

**Financial Applications of  
Human Perception of Fractal Time Series**

Daphne Sobolev

A dissertation submitted for the degree of

**Doctor of Philosophy**

of the

**University College London.**

Division of Psychology and Language Sciences

University College London

2014

## **Declaration**

I, Daphne Sobolev, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:

## Abstract

The purpose of this thesis is to explore the interaction between people's financial behaviour and the market's fractal characteristics. In particular, I have been interested in the Hurst exponent, a measure of a series' fractal dimension and autocorrelation.

In Chapter 2 I show that people exhibit a high level of sensitivity to the Hurst exponent of visually presented graphs representing price series. I explain this sensitivity using two types of cues: the illuminance of the graphs, and the characteristic of the price *change* series. I further show that people can learn how to identify the Hurst exponents of fractal graphs when feedback about the correct values of the Hurst exponent is given.

In Chapter 3 I investigate the relationship between risk perception and Hurst exponent. I show that people assess risk of investment in an asset according to the Hurst exponent of its price graph if it is presented along with its price change series. Analysis reveals that buy/sell decisions also depend on the Hurst exponent of the graphs.

In Chapter 4 I study forecasts from financial graphs. I show that to produce forecasts, people imitate perceived noise and signals of data series. People's forecasts depend on certain personality traits and dispositions. Similar results were obtained for experts.

In Chapter 5 I explore the way people integrate visually presented price series with news. I find that people's financial decisions are influenced by news more than the average trend of the graphs. In the case of positive trend, there is a correlation between financial forecasts and decisions.

Finally, in Chapter 6 I show that the way people perceive fractal time series is correlated with the Hurst exponent of the graphs. I use the findings of the thesis to describe a possible mechanism which preserves the fractal nature of price series.

# **Acknowledgements**

I would like to thank my supervisor, Nigel Harvey, for unveiling for me the challenge of psychology, for his infinite patience, guidance, encouragement, and assistance.

I would like to thank my second supervisor, Alan Johnston, for his original ideas, help, and support.

I am also grateful to my project student Bryan Chan. Chapter 5 is based on experiments performed by Bryan under my supervision.

# Contents

Declaration.....	2
Abstract.....	3
Acknowledgements.....	4
List of Figures.....	10
List of Tables.....	15
Chapter 1: Background.....	18
Introduction.....	18
Part I: The role of fractals in finance.....	20
Part II: Studies in psychology and behavioural finance.....	25
Perception of fractal time series.....	26
Risk perception and financial decisions.....	34
Judgmental forecasting from fractal time series: The effects of task instructions, personality traits, sense of power, and expertise on noise imitation.....	39
The effects of news valence, price trend and individual differences on financial behaviour.....	46
Mechanisms preserving asset price graph structure.....	53
Part III: Mathematical aspects.....	64
Definition of fBm and fGn series.....	64
Fractal series as experimental stimuli.....	66
Part IV: General experimental remarks.....	78
Choice of incentives across experiments.....	78

Outlier removal criteria.....	78
Chapter 2: Perception of fractal time series .....	80
Experiment 1 .....	80
<i>Method</i> .....	80
<i>Results</i> .....	82
<i>Discussion</i> .....	87
Experiment 2.....	87
<i>Method</i> .....	88
<i>Results</i> .....	88
<i>Discussion</i> .....	92
Experiment 3.....	93
<i>Method</i> .....	94
<i>Results</i> .....	96
<i>Discussion</i> .....	99
Experiment 4.....	101
<i>Method</i> .....	102
<i>Results</i> .....	104
<i>Discussion</i> .....	108
Experiment 5.....	109
<i>Method</i> .....	109
<i>Results</i> .....	112
<i>Discussion</i> .....	115

Conclusions.....	115
<i>Limitations</i> .....	116
Chapter 3: Risk perception and financial decisions .....	117
Experiment 1.....	117
<i>Method</i> .....	118
<i>Results</i> .....	122
<i>Discussion</i> .....	134
Experiment 2.....	135
<i>Method</i> .....	135
<i>Results</i> .....	137
<i>Discussion</i> .....	143
Experiment 3.....	144
<i>Method</i> .....	145
<i>Results</i> .....	147
<i>Discussion</i> .....	150
Conclusions.....	150
<i>Limitations</i> .....	152
Chapter 4: Judgmental forecasting from fractal time series: The effect of task instructions, individual differences, and expertise on noise imitation.....	154
Experiment 1.....	154
<i>Method</i> .....	155
<i>Results</i> .....	158

<i>Discussion</i> .....	167
Experiment 2 .....	168
<i>Method</i> .....	168
<i>Results</i> .....	171
<i>Discussion</i> .....	173
Experiment 3 .....	173
<i>Method</i> .....	174
<i>Results</i> .....	177
<i>Discussion</i> .....	181
Conclusions .....	181
<i>Limitations</i> .....	184
Chapter 5: The effects of news valence, price trend and individual differences on financial behaviour .....	185
Experiment 1 .....	185
<i>Method</i> .....	187
<i>Results</i> .....	191
<i>Discussion</i> .....	196
Experiment 2 .....	197
<i>Method</i> .....	198
<i>Results</i> .....	199
<i>Discussion</i> .....	204
Conclusions .....	205

<i>Limitations</i> .....	208
Chapter 6: Psychological Mechanisms Supporting Preservation of Asset Price	
Characterisations .....	210
Experiment 1 .....	210
<i>Method</i> .....	212
<i>Results</i> .....	215
<i>Discussion</i> .....	231
Experiment 2 .....	232
<i>Method</i> .....	234
<i>Results</i> .....	237
<i>Discussion</i> .....	247
Conclusions .....	250
<i>Limitations</i> .....	252
Chapter 7: General Discussion .....	254
Summary .....	254
Implications .....	256
Limitations .....	258
Directions for future research .....	258
Bibliography .....	262
Appendices .....	287
Appendix A: question list for Experiment 5 in Chapter 2 .....	287
Appendix B: Interactions and tests of simple effects in Experiments in chapter 6 .....	289

# List of Figures

## Chapter 1

Figure 1.1 Examples of fBm price series with Hurst coefficients ranging from 0.1 (anti-persistent) through 0.5 (random walk) to 0.9 (persistent) in 0.1 increments. ....	22
Figure 1.2 Example of fBm series with $H=0.3$ (left panels) and $0.7$ (right panels). Graphs in the first row show data referring to 30000 days, graphs in the second row show data referring to 6000 days, and graphs in the third row show data referring to 1000 days. All graphs are plotted on intervals of the same length along the x-axis. ....	58
Figure 1.3 Illustration of the mechanism which preserves geometrical properties of price graphs. The left column illustrates graphs with low local steepness and oscillation, and the right column presents graphs with high local steepness and oscillation. People observe data characterised by different properties (panels on the first row). They choose smaller scaling factors and time periods to present graphs with higher local steepness and oscillation. However, the scaled graphs still preserve properties of the original graphs (panels on the second row). Next, people make forecasts from the graphs (forecasts are marked with stars). Correspondingly, forecast dispersions are higher for the steeper graphs (panels on the third row). This process results in price graphs with properties that are correlated with those of the original data. ....	61
Figure 1.4 Examples of price change series with Hurst coefficients ranging from 0.1 (anti-persistent) through 0.5 (random walk) to 0.9 (persistent) in 0.1 increments. ....	65
Figure 1.5 FBm series with $H = 0.3, 0.5, 0.7$ (left column) and their corresponding fGn series (right column). ....	66

## Chapter 2

Figure 2.1 Experiment 1: Graphical user interface .....	83
Figure 2.2 Bar graph showing mean absolute errors for $H < .5$ (shaded) and $H > .5$ (unshaded) for raw price series from Experiment 1 (left) and price change series from Experiment 2 (right).....	92
Figure 2.3 Graphical user interface for Experiment 3.....	95
Figure 2.4 Experiment 3: Main effects of darkness of exemplar graph lines on absolute error scores (upper panel) and signed error scores (lower panel).....	100
Figure 2.5 Graphical user interface for Experiment 4.....	103
Figure 2.6 Experiment 4: Main effects of thickness of exemplar graph lines on absolute error scores (upper panel) and signed error scores (lower panel).....	105
Figure 2.7 The task window of Experiment 5.....	110
Figure 2.8 Absolute error versus trial number in Experiment 5. Exponential regression line is presented in the upper panel, and the regression line of the model $\text{Mean absolute error} = a/\text{trial number} + b + e$ is presented in the lower panel. ....	113

## Chapter 3

Figure 3.1 Task windows from Experiment 1: Risk rating task in fBm condition (upper panel) and randomness rating task in fBm&fGn condition (lower panel). ....	119
Figure 3.2 Percentage of choices of graphs with low Hurst exponent at the risk comparison task in the fBm condition in Experiment 1 against $\Delta H$ , presented for participant sections with different self-ratings of agreeableness (first row) and emotional stability (second row)....	131
Figure 3.3 Percentage of choices of graphs with low Hurst exponent at the randomness comparison task in the fBm condition in Experiment 1 against $\Delta H$ , presented for participant sections with different self-ratings of agreeableness (first row) and emotional stability (second row).....	132

Figure 3.4 The task window of Experiment 2.....	136
Figure 3.5 Mean risk assessment plotted against the Hurst exponents of the presented graphs. .....	140
Figure 3.6 The task window of Experiment 3. Upper panel: the buy condition; Lower panel: the sell condition. ....	146

## Chapter 4

Figure 4.1 Prediction program main window. The data are presented on the left of the line at $t = 63[\text{days}]$ , and a participant's prediction points are on its right. ....	157
Figure 4.2 A participant's predictions (dots connected by a line) and data (line) for graphs with $H = 0.1, 0.5, 0.9$ . This participant appears to have imitated noise. ....	159
Figure 4.3 A participant's predictions (dotted line) and data (line) for graphs with $H = 0.1,$ $0.5, 0.9$ . ....	160
Figure 4.4 Histograms showing the distribution of added points in Experiment 4 for computer generated graphs (upper panel) and real asset price series (lower panel). ....	163
Figure 4.5 Prediction and memory test window. The figure shows one word from the neutral word list ("Sphere") and two of the 9 words in the list box ("Insecure", "Unimportant") used for the low power condition. ....	170
Figure 4.6 Data, predictions and probability estimates made by a participant from the expert group in Experiment 3, for graphs with low (first panel), medium (second panel), and high (third panel) Hurst exponents. ....	176

## Chapter 5

Figure 5.1 A typical task window from Experiment 1. The figure shows the non-conflicting condition with bad news and a negative trend. ....	190
--	-----

Figure 5.2 A typical task window from Experiment 2. The figure shows the conflicting condition with bad news and a positive trend. ....	200
---	-----

## Chapter 6

Figure 6.1 The task window of Experiment 1.....	213
Figure 6.2 . Chosen time-scales with respect to the conditions of Experiment 1. ....	217
Figure 6.3 Mean steepness (upper panel) and oscillation (lower panel) of time-scaled graphs in Experiment 1.....	221
Figure 6.4 An illustration of the reference points used for the calculation of FD1, FD2, and FError when forecast horizon of 100 days: price graph against time (solid line: the data which was presented to the participant, dashed line: the continuation of the series which was not presented to the participant), participants forecasts (stars), the last data point which was presented to the participants (square), price at the required forecast date (circle), and the mean of participants' forecasts (triangle).....	224
Figure 6.5 Forecast dispersion measures in Experiment 1. Upper panel: FD1. Central panel: FD2. Lower panel: FError.....	228
Figure 6.6 The task window of Experiment 2: a price graph (the jagged lined) and a corresponding smoothed graph (the smoother line).....	235
Figure 6.7 Mean of chosen smoothness levels against the Hurst exponent of the given graphs (upper panel) and forecast density, measured by the number of required forecast points in the forecasting period (lower panel). Standard error is indicated with the bars. ....	239
Figure 6.8 The mean local steepness (upper panel) and oscillation (lower panel) of smoothed data graphs for each of the experimental conditions.....	245
Figure 6.9 The mean steepness of forecasts plotted against the Hurst exponent of the graphs (upper panel) and plotted against the number of required forecast points in the forecasting period (lower panel). Bars show standard error measures. ....	248

Figure 6.10 The mean steepness (upper panels) and oscillation (lower panels) of forecasts plotted against the Hurst exponent of the graphs (left panels) and plotted against the number of required forecast points in the forecasting period (right panels). Bars show standard error measures..... 249

# List of Tables

## Chapter 1

Table 1.1 The standard deviation of different methods of evaluation of the Hurst exponent of time series for different series lengths (from Delignieres et al., 2006).....	70
Table 1.2 The results of Hurst exponent analysis of real financial time series. The classification criterion was $\Delta H < 0.055$ . .....	72

## Chapter 2

Table 2.1 Experiment 1: Average values for absolute error (first panel) and signed error (second panel) for each combination of four ranges of Hurst coefficients, six different series lengths, and first and second instances. Standard deviations are denoted by parentheses.....	84
Table 2.2 Experiment 2: Average values for absolute error (first panel) and signed error (second panel) for each combination of four ranges of Hurst coefficients, six different series lengths, and first and second instances. Standard deviations are denoted by parentheses.....	89
Table 2.3 Experiment 3: Average values for absolute error (first panel) and signed error (second panel) for each combination of Hurst coefficient range, darkness level, and instance for the darkness condition. Standard deviations are denoted by parentheses.....	97
Table 2.4 Experiment 4: Average values for absolute error (first panel) and signed error (second panel) for each combination of Hurst coefficient range, thickness level, and instance for the thickness condition. Standard deviations are denoted by parentheses. ....	106
Table 2.5 Absolute and signed errors in Experiment 5.....	114

### Chapter 3

Table 3.1 The percentage of participants' answers, in which participants chose the asset with the low Hurst exponent (RiskLowHPerc and RandLowHPerc) in Experiment 1.....	123
Table 3.2 Mean values of $d'$ and $\beta$ in conditions fBm (first panel) and fBm&fGn (second panel) in Experiment 1.....	126
Table 3.3 Correlations between percentage of H-correlated answers in the fBm condition (first panel) and fBm&fGn condition (second panel) of Experiment 1. Statistically significant correlations are marked with a star. ....	129
Table 3.4 Variable notation.....	138
Table 3.5 Correlations and partial correlations between risk assessment and graph variables, and the beta values in multiple regression of risk assessment with the seven variables in Experiment 2.....	141
Table 3.6 Correlations between the variables examined in Experiment 2 for the stimuli sample. ....	142
Table 3.7 The percentage of participants' answers, in which participants chose the asset with the lower Hurst exponent in Experiment 3, and the associated confidence ratings. ....	148

### Chapter 4

Table 4.1 Correlation between geometrical characteristics of data and prediction graphs in the 'no limit' condition (first panel) and in the 'up to 4 points' condition (the second panel) in Experiment 1.....	165
Table 4.2 Correlation between geometrical characteristics of data and prediction graphs in Experiment 2.....	172
Table 4.3 Average prediction errors for each. prediction horizon in Experiment 3.....	178
Table 4.4 Correlation between geometrical characteristics of data and prediction graphs in Experiment 3.....	180

## **Chapter 5**

Table 5.1 Results of Experiment 1 for the western group (first panel) and the Eastern group (second panel) .....	192
Table 5.2 Results of Experiment 2, including trading latencies, share numbers, plausibility ratings (first panel), forecast differences and returns (second panel). .....	201

## **Chapter 6**

Table 6.1 The mean local steepness (first panel) and oscillation (second panel) of time-scaled graphs in Experiment 1. ....	222
Table 6.2 The mean forecast dispersions FD1 (first panel), FD2 (second panel), FError (third panel). ....	225
Table 6.3 Correlations between forecast dispersion measures, local steepness of graphs, oscillation, and forecast horizon. ....	230
Table 6.4. The mean chosen smoothness levels (first panel), local steepness of forecasts (second panel), and oscillation of participants' forecasts (third panel) in Experiment 2. ....	240

## **Appendix**

Table B.1 Interaction and simple tests of simple effects in Experiment 1 in Chapter 6. DV denotes dependent variables, and IV – independent variables.....	289
Table B.2 The results of a three-way repeated measures ANOVA on FD2 and FError. First panel: main effects. Second panel: interaction and tests of simple effects in Experiment 1 in Chapter 6. DV denotes dependent variables, and IV – independent variables. ....	293
Table B.3 Interactions and tests of simple effects in Experiment 2 in Chapter 6. ....	297

# Chapter 1: Background

## Introduction

*“Citigroup runs one of the biggest foreign-exchange operations at Canary Wharf. On a typical day in 2003, it is crowded, busy, and self-absorbed. The Citigroup trading room is vast, with hundreds of computers, ceilings, track lighting, and 130 currency traders and salespeople arrayed along rows of desks, six to a side...But consider the “mistakes” on this floor. Seated at one row of desks, a pair of analysts spend their days studying the orders of the bank’s own costumers. They are looking at broad patterns they can report back to the clients in regular newsletters. Theirs is the sort of market-insider information that, one form of the Efficient Market Hypothesis says, should not be useful... A few desks below is a math Ph.D. from Cambridge. He spends much of each day studying the fast-changing “volatility surface” of the option market – an imaginary 3-D graph of how price fluctuations widen and narrow... By the Black-Scholes formula, there should be nothing of interest in such a surface; it should be flat as a pancake. In fact it is wild, complex shaped. Tracking it and predicting its next changes are fundamental ways in which Citigroup’s option traders make money” (Mandelbrot and Hudson, 2004, pages 80-81).*

The power that financial markets encapsulate is difficult to apprehend. For instance, the daily average turnover in the Foreign Exchange market alone was nearly \$4 trillion in 2010. This market’s value was higher by 20% from that estimated during the crises of 2007, mostly due to online trading, high-frequency traders, and bank investments (King and Rime, 2010). To reach such an outstanding value, millions of investment decisions have to be made each and every day. A large percentage of traders employ judgmental methods (Cheung and Chinn, 2001; Taylor and Allen, 1992). Mathematically, the nature of the data used for these

decisions is highly controversial; some assume that it is entirely random, whereas others believe that it is has statistically self-similar fractal structure. The purpose of this thesis is to reveal how people perceive and react to financial time series. In particular, I am interested in questions about whether people are sensitive to fractal structure, how they make forecasts from financial data, the ways they estimate investment risks, and how they decide whether to buy or sell assets. Furthermore, I explore the ways human personality traits and dispositions interact with the data. Finally, I discuss the way people may influence price series.

I begin the introduction chapter by discussing the nature of financial data and fractals. I then describe the psychological and financial background of the study. I conclude the introduction with a description of the mathematical aspects of the thesis.

## **Part I: The role of fractals in finance**

*“Numbers are often used as a way of demonstrating objectivity and value neutral judgements when in fact, like any other mode of information transfer, they contain within them a whole series of judgements, rationalities, expectations and hopes” (Hall, 2006, page 673).*

Stories are at the heart of any human society; since childhood they nurture our understanding of causality, thereby endowing us with a certain sense of security and control. In particular, narratives are used by financial practitioners to create a sense of conviction, which enables functioning under conditions of severe uncertainty (Tuckett, 2012). But, in the case of financial markets, can stories really provide a mean of power acquisition, or are they merely cynical illusions? And if one were to believe that stories can yield control, which type of narrative should one follow?

According to the Efficient Market Hypothesis (EMH), such a form of control is impossible (Hodnett and Heng-Hsing, 2012; Mehrara and Oryoie, 2012). The strong form of the EMH, originally developed by Fama in the 1960s (Malkiel and Fama, 1970), states that market prices reflect all types of available information about all asset fundamental values. As future information cannot be forecasted, neither can future prices. Therefore, one cannot “beat the market”. Since the sixties, different versions of the EMH have been formulated. In particular, the weak version of the EMH limits the conclusion of the strong form to historical price data alone; as inferring future prices from past and present prices is termed *Technical Analysis*, the weak version of the EMH invalidates this type of trading method. Validity of the different versions the EMH has been challenged over and over again during the years. In fact, the EMH is one of the most tested hypotheses in finance (Yen and Lee, 2008).

According to the contradictory evidence that emerged, different investment recommendations were developed (see e.g. Ang, Goetzmann and Schaefer, 2010). The debate over the EMH is far from being settled. It is, therefore, bewildering that most traders use financial tools which are based on one of the most important results derived from the EMH: the random-walk hypothesis.

The random walk hypothesis consists of the narrative that the market has no memory: changes in prices at any moment are entirely independent of the history of asset prices. Wrapped in thick layers of mathematical formulae, this narrative supplied to the masses of traders, investors, and other financial practitioners a method to price assets and manage their portfolios in a way that was supposed to guarantee (up to some pre-determined level of risk) that they would make profits. The random-walk narrative is attractive in its simplicity; it is friendly as it lends itself to mathematical analysis; and it is comforting, as it assigns small probabilities to financial crises (Mandelbrot and Hudson, 2004). In addition, it is powerful: Black-Scholes formula, described by Berkowitz (2010, page 1) as “one of the most widely used option valuation procedures among practitioners”, is based on the random walk assumption. Furthermore, a different model based on the random walk hypothesis – the Capital Asset Pricing Model (CAPM) - was shown to be used for investment decisions by 73.5% of the respondents to the survey of Graham and Harvey (2002). But is the random walk model correct?

As with the EMH, validity of the random-walk hypothesis has been tested in numerous contexts and geographic location. For instance, Mehmood, Mehmood and Mujtaba (2012), and Narayan and Smyth (2006) supported the random walk hypothesis, Umanath (2012) and Al-Jafari (2011) rejected it, whereas Righi and Ceretta (2011) and Otto (2010) found that its validity depends on a wide range of different factors.

Fractal models provided an alternative narrative of the nature of the market. Primitive versions of fractal formulation of the market had already been suggested at the beginning of

the 20<sup>th</sup> century. For instance, in the 1930s, Elliot (Frost and Prechter, 1998) defied the assumption that the market has no memory by noticing that certain patterns (“waves”) tend to appear in it. Each of Elliot’s waves could be decomposed into parts which resembled the original wave, hence giving rise to a fractal-like self-similarity. Existence of structure in price graphs is impossible according to the random walk hypothesis. However accurate his observation was, Elliot did not construct any statistical or mathematical theory that could support his views rigorously. On the other hand, in the 1960s, Mandelbrot (Mandelbrot and Hudson, 2004) proved that some assets do not obey the Gaussian statistics imposed by the random walk hypothesis. Mandelbrot showed that, instead, asset prices exhibited a statistically self-similar behaviour.

More precisely, Mandelbrot argued that financial time series could be modelled as *fractional Brownian motions* (fBm), series whose roughness can be characterised by a constant termed the *Hurst Exponent* (H). For fBm series, the values of the Hurst exponent range between 0 and 1. Loosely speaking, as the H value of a time series approaches 0, the series seems to be noisier, and as H approaches 1 it seems to be more regular. Figure 1.1 presents graphs of fBm series with different Hurst exponents.

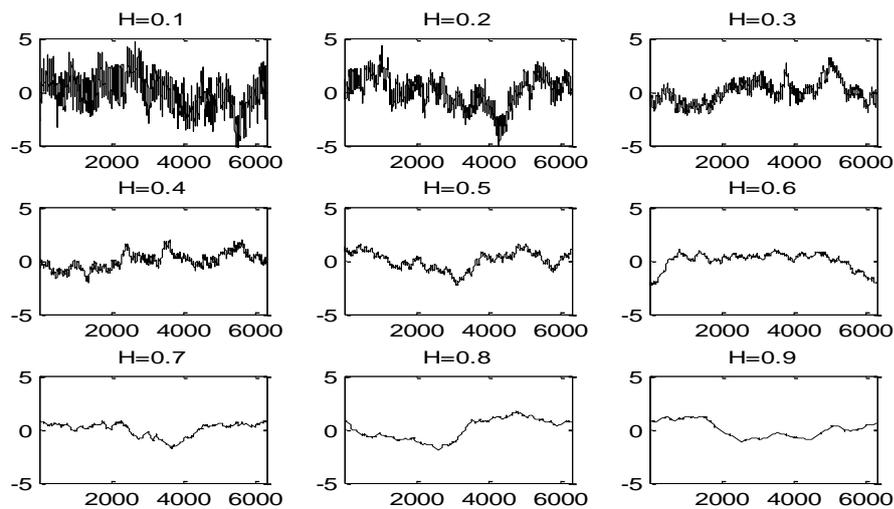


Figure 1.1 Examples of fBm price series with Hurst coefficients ranging from 0.1 (anti-persistent) through 0.5 (random walk) to 0.9 (persistent) in 0.1 increments.

The Hurst exponent is related to the dimension of the fractal, and it can be shown that it is also a measure of a series' memory: for  $H > 0.5$ , the increments of the series are positively autocorrelated, whereas for  $H < 0.5$ , they are negatively autocorrelated. The case of  $H = 0.5$  corresponds to a random walk. It has been shown that Hurst exponents of real assets typically vary between 0.35 and 0.65, rather than being exactly 0.5 (Sang, Ma, and Wang, 2001). Therefore, one can consider Mandelbrot's theory to be a generalisation of the random walk hypothesis. Mandelbrot's model was much more complicated than the random-walk theory, and therefore did not lend itself to mathematical handling. Indeed, Mandelbrot and Hudson wrote in 2004 (Mandelbrot and Hudson, 2004) that the fractal theory was not a forecasting tool. Financial trading methods relying on fractal assumptions have been developed only recently and are, in general, rare. The fractal narrative does not possess the properties required to allure most investors: it cannot supply immediate answers, and the answers it does give are not reassuring. Instead of pacifying investors by depicting a relatively safe world the way the random-walk model does, it presents the market as dangerous and unpredictable (Mandelbrot and Hudson, 2004). In addition, the fractal model of the market is still considered to be highly controversial.

Nevertheless, during the last few years, and especially after the last series of global financial crises, a new interest in the model has emerged. Investors and traders started to suspect that the tools they use might not be adequate to describe extreme phenomena, such as the creation of bubbles and acute price falls, which occurred much more frequently than classical theories predicted. This new interest manifested itself through a large body of research aiming to answer the question whether the market is fractal or not (e.g. Parthasarathy, 2013; Malavoglia, Gaio, Júnior and Lima, 2012; Ling-Yun, 2011; Onali and Goddard, 2011; Sun, Rachev and Fabozzi, 2007; In and Kim, 2006). Attempts to predict the market based on its fractal properties were made as well (Duchon, Robert, and Vargas, 2012; Richards, 2004; Cui and Yang, 2009). Furthermore, new theories and investment strategies, combining fractal models with previous formulations, such as the Black-Scholes formula,

were developed (Bayraktar and Poor, 2005). Fractal analysis also has applications in macroeconomics (see e.g. Blackledge, 2008). Recently, innovative approaches, such as multifractals have been developed (Dezsi and Scarlat, 2012; Schmitt, Ma, and Angounou, 2011).

For the purpose of this thesis, I adopted the fractal model. On the one hand, it offers a wider view on the market than that obtained from the random walk model. On the other, it constitutes a practical source of stimulus material for experiments in psychology.

Demonstration of fractal-related psychological phenomena does not require accuracy to the degree that the multifractal model might offer. Furthermore, the Hurst exponent of computer generated series is correlated with other measures of graphs sets (provided that the series were generated by the same algorithm). Among the variables which are correlated with the Hurst exponent of a graph are its local steepness, defined as the average of the absolute value of the gradients between successive points in the graph, the graph's oscillation, defined as the difference between the maximum and minimum values of the graph over a given interval, and the series' standard deviation (historical volatility). Examining people's reactions to fractal stimuli can, therefore, yield information about properties other than the Hurst exponent, which are of financial interest. Finally, I know of no previous work on financial implications of human perception of fractal series.

## **Part II: Studies in psychology and behavioural finance**

*“Bulls are like the giraffe which is scared by nothing, or like the magician of the Elector of Cologne, who in his mirror made the ladies appear much more beautiful than they were in reality. They love everything, they praise everything, they exaggerate everything [...] the bulls make the public believe that their tricks signify wealth and that crops grow on graves. When attacked by serpents, they, like the Indians, regard them as both delicate and a delicious meal... The bears, on the contrary, are completely ruled by fear, trepidation, and nervousness. Rabbits become elephants, brawls in a tavern become rebellions, faint shadows appear to them as signs of chaos. But if there are sheep in Africa that are supposed to serve as donkeys and wethers to serve even as horses, what is there miraculous about the likelihood that every dwarf will become a giant in the eyes of the bears?” (Joseph de la Vega, 1688, pages 162-163).*

The understanding that market participants are people who exhibit a wide spectrum of the human properties is rooted in ancient times. Nevertheless, it was only close to the end of the 20<sup>th</sup> century that it became evident that traders' human advantages and drawbacks influence the way markets behave. This realisation made certain fields in psychology highly relevant to finance. In the following section I review the studies in psychology and behavioural finance which form the basis for the financial applications discussed in this thesis. In particular, I will discuss studies concerning the perception of fractal time series, risk perception, buy/sell decisions, judgmental forecasts, the effects of news on financial decisions and forecasts, and mechanisms preserving asset price graph structure.

## **Perception of fractal time series**

As emphasised by Batchelor (2013), Cheung and Chinn (2001) and Taylor and Allen (1992), a large number of traders employ chartist methods, which are based on extrapolation and pattern recognition of graphically presented financial time series. It is, therefore, important to understand what people actually see in this type of data: are people sensitive to the Hurst exponent of fractal time series? If they are, how sensitive are they? Performance of numerical algorithms which estimate the Hurst exponent of a series depends on the length of the series (Delignières, Ramdani, Lemoine, Torre, Fortes and Ninot, 2006). How is people's sensitivity affected by the length of the given series? What type of fractal data are people more sensitive to? Do people treat fractal series as if they were produced as a sum of signal and noise series? Can people create mental representations of the Hurst exponent of time series? And what meaning do they attribute to them?

Apart from immediate perceptual implications, answering these questions is an imperative, initial stage that must precede any attempt to answer questions of a direct financial importance. For instance, as will be described in the following chapters, knowing that people are sensitive to price *change* series (as well as to price series), enabled me to investigate the conditions in which investment risk assessments depend on the Hurst exponents of the given graphs. Understanding the range of series lengths within which people exhibited high sensitivity to the Hurst exponent of given graphs enabled me to choose experimental stimuli of reasonable lengths. Knowing the resolution at which people can distinguish between Hurst exponent values was essential in order to design the stages of an experiment about buy/sell decisions. Also, answering the question about the perception of signal and noise of data series inspired the question whether people's forecasts from fractal graphs could be also separated into signal and noise.

### *Historical background: perception of random series*

Perception of time series was already being studied in psychology during the 1950s (e.g. Jarvik, 1951). However, it was only in the 1970s that a wider interest in perception of randomness in time series emerged. Kahneman and Tversky (1972) showed that participants judged sequences of unbiased coin tosses as more random if they contained more alternations in the order of appearance of the heads and tails. They explained their results in terms of people's use of the representativeness heuristic. They conjectured that a sequence is judged random if it resembles locally the global characterizations of a random sequence. Gilovich, Vallone, and Tversky, (1985) studied beliefs of basketball fans and players. They found that, in spite of the fact that the outcomes of a field goal and free throw are largely independent of the outcome on the previous attempt, fans believed that a player's chances of hitting a basket are greater following a hit than following a miss. They speculated that the reason for this misperception of randomness could be a memory bias or a problem of analyzing data. Falk and Konold (1994) suggested that when people are asked to evaluate the degree of randomness in binary sequences, they base their estimates on difficulty of encoding. More precisely, they used Shannon entropy as a normative measure of the degree of randomness in the sequences. They showed that the correlation between the evaluated randomness and the entropy of the sequences was much smaller than the correlation between evaluated randomness and the time required for participants to memorise the sequences.

Although these accounts give extremely important insights into randomness perception, none of them can be applied directly to explain how people perceive fractal time series. The global and local structural similarity account, which Kahneman and Tversky used along within their representativeness account, might seem compelling for its self-similar fractal-like nature. However, it is unlikely that people use it to perform tasks such as discrimination between the Hurst exponents in different graphs that are merely statistically self-similar. Furthermore, Gigerenzer (1991, page 102) has claimed that Kahneman and Tversky's "heuristics such as representativeness have little to say about how the mind adapts to the

structure of a given environment”. In his opinion, “all three heuristics [...] are largely undefined concepts and can post hoc be used to explain almost everything”. Gilovich et. al’s (1985) memory bias account is irrelevant to the explanation of performance in a task in which people are presented with graphs. Also, although Falk and Konold’s (1994) complexity account is useful for very short sequences that can be memorised, it is not for real-life financial time series that consist of hundreds or thousands of elements. Recently, however, studies focusing on the perception of fractal patterns have been performed.

### *Perception of fractal patterns*

People seem to be predisposed to analysing fractals. For instance, Mitina and Abraham (2003) showed that people are sensitive to fractal geometric pictures. In particular, they found that the aesthetic attractiveness of the patterns and its fractal dimension were correlated. Cutting and Garvin (1987) presented participants with simple geometric fractal-like patterns, and asked them to evaluate their complexity. They found that the fractal dimension and the recursion depth correlated with complexity estimates. Forsythe, Nadal, Sheehy, Cela-Conde, and Sawey (2011) showed that the fractal dimension of pictures of the natural environment, abstract art, and figurative art by acclaimed artists varied with ratings of their beauty. In particular, Spehar, Clifford, Newell, and Taylor (2003) found that participants’ preferences among natural images peaked at a value of  $H = 0.7$ . Redies, Hasenstein, and Denzler (2007) demonstrated that graphic art from the western hemisphere exhibited fractal-like statistics, and that these findings were universal beyond culture or era. This natural inclination towards fractals might be rooted in the process of evolution, since many natural phenomena have a fractal character. For example, woody plants, trees, waves, clouds, cracks in materials, snowflakes, mineral patterns, coastlines, galaxies, and retinal blood vessels, are fractals (Taylor, Spehar, Van Donkelaar and Hagerhall, 2011). The fractal dimensions of images are between 1 and 2, and their Hurst exponents are between 0 and 1.

This natural predisposition might be facilitated by physiological mechanisms. Although Taylor, Spehar, Van Donkelaar and Hagerhall (2011) did not find a correlation between the fractal dimension of Jackson Pollock's paintings and the fractal dimension of eye-tracking patterns of observers, there is evidence for fractal functioning in other parts of the human visual system. In 1982, De Valois, Albrecht, and Thorell found that striate cells have a narrow spatial bandwidth, covering a wide range of frequencies. Their results support the idea that the visual system has multi-scale properties and is, therefore, adaptive for fractal environments. More recently, Georgeson, May, Freeman and Hesse (2007) developed a multi-scale model for human edge analysis. Taylor (2006) found that skin conduction changed as the Hurst exponent of observed images was manipulated. Taylor et al. (2011) showed that participants' EEG responses to images depend on their Hurst exponents.

#### *Perception of fractal time series*

In *The (Mis)behaviour of Markets*, Mandelbrot and Hudson (2004, pages 17-19) invited the reader to participate in an experiment. Mandelbrot and Hudson presented the readers with two graphs of real-life price series "of the kind you would find in a brokerage-house report," one graph depicting a computer generated fractal series, and one graph of a random walk series. The question they challenged the reader to answer was: "Ignore whether they trend up or down. Focus on how they vary from one moment to the next. Which are real? Which are fake? What rules were used to draw the fake?" On the following page, they depicted the corresponding price *change* graphs. Mandelbrot and Hudson asserted that the price change graphs were easier to distinguish between than price graphs. This experiment suggested that people are more sensitive to properties of price change graphs than to properties of price graphs.

Mandelbrot and Hudson did not perform any experiment to validate their views. As far as I am aware, people's ability to distinguish graphical depictions of fractal time series from those produced by a random walk has not been the subject of a statistically valid study.

There have been three reports of investigations into discrimination of the fractal structure of unidimensional spatial graphs or contours (Gilden, Schmuckler and Clayton, 1993; Kumar, Zhou and Glaser, 1993; Westheimer, 1991). However, all three papers confined their experiments to either trained human observers or to a very small number of participants (two or eight), which do not allow generalisation of the results to large populations.

Kumar, Zhou and Glaser (1993) compared people's sensitivity to fractal dimension of graphs to the performance of five numerical algorithms including the grid-dimension method (described on page 1140 of their paper). Kumar et al. noted that trained human observers participated in their experiments, but no further descriptions of the nature of the training or the observers were mentioned. They showed their participants graphs with different Hurst exponents. The task in their experiment was to determine whether the roughness of the target graph was higher than that of a reference graph. Kumar et al. took into account also the luminance of the screen they used for presentation. The authors found that people usually have lower discrimination thresholds than numerical codes. People's thresholds depended on fractal production method. For some of the methods, discrimination threshold was as low as 0.03 (Hurst exponent of a graph ranges between 0 and 1).

Westheimer (1991) presented himself and another highly experienced psychophysicist who was familiar with fractals with sequences of 256-point unidimensional fractal contours drawn from an ensemble of seven equally spaced in terms of their fractal dimension. Their task was to decide whether each stimulus "was more ragged, corrugated, jagged or fractured than the average of the series" (page 216). Their performance was good: differences in the second decimal of the fractal dimension of the stimuli could be distinguished and sensitivity increased as H increased from 0.75 through 0.80 to 0.85.

Gilden et al. (1993) investigated the question of whether people are adapted to the fractal characteristics of contours in their observable environment. To investigate this question, Gilden et al (1993, experiment 1) generated 200 unidimensional fractal graphs, each made

up of 256 points, for each of 14 families of stimuli that differed in terms of their fractal dimension. Eight participants were trained with feedback so that they understood “the sense in which random contours could belong to the same family without appearing identical on a point-to-point basis” (page 466). These participants were then presented with simultaneous pairs of stimuli and asked to decide whether they had been drawn from the same family. Sensitivity rose as  $H$  increased up to 0.78 but then dropped again as  $H$  was increased further. In a follow-up experiment with three participants, they replicated this finding but also found that peak sensitivity dropped to  $H = 0.5$  when the vertical extent of the display was doubled in size. They conclude that their results show that people are adapted to the fractal characteristics of the contours found in their natural environment.

Though all three studies provided important insights, results from the latter two are not fully compatible. Gilden et al (1993) found that sensitivity peaked between  $H = 0.5$  and  $H = 0.78$  whereas Westheimer (1991) found that it continued to increase between  $H = 0.75$  and  $H = 0.85$ . There could be a number of reasons for this discrepancy: for example, both expertise of participants and details of procedure differed. More importantly, both studies were statistically underpowered. The effect of series length was not studied and neither was the process by which people learn how to discriminate between series with different Hurst exponents.

Given these contradicting results, I hypothesised that people’s sensitivity to the Hurst exponent of graphically presented fBm series depended on the Hurst exponent (Hypothesis  $H_{1,1}$ ). However, *a priori* it is not possible to propose a directional hypothesis.

How do people discriminate between graphs with different Hurst exponents? Both Gilden et al (1993) and Westheimer (1991) attempted to answer this question. Gilden et al (1993) suggested that the discrimination involved the extraction of statistical features of the series. I expected assessments of any statistical features of the series to improve as the length of the series increases, as information forming the basis for the judgement would increase, too. I,

therefore, hypothesised that people's discriminability of the Hurst exponent of fBm series increased with the series length (Hypothesis H<sub>1,2</sub>).

In addition, Gilden et al's (1993) suggested that, although ideal fractals do not have a natural decomposition into signal and noise components, people process fractal stimuli as if they do. In particular, they suggested that, to assess noise, people extract changes between successive series points and then assess the width of the distributions of those changes: "the observer that discriminates in terms of the width of the increment distribution is generally more sensitive over the domain of fBm families" (Gilden et al, 1993, page 475). However, Gilden et al did not provide human evidence supporting this conjecture. If indeed people do use this approach, I would expect that presenting participants with fGn sequences would enhance their discriminability, because I externally perform one of the (presumably error-prone) operations that people otherwise have to perform internally. I, therefore, hypothesise that people exhibit a higher degree of sensitivity to fGn graphs than to fBm graphs (Hypothesis H<sub>1,3</sub>). As the accuracy of the assessment of distribution width should be higher when more data is processed, I conjecture that discriminability of the Hurst exponent of fGn sequences is higher when the series are longer (Hypothesis H<sub>1,4</sub>).

In addition, I expect that change series derived from series with H values less than 0.5 will be harder to discriminate than those derived from series with H values greater than 0.5 (Hypothesis H<sub>1,5</sub>). This is because difficulty in discriminating widths of distributions in change series is what Gilden et al (1993) argue drives the patterns of discrimination in the original series.

In particular, Gilden et al's (1993) conjecture implies that people use local gradients as a cue in discrimination tasks of the Hurst exponents of fBm series. However, another possible cue is the graphs' illuminance. Westheimer (1991, page 215) made the following point in his discussion of the cues that people may use to discriminate fractal contours: "By definition, an increase in the fractal dimension of a line also produces an increase in line length and that

can produce a change in retinal illuminance. For example, a bright line on a dark background becomes brighter as its fractal dimension increases and that itself would be a visual clue”.

In pilot work prior to the series of studies reported in the thesis, I presented participants with a screen of nine cells showing series with different H values between 0.1 and 0.9 in increments of 0.1. (The display was similar to Figure 1.1. The presented graphs were randomised and there was no indication of the Hurst exponents of the graphs.) I asked them to identify ways in which these graphs were different. A variety of answers was given in response to this question but some participants mentioned that the graphs appeared to vary in terms of the “darkness” or “thickness” of the line. As Figure 1.1 suggests, graphs of series with lower Hurst exponents can be seen as being darker or thicker than those with higher Hurst exponents.

I conjectured that participants’ perception of graphs’ “thickness” was a result of their sensitivity to the local gradients of the graphs, as suggested by Gilden et al. (1993). Graphs with low values of Hurst exponents are locally steep and have very frequent change of directions. Thickening the line of a graph may mask small fluctuations and affect its perceived smoothness. I, therefore, hypothesised that people use graphs’ gradients as a cue assisting in discrimination of the Hurst exponents of fBm graphs (Hypothesis H<sub>1,6</sub>).

Following Westheimer (1991), I hypothesised that people use graphs’ illuminance as a cue assisting in discrimination of the Hurst exponents of fBm graphs (Hypothesis H<sub>1,7</sub>).

The experimental paradigms used by Gilden et al (1993) and Westheimer (1991) were similar to each other, in the sense that participants were asked to discriminate between the Hurst exponents of two graphs. However, I argue that people can also learn to identify the Hurst exponents of given graphs through feedback (Hypothesis H<sub>1,8</sub>). This is because a large body of research supports the hypothesis that feedback facilitated learning of categories. In particular, Maddox, Love, Glass and Filoteo (2008) have shown that feedback assisted people learn category structures when optimal rules were not verbalised. Finally, following

Mandelbrot and Hudson (2004), I hypothesise that people perceive investments in assets whose price graphs have lower Hurst exponent values riskier than investments in assets whose graphs have higher Hurst exponents (Hypothesis H<sub>1,9</sub>).

The study of the way people perceive fractal time series is reported in Chapter 2.

### **Risk perception and financial decisions**

Risk assessment is one of the most important tasks that financial analysts perform. In particular, it is used for portfolio optimisation (Markowitz, 1952; Holton, 2004). A large number of techniques designed to help practitioners deal with financial risk have, therefore, been developed. For instance, the Black-Scholes formula provides investors with a hedging strategy, which should, theoretically, yield a risk-free portfolio (Black and Scholes, 1973). However, Haug and Taleb (2011, page 98) claim that investors do not evaluate risk using theories of this type: “Option traders do not “buy theories”, particularly speculative general equilibrium ones, which they find too risky for them and extremely lacking in standards of reliability. A normative theory is, simply, not good for decision-making under uncertainty (particularly if it is in chronic disagreement with empirical evidence). People may take decisions based on speculative theories, but avoid the fragility of theories in running their risks”. Mandelbrot and Hudson (2004, page 231) foresaw Haug and Taleb’s arguments: “Real investors know better than economists. They instinctively realize that the market is very, very risky, riskier than the standard [normative] models say”. How do people assess risk of investments in assets, based on graphical presentations of their price series?

#### *Factors affecting Judgmental risk assessment*

In experimental settings in which the experimenter manipulates only properties of time series, participants’ risk assessments should depend on the properties of the presented graphs. However, a series of studies in behavioural finance has shown that financial risk

perception depends on a large number of variables, including: the controllability of an asset and how worrying it is (Koonce, McAnally and Mercer 2005), tension experienced by financial leaders (Woollen, 2011), probability of gain, loss and status quo (Holtgrave and Weber, 1993). As a large number of factors is involved in risk assessment, I hypothesise that, in a pure technical-analysis condition, in which people rate risk of assets based on graphs of fBm price series, and with no additional cues, risk assessments would depend only weakly on the Hurst exponent (Hypothesis H<sub>2,1</sub>).

However, Mandelbrot and Hudson (2004, pages 169-170) asserted that investors can intuitively sense the fact that the fractal nature of financial series makes them more risky than orthodox approaches predict: “Instinctively, most people regard a cotton contract as a riskier proposition than a Blue Chip stock — despite the fact that, by the standard analysis, commodity investments should play a bigger role in the portfolios of the wealthy. Most people sense the greater risk, and shun it. Perhaps no great statistical analysis was needed at all: This fact of mass psychology, alone, might have been sufficient evidence to suggest there is something amiss with the standard financial models”.

Mandelbrot and Hudson did not specify the cognitive processes underlying people’s risk assessments, nor did they study the conditions in which risk perception is correlated with fractal parameters of price series. However, as noted before, they suggested that people are more sensitive to geometrical properties of fGn price change series than to their corresponding fBm price series. They presented the readers sets of price and price change graphs, and wrote: “All fairly similar, many readers will say [about the price graphs]. Indeed, stripped of legends, axis labels, and other clues to context, most price “fever charts,” as they are called in the financial press, look much the same. But pictures can deceive better than words. For the truth, look at the next set of charts. These show, rather than the prices themselves, the change in price from moment to moment. Now, a pattern emerges, and the eye is smarter than we normally give it credit for - especially at perceiving how things change” (Mandelbrot and Hudson, 2004, pages 17-18). In addition, in different contexts, risk

perception has been found to be highly susceptible to the means by which the information is conveyed (Weber, Siebenmorgen, and Weber, 2005; Gaissmaier, Wegwarth, Skopec, Müller, Broschinski and Politi, 2012; Stone, Yates, Parker and Andrew, 1997; Raghurir and Das, 2010). For instance, Stone, Yates, Parker and Andrew (1997) showed that different display formats of low probability risk information (numerical format, bars, stick figures, and people's faces sketches) affected risk-related behaviour. Given Mandelbrot and Hudson's views on judgmental risk perception and the latter's dependence on communication means, I hypothesise that providing people with cues about the Hurst exponents of the given series would affect their risk judgement. More precisely, I expect that, when both price series (fBm) and its corresponding price change series (fGn) are presented, risk assessments are negatively correlated with the Hurst exponent of price series (Hypothesis H<sub>2,2</sub>). In fact, many financial data providers enable participants to display price change series in addition to price series. For instance, the website of Yahoo! Finance (<http://finance.yahoo.com>) enables investors to see graphs of price change (using the option "Rate of Change (ROC) indicator"). Situations in which traders are exposed to both fBm and fGn series are, therefore, prevalent.

Another important factor affecting risk perception is that of individual differences. Indeed, nationality (Weber and Hsee, 1998), gender (Walia and Kiran, 2012), testosterone level (Stenstrom and Saad, 2011), financial literacy (Sachse, Jungermann, and Belting, 2012), and life history (Griskevicius, Tybur, Delton and Robertson 2011) have been shown to have a significant effects on risk perception. I was especially interested in the effect of the Big Five personality traits on risk assessment. The Big Five personality traits comprise emotional stability, extraversion, openness to experience, agreeableness, and conscientiousness. Emotional stability has been shown to have a significant effect on risk perception (Sjöberg, 2003). Furthermore, Jakes and Hemsley (1986) found that people who have high scores on the Psychoticism ('P') and Neuroticism ('N') on the Eysenck Personality Questionnaire (EPQ) perceived a larger number of meaningful objects (but not a larger number of simple

geometrical shapes) in random dot stimuli than those low on 'N' and 'P'. This implies that people with higher Neuroticism scores attributed more meaning to the patterns that they found. Neuroticism ratings on the EPQ are strongly, negatively correlated with the Big Five's emotional stability (Van der Linden, Tsaoasis, and Petrides, 2012). I, therefore, conjectured that, in the context of risk assessment tasks, people with lower emotional stability would be more likely to attribute the meaning of risk to patterns found in fractal graphs. More precisely, risk ratings of people lower on emotional stability should be correlated with the Hurst exponent of the presented graphs more strongly (Hypothesis H<sub>2,3</sub>).

Risk perception depends on graphical mathematical properties other than the Hurst exponents of the series as well. Particularly important is the standard deviation of the series, as it represents the historical volatility of the asset. Historical volatility is used as a volatility measure in many classical financial theories (Amilon, 2003; Kala and Pandey, 2012).

However, the dependence of risk assessment on the standard deviation of the series is not well-understood. Klos, Weber and Weber (2005) found that risk assessment is only weakly correlated with estimates of standard deviation of price series. On the other hand, Sachse et al. (2012) and Weber, Siebenmorgen, and Weber (2005) found a high correlation between risk and volatility. These inconsistencies might be partially explained by other differences in the stimulus materials used in these studies. Another important factor is the graphs' mean run length (the number of successive points in which prices move in the same direction).

Risk judgements have been shown to be correlated with the run lengths of price graphs (Raghubir and Das, 2010). Finally, Duxbury and Summers (2004) found that traders are more loss averse than variance averse. However, none of these authors examined the relative importance of these variables to risk estimation when price change graphs are presented in addition to price graphs.

I performed a systematic examination of mathematical factors which could affect risk perception. To do so, apart from the Hurst exponent, the standard deviation of the series, and its mean run-length, I studied the effects of the graph's oscillation (the difference between its

maximum and minimum values), the difference between the values of the last and first points of the series, the absolute value of the difference between the values of the last and first points of the series, and the difference between the first point of the series and its minimum.

Oscillation and the absolute value of the difference between the values of the last and first points of the series could serve as proxy measures of the graphs' volatility. In line with Sachse et al. (2012) I hypothesised that these variables, as well as the standard deviation and the mean run length, are positively correlated with risk assessments (Hypothesis H<sub>2,4,a</sub>).

The difference between the values of the last and first points of the series and the difference between the first series point and its minimum could also be proxy measures for the amount of money which can be lost. I, therefore, hypothesised that they would be negatively correlated with risk assessments (Hypothesis H<sub>2,4,b</sub>).

Risk assessment is likely to be correlated with the standard deviation of the series. When fBm graphs are produced by a single algorithm, their Hurst exponent is correlated with their standard deviation. However, Mandelbrot and Hudson (2004) and Haug and Taleb (2011) asserted that people do not assess risk according to normative measures such as the standard deviation of the series, and that they are sensitive to the occurrence of rare event. The probability of rare events is assessed more accurately by the Hurst exponent of the series than by its standard deviation. I, therefore, hypothesise that the effect of the Hurst exponent on risk assessment will be stronger than that of the standard deviation (Hypothesis H<sub>2,5</sub>).

#### *Factors affecting financial decisions*

Nosić and Weber (2010) and Weber, Weber, and Nosić (2013) argued that historical volatility (standard deviation) affects financial decisions. Raghurir and Das (2010) showed that the mean run length affects risk perception. Following these studies, I hypothesised that the standard deviation of an asset's price series and its mean run length would affect buy/sell decisions (Hypothesis H<sub>2,6</sub>).

As the Hurst exponent is correlated with the standard deviation, I expected that buy/sell choices would be affected by the Hurst exponent. More precisely, I hypothesised that the lower the Hurst exponent of an asset is, the higher people's tendency to sell it would be, and the higher the Hurst exponent of an asset is, the higher people's tendency to buy it would be (Hypothesis H<sub>2,7</sub>).

### **Judgmental forecasting from fractal time series: The effects of task instructions, personality traits, sense of power, and expertise on noise imitation**

Forecasting is as fundamental a task for financial analysts as risk assessment is. In fact, forecasting and risk assessment are often inseparable. For instance, some banks issue risk statements along with their macroeconomic forecasts (Knuppel and Schultefrankenfeld, 2011), and some algorithmic methods of evaluating risk involve forecasting (Liu and Hung, 2010). It has been shown that commonly used judgmental forecasting methods rely not only on the heuristics specified by technical analysis but also on intuition (Batchelor and Kwan, 2007). Here I explore the human aspect of forecasting from fractal graphs. In particular, I will focus on the phenomenon of noise imitation.

#### *Literature about judgmental forecasting and noise imitation*

The research of judgmental forecasting is rooted in studies about the similarity between properties of binary time series and forecasts. For instance, Edwards (1961) demonstrated positive recency in participants' forecasts. Modern studies of judgmental forecasting have typically either used artificial time series generated by adding random noise to a signal or used real series assumed to be decomposable into signal and noise (Lawrence, Goodwin, O'Connor and Önköl, 2006). High levels of performance have been taken to reflect forecasters' ability to separate signal from noise and to forecast on the basis of the signal alone (Harvey, 1988). This work has revealed that forecasters are subject to a number of systematic biases. These include tendencies to overestimate sequential dependence (Bolger

and Harvey, 1993; Reimers and Harvey, 2011), and to make higher forecasts for desirable outcomes (Eggleton, 1982; Harvey and Bolger, 1996; Harvey and Reimers, 2013; Lawrence and Makridakis, 1989). People also tend to include rather than exclude the noise component of the data series in their forecasts (Harvey, 1995; Harvey, Ewart and West, 1997). Noise level of the prediction has been found to be correlated with the noise level of the data (Harvey, 1995). Bolger and Harvey (1993) hypothesised that people imitated noise in order to make their forecasts representative of the data series. The results of Harvey et al (1997) supported this explanation.

The tendency to imitate noise in forecasts has been found to be difficult to control. For instance, Harvey, Ewart, and West (1997, page 126) provided participants in one of their experiments with highly detailed explanation about the nature of the task: “Put six crosses on the graph to show us your forecasts. Obviously you cannot be certain where these future points will be but try to ensure that your forecasts show the *most likely* positions for them. For example, if you feel that a particular point could lie within a range of values, put your cross in the centre of that range if you feel that this is the most likely position for the true point within the range. Your aim is to maximise the probability that your forecasts will be correct. Your six crosses need to be placed on the six vertical lines to the right of the last data point”. Nevertheless, participants in this experiment imitated noise. Harvey, Ewart, and West (1997) did not use fractal series as their experimental stimuli. I do not know of any study which examined the way people make forecasts from fractal series. Do people imitate noise of fractal time series? Do task instructions, high levels of certain personality traits, sense of power, or expertise act to reduce it? The experiments reported in Chapter 4 were designed to answer these questions.

#### *Decomposition of series into signal and noise*

Not all time series comprise linear combinations of signal components and noise. As mentioned above, fractals do not have a natural decomposition into signal and noise

components. This is because they typically have a degree of self-similarity. For instance, *exact self-similar fractals* are geometric shapes in which exactly the same structure appears independently of the observed scale. *Fractional Brownian motions* (fBm) are *statistically self-similar fractals*, and therefore exhibit self-similarity in a weaker way than exact fractals. fBm series are, therefore, sometimes referred to also as coloured noise (Stoyanov, Gunzburger, and Burkardt, 2011). However, due to their statistical self-similarity, seemingly small fluctuations of fBm series carry statistical information about the global structure of the series. It is important to note that, in spite of this, for practical reasons, methods for the decomposition of fractals into signal and noise components have been developed in a few studies (Azami, Bozorgtabar, and Shiroie, 2011; Wornell and Oppenheim, 1992).

Gilden, Schmuckler and Clayton (1993) proposed that people treat fractal patterns *as if* they had a natural decomposition into signal and noise components. In Chapter 2 I provide evidence supporting this view. I, therefore, expected people to make forecasts from fractal series in a way that is similar to that they use to make forecasts from series that can be decomposed into signal and noise. That is, I expected people to extrapolate from them in a way that suggests that they imitate the ‘noise’ component of the data series. Furthermore, consistently with Harvey (1995), I hypothesise that the noise level in a sequence of forecasts is negatively correlated with the Hurst exponent of the time series (Hypothesis H<sub>3,1</sub>).

#### *Task instructions*

In judgmental forecasting papers, the number of required forecasts has typically been predetermined. For instance, in one of their experiments, Harvey, Ewart, and West (1997) asked participants to provide six forecasts for each graph.

It is possible that asking participants to provide a fixed number of forecasts, which is larger than one, could affect noise imitation. Consider, for instance, a task in which they are instructed to add five forecast points at pre-determined places (dates) to a given graph.

People might add noise to their forecasts because a straight line is determined merely by two

points: participants might think that had the experimenter thought that the correct answer is a straight line, the experimenter would not have asked them to give five prediction points but merely two. On the other hand, asking participants to provide a pre-determined number of forecast points prevents them also from adding many more points to the given graph. I, therefore, argue that, in order to evaluate the scale of the phenomenon correctly, instructions should not include a pre-determined number of forecasts.

Harvey, Ewart, and West (1997) manipulated the number of required forecasts in their study. Participants were asked to provide either one or six forecast points. They found that the number of required forecast points did not affect forecast accuracy. They suggested that this implied that people added noise to their forecast independently of the number of the required points. To explain their conclusion, they argued that “Patterns cannot be expressed when single forecasts are made [...] [in this case] there are no patterns to mask”. However, I argue that, in fact, participants could consider the pattern formed by their single forecast point and the last data point a signal. Indeed, there are many examples in Gestalt in which people appear to see patterns where there are none. For instance, the gambler’s fallacy was explained using the Gestalt approach by Roney and Trick (2003) and by Du, Zhang, Zeng, Gui, Luo, and Ruan (2008). In addition, it was found that judgmental forecasts depended on the format of the presentation (Harvey and Bolger, 1996). Therefore, participants in Harvey, Ewart, and West’s (1997) study might have referred in their one-point forecasts to the straight line between their forecast point and the last point of the data. I argue that presenting a line between forecast points produced by the participants could reduce uncertainty about this effect.

I do not know of any other study in which the effect of the number of forecast points on noise forecast has been examined. However, if noise imitation is a bias arising from the number of forecast points, I would expect large numbers of forecast points to be associated with noisy forecasts. More precisely, I hypothesise that added noise is correlated with the number of points participants choose to add to the graphs (Hypothesis H<sub>3,2</sub>)

### *Personality traits*

There have been a number of recent reports that traders' financial performance depends on personality variables (Frijns, Koellen, and Lehnert, 2008; Kapteyn and Teppa, 2011; Fenton-O'Creevy, Lins, Vohra, Richards, Davies and Schaaff, 2012; Fenton-O'Creevy, Soane, Nicholson and Willman, 2011; Peterson, Murtha, Harbour, Friesen, 2011; Robin and Strážnicka, 2012). However, to date, there appears to be just one study relating judgmental forecasts from time series to personality traits.

Eroglu and Croxton (2010) examined the effects of personality on judgmental forecasts of daily sales in a fast-food restaurant chain. They assessed personality in terms of the 'Big Five' traits (openness to experience, conscientiousness, extraversion, agreeableness, emotional stability) that have been found to explain much between-individuals variance in a wide variety of tasks (Lang, John, Lüdtke, Schupp, and Wagner, 2011). Eroglu and Croxton found that use of anchoring heuristics (which appears to underlie the trend-damping and overestimation of sequential dependence effects outlined above) increased with conscientiousness but decreased with extraversion. Anchoring is one of the three cognitive heuristics identified by Tversky and Kahneman (1974) as forming the basis of a wide variety of human judgments.

Harvey (1995) argued that people's tendency to imitate noise as well as signal when extrapolating from past data arises because they use another of the heuristics identified by Tversky and Kahneman (1974). Specifically, forecasters use the representativeness heuristic: this heuristic is based on the reasonable assumption that outputs of the same system are more likely to be similar than outputs of different systems. Hence, when forecasting, people attempt to ensure that the sequence of forecasts that they produce closely represents (looks like) the data series. If the conclusions that Eroglu and Croxton (2010) draw from their

findings apply generally to cognitive heuristics (rather than applying only to the anchoring heuristic), I expect imitation of ‘noise’ also to increase with conscientiousness but to decrease with extraversion (Hypothesis H<sub>3,3</sub>).

### *Sense of power*

Another factor known to influence people’s judgement is their current disposition (i.e. way of approaching issues that are more temporary and context-dependent than personality traits). Here I focussed on the effects of sense of power on noise imitation. Power is usually defined in psychology as control over resources or decision processes (Anderson, John, and Keltner, 2012). Anderson et al (2012) added to this definition the ability to influence other people.

No studies appear to have explored the effect of sense of power on financial forecasting. However, Hassoun (2005) studied traders’ emotions and dispositions and found evidence that traders often use expressions describing high or low sense of power. As an example, consider the following quote from a trader in Hassoun’s (2005, page 105-106) study:

“One day I bought 5600 contracts in one hour. For the same client. He’s THE client, you use the formal with him... I once sold 4000 contracts with him, another time 4800; once I bought 3000. But [that one] was the biggest [trade] I’ve done... You’ve got everybody watching you, they can’t believe their eyes. And it was unbelievable - you’d’ve thought we were on the Notionnel. In the space of a minute he’s going, ‘Buy200’, ‘You got it!’, ‘Buy 300’, ‘I’ll give ya 200!’ The NIPs were staring at us, it showed up on the CAC —we were creatures from outer space, there’s no other word for it... Keep in mind that the CAC [Futures] record is 73 000 contracts in one day. Once, at the Sirap, we did 43 000 contracts on the CAC in a single day. We were way over 50% [of pit volume] - we were the kings of the universe! There was nobody but us. You couldn’t do a trade without going to see the Sirap— impossible! I was all over the place. In all the commentaries it was ‘Sirap, Sirap’ all day long.”

Hassoun concluded that sense of power is an especially important disposition on the trading floor. The question here is whether it affects forecasts?

Galinsky, Magee, Gruenfeld, Whitson, and Liljenquist (2008) studied the effects of sense of power on people's performance. They found two main effects. Firstly, they found that people with a high sense of power tend to be more creative and less influenced by environmental cues if they were irrelevant. Secondly, they found that powerful people are influenced by situation cues more if they are perceived to facilitate goals. In the special case of forecasting from fractal time series, these two effects might have opposite influences. If one is to accept the first account, and if noise imitation is considered a type of non-creative conformity with data, I would expect people with a high sense of power to imitate noise less than others. However, if the second argument can be applied to forecasts, and if noise is considered a situational cue facilitating forecasts, then powerful people might imitate noise more than powerless people. I, therefore, expected that sense of power would affect noise imitation (Hypothesis H<sub>3,4</sub>). However, *a-priori*, I cannot say which of these two competing effects would be the dominant one.

#### *Expert forecasts*

In different contexts, it has been shown that financial experts exhibit similar behaviour to that of lay people (Zaleskiewicz, 2011; Muradoğlu and Önköl, 1994). I, therefore, hypothesised that experts would exhibit similar biases to those exhibited by lay people. In particular, I hypothesised that, when asked to produce judgmental forecasts from graphically presented price series, experts would imitate the perceived noise component of the given graph (Hypothesis H<sub>3,5</sub>).

## **The effects of news valence, price trend and individual differences on financial behaviour**

Modern behavioural theories developed to simulate markets typically employ models of agents that exhibit some aspects of human behaviour. By so doing, they provide insight into phenomena that are not explained by classical theories. However, the assumptions underlying agents' behaviour do not always reflect results of psychological studies. There are a number of examples of this.

Harras and Sornette (2011) constructed a market model, in which agents choose at each time step whether to trade or not. Traders in their model use information from three sources: private information, public information, and the expected decisions of other traders. However, their model does not take into account news valence, even though I know that the importance that people attribute to information depends on its valence (Kahneman and Tversky, 1979).

Pfajfar (2013) constructed a model with two agent types: a rational group and a bounded rational group. Agents' forecasts were limited to just two options, perfect foresight or the naive predictor (for whom the forecast was the same as the last data point). However, numerous psychological studies have shown that people exhibit various forecasting biases, including trend damping and adding noise to forecasts (Harvey and Reimers, 2013; Harvey, 1995): human forecasts are rarely perfect or naive.

Anufriev and Panchenko (2009) modeled a market with fundamentalists and trend-following agents, assuming that all agents were risk averse. However, psychological studies have shown that some people are risk seeking rather than risk averse (Nicholson, Soane, Fenton-O'Creevy, Willman, 2005; Cheung and Mikels, 2011).

To some extent, these mismatches reflect the simplifications necessary to ensure that mathematical manipulation of the equations within the models is tractable (De Grauwe, 2010). Inappropriate assumptions may also reflect lack of communication between those

working within behavioral finance and psychology. However, financial modelers could also legitimately point out that the psychological literature typically supplies disconnected principles for human behavior that are not always easy to apply to trading environments. My first aim was to provide data that is more specifically relevant to the concerns of those developing agent-based simulations of market behavior.

I focus on three main topics: the way people incorporate news and graphically presented price series into their financial decisions, the time they take to make those decisions, and the effect of individual differences on their decisions. All three topics are addressed, explicitly or implicitly, in behavioural models of the market. Related assumptions are also present in classical models. For example, the Efficient Market Hypothesis (EMH) requires that news is incorporated into asset prices immediately and in an unbiased manner (Malkiel and Fama, 1970). Incorporation of news should, therefore, be independent of individual differences (Findlay and Williams, 2000 - 2001). My second aim is to test these assumptions and develop an account of trading that can accommodate the findings.

#### *The effect of news on financial decisions*

Different versions of the EMH define the scope of the information to be included in prices. This information varies from the previous price series (the weak version) through all publicly available information (the semi-strong version) to all information (the strong version). The semi-strong and strong versions of the EMH therefore assert that news cannot be used by investors in order to make profit (Findlay and Williams, 2000 - 2001).

Nevertheless, a large number of studies have demonstrated that news has a large effect on investment decisions and price series (Hayo and Neuenkirch, 2012; Engelberg and Parsons, 2011; Cecchini, Aytug, Koehler, and Pathak, 2010; Barber and Odean, 2008; Reeves and Sawicki, 2007; Tetlock, 2007).

