Entropy Analysis of Financial Time Series

Citation for published version (APA):

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
Entropy Analysis of Financial Time Series

A thesis submitted to The University of Manchester for the degree of Doctor in Business Administration.

2015

Stephan Schwill

Manchester Business School
Contents

1 Introduction 11
   1.1 Bivariate Analysis of Drawdowns 12
   1.2 Information Flows between FX Volatility Regimes 14
   1.3 Volatility Timing of Hedge Funds and Entropy 16

2 Entropy Measure and Information Theory 19
   2.1 Introduction 20
   2.2 Entropy 20
   2.3 Mutual Information and Transfer Entropy 23
      2.3.1 Definition 23
      2.3.2 Entropy Estimation, Statistical Features 26
      2.3.3 Process Entropy 31
      2.3.4 Entropy Magnitude 32
   2.4 Approximate Entropy (ApEn) 33
      2.4.1 ApEn Definition 33
      2.4.2 ApEn Estimation 36

3 Bivariate Analysis of Drawdowns 38
   3.1 Introduction 39
   3.2 Related Literature 41
      3.2.1 Drawdowns Literature 41
      3.2.2 Entropy Econometric Literature 46
   3.3 Data 48
   3.4 Information Theoretic Method 51
      3.4.1 Entropy Measures 51
   3.5 Empirical Results 57
      3.5.1 Dependency between Draws in EUR/USD and GBP/USD 57
      3.5.2 Information Flows between Draws in EUR/USD and GBP/USD 61
   3.6 Conclusion 66

4 Information Flows between FX Volatility Regimes 69
   4.1 Introduction 70
   4.2 Hidden Markov Models 72
      4.2.1 Foundations 72
      4.2.2 HMM Estimation 74
   4.3 Data 77
   4.4 HMM Estimation Results 79
   4.5 Dependence between Hidden States 82
      4.5.1 Hidden State Dependence 84
      4.5.2 Hidden State Information Flow 86
5 Volatility Timing of Hedge Funds and Entropy

5.1 Introduction

5.2 Entropy of Volatility

5.2.1 GARCH(1,1)

5.2.2 Changing Volatility Level

5.2.3 Markov Switching Volatility

5.3 Data

5.4 Entropy of Realized and Implied Volatility

5.4.1 Realised Volatility

5.4.2 VIX

5.4.3 Spread

5.5 Empirical Results

5.5.1 Timing Model

5.5.2 Panel Regression

5.5.3 Strategy and Survivorship

5.5.4 Conditional Fama French Model

5.5.5 VIX and Volatility Spread

5.6 Conclusion

6 Conclusion
# List of Tables

3.1 Draws in FX, Equity and Interest Rate markets .................................. 45
3.2 EUR/USD and GBP/USD return and draw statistics. ............................. 49
3.3 Stretched exponential distribution fit .................................................. 50
3.4 Entropy $\hat{H}$ of daily and hourly EUR/USD and GBP/USD ................. 56
3.5 Mutual Information (daily returns) ....................................................... 58
3.6 Mutual Information of hourly returns .................................................. 60
3.7 Transfer entropy EUR/USD $\rightarrow$ GBP/USD ....................................... 62
3.8 Transfer entropy GBP/USD $\rightarrow$ EUR/USD ....................................... 63
3.9 Transfer entropy EUR/USD $\rightarrow$ GBP/USD (hourly prices) ................. 64
3.10 Transfer entropy GBP/USD $\rightarrow$ EUR/USD (hourly prices) ................. 66
4.1 Correlation Matrix ......................................................................... 78
4.2 Summary Statistics ......................................................................... 78
4.3 Hidden Markov Model Estimates ....................................................... 80
4.4 Entropy rate $h_{\infty}$ and sample $\hat{H}(X_2|X_1)$ in (bit) for the stationary distribution of HMM process. ....................................................... 84
4.5 Volatility States Co-Movement and Interdependence ............................ 85
4.6 Volatility States Information Flows ..................................................... 90
4.7 Summary Statistics ......................................................................... 92
4.8 Expected Time State Duration .......................................................... 93
5.1 Summary statistics .......................................................................... 111
5.2 Panel Regression Conditional CAPM ............................................... 120
5.3 Panel Regression Conditional CAPM: Live, Dead and LS Equity Funds 124
5.4 Panel Regression Conditional 3F model: LS Equity funds. .................... 126
5.5 Panel Regression: RV, VIX, $\sqrt{RV} - VIX$. .................................... 129
List of Figures

2.1 Coin Toss Entropy ........................................... 22
2.2 Entropy Measures .......................................... 25
2.3 Entropy Estimation .......................................... 30
2.4 ApEn for AR(1) process. ................................. 35
2.5 ApEn with $m = 1, \ldots, 5$ and $N = 20, \ldots, 500$. ... 37
3.1 EUR/USD, GBP/USD time series ....................... 48
3.2 Discretization example .................................. 53
3.3 Block Entropy EUR/USD ................................. 56
3.4 Mutual Information (daily returns) ..................... 59
3.5 $\hat{E}_T^{EUR\rightarrow GBP}$ and $\hat{E}_T^{GBP\rightarrow EUR}$ (daily returns) ... 62
3.6 Draws Information Flow from hourly EUR/USD to GBP/USD ... 64
3.7 Draws Information Flow from hourly GBP/USD to EUR/USD ... 65
4.1 A Hidden Markov Model, where $\{q_t\}$ represents the hidden states, $\{y_t\}$ the emissions and the arrows describe dependence relations... 73
4.2 Currency Time Series ................................... 77
4.3 State Sequence for the daily EUR/USD exchange rate ....... 81
4.4 EUR/USD state dependent return distributions ............ 81
4.5 Duration of sequence remaining in the high volatility state 1 ... 93
4.6 Duration of sequence remaining in the low volatility state 2 .... 93
5.1 ApEn: GARCH(1,1) .......................................... 103
5.2 ApEn: Volatility regime change .......................... 105
5.3 ApEn: Regime change versus mixed normal distribution. .... 106
5.4 ApEn: Markov switching process .......................... 108
5.5 Hedge Funds: AuM, live funds ........................... 110
5.6 Realized Variance .......................................... 113
5.7 RV: Maximum entropy episode ........................... 114
5.8 RV: Minimum entropy episode ............................ 115
5.9 VIX ....................................................... 116
5.10 Volatility Spread ($\sqrt{RV} - VIX$) ...................... 116
Entropy Analysis of Financial Time Series

The University of Manchester

Stephan Schwill

Doctor of Business Administration

2015

Abstract

This thesis applies entropy as a model independent measure to address research questions concerning the dynamics of various financial time series. The thesis consists of three main studies as presented in chapters 3, 4 and 5. Chapters 3 and 4 apply an entropy measure to conduct a bivariate analysis of drawdowns and drawups in foreign exchange rates. Chapter 5 investigates the dynamics of investment strategies of hedge funds using entropy of realised volatility in a conditioning model.

In all three studies, methods from information theory are applied in novel ways to financial time series. As Information Theory and its central concept of entropy are not widely used in the economic sciences, a methodology chapter was therefore included in chapter 2 that gives an overview on the theoretical background and statistical features of the entropy measures used in the three main studies. In the first two studies the focus is on mutual information and transfer entropy. Both measures are used to identify dependencies between two exchange rates. The chosen measures generalise, in a well defined manner, correlation and Granger causality. A different entropy measure, the approximate entropy, is used in the third study to analyse the serial structure of S&P realised volatility.

The study of drawdowns and drawups has so far been concentrated on their univariate characteristics. Encoding the drawdown information of a time series into a time series of discrete values, Chapter 3 uses entropy measures to analyse the correlation and cross correlations of drawdowns and drawups. The method to encode the drawdown information is explained and applied to daily and hourly EUR/USD and GBP/USD exchange rates from 2001 to 2012. For the daily series, we find evidence of dependence among the largest draws (i.e. 5% and 95% quantiles), but it is not as strong as the correlation between the daily returns of the same pair of FX rates. There is also dependence between lead/lagged values of these draws. Similar and stronger findings were found among the hourly data. We further use transfer entropy to examine the spill over and lead-lag information flow between drawup/drawdown of the two exchange rates. Such information flow is indeed detectable in both daily and hourly data. The amount of information transferred is considerably higher for the hourly than the daily data. Both daily and hourly series show clear evidence of information flowing from EUR/USD to GBP/USD and, slightly stronger, in the reverse direction. Robustness tests, using effective transfer entropy, show that the information measured is not due to noise.

Chapter 4 uses state space models of volatility to investigate volatility spill overs between exchange rates. Our use of entropy related measures in the investigation
of dependencies of two state space series is novel. A set of five daily exchange rates from emerging and developed economies against the dollar over the period 1999 to 2012 is used. We find that among the currency pairs, the co-movement of EUR/USD and CHF/USD volatility states show the strongest observed relationship. With the use of transfer entropy, we find evidence for information flows between the volatility state series of AUD, CAD and BRL.

Chapter 5 uses the entropy of S&P realised volatility in detecting changes of volatility regime in order to re-examine the theme of market volatility timing of hedge funds. A one-factor model is used, conditioned on information about the entropy of market volatility, to measure the dynamic of hedge funds equity exposure. On a cross section of around 2500 hedge funds with a focus on the US equity markets we find that, over the period from 2000 to 2014, hedge funds adjust their exposure dynamically in response to changes in volatility regime. This adds to the literature on the volatility timing behaviour of hedge fund manager, but using entropy as a model independent measure of volatility regime. Finally, chapter 6 summarises and concludes with some suggestions for future research.
Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright

The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on Presentation of Theses.
Acknowledgments

My utmost thanks and sincere gratitude go to my supervisor, Prof. Dr. Ser-Huang Poon, who guided and supported me throughout this journey.

I would like to dedicate this work to my love, Doreen, whose unwavering support, and encouragement during the ups and downs of a long study have made this thesis possible. And to my parents, who instilled me with a sense of wonder and curiosity about the world around me.
Chapter 1

Introduction

The work of Shannon (1948) on the theory of communication, and Jaynes (1963) ideas on the maximum entropy principle, to name just two authors in this area, has led to the development of concepts that have been applied to many research areas. Questions about information, uncertainty, entropy and ignorance play a role in the theory at a fundamental level; information is a decrease in ambiguity, uncertainty is a state of knowledge in which logical reasoning is impossible, entropy is expected information and ignorance is not knowing that there is uncertainty. At a more concrete level, these concepts have been developed and used to generalize important results in statistics, such as the Cramér-Rao inequality Kullback (1954)\(^1\) and led to a better understanding of non-linear time series, dynamical systems and complexity.

This thesis consists of three studies in two research topics of finance, which have, as a common thread, the application of tools from information theory. Extreme market events, spillover and interdependence between foreign exchange (FX) markets is the first research topic. In chapter 3, crashes in currency pairs are analysed via drawdowns and drawups, for which a concept for the bivariate analysis is developed that uses mutual information and transfer entropy as information theoretic tools. FX markets are also studied in chapter 4, focusing on volatility spillovers between currency pairs, which are measured using the same entropy tools. Chapter

\(^{1}\)For an excellent and comprehensive review of Information Theory and Entropy Econometrics with its history, please consult Golan (2008).
5 is devoted to analysing a hedge fund manager’s ability to time changes in market volatility regimes. A conditional factor model is built, using approximate entropy to identify volatility regime changes in equity markets. As information theory and its central concept of entropy are not widely used in the economic sciences, we provide in chapter 2 an overview of the theoretical background and statistical features of the entropy measures used in the three main studies. The following subsections explain the research motivation and findings of chapters 3, 4 and 5.

1.1 Bivariate Analysis of Drawdowns

In financial time series, drawdowns (drawups) are defined as the cumulative return from a local maximum (minimum) to the next local minimum (maximum). Draws, including both drawdowns and drawups, is a series that is irregularly spaced in time. It provides a statistical measure for dependence in different market regimes. The concept of draws is important, not only from a theoretical point of view but in its use in various contexts in the finance industry, e.g. in asset management, where drawdowns tend to drive redemptions. For buy and hold portfolios that are subject to mark to market limits, extended drawdowns can force unwanted liquidations and readjustments of portfolios. Furthermore, large drawdowns could trigger feedback mechanisms through portfolio insurance and fire sales at funds, which have the tendency to drive prices further down. Many empirical properties of drawdowns in financial time series of various asset classes have been documented in a series of papers, such as Sornette (2003), Johansen and Sornette (2000), and Johansen and Sornette (2001). The characteristic size of draws is stable within an asset class but varies across asset classes. For virtually all markets, the authors found large draws are outliers to the parametric model (Johansen and Sornette (2001)) of the draw distribution. Johansen and Sornette (2001) lay out a crash theory that explains drawdown/drawup that appeals to fit almost all time series. They claim that drawdown/drawup can be explained by the onset of a well-defined microstructural phase transition, which is characterized by the emergence of the sudden
persistence of daily drops, coupled with an increase in the correlated amplitude of
the drops. Rebonato and Gaspari (2006) confirm these findings also for the US
Treasury market.

**Research Motivation and Findings**

Many cases are known where a large market event triggered a reaction in another
market. Volatility spillover and contagion are well documented phenomena in finan-
cial markets (see Engle, Ito, and Lin (1990)). The question of cross-correlation and
interdependence is also a natural question in the case of draws. To our knowledge,
there has been no study done on the relationship of large drawdowns (drawups)
with financial time series. Rebonato and Gaspari (2006) qualitatively examined on
a case-by-case basis if large draw-downs/drawups from different maturity blocks of
US interest rates coincide with each other. But no quantitative, systematic mea-
sure has been proposed so far on how to analyse them. One obstacle arising when
moving from fixed interval returns to draws is that time is compressed and is no
longer equally spaced. Instead of having time series of synchronous measurements,
the draws resulting from two time series will in general be clocked very differently.
A further obstacle is that it is often argued that large drawdowns/drawups are gen-
erated from nonlinear mechanisms (see Sornette (2003)), thus making conventional
parametric models and statistical inference no longer applicable. Converting draws
into symbolic time series and using entropy quantities as non-linear, model inde-
pendent and flexible tools, we were able to answer the question: can a large draw
over possibly multiple periods that is unfolding in one market help to predict a large
draw in another market around the same period?

Encoding the drawdown information of a time series into a time series of discrete
values, we use entropy to analyse the correlation and cross correlations of draw-
downs and drawups between two key exchange rates. The method we used to encode
the drawdown information is novel and enables us to study, for the first time, draws
in a bivariate setting. We investigate daily and hourly EUR/USD and GBP/USD
exchange rates from 2001 to 2012. For the hourly and the daily data, we found evidence of dependence among the largest draws. We were also able to identify information flows between drawup/drawdown of the two exchange rates. Both daily and hourly series show clear evidence of information flowing from EUR/USD to GBP/USD and, slightly more pronounced, in the reverse direction. Robustness tests, using effective transfer entropy, show that the information measured is not due to noise. By using the ‘relative explanation added’ for the information gain of the entropy of the process, we conclude that there is a measurable information transfer between the two exchange rates, which can be potentially useful for forecasting and risk management.

1.2 Information Flows between FX Volatility Regimes

History provides many examples of financial crisis and how turbulence in one financial market spread across other markets within a short time period. Evidence for volatility transmissions has been found in many markets. One of the first studies on volatility spillover focussing on foreign exchange markets is Engle, Ito, and Lin (1990). Using the Japanese yen and US dollar exchange rate, the authors found evidence of intraday volatility spillover across different markets. Hong (2001) investigated the German mark and Japanese yen denominated in US dollars and found evidence of unidirectional volatility spillover from the German mark to the Japanese yen.

The methods to identify spillover, co-movement or interdependence vary considerably. Engle, Ito, and Lin (1990) employed a GARCH model in a vector autoregression model to test if conditional variances are affected by the squared innovations (i.e. news, information) in other markets. Cheung and Ng (1996) used the cross-correlation function between squared residuals, that are standardized by their individual conditional variance estimators, as test for volatility spillover. Hamilton
(1989) introduced a model in which changes in regimes are governed by a hidden state process, which is Markovian. It gives more flexibility to the different adjustment speeds of volatility, and the persistence of the estimated volatility appears to be significantly different from that of the standard single-regime GARCH estimated volatility. Further models have been proposed, e.g. the SWARCH model in Hamilton and Susmel (1994) that has a Markov-modulated GARCH process, and the bivariate Markov switching process in Bialkowski and Serwa (2005), to name a few.

