phi}_t(\theta) &\sim e^{\beta U_t(\theta)} \sim e^{\beta(1-\alpha)\pi_t(\theta)}[\tilde{\phi}_t(\theta)]^{1-\alpha}. \end{aligned} \tag{6.16}$$

Here $\tilde{\phi}_t(\theta)$ represents the up-to-date distribution of choices that would be made based on information up to the recent past, while the actual beliefs distribution $\phi_t(\theta)$ is still partly based on old information. In simple cases, however, where the mean and variance of the mixture depend on the past utility only through the previous mean and variance, a closed form updating rule for the new mean and variance can still be given. This is demonstrated in the following example.

The average belief and the dispersion of beliefs among the agents that have decided to update are denoted by $\{\mu_t, \sigma_t^2\}$, where $\sigma_t^2 = \sigma^2$ is still constant over time. The average and the dispersion of beliefs for the entire population is now denoted by $\mu_{\delta,t}$ and $\sigma_{\delta,t}^2$. Using Eq.(6.16) it can be verified that the dynamics of the first two moments of the beliefs distribution is now governed by:

$$\begin{aligned} \mu_{\delta,t} &= \delta\mu_{\delta,t-1} + (1-\delta)\,\mu_t \\ \left(\mu_{\delta,t}^2 + \sigma_{\delta,t}^2\right) &= \delta\left(\mu_{\delta,t-1}^2 + \sigma_{\delta,t-1}^2\right) + (1-\delta)\left(\mu_t^2 + \sigma^2\right), \end{aligned} \tag{6.17}$$

where as before (Eq. 6.8), we have:

$$\mu_t = \alpha\mu_{t-1} + (1-\alpha)\, x_t. \tag{6.18}$$

If we take squares on both sides of the first relation in Eq. (6.17) and subtract both relations to eliminate $\mu^2_{\delta,t}$, we find the following dynamics for the dispersion of beliefs $\sigma^2_{\delta,t}$:

$$\sigma^2_{\delta,t} = \sigma^2 + \delta\left(\sigma^2_{\delta,t-1} - \sigma^2\right) + \delta\,(1-\delta)\left(\mu_t - \mu_{\delta,t-1}\right)^2. \tag{6.19}$$

Thus the dispersion of beliefs exhibits fluctuations when agents do not move together, even when the dispersion of beliefs among those who adapt is constant in each period. The movements are driven by the differences in average beliefs between those who decide to adapt and those who wait.

The time evolution for the price is now described by the following analogue of Eq. (6.12):

$$x_t = \mu_{\delta,t-1} + \lambda\sigma_{\delta,t-1}u_t, \tag{6.20}$$

where λ is our measure for herding and $\sigma_{\delta,t-1}$ is the time varying standard deviation of the beliefs distribution. Let us first focus on the average beliefs $\left\{\mu_t, \mu_{\delta,t}\right\}$. Note that Eqs (6.17) and (6.18) give, respectively,

$$(1-\delta)\,\mu_t = \mu_{\delta,t} - \delta\mu_{\delta,t-1}$$

and

$$(1-\delta)\,\mu_t = \alpha\,(1-\delta)\,\mu_{t-1} + (1-\alpha)\,(1-\delta)\,x_t,$$

which together yield:

$$\mu_{\delta,t} - \delta\mu_{\delta,t-1} = \alpha\left(\mu_{\delta,t-1} - \delta\mu_{\delta,t-2}\right) + (1-\alpha)\,(1-\delta)\,x_t.$$

If we substitute $\mu_{\delta,t} = x_{t+1} - (\lambda\sigma_{\delta,t}u_{t+1})$ (Eq. 6.20), and rearrange terms, we obtain:

$$\Delta x_t = \alpha\delta\Delta x_{t-1} + \lambda\left(\sigma_{\delta,t-1}u_t - (\alpha+\delta)\sigma_{\delta,t-2}u_{t-1} + \alpha\delta\sigma_{\delta,t-3}u_{t-2}\right),$$

where, as before, $\{u_t\}$, is again a sequence of IID standard normal random variables.

For positive memory ($\alpha > 0$) and a-synchronous updating ($\delta > 0$), the dynamics of the differences $\{\Delta x_t\}$ is described by an ARMA(1,2) process. Interestingly, and in contrast with our previous example, the (conditional) variance of the disturbances is no longer constant. The conditional variance is proportional to the dispersion of beliefs, which now exhibits endogenous fluctuations. Intuitively, these fluctuations in the degree of heterogeneity can be ascribed to the differences in opinion between those who decide to update their beliefs and those who believe that updating can wait. According to Eq. (6.19), the dynamics of heterogeneity is described by an autoregressive process of order one, from which one may correctly expect a GARCH(1,1) model to emerge for the conditional variances. For clarity of exposition we set memory to zero i.e. $\alpha = 0$, so that the initial ARMA(1,2) process for $\{\Delta x_t\}$ is stripped down to a MA(1) process:

$$\begin{aligned} \Delta x_t &= \lambda\sigma_{\delta,t-1}u_t - \lambda\delta\sigma_{\delta,t-2}u_{t-1} \\ &\equiv \epsilon_t - \delta\epsilon_{t-1} \end{aligned} \tag{6.21}$$

where the conditional variance $h_t = \text{Var}_t(\epsilon_t) = \lambda^2\sigma^2_{\delta,t-1}$ is updated according to:

$$h_t = (1-\delta)\lambda^2\sigma^2 + \delta(1-\delta)\lambda^2\epsilon^2_{t-1} + \delta h_{t-1}, \tag{6.22}$$

where we have used Eq. (6.19) together with the fact that for $\alpha = 0$, Eq. (6.18) implies $\mu_t = x_t$, so that Eq. (6.20) gives $\mu_t - \mu_{\delta,t-1} = x_t - \mu_{\delta,t-1} = \epsilon^2_{t-1}$. Because $0 < \lambda^2 < 1$ a given GARCH parameter δ implies an upper limit of $\delta(1-\delta)$ for the ARCH parameter. Note that the sum $s = \delta + \lambda^2\delta(1-\delta)$ of the ARCH and GARCH parameters satisfies $0 < s < 1$ for $0 < \delta < 1$, so that the resulting GARCH process is stable.

6.3.3 Heterogeneity in memory

In this section we consider the third and final behavioral extension of the benchmark model: heterogeneity in memory. As the diversity in future time horizons seems relevant, also the diversity in past time horizons may play an important role in the co-evolution of prices and beliefs. With past time horizons we refer to the history of price observations used to test candidates for future trading. Agents who are more concerned with price behaviour on the short-term are more likely to compare performances over a smaller sample of recent observations than agents who have a clear interest in the long-term dynamics. Another reason for agents to use small memory parameters might be that they believe that the world they live in is inherently non-stationary, and only recent performances can be considered relevant.

Recently, a number of studies addressed the issue of different (future) time horizons among agents (see e.g. Muller *et al.*, 1997; Dacorogna, 2002). This generalization is supported by the observation that agents trade at different frequencies. That is, the population of traders often consists of both long-term traders and short-term traders. Heterogeneity in agents' time scale are believed to be responsible for a number of stylized facts. Long term traders naturally focus on long-term behavior of prices thereby neglecting fluctuations at the smallest time scale, whereas short-term traders are not concerned with price movements on the long-run but rather aim to exploit short term predictability. The effects of the diversity in time horizons on price dynamics have been studied by LeBaron (2001) in an artificial stock market model. He concluded that the presence of heterogeneity in horizons may lead to an increase in return variability, as well as volatility-volume relationships similar to those of actual markets. Dacorogna (2002) found evidence supporting trading horizon heterogeneity to be responsible for the slow decay of correlations found empirically.