How do people respond to news? Chapter 5 reports studies designed to address this issue. Caginalp, Porter and Hao (2010) have produced evidence implying that people underreact to news when valuing asset prices. However, De Bondt and Thaler (1985) argued on the basis

of their analysis of winner and loser portfolios that they over-react to news. Moreover, Tuckett (2012) has shown that investors construct narratives in order to give their world meaning and to enable them to function under conditions of extreme uncertainty. Narratives were shown to be essential; for instance, Taffler and Tuckett (2012) interviewed 134 fund managers. Fund managers were asked to tell about their successes and failures. Taffler and Tuckett showed that fund managers' narratives reflected a meta-narrative, which is a core belief about the way the market functions. Narratives served as a tool to preserve the meta-narrative intact, even in face of a contradicting reality. Thus, I argue that people may attribute high importance to news because news items *are* the narratives of the financial world: they describe, or at least give the illusion, of causality, whereas price graphs that appear largely random may not offer the same degree of psychological comfort. I therefore hypothesize that people will choose to base their trading strategy on news more than they do on price graphs (Hypothesis H<sub>4,1</sub>).

Andreassen (1990) used experiments to study the conditions under which overreaction to news occurred, and, in particular, the effect of contradiction between news items and stock price trends on financial decisions. He presented his participants with 60 experimental trials, each consisting of a display of the current price of a stock, the price change from the previous trial, and a news item about the stock. Participants were instructed to "buy shares for less than you sell them" and "sell them before they do down". There were three experimental conditions. In the first condition, participants saw no news; in the second, they saw 'normal' news; in the third, they saw 'reversed news'. 'Normal' news items were positive when price trend was positive and negative when price trend was negative. The valence of 'reversed' news was opposite to the sign of the price series trend. Trends in the series were manipulated as well. The main dependent variable was participants' 'tracking', measured by the correlation between the number of shares held at the end of each trial and the concurrent price. Andreassen (1990) found that tracking was the highest in the reverse-news and no news conditions, and weakest in the normal news condition. That is, buy/sell

decisions depended on prices more when news valence contradicted the trend of the price series than when prices movements were in agreement with news valence.

Oberlechner and Hocking (2004) performed a large-scale survey to examine the views that foreign exchange traders hold on news available to market participants. In line with the results of Andreassen (1990), they found that news items that were consistent with market expectations were considered less important than those that were inconsistent with them. Hence, I hypothesize that participants will track prices more and show more active trading (buying or selling rather than holding their assets) in non-contradicting conditions than in contradicting ones (Hypothesis H<sub>4,2</sub>).

Andreassen (1990) did not examine the effect of each of the four possible combinations of news valence and price trend separately. Considering only contradicting versus non-contradicting results masks any effects of news valence. However, it is known that people react to good and bad news in an asymmetric way. For instance, Galati and Ho (2003) found that people sometimes ignore good news but react to bad news. Hence, on the basis of their results, I hypothesized that people will sell more assets when the news is bad than they will buy when it is good (Hypothesis H<sub>4,3</sub>).

#### *The timing of financial decisions*

The second assumption of the EMH deals with trading latencies of market participants. Trading latency is a measure for the time required for an investor to make a buy or sell decision. Nearly all behavioral models have to make some assumptions about agents' trading latencies. For instance, Kuzmina (2010) assumed that all market participants submit their trades simultaneously. In addition to modeling considerations, investment timing affects market behavior. Indeed, Odean (1998) showed that traders tend to sell winning assets too early and hold losing assets too long.

The psychological basis for the timing of financial decisions has not been subject to intensive investigation. However, Lee and Andrade (2011) found that participants in whom

they had induced a sense of fear tended to sell stock earlier than participants in a control condition. They chose to manipulate fear because it is increased by risk and uncertainty. Their results therefore imply that financial risk and uncertainty reduces trading latency.

In our task, trading latency was defined as the number of data points that participants saw before they made a buy/sell decision. In those cases in which participants chose to hold their shares until the end of the series, trading latency was defined as the maximum number of series points<sup>1</sup>. On the basis of Lee and Andrade's (2011) findings, I hypothesized that trading latency would be shorter when uncertainty is higher, that is, when there is an inconsistency between news valence and price trends (Hypothesis H<sub>4,4</sub>). Also, if I am correct in hypothesizing that people rely more on news than on price trend data when making financial decisions, then I would expect that the effect of news on trading latencies will be stronger than that of the price trend, and that trading latency will be shorter when news is bad (Hypothesis H<sub>4,5</sub>).

#### *Individual differences: Effects of culture*

The trader rationality assumption of the EMH requires homogeneous trader groups. However, this assumption does not hold (Lo, Repin and Steenbarger, 2005). Ackert, Church, and Zhang (2002) conducted experimental markets in the US, Canada, and China in order to examine the effect of imperfect private information on information dissemination. In their markets, traders were given information about period-end dividend. The researchers manipulated the accuracy level of the information given to traders. They defined degree of information dissemination as the movement in transaction price towards the price given to well-informed agents. They found that degree of information dissemination depended on the accuracy of the given information and on participants' nationality. When accuracy of information was 90%, news dissemination was greater in the USA and Canada than in

---

<sup>1</sup> The graphs that participants saw showed asset price as a function of time. Hence, trading latency represented the date on which participants made their financial decision in the virtual trading task rather than the actual duration of each trial.

China. However, when information accuracy was 75%, it was higher in China than in Canada and similar to that observed in the USA.

Inaccurate or misleading information can be represented by a mismatch between news items and price graph trend. In line with the findings of Ackert et al (2002), I hypothesize that participants from Western culture will react to news more than participants from Eastern countries in consistent conditions (good news with positive price trend or bad news with negative trend) but that participants from Eastern Asian countries will react to news more than participants from Western countries in inconsistent conditions (good news with negative trend or bad news with positive trend) (Hypothesis H<sub>4,6</sub>).

Nisbett (2003) has carried out a program of work that indicates that people in Eastern cultures think more holistically and less analytically than those in Western ones. They make greater attempts to pull all available evidence into a single holistic framework.

Consequently, I expect them to require more time to produce a narrative that meets their adequacy criteria. If trading requires development of such narratives, they should exhibit longer trading latencies (Hypothesis H<sub>4,7a</sub>) that would, in turn, result in higher degrees of dispersion in their returns (Hypothesis H<sub>4,7b</sub>).

#### *Individual differences: Effects of personality*

Only Durand, Newby and Sanghani (2008) and Durand, Newby, Peggs and Siekierka (2013) have systematically studied how trading decisions are affected by the big five personality traits (McCrae and Costa, 1987; Norman, 1963). Based on results from their investor survey, Durand et al (2008) argued that people with different personalities are attracted to different types of security: for example, those who were more extraverted had a greater preference for innovation. Based on results from their trading experiment, Durand et al (2013) went on to argue that personality influences not only what people trade in but also how they trade. For example, people more open to experience developed more diversified portfolios.

The trading task used here was simpler than the one used by Durand et al (2013).

Participants were not required to form portfolios of investments. They merely had to decide whether to sell, hold, or buy a series of 12 assets. I ask whether personality influences performance even in this basic trading task. From a sense-making perspective, I expected that it would do so.

It is known that people more open to experience have shorter reaction times in a variety of (non-financial) tasks (Fiori and Antonakis, 2012). This is probably because those who are more open to experience have a greater need for cognition (Sadowski and Cogburn, 1997). People with higher need for cognition put more cognitive effort into tasks and hence process the information they are given more selectively and effectively (Cacioppo, Petty and Morris, 1983). This implies that people more open to experience will put more effort into making sense of trading-related information and succeed in doing so sooner. As a result, they will have shorter trading latencies (Hypothesis H<sub>4,8</sub>). Faster trading may, in turn, influence share buying and resulting returns, as buy/sell decisions may be made in different market conditions.

#### *News relevance*

In their survey, Oberlechner and Hocking (2004) found that foreign exchange traders attributed high relevance to news items which were perceived as being able to influence the market. Thus, in the trading task, I expected a positive correlation between views about the extent to which an event would affect prices and final share number (Hypothesis H<sub>4,9</sub>).

#### *The effects of news and graphs trend on forecasts and financial decisions*

Despite a large literature on judgmental forecasting (Lawrence, Goodwin, O'Connor and Önkal, 2006), Harvey's (2010) study appears to be the only one that has established a connection between financial forecasts and decisions – and those were managerial rather than financial decisions. Andreassen (1990) merely conjectured that forecasts mediate between data and decisions. I hypothesise that forecasts mediate between data and decisions.

In other words, they are affected by news and graph trends. Hence, the difference between a participant's forecast and the last data point should depend on the news valence and the direction of the trend in the price data (Hypothesis H<sub>4,10</sub>). Furthermore, there should be a positive correlation between that difference and final share number (Hypothesis H<sub>4,11</sub>).

### **Mechanisms preserving asset price graph structure**

Economic systems are extremely complex: they involve millions of traders and investors, and are non-deterministic (Matilla-García and Marín, 2010). Nevertheless, the theoretical justification for many forecasting methods and financial models is that certain parameters of the system are constant. For example, in the context of forecasts, Hyndman and Athanasopoulos (2013, Section 1.1) wrote: “What is normally assumed is that *the way the environment is changing* will continue into the future. That is, that a highly volatile environment will continue to be highly volatile; a business with fluctuating sales will continue to have fluctuating sales; and an economy that has gone through booms and busts will continue to go through booms and busts”. Similar assumptions on the stability of the variance were made by Black and Scholes in the context of option pricing (“The variance rate of the return on the stock is constant”, Black and Scholes (1973). page 640).

What mechanisms enable financial markets to maintain stability of certain parameters, at least for periods long enough to make forecasts and financial modelling feasible?

I suggest that traders' behaviour depends on the way that they perceive financial time series and make forecasts from them. Their perception of, forecasting from, and trading on these series may be one of the mechanisms which stabilises markets. I examine people's perception through the way they employ two frequently used data presentation techniques: time-scaling and moving average filters.

I investigated these ideas using fractal time series, as certain fractal properties of time series have been shown to remain stable in financial data over long periods of time (Parthasarathy, 2013; Malavoglia, Gaio, Júnior and Lima, 2012; Sun, Rachev and Fabozzi, 2007; In and Kim, 2006). Furthermore, as explained before, among the variables which are correlated with the Hurst exponent in graphically presented series (provided that they were generated by the same algorithm) are local steepness, defined as the average of the absolute value of the gradients of the graph, oscillation, defined as the difference between the maximum and minimum values of the graph over a given interval, and the standard deviation, which corresponds to historical volatility. Knowledge of the way people respond to these properties of fractal stimuli is likely to have financial implications.

*Models and theories about stability of market parameters: the effects of time-scaling*

Referring to the question of why markets sustain stable fractal qualities for long durations, Mandelbrot and Hudson (2004, page 239) wrote: “In the case of cotton, I found all the price variations followed the same statistical properties for days over a few decades and for months over eighty years. All of the lines were equally wiggly. Why would this be? First, I surmise, economics differs from physics in having no intrinsic time scales. The chart of a day’s activity looks like that of a month because, from the narrow viewpoint of the probability of losses or gains, a day really is like a month. Yes, some time-scales have some meaning: Companies report their financial results quarterly and annually. A trading day has its own internal rhythm [...] These differences are nothing like the immutable, fundamental differences in time scale that arise in physics. There is, in finance, no barrier like that between the subatomic laws of quantum physics and the macroscopic laws of mechanics”. Mandelbrot and Hudson’s account is compelling. However, it does not provide an insight into the human factors which accumulate to produce the market’s behaviour. I do not know of any psychological study examining this question.

It has been recognised for some time that market participants are heterogeneous (e.g., Müller, Dacorogna, Davé, Pictet, Olsen, and Ward, 1993). However, Peters (1994, pages 44-46) went further in suggesting that people's varying perspectives and the manner in which they perceive price series are sources of both the liquidity and the fractal behaviour of the market: "Markets remain stable when many investors participate and have many different investment horizons. When a five-minute trader experiences a six-sigma event, an investor with a longer investment horizon must step in and stabilize the market. The investor will do so because, within his or her investment horizon, the five-minute trader's six-sigma event is not unusual. For this reason, investors must share the same risk levels (once an adjustment is made for the scale of the investment horizon), and the shared risk explains why the frequency distribution of returns looks the same at different investment horizons... The fractal statistical structure exists because it is a stable structure". Some of Peters' predictions have been verified (Kristoufek, 2012).

Inspired by these ideas, Corsi (2009) constructed a model that takes into account the different volatility components that result from the actions of short, medium, and long term traders. He wrote (page 178): "Typically, a financial market is composed of participants having a large spectrum of trading frequency. At one end of the spectrum we have dealers, market makers, and intraday speculators, with very high intraday frequency as a trading horizon. At the other end, there are institutional investors, such as insurance companies and pension funds who trade much less frequently and possibly for larger amounts. The main idea is that agents with different time horizons perceive, react to, and cause different types of volatility components". Corsi's (2009) model produced financial return series that exhibited fractal properties such as self-similarity, long memory, and fat tail distributions. In addition, Corsi claimed that short-term traders use both short and long term considerations to make their decisions whereas long term traders take account of only long term volatility considerations.

These authors did not examine human behaviour: they did not test their assumptions and models. Within psychology, the effect of forecast horizon on forecasts has been investigated (Bolger and Harvey, 1993; Lawrence and Makridakis, 1989) but no studies have been reported on the effects of forecast horizon on people's choice of the length of series they wish to display as a basis for their forecasts. Here I allow people to vary temporal scaling between small scale (presentation of asset prices of long period of time over an interval on the x-axis of a certain length) and large scale (presentation of asset prices of short period of time over an interval of the same length on the x-axis).

*The effects of forecast horizon on chosen time scaling, properties of scaled graphs, and forecasts*

Many financial data services (e.g. Yahoo! Finance, <http://finance.yahoo.com>) enable traders to scale presented price graphs. (For instance, Yahoo! Finance allows the viewers to scale graphs by either setting their time-domain or by continuously dragging the mouse on the graphs). Following the Heterogeneous Market approach of Peters (1995), Müller et al. (1993), and Corsi (2009), I hypothesise that people will exhibit a large degree of variation in their choice of temporal scaling (Hypothesis  $H_{5,1a}$ ) and that this variability will be greater for more distant trading horizons (Hypothesis  $H_{5,1b}$ ).

The resolution of financial data is high, but finite. Therefore, scaling-down (that is, zooming-in along the x-axis and presenting data representing a shorter period of time over the same actual interval length) typically decreases the local gradients of the graphs. In addition, the maximal values of a subset are smaller or equal to those of any including set, and its minimal values are larger or equal to those of any including set. Therefore, the oscillations of scaled-down graphs are smaller or equal to those of the original graphs. Examples of the effect of scaling-down of graphs with low and high Hurst exponents are presented in Figure 1.2. Because of these effects, I hypothesise that the effect of forecast horizon on chosen time scales suggested in Hypothesis  $H_{5,1b}$  would result in a corresponding

effect on the geometrical properties of the presented graphs. That is, I suggest that there should be a positive correlation between forecast horizon and the local steepness and oscillation of the time-scaled data graphs (Hypothesis H<sub>5,2</sub>).

Although the effect of scaling the vertical axis of a graph (Lawrence and O'Connor, 1992) has been studied by researchers of judgmental forecasting, scaling of the horizontal time axis has not. However, Athanassakos and Kalimipalli (2003) found a strong correlation between analysts' forecast dispersion and future return volatility. If forecast horizon affects market's volatility through financial forecasts, I expect dispersion of participants' forecasts to be positively correlated with the required forecast horizon (Hypothesis H<sub>5,3</sub>).

The above hypotheses address Corsi's (2009) model and thus also the formation of fractal price series. However, I still need to consider what processes stabilise the geometric properties of the resultant time series.

*The effects of the Hurst exponent on chosen time scaling, properties of scaled graphs, forecasts, and financial decisions*

Mandelbrot and Hudson (2004) emphasised that the way that investors perceive geometric properties of price graphs is likely to affect their risk perception. In line with this, Manzan and Westerhoff (2005) found that inclusion of agents' reactions to volatility in a market model resulted in realistic estimates of exchange rates.

These studies lead me to expect that people react to the geometric structure of the price series in addition to trading horizons. As mentioned before, scaling a graph changes the visual properties of the graph, and, in particular, the perceived noise level (see Figure 1.1 and Figure 1.2). In light of Gilden et al's (1993) findings, I anticipate that people will prefer to make forecasts from graphs with lower perceived noise levels because it is easier to decompose the data series into perceived signal and noise components. Thus I expect that chosen time scaling factors will be smaller for graphs that have smaller Hurst exponents.

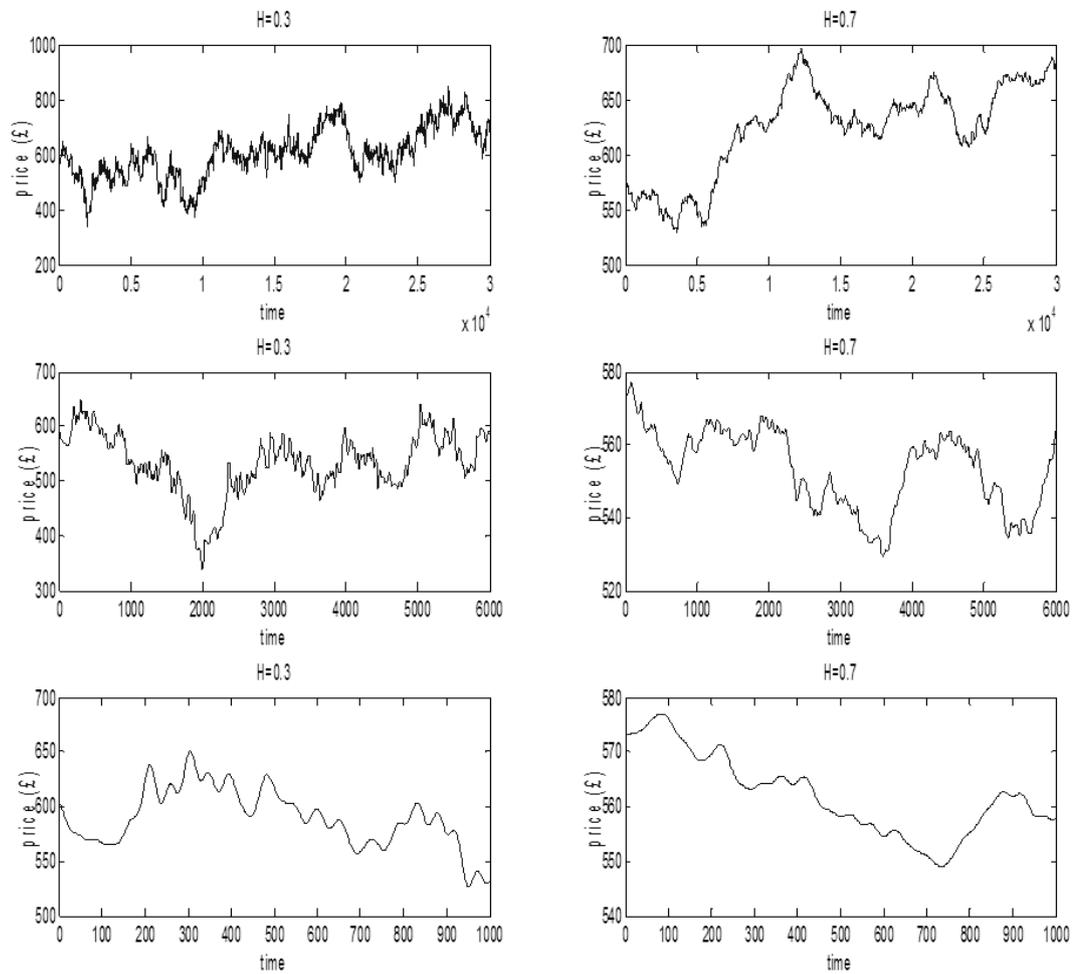


Figure 1.2 Example of fBm series with  $H=0.3$  (left panels) and  $0.7$  (right panels). Graphs in the first row show data referring to 30000 days, graphs in the second row show data referring to 6000 days, and graphs in the third row show data referring to 1000 days. All graphs are plotted on intervals of the same length along the x-axis.

That is, people prefer presentation of data corresponding to shorter periods of time when dealing with graphs with smaller Hurst exponents. (Hypothesis H<sub>5,4</sub>).

In contrast to Manzan and Westerhoff (2005) and Mandelbrot and Hudson (2004), I focus on the way people's geometric perception acts to preserve the structure of price graphs. Gilden et al's (1993) argument may lead one to conjecture that the attempt to reduce graphs' noise by scaling graphs could result in graphs which have the same perceived noise level, independently of their original Hurst exponents. However, equating the local steepness of graph with low Hurst exponent to that of graphs with high Hurst exponents requires a very large change of scale. For example, see figure 1.2; scaling a graph with  $H = 0.3$  presented on the interval  $[0, 30000]$  to the interval  $[0, 6000]$  yields a graph which still looks locally steeper than a graph with  $H = 0.7$ , presented on the interval  $[0, 30000]$ . In order to equate the local steepness of the graph with  $H = 0.7$  to that of the graph with  $H = 0.3$ , time-scaling ratio of more than 5 is required. In a different context, it was found that people do not match perfectly their performance with data (e.g., when making forecasts from trended data, they damp the trend (Harvey and Reimers, 2013)). Therefore, I hypothesise that people will not equate properties of scaled graphs of data with low Hurst exponents to that of data with high Hurst exponents. Consequently, the time scales that people choose result in a negative correlation between the local steepness and oscillation of the time-scaled graph and the Hurst exponent of the original data (Hypothesis H<sub>5,5a</sub>) and in a positive correlation between the local steepness and oscillation of the time-scaled graphs and of the original graphs (Hypothesis H<sub>5,5b</sub>). Furthermore, as forecast quality depends on the noise level of the data and people try to imitate properties of data in their forecasts (see Chapter 4), I hypothesise that the dispersion of people's forecasts will be negatively correlated with the Hurst exponents of the original graphs and positively correlated with the local steepness and oscillation of the data graphs (Hypothesis H<sub>5,6</sub>).

Finally, I expect that people's trading behaviour to depend on their forecasts (Hypothesis H<sub>5,7</sub>).

Notice that, along with the results of Athanassakos and Kalimipalli (2003), the process described in Hypotheses  $H_{5,4}$ ,  $H_{5,5}$ ,  $H_{5,6}$ , and  $H_{5,7}$  provides a mechanism that preserves the properties of price series. The suggested process is shown in Figure 1.3. Indeed, I assume that at any given moment, people examine financial series which exhibit certain geometrical properties, such as local gradients and oscillations. I argue that people actively choose the way they perceive such graphs through their choices of scales. In line with previous literature (Corsi, 2009), I suggest that people's scaling choices are highly variable. However, I hypothesise that their scaling means are correlated with the geometrical properties of the data. I suggest that these scaling choices result in scaled graphs, which have properties that are correlated with those of the original graphs. People, then, make forecasts from the scaled graphs. I further suggest that the dispersions of these forecasts depend on the properties of the data. That is, I hypothesise that forecasts from data that are characterised by larger local gradients and oscillations, will exhibit larger dispersion. According to Athanassakos and Kalimipalli (2003), large forecast dispersion is associated with larger future return volatility, which is, in turn, correlated with larger local gradients and oscillations of price series. The latter relies on a connection between forecasts and financial decisions. Hence, actions based on data with large local gradients and oscillations will result in future asset price series, which have the same properties.

#### *Moving average filter models*

De Grauwe and Grimaldi (2005) constructed a market model, which included agents acting as fundamentalists and chartists. Chartists in their model computed moving averages of past exchange rate changes and used the results of these calculations to produce forecasts.

Indeed, moving average filters are a commonly offered option in financial data analysis programmes (e.g. Yahoo! Finance, <http://finance.yahoo.com/>) and are highly popular among traders (Glezakos and Mylonas, 2003). De Grauwe and Grimaldi managed to demonstrate evolution of fat-tailed distributions. Their findings indicate that the way people use moving average filters might have a role also in preservation of price series properties.

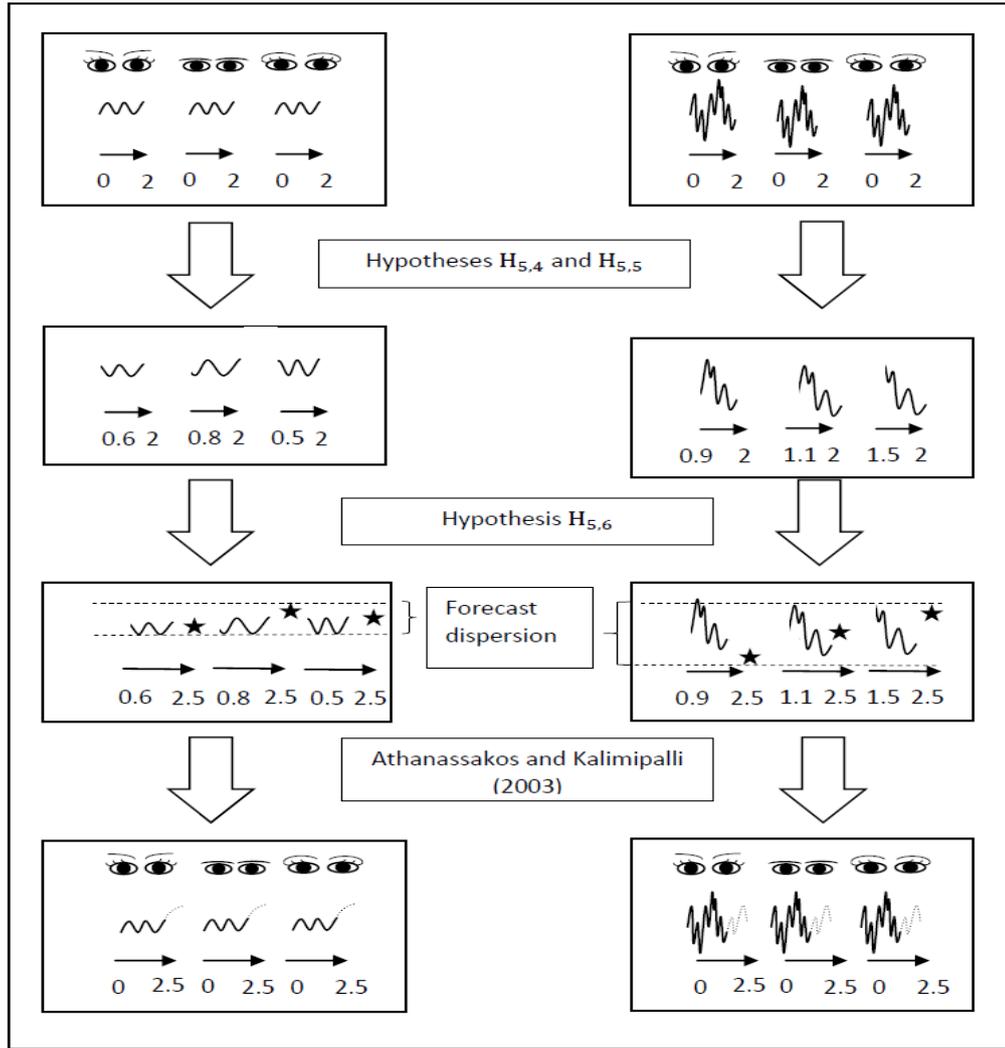


Figure 1.3 Illustration of the mechanism which preserves geometrical properties of price graphs. The left column illustrates graphs with low local steepness and oscillation, and the right column presents graphs with high local steepness and oscillation. People observe data characterised by different properties (panels on the first row). They choose smaller scaling factors and time periods to present graphs with higher local steepness and oscillation. However, the scaled graphs still preserve properties of the original graphs (panels on the second row). Next, people make forecasts from the graphs (forecasts are marked with stars). Correspondingly, forecast dispersions are higher for the steeper graphs (panels on the third row). This process results in price graphs with properties that are correlated with those of the original data.

However, I do not know any study on the effect of moving average filters on forecasts. Furthermore, De Grauwe and Grimaldi chose geometrically declining weights for the filters used by their agents, though people may use different filters in different situations. I was interested in two factors which could affect individual choices of sizes of moving average filter. The first factor was the geometrical properties of the price series. The second factor was the required forecast density. Stock market investors are required many times to make forecasts for multiple time horizons (Pesaran, Pick, and Timmermann, 2011).

*The effect of the Hurst exponent on the window size of a moving average filter and financial forecasts*

In line with the Heterogeneous Market Hypothesis of Müller et al. (1993) I, firstly, hypothesise that, when people are presented with fractal price graphs and are given an opportunity to vary the width of a moving average filter applied on the graph, the variance of the choices of averaging windows is substantial (Hypothesis H<sub>5,8a</sub>).

Application of a moving average filter acts to smooth graphs, and, in fact, is considered a method of noise elimination. Therefore, as in the case of graph scaling, I follow Gildea et al (1993) and hypothesize that the Hurst exponent affects the choice of filter size. More specifically, I hypothesise that chosen smoothing factors are smaller when Hurst exponents are smaller (Hypothesis H<sub>5,8b</sub>). (That is, people zoom-in more when graphs with low H values are presented than when graphs with high H values are presented),

As before, I suggest that chosen smoothing factors result in graphs whose properties are correlated with those of the original graphs. That is, there is a negative correlation between the Hurst exponent of the original data and the local steepness and oscillation of the smoothed graphs (Hypothesis H<sub>5,9a</sub>), and that there is a positive correlation between the local steepness and oscillation of the smoothed data graphs and the original ones (hypothesis H<sub>5,9b</sub>).

People imitate noise of data series (Harvey, 1995). I suggest that when people are asked to make a sequence of forecasts from fractal graphs, the local steepness and oscillation of the forecast sequence are positively correlated with the local steepness and oscillation of the smoothed graphs, respectively, and negatively correlated with the Hurst exponent of the data graphs (Hypothesis  $H_{5,10}$ ). Hence, volatile price series result in noisy forecasts, which, in turn, may increase market's volatility.

*The effect of forecast density on the window size of a moving average filter and financial forecasts*

Though the judgmental forecasting literature includes many studies on multi-period forecasts (Harvey, 1995; Harvey and Reimers, 2013), I know of no research examining the effects of forecast density on the forecasts. I hypothesise that people use the required forecast dates as a forecast cue and, hence, try to match the resolution of the data to that of the required forecast grid. More precisely, I hypothesise that chosen smoothing factors are smaller when forecast densities are larger (Hypothesis  $H_{5,11}$ ), and that there is a positive correlation between the local steepness and oscillation of the smoothed data graphs and the required density of forecasts (Hypothesis  $H_{5,12}$ ). As data which is perceived to be noisier would result in noisier forecasts, I conjecture that local steepness and oscillation of the forecasts is positively correlated with the required density of the forecast (Hypothesis  $H_{5,13}$ ).

In Chapter 6, I report the effects of scaling, forecast horizons, size of moving filter averaging, and the density of the required forecast on forecasts. I examine the question whether the way people perceive data and make forecasts from it could be one of the mechanisms that preserve the structure of financial time series. Moreover, I examine the correlation between forecasts and financial decisions.

### **Part III: Mathematical aspects**

*“Unfortunately, the world has not been designed for the convenience of mathematicians”*  
(Mandelbrot and Hudson, 2004, page 41).

In the following section, I present the formal definition of fBm and fGn series. In addition, I discuss different aspects of work with fractal graphs in the psychology laboratory: the advantages and disadvantages of computer-generated and real-life series as experimental stimuli; the method I employed in order to generate fractal graphs; methods for Hurst exponent analysis; criteria for the choice of financial time series; notes about the way I presented fractal graphs in the experiments; and the effects of normalisation of fractal graphs.

#### **Definition of fBm and fGn series**

Fractional Brownian motion, with a Hurst exponent,  $H$ , is a series which satisfies the condition that the variance of the differences between outputs  $X(t)$  at times  $t_1$  and  $t_2$  is proportional to the difference between those times to the power  $2H$ :

$$(1) \text{ var}(X(t_2) - X(t_1)) \propto |t_2 - t_1|^{2H}, \text{ where } 0 < H < 1$$

(Peitgen and Saupe, 1988). For a random walk, the differences  $(X(t_2) - X(t_1))$  have a Gaussian distribution and satisfy (1) with  $H = 0.5$ . When  $H$  is above 0.5, series are termed persistent: outputs change their direction less frequently than they do in a random walk. When  $H$  is below 0.5, series are called anti-persistent: outputs reverse their direction more frequently than they do in a random walk.

An important property of fBm series is that they are statistically self-similar with respect to  $H$ : in other words,  $X(t_0 + t) - X(t_0)$  and  $\frac{1}{r^H}(X(t_0 + rt) - X(t_0))$  have the same distribution

functions for any  $t_0$  and  $r > 0$ . It can be shown that the fractal dimension ( $D$ ) of an fBm series with Hurst exponent  $H$  is given by  $D = 2 - H$  (see Peitgen and Saupe, 1988).

The Hurst exponent values of many financial series lie in the interval  $H \in [0.4, 0.6]$  (Sang, Ma and Wang, 2001). Figure 1.1 shows fBm series with nine different  $H$  exponents from 0.1 (anti-persistent) through 0.5 (random walk) to 0.9 (persistent).

If  $X(t_n)$  is an fBm series, then the increment process,  $Y(t_n) = X(t_{n+1}) - X(t_n)$  is termed the fractional Gaussian noise (fGn series). Figure 1.4 presents fGn graphs with different Hurst exponent values. Figure 1.5 presents fBm series with  $H = 0.3, 0.5, 0.7$  and their corresponding fGn series.

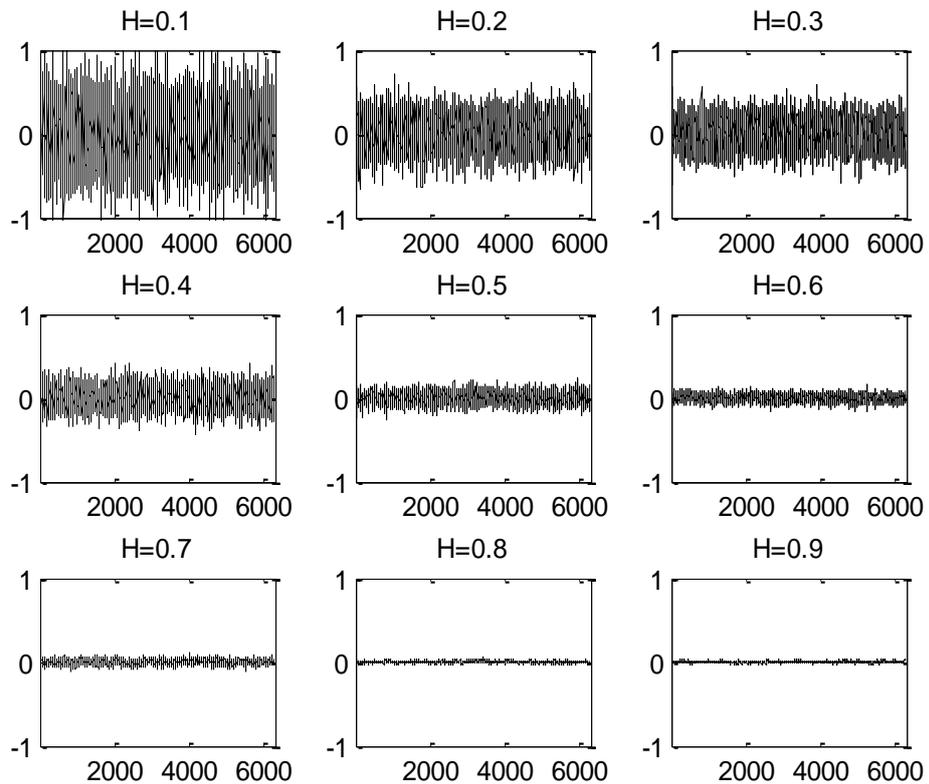


Figure 1.4 Examples of price change series with Hurst coefficients ranging from 0.1 (anti-persistent) through 0.5 (random walk) to 0.9 (persistent) in 0.1 increments.

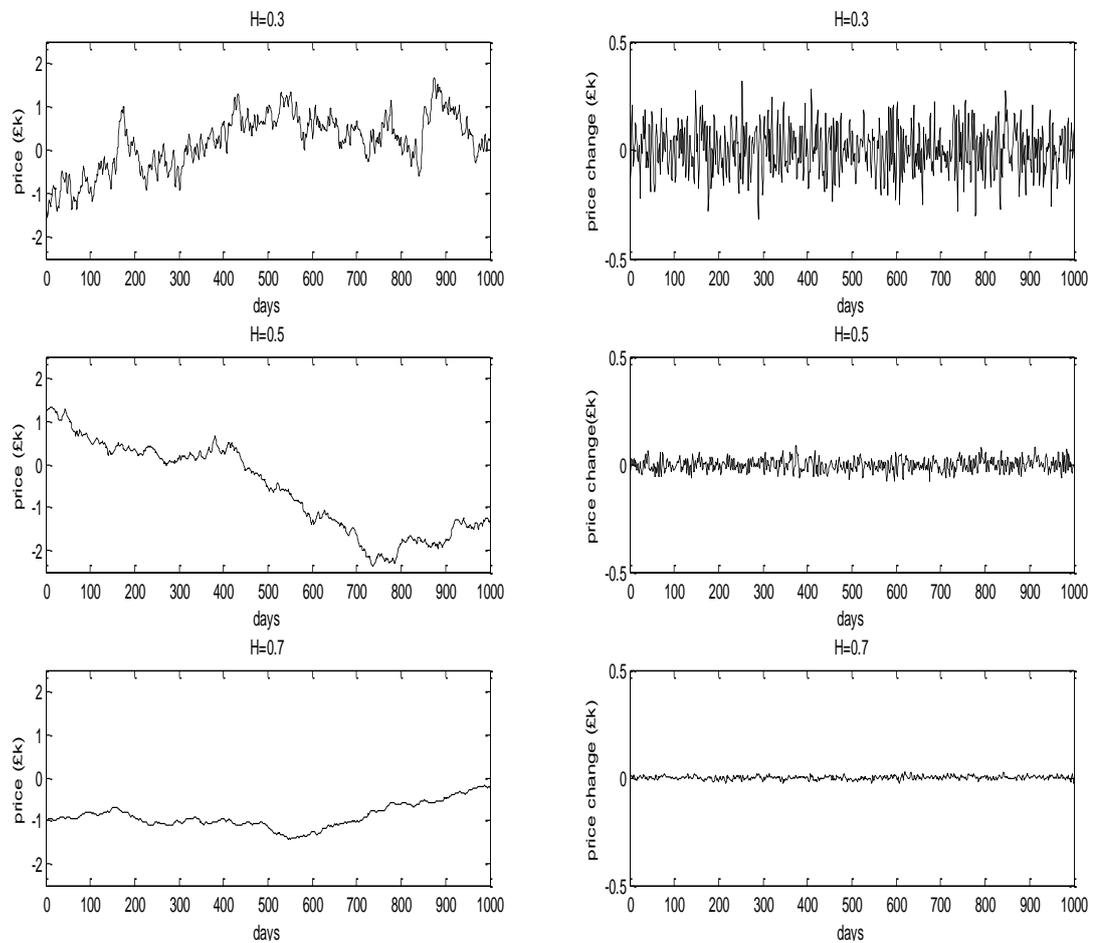


Figure 1.5 FBM series with  $H = 0.3, 0.5, 0.7$  (left column) and their corresponding fGn series (right column).

### Fractal series as experimental stimuli

#### *Advantages and disadvantages of computer-generated and real-life fractal series as experimental stimuli*

Fractal time series can be categorised according to their source: computer-generated graphs (artificial fractals), and real-life asset price graphs. Fractal generation programmes allow accurate control of the Hurst exponent in artificial series (Peitgen and Saupe, 1988). In addition, a large number of graphs with a wide range of Hurst exponents (e.g.

$H \in [0.1, 0.9]$ ) can be produced in short periods of time. Therefore, fractal generation programmes can be used to produce convenient experimental stimuli. Furthermore, the ease of production of experimental stimuli contributes to the robustness of the statistical analysis of the results.

On the other hand, the ecological validity of computer-generated series is lower than that of real-life asset price series. The Hurst exponents of real-life assets usually satisfy  $H \in [0.4, 0.6]$ , and therefore, an attempt to strengthen statistical analysis by using artificial series reflecting a wide range of Hurst exponent ( $H \in [0.1, 0.9]$ ) might result in a lower external validity. Moreover, it is difficult to construct reliable measures for accuracy of prediction from artificial graphs (Armstrong and Fildes, 1995). Quality of forecasts from real asset price graphs can be assessed by comparing the participant's predictions to the historical evolution of prices.

However, the methods that are available for evaluating the Hurst exponents of real fractal series are inaccurate (Delignières, Ramdani, Lemoine, Torre, Fortes and Ninot, 2006). In addition, it is difficult to find real series that meet accepted stability criteria (Sang, Ma and Wang, 2001).

I, therefore, decided to employ both computer-generated and real asset time series in the experiments. Computer-generated series were employed whenever stimuli with accurately known values of Hurst exponents were required. I used real asset price graphs for the evaluation of the quality of participants' forecasts.

#### *Generation of fractal time series*

All computer-generated time series used as experimental stimuli in the studies were fBm series. They were generated in Matlab using the spectral method described by Saupe (Peitgen and Saupe, 1988).

According to Saupe, a discrete approximation of fBm process with  $H = \frac{\beta-1}{2}$  can be generated by the random function

$$(1) X_N(t) = \sum_{k=1}^{N/2} [A_k \cos(kt) + B_k \sin(kt)], X(t) = \lim_{N \rightarrow \infty} X_N(t)$$

where

$$A_k = g_k \cos(2\pi \cdot r_k), B_k = g_k \sin(2\pi \cdot r_k),$$

$$r_k = \text{rand}(\cdot), g_k = \text{Gauss} \left[ (k+1)^{-\frac{\beta}{2}} \right],$$

$\text{rand}(\cdot)$  is a function that generates uniformly distributed numbers between 0 and 1, and  $\text{Gauss}(m)$  is a function that generates normally distributed numbers with mean  $m$ . I chose for the experimental series  $N = 3000$  for the calculation of each series point  $X_N(t)$ . The spectral algorithm that I used generated periodic functions, with period length  $2\pi$ . I calculated 6284 ( $\cong 1000 \cdot 2\pi$ ) points for each series.

*Using real asset price graphs in experiments*

*Analysis of Hurst exponents* Many numerical methods have been developed in order to evaluate the Hurst exponent of a given time series. Commonly used methods are rescaled range analysis (R/S), power spectral density analysis (PSD), detrended fluctuation analysis (DFA), maximum likelihood estimation (MLE), dispersional analysis (Disp), and scaled windowed variance methods (SWV) (see Delignieres et al, 2006, for a comprehensive review of these methods).

In 2003, Katsev and L'Heureux showed that accuracy of estimation of the Hurst exponent by numerical codes depends greatly on the length of the series. They concluded (page 1085) "...that the uncertainty in the Hurst exponent values measured from short data sets (less than 500 points) is usually too large for most practical purposes". Delignieres et al (2006) studied the dependence of the accuracy of Hurst evaluation methods on the length of a given series.

They generated fBm and fGn sequences using the algorithm suggested by Davies and Harte (1987) and then systematically evaluated the errors of the calculated Hurst exponent and other parameters found by different methods. Delignieres et al recommended using different evaluation methods for each range of Hurst exponents. (Clearly, for practical applications, in which the value of  $H$  is *a priori* unknown, one should estimate its value using any of these methods, and then refine the estimation by using the method which is relevant to the series' Hurst exponent range.) However, they found that the variance of these methods is considerable for relatively short series. The variances obtained when applying these recommended algorithms to 100 series of different lengths, are given in Table 1.1. It is especially important to note that no single method has been recommended for evaluation of Hurst exponent of both fBm and fGn series (Caccia, Percival, Cannon, Raymond and Bassingthwaigthe, 1997). Cannon, Percival, Caccia, Raymond and Bassingthwaigthe (1997, page 606) wrote: "To have a 0.95 probability of distinguishing between two signals with true  $H$  differing by 0.1 (by numerical codes), more than  $2^{15}$  (32768) points are needed."

Following Delignieres et. al (2006), I used the ldSWV (Scaled Windowed Variance) method to calculate the Hurst exponent of real asset time series. I realised the algorithm described by Cannon et al. (1997) in Matlab. As can be seen in Table 1.1, estimation error could exceed 0.1.

#### *Choice of real-life series*

I used financial time series downloaded from "Yahoo! Finance" (<http://finance.yahoo.com/>). I calculated the Hurst exponents of a large number ( $N > 100$ ) of financial time series over a large range of periods before choosing the stimulus time series. The Hurst exponent was evaluated using the ldSWV algorithm (Cannon et al., 1997). Most of the examined time series were characterised by frequent stock splits and variable Hurst exponents.

Table 1.1 The standard deviation of different methods of evaluation of the Hurst exponent of time series for different series lengths (from Delignieres et al., 2006)

Standard deviation			
Method	Series Lengths		
	128 elements	512 elements	1024 elements
SWV (fBm)	0.03-0.17	0.03-0.16	0.02-0.11
DFA (fGn, $H < 0.5$ )	0.03-0.12	0.03-0.1	0.02-0.075
R/S analysis (fGn, $H > 0.5$ )	0.1-0.12	0.06-0.1	0.06-0.075
MLE (fGn, $H > 0.5$ )	0.07-0.04	0.04-0.02	-

Stock split is an adjustment of the price of an asset which occurs when there is an increase in the number of shares. The price is adjusted in a way that guarantees that the value of the company (number of shares time share price) remains constant. The effect of a stock split is a sharp discontinuity in prices. Although it was possible to adjust the graphs by multiplying the value by the split ratio, I preferred to present the participants actual price sequences.

Large variations in Hurst coefficients were also found to be common. Mandelbrot found that the cotton price maintained a Hurst coefficient which was close to constant value over a period of 100 years (Mandelbrot, 2004). However, Sang et al (2001, page 270) demonstrated that Hurst coefficients of Boeing and IBM changed significantly every few years. For instance, they found that for IBM,  $H$  was 0.37 between 1977 and 1982 but was 0.67 between 1974 and 1976. Sang et al used R/S analysis, which is considered inaccurate (Delignieres et

al. 2006). However, my calculations using the ldSWV algorithm also revealed a high instability in the values of H.

I divided the Hurst exponent range into three sets: Low, Medium, and High Hurst sets. The Low H set was  $H < 0.49$ , the Medium H set was  $H \in [0.5, 0.56]$ , and the High H set was  $H > 0.57$ .

The chosen data consisted of the close prices of financial time series which satisfied all of the following conditions:

- i. The time series had at least 2500 consecutive work days without a stock split.
- ii. The Hurst exponent of the series, as calculated by ldSWV algorithm for the first 1000, 1500, 2000, and 2500 elements of the series, belonged to one of the H-sets described above (Low, Medium, and High Hurst set).
- iii. I denote by  $H(n)$  the value of Hurst coefficient as calculated by ld-SWV algorithm over a period of  $n$  days. During these 2500 days, the value of calculated H did not change substantially, that is:

$$\Delta H = \max_{n \in \{1000, 1500, 2000, 2500\}} |H(n) - \bar{H}| \leq 0.055,$$

where

$$\bar{H} = \frac{1}{4} \sum_{n \in \{1000, 1500, 2000, 2500\}} H(n).$$

The chosen time series reflect wide sections of the market and include, for example, General Electric Co. (GE), Walt Disney Co., Ford, The Children's Place Retail Stores, EUR/USD, FTSE 100, NASDAQ Composite, and Dow Jones Industrial Average. The sampled period of times were also diverse, with starting dates between 1928 (Dow Jones Industrial Average) and 2001 (Ford). The results of the financial time series analysis are given in Table 1.2.

*Presentation of the series* I performed both laboratory and online experiments. I did not control for the number of pixels with which participants saw the graphs in online experiments. On the other hand, in laboratory experiments, I controlled the ratio of the

Table 1.2 The results of Hurst exponent analysis of real financial time series. The classification criterion was  $\Delta H < 0.055$ .

	Series number	Time series	H(1000)	H(1500)	H(2000)	H(2500)	$\Delta H$
H (2500) < 0.485	1	Merck	0.4520	0.4542	0.4588	0.4312	0.0097
	2	Caterpillar	0.4486	0.4180	0.4382	0.4320	0.0144
	3	EI DuPont de Nemours & Co.	0.4620	0.4477	0.4549	0.4462	0.0093
	4	PG	0.4286	0.4591	0.4591	0.4782	0.0220
	5	General Electric Co. (GE)	0.4482	0.4679	0.4520	0.4846	0.0214
	6	Barrick Gold Corporation (ABX)	0.4466	0.4692	0.4601	0.4605	0.0101
	Mean		0.4477	0.4527	0.4539	0.4554	
	Max		0.4620	0.4692	0.4601	0.4846	
	Std		0.0109	0.0188	0.0083	0.0229	
H (2500) < 0.556	1	Ford	0.5171	0.5227	0.5481	0.5364	0.0170
	2	Walt Disney Co.	0.5393	0.5392	0.5517	0.5477	0.0072
	3	Juniper Networks, Inc.	0.5406	0.5252	0.5471	0.5510	0.0100
	4	IBM International Business Machines	0.5195	0.5344	0.5552	0.5360	0.0189

		Corp.					
	5	The Children's Place Retail Stores	0.5097	0.5095	0.5497	0.5347	0.0238
	6	EUR/USD	0.5115	0.5236	0.5052	0.5001	0.0135
		Mean	0.5229	0.5258	0.5428	0.5343	
		Min	0.5097	0.5095	0.5052	0.5001	
		Max	0.5406	0.5392	0.5552	0.5510	
		Std	0.0137	0.0103	0.0187	0.0181	
H (2500) > 0.57	1	FTSE 100	0.6293	0.6361	0.6092	0.5876	0.0205
	2	NASDAQ Composite	0.6135	0.6163	0.6566	0.6954	0.0499
	3	Russell 2000	0.7417	0.6988	0.6621	0.6536	0.0526
	4	Dow Jones Industrial Average	0.6061	0.5753	0.5673	0.5720	0.0259
	5	Composite Index (^JKSE)	0.5830	0.5839	0.5931	0.6055	0.0141
	6	Value Line Arithmetic Index,RTH	0.6432	0.6300	0.6275	0.6250	0.0118
		Mean	0.6361	0.6234	0.6193	0.6232	
		Min	0.5830	0.5753	0.5673	0.5720	
		Std	0.0556	0.0443	0.0368	0.0455	

number of elements series per pixel with which each graph was presented. The programmes of the laboratory experiments were written in Matlab.

It is important to note that, in some of the experiments, I used a whole period of the produced series. This set the difference between the first and last data point to zero. In other experiments, I presented only a part of a period (half a period or a quarter of a period). That enabled me to study the effect of the difference between the first and last data points on the examined variables.

*The effect of normalisation of fractals on their Hurst exponents*

The oscillation (difference between maximum and minimum values) of fBm series is confounded with their Hurst exponent. In some experiments, I wanted to examine the hypothesis that participants react to the Hurst exponents of the presented graphs rather than to their oscillations. For this reason, in those experiments, I normalised fBm series in a way that ensured that all graphs had the same oscillation. Below, I explain why normalisation had only a minor effect on the results of certain experimental procedures.

*Normalisation and assumptions.* In order to normalise a non-constant series  $X_N(t)$  defined on  $[t_0, t_1]$  to an interval  $[y_0, y_1]$ , I multiplied it by the factor

$$(2) \quad a = \frac{y_1 - y_0}{\max_{t \in [t_0, t_1]} X_N(t) - \min_{t \in [t_0, t_1]} X_N(t)}$$

I denote the normalised series by  $\tilde{X}_N(t) = a \cdot X_N(t)$ . I normalised data series to the interval  $[1, 10]$ , and therefore I multiplied them by

$$a = \frac{9}{\max_{t \in [t_0, t_1]} X_N(t) - \min_{t \in [t_0, t_1]} X_N(t)}$$

For example, for  $H = 0.9$  I obtained an average value of  $a = 3$ , and for  $H = 0.1$ ,  $a = 1.125$ .

In order to simplify the following calculation, I assume in this section that

$$(3) A_k \cong (k + 1)^{-\beta/2}.$$

For infinite series, one can derive  $\beta$  from relation (3) by the limit

$$(4) \beta \cong \lim_{k \rightarrow \infty} \frac{-2\log(A_k)}{\log(k + 1)}.$$

The Hurst exponent can then be calculated as  $H = \frac{\beta-1}{2}$ .

*The Hurst exponent of truncated, normalised series.* Clearly, for practical reasons, one cannot generate fractals with infinitely many elements ( $N \rightarrow \infty$ ). Therefore, estimate of the Hurst exponent of truncated, normalised series cannot be performed using the limit process given in equation (4). In particular, for finite series, the expression  $\frac{-2\log(A_k)}{\log(k+1)}$  depends on  $k$ . I estimate the  $\beta$  value of a truncated, normalised series by its value for  $k = N$ .

To estimate the effect of normalisation by a factor  $a$  on a truncated series generated by summing  $N$  elements in equation (1), I denote:

$$(5) \tilde{A}_k = a A_k,$$

and

$$(6) \tilde{A}_k \cong (k + 1)^{-\tilde{\beta}_k/2}.$$

Then, by equation (3),

$$(7) \log(A_k) \cong -\frac{\beta}{2} \log(k + 1).$$

Similarly, by equation 6,

$$(8) \log(\tilde{A}_k) \cong -\frac{\tilde{\beta}_k}{2} \log(k + 1).$$

By equations (5), (7) and (8),

$$\log(\tilde{A}_k) = \log(a) + \log(A_k)$$

$$\Rightarrow -\frac{\tilde{\beta}_k}{2} \log(k+1) = \log(a) - \frac{\beta}{2} \log(k+1),$$

or,

$$(8) \quad \tilde{\beta}_k = \beta - 2 \frac{\log(a)}{\log(k+1)}.$$

Notice that, if  $k \rightarrow \infty$ , then  $\lim_{k \rightarrow \infty} \frac{\log(a)}{\log(k+1)} = 0$ , hence

$$\lim_{k \rightarrow \infty} \tilde{\beta}_k = \beta.$$

Therefore, normalisation of accurate (infinite) series does not change their  $\beta$  or their Hurst exponents. However, for finite values of  $k$ ,

$$\tilde{\beta}_k - \beta = -2 \frac{\log(a)}{\log(k+1)},$$

and

$$\tilde{\beta}_N - \beta = -2 \frac{\log(a)}{\log(N+1)}.$$

Therefore, normalisation distorts the Hurst exponent of finite series.

*Implications of time series normalisation on the experiments* In a few of the experiments, all fBm series were normalised to the same interval [1, 10]. As each series had different extremum values, each series was multiplied by a different constant. For example, as noted above, I normalised fBm series with  $H = 0.9$  by a factor  $a = 3$ . This normalisation distorted the Hurst exponent by approximately  $\frac{\log(3)}{\log(3001)} = 0.14$ . I normalised series with  $H = 0.1$  by a factor of 1.13. That distorted the Hurst exponent by  $\frac{\log(1.125)}{\log(3001)} = 0.015$ .

However, in experiments with normalised series, participants were asked to compare target graphs of similar Hurst exponents. The variance in normalisation constants for a given value of the Hurst exponent was small (the maximal difference was less than 0.2). Therefore the normalisation process had a negligible effect on the evaluation of participants' performance at a given Hurst exponent value. For example, for the extreme case of  $H=0.9$ ,

$$\frac{\log(3.2)}{\log(3001)} - \frac{\log(3)}{\log(3001)} = 0.0081 < 0.01.$$

For fGn series, the quotient of amplitudes of series corresponding to  $H = 0.1$  and  $H = 0.9$  is much higher than for fBm series, and can reach 100. Normalisation by a factor of order 100 would have resulted in a distortion of Hurst exponent by  $\frac{\log(100)}{\log(3001)} = 0.57$  for  $H = 0.9$ .

Furthermore, variance of normalisation constants for a given value of Hurst exponent is much higher for fGn series than for fBm series. For these reasons, I did not normalise fGn series in any of the experiments.

## **Part IV: General experimental remarks**

### **Choice of incentives across experiments**

A small number of principles guided my choice of incentives across experiments. I list these principles below.

1. Participants who were students at UCL were paid UCL's standard fees for participants in experiments (£1 per 10 minutes and at least £2).
2. Whenever I felt that additional incentive is required to motivate participants to make efforts, a prize for performance was advertised along with the standard fee.
3. As, theoretically, the number of participants in online experiments is unlimited, incentives offered in online experiments did not consist of a flat fee. Instead, I advertised a prize draw. The prize consisted of  $N/10$  USB sticks, where  $N$  was the number of participants required for the experiment. The advertisement stated clearly that  $N/10$  USB sticks will be given to  $N/10$  participants chosen randomly from the first  $N$  participants.

### **Outlier removal criteria**

Similarly, a small number of principles guided the choice of outlier removal procedure.

These principles are listed below.

1. As a default, any measurements more than two standard deviations larger or smaller than the groups' mean were removed.
2. In a few cases, application of the two standard deviation criterion resulted in a very large number of removed measurements. Such cases may indicate a non-linear relation between variables. To avoid removal of a large number of measurements, a few authors applied a natural logarithm on the results (Lin, Murphy and Shoben, 1997). Application of a natural logarithm on our results did not reduce sufficiently

outlier number when the two-standard deviation criterion was used. Therefore, instead of applying a natural logarithm on the results, I applied a stricter criterion, namely, used three or four standard deviations to define the outlier region.

## Chapter 2: Perception of fractal time series

This chapter explores the way people perceive graphically presented fractal time series. The study consisted of five experiments. It characterises people's sensitivity to fBm and fGn graphs. I examined the cues they used when performing identification tasks. Finally, I investigated people's ability to learn to identify the Hurst exponent, and the financial meaning they attributed to it.

### Experiment 1

The aim of Experiment 1 was to examine the following hypotheses:

$H_{1,1}$ : people's sensitivity to the Hurst exponent of graphically presented fBm series depends on the Hurst exponent.

$H_{1,2}$ : discriminability of the Hurst exponent of fBm series increases with the series length.

To achieve this, I presented participants with fractal task graphs. I manipulated the Hurst exponents of the series and their lengths (number of presented elements). In addition, I provided participants with example graphs which depicted graphs with different Hurst exponents. A measure,  $M$ , linearly dependant on the Hurst exponent of each example graph, was indicated. Participants were asked to estimate the  $M$  value of each of the task graphs using the example set.

#### *Method*

*Participants* Thirty-two undergraduates (17 men and 15 women) acted as participants. Their average age was 22.7 years. They were paid a fee of £6.00 per hour.

*Stimulus materials* I generated six sets of target graphs and four sets of example graphs, each with 33 different H values ranging from 0.1 to 0.9 in steps of 0.125. I then divided this range into four sub-ranges:  $0.1 \leq H \leq .275$ ;  $0.3 \leq H \leq 0.475$ ;  $0.5 \leq H \leq 0.675$ ;  $0.7 \leq H \leq 0.9$ . (I shall refer to these as sub-ranges H<sub>1</sub>, H<sub>2</sub>, H<sub>3</sub> and H<sub>4</sub>, respectively.) Finally, to provide target graphs for each participant, I randomly sampled two H values from each sub-range for each of six series lengths. This gave a total set of 48 different target series for each participant (two graphs x six lengths x four H sub-ranges). The same four example graphs for each of the 33 different values of the H exponent were available to all participants.

Series with 6284 points were generated with the spectral algorithm described by Saupe (Peitgen and Saupe, 1988). Details of this procedure are provided in the Chapter 1. Segments of the generated series were presented in lengths of 100, 250, 500, 750, 1000, and 1250 elements as the target series. Example graphs always included 1250 points. The graphs' point density was set to one point per pixel. Thus, I was able to specify the quality of the visual image and ensure it was the same for all participants. In order to avoid confounding of results with amplitude effects, vertical ranges of all graphs were normalised to the interval [1, 10]. This may have distorted the Hurst exponent of a given series significantly but it is unlikely to have distorted the *difference* between the H exponents of two different series, which had originally the same H value, by more than 0.01 (see Chapter 1, Part III).

*Design* During the familiarisation task, participants were presented with three randomly ordered graphs and shown how to use the graphical user interface. The experimental task followed immediately afterwards.

*Procedure* Participants were told that graphs differed in terms of a property, M, that could vary between zero and 100 (M was the H exponent multiplied by 100.) They had to inspect each of the target graphs carefully in order to estimate its M value. To assist them, they had access to a set of 132 example graphs that could be displayed one at a time by clicking on the appropriate button in the display. Figure 2.1 shows the graphic user interface. To select

examples for display, participants first scrolled down to the M value of their choice and then clicked on as many examples as they wished to see.

Participants were told that they could view the example graphs at any time by clicking on the appropriate button and that there was no limit to the number of times they could view any example. They were instructed as follows: ‘Please search the example list for graphs which resemble the target graph. Your estimation should be based on the “M” value of the graph groups that most resemble the target graph. Please estimate the “M” value of each of the graphs as a number between 0 and 100.’ They were also told that the “M” values of the target graph were not necessarily the same as the M values that appeared in the example table and that target graphs could have M values such as 23 or 97. Finally, they were alerted to the fact that the lengths of target graphs would vary and sometimes be short compared with the lengths of example graphs.

### *Results*

Participants’ estimates of the M value of target series were transformed into H estimates by dividing them by 100. One participant whose mean absolute error was more than two standard deviations greater than that of the average for the rest of the group was excluded from the analysis. I extracted both absolute and signed error scores for each combination of variables (Table 2.1). Signed error measures bias whereas absolute error is influenced both by bias and by response variability. As response variability can be interpreted as a reflection of task difficulty, absolute error is of primary interest here. However, I also analysed signed error as this can lead to additional insights into factors influencing discrimination. Mean values of both types of error score were low: for absolute error, 0.055; for signed error, 0.023

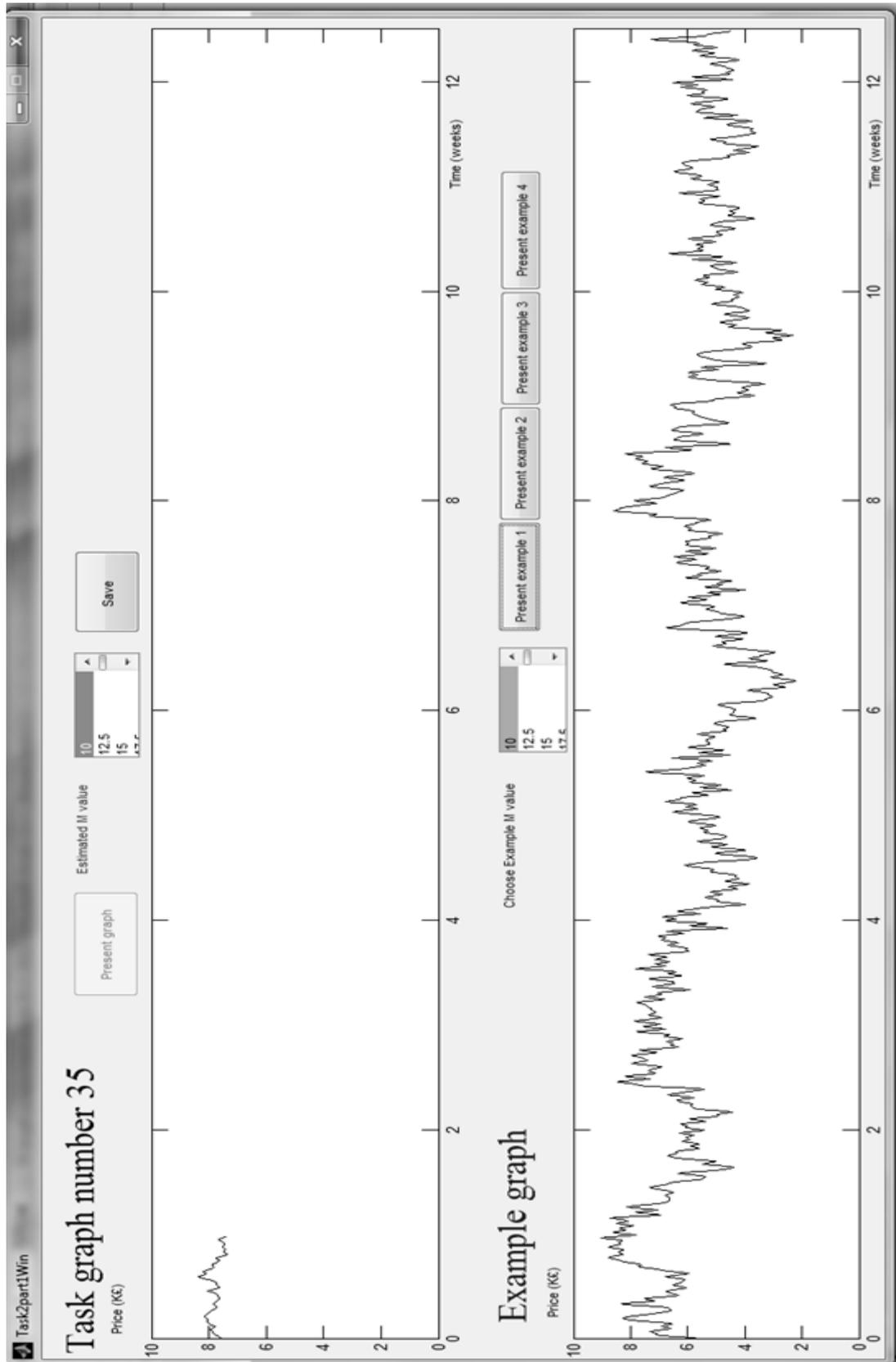


Figure 2.1 Experiment 1: Graphical user interface

Table 2.1 Experiment 1: Average values for absolute error (first panel) and signed error (second panel) for each combination of four ranges of Hurst coefficients, six different series lengths, and first and second instances. Standard deviations are denoted by parentheses.