Research Motivation and Findings

In this study, we use a model similar to Hamilton (1989) to identify volatility regimes in a set of currencies all against the US dollar. We define a two-state hidden Markov model with high and normal volatility states. The estimated state process contains discrete values, which lends itself to an analysis with information theoretic tools. Following the procedures in the previous chapter, we use the entropy framework with mutual information and transfer entropy to examine the interdependence and volatility spillover among the exchange rates in our sample in a novel way.

We found evidence for various volatility regime relationships between the currency pairs in the sample. Among the European currencies (viz. EUR/USD, GBP/USD, CHF/USD), volatility regime co-movements are identified. CAD/USD and AUD/USD were found to exhibit volatility co-movement and, to a lesser extent, interdependency with a time lag. We were able to find evidence for information flows between the volatility states of currency pairs. Most notable is an information flow from EUR/USD to GBP/USD, which indicates a causal volatility spillover relationship, confirming the findings in Inagaki (2006) using a different model and set up. The work showcases the usefulness of the concepts of mutual information and transfer entropy when studying volatility spillover between markets.
1.3 Volatility Timing of Hedge Funds and Entropy

Hedge fund managers are in general free to change trading strategies, to allocate capital to asset classes and to choose leverage. Hedge fund managers often reallocate their capital dynamically in response to a change in the market environment. To pursue their investment goals, hedge fund managers scan the markets for signals and information that can give support to their analyses regarding the current and the future states of the market. From the information gathered, the fund manager forms an expectation of the future states of the market and develops a trading strategy to make profitable use of this knowledge. Absolute and relative price levels, market volatility and trend are information that fund manager uses in determining the investment opportunity set.

Treynor and Mazuy (1966) were among the first to study whether investment managers are able to profitably time changes in market conditions. The authors extended the CAPM by a quadratic term to capture the nonlinear exposure to the market depending on an expected high or low market return. They found, from a sample of 57 mutual funds, only one fund for which the hypothesis of market timing ability was acceptable. Henriksson and Merton (1981) formulated a market timing model modelled as an option payoff on the market return, with a strike price equal to the risk free rate. Ferson and Schadt (1996) defined a model of market timing where the risk exposure evolves linearly with the market observables. In their analysis, various variables that are public information are used as market condition. Using a monthly data set of mutual funds, the authors found evidence for changes in risk exposure in response to public information on the economy for several state variables.

Market volatility timing has also been actively researched. Busse (1999) uses daily returns of mutual funds to investigate the fund’s ability to time market volatility. His findings suggest that funds decrease market exposure when market volatility is high. The value of this timing ability is in higher Sharpe ratios for successful
volatility timing managers. The author further showed that surviving funds show sensitivity to market volatility, where non-surviving funds do not show this sensitivity. The economic value of volatility timing for investors has been investigated by Fleming, Kirby, and Ostdiek (2001). Using conditional analysis on short horizon asset allocation strategies, the authors found volatility timing strategies outperformed unconditional static strategies.

Research Motivation and Findings

Market environments where the prices and volatility develop with a high degree of irregularity inhibit managers’ ability to form a hypothesis on the further development of the market. Fund managers are therefore more likely to choose a lower level of exposure when there is no clear view on market expected level of volatility.

We extend the market timing model of Ferson and Schadt (1996), which assumes risk exposures evolve as a linear function of market observables. Various modifications of this baseline model are tested using the approximate entropy (ApEn) of market volatility series as the chosen state variable for the market condition. ApEn, which quantifies serial correlation patterns, is sensitive to volatility regime changes. Depending on the state of the market, the manager then adjusts the fund’s equity exposure via its market beta. Volatility is an important determinant for the level of market beta that a fund is choosing. Funds wishing to keep their volatility stable over time often target a specific Sharpe ratio and are more likely to reduce the fund’s beta when a higher volatility regime is expected.

The current analysis adds to the market volatility timing research on hedge funds. Through simulations, we demonstrate the use of ApEn in measuring changes in the serial structure of volatility. Popular econometric models of volatility are used to show the measure’s usefulness in distinguishing different volatility patterns. In a sample of 1,903 hedge funds, we found that managers adjust a fund’s market beta when volatility is settling into a new regime. Dead funds show no such market beta adjustment in response to a change in volatility regularity, differing from live funds.
where such an adjustment can be identified. The results are consistent with the findings from Busse (1999) on mutual funds. The results also show the usefulness of entropy as a model independent measure of volatility regime and patterns.
Chapter 2

Entropy Measure and Information Theory

Abstract

As information theory and the concept of entropy are not widely used in the economic sciences, we provide in this chapter an overview of the theoretical background and statistical features of the entropy measures used in the three main chapters. A brief introduction to the concept of entropy, as Shannon (1948) developed it, is given. This is followed by an overview of the basic features of Mutual Information and Transfer Entropy, which are used in chapters 3 and 4. Both measures are employed to identify dependencies between two exchange rates. A different entropy measure, the Approximate Entropy, is outlined in the last section. It is used in chapter 5 to analyse the serial structure of S&P 500 volatility.
2.1 Introduction

The seminal work of Shannon (1948) on a theory of communication and Jaynes (1963) on the maximum entropy principle form one line of research concerning information and uncertainty that has led, in recent years, to rich research in statistical sciences. Applications have touched a wide array of sciences including Econometrics, where it is sometimes termed Information and Entropy Econometrics Golan (2008).

Shannon (1948) developed his theory of communication in thermodynamics, where entropy is a measure of disorder in a physical system. In information theory, the interest is in finding an ‘optimal’ coding given the source data. Shannon solved this by introducing a quantity, later known as the Shannon Entropy, that corresponds to the average number of bits needed to optimally encode the source data. This concept and the ideas of Jaynes were used to enrich and complement the classical toolbox in econometrics. In our research, we are particularly interested in two developments in this research area, viz. Mutual Information and Approximate entropy, which we will introduce later in this chapter. Mutual information and its generalisation, transfer entropy, captures any dependency relations between two time series. Approximate entropy captures changes in a systems complexity.

The structure of this chapter is as follows: In the first section, we introduce the concept of Shannon Entropy, which forms the basis for all other tools we use. The second section introduces Mutual Information and Transfer Entropy. Both concepts are used as generalized cross-correlation measures, which we apply to time series of exchange rates. In the last section, the complexity measure, Approximate Entropy, is explained. We use it to measure patterns in time series and classify the time series by regularity or irregularity.

2.2 Entropy

Shannon (1948) developed his theory of communication based on concepts in thermodynamics, where entropy is a measure of disorder in a physical system. Entropy,
in a physical system, is the number of allowed microscopic states compatible with a
given observed macro state. Shannon used this in his work to relate to the average
number of bits to optimally encode a source.

Entropy is a measure of uncertainty in random variables. For a random variable
$X$ over a probability space $\Omega$ and probability distribution $p(x)$, entropy is defined,
according to Shannon (1948), as:

$$H(X) = -\int_{\Omega} p(x) \cdot \log_2(p(x)) dx \quad (2.1)$$

$$H(X) = -\sum_{x \in \Omega} p(x) \cdot \log_2(p(x)) \quad (2.2)$$

where $H(X)_{cont}$ and $H(X)_{discr}$ are for, respectively, continuous random variables
and discrete probability distribution. Entropy, as defined in the above equations,
and all the concepts described in this chapter can be applied to random variables
on any probability space. The formulation on discrete probability spaces is a special
case, but the concepts which we introduce in this section for discrete values can
be generalised for continuous random variables in a straightforward way. We will
therefore use the discrete version of the concepts only. A neat and brief version
of equations (2.1) and (2.2), is shown below:

$$H(X) = E_p \log \frac{1}{p(X)} \quad (2.3)$$

$H(X)$ in (2.3) expresses entropy as the expected value of the random variable
$\log(\frac{1}{p(X)})$. There are two important conventions implicit in the formula. First,
historically, the logarithm has been taken with base two, but any base would fulfil
the requirements of an information measure, as described by Shannon (1948). Sec-
ond, when computing entropy, the state with $p(x) = 0$ is set to 0, which can be
justified by $\lim_{p \to 0} p \cdot \log_2(p) = 0$. For entropy, the units of measurement are bits.

As an example, a coin that has an equal probability for head and tail has $H(X) =
-2 \cdot \frac{1}{2} \log_2(\frac{1}{2}) = 1$. The maximum entropy of 1 for a discrete variable with two states is
reached in this case. On the other hand, if the coin was biased and always produced
head (or tail), then $H(X) = -1 \cdot \log_2(1) - 0 \cdot \log_2(0) = 0$; the uncertainty is zero. Figure 2.1 shows the level of entropy dependent on the probability of head (or tail) coming up when a coin is tossed.

The uncertainty, as measured by entropy, can be understood as the average number of questions, each with 'yes/no' answer, needed to determine the value of the random variable. In the coin example, the questions are 'head?' (or 'tail'?). For 50% of the cases with a fair coin, the answer will be 'yes' and the questioning stops. In the other half of the cases, we are required to ask again. Entropy measures, in this sense, represents the average number of binary (yes/no) questions, expressed in units of bits. The higher the entropy the higher the uncertainty and the more questions we need to ask to determine the correct answer.

Entropy is bounded and the following inequality holds for discrete distributions $0 \leq H(X) \leq \log_2(|X|)$, where $|X|$ denotes the number of discrete values of $X$, and the equality holds if $X$ has a uniform distribution.

The joined entropy and conditional entropy of two variables $X$ and $Y$ with prob-
Ability space $\mathcal{X}, \mathcal{Y}$ are defined analogously:

$$H(X, Y) = - \sum_{x \in X, y \in Y} p(x, y) \cdot \log_2(p(x, y))$$

(2.4)

$$H(X|Y) = - \sum_{x \in X, y \in Y} p(x|y) \cdot \log_2(p(x|y))$$

(2.5)

The general case of joined entropy for multiple variables $X_1, \ldots, X_n$, $H(X_1, \ldots, X_n)$ is of importance when introducing transfer entropy below.

If $X$ and $Y$ are independent random variables, the conditional entropy is the same as the entropy of the variable:

$$H(X|Y) = - \sum_{x \in X, y \in Y} p(x|y) \cdot \log_2(p(x|y))$$

$$= - \sum_{x \in X, y \in Y} p(x) \cdot \log_2(p(x)) = H(X)$$

(2.6)

Conditional and joined entropy are related as follows:

$$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$$

(2.7)

In the case of independent variables $X$ and $Y$, equation (2.6) is a special case of this general chain rule for entropy (Cover and Thomas (2006)).

2.3 Mutual Information and Transfer Entropy

2.3.1 Definition

The joint and conditional entropy enable us to define simple measures of dependency between $X$ and $Y$. In particular, Mutual Information $I(X; Y)$ of two random variables, $X$ and $Y$, with joint distribution $p(x, y)$ is defined below. We also show, in the deduction below, how mutual information is related to the entropies of $X$ and $Y$. 

23
\[ I(X; Y) \equiv - \sum_{x \in X, y \in Y} p(x, y) \cdot \log_2 \frac{p(x, y)}{p(x)p(y)} \]  
\[ = - \sum_{x \in X, y \in Y} p(x, y) \cdot \log_2 \frac{p(x|y)}{p(y)} \]  
\[ = - \sum_{x \in X, y \in Y} p(x, y) \cdot \log_2 p(x) + \sum_{x \in X, y \in Y} p(x, y) \cdot \log_2 p(x|y) \]  
\[ = - \sum_{x \in X} p(x) \cdot \log_2 p(x) - \left( - \sum_{x \in X, y \in Y} p(x, y) \cdot \log_2 p(x|y) \right) \]  
\[ = H(X) - H(X|Y) \]  
\[ (2.8) \]

\[ (2.9) \]

\[ (2.10) \]

\[ (2.11) \]

\[ (2.12) \]

Mutual information measures the reduction of uncertainty about \( X \) from observing \( Y \). In the extreme case, where \( X \) and \( Y \) are independent such that \( p(x, y) = p(x) \cdot p(y) \) then \( I(X, Y) = 0 \), there is no mutual information. So the information gained is due to some dependence between \( X \) and \( Y \). In fact for Gaussian variables \( X \) and \( Y \sim \mathcal{N}(0, \sigma) \) correlated at \( \rho \), mutual information \( I(X, Y) \) has a simple analytic solution \(-\frac{1}{2} \log(1 - \rho^2)\) (see Cover and Thomas (2006), p.252).

With the chain rule, one can easily show that mutual information is symmetric \( I(X; Y) = I(Y; X) \) and, hence, mutual information cannot be used to determine the direction of an information flow. Mutual information is positive \( I(X; Y) \geq 0 \) and equal to 0 if and only if \( X \) and \( Y \) are independent. This important feature is used in a closely-related concept. The Kullback-Leibler distance, or the relative entropy, between two probability distributions is zero if and only if the two distributions are equal.

Mutual information can be modified to include lead-lag relationships:

\[ I(X; Y)_\tau = - \sum_{x_{n-\tau} \in X, y_n \in Y} p(x_{n-\tau}, y_n) \cdot \log_2 \frac{p(x_{n-\tau}, y_n)}{p(x_n)q(y_n)} \]  
\[ (2.13) \]

Mutual information with and without lag is depicted in Figure 2.2 (b).

It is important to note that mutual information does not imply causality. As pointed out by Schreiber (2000), introducing a time delay in one of the observa-
tions does not distinguish information that is actually generated from a common response to input signal or a common history driven by an external factor. Transfer entropy was introduced by the author as an information measure that corrects for this shortcoming of mutual information.

![Figure 2.2: Entropy Measures](image)

The left diagram shows the relationship between the various entropies and mutual information. The right diagrams show a) $h_\infty$ and $H(X_0|X_{-1})$, b) $I(X_t;Y_t), I(X_t;Y_{t-1})$, and c) $T_{Y\rightarrow X}(1, 1)$

Let $p(x_1, \ldots x_n)$ denote the probability of observing the subsequence $(x_1, \ldots x_n)$, Schreiber (2000) defines transfer entropy as

$$T_{Y\rightarrow X}(m, l) = \sum p(x_{t_1}, \ldots x_{t_m}, y_{t_{m-l+1}}, \ldots, y_{t_m})$$

$$\cdot \log_2 \frac{p(x_{t_{m+1}}|x_{t_1}, \ldots x_{t_m}, y_{t_{m-l+1}}, \ldots, y_{t_m})}{p(x_{t_{m+1}}|x_{t_1}, \ldots x_{t_m})}$$  \hspace{1cm} (2.14)

where $x_t$ and $y_t$ represent the discrete states of $X$ and $Y$ at time $t$. The parameters $m$ and $l$ indicate the number of past observations included in $X$ and $Y$ respectively (see Figure 2.2 (c)). In the absence of an information flow from $Y$ to $X$, i.e. $X$ and $Y$ are independent, then

$$p(x_{t_{m+1}}|x_{t_1}, \ldots x_{t_m}, y_{t_{m-l+1}}, \ldots, y_{t_m}) = p(x_{t_{m+1}}|x_{t_1}, \ldots x_{t_m})$$

and $T_{Y\rightarrow X}(m, l) = 0$. Transfer entropy in (2.14) can also be expressed as conditional
entropies of the different blocks of history.

\[ T_{Y \rightarrow X}(m, l) = H(x_{t_{m+1}|x_{t_m}, \ldots, x_{t_1}}) - H(x_{t_{m+1}|x_{t_m}, \ldots, x_{t_1}, y_{t_m}, \ldots, y_{t_{m-l+1}}}) \]

\[ = (H(x_{t_{m+1}|x_{t_m}, \ldots, x_{t_1}}) - H(x_{t_m, \ldots, x_{t_1}})) \]

\[ - (H(x_{t_{m+1}|x_{t_m}, \ldots, x_{t_1}, y_{t_m}, \ldots, y_{t_{m-l+1}}}) - H(x_{t_m, \ldots, x_{t_1}, y_{t_m}, \ldots, y_{t_{m-l+1}}})) \]

(2.15)

In this way, transfer entropy can be understood as difference in the gain of information about \( x_{t_{m+1}} \) conditional on both its own history and the history of \( y \), and that conditional on its own history only. If the history of \( X \) does not inform us of \( x_{t_{m+1}} \), but the history of \( Y \) completely determines \( x_{t_{m+1}} \), then \( T_{Y \rightarrow X}(m, l) = H(x_{t_{m+1}|x_{t_m}, \ldots, x_{t_1}}) = H(x_{t_{m+1}}) \). In the absence of information flow from \( Y \) to \( X \), and the difference between the third and fourth terms in (2.15) is zero. Hence we have \( 0 \leq T_{Y \rightarrow X}(m, l) \leq H(X) \). While mutual information quantifies the deviation from independent \( X \) and \( Y \), transfer entropy quantifies the deviation from \( X \) being determined by its own history only (via conditional probabilities). Unlike mutual information, transfer entropy \( T_{Y \rightarrow X}(m, l) \) is not symmetric and takes into account only the statistical dependencies originating from the variable \( Y \) and not those from a common signal.