In the case where the memory parameter α is distributed among agents according to the pdf $\nu(\alpha)$ with $0 < \alpha < 1$, the most recent beliefs distribution for each memory type can be denoted by $\tilde{\phi}_{\alpha,t}(\theta)$. The overall beliefs distribution is then given by

$$\phi_t(\theta) = \int_0^1 \tilde{\phi}_{\alpha,t}(\theta)\nu(\alpha)\,\mathrm{d}\alpha.$$

Again, in general it can be very difficult to derive the updating rules for $\phi_t(\theta)$ analytically. However, if closed form expressions for the moments $E[\alpha^k]$ with respect to $\nu(\alpha)$ are known, for example, it is possible to derive analytic expressions for some of the moments, such as the mean μ_t and variance σ_t^2, of the beliefs distribution. This is demonstrated in some examples below.

Let α_i denote the memory of agent i, and $\mu_{i,t}$ and $\sigma_{i,t}^2$ represent the first two moments of the beliefs distribution corresponding to agents with memory parameter α_i. In the case of synchronous updating, it can be verified that $\sigma_{i,t}^2 = \sigma^2$ for all i, and that the dynamics of the average $\{\mu_t\}$ over the entire population is described by:

$$\mu_t = \int_0^1 \mu_t(\alpha)\nu(\alpha)\mathrm{d}\alpha. \tag{6.23}$$

For the mean over agents with memory parameter α_i we find

$$\begin{aligned}\mu_{i,t} &= \alpha_i \mu_{i,t-1} + (1-\alpha_i)\, x_t \\ &= (1-\alpha_i) \sum_{k=0}^{\infty} \alpha_i^k x_{t-k} \\ &= \sum_{k=0}^{\infty} \alpha_i^k x_{t-k} - \sum_{k=0}^{\infty} \alpha_i^{k+1} x_{t-k}.\end{aligned}$$

Taking expectations on both sides yields:

$$\mu_t = \sum_{k=0}^{\infty} E_\nu \left[\alpha^k\right] x_{t-k} - \sum_{k=0}^{\infty} E_\nu \left[\alpha^{k+1}\right] x_{t-k}, \tag{6.24}$$

where $E_\nu[\cdot]$ stands for the expectation with respect to the distribution of the memory parameter among agents, which has pdf $\nu(\alpha)$. For a large but finite number of agents, endogenous noise enters the price equation via

$$x_t = \mu_{t-1} + \sigma_t \lambda u_t \tag{6.25}$$

where λ is a measure for the scale of endogenous noise which can either result from the presence of herding or a finite number of agents, and $\{u_t\}$ a sequence of independent random variables with mean zero and unit variance (cf. Eqs (6.12) and (6.20)). From Eqs (6.24) and (6.25) we obtain:

$$\begin{aligned}x_t &= \sum_{k=0}^{\infty} E_\nu \left[\alpha^k\right] x_{t-k-1} - \sum_{k=0}^{\infty} E_\nu \left[\alpha^{k+1}\right] x_{t-k-1} + \sigma_t \lambda u_t \\ &= x_{t-1} - \sum_{k=1}^{\infty} E_\nu \left[\alpha^k\right] (x_{t-k} - x_{t-k-1}) + \sigma_t \lambda u_t.\end{aligned}$$

It follows that the increments $\Delta x_t = x_t - x_{t-1}$ are described by:

$$\Delta x_t = -\sum_{k=1}^{\infty} E_\nu[\alpha^k] \Delta x_{t-k} + \sigma_t \lambda u_t.$$

Upon defining L as the lag operator i.e. $Lx_t = x_{t-1}$, we can reformulate to obtain:

$$\left(\sum_{k=0}^{\infty} E_\nu \left[\alpha^k \right] L^k \right) \Delta x_t = \sigma_t \lambda u_t. \tag{6.26}$$

If α is assumed to be uniformly distributed on $[0, 1]$, then we have $E_\nu \left[\alpha^k \right] = \frac{1}{1+k}$ (see e.g. Linden, 1999, who studies the aggregated AR(1) process with uniformly distributed coefficients). As the latter does not decay exponentially when k tends to infinity, we observe long memory behavior for $\{\Delta x_t\}$. Thus for $\alpha \sim \text{UNIF}(0, 1)$, we have:

$$\left(\sum_{k=0}^{\infty} \frac{L^{k+1}}{k+1} \right) \Delta x_t = L \sigma_t \lambda u_t,$$

or equivalently:

$$-\ln(1 - L) \Delta x_t = \sigma_{t-1} \lambda u_{t-1},$$

where the operator $\ln(1 - L)$ on the left hand side should be interpreted in terms of its power series in L.

More generally, we may consider a memory parameter distributed according to a BETA(a, b) distribution on $(0, 1)$, with pdf

$$\nu(\alpha) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \alpha^{a-1} (1-\alpha)^{b-1}, \qquad 0 < \alpha < 1.$$

The expectation $E_\nu[\alpha^k]$ can then be expressed as

$$E_\nu[\alpha^k] = \frac{\Gamma(a+b)}{\Gamma(a)} \frac{\Gamma(a+k)}{\Gamma(a+b+k)}.$$

Although we will leave a detailed analysis of the dynamics for some of these more general cases for future research, we will make some remarks here.

For $a = d$, $b = 1 - d$, with $d \in (0, 1)$, one finds

$$E_\nu[\alpha^k] L^k = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)} L^k.$$

Since these correspond to the terms of a Taylor expansion of $(1 - L)^{-d}$ around $L = 0$, the sum in Eq. (6.26) can be written in terms of the following fractionally integrated process:

$$(1-L)^{-d}\Delta x_t = \sigma_t \lambda u_t.$$

Similar processes, when driven by IID normal innovations, are known as fractionally integrated brownian motion. For $d \in (0, \frac{1}{2})$ the latter is invertible and stationary, and the autocorrelation coefficients are negative (see Hosking, 1981), a property which is commonly observed in exchange rate returns, for example. Although Hosking refers to these processes as being short-memory, both the autocorrelation function and partial autocorrelation function decay hyperbolically to zero. Therefore fractionally integrated brownian motion is often referred to as being long-range dependent (see e.g. Campbell *et al.*, 1997). The fact that we find a related process is in line with the results of Granger (1980), who showed that aggregation of simple processes with different characteristic time scales can easily lead to long memory.

However, considering σ_t to be time independent would be an oversimplification. Although the variance σ_t^2 of beliefs among agents using the same memory parameter α converges to a constant σ^2, the belief parameters θ_t for agents with different memory parameters α are centered around different values of μ_t, which depend on α. The conditional variance of the θ_t for the population as a whole is therefore given by

$$\sigma_t^2 = \sigma^2 + \mathrm{Var}_\nu(\mu_t(\alpha))$$

where the variance of $\mu_t(\alpha)$ is taken with respect to the distribution $\nu(\alpha)$ of the memory parameter α among the population. For this variance one finds

$$\begin{aligned}\mathrm{Var}_\nu(\mu_t(\alpha)) &= \mathrm{Var}_\nu\left((1-\alpha)\textstyle\sum_{k=0}^{\infty}\alpha^k x_{t-k}\right)\\ &= E_\nu\left[\left((1-\alpha)\textstyle\sum_{k=0}^{\infty}\alpha^k x_{t-k}\right)^2\right] - \mu_t^2,\end{aligned}$$

where, as before, μ_t is given by Eq. (6.23). The first term on the right hand side can alternatively be expressed as

$$E_\nu\left[\left((1-\alpha)\sum_{k=0}^{\infty}\alpha^k x_{t-k}\right)^2\right] = E_\nu\left[(1-\alpha)^2\sum_{m=0}^{\infty}\alpha^m\sum_{n=0}^{m}x_{t-n}x_{t-m+n}\right].$$

Although this is a closed form expression of the second moment of $\mu_t(\alpha_i)$ with respect to $\nu(\alpha)$, it seems that this term cannot be expressed in a form which permits simple recursive updating as new prices x_t become available. It does however show that the dispersion of beliefs among agents obeys a non-trivial time dependence in the presence of heterogeneity in memory, and that the second moment of the innovations governing the price dynamics will exhibit a likewise complexity.