Absolute error		Series length							Mean	Mean
H - range	Instance	100	250	500	750	1000	1250			
1	1	0.071 (0.070)	0.046 (0.040)	0.057 (0.043)	0.040 (0.044)	0.054 (0.058)	0.063 (0.052)	0.055 (0.052)	0.059	
	2	0.070 (0.062)	0.067 (0.051)	0.088 (0.112)	0.061 (0.051)	0.044 (0.042)	0.046 (0.051)	0.063 (0.067)	(0.060)	
2	1	0.067 (0.043)	0.052 (0.072)	0.082 (0.067)	0.048 (0.047)	0.069 (0.065)	0.058 (0.050)	0.063 (0.058)	0.061	
	2	0.061 (0.053)	0.067 (0.050)	0.057 (0.051)	0.051 (0.061)	0.066 (0.044)	0.054 (0.040)	0.059 (0.050)	(0.054)	
3	1	0.062 (0.048)	0.050 (0.044)	0.079 (0.048)	0.043 (0.043)	0.061 (0.083)	0.055 (0.052)	0.058 (0.055)	0.051	
	2	0.040 (0.033)	0.040 (0.034)	0.065 (0.064)	0.035 (0.038)	0.049 (0.045)	0.037 (0.034)	0.044 (0.043)	(0.050)	
4	1	0.057 (0.057)	0.044 (0.035)	0.042 (0.036)	0.048 (0.035)	0.048 (0.034)	0.062 (0.052)	0.050 (0.043)	0.050	
	2	0.061 (0.033)	0.048 (0.040)	0.055 (0.040)	0.044 (0.034)	0.042 (0.039)	0.047 (0.033)	0.049 (0.049)	(0.046)	
Mean		0.061 (0.059)	0.052 (0.047)	0.066 (0.063)	0.046 (0.046)	0.054 (0.053)	0.053 (0.047)		0.055 (0.053)	

Signed error		Series length							
H - range	Instance	100	250	500	750	1000	1250	Mean	Mean
1	1	0.039 (0.093)	0.036 (0.051)	0.051 (0.051)	0.033 (0.049)	0.046 (0.065)	0.060 (0.055)	0.044 (0.062)	0.048
	2	0.067 (0.066)	0.062 (0.057)	0.069 (0.125)	0.050 (0.063)	0.035 (0.051)	0.033 (0.060)	0.053 (0.075)	(0.069)
2	1	0.033 (0.073)	0.042 (0.078)	0.061 (0.087)	-0.007 (0.067)	0.002 (0.095)	0.037 (0.068)	0.028 (0.081)	0.030
	2								(0.076)
3	1	0.019 (0.077)	0.007 (0.067)	0.057 (0.074)	-0.014 (0.060)	0.020 (0.101)	0.029 (0.070)	0.020 (0.078)	0.013
	2	0.027 (0.044)	-0.017 (0.050)	0.015 (0.091)	0.004 (0.051)	0.022 (0.063)	-0.008 (0.050)	0.007 (0.062)	(0.070)
4	1	-0.035 (0.074)	0.015 (0.055)	0.002 (0.056)	-0.018 (0.058)	0.019 (0.056)	0.025 (0.078)	0.002 (0.066)	0.000
	2	-0.027 (0.103)	0.0185 (0.055)	0.007 (0.068)	-0.007 (0.060)	0.018 (0.052)	-0.013 (0.060)	-0.001 (0.069)	(0.068)
	Mean	0.020 (0.082)	0.027 (0.064)	0.034 (0.084)	0.007 (0.065)	0.025 (0.071)	0.024 (0.066)		0.023 (0.073)

*Absolute error scores* I carried out a three-way repeated-measures analysis of variance (ANOVA) on absolute error scores using three within-participant variables: the four H sub-ranges, the six series lengths, and the first and second instances of each combination of H sub-range and series length (Table 2.1). Here and elsewhere, I report effects with Greenhouse-Geisser corrections when Mauchly's test showed that the sphericity assumption was violated. There was a significant main effect of target H value ( $F(2.22, 66.50) = 3.59$ ;  $p = .03$ ;  $\eta^2 = .11$ ). Orthogonal contrasts showed that errors for H below 0.5 were significantly higher than those for errors for H above 0.5 ( $t(371) = 3.56$ ;  $p < .001$ ) but failed to show that errors for H between 0.5 and 0.675 were greater than those for H above 0.7 ( $t(371) = 0.44$ ; NS).

There was also an effect of series length ( $F(3.46, 103.65) = 4.24$ ;  $p = .005$ ;  $\eta^2 = .12$ ). Orthogonal contrasts showed that errors for shorter series (500 points or fewer) were higher than errors for longer ones ( $t(317) = 3.16$ ;  $p < .001$ ). However, the error depended weakly on series length: for series with 1250 elements, the mean error was 0.05 (std: 0.08), whereas for series length of 100 elements, the mean error was 0.06 (std: 0.06).

*Signed error scores* Signed error scores show that, overall, participants tended to overestimate the H values of the series. A three-way repeated-measures ANOVA on these scores, using the same variables as before, showed a main effect of target H value ( $F(2.34, 70.19) = 27.42$ ;  $p < .001$ ;  $\eta^2 = .48$ ). As Table 2.1 shows, estimates were too high when H was very low ( $H \leq 0.275$ ).

There was also a main effect of series length ( $F(3.60, 108.13) = 3.30$ ;  $p = .02$ ;  $\eta^2 = .10$ ) and an interaction between it and H value ( $F(8.78, 263.26) = 2.92$ ;  $p < .01$ ;  $\eta^2 = .09$ ). Whereas estimates for very low values of H remained too high as series length increased, estimates for other values of H became increasingly accurate. This improvement in accuracy with longer series can be partly attributed to practice: an interaction between series length and instance showed that, while the average decrease in mean overestimation of H values over

the session was small (.003), the decrease for the longest series (.027) was much higher ( $F(3.68, 110.24) = 4.54; p < .01; \eta^2 = .13$ ).

### *Discussion*

Participants were more sensitive to differences in series with  $H > 0.5$  than series with  $H < 0.5$ . This pattern of results replicates the one that Gildea et al (1993) reported for visuo-spatial contours in a new context (visual representation of time series). However, as  $H$  values increased within the range  $[0.5, 1]$ , there was no evidence that sensitivity either dropped off (Gildea et al, 1993) or increased further (Westheimer, 1991).

Sensitivity improved as the number of displayed points increased beyond 500. This implies that discrimination depended on extraction of some statistical feature from the series just as Gildea et al (1993) suggest. With more data points, values of that feature became a more reliable guide to discrimination. However, for a given series length, it was a less reliable guide for series that were negatively autocorrelated ( $H < 0.5$ ) than for those that were positively autocorrelated ( $H > 0.5$ ). I, therefore, accepted Hypothesis  $H_{1,1}$  and Hypothesis  $H_{1,2}$ .

## **Experiment 2**

The aim of Experiment 2 was to examine the following hypotheses:

Hypothesis  $H_{1,3}$ : people exhibit a higher degree of sensitivity to fGn graphs than to fBm graphs.

Hypothesis  $H_{1,4}$ : discriminability of the Hurst exponent of fGn sequences is higher when the series is longer.

Hypothesis  $H_{1,5}$ : change series derived from series with  $H$  values less than 0.5 are harder to discriminate than those derived from series with  $H$  values greater than 0.5.

The details of the task were similar to those of Experiment 1. However, in Experiment 2, I presented participants with series of price changes, rather than series of prices themselves.

### *Method*

*Participants* Thirty undergraduates (10 men and 20 women) acted as participants. Their average age was 24.6 years. They were paid a fee of £6.00 per hour.

*Stimulus materials* The target and example series were produced from series used in Experiment 1 by calculating the difference between successive values. The graphical user interface in this experiment was identical to the one used before (Figure 2.1) except that the vertical axes of graphs were labelled 'Price change (K£)' rather than 'Price (K£)'. As I was interested in testing Gildea et al's (1993) claim that the width of the distribution of increments (i.e. price changes) is the primary cue that participants use to discriminate H values, I did not normalise series in this experiment.

*Design and procedure* Both design and procedure were identical to those used for Experiment 1.

### *Results*

As before, participants' estimates of the M value of target series were transformed into H estimates by dividing them by 100. One participant whose mean absolute error was more than two standard deviations greater than that of the average for the rest of the group was excluded from the analysis. Again, I extracted both absolute error scores (mean = 0.037) and signed error scores (mean = 0.005) for each combination of variables (Table 2.2).

*Absolute error scores* To analyse absolute error scores, I carried out a three-way repeated measures ANOVA using the same three within-participant variables as before. Although the overall effect of H level was not significant, orthogonal contrasts showed that error for series with H less than 0.5 was significantly *lower* than that for series with H higher than 0.5 ( $t(347) = 2.73$ ;  $p < .01$ ). There was also a main effect of series length ( $F(3.87, 108.45) =$

2.94;  $p = .03$ ;  $\eta^2 = .10$ ): orthogonal contrasts showed that, as for Experiment 1, error scores for shorter series (500 points or fewer) were higher than those for longer ones ( $t(247) = 3.18$ ;  $p < .01$ ).

Table 2.2 Experiment 2: Average values for absolute error (first panel) and signed error (second panel) for each combination of four ranges of Hurst coefficients, six different series lengths, and first and second instances. Standard deviations are denoted by parentheses.

Absolute error		Series length							
H	Instance	100	250	500	750	1000	1250	Mean	Mean
range									
1	1	0.031	0.024	0.041	0.029	0.040	0.029	0.032	0.034
		(0.031)	(0.029)	(0.038)	(0.038)	(0.037)	(0.034)	(0.035)	
	2	0.041	0.033	0.041	0.050	0.024	0.021	0.035	(0.035)
		(0.040)	(0.032)	(0.037)	(0.047)	(0.027)	(0.023)	(0.036)	
2	1	0.030	0.041	0.043	0.035	0.032	0.022	0.034	0.032
		(0.030)	(0.036)	(0.040)	(0.029)	(0.030)	(0.025)	(0.032)	
	2	0.040	0.020	0.028	0.025	0.035	0.030	0.030	(0.032)
		(0.037)	(0.025)	(0.029)	(0.026)	(0.032)	(0.033)	(0.031)	
3	1	0.050	0.039	0.050	0.048	0.022	0.027	0.039	0.041
		(0.089)	(0.086)	(0.050)	(0.037)	(0.022)	(0.049)	(0.061)	
	2	0.040	0.036	0.061	0.035	0.024	0.054	0.042	(0.057)
		(0.040)	(0.036)	(0.068)	(0.022)	(0.028)	(0.087)	(0.053)	
4	1	0.050	0.035	0.037	0.034	0.040	0.032	0.038	0.037
		(0.043)	(0.025)	(0.033)	(0.027)	(0.034)	(0.031)	(0.033)	
	2	0.032	0.041	0.034	0.030	0.039	0.038	0.036	(0.034)
		(0.030)	(0.036)	(0.034)	(0.023)	(0.040)	(0.045)	(0.035)	
Mean		0.039	0.034	0.042	0.036	0.032	0.032	0.036	0.036
		(0.046)	(0.042)	(0.043)	(0.033)	(0.032)	(0.045)		(0.041)

Signed		Series length								
error	H	Instance	100	250	500	750	1000	1250	Mean	Mean
		range								
	1	1	0.016 (0.041)	0.017 (0.034)	0.023 (0.051)	0.017 (0.045)	0.016 (0.052)	0.017 (0.041)	0.018 (0.044)	0.022 (0.044)
		2	0.037 (0.044)	0.027 (0.038)	0.035 (0.044)	0.047 (0.050)	0.005 (0.036)	0.002 (0.031)	0.025 (0.044)	(0.044)
	2	1	0.008 (0.042)	0.022 (0.050)	-0.005 (0.059)	-0.002 (0.045)	0.004 (0.044)	0.007 (0.033)	0.006 (0.046)	0.001 (0.045)
		2	-0.009 (0.054)	0.001 (0.032)	0.003 (0.041)	-0.003 (0.036)	0.011 (0.047)	-0.023 (0.038)	-0.003 (0.043)	
	3	1	0.002 (0.103)	-0.001 (0.095)	0.028 (0.065)	-0.031 (0.053)	0.005 (0.031)	0.008 (0.055)	0.002 (0.073)	-0.005 (0.070)
		2	-0.002 (0.057)	0.002 (0.052)	-0.015 (0.091)	-0.019 (0.036)	-0.002 (0.037)	-0.032 (0.098)	-0.011 (0.067)	
	4	1	-0.028 (0.060)	0.003 (0.043)	0.006 (0.050)	0.003 (0.044)	0.019 (0.048)	0.008 (0.044)	0.0019 (0.050)	0.002 (0.050)
		2	-0.006 (0.044)	0.032 (0.044)	-0.013 (0.046)	0.001 (0.038)	0.013 (0.055)	-0.015 (0.057)	0.002 (0.050)	
		Mean	0.002 (0.061)	0.013 (0.053)	0.008 (0.060)	0.002 (0.048)	0.009 (0.044)	-0.004 (0.055)		0.049 (0.054)

*Signed error scores* Signed error scores show that, overall, participants had a slight tendency to overestimate H values of series. A repeated measures ANOVA on these scores, using the same three variables as before, showed a significant effect of target H value ( $F(2.90, 61.55) = 12.51$ ;  $p < .001$ ;  $\eta^2 = .31$ ): on average, participants overestimated H values below 0.5 by 0.01. An interaction between length and instance arose because this effect increased over the session – presumably as participants learned more about the range over which H values varied ( $F(2.50, 69.99) = 3.42$ ;  $p = .03$ ;  $\eta^2 = .11$ ). An interaction between H value and series length arose because the relatively high level of overestimation for the lowest H value obtained when series had fewer than 1000 points was much reduced for series when they had more than 1000 points, whereas signed error scores for series with higher H values was comparatively unaffected by series length ( $F(7.26, 203.31) = 2.63$ ;  $p = .01$ ;  $\eta^2 = .09$ ). Finally, as in Experiment 1, an interaction between series length and instance showed that, while the average decrease in mean overestimation of H values over the session was small (0.002), the decrease for the longest series (0.027) was much higher ( $F(3.58, 100.16) = 3.96$ ;  $p < .01$ ;  $\eta^2 = .12$ ).

*Cross-experiment comparison* In Experiment 1, mean absolute error score was .06 whereas here it was .04. This difference was significant ( $F(1, 28) = 39.83$ ;  $p < .001$ ;  $\eta^2 = .59$ ).

In Experiment 1, people were better at discriminating H values above 0.5 than at discriminating H values below 0.5. In this experiment, I changed the stimuli by presenting series of price changes or increments rather than the price series themselves. However, the target H values were exactly the same as before. This change had a clear effect on the pattern of discriminability: people were now *poorer rather than better* at discriminating H values above 0.5 than at discriminating H values below 0.5. To confirm the significance of this change, I carried out a four-way ANOVA using the same three within-participant variables as before but now also including Experiment as a between-participant variable. This showed a significant cross-over interaction between Experiment and target H value ( $F(2.64, 73.97) = 4.25$ ;  $p = .01$ ;  $\eta^2 = .13$ ). This effect is shown in Figure 2.2.

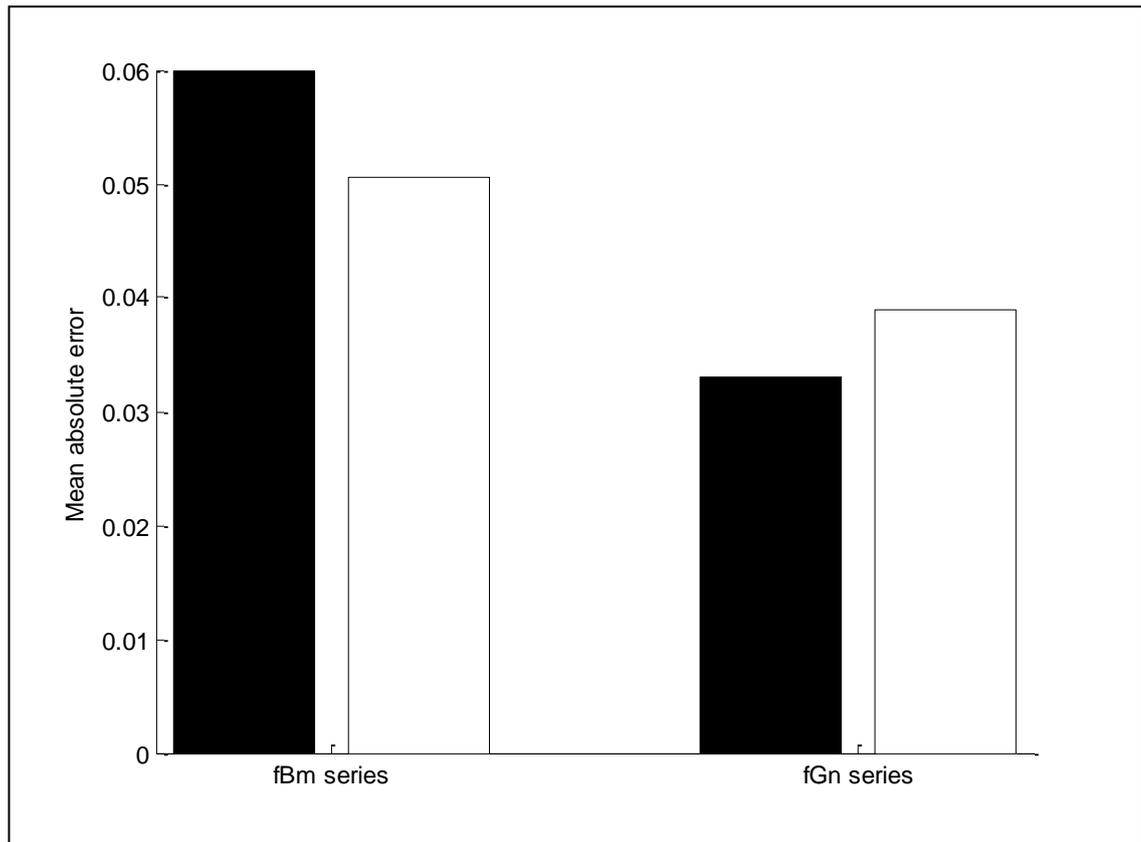


Figure 2.2 Bar graph showing mean absolute errors for  $H < .5$  (shaded) and  $H > .5$  (unshaded) for raw price series from Experiment 1 (left) and price change series from Experiment 2 (right).

### *Discussion*

As expected, discriminability of  $H$  values was better for price change series than for raw price series. This is consistent with Gilden et al's (1993) view that people extract information about the increments between successive points in order to discriminate fractal stimuli. By performing the increment extraction task for the participants, I removed one possible source of error. This made it easier for people to assess the amplitude of the apparent noise in the series and thereby discriminate series with different  $H$  values. I, therefore, accepted Hypothesis  $H_{1,3}$ . Furthermore, accuracy increased with series length. I accepted Hypothesis  $H_{1,4}$ .

In contrast to the previous experiment, discriminability was better with negatively autocorrelated series ( $H < 0.5$ ) than with positively autocorrelated ones. This is the opposite from what is implied by Gilden et al's (1993) argument. If extraction of price change information to use for discrimination between  $H$  values leads to better performance with positively autocorrelated series, then being presented with price change information to use for discrimination between  $H$  values should also lead better performance with positively autocorrelated series. I, therefore rejected Hypothesis  $H_{1,5}$ .

What could explain this unexpected reversal in the pattern of results? One possibility is that it is much harder to extract price change information from raw price series that are negatively autocorrelated. This seems unlikely: the individual price changes appear much larger and easier to identify in Figure 1.1 for lower  $H$  values. On the other hand, price change series in Figure 1.2 appear more distinct for lower  $H$  values: the difference in distribution widths is much larger between  $H = 0.1$  and  $H = 0.2$  than between  $H = 0.8$  and  $H = 0.9$ . Thus it is possible that participants used distribution widths to discriminate between  $H$  values for price change series but used some other feature to discriminate between  $H$  values for raw price series. In the following experiments, I explored these other perception cues could be.

### **Experiment 3**

Experiment 3 was designed to explore the effect of darkness, or brightness, of a fractal graph as a cue guiding the discrimination of Hurst exponents of fBm graphs. In particular, I was interested in Hypothesis  $H_{1,7}$ : people use graphs' illuminance as a cue assisting in discrimination of the Hurst exponents of fBm graphs.

In order to examine this, I manipulated the darkness of the example graphs that participants saw. This would be expected to change retinal illuminance without affecting the Hurst exponent of the graphs. Target graphs were always presented in the way that they had been in previous experiments but example graphs varied in terms of their darkness. Four

randomly ordered blocks of trials contained example graphs that were 1) darker than target graphs, 2) of the same darkness as target graphs, 3) somewhat lighter than target graphs, 4) considerably lighter than target graphs.

In line with Westheimer's (1991) argument, I expected absolute error to be higher when target and example graphs had different levels of darkness. However, my primary focus here is on signed error. If H values are discriminated on the basis of retinal illuminance, I would expect that using different levels of darkness for target and example graphs would bias H estimates. For example, making example graphs darker would make their H values appear to be smaller. As a result, a target correctly matched to an example graph with an H value of, say, 0.4 when target and example graphs are equally dark would be matched to an example graph with an H value that is greater than 0.4 when example graphs are darker than target graphs. Consequently, signed error would become more positive. Conversely, the same target graph would be matched to an example graph with an H value that is less than 0.4 when example graphs are less dark than target graphs. Consequently, signed error would become more negative.

### *Method*

*Participants* Thirty-three undergraduates (13 men and 20 women) with an average age of 25.5 years acted as participants. They were paid a flat fee of £3.00. In addition, they were (truthfully) told that the two individuals with the best results would receive an additional £10.

*Stimulus materials* The series were generated in the same way as they were in Experiment 1. Selection of H values for target and example graphs was also carried out in the same way as it was in that experiment. All target graphs were presented with a brightness of 0.2 on a grey scale that ranged from zero (black) to one (white). Example graphs were presented with a brightness of 0, 0.2, 0.4, or 0.6 on the same scale. Both target and example graphs had a constant thickness of one pixel. Figure 2.3 shows a typical task screen from the experiment.

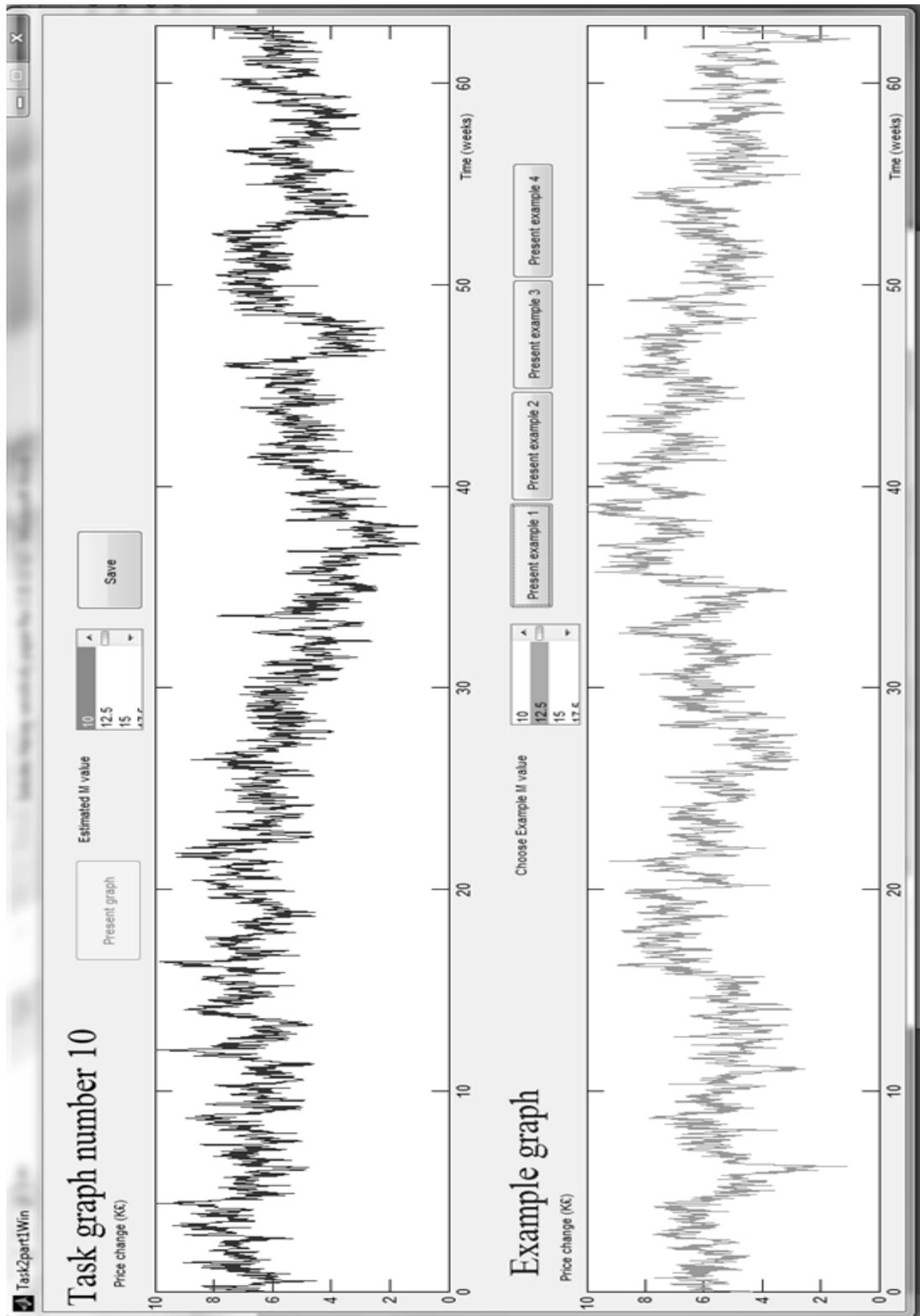


Figure 2.3 Graphical user interface for Experiment 3

*Design* After task familiarisation, which, as in previous experiments, involved practice with three graphs, participants were presented with 32 target graphs. These were divided into four blocks of eight graphs. In each of these blocks, example graphs had a different level of darkness. Order of presentation of blocks was determined randomly for each participant. Within each of the blocks, participants were presented with two instances of target graphs that had H values drawn from each of the four ranges of H values used in previous experiments. Ordering of trials within blocks was random.

*Procedure* Procedure was the same as in previous experiments except that, after familiarisation but before the experimental trials, participants were warned that example graphs would sometimes be presented with lines having a different darkness from those of the target graphs. They were explicitly told that “any such difference is not relevant to your task. Please ignore it and make your decision solely on the basis of the M values of the graphs.”

### *Results*

Participants’ estimates of the M value of target series were again transformed into H estimates by dividing them by 100. As before, participants whose mean absolute error scores were more than two standard deviations greater than that of the average for the rest of the group were excluded from the analysis. This reduced the size of the sample to 29 participants. I extracted both absolute error scores (mean = 0.045) and signed error scores (mean = 0.007) for each combination of variables in each condition (Table 2.3).

*Absolute error scores* To analyse absolute error scores, a three-way repeated measures ANOVA was performed using the same three within-participant variables as before. There was a main effect of the darkness of the example graphs ( $F(3, 84) = 6.34; p = .001; \eta^2 = .19$ ) and tests of linear contrasts showed that it arose because absolute error was lower when target and example graphs had the same darkness than when they did not ( $t(231) = 4.28; p < .001$ ).

In this experiment, the main effect of target H value that was obtained in previous experiments failed to attain significance. The absolute error scores for the highest target H value were inexplicably elevated for the middle two darkness levels: as a result, there was an interaction between target H level and darkness level ( $F(5.07, 141.95) = 2.47$ ;  $p = .04$ ;  $\eta^2 = .08$ ).

Table 2.3 Experiment 3: Average values for absolute error (first panel) and signed error (second panel) for each combination of Hurst coefficient range, darkness level, and instance for the darkness condition. Standard deviations are denoted by parentheses.

Absolute error	H						Mean	Mean
	range	Instance	0 (black)	0.2	0.4	0.6		
1	1	1	0.053	0.036	0.045	0.063	0.049	0.047
			(0.051)	(0.030)	(0.036)	(0.063)	(0.047)	
1	2	1	0.045	0.035	0.050	0.048	0.044	(0.043)
			(0.040)	(0.027)	(0.037)	(0.044)	(0.038)	
2	1	1	0.064	0.041	0.050	0.054	0.052	0.049
			(0.059)	(0.046)	(0.044)	(0.051)	(0.050)	
2	2	1	0.052	0.023	0.041	0.062	0.045	(0.045)
			(0.044)	(0.024)	(0.031)	(0.047)	(0.040)	
3	1	1	0.037	0.035	0.055	0.053	0.045	0.043
			(0.048)	(0.038)	(0.049)	(0.040)	(0.045)	
3	2	1	0.040	0.033	0.040	0.050	0.041	(0.041)
			(0.039)	(0.035)	(0.033)	(0.043)	(0.038)	
4	1	1	0.032	0.039	0.040	0.033	0.036	0.040
			(0.035)	(0.031)	(0.040)	(0.026)	(0.033)	
4	2	1	0.038	0.047	0.060	0.035	0.045	(0.036)
			(0.032)	(0.037)	(0.048)	(0.031)	(0.038)	
Mean			0.045	0.036	0.048	0.050	0.045	
			(0.045)	(0.034)	(0.040)	(0.045)	0.040	

Signed	H							
error	range	Instance	0 (black)	0.2	0.4	0.6	Mean	Mean
	1	1	0.028 (0.068)	0.012 (0.046)	0.022 (0.053)	0.039 (0.080)	0.025 (0.063)	0.025 (0.025)
		2	0.033 (0.051)	0.017 (0.041)	0.024 (0.058)	0.024 (0.061)	0.025 (0.053)	(0.058)
	2	1	0.053 (0.069)	-0.003 (0.063)	-0.003 (0.067)	0.008 (0.075)	0.014 (0.072)	0.014 (0.014)
		2	0.041 (0.054)	0.001 (0.034)	0.007 (0.052)	0.007 (0.078)	0.014 (0.058)	(0.065)
	3	1	0.015 (0.059)	-0.019 (0.048)	-0.038 (0.064)	-0.020 (0.064)	-0.016 (0.061)	-0.008 (-0.008)
		2	0.029 (0.047)	-0.005 (0.048)	-0.009 (0.051)	-0.016 (0.065)	-0.000 (0.056)	(0.059)
	4	1	0.015 (0.046)	0.016 (0.047)	-0.019 (0.054)	-0.012 (0.040)	0.000 (0.05)	-0.004 (-0.004)
		2	0.008 (0.049)	-0.021 (0.056)	-0.018 (0.075)	-0.001 (0.048)	-0.008 (0.059)	(0.054)
		Mean	0.028 (0.057)	-0.000 (0.050)	-0.004 (0.062)	0.004 (0.067)		0.007 (0.059)

*Signed error scores* Signed error scores were analysed in a similar manner. A main effect of target H value ( $F(2.32, 64.85) = 12.09$ ;  $p < .001$ ;  $\eta^2 = .30$ ) arose because range effects (Parducci, 1965) led to a response contraction bias (Poulton, 1989).

There was also a main effect of the darkness of the example graphs ( $F(3, 84) = 12.80$ ;  $p < .001$ ;  $\eta^2 = .31$ ). Tests of linear contrasts showed that it arose solely because overestimation of H values was greater when example graphs were darker than when they had the same characteristics as target graphs ( $t(231) = 5.73$ ;  $p < .001$ ). Thus, as predicted, signed error became more positive when example graphs were made darker than target graphs. However, in contrast to the predictions, there was no evidence that signed error became more negative when example graphs were made less dark than target graphs.

Finally, there was a marginally significant interaction between H value and darkness of example graphs ( $F(9, 252) = 2.20$ ;  $p = .04$ ;  $\eta^2 = .07$ ). This arose because the degree of overestimation that was obtained when example graphs were darker than target graphs was somewhat less for the highest and lowest ranges of H values than for the middle two.

Figure 2.4 shows main effects of darkness of example graphs on absolute error scores (upper panel) and signed error scores (lower panel).

### *Discussion*

Predictions focussed on signed error scores. Making example graphs darker than target graphs made signed error more positive in a manner consistent with Westheimer's (1991) argument that retinal illuminance can be used to discriminate between the H coefficients of different fractal contours. This result implies that retinal illuminance provides an important cue for discriminating between visual representations of fBm time series varying in terms of their Hurst coefficients. I, therefore, accepted Hypothesis  $H_{1,7}$ .

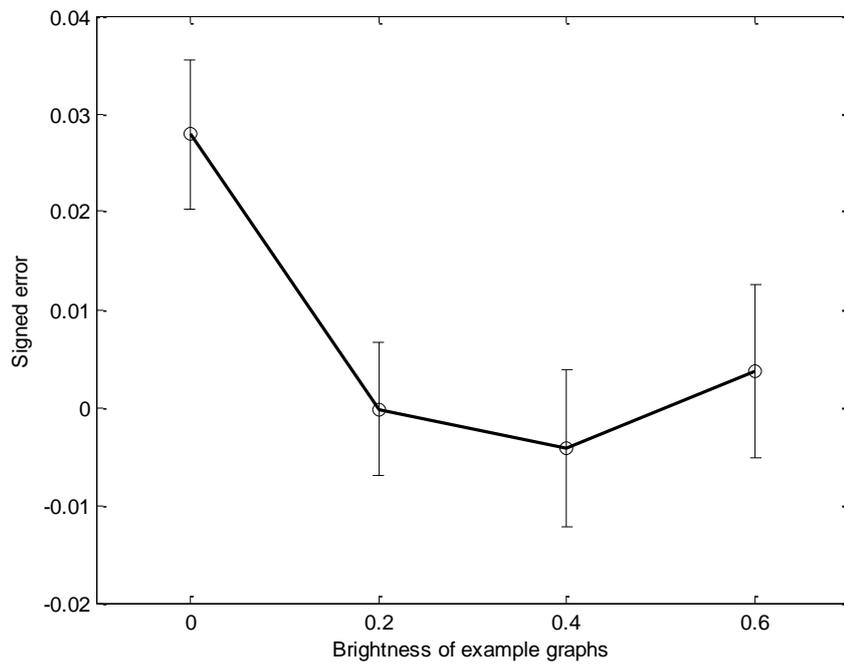
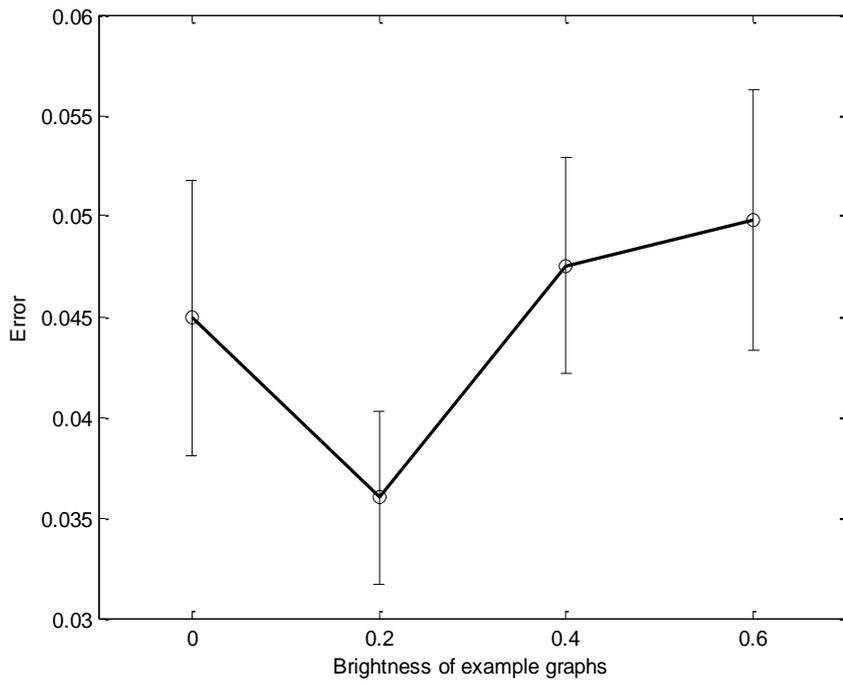


Figure 2.4 Experiment 3: Main effects of darkness of exemplar graph lines on absolute error scores (upper panel) and signed error scores (lower panel).

In contrast to what was expected, making example graphs brighter than target graphs did not make signed error more negative. It is clear that, at some level, the differences between the target graph shade of grey (0.2) and the other two shades of grey used for the example graphs (0.4, 0.6) had a psychological impact because they affected absolute error. So why did they not produce the expected effect on signed error? Perhaps the differences in retinal illuminance associated with them were insufficient to bias estimates of the Hurst exponent. In contrast, the difference in retinal illuminance between the black example graph and the darkest grey used for the target graphs was sufficient to have such an effect.

## **Experiment 4**

Experiment 4 was designed to examine Hypothesis  $H_{1,6}$ : the gradients of fractal graphs serve as a cue that assists discrimination of the Hurst exponents of the graphs.

I investigated the effect of smoothness on discriminability by manipulating the graphs' thickness. This masked fine fluctuations in the series by smoothing out differences between successive points. Therefore, making example graphs thicker should result in their perceived gradients being smaller. That, in turn, should cause their H values to seem too high. As a result, a target correctly matched to an example graph with an H value of, say, 0.4 when target and example graphs are depicted using lines that are equally thick would be matched to an example graph with an H value less than 0.4 when lines used to depict example graphs are thicker than those used to depict the target graph. Consequently, signed error should become more negative. Conversely, the same target graph would be matched to an example graph with an H value that is greater than 0.4 when example graphs are depicted using lines that are thinner than those used to depict the target graph. Consequently, signed error should become more positive.

Of course, making the lines of example graphs thicker would also have the same effect as making them darker: it would change their retinal illuminance. However, this effect is just the opposite of the one predicted by smoothing: if retinal illuminance is important, making

example graphs thicker should increase rather than decrease the H value of the example graph that is matched to the target graph. Obtaining the pattern of results predicted by retinal illuminance would not show that people do not use series autocorrelation as a cue: it would merely show that, under the experimental conditions, it is a relatively unimportant cue compared to retinal illuminance. On the other hand, obtaining the pattern of results predicted by use of series autocorrelation as a cue would show that it is relatively important compared to retinal illuminance.

### *Method*

*Participants* Thirty-five undergraduates (16 men and 19 women) with an average age of 26.8 years acted as participants. They were paid a flat fee of £3.00. In addition, they were (truthfully) told that the two individuals with the best results would receive an additional £10.

*Stimulus materials* The series were generated in the same way as they were in Experiment 1. Selection of H values for target and example graphs was also carried out in the same way as it was in that experiment. All target graphs were presented with a thickness of two pixels and example graphs were presented with a thickness of one, two, three, or four pixels. Both target and example graphs had a constant brightness of 0 (black) on the scale of brightness used in Experiment 3. Figure 2.5 shows a typical task screen from the experiment.

*Design* Design was identical to that used for Experiment 3 except that the four blocks of trials varied in terms of the thickness of the lines used to depict the example graphs rather than in terms of the brightness of those lines.

*Procedure* Procedure was the same as in previous experiments, except that participants were warned that example graphs would sometimes be presented with lines having a different thickness from those of the target graphs. They were told that “any such difference is not relevant to your task. Please ignore it and make your decision solely on the basis of the M values of the graphs.”

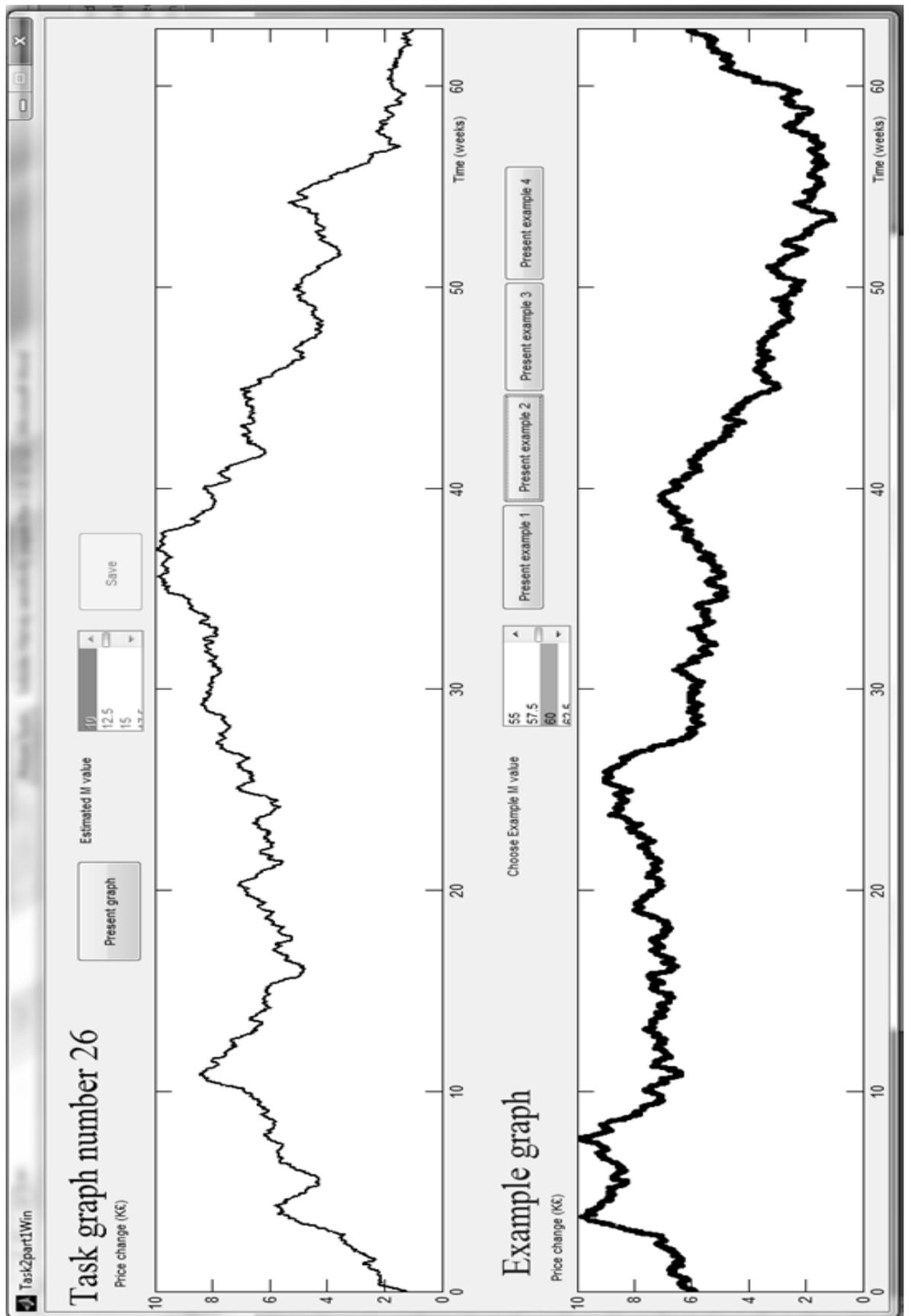


Figure 2.5 Graphical user interface for Experiment 4

## *Results*

Participants' estimates of the M value of target series were again transformed into H estimates by dividing them by 100. As before, participants whose mean absolute error scores were more than two standard deviations greater than that of the average for the rest of the group were excluded from the analysis. This reduced the size of the sample to 30 participants. Both absolute error scores (mean = .067) and signed error scores (mean = .009) for each combination of variables in each condition were extracted (Table 2.4).

*Absolute error scores* A three-way repeated measures ANOVA using the same three within-participant variables as before showed that there was a main effect of the thickness of the example graphs ( $F(3, 84) = 8.15; p < .001; \eta^2 = .23$ ). Tests of linear contrasts showed that it arose because absolute error was lower when target and example graphs had the same thickness than when they did not ( $t(239) = 5.86; p < .001$ ).

There was also a main effect of target H value ( $F(2.29, 64.13) = 10.32; p < .001; \eta^2 = .27$ ). As in Experiment 1, absolute error was lower for positively autocorrelated series ( $H > 0.5$ ) than for negatively autocorrelated ones ( $t(239) = 2.99; p < .05$ ).

*Signed error scores* A main effect of target H value ( $F(3, 84) = 11.04; p < .001; \eta^2 = .28$ ) arose because range effects (Parducci, 1965) led to a response contraction bias (Poulton, 1989).

There was a main effect of the thickness of the example graphs ( $F(3, 84) = 13.93; p < .001; \eta^2 = .33$ ). Tests of linear contrasts showed that it arose solely because overestimation of H values was greater when example graphs were not as thick as target graphs than when example and target graphs were of the same thickness ( $t(239) = 6.39; p < .001$ ). Thus, as predicted by the argument that people use series autocorrelation as a cue, signed error became more positive when example graphs were made less thick than target graphs. However, contrary to predictions, signed error did not become more negative when example graphs were made thicker than target graphs.

Figure 2.6 shows main effects of line thickness of example graphs on absolute error scores (upper panel) and signed error scores (lower panel).

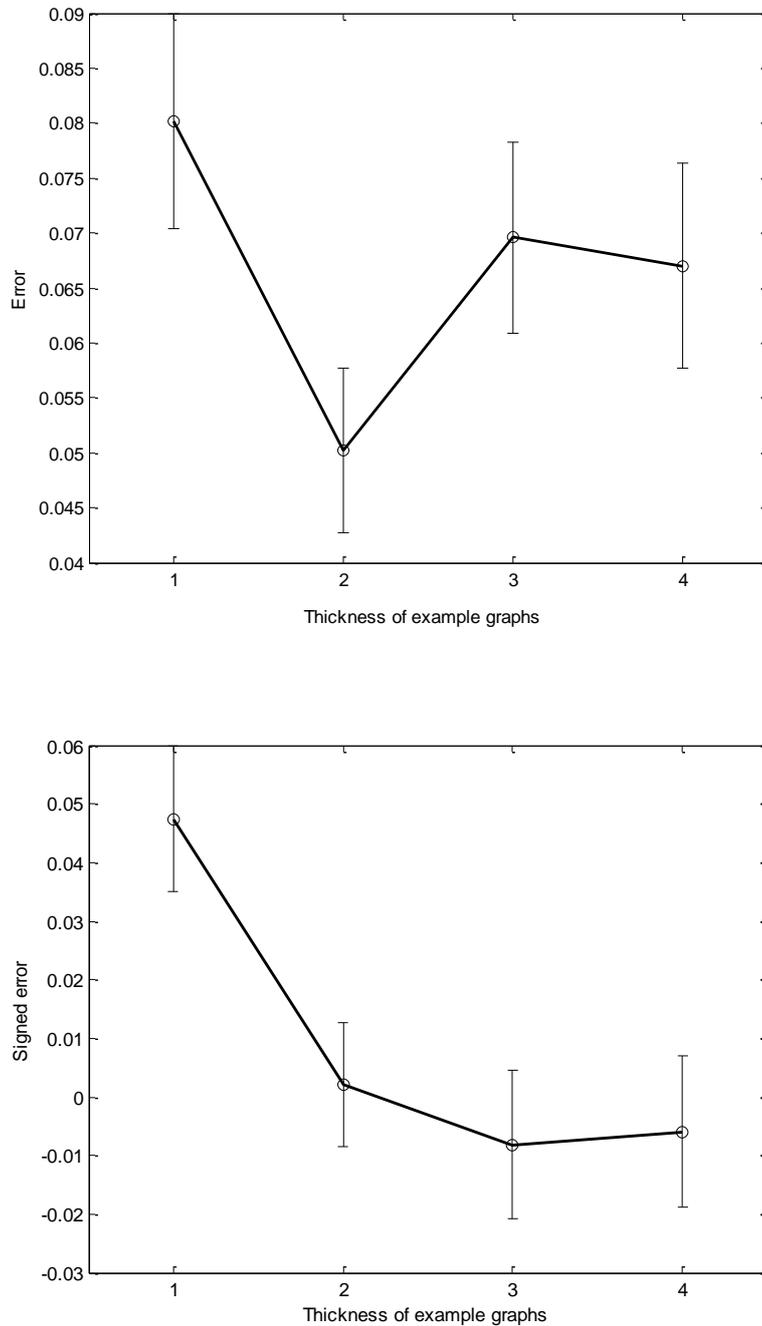


Figure 2.6 Experiment 4: Main effects of thickness of exemplar graph lines on absolute error scores (upper panel) and signed error scores (lower panel)

Table 2.4 Experiment 4: Average values for absolute error (first panel) and signed error (second panel) for each combination of Hurst coefficient range, thickness level, and instance for the thickness condition. Standard deviations are denoted by parentheses.

Absolute error	H						Mean	Mean
	range	Instance	1	2	3	4		
	1	1	0.073	0.051	0.071	0.050	0.061	0.064 (0.057)
			(0.063)	(0.056)	(0.057)	(0.052)	(0.057)	
		2	0.098	0.043	0.061	0.065	0.067	
			(0.073)	(0.046)	(0.050)	(0.053)	(0.059)	
	2	1	0.117	0.054	0.086	0.075	0.083	0.081 (0.080)
			(0.078)	(0.053)	(0.084)	(0.091)	(0.080)	
		2	0.107	0.054	0.073	0.080	0.079	
			(0.096)	(0.036)	(0.068)	(0.066)	(0.071)	
	3	1	0.076	0.048	0.080	0.093	0.074	0.072 (0.066)
			(0.060)	(0.040)	(0.059)	(0.090)	(0.066)	
		2	0.078	0.055	0.072	0.074	0.070	
			(0.072)	(0.054)	(0.081)	(0.062)	(0.068)	
4	1	0.051	0.046	0.072	0.053	0.055	0.051 (0.054)	
		(0.038)	(0.054)	(0.061)	(0.059)	(0.054)		
	2	0.043	0.052	0.043	0.046	0.046		
		(0.054)	(0.061)	(0.038)	(0.035)	(0.048)		
Mean		0.080	0.050	0.070	0.067	0.067	(0.063)	

Signed	H							
error	range	Instance	1	2	3	4	Mean	Mean
1	1	1	0.058	0.024	0.039	0.032	0.038	0.040
			(0.078)	(0.072)	(0.082)	(0.065)	(0.075)	
	2	1	0.071	0.023	0.033	0.038	0.041	(0.077)
			(0.100)	(0.059)	(0.072)	(0.075)	(0.079)	
2	1	1	0.070	0.001	-0.011	0.007	0.017	0.013
			(0.120)	(0.076)	(0.120)	(0.119)	(0.114)	
	2	1	0.047	0.013	-0.017	-0.007	0.009	(0.110)
			(0.14)	(0.065)	(0.099)	(0.104)	(0.106)	
3	1	1	0.044	-0.008	-0.052	-0.055	-0.018	-0.005
			(0.087)	(0.062)	(0.086)	(0.118)	(0.098)	
	2	1	0.053	0.010	-0.003	-0.026	0.008	(0.098)
			(0.092)	(0.077)	(0.109)	(0.094)	(0.097)	
4	1	1	0.018	-0.024	-0.047	-0.034	-0.022	-0.013
			(0.062)	(0.067)	(0.083)	(0.072)	(0.074)	
	2	1	0.020	-0.022	-0.008	-0.003	-0.003	(0.071)
			(0.067)	(0.077)	(0.057)	(0.059)	(0.066)	
Mean			0.047	0.002	-0.008	-0.006		0.009
			(0.097)	(0.071)	(0.094)	(0.095)		(0.089)

## *Discussion*

Absolute error scores showed an analogous pattern to the one found in the previous experiment. They were higher when the thickness of the example graphs was different from the thickness of the target graphs. This pattern is again what was expected on the basis of previous work (Egeth, 1966; Ballesteros, 1996; Watanabe, 1988; Williams, 1974) and is likely, at least in part, to reflect the fact that the absolute size of the biases revealed by the analysis of signed error (discussed next) was greater when example and target graphs were of different thicknesses.

Analysis of signed error showed that making the example graphs thinner than the target graphs produced a bias in the direction to be expected if this manipulation reduced the gradients of the series by masking differences between successive points. This bias was in the opposite direction to that expected on the basis of changes in retinal illuminance. Thus I accepted Hypothesis H<sub>1,6</sub>.

However, making example graphs thicker than target graphs did not have either the effect predicted by masking of gradients or the opposite effect by changes in retinal illuminance. One possibility is that participants used both cues and that their effects on signed error cancelled one another out.

Taken together, results of Experiments 3 and 4 imply that people use more than one cue to discriminate between graphs of fBm series. The present experiment implies that people are sensitive to the Hurst exponent of time series. The previous experiment showed that they also use retinal illuminance to discriminate between such series. However, when these two cues were pitted against one another in the way that they were in the present experiment, the effects of the gradient cue may dominate those of the retinal illuminance cue (example graphs thinner than target graphs) or the effects of the two cues may cancel each other out (example graphs thicker than target graphs).

## Experiment 5

Experiment 5 was designed to explore the following hypotheses:

Hypothesis H<sub>1,8</sub>: people can learn to identify the Hurst exponents of given graphs.

Hypothesis H<sub>1,9</sub>: people perceive investments in assets that have price graphs with a low Hurst exponent to be riskier than investments in assets that have price graphs with a high Hurst exponent.

Participants were presented with a sequence of 96 time series. They were asked to identify a measure that was linearly dependent on the Hurst exponent of each graph. In order to facilitate learning during the learning stages, they were given feedback that included the correct value of this measure. Each learning stage was followed by a test, in which no feedback was given. In contrast to Experiments 1 - 4, no example graphs were presented to the participants: learning was based only on feedback. At the end of the experiment, participants were asked to answer a questionnaire that included a question about the risk level of investment in an asset that had a price series with a Hurst exponent higher or lower than 0.5.

### *Method*

*Participants* Thirty-five undergraduates (13 men and 22 women) acted as participants. Their average age was 22.9 years. They were paid a fee of £3.00. In addition, two prizes of £10.00 each were awarded to the two participants whose average error was smallest. The prize was advertised in the advertisement for the experiment and was mentioned in the instructions.

*Stimulus materials* I generated six sets of fBm graphs each with 32 different H values ranging from 0.1 to 0.875 in steps of 0.025. This range was then divided into eight sub-ranges:  $0.1 \leq H \leq 0.175$ ;  $0.2 \leq H \leq 0.275$ ;...;  $0.8 \leq H \leq 0.875$ .

*Design* Each participant was presented with 96 graphs, which were separated into two main stages, each comprising 48 graphs. Each stage included 5 learning sub-stages and a test stage, each consisting of eight graphs. At each sub-stage, graphs were randomly chosen from the six possible sets in each H-range. Presentation order of graphs in each sub-stage was random. All graphs were presented using a Matlab code. The graphs were not normalised. The task window of the programme is shown in Figure 2.7.

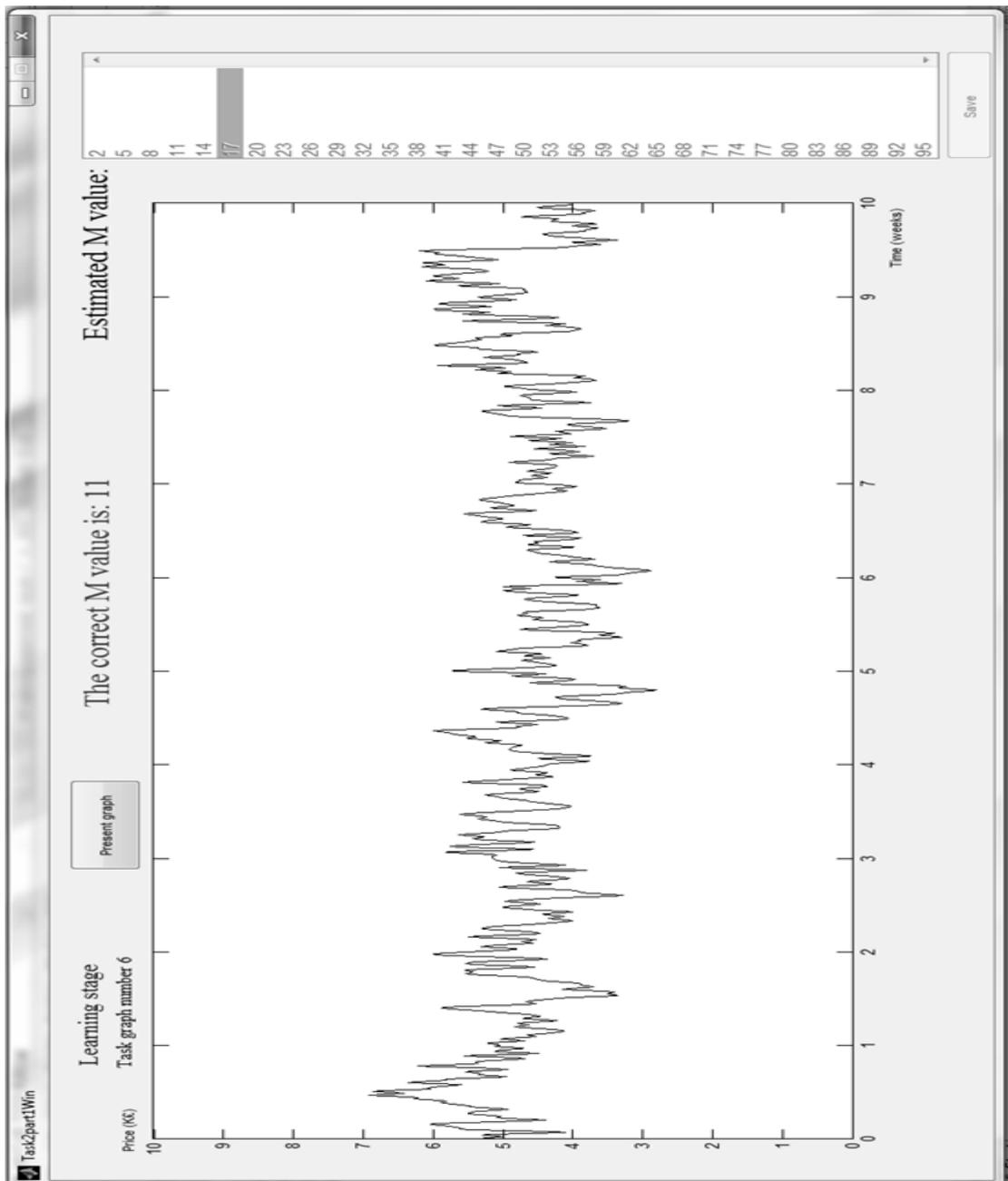


Figure 2.7 The task window of Experiment 5

Presentation of each graph in the learning sub-stages was followed by immediate feedback. Feedback referred to a variable denoted by “M”, defined by  $M = 3 * ((H - 0.1) / 0.025 + 1) - 1$ . This transformation was chosen in order to ensure that all M values were integers. In addition, the range of M was 2 to 95 and, therefore, close to the natural range of percentages. Furthermore,  $M(0.5) = 50$ , which enabled natural formulation of questions about the differences between the risk level of investment in assets whose price series have  $M < 50$  or  $M > 50$ .

During the test sub-stages no feedback was given.

*Procedure* Participants were asked to look at each graph of the 96 presented graphs, estimate its M value by choosing a value from a given list of values between 2 and 95, and save their selection. After completing this task, participants were asked to complete question list.

The experiment instructions were:

“In the following task, you will be presented with a sequence of 96 graphs. The graphs differ by a property called “M”. M values of presented graphs will range between 1 and 96.

You will be asked:

1. to look at the graphs carefully,
2. to estimate the value of the “M” property of the graphs as a number between 1 and 96.
3. to enter your estimation and then save it. [...]

In order to complete the task, the experiment includes learning stages, in which you will get feedback on your estimates. The feedback includes the M value. [...]

Initially, you will not have any idea of the correct M value. So you need to use the feedback that you will get after each graph to understand what is meant by the M value so that you can make better estimates in the future.”

The questions participants were asked are listed in Appendix A.

### *Results*

Participants whose mean absolute error scores were more than two standard deviations greater than that of the average for the rest of the group were excluded from the analysis. This reduced the size of the sample to 33 participants. Absolute error scores and signed error scores for each participant at each of the experiment stages were extracted. The answers to the questionnaire were also analysed.

*Absolute error scores* Over all, the mean value of participants absolute error was 0.079 (min = 0.051, max = 0.122, std = 0.022). A two-way repeated measures ANOVA using the variables experiment stage and experiment sub-stage showed that there was a main effect of the stage of the experiment ( $F(1, 32) = 34.26; p < .001$ ) and sub-stage ( $F(4, 128) = 26.71; p < .001$ ). There was also a significant interaction effect between stage and sub-stage ( $F(4, 128) = 17.19; p < .001$ ).

Paired-t-tests revealed significant differences between participants' errors in sub-stages 1 and 5 of the first test stage ( $t(32) = 6.56; p < 0.001$ ), sub-stage 1 of the first stage and test 1 ( $t(32) = 8.02; p < 0.001$ ) and sub-stage 1 of stage 2 and test 2 ( $t(32) = 2.97; p = .006$ ).

There were no significant differences between sub-stages 1 and 5 of stage 2, indicating that there was no significant improvement of performance during the second stage ( $t(32) = 1.18, p = .25$ ). There were no significant differences between performance in the fifth sub-stage and test stage in any of the experimental stages. This indicates that feedback did not affect results as an incentive. Dependence of mean absolute error on trial number is shown in Figure 2.8. As participants' errors do not seem to converge to zero, a regression with respect to the model Mean error =  $ae^{bt} + \text{error}$  yielded a relatively small  $R^2$  value ( $a = 0.11; b = -0.008; p < .01; R^2 = .41$ ). Translating the mean error by subtracting from it its minimum value did not improve  $R^2$  significantly. However, regression with respect to the model Mean error =  $a + b/t + \text{error}$  yielded  $a = 0.06; b = 0.28; p < .01; R^2 = .85$ . Therefore, although

learning error is usually modelled by an exponent (Castro, Kalish, Nowak, Qian, Rogers and Zhu, 2008), in this case, a model for the mean error, which predicts that the error is inverse-proportional to the time, fits the results better than an exponential model.

*Signed error scores* Apart from sub-stage 1 of stage 1, all mean signed errors were insignificantly different than 0.

Table 2.5 shows participants' mean errors and signed errors in all sub-stages of stages 1 and 2 and the test stages.

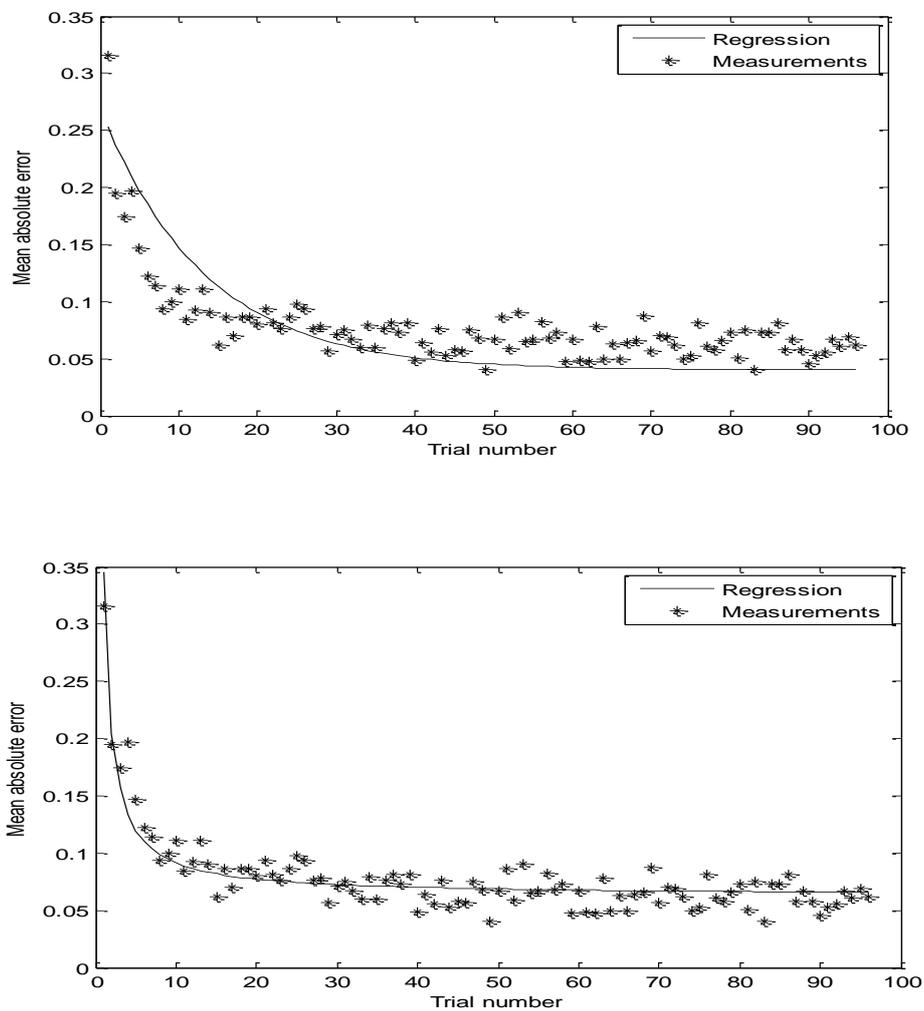


Figure 2.8 Absolute error versus trial number in Experiment 5. Exponential regression line is presented in the upper panel, and the regression line of the model Mean absolute error= $a/\text{trial number}+b+e$  is presented in the lower panel.

Table 2.5 Absolute and signed errors in Experiment 5.

Measure	Sub- stage	1	2	3	4	5	Test
Error	Stage1	0.170	0.095	0.085	0.080	0.072	0.065
		(0.074)	(0.053)	(0.047)	(0.049)	(0.036)	(0.027)
	Stage 2	0.072	0.061	0.066	0.061	0.065	0.059
		(0.025)	(0.023)	(0.026)	(0.025)	(0.035)	(0.023)
Signed error	Stage 1	-0.280	0.011	0.015	0.012	0.003	-0.011
		(0.067)	(0.056)	(0.046)	(0.043)	(0.050)	(0.046)
	Stage 2	-0.001	-0.001	-0.007	0.004	0.008	0.000
		(0.043)	(0.027)	(0.037)	(0.039)	(0.041)	(0.027)

*Analysis of answers to the questionnaire* Answers to questions revealed that, on average, participants did not consider graphs with  $H < 0.5$  more difficult to identify than graphs with  $H > 0.5$  (16/33 = 49% of the participants chose the former and 17/33 = 51% chose the latter). However, the vast majority of the participants (28/33 = 85%) identified assets with Hurst exponents that were smaller than 0.5 as riskier to invest in. Accordingly, most participants answered that they would prefer investing money in assets whose Hurst exponent was higher than  $H = 0.5$  (25/33 = 76%). Interestingly, many of those who said that they would prefer investing in assets with  $H > 0.5$  rationalised their preference by using arguments such as: “Price is stable”, “Greater stability and predictability”, “Less fluctuation, lower risk”, “If I make a loss, it would be a small loss”, and “Safer”, whereas participants who preferred investing in assets with  $H < 0.5$  used arguments as: “More chances that the asset will go up. Buy low and sell high”, “Price changes frequently and I will get a good deal”. Therefore, answers reflected mainly personal risk-taking preferences rather than any difference in the

perception of risk level of the assets. Indeed, the features that participants in both groups typically used to distinguish graphs with high and low Hurst exponents were: “Degree of fluctuations”, “Smoothness”, “Overall height of the graphs”, “Overall trend”, and “Shape”.