Another way to look at transfer entropy it is to understand it as the resolution of uncertainty. This is similar to the interpretation of Granger causality in terms of prediction. The transfer entropy from \( Y \) to \( X \) is the degree to which \( Y \) makes less uncertain the future of \( X \) beyond the degree that \( X \) is already giving information on resolving the future of \( X \). One can actually show that Granger causality for Gaussian variables is equivalent to transfer entropy (Barnett, Barrett, and Seth (2009)).

### 2.3.2 Entropy Estimation, Statistical Features

There is a multitude of methods for estimating entropy measures. One technique uses the kernel density method. In particular, transfer entropy and mutual informa-
tion have both been estimated in the literature based on the kernel density method (see Schreiber (2000); and Blumentritt and Schmid (2011)). The maximum likelihood method (Paninski (2003)) has also been applied successfully. In the following, we will focus on estimation methods which assume a discrete co-domain of the random variables, or the domains have been 'discretized'.

The discretization of the values in a given time series (e.g. return series, drawdown series) is a mapping of the continuous values of a random variable $X$ to a discrete set by partitioning the support of $X$. These discretized values are often referred to as 'letters' or symbols. Entropy is then approximated by the finite sum:

$$H(X) \approx \hat{H}_{\text{binned}}(X) = - \sum_i \hat{p}(X \in \Delta_i) \cdot \log_2(\hat{p}(X \in \Delta_i))$$

where $\hat{p}(X \in \Delta_i)$ denotes the estimated probability of the random variable having values in $\Delta_i$. As the partitions get finer with smaller bin sizes, $\hat{H}_{\text{binned}}(X)$ converges to $H(X)$ for well behaved distributions (see Grassberger (2003) for a discussion).

In the literature\(^1\) different ways of partitioning discrete data have been introduced. In some instances, the partitions are formed by dividing the distribution into equal parts, producing an equal marginal probability for every symbol defined. Marschinski and Kantz (2002) chose such a partition scheme to avoid 'undesirable' effects due to very inhomogenous histograms. As an example of an equal marginal probability discretization, consider a return series $\{r_1, \ldots, r_n\}$ with quantiles $q(r_t, c)$ for $c \in (0, 1)$. The following maps all returns into three symbols depending on whether the returns are below, between or above the 33% and 66% quantiles, forming an equiprobable binning scheme:

\[
d_r(r_t) = \begin{cases} 
0 & \text{if } r_t < q(r_t, 0.33) \\
1 & \text{if } q(r_t, 0.33) \leq r_t \leq q(r_t, 0.66) \\
2 & \text{if } r_t > q(r_t, 0.66)
\end{cases}
\] (2.16)

\(^1\)See for example Hlavackova-Schindler, Palus, and Vejmelka (2007).
Other authors\(^2\) prefer unequal marginal probabilities partition scheme. In our studies, equal marginal probability is not guaranteed and will depend on the distribution of the draw size.

Let \(X \rightarrow A\) be the discretized random variable where all values are binned into \(M = |A|\) boxes. The probability \(p_i\) of \(X\) falling into bin \(i\), \(1 \leq i \leq M\), would be estimated by counting the number of times \(X\) is in bin \(i\), \(n_i = |\{X \in \Delta_i\}|\), divided by the sample size \(N\), so \(p_i = \frac{n_i}{N}\). The estimator, which we will call the naïve entropy estimator, can be calculated as follows.

\[
\hat{H}_{\text{naive}} = - \sum_{i=1}^{M} p_i \cdot \log_2(p_i) = - \sum_{i=1}^{M} \frac{n_i}{N} \cdot \log_2\left(\frac{n_i}{N}\right)
\]

\[
= \log_2(N) - \frac{1}{N} \sum_{i} n_i \cdot \log_2(n_i) \tag{2.17}
\]

For small samples, the binned distribution is likely to be less uniform and, hence, has a lower and downward-biased entropy \(E[\hat{H}_{\text{naive}}] - H < 0\). In figure 2.3, the average naïve entropy estimate for a fair coin (\(H = 1\) bit) is shown for different samples sizes. In a Monte Carlo simulation, a fair coin is tossed over sequences with lengths ranging from 50 to 1000, each sampled 5000 times. For each of the simulated samples of a specific size, the entropy is estimated with the naïve estimator in equation (2.17), and the mean of the estimate \(E[\hat{H}_{\text{naive}}]\) is calculated. \(\hat{H}_{\text{naive}}\) turns out to be below 1 for all sample sizes, which illustrates the downward bias of the naïve estimator. The effect is more pronounced for small sample sizes (left diagram of figure 2.3). Grassberger (2003) offers a correction to this ‘small sample’ bias. Assuming that all \(p_i \ll 1\), the proposed new estimator \(\hat{H}_\psi\) is as follows:

\[
\hat{H}_\psi = \ln N - \frac{1}{N} \sum_{i=1}^{M} n_i \cdot \psi(n_i) \tag{2.18}
\]

\[
\psi(x) = \frac{d(ln\Gamma(x))}{dx}, \quad \Gamma(0, x) = \int_1^\infty \frac{e^{-xt}}{t} \, dt \tag{2.19}
\]

We will use this estimator \(\hat{H}_\psi\), in particular when estimating transfer entropy.

---

\(^{2}\)See for instance Jizba, Kleinert, and Shefaat (2011) and Peter, Dimpfl, and Huergo (2010).
for larger blocks of history. Transfer entropy, as we have seen in equation (2.15), is calculated with block entropies. The block time series is a series of symbol set representing the combinations of letters in each cell of the block. Since the blocks overlapped in the original time series, the block time series has the same size as the original sequence minus the block length. For alphabets of two to three letters that we use here, the assumption $p_i \ll 1$ will hold when estimating block entropies, and hence the Grassberger estimator can be used.

Statistical fluctuations in samples, and in particular small samples, will artificially induce deviations to the entropy estimates. The magnitude of the ‘noise’ depends on the sample size and the size of the symbol set $M$, or in the case of more than one discrete random variable, the size of the co-domain. For some distributions, there exists analytical expressions for difference between the expected value of $\hat{H}$ and the actual value $H$ (Schürmann (2004)). At this point, we would like to highlight the variance of the naïve estimator with the coin tossing example we used above. For the Monte Carlo simulations of a specific length, the standard error $\hat{se}(\hat{H}_{\text{naive}})$ of the entropy estimates is calculated. The standard errors are plotted in figure 2.3. Also plotted in figure 2.3 are the standard errors of entropy estimates of a uniform distributed random variable with three and four symbols (analogous to a coin with three or four sides). The figure shows that the standard error is slightly higher for a given sample size when the symbol set is larger. The difference is greater the smaller the sample size.

For a given size of alphabet, it seems logical to choose the largest possible block length in order to find invariant values and to detect the patterns in history that offer the most information in forecasting. But for block entropies, the length of the block is limited by the sample size (and the size of the symbol set). The larger the block length, the larger the noise. As block length approaches the sample size, the estimated block entropy drops to zero as the symbol combinations concentrate in fewer and fewer symbols.

As a way to control the noise in the transfer entropy estimates, an approach pioneered by Marschinski and Kantz (2002) is to estimate the noise itself by a boot-
strapping method. For two processes $X$ and $Y$, the information flow from $Y$ to $X$ is measured by Transfer Entropy $T_{Y \rightarrow X}$. For the information source $Y$, we form a second set by shuffling the original time series. If $\{Y_1, \ldots, Y_n\}$ is the original sample and $\pi$ is a permutation of the tuple $(1, \ldots, n)$, then a shuffled series is $\{Y_{\pi(1)}, \ldots, Y_{\pi(n)}\}$. Through shuffling the information source, all potential correlations between the two time series are destroyed and hence the observed transfer entropy should be zero. Repeating the shuffling process many times, the mean $\mu(\hat{T}_{sh})$ and the standard error $\sigma(\hat{T}_{sh})$ of the transfer entropy is estimated. Since there is no structure in the reshuffled data, any non-zero estimate must be an artefact of the finite sample size.

To account for the ‘noise’ in the Transfer Entropy calculation, the Effective Transfer Entropy (ET) is calculated by deducting the mean $\mu(\hat{T}_{sh})$ from the Transfer Entropy estimate as follows:

$$ET_{Y \rightarrow X}(m, l) \equiv T_{Y \rightarrow X}(m, l) - \mu(\hat{T}_{sh}) \quad (2.20)$$

For $ET_{Y \rightarrow X}(m, l)$, $m$ amount of past history of $X$ and $l$ amount of past history of $Y$ are used in predicting $X$. 

Figure 2.3: Entropy Estimation
The graphs show the entropy estimate and standard error in relation to sample size between 50-1000 data points. The graph on the left shows standard error $\hat{se}$ for uniform distribution with 2, 3, 4 symbols. On the right, the entropy estimate $\hat{H}_{naive}$ for a fair coin (uniform distribution with two symbols) is shown.
2.3.3 Process Entropy

For a sequence of identical distributed and independent random variables, the entropy of one variable conditional on the others has a particular simple form. We have seen for two independent variables $X$ and $Y$ that the conditional entropy is $H(X|Y) = H(X)$. Generalizing this to the case of $n$ iid variables gives $H(X_0|X_1,\ldots X_n) = H(X_0)$. The block entropy in this case is $H(X_0, X_1,\ldots X_n) = n \cdot H(X_i)$ with $0 \leq i \leq n$.

For non-independent random variables, which is often the case in empirical studies, we would like to have some understanding of how both conditional entropy and block entropy behave over different block lengths. Given a process $\{X_i\}$, the limit of the conditional entropy

$$h_\infty = \lim_{n \to \infty} H(X_0|X_{-1},\ldots X_{-(n+1)})$$

(2.21)

is, in the literature, referred to as **entropy rate** or **entropy of the source** (Cover and Thomas (2006)). It quantifies the average information needed to predict a future observation $X_0$ given an $m$ period of history ($X_{-1},\ldots X_{-(m+1)}$). In the univariate case, it quantifies the level of predictability of the series conditional on a specific pattern in the history. Similarly, the transfer entropy $T_{Y \rightarrow X}$ can be expressed as the entropy of $X$ conditional on its own history and on the history of $Y$. See figure 2.2 for a depiction of process entropy and its relation to other entropy measures. We will discuss here some special cases that are relevant in the subsequent chapters.

If a process is periodic, then $h_\infty = 0$ for block length longer than the periodicity, and the conditional entropy goes to 0. For any iid process $\{X_i\}$, $h_\infty = H(X)$, with $H(X)$ being the entropy of the symbol distribution. This means that there is no information gain no matter how much history is included in the calculation. The entropy rate for any given process lies between these two extremes: $0 \leq h_\infty \leq H(X)$.

For Markov processes, there is an analytic expression for $h_\infty$. Let $X$ be a Markov

---

3A process is periodic if there exists a $k \in \mathbb{N}$, the periodicity for which $X_t = X_{t+k}$ for all $t$. 

31
process with transition matrix $P = (P_{ij})$. The stationary distribution $\pi$ of $X$, is the limiting distribution (if it exists) to which the Markov process $X$ converges. It is the solution of the eigenvalue problem $\pi = P \cdot \pi$ and has the form $\pi_j = \sum_i \pi_i \cdot P_{ij}$ ($\pi_i$ is the probability of $X$ being in state $i$). The stationary distribution is so called because if the initial state of a Markov chain is drawn according to a stationary distribution, the Markov chain forms a stationary process. One can show that if the Markov chain is irreducible and aperiodic, the stationary distribution is unique, and from any starting distribution, the distribution of $X_n$ tends to the stationary distribution as $n \to \infty$.

For a stationary distribution of a Markov process, the entropy rate can be simplified as

$$h_\infty = \lim_{n \to \infty} H(X_n|X_{n-1}, \ldots, X_1)$$

$$= \lim_{n \to \infty} H(X_n|X_{n-1}) = H(X_2|X_1)$$

(2.22)

That is, only the immediate history $X_{n-1}$ is needed for predicting $X_n$. Combining this with the expression for the stationary distribution, the entropy rate of a stationary markov process with stationary distribution $\pi_i$ and transition matrix $P_{ij}$ can be simplified as:

$$h_\infty = \sum_i \pi_i \cdot \left( \sum_j -P_{ij} \cdot \log_2 P_{ij} \right)$$

(2.23)

### 2.3.4 Entropy Magnitude

To measure the amount of information gain due to the transfer entropy, we follow Marschinski and Kantz (2002) by relating the information gain to the entropy of the series based on the $REA$ (relative explanation added) measure defined as follows:

$$REA(m,l) \equiv \frac{T_{Y \to X}(m,l)}{H(X_0, X_1, \ldots, X_{-(m)})}$$

(2.24)

Here the information flow from $Y$ to $X$, measured by the transfer entropy, is related to the total flow of information based the conditional block entropy of $X$. Put
differently, it measures the additional information gain by observing the history of $X$ and $Y$ when already observing $X$. Using the definition of transfer entropy, we arrive at the following form

$$REA(m, l) = \frac{H(X_0|X_1, \ldots, X_m, Y_1, \ldots, Y_l)}{H(X_0|X_1, \ldots, X_m)} - 1 \quad (2.25)$$

If the past history of $Y$ is not adding information to predicting $X$, then $H(X_0|X_1, \ldots, X_m, Y_1, \ldots, Y_l) = H(X_0|X_1, \ldots, X_m)$ and the $REA$ is zero.

To give a sense of the levels expected for $REA$, we cite directly the results of Marschinski and Kantz (2002). The $REA$ for using DAX to predict the Dow Jones Industrial, using one minute tick data, is estimated to be 0.4% for a three symbol discretization.

### 2.4 Approximate Entropy (ApEn)

Various information measures have been introduced to complement our understanding of stochastic time series. In a series of articles,$^4$ Pincus (1991) developed a new mathematical approach to measuring regularity in a time series based on the Approximate Entropy.

#### 2.4.1 ApEn Definition

Let $u$ be a sequence of numbers, $u(1), \ldots, u(N)$, of length $N$. Given a non-negative number $m$ with $m \leq N$, we form $m$-blocks of subsequences $x(i) \equiv (u(i), u(i + 1), \ldots, u(i + m - 1))$. The distance between two blocks is measured by $d(x(i), x(j)) \equiv \max_{k=1,2,\ldots,m}(|u(i + k - 1) - u(j + k - 1)|)$, so it is the maximum of the point-wise difference between the blocks. Given a positive real number $r$, we count for a given block $x(i)$ the fraction of blocks $x(j)$ that have a distance of less than $r$ and name it $C_i^m(r)$. The formal definition is shown in equation (2.26), which is based on

the Heaviside function \( \Theta \) counting the instances where the distance \( d \) is below the threshold \( r \):

\[
C_i^m(r) = \frac{1}{N - m + 1} \sum_{j=1}^{N-m+1} \Theta(r - d(x(i), x(j))) \quad (2.26)
\]

\[
\Phi^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \log C_i^m(r) \quad (2.27)
\]

With \( \Phi_m(r) \) defined in equation (2.27), the approximate entropy, \( \text{ApEn}(m, r, N) \), is defined as:

\[
\text{ApEn}(m, r, N)(u) = \Phi^m(r) - \Phi^{m+1}(r) \quad , \quad m \geq 0 \quad (2.28)
\]

\[
\text{ApEn}(0, r, N)(u) = -\Phi^1(r) \quad (2.29)
\]

\( \text{ApEn}(m, r, N) \) is always well defined. In equation (2.26), when a fraction of block is \( < r \), the distance of the block with itself, \( d(x_i, x_i) \), is always counted. So \( C_i^m(r) \geq 1 \), and hence the logarithmic of \( C_i^m(r) \) is never undefined. The following expansion shows that \( \text{ApEn} \) measures how close specific patterns are.

\[
-\text{ApEn}(m, r, N)(u) = \Phi^{m+1}(r) - \Phi^m(r)
\]

\[
= \frac{1}{N - m} \sum_{i=1}^{N-m} \log C_i^{m+1}(r) - \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \log C_i^m(r)
\]

\[
\approx \frac{1}{N - m} \sum_{i=1}^{N-m} \left( \log C_i^{m+1}(r) - \log C_i^m(r) \right)
\]

\[
= \frac{1}{N - m} \sum_{i=1}^{N-m} \log \left( \frac{C_i^{m+1}(r)}{C_i^m(r)} \right) \quad (2.30)
\]

The last line is the average, over all \( m \)-blocks, of the log of the conditional probability of \( |u(j+m) - u(i+m)| < r \), given that \( |u(j+k)-u(i+k)| < r \) for all \( k = 0, 1, \ldots m-1 \). So, approximate entropy measures the logarithmic likelihood that sequences of patterns that are close for \( m \) observations remain close on next comparisons.