6.4 Empirical validation

As a framework for modelling the stochastic dynamics of prices and beliefs explicitly, the CBS aims to make insightful how the stylized facts of observed prices may be traced back to the unobservable diversity of beliefs, and vice versa. In the previous section we examined the specific role of three different behavioural aspects. By keeping the basic CBS as simple as possible, the three extended behavioral models were shown to relate directly to familiar econometric time-series models. The basis of the models was found to be the ARIMA model. Interestingly, a-synchronous updating of beliefs was found to translate into GARCH-type structure, and long range dependence (or long memory) emerged for agents employing different time horizons in evaluating past performance of beliefs (i.e. heterogeneity in memory).

As an immediate result, we find that the parameters of these well-known models from econometrics suddenly have an economic and behavioral interpretations attached. Although tempting, interpreting these hidden behavioral parameters directly from the data seems optimistic in this stage. We are aware of the fact that several relevant behavioral model aspects and economic variables have been ignored. The addition of relevant other features to the model is likely to affect the GARCH structure, or to alter the long range dependence. Our models do not capture, for example, longer range effects as described by Taylor and Allen (1992). Note, however, that the idea that the degree of fractional integration is a proxy for the degree of heterogeneity in memory finds support in Dacorogna (2002).

In this light there is little interest in focusing on the *exact* values of the parameter estimates. However, it will be interesting to see whether the signs of the MA coefficients and roughly the size of the GARCH parameters are both consistent with what is predicted by the theory i.e. the stylized CBSes. Finding negative values for the memory parameter, for example, is likely not to win support for our CBS. The predicted effects of the behavioral assumptions are put to the test in this empirical validation. Focus is in particular on the model obtained in the presence of a-synchronous updating (see Eqs. 6.21 and 6.22). The

Country	MA(1)	ARCH	GARCH	$\text{ARCH}_{\max}^{\textit{Implied}}$
Belgium	-0.05	0.17	0.24	0.18
Germany	0.00	0.15	0.60	0.24
Finland	-0.05	0.05	0.95	0.05
France	-0.03	0.05	0.91	0.08
Austria	-0.03	0.06	0.90	0.09
Singapore	-0.08	0.08	0.91	0.08

Table 6.1: *Estimates of an MA(1) model with GARCH(1,1) disturbances for six exchanges rate returns, and implied ARCH upper limit (last column). The predicted MA coefficient was $-\psi$ with ψ non-negative (Eq. 6.15), implying non-positive MA coefficients, as found in all cases.*

ARCH and GARCH coefficients that emerge in that example are given by $\delta(1-\delta)$ and δ, respectively, where $(1-\delta)$ denotes the probability for each agent to update his belief at any given period in time. This behavioural aspect of the model thus predicts a relation between the ARCH and the GARCH coefficients, which can be checked empirically. Moreover, the model predicts a negative MA(1) coefficient equal to $-\delta$. Hereby we have set memory to zero, leaving us with a model that is more tractable, but also with one degree of freedom less. Keep in mind, however, that in the case of herding we also predicted a negative MA(1) coefficient, equal to $-\alpha$ (see Eq. 6.13).

The data used to illustrate the model are quotes from the foreign exchange market. We consider the daily exchange rate of the US dollar against the local currencies of six different countries. Five of these countries are European: Belgium, Germany, Finland, France, and Austria, while the last is from South East Asia, namely Singapore. Let p_t denote the amount of US dollars one can buy for one unit of the local currency at time t. Instead of absolute prices, however, we consider the daily logarithms i.e. $x_t = \log p_t$, so that Δx_t denotes the daily log return on the exchange rate. The sample period, from 5-11-1987 until 5-11-2002, includes a total of 3914 observations.

Table 6.1 reports the estimates of an MA(1) model with GARCH(1,1) disturbances for all six of the exchanges rate returns. The first three columns list the actual estimates, whereas the fourth column reports the *implied* upper limit for the ARCH coefficient i.e. the upper limit implied by our behavioral model which predicts a relation between the ARCH and GARCH coefficients (see Eq. 6.22, and the discussion following it). According to the latter equation, a GARCH coefficient of δ implies an ARCH coefficient no larger than $\delta(1-\delta)$. A first look at these results shows that the size of the estimated ARCH effect indeed satisfies this criterion obtained in the case of a-synchronous updating of beliefs.

With only one degree of freedom, namely the extent to which beliefs are updated in an a-synchronous fashion, it should not be surprising that the exact size of the estimated MA coefficient is not consistent with the (G)ARCH parameters. Apart from having set α to zero, the setup is also stylized in other respects. With that in mind, it is encouraging to see that the signs of the estimated MA coefficient are also consistent with what we would expect from the theory. In sum, in light of all limitations, we consider the empirical validation successful and hope that it may open doors to more informative empirical assessments by means of more sophisticated CBSes.

6.5 Concluding remarks

The objective of this chapter was to make insightful how certain behavioral features may render some of the well-established stylized facts of observed prices. In order of treatment, we considered: (i) herd behaviour; (ii) a-synchronous updating of beliefs; and (iii) heterogeneity in the time horizons (memory) of agents. The adopted framework is that of the Continuous Beliefs System (CBS) (see Diks and van der Weide, 2003). We used two qualities of a CBS that are particularly useful. First, it enables modelling the evolution of the distribution of beliefs in an explicit fashion. Second, it does so by relating the unpredictable nature of prices with that of the unpredictable nature of individual preferences. In sum, we used the fact that a CBS allows for describing the stochastic dynamics of both prices and beliefs explicitly by means of a Random Dynamical System.

By keeping our benchmark CBS illustrative but deliberately simple, the stochastic dynamics of each of the three behavioral models was shown to correspond directly to a model well-known from time-series econometrics. We came across ARIMA, ARFIMA, and GARCH. This marks our first result, as the stylized facts for which these models are celebrated may now be traced back to the hidden features of beliefs which we hold responsible. Here, the ultimate simplicity of the benchmark CBS helped us to isolate the role of each of the three behavioural features considered.

As we summarize the results for the three models seperately, keep in mind that our point of departure was a model in which the degree of heterogeneity is constant, and where price dynamics is described by a random walk. For each behavioral extension we investigated how it modifies both the dynamics of prices and that of heterogeneity.

The effect of herd behaviour is roughly that of reducing the number of agents in the market. As such, it increases the amount of endogenous uncertainty, and thereby the level of volatility. Moreover, it adds moving average structure, leading to an ARIMA specification in levels or MA in differences. Notably, it does not alter the dynamics of heterogeneity. The latter remains constant.

Fluctuations do enter the dynamics of heterogeneity when we allow agents to update their beliefs at different frequencies. These fluctuations are found to be driven by the differences in average beliefs between those who decide to evaluate and update, and those who believe that updating can wait. If endogenous uncertainty persists in the market, for example as a result of herd behaviour, a time-varying degree of heterogeneity results into a time-varying level of volatility. The extent of persistence in volatility is governed by the likelihood that individual agents delay the updating of their beliefs. This result finds support in an earlier study by Brock and LeBaron (1996), who saw the persistence in volatility of simulated price returns rise to a more realistic level when allowing for less frequent updating. Interestingly, we obtained an explicit description of how the updating of beliefs interacts with the volatility process. For our stylized but yet intuitive example we found the dynamics that to be exactly that of a GARCH(1,1) model.