### *Discussion*

Experiment 5 showed that, given merely feedback, people can learn to identify the Hurst exponent of time series with some accuracy. Furthermore, they do not exhibit any significant bias, and their standard deviation is small. Importantly, people attribute to different H-ranges ( $H < 0.5$ ,  $H > 0.5$ ) a financial meaning: assets that had price graphs with a Hurst exponent lower than 0.5 are considered riskier to invest in than those with a Hurst exponent higher than 0.5. These results affected participants’ investment preferences. I accepted Hypotheses  $H_{1,8}$  and  $H_{1,9}$ .

### **Conclusions**

The study of randomness of binary sequences has many psychological and educational applications. For instance, Falk and Konold (1997, page 301) wrote: “Judging a situation as more or less random is often the key to important cognitions and behaviours. Perceiving a situation as nonchance calls for explanations [...] Lawful environments encourage a coping orientation [...] In contrast, there seems to be no point in patterning our behaviour in a random environment.” However, in real-life, people have to deal many times with time series describing threatening events: for instance, traders have to react to price swings (Mandelbrot and Hudson, 2004). The threat encapsulated in financial series is termed ‘risk’ rather than ‘randomness’.

Price series are rich in detail and can behave very unpredictably. To be able to understand their behaviour, graphical representations are used. Previous studies on human perception of fractal time series have suggested that people use the gradients and illuminance of series to assess the Hurst exponent of graphically presented time series. The experiments reported

here confirmed these suggestions: using these cues enabled people to reach a high level of accuracy in the discrimination and identification of the Hurst exponent of different series. However, the results indicated that biases arise from the use of these same cues: the darkness and the thickness of the lines with which the graph is presented may affect perception of the Hurst exponent. The results also show that people can learn to identify the Hurst exponent of graphs and suggest that, in financial contexts, the meaning that they attribute to it is related to risk.

### *Limitations*

The conditions of Experiment 1 and Experiment 2 were not identical: in Experiment 1, I normalised all presented graphs, whereas in Experiment 2, I did not. The main consideration for normalising fBm series in Experiment 1 was to eliminate amplitude cues. The main consideration against normalising fGn series in Experiment 2 was to avoid a large distortion of their Hurst exponents (normalisation of fGn series with  $H$  in the domain  $[0.1, 0.9]$  to the same interval results in larger distortions in the Hurst exponent than the distortion caused to the Hurst exponents of fBm series by normalisation). However, that difference suggests caution if I am to generalise the results of the comparison between participants' performances in Experiment 1 and 2 beyond the conditions of the experiments.

## Chapter 3: Risk perception and financial decisions

This chapter explores the way people assess risk and make financial decisions when presented with graphs of financial time series. The study described in this chapter consisted of a series of four experiments.

Experiment 5 in Chapter 2 revealed that participants related the Hurst exponent of the time series with risk of investment in the corresponding asset. However, that experiment gave only a rough estimate for the dependence of risk assessment on the Hurst exponent. The research reported in this chapter was designed to develop greater understanding of the way people assess the risk of investments, based on their price graphs.

### Experiment 1

Experiment 1 was designed to explore the following hypotheses:

Hypothesis  $H_{2,1}$ : when no additional cues are presented, risk perception of investment in assets, based on their price graphs, depends weakly on the Hurst exponent of the price series.

Hypothesis  $H_{2,2}$ : when both price series (fBm) and its corresponding price change series (fGn) are presented, risk assessments are negatively correlated with the Hurst exponent of price series.

Hypothesis  $H_{2,3}$ : risk ratings of people who are low on emotional stability are correlated with the Hurst exponent of the presented graphs stronger than those of people who are high on emotional stability.

To examine these hypotheses, I presented participants with pairs of graphs of computer-generated fractal series. I manipulated graph presentation format. In one condition,

participants were presented with fBm series, whereas in the second condition, they were presented with fBm series as well as their corresponding fGn series. They were told that the fBm graphs represented asset prices. fGn series were presented as the corresponding price change series. The difference in the Hurst exponents between the graphs in each pair was manipulated. Participants were asked to compare risk or randomness levels of graphs. They completed a personality questionnaire at the end of the experiment.

### *Method*

*Design* All the experiments in this study were performed on the internet. Online experiments are recommended as they reduce experimenter effects and volunteer bias while increasing access to demographically and culturally diverse participant groups (Reips, 2002). In addition, they have similar internal and external validity as those of laboratory or field experiments (Horton, Rand and Zeckhauser, 2011).

Two sets of 50 fBm graph pairs were randomly chosen for each participant. In Condition fBm, only fBm graphs were presented. In Condition fBm&fGn, fBm graphs were presented along with their corresponding fGn graphs. The graphs were presented using a graphic user interface program written in Matlab. Figure 3.1 shows a typical task windows from Condition fBm and from Condition fBm&fGn.

Participants were asked to discriminate between the risk levels of investments in asset pairs in one of the graph sets (risk-discrimination task) and to discriminate between the randomness levels of the behavior of each of the graphs in pairs in the other set (randomness-discrimination task). The order of the tasks was randomly chosen for each participant. The randomness task served as a control, verifying whether participants could discriminate between graphs with different Hurst exponents.

The Hurst exponents of the graphs in each pair were different. I denote the differences between the Hurst exponents of the graphs in each pair by  $\Delta H$ . Each set of fifty graph pairs included 15 pairs with  $\Delta H = 0.1$ , 15 pairs with  $\Delta H = 0.05$ , and 20 pairs with  $\Delta H = 0.025$ .

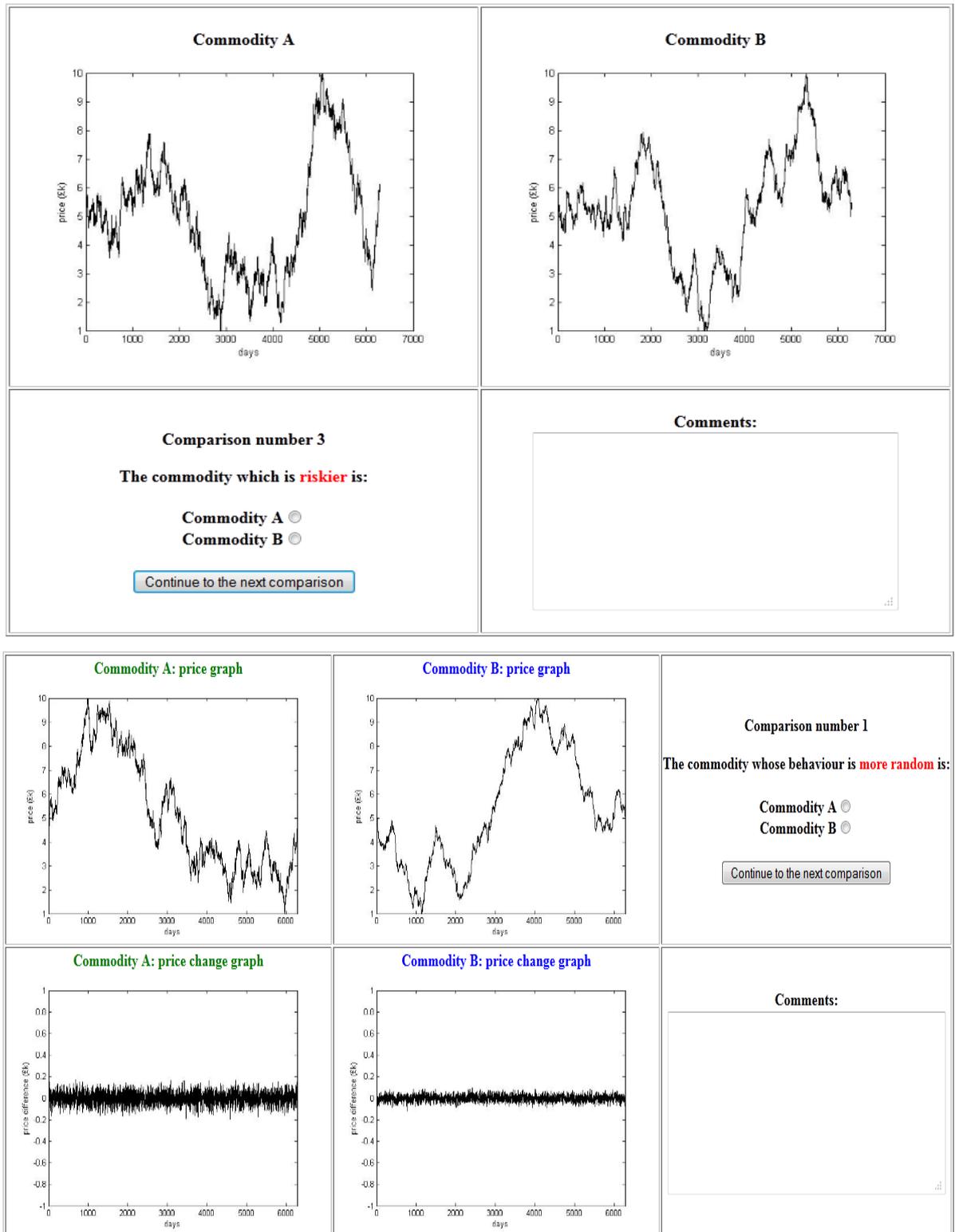


Figure 3.1 Task windows from Experiment 1: Risk rating task in fBm condition (upper panel) and randomness rating task in fBm&fGn condition (lower panel).

In Chapter 2, I showed that most people can distinguish between the Hurst exponents of graphs when  $\Delta H = 0.1$  and  $\Delta H = 0.05$ , but that, when  $\Delta H = 0.025$ , accuracy is lower. The order of presentation of the graphs in each pair on the screen was randomized.

These manipulations resulted in a two (fBm or fBm&fGn condition) by two (risk or randomness discrimination task) by three ( $\Delta H = 0.1, 0.05, 0.025$ ) design.

*Participants* I was interested in answers of both experts and non-experts. Muradoglu and Harvey (2012) and Barber and Odean (2008) noted that a large number of lay people have started to trade online over the past few years because of increased access to internet trading sites.

Experiment 1 was advertised on financial analyst and economist groups on LinkedIn. A prize draw was announced in order to encourage participation. The prize consisted of three memory sticks.

Over a period of one month, 77 people participated in Condition fBm. The answers of 41 people who completed all tasks (21 men and 20 women, average age: 45.3) were included in the analysis. All participants but one had academic degrees or were students. Twelve participants had a PhD, nine had an MSc, 14 had a BSc/BA, and five were students.

Over a period of one month, 81 people participated in Condition fBm&fGn. 47 people (16 women, 31 men, average age: 46.1) completed all tasks. Apart from three of them, all participants had academic degrees or were students. Four participants had a PhD, 19 had an MSc, and 21 had a BA/BSc.

Participants included people from Australia, New Zealand, Malaysia, India, Philippines, Canada, USA, Argentina, UK, the Netherlands, Norway, France, Luxembourg, Italy, Greece, Israel, Poland, and Ukraine.

Participants were asked whether they were financial analysts. In the fBm condition, seven participants answered positively. In the fBm&fGn condition, ten answered positively.

*Materials* Stimuli consisted of 54 (9 x 6) fBm graphs with Hurst coefficients  $H = 0.1, 0.2, 0.3, \dots, 0.9$ , 54 (9 x 6) fBm graphs with Hurst coefficients  $H = 0.35, 0.4, 0.45, \dots, 0.75$ , 54 (9 x 6) fBm graphs with Hurst coefficients  $H = 0.4, 0.425, 0.45, \dots, 0.6$ , and their corresponding fGn graphs.

All fBm series and their corresponding fGn series were produced in Matlab as described in Chapter 1, Part III. To avoid confounding of results with the difference between the first and last data points, all graphs depicted one period of the produced fractals. Therefore, the first and last point in each of the graphs was identical. Similarly, to avoid confounding of results with the graphs' ranges, I normalised all graphs to have the same range (the interval [1, 10]). Normalisation of graph pairs for which the Hurst exponent differs by not more than 0.1 changes only slightly the differences between their Hurst exponents (see Chapter 1, Part III).

Each series consisted of 6284 points. The graphs were saved in jpg format. These jpg images were presented over a third of a 15-inch computer screen with 1366 x 768 pixels. I, therefore, estimate that the number of points that participants could see was 500. However, as shown in Chapter 2, participants' sensitivity to Hurst exponents depends only weakly on the length of the given series over a wide range of series lengths.

Participants' personalities were assessed using the TIPI instrument, a ten-item standardised personality questionnaire (Gosling, Rentfrow, and Swann, 2003). The TIPI evaluates personality along the dimensions of the Big Five traits.

*Procedure* The experiment consisted of three tasks. In Task A, participants were presented with 50 pairs of graphs. They were asked to determine which of the graphs presented in each pair represented an asset in which it was riskier to invest. Task B was similar to Task A, except that participants were asked to determine which of the two graphs represented an

asset which behaved more randomly. After completing tasks A and B, participants were asked to fill in the TIPI questionnaire.

### *Results*

Primary dependent variables were the percentage of each participant's answers, in which they designated as riskier the asset with a lower Hurst exponent (RiskLowHPerc) and the percentage of their answers, in which they designated as behaving more randomly the asset with a lower Hurst exponent (RandLowHPerc). A high value of RiskLowHPerc (close to 1) indicated that participants assessed the assets' risk according to the Hurst exponents of the corresponding graphs, whereas medium values (close to 0.5) indicated that the dependence of risk assessments on the Hurst exponent was close to chance level. Similar indications are applicable for RandLowHPerc.

*Inclusion criteria* For each condition separately, I performed a regression between RiskLowHPerc and RandLowHPerc for  $\Delta H = 0.1$  and the results of participants' self assessment in the TIPI questionnaire (taking into account all the personality traits in the Big Five decomposition). In the fBm condition, the Cook's distance (Cook, 1977) of one of the participants was more than two standard deviations larger than the group's mean. I, therefore discarded the results of this participant and used the answers of  $N = 40$  participants for the analysis of the results of the fBm condition.

In the fBm&fGn group, the Cook's distance of one of the participants was more than two standard deviations larger from the group's mean. In addition, the percentages of choices of graphs with low H or low standard deviation of four people were more than two standard deviations larger than the group's mean. I, therefore discarded the results of five participants from this group, and used the answers of  $N = 42$  participants for the analysis.

### *Dependence of participant performance on the experimental condition, task type and on $\Delta H$*

Table 3.1 presents the percentage of participants' answers, in which participants chose the graph with the lower Hurst exponent (RiskLowHPerc and RandLowHPerc averaged over all

participants in each group). In the fBm condition, the correspondence between participants' answers to the risk comparison task and Hurst exponent was close to chance level in all stages. T-tests showed that in the fBm condition, none of the RiskLowHPerc values was significantly different from chance level (0.5). However, RandLowHPerc were significantly different than 0.5 (for H = 0.1:  $t(39) = 8.62$ ;  $p < .01$ , for H = 0.05:  $t(39) = 6.70$ ;  $p < .01$ , and for H = 0.025:  $t(39) = 4.91$ ;  $p < .01$ ). The latter served as an indication that participants were sensitive to changes in the Hurst exponents of the graphs.

Table 3.1 The percentage of participants' answers, in which participants chose the asset with the low Hurst exponent (RiskLowHPerc and RandLowHPerc) in Experiment 1.

Condition	Task	$\Delta H$	Mean	Std
FBm	Risk comparison	0.1	0.55	0.21
		0.05	0.54	0.19
		0.025	0.51	0.17
	Randomness comparison	0.1	0.76	0.19
		0.05	0.67	0.16
		0.025	0.59	0.12
Fbm&fGn	Risk comparison	0.1	0.82	0.21
		0.05	0.70	0.20
		0.025	0.61	0.12
	Randomness comparison	0.1	0.87	0.14
		0.05	0.77	0.16
		0.025	0.65	0.14

In the fBm&fGn condition, higher levels of RandLowHPerc were obtained (for  $\Delta H = 0.1$ , the increase was 11%). These values were significantly different than 0.5 (for  $H = 0.1$ :  $t(41) = 17.86$ ;  $p < .01$ , for  $H = 0.05$ :  $t(41) = 10.53$ ;  $p < .01$ , and for  $H = 0.025$ :  $t(41) = 7.23$ ;  $p < .01$ ). However, the differences in risk assessments between fBm and fBm&fGn conditions were higher (for  $\Delta H = 0.1$ , the increase was of nearly 30% from 55% (std: 0.21) to 82% (std: 0.21)). All RiskLowHPerc values in the fBm&fGn condition were significantly different from 0.5 (for  $H = 0.1$ :  $t(41) = 9.96$ ;  $p < .01$ , for  $H = 0.05$ :  $t(41) = 6.66$ ;  $p < .01$ , and for  $H = 0.025$ :  $t(41) = 5.91$ ;  $p < .01$ ).

*Analysis of sensitivity and biases* I performed a signal detection analysis on participants' choices<sup>2</sup>. The different categories of the analysis were defined as follows:

1. A 'hit' - a case in which the participant chose the first graph and the Hurst exponent of that graph was smaller than that of the second graph.
2. A 'miss' - a case in which the participant chose the second graph, and the Hurst exponent of the first graph was smaller than that of the second graph.
3. A 'False alarm' - a case in which the participant chose the first graph, and the Hurst exponent of that graph was larger than that of the second graph.
4. A 'correct rejection' - a case in which the participant chose the second graph, and the Hurst exponent of the first graph was larger than that of the second graph.

For each participant, I calculated  $d'$  (sensitivity) and  $\beta$  (bias) (Macmillan and Creelman, 2005). To avoid a case in which  $d'$  is infinite (perfect accuracy), I converted proportions of 0 and 1 to  $1/(2N)$  and  $1-1/(2N)$  (as suggested in Macmillan and Creelman, 2005, page 8).

$d'$  is usually referred to as a sensitivity measure. In the current setting, it can be regarded as reflecting a participant's understanding of the notions of risk and randomness. For instance,

---

<sup>2</sup> An ANOVA on RiskLowHPerc and RandLowHPerc led to similar conclusions to those of the signal detection analysis.

participant with hit-rate of 1 and false-alarm rate of 0 at the randomness rating task is considered perfectly sensitive (see Macmillan and Creelman, 2005). However, such results reveal also that participant's definition of randomness coincides with the way that it has been defined here in terms of the Hurst exponent.  $d'$  was, therefore, of primary interest here.  $\beta$  was also analysed as it is a bias measure for decision criteria.

Descriptive statistics for  $d'$  and  $\beta$  are presented in Table 3.2. As can be seen in the table, all  $d'$  values were significantly different than 0, apart from those of the risk assessment in the fBm condition. The analysis failed to find differences between most  $\beta$  values and 1. A three-way ANOVA using the same variables as before on  $d'$  revealed that  $d'$  was larger in the fBm&fGn condition than in the fBm condition ( $F(1, 39) = 41.80$ ;  $p < .01$ ; partial  $\eta^2 = .52$ ), when participants assessed randomness ( $F(1, 39) = 23.11$ ;  $p < .01$ ; partial  $\eta^2 = .37$ ), and when  $\Delta H$  was larger ( $F(2, 78) = 64.48$ ;  $p < .01$ , partial  $\eta^2 = .62$ ). These results support Hypotheses  $H_{2,1}$  and  $H_{2,2}$ : the analysis failed to show any effect of the Hurst exponent on risk assessment in the fBm condition. However, there was a significant effect of the Hurst exponent on risk assessment when price change graphs were presented alongside the corresponding price series.

The effect of the interaction of Condition and Task type on  $d'$  was significant ( $F(1, 39) = 5.83$ ;  $p = .02$ , partial  $\eta^2 = .13$ ). Tests of simple effects showed that  $d'$  was higher in the randomness task than in the risk task in the fBm condition ( $F(1, 39) = 25.06$ ;  $p < .01$ ; partial  $\eta^2 = .39$ ) and in the fBm&fGn condition ( $F(1, 39) = 6.51$ ;  $p = .02$ ; partial  $\eta^2 = .14$ ). In addition,  $d'$  was larger in the fBm condition in the randomness task ( $F(1, 39) = 33.81$ ;  $p < .01$ ; partial  $\eta^2 = .46$ ) and in the risk rating task ( $F(1, 39) = 21.65$ ;  $p < .01$ ; partial  $\eta^2 = .36$ ).

A significant interaction between Condition and  $\Delta H$  was found ( $F(2, 78) = 8.49$ ;  $p < .01$ , partial  $\eta^2 = .18$ ). Tests of simple effects showed that  $d'$  was larger when  $\Delta H$  was larger in the fBm condition ( $F(2, 38) = 13.77$ ;  $p < .01$ ; partial  $\eta^2 = .42$ ) and in the fBm&fGn condition ( $F(2, 38) = 47.41$ ;  $p < .01$ ; partial  $\eta^2 = .71$ ). In addition,  $d'$  was larger in the fBm condition

when  $\Delta H = 0.1$  ( $F(1, 39) = 44.63$ ;  $p < .01$ ; partial  $\eta^2 = .54$ ), when  $\Delta H = 0.05$  ( $F(1, 39) = 24.66$ ;  $p < .01$ ; partial  $\eta^2 = .39$ ), and when  $\Delta H = 0.025$  ( $F(1, 39) = 15.17$ ;  $p < .01$ ; partial  $\eta^2 = .28$ ).

Table 3.2 Mean values of  $d'$  and  $\beta$  in conditions fBm (first panel) and fBm&fGn (second panel) in Experiment 1.

Condition	Task	$\Delta H$	$d'$			B		
			Mean	Std	t-test	Mean	Std	t-test
					comparing $d'$ to 1			comparing B to 1
FBm (N=40)	Risk	0.1	0.29	1.24	t(39) = 1.49; p = .15	1.08	0.42	t(39) = 1.15; p = .26
		0.05	0.22	1.12	t(39) = 1.25; p = .22	1.07	0.51	t(39) = 0.81; p = .42
		0.025	0.08	0.97	t(39) = 0.51; p = .62	1.10	0.41	t(39) = 1.47; p = .15
	Randomness	0.1	1.53	1.15	t(39) = 8.43; p < .01	0.96	0.32	t(39) = -0.79; p = .43
		0.05	0.91	0.90	t(39) = 6.38; p < .01	1.24	0.68	t(39) = 2.21; p = .03
		0.025	0.49	0.68	t(39) = 4.56; p < .01	1.25	0.81	t(39) = 1.96; p = .06

Condition	Task	$\Delta H$	d'			B			
			Mean	Std	t-test	Mean	Std	t-test	
					comparing d' to 1			comparing B to 1	
Fbm&fGn (N=42)	Risk	0.1	1.88	1.18	t (41) = 10.37; p < .01	1.24	0.48	t (41) = 3.18; p = .003	
		0.05	1.16	1.16	t (41) = 6.47; p < .01	1.11	0.49	t (41) = 1.47; p = .15	
		0.025	0.63	0.75	t (41) = 5.47; p < .01	1.19	0.54	t (41) = 2.30; p = .03	
		Randomness	0.1	2.18	0.83	t (41) = 17.12; p < .01	1.18	0.67	t (41) = 1.73; p = .09
			0.05	1.55	0.94	t (41) = 10.63; p < .01	1.25	0.62	t (41) = 2.56; p = .01
			0.025	0.88	0.78	t (41) = 7.30; p < .01	1.11	0.41	t (41) = 1.75; p = .09

There was also a significant interaction between Task and  $\Delta H$  ( $F(2, 78) = 3.31$ ;  $p = .042$ , partial  $\eta^2 = .08$ ). Tests of simple effects showed that  $d'$  was larger when  $\Delta H$  was larger in the risk task ( $F(2, 38) = 20.03$ ;  $p < .01$ ; partial  $\eta^2 = .51$ ) and in the randomness task ( $F(2, 38) = 40.44$ ;  $p < .01$ ; partial  $\eta^2 = .68$ ). Percentage of low-H choices was higher in the randomness task when  $\Delta H = 0.1$  ( $F(1, 39) = 23.48$ ;  $p < .01$ ; partial  $\eta^2 = .38$ ), when  $\Delta H = 0.05$  ( $F(1, 39) = 11.60$ ;  $p = .02$ ; partial  $\eta^2 = .23$ ), and when  $\Delta H = 0.025$  ( $F(1, 39) = 6.86$ ;  $p = .012$ ; partial  $\eta^2 = .15$ ).

A three-way ANOVA on  $\beta$  using the same variables as before failed to find any significant effect of Condition, Task type, or H difference on  $\beta$ .

*Correlation between individual characteristics and risk/randomness judgment* There were statistically significant correlations between participants' performance at different  $\Delta H$ -levels of the risk and randomness comparison task. Correlation results are presented in Table 3.3. These correlations suggest that individual differences (e.g., personality traits) might affect risk and randomness ratings.

I calculated the correlations between personality trait ratings, RandLowHPerc, and RiskLowHPerc. For the fBm condition, when  $\Delta H$  was 0.05, RandLowHPerc increased with self-rating of Agreeableness ( $r = .34$ ;  $p = .03$ ). Agreeableness was also correlated with RiskLowHPerc when  $\Delta H$  was 0.1 ( $r = .39$ ;  $p = .01$ ). Correlations of performance with agreeableness may indicate more agreeable participants tended to cooperate more with the task requirements (as they perceived them).

Risk assessment depended also on emotional stability: investment risks judged by participants with lower emotional stability showed greater dependence on the Hurst exponent (for RiskLowHPerc in the fBm condition, when  $\Delta H$  was 0.1,  $r = -.32$ ;  $p = .046$ , and when  $\Delta H$  was 0.05,  $r = -.31$ ;  $p = .050$ ). The traits agreeableness and emotional stability were not significantly correlated. The results are presented in Figures 3.2 and 3.3.

Table 3.3 Correlations between percentage of H-correlated answers in the fBm condition (first panel) and fBm&fGn condition (second panel) of Experiment 1. Statistically significant correlations are marked with a star.

fBm condition	Task						
		Risk			Randomness		
Task	$\Delta H$	0.1	0.05	0.025	0.1	0.05	0.025
Risk	0.1	1	$r = .58^*$ , $p < .001$	$r = .64^*$ , $p < .001$	$r = .20$ , $p = .21$	$r = .35^*$ , $p = .026$	$r = .07$ , $p = .65$
	0.05		1	$r = .56^*$ , $p < .001$	$r = .19$ , $p = .25$	$r = .29$ , $p = .07$	$r = .04$ , $p = .82$
	0.025			1	$r = .18$ , $p = .27$	$r = .42^*$ , $p = .006$	$r = -.010$ , $p = .96$
Randomness	0.1				1	$r = .46^*$ , $p = .003$	$r = .31$ , $p = .05$
	0.05					1	$r = .22$ , $p = .18$
	0.025						1

Fbm&fGn		Task					
condition		Risk			Randomness		
Task	$\Delta H$	0.1	0.05	0.025	0.1	0.05	0.025
Risk	0.1	1	$r = .56^*$ , $p < .001$	$r = .19$ , $p = .23$	$r = .18$ , $p = .25$	$r = .23$ , $p = .15$	$r = .22$ , $p = .17$
	0.05		1	$r = .60^*$ , $p < .001$	$r = .11$ , $p = .47$	$r = .13$ , $p = .42$	$r = .35^*$ , $p = .02$
	0.025			1	$r = .08$ , $p = .61$	$r = .18$ , $p = .26$	$r = .36^*$ , $p = .02$
Randomness	0.1				1	$r = .69^*$ , $p < .001$	$r = .24$ , $p = .12$
	0.05					1	$r = .31^*$ , $p = .04$
	0.025						1

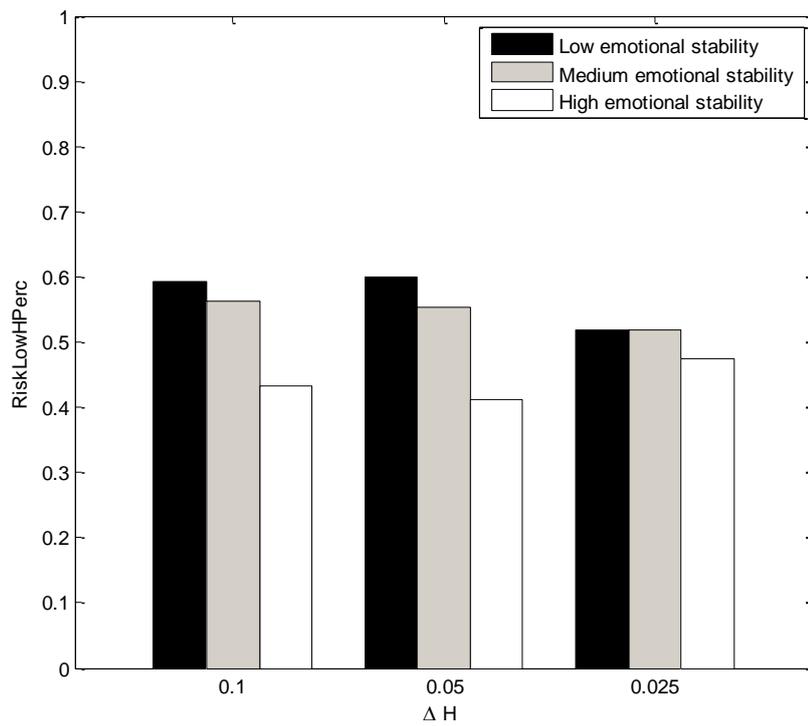
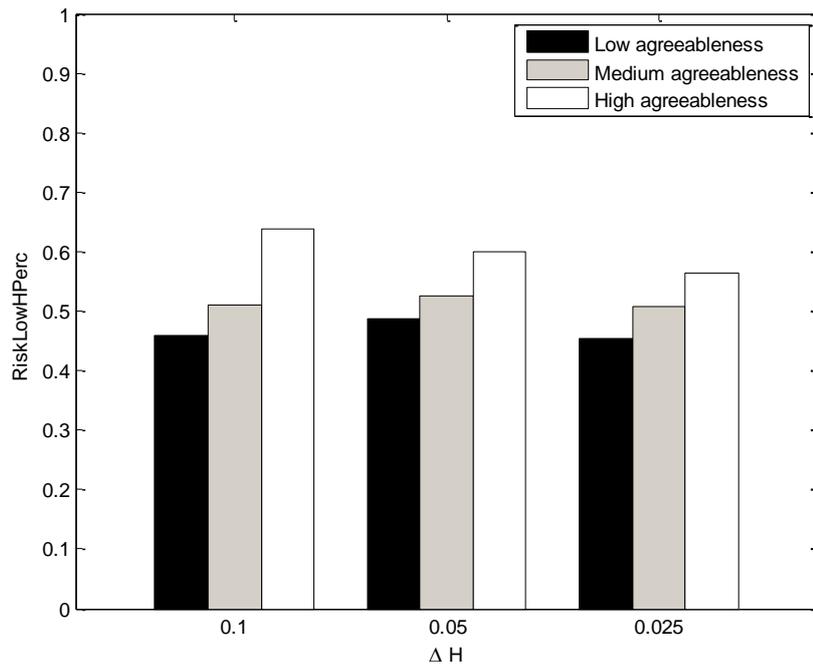


Figure 3.2 Percentage of choices of graphs with low Hurst exponent at the risk comparison task in the fBm condition in Experiment 1 against  $\Delta H$ , presented for participant sections with different self-ratings of agreeableness (first row) and emotional stability (second row).

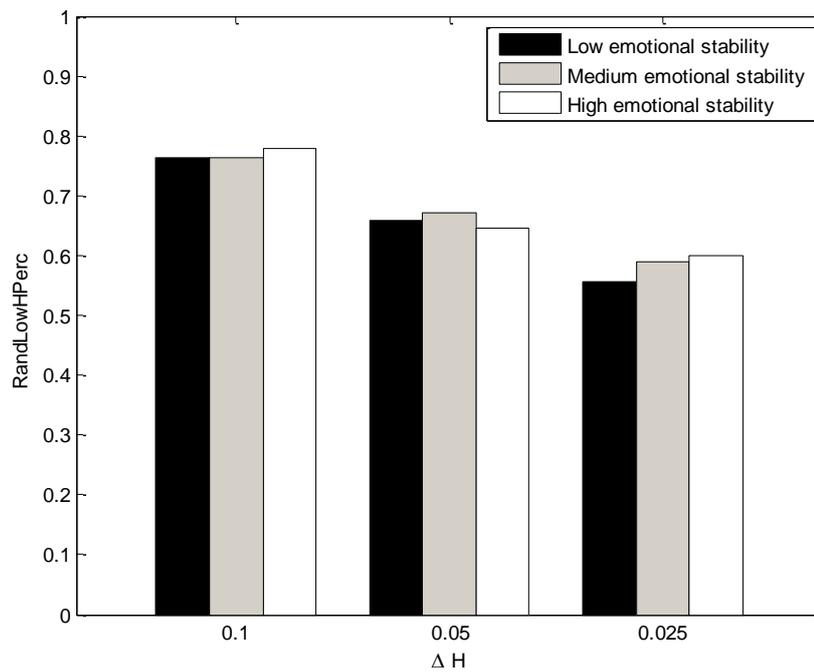
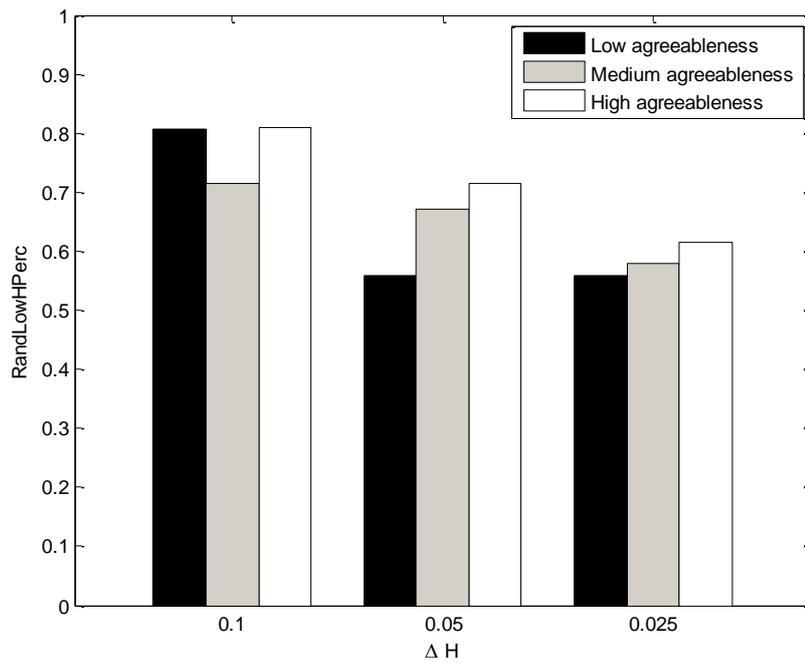


Figure 3.3 Percentage of choices of graphs with low Hurst exponent at the randomness comparison task in the fBm condition in Experiment 1 against  $\Delta H$ , presented for participant sections with different self-ratings of agreeableness (first row) and emotional stability (second row).

Although risk discrimination in the fBm condition did not depend on the Hurst exponent of the graphs, in nearly 60% of the trials participants with low emotional stability designated the asset with the lower Hurst exponent as riskier to invest in (Figure 3.2).

The results for  $d'$  were similar: with a large  $\Delta H$ ,  $d'$  for the risk task was significantly correlated with agreeableness ( $r = .37$ ;  $p = .02$ ) and emotional stability ( $r = -.32$ ;  $p = .04$ ).

With a medium  $\Delta H$ , the correlation between  $d'$  for the risk task agreeableness was  $r = .33$ ;  $p = .04$ ) and with self rating of emotional stability was  $r = -.33$ ;  $p = .04$ . Agreeableness was also correlated with  $d'$  for the randomness task with medium  $\Delta H$  ( $r = .34$ ;  $p = .03$ ) and with  $\beta$  of the randomness task at stage 3 ( $r = .32$ ;  $p = .047$ ). No other correlations between  $d'$  or  $\beta$  and personality traits were found for the fBm condition. Results supported Hypothesis  $H_{2,3}$ , according to which risk ratings of people who are low on emotional stability are correlated with the Hurst exponent of the presented graphs more strongly than those of people who are high on emotional stability.

In the fBm&fGn condition, people who rated their extraversion lower had higher values of RiskLowHPerc for all  $\Delta H$  values (for  $\Delta H = 0.05$ ,  $r = -.50$ ;  $p = .001$ , for  $\Delta H = .05$ ,  $r = -.48$ ;  $p = .001$ , and for  $\Delta H=0.025$ ,  $r = -.33$ ;  $p = .04$ ). No other correlations were found between personality traits ratings, RandLowHPerc, and RiskLowHPerc.

The correlation between  $d'$  and extraversion was significant for the risk task (for stage 1:  $r = -.50$ ;  $p < .01$ , for stage 2:  $r = -.48$ ;  $p < .01$ , for stage 3:  $r = -.34$ ;  $p = .03$ ). For the same task, the correlation between  $\beta$  and extraversion at stage 1 was  $r = -.46$ ;  $p < .01$ . No other correlations were found between personality traits and the  $d'$  or  $\beta$  at the risk or randomness tasks.

One-way ANOVAs on the variables RiskLowHPerc, RandLowHPerc,  $d'$ , and  $\beta$ , with respect to expertise failed to find differences in risk or randomness assessments of experts and non-experts of participants in the fBm condition. However, in the fBm&fGn condition, experts had higher values of RandLowHPerc(fGn,3) ( $F(1, 40) = 8.21$ ;  $p < .01$ ) and  $d'$  ( $F(1, 40) =$

8.70;  $p = .005$ ). This finding suggests that, although experts were more sensitive to differences in the Hurst exponents of the graphs, they did not use this information in their risk assessment differently than non-experts did.

### *Discussion*

Experiment 1 showed that, given no further cues, risk assessment of assets for which prices were represented by fractal graphs did not depend on the Hurst exponent of those graphs in most of the participants. This supports Hypothesis  $H_{2,1}$ . Furthermore, the experiment revealed that this lack of dependence was not a result of inability of discriminating Hurst exponent of the given graphs: 76% (std: 0.19) of participants' randomness ratings were correlated with the Hurst exponents of each graph pair at the first stage of the experiment. This percentage is far above chance level.

However, when price change graphs were presented with the corresponding price graphs, 82% (std: 0.21) of participants' answers designated assets with the lower Hurst exponent as the riskier investments. Beyond emphasising the fragility of notions of human risk perception, this result suggests that people indeed have the ability to relate to fractal properties when assessing risk. In particular, it supports Hypothesis  $H_{2,2}$ .

Experiment 1 also demonstrated that personality traits influence risk assessment. When price change information was not explicit, emotional stability and agreeableness affected risk perception. Emotional stability did not affect randomness judgements. This result corresponds to that of Jakes and Hemsley (1986), who showed that people high in neuroticism tend to attribute meanings to complex patterns they find in presented stimuli. Furthermore, Steger, Kashdan, Sullivan, and Lorentz, (2008) showed that, in non-financial contexts, people low in emotional stability and high in agreeableness tend to search for meaning more than others. Therefore, these results suggest that the search for meaning guided participants to interpret the Hurst exponent as a risk measure. On the other hand, when price change information was explicit, and provided participants with a clear cue for

the meaning of the task, risk discrimination was no longer affected by emotional stability. Instead, it was affected by extraversion, a personality trait related to risk-propensity (Nicholson, Soane, Fenton-O'Creevy, and Willman, 2005).

## **Experiment 2**

Experiment 2 was designed to replicate of the results obtained in Experiment 1 for Hypothesis  $H_{2,2}$  and to examine the following hypotheses:

Hypothesis  $H_{2,4,a}$ : the series Hurst exponent, standard deviation, mean run length, oscillation, and absolute value of the difference between the values of the last and first points of the series are correlated with risk assessments.

Hypothesis  $H_{2,4,b}$ : the difference between the values of the last and first points of the series and the difference between the first series point and its minimum are negatively correlated with risk assessments.

Hypothesis  $H_{2,5}$ : the effect of the Hurst exponent on risk assessment is stronger than that of the standard deviation.

To test these hypotheses, I presented participants on each trial with a single price graph and its corresponding price change graph. Participants were asked to rate the risk level of investment in the described asset rather than to compare risk levels as in Experiment 1.

### *Method*

*Design* Participants were asked to assess the risk level of investment in a single asset at each trial. Price graphs were presented with their corresponding price change graphs.

For each participant, two sets of nine graphs with  $H = 0.1, 0.2, \dots, 0.9$  were randomly chosen from six sets of fBm graphs, resulting in a set of 18 graphs. This manipulation resulted in a two (graph instance) by nine (H values) design.

*Participants* Forty-two people (29 men and 13 women, average age: 35.8 years) acted as participants. They were recruited through professional groups of financial analysts and economists on LinkedIn, and the departmental participant pool. All participants were offered participation in a prize draw of four USB sticks, and information about the experiment. Students from UCL were offered, in addition, 0.25 academic credit points.

Participants were asked whether they were financial analysts. Thirteen participants gave a positive answer to this question.

*Materials* I generated six sets of target graphs each with nine different H values ranging from 0.1 to 0.9 in steps of 0.1, using the spectral algorithm described by Saupe (Peitgen and Saupe, 1988). Each of the series had 6284 points consisting of one period. The target graphs consisted of a quarter of a period (1571 points). Hence, the differences between the values of the first and last presented points were random. No scaling was performed on the stimulus series. For each of these graphs, a corresponding fGn series was calculated as in Experiment 1. The task window is presented in Figure 3.4.

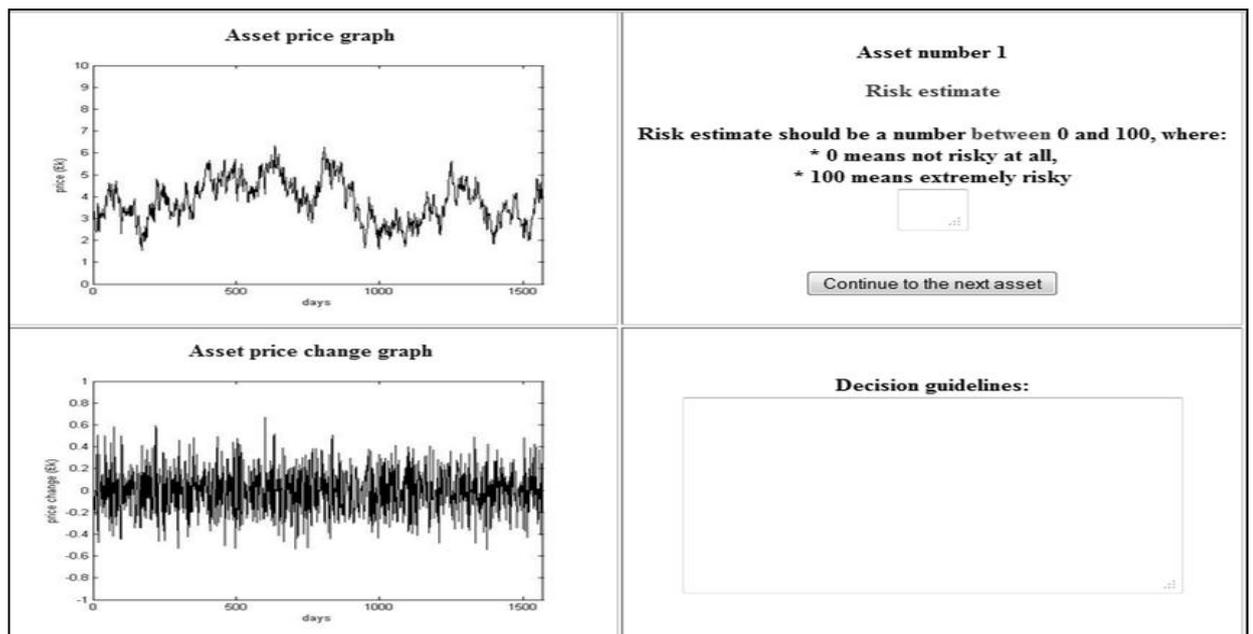


Figure 3.4 The task window of Experiment 2.

### *Procedure*

Each participant was presented with 18 computer-generated graphs and their corresponding change series. Participants were told that these graphs represent prices and price daily changes. Participants were asked to look at each of the graphs carefully and to assess the risk level of investment in the given asset as a number between 0 and 100, where 0 meant: "not risky at all" and 100 meant "extremely risky".

### *Results*

Primary dependent variables were participants' risk assessments and the following seven variables: Hurst exponent, standard deviation, the series mean run length, the series oscillation, the difference between the values of the last and first points of the series, the absolute value of the difference between the values of the last and first points of the series, and the difference between the first series point and its minimum. I was interested in the standard deviation as it is a basic measure for risk according to normative theories (Hendricks, 1996). The effect of mean run length on risk assessment was studied by Raghurir and Das (2010). Oscillation and the absolute value of the difference between the values of the last and first points of the series are measures for the size of the changes in the series. The difference between the values of the last and first points of the series indicates the general direction of the trend. The difference between the first series point and the minimum of the series may indicate how much money can be lost. Notations of these variables are given in Table 3.4.

Correlations between risk assessments and the seven series variables may indicate the importance participants attributed to the latter as risk indicators.

*Inclusion criteria* I performed a regression between the mean risk assessment of each participant and participants' responses to the TIPI questionnaire (taking into account all the personality traits in the Big Five decomposition). The Cook's distance (Cook, 1977) of two

Table 3.4 Variable notation

Notation	Description
H	The Hurst exponent
STD	The standard deviation
MeanRun	The mean run length of the series. Run length is the number of consecutive elements in the series, in which the series does not change its direction.
Osc	The series oscillation (the difference between its maximum and minimum values)
Diff	The difference between the values of the last and first points of the series
AbsDiff	The absolute value of the difference between the values of the last and first points of the series
FirstMinDiff	The difference between the first series point and its minimum

of the participants was more than two standard deviations larger than the group's mean. I, therefore, excluded their results from the analysis.

In addition, for each participant, I calculated the correlation between risk assessment and the Hurst exponents of the graphs, and between risk assessment and the standard deviations of the graphs. Participants whose mean scores of both correlations were smaller by more than two standard deviations than those of the average for the group were excluded from the analysis. This resulted in the exclusion of two additional participants from the analysis, reducing the size of the sample to 38 participants.

*The effect of the Hurst exponent on risk assessments* I performed a two-way repeated measures ANOVA on participants' risk assessments, using Hurst exponent (0.1, 0.2, ..., 0.9) and Instance (first or second presentation) as within-participant variables. The Hurst exponent violated Mauchly's test of sphericity and hence I report the results of a Huynh-Feldt test. Risk assessment was higher when the Hurst exponent was smaller ( $F(5.38, 199.21) = 32.44$ ;  $p < .01$ ; partial  $\eta^2 = .47$ ). No other effect was significant.

I was particularly interested in participants' risk assessments for the range  $H \in [0.3, 0.7]$ , as the Hurst exponent of most real assets is included in this range. The difference in risk estimates between graphs with  $H = [0.3, 0.4]$  and graphs with  $H = [0.6, 0.7]$  was statistically significant ( $t(159) = 8.70$ ;  $p < .01$ ) and so were the differences in risk estimates between graphs with  $H = [0.1, 0.3]$  and graphs with  $H = [0.4, 0.6]$  ( $t(239) = 0.15$ ;  $p < .01$ ), and between graphs with  $H = [0.4, 0.6]$  and graphs with  $H = [0.7, 0.9]$  ( $t(239) = 6.00$ ;  $p < .01$ ). Figure 3.5 presents these results.

Experiment 2, therefore, provided additional support for Hypothesis  $H_{2.2}$ .

*Correlations between graph variables and risk assessment* The correlations between participants' risk assessments and the variables are presented in Table 3.4. Correlations between risk estimates and these variables are given in Table 3.5 (first row). The correlations between risk assessments and H, Std, Osc and FirstMinDiff were the highest (their absolute values were in the range  $[0.46, 0.49]$ ;  $p < 0.01$ ). Participants judged a series to be riskier when its Hurst exponent was smaller.

The similarity of the correlations between risk estimates and H, Std, Osc and FirstMinDiff was expected, as these variables were correlated. Table 3.6 presented the correlations between the examined variables.

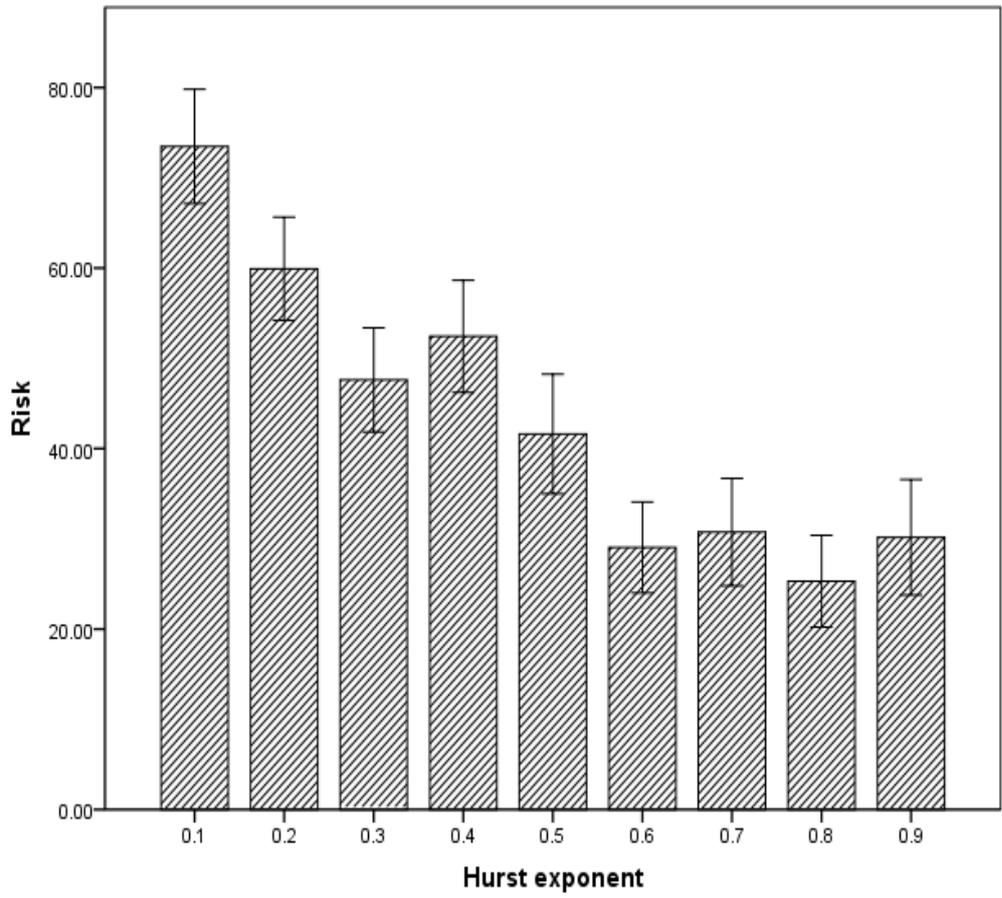


Figure 3.5 Mean risk assessment plotted against the Hurst exponents of the presented graphs.

Table 3.5 Correlations and partial correlations between risk assessment and graph variables, and the beta values in multiple regression of risk assessment with the seven variables in Experiment 2.

$R^{2*}$  denotes the  $R^2$  of regression of all variables but the variable in each column.

Diff  $R^2$   $R^{2*}$  denotes the difference between  $R^2$  of regression of all variables together ( $R^2 = .31$ ) and  $R^{2*}$ .

	H	Std	MeanRun	Osc	Diff	AbsDiff	FirstMinDiff
Correlation with risk	$r = -.47,$ $p < .01$	$r = .46,$ $p < .01$	$r = -.39,$ $p < .01$	$r = .49,$ $p < .01$	$r = -.19,$ $p < .01$	$r = .24,$ $p < .01$	$r = .46,$ $p < .01$
Partial correlation with risk, with respect to control variables	$r = -.16,$ $p < .01$	$r = -.05,$ $p = .168$	$r = .09,$ $p = .01$	$r = .10,$ $p < .01$	$r = -.22,$ $p < .01$	$r = .11,$ $p < .01$	$r = -.12,$ $p < .01$
	Std,	H,	H,	H,	H,	H,	H,
	MeanRun,	MeanRun,	Std,	Std,	Std,	Std,	Std,
	Osc,	Osc,	Osc,	MeanRun,	MeanRun,	MeanRun,	MeanRun,
	Diff,	Diff,	Diff,	Diff,	Osc,	Osc,	Osc,
	AbsDiff,	AbsDiff,	AbsDiff,	AbsDiff,	AbsDiff,	Diff,	Diff,
	FirstMinDiff	FirstMinDiff	FirstMinDiff	FirstMinDiff	FirstMinDiff	FirstMinDiff	AbsDiff
Beta values	$\beta = -.67$ $p < .01$	$\beta = -.19$ $p = .17$	$\beta = .27$ $p = .01$	$\beta = .49$ $p < .01$	$\beta = -.45$ $p < .01$	$\beta = .11$ $p < .01$	$\beta = -.37$ $p < .01$
$R^{2*}$	$R^2 = .29$	$R^2 = .31$	$R^2 = .30$	$R^2 = .30$	$R^2 = .27$	$R^2 = .30$	$R^2 = 0.30$
Diff $R^2$ $R^{2*}$	0.017	0.001	0.006	0.006	0.034	0.007	0.01

Table 3.6 Correlations between the variables examined in Experiment 2 for the stimuli sample.

	H	Std	MeanRun	Osc	Diff	AbsDiff	FirstMinDiff
H		$r = -.78^*$ , $p < .01$	$r = .92^*$ , $p < .01$	$r = -.86^*$ , $p < .01$	$r = -.08^*$ , $p = .04$	$r = -.32^*$ , $p < .01$	$r = -.63^*$ , $p < .01$
Std			$r = -.67^*$ , $p < .01$	$r = .95^*$ , $p < .01$	$r = -.10^*$ , $p = .01$	$r = .55^*$ , $p < .01$	$r = .73^*$ , $p < .01$
MeanRun				$r = -.70^*$ , $p < .01$	$r = -.08^*$ , $p = .04$	$r = -.29^*$ , $p < .01$	$r = -.49^*$ , $p < .01$
Osc					$r = -.03$ , $p = .48$	$r = .45^*$ , $p < .01$	$r = .75^*$ , $p < .01$
Diff						$r = -.003$ , $p = .95$	$r = -.60^*$ , $p < .01$
AbsDiff							$r = .33^*$ , $p < .01$

In order to estimate the relative contributions of each of the variables, I calculated the correlations again, this time controlling for all other six variables at each calculation. As Table 3.5 (second row) shows, controlling for the variables Std, MeanRun, Osc, Diff, AbsDiff, and FirstMinDiff, the correlation between risk assessment and the Hurst exponent of the graph was  $r = -.16$ ,  $p < .01$ . This correlation was second only to the correlation of risk assessment with Diff. The partial correlation of risk assessment with Std was insignificant.

A regression of risk assessment with respect to these seven variables yielded  $R^2 = 0.31$  (  $F(7, 683) = 45.38$ ;  $p < .01$ ). The beta values ( $\beta$ ) corresponding to each of the variables are presented in Table 3.5 (third row). The absolute value of the  $\beta$  of the Hurst exponent was the highest: 0.67 ( $p < .01$ ), whereas the  $\beta$  of MeanRun was smaller ( $\beta = 0.27$ ;  $p = .013$ ) and the beta value of the std was insignificant. Regressing risk with respect to the variables Hurst exponent alone yielded  $R^2 = 0.22$  (  $F(1, 683) = 202.64$ ;  $p < .01$ ).

Furthermore, I calculated the difference between the  $R^2$  values of a regression model containing all seven variables, and the  $R^2$  values of a regression model containing all seven variables apart from each of the seven variables separately (Cooksey, 1996, page 165-166). This difference is termed ‘usefulness coefficient’. It is used as a measure for the contribution of each variable over the contributions of the other variables. The results are presented in Table 3.5 (the last two rows). This difference measures the contribution of each of the seven variables beyond the contribution common to of all predictors and is termed ‘usefulness index’. I found that the difference between the last and first points of the series had the largest independent contribution to risk assessment. However, as before, I found that the effect of the Hurst exponent on risk ratings was larger than that of the standard deviation or the mean run length.

I, therefore, conclude that the effect of the Hurst exponent on risk assessment is stronger than that of the standard deviation and the mean run-length. The difference between the last and first points of the series affects risk assessment, too. I, therefore, accept Hypotheses  $H_{2,4}$  and  $H_{2,5}$ .

### *Discussion*

Experiment 2 showed that, when price graphs are presented along with price change graphs, the Hurst exponent affects risk judgements. More precisely, the lower the Hurst exponent was, the higher the perceived risk was. This provides further support for Hypothesis  $H_{2,2}$ .

The dependence of risk assessment on the Hurst exponent was stronger than on the standard deviation of the graphs, a measure used to estimate the historical volatility in normative financial models (Hendricks, 1996). That supports Mandelbrot and Hudson's (2004) views about people's reaction to fractal characteristics of price series and Hypothesis H<sub>2,5</sub>.

The standard deviation of the graphs, their oscillation (the difference between its maximum and minimum values), and the differences between the first and last presented points also had effects on risk assessment, supporting Hypothesis H<sub>2,4</sub>.

The results complement those of Duxbury and Summers (2004). In spite of the differences between the experimental settings used here and those of Duxbury and Summers, I showed that the difference between the first and last elements of the presented series was negatively correlated with risk assessments. This difference could be considered as a measure of the amount of money which was likely to be lost when investing in an asset.

### **Experiment 3**

Experiments 1 and 2 showed that risk perception is affected by mathematical properties of the presented data. But are financial decisions affected by it?

Experiment 3 was designed to address the following hypotheses:

Hypothesis H<sub>2,6</sub>: the standard deviation of an asset's graph and its mean run length affect buy/sell decisions.

Hypothesis H<sub>2,7</sub>: the lower the Hurst exponent of an asset's price series is, the higher people's tendency to sell it is. The higher the Hurst exponent of the price series is, the higher people's tendency to buy it is.

To examine these hypotheses, I presented participants with pairs of graphs of fBm series representing different assets, along with their corresponding fGn graphs in a similar way to that used in the fBm&fGn condition in Experiment 1. However, in Experiment 3, I asked

participants to decide which of the assets they would have liked to buy or sell. In addition, I asked them to rate their confidence level in their decision.

### *Method*

*Design* Participants were randomly allocated to the buy or sell condition. Fifty graphs were chosen randomly for each participant. I denote the Hurst exponent difference between graphs in each set by  $\Delta H$ . As in Experiment 1, graph pairs were chosen according to the required  $\Delta H$ . The first stage included 15 graphs with  $\Delta H = 0.1$ , the second stage included 15 graphs with  $\Delta H = 0.05$ , and the third stage comprised 20 graphs with  $\Delta H = 0.025$ .

These manipulations resulted in a two (buy or sell condition) by three ( $\Delta H = 0.1, 0.05, 0.025$ ) design.

*Participants* Eighty four people participated in the experiment (24 women, 60 men, average age: 45.4 years). They were randomly allocated to two groups: the Buy group and the Sell group. The Buy group included 40 participants (13 women and 27 men, average age: 45.2 years) and the Sell group included 44 people (11 women and 33 men, average age: 45.6 years). As in the previous experiments, participants represented wide cultural spectrum.

Participants were recruited through professional groups of financial analysts and economists on LinkedIn and through student websites. They were asked whether they work as financial analysts. Eleven of them replied positively within the Buy group, and 12 of them replied positively within the Sell group.

*Materials* Stimulus materials comprised the same graph sets that were used for the fGn condition in Experiment 1. For each participant, presented graphs were chosen randomly from the six graph sets. They were presented in a random order.

The task window of Experiment 3 enabled participants to choose the asset they wanted to buy and to rate their confidence level in their decision. The task window of Experiment 3 is shown in Figure 3.6.

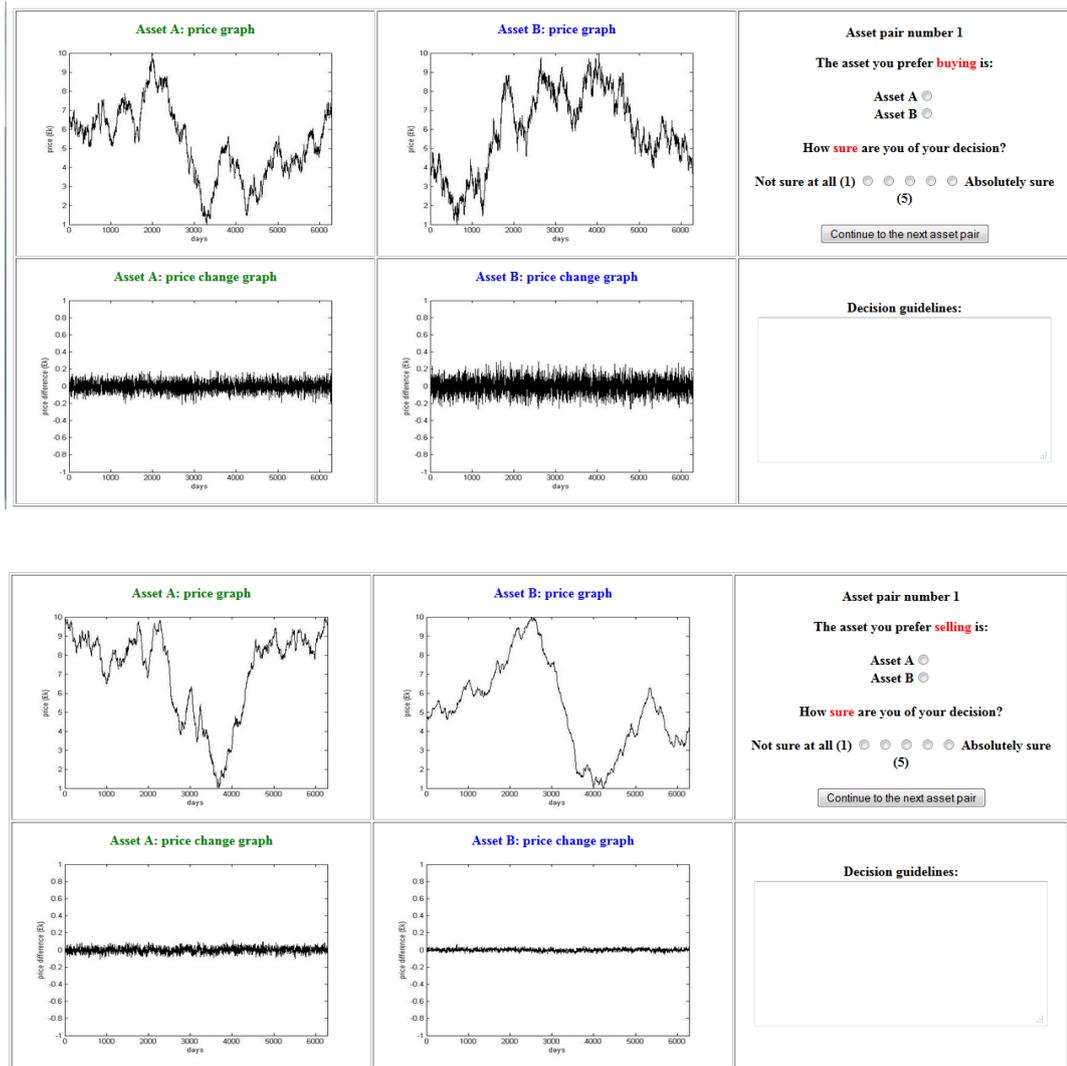


Figure 3.6 The task window of Experiment 3. Upper panel: the buy condition; Lower panel: the sell condition.

*Procedure* Participants in both buy and sell conditions were told that they would be presented with a sequence of 50 sets of graphs, each of which would include two graphs describing prices of different assets, A and B, versus time, and two corresponding graphs describing the daily price changes of the same assets versus time. Participants in the Buy condition were asked to imagine that they had £1000 and would like to buy shares of an asset for £500. Then, they were asked to decide which of the assets they would like to buy. Participants in the Sell condition were asked to imagine that they had £500 worth shares of

asset A and £500 worth shares of asset B, and that they wanted to sell one of these assets. They were asked to decide which of these assets they would like to sell. Participants in both conditions were asked to provide confidence judgments. To do so, they were required to assess how sure they were about each of their decisions on a scale of 1 to 5, where 1 meant "not sure at all", and 5 meant "absolutely sure".

### *Results*

The primary dependent variable was the percentage of each participant's answers, in which the asset with the lower Hurst exponent was bought (BuyLowHPerc) or sold (SellLowHPerc). A high value of LowHPerc for participants in the buy condition (closer to 1) indicates that participant chose to buy assets with low Hurst exponent, whereas medium values (close to 0.5) indicates that the dependence of buying choices on the Hurst exponent is close to chance level. Similar interpretation is applicable for the sell condition.

*Inclusion criteria* For each participant, I calculated BuyLowHPerc or SellLowHPerc. Participants whose mean score of percentage of low-H choices was two standard deviations smaller or larger than those of the average of their group were excluded from the analysis. This resulted in the exclusion of two participants, reducing the size of the sample to 82 participants (39 participants in the Buy group and 43 participants in the Sell group).

*Percentage of choices of assets with low Hurst exponent* A two-way repeated measure ANOVA was performed on the percentage of low-H choices, using condition (buy or sell) as a between-participant variable, and  $\Delta H$  ( $\Delta H = 0.1$ ,  $\Delta H = 0.05$ , or  $\Delta H = 0.025$ ) as a within-participant variable. None of the variables violated Mauchly's test of sphericity. Percentage of low-H choices was higher in the buy condition ( $F(1, 37) = 5.39$ ;  $p = .03$ ; partial  $\eta^2 = .13$ ) but there was no effect of  $\Delta H$  on the percentage of low-H choices. The results are shown in Table 3.7.

These results show that people prefer buying assets with a higher Hurst exponent and selling assets with a lower Hurst exponent. I, therefore, accept Hypothesis H<sub>2,7</sub>.

Table 3.7 The percentage of participants' answers, in which participants chose the asset with the lower Hurst exponent in Experiment 3, and the associated confidence ratings.