\( \text{ApEn} \) measures persistence, correlation and regularity. If \( C_i^{m+1}(r) \) equal \( C_i^m(r) \), then \( \text{ApEn}=0 \) and there is strong serial dependence. Hence, lower \( \text{ApEn} \) values correspond to higher persistence and autocorrelation. Higher \( \text{ApEn} \) values imply
less persistence and more independent observations.

Figure 2.4: ApEn for AR(1) process.
Approximate Entropy ApEn(N = 200, m = 1) for AR(1) process with various autoregressive coefficients $\rho = 0, \ldots, 0.9$.

We can derive another useful representation of approximate entropy. We have shown, in equation (2.30), that the approximate entropy for $m = 1$, ApEn($m = 1, r, N$), can be reformulated as expected log-likelihood of two blocks staying close when extended by one element. Assuming the stationarity of the stochastic process, we can reduce the number of calculations in ApEn($m = 1, r, N$), as follows:

$$ApEn(1, r, N)(u) = \Phi^1 - \Phi^2$$

$$= \frac{1}{N - 1} \sum_{i=1}^{N-1} \log \left( \frac{C_2^i(r)}{C_1^i(r)} \right)$$

Furthermore, Pincus (1991) found analytical expression for ApEn for a number of processes. For $iid$ distributed random variables, ApEn($m, r$) is given by the following theorem.

**Theorem** For an $iid$ process with density function $f(x)$, for any $m$,

$$ApEn(m, r) = - \int f(y) \cdot \log \left( \int_{z=y-r}^{z=y+r} f(z) \, dz \right) \, dy$$

35
For example, consider the autoregressive process of order one, AR(1), below:

$$X_k = \rho \cdot X_{k-1} + \epsilon_k \quad k = 0, 1, 2 \ldots \quad (2.35)$$

For different values of $\rho = 0, 0.1, \ldots 0.9$, a Monte Carlo simulation of 1000 paths of length 200 are drawn. The resulting approximate entropy estimation with standard error is depicted in figure 2.4. For higher correlation values, $\rho$, the approximate entropy is lower. The highest ApEn value is for the random walk process with $\rho = 0$. For $\rho = 0.9$, ApEn declined to 0.65 [bit]. This shows that persistence is a feature that the approximate entropy measure is sensitive to.

### 2.4.2 ApEn Estimation

Rukhin (2000) shows that ApEn converges asymptotically to a chi-squared distribution for $N \to \infty$. Specifically for an iid binary-valued sequence of size $N$, a block size $m$ and a constant $c = \frac{3}{\ln 2}$, ApEn($m, N$) converges to a $\chi^2$ distribution with $2^{m-2}$ degrees of freedom as $N \to \infty$. For finite sample, ApEn is a biased statistic. The small sample bias can be illustrated with a normal distributed random variable, $N(0, \sigma)$, for sample length ranging from 20 to 500, each with 1000 simulated paths. In the right of figure 2.5, the mean and standard error of the ApEn estimates for this random walk are presented. The threshold $r$ is set equal to 20% of the sample standard deviation. The mean value of ApEn increases from $\overline{\text{ApEn}}(m = 2, r = 0.2\hat{\sigma}, N = 20) = 0.11 \text{ [bit]}$ to $\overline{\text{ApEn}}(m = 2, r = 0.2\hat{\sigma}, N = 500) = 1.44 \text{ [bit]}$ as the sample length, $N$, increases.

For small samples, the underestimation of ApEn is visible. We can see from figure 2.5 that as sample length approaches 1, ApEn approaches 0. In view of this small sample bias, one may increase the threshold $r$ as a counter measure. But there is a tradeoff, with a larger $r$ information is lost from the distribution. The small sample bias also implies that comparisons between ApEn make sense only if the sample length $N$ is the same. Furthermore, Pincus (2008) studied ApEn($m, r, N$) for various equity indices and concluded that the parameters $m = 1, 2$ and the filter threshold
Figure 2.5: ApEn with $m = 1, \ldots, 5$ and $N = 20, \ldots, 500$.
Left: rolling 90-day window ApEn($m = 1, \ldots, 5$, $r = 0.2\hat{\sigma}$)(SP500) time series. Right: graph of estimated ApEn of a random walk with different sample sizes and standard error.

$0.1\hat{\sigma} \leq r \leq 0.25\hat{\sigma}$, with $\hat{\sigma}$ being the standard deviation of the sample, produce good results.

To construct statistical bounds, the bootstrapping method is employed in chapter 5. The bootstrapping method normally involves reshuffling the data several times, and the approximate entropy estimate averaged over the bootstrapped simulations is used to estimate the bias and calculate the effective ApEn. However, reshuffling the data destroys the dependencies within the univariate time series as well. The more appropriate method is a standard block bootstrap, where block resampling with a fixed block length is used. This is the approach adopted here. In estimating ApEn for a random walk in figure 2.5, the bootstrapped standard error is shown as grey band around the ApEn estimate in the figure. The smaller sample lengths are associated with bigger standard errors.
Chapter 3

Bivariate Analysis of Drawdowns

Abstract

In this chapter, we use mutual information and transfer entropy to analyse the dependence relation between large drawdowns/drawups in EUR/USD and GBP/USD. Despite various crisis episodes documented in our sample period, we did not find evidence of extreme dependence in this currency pair. Large drawdowns/drawups in EUR/USD and GBP/USD are weakly dependent. Transfer entropy shows information flows between EUR/USD and GBP/USD however.
3.1 Introduction

Financial time series drawdowns (drawups) are defined as the cumulative return from a local maximum (minimum) to the next local minimum (maximum). Draws, which include both drawdowns and drawups, are a time series feature that has no regular time stamp. It provides aspect of statistical dependence in different market regimes. The concept of draws is important, not only from a theoretical point of view but also for its use in various contexts in the finance industry, e.g. in asset management where drawdowns tend to drive redemptions. For buy and hold portfolios that are subject to mark to market limits, extended drawdowns can force unwanted liquidations and the readjustments of portfolios. Furthermore, large drawdowns could trigger feedback mechanisms through portfolio insurance and fire sales at funds, which have the tendency to drive prices further down. The study of drawdowns has drawn considerable attention, and many empirical properties of drawdowns in financial time series have been documented.

Johansen and Sornette (2001) lay out a crash theory that explains drawdowns/drawups that appears to fit almost all time series. They claim that drawdowns/drawups can be explained by the onset of a well-defined, micro-structural transition phase, which is characterized by the emergence of the sudden persistence of daily drops, coupled with an increase in the correlated amplitude of the drops. This explains why price changes appear to be independent for most of the time, but in exceptional circumstances, serial co-dependence and amplification may set in, creating drawdown/drawup outliers. To our knowledge, there has been no study conducted on the relationship of drawdowns (drawups) in financial time series. Rebonato and Gaspari (2006) examined qualitatively, on a case-by-case basis, if large drawdowns/drawups from different maturity blocks of US interest rates coincide with each other. But no quantitative measure has been proposed so far to analyse these drawdowns/drawups.

Detecting and measuring interactions between financial time series is a key area of research in finance and economics. With regard to linear dependence, there are,
for instance, correlation and cross-correlation measures and their more general formulations in vector autoregressive models.\textsuperscript{1} More recently, copula methods have been used, with some degree of success, to model the more general form of dependence between time series (McNeil, Frey, and Embrechts (2005)). But these tools are not suited for modelling the dependence of drawdowns/drawups. One obstacle, arising when moving from returns to draws, is that time is compressed and is no longer equally spaced. Instead of having time series of synchronous measurements, the draws resulting from two time series will in general be clocked very differently. A further obstacle is that it is often argued that large drawdowns/drawups are generated from nonlinear mechanisms. The herding effects of market participants under stressed market conditions, positive feedback from portfolio insurance and liquidity squeeze in times of market stress can all lead to non-linear price dynamics (Sornette (2003)).

As an alternative to the standard models, information theoretic measures have been applied successfully to finance and other areas to cope with the aforementioned problems.\textsuperscript{2} Transfer entropy, in particular, measures information flows between series and detects the subtle dependencies between them. It is a model-free measure, which is based on the Kullback-Leibler distance (Schreiber (2000)) of transition probabilities. Mutual information, on the other hand, quantifies correlation and cross-correlation type dependencies of time series. It has also been used in finance extensively (Golan (2008), Dionisio, Menezes, and Mendes (2004)). In this chapter, we use information theoretic tools to analyse how large drawdowns and large drawups in EUR/USD and GBP/USD are related to each other. Specifically, the mutual information measure is used to analyse the contemporaneous dependence (i.e. same hour or same day) or for a specific time lag. Transfer entropy is used to measure how large drawdowns and drawups in one time series influence large drawdowns and drawups in another time series. Our results show that the relationship between large draws of EUR/USD and GBP/USD is weaker than the dependence

\textsuperscript{1}For more details see Tsay (2010).
between the returns of the same series. The strength of dependence drops markedly when time lags between the series are considered. On the other hand, transfer entropy suggests an information flow between the large draws in the two series. Such information spillover is slightly stronger from GBP/USD to EUR/USD than in the reverse direction.

This chapter is structured as follows. Section 3.2 reviews the relevant literature pertaining to the analysis. It covers the major research on the drawdown/drawup and some recent finance applications of entropy measure and information theoretic tools. Section 3.3 describes the data used and discusses its statistical properties. Various information theoretic tools used in this study, which have been introduced in Chapter 2, are briefly reviewed in section 3.4. The method used to translate draws into sequences of symbols is also explained in this section. The results from the dependence analysis are presented and discussed in Section 3.5. We summarise and conclude in Section 3.6 and suggest possible directions for future research.

3.2 Related Literature

3.2.1 Drawdowns Literature

In the microstructure model proposed by Johansen, Ledoit, and Sornette (2000), agents respond differently during normal and crisis market conditions, producing different patterns of imitation, herding and positive feedback during market rallies and crashes, leaving behind different signatures of ‘auto-correlation’ and dependencies. The authors claim that outliers in drawdowns commence with the emergence of sudden and persistent daily price drops, coupled with an increase in the correlated amplitude of the drops. So price changes appear to be independent for most of the time and under ‘normal’ market conditions, but serial co-dependence and amplification may set in in exceptional circumstances, creating the drawdown outliers.

Given a time series, a drawdown is defined as a consecutive series of negative returns. A drawup is accordingly defined as an uninterrupted sequence of positive
returns. Specifically, a drawdown (drawup) is the cumulative return from one local maximum (minimum) to the next local minimum (maximum). For the remainder of this chapter, unless otherwise stated, we use draws to refer to both drawdowns and drawups in general.

Draws play a significant role in Sornette’s theory in capturing market dependence under different market regimes, as laid out by Johansen, Ledoit, and Sornette (2000) and Johansen and Sornette (2001). Following Johansen and Sornette (2001), and starting with drawdowns of a specific length, the probability of observing a drawdown of magnitude $D$ having lasted exactly $n$ units of time is proportional to

$$p_n(D) \propto \int_{-\infty}^{0} dx_1 p(x_1) \cdots \int_{-\infty}^{0} dx_n p(x_n) \delta \left( D - \sum_{i=1}^{n} x_i \right)$$

(3.1)

where $\delta$ is the delta function,\(^1\) $\delta \left( D - \sum_{i=1}^{n} x_i \right)$ on the right hand side of equation (3.1) ensures that the sum is over all possible run durations $n \in \mathbb{N} - \{0\}$.

As drawdowns can, in principle, last any number of periods, we need to sum over all possible durations to derive the probability of a drawdown of magnitude $D$ unconditional on $n$ as follows:

$$p(D) = \frac{p_u}{p_d} \left[ \sum_{n=1}^{\infty} \int_{-\infty}^{0} dx_1 p(x_1) \cdots \int_{-\infty}^{0} dx_n p(x_n) \delta \left( D - \sum_{i=1}^{n} x_i \right) \right]$$

(3.2)

where $p_d = \int_{-\infty}^{0} p(x)dx$ is the probability of observing a negative price change, $p_u = 1 - p_d$ is the positive counterpart and the term $\frac{p_u}{p_d}$ ensures the normalisation of the distribution.

Through asymptotic analysis and assuming $i.i.d$ returns, the analytical expression for the drawdown probability density function $p(D)$ has an exponential distribution (Johansen and Sornette (2001)):

$$p(D) = \frac{1}{D_0} \exp\left( -\frac{|D|}{D_0} \right); \quad D_0 = -\frac{1}{p_u p_d} \int_{-\infty}^{0} x p(x)dx$$

(3.3)

\(^1\) $\delta$ is a function with $\int_{-\infty}^{+\infty} \delta(x-a)f(x)dx = f(a)$
\( D_0 \) has a natural interpretation as the characteristic size of a drawdown. Sornette et al. extend the exponential distribution to the stretched exponential distribution in equation (3.4) below to account for the slower decaying tails using extra terms in the Taylor series expansion in the asymptotic analysis:

\[
P_{se}(D) = D_0 e^{-\left(\frac{|D|}{\chi}\right)^z}
\] (3.4)

Equation (3.4) can fit distributions with smaller \((z > 1)\) or larger \((z < 1)\) kurtosis than the simple exponential distribution.

The parameters \( z \) and \( \chi \) characterize the distribution concisely: the larger the value of \( \chi \), the larger the 'typical' drawdowns; the smaller the value of \( z \), the fatter the tails and the greater the relative likelihood of large drawdowns. Rebonato and Gaspari (2006) show, for identical and independent Gaussian returns, the expected drawdown magnitude is \( E[D]_{Gauss} = \frac{4\sigma}{\sqrt{2\pi}} \), and the expected daily drop \( d \) within a draw, \( E[d]_{Gauss} = \frac{2\sigma}{\sqrt{2\pi}} \).

A concept closely related to drawdowns and drawups is that of a run, defined as a sequence of positive (or negative) price changes. Results on the length of a run (which is independent of the duration of a draw) have been obtained by Mood (1940) and others. We denote the length of a run or drawdown by \( l_d \). The probability that an arbitrarily chosen drawdown consists of exactly \( l_d = n \) price moves can be calculated as the probability, given an initial down move, of further \( n-1 \) moves and a concluding up move as follows:

\[
P(l_d = n) = p_u \cdot p_d^{n-1}
\] (3.5)

where \( p_u \) the probability of an up move and \( p_d \) is the probability of a down move, with \( p_u + p_d = 1 \), in a return series \( \{x_0, \ldots, x_n\} \).

Applying the more general results from Mood (1940), one can calculate the expected number of runs \( E[N_{run}] \) in a time series of \( n \) price changes. This is given
as
\[ E[N_{\text{run}}] = 2np_u(1 - p_u) + p_u^2 + (1 - p_u)^2 \]  
(3.6)

For the case of a fair coin, \( p_u = p_d = \frac{1}{2} \), the expected number of runs is therefore given as \( E[N_{\text{run}}] = 2n\frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{n+1}{2} \). For large \( n \) this is approximately \( \frac{n}{2} \).