Heterogeneity in the time horizons (memory) of agents was found to enrich the dynamics of both the first and second moment. Assuming a distribution in memory rather than identical memory for each agent, we encountered one more stylized fact, namely that of long-range dependence. Leaving aside the volatility process for a moment, taking a BETA$(d, 1-d)$ distribution for memory, we obtained the well- known fractionally integrated process documented by e.g. Hosking (1981) for the first moment. When delving into greater detail, we quickly stumbled upon greater complexity. In particular, the degree of heterogeneity was found to move in mysterious ways. Although the heterogeneity among those who employ identical memory is constant over time, additional dispersion of opinions is present which can be attributed exclusively to the fact that not all agents operate at the same time horizon. The latter does not only add to the overall degree of heterogeneity, it was also found to be time-varying.

To understand the complexity, note that the supply of fluctuations is provided by the dispersion of average beliefs across groups of agents that are operating at different time horizons. As memory plays an important role in the updating of beliefs, we find that the evolution of average beliefs for each of those groups of agents interacts differently with the price dynamics, given that their memories are different. Heterogeneity being driven by the diversity of these evolutions, it requires aggregation of these different dynamics to obtain the dynamics of heterogeneity. The latter, although interesting, easily becomes convoluted.

At the end of the chapter we included a modest empirical validation to address the intimate relation between our behavioural models and the more familiar models from econometrics. As a corollary, the parameters of these time-series models now have an economic and/or behavioural interpretation attached. For example, our stylized model predicts a relation between the ARCH and GARCH parameters. Merely as a 'reality check', we verified whether both that relation, and the signs of the other coefficients, finds support in the empirical data. In light of all simplifications, we consider the empirical validation successful, and hope that more informative empirical assessments by means of more sophisticated CBSes can be made in the future.

As a final remark we would like to comment on the difference with the recently proposed LTL approach of Brock *et al.* (2005). Although a CBS is more explicit in terms of the beliefs distribution, the price dynamics of a CBS and an LTL are closely related. (For a more elaborate comparison, see Diks and van der Weide (2003).) In case the opportunity function $\varphi(\theta)$ (see footnote 2) is normalizable, the LTL coincides with the deterministic part of the corresponding CBS. However, there are two key differences. For one, it is not clear how an LTL can be set up with a non-normalizable opportunity function (which for an LTL corresponds to the pdf according to which strategies are drawn at random). More importantly, an LTL is a deterministic dynamical system by construction, and up until now not tailored to dealing with endogenous randomness. It is exactly the associated stochastic dynamics that plays a leading role in this chapter.

Samenvatting (Summary in Dutch)

Het proefschrift bestaat, naast het inleidende hoodstuk, uit vijf hoofdstukken. Het eerste gedeelte introduceert een model met als doel om empirische data van financiële markten beter te beschrijven, het tweede gedeelte introduceert een model met als doel om dezelfde data beter te begrijpen. Hoofdstuk twee en drie concentreren zich metname op het econometrisch modelleren van (multivariate) financiële volatiliteit. Hoofdstuk vier fungeert als een brug tussen de twee verschillende delen van mijn proefschrift. Dit hoofdstuk doet een poging om marktverwachtingen over toekomstige financiële volatiliteit te schatten uit empirische optieprijs data. In de laatste twee hoofdstukken ontwikkelen we een structureel economisch model waarin de manier waarop economische agenten verwachtingen vormen over toekomstge prijzen centraal staat. Hoofdstuk vijf introduceert een analytisch raamwerk dat is ontworpen om de co-evolutie van marktverwachtingen en prijzen te modelleren. Hoofdstuk zes behandelt een aantal speciale gevallen van dit raamwerk door de nadruk te leggen op een selectie van verschillende gedragskenmerken van de economische agenten (in het bijzonder de manier waarop ze verwachtingen vormen).

In hoofdstuk twee wordt een nieuw multivariaat volatiliteitmodel geintroduceert, naar welke gerefereerd zal worden als Gegeneraliseerd Orthogonaal GARCH (GO-GARCH). Het is een lid van de ARCH-familie, waar ARCH voor Auto-Regressieve Conditionele Heteroskedasticiteit staat. Het ARCH-model is tot op heden waarschijnlijk het meest populaire model dat wordt gebruikt voor het beschrijven van de bewegelijkheid van financiële volatiliteit over de tijd. Het model werd geïntroduceerd door Engle in 1982 en is kort daarna gegeneraliseerd tot GARCH door Bollerslev (in 1986). De multivariate uitbreiding, multivariate GARCH, voorziet in een modelspecificatie voor de tijd-variatie van de covariantie matrix (met tijd-variërende varianties op- en covarianties van de diagonaal). Schattingen van multivariate GARCH modellen kunnen echter problematisch zijn omdat het aantal onbekende model parameters die moeten worden geschat snel toenemen met de dimensie.

De eerste algemene multivariate GARCH modellen, welke naar voren zijn gebracht in de literatuur, gebruiken een relatief groot aantal parameters welke tot convergentie moeilijkheden leiden bij het schatten (zie bv. Bauwens e.a., 2006). Nieuwe specificaties worden vaak bepaald door middel van praktische overwegingen. De uitdaging is om een parameterisatie van de covariantie matrix te vinden die uitvoerbaar is in schattingstermen tegen een minimaal verlies van algemeenheid.

Het GO-GARCH model verwijst naar een natuurlijke generalisatie van het O-GARCH model, en nestelt zich als een speciaal geval in het meer algemene BEKK model. Het geeft ruimte aan de belangrijkste gestileerde feiten van multivariate volatiliteit terwijl de uitvoerbaarheid behouden blijft. Zowel kunstmatige als empirische voorbeelden zijn opgenomen om het model te illustreren. Voor de gepubliceerde versie van het hoofdstuk zie van der Weide (2002).

In hoofdstuk drie introduceren we een nieuwe schattingsmethode voor voor het GO-GARCH model. De methode maakt gebruik van de eigenvectoren van een geschikte selectie van steekproef autocorrelatie matrices van gekwadrateerde en kruisvermenigvuldigingen van de rendementen data. De methode is numeriek aantrekkelijker dan de originele schattings methode. In het bijzonder vereist de nieuwe methode geen strikte veronderstellingen omtrent de volatiliteitmodellen van de latente factoren, en is daarom minder gevoelig voor model misspecificatie. Het hoofdstuk voorziet zowel in Monte-Carlo simulaties, en in een studie van empirische data van Europese sector rendementen. Voor de gepubliceerde versie zie Boswijk en van der Weide (2011).

Ingebed in optieprijzen zitten marktverwachtingen over toekomstige volatiliteit. Hoewel de veronderstelling van rationele verwachtingen een populair paradigma is, is het moeilijk om de subjectieve aard van verwachtingen te negeren. Het doel van hoofdstuk vier is om marktverwachtingen zichtbaar te maken terwijl zij zich ontwikkelen over de tijd, en om opties te prijzen consistent met de heersende verwachtingen, of deze nu rationeel of niet-rationeel zijn. We werken met een analytisch passend raamwerk voor het prijzen van opties dat zowel voorziet in stochastische volatiliteit als a-symmetrische volatiliteit. Dagelijkse schattingen van de geïmpliceerde verdeling van volatiliteit zijn verkregen door, dag voor dag, het model op optie prijzen te schatten. We leggen geen structuur op hoe verwachtingen worden bijgesteld door de tijd heen, maar laten toe dat marktverwachtingen vrij kunnen bewegen. Zie Peters en van der Weide (2011) voor de working paper versie.