Variable	Condition	$\Delta H$	Mean	Std
Percentage of low H choices	Buy	0.1	0.44	0.18
		0.05	0.46	0.15
		0.025	0.48	0.12
	Sell	0.1	0.53	0.16
		0.05	0.50	0.15
		0.025	0.51	0.12
Confidence in low H choices	Buy	0.1	5.76	3.08
		0.05	5.74	3.03
		0.025	5.58	2.97
	Sell	0.1	5.28	3.40
		0.05	5.45	3.34
		0.025	5.56	3.27

*Confidence level in choice of the asset with the lower Hurst exponent* Using participants' confidence ratings, I constructed a score representing the confidence level of participants' decisions in a choice of the asset with a lower Hurst exponent. The range of the score was 1-10, where:

- 1 represented a confident choice of the asset with the lower Hurst exponent, corresponding to cases in which participants rated their confidence level as 5 (“extremely sure”),
- 5 represented an unconfident choice of the asset with the lower Hurst exponent, corresponding to cases in which participants rated their confidence level as 1 (“extremely unsure”),
- 6 represented an unconfident choice of the asset with the higher Hurst exponent, corresponding to cases in which participants rated their confidence level as 1 (“extremely unsure”),
- 10 represented a confident choice of the asset with the higher Hurst exponent, corresponding to cases in which participants rated their confidence level as 5 (“extremely sure”).

I performed a two-way repeated measure ANOVA for this confidence score, using Condition (buy or sell) as a between-participant variable, and  $\Delta H$  ( $\Delta H = 0.1$ ,  $\Delta H = 0.05$ , or  $\Delta H = 0.025$ ) as a within-participant variable. None of the variables violated Mauchly’s test of sphericity. Confidence in low-H choices was higher in the buy condition ( $F(1, 584) = 8.41$ ;  $p = .004$ ; partial  $\eta^2 = .014$ ).  $\Delta H$  did not affect the percentage of low-H choices. The results are presented in Table 3.7.

*The effect of Std and MeanRun on choices* For each participant, I calculated the percentages of answers in which participants chose the graph with the smaller value of the variable Std and MeanRun. For each of these variables, I performed a two-way repeated measure ANOVA using the same variables as before. The analysis failed to show a significant effect of Condition or  $\Delta H$  on the percentages of answers in which participants chose the smaller value of Std or MeanRun.

## *Discussion*

Experiment 3 showed that people prefer buying assets with a high Hurst exponent and selling assets with a low Hurst exponent. This result remained statistically significant when confidence ratings were taken into account. In contrast, the standard deviation and mean run length of the series did not affect trading behaviour.

## **Conclusions**

Holton (2004) asserted that “it is impossible to operationally define risk. At best, we can operationally define our *perception* of risk... Perceived risk takes many forms”. This study aimed to elucidate the way people perceive risk of assets when their prices are presented graphically.

The experiments supported Mandelbrot and Hudson’s (2004) argument that people are sensitive to the fractal characteristics of price graphs. Risk assessments were found to be correlated with the Hurst exponent of the presented graphs. This correlation was similar to that between risk assessments and the standard deviation of the graphs. However, controlling for all other variables, the correlation between risk assessments and the Hurst exponent of the graphs was much stronger than that between risk assessments and the standard deviation or the mean run length of the graphs. Furthermore, financial buy/sell decisions were correlated with the Hurst exponent: participants preferred buying assets with high Hurst exponents and selling assets with low Hurst exponents. There is a large body of evidence showing that the majority of people exhibits risk aversion through their choices (Simonsohn, 2009; Mattos, Garcia, and Pennings Joost, 2007). If participants attributed higher risk to graphs with lower Hurst exponents, then they should prefer to buy assets with higher Hurst exponents. Indeed, participants’ trading choices fitted this model. The analysis failed to find significant correlations between financial decisions and the standard deviation of the graphs or between those decisions and mean run length of the graphs.

The results depended on the task's characteristics: when price graphs alone were presented, most participants did not attribute higher risk to lower Hurst exponents. That was in spite of their sensitivity to the Hurst exponent, as exhibited by the correlation between Hurst exponents of presented graphs and randomness ratings, obtained with the graphs having the same characteristics as they did in the risk assessment task. Sensitivity to the Hurst exponent was observed also by Westheimer (1991) and Gilden, Schmuckler and Clayton (1993). On the other hand, when price graphs were presented along with their corresponding price change graphs, participants' risk assessments were significantly correlated with the Hurst exponent of the graphs. As participants exhibited high levels of sensitivity to the Hurst exponent in the condition in which no price change graphs were exhibited, I argue that dependence of risk perception on the Hurst exponent cannot be fully explained by a perceptual improvement due to the presence of fGn graphs, or by participants' attempts to guess what the experimental manipulation was. I suggest that, rather than providing only perceptual information, price change graphs are used also as verification cues: presentation of price change graphs validated the meaning of the Hurst exponent, of which participants were aware with or without the price change graphs, as a risk measure.

When price change graphs were not presented, the extent to which participants' risk assessments depended on the Hurst exponent was negatively correlated with participants' emotional stability and positively correlated with their agreeableness. Studies concerned with search for meaning in non-financial contexts have revealed that people low in emotional stability and high in agreeableness and openness to experience tend to search for meaning more than people who have high emotional stability and low agreeableness and openness (Steger, Kashdan, Sullivan, and Lorentz, 2008). If a search for meaning guided participants in the above experiments, I would expect that those with these personality traits would try to use observed patterns to explain risk more than others. Indeed, when price changes were not explicitly presented, participants whose emotional stability was lower and

agreeableness was higher did tend to judge investment risk more in accordance with the Hurst exponents of the price graphs.

Our results do not follow from Weber, Siebenmorgen, and Weber's (2005) 'risk-as-feelings' hypothesis. According to their approach, different communication methods elicit different emotions and these, in turn, trigger different assessment of different degrees of risk. For instance, they argued that providing participants with company names in addition to other data types affects risk assessment through the valence of participants' emotions towards the company. However, here the data indicate that presentation of information can affect risk assessment beyond the additional information it provides: it can cater for people's need of validation of the hypothesis they construct about risk.

To conclude, the results are in line with Mandelbrot's and Hudson's view (2004) that people use their sensitivity to fractal characteristics of price graphs to assess financial risk.

However, they appear to need validation of their interpretation of these characteristics. This validation can be provided by explicit presentation of price change information. In other words, people's need for meaning has a role in guiding their risk assessments. In particular, different communication patterns can emphasise information relevant to people's conjectures about the nature of financial risk, and thus serve as validation cues.

### *Limitations*

Online experiments do not allow verification of the identities of participants. Thus, for example, I could not ensure that participants who declared that they were financial analysts were indeed financial analysts. Though recent studies suggested that due to the Internet, a large percentage of traders are lay people (Barber and Odean, 2008; Muradoglu and Harvey, 2012), it would be important to replicate the results using a larger number of experts.

Prices in the experiments were not updated in real time; participants were presented with static price graphs. Real-life situations, involving a constant stream of prices and news items

pose higher cognitive demands on investors and hence might alter their risk perception. It would be useful to study risk perception in dynamical settings. This is what I do in Chapter 5. Next, I turn to discuss financial forecasts.

# **Chapter 4: Judgmental forecasting from fractal time series: The effect of task instructions, individual differences, and expertise on noise imitation**

In this chapter, I examine the way people make forecasts from fractal time series, and, in particular, the effects of task instructions, personality, sense of power, and expertise on noise imitation. In particular, I am interested in factors that could reduce noise imitation.

## **Experiment 1**

Experiment 1 was designed to examine the following hypotheses:

H<sub>3,1</sub>: People forecast from fractal series in a way that suggests that they perceive series with  $H < 0.5$  as noisier than series with  $H > 0.5$  and that they attempt to imitate this noise in their sequence of forecasts in all examined ranges of Hurst exponents.

H<sub>3,2</sub>: The amount of added noise, as measured by the local steepness of the forecasts and by the number of forecast extremal points, is correlated with the number of points that participants choose to forecast.

H<sub>3,3</sub>: Imitation of noise increases with conscientiousness but decrease with extraversion.

I presented participants with a sequence of nine simulated fractal price graphs and three real asset price graphs. They made forecasts from these time series. There were two experimental conditions ('no limit' and 'up to 4 points'). In both conditions, the number of forecast points

participants were asked to provide was not fixed. However, in the ‘no limit’ condition, participants could add as little or as many points as they wanted, whereas in the ‘up to 4 points’ condition, the number of required points was limited to four. At the end of the experiment, participants completed a personality and view questionnaire. The Hurst exponent of the graphs was the manipulated variable.

### *Method*

*Participants* In the ‘no limit’ condition there were 37 participants (25 women, 12 men). Their average age was 24.7 years. In the ‘up to 4 point’ condition there were 33 participants (18 Women, 15 men). Their average age was 24.18 years. All participants were recruited through the departmental subject pool. They were paid the standard participation fee (£3).

*Stimulus materials* A set of 54 simulated fractal price series, comprising six sets of nine graphs with Hurst exponents ranging from 0.1 to 0.9 in 0.1 increments, were generated using the spectral method described by Saupe (Peitgen and Saupe, 1988). The data series in all presented graphs comprised a single period produced by the generating algorithm. A further 18 real financial time-series were selected from data available at <http://finance.yahoo.com/> as described in Chapter 1. These series were also divided into three sets, each comprising six series having a low ( $H < .49$ ), a medium ( $0.5 < H < 0.56$ ), and a high ( $0.57 < H < 0.7$ ) Hurst exponent. All simulated graphs were normalised to the same interval ([1, 9]).

Participants completed the TIPI Big Five personality questionnaire (Gosling et al, 2003). To assess their views about the morality of the world and the people in it, they also rated on a seven-point scale from strongly disagree (1) to strongly agree (7) their beliefs that the world is fair/just, that it is corrupt/cruel, that people are trustworthy/decent, and that they are immoral/sinful. Finally, to assess their views about the predictability of the world and the people in it, they rated on a seven-point scale from strongly disagree (1) to strongly agree (7) their beliefs that the world is random/arbitrary, that it is organised/deterministic, that people

are unreasonable/irrational, and that they are thoughtful/predictable. For all these questionnaires, reverse scoring was applied to questions where it was appropriate.

*Design.* For each participant, nine artificial graphs from each of six sets of nine artificial graphs, and three real-life price graphs from each of the three sets of six real series were chosen randomly. The simulated series (presented in random order) were followed by the real series (presented in random order). Series were presented graphically. Participants added points to the right-hand side of each graph to make their forecasts. As they did so, their forecast points were connected by lines. An additional line connected their first forecast point with the last data point. Participants could edit their predictions by changing the location of points or deleting them. The interval between which predictions were made was bounded by red and green vertical lines. Figure 4.1 shows a typical task window from the experiment.

*Procedure* The experiment comprised four stages. First, to familiarise participants with the forecasting task, they practised making forecasts from three series. Second, they made forecasts from the nine simulated series. Third, they made forecasts from the three real series. Fourth, they completed the TIPI and world views questionnaires.

Participants were told that they would be presented with graphs of prices of different commodities and then be asked to look at them carefully to predict the prices for the required period, and to answer questions about their predictions. They were also told that there would be a short list of self-ratings for them to complete at the end of the experiment. In the 'no limit condition', detailed instructions for forecasting the simulated series then continued as follows: "The data in the graph refers to the first 63 days of the given period. You are asked to give your predictions for the period from day 63 to day 82.

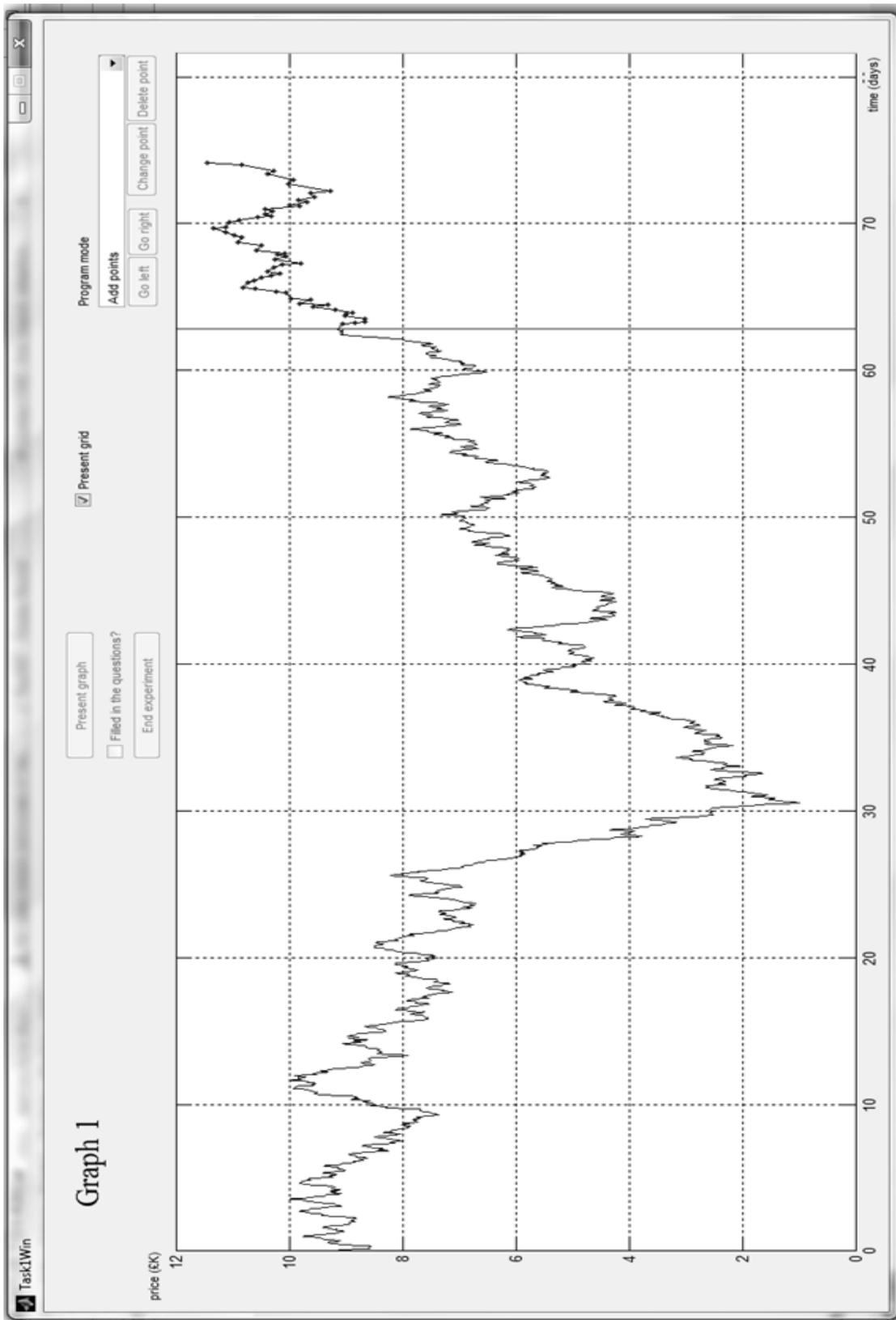


Figure 4.1 Prediction program main window. The data are presented on the left of the line at  $t = 63$ [days], and a participant's prediction points are on its right.

In order to complete your predictions, please add points to the graph in the area between the red and green vertical lines. Adding prediction points is done by pressing the left button of the mouse at the area between the red and green lines. The last point must be on (or very close to) the green line. You can add as many points as you consider appropriate.”

Instructions for the ‘up to 4 points’ condition were similar. However, participants were instructed as follows: “Please forecast from the data series by placing points on the graph in the most likely positions in which they would appear. Please add up to 4 points for each graph.” These instructions were printed in font larger than the first lines.

In both conditions, if participants asked for clarification about how many forecasts to make, they were told to add as many or as few as they wished. Instructions for the real series were similar except that data series extended over days 1-200 and the interval over which forecasts could be made covered days 200-250.

### *Results*

In the ‘no limit’ condition, most participants produced small scale-fluctuations in their sequences of forecasts, indicating that they were attempting to imitate the ‘noise’ in the data series. Examples are shown in Figure 4.2. There were, however, a few participants who appeared not to imitate the ‘noise’. Examples of this type of behaviour are shown in Figure 4.3. A qualitative analysis revealed that predictions of 32 participants exhibited noise imitation whereas predictions of five participants did not. In the ‘up to 4 point’ condition, forecasts similar to those presented in Figure 4.3 were obtained.

*Measuring imitation of ‘noise’ in fractal series* If people imitate noise when they make forecasts from fractal graphs, then they should produce series with similar Hurst exponents to those used to generate the series. However, forecast sequences that people produced were too short to allow  $H$  to be reliably estimated. Hence, I used proxy measurements to assess ‘noise’ in forecast sequences.

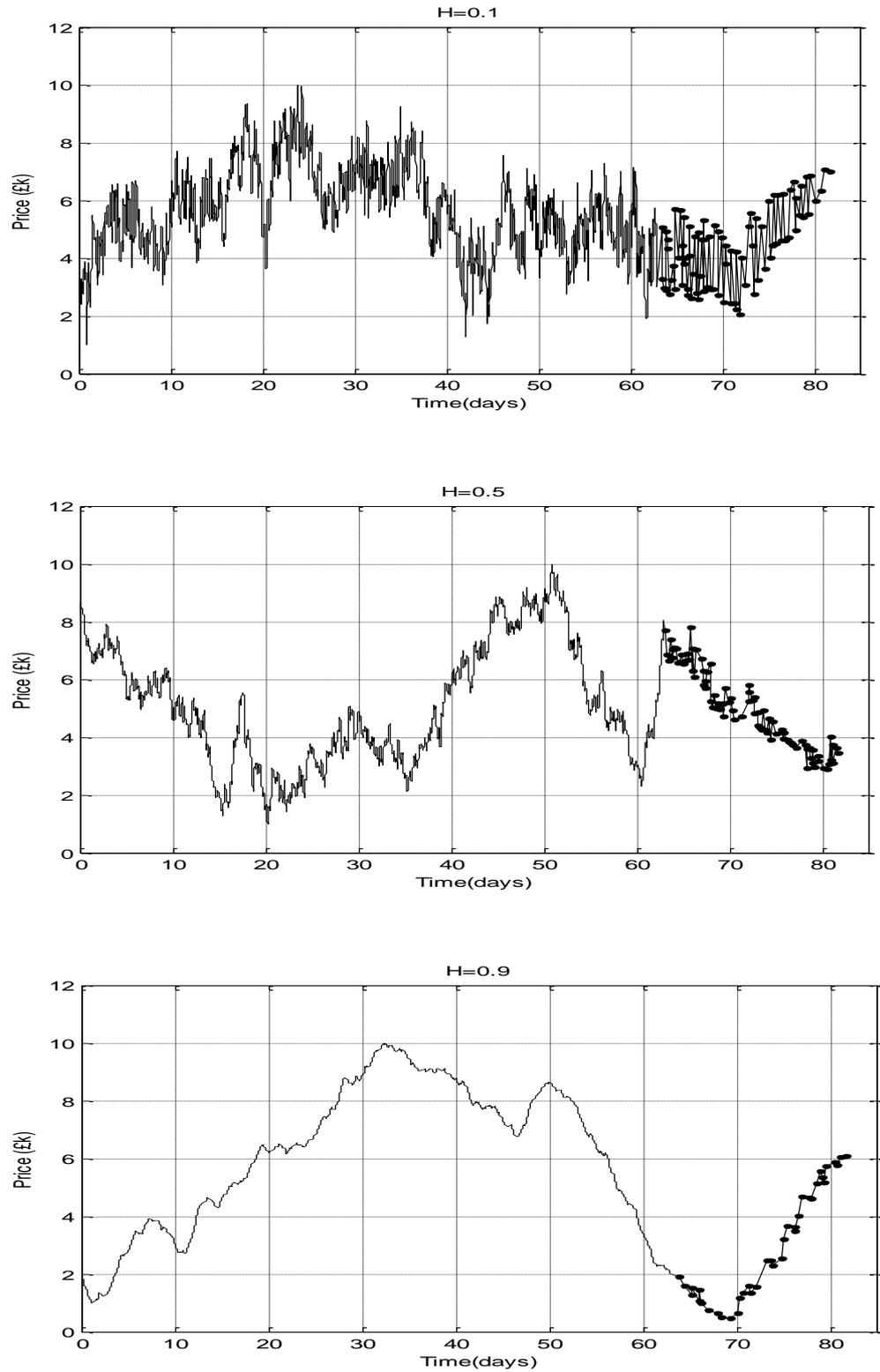


Figure 4.2 A participant's predictions (dots connected by a line) and data (line) for graphs with  $H = 0.1, 0.5, 0.9$ . This participant appears to have imitated noise.

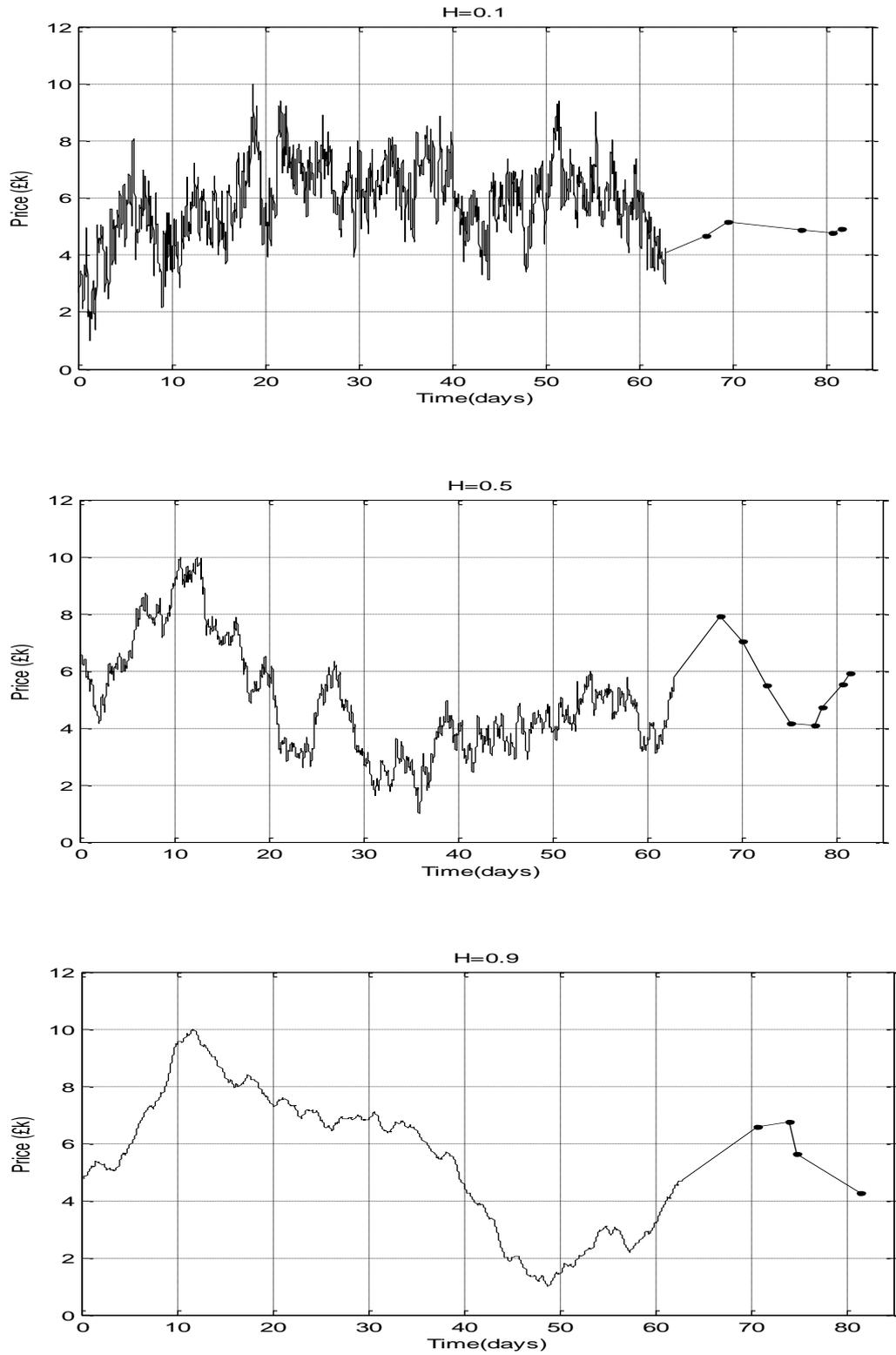


Figure 4.3 A participant's predictions (dotted line) and data (line) for graphs with  $H = 0.1$ ,  $0.5$ ,  $0.9$ .

For the primary measurement, I extracted the average absolute value of the local gradient between successive forecasts for each series seen by each forecaster. Higher values of this measure are associated with forecast sequences that look noisier and are more jagged. The same measure was used to assess 'noise' level of the given data.

In both conditions, the number of forecasts that people had to make was left unspecified I reasoned that those who wished to imitate the 'signal' without the 'noise' would need fewer forecasts to describe the future trajectory of the series than those who wished to imitate both 'signal' and 'noise'. Hence, number of forecasts provided a secondary measure of the 'noise' added to forecasts.

I also measured the number of maxima and minima (i.e. reversals in direction) in the forecast sequence. As Figure 1 shows, there tend to be more reversals as the Hurst coefficient decreases (because of increasing negative autocorrelation) and this, according to Gildea et al (1993), is interpreted as higher 'noise'. Hence, number of maxima and minima provided another secondary measure of level of noise added to forecasts.

To measure 'noise' imitation, these three measures were correlated with the average absolute value of the local gradient of presented data series and with the Hurst exponent.

*Inclusion criteria* In the 'no limit' condition, the primary measure of noise imitation (the correlation between mean absolute value of the gradient in the forecast sequence and the Hurst exponent of the data series) yielded two participants whose imitation level differed from the average by more than two standard deviations. They were excluded from the analysis. Three additional participants were excluded because regression of the above correlation on to the five personality variables produced Cook's distances (Cook, 1977) which were more than two standard deviations larger than those of the average for the rest of the group. The remaining 32 participants were entered into the analyses reported below.

In the 'up to 4 points' condition, the primary measure of noise imitation yielded one participant whose imitation level differed from the average by more than two standard

deviations. This participant was excluded from the analysis. The remaining 32 participants (288 computer generated graphs and 96 real asset series) were entered into the analyses reported below.

First I discuss the number and quality of forecasts before turning to tests of our three hypotheses.

*Number and quality of forecasts* In the ‘no limit’ condition, on average, participants added a large number of points to each graph in the simulated series (M: 40.79, std: 23.4, min: 4, max: 147) and in the real series (M: 33.09, SD: 22.51, min: 5, max: 87). Of these points, about half were maxima or minima, both in the simulated series (M: 21.42, SD: 17.02, min: 1, max: 101) and in the real series (M: 17.57, SD: 14.66, min: 0, max: 60). This proportion was sufficiently large to produce locally steep prediction gradients: for simulated series, the average of the absolute value of these gradients was 2.43 (SD: 2.19, min: 0.06, max: 12.73) and, for real series, it was 2.04 (SD: 2.10, min: 0, max: 9.79).

In the ‘up to 4 points’ condition, for computer generated graphs, participants added on average 3.84 points to each graph in the simulated series (std: 0.53, min: 2, max: 5). As Figure 4.4 shows, for most of the graphs (239/288), participants chose to add four forecast points. In spite of the instructions, participants added five points to nine graphs. For real asset price graphs, participants added, on average, 3.62 points to each graph (std: 0.53, min: 2, max: 5). As with the computer generated graphs, participants chose to add 4 forecast points to most graphs. In spite of the instructions, participants added five points to two graphs.

Of the added points, about a third were maxima or minima (M: 1.28, SD: .77, min: 0, max: 3) in the case of computer generated graphs, and more than a quarter in the case of real asset series (M: 1.01, SD: 0.84, min: 0, max: 2).

The average of the absolute value of predictions' gradients was 0.35 for both graph types (for computer generated graphs: SD: 0.28, min: 0.02, max: 2.78. For real asset series: SD=0.24, min: 0, max: 1.15).

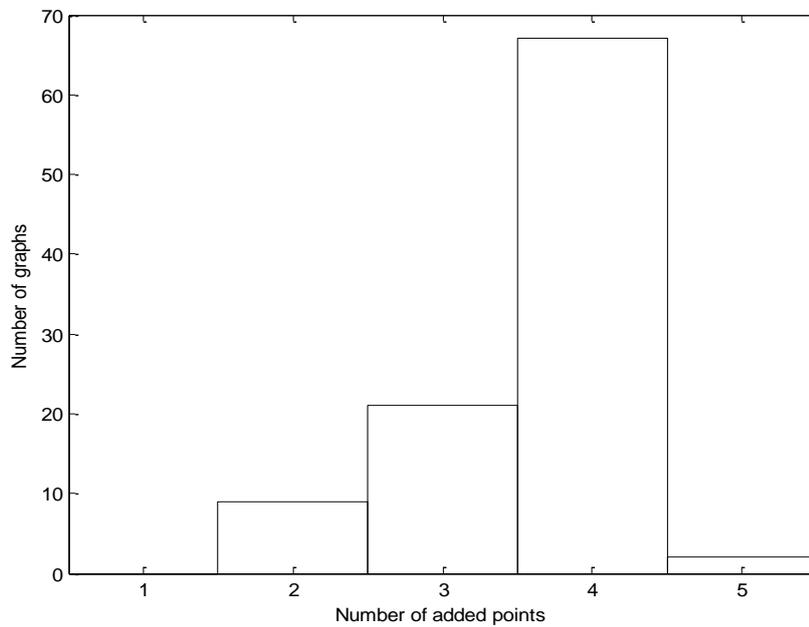
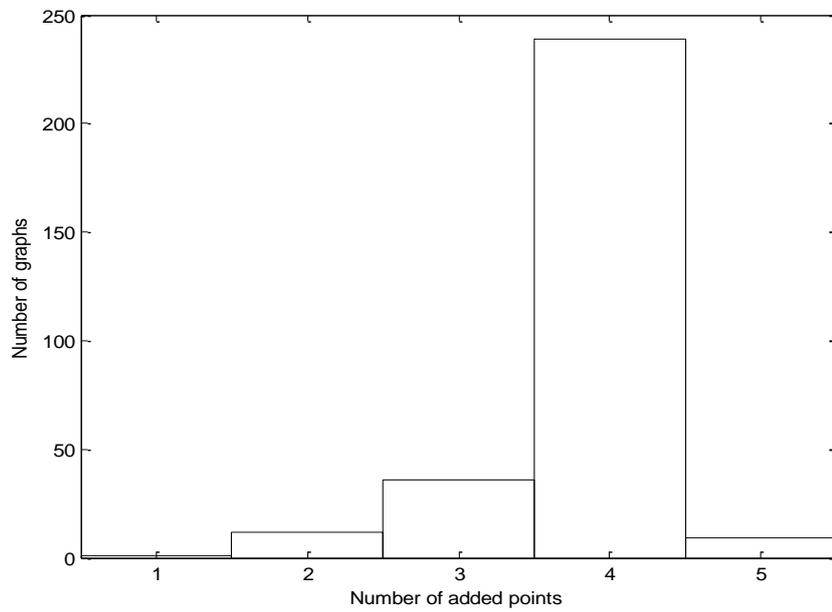


Figure 4.4 Histograms showing the distribution of added points in Experiment 4 for computer generated graphs (upper panel) and real asset price series (lower panel).

These results suggest that participants had a strong tendency to give detailed forecasts.

*Imitating 'noise' when forecasting from fractal series* In the 'no limit' condition, participants clearly attempted to imitate the 'noise' in the data series. Table 4.1 reveals that the primary measure of this, the correlation between the local gradient in the data series and the local gradient in the forecast sequence, was significant for both simulated and real series. Local gradient of the forecast sequence also correlated strongly with the Hurst exponent of the data series in both types of series. For simulated series, the secondary measures (mean number of added points, mean number of maxima and minima) also correlated with local gradients in the data series and with Hurst exponents, thereby providing further evidence of 'noise' imitation.

The correlation between Hurst exponent of the data series and local steepness of the data series was  $r = -.93$  ( $p < .01$ ) for simulated series and  $r = -.82$  ( $p < .01$ ) for real series and therefore only small differences were observed between correlations of prediction variables with Hurst exponent and local steepness of data graphs.

The results showed a significant correlation between the number of added points and the local steepness of the forecasts ( $r = .56$ ;  $p < .01$ ) and between the number of added points and the number of extremal points ( $r = .85$ ;  $p < .01$ ). This correlation supports Hypothesis  $H_{3,2}$ .

Table 4.1 Correlation between geometrical characteristics of data and prediction graphs in the ‘no limit’ condition (first panel) and in the ‘up to 4 points’ condition (the second panel) in Experiment 1.

‘No limit’ condition		Prediction’s parameters		
Data set	Data parameters	Mean number of added points	Mean number of extreme points	Local steepness
Simulated graph set	Hurst exponent	$r = -.31 (p < .01)$	$r = -.39 (p < .01)$	$r = -.58 (p < .01)$
	Local steepness	$r = .30 (p < .01)$	$r = .36 (p < .01)$	$r = .61 (p < .01)$
Real asset price graph set	Hurst exponent	Insignificant	Insignificant	$r = -.49 (p < .01)$
	Local steepness	Insignificant	Insignificant	$r = .63 (p < .01)$
‘Up to 4 points’ condition		Prediction’s parameters		
Data set	Data parameters	Mean number of added points	Mean number of extreme points	Local steepness
Computer generated graph set	Hurst exponent	Insignificant	$r = -.18 (p = .02)$	Insignificant
	Local steepness	Insignificant	$r = .21 (p < .01)$	Tendency to significance  $(r = .11, p = .54)$
Real asset price graph set	Hurst exponent	$r = -.25 (p = .01)$	Insignificant	Insignificant
	Local steepness	Insignificant	Insignificant	$r = 0.28 (p < .01)$

A t-test comparing the results of both conditions showed that in the 'up to 4 points' condition the number of added points was smaller ( $t(278) = 26.13; p < .01$ ), and as a result, the average steepness of forecasts was smaller ( $t(278) = 15.72; p < .01$ ), and the number of extremal points was smaller ( $t(278) = 19.63; p < .01$ ).

As all three noise measures in the 'up to 4 points' condition were much smaller, on average, than those obtained in the 'no limit' condition, I obtain further support for the Hypothesis  $H_{3,2}$ . I, therefore, accepted Hypothesis  $H_{3,2}$ .

*Effects of personality on forecasting* In the 'no limit' condition, for simulated series, extraversion was correlated with the mean number of added points ( $r = -.40; p < .01$ ) and with the mean number of the mean number of maxima and minima in the forecast sequence ( $r = -.36; p < .01$ ).

Taking the correlation between the mean absolute value of the local gradients in the data series and the mean absolute value of the local gradients in the forecast sequence as a measure of strength of 'noise' imitation, the data indicate that, for Hurst exponents between 0.4 and 0.6 (the range relevant to asset prices), conscientiousness was correlated with strength of noise imitation in simulated series ( $r = -.41; p = .02$ ): more conscientious people showed more evidence of imitating noise.

For real series, the same measure revealed that extraversion correlated with strength of noise imitation ( $r = .38; p = .04$ ): more extraverted participants showed less evidence of imitating noise. That might have been due to the smaller number of forecasts produced by people with higher extraversion.

In the 'up to 4 point' condition, there was no significant correlation between extraversion or conscientiousness and the mean number of added points or with the mean number of maxima and minima in the forecast sequence. Thus there was no evidence that personality influenced noise imitation level.

I, therefore, accepted Hypothesis H<sub>3,3</sub>.

### *Discussion*

There was clear evidence that two of the effects that have been reported for non-fractal series also occur with fractal series. First, in line with Gildea et al (1993), participants appear to treat differences between successive points as ‘noise’ and attempt to imitate this noise when forecasting, supporting H<sub>3,1,a</sub>. Second, in line with Harvey (1995), forecast noise level was negatively correlated with the Hurst exponent of the time series. I, therefore, accepted Hypothesis H<sub>3,1,b</sub>.

In particular, most participants added a few tens of points to each graph (though participants’ fees were independent of their performance), and this resulted in high noise levels. In fact, even when number of points was limited to four, most participants added four or five points to the graphs. This implies that participants felt a need to provide detailed forecasts.

On the other hand, noise level, as measured by the local steepness of the forecasts and by the number of extremal points, was positively correlated with the number of added points, supporting Hypothesis H<sub>3,2</sub>. Eroglu and Croxton (2010) found that biases arising from anchoring were higher in more conscientious people but lower in those who are more extraverted. I argued that, if biases arising from other heuristics show the same pattern, then conscientious people should show greater imitation of ‘noise’ in fractal series and extraverted people should show less. I did indeed find that more conscientious people imitated noise more – though this result was restricted to real series and simulated series having similar characteristics as the real series (i.e.  $0.4 < H < 0.6$ ). Also, extraversion decreased the level of noise in forecast sequences though the degree of reduction was significantly related to the ‘noise’ in the data series. I accepted Hypothesis H<sub>3,3</sub>.

## Experiment 2

Experiment 2 was designed to test Hypothesis H<sub>3,4</sub>: Sense of power affects the degree to which forecasters imitate the noise that they perceive in data series.

Though Hypothesis H<sub>3,4</sub> is non-directional, Experiment 1 indicated that more conscientious people imitate noise more. This implies that forecasters do perceive noise imitation as the appropriate way of making forecasts and that powerful people should imitate forecasts more than those who are less powerful.

Experiment 2 consisted of two main stages: a priming stage, which included a word memory test, and a combined memory test and forecasting task. I manipulated the words participants were asked to memorise so that in one condition the word list included expressions related to situations of high sense of power and, in the other, it included expressions related to situations of low sense of power. The purpose of this stage was to prime participants to hold one of these dispositions. The combined memory test and forecasting task consisted of nine trials. On each trial, participants were first asked to recall a word from a pair that had been previously memorised as part of a set of paired associates. Then, they made predictions from fractal graphs with different Hurst exponents in the same way as in Experiment 1. . Instructions given to participants were similar to those of the ‘no limit’ condition in Experiment 1.

### *Method*

*Participants* Sixty-one participants were recruited and paid in the same way as before. Their average age was 24.4 years and they comprised 40 women and 21 men. Twenty-nine participants were randomly allocated to the high power condition and the remaining 32 were allocated to the low power condition.

*Design and stimulus materials* The priming manipulation comprised a memory test. In the encoding stage, participants were asked to memorise a set of nine word pairs. Each word

pair consisted of one neutral word and one word intended to prime either a sense of power or a lack of it. The neutral words were chosen randomly from six sets obtained from an online random word generator (<http://watchout4snakes.com/creativitytools/randomword/randomwordplus.aspx>). Words in the high-power condition were powerful, strong, influential, authority, commanding, dominant, ruling, leading, and control. Those in the low power condition were powerless, weak, unimportant, insecure, obeying, subject, helpless, incapable, and small. The order in which the powerful/powerless condition words were presented was random, and so was their pairing with a neutral word. Participants were asked to spend about two minutes memorising the nine word pairs so that they could recall them later in the experiment. They then pressed a button to advance to the recall stage.

The recall stage of the memory task was combined with the forecasting task (Figure 4.5). One word of each pair was presented. It was chosen at random as either a neutral word or one from the powerful or powerless sets. Participants were asked to retrieve the word it had been paired with during encoding from a list box containing nine options. When they were wrong, they were required to correct themselves. As they could not proceed before correctly recalling the word pair, those who made more mistakes were exposed to the experimental manipulation for a longer time.

After participants had retrieved the correct word, a graphical representation of a fractal price series was presented to them in the same way as in Experiment 1. They made their forecasts in the same way as before. After they had done so, they continued to the recall stage for the next word pair, and so on. A total of nine words and nine graphs were presented. As before, graphs were chosen at random from six sets of nine graphs with  $H = 0.1, 0.2, \dots, 0.9$ , produced using Saupe's spectral algorithm (Peitgen and Saupe, 1988).

As before, I used the TIPI personality questionnaire to measure the Big Five personality traits (Gosling et al, 2003). However, in order to check the effectiveness of the sense of

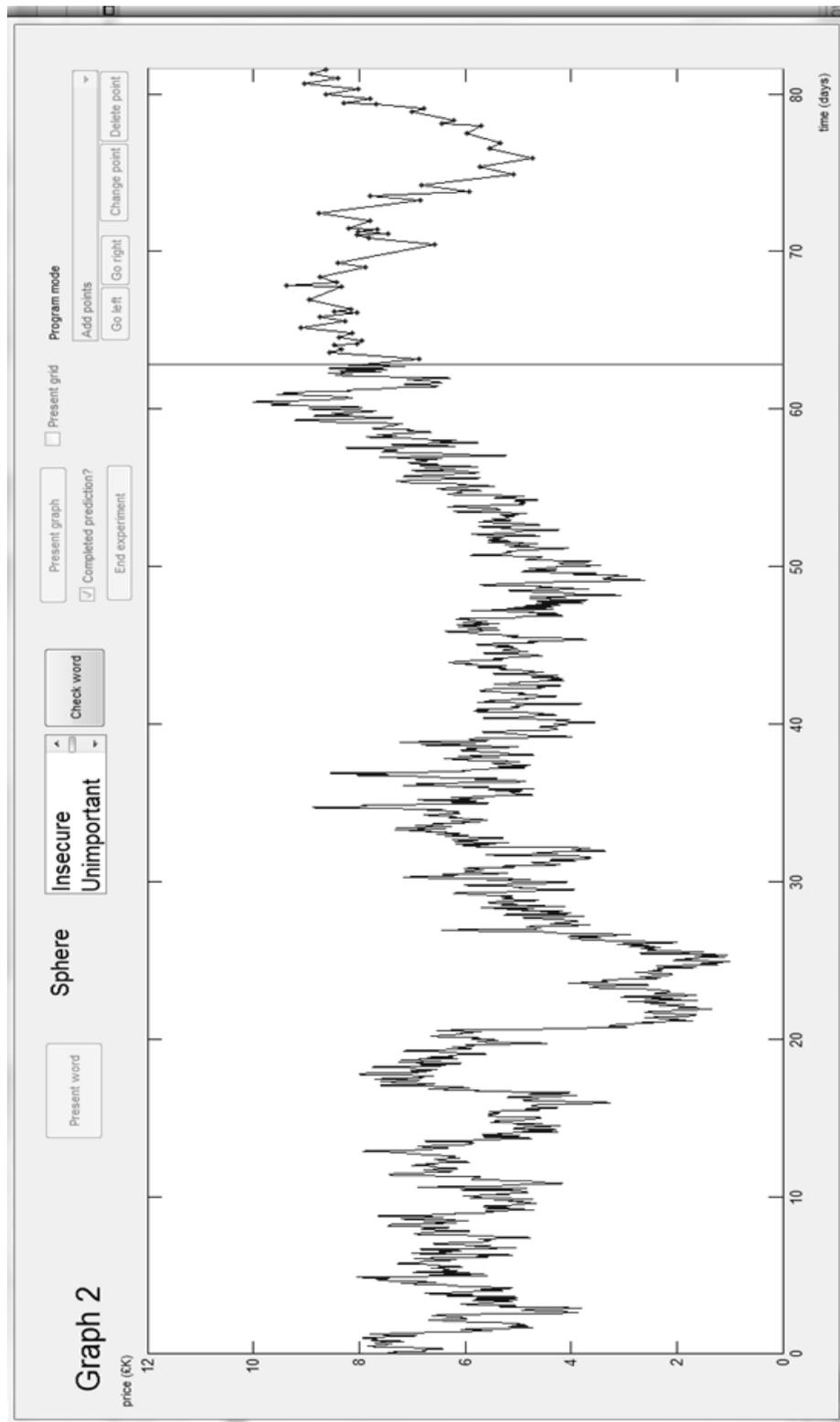


Figure 4.5 Prediction and memory test window. The figure shows one word from the neutral word list (“Sphere”) and two of the 9 words in the list box (“Insecure”, “Unimportant”) used for the low power condition.

power manipulation, I added two items referring to it: *forceful, strong* and *powerless, weak*. They were added to the TIPI as items number six and twelve.

*Procedure* Initially, participants were given three graphs from which to make forecasts in order to give them some experience to familiarise them with the task. (No memory test was combined with the forecasting task at this stage.) Then, once they had spent two minutes memorising the nine word pairs, they performed the combined memory recall and forecasting task. After that, they were given another test of their recall of the nine word pairs. Finally, they completed the personality questionnaire (including the two sense-of-power items). Instructions for forecasting were the same as those given in Experiment 1.

### *Results*

Informal inspection suggested that most participants tended to imitate the data. Typical predictions were similar to those shown in Figure 4.2.

I excluded outlying participants using the same criteria as before. This resulted in six participants being dropped from the analysis, leaving 26 in the high power condition and 29 in the low power one (a total of 495 graphs).

I discuss the number and quality of forecasts before turning to tests of our hypotheses.

*Number and quality of forecasts* As before, participants tended to make a large number of forecasts from each graph (M: 48.25, SD: 27.33, min: 3, max: 148). On average, more than half of these points were maxima or minima (M: 27.35, SD: 19.04, min: 0, max: 100).

Again, this resulted in steep gradients between predictions (M: 4.23, SD: 3.50, min: 0.55, max: 12.09).

*Imitating 'noise' when forecasting from fractal series* As can be seen from Table 4.2, all three of the measures of noise in the forecast sequence correlated significantly with both the Hurst exponent and the mean local gradient in the data series. These findings provide evidence that participants imitated the 'noise' in the series.

Table 4.2 Correlation between geometrical characteristics of data and prediction graphs in Experiment 2.

		Prediction's parameters		
Data set	Data parameters	Mean number of added points	Mean number of extremal points	Local steepness
Simulated graph set	Hurst exponent	$r = -.40$ ( $p < .01$ )	$r = -0.48$ ( $p < .01$ )	$r = -.59$ ( $p < .01$ )
	Local steepness	$r = .33$ ( $p < .01$ )	$r = 0.41$ ( $p < .01$ )	$r = 0.60$ ( $p < .01$ )

*Effects of personality on forecasting* The correlation Hurst exponent of the data series and the mean absolute value of the local gradients in the forecast sequence had an average value of  $-0.75$  (SD:  $0.22$ ), indicating that most participants produced noise in their sequence of forecasts similar to the 'noise' in the data series. Using size of this correlation as a measure of strength of 'noise' imitation, I found that strength of noise imitation increased with conscientiousness ( $r = -.38$ ,  $p < .01$ ). This replicates the result that was obtained in Experiment 1 for values of the Hurst exponent between  $0.4$  and  $0.6$ . (In the present experiment, the finding still held when values of the Hurst exponent were restricted to that range:  $r = -.30$ ,  $p = .03$ ).

*Effects of sense of power on forecasting* First, I performed a manipulation check to determine whether the priming manipulation had achieved its aims; I compared people's self-assessments on the two items referring to sense of power that had been added to the TIPI questionnaire (i.e. forceful versus powerless, strong versus weak). This showed that the mean power rating of participants in the high power condition was  $5.37$  and that that of those

in the low power condition was 4.70 ( $F(1, 54) = 4.67; p = .04$ ). This indicated that the priming manipulation was effective.

For Hurst exponents between 0.4 and 0.6, strength of noise imitation in the high power condition (M: -0.67, SD: 0.53) and in the low power condition (M: -0.52, SD: 0.54) were significantly different ( $F(1, 494) = 10.64, p < .01$ ). Participants in the high power condition imitated 'noise' in the data series more than those in the low power condition. I, therefore, accepted Hypothesis H<sub>3,4</sub>.

### *Discussion*

Results from this experiment replicated the main findings obtained in the previous one. First, various measures indicated that participants tended to imitate the 'noise' that they perceived in the data series. Second, the tendency to imitate noise was greater in conscientious people.

In addition, this experiment showed that participants with a high sense of power imitated the 'noise' they perceived in the data series more than those in a low sense of power. This is to be expected on the basis of Galinsky et al's (2008) analysis if forecasters consider 'noise' imitation as the correct way of making predictions. Together with the finding that more conscientious people imitate 'noise' more, these findings concerning the effects of sense of power imply that people do indeed consider noise imitation to be appropriate. I accepted Hypothesis H<sub>3,4</sub>.

### **Experiment 3**

Experiment 3 was designed to test Hypothesis H<sub>3,5</sub>: Noise imitation occurs in both forecasts of professionals in finance and in those of lay people.

A secondary aim of the experiment was to assess the quality of these forecasts. Therefore, I compared forecast errors of expert and non-professional groups. I was interested in the question whether financial predictions and probability estimates made by experts ("expert group") are different from those of participants who had no academic background in finance

or economics (“non-professional group”). Non-professional participants were recruited through the departmental participant pool. Participants from the expert group were recruited at a conference on financial modelling. The experiment was coordinated with the organisers of the conference.

Due to constraints resulting from the settings of the expert condition of the experiment, Experiment 3 was a pen-and-paper experiment. Participants were given graphs of prices of real assets, and were asked to make price predictions (see Figure 4.6). In addition, they were asked to assess probabilities of their predictions being correct and to fill in the Ten Item Personality Inventory (TIPI questionnaire, Gosling et al, 2003).

This study involved only a small number of participants in each group (N=13). Therefore, its results should be treated merely as an indicator of the tendencies among finance professionals and non-professional people.

### *Method*

*Participants* There were 13 participants in the expert group (one woman, 12 men). Their average age was 45.5 years. Twelve out of the 13 participants had a PhD in economics, finance, or related topics. The thirteenth participant was a final year Finance PhD student. Only 11 of the participants completed the TIPI questionnaire.

There were 13 participants in the non-professional group (9 women, 4 men). Their average age was 22.8 years. They were recruited via the local departmental participant pool website. They were paid the standard participation fee (£2).

*Stimulus materials* I employed the same real financial series as in Experiment 1. Participants completed the TIPI Big Five personality questionnaire (Gosling et al, 2003).

*Design* Each participant was presented with three graphs, one from each H range. (These ranges were the same as those used in Experiment 1.) The graphs were randomly chosen and ordered. Each graph contained 2000 points, and was presented on the axes  $0 \leq x \leq 250, 0 \leq$

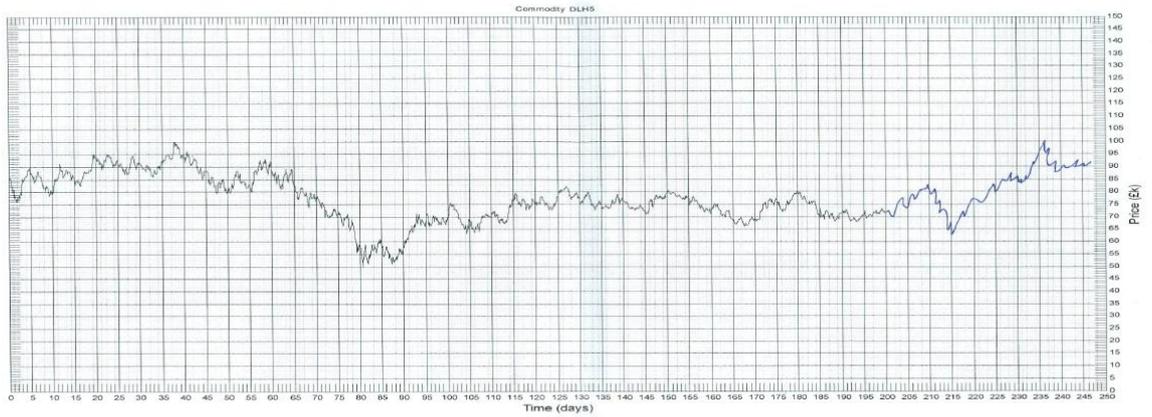
$y \leq 150$ . The y axis range was chosen to allow participants to make predictions with high gradients, as the data were bounded between 50 and 100 (£k). The 2000 data points were presented on the range  $0 \leq x \leq 200$ .

Examples for graphs of different H ranges are presented in Figure 4.6. The names of the assets were coded. The graphs were presented with fine grids to facilitate accurate extraction of points.

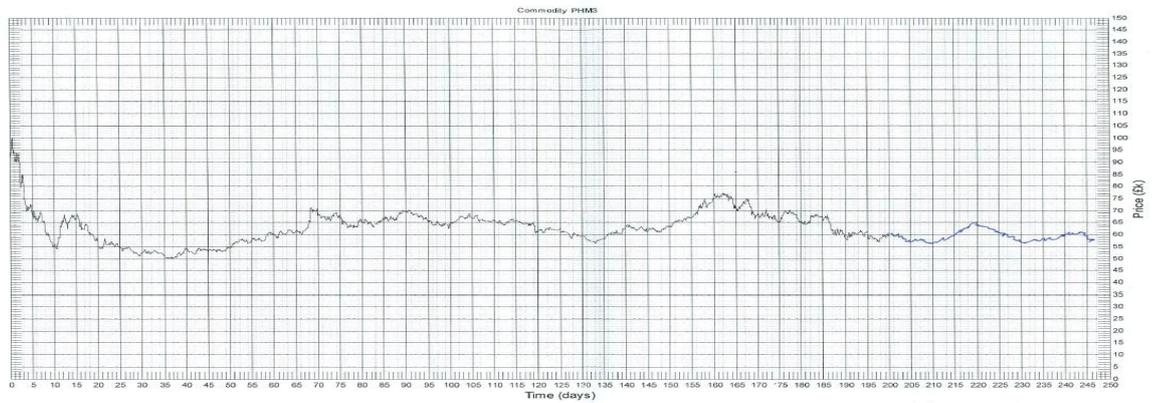
*Procedure* Participants were given a two minute presentation about the experiment instructions, after which they were handed forms containing the experimental materials. These forms consisted of three graphs of prices, a probability assessment table, and the TIPI questionnaire.

Participants were informed that they would be presented with three graphs of prices for a period of 200 days. They were asked:

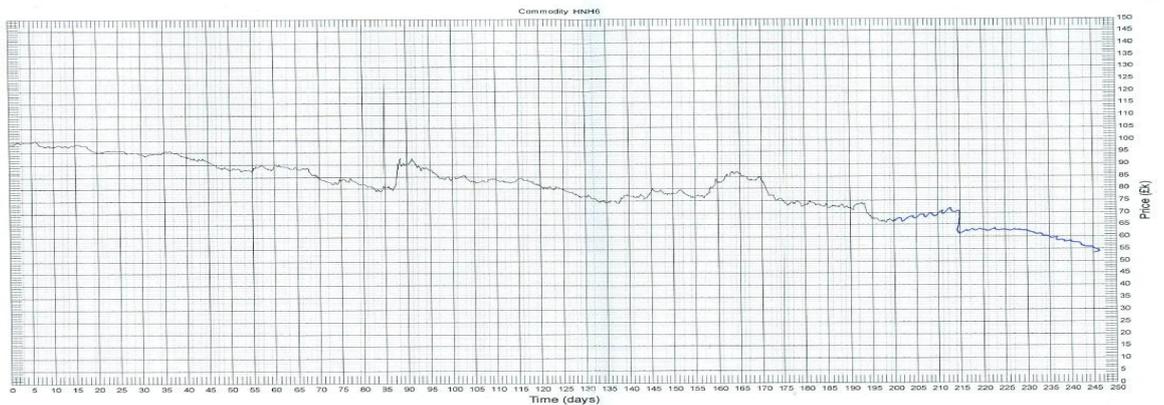
1. to look at the graphs carefully, and then predict the prices of the commodities at days 201-250 by continuing the price curve on each of the graphs,
2. to assess the probability that the actual outcome would fall within a range of  $\pm 10$  points (£1000) of their forecast for days 215, 230 and 245. These probability estimates should be expressed as a number between 0 and 1, where 0 means complete uncertainty, and 1 means certainty of 100%,
3. to indicate whether the commodity described reminds them of any familiar commodity. If yes, participants were asked to specify the name of this commodity and the approximate period depicted,
4. to complete the TIPI question list.



The probability that the actual outcome will fall within $\pm 10$ points of your forecast	Day 215	Day 230	Day 245	Is the commodity familiar to you? Yes <input type="checkbox"/> No <input checked="" type="checkbox"/> If you answered the previous question with "Yes", please specify the name of this commodity and the approximate period depicted: <u>Lufthansa</u>
	0,8	0,7	0,6	



The probability that the actual outcome will fall within $\pm 10$ points of your forecast	Day 215	Day 230	Day 245	Is the commodity familiar to you? Yes <input type="checkbox"/> No <input checked="" type="checkbox"/> If you answered the previous question with "Yes", please specify the name of this commodity and the approximate period depicted:
	0,6	0,5	0,4	



The probability that the actual outcome will fall within $\pm 10$ points of your forecast	Day 215	Day 230	Day 245	Is the commodity familiar to you? Yes <input type="checkbox"/> No <input checked="" type="checkbox"/> If you answered the previous question with "Yes", please specify the name of this commodity and the approximate period depicted:
	0,5	0,4	0,3	

Figure 4.6 Data, predictions and probability estimates made by a participant from the expert group in Experiment 3, for graphs with low (first panel), medium (second panel), and high (third panel) Hurst exponents.

## *Results*

Most participants in the expert group (10/13) and all participants in the non-professional group produced graphs with small fluctuations, suggesting an attempt to imitate the noise of the data. However, three participants from the expert group continued the price graphs by sketching a constant or a trend line. In the following sections, I denote the subgroup of the expert group, consisting of the three participants who sketched a constant or a trend line “the E3 group”, and the remaining participants of the expert “the E10 group”. Examples for typical predictions of a participant from the E10 group are given in Figure 4.6.

*Quality of forecasts* I sampled a point every 0.5 day from each of the 78 resultant graphs (2\*3\*13). This sampling procedure produced 100 points when participants made their predictions for the whole required period. However, not all graphs contained forecasts up to day 250 (see Figures 4.6). The minimum number of points sampled from a single graph was 89 points for the expert group, and 94 for the non-professional group.

On average, participants from the expert group depicted 25.49 extremum (minimum or maximum) points (SD: 17.07, min: 0, max: 59), and the non-expert group depicted 43.08 extremum points (SD: 17.72, min: 10, max: 78). The resultant graphs were locally steep: for the expert group, the average of the absolute value of the local gradients between predictions was 1.06 (SD: 0.87, min: 0, max: 3.39) and for the non-expert group it was 1.74 (SD: 0.93, min: 0.50, max: 4.39). The average number of extremum points of participants from the E10 group, who did not sketch constant or trend line (N = 10), was 33.13 (std: 19.94, min: 15, max: 59), and the average of the absolute value of their prediction gradients was 1.35 (std: 0.77, min: 0.44, max: 3.39). A t-test failed to find a significant difference between the average steepness of this expert sub-group and the non-professional subgroup ( $t(29) = 1.62$ ;  $p = .115$ ), though a significant difference was found between the number of extremum points of the expert group and the non-professional group ( $t(29) = 2.5$ ;  $p = .02$ ).

I calculated the root mean squared error scores relative to the actual outcome of the real series over the forecast interval for each forecast series and for naive forecasts, consisting of the constant value of the last presented data point over all forecast horizons. The averages of raw error scores and normalised error scores (raw error divided by the range of prices in the data series) for the expert group, the non-professional group, the naive forecasts, and E3 are presented in Table 4.3. As can be seen, in general, the average errors are high. Furthermore, as expected, a repeated measure ANOVA showed a main significant effect of forecast horizon ( $F(2, 76) = 73.48, p < .001$ ): forecasts became worse as its horizon was larger. However, there was no significant effect of the forecaster group variable.

The averages of the normalized errors of participants from the E3 group were smaller than those of the naive forecaster. However, due to the small number of members in this group, no further statistical analysis could be made.

Table 4.3 Average prediction errors for each. prediction horizon in Experiment 3

Error measure	Group						
	Expert group		Non-professional		Naive forecaster		E10
	Mean	Std. deviation	Mean	Std. deviation	Mean	Std. deviation	Mean
Error <sub>215</sub>	5.47	2.26	6.30	1.78	4.17	1.80	5.08
Error <sub>230</sub>	8.60	2.97	8.88	2.11	6.87	3.01	6.40
Error <sub>245</sub>	10.25	3.39	10.50	2.92	8.93	3.28	6.89
NormError <sub>215</sub>	0.13	0.05	0.15	0.04	0.11	0.05	0.11
NormError <sub>230</sub>	0.21	0.07	0.22	0.07	0.18	0.10	0.14
NormError <sub>245</sub>	0.27	0.10	0.27	0.11	0.24	0.15	0.15

Assessed probabilities of forecasts being within £1,000 of the outcome decreased or remained constant as forecast horizon increased in 84.6% (33/39) of the series for the expert group and in 74.4% (29/39) of the series in the non-professional group. The analysis failed to find a significant difference in the percentage of probability estimates which decreased as forecast horizon increased between the groups ( $\chi^2(1, 78) = 1.26; p = .26$ ).

*'Noise' imitation* The primary measure for noise imitation was the correlation between the absolute values of the local gradient (local steepness) of the data series and the local steepness of the forecast sequence.

Table 4.4 shows that for participants in the E10 and the non-professional groups, these correlations were highly significant. The secondary measure for noise imitation was the correlation between Hurst exponents of the data graph and the local steepness of the forecasts. For the expert, E10 and the non-professional groups, highly significant correlations were obtained for the secondary measure as well. These results suggest that participants attempted to imitate 'noise'. I accepted Hypothesis H<sub>3,5</sub>.

The correlation between the Hurst exponent and local steepness of the data was  $r = -.90$  ( $p < .001$ ) for graphs presented to the expert group, and  $r = -.89$  ( $p < .001$ ) for graphs presented to the non-professional group. Therefore, only small differences between the groups were observed between correlations of forecast variables with Hurst exponent and local steepness of data graphs.

*Effects of personality on forecasting* As before, the measures for strength of noise imitation were the correlation between Hurst exponent of the local steepness of the data series, and the local steepness of the forecast series. As local steepness of forecasts of members of E3 was constant, strength of noise imitation could not be calculated for members of the E3 group. Therefore this section concerns analysis of the results of E10 and the non-professional groups.

In spite of the small number of participants in E10, there was a significant negative correlation between conscientiousness and the strength of noise imitation, defined as correlation of local steepness of the forecasts with the H exponent of the data series ( $r = -.71, p = .03$ ). This negative correlation indicates that experts who were more conscientiousness tended to imitate ‘noise’ in the data series more.

In addition, there were significant positive correlations between agreeableness, conscientiousness, and emotional stability, and the number of extremum points ( $r = .58, p = .001$ ;  $r = .46, p = .02$ ;  $r = .49, p = .010$  respectively). The more experts were agreeable, conscientious, and emotional stable, the more ‘dramatic’ their forecasts appeared.

Table 4.4 Correlation between geometrical characteristics of data and prediction graphs in Experiment 3

Group	Data parameters	Correlation between data parameter and local steepness of forecasts
Expert group	Hurst exponent	$r = -.33 (p = .04)$
	Local steepness	Insignificant
E10 group	Hurst exponent	$r = -.46 (p = .01)$
	Local steepness	$r = .50 (p = .005)$
Non-professional group	Hurst exponent	$r = -.55 (p < .01)$
	Local steepness	$r = .66 (p < .01)$

In the non-professional group, there were significant correlations between emotional stability and the local steepness of the forecasts ( $r = .36, p = 0.03$ ), and emotional stability, and the number of extremum points ( $r = .35, p = .03$ ).

### *Discussion*

Experiment 3 showed that noise added to forecasts was correlated with Hurst exponent of the presented data series. This finding is in line with that of Gilden et al (1993). However, here it was shown that it extends to experts' forecasts as well.

Forecasts of professionals and lay people share many features. In particular, most participants in both groups imitated the noise in the data series. There were no significant differences between forecast errors of lay people, professionals who imitated data's noise, and naive forecasts. I accepted Hypothesis H<sub>3,5</sub>.

On the other hand, there were a few differences between forecasts of experts and non-experts. In general, experts tended to imitate noise less than lay people, and their noise imitation level was correlated with self-rating of conscientiousness (unlike that of the non-expert group).

## **Conclusions**

Evidence is accumulating that price series have a fractal structure (Mandelbrot and Hudson, 2004; Coen and Torluccio, 2012; Onali and Goddard, 2011; Bianchi et al, 2010; Hai-Chin and Ming-Chang, 2004). Unlike the series that have previously been studied by those interested in judgmental forecasting, fractal series cannot be naturally decomposed into signal and noise. Despite this, Gilden et al (1993) have argued from results of their studies on the discrimination of fractal contours that people analyse fractals as if they can be decomposed in this way: changes in successive prices (related to autocorrelation) are treated as if they are noise. This interpretation is consistent with the results reported in Chapter 2.

If Gildea et al (1993) are correct, previous findings concerning judgmental forecasting from series that can be decomposed into signal and noise components should generalise to fractal series. In particular, noise in a sequence of forecasts should increase with the noise in the data series (Harvey, 1995; Harvey et al, 1997). All of the experiments reported here produced findings that fulfilled these expectations: the mean absolute size of local gradients in the forecast sequence increased with the mean absolute size of the local gradients in the data series and final forecasts were higher than initial ones even though there was no overall trend in data series.

Recent reports have indicated that personality traits affect traders' performance (Frijns et al, 2008; Kapteyn and Teppa, 2011; Robin and Stráznicka, 2012; Fenton-O'Creevy et al, 2012; Fenton-O'Creevy et al, 2011). Eroglu and Croxton (2008) attributed effects that they obtained to people being more or less susceptible to biases arising from use of anchoring heuristics (Tversky and Kahneman, 1974). Harvey (1995) argued that noise imitation is a bias that arises from use of another of the three heuristics identified by Tversky and Kahneman (1974): representativeness. Thus, if the effects of personality obtained by Eroglu and Croxton (2008) apply not just to biases arising from anchoring but also to biases arising from use of other heuristics, noise imitation effects should be more evident in those who are conscientious and less evident in those who are extraverted. I did indeed find that sequences of forecasts made by more conscientious people showed stronger evidence of imitation of 'noise' in the data series in the experiments reported here. Also, for the real series used in Experiment 1, I found that more extraverted people showed weaker evidence of imitation of the 'noise' in the data series.

Individual differences in forecasting behaviour may also be produced by differences in temporary dispositions. Of these, a sense of power is thought to be particularly important on the trading floor (Hassoun, 2005). I used a priming task to induce either a sense of power or of powerlessness and found that those who felt more powerful showed a stronger tendency to imitate the 'noise' in the data series. This finding can be seen as consistent with Galinsky

et al's (2008) analysis of the effects of a sense of power if I assume that forecasters consider 'noise' imitation as the correct way of making predictions. The fact that more conscientious people show a greater tendency to imitate 'noise' suggests that they do.