For the independent tossing of a fair coin with \( p_u = p_d = \frac{1}{2} \), the expected length and variance of a draw are
\[
\begin{align*}
E[l_d]_{p_u=p_d=1/2} &= 2 \\
\text{var}[l_d]_{p_u=p_d=1/2} &= 2
\end{align*}
\]  
(3.7)  
(3.8)

Having reviewed the fundamental theoretical results of drawdowns, we review the empirical findings on financial market drawdowns as documented in a series of papers by Johansen, Ledoit and Sornette (JLS)\(^3\) on the daily changes of Dow Jones Industrial Average, FTSE 100, some single-name stocks in different markets, foreign exchange US$/DM, US$/Yen, US$/CHF and Gold price. JLS found the modified exponential distribution introduced in equation (3.4) is appropriate for up to 98% of drawdowns and drawups in equity and most financial returns. The remaining 1% to 2% of the largest drawdowns are not explained by the simple exponential or by the stretched exponential. Drawdowns up to three times larger than expected are prevalent in equity indices, stock prices, FX rates and commodity prices, with the CAC40 being among the very few exceptions. JLS suggest that these draws are outliers which belong to a different class. Only half of the time series studied exhibit outliers in the drawups. Furthermore, in contrast to drawdowns drawups have a \( \hat{z} \) (see Table 3.1) closer to the simple exponential case. The characteristic size of draws, \( \chi \), is stable within an asset class but varies across asset classes. FX, as a group, has the smallest \( \chi \), and has \( z \) values \( \approx 0.84 \) to 0.91 (see Table 3.1), which corresponds to the fat tail distribution. For equity indices, the typical price movement has \( \chi \approx 1.05\% \) to 2.1% for drawdowns and drawups, which is larger than that for FX.

\(^3\)See, for example, Sornette (2003), Johansen and Sornette (2001) and Johansen and Sornette (2000).
The $z$ coefficient for drawdowns is in the same range as for currencies i.e. 0.80 to 0.90, but for drawups, it is close to 1, i.e. the simple exponential distribution. For individual stocks, the characteristic size is much larger, with $\chi \approx 1.91\%$ to 7.61%. The exponents $z$ are close to 1 for both drawups and drawdowns. The $z < 1$ exponent of the equity indices, gold and FX rates, is compatible with a daily return distribution with fat tails. Diverging from this is the distribution of drawdowns of large US stocks, with $z \approx 1$, compatible with a daily return distribution with normal tail size.

<table>
<thead>
<tr>
<th></th>
<th>Drawdowns</th>
<th>Drawups</th>
<th>US Interest Rate</th>
<th>Drawdowns</th>
<th>Drawups</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FX and US Equity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FX Japanese Yen/US Dollar</td>
<td>0.90 ± 0.02</td>
<td>0.89 ± 0.02</td>
<td>3m 1.006 ± 0.034</td>
<td>0.999 ± 0.032</td>
<td></td>
</tr>
<tr>
<td>SP 500</td>
<td>0.90 ± 0.01</td>
<td>1.03 ± 0.02</td>
<td>6m 1.058 ± 0.032</td>
<td>1.053 ± 0.031</td>
<td></td>
</tr>
<tr>
<td>NASDAQ Composite</td>
<td>0.80 ± 0.02</td>
<td>0.90 ± 0.02</td>
<td>1y 1.188 ± 0.035</td>
<td>1.211 ± 0.030</td>
<td></td>
</tr>
<tr>
<td>Dow Jones</td>
<td>0.84 ± 0.01</td>
<td>0.99 ± 0.01</td>
<td>2y 1.053 ± 0.023</td>
<td>0.987 ± 0.022</td>
<td></td>
</tr>
<tr>
<td>Microsoft</td>
<td>1.04 ± 0.03</td>
<td>1.01 ± 0.03</td>
<td>5y 1.127 ± 0.025</td>
<td>1.050 ± 0.024</td>
<td></td>
</tr>
<tr>
<td>Cisco</td>
<td>1.16 ± 0.04</td>
<td>1.21 ± 0.04</td>
<td>10y 1.154 ± 0.025</td>
<td>1.052 ± 0.023</td>
<td></td>
</tr>
<tr>
<td>General Electric</td>
<td>1.02 ± 0.02</td>
<td>1.02 ± 0.02</td>
<td>20y 1.188 ± 0.025</td>
<td>1.060 ± 0.024</td>
<td></td>
</tr>
<tr>
<td>Intel</td>
<td>1.06 ± 0.03</td>
<td>1.21 ± 0.03</td>
<td>30y 1.144 ± 0.026</td>
<td>1.063 ± 0.023</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1: Draws in FX, Equity and Interest Rate markets

Reported fit of the exponential distribution $\hat{z}$ (± standard error) for drawdowns and drawups on FX, selected US equities and interest rates daily returns from Johansen and Sornette (2001).

JLS argue that the anomalously large drawdowns can only be explained by rare and successive daily drops with correlated magnitude. These occur when the market enters a critical stage, where a phase transition leads to a different behaviour among rational, informed traders and noisy, imitative traders. Herding behaviour and market participants imitating the majority as an optimal strategy are cited in Sornette (2003) as examples of different traits of the market participants. It is the change of mode of interaction between market participants that triggers a regime change in price moves, producing different serial co-dependence and magnitudes. Other industry practices, such as portfolio insurance and margin requirements, further aggravate the herding behaviour during periods of market stress.

Rebonato and Gaspari (2006) conducted an extensive study of draws in US LIBOR and swap rates and found a significantly larger number of outliers among the interest rates with shorter maturities, for both drawdowns and drawups. By and large, when the $z$ coefficient for drawups is close to 1 and for drawdowns higher than
1, thin tails are suggested (see Table 3.1). The runs of length 1 (immediate price reversals) are more often than expected, when compared to the null hypothesis of independent price moves. Runs of lengths 2 and 3 are, on the other hand, less often than expected.

Rebonato and Gaspari (2006) study in the USD interest rate market is the first and perhaps also the only bivariate analysis of draws. It is well known that the first eigenvector from a PCA analysis accounts for 80% - 90% of the movements of the interest rate term structure. The study provides some evidence of coinciding drawdowns within the blocks of the short rates (3m, 6m and 1y), the mid-term rates (2y, 5y and 10y) and the long-term interest rates (20y and 30y), but not between the blocks. In particular, none of the nine large drawdowns among the short rates coincides with any of the large drawdowns in the other two blocks.

### 3.2.2 Entropy Econometric Literature

An important time series tool based on the information theory of Shannon (1948) has started to gain attention among the finance research community. The key in this line of research is to find optimal encodings of data with respect to the noise. It is a non-parametric way to analyse information or relationships under uncertainty.

Darbellay and Wuertz (2000) use mutual information to estimate the dependencies in the 30-minute returns on the DEM/USD exchange rate, and the daily returns on Dow Jones industrial average stock index. The authors show that returns and volatility, measured over a rolling fixed window, are not independent.

Schreiber (2000) extends mutual information into the concept of transfer entropy, which takes into account transition probabilities. Transfer entropy is particularly useful in detecting information flows between systems and has, since then, found applications in the multivariate analysis of time series in different epistemic domains, e.g. neurology and ecology.

Transfer entropy has been applied to financial time series by Marschinski and
Kantz (2002), who find evidence of information flow from the Dow Jones Industrial Average index to the DAX stock index, sampled at the one-minute interval between May 2000 and June 2001. Any information flow from the reverse direction is considerably less pronounced.

Kwon and Yang (2008b) use transfer entropy and find empirical evidence for information flows between the daily returns of 25 composite stock indices. The value of outgoing transfer entropy from the American and European markets is the highest and appears to flow mainly to the Asian and Pacific regions. Intra-continent information flows are less pronounced than inter-continent flows. Similarly, Kwon and Yang (2008a) found empirical evidence for information flows between daily returns of 25 stock indices and individual single-name stock returns in the respective markets.

Dimpfl and Huergo (2011) used transfer entropy and found evidence of bi-directional information flows between CDS and bond spread, with the information transfer from the CDS market to the bond markets being slightly higher. The data could not be modelled using a Vector Error Correction Model due to the lack of a cointegration relation. Furthermore, a unidirectional information flow from the VIX to the iTraxx is also established.

Jizba, Kleinert, and Shefaat (2011) extend the transfer entropy into the Rényi transfer entropy, using Rényi Entropy, $S_q^{(R)}$. $S_q^{(R)}$, defined below, is a family of entropy measures for a distribution $P$ on a finite set $X$ and indexed by an order $q > 0$:

$$S_q^{(R)}(P) = \frac{1}{1 - q} \log_2 \sum_{x \in X} p_q^q(x)$$

(3.9)

Unlike the Shannon entropy, with the control of the parameter $q$, the Rényi entropy can place emphasis on different parts of the distribution, not necessarily taking into account equally the entire underlying empirical price distribution. A $q \in (0, 1)$ focuses on the tail of the distribution, and with $q > 1$, the focus is on the main part of the distribution. The Rényi transfer entropy was estimated on daily returns of 11 stock indices from Europe, the US and Asia over the period from 1990 to 2009.
The authors found that US and European markets are more influenced by price turbulence in the Asia-Pacific region than the other way around.

### 3.3 Data

Log returns of daily closing and hourly EUR/USD and GBP/USD exchange rates were calculated for the period from 3rd January 2001 to 27th July 2012. There are 3,017 daily and 71,039 hourly observations.\(^4\)

![EUR/USD and GBP/USD time series](image)

Figure 3.1: EUR/USD, GBP/USD time series

Daily EUR/USD and GBP/USD exchange rates for the period from 3rd January 2001 to 27th July 2012.

The sample period includes different market regimes and a number of significant economic events, such as the financial crisis of 2007. Table 3.2 presents statistics for exchange rate returns and drawdowns/drawups. The first four lines of Table 3.2 report the four first central moments \((\mu_1, \mu_2, \mu_3, \mu_4)\) of the respective return series. The mean for the daily series are comparable to the mean of the respective hourly series for both currency pairs, sampled 24 hours within a day. Within the chosen sample period, both EUR and GBP appreciated against the USD (more so

\(^4\)Daily closing rates were from Bloomberg. The hourly data were downloaded from www.fxhistoricaldata.com.
Table 3.2: EUR/USD and GBP/USD return and draw statistics.

The first four sample moments $\hat{\mu}_1, \ldots, \hat{\mu}_4$ of the hourly (h) and daily (d) time series are estimated. $\hat{\rho}$ is correlation between the currency pairs in the hourly and daily series. In panel B we report the expected draw magnitude $E[D]_{\text{Gauss}}$ and daily drop $E[d]_{\text{Gauss}}$ assuming iid Gaussian returns. Panel C lists for drawdowns (drawups) expected length $E[l_d]$ ($E[l_u]$), standard deviation $\sigma(D)$ ($\sigma(U)$) and standard deviation of length $\sigma(l_d)$. The largest drawdowns and drawups in the time series with magnitude and start time are reported in panel D.

for the EUR than the GBP) resulting in positive means for both currency pairs. The Volatility ($\mu_2$) of the EUR/USD hourly series is comparable to the daily series (applying a square root of time scaling). For the GBP/USD series, the hourly series is more volatile than the daily series. EUR/USD is positive skewed, reflecting the general appreciation of the EUR during the sample period despite the European sovereign crisis. GBP/USD is negatively skewed, mainly due to the large movements in 2008. All the series have a kurtosis higher than four, which is more pronounced for the hourly than the daily returns. The unconditional correlation $\hat{\rho}$ between the two exchange rates is 0.673 for both the daily and hourly data.

The statistics concerning the drawdowns and drawups in the four time series are reported in the lower panel of Table 3.2. For all series, the average draw magnitude is lower than if the series were a normally distributed iid process, i.e. $E[D] \leq E[D]_{\text{Gauss}}, E[U] \leq E[D]_{\text{Gauss}}$. The average length of the drawdown is lower than the Gaussian case with $E[l_d] \leq 2$, but higher for drawups, with the exception of the
Table 3.3: Stretched exponential distribution fit

For drawdowns (drawups), the stretched exponential $P_{se}(D) = D_0 e^{-\left(\frac{|D|}{\chi}\right)^z}$ is estimated using the maximum likelihood method. The shape parameter, $\hat{z}$, and size parameter, $\hat{\chi}$, of the hourly and daily series are reported for drawdowns in Panel A and drawups in Panel B. Following Rebonato and Gaspari (2006), we calculate the standard error of the estimates using a Jackknife procedure.

The bulk of the drawdowns have a length $|l_d| \leq E[l_d] + \sigma(l_d) \approx 3.3$ periods, and similar numbers hold for drawups. The average drawdown magnitude is marginally lower than that of the average drawup $E[D] < E[U]$. For both currency pairs, the largest draws in the daily series occurred in autumn 2008, at the height of the financial crisis, which started in 2007 (Panel D in Table 3.2). The largest drawdown of the daily EUR/USD series started on September 29, 2008, lasting six days, and was triggered by renewed concerns over the European financial system. Two months later, the largest drawup, lasting six days, started on December 10, 2008, in the week when the Federal Reserve Bank cut its short term interest rates to historical lows. On October 17, 2008, the largest drawdown of the daily GBP/USD was observed. It lasted seven days and was set in a period of sell off of the pound due to the macroeconomic environment and concerns over the British banking sector. Immediately following it, there was a market correction on October 28 lasting three days, which formed the largest drawup in the GBP/USD daily return series. The expected drawdown length for all series is lower than the Gaussian case with $E[l_d] < 2$, which shows a tendency for price reversal. With the exception of the daily EUR/USD, all drawups had a marginally higher expected length than in the Gaussian case with $E[l_d] > 2$.

Table 3.3 shows the shape parameter, $\hat{z}$, and the size parameter, $\hat{\chi}$, of the modified exponential distribution fitted the drawdowns and drawups, and based on the
maximum likelihood method. This method is known to be asymptotically unbiased but requires large samples. The estimation error $\hat{\sigma}$ is calculated by the bootstrapping technique (see Rebonato and Gaspari (2006)). New data sets are generated repeatedly by replacing a subset of the data by data drawn from an exponential distribution specified by the estimated parameters. The daily returns have a draw shape parameter $\hat{z} > 1$ by more than two standard deviations for drawdowns. This suggests thinner tails than the simple exponential distribution. On the other hand, the drawups of the daily return series exhibit a shape $\hat{z}$ which is within two standard deviations of the simple exponential distribution. Similarly, the estimated shape $\hat{z}(D)$ of the hourly drawdown is within two standard deviations of the simple exponential. For the drawups of the hourly series, $\hat{z}(U) < 1$ by more than two standard deviations, which indicates a thicker tail than the simple exponential.

For drawdowns in both daily and hourly series, the characteristic scale $\hat{\chi}(D)$, determining the expected draw $E[D]$, is within two standard deviations to the mean, i.e. $|E[D] - \hat{\chi}(D)| < 2 \cdot \sigma_{\chi}(D)$. The same is true for the daily drawups, but not for the hourly drawups, which is lower than the expected drawup $E[U]$.

### 3.4 Information Theoretic Method

#### 3.4.1 Entropy Measures

In Chapter 2, we introduced the entropy measures, *Mutual Information* and *Transfer Entropy*. These allows us to detect correlation and cross correlation effects between time series. Mutual Information $I(X;Y)$ of two random variables $X$ and $Y$ with joint distribution $p(x,y)$ is defined as

$$I(X;Y) = - \sum_{x \in X, y \in Y} p(x, y) \cdot \log_2 \frac{p(x, y)}{p(x)p(y)} \quad (3.10)$$

Mutual information measures the reduction of uncertainty about $X$ from observing $Y$. We further saw in Chapter 2 that introducing a time delay in one of the obser-
vations does not distinguish information that is actually exchanged from a common response to input signal or a common history driven by an external factor. To detect information spillover, we use transfer entropy instead. Let \( p(x_1, \ldots, x_n) \) denote the probability of observing the subsequence \((x_1, \ldots, x_n)\), then transfer entropy is defined as

\[
T_{Y \rightarrow X}(m, l) = \sum p(x_{t_1}, \ldots, x_{t_m}, y_{t_{m-l+1}}, \ldots, y_{t_m}) \cdot \log_2 \frac{p(x_{t_{m+1}} | x_{t_1}, \ldots, x_{t_m}, y_{t_{m-l+1}}, \ldots, y_{t_m})}{p(x_{t_{m+1}} | x_{t_1}, \ldots, x_{t_m})}
\]  

(3.11)

where \(x_t\) and \(y_t\) represent the discrete states of \(X\) and \(Y\) at time \(t\). The parameters \(m\) and \(l\) indicate the number of past observations included in \(X\) and \(Y\) respectively.