Met hoofdstuk vijf concentreren we ons op het economisch modelleren van het proces waarmee agenten hun verwachtingen vormen. We introduceren een nieuw analytisch raamwerk voor het bestuderen van de co-evolutie van verwachtingen en prijzen. Agenten baseren hun keuzes op resultaten uit het verleden en her-evalueren strategieën zodra nieuwe informatie beschikbaar komt. Door individuele keuzes te beschouwen als stochastische variabelen, wat natuurlijk is in een stochastische preferentie raamwerk, kan heterogeniteit gezien worden als een natuurlijke bron van stochasticiteit in de dynamica van prijzen. We beschouwen enkele voorbeelden waarvoor deze stochasticiteit niet verdwijnt zelfs als het aantal agenten naar oneindigheid gaat. Het hoofdstuk is gepubliceerd als een working paper, zie Diks en van der Weide (2003).

In hoofdstuk zes beschouwen we een specifiek eenvoudig voorbeeld van het raamwerk dat geïntroduceerd is in hoofdstuk vijf met het doel om de effecten te onderzoeken op prijsdynamieken van enkele gedrags veronderstellingen: (i) kuddegedrag; (ii) a-synchroon herzien van verwachtingen; en (iii) heterogeniteit in tijds-horizonnen (geheugen) onder agenten. Het benchmarkmodel met veel handelaren geeft een “random walk” gedreven door nieuws. Door kuddegedrag te introduceren wordt dit proces aangepast tot een ARIMA (0,1,1) proces. In termen van rendementen voorspelt het model MA(1) structuur met een negatieve coëfficiënt. A-synchroon herzien van verwachtingen leidt tot een MA(1) model voor rendementen met GARCH(1,1) innovaties, en voorspelt een relatie tussen de ARCH en GARCH coëfficiënten. Heterogeniteit in geheugen leidt tot lange termijn afhankelijkheid in rendementen. In het empirische deel van het hoofdstuk voeren we een bescheiden ‘reality-check’ uit met betrekking tot het verwachte teken van de MA coëfficiënt en de relatie tussen de ARCH en GARCH coëfficiënten voor wisselkoersdata. Voor de gepubliceerde versie van het hoofdstuk, zie Diks en van der Weide (2005).

References

Alexander, C. (1998). Volatility and correlation: Methods, models and applications. *In Risk Management and Analysis: Measuring and Modelling Financial Risk, John Wiley, New York* 125–172.

Alexander, C. (2001). Orthogonal GARCH. *In Mastering Risk, Financial TimesPrentice Hall, London*, **2,** 21–38.

Alexander, C. (2002). Principal component models for generating large GARCH covariance matrices. *Review of Banking, Finance and Monetary Economics*, **31 (2),** 337–359.

Alexander, C. and Chibumba, A. (1996). Multivariate orthogonal factor GARCH. *University of Sussex Discussion Papers in Mathematics.*

Alghalith, M. (2010). The interaction between food prices and oil prices. *Energy Economics*, **32 (6),** 1520–1522.

Allen, H. and Taylor, M. (1990). Charts, noise and fundamentals in the london foreign exchange market. *The Economic Journal*, **100,** 49–59.

Andersen, L. Benzoni T. and Lund, J. (2002). An empirical investigation of continuous-time equity return models. *Journal of Finance*, **57,** 1239–1284.

Anderson, T. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *Journal of Finance*, **51 (1),** 169–204.

Anderson, T., Bollerslev, T., Diebold, F. and Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, **61 (1),** 43–76.

Anderson, T., Bollerslev, T., Diebold, F. and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, **71 (2),** 579–625.

Ang, A. and Bekaert, G. (2002). International asset allocation under regime switching. *Review of Financial Studies*, **15,** 1137–1187.

Ang, A. and Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, **63,** 443–494.

Arnold, L. (1998). *Random Dynamical Systems.* Berlin: Springer Verlag.

Avery, C. and Zemsky, P. (1998). Multidimensional uncertainty and herd behavior in financial markets. *American Economic Review*, **88,** 724–748.

Bakshi, C. Cao G. and Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of Finance*, **94,** 277–318.

Ball, C. and Roma, A. (1994). Stochastic volatility option pricing. *Journal of Financial and Quantitative Analysis*, **29,** 589–607.

Barberis, N. and Thaler, R. (2002). A survey of behavioral finance. *NBER Working Paper Series*, **9222,** 1–77.

Barndorff-Nielsen, O. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society - Series* B, **64,** 253–280.

Bates, D. (2000). Post-'87 crash fears in the s&p 500 futures option market. *Journal of Econometrics*, **94 (1),** 181–238.

Bauwens, L., Laurent, S. and Rombouts, J. (2006). Multivariate garch models: A survey. *Journal of Applied Econometrics*, **21 (1),** 79–109.

Beaghen, M. (1997). *Canonical Variate Analysis and Related Methods with Longitudinal Data.* PhD Dissertation, Virginia Polytechnic Institute and State University.

Bekaert, G., Campbell, R. and Ng, A. (2005). Market integration and contagion. *Journal of Business*, **78,** 39–69.

Ben-Akiva, M. and Watanatada, T. (1981). Application of a continuous spatial choice logit model. In *Structural Analysis of Discrete Data and Econometric Applications* (eds C. F. Manski and D. L. McFadden), pp. 320–343. MIT Press, Cambridge, MA.

Bernanke, B. (1983). Irreversibility, uncertainty, and cyclical investment. *Quarterly Journal of Economics*, **98 (1),** 85–106.

Bikhchandani, S., Hirschleifer, D. and Welch, I. (1992). A theory of fads, fashion, custom and cultural change as informational cascases. *Journal of Political Economy*, **100,** 992–1027.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, **81,** 637–654.

Bollerslev, Chou R. T. and Kroner, K. (1992). Arch modeling in finance: A selective review of the theory and empirical evidence. *Journal of Econometrics*, **52,** 5–59.

Bollerslev, Engle R. T. and Wooldridge, J. (1988). A capital asset pricing model with time varying covariances. *Journal of Political Economy*, **96,** 116–131.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31 (3),** 307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach. *Review of Economics and Statistics*, **72,** 498–505.

Bollerslev, T. (2008). Glossary to ARCH (GARCH). *CREATES Research Paper*, **2008-49**.

Boswijk, H.P., Hommes, C. and Manzan, S. (2007). Behavioral heterogeneity in stock prices. *Journal of Economic Dynamics and Control*, **31 (6),** 1938–1970.

Boswijk, H.P. and van der Weide (2009). Method of moments estimation of GO-GARCH models. *UvA Econometrics Discussion Paper*, **2009/05**.

Boswijk, H.P. and van der Weide, R. (2006). Wake me up before you GO-GARCH. *UvA-Econometrics Discussion Paper*, **2006/03**.

Boswijk, H.P. and van der Weide, R. (2008). Testing the number of factors in GO-GARCH models. *Working Paper, University of Amsterdam.*

Boswijk, H.P. and van der Weide, R. (2011). Method of moments estimation of GO-GARCH models. *Journal of Econometrics*, **163 (1),** 118–126.

Boussama, F. (1998). Ergodicity, mixing and estmation in GARCH models. *Ph.D. dissertation, University Paris* 7.

Brennan, M. (1979). The pricing of contingent claims in discrete-time models. *Journal of Finance*, **34,** 53–68.

Brock, W. and Hommes, C. (1997). A rational route to randomness. *Econometrica*, **65,** 1059–1095.