Next, I showed that many of the results obtained for lay people can be generalized to experts. The expert sample consisted of 12 people who had a PhD in Finance or Economics, and one Finance PhD student. Most experts worked as professors in finance, economy or related topics. Nevertheless, when asked to make forecasts from graphs depicting the price series of real assets, 10 out of 13 of them produced forecasts which included noise. Noise was significantly correlated with the Hurst exponent of the given data graphs. Furthermore, the average accuracy of the experts' forecasts, as measured with respect to the historical evolution of prices, could not be distinguished from that of participants in the non-expert group and it was lower than that of a naive forecaster. Only three experts, whose forecasts depicted a straight line showed accuracy that was higher than that of the naive forecaster.

Generalizing the findings of Experiment 1, Experiment 3 showed that, among experts, higher degrees of noise imitation were associated with higher conscientiousness. A large percentage of traders use technical analysis techniques, or define themselves as technical analysts (Cheung and Chinn, 2001; Taylor and Allen, 1992) and so this could have important implications.

To conclude, the results indicate that people have a tendency to elaborate when performing forecasting tasks. Even though participants were not asked to provide a specific number of forecasts (and could make a single point forecast had they wanted to), they chose to make many of them. This was independent on the experimental design and whether the task was computer-based or used pen-and-paper. Noise imitation was found in both lay people and experts. In most experiments, it did not increase with agreeableness, suggesting that participants were not motivated by the need to comply with the way they might have perceived the experimenter's goals. On the other hand, it increased with participants'

conscientiousness and sense of power. This might indicate that they imitated noise because they thought that this was the correct way to make forecasts from the graphs.

### *Limitations*

I attempted to avoid encouraging participants to imitate noise through our experiments' instructions. For instance, I asked the expert group "to look at the graphs carefully, and then predict the prices of the commodities at days 201-250 by continuing the price curve on each of the graphs". Furthermore, Harvey et al (1997) showed that noise imitation occurred even when instructions were very detailed. Nevertheless, it is important to continue to examine the wording chosen for the task. For example, it would be interesting to examine how much noise imitation can be reduced by informing people about it.

I used TIPI questionnaire to assess participants' personality traits. TIPI is a standardised questionnaire, but it is short and less accurate than longer personality questionnaires that measure the Big Five personality traits. Gosling et al (2003) recommend using these longer versions when time permits. The additional power resulting from this approach may reveal additional influences of personality on forecasting behaviour.

The results reported here prompt the question as to whether, apart from imitating the perceived noise component of the graphs, people also imitate its perceived signal. However, the experiments that I described here were not designed to answer that. In particular, I did not investigate factors that determine the characteristics of any signal that people include in their forecast sequence. However, this issue is touched on in Chapter 6 where I examine the size of the averaging window that people consider appropriate to apply to financial series in order to make financial forecasts from them. In the next chapter, I study the way people make forecasts when news is given in addition to price graphs.

## **Chapter 5: The effects of news valence, price trend and individual differences on financial behaviour**

*“I made my money by selling too soon” (Bernard Baruch, cited in Katsenelson, 2007, page 252).*

*“When good news about the market hits the front page of the New York Times, sell” (Bernard Baruch, cited in Hill, Franklin, Clason and Mackay, 2009, page 195).*

**Remark:** The experiments described in this chapter were performed in collaboration with Bryan Chan.

In this chapter, I examine the way that people incorporate news items and price graphs in order to make financial decisions. In particular, I characterise the conditions in which people prefer attributing more weight to news than to price graphs. I study decision times in each of these conditions. I also investigate the effects of culture and personality traits on financial decisions. Finally, I examine the way people make forecasts from the data and use their forecasts to decide whether to buy, sell, or hold assets.

### **Experiment 1**

In Experiment 1 I investigated the following hypotheses:

H<sub>4,1</sub>: people choose to base their trading strategy on news more than they do on price graphs.

H<sub>4,2</sub>: people track prices more and show more active trading (buying or selling rather than holding their assets) in non-conflicting conditions than in conflicting ones.

H<sub>4,3</sub>: people sell more assets when the news is bad than they buy when it is good.

H<sub>4,4</sub>: trading latency is shorter when uncertainty is higher, that is, when there is an inconsistency between news valence and price trends.

H<sub>4,5</sub>: the effect of news on trading latencies is stronger than that of the price trend, and trading latency is shorter when news is bad.

H<sub>4,6</sub>: people from Western culture react to news more than people from Eastern countries in consistent conditions (good news with positive price trend or bad news with negative trend). People from Eastern Asian countries react to news more than people from Western countries in inconsistent conditions (good news with negative trend or bad news with positive trend)

H<sub>4,7a</sub>: people from Eastern culture exhibit longer trading latencies.

H<sub>4,7b</sub>: people from Eastern culture have higher degrees of dispersion in their returns.

H<sub>4,8</sub>: people more open to experience have shorter trading latencies.

I presented participants with a sequence of 12 graphs of real asset prices. Participants were told that they would be initially endowed with one share of each of the assets and a virtual sum of money large enough to buy one additional share of each of those assets.

Graphs of each asset were updated gradually so that a new point was added to the graphs every 0.2 seconds. After each block of 20 points, participants were asked to decide whether to buy, sell, or hold their asset. After every block of 40 points, participants were presented with a news item. The direction of the trends in the price graphs and valence of news were manipulated to form a two (positive versus negative trend) by two (good versus bad news) within-participant design. U-shaped and inverse-U-shaped graphs were added as fillers to mask the rationale of the experiment.

I recorded the number of shares that participants had in each of the experimental conditions after deciding to buy another share of each asset, sell their share, or hold their share. I refer

to this variable as the *final share number*. I also recorded the number of points that were displayed before decisions to buy or sell were made. I refer to this as *decision latency*.

### *Method*

*Participants* Sixty people (28 men and 32 women) acted as participants. Their average age was 25 years. All participants were recruited through a participant recruitment website at University College London. Participants from Western and Eastern cultures were recruited separately to ensure that there were equal numbers in the two groups.

The Western group comprised thirty people (17 men and 13 women) with an average age of 29 years. The majority of them had an undergraduate degree or above and came from a wide range of occupational backgrounds (ranging from students to a retired engineer).

The Eastern group comprised thirty people (11 men and 19 women) with an average age of 21 years. Twenty of these participants were from Hong Kong, nine from China and one from Singapore. All of them had spent most of their lives in their country of origin. Most of them were undergraduate or postgraduate students.

All participants were paid a fixed fee of £2.00. An additional £2 was available as performance-related pay: if the value of a participant's portfolio at the end of the experiment was at least £15 more than its initial value, an additional £1 was paid: if the value of that portfolio was at least £30 more than its initial value, an additional £2 was paid.

*Stimulus materials* I used the real-life time series documented in Chapter 1, Part III.

Eighteen price series were downloaded from Yahoo! Finance (<http://finance.yahoo.com/>).

Each series consisted of 2500 close prices. To avoid confounding variables, I chose six time series with a Hurst exponent that was close to a constant and in the interval  $[0.50, 0.56]^3$ .

The Hurst exponent of time series is correlated with variables such as the series oscillation, variance and autocorrelation. I then chose 40 subsets of 220 consecutive elements from the

---

<sup>3</sup> This interval ensured that successive price changes were independent, thereby making series consistent with the random walk behaviour expected from the EMH. This allows the results to be compared with predictions derived from that approach.

original series. Each group of 10 subsets had a positive average trend, a negative average trend, a U-shape, or an inverse U-shape. The criterion for selection as a U-shape or inverse U-shape subsets was that the first and last points were not different by more than half a point. I then reflected subsets with negative and positive average trends about day 110 to create 10 more subsets of positive and negative trends, respectively. U-shaped and inverse U-shaped subsets were reflected about the time axis. Finally, all 80 resultant series were normalized to fit the same price range of [£2, £10]. This procedure for the construction of the series ensured that the average trend of the graphs in the positive and negative trend sets was the same.

Presented news items were based on real items, published on BBC

(<http://www.bbc.co.uk/news/>) and Yahoo! Finance (<http://finance.yahoo.com/>). News was of two types, good and bad. Each news item was formulated as a single sentence. A total of 30 news items evaluated as good were downloaded. Bad news was generated from the good news by inverting its meaning. For instance, in order to generate a bad news item from the good news item “Company awarded \$115 Million in Patent-Infringement lawsuit”, I transformed it into “Company asked to pay \$115 Million in Patent-Infringement lawsuit”.

Participants’ personality traits were assessed using the Ten Item Personality Inventory (TIPI), a standardized personality questionnaire (Gosling et al, 2003). The TIPI measures the Big-Five personality traits: Extraversion, Agreeableness, Conscientiousness, Emotional stability, and Openness to experience.

*Design* Twelve graphs were chosen at random for each participant, four from the positive trend group, four from the negative trend group, two from the U-shaped group, and two from the inverse U-shaped group. For each graph, five news items, which were either all good or all bad, were chosen and randomly assigned to time points. News items were sampled without repetition, so that each news item was viewed by each participant only once. Two of the graphs with the positive trend were assigned to good news sets and two of them to bad news sets. Similar choices were made for the graphs with the negative trend, resulting in a

two (positive or negative trend) by two (good or bad news valence) design. Every condition was tested using two graphs per participant.

The purpose of the U-shaped series and inverse U-shaped series was to mask the manipulations, and so participants' results in these conditions were not analyzed. However, each of them was also paired with either good or bad news group.

Graphs and news were presented using a graphic user interface program written in Matlab.

Figure 5.1 shows a typical task window from the experiment.

*Procedure* The experiment comprised three stages. First, in a familiarization task, participants were asked to make financial decisions with respect to three practice graphs. Results of familiarization task were not taken into account in the analysis. Second, they were asked to make financial decisions with respect to the randomly chosen 12 experimental graphs. Third, they were asked to complete the TIPI questionnaire (Gosling et al, 2003).

Participants were endowed with a virtual sum of money and one share of each of the 12 different assets. They were instructed to increase the total value of their portfolio above its initial value as much as possible. Participants were also told that they would be presented with the price graphs of each of these assets, one at a time. Prices were updated at a rate of one point per 0.2 second. The total value of the portfolio and each of the assets was updated after every point as well. These values were presented to the participants in a table.

Additional instructions informed them that, after every 20 points, they would be asked to decide whether to 1) buy another share of the asset, resulting in them having another share of the stock but less money to buy more stocks, 2) sell their share of the asset, resulting in them having no shares in it but more money, or 3) hold their share of the asset. They were informed that, if they decided to buy or sell, they would then move on to consider the next asset. However, if they decided to hold, the price graph of the current asset would continue to be updated until they were asked to make another decision about it at the next decision point or until day 220.

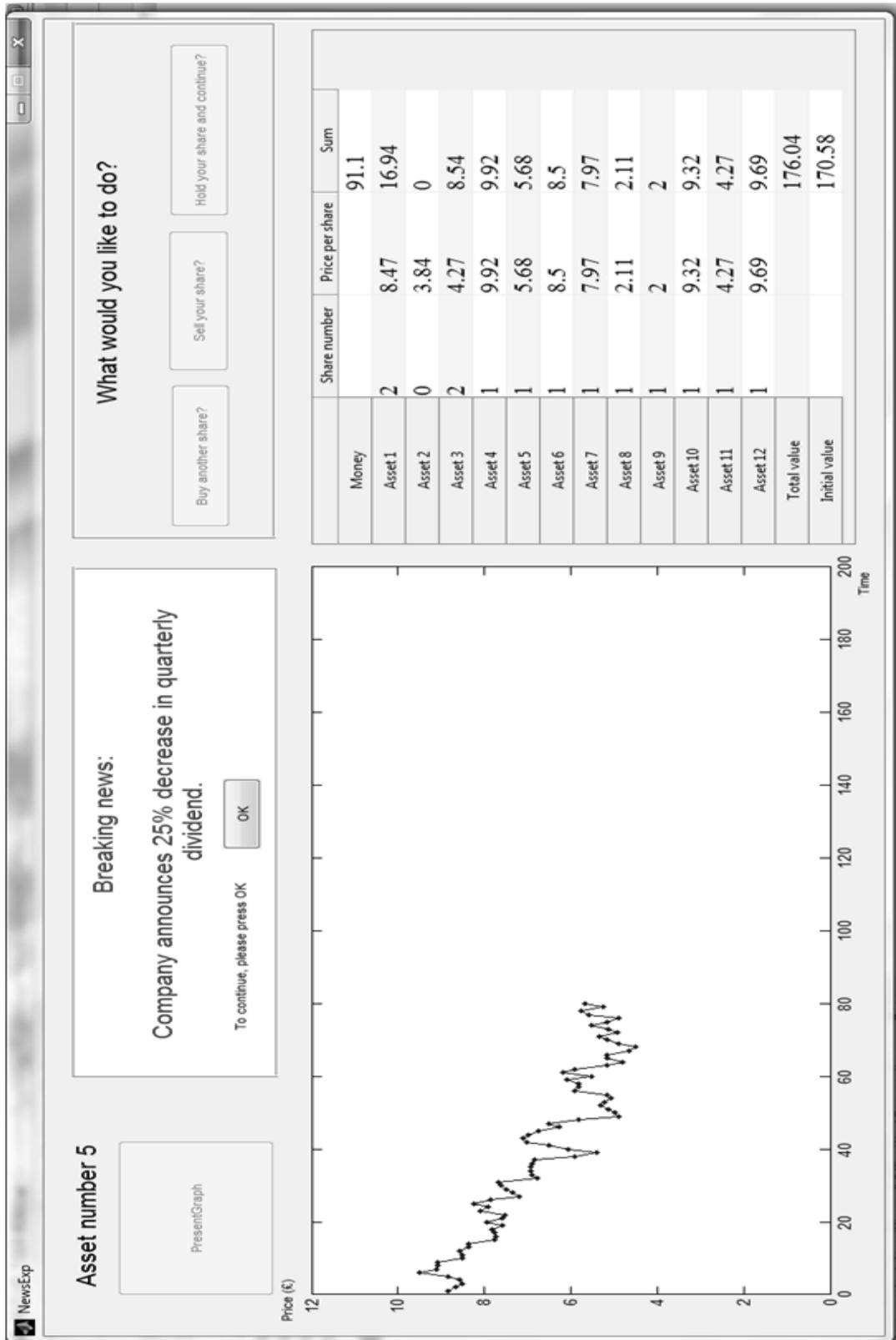


Figure 5.1 A typical task window from Experiment 1. The figure shows the non-conflicting condition with bad news and a negative trend.

After every 40 price points, participants were presented with a piece of news that was related to the current asset, together with a message emphasizing that they should read it carefully. Participants were also told that there might be a “Possible additional investment task” and that the experimenter may ask them to use their portfolio (money and assets left from the second stage of the experiment) for another investment task. The reason for this was that performing any action – buying, selling, or holding an asset – did not change the total value of the portfolio. The total value of participants’ portfolio changed only as asset prices changed. Possible future use of assets chosen to be held or bought endowed these actions with financial meaning.

Participants were informed how their fees depended on their performance. However, they were not provided with any trading strategy of the type Andreassen (1990) used to instruct his participants.

At the end of the experiment, participants completed the TIPI questionnaire.

### *Results*

Results are shown in Table 5.1. Primary dependent variables were trading latency and final share number. Trading latency was measured by the number of data points participants saw before making the decision to buy or sell each asset, or the maximum number of presented points (220) if participants made their decision to buy, sell, or hold their asset after all point series had been presented on the graph. A final share number of zero indicated that participants had sold their share, one meant that participants chose to hold their share, and two showed that participants had chosen to buy an additional share. I also analyzed participant returns (defined as the difference between the asset price at decision time and at the time of initial presentation of the series).

*The effect of news on financial decisions* To examine hypothesis  $H_{4,1}$ , I carried out a four-way analysis of variance (ANOVA) on final share number using culture (Western or

Table 5.1 Results of Experiment 1 for the western group (first panel) and the Eastern group (second panel).

Western group, N=30		Trend	
		Positive	Negative
Trading latency			
News valence	Good	48.33 (48.89)	60.00 (61.84)
	Bad	45.67 (38.28)	36.67 (24.75)
Share number			
News valence	Good	1.35 (0.92)	0.83 (0.96)
	Bad	1.03 (1.01)	0.4 (0.81)
Returns			
News valence	Good	3.14 (2.00)	-3.31 (2.22)
	Bad	2.12 (1.38)	-2.68 (1.16)

Eastern group, N=30		Trend	
		Positive	Negative
Trading latency			
News valence	Good	92.33 (71.98)	106.34 (70.78)
	Bad	66.00 (55.27)	73.66 (61.45)
Share number			
News valence	Good	1.13 (0.96)	1.15 (0.917)
	Bad	0.72 (0.96)	0.60 (0.87)
Returns			
News valence	Good	4.24 (2.25)	-4.35 (2.25)
	Bad	2.94 (1.89)	-3.44 (2.02)

Eastern) as a between-participant variable and trend (positive or negative), news valence (good or bad), and instance (first or second presentation of series in each condition) as within-participant variables. This revealed that final share number was larger when news was good ( $F(1, 29) = 29.35; p < .001; \text{partial } \eta^2 = .50$ ) and when trend was positive ( $F(1, 29) = 7.56; p = .01; \text{partial } \eta^2 = .21$ ). The size of effect of news valence was larger than that of the trend in the graphs, a finding that is consistent with hypothesis  $H_{4.1}$ . There was also an interaction between group and trend ( $F(1, 29) = 5.40; p = .03; \text{partial } \eta^2 = .16$ ). Tests of simple effects showed that in Western participants, final share number was higher when the trend was positive ( $F(1, 29) = 11.27; p = .002; \text{partial } \eta^2 = .28$ ).

To examine hypothesis  $H_{4.2}$ , I put participants' results into two groups: the conflicting conditions (good news, negative trend and bad news, positive trend) and the non-conflicting conditions (good news, positive trend and bad news, negative trend). For each group, I extracted the deviation of the final share number from 1 (the 'hold' option). ANOVA failed to yield a significant difference in this variable between the conflicting and non-conflicting conditions. Next, following Andreassen (1990), I calculated participants' price tracking (the correlation between the price of an asset at decision time with the final share number) for the conflicting and non-conflicting sets of results. An ANOVA showed neither an effect of culture nor of conflict between trend type and news type. Hence, I failed to replicate Andreassen's (1990) results: the data are not consistent with hypothesis  $H_{4.2}$ .

To examine hypothesis  $H_{4.3}$ , I grouped all participants' results together, and extracted two new variables. The first one was the difference between final share number and one share (the result of a 'hold' choice) when news was good. The second variable was the difference between one share and final share number when news was bad. These variables indicate the signed choice deviation from a 'hold' decision. ANOVA revealed that when news was good people bought fewer shares (mean: 0.12; std: 0.95) than they sold when news was bad (mean: 0.31; std: 0.95). This difference ( $F(1, 479) = 5.16; p = .02$ ) is consistent with hypothesis  $H_{4.3}$ .

*The timing of financial decisions* To examine hypothesis H<sub>4,4</sub>, I performed a t-test to compare trading latencies in the conflicting and non-conflicting conditions. No difference was found: the data are not consistent with hypothesis H<sub>4,4</sub>.

To examine hypothesis H<sub>4,5</sub>, I carried out a four-way ANOVA on trading latency with culture (Western or Eastern) as a between-participant variable and trend (positive or negative), news valence (good or bad), and instance (first or second presentation of series in each condition) as within-participant variables. This showed that trading latency was longer when news was good ( $F(1, 29) = 29.05$ ;  $p < .01$ ; partial  $\eta^2 = .50$ ) but that the effect of trend was insignificant. This pattern of results is consistent with hypothesis H<sub>4,5</sub>.

*Effects of culture* To investigate Hypothesis H<sub>4,6</sub>, I performed three separate two-way ANOVAs on number of shares, trading latency and returns, each with culture (Western or Eastern) as a between-participant variable and condition (non-conflicting or conflicting) as a within-participant variable. In no case was an interaction effect between culture and condition found. I therefore reject hypothesis H<sub>4,6</sub>.

To examine hypothesis H<sub>4,7a</sub>, I carried out a four-way ANOVA on trading latency with culture as a between-participant variable and trend, news valence, and instance as within-participant variables. This showed that trading latency was shorter for Western participants ( $F(1, 29) = 17.23$ ;  $p < .01$ ; partial  $\eta^2 = .37$ ), a finding that is consistent with hypothesis H<sub>4,7a</sub>.

I performed a four-way analysis of variance (ANOVA) on returns using the same variables as before. As expected, returns were larger when trends were positive ( $F(1, 29) = 417.32$ ,  $p < .001$ ; partial  $\eta^2 = .94$ ). Table 5.1 shows that return variances of participants from the Eastern group were higher than those of participants from the Western group. To compare these, I defined return dispersion as the absolute value of the difference between the return of each asset of each participant and the mean return in participant's group. A t-test revealed that return dispersion in the Eastern group was larger than that of Western group ( $t(239) = 5.60$ ;  $p < .001$ ). These results are consistent with hypothesis H<sub>4,7a</sub>.

I did not match the age or gender of participants in the Western and Eastern groups.

However, these variables had no significant effects on trading latency or return dispersion and so could not provide an alternative account for the differences observed between the two groups.

*Effects of personality* For each of the participants and for each of the experimental conditions (good or bad news, positive or negative trend), I extracted the mean trading latency, mean final share number and mean returns. Participants with greater openness to experience had lower trading latencies ( $r = -.28$ ;  $p = .03$  when news was good and the trend was positive;  $r = -.32$ ;  $p = .01$  when news was good and the trend was negative;  $r = -.37$ ;  $p = .004$  when news was bad and the trend was positive;  $r = -.33$ ;  $p = .01$  when news was bad and the trend was negative). They also bought more shares but only when bad news was combined with a positive trend in the price data ( $r = .36$ ;  $p = .005$ ). Finally, their returns were higher when the trend in the price data was negative ( $r = .34$ ;  $p = .008$  for good news;  $r = .31$ ;  $p = .02$  for bad news) but lower when it was positive and the news was bad ( $r = -.27$ ;  $p = .04$ ). These results are consistent with hypothesis H<sub>8</sub>. Correlations between remaining four personality traits and the task variables were not statistically significant.

### *Discussion*

Participants made faster decisions (H<sub>4,5</sub>) and bought fewer shares when news was bad than when it was good. They also sold more shares when the news was bad than they bought when it was good (H<sub>4,3</sub>). In addition, they bought more shares when the trend in the price data was positive but this effect was weaker than that of the news valence (H<sub>4,1</sub>).

Why was the effect of news valence on share number stronger than that of the trend in the price graphs? Though participants were instructed to pay attention to the news items, their presentation was no more visually salient than that of the trend in the price series (Figure 5.1). Furthermore, portfolio values were continuously updated in a manner that matched the prices changes in the graph. Participants could, therefore, see that their losses (or gains)

corresponded directly to changes in the price series rather than to the news items. Hence, I interpret the greater influence of news on trading in light of Tuckett's (2012) arguments that people need to find meaning in their environment. News offers narratives and therefore people tend to focus on it.

None of the hypotheses ( $H_{4,2}$ ,  $H_{4,4}$ ,  $H_{4,6}$ ) based on putative effects of a conflict between news and price data were supported. Although share buying was affected both by news and by price trend, effects of these variables did not interact in the manner expected on the basis of conflict effects.

Participants in the Eastern group made their trades much later than those in the Western one, and, as a result, their return dispersions were larger ( $H_{4,7}$ ). This finding is consistent with the notion that they developed more complex narratives that pulled together the different pieces of information they had encountered into a more holistic framework (Nisbett, 2003).

The finding that participants with greater openness to experience had shorter trading latencies is consistent with results obtained by Fiori and Antonakis (2012) in a variety of non-financial tasks. However, from a risk taking perspective, it is perhaps surprising. Nicholson et al (2005) found that propensity to take risks was greater in extraverts and in those who are more open to experience. As shorter trading latencies indicate lower risk propensity, their findings would lead me to expect longer rather than shorter decision latencies in those with high levels of openness to experience. Hence, it appears unlikely that the relation between trading latency and openness to experience was mediated by risk propensity. Instead, it is more likely that people open to experience put more cognitive effort into their task and thereby made more effective use of the information they received. As a result, they were able to produce a satisfactory narrative for it sooner.

## **Experiment 2**

Experiment 2 was designed to test the following hypotheses:

H<sub>4,9</sub>: there is a positive correlation between views about the extent to which an event will affect prices and final share number.

H<sub>4,10</sub>: the difference between a participant's forecast and the last data point depends on the news valence and the direction of the trend in the price data.

H<sub>4,11</sub>: there is a positive correlation between that difference and final share number.

Experiment 2 also provided an opportunity for confirming the conclusions pertaining to hypotheses H<sub>4,1</sub>- H<sub>4,5</sub>.

### *Method*

In addition to making trading decisions, this experiment required participants to make forecasts and to assess how plausible it was that each news event would affect asset prices.

*Participants* Thirty people (11 men and 19 women) recruited in the same way as before acted as participants. They were all from Western culture and their average age was 25 years. Twenty eight of them were undergraduate or postgraduate students. They were paid a fixed fee of £2.00. Up to an additional £2 was paid according to their performance in the same way as in Experiment 1.

*Materials and design* These were the same as in Experiment 1.

*Procedure* The procedure was similar to Experiment 1, except that participants were presented with a news item every 40 points starting from point 20 (rather than every 40 points starting from point 40). This was to ensure that all participants, including those who decided to buy or sell their assets after 20 points, saw at least one news item.

In addition, after every 20 points, participants were asked, before making their decision, to make a single forecast for the point that was 20 points ahead of the current one. Forecasts were made by clicking the mouse on a vertical line designating the required forecast date. Until participants pressed the button "save forecast", they could edit their forecast by clicking the mouse again on the line. Moreover, whenever a news item was presented, they

were asked to rate how plausible it was that such a news event would affect asset prices. Plausibility ratings were performed using a slider and they ranged between 0 and 100, where 0 meant “not plausible at all” and 100 meant “extremely plausibly”. Figure 5.2 presents a typical task window in Experiment 2.

### *Results*

Results are shown in Table 5.2. In addition to analyzing the data as before, I extracted participants’ plausibility ratings and forecasts. (One forecast of one of the participants in the condition bad news, negative average trend was removed because it was more than four standard deviations from the mean of the forecasts in that condition).

*The effect of news on financial decisions* A three-way ANOVA, using trend (positive or negative), news valence (good or bad), and instance (first or second presentation of series in each condition) as within-participant variables, showed that final share number was higher when news was good ( $F(1, 29) = 11.47$ ;  $p = .002$ ; partial  $\eta^2 = .28$ ) and when price graphs had a positive trend ( $F(1, 29) = 4.54$ ;  $p = .04$ ; partial  $\eta^2 = .14$ ). These results are consistent with hypothesis  $H_{4,1}$  and replicate those obtained in Experiment 1.

As before, trials were classified into those in which the news valence and price trend were conflicting and non-conflicting. ANOVAs comparing the final number of shares and the deviation of final number of shares from 1 (‘hold’ decision) failed to find any significant effect of conflict. Thus, as in Experiment 1, I reject Hypothesis  $H_{4,2}$ .

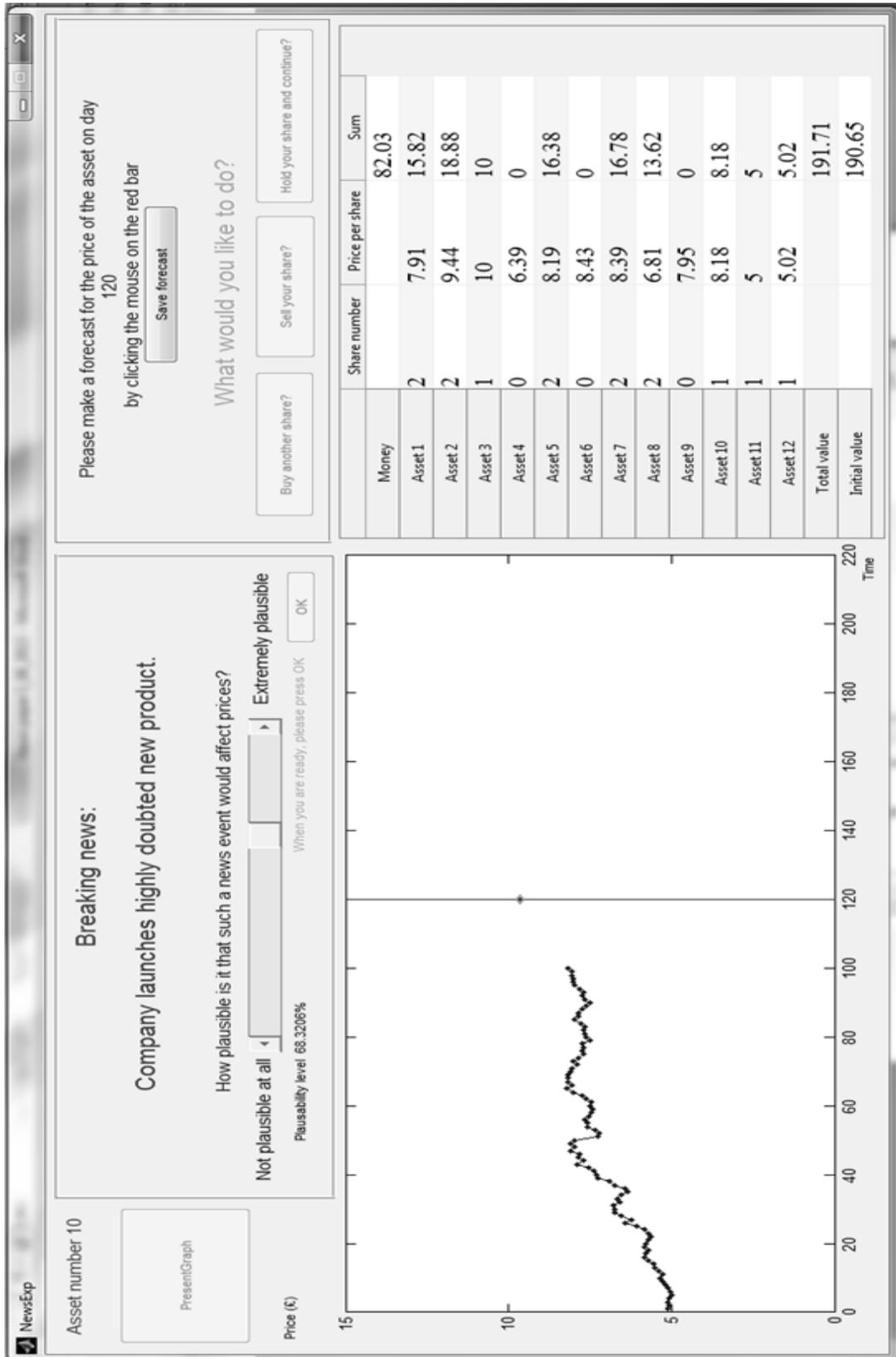


Figure 5.2 A typical task window from Experiment 2. The figure shows the conflicting condition with bad news and a positive trend.

Table 5.2 Results of Experiment 2, including trading latencies, share numbers, plausibility ratings (first panel), forecast differences and returns (second panel).

Western participants, N=30			Trend	
			Positive	Negative
Trading latency				
	News	Good	75.00	69.00
	valence		(58.87)	(60.78)
		Bad	67.33	45.00
			(55.11)	(38.90)
Share number				
	News	Good	1.15	0.85
	valence		(0.97)	(0.95)
		Bad	0.55	0.31
			(0.87)	(0.72)
Plausibility				
	News	Good	0.67	0.65
	valence		(0.16)	(0.17)
		Bad	0.65	0.68
			(0.18)	(0.17)

---

Forecasts				
	News valence	Good	0.76	0.69
			(0.66)	(1.05)
		Bad	-0.08	-0.49
			(0.94)	(1.07)
Returns				
	News valence	Good	2.20	-2.23
			(1.29)	(1.36)
		Bad	1.57	-1.59
			(1.23)	(0.91)

---

To test hypothesis H<sub>4,3</sub>, I proceeded in the same way as before. The ANOVA revealed an asymmetry in final share number with respect to news and trend ( $F(1, 119) = 11.62$ ;  $p = .001$ ). Participants sold more shares when news was bad and the trend in the price data negative than they bought when news was good and the trend in the price data was positive. Similar results were obtained when I compared deviation from ‘hold’ option for good news and bad news ( $F(1, 239) = 24.20$ ;  $p < .001$ ). As in Experiment 1, the results are consistent with hypothesis H<sub>4,3</sub>.

*The timing of financial decisions* An ANOVA comparing differences between trading latencies in conflicting and non-conflicting conditions failed to reveal any effects of conflict. Thus, as in Experiment 1, the data do not support hypothesis H<sub>4,4</sub>.

A three-way ANOVA using trend, news valence, and instance as within-participant variables showed that trading latency was longer when the news was good ( $F(1, 29) = 8.23$ ;

$p = .008$ ; partial  $\eta^2 = .22$ ). As no main effect of trend was obtained, the results are again consistent with hypothesis  $H_{4,5}$ .

There was an interaction between news and trend ( $F(1, 29) = 5.68$ ;  $p = .02$ ; partial  $\eta^2 = .16$ ). Tests of simple effects showed that, when the trend was negative, trading latency was longer in the good news condition ( $F(1, 29) = 14.27$ ;  $p = .001$ ; partial  $\eta^2 = .33$ ) and that, when the news was bad, trading latency was longer when the trend was positive ( $F(1, 29) = 11.44$ ;  $p = .002$ ; partial  $\eta^2 = .28$ ). Further analysis showed that trading latency was longer when the news was good and the trend positive than when the news was bad and the trend negative ( $t(59) = 3.43$ ;  $p = .001$ ).

*Plausibility ratings* A three-way ANOVA on plausibility estimates using the same three variables as before failed to find any significant effects. Thus, the data failed to support for Hypothesis  $H_{4,9}$ .

*Forecast quality* Before examining how forecasts depended on trading information ( $H_{4,10}$ ) and how they influenced trading decisions ( $H_{4,11}$ ), I examined their quality by extracting two variables. The first was the mean absolute difference (MAD), defined as the absolute value of the mean of the difference between the forecasts each participant made for each graph at time  $t$  and the prices at time  $t$ . MAD ( $M = 0.82$ ;  $SD = 0.51$ ) measures the deviation of participants' forecasts from naive forecasts. The second was the mean absolute error (MAE), defined as the absolute value of the mean of the difference between the forecasts each participant made for each graph at time  $t$  (for time  $t+20$ ) and the prices at time  $t+20$ . MAE ( $M = 0.74$ ;  $SD = 0.67$ ) measures the deviation of participants' forecasts for each graph from forecasts that would have produced zero error. T-tests showed that both these variables were significantly different from zero (for MAD,  $t(238) = 24.80$ ;  $p < .001$ ; for MAE,  $t(238) = 25.67$ ;  $p < .001$ ). Thus, in line with Harvey and Reimers (2013), Harvey (1995), and Reimers and Harvey (2011) but in contrast to the assumption made by Pfajfar (2013), forecasts were neither naïve nor perfect.

*Dependence of forecasts on news valence and trends in price data* Participants could produce up to 10 forecasts for each asset. For each time,  $t$ , at which participants made a decision regarding an asset, I extracted the differences between their forecasts for the price of the asset at time  $t+20$  and the price of the asset at time  $t$ . I then averaged these differences for each graph. An ANOVA, using the variables trend (positive or negative), news valence (good or bad), and instance (first or second presentation of series in each condition), showed that the difference between forecasts and asset prices was higher when news was good ( $F(1, 29) = 38.93$ ;  $p < .001$ ; partial  $\eta^2 = .57$ ), and when the trend was positive ( $F(1, 29) = 14.76$ ;  $p = .001$ ; partial  $\eta^2 = .34$ ). These results provide support hypothesis  $H_{4,10}$ .

*Correlation between forecasts and financial decisions* To examine Hypothesis  $H_{4,11}$ , I calculated the correlation of the number of shares participants had at the end of each trial with the difference between participants' forecasts at the time of their final trading decisions and the value of the last price they saw. A positive correlation between these two variables shows that participants tended to buy more shares when they thought that the prices would rise. Calculated for each condition separately, I found positive correlations when the trends were positive, whether the news items were good ( $r = .53$ ;  $p < .001$ ) or bad ( $r = .48$ ;  $p < .001$ ). No significant correlations were obtained for conditions with negative trends. These results suggest that forecasts mediated between the data and trading decisions only when prices were rising. Thus, the results partially support hypothesis  $H_{4,11}$ .

### *Discussion*

Just as in Experiment 1, results were consistent with hypotheses  $H_{4,1}$ ,  $H_{4,3}$ , and  $H_{4,5}$  but not with  $H_{4,2}$  and  $H_{4,4}$ . Thus the findings here provide confirmation of the conclusions drawn from the earlier experiment.

Experiment 2 supported Andreassen's (1990) claim that forecasts mediate between data and decisions. Forecasts depended strongly on news valence. Their dependence on the trends in the price series was weaker. Yet many experiments have shown that, in the absence of any news, forecasts depend strongly on the trends in data series (e.g., Harvey and Reimers, 2013;

Lawrence et al, 2006). It appears that the presence of news dominates information relating to the trend in the price series: as I argued above, the appeal of the narrative structure of news is so strong that people prefer to act on it rather than on the trend cues<sup>4</sup>.

Once forecasts had been made, their influence on trading was affected by the trend in the price series. When that trend was positive, forecasts were taken into account when making decisions to buy or sell.

Finally, the results indicate that forecasts were neither naive nor perfect. This finding implies that the forecasting assumption underlying Pfajfar's (2013) behavioral model of markets is not realistic.

## **Conclusions**

During the past few years, a large body of research on agent-based market models has accumulated. A search using the key words "agent", "model" and "market" of the EconLit database between the years 2000 and 2013 yielded 3,946 papers, of which 1,911 were published between 2008 and 2013. The cumulative behavior of individuals has become a centre of attention within finance; there is now a bridge between the scale of a single person, which traditionally has been of interest only within psychology, and the scale of the masses, as classically modeled in finance.

However, many behavioral models of market behavior include assumptions which are not based on psychological findings. This study has supplied data relevant to these models and cast new light on the way people react to financial data in trading tasks. Specifically, I chose to examine three factors that are relevant to EMH and frequently involved in modern financial models: the effect of news on financial decisions, trading latency, and individual differences between investors. Superficially, these three factors may appear to be diverse and

---

<sup>4</sup> Inclusion of filler series with U-shaped and inverted U-shaped trends may have acted to reduce the weight that participants put on price trend data when making their trading decisions. However, inclusion of filler series ensured high external validity of the experiments: clearly, in real-life, not all trends are easy to identify.

unconnected. However, the effects related to them can all be accommodated within a single coherent approach.

Though results are consistent with previous work on the inadequacy of the EMH (Findlay and Williams, 2000 - 2001), they are best understood within a framework for understanding and modeling trader behavior that takes into account the natural, human search for meaning.

First, though participants in the experiments could always see that the value of their portfolio changed according to the trend of the presented price graphs, most of them still chose to base their decisions on news items rather than on the price series. Trading latencies also depended on news rather than on the trend in price series. News provides narratives for those searching for meaning more easily than price trends do. In fact, news items may allow people to make sense of the price trends by supplying 'cognitively comforting' causal interpretations of them in the way that Tuckett (2012) suggests. Causal interpretations within a narrative also underlie fundamental analysis and so this may also help to explain why many analysts prefer it to technical analysis.

Second, openness to experience is correlated with need for cognition (Sadowski and Cogburn, 1997). Cacioppo et al (1983) have shown that those with higher need for cognition put more cognitive effort into tasks and, as a result, are better able to focus their attention on the most relevant information. This implies that people in our task who were more open to experience put more cognitive effort into selectively processing and integrating the information they received. As a result, they produced adequate narratives more quickly and were able to act on them sooner: they had shorter trading latencies.

Third, trading latencies of participants from Eastern cultures were much longer than those of Western participants. This difference resulted in a significantly higher dispersion of returns in the Eastern group. The work of Nisbett and his colleagues (e.g., Nisbett, 2003; Nisbett, Peng, Choi and Norenzayan, 2001) has shown that those in Eastern cultures think more holistically and less analytically than those in Western ones. They make greater attempts to pull all available evidence into a single holistic framework. Narratives provide the primary

means for bringing evidence into a coherent framework (Pennington and Hastie, 1993).

Finding more coherent narratives requires additional processing. According to this line of reasoning, the Eastern participants had higher trading latencies because they spent more time make sensing of the evidence by generating more coherent narratives to explain it.

Fourth, forecasts may provide some insight into how participants selectively incorporated price trend information into their narratives. Forecasts were indeed higher when news was good and price trend was positive. Thus, even though forecasts were not optimal, they were in the right direction, a finding consistent with previous work (Harvey and Reimers, 2013). However, these forecasts influenced trading only when price trends were positive. Even though participants had forecast a drop in price when the price trend was down, they tended not to sell (c.f. Odean, 1998). One interpretation, derived from one originally proposed by Lawrence and Makridakis (1989), is that people had contrasting narratives for up trends and down trends. If prices were increasing, no agency would intervene to stop them from increasing and hence, trades could be consistent with forecasts. However, if prices were forecast to decrease, there would be at least a possibility that some agency (e.g., the company owned by the shareholders) would intervene in an attempt to prevent any further decrease. As a consequence, it would be sensible not to act on or to delay acting on the forecast.

In summary, the findings reported here are best understood within an approach that sees traders as trying to make sense of information by incorporating it within a narrative that provides a causal interpretation of events. Given research in other domains (Pennington and Hastie, 1993), I suggest that people select between different possible narratives by choosing the one that has the greatest degree of coherence. Other approaches, such as the EMH or behavioural models that incorporate a number of disconnected cognitive biases, do not appear to be capable of providing a satisfactory explanation for our findings.

### *Limitations*

The experiments were designed to provide the control needed to test hypotheses while still providing participants with a task scenario that captured the essential features of the sort of computer-based trading experienced by small investors. However, there were some features of real trading that were not incorporated within the paradigm. For example, I presented participants with information typical of that likely to be relevant to the trading task. In real trading, however, people are likely to actively seek out information. As a result, they will be subject to confirmation bias (Hilton, 2001): they will selectively gather information that is consistent with the narrative that they have developed while making little effort to obtain information inconsistent with it. The paradigm used here did not allow effects of this bias to be studied.

In addition, in our experimental settings, participants could only buy or sell a single share of each company. After making a buy or sell decision, participants could no longer see how the price of the company evolved. This setting was chosen in order to make the experiment as simple as possible. However, this manipulation could have affected participants' financial decisions. Furthermore, informing participants about a possible additional investment task could have affected their buy/sell/hold decisions. I consider it important to try to replicate the results presented using different trading tasks and incentive mechanisms. An alternative design could, for instance, allow participants to buy or sell more than one asset. Participants would be able to see price evolution of each asset for the same duration, and continue buying or selling shares throughout this period. The incentive mechanism could be based only on the value of the portfolio.

Participants were not professional traders. I was interested in obtaining results from lay people: the Internet has greatly facilitated non-professional trading (Barber and Odean, 2008; Muradoglu and Harvey, 2012). Nevertheless, it is worth emphasizing that studies contrasting the financial behavior of lay people and experts have rarely found differences between them (Zaleskiewicz, 2011; Muradoğlu and Önköl, 1994). Furthermore, the present

results coincide with those obtained from studies of professional traders (e.g. Odean, 1998). However, it would still be valuable to replicate them on that population.

I focused on one characteristic of news and price graphs: their valence or sign. However, both news and price graphs have other features that could be important (Nelson, Bloomfield, Hales and Libby, 2001). For example, the degree of relevance of the news to the asset may affect financial decisions and the volatility of price graphs may influence trading latency.

In both experiments, participants were exposed to both graphical and verbal data. In future work, these could be studied separately. This would allow examination of the way that news dominates price information more systematically and may throw light on how people perform in situations that require 'pure' technical or 'pure' fundamental analysis.

Finally, it is important to note that participants were not asked to produce narratives or tell us possible narratives. I consider it important to examine the narrative hypothesis further and will discuss this issue in the final chapter.

## **Chapter 6: Psychological Mechanisms Supporting Preservation of Asset Price Characterisations**

*“Fractal geometry is not just a chapter of mathematics, but one that helps Everyman to see the same old world differently” (Mandelbrot, cited in Aufmann, Lockwood, Nation and Clegg, 2010, page 551).*

In this chapter, I examine the question of whether the way people perceive financial data sequences and make forecasts from them has a role in the stabilisation of market parameters. Athanassakos and Kalimipalli (2003) have shown that future volatility is correlated with forecast dispersion. Therefore, a correlation between forecast dispersion and measures of the volatility of past data could serve as a part of the mechanism that preserves data properties for durations long enough to enable the use of forecasting methods and financial algorithms.

### **Experiment 1**

The primary aim of Experiment 1 was to investigate the effects of the Hurst exponents of price graphs on financial forecasts and decisions: this is because such effects may be one of the mechanisms that directly stabilises properties of price graphs (see the section Mechanisms preserving asset price graph structure in Chapter 1). A secondary aim was to explore the effects of forecast horizon on the same variables, as these effects could provide support for Corsi’s (2009) approach. According to Corsi, prices exhibit fractal behaviour due to the heterogeneity of investor forecast horizon (see the section: *Models and theories about stability of market parameters: the effects of time-scaling*). In particular, I tested the following hypotheses:

H<sub>5,1a</sub>: People use scaling when making financial forecasts and decisions. In particular, they exhibit a large degree of variation in their choice of temporal scaling of fractal graphs (consistently with the heterogeneity hypothesis of Müller, Dacorogna, Davé, Pictet, Olsen, and Ward, 1993).

H<sub>5,1b</sub>: Variation of choices of temporal scaling is greater for more distant trading horizons.

H<sub>5,2</sub>: There is a positive correlation between forecast horizon and the local steepness and oscillation of the time-scaled data graphs.

H<sub>5,3</sub>: Dispersion of forecasts is positively correlated with the required forecast horizon.

H<sub>5,4</sub>: Selected time scaling factors are smaller for graphs that have smaller Hurst exponents: people prefer presentation of data corresponding to shorter periods of time when dealing with graphs with smaller Hurst exponents.

H<sub>5,5a</sub>: The time scales that people choose result in a negative correlation between the local steepness and oscillation of the time-scaled graph and the Hurst exponent of the original data.

H<sub>5,5b</sub>: The time scales that people choose result in a positive correlation between the local steepness and oscillation of the time-scaled graphs and the original graphs.

H<sub>5,6</sub>: The dispersion of forecasts is negatively correlated with the Hurst exponents of the original graphs and positively correlated with the local steepness and oscillation of the data graphs.

H<sub>5,7</sub>: People's trading behaviour depends on their forecasts.

I presented participants with a sequence of fractal time series representing price graphs. At the beginning of each trial, each graph was presented on the time interval of  $t = [100, 200]$  days. Participants could control the time interval of the presented graph by using a slider. Possible time intervals ranged between  $[0, 200]$  days at the maximal zoom-out limit of the

slider, and [196, 200] days at the maximal zoom-in limit of the slider. Participants were asked to choose the time interval they considered the most appropriate for making financial forecasts and decisions, and then to make forecasts and decisions based on the time-scaled graph. I manipulated two variables: the Hurst exponent of the original data graphs (and thus also their local steepness and oscillation), and the required forecast horizon. Figure 6.1 depicts the task window of Experiment 1.

### *Method*

*Participants* Thirty-four people (15 men and 19 women) with an average age of 23.29 years acted as participants. They were paid a flat fee of £3.00 and a further £1.00 if their financial decisions were more than 65% correct. Correctness was determined by participants' performance with respect to the generated graphs. For instance, if prices at the forecast horizon were higher than the price on day 200 by more than 5%, a 'buy' decision was considered correct and both 'sell' and 'hold' decisions were considered wrong.

*Stimulus materials* Stimulus graphs comprised five sets of three time series with Hurst exponents  $H = 0.3, 0.5, \text{ and } 0.7$ . Time series were produced using the Spectral method described by Saupe (Peitgen and Saupe, 1988). All series included 62831 ( $\sim 2000\pi$ ) points. They were presented to the participants as asset price graphs. A constant was added to them to ensure that they were positive. To increase measurement precision, they were also multiplied by 100 to encourage participants to make forecasts using more than one significant digit.

*Stimulus presentation and control* Stimulus graphs were presented using a Matlab programme that enabled participants to scale the data along the time axis, to make forecasts for a specified horizon, and to express their financial decisions.

Time scaling was accomplished using a slider. At the beginning of each trial, each graph was presented on the time interval [100, 200]. The scaling slider's range varied from a time interval of four days at the maximal zoom-in side of the slider (presentation of price data

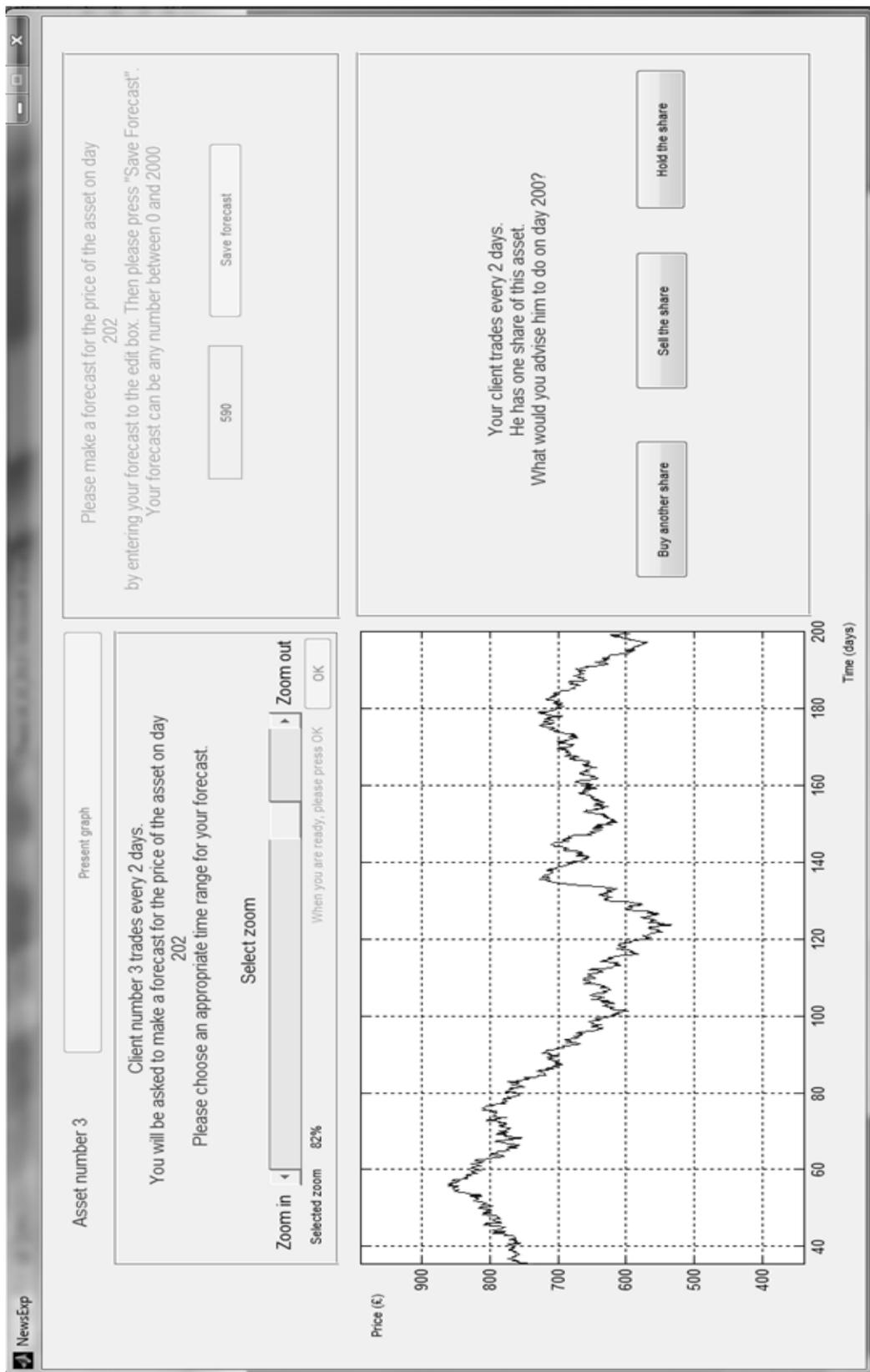


Figure 6.1 The task window of Experiment 1.

from days 196 to 200) to 200 days at the maximal zoom-out side of the slider (presentation of price data from days 0 to 200). Thus they could scale the graphs by a factor of 50 (i.e.  $200/4$ ).

Participants made single point forecasts by entering a number into a text box. Forecast horizon was set to 2, 15, or 100 days, making the factor by which horizons varied (i.e.  $100 / 2$ ) identical to that by which scaling could vary (i.e.  $200 / 4$ ).

Participants then made a financial decision to buy another share of the presented asset, to sell their share, or to do neither of these.

On each trial, they could change the time interval shown on the graph until they clicked the button “When you are ready, please press OK”. They could edit their forecasts until they clicked the button “Save forecast”.

*Design* Participants were presented with 48 graphs: three familiarisation graphs and 45 experimental graphs. Only experimental graphs were included in the analysis. Each graph required three responses: the first was choice of time interval; the second was to forecast the asset’s future price; the third was to make a financial decision.

Each participant saw all 15 graphs. Each one was presented three times in different contexts that varied according to the required forecast horizon (2, 15, and 100 days). The order of the graphs and the required forecast horizons were randomly chosen. This combination produced a three (forecast horizons) by three (Hurst exponent values) by five (instances of time series with the same Hurst exponent values) within-participants design.

*Procedure* Participants were instructed to assume that the experiment day was day 200 and asked to read the following instructions:

“In the following experiment, you are asked to imagine that you are a financial analyst. You have 45 clients. Each of your clients has one share of a single asset. Clients differ in their

trading frequency: some clients trade every two days, some trade every 15 days, and some every 100 days. Your aim should be to increase the total value of their portfolios as your fees will depend on your performance.

In order to make your decisions, you will be presented with the price graphs of each of these assets. You will be able to control the time range of each graph by changing its zoom.

For each asset you will be asked to:

1. Notice the trading frequency of your client and the day you will be asked to make financial forecast for. Look at the price graph of the asset carefully.
2. Choose for each graph a time range which you consider the most appropriate for the purpose of making a financial forecast.
3. Write your forecast for the price of the asset on the required day.
4. Advise to your client whether to buy another share of the asset, sell their share, or hold it.”

Participants could choose the time range of the data graphs by dragging a slider.

Forecasts were made by entering a number to a text box. Participants could advise their clients whether to buy, sell, or hold their shares by clicking one of three buttons.

All tasks had to be completed before participants could continue to the next graph.

### *Results*

I excluded from the analysis participants whose means of choices of time scaling factor were more than three standard deviations greater than that of the average for the rest of the group and those whose forecasts were different from the mean of the group by more than two standard deviations. This reduced the size of the sample from 34 to 30 participants, leaving a total of 1350 graphs for the analysis. Variables of primary interest were the chosen time

scaling factor, the local steepness and oscillation of the scaled graphs, the dispersion of participants' forecasts, and the resultant share number.

*Choice of time-scaling factor* I refer to the location on the scaling-slider which participants chose for each graph as the *time-scaling factor*. This measurement could vary between 0, corresponding to four days and 1, corresponding to 200 days (the transformation used to translate time-scaling factors to the actual day number presented on the graphs was: day number =  $196 * (\text{time-scaling factor}) + 4$ ). The mean time-scaling participants chose was 0.40, and the standard deviation was 0.37. A t-test performed on participants' choices of smoothness levels showed that the mean value was significantly different from 0.5 (the initial setting):  $t(1349) = 9.74, p < .001$ , from 0.0 (maximal zoom-in):  $t(1349) = 40.05, p < .001$ , and from 1.0 (using information from the maximum time-interval that was available):  $t(1349) = 59.53, p < .001$ . As the standard deviation was quite large (0.37 – close to the mean and larger than a third of the possible range), I accept Hypothesis  $H_{5,1a}$  (people use scaling to make financial forecasts and decisions, and they exhibit a large degree of variation in their choice of temporal scaling of fractal graphs). This result supports also the heterogeneity hypothesis of Müller, Dacorogna, Davé, Pictet, Olsen, and Ward (1993).

To examine Hypotheses  $H_{5,1b}$ , and  $H_{5,4}$ , I carried out a three-way repeated measures ANOVA on the chosen time scale using the forecast horizon, Hurst exponent, and graph instance as within-participant variables. Mauchly's sphericity assumption was violated for the horizon variable but not for the other variables. Here and everywhere else, I report the results of the Huynh-Feldt test whenever Mauchly's sphericity assumption is violated. The results showed that the chosen scaling factor was larger when forecast horizons were longer ( $F(1.52, 42.44) = 148.97; p < .001; \text{partial } \eta^2 = .84$ ). That means that when forecast horizons were longer, participants chose to present data from longer periods of time. However, the effect of forecast horizon on chosen scaling factor was quadratic ( $F(1, 28) = 27.31; p < .001; \text{partial } \eta^2 = .49$ ). The latter had a significant linear component as well ( $F(1, 28) = 221.22; p < .001; \text{partial } \eta^2 = .89$ ).

The correlation between chosen time-scaling factor and forecast horizon was  $r = .77$  ( $p < .001$ ). I accepted Hypothesis  $H_{5,1b}$  (variation of choices of temporal scaling is greater for more distant trading horizons).

The ANOVA reported above showed also that the chosen scaling factor was smaller when  $H$  was smaller ( $F(2, 56) = 5.76$ ;  $p = .005$ ; partial  $\eta^2 = .17$ ). This means that participants zoomed-in more when  $H$  was smaller; they viewed data relating to shorter time periods when the Hurst exponents of the graphs were smaller. I, therefore, accepted Hypothesis  $H_{5,4}$  (people prefer presentation of data corresponding to shorter periods of time when dealing with graphs with smaller Hurst exponents). The effect of the Hurst exponent on chosen scaling factor was linear:  $F(1, 28) = 9.97$ ;  $p = .004$ ; partial  $\eta^2 = .26$ . Figure 6.2 depicts the mean selected scaling factor against the Hurst exponent of the graphs for the different experimental conditions.

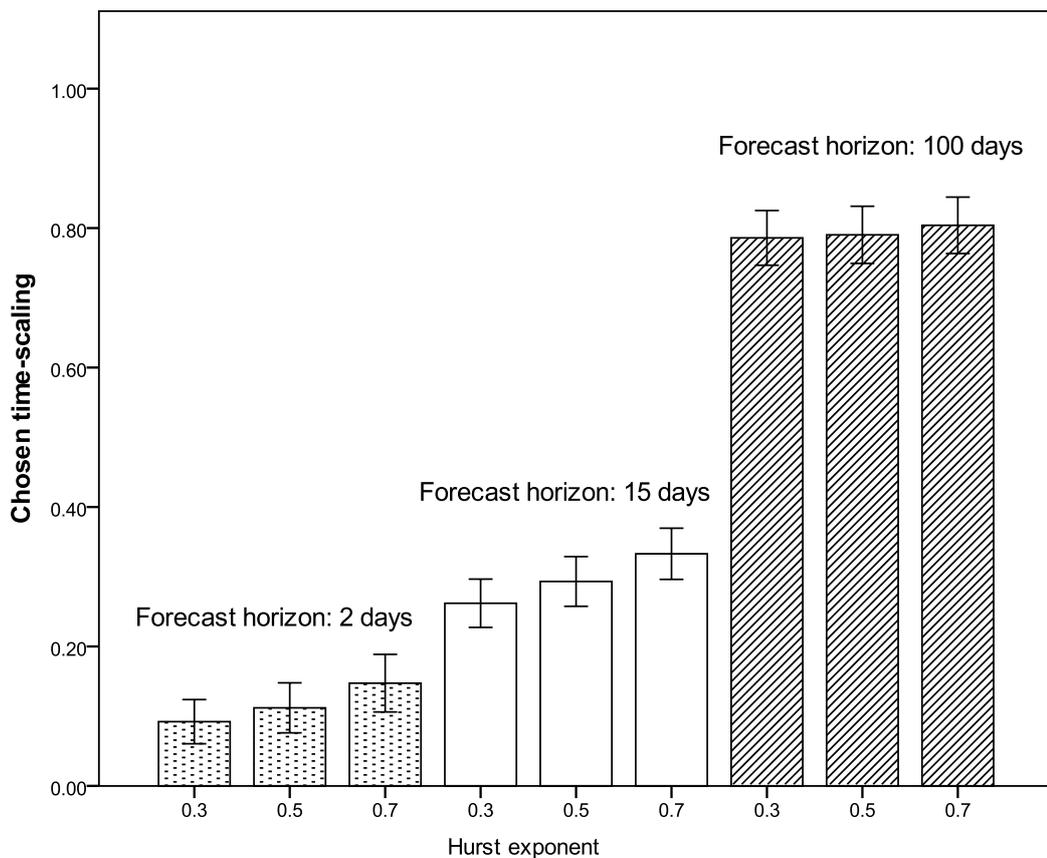


Figure 6.2 . Chosen time-scales with respect to the conditions of Experiment 1.

Participants' selections of scaling factors affected the geometric properties of the graphs participants based their forecasts and decisions on. How did the resultant, scaled graphs look?

*Properties of scaled graphs* To measure the perceived local steepness of a scaled time series, I extracted the average of the absolute value of the gradient at each series point. I then multiplied this value by the ratio of the observed time interval and the number of pixels along the time axes of the graph (600). I calculated local steepness measures for the original data series and for the data series after participants' scaling.

To examine Hypotheses  $H_{5,2}$  and  $H_{5,5}$  with respect to the graphs' local steepness, I carried out a four-way repeated measures ANOVA on the local steepness of the data graphs, using the variables state (before/after scaling), the forecast horizon, the Hurst exponent, and the instance of the graphs as within-participant variables. Mauchley's test of sphericity assumption was violated for all variables except for instance. As expected, scaling reduced the local steepness of the graphs: the state variable was significant ( $F(1, 29) = 29.66$ ;  $p < .001$ ; partial  $\eta^2 = .51$ ). Local steepness was larger when forecast horizon was longer ( $F(1.50, 43.47) = 159.79$ ;  $p < .001$ ; partial  $\eta^2 = .85$ ) and when the Hurst exponent was smaller ( $F(1.07, 31.15) = 2307.99$ ;  $p < .001$ ; partial  $\eta^2 = .99$ ). A small effect of instance was also found ( $F(4, 116) = 2.54$ ;  $p = .04$ ; partial  $\eta^2 = .08$ ). That means that scaling depended on the specific realisation of graphs used for the experiment. However, this effect was smaller than the other effects.

There were significant interactions between all variables. Tests of simple effects yielded results which were in line with all our hypotheses or did not contradict them. I report the results of the interactions and of the corresponding simple tests in Table B.1 in Appendix B.

The local steepness values of the original and of the scaled graphs were significantly correlated ( $r = .58$ ;  $p < .01$ ). Both steepness variables were correlated with the Hurst exponents of the original graphs ( $r = -.95$ ;  $p < .01$ ,  $r = -.55$ ;  $p < .01$ , respectively). These

results support Hypotheses  $H_{5,2}$  and  $H_{5,5}$  with respect to the graphs' local steepness (that is, there is a positive correlation between forecast horizon and the local steepness of the time-scaled data graphs, and there is a negative correlation between the local steepness of the time-scaled graphs and the Hurst exponent of the original data).

To examine Hypotheses  $H_{5,2}$  and  $H_{5,5}$  with respect to the graphs' oscillation (the difference between the minimum and the maximum of each graph), I carried out a four-way repeated measures ANOVA on the oscillation of the data graphs, using the same variables as before. Mauchley's test of sphericity assumption was violated only for the variable instance.

The results showed that oscillation was smaller in the scaled graphs ( $F(1, 29) = 98.49$ ;  $p < .001$ ; partial  $\eta^2 = .77$ ). The analysis also revealed that oscillation was larger when forecast horizon was longer ( $F(2, 58) = 204.46$ ;  $p < .001$ ; partial  $\eta^2 = .88$ ), and when the Hurst exponent was smaller ( $F(2, 58) = 6106.67$ ;  $p < .001$ ; partial  $\eta^2 = .99$ ). There was also a significant effect of graph's instance on oscillation ( $F(4, 116) = 547.22$ ;  $p < .01$ ; partial  $\eta^2 = .95$ ).

All possible interactions of these variables were significant as well, with  $F > 10.27$  ( $p < .001$ ; partial  $\eta^2 > .26$ ). I report the results of the interactions and of the corresponding simple tests in Table B.1 in Appendix B.

Like with the local steepness, the oscillation of the original graphs and the oscillation of the scaled graphs were significantly correlated ( $r = .58$ ;  $p < .01$ ). Both oscillation variables were correlated with the Hurst exponents of the original graphs, though not as strongly as the local steepness ( $r = -.50$ ;  $p < .01$ ,  $r = -.72$ ;  $p < .01$ , respectively). These results support Hypotheses  $H_{5,2}$  and  $H_{5,5}$  with respect to the graphs' oscillation (that is, there is a positive correlation between forecast horizon and the oscillation of the time-scaled data graphs, and there is a negative correlation between the oscillation of the time-scaled graphs and the Hurst exponent of the original data).

The mean values of local steepness and oscillation of the scaled graphs are presented in Table 6.1. Figure 6.3 depicts the mean local steepness and oscillation of the time-scaled graphs for the different conditions of the Hurst exponent and the forecast horizon.

To conclude, I accepted Hypotheses  $H_{5,2}$  and  $H_{5,5}$ .

*Forecast dispersion* Forecast dispersion measures can indicate how unstable the market is. I extracted three dispersion measures:

1. FD1 - forecast dispersion with respect to the mean forecast of participants in each of the conditions of the experiment (the standard deviation of the absolute value of the difference between the forecast of each participant in a certain condition and the mean of all participants' forecasts in the same condition). FD1 provides information about forecast dispersion over the group.
2. FD2 - forecast dispersion with respect to the last data point in each of the conditions of the experiment (the standard deviation of the absolute value of the difference between the forecast of each participant in a certain condition and the value of the time series on day 200). FD2 provides information about dispersion with respect to the present price of each asset.
3. FError - forecast dispersion with respect to price of the time series on the required forecast day (the standard deviation of the absolute value of the difference between the forecast of each participant in a certain condition and the value of the time series on the forecast date). FError indicates participants' forecast error with respect to the produced time series.

Figure 6.4 illustrates the reference points used for the calculation of each of these error measures.