While mutual information quantifies the deviation from \(X\) (or \(Y\)) being independent, transfer entropy quantifies the deviation from \(X\) being determined by its own history only (via conditional probabilities) or by the history of \(Y\). Unlike mutual information, transfer entropy \(T_{Y \rightarrow X}(m, l)\) is not symmetric, and takes into account only the statistical dependencies originating in the variable \(Y\) and not those from a common signal.

In the literature, both transfer entropy and mutual information have been estimated for continuous variables based on the kernel density methods Schreiber (2000), Blumentritt and Schmid (2011) and the maximum likelihood methods Paninski (2003). In this chapter, we are dealing with market-observed discrete returns, and the number of discrete values is too high for any practical purposes and needs to be considerably reduced before calculating the entropy measures. In this chapter, the partition used is motivated by economic considerations focussing on specific market situations, such as large drawdowns and drawups, which may or may not lead to a equal marginal probability for each partition.

Take the example of a return series, displayed in Figure 3.2 with nine time steps. The returns \(\{r_1, \ldots, r_9\}\) form drawdowns, drawups \(\{D_1, \ldots, D_6\}\), which are depicted in the second line of the figure. For instance, the consecutive negative returns \(r_3, r_4, r_5\) form one drawdown \(D_3\). These nine returns are part of a larger time series that
we wish to put into partitions. Here, there are at least two different approaches to partitioning the data. The first approach is to follow the literature (such as discussed in section 3.2.2) by mapping individual returns to a partition (or letter) according to the size of the returns. Let \( q(a, b) \) denote the quantile of \( a \) at \( b \) confidence level.

The first approach is to map the returns as follows:

\[
d_t(r_t) = \begin{cases} 
0 & \text{if } r_t < q(r_t, 0.33) \\
1 & \text{if } q(r_t, 0.33) \leq r_t \leq q(r_t, 0.66) \\
2 & \text{if } r_t > q(r_t, 0.66) 
\end{cases}
\]

(3.12)

Since \( r_3 \) is the only return that is negative and large in absolute term, it is likely to be the only return in the group of nine returns to be mapped to the letter ‘0’. The other eight returns are small in absolute term, and are likely to be mapped to letter ‘1’. The outcome of this mapping is shown in Figure 3.2 as ‘Discretized Returns’.

Alternatively, since we are interested in large drawdowns and drawups, we could map returns according to their membership of a particular type of draw instead of
the magnitude of the individual returns:

\[ d_D(r_t) = \begin{cases} 
0 & \text{if } D(r_t) < q(D, 0.05) \\
1 & \text{if } q(D, 0.05) \leq D(r_t) \leq q(D, 0.95) \\
2 & \text{if } D(r_t) > q(D, 0.95) 
\end{cases} \]  

(3.13)

where \( D(r_t) \) denotes the size of the draw, \( q(D, 0.05) \) is the 5% quantile of the draw distribution and similarly \( q(D, 0.95) \) is in the 95% percentile. This means that the letter '0' is associated with a large drawdown, the letter '2' is associated with a large drawup and the letter '1' has all the draws with small draw size. So, it is likely that in our example, all the returns in \( D_3 \) (i.e. \( r_3, r_4 \) and \( r_5 \)) will be mapped to the letter '0' and the other six returns \( (r_1, r_2, r_6, r_7, r_8 \) and \( r_9) \) will be mapped to the letter '1'. There is no large drawup among the nine returns. The outcome of this mapping is shown in Figure 3.2 as 'Discretized Draw'. In the previous case, the marginal probability for each partition (or letter) will be the same. For the second approach, which is the approach we have adopted here, equal marginal probability is not guaranteed and will depend on the distribution of the draw size.

Let \( X \rightarrow A \) be the discretized random variable where all values are binned into \( M = |A| \) boxes. The entropy of a discrete probability distribution can be calculated by counting the relative frequencies of \( X \) falling into each bin. The estimator, which we will call the naïve entropy estimator, can be calculated as follows.

\[
\hat{H}_{\text{naive}} = -\sum_{i=1}^{M} p_i \cdot \log_2(p_i) = -\sum_{i=1}^{M} \frac{n_i}{N} \cdot \log_2\left(\frac{n_i}{N}\right) \\
= \log_2(N) - \frac{1}{N} \sum_i n_i \cdot \log_2(n_i)
\]  

(3.14)

If the sample size, \( N \), is small, statistical fluctuations will artificially induce deviations to the entropy estimates producing a downward bias. Grassberger (2003) offers a correction to this 'small sample' bias. Assuming that all \( p_i \ll 1 \), the proposed
new estimator $H_\psi$ is as follows:

$$
\hat{H}_\psi = \ln N - \frac{1}{N} \sum_{i=1}^{M} n_i \cdot \psi(n_i) 
$$

$$
\psi(x) = \frac{d(\ln \Gamma(x))}{dx}, \quad \Gamma(0, x) = \int_{1}^{\infty} e^{-xt} \frac{dt}{t}
$$

We will use this estimator in the Section 3.5.2.

Depending on the particular choice of partition, the entropies of the individual processes will change. Herein, the choice of quantile $q_1$ in the discretization scheme (3.13) determines which symbols the draw states in the process will get mapped. With a low $q_1$ the individual processes are expected to have low entropies as the symbols are more unequally distributed. Higher $q_1$ quantiles lead to correspondingly higher entropies. The maximal entropy in this set up of symbol sequences with three letters is given by the entropy of the uniform three symbol distribution:

$$
-3 \cdot \frac{1}{3} \cdot \log_2(\frac{1}{3}) = 1.585.
$$

The entropy estimates for the two exchange rates are shown in Table 3.4. From Table 3.4, the maximal entropy is reached with partitions of quantiles $q_1 \approx 0.225$ and $q_2 = 1 - q_1 \approx 0.775$ which are lower than $q_1 = \frac{1}{3}$. This is the case as the quantiles of the draw distribution were used in the discretization and not the quantile of the draw state distribution. As larger draws tend to be longer than smaller draws, moving from a partition defined by draw magnitude to that based on the discretized draw states will lead to a partition that is ‘skewed’ towards the larger draws.

Transfer entropy, as defined in equation (3.11), uses entropy that is conditional on blocks of history to determine if processes influence each other. In the remainder of this section, we discuss some structural features of $H(X_0|X_{-1}, \ldots X_{-(m+1)})$ of the exchange rate processes defined by the discretization. In Figures 3.3 the conditional block entropies and block entropies of the daily EUR/USD draw states are shown. The left figure is for a partition with quantiles $q_1 = 0.05$, $q_2 = 0.95$ focusing on the large draws. The right figure is for a more equiprobable partition, with quantiles $q_1 = 0.15$, $q_2 = 0.85$. The absolute levels are differ by a factor of two, which is explained by the entropies of the symbol distribution, which are $\hat{H}_{q_1=0.05}(EUR/USD) = 0.968$
Quantile $q_1$: 0.025 0.050 0.075 0.100 0.125 0.150 0.175 0.200 0.225 0.250

Panel A: Daily Series

$\hat{H}_{EUR/USD}^{(d)}$ 0.674 0.968 1.164 1.308 1.411 1.483 1.531 1.561 1.58 1.585
$\hat{H}_{GBP/USD}^{(d)}$ 0.66 0.968 1.163 1.315 1.412 1.478 1.532 1.562 1.58 1.585

Panel B: Hourly Series

$\hat{H}_{EUR/USD}^{(h)}$ 0.597 0.89 1.095 1.247 1.362 1.446 1.507 1.549 1.574 1.585
$\hat{H}_{GBP/USD}^{(h)}$ 0.593 0.888 1.088 1.242 1.357 1.442 1.504 1.547 1.572 1.584

Table 3.4: Entropy $\hat{H}$ of daily and hourly EUR/USD and GBP/USD

The partitions are defined as in equation (3.13), with a range of values for quantile $q_1$:

$$d_D(r_t) = \begin{cases} 
0 & \text{if } D(r_t) < q(D, q_1) \\
1 & \text{if } q(D, q_1) \leq D(r_t) \leq q(D, 1 - q_1) \\
2 & \text{if } D(r_t) > q(D, q_1) 
\end{cases}$$

and $H_{q_1=0.15}(EUR/USD) = 1.48$ (see Table 3.4).

![Figure 3.3: Block Entropy EUR/USD](image)

The figure shows the unconditional $H(X_0, X_{-1}, \ldots, X_{-(m+1)})$ and conditional block entropies $H(X_0|X_{-1}, \ldots, X_{-(m+1)})$ for history lengths of $m = 1$ to $m = 9$ on the discretized time series of EUR/USD. Two discretization schemes are used. On the left $q_1 = 0.05$ and on the right $q_1 = 0.15$.

The conditional entropy for the case $q_1 = 0.05$ drops slower than that for $q_1 = 0.15$. This is consistent with the finding that the length of draws is $E[l_D] \approx 2$ (see Table 3.2 the lines for $E[l_d], E[l_u]$) but conditional on the draw being in the higher quantiles $q = 0.05, 0.95$ the average length of EUR/USD drawdowns can be estimated as $E[l_D|D < q(D, 0.1)] = 4.03 \pm 1.65$ and for drawups as $E[l_U|U > q(U, 0.9)] = 4.21 \pm 1.93$. These results taken together, show that the bulk of large draws have a length not exceeding $4.03 + 1.65 \approx 5.68$ $4.21 + 1.93 \approx 6.14$. Therefore, we would expect to see the conditional block entropy to reach a plateau at lengths...
of over 6.14. We confined the discussion here to the daily EUR/USD series as the picture is very similar for the GBP/USD series and for both hourly series.

3.5 Empirical Results

3.5.1 Dependency between Draws in EUR/USD and GBP/USD

We estimated the mutual information in equation (3.10), with \( \tau = -8, \ldots, 8 \) for daily and hourly exchange rate returns using the discretization scheme in (3.13). With a sample size of around 3,000 for the daily series and 70,000 for the hourly series, the naïve estimator in (3.14) with \( M = 3 \) (for the three letters '0', '1', '2') is used to estimate the mutual information \( I \). As a robustness check against the threshold choice \((q_1 = 0.05, q_2 = 1 - q_1 = 0.95)\) for the discretization scheme, we have also calculated mutual information based on other threshold choices. The results for the daily exchange rate returns are reported in Table 3.5 for \((q_1 = 0.045, 0.050, 0.055)\) and plotted in Figure 3.4. It is clear from Table 3.5 and Figure 3.4 that the mutual information is the highest at \( \tau = 0 \) and drops off sharply for \( \tau \neq 0 \). The mutual information for \( \tau < 3 \) is greater than \( \tau > 0 \) for the same \(|\tau|\), indicating that the EUR/USD is leading with respect to GBP/USD.

To understand how sensitive the mutual information estimates are as regards to the particular choice of partition, various quantiles were chosen for which the mutual information was estimated. In columns two and four in Table 3.5, we report the estimated mutual information for the quantile sets \((q_1 = 0.045, q_2 = 0.955)\) and \((q_1 = 0.055, q_2 = 0.945)\). In the right part of Figure 3.4 we report the mutual information estimates for the lags \( \tau = -2, -1, 0, +1, +2 \) and quantiles ranging from \( q_1 = 0.025 \) to \( q_1 = 0.35 \), which approximates to the equiprobable case. The curves representing the mutual information estimates are continuous, with no spikes at particular quantiles. This gives some assurance that the choice of quantile does not alter the conclusion substantially. It is also clear from the right part of Figure 3.4
I(X_t,Y_{t-r}) I(X_t,Y_{t-r}), q = 0.050

| lag τ | q=0.045 | q=0.050 | q=0.055 | μ(̂I(X,Y_{shuffled})) ± δ | μ(̂I(X_B,Y_B)|p=0.67 ± δ | μ(̂I(X_B,Y_B)|p=0.5 ± δ |
|-------|---------|---------|---------|------------------------|------------------|------------------|
| +4    | 0.00842 | 0.00871 | 0.00973 | 0.00289 ± 0.00203      | 0.05831 ± 0.04773 | 0.03105 ± 0.02221 |
| +3    | 0.01703 | 0.01806 | 0.02037 | 0.00288 ± 0.00203      | 0.07152 ± 0.04389 | 0.03766 ± 0.01956 |
| +2    | 0.03545 | 0.03876 | 0.04383 | 0.00289 ± 0.00203      | 0.08884 ± 0.03524 | 0.04595 ± 0.01427 |
| +1    | 0.05595 | 0.06266 | 0.07181 | 0.00286 ± 0.00198      | 0.10822 ± 0.02336 | 0.05398 ± 0.00901 |
| 0     | 0.07187 | 0.0832  | 0.09617 | 0.00283 ± 0.00194      | 0.12536 ± 0.01992 | 0.06029 ± 0.00897 |
| -1    | 0.04303 | 0.04926 | 0.05633 | 0.00275 ± 0.00192      | 0.10822 ± 0.02336 | 0.05398 ± 0.00901 |
| -2    | 0.01879 | 0.02987 | 0.02278 | 0.00273 ± 0.00191      | 0.08884 ± 0.03524 | 0.04595 ± 0.01427 |
| -3    | 0.00657 | 0.00718 | 0.00771 | 0.00273 ± 0.00189      | 0.07152 ± 0.04389 | 0.03766 ± 0.01956 |
| -4    | 0.00415 | 0.00377 | 0.00398 | 0.00272 ± 0.00188      | 0.05831 ± 0.04773 | 0.03105 ± 0.02221 |

Table 3.5: Mutual Information (daily returns)

In the first three columns, the estimated mutual information ̂I(EUR/USD_t, GBP/USD_{t+τ}) on the discretized time series of the daily return series is reported. The discretization scheme (3.13) has been used with quantile q = 0.050 and as a robustness check with quantiles q = 0.045, 0.055. In the last three columns, the results of numerical simulations of three processes is shown. μ(̂I(X,Y_{shuffled})) shows the value, where one discretized time series is shuffled, destroying any correlation or cross correlation information. In columns μ(̂I(X_B,Y_B)|p=0.67,0.5, the mutual information of correlated processes, with correlation ρ = 0.67, 0.5 and using the discretization scheme (3.13) is calculated.

that mutual information is the highest for lag τ = 0 and substantially lower for all other lags. The increase in information is higher between q_1 = 0.055 and q_1 = 0.125 than for other quantile values.

Next, we compare the empirical estimates with simulated results produced by two correlated Gaussian processes. For each simulated time series, draw states were calculated based on the discretization scheme (3.13) and the naïve estimator was applied. The correlation factor used in one simulation set is the estimated unconditional correlation of the two return series EUR/USD and GBP/USD which is ̂ρ = 0.673 (see Table 3.2). This allows us to quantitatively describe the exclusive influence of correlation on the draw dependence in the time series under the simplifying assumption of Gaussian returns. The results are depicted in Figure 3.4 and in the sixth column of Table 3.5. The mutual information for the simulated draw states is consistently and markedly higher than that for the empirical series at all lags. The lower mutual information of the pair of empirical draw state series indicates that large draws in the empirical series show a higher degree of independence than what would be expected from a pair of correlated Gaussian processes with the same correlation coefficient. The difference between simulated draws and empiri-
A further pair of correlated Gaussian processes were simulated with a lower correlation of $\hat{\rho} = 0.5$. The specific factor has been set after experimenting with different correlations, with the aim of having one lower bound on the mutual information at lag $\tau = 0$. The results of this simulation are also shown in Figure 3.4, and the seventh column of Table 3.5. At lag 0, the mutual information is lower than that for the empirical series. Also, in this set-up, the mutual information drops much slower for lags $\tau \neq 0$ than the case for the empirical series. At lags $\tau \in \{0,1\}$, the mutual information of the simulated pair is higher than the corresponding mutual information of the empirical series.

To quantify the error when estimating entropy (mutual information), we re-estimate the mutual information of the two given time series, but with one underlying return series being shuffled. The result of the shuffle is to strip off any correlation and cross-correlation between the series. Furthermore, any auto-correlation information is also destroyed by the shuffling, leading to a different draw distribution. This is done iteratively, generating a sample of shuffled series. The discretization procedure is applied to all shuffled series, and the mutual information is calculated and presented in the figure as $I(X_{\text{shuffled}}, Y)$. The mean and standard error of the
In the first three columns, the estimated mutual information \( \hat{I}(\text{EUR/USD}_t, \text{GBP/USD}_{t+\tau}) \) on the discretized hourly return series is reported. The discretization scheme (3.13) has been used with quantile \( q = 0.050 \) and as a robustness check with quantiles \( q = 0.045, 0.055 \). \( \mu(\hat{I}(X, Y_{\text{shuffled}})) \) shows the value, where one discretized time series is shuffled, destroying any correlation or cross correlation relations.