Brock, W. and Hommes, C. (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics and Control*, **22,** 1235–1274.

Brock, W., Hommes, C. and Wagener, F. (2005). Evolutionary dynamics in markets with many trader types. *Journal of Mathematical Economics*, **41,** number 1-2, 7–42.

Brock, W., Hommes, C. and Wagener, F. (2009). More hedging instruments may destabilize markets. *Journal of Economics Dynamics and Control*, **33,** 1912–1928.

Brock, W. A. and Durlauf, S. N. (2001). Discrete choice with social interactions. *Review of Economic Studies*, **68,** 235–260.

Brock, W. A. and LeBaron, B. (1996). A dynamic structural model for stock return volatility and trading volume. *Review of Economics and Statistics*, **78,** 94–110.

Buraschi, A. and Jiltsov, A. (2006). Model uncertainty and option markets with heterogeneous beliefs. *Journal of Finance*, **61,** 2841–2898.

Burnside, C., Eichenbaum, M. and Robelo, S. (2011). Understanding booms and busts in housing markets. *NBER Working Paper*, **16734**.

Busch, B. Christensen T. and Nielsen, M. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics*, **160 (1),** 48–57.

Cabrales, A. and Hoshi, T. (1996). Heterogeneous beliefs, wealth accumulation, and asset price dynamics. *Journal of Economic Dynamics and Control*, **20,** 1073–1100.

Cameron, K. and Schnusenberg, O. (2009). Oil prices, SUVs, and Iraq: An investigation of automobile manufacturer oil price sensitivity. *Energy Economics*, **31,** 375–381.

Campbell, J. Y., Lo, A. W. and MacKinley, A. C. (1997). *The Econometrics of Financial Markets.* Princeton University Press, Princeton, NJ.

Christoffersen, P. and Jacobs, K. (2004). The importance of the loss function in option valuation. *Journal of Financial Economics*, **72,** 291–318.

Comte, F. and Lieberman, O. (2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis*, **84,** 61–84.

Constantinides G., J. Jackwerth and Perrakis, S. (2009). Mispricing of s&p 500 index options. *The Review of Financial Studies*, **22 (3),** 1247–1277.

Cox, S. Ross J. and Rubinstein, M. (1979). Optioin pricing: A simplified approach. *Journal of Financial Economics*, **7,** 229–263.

Cyert, R. and DeGroot, M. (1974). Rational expectations and bayesian analysis. *Journal of Political Economy*, **82 (3),** 521–536.

Dacorogna, M. (2002). Scaling analysis to characterize financial markets. Technical Report. Paper presented at the CeNDEF workshop, Leiden, The Netherlands, June 2002. See http://www.fee.uva.nl/cendef.

Dagsvik, J. K. (1994). Discrete and continuous choice, max-stable processes, and independence from irrelevant attributes. *Econometrica*, **62,** number 5, 1179–1205.

Dagsvik, J. K. (2002). Discrete choice in continuous time: Implications of an intertemporal version of the IIA property. *Econometrica*, **70,** 817–831.

Das, S. and Uppal, R. (2004). Systematic risk and international portfolio choice. *Journal of Finance*, **59,** 2809–2834.

Deaton, A. and Laroque, G. (1992). On the behaviour of commodity prices. *Review of Economic Studies*, **59 (1),** 1–23.

Diebold, F. and Nerlove, M. (1989). The dynamics of exchange rate volatility: A multivariate latent factor ARCH model. *Journal of Applied Econometrics*, **4,** 1–21.

Diether, K. B., Malloy, C. J. and Scherbina, A. (2002). Differences of opinion and the cross section of stock returns. *Journal of Finance*, **57,** 2113–2141.

Diks, C. and van der Weide, R. (2003). Heterogeneity as a natural source of randomness. *Tinbergen Discussion Paper*, **073/1**.

Diks, C. and van der Weide, R. (2005). Herding, a-synchronous updating and heterogeneity in memory in a cbs. *Journal of Economic Dynamics and Control*, **29,** 741–763.

Diks, C. and van der Weide, R. (2011). A role for prior beliefs in explaining price movements. *mimeo.*

Ding, Z. (1994). *Time Series Analysis of Speculative Returns*. PhD Dissertation, University of California at San Diego.

Dotsey, M. and Reid, M. (1992). Oil shocks, monetary policy, and economic activity. *Economic Review*, **78,** 14–27.

Doz, C. and Renault, E. (2006). Factor stochastic volatility in mean models: A GMM approach. *Econometric Reviews*, **25,** 275–309.

Drost, F. and Nijman, T. (1993). Temporal aggregation of GARCH processes. *Econometrica*, **61,** 909–927.

Dufwenberg, M., Lindqvist, T. and Moore, E. (2005). Bubbles and experience: An experiment. *American Economic Review*, **95 (5),** 1731–1737.

Dumas, B., Fleming, J. and Whaley, R. (1998). Implied volatility functions: Empirical tests. *Journal of Finance*, **53 (6),** 2059–2106.

Elder, J. and Serletis, A. (2009). Oil price uncertainty in canada. *Energy Economics*, **31,** 852–856.

Elder, J. and Serletis, A. (2010). Oil price uncertainty. *Journal of Money, Credit and Banking*, **42 (6),** 1137–1159.

Engle, Ng V. R. and Rothschild, M. (1990). Asset pricing with a factor ARCH covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics*, **45,** 213–238.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50 (4),** 987–1007.

Engle, R. (2002). Dynamic conditional correlation. *Journal of Business and Economic Statistics*, **20 (3),** 339–350.

Engle, R. (2004). Risk and volatility: Econometric models and financial practice. *American Economic Review*, **94 (3),** 405–420.

Engle, R., Ghysels, E. and Sohn, B. (2009). On the economic sources of stock market volatility. *NYU Finance Working Papers*, **FIN-08-043**.

Engle, R. and Kroner, K. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, **11,** 122–150.

Engle, R. and Rangel, J. (2008). The spline GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies*, **21 (3),** 1187–1222.

Evans, G. and Honkapohja, S. (2003). Expectations and the stability problem for optimal monetary policies. *Review of Economic Studies*, **70 (4),** 807–824.

Fan, J., Wang, M. and Yao, Q. (2008). Modelling multivariate volatilities via conditionally uncorrelated components. *Journal of the Royal Statistical Society Series* B, **70,** 679–702.

Flood, R. and Garber, P. (1980). Market fundamentals versus price-level bubbles: The first tests. *Journal of Political Economy*, **88 (4),** 745–770.

Flury, B. (1984). Common principal components in k groups. *Journal of the American Statistical Association*, **79,** 892–898.

Flury, B. and Gautschi, W. (1986). An algorithm for simultaneous orthogonal transformation of several positive definite matrices to nearly diagonal form. *SIAM Journal of Scientific and Statistical Computing*, **7,** 169–184.

Frankel, J. and Froot, K. (1987). Using survey data to test standard propositions regarding exchange rate expectations. *American Economic Review*, **77 (1),** 133–153.

Frankel, J. and Froot, K. (1990). Chartists, fundamentalists, and trading in the foreign exchange market. *American Economic Review*, **80 (2),** 181–185.

Frankel, J. A. and Froot, K. A. (1988). Chartists, fundamentalists and the demand for dollars. *Greek Economic Review*, **10,** 49–102.

Froot, K., Scharfstein, D. and Stein, J. (1992). Heard on the street: Information inefficiencies in a market with short-term speculators. *Journal of Finance*, **47,** 1461–1484.

Gallant, A. R., Rossi, P. E. and Tauchen, G. (1992). Stock prices and volume. *Review of Financial Studies*, **5,** 199–242.