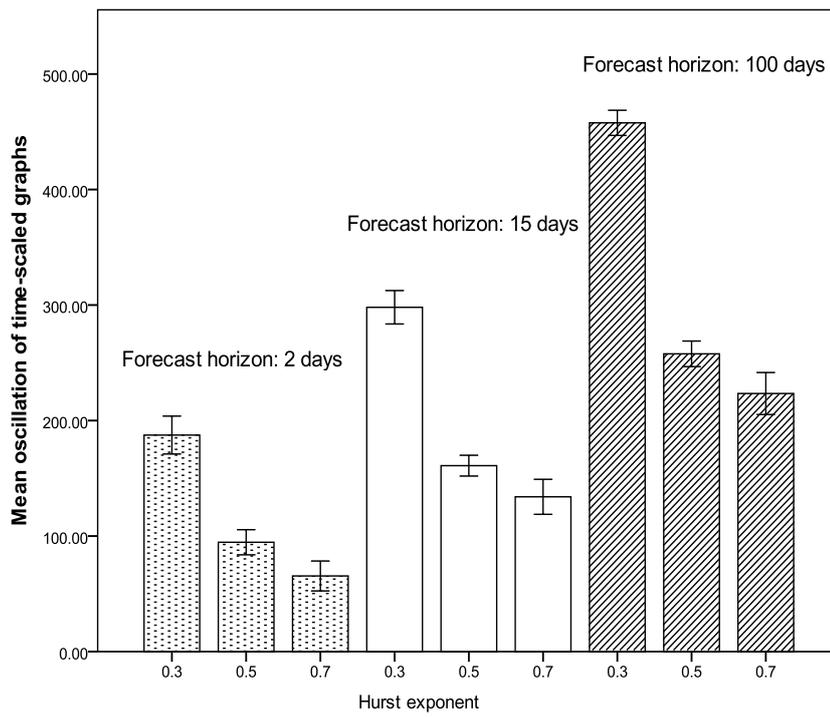
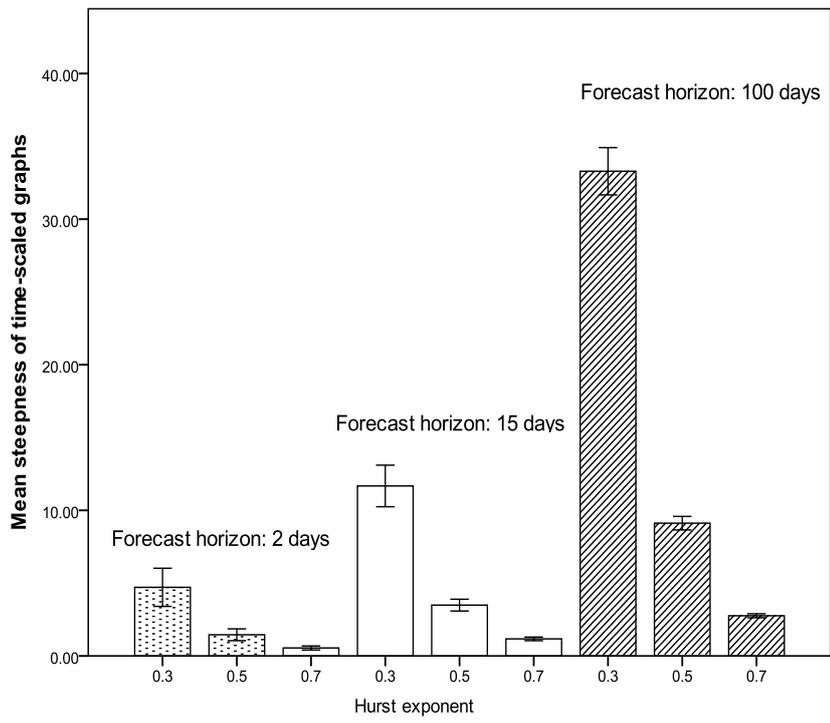


Figure 6.3 Mean steepness (upper panel) and oscillation (lower panel) of time-scaled graphs in Experiment 1.

Table 6.1 The mean local steepness (first panel) and oscillation (second panel) of time-scaled graphs in Experiment 1.

Mean local steepness		Hurst exponent			
		0.3	0.5	0.7	Mean
Forecast horizon (days)	2	4.71	1.46	0.55	2.24
		(8.12)	(2.51)	(0.86)	(5.24)
	15	11.68	3.49	1.17	5.45
		(8.85)	(2.50)	(0.77)	(6.97)
	100	33.29	9.12	2.76	15.06
		(10.06)	(2.88)	(0.86)	(14.49)
Mean		16.56	4.69	1.49	7.58
		(15.16)	(4.18)	(1.25)	(11.17)

Mean oscillation		Hurst exponent			
		0.3	0.5	0.7	Mean
Forecast horizon (days)	2	187.43 (101.74)	94.62 (67.46)	65.46 (80.33)	115.84 (98.98)
	15	298.05 (89.45)	160.95 (55.99)	133.98 (94.02)	197.66 (108.64)
	100	457.74 (67.54)	257.74 (223.41)	223.41 (112.75)	312.97 (134.15)
Mean		314.40 (141.23)	171.11 (92.79)	140.95 (116.13)	208.82 (140.44)

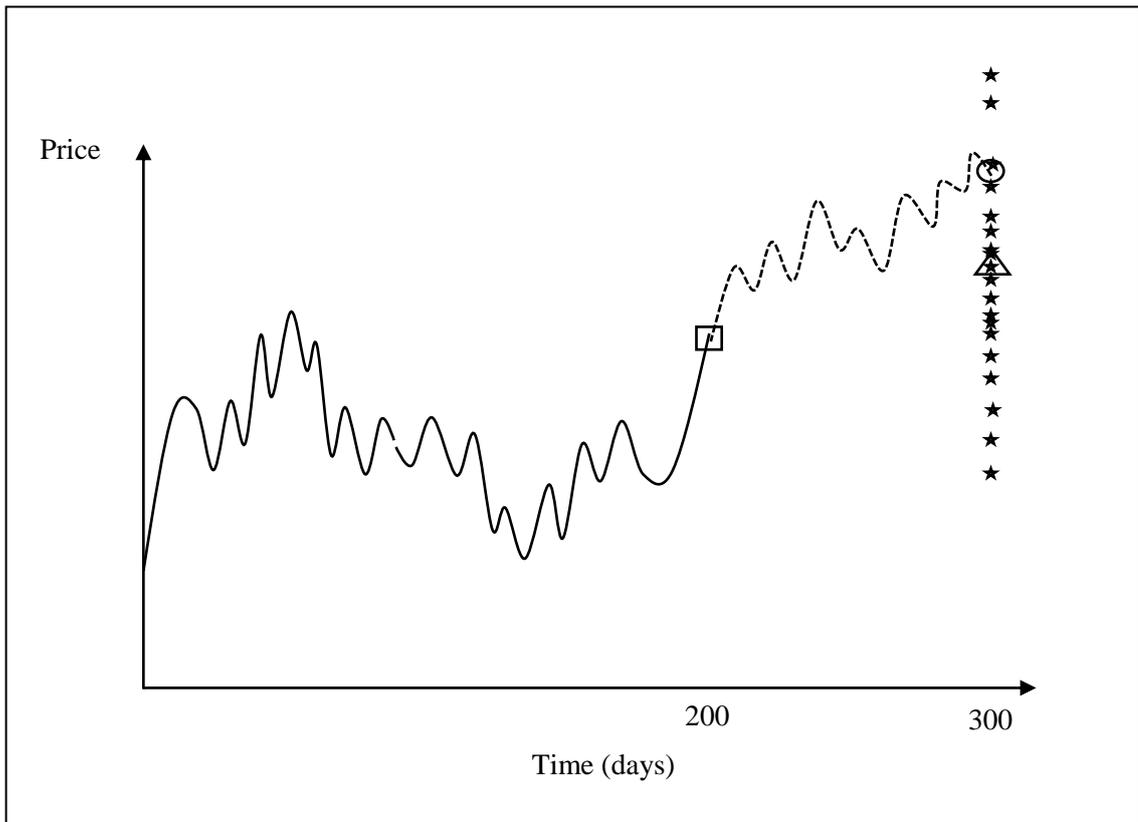


Figure 6.4 An illustration of the reference points used for the calculation of FD1, FD2, and FError when forecast horizon of 100 days: price graph against time (solid line: the data which was presented to the participant, dashed line: the continuation of the series which was not presented to the participant), participants forecasts (stars), the last data point which was presented to the participants (square), price at the required forecast date (circle), and the mean of participants' forecasts (triangle).

FD1 was calculated using the differences between participants' forecasts and the mean of participants' forecasts (triangle), FD2 was calculated using the differences between participants' forecasts and the last data point (square), and FError was calculated using the differences between participants' forecasts and the price at the required forecast date (circle).

The mean values of the three dispersion measures are presented in Table 6.2. Figure 6.5 depicts the means of the dispersion measures for the different experimental conditions.

Table 6.2 The mean forecast dispersions FD1 (first panel), FD2 (second panel), FError (third panel).

Forecast dispersion FD1		Hurst exponent			
		0.3	0.5	0.7	Mean
Forecast horizon (days)	2	26.82 (26.97)	23.81 (35.64)	19.05 (34.67)	23.23 (32.74)
	15	49.07 (42.78)	36.34 (27.31)	28.26 (23.38)	37.88 (33.31)
	100	99.61 (88.12)	65.20 (52.24)	84.76 (99.42)	83.19 (83.43)
Mean		58.50 (65.99)	41.78 (43.31)	44.02 (68.60)	48.10 (60.79)

Forecast dispersion FD2		Hurst exponent			
		0.3	0.5	0.7	Mean
Forecast horizon (days)	2	28.61	23.35	20.75	24.24
		(26.81)	(37.31)	(35.08)	(33.46)
	15	50.98	36.81	30.25	39.35
		(45.02)	(27.18)	(23.00)	(34.18)
	100	106.90	66.50	86.67	86.69
		(89.15)	(51.46)	(99.13)	(83.96)
Mean		62.16	42.22	45.89	50.09
		(68.08)	(43.72)	(68.51)	(65.22)

Forecast error		Hurst exponent			
FError		0.3	0.5	0.7	Mean
Forecast horizon (days)	2	50.74	27.89	29.03	35.88
		(45.48)	(35.03)	(33.39)	(39.68)
	15	114.16	41.50	61.12	72.26
		(110.39)	(28.94)	(33.37)	(75.07)
	100	174.23	167.10	134.38	158.57
		(149.10)	(114.93)	(105.49)	(125.51)
Mean		113.04	78.83	74.84	88.91
		(121.06)	(94.89)	(79.88)	(101.45)

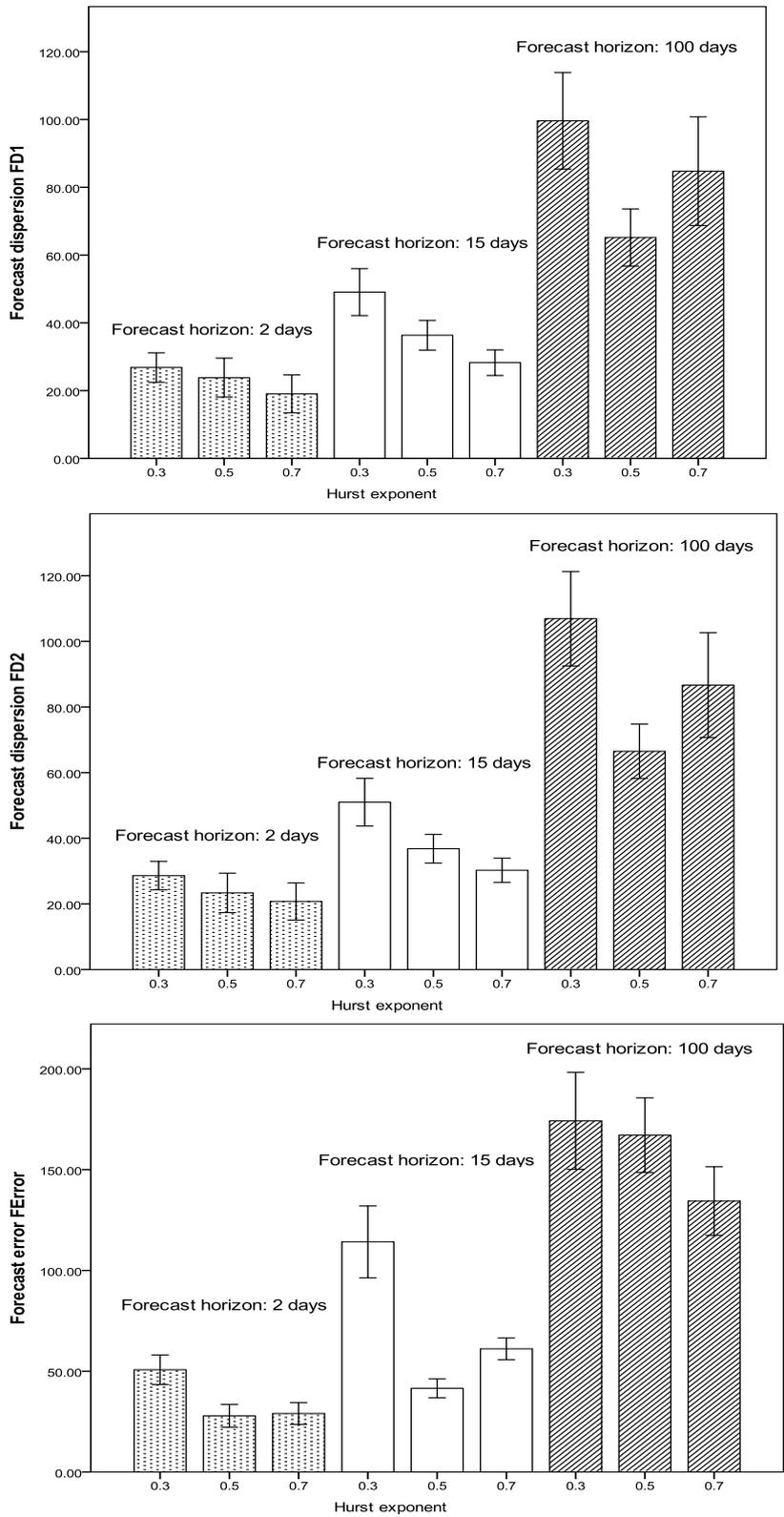


Figure 6.5 Forecast dispersion measures in Experiment 1. Upper panel: FD1. Central panel: FD2. Lower panel: FError.

To examine Hypotheses  $H_{5,3}$  and  $H_{5,6}$ , I carried out for each of the dispersion measures a three-way repeated measures ANOVA using the variables Horizon, Hurst exponent, and Instance as within-participant variables. I report here the results of the ANOVA of FD1. The results of the ANOVAs of FD2 and FError were similar. I report them in Table B.2 in Appendix B.

For FD1, sphericity assumption was violated for all variables apart from the Hurst exponent and the instance. The analysis revealed that FD1 was larger when the Hurst exponent was smaller ( $F(2, 58) = 10.32$ ;  $p < .001$ ; partial  $\eta^2 = .26$ ) and when forecast horizon was longer ( $F(1.39, 40.42) = 84.67$ ;  $p < .001$ ; partial  $\eta^2 = .75$ ). There was also a significant effect of instance on forecast dispersion, indicating that participants reacted to graph characteristics other than the Hurst exponent as well ( $F(4, 116) = 16.91$ ;  $p < .001$ ; partial  $\eta^2 = .37$ ).

All possible interactions between these variables were significant, with  $F > 5.44$  ( $p \leq .002$ ; partial  $\eta^2 > 0.16$ ). I report the results of the interactions and of the corresponding simple tests in Table B.1 in Appendix B.

The correlations between forecast dispersion measures and local steepness or oscillation of the scaled graphs were higher than those with the same properties of the original graphs. Significant correlations were obtained also between forecast dispersion measures and forecast horizon. The correlations are summarised in Table 6.3. These results support Hypotheses  $H_{5,3}$  and  $H_{5,6}$  (dispersion of forecasts is positively correlated with the required forecast horizon, and the dispersion of forecasts is negatively correlated with the Hurst exponents of the original graphs and positively correlated with the local steepness and oscillation of the data graphs).

Table 6.3 Correlations between forecast dispersion measures, local steepness of graphs, oscillation, and forecast horizon.

Forecast dispersion measure	Original graphs			Time-scaled graphs		Forecast horizon
	Hurst exponent	Local steepness	Oscillation	Local steepness	Oscillation	
FD1	r = -.10	r = .12	r = .22	r = .29	r = .44	r = .42
	p < .01	p < .01	p < .01	p < .01	p < .01	p < .01
FD2	r = -.11	r = .13	r = .23	r = .31	r = .45	r = .43
	p < .01	p < .01	p < .001	p < .01	p < .01	p < .01
FError	r = -.15	r = .17	r = .15	r = .34	r = .46	r = .50
	p < .01	p < .01	p < .01	p < .01	p < .01	p < .01

*Decision parameters* To examine Hypotheses H<sub>5,7</sub>, I extracted the resultant share number.

For each asset, resultant share number was defined to be 0 if participant chose the option ‘sell’, 1 if participant chose the option ‘hold’, and 2 if participant chose the option ‘buy’. I carried out a three-way repeated measures ANOVA using the same variables used before as within-participant variables. The analysis failed to find a significant effect of forecast horizon and the Hurst exponent on the resultant share number. I found a significant effect of graph instance on the resultant share number ( $F(2.25, 62.92) = 7.02$ ;  $p < .001$ ; partial  $\eta^2 =$

.2) and a weak but significant interaction between graph instance and the Hurst exponent ( $F(2.89, 80.93) = 2.88$ ;  $p = .04$ ; partial  $\eta^2 = .09$ ). Tests of simple effects showed that the effect of instance was significant only for low and high Hurst exponents (for  $H = 0.3$ ,  $F(4, 25) = 2.99$ ;  $p = .04$ ; partial  $\eta^2 = .32$ , for  $H = 0.7$ ,  $F(4, 25) = 2.92$ ;  $p = .03$ ; partial  $\eta^2 = .32$ ).

I expected resultant share number to depend on participants' forecasts. The analysis revealed that resultant share number was significantly and positively correlated with the difference between the participant's forecast and the last data point ( $r = .53$ ;  $p < .01$ ). This establishes a connection between participants' expectations and actions: the higher the difference between the forecast and the price at present was, the larger was participants' tendency to advise buying more shares. When participants thought that the prices would decrease, they tended to advise that shares be sold. This provides support for Hypothesis H<sub>5,7</sub> (people's trading behaviour depends on their forecasts).

### *Discussion*

Experiment 1 was performed to analyse the effects of the Hurst exponent and forecast horizon on financial forecasts and decisions. Participants were asked to imagine that they were financial analysts and that they had clients with different trading frequencies.

Participants were presented with a set of 45 graphs, each representing the price series of each of their client's assets. On each trial, participants were informed that they would have to make a forecast for a certain date and were asked to scale the graph in the way that they considered most appropriate for that purpose. Afterwards, they were asked to make the forecast and to advise their clients whether to buy, sell, or hold their assets. I manipulated the Hurst exponent of the data graphs and the forecast horizons.

The results indicated that, when asked to make financial forecasts, participants chose to scale the graphs rather than leave them with the initially presented time interval. Their choices had a relatively large variance. I, therefore, accepted Hypothesis H<sub>5,1a</sub>, supporting the Heterogeneous Market approach of Peters (1995) and Müller et al. (1993).

In line with Corsi's argument (2009), I found that participants chose to scale the graphs in a way that was correlated with the required forecast horizon and that, when forecast horizons were larger, scaled graphs had higher local steepness and oscillation than the originals.

These results supported Hypotheses  $H_{5,1b}$  and  $H_{5,2}$ . In addition, the results indicate that longer forecast horizons result in larger forecast dispersions, and so support Hypothesis  $H_{5,3}$ .

The results indicate that the geometric properties of the data graphs affect people's scaling and decisions as well. People's chosen time-scale depended on the Hurst exponents of the graphs. In particular, they tended to "zoom-in" more when Hurst exponents were smaller.

That is, when the Hurst exponent was small, people chose to look at a smaller time-period. I, therefore, accept Hypothesis  $H_{5,4}$ .

The local steepness and oscillation of the scaled graphs were positively correlated with the local steepness and oscillation of the original graphs, and negatively correlated with the Hurst exponents of the original graphs. Therefore, I accept Hypothesis  $H_5$ , which suggests that the way that participants choose to see the market preserves geometric properties of the data.

As a result, forecast dispersion measures were negatively correlated with the Hurst exponents of the data graphs. Thus, I accepted Hypothesis  $H_{5,6}$ . According to Athanassakos and Kalimipalli (2003), there is a strong correlation between analysts' forecast dispersion and future return volatility. Therefore, the way people choose to see price series serves as one of the mechanisms that preserve their structure.

Finally, there was a significant correlation between participants' forecasts and final share number. I accepted Hypothesis  $H_{5,7}$ .

## **Experiment 2**

The primary aim of Experiment 2 was to examine the effect of the Hurst exponent of a time series on the size of a chosen moving average filter and on financial forecasts from fractal

graphs. I hypothesised that the way that people perceive fractal graphs has a role in stabilising the market. More precisely, I hypothesised that people select moving average filters which preserve the geometric properties of the price graphs. The secondary aim of the experiment was to examine the effect of the density of the required forecasts on the chosen sizes of a moving average filter. I hypothesised that chosen smoothing factors are smaller when required forecast densities are larger

I tested the following hypotheses:

H<sub>5,8a</sub>: People use smoothing when making financial forecasts and decisions. In particular, the variance of the choices of averaging windows is substantial with respect to the mean, that is, at least 50% of the mean

H<sub>5,8b</sub>: Chosen smoothing factors are smaller when Hurst exponents are smaller. That is, people zoom-in more and present shorter time intervals when graphs with lower Hurst exponents are presented.

H<sub>5,9a</sub>: There is a negative correlation between the Hurst exponent of the original data and the local steepness and oscillation of the smoothed graphs.

H<sub>5,9b</sub>: There is a positive correlation between the local steepness and oscillation of the smoothed data graphs and the original ones.

H<sub>5,10</sub>: The local steepness and oscillation of forecast sequences made from fractal graphs are positively correlated with the local steepness and oscillation of the smoothed graphs, respectively, and negatively correlated with the Hurst exponent of the data graphs.

H<sub>5,11</sub>: Chosen smoothing factors are smaller when required forecast densities are larger (people zoom-in more when forecast densities are high).

H<sub>5,12</sub>: There is a positive correlation between the local steepness and oscillation of the smoothed data graphs and the required density of forecasts.

$H_{5,13}$ : Local steepness and oscillation of the forecasts is positively correlated with the required density of the forecast.

I presented participants with a sequence of time series. Each one was presented on a separate trial. At the beginning of each trial, two identical copies of the same time series were presented on the same axes. Both copies remained visible during the whole duration of each trial. However, the task window enabled participants to smooth one of the graphs. The other graph remained fixed. That made it possible for the participants to smooth each price data graph while seeing the original data. Participants were asked to choose the smoothness level they considered the most appropriate for making financial decisions from it, and then to make a forecast series based on the smoothed graph. I manipulated two main variables: the Hurst exponent of the original data graphs (and thus also their local steepness and oscillation), and the number of required forecast points, or, equivalently, the forecast density. Figure 6.6 depicts the task window of Experiment 2. It shows a graph of the original data and the corresponding smoothed graph (on the same axis).

### *Method*

*Participants* Thirty-four people (15 men and 19 women) with an average age of 26.4 years acted as participants. They were paid a flat fee of £3.00.

*Stimulus materials* Stimulus graphs included six sets of five time series with Hurst exponents  $H = 0.3, 0.4, 0.5, 0.6, \text{ and } 0.7$ . The time series were produced using the Spectral method described by Saupe (Peitgen and Saupe, 1988). All of the time series included 3600 points and were presented to participants as asset price graphs.

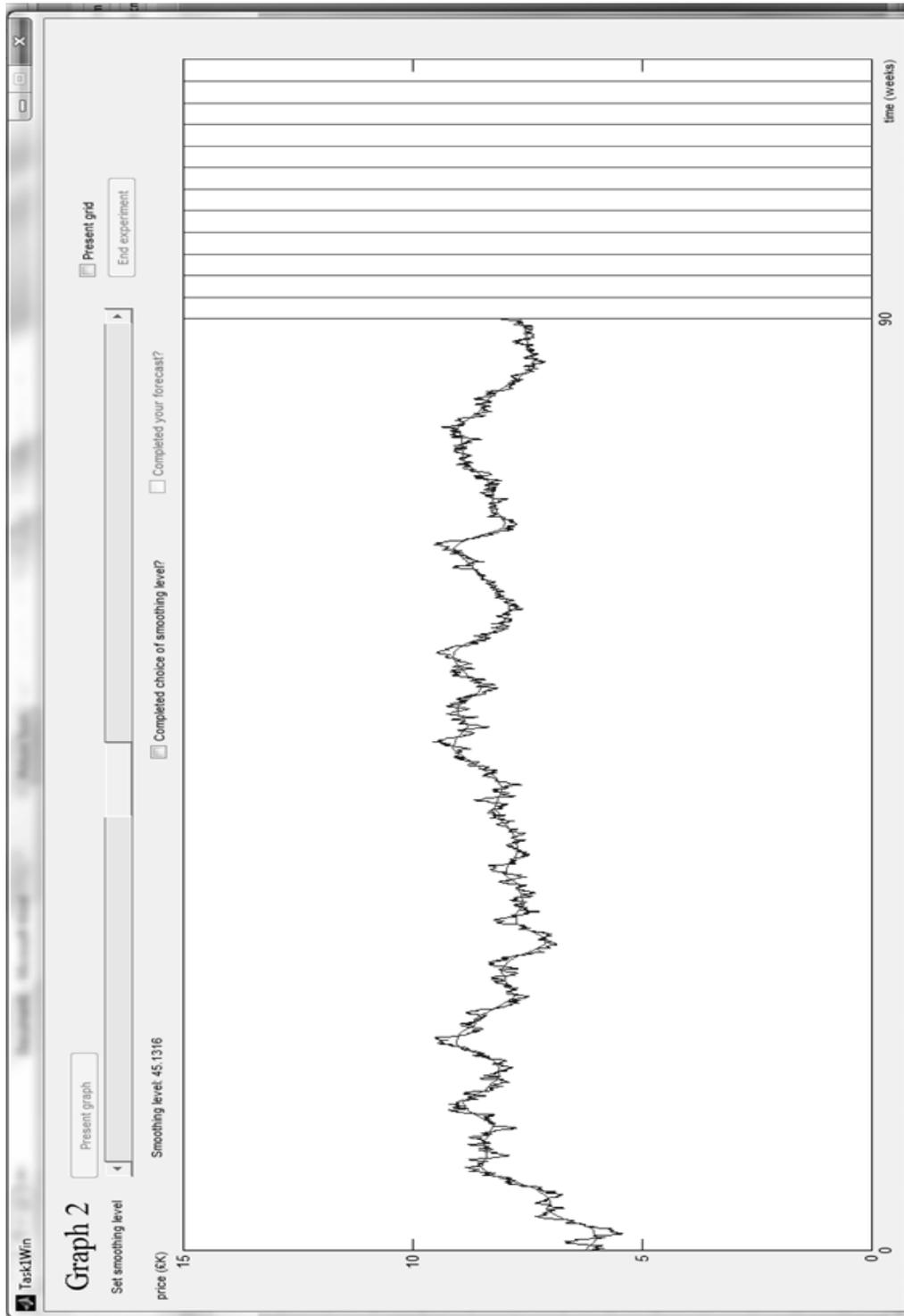


Figure 6.6 The task window of Experiment 2: a price graph (the jagged lined) and a corresponding smoothed graph (the smoother line).

*Experimental programme* Stimulus graphs were presented using a Matlab programme. The experimental programme enabled participants to apply an averaging filter to the price graphs, while viewing the original price graphs and to make forecasts on pre-specified dates.

Application of the averaging filter was done using a slider. The filter's range was from an averaging window of size 2 (averaging over every two adjacent elements of the series) to averaging over the whole series, the latter resulting in a constant line. To enable participants to both express fine details at the lower end of the scale and reach the maximum averaging, the slider was exponentially calibrated.

The experimental programme required participants to make forecasts on dates designated by vertical lines. There were 6, 12, 24, or 36 lines. In each task, participants could change smoothing level until they clicked the button "Completed choice of smoothing level?". They could edit their forecasts by clicking the mouse again on any bar, until they clicked the button "Completed your forecast?" (Figure 6.6).

*Design* Participants were presented with 23 graphs: three familiarisation graphs and 20 experimental graphs. Only experimental graphs were taken into account during the analysis stage. Each graph required two responses. The first response was a choice of smoothing level. The second response was to forecast the asset's future prices.

For each participant, four graphs with each value of Hurst exponent ( $H=0.3, 0.4, 0.5, 0.6, 0.7$ ) were randomly chosen from the stimulus sets. For each value of Hurst exponent, the density of the required forecast was manipulated, and was set to a value of 6, 12, 24, or 36 forecasts within a three-year period. That gave rise to a five (Hurst exponent) by four (forecast density) design. Ordering of trials with different Hurst exponents and forecast densities was random.

*Procedure* Participants were asked to read the following instructions:

“In the following task, you are asked to imagine that you are a financial analyst working at an investment company. Your clients ask you to give them a three year forecast. Each client asks for a forecast of a different resolution: some clients need a monthly forecast (a total of 36 points), some require a forecast point every 6 months (a total of 6 points), and some are interested in an intermediate number of forecast points (a total of 12 points or 24 points).

You will be presented with a series of 3 practice graphs and 20 experiment graphs representing prices of different assets. The programme will enable you to set the smoothness level of the data graphs. You are asked:

1. to look at the graphs carefully,
2. for each of the graphs, to determine the smoothness level you consider the most appropriate for making financial decisions from it,
3. to predict the prices on a series of time points based on the smoothed graph. The number of forecasts will be 6, 12, 24, or 36 points according to the request obtained from each of your clients.”

Participants chose a smoothness level of data graphs by dragging a slider. The smoothed graph was presented in red. The original graph was presented in blue.

Forecasts were made by clicking a mouse at specific dates, designated by vertical lines. Participants had to complete the forecasts on all vertical lines (dates) before they could continue to the next graph.

### *Results*

Participants whose means of smoothing level choices were more than two standard deviations greater than that of the average for the rest of the group were excluded from the analysis. This reduced the size of the sample from 34 to 32 participants. Three additional extreme measurements (out of the original  $20 * 34 = 680$  measurements), in which

participants chose smoothing levels more than four standard deviation greater than that of the mean of the experimental condition were removed as well. Therefore, I used 637 graphs for the analysis.

The variables of primary interest were chosen the smoothing factors, the local steepness and oscillation of smoothed data graphs and participants' forecasts. Chosen smoothing factors indicate the resolution at which participants preferred to perceive the market. Local steepness and the oscillation of graphs can be used to measure similarity between forecasts and the original and smoothed data. Such correlations may suggest perception as a mechanism of preservation of parameters of asset graphs. The results are presented in Table 6.5.

*Choice of smoothness level* The mean smoothness level participants chose was 59.09. The standard deviation was larger than the mean: 82.61. A t-test performed on participants choices of smoothness levels showed that it was significantly larger than 1 (a trivial filter):  $t(636) = 17.76$  ( $p < .01$ ). These results support Hypothesis  $H_{5,8a}$  (The variance of the choices of averaging windows is substantial with respect to the mean).

To examine Hypotheses  $H_{5,8b}$  and  $H_{11}$ , I carried out a two-way repeated measures ANOVA on chosen smoothness level using the Hurst exponent and the forecast density as within-participant variables. Here, and everywhere else, when Mauchly's sphericity assumption is violated, I report results of the Huynh-Feldt test. Mauchly's sphericity assumption was violated for both the Hurst exponent and the required number of points. The Huynh-Feldt test showed that Hurst exponent had a significant effect on the chosen smoothing factor ( $F(4, 65.93) = 3.12$ ;  $p = .045$ ; partial  $\eta^2 = .10$ ). However, this effect was quadratic and not linear ( $F(1, 29) = 9.54$ ;  $p = .04$ ; partial  $\eta^2 = .25$ ). The chosen smoothing factor was larger for  $H > 0.5$  and  $H < 0.5$  than for  $H = 0.5$ . That means that participants applied larger smoothing factors on graphs that did not satisfy the assumptions of the random walk model than on those that did satisfy those assumptions. These results support Hypothesis  $H_{5,8b}$  (people

zoom-in more and present shorter time intervals when graphs with lower Hurst exponents are presented) only for H values smaller than or equal to 0.5. The chosen smoothing factor was larger when forecast density was smaller ( $F(3, 53.12) = 6.54$ ;  $p = .004$ ; partial  $\eta^2 = .18$ ). This was a linear effect ( $F(1, 29) = 10.17$ ;  $p = .003$ ; partial  $\eta^2 = .26$ ) and supports Hypothesis H<sub>11</sub> (people zoom-in more when forecast densities are high). Figure 6.7 depicts the mean chosen smoothing factors against the Hurst exponent of the graphs and the required forecast density.

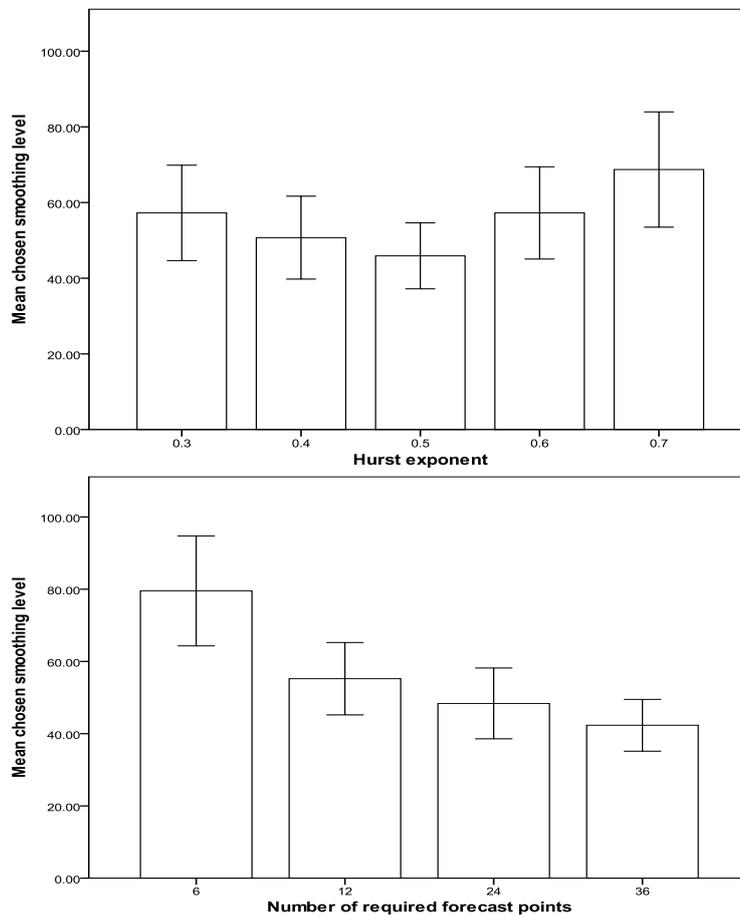


Figure 6.7 Mean of chosen smoothness levels against the Hurst exponent of the given graphs (upper panel) and forecast density, measured by the number of required forecast points in the forecasting period (lower panel). Standard error is indicated with the bars.

Table 6.4. The mean chosen smoothness levels (first panel), local steepness of forecasts (second panel), and oscillation of participants' forecasts (third panel) in Experiment 2.

The mean chosen smoothness levels		Hurst exponent					Mean
		0.3	0.4	0.5	0.6	0.7	
Required number of forecast points	6	65.68 (74.25)	86.21 (99.93)	54.31 (53.52)	73.35 (63.09)	116.62 (150.47)	79.5 (96.56)
	12	57.05 (54.19)	43.59 (49.41)	59.80 (57.20)	70.00 (99.12)	44.94 (37.93)	55.07 (63.09)
	24	61.25 (95.47)	36.04 (33.86)	34.50 (36.66)	41.73 (33.90)	66.40 (76.50)	48.00 (61.71)
	36	41.61 (41.79)	40.30 (43.07)	32.90 (31.16)	40.00 (46.31)	59.09 (61.28)	42.78 (46.022)
Mean		56.32 (69.21)	51.54 (64.63)	45.31 (46.89)	56.12 (66.45)	71.76 (94.67)	56.21 (82.61)

The mean local steepness of forecasts		Hurst exponent					
		0.3	0.4	0.5	0.6	0.7	Mean
Required number of forecast points	6	0.41 (0.23)	0.28 (0.14)	0.22 (0.10)	0.21 (0.11)	0.18 (0.07)	0.26 (0.16)
	12	0.58 (0.25)	0.40 (0.17)	0.30 (0.11)	0.28 (0.13)	0.20 (0.09)	0.35 (0.20)
	24	0.77 (0.33)	0.60 (0.33)	0.48 (0.20)	0.39 (0.24)	0.29 (0.14)	0.51 (0.31)
	36	0.82 (0.45)	0.67 (0.32)	0.56 (0.31)	0.58 (0.65)	0.39 (0.21)	0.60 (0.43)
Mean		0.65 (0.37)	0.49 (0.30)	0.39 (0.24)	0.37 (0.38)	0.26 (0.16)	0.43 (0.31)

Forecasts'		Hurst exponent					
		0.3	0.4	0.5	0.6	0.7	Mean
oscillation							
Required	6	1.69	1.48	1.17	1.01	1.01	1.27
number		(0.84)	(0.90)	(0.57)	(0.48)	(0.43)	(0.72)
of							
forecast	12	2.19	1.60	1.22	1.18	1.00	1.45
points		(0.85)	(0.93)	(0.61)	(0.35)	(0.47)	(0.83)
	24	2.01	1.79	1.48	1.21	1.19	1.54
		(0.75)	(0.88)	(0.54)	(0.65)	(0.71)	(0.78)
	36	1.91	1.82	1.57	1.49	1.34	1.63
		(0.83)	(0.71)	(0.69)	(0.98)	(0.517)	(0.78)
Mean		1.95	1.67	1.36	1.22	1.14	1.47
		(0.83)	(0.86)	(0.62)	(0.67)	(0.55)	(0.72)

Application of similar smoothing filters on graphs with high and low Hurst exponents may result in different local gradients and oscillation. What were the local steepness and oscillation of the resultant, smoothed graphs and how did they correlate with the Hurst exponent of the data?

*Properties of smoothed data graphs* To examine Hypotheses  $H_{5,9}$  and  $H_{5,12}$ , I extracted the local steepness and oscillation of the original data graphs and of the smoothed graphs. The measure for local steepness of a time series was the average of the absolute value of the gradient at each series point. The oscillation of each series was defined as the difference between its maximum and minimum values.

To assess the effect of the smoothing task on the data, I carried out a three-way repeated measures ANOVA on the local steepness of the data graphs, using state (before/after smoothing), the Hurst exponent and forecast density as within-participant variables. Mauchley's test of sphericity assumption was violated for the Hurst exponent and forecast density. The local steepness of graphs was smaller after smoothing ( $F(1, 29) = 346.9$ ;  $p < .001$ ; partial  $\eta^2 = .92$ ) and when Hurst exponent was larger ( $F(4, 37.06) = 825.60$ ;  $p < .001$ ; partial  $\eta^2 = .97$ ). No other significant effects were found. I report results about the interaction obtained in Table B.3 in Appendix B.

The correlation between the Hurst exponent and the local steepness of the smoothed graphs was  $r = -.51$ ;  $p < .01$  (the correlation between Hurst exponent and local steepness of the original data graphs was  $r = -.94$ ;  $p < .01$ ). The correlation between the local steepness of the original and smoothed data graphs was  $r = .52$ ;  $p < .01$ .

To assess the effect of the task variables on the oscillation of the data, I carried out a three-way repeated measures ANOVA on the oscillation of the data graphs, using the same variables as before. Mauchley's sphericity assumption was violated for the Hurst exponent and number of required forecast points. The analysis revealed that oscillation was larger in the original data ( $F(1, 29) = 163.82$ ;  $p < .001$ ; partial  $\eta^2 = .85$ ) and when Hurst exponent

was smaller ( $F(2.5, 72.49) = 188.91; p < .001; \text{partial } \eta^2 = .87$ ). There was a significant interaction between state and the Hurst exponent ( $F(1.71, 49.55) = 129.45; p < .001; \text{partial } \eta^2 = 0.82$ ). In addition, there were small interaction effects between forecast density and state ( $F(2.36, 68.29) = 3.46; p = .03; \text{partial } \eta^2 = .11$ ) and between forecast density and the Hurst exponent ( $F(5.12, 148.45) = 5.38; p = .03; \text{partial } \eta^2 = .16$ ). I report the relevant tests of simple effects in Table B.3 in Appendix B.

The correlation between Hurst exponent and the oscillation of the smoothed data graphs was  $r = -.61; p < .01$ . (The correlation between Hurst exponent and the oscillation of the original data graphs was  $r = -.80; p < .01$ ). The correlation between the oscillations of the smoothed and original data graphs was  $r = .88; p < .01$ .

Figure 6.8 depicts the local steepness and mean oscillation of the smoothed data graph for the different values of the Hurst exponent and the different numbers of required forecast points. These results support Hypotheses  $H_{5,9a}$  and  $H_{5,9b}$  (there is a negative correlation between the Hurst exponent of the original data and the local steepness and oscillation of the smoothed graphs). However, due to the lack of main effects of forecast density on properties of the smoothed graphs, I reject Hypothesis  $H_{5,12}$  (about the correlation between the local steepness and oscillation of the smoothed data graphs and the required density of forecasts).

*Properties of participants' forecasts* To examine Hypotheses  $H_{5,10}$  and  $H_{5,13}$ , I extracted local steepness and oscillation of the forecast series and compared them to those of the data and the smoothed data.

To analyse local steepness of the forecasts, I carried out a two-way repeated measures ANOVA on the steepness of participants' forecasts using the Hurst exponent and forecast density as within-participant variables. Huynh-Feldt test showed that local steepness of forecasts was larger when Hurst exponent of the data graphs was smaller ( $F(3.05, 88.54) = 41.15; p < .01; \text{partial } \eta^2 = .59$ ) and when the forecast density was larger ( $F(1.78, 51.65) = 30.94; p < .01; \text{partial } \eta^2 = .52$ ).

The correlation between the local steepness of the forecasts and the Hurst exponent of the smoothed graphs was  $r = -0.39$  ( $p < .01$ ). Similar (positive) correlations were found between the steepness of the forecasts and the local steepness of the data before or after the smoothing ( $r = 0.39$ ;  $p < .01$ , and  $r = .33$ ;  $p < .01$  respectively).

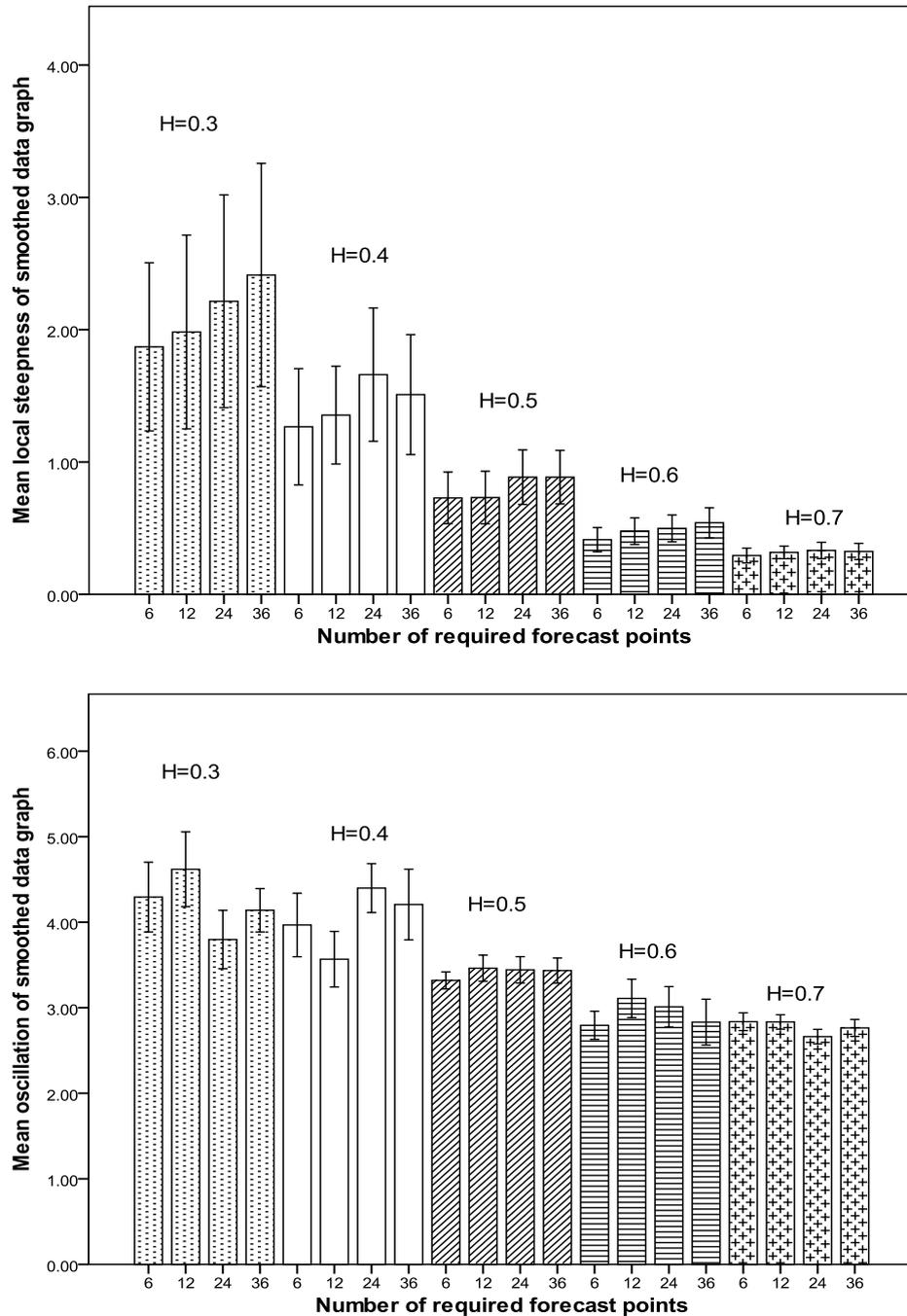


Figure 6.8 The mean local steepness (upper panel) and oscillation (lower panel) of smoothed data graphs for each of the experimental conditions

Controlling for the Hurst exponent (and local steepness) of the data graphs, the correlation between the steepness of the forecasts and the steepness in the smoothed data was significant ( $r = .16$ ;  $p < .01$ ). However, controlling for the Hurst exponent of the data graphs and the local steepness in the smoothed data, the correlation between the steepness in the forecasts and the steepness in the original data graphs was insignificant ( $p = .13$ ). That suggests that participants did indeed to make their forecasts according to the smoothed graphs, as the instructions required them to do.

The correlation between forecast density and the local steepness of the forecasts was  $r = 0.41$  ( $p < .01$ ).

To analyse the oscillation of the forecasts, I carried out a two-way repeated measures ANOVA using the same variables. Mauchley's sphericity assumption was violated for the Hurst exponent, but not for the number of required forecast points. Huynh-Feldt test showed that the oscillation of the forecasts was larger when the Hurst exponent of the data was smaller ( $F(3.42, 99.08) = 37.02$ ;  $p < .01$ ; partial  $\eta^2 = .56$ ). In addition, the oscillation of the forecasts was larger when the required forecast density was larger ( $F(3, 87) = 8.80$ ;  $p < .01$ ; partial  $\eta^2 = .23$ ).

The correlation between the oscillation of the forecasts and the Hurst exponent of the smoothed graphs was  $r = -.38$  ( $p < .01$ ). Similar (positive) correlations were found between the oscillation in the forecasts and the oscillation in the data both before and after smoothing ( $r = .43$ ;  $p < .01$ , and  $r = .40$ ;  $p < .01$  respectively). Controlling for the Hurst exponent of the data graphs and the data oscillation, the correlation between the oscillation of the forecasts and smoothed data was small but significant ( $r = .08$ ;  $p = .04$ ). However, controlling for the Hurst exponent of the data graphs and the oscillation of the smoothed data, the correlation between the steepness of the forecasts and original data graphs was insignificant ( $p = .08$ ). As with the case of the local steepness, these results support the hypothesis that participants

indeed made their forecasts according to the smoothed graphs, as the instructions required them to.

The correlation between the number of required forecast points and the oscillation in the forecast sequence was  $r = .16$  ( $p < .01$ ).

These results support Hypotheses  $H_{5,10}$  and  $H_{5,13}$ . (That is, the local steepness and oscillation of forecast sequences are positively correlated with the local steepness and oscillation of the smoothed graphs, negatively correlated with the Hurst exponent of the data graphs, and positively correlated with the required density of the forecast).

Figure 6.9 and Figure 6.10 presents the mean values of the local steepness and oscillation in the forecasts against the Hurst exponent of the data and the number of required forecast points.

### *Discussion*

Experiment 2 aimed to elucidate the way that people perceive financial graphs and make financial forecasts from them. Participants were presented with a set of 20 graphs, and were asked to look at each one to determine the smoothness level they considered the most appropriate for making financial decisions. They were then asked “to predict the prices on a series of time points based on the smoothed graph”. I manipulated both the Hurst exponent of the data graphs, and the density of required forecast points.

The results showed clearly that participants considered graphs smoothed with a non-trivial averaging filter more appropriate for making financial decisions than the raw data. Chosen window sizes had a large variance, thereby supporting hypothesis  $H_{5,8a}$ .

In spite of the large variance of chosen smoothness factors, participants’ choices of smoothness levels were far from random: they depended linearly on forecast density, and exhibited a U-shape dependence on the Hurst exponents of the given graphs. I, therefore accepted Hypothesis  $H_{5,8b}$  for H values smaller or equal to 0.5 and Hypothesis  $H_{5,11}$ .

However, the most important aspect of the smoothing process was not the size of the chosen filter, but rather the visible properties it produced in the resulting smoothed graphs. The analysis revealed that the local steepness and oscillation of the smoothed graphs were significantly different than those in the original data. Furthermore, they were correlated with the Hurst exponent, local steepness and oscillation of the data graphs. This supports both parts of hypotheses  $H_{5,9}$ .

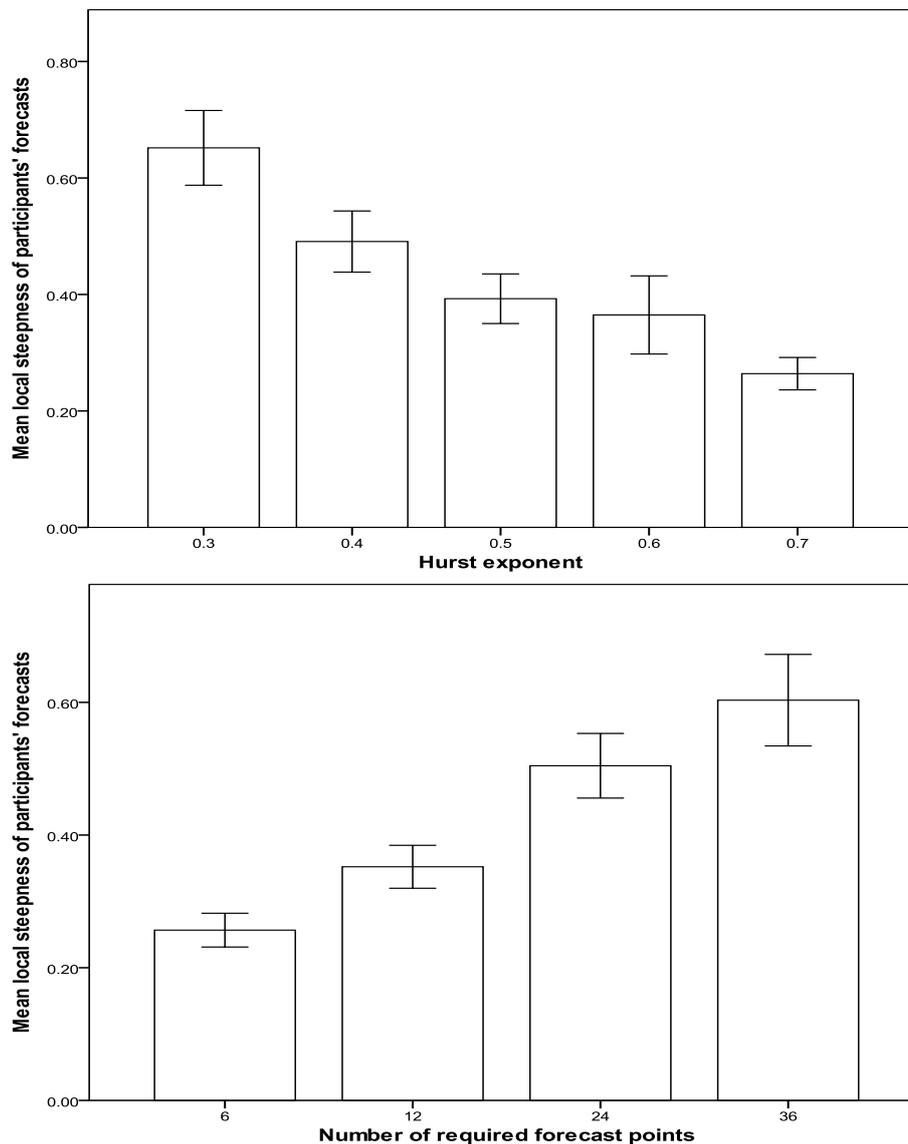


Figure 6.9 The mean steepness of forecasts plotted against the Hurst exponent of the graphs (upper panel) and plotted against the number of required forecast points in the forecasting period (lower panel). Bars show standard error measures.

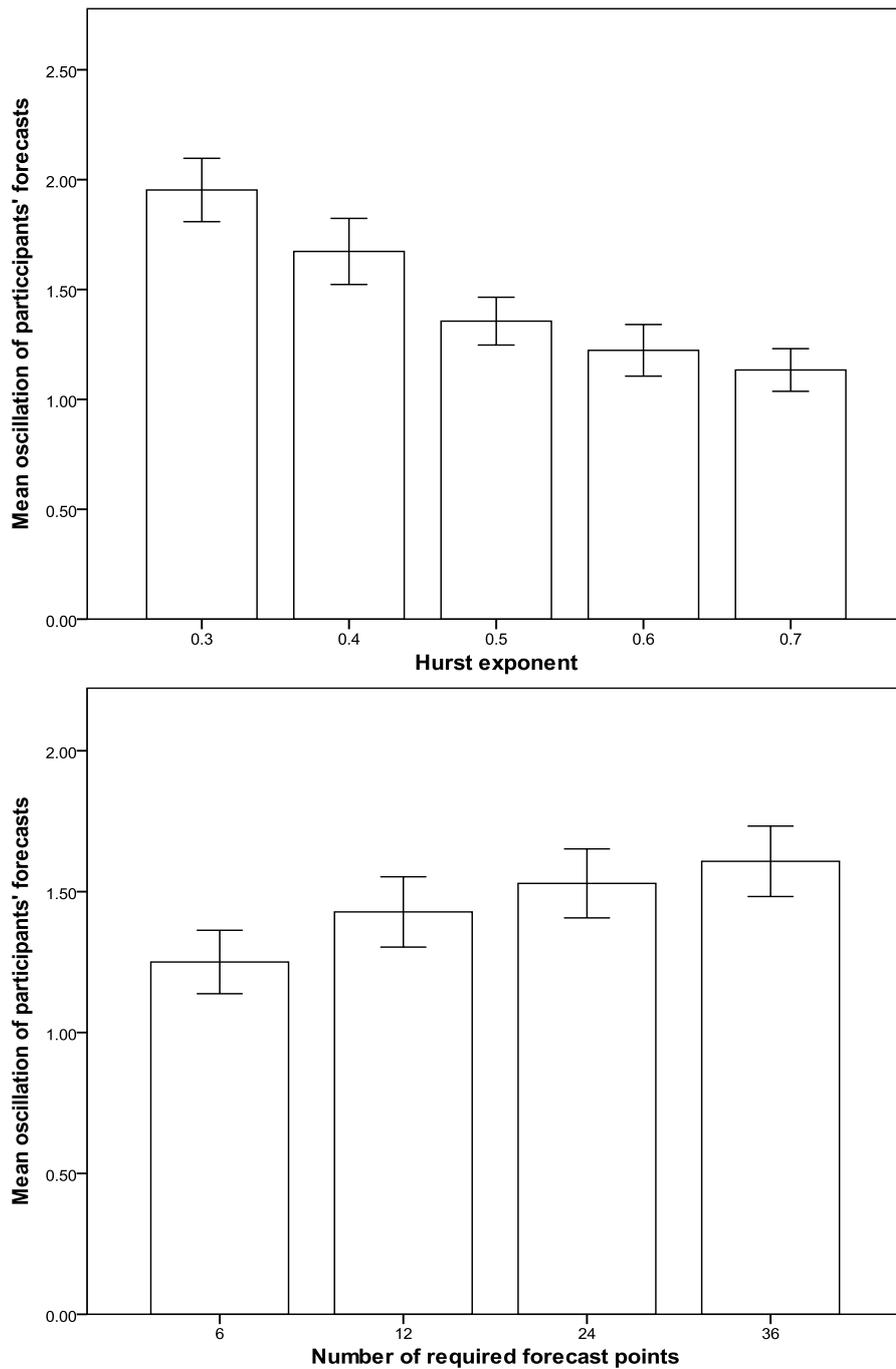


Figure 6.10 The mean steepness (upper panels) and oscillation (lower panels) of forecasts plotted against the Hurst exponent of the graphs (left panels) and plotted against the number of required forecast points in the forecasting period (right panels). Bars show standard error measures.

On the other hand, the analysis failed to show a significant effect of the number of required forecasts on the local steepness of the smoothed graphs or their oscillation. That means that the way people perceived the graphs did not depend on the density of the required forecasts. I, therefore, rejected Hypothesis H<sub>5,12</sub>.

Nevertheless, both manipulated variables – the Hurst exponent of the data graphs and forecast density – affected properties of participants' forecasts. Their average steepness and oscillation were positively correlated with those in the data, and negatively correlated with the Hurst exponents of the original graphs. I, therefore, accepted Hypothesis H<sub>5,10</sub>. As with scaling, the way people used moving window averaging and then made forecasts preserved the geometric properties of the data.

Local steepness and oscillation of forecasts were positively correlated with forecast density. I, therefore, accepted Hypothesis H<sub>5,13</sub>. However, as Hypothesis H<sub>5,12</sub> was rejected, I interpret the dependence of forecasts on forecast density as a bias resulting from the task rather than from the way participants perceived the data: a larger number of required forecasts encouraged participants to produce steeper forecasts with larger amplitudes. This result is in line with the correlation that I found between forecast noise and the number of forecast points in Chapter 4.

## **Conclusions**

In the book “An Engine, Not a Camera, How Financial Models Shape Markets”, MacKenzie (2006, page 12) wrote: “Financial economics, I argue, did more than analyze markets; it altered them. It was an “engine” [...]: an active force transforming its environment, not a camera passively recording it”. MacKenzie analyses the way financial theories developed and affected the markets. However, I argue that not only theories affect markets. Rather, I suggest that the way people perceive and react to financial data can affect price series. In particular, this behaviour stabilises markets enough to make financial theories and forecast methods feasible.

This research has dealt with the way that people use highly popular financial data presentation techniques – scaling and moving window averaging. Both techniques have been related via financial models to the formation of fractal or fat-tail price series (Peters, 1995; Müller et al. 1993; Corsi, 2009; De Grauwe and Grimaldi, 2005). Scaling was discussed in the context of trading horizons. I showed here that, apart from the trading horizon, scaling and moving window averaging depend on geometrical properties of the perceived data graphs. Indeed, the effect of the perception of volatility in price series on the market has been shown to be important by Manzan and Westerhoff (2005). However, they studied this effect in the context of over- and under-reaction. My results indicate that, though there is a large variability among participants in choice of scaling and moving window averaging parameters, there is still a correlation between the local steepness and oscillation of the transformed data graphs, and the local steepness, oscillation, and the Hurst exponents of the original price graphs. This emphasises that the way that people perceive the market is not as passive as a camera – yet, it does preserve important qualities of the data.

However, people are more than data preservation machines; they are the engine of the market. These experiments reveal that the way people make forecasts from data presented according to their own choice, corresponds to properties of the data as well. Three different forecast dispersion measures (Experiment 1) and noise measures (Experiment 2) were positively correlated with the local steepness and oscillation of the data graphs. However, forecast dispersion is correlated with volatility of returns (Athanasakos and Kalimipalli, 2003). I, therefore, conclude that the way people perceive data stabilises its properties and suggest that this process could have a role in making forecasting methods and investment algorithms possible.

Scaling has been examined in the financial literature in the context of forecast horizon (Peters, 1995; Müller et al. 1993; Corsi, 2009). However, the assumptions of these models had not been previously tested. I accepted the hypothesis about the connection between trading horizons and scaling and, hence, support these models.

In addition, I examined the effect of forecast density on the size of the moving average window that people select. Although such an effect was present, the analysis failed to show a correlation between properties of the perceived graphs and forecast density.

Correlations were significantly less than one. This suggests that the market's constants are not accurately preserved, and can provide a reason for the lack of improvement in forecasting accuracy despite advances in computational power over the past few decades (Armstrong, Green, and Graefe, 2014). Forecasting accuracy depends on, among other variables, the validity of its assumptions: if these assumptions do not hold accurately, its success is not guaranteed.

### *Limitations*

The results of these experiments are consistent with findings in finance literature. For instance, in line with Corsi's hypothesis (2009), when participants had to make short-term decisions, they used information from longer periods of time than a linear model would have predicted. Furthermore, in spite of the fact that participants were not instructed to use any specific trading strategy, they recommended that more shares be bought when they thought that prices would increase and that more be sold when they thought that prices would decrease. Indeed, research comparing financial forecasts of lay people and practitioners has typically found only small differences between the two groups (Zaleskiewicz, 2011; Muradoğlu and Önköl, 1994). Moreover, during the last years, the internet has made trading easier for lay people (Muradoglu and Harvey, 2012) and inexperienced investors (Barber and Odean, 2001). Nevertheless, study of the effects of expertise on performance in the tasks employed here could be worthwhile.

In Experiment 1, trading horizon and the Hurst exponents of the graphs were treated as independent variables. However, Vácha and Vošvrda (2005) have shown that presence of traders with different trading forecast horizons in a model can result in price series with different Hurst exponents. Vácha and Vošvrda (2005) showed that larger percentages of

short-term traders were associated with lower Hurst exponents. Given the different paradigm, these results do not contradict those reported here but it would still be interesting to develop a psychological account of them.

## Chapter 7: General Discussion

### Summary

Using the notions that MacKenzie termed in his book “An Engine, Not a Camera, How Financial Models Shape Markets” (2006), this thesis has explored a wide spectrum of human financial behaviour, ranging from the ‘camera’ aspect – people’s perception of financial stimuli, to the ‘engine’ aspect - the characterisation of people as the driving force of the markets.

The ‘market’ was predominantly represented in the experiments by graphically visualised fractional Brownian motions (fBm) or real asset price time series (Chapters 2-6). These designs represented settings corresponding to pure technical analysis. It is known that a large percentage of traders use technical analysis techniques to make financial decisions (Batchelor, 2013; Cheung and Chinn, 2001; Taylor and Allen, 1992). However, when examining the effects of the market on people, I also referred to verbal descriptions relevant to the market, formulated as news items (Chapter 5). Incorporating verbal news in the experiments helped me understand the way people trade beyond technical analysis considerations. Though financial models usually do not take into account differences in human reaction to verbal news and price graphs, I conjectured that, in fact, this difference may affect financial decisions. The media has been shown to have a significant effect on investment patterns (Engelberg and Parsons, 2011).

People’s perception of the market can be examined in different levels. The most fundamental level is that of sensory perception. In Chapter 2 I studied the way people *see* fBm series: whether they were sensitive to the Hurst exponent of the series, what cues they used when assessing them, and whether they could learn to identify them. The results showed that people are highly sensitive to the Hurst exponent of fractal graphs. To discriminate between the Hurst exponents of different graphs, people used cues such as the

perceived ‘width’ and ‘overall darkness’ of the graphs, as well as estimates of their local steepness. Participants learnt to identify the Hurst exponent of fractal graphs from feedback alone.

At the end of Chapter 2 and through Chapter 3 I report studies involving a higher level of analysis: the *meaning* that people attributed to fractal graphs, and in particular, the risk that they perceive in them. I found that, under certain conditions, people assess the risk of investing in an asset in line with the Hurst exponent of the corresponding price series. Furthermore, dependence of risk perception on the Hurst exponent was stronger than it was on other potentially relevant measures, such as the standard deviation of the graphs (their historical volatility) and their mean run-length.

In Chapters 4 and 5, I investigated the effects of people’s perception of the market on price series through two inseparable “engines”: financial forecasts and buy/sell decisions. I assumed that buy/sell decisions affected the market directly; financial forecasts affected the market indirectly, through the buy/sell decisions they implied. I showed that, when making forecasts, people attempt to imitate the noise component of the graphs that they were given. Participants’ forecasts were neither optimal nor naïve. When making financial decisions, they were influenced by properties of both news items and price series. However, they relied more on the former. They bought more shares when they forecast that prices would rise but failed to sell more when they forecast that prices would fall.

Finally, in chapter 6 I studied the interaction between the ‘camera’ and the ‘engine’ - perception of graphical data, forecasts, and buy/sell decisions. Participants in the experiments were presented with sequences of fractal graphs. They could subject them to scaling and smoothing transformations, in a manner similar to the way that financial data providers enable the users of their programmes to select the graph presentation parameters. I found that both scaling and smoothing resulted in graphs, in which local steepness and oscillation were correlated with those of the original graphs. Forecast dispersion was also

correlated with geometric properties of the data graphs. As forecast dispersion was found to be correlated with future price volatility (Athanasakos and Kalimipalli, 2003), I concluded that people's perceptions and actions had a role in the preservation of the parameters of price graphs.

## **Implications**

The results have potential applications in risk communication, forecasting, financial modelling, psychology, and medicine.