We repeat the procedure above for the hourly exchange rates and report the results in Table 3.6. The findings for the hourly relationship are similar to that reported above for the daily returns, in that mutual information, \( I \), peaked at \( \tau = 0 \), and dropped off for \( \tau \neq 0 \). The \( I \) at \( \tau < 0 \) is marginally higher than the equivalent \( |\tau| \) for \( \tau > 0 \). As in the case of the daily returns, we test the sensitivity of our results to the quantile choice. Table 3.6 shows the results for the quantiles \( q_1 = 0.045, 0.050, 0.055 \) for different lags \( \tau = -5, \ldots +5 \). The estimated \( \hat{I} \) for the hourly series contains higher levels of information than that for the daily draw state series at \( \tau = 0 \) and other short \( |\tau| \). So there is a stronger dependency among the hourly returns at contemporaneous term and at short lags.
3.5.2 Information Flows between Draws in EUR/USD and GBP/USD

Transfer entropy can be expressed as the difference in two block entropies, as shown in equation (3.11). Here, we use the three letter partition defined in equation (3.13) and the Grassberger estimator \( \hat{H}_\phi \) in equation (3.15) for the estimation of transfer entropy.

Since the transfer entropy is not zero for small sample, we follow Marschinski and Kantz (2002) to estimate the mean and standard error by the bootstrapping method, involving reshuffling the predictor series, and destroying all its time series pattern. This process is repeated enough times, after which a mean, \( \mu(\hat{T}_{sh}) \), and a standard deviation, \( \sigma \), is calculated, and the adjusted **Effective Transfer Entropy** (ET) is calculated as:

\[
ET_{Y \rightarrow X}(m, l) = T_{Y \rightarrow X}(m, l) - \mu(\hat{T}_{sh})
\]  

(3.17)

For \( ET_{Y \rightarrow X}(m, l) \), \( m \) refers to the length of own past history in \( X \), and \( l \) to the length of past history of \( Y \), used in the prediction of \( X \). The findings of \( ET \) for \( m = 1, \ldots, 4 \) and \( l = 1, \ldots, 4 \) are presented in Figure 3.5. The detailed statistics for the more significant cases (i.e. represented by the darkest block) are presented in Tables 3.7 and 3.8. From Figure 3.5, we note that for a fixed \( m \), the \( ET \) generally increases as \( l \) increases. On the other hand, for the same \( l \), \( ET \) may decrease as \( m \) increases if \( X \) can be forecast more efficiently using its own past.

For the results presented in Figure 3.5 and Tables 3.7 and 3.8, we conclude that the maximum \( \hat{ET}_{EUR/USD \rightarrow GBP/USD} \), is at \( m = 2 \) and \( l = 4 \), whereas for \( \hat{ET}_{GBP/USD \rightarrow EUR/USD} \), \( m = 1 \) and \( l = 4 \). From the amount of \( \hat{ET} \) for the two exchange rates, there is slightly more information flow from GBP/USD to EUR/USD than the other way around. Note that the negative values of effective transfer entropy in Figure 3.5 are due to the fact that the estimated transfer entropy resulting from the shuffling is higher than that of the original time series. More noise is in-
Figure 3.5: $\hat{ET}_{\text{EUR}} \rightarrow \text{GBP}$ and $\hat{ET}_{\text{GBP}} \rightarrow \text{EUR}$ (daily returns)

On the discretized series of daily prices, effective transfer entropy, as specified in equation (3.17), is shown for block lengths of one to four days’ history. The left figure shows the estimated effective transfer entropy $\hat{ET}_{\text{EUR}/\text{USD}} \rightarrow \text{GBP}/\text{USD}(m, l)$. In the right figure, the information flow, as measured by the effective transfer entropy in the other direction, GBP/USD to EUR/USD, is shown.

<table>
<thead>
<tr>
<th>History</th>
<th>Transfer Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>m ERP/USD 1 EUR/USD</td>
<td>$\hat{T}_{\text{naive}}$ [10^{-3}]</td>
</tr>
<tr>
<td>1 4</td>
<td>40.774</td>
</tr>
<tr>
<td>2 4</td>
<td>45.569</td>
</tr>
<tr>
<td>3 4</td>
<td>54.445</td>
</tr>
<tr>
<td>4 4</td>
<td>68.163</td>
</tr>
</tbody>
</table>

Table 3.7: Transfer entropy EUR/USD → GBP/USD

The information flow between draws from EUR/USD to GBP/USD is estimated using daily prices. These are calculated for block lengths of up to four days history. Transfer Entropy is estimated with the naïve estimator $\hat{T}_{\text{naive}}$, and the Grassberger estimator $\hat{T}$ are reported. With the predictor series EUR/USD reshuffled repeatedly, $\mu(T_{nh}) \pm \sigma$ reports the bias due to small sample effects. Effective Transfer Entropy $\hat{ET}$ and the relative explanation added (REA) is reported in the last columns.

introduced as one increases $m$ and $l$. When comparing the $ET$ estimates with the mutual information, $\hat{I}$, reported in the previous section, we notice that the amount of information gain is much lower.

To understand the relative information gain due to the transfer entropy we follow Marschinski and Kantz (2002) by relating this information gain to the entropy of
Table 3.8: Transfer entropy GBP/USD → EUR/USD
The information flow between draws from GBP/USD to EUR/USD is estimated using daily returns. These are calculated for block lengths of up to four days’ history. \( \hat{T} \) is the Grassberger estimator for Transfer Entropy. With the predictor series EUR/USD reshuffled repeatedly, \( \mu(T_{sh}) \pm \sigma \) reports the bias due to small sample effects. Effective Transfer Entropy \( \hat{ET} \) and the relative explanation added (REA) is reported in the last columns.

<table>
<thead>
<tr>
<th>History</th>
<th>Transfer Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR/USD</td>
<td>GBP/USD</td>
</tr>
<tr>
<td>m</td>
<td>( \hat{T} ) (bit)</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

The information flow from \( Y \) to \( X \), measured by the transfer entropy, is related to the total flow of information measured in \( X \) alone, based on the conditional block entropy of \( X \). Put differently, it measures the additional information gain by observing the history of \( X \) and \( Y \) when already observing \( X \). The results in Tables 3.7 and 3.8 suggest that the amount of information gain from the predictive exchange rate is about 2.5% for both exchange rates. For a comparison, the REA for using DAX to predict the Dow Jones Industrial, as estimated by Marschinski and Kantz (2002), using the one-minute tick data, is only 0.4% for a three symbol discretization.

Finally, we can conclude that the maximum information gain from the given draw states of a series is to use one or two days of own history (\( m = 1, 2 \)) and up to four days of the predictor’s history (\( l = 4 \)). Using longer own history beyond two days may contain more information but also induces more noise.

We now repeat the estimation using the hourly returns of EUR/USD and GBP/USD. The \( \hat{ET} \) estimated for \( m = 1, \ldots, 8 \) and \( l = 1, \ldots, 8 \) are presented in...
History Transfer Entropy

\[ \hat{T}_{sh} (\text{bit}) \mu (\hat{T}_{sh}) \pm \sigma (\text{bit}) \hat{E}T (\text{bit}) \text{ REA} \quad \text{[bit]} \quad \text{[10}^{-3}] \quad \text{[10}^{-3}] \quad \% \]

<table>
<thead>
<tr>
<th>m</th>
<th>1</th>
<th>GBP/USD</th>
<th>EUR/USD</th>
<th>( \hat{T} ) (bit)</th>
<th>( \mu (\hat{T}_{sh}) \pm \sigma ) (bit)</th>
<th>( \hat{E}T ) (bit)</th>
<th>REA</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>8</td>
<td>66.219</td>
<td>35.427</td>
<td>1.253</td>
<td>30.792</td>
<td>6.849 ± 0.279</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>77.439</td>
<td>46.802</td>
<td>1.467</td>
<td>30.637</td>
<td>6.858 ± 0.329</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>88.898</td>
<td>58.035</td>
<td>1.612</td>
<td>30.863</td>
<td>6.963 ± 0.364</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>96.159</td>
<td>67.108</td>
<td>1.764</td>
<td>29.052</td>
<td>6.631 ± 0.403</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>88.994</td>
<td>63.064</td>
<td>1.672</td>
<td>25.931</td>
<td>6.015 ± 0.388</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>100.506</td>
<td>73.534</td>
<td>1.806</td>
<td>27.402</td>
<td>6.356 ± 0.419</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.9: Transfer entropy EUR/USD → GBP/USD (hourly prices)
The information flow between draws from EUR/USD to GBP/USD is estimated using hourly prices. These are calculated for block lengths of up to eight hours’ history. \( \hat{T} \) is the Grassberger estimator for Transfer Entropy. With the predictor series EUR/USD reshuffled repeatedly, \( \mu (\hat{T}_{sh}) \pm \sigma \) reports the bias due to small sample effects. Effective Transfer Entropy \( \hat{E}T \) and the relative explanation added (REA) is reported in the last columns.

Figures 3.6 and 3.7. The levels of \( \hat{E}T \) are higher than those for the daily returns. The statistics for the more significant results are summarised in Table 3.9 and Table 3.10.

Figure 3.6: Draws Information Flow from hourly EUR/USD to GBP/USD
Effective transfer entropy as specified in equation (3.17) is shown for block lengths of one to eight hours history. The figure shows the estimated effective transfer entropy \( \hat{E}T_{EUR/USD\rightarrow GBP/USD}(m,l) \).

The maximum \( \hat{E}T \) 0.0353 bits for the hourly pair is reached in the direction from GBP/USD, considering a history of four hours of EUR/USD and eight hours of GBP/USD, with \( \text{REA} = 7.828 \pm 0.25 \). So marginally more information flows from...
Figure 3.7: Draws Information Flow from hourly GBP/USD to EUR/USD

Effective transfer entropy as specified in equation (3.17) is shown for block lengths of one to eight hours history. The figure shows the estimated effective transfer entropy $\hat{ET}_{GBP/USD\rightarrow EUR/USD}(m, l)$.

GBP/USD to EUR/USD than the other way around. It is clear from the results that the greatest information gain is when $l = 8$, i.e. up to eight hours of the past history of the predictive variable are used. The amount of own history needed is less clear cut and the difference in $\hat{ET}$ is small. Based on the maximum $\hat{ET}$, optimal $m = 6$ for GBP/USD and $m = 4$ for EUR/USD.

Based on REA the predictive variable contribute another 7% to 8% of forecasting power compared to using only own past history. Given the importance of the EUR/USD pair, its liquidity and volume, the result of a marginally higher information flow from GBP/USD to EUR/USD than the other way around, seems a little surprising. However, we should note that the difference in information for every pair of history blocks $(m, l)$ is very small for the daily return series and is within one standard error $\hat{\sigma}(T_{sh})$ (see the fifth column in Tables 3.7 and 3.8). Moreover, the draws considered are large in magnitude (0.05, 0.95 quantiles of the respective drawdown and drawup distributions), so they belong to a part of the return/drawstate distribution where dependency between the two series could be different from that for the main part of the distribution, hence the result may not necessary match
Table 3.10: Transfer entropy GBP/USD → EUR/USD (hourly prices)
The information flow between draws from GBP/USD to EUR/USD is estimated using hourly prices. These are calculated for block lengths of up to eight hours history. $\hat{T}$ is the Grassberger estimator for Transfer Entropy. With the predictor series GBP/USD reshuffled repeatedly, $\mu(T_{sh}) \pm \sigma$ reports the bias due to the small sample effect. Effective Transfer Entropy $\hat{ET}$ and the relative explanation added (REA) is reported in the last columns.

<table>
<thead>
<tr>
<th>History</th>
<th>Transfer Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>$\hat{T}$ (bit)</td>
</tr>
<tr>
<td>EUR/USD</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>8</td>
</tr>
</tbody>
</table>

common observations. An impact analysis by Omrane and Hafner (2009) suggests that in the very short term (in hours), positive macroeconomic news announcements in the US increase the GBP/USD volatility more strongly than that of EUR/USD. Moreover, during the crisis of 2007, in the foreign exchange markets, Melvin and Taylor (2009) found evidence that the volatility of bid-ask spreads for spot trades increased for all major currency pairs, and the increase in the GBP/USD pair was the most dramatic (+5,000% increase in average bid-ask spread).

We saw that transfer entropy is able to detect information flows between the draw state of the FX series. The level of information flow is three times greater for the hourly data than for the daily data. Using relative explanation added (REA), relating the information gained to the entropy of the block history, we were able to show that this information gain is considerable, especially when comparing the results to findings reported elsewhere in the literature.

3.6 Conclusion

Using entropy measures, this chapter investigates the dependence and spillover relationship between the drawups and drawdowns of EUR/USD and GBP/USD. For the daily series, we found evidence of dependence among the largest draws (i.e. 5% and 95% quantiles), but it is not as strong as that expected from the correlated
Gaussian series at the same level of correlation. There is a dependence between lead or lagged values of these draws. Similar findings were obtained from the hourly data, although the dependence is much stronger in the hourly data. Next we used transfer entropy to examine the spillover and lead-lag information flow between the drawup/drawdown of the two exchange rates. Such information flow is indeed detectable in both daily and hourly data. The amount of information transferred is considerably higher for the hourly than the daily data. The data demands in estimating the transfer entropy measure are considerable. Given the limited number of observations in the daily series, the lead-lag terms considered was restricted to four by four (days). For the hourly data we were able to extend our analysis to eight by eight (hours). Both daily and hourly series show clear evidence of information flowing from EUR/USD to GBP/USD and in the reverse direction. Robustness tests, using effective transfer entropy $ET$, show that the information measurable is not due to noise. By using the quantity ‘relative explanation added’ (REA), which relates the information gain to the entropy of the process and comparing our results to documented cases in the literature (see Marschinski and Kantz (2002), Dimpfl and Huergo (2011), Kwon and Yang (2008a)), we conclude that there is a measurable information transfer between the two exchange rates that is potentially useful for forecasting and risk management. Many questions remain open, and these will be addressed in future analyses. For the moment, we do not have a model for the data generating process to understand how strong the draw state dependence between two time series is. The other immediate improvement planned is to employ the bootstrapping method from Horowitz (2003) in order to make inferences. One such solution is a standard block bootstrap. However, as has been outlined by Lahiri (1999), standard methods such as non-overlapping-block bootstraps, moving-block bootstraps or bootstraps with random block length produce biased estimates. Horowitz (2003) introduced a bootstrapping method based on the transition probabilities of the underlying Markov process. In this the process, $Y$ is simulated based on the calculated transition probabilities, thereby destroying the dependencies between $Y$ and $X$, but retaining the dynamics of the series $Y$. Transfer entropy is
then estimated again using the simulated time series. Repeating this procedure yields the distribution of the transfer entropy estimate under the null hypothesis of no information flow. The third extension which we plan to follow is the use of Renyi’s entropy instead of Shannon’s entropies in calculating transfer entropy. The Renyi entropy lends itself readily to the analysis of a tail event, as shown by Jizba, Kleinert, and Shefaat (2011), and would allow for a more robust way to examine different parts of the draw tail distributions.
Chapter 4

Information Flows between FX Volatility Regimes

Abstract

We use state space models of volatility to investigate volatility spillovers between exchange rates. Our use of entropy-related measures in the investigation of dependencies of two state space series is novel. A set of five daily exchange rates from emerging and developed economies against the dollar from 1999-2012 is used. We find that among the currency pairs, the co-movement of EUR/USD and CHF/USD volatility states show the strongest observed relationship. With the use of transfer entropy, we find evidence for information flows between the volatility state of AUD, CAD and BRL. We can further measure an information flow from EUR/USD to GBP/USD, which indicates a causal volatility spillover relationship.
4.1 Introduction

Relations between large drawdowns and drawups of two currency pairs were the focus of the previous chapter. Tumultuous periods, which in most cases are the market regimes generating large draws, are the topic of this study. Using hidden Markov models to identify volatile periods, we investigate volatility transmissions between foreign exchange rates applying the information theoretic tools that we used in the previous chapter.