Garber, P. (1989). Tulipmania. *Journal of Political Economy*, **97 (3),** 535–560.

Garber, P. (1990). Famous first bubbles. *Journal of Economic Perspectives*, **4 (2),** 35–54.

Garcia, E. Ghysels R. and Renault, E. (2003a). The econometrics of option pricing. *Working Paper, CIRANO, CIREQ.*

Garcia, E. Ghysels R. and Renault, E. (2010). The econometrics of option pricing. *Handbook of Financial Econometrics, Elsevier-North Holland, Amsterdam* 479–552.

Garcia, R. and Renault, E. (2001). Latent variable models for stochastic discount factors. *Handbooks in Mathematical Finance: Topics in Option Pricing, Interest Rates and Risk Management, Cambridge University Press.*

Garcia, R. Luger R. and Renault, E. (2003b). Empirical assessment of an intertemporal option pricing model with latent variables. *Journal of Econometrics*, **116,** 49–83.

Gaunersdorfer, A., Hommes, C. and Wagener, F. (2008). Bifurcation routes to volatility clustering under evolutionary learning. *Journal of Economic Behavior and Organization*, **67,** 27–47.

Granger, C. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, **14,** 227–338.

Greenwood, R. and Nagel, S. (2009). Inexperienced investors and bubbles. *Journal of Financial Economics*, **93,** 239–258.

Guesnerie, R. (2002). Anchoring economic prediction in common knowledge. *Econometrica*, **70,** 439–480.

Hafner, C. (2003). Fourth moment structure of multivariate GARCH models. *Journal of Financial Econometrics*, **1,** 26–54.

Hafner, C. (2008). Temporal aggregation of multivariate GARCH processes. *Journal of Econometrics*, **142,** 467–483.

Hamilton, J. (1983). Oil and the macroeconomy since World War II. *Journal of Political Economy*, **91 (2),** 228–248.

Hamilton, J. and Lin, G. (1996). Stock market volatility and the business cycle. *Journal of Applied Econometrics*, **11 (5),** 573–593.

Han, B. (2008). Investor sentiment and option prices. *The Review of Financial Studies*, **21 (1),** 387–414.

Hanemann, W. M. (1984). Discrete/continuous models of consumer demands. *Econometrica*, **52,** number 3, 541–561.

Haruvy, E., Lahav, Y. and Noussair, C. (2007). Traders' expectations in asset markets: Experimental evidence. *American Economic Review*, **97 (5),** 1901–1920.

He, C. and Terasvirta, T. (1999). Properties of moments of a family of GARCH processes. *Journal of Econometrics*, **92,** 173–192.

Heady, E. and Kaldor, D. (1954). Expectations and errors in forecasting agricultural prices. *Journal of Political Economy*, **62 (1)**.

Heemeijer, P., Hommes, C., Sonnemans, J. and Tuinstra, J. (2009). Price stability and volatility in markets with positive and negative expectations feedback: An experimental investigation. *Journal of Economic Dynamics and Control*, **33 (5),** 1052–1072.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, **6,** 327–343.

Hommes, C. (2001). Financial markets as nonlinear adaptive evolutionary systems. *Quantitative Finance*, **1,** 149–167.

Hommes, C. (2002). Modeling the stylized facts in finance through simple nonlinear adaptive systems. *Proceedings of the National Soceity of Science*, **99,** 7721–7728.

Hommes, C. (2006). Heterogeneous agent models in economics and finance. *Handbook of Computations Economics*, **2,** 1109–1186.

Hommes, C. and Rosser Jr, J. (2001). Consistent expectations equilibria and complex dynamics in renewable resource markets. *Macroeconomic Dynamics*, **5,** 180–203.

Hommes, C., Sonnemans, J., Tuinstra, J. and van de Velden, H. (2005). Coordination of expectations in asset pricing experiments. *Review of Financial Studies*, **18,** 955–980.

Hommes, C. and Sorger, G. (1998). Consistent expectations equilibria. *Macroeconomic Dynamics*, **2,** 287–321.

Hommes, C. and Wagener, F. (2009). Complex evolutionary systems in behavioral finance. *Handbook of Financial Markets: Dynamics and Evolution, Edited by T. Hens and K.R. Schenk-Hoppe* 217–276.

Hong, H., Kubik, J. D. and Solomon, A. (2000). Security analysts' career concerns and herding of earnings forecasts. *RAND Journal of Economics*, **31,** 121–144.

Horn, R. and Johnson, C. (1999). *Matrix Analysis*. Cambridge: Cambridge University Press.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, **68,** 165–176.

Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, **42,** 281–300.

Hussam, R., Porter, D. and Smith, V. (2008). Thar she blows: Can bubbles be rekindled with experienced subjects? *American Economic Review*, **93 (3),** 924–937.

Ito, T. (1990). Foreign exchange rate expectations: Micro survey data. *American Economic Review*, **80,** 434–449.

Jackwerth, J. and Rubinstein, M. (1996). Recovering probability distributions from option prices. *Journal of Finance*, **51 (5),** 1611–1632.

Jeantheau, T. (1998). Strong consistency of estimators for multivariate ARCH models. *Econometric Theory*, **14,** 70–86.

Jensen, S. and Rahbek, A. (2007). On the law of large numbers for (geometrically) ergodic Markov chains. *Econometric Theory*, **23,** 761–766.

Jones, C. and Kaul, G. (1996). Oil and the stock markets. *Journal of Finance*, **51,** 463–491.

Jury, E. I. (1974). *Inners and Stability of Dynamical Systems*. John Wiley & Sons, New York.

Kandel, E. and Pearson, N. D. (1995). Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy*, **4,** 831–872.

Keane, M. P. and Wolpin, K. I. (1994). The solution and estimation of discrete choice dynamic-programming models by simulation and interpolation – Monte-Carlo evidence. *Review of Economics and Statistics*, **76,** 648–672.

Kilian, L. and Murphy, D. (2010). The role of inventories and speculative trading in the global market for crude oil. *CEPR Discussion Paper*, **7753**.

Kilian, L. and Taylor, M. P. (2001). Why is it so difficult to beat the random walk forecast of exchange rates? Technical Report. ECB Working paper No. 88. European Central Bank.

Kirman, A. (1993). Ants, rationality and recruitment. *Quarterly Journal of Economics*, **108,** 137–156.

Klaassen, F. (1999). Have exchange rates become more closely tied? Evidence from a new multivariate GARCH model. *CentER Working Paper, Tilburg University.*

Kroner, K. and Ng, V. (1998). Modeling asymmetric comovements of asset returns. *The Review of Financial Studies*, **11,** 817–844.

Kurz, M. (2001). Endogenous uncertainty and market volatility. *Economic Theory*, **17,** 497–544.

Lahiri, K. and Sheng, X. (2008). Evolution of forecast disagreement in a bayesian learning model. *Journal of Econometrics*, **144 (2),** 325–340.

Lanne, M. and Saikkonen, P. (2007). A multivariate generalized orthogonal factor GARCH model. *Journal of Business and Economic Statistics*, **25,** 61–75.

LeBaron, B. (1999). Technical trading rule profitability and the foreign exchange intervention. *Journal of International Economics*, **49,** 125–143.

LeBaron, B. (2000). Agent based computational finance: Suggested readings and early research. *Journal of Economic Dynamics and Control*, **24,** 679–702.

LeBaron, B. (2001). Evolution and time horizons in an agent-based stock market. *Macroeconomic Dynamics*, **5,** 225–254.

LeBaron, B., Arthur, W. and Palmer, R. (1999). The time series properties of an artificial stock market. *Journal of Economic Dynamics and Control*, **23,** 1487–1516.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, **88,** 365–411.