### *Risk communication in finance*

The experiments performed in Chapter 3 are consistent with previous findings (Stone, Yates, Parker and Andrew, 1997) concerning the fragile nature of human risk perception: when price graphs were presented without additional cues, risk assessment did not depend on the Hurst exponents of the presented graphs but, when price change graphs were presented with price graphs, risk assessment did depend on them. At present, there is no standard for the presentation of price graphs. Weber, Siebenmorgen, and Weber (2005) have suggested that it could be useful to formulate such a standard for the presentation of graphs. In addition, I suggest that an emphasis on data analysis techniques may also alter perceived risk.

Furthermore, I showed that thickness and darkness of line in graphs affects perception of the Hurst exponent (see Chapter 2): this could, in turn, distort risk perception and so maybe the format in which line price graphs are presented (line width and colour) should be standardised as well.

### *Forecasting*

The experiments showed that, when people make forecasts from fractal graphs, they imitate the noise that they perceive in the data (see Chapter 4). It might be sensible to warn professionals about their tendency to imitate noise, as was established by Harvey (1995).

The analyses failed to find important differences between forecasts of experts in finance and lay people. This is in line with the results of Zaleskiewicz (2011) and Muradoğlu and Önkal (1994) and it emphasises the importance of using algorithmic forecasting methods rather than judgmental forecasts.

#### *Financial models and simulation*

I showed that assumptions which are commonly used in financial models and simulations are inaccurate. Financial models should include realistic assumptions on the way people incorporate data of different types when making financial decisions, allow variability in trading latencies, and take into account individual differences (see Chapter 5). In addition, the analyses depicted participants as people who try to find the meaning of the data they perceive. Financial models and simulations should attempt to exploit this interpretation of traders' performance rather than focussing exclusively on the cognitive bias approach.

#### *Psychological research on judgmental forecasting*

Research on judgmental forecasting has tended to focus on relatively short and simple series. Typically, participants have been required to make forecasts from series with a relatively small number of elements (Reimers and Harvey, 2011). However, in many modern contexts such as finance, people have to deal with complex time series containing many elements. Results reported here suggest that people can deal with series consisting of thousands of elements; they can learn their statistical properties and remember them. In fact, the longer the series is, the better people understand its properties. I hope that this thesis will encourage researchers to perform studies with a high degree of external validity and to use, in appropriate contexts, realistic experimental stimuli.

#### *Medicine*

I have shown that people are highly sensitive to fractal graphs. This sensitivity may have applications in fields other than finance. For instance, many medical signals which

physicians see on a daily basis, such as heart rate and EEG patterns, have been shown to have fractal properties (see e.g. Goldberger, Amaral, Hausdorff, Ivanov, Peng, and Stanley, 2002). People's ability to learn to identify the Hurst exponent of fractal series could help practitioners with diagnosis of certain medical conditions.

## **Limitations**

As noted before, participants in most of the experiments were mainly lay people. Although results were generally in line with those obtained in studies using experts, it remains important to replicate them on finance practitioners and in real trading environments. An exemplary study which achieved a high level of external validity is that of Fenton-O'Creevy, Soane, Nicholson and Willman (2011). They worked with traders in banks in The City of London, where risk perception and reaction to news are integral to the tasks that are performed.

## **Directions for future research**

Throughout this thesis, two human needs were found to affect financial behaviour: the need for validation, or reassurance, and the search for meaning. The need for reassurance was demonstrated in Chapter 3: I showed that people are sensitive to the Hurst exponent of price series but that they used the Hurst exponent as a risk measure only if cues validating its relevance as a risk measure were provided. The search for meaning was used to explain participants' preference of news to price graphs in Chapter 5.

In the experiments, information to (partially) satisfy these needs was given to the participants: in Experiments 2-4 in Chapter 3, I presented participants with price change graphs in addition to the price graphs. In Chapter 5, I let participants read one news item at a

time. All news items related to a single asset were either positive or negative. However, in real life situations, information is rich, abundant, and often includes internal contradictions. How do people try to satisfy these needs in real-life situations? How do people react when there are conflicts between them? How do social factors affect people's search for meaning and need for validation? What part do price graphs have in satisfying these needs?

#### *Academic background*

*The search for meaning* Tuckett (2011) performed a sequence of interviews with investors and managers. He found that they tried to give meaning to their environment through the creation of narratives: “fund managers build conviction by telling stories and [...] these stories contain specific repetitive elements so that we can think of them as following a predetermined script. Such scripts establish conviction both that something exceptional is available and it's safe to invest in it” (page 105). Tarim (2013) analysed narratives present in conversations of investors in the headquarters of three brokerage firms in Istanbul. Investment advisors worked with computers which presented continuously news and other types of data, including prices. Tarim used a stream categorisation system based on that of Boje (2001), consisting of four types: ‘cause-effect’, ‘correlation’, ‘randomness’ and ‘proto-story’. The latter was used in cases where a narrative could not be categorised into one of the first three categories for lack of logical compatibility because events were not connected in a meaningful way. Tarim found that most narratives could be classified as ‘cause-effect’ or ‘proto-story’, whereas only a small percentage of them could be categorised as referring to correlations or randomness. Finally, a large percentage of the stories involved not only the past and the present, but also the present and the future, implying that the traders used forecasts in their narratives. Goodhart (2013) suggested that situations which raise emotional reactions, such as the financial crisis of 2008, produce narratives that are inaccurate and create a misleading picture of the market.

Tuckett's (2011), Tarim's (2013), and Goodhart's (2013) studies describe the meaning people attribute to market events. However, they do not predict what narratives people would create in different situations, and how these narratives are related to news, price graphs, and the Hurst exponent of the graphs. I do not know of any study that characterises the narratives people create using these terms.

*The need for reassurance* Apart from meaning, Tuckett (2011) suggested that investors search for validation of their decisions in the form of non-contradicting pieces of information: "Hypotheses supported by different methods, and particularly those supported by unobtrusive measures, have a stronger claim" (page 105). Tuckett emphasised the psychological discomfort investors felt when the need for reassurance was not met. For instance, one of the investors he interviewed said (about a controversial decision he had made) that: "It was not easy going against consensus sentiment" (page 35). In a different situation, the investor "was not able to develop confidence in his thesis when the stock price kept falling" (page 37).

The way people combine different data items has been studied by De Bondt and Thaler (1985) and by Andreassen (1990). De Bondt and Thaler (1985) hypothesised that people over-react to news when making financial decisions. Andreassen (1990) studied the effect of contradiction between news items and stock price trends. He showed that people tend to use news items more in their decisions when they contradict price trends. Oberlechner and Hocking (2004) found that contradicting news was considered more important than non-contradicting news and that information received at times of high volatility is more important than information obtained after a long period of stability. Recently, Goodwin (2014) investigated forecast adjustments that participants make when news items with different valences are presented simultaneously. He found that people treat news in a compensatory manner, so that good and bad news tend to cancel each other out.

From the perspective of reassurance, Andreassen (1990) and Oberlechner and Hocking (2004) seem to imply that data that does not offer reassurance is considered more important than data that does. However, I did not succeed in replicating Andreassen's (1990) and Oberlechner and Hocking's (2004) findings within the paradigm used here (Chapter 5).

Individual differences have been found to affect reassurance seeking and its consequences. For instance, it has been shown that reassurance seeking predicted stress in women but not in men (Shih and Auerbach, 2010).

I know of no study that examines the conditions in which the need for reassurance dominates people's behaviour in the financial context, or the interaction between the need for reassurance and the market's volatility. Neither am I aware of any study examining the effects of individual differences on reassurance-seeking behaviour among traders.

*Interactions between the search for meaning and the need for validation* Gonzalez, Lerch and Lebiere (2003) studied the way that people make decisions in ever-changing complex, dynamic environments. They argued that decision makers used their past knowledge and heuristics and that they adapted them to fit the given situation. Then they refined their strategies according to the feedback they received.

The financial world is an example of such an environment. The search for meaning can be viewed as the motivation that drives people to use the sort of cognitive strategies that Gonzales et al (2003) describe. Need for validation can be related to people's anticipation of feedback that they receive. However, the financial world is an especially illusory one: the feedback that is received can be the result of a nearly random price movement and, hence, misleading, and the information that is obtained can be inaccurate or wrong. Therefore, in certain cases, the need for validation can be in conflict with the need for meaning. What would a trader do when different news items contradict each other? How do traders choose information items? These are general issues for future work.

# Bibliography

- Ackert, L. F., Church, B. K., & Zhang, P. (2002). Market behavior in the presence of divergent and imperfect private information: experimental evidence from Canada, China, and the United States. *Journal of Economic Behavior and Organization*, 47 (4), 435-450. doi: 10.1016/S0167-2681(01)00212-8.
- Al-Jafari, M. J. (2011). Random walks and market efficiency tests: evidence from emerging equity market of Kuwait. *European Journal of Economics, Finance & Administrative Sciences*, 36, 19-28.
- Amilon, H. (2003). A neural network versus Black–Scholes: a comparison of pricing and hedging performances. *Journal of Forecasting*, 22 (4), 317-335. doi: 10.1002/for.867.
- Anderson, C., John, O. P., & Keltner, D. (2012). The personal sense of power. *Journal of Personality*, 80 (2), 313-344. doi: 10.1111/j.1467-6494.2011.00734.x.
- Andreassen, P. B. (1990). Judgmental extrapolation and market overreaction: On the use and disuse of news. *Journal of Behavioral Decision Making*, 3 (3), 153-174. doi: 10.1002/bdm.3960030302.
- Ang, A., Goetzmann, W. N., & Schaefer, S. M. (2010). The efficient market theory and evidence: implications for active investment management. *Foundations and Trends in Finance*, 5 (3), 157–242. doi: 10.1561/05000000034.
- Anufriev, M., & Panchenko, V. (2009). Asset prices, traders' behavior and market design. *Journal of Economic Dynamics and Control*, 33 (5), 1073-1090. doi:10.1016/j.jedc.2008.09.008.

- Armstrong, S. J., & Fildes, R. (1995). On the selection of error measures for comparisons among forecasting methods. *Journal of Forecasting*, *XIV* (1), 67-71. doi: 10.1002/for.3980140106.
- Armstrong, S. J., Green, K. C., & Graefe, A. (2014). Golden rule of forecasting: Be conservative. Working paper draft, available on <http://www.kestengreen.com/GoldenRule.pdf>.
- Athanassakos, G. & Kalimipalli, M. (2003). Analyst forecast dispersion and future stock return volatility. *Quarterly Journal of Business and Economics*, *42* (1/2), 57-78. ISSN: 07475535.
- Aufmann, R. N., Lockwood, J., Nation, R. D. & Clegg, D. K. (2010). *Mathematical Excursions*. Boston, USA: Brooks/Cole.
- Azami, H. Bozorgtabar, B., & Shiroie, M. (2011). Automatic signal segmentation using the fractal dimension and weighted moving average filter. *International Journal of Electrical & Computer Sciences*, *11* (6), 8-15.
- Ballesteros, S., & Manga, D. (1996). The effects of variation of an irrelevant dimension on same-different visual judgments. *Acta Psychologica*, *92* (1), 1-16. doi: 10.1016/0001-6918(95)00003-8.
- Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, *21* (2), 785-818. doi:10.1093/rfs/hhm079.
- Batchelor, R. (2013). Forecasting financial markets: Some light from the dark side. Forecasting with big data, The 33<sup>rd</sup> International Symposium on Forecasting, Proceedings.

- Batchelor, R., & Kwan, T. Y. (2007). Judgemental bootstrapping of technical traders in the bond market. *International Journal of Forecasting*, 23, 427–445. doi: 10.1016/j.ijforecast.2007.05.007.
- Bayraktar, E., & Poor, H. V. (2005). Arbitrage in fractal modular Black-Scholes models when the volatility is stochastic. *International Journal of Theoretical & Applied Finance*, 8 (3), 283-300.
- Berkowitz, J. (2010). On justifications for the ad hoc Black-Scholes method of option pricing. *Studies in Nonlinear Dynamics & Econometrics*, 14 (1), 1-25. doi: 10.2202/1558-3708.1683.
- Bianchi, S., De Bellis, I., & Pianese, A. (2010). Fractal properties of some European electricity markets. *International Journal of Financial Markets and Derivatives*, 1 (4), 395-421. doi: 10.1504/IJFMD.2010.035766.
- Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81 (3), 637–654. doi: 10.1086/260062.
- Blackledge, J. M. (2008). Application of the Fractal Market Hypothesis for macroeconomic time series analysis. *ISAST Transactions on Electronics and Signal Processing*, 1 (2), 1-22.
- Boje, D. (2001). *Narrative Methods for Organizational and Communication Research*. London, UK: Sage.
- Bolger, F., & Harvey, N. (1993). Context-sensitive heuristics in statistical reasoning. *The Quarterly Journal of Experimental Psychology*, 46 (4), 779–811. doi: 10.1080/14640749308401039.

- Caccia, D. C., Percival, D., Cannon, M. J., Raymond, G., & Bassingthwaigthe, J. B. (1997). Analyzing exact fractal time series: Evaluating dispersional analysis and rescaled range methods. *Physica A*, *246* (3), 609–632. doi: 10.1016/S0378-4371(97)00363-4.
- Cacioppo, J. T., Petty, R. E., & Morris, K. J. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology*, *45* (4), 805-818. doi: 10.1037/0022-3514.45.4.805.
- Caginalp, G., Porter, D., & Hao, L. (2010). Asset market reactions to news: an experimental study. Available at SSRN: <http://ssrn.com/abstract=1988413> or <http://dx.doi.org/10.2139/ssrn.1988413>.
- Cannon, M. J., Percival, D. B., Caccia, D. C., Raymond, G. M., & Bassingthwaigthe, J. B. (1997). Evaluating scaled windowed variance methods for estimating the Hurst coefficient of time series. *Physica A*, *241* (3), 606–626. doi: 10.1016/S0378-4371(97)00252-5.
- Castro, R., Kalish, C., Nowak, R., Qian, R., Rogers, T., & Zhu, X. (2008). Human active learning, in *Advances in Neural Information Processing Systems (NIPS)*, Vancouver, Canada, December 2008.
- Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010). Making words work: Using financial text as a predictor of financial events. *Decision Support Systems* *50*, 164–175. doi: 10.1016/j.dss.2010.07.012.
- Cheung, Y. W., & Chinn, M. D. (2001). Currency traders and exchange rate dynamics: a survey of the US market. *Journal of International Money and Finance* *20*, 439–471. doi: 10.1016/S0261-5606(01)00002-X.
- Cheung, E., & Mikels, J. A. (2011). I'm feeling lucky: The relationship between affect and risk-seeking in the framing effect. *Emotion*, *11* (4), 852-859. doi: 10.1037/a0022854.

- Coen, T., & Torluccio, G. (2012). Self-Similarity in the Analysis of Financial Markets' Behaviour. *International Research Journal of Finance & Economics*, 87, 176-184.
- Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, 19, 15-18.
- Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications*. London, UK: Academic Press.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7 (2), 174-96. doi: 10.1093/jjfinec/nbp001.
- Cui, H., & Yang, L. (2009). The Analysis and Improvement of the Price Forecast Model Based on Fractal Theory. International Conference on Computer Engineering and Technology, ICCET via IEEE SERIES (IEL), 1, 395-399. doi: 10.1109/ICCET.2009.75.
- Cutting, J., E., & Garvin, J. J. (1987). Fractal curves and complexity. *Perception & Psychophysics*, 42 (4), 365-370. doi: 10.3758/BF03203093.
- Davies, R. B., & Harte, D. S. (1987). Tests for Hurst effect. *Biometrika*, 74 (1), 95-101. doi: 10.1093/biomet/74.1.95.
- De Bondt, W. F. M., & Thaler, R. (1985). Does the stock market overreact? *Journal of Finance* 40 (3), 793-805. doi: 10.1111/j.1540-6261.1985.tb05004.x.
- De Grauwe, P. (2010). Top-Down versus Bottom-Up Macroeconomics, *CESifo Economic Studies*, 56 (4), 465-497. doi: 10.1093/cesifo/ifq014.
- De Grauwe, P., & Grimaldi, M. (2005). Heterogeneity of agents, transactions costs and the exchange rate, *Journal of Economic Dynamics & Control*, 29, 691-719. doi: 10.1016/j.jedc.2004.01.004.

- De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22, 545–559. doi: 10.1016/0042-6989(82)90113-4.
- Delignières, D., Ramdani, S., Lemoine, L., Torre, K., Fortes, M., & Ninot, G. (2006). Fractal analyses for 'short' time series: Are-assessment of classical methods. *Journal of Mathematical Psychology*, 50 (6), 525-544. doi: 10.1016/j.jmp.2006.07.004.
- DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences*, 33 (4), 533-552. doi: 10.1016/S0191-8869(01)00171-4.
- Dezsi, D., & Scarlat, E. (2012). A multifractal model of asset returns in the context of the new economy paradigm. *Timisoara Journal of Economics and Business*, 5 (1), 23-32.
- Du, X. M., Zhang, Q. L., Zeng, J. M., Gui, Q., Luo, J., & Ruan, X. L. (2008). Hot hand fallacy or gambler's fallacy? A research on the Gestalt phenomena in random sequence recency effect. *Acta Psychologica Sinica*, 40 (8), 853-861. doi: 10.3724/SP.J.1041.2008.00853.
- Duchon, J., Robert, R., & Vargas, V. (2012). Forecasting volatility with the multifractal random walk model. *Mathematical Finance*, 22 (1), 83-108. doi: 10.1111/j.1467-9965.2010.00458.x.
- Durand, R. B., Newby, R., Peggs, L., & Siekierka, M. (2013). Personality. *The Journal of Behavioral Finance*, 14 (2), 116-133. doi: 10.1080/15427560.2013.791294.
- Durand, R. B., Newby, R., & Sanghani, J. (2008). An intimate portrait of the individual investor. *The Journal of Behavioral Finance*, 9 (4), 193-208. doi: 10.1080/15427560802341020.

- Duxbury, D., & Summers, B. (2004). Financial risk perception: Are individuals variance averse or loss averse? *Economics Letters*, 84 (1), 21-28. doi: 10.1016/j.econlet.2003.12.006.
- Edwards, W. (1961) Probability learning in 1,000 trials. *Journal of Experimental Psychology*, 62, 381-390.
- Egeth, H. E. (1965). Parallel versus serial processes in multidimensional discrimination. *Perception and Psychophysics*, 1, 245-252.
- Eggleton, I. R. C. (1982). Intuitive time-series extrapolation. *Journal of Accounting Research*, 20, 68-102.
- Engelberg, J. E., & Parsons, C. A. (2011). The causal impact of media in financial markets. *Journal of Finance*, 66 (1), 67-97. doi: 10.1111/j.1540-6261.2010.01626.x.
- Eroglu, C., & Croxton, K. L. 2010. Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting*, 26 (1), 116-133. doi: 10.1016/j.ijforecast.2009.02.005.
- Falk, R., & Konold, C. (1994). Random means hard to digest. *Focus on Learning Problems in Mathematics*, 16, 2-12.
- Fenton-O'Creevy, M., Lins, J., Vohra, S., Richards, D., Davies, G., & Schaaff, K. (2012). Emotion regulation and trader expertise: Heart rate variability on the trading floor. *Journal of Neuroscience, Psychology and Economics*, 5 (4), 227-237. doi: 10.1037/a0030364.
- Fenton-O'Creevy, M., Soane, M., Nicholson, N., & Willman, P. (2011). Thinking, feeling and deciding: The influence of emotions on the decision making and performance of traders. *Journal of Organizational Behavior*, 32 (8), 1044-1061. doi: 10.1002/job.720.

- Findlay, M. C., & Williams, E. E. (Winter 2000-2001). A fresh look at the Efficient Market Hypothesis: How the intellectual history of finance encouraged a real 'Fraud-on-the-Market'. *Journal of Post Keynesian Economics*, 23 (2), 181-199. Article Stable URL: <http://www.jstor.org/stable/4538722>.
- Fiori, M., & Antonakis, J. (2012). Selective attention to emotional stimuli: What IQ and openness do, and emotional intelligence does not. *Intelligence*, 40 (3), 245–254. doi: <http://dx.doi.org/10.1016/j.intell.2012.02.004> .
- Forsythe, A., Nadal, M., Sheehy, N., Cela-Conde, C. J., & Sawey, M. (2011). Predicting beauty: Fractal dimension and visual complexity in art. *British Journal of Psychology*, 102 (1), 49-70. doi: 10.1348/000712610X498958.
- Frijns, B., Koellen, E., & Lehnert, T. (2008). On the determinants of portfolio choice. *Journal of Economic Behavior & Organization*, 66 (2), 373–386. doi: 10.1016/j.jebo.2006.04.004.
- Frost, A. J. & Prechter, R. R. (1998). *Elliott wave principle, 20<sup>th</sup> anniversary edition*. Georgia, USA: New Classic Library.
- Gaissmaier, W., Wegwarth, O., Skopec, D., Müller, A. S., Broschinski, S., & Politi, M. C. (2012). Numbers can be worth a thousand pictures: Individual differences in understanding graphical and numerical representations of health-related information. *Health Psychology*, 31 (3), 286-296. doi: 10.1037/a0024850.
- Galati, G., & Ho, C. (2003). Macroeconomic news and the Euro/Dollar exchange rate. *Economic Notes*, 32 (3), 371-398. doi: 10.1111/1468-0300.00118.
- Galinsky, A. D., Magee, J. C., Gruenfeld, D. H., Whitson, J. A., & Liljenquist, K. A. (2008). Power reduces the press of the situation: Implications for creativity, conformity, and dissonance. *Journal of Personality and Social*, 95 (6), 1450–1466. doi: 10.1037/a0012633.

- Gençay, R., Gradojevic, N., Selçuk, F., & Whitcher, B. (2010). Asymmetry of information flow between volatilities across time scales. *Quantitative Finance*, *10* (8), 895-915. doi: 10.1080/14697680903460143 .
- Georgeson, M. A., May, K. A., Freeman, T. C. A., & Hesse, G. S. (2007). From filters to features: Scale–space analysis of edge and blur coding in human vision. *Journal of Vision* *7* (13), 1-21. ISSN: 1534-7362.
- Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond “heuristics and biases”. *European Review of Social Psychology*, *2*, 83-115. doi: 10.1080/14792779143000033.
- Gilden, D. L., Schmuckler, M. A., & Clayton, K. (1993). The perception of natural contour. *Psychological Review*, *100* (3), 460-478. doi: 10.1037/0033-295X.100.3.460.
- Gilovich, T., Vallone, R., & Tversky, A., (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, *17*, 295-314. doi: 10.1016/0010-0285(85)90010-6 .
- Glezakos, M., & Mylonas, P. (2003). Technical analysis seems to be a valuable investment tool in the Athens and Frankfurt stock exchanges. *European Research Studies*, *6* (1-2), 169-192.
- Goldberger, A. L., Amaral, L. A. N., Hausdorff, J. M., Ivanov, P. Ch., Peng, C.-K., & Stanley, H. E. (2002). Fractal dynamics in physiology: Alterations with disease and aging. *Proceedings of the National Academy of Sciences of the United States of America*, *99* (Suppl 1), 2466-2472. ISSN: 0027-8424.
- Goodhart, C. A. E. (2013). Narratives of the Great Financial Crisis (GFC): Why I am out of step. *The Journal of Financial Perspectives*, *1* (3), 1-6.
- Goodwin, P. (2014). Judgmental adjustments to forecasts when special events are due to occur: An analysis of information use. ISF 2014, Programme Book, available on

[http://forecasters.org/wp/wp-content/uploads/ISF2014\\_Program-Book\\_06232014.pdf](http://forecasters.org/wp/wp-content/uploads/ISF2014_Program-Book_06232014.pdf).

- Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27* (4), 591-635. doi: 10.1207/s15516709cog2704\_2.
- Gosling, S. D, Rentfrow, P. J., & Swann, W. B. Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*, 504–528. doi: 10.1016/S0092-6566(03)00046-1.
- Graham, J., & Harvey, C. (2002). How do CFOs make capital budgeting and capital structure decisions? *Journal of Applied Corporate Finance*, *15* (1), 8-23. doi: 10.1111/j.1745-6622.2002.tb00337.x.
- Griskevicius, V., Tybur, J. M., Delton, A. W., & Robertson, T. E. (2011). The influence of mortality and socioeconomic status on risk and delayed rewards: A life history theory approach. *Journal of Personality and Social Psychology*, *100* (6), 1015–1026. doi: 10.1037/a0022403.
- Hai-Chin, Y., & Ming-Chang, H. (2004). Statistical properties of volatility in fractal dimensions and probability distribution among six stock markets. *Applied Financial Economics*, *14* (15), 1087-1095. doi: 10.1080/09603100412331297694.
- Hall, S. (2006). What counts? Exploring the production of quantitative financial narratives in London's corporate finance industry. *Journal of Economic Geography*, *6* (5), 661-678. doi :10.1093/jeg/lbl008.
- Harras, G., & Sornette, D. (2011). How to grow a bubble: A model of myopic adapting agents. *Journal of Economic Behavior and Organization*, *80* (1), 137-152. doi: 10.1016/j.jebo.2011.03.003.
- Harvey, C. R. (2010). *Managerial Miscalibration*. National Bureau of Economic Research, Inc.

- Harvey, N. (1995). Why are judgments less consistent in less predictable task situations? *Organizational Behavior and Human Decision Processes*, 63 (3), 247–263. doi: 10.1006/obhd.1995.1077.
- Harvey, N. (1988). Judgmental forecasting of univariate time series. *Journal of Behavioral Decision Making*, 1 (2), 95–110. doi: 10.1002/bdm.3960010204.
- Harvey, N., & Bolger, F. (1996). Graphs versus tables: effects of data presentation format on judgemental forecasting. *International Journal of Forecasting*, 12 (1), 119–137. doi: 10.1016/0169-2070(95)00634-6.
- Harvey, N., Ewart, T., & West, R. (1997). Effects of data noise on statistical judgement. *Thinking & Reasoning*, 3 (2), 111-132. doi: 10.1080/135467897394383.
- Harvey, N., & Reimers, S. (2013). Trend damping: Under-adjustment, experimental artifact, or adaptation to features of the natural environment? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39 (2), 589-560. doi: 10.1037/a0029179.
- Harvey, N., & Reimers, S. (2014). Bars, lines, and points: The effect of graph format on judgmental forecasting. Economic forecasting – past, present, and future. The 34<sup>th</sup> International Symposium on Forecasting, Proceedings.
- Hassoun, J., P. (2005). Emotions on the trading floor: Social and symbolic expressions. In Cetina, K. K., & Preda, A (Ed.), *The sociology of Financial Markets*. New York, NY: Oxford University Press.
- Haug, E. G., & Taleb, N. N. (2011). Option traders use (very) sophisticated heuristics, never the Black–Scholes–Merton formula. *Journal of Economic Behavior and Organization*, 77 (2), 97-106. doi: 10.1016/j.jebo.2010.09.013.

- Hayo, B., & Neuenkirch, M. (2012). Bank of Canada communication, media coverage, and financial market reactions. *Economics Letters*, *115*( 3), 369-372. doi:10.1016/j.econlet.2011.12.086.
- Hendricks, D. (1996). Evaluation of value-at-risk models using historical data. *FRBNY Economic Policy Review*, *2* (1), 39-69. Available at SSRN: <http://ssrn.com/abstract=1028807> or <http://dx.doi.org/10.2139/ssrn.1028807>.
- Hill, N., Franklin, B., Clason, G. S., & Mackay, C. (2009). Make more money: Secrets from the world's greatest financial classics. Oxford, UK: Infinite Ideas Limited.
- Hilton, D. (2001). The psychology of financial decision making: Applications to trading, dealing, and investment analysis. *The Journal of Psychology and Financial Markets*, *2* (1), 37-53. doi: 10.1207/S15327760JPFM0201\_4.
- Hodnett, K., & Heng-Hsing H. (2012). Capital market theories: Market efficiency versus investor prospects. *International Business & Economics Research Journal*, *11* (8), 849-862.
- Holtgrave, D. R., & Weber, E. U. (1993). Dimensions of risk perception for financial and health risks. *Risk Analysis*, *13*, 553–558. doi: 10.1111/j.1539-6924.1993.tb00014.x.
- Holton, G. A. (2004). Defining Risk. *Financial Analysts Journal*, *60* (6), 19-25.
- Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, *14* (3), 399-425.
- Hyndman, R. J., & Athanasopoulos, G. *Forecasting: Principles and Practice*. Available on <http://otexts.com/fpp/1/3/>. To appear in print in 2014.
- In, F., & Kim, S. (2006). Multiscale hedge ratio between the Australian stock and futures markets: Evidence from wavelet analysis. *Journal of Multinational Financial Management* *16* (4), 411-423. doi:10.1016/j.mulfin.2005.09.002.

- Jakes, S., & Hemsley, D. R. (1986). Individual differences in reaction to brief exposure to unpatterned visual stimulation. *Personality and Individual Differences*, 7 (1), 121-123. doi: 10.1016/0191-8869(86)90118-2.
- Jarvik, M. E. (1951). Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *Journal of Experimental Psychology*, 41, 291-297. doi: 10.1037/h0056878.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430-454. doi: 10.1016/0010-0285(72)90016-3.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- Kala, R. A. M., & Pandey, S. L. D. (2012). A feasibility analysis of Black-Scholes-Merton differential equation model for stock option pricing by using historical volatility: With reference to selected stock options traded in NSE. *Asian Journal of Finance & Accounting*, 4 (2), 86-106. ISSN: 1946052X.
- Kapteyn, A., & Teppa, F. (2011). Subjective measures of risk aversion, fixed costs, and portfolio choice. *Journal of Economic Psychology*, 32 (4), 564–580. doi: 10.1016/j.joep.2011.04.002.
- Katsenelson, V. N. (2007). *Active value investing: Making money in range-bound markets*. Hoboken, New Jersey: John Wiley & Sons.
- Katsev, S., & L'Heureux, I. (2003). Are Hurst exponents estimated from short or irregular time series meaningful? *Computers & Geosciences* 29 (9), 1085–1089. doi: 10.1016/S0098-3004(03)00105-5.

- Kazoleas, D. C. (1993). A Comparison of the persuasive effectiveness of qualitative versus quantitative evidence: A test of explanatory hypotheses. *Communication Quarterly*, 41 (1), 40-50. ISSN: 0146-3373.
- Klos, A., Weber, E. U., & Weber, E. (2005). Investment decisions and time horizon: risk perception and risk behavior in repeated gambles. *Management Science* 51 (12), 1777-1790. doi: 10.1287/mnsc.1050.0429.
- Knuppel, M., & Schultefrankenfeld, G. (2011). How informative are central bank assessments of macroeconomic risks? *Deutsche Bundesbank, Research Centre, Discussion Paper Series 1: Economic Studies*, 13.
- Koonce, L. L., McAnally, M. L., & Mercer, M. (2005). How do investors judge the risk of financial items? *The Accounting Review*, 80 (1), 221-241. doi: <http://dx.doi.org/10.2308/accr.2005.80.1.221>.
- Kristoufek, L. (2012). Fractal markets hypothesis and the global financial crisis: scaling, investment horizons and liquidity. *Advances in Complex Systems*, 15, 1250065. doi: 10.1142/S0219525912500658.
- Kumar, T., Zhou, P., & Glaser, D. A. (1993). Comparison of human performance with algorithms for estimating fractal dimension of fractional Brownian statistics. *Journal of the Optical Society of America. A, Optics and image science*, 10 (6), 1136-1146. ISSN: 0740-3232.
- Kuzmina, J. (2010). Emotion's component of expectations in financial decision making. *Baltic Journal of Management*, 5 (3), 295 – 306. doi: 10.1108/17465261011079721.
- Lang F.R., John D., Lüdtke O., Schupp J., & Wagner, G. G. (2011). Short assessment of the Big Five: robust across survey methods except telephone interviewing. *Behavior Research Methods*, 43(2), 548–567. doi: 10.3758/s13428-011-0066-z.

- Lawrence, M., Goodwin, P., O'Connor, M., & Önkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, 22 (3), 493-518. doi: 10.1016/j.ijforecast.2006.03.007.
- Lawrence, M., & Makridakis, S. (1989). Factors affecting judgemental forecasts and confidence intervals. *Organizational Behavior and Human Decision Processes*, 43 (2), 172–187. doi: 10.1016/0749-5978(89)90049-6.
- Lawrence, M., & O'Connor, M. (1992). Exploring judgemental forecasting. *International Journal of Forecasting*, 8 (1), 15-26. doi: 10.1016/0169-2070(92)90004-S.
- Lee, C. J., & Andrade, E. B. (2011) Fear, social projection, and financial decision making. *Journal of Marketing Research* 48 (SPL), S121-S129. Available at SSRN: <http://ssrn.com/abstract=1866568>.
- Lin, E.L., Murphy, G.L., & Shoben, E.J. (1997). The effects of prior processing episodes on basic-level superiority. *Quarterly Journal of Experimental Psychology*, 50A, 25-48. doi: 10.1080/713755686.
- Ling-Yun H. (2011). Chaotic structures in Brent & WTI crude oil markets: Empirical evidence. *International Journal of Economics & Finance*, 3 (5), 242-249. ISSN: 1916971X.
- Liu, H. C., & Hung, J. C. (2010). Forecasting volatility and capturing downside risk of the Taiwanese futures markets under the financial tsunami. *Managerial Finance*, 36 (10), 860-875. doi: 10.1108/03074351011070233.
- Lo, A. W., Repin, D. V., & Steenbarger, B. N. (2005). Fear and greed in financial markets: A clinical study of day-traders. *American Economic Review*, 95 (2), 352-359. ISSN: 0002-8282.

- MacGregor, D., Slovic, P., Berry, M., & Evensky, H. R. (1999). Perception of financial risk: A survey study of advisors and planners. *Journal of Financial Planning*, 12, 68–86. Available at SSRN: <http://ssrn.com/abstract=1860403>.
- MacKenzie, D. (2006). *An engine, not a camera: How financial models shape markets*. Massachusetts, USA: The MIT Press.
- MacMillan, N. A., & Creelman, C. D. (2005). *Detection theory, a user's guide*. New Jersey, Mahwan: Lawrence Erlbaum Associates.
- Maddox, W. T., Love, B. C., Glass, B. D., & Filoteo, J. V. (2008). When more is less: Feedback effects in perceptual category learning. *Cognition*, 108 (2), 578-589. doi: 10.1016/j.cognition.2008.03.010.
- Malavoglia, R. C., Gaio, L. G., Júnior, T. P., & Lima, F. G. (2012). The Hurst exponent: a study of the major international stock markets. *Journal of International Finance & Economics*, 12 (1), 113-121. ISSN: 1555-6336.
- Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417. doi: 10.1111/j.1540-6261.1970.tb00518.x.
- Mandelbrot, B., & Hudson R. L. (2004). *The (mis)behaviour of markets, a fractal view of risk, ruin and reward*. London, UK: Profile Books LTD.
- Manzan, S., & Westerhoff, F. (2005). Representativeness of news and exchange rate dynamics. *Journal of Economic Dynamics and Control*, 29 (4), 677-689. doi:10.1016/j.jedc.2003.08.008.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77-91. doi: 10.1111/j.1540-6261.1952.tb01525.x.

- Matilla-García, M., & Marín, M. R. (2010). A new test for chaos and determinism based on symbolic dynamics. *Journal of Economic Behavior and Organization*, 76 (3), 600-614. doi: 10.1016/j.jebo.2010.09.017 .
- Mattos, F., Garcia, P., & Pennings, Joost M. E. (2007). Insights into trader behavior: Risk aversion and probability weighting. NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management, 2007 Conference, Chicago, Illinois.
- McCrae, R.R., & Costa, P. T. Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology* 52 (1), 81–90. doi:10.1037/0022-3514.52.1.81.
- Mehmood, M. S., Mehmood, A., & Mujtaba, B. G. (2012). Stock market prices follow the random walks: Evidence from the efficiency of Karachi stock exchange. *European Journal of Economics, Finance & Administrative Sciences*, 51, 71-80. Available at SSRN: <http://ssrn.com/abstract=2131719>.
- Mehrara, M., & Oryoie, A. R. (2012). Efficient markets hypothesis in foreign exchange market before and after the global financial crisis of 2007-08. *International Journal of Business & Social Science*, 3 (9), 165-167.
- Mitina O. V., & Abraham F. D. (2003). The use of fractals for the study of the psychological perception: Psychophysics and personality factors. *International Journal of Modern Physics C*, 14 (8) 1047-1060. doi: 10.1142/S0129183103005182.
- Müller, U. A., Dacorogna, M. M., Davé, R. D. ,Pictet, O. V., Olsen, R. B., & Ward, J. R.(1993). Fractals and intrinsic time – A challenge to econometricians. 39th International AEA Conference on Real Time Econometrics, 14–15 October 1993, Luxembourg.

- Muradoglu, G., & Harvey, N. (2012). Behavioural finance: the role of psychological factors in financial decisions. *Review of Behavioral Finance*, 4 (2), 68 – 80. doi: 10.1108/19405971211284862.
- Muradoğlu, G. & Önköl, D. (1994). An exploratory analysis of portfolio managers' probabilistic forecasts of stock prices. *Journal of Forecasting*, 13, 565–578. Available at SSRN: <http://ssrn.com/abstract=1296864>.
- Narayan, P. K., & Smyth, R. (2006). Random walk versus multiple trend breaks in stock prices: Evidence from 15 European markets. *Applied Financial Economics Letters*, 2 (1), 1-7. doi: 10.1080/17446540500424784.
- Nelson, M. W., Bloomfield, R., Hales, J. W., & Libby, R. (2001). The effect of information strength and weight on behavior in financial markets. *Organizational Behavior and Human Decision Processes*, 86 (2), 168–196. doi: 10.1006/obhd.2000.2950.
- Nicholson, N., Soane, E., Fenton-O'Creevy, M., & Willman, P. (2005). Personality and domain-specific risk taking. *Journal of Risk Research*, 8 (2), 157-176. doi: 10.1080/1366987032000123856.
- Nisbett, R. E. (2003). *The geography of thought: How Asians and Westerners think differently ... and why*. New York, NY: The Free Press.
- Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, 108 (2), 291-310. doi: 10.1037//0033-295X.108.2.291.
- Norman, W. T. (1963). Toward an adequate taxonomy of personality affect. *Journal of abnormal and Personality Psychology*, 66 (6), 574-583. doi: 10.1037/h0040291.
- Nosić, A., & Weber, M. (2010). How riskily do I invest? The role of risk attitudes, risk perceptions, and overconfidence. *Decision Analysis*, 7(3), 282-301. Available at SSRN: <http://ssrn.com/abstract=1004002>.

- Oberlechner, T., & Hocking, S. (2004). Information sources, news, and rumors in financial markets: Insights into the foreign exchange market. *Journal of Economic Psychology*, 25 (3), 407-424. doi: 10.1016/S0167-4870(02)00189-7.
- Odean, T. (1998). Are investors reluctant to realize their losses? *The Journal of Finance*, 53 (5), 1775-1798. doi: 10.1111/0022-1082.00072.
- Onali, E., & Goddard, J. (2011). Are European equity markets efficient? New evidence from fractal analysis. *International Review of Financial Analysis*, 20 (2), 59-67. doi: 10.1016/j.irfa.2011.02.004.
- Otto, S. (2010). Does the London metal exchange follow a random walk? Evidence from the predictability of futures prices. *Open Economics Journal*, 3, 25-42. doi: 10.2174/1874919401003010025.
- Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, 72, 407-418. doi: 10.1037/h0022602.
- Parthasarathy, S. (2013). Long range dependence and market efficiency: Evidence from the Indian stock market. *Indian Journal of Finance*, 7 (1), 17-25.
- Peitgen, H. O., & Saupe, D. (1988). *The science of fractal images*. New York, NY: Springer-Verlag.
- Pennington, N., & Hastie, R. (1993). The story model for juror decision making. In R. Hastie (Ed.) *Inside the juror* (pp.192–224). Cambridge, UK: Cambridge University Press.
- Pesaran, M. H., Pick, A., & Timmermann, A. (2011). Variable selection, estimation and inference for multi-period forecasting problems. *Journal of Econometrics*, 164 (1), 173-187. doi:10.1016/j.jeconom.2011.02.018 .
- Peters, E., E. (1994). *Fractal market analysis*. New York, NY: John Wiley & Sons.

- Peterson, R. L., Murtha, F. F., Harbour, A. M., & Friesen R. (working paper, draft, 2011).  
The personality traits of successful investors during the U.S. stock market's "lost decade" of 2000-2010. MarketPsych LLC New York ~ Los Angeles .  
info@marketpsych.com www.marketpsych.com.
- Pfajfar, D. (2013). Formation of rationally heterogeneous expectations. *Journal of Economic Dynamics and Control*, 37 (8), 1434-1452. doi: 10.1016/j.jedc.2013.03.012.
- Poulton, E.C. (1989). *Bias in quantifying judgments*. Hove, UK: Erlbaum.
- Raghubir, P., & Das, S. R. (2010). The long and short of it: Why are stocks with shorter runs preferred? *Journal of Consumer Research*, 36 (6), 964-982. doi: 10.1086/644762.
- Redies, C., Hasenstein, J., & Denzler, J. (2007). Fractal-like image statistics in visual art: Similarity to natural scenes. *Spatial Vision*, 21, 1–2, 137–148. ISSN: 0169-1015.
- Reeves, R., & Sawicki, M. (2007). Do financial markets react to Bank of England communication? *European Journal of Political Economy*, 23, 207–227. doi: 10.1016/j.ejpoleco.2006.09.018.
- Reimers, S. & Harvey, H. (2011). Sensitivity to autocorrelation in judgmental time series forecasting. *International Journal of Forecasting* 27 (4), 1196-1214. doi: 10.1016/j.ijforecast.2010.08.004.
- Reips, U. D. (2002). Standards for internet-based experimenting. *Experimental Psychology*, 49 (4), 243-256. doi: 10.1026//1618-3169.49.4.243.
- Richards, G. R. (2004). A fractal forecasting model for financial time series. *Journal of Forecasting*, 23, 587-602. doi: 10.1002/for.927.
- Righi, M. B., & Ceretta, P. S. (2011). Random walk and variance ratio tests for efficiency in the sub-prime crisis: Evidence for the U.S. and latin markets. *International Research Journal of Finance & Economics*, 72, 25-32.

- Robin, S., & Strážnicka, K. (2012). Personality characteristics and asset market behavior. *29èmes Journées de Microéconomie Appliquée (JMA)*, Brest, France.  
<http://halshs.archives-ouvertes.fr/halshs-00725976>.
- Roney, C. J. R., & Trick, L. M. (2003). Grouping and gambling: A Gestalt approach to understanding the gambler's fallacy. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *57* (2), 69-75. doi: 10.1037/h0087414.
- Sachse, K., Jungermann, H., & Belting, J. M. (2012). Investment risk – The perspective of individual investors. *Journal of Economic Psychology*, *33* (3), 437-447. Doi: 10.1016/j.joep.2011.12.006.
- Sadowski, C. J. & Cogburn, H. E. (1997). Need for cognition in the Big-Five Factor structure. *The Journal of Psychology: Interdisciplinary and Applied*, *131* (3), 307-312. doi: 10.1080/00223989709603517.
- Sang, H-w., Ma, T., & Wang, S-z. (2001). Hurst exponent analysis of financial time series. *Journal of Shanghai University (English Edition)*, *5* (4), 269-272. ISSN: 1007-6417.
- Schmitt, F. G., Ma, L., & Angounou, T. (2011). Multifractal analysis of the Dollar-Yuan and Euro-Yuan exchange rates before and after the reform of the peg. *Quantitative Finance*, *11* (4), 505-513. doi: 10.1080/14697681003785983.
- Shih, J. H., & Auerbach, R. P. (2010). Gender and stress generation: An examination of interpersonal predictors. *International Journal of Cognitive Therapy*, *3* (4), 332-344. doi: 10.1521/ijct.2010.3.4.332.
- Simonsohn, U. (2009). Direct risk aversion: evidence from risky prospects valued below their worst outcome. *Psychological science*, *20* (6), 686-92. doi: 10.1111/j.1467-9280.2009.02349.x.

- Sjöberg, L. (2003). Distal factors in risk perception. *Journal of Risk Research*, 6 (3), 187-211. doi: 10.1080/1366987032000088847.
- Spehar, B., Clifford, C. W. G., Newell, B. & Taylor, R. P. (2003). Universal aesthetic of fractals. *Computers and Graphics*, 27, 813-820. doi: 10.1016/S0097-8493(03)00154-7.
- Steger, M. F., Kashdan, T. B., Sullivan, B. A., & Lorentz, D. (2008). Understanding the search for meaning in life: Personality, cognitive style, and the dynamic between seeking and experiencing meaning. *Journal of Personality*, 76 (2), 199-228. doi: 10.1111/j.1467-6494.2007.00484.x.
- Stenstrom, E., & Saad, G. (2011). Testosterone, financial risk-taking, and pathological gambling. *Journal of Neuroscience, Psychology, and Economics* 4 (4), 254–266. doi: 10.1037/a0025963.
- Stone, E. R., Yates, J., Parker, F. & Andrew, M. (1997). Effects of numerical and graphical displays on professed risk-taking behaviour. *Journal of Experimental Psychology: Applied*, 3 (4), 243-256. doi: 10.1037/1076-898X.3.4.243.
- Stoyanov, M., Gunzburger, M., & Burkardt, J. (2011). Pink noise,  $1/f^\alpha$  noise, and their effect on solutions of differential equations. *International Journal for Uncertainty Quantification*, 1 (3), 257–278. doi: 10.1615/Int.J.UncertaintyQuantification.2011003089.
- Sun, W., Rachev, S., & Fabozzi, F. J. (2007). Fractals or I.I.D.: Evidence of long-range dependence and heavy tailedness from modeling German equity market returns. *Journal of Economics and Business*, 59 (6), 575-95. doi:10.1016/j.jeconbus.2007.02.001.
- Taffler, R. J., & Tuckett, D. (2012). Fund management: an emotional finance perspective. UK: Research Foundation of CFA Institute.

- Tarim, E. (2013). Situated cognition and narrative heuristic: evidence from retail investors and their brokers. *The European Journal of Finance*, 1-24. doi: 10.1080/1351847X.2013.858054.
- Taylor, M. P., & Allen, H. (1992). The use of technical analysis in the foreign exchange market. *Journal of International Money and Finance*, 11 (3), 304–314. doi: 10.1016/0261-5606(92)90048-3.
- Taylor, R. P. (2006). Reduction of physiological stress using fractal art and architecture. *Leonardo*, 39 (3), 245–251. doi: 10.1162/leon.2006.39.3.245.
- Taylor, R. P., Spehar, B., Van Donkelaar, P., & Hagerhall, C. M. (2011). Perceptual and physiological responses to Jackson Pollock's fractals. *Frontiers in Human Neuroscience*, 5 (60), 1-13. doi: 10.3389/fnhum.2011.00060.
- Tetlock, P. C. (2007). Giving content to investor sentiment: the role of media in the stock market. *Journal of Finance*, 62 (3), 1139-1168. doi: 10.1111/j.1540-6261.2007.01232.x.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, New Series*, 185 (4157), 1124-1131.
- Tuckett, D. (2011). *Minding the markets: An emotional finance view of financial instability*. New York, NY: Palgrave Macmillan.
- Tuckett, D. (2012). Financial markets are markets in stories: Some possible advantages of using interviews to supplement existing economic data sources. *Journal of Economic Dynamics & Control*, 36, 1077–1087. doi: 10.1016/j.jedc.2012.03.013.
- Umanath, K. (2012). Random walk in return on banking stocks – empirical evidence from India. *Global Management Review*, 6 (4), 51-60.

- Vácha, L. & Vošvrda, M. S. (2005). Dynamical agents' strategies and the fractal market hypothesis. *Prague Economic Papers*, 14(2), 163-70.
- Van der Linden, D., Tsaousis, I., & Petrides, K.V. (2012). Overlap between general factors of personality in the Big Five, Giant Three, and trait emotional intelligence. *Personality and Individual Differences*, 53 (3), 175-179. doi: 10.1016/j.paid.2012.03.001.
- Vega, J. D. L. (1688). Confusión de Confusiones. In: M. S. Fridson (Ed.) *Extraordinary Popular Delusions and the Madness of Crowds & Confusión de Confusiones*. (1996). New York, NY: John Wiley & Sons.
- Walia, N., & Kiran, Ravi. (2012). Understanding the risk anatomy of experienced mutual fund investors. *Journal of Behavioral Finance*, 13 (2), 119-125. doi: 10.1080/15427560.2012.673517.
- Watanabe, T. (1988). Effect of irrelevant differences as a function of the relations between relevant and irrelevant dimensions in the same-different task. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 132-142. doi: 10.1037/0096-1523.14.1.132.
- Weber, E. U., & Hsee, C. (1998). Cross-cultural differences in risk perception, but cross-cultural similarities in attitudes towards perceived risk. *Management Science*, 44 (9) 1205-1217. doi: 10.1287/mnsc.44.9.1205.
- Weber, E. U., Siebenmorgen, N., & Weber, M. (2005). Communicating asset risk: How name recognition and the format of historic volatility information affect risk perception and investment decisions. *Risk Analysis*, 25 (3), 597-609. doi: 10.1111/j.1539-6924.2005.00627.x.

- Weber, M., Weber, E. U., & Nasic, A. (2013). Who takes risks when and why: Determinants of changes in investor risk taking. *Review of Finance*, 17, 847–883. Available at SSRN: <http://ssrn.com/abstract=1441273>.
- Westheimer, G. (1991). Visual discrimination of fractal borders. *Proceedings of the Royal Society of London B*, 243 (1308), 215-219. doi: 10.1098/rspb.1991.0034.
- Williams, C. (1974). The effect of an irrelevant dimension on “same”-“different” judgments of multidimensional stimuli. *Quarterly Journal of Experimental Psychology*, 26, 26-31.
- Woollen, B. (2011). Investment risk and the mind of the financial leader. *Consulting Psychology Journal: Practice and Research*, 63 (4), 254-271. doi: 10.1037/a0025245.
- Wornell, G. W., & Oppenheim, A. V. (1992). Estimation of fractal signals from noisy measurements using wavelets. *IEEE Transactions on Signal Processing*, 40 (3), 611-623. doi: 10.1109/78.120804.
- Yen, G., & Lee, C. F. (2008). Efficient Market Hypothesis (EMH): past, present and future. *Review of Pacific Basin Financial Markets and Policies*, 11 (2), 305–329. doi: 10.1142/S0219091508001362.
- Zaleskiewicz, T. (2011). Financial forecasts during the crisis: Were experts more accurate than laypeople? *Journal of Economic Psychology*, 32 (3), 384–390. doi: 10.1016/j.joep.2011.02.003.

# Appendices

## Appendix A: question list for Experiment 5 in Chapter 2

### Question list

1. List three features that distinguished high M graphs from low M graphs:
  - a. \_\_\_\_\_
  - b. \_\_\_\_\_
  - c. \_\_\_\_\_
  
2. How would you describe graphs with  $M < 50$ ?  
\_\_\_\_\_
  
3. How would you describe graphs with  $M > 50$ ?  
\_\_\_\_\_
  
4. Was it easier for you to assess the “M” value of graphs with  $M < 50$ , or of graphs with  $M > 50$ ? (please circle a or b)
  - a. Easier to assess M value for  $M < 50$
  - b. Easier to assess M value for  $M > 50$
  
5. What, do you think, was your average error at the test stages?  
\_\_\_\_\_
  
6. What is the likelihood (0-100) that your mean error in the test stages was less than .05?  
\_\_\_\_\_

7. Would you prefer investing money in assets whose price graphs have a relatively high “M” value (higher than 50) or a low “M” value (lower than 50)?(please circle a or b)
- a. I would prefer investing money in assets with  $M < 50$ .
  - b. I would prefer investing money in assets with  $M > 50$ .

Why?

Reason: \_\_\_\_\_  
\_\_\_\_\_

8. Which graphs, do you think, represent prices of assets which are riskier to invest in, graphs with  $M < 50$  or graphs with  $M > 50$ ? (please circle a or b)
- a. Graphs with  $M < 50$  represents riskier assets.
  - b. Graphs with  $M > 50$  represents riskier assets.

Thank you for your participation ☺

**Appendix B: Interactions and tests of simple effects in Experiments in chapter 6.**

Table B.1 Interaction and simple tests of simple effects in Experiment 1 in Chapter 6. DV denotes dependent variables, and IV – independent variables.

Repeated measures ANOVA		Interaction	Results of tests of simple effects
DV	IV		
local steepness of the data graphs	State, forecast horizon, the Hurst exponent, instance	State and horizon (F (2, 58) = 159.79; p < .001; partial $\eta^2$ = .85)	For each horizon level, steepness of the data was smaller after scaling than before it (for horizon of 2 days, F (1, 29) = 247.16; p < .001; partial $\eta^2$ = .90, for horizon of 15 days, F (1, 29) = 60.95; p < .001; partial $\eta^2$ = .68, and for horizon of 100 days, F (1, 29) = 68.80; p < .001; partial $\eta^2$ = .70). After scaling, longer forecast horizons resulted in graphs with higher local steepness (F (2, 28) = 127.51; p < .001; partial $\eta^2$ = .90).
		State and the Hurst exponent (F (2, 58) = 36.40; p < .001; partial $\eta^2$ = .56)	At each Hurst exponent value, scaling reduced the local steepness of the graphs (for H = 0.3, F (1, 29) = 34.44; p < .001; partial $\eta^2$ = .54, for H = 0.5, F (1, 29) = 18.27; p < .001; partial $\eta^2$ = .39, and for H = 0.7, F (1, 29) = 5.23; p < .001; partial $\eta^2$ = .15). After scaling, local steepness of graphs with higher Hurst exponents was still lower (F (2, 28) = 222.37; p < .001; partial $\eta^2$ = .94). (Before the scaling, local steepness of graphs with higher Hurst exponents was lower, as expected from the definition of H).
		Forecast	For each horizon, the steepness of the graphs was larger

<p>horizon and the Hurst exponent (F (4, 116) = 136.69; p &lt; .001; partial <math>\eta^2</math> = .83)</p>	<p>when H was smaller (in both the data and the scaled graphs). This effect increased as Hurst exponent increased (for forecast horizon of 2 days, F (2, 28) = 331.41; p &lt; .001; partial <math>\eta^2</math> = .96, for forecast horizon of 15 days, F (2, 28) = 374.30; p &lt; .001; partial <math>\eta^2</math> = .96, for forecast horizon of 100 days, F (2, 28) = 628.40; p &lt; .001; partial <math>\eta^2</math> = .98). For each value of the Hurst exponent, the local steepness of the graphs increased with the horizon (for H=0.3, F (2, 28) = 124.71; p &lt; .001; partial <math>\eta^2</math> = .90, for H=0.5, F (2, 28) = 108.94; p &lt; .001; partial <math>\eta^2</math> = .87, and for H=0.3, F (2, 28) = 95.86; p &lt; .001; partial <math>\eta^2</math> = .87).</p>
<p>Oscillation of the data graphs (State, forecast horizon, the Hurst exponent, instance) (F (2, 58) = 204.46; p &lt; .001; partial <math>\eta^2</math> = .88).</p>	<p>For horizon of two days, oscillation was smaller in the scaled graphs than in the original graphs (F (1, 29) = 239.69; p &lt; .001; partial <math>\eta^2</math> = .89). The same phenomenon occurred for forecast horizon of 15 days (F (1, 29) = 70.04; p &lt; .001; partial <math>\eta^2</math> = .71). However, for the long time horizon (100 days), oscillation was larger in the scaled graphs than in the original graphs (F (1, 29) = 55.81; p &lt; .001; partial <math>\eta^2</math> = .66). In the scaled graphs, the oscillation was higher when horizon was longer (F (2, 28) = 161.63; p &lt; 0.001; partial <math>\eta^2</math> = 0.92). (In unscaled data graphs oscillation was the same whether forecast horizon was large or small).</p>
<p>State and the Hurst exponent (F (2, 58) =</p>	<p>For each H value, oscillation was larger in the original data than in the scaled graphs (for H = 0.3, F (1, 29) = 188.85; p &lt; .001; partial <math>\eta^2</math> = .87, for H = 0.5, F (1, 29) = 30.70; p &lt; .001; partial <math>\eta^2</math> = .51, and for H = 0.7, F (1, 29) = 54.63; p</p>

		181.29; p < .001; partial $\eta^2 =$ .86).	< .001; partial $\eta^2 = .65$ ).  In the scaled graphs, when the Hurst exponent was smaller, the oscillation was larger (F (2, 28) = 890.57; p < .001; partial $\eta^2 = .99$ ).  (In the data graphs, when the Hurst exponent was smaller, the oscillation was larger).
	Hurst exponent and forecast horizon	(F (4, 116) = 43.89; p < .001; partial $\eta^2 =$ .60)	At each of the forecast horizons, oscillation was larger when H was smaller (for the horizon of two days, F (2, 28) = 1404.68; p < .001; partial $\eta^2 = .99$ , for the horizon of 15 days, F (2, 28) = 1175.87; p < .001; partial $\eta^2 = .99$ , and for the forecast horizon of 100 days, F (2, 28) = 3569.82; p < .001; partial $\eta^2 = 0.99$ ).  For each Hurst exponent values, oscillation was higher when horizon was longer (for H = 0.3, F (2, 28) = 169.40; p < .001; partial $\eta^2 = .92$ , for H = 0.5, F (2, 28) = 115.43; p < .001; partial $\eta^2 = .89$ , and for H = 0.7, F (2, 28) = 108.00; p < .001; partial $\eta^2 = .89$ ).
FD1	Horizon, Hurst exponent, and instance	Hurst exponent and horizon (F (3.09, 89.73) = 5.44; p = .002; partial $\eta^2 =$ .16)	For each H value, FD1 was larger when forecast horizon was larger (for H = 0.3, F (2, 28) = 33.00; p < .001; partial $\eta^2 = .70$ , for H = 0.5, F (2, 28) = 24.68; p < .001; partial $\eta^2 = .64$ , and for H = 0.7, F (2, 28) = 31.75; p < .001; partial $\eta^2 = .69$ ).  For forecast horizons of 15 and 100 days, FD1 was larger when Hurst exponent was smaller (for forecast horizon of 15 days F (2, 28) = 11.16; p < .001; partial $\eta^2 = .44$ , for forecast horizon of 100 days F (2, 28) = 6.68; p = .004; partial $\eta^2 = .32$ ).

<p>Hurst exponent and Instance (F (6.79, 196.97) = 7.67; p = .002; partial <math>\eta^2</math> = .21),</p>	<p>For small and medium H values, the effects of instance on FD1 were smaller than those obtained for large H values (for H = 0.3, F (4, 26) = 5.41; p = .003; partial <math>\eta^2</math> = .45, and for H = 0.5, F (4, 26) = 5.73; p = .002; partial <math>\eta^2</math> = .47, for H = 0.7, F (4, 26) = 12.55; p &lt; .001; partial <math>\eta^2</math> = .66).</p>
<p>Horizon and Instance (F (4.41, 127.89) = 18.28; p = .002; partial <math>\eta^2</math> = .39)</p>	<p>For small and medium forecast horizon, the effect of instance on FD1 was insignificant. However, for forecast horizon of 100 days, a strong effect of instance on FD1 was obtained (F (4, 26) = 14.93; p &lt; .001; partial <math>\eta^2</math> = .70).</p>

Table B.2 The results of a three-way repeated measures ANOVA on FD2 and FError. First panel: main effects. Second panel: interaction and tests of simple effects in Experiment 1 in Chapter 6. DV denotes dependent variables, and IV – independent variables.

Repeated measures ANOVA		Results: main effects
DV	IV	
FD2	Horizon, Hurst exponent, and instance	FD2 was larger when the forecast horizon was larger ( $F(1.37, 39.64) = 86.38$ ; $p < .001$ ; partial $\eta^2 = .75$ ) and when the Hurst exponent was smaller ( $F(2, 59) = 13.58$ ; $p < .001$ ; partial $\eta^2 = .32$ ). Graph instance had a significant effect on FD2 ( $F(3.70, 107.42) = 15.55$ ; $p < .001$ ; partial $\eta^2 = .35$ ). All interactions were significant. I report the results of the interactions and the corresponding simple tests below.
FError	Horizon, Hurst exponent, and instance	FError was larger when the Hurst exponent was smaller ( $F(2, 58) = 57.15$ ; $p < .001$ ; partial $\eta^2 = .66$ ) and when the forecast horizon was larger ( $F(1.2, 34.81) = 246.25$ ; $p < .001$ ; partial $\eta^2 = .90$ ). Instance had a significant effect on FError ( $F(4, 116) = 35.45$ ; $p < .001$ ; partial $\eta^2 = .55$ ). As before, all interactions were significant. I report the results of these interactions and the corresponding simple tests below.

Repeated measures ANOVA		Interaction	Results of tests of simple effects
DV	IV		
FD2	Horizon, Hurst exponent, and instance	Hurst exponent and horizon (F (3.05, 88.39) = 6.49; p < .001; partial $\eta^2$ = .18)	For each Hurst exponent, FD2 was larger when forecast horizon was larger (for H = 0.3, F (2, 28) = 34.17; p < .001; partial $\eta^2$ = .71, for H = 0.5, F (2, 28) = 26.32; p < .001; partial $\eta^2$ = .65, and for H = 0.7, F (2, 28) = 34.20; p < .001; partial $\eta^2$ = .71). For forecast horizon of 15 days FD2 was larger when Hurst exponent was smaller (F (2, 28) = 9.29; p < .001; partial $\eta^2$ = .40).
		Hurst exponent and instance (F (6.042, 175.21) = 9.54; p < .001; partial $\eta^2$ = .25)	The effects of instance on FD2 increased with H (for H = 0.3, F (4, 26) = 5.88; p = .002; partial $\eta^2$ = .48, for H = 0.5, F (4, 26) = 6.92; p = .001; partial $\eta^2$ = .52, and for H = 0.7, F (4, 26) = 9.64; p < .001; partial $\eta^2$ = .60).
		Horizon and instance (F (4.73, 137.05) =	For medium and large forecast horizons, I obtained significant simple effects of instance on FD2 (for forecast horizon of 15 days, F (4, 26) = 4.39; p = .008; partial $\eta^2$ = .40, and for forecast horizon of 100 days, F (4, 26) =

		15.61; p < .001; partial $\eta^2 =$ .35)	11.18; p = .008; partial $\eta^2 = .63$ ).
FError	Horizon, Hurst exponent, and instance	Hurst exponent and horizon (F (3.4, 98.60) = 16.68; p < .001; partial $\eta^2 =$ .37)	For each Hurst exponent value, FError was larger when forecast horizon was longer (for H = 0.3, F (2, 28) = 145.07; p < .001; partial $\eta^2 = .91$ , for H = 0.5, F (2, 28) = 201.41; p < .001; partial $\eta^2 = .94$ , and for H = 0.7, F (2, 28) = 54.67; p < .001; partial $\eta^2 = .80$ ).  For medium forecast horizons, the effect of H on FError was larger than for small and large forecast horizons (for forecast horizon of 2 days, F (2, 28) = 17.61; p < .001; partial $\eta^2 = .56$ , for forecast horizon of 15 days, F (2, 28) = 59.92; p < .001; partial $\eta^2 = .81$ , for forecast horizon of 100 days, F (2, 28) = 10.24; p < .001; partial $\eta^2 = .42$ ).
		Hurst exponent and instance (F (7, 202) = 19.82; p < .001; partial $\eta^2 =$ .41).	The effect of graph instance on FError was the largest for H = 0.5 (for H = 0.3, F (4, 26) = 38.45; p < .001; partial $\eta^2 =$ .86, for H = 0.5, F (4, 26) = 75.21; p < .001; partial $\eta^2 =$ .92, and for H = 0.7, F (4, 26) = 22.82; p < .001; partial $\eta^2 =$ .78).
	Horizon and instance	(F (3.42, 99.13) =	The effect of instance increased with forecast horizon (for forecast horizon of 2 days, F (4, 26) = 19.68; p < .001; partial $\eta^2 = .75$ , for forecast horizon of 15 days, F (4, 26) = 39.65; p < .001; partial $\eta^2 = .86$ , for forecast horizon of 100

---

41.64; days,  $F(4, 26) = 59.17$ ;  $p < .001$ ; partial  $\eta^2 = .90$ .  
 $p < .001$ ;  
partial  $\eta^2$   
 $= .59$ ).

---

Table B.3 Interactions and tests of simple effects in Experiment 2 in Chapter 6.

Repeated measures ANOVA		Interaction	Results of tests of simple effects
DV	IV		
local steepness	State, the Hurst exponent and the forecast density	state and the Hurst exponent (F (4, 37.06) = 308.98; p < .001; partial $\eta^2$ = 0.91).	For all H values, local steepness was significantly smaller when H was larger (for H = 0.3, F (1, 29) = 364.29; p < .001; partial $\eta^2$ = .93, for H = 0.4, F (1, 29) = 230.19 ; p < .001; partial $\eta^2$ = .89, for H=0.5, F (1, 29) = 291; p < .001; partial $\eta^2$ = .91, for H=0.6, F (1, 29) = 348.08 ; p < .001; partial $\eta^2$ = .92, for H=0.7, F (1, 29) = 225.09 ; p < .001; partial $\eta^2$ = .89).  In the original graphs, local steepness was larger when H was smaller (F (4, 26) = 563525; p < 0.001; partial $\eta^2$ = 1). The same relation was preserved after participants smoothed the data graphs (F (4, 26) = 13.71; p < .001; partial $\eta^2$ = .68).
Oscillation	State, the Hurst exponent and the forecast density	state and the Hurst exponent (F (1.71, 49.55) = 129.45 ; p < .001; partial $\eta^2$ = 0.82).	For all H values, the oscillation of the data was larger before the smoothing than after smoothing (for H = 0.3, F (1, 29) = 181.40; p < .001; partial $\eta^2$ = .86, for H = 0.4, F (1, 29) = 115.73; p < .001; partial $\eta^2$ = .80, for H = 0.5, F (1, 29) = 116.15; p < .001; partial $\eta^2$ = .80, for H = 0.6, F (1, 29) = 133.64; p < .001; partial $\eta^2$ = .82, for H=0.7, F (1, 29) = 75.35; p < .001; partial $\eta^2$ = .72).  Before the smoothing, oscillation of graphs was

---

larger when H was smaller ( $F(4, 26) = 304.79$ ;  $p < .001$ ; partial  $\eta^2 = .98$ ). The same relation was observed after smoothing data graphs ( $F(4, 26) = 79.93$ ;  $p < .001$ ; partial  $\eta^2 = .92$ ).

---