Lee, R. (2004). The moment formula for implied volatility at extreme strikes. *Mathematical Finance*, **14,** 469–480.

Lehmann, B. N. (1991). Asset pricing and intrinsic values. *Journal of Monetary Economics*, **28,** 485–500.

Levich, R. M. and Thomas, L. R. (1993). The significance of technical trading-rule profits in the foreign exchange market: A bootstrap approach. *Journal of International Money and Finance*, **12,** 451–474.

Linden, M. (1999). Time series properties of aggregated AR(1) processes with uniformly distributed coefficients. *Economics Letters*, **64,** 31–36.

Ling, S. and McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory*, **19,** 280–310.

Lo, A. W. and MacKinlay, A. C. (1997). *A Non-random Walk Down Wall Street.* Princeton University Press, Princeton, NJ.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets. *Journal of Finance*, **56,** 649–676.

Lucas, R. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, **4,** 103–124.

Lux, T. (1995). Herd behaviour, bubbles and crashes. *Economic Journal*, **105,** 881–896.

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

Malkiel, B. (1990). *A Random Walk Down Wall Street.* Norton, New York.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, *New York* (ed. P. Zarembka), pp. 105–142. Academic Press, NY.

Merton, R. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, **4,** 141–183.

Michaely, R. and Vila, J.-L. (1996). Trading volume with price valuation: Evidence from the ex-dividend day. *Review of Financial Studies*, **9,** 471–509.

Milgrom, P. and Stokey, N. (1982). Information, trade, and common knowledge. *Journal of Economic Theory*, **26,** 17–27.

Mills, T. (1999). *The Econometric Modelling of Financial Time Series*, 2nd edn. Cambridge University Press, Cambridge, UK.

Muller, U. A., Dacorogna, M. M., Dave, R. R., Olsen, R. B., Pictet, O. V. and Weizsacker, J. E. Von (1997). Volatilities of different time resolutions - analyzing the dynamics of market components. *Journal of Empirical Finance*, **4,** 213–239.

Muth, J. (1961). Rational expectations and the theory of price movements. *Econometrica*, **29 (3),** 315–325.

Nelson, L. (2002). Persistence and reversal in herd behavior: Theory and application to the decision to go public. *Review of Financial Studies*, **15,** 65–95.

Nerlove, M. (1979). The dynamics of supply: Retrospect and prospect. *American Journal of Agricultural Economics*, **61 (5),** 874–888.

O'Hara, M. (2008). Bubbles: Some perspectives (and loose talk) from history. *Review of Financial Studies*, **21 (1),** 11–17.

Peters, R. and de Vilder, R. (2006). Testing the continuous semimartingale hypothesis for the s&p 500. *Journal of Business and Economic Statistics*, **24,** 444–454.

Peters, R. and van der Weide, R. (2011). Volatility: Expectations and realizations. *mimeo.*

Poon, S.H. and Granger, C. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, **41 (2),** 478–539.

Renault, E. and Touzi, N. (1996). Option hedging and implied volatilities in a stochastic volatility model. *Mathematical Finance*, **6,** 279–302.

Resnick, S. I. and Roy, R. (1994). Superextremal processes, max-stability and dynamic continuous choice. *Annals of Applied Probability*, **4,** 791–811.

Resnick, S.I. (1998). *A Probability Path.* Birkhäuser, Boston.

Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics*, **85,** 777–792.

Rubinstein, M. (1976a). The strong case for the generalized logarithmic utility model as the premier model of financial markets. *Journal of Finance*, **31**.

Rubinstein, M. (1976b). The valuation of uncertain income streams and the pricing of options. *Bell Journal of Economics and Management Science*, **7,** 407–425.

Russell, B. (1947). *Am I An Atheist Or An Agnostic?*

Sargent, T. (1993). *Bounded rationality in macroeconomics: The Arne Ryde memorial lectures*. Oxford University Press.

Scharfstein, D. and Stein, J. (1990). Herd behavior and investment. *American Economic Review*, **80 (3),** 465–479.

Schwert, G.W. (1989). Why does stock market volatility change over time? *Journal of Finance*, **44 (5),** 1115–1153.

Sentana, E. and Fiorentini, G. (2001). Identification, estimation and testing of conditionally heteroskedastic factor models. *Journal of Econometrics*, **102,** 143–164.

Shalen, C. T. (1993). Volume, volatility, and the dispersion of beliefs. *Review of Financial Studies*, **6,** 405–434.

Shiller, R. (1990). *Market volatility*. MIT Press.

Shiller, R. (2003). From efficient markets theory to behavioral finance. *Journal of Economic Perspectives*, **17 (1),** 83–104.

Shiller, R. (2005). *Irrational exuberance*. Princeton University, Princeton University Press.

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

Silvennoinen, A. and Terasvirta, T. (2009). Multivariate GARCH models. *T.G. Andersen, R.A. Davis, J.P. Kreiss and T. Mikosch, (Eds.), Handbook of Financial Time Series, Springer, New York* 201–229.

Simon, H. (1957). *Models of man, social and rational: Mathematical essays on rational human behavior in a social setting.* Garland Pub. (New York).

Smith, V., Suchanek, G. and Williams, A. (1988). Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica*, **56 (5),** 1119–1151.

Sogner, L. and Mitlohner, H. (2002). Consistent expectations equilibria and learning in a stock market. *Journal of Economic Dynamics and Control*, **26 (2),** 171–185.

Sweeney, R. J. (1986). Beating the foreign exchange market. *Journal of Finance*, **41,** 163–182.

Taylor, M. and Allen, H. (1992). The use of technical analysis in the foreign exchange market. *Journal of International Money and Finance*, **11 (3),** 304–314.

ter Ellen, S. and Zwinkels, R. (2010). Oil price dynamics: A behavioral finance approach with heterogeneous agents. *Energy Economics*, **32 (6),** 1427–1434.

Times, New York (2008). Price volatility adds to worry on U.S. farms. *The 'Food Chain' series*, **April 22**.

Tirole, J. (1982). On the possibility of speculation under rational expectations. *Econometrica*, **50 (5),** 1163–1181.

Trueman, B. (1994). Analysts forecasts and herding behavior. *Review of Financial Studies*, **7,** 97–124.

Tse, Y. (2000). A test for constant correlations in a multivariate GARCH model. *Journal of Econometrics*, **98,** 107–127.

Tse, Y. (2002). Residual-based diagnostics for conditional heteroscedasticity models. *Econometrics Journal*, **5 (2),** 358–374.

Tse, Y. and Tsui, A. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business and Economic Statistics*, **20 (3),** 351–362.

Urga, G. (2007). Common features in economics and finance: An overview of recent developments. *Journal of Business and Economic Statistics*, **25,** 2–11.

van der Weide, R. (2002). GO-GARCH: A multivariate orthogonal GARCH model. *Journal of Applied Econometrics*, **17 (5),** 549–564.

Vilenkin, N. (1968). Special functions and the theory of group representation, translations of mathematical monographs. *American Math. Soc., Providence, Rhode Island, U.S.A.*, **22**.

Vrontos, I., Dellaportas, P. and Politis, D. (2003). A full-factor multivariate GARCH model. *Econometrics Journal*, **6,** 312–334.

Yoon, K.H. and Ratti, R. (2010). Energy price uncertainty, energy intensity and firm investment. *Energy Economics*, **33 (1),** 67–78.

Ziegler, A. (2002). State-price densities under heterogeneous beliefs, the smile effect, and implied risk aversion. *European Economic Review*, **46,** 1539–1557.