History provides many examples of financial crises and how turbulence in financial markets spread across markets within a short time. Evidence for volatility transmissions has been found in many markets. Edwards and Susmel (2000) analysed weekly stock market data for a group of Latin American countries with a switching GARCH model (SWARCH). A univariate version of the SWARCH model is used to identify breakpoints in the ARCH model of conditional variance and highly volatile periods. The multivariate SWARCH is used to find evidence of volatility co-movements across countries. Volatility spillover effects have also been identified between different asset classes. Baba, Packer, and Nagano (2008) found that during the second half of 2007, turmoil in the money markets spilled over to the FX swap markets and the cross-currency basis swap market. The mechanism underlying the spilling over was that at the time, the FX swap market was increasingly used by financial institutions to cover US dollar funding shortages.


From a methodological point of view, the extent to which the approaches identify
spillover, co-movement or interdependence vary considerably. Engle, Ito, and Lin (1990) employed a GARCH model in a vector autoregression to test if conditional variances were affected by the squared innovations (i.e. news, information) in other markets. Cheung and Ng (1996) used the cross-correlation function between squared residuals, which were standardized by their individual conditional variance estimators, as a test for volatility spillover. Empirical studies Poon and Granger (2003) have shown that GARCH-estimated volatilities appear to be too persistent following large shocks. In general, the standard single-state GARCH model does not provide for the different speeds of volatility adjustment in different states. Hamilton (1989) introduced a model in which changes in regimes are governed by a Markovian hidden state process. It gives more flexibility to the different adjustment speeds of volatility, and the persistence of the estimated volatility for the different states are very different from that estimated using the standard single-state GARCH. Later, Hamilton and Susmel (1994) adapted this model and introduced the SWARCH model, which they applied to identify volatility regimes in stock returns. SWARCH models has a Markov-modulated GARCH process, where the conditional variance is state dependent. Engle, Ito, and Lin (1990) used this model in their analysis of volatility spillover. Bialkowski and Serwa (2005) used a bivariate Markov switching process, where each state corresponds to a bivariate normal distribution. They find evidence of feedback spillovers (based on the Granger causality test) between the Japanese and Hong Kong equity indices during the Asian crisis in 1997.

Volatility spillover is a situation whereby a change in volatility in one market is related to a change in volatility in another market. When the change in volatility regime of two markets is contemporaneous, it is termed a co-movement, or inter-dependent when changes in two markets at different periods are related. In the literature, the term spillover is sometimes specifically reserved for causality in variance. When drawing conclusions from our analysis, we will highlight which of the specific situations we refer to.

In this study, we follow Hamilton (1989) in identifying volatility regimes among a set of currencies quoted against the US dollar. We use squared currency returns as
a proxy for daily volatility. A two-state hidden Markov model is defined, with which we model high and normal volatility states. Following the procedures in Chapter 3, we use the entropy framework with mutual information and transfer entropy to evaluate various interdependency hypotheses. The estimated state process is a discrete Markov process, which lends itself to the analysis with information theoretic tools. We find evidence of various volatility regime relationships between the currency pairs in our sample. Among the European currencies (EUR/USD, GBP/USD, CHF/USD), there are volatility regime co-movement relationships that are measurable. CAD/USD and AUD/USD exhibit a volatility co-movement relationship and, to a lesser extent, interdependency with time lag. We also find evidence for information flows between the volatility states of currency pairs. Most notable is an information flow from EUR/USD to GBP/USD, which indicates a causal volatility spillover relationship, in line with the findings in Inagaki (2006), who reported evidence for such a unidirectional volatility spillover from EUR to GBP between 1999 and 2004. Our work showcases also the usefulness of the concepts of mutual information and transfer entropy when applying them to analyse volatility spillover relationships.

4.2 Hidden Markov Models

4.2.1 Foundations

A Hidden Markov Model (HMM) consists of a random process \( \{y_1, \ldots, y_n\} \), where the probability distribution depends on the realisation of a 'hidden' markov chain \( \{q_1, \ldots, q_n\} \) with finite state space \( I = \{1, \ldots, N\} \). The distribution of \( \{y_t\} \) depends on the state \( q_t \), so \( P(y_t|q_t) \). A Hidden Markov Model is denoted by \( \theta = \{P(q_1), P(q_{t+1}|q_t), P(y_t|q_t)\} : \)

- \( P(q_1) \) is an initial state probability.

- \( a_{ij} \equiv P(q_{t+1} = i|q_t = j) \) for \( 1 \leq i, j \leq N \) are the transition probabilities between the states.
\( P(y_t|q_t) \) are the emission probabilities.

There are no restrictions as to the choice of the emission probability distributions. In many applications, normal distributions, a mixture of normal distributions and Poisson distributions have been used (Zucchini and MacDonald (2009)). The set-up of a Hidden Markov model is depicted in Figure 4.1. Each node in the figure represents the value of a random variable. A line connecting two nodes represents dependence and the absence of a connecting line implies conditional independence. The hidden state \( q_{t+1} \) at time \( t+1 \) depends only on \( q_t \) in the previous state as shown in equation (4.1), and the emission at time \( t+1 \) depends only on \( q_{t+1} \) at time \( t+1 \) as shown in equation (4.2):

\[
P(q_{t+1}|q_1, \ldots, q_t, y_1, \ldots, y_t) = P(q_{t+1}|q_t) \quad (4.1)
\]
\[
P(y_{t+1}|q_1, \ldots, q_t, y_1, \ldots, y_t) = P(y_{t+1}|q_{t+1}) \quad (4.2)
\]

Equation (4.1) implies that the Markov process \( \{q_1, \ldots, q_n\} \) is fully determined by the initial state probability \( P(q_1) \) and the transition probabilities \( P(q_t|q_{t-1}) \). These relations allow for various conditional independence relations that are used in the estimation of HMM parameters. See Barber (2012) for a more general discussion on the conditional independence relations in graphical models and the d-separation feature.
4.2.2 HMM Estimation

We will explain in this section two central methods for estimating the HMM models. The Expectation Maximisation Algorithm (EM), otherwise known as the Baum-Welch Algorithm, enables us to find the maximum likelihood estimates of the HMM model parameters when we cannot observe all the data (Barber (2012)). There is also the Viterbi algorithm that provides the inference of the most likely state sequence of the hidden states, given the observations.

First of all, we note that there are two ways of indexing a time series, depending on the context. Starting at time $t_1$ and ending at time $t_2$, a sequence $\{y_t\}$ is denoted $y_{t_1}^{t_2} = y_{t_1}, y_{t_1+1}, \ldots, y_{t_2}$. For a particular reference point in time with the observation denoted by $y_0$, we look back with $y_{-n}, \ldots, y_{-1}, y_0$ and forward with $y_0, y_1, \ldots y_n$.

The forward-backward algorithm is used to calculate the conditional probability $P(q_t, y_1^T)$ of a hidden state at a specific time, given all the observations in the time series $\{y_t\}_{t \leq T}$. The forward algorithm is concerned with the probability of $P(q_t, y_1^t)$ for all $t \in \{1, \ldots T\}$ and state space $I = \{1, \ldots, N\}$. In the backward algorithm the conditional probability $P(y_{t+1}^T | q_t)$ is computed for all $t \in \{1, \ldots T\}$. Using the backward and the forward algorithm together, the conditional probability $P(q_t | y_1^T)$ can be computed, assuming that the emission probabilities $P(y_t | q_t)$, transition probabilities $P(q_{t+1} | q_t)$ and initial distribution $P(q_1)$ are known. This follows from the independence relationship in equations (4.1) and (4.2) as shown below:

$$P(q_t | y_1^T) \propto P(q_t, y_1^T) = P(q_t, y_1^t, y_{t+1}^T)$$

$$= P(y_{t+1}^T | q_t, y_1^t) \cdot P(q_t, y_1^t)$$

$$= P(y_{t+1}^T | q_t) \cdot P(q_t, y_1^t)$$

$$= \beta_t(q_t) \cdot \alpha_t(q_t) \quad (4.3)$$

For the second last line in equation (4.3) above, the first item $P(y_{t+1}^T | q_t)$ is calculated using the backward algorithm and the second item $P(q_t, y_1^t)$ is calculated using the forward algorithm. The terms of the last line in equation (4.3) above are explained
The probabilities \( P(q_t, y_t^1) \equiv \alpha_t(q_t) \) for a given set of parameters \( \theta = \{P(q_1), P(q_{t+1}|q_t), P(y_t|q_t)\} \) obey a recursive relationship that can be calculated using the independence relations (4.1) and (4.2) as follows:

\[
P(q_t, y_t^1) = \sum_{q_{t-1}=1}^{N} P(q_t, q_{t-1}, y_t^1) = \sum_{q_{t-1}=1}^{N} P(y_t|q_t, q_{t-1}, y_t^{1-1})P(q_{t-1}, q_{t-1}, y_t^{1-1})P(q_{t-1}, y_t^{1-1}) = \sum_{q_{t-1}=1}^{N} P(y_t|q_t)P(q_{t-1}, q_{t-1}) \cdot \alpha_{t-1}(q_{t-1}) = \alpha_t(q_t)
\]

\[\text{(4.4)}\]

For a given set of parameters \( \theta = \{P(q_1), P(q_{t+1}|q_t), P(y_t|q_t)\} \), the probabilities \( P(y_{t+1}^T|q_t) \equiv \beta_t(q_t) \) are similarly calculated based on a backward algorithm as follows:

\[
P(y_{t+1}^T|q_t) = \sum_{q_{t+1}=1}^{N} P(y_{t+1}^T, q_{t+1}|q_t) = \sum_{q_{t+1}=1}^{N} P(y_{t+2}^T|q_{t+1}, q_t, y_{t+1})P(y_{t+1}|q_{t+1}, q_t)P(q_{t+1}|q_t) = \sum_{q_{t+1}=1}^{N} P(y_{t+2}^T|q_{t+1})P(y_{t+1}|q_{t+1})P(q_{t+1}|q_t) = \sum_{q_{t+1}=1}^{N} \beta_{t+1}(q_{t+1}) \cdot P(y_{t+1}|q_{t+1}) \cdot P(q_{t+1}|q_t) = \beta_t(q_t)
\]

\[\text{(4.5)}\]

Both algorithms have a computational complexity of order \( O(N^2 \cdot T) \). We explain this briefly for the backward algorithm. In equation (4.5), \( \beta_t(q_t) \) is a sum over all states \( 1, \ldots, N \) repeated at each point in time \( t \in \{1, \ldots, T\} \) and for each of the possible states at time \( t \), giving \( N^2 \cdot T \) calculations in total. In comparison, if \( P(q_t, y_t^1) \)
is summed over all combinations of state sequences as follows:

\[
P(q_t, y_t^T) = \sum_{q_1=1}^{N} \cdots \sum_{q_T=1}^{N} P(q_t, q_{t-1}, y_t^T), \quad (4.6)
\]

the calculation involves all possible state paths from time 1 to time \( t \) and the joint probability of a certain state. This is repeated for all \( t \in \{1, \ldots, T\} \) giving \( N^T \cdot T \) calculations in total.

The parameters \( \theta \) in a HMM model are estimated based on the Expectation Maximisation Algorithm below:

\[
\theta = \arg \max_{\theta} P(y_1, \ldots, y_T | \theta) \quad (4.7)
\]

The algorithm proceeds iteratively in two steps. First, the estimation step computes the conditional expectations of the missing data given the observations and the current estimate of \( \theta \). Next, the maximisation step maximizes the log likelihood with respect to \( \theta \) for the complete data set, including the hidden states.

\[
\xi_{ij}(t) = P(q_t = i, q_{t+1} = j | y_T^T, \theta) = \frac{\alpha_t(q_t = i) \cdot a_{ij} \cdot P(y_{t+1} | q_{t+1} = j) \cdot \beta_{t+1}(q_{t+1} = j)}{P(y_T^T | \theta)} \quad (4.8)
\]

\[
= \frac{\alpha_t(q_t = i) \cdot a_{ij} \cdot P(y_{t+1} | q_{t+1} = j) \cdot \beta_{t+1}(q_{t+1} = j)}{\sum_{1 \leq t \leq T} \beta_t(q_t) \cdot \alpha_t(q_t)} \quad (4.9)
\]

\[
\xi_{ij}(T) = \frac{\alpha_T(q_T = i) \cdot a_{ij} \cdot P(y_T | q_T = j) \cdot \beta_{T+1}(q_{T+1} = j)}{P(y_T | \theta)} \quad (4.10)
\]

Finally, the estimated likelihood is used to classify the state, specifically given the estimated parameters (\( \hat{\theta} \)) from equation (4.7):

\[
q_T^T = \arg \max_{q_T^T} P(y_T^T, q_T^T | \hat{\theta}) \quad (4.11)
\]
4.3 Data

The data set consists of log returns of Bloomberg daily fixing for AUD (Australian Dollar), BRL (Brazilian Real), CAD (Canadian Dollar), CHF (Swiss Franc), EUR (European Monetary Union currency unit) and the GBP (British Pound) quoted against the US dollar for the period from January 1, 1999 to March 11, 2012.\footnote{In the text, we will most of the times refer to the currency pairs AUD/USD, BRL/USD, CAD/USD, CHF/USD, EUR/USD and GBP/USD by their base currency AUD, BRL, CAD, EUR, GBP only.} There are 3,180 data points for each currency pair. The sample period includes several market regimes and significant economic events, such as the post dot.com crash, the financial crisis of 2007, and the beginning of the European debt crisis in 2010. Table 4.1 shows the unconditional correlation of all currency pairs. The correlation between the EUR and CHF is the highest with \( \hat{\rho} = 0.875 \). Higher correlations are measured between the European currencies, EUR, GBP and CHF, with AUD and CAD also exhibiting a higher correlation. BRL has the weakest correlations with all the other currency in the sample.

![Currency Time Series](image)

Figure 4.2: Currency Time Series

All the exchange rates in the sample period from January 1999 to March 2012 are quoted against US dollar.
Table 4.1: Correlation Matrix

Matrix of unconditional correlations between the daily returns of exchange rates in the sample.

<table>
<thead>
<tr>
<th></th>
<th>AUD</th>
<th>BRL</th>
<th>CAD</th>
<th>CHF</th>
<th>EUR</th>
<th>GBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUD</td>
<td>-</td>
<td>0.333</td>
<td>0.623</td>
<td>0.395</td>
<td>0.543</td>
<td>0.505</td>
</tr>
<tr>
<td>BRL</td>
<td>0.333</td>
<td>-</td>
<td>0.623</td>
<td>0.395</td>
<td>0.543</td>
<td>0.505</td>
</tr>
<tr>
<td>CAD</td>
<td>0.623</td>
<td>0.623</td>
<td>-</td>
<td>0.395</td>
<td>0.543</td>
<td>0.505</td>
</tr>
<tr>
<td>CHF</td>
<td>0.395</td>
<td>0.395</td>
<td>0.395</td>
<td>-</td>
<td>0.543</td>
<td>0.505</td>
</tr>
<tr>
<td>EUR</td>
<td>0.543</td>
<td>0.543</td>
<td>0.543</td>
<td>0.543</td>
<td>-</td>
<td>0.505</td>
</tr>
<tr>
<td>GBP</td>
<td>0.505</td>
<td>0.505</td>
<td>0.505</td>
<td>0.505</td>
<td>0.505</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 4.2: Summary Statistics

This table shows the basic statistics for exchange rate returns in the sample. Panel A shows the various quantiles of the return distribution, including the maximum and minimum. Panel B shows the first and second moments, the standard deviation ($\sigma$), the standardized skewness ($\mu_3/\sigma^3$) and the excess kurtosis ($\mu_4/\sigma^4 - 3$).

The return statistics for the sample are summarized in Table 4.2. From Table 4.2, it can be seen that AUD and BRL have daily volatility at 0.87% and 1.16% respectively; the highest volatility in the sample. These two currencies also stand out in terms of extreme returns (min, max), and the kurtosis indicates a significant fat tailedness. The other currencies have comparable return distributions, as measured by the quantiles and moments, with EUR/USD being the world’s most important currency pair, having the lowest volatility and the least fat tails.
4.4 HMM Estimation Results

Here we are interested in market states associated with higher or lower volatility. As a proxy for volatility, we use the daily return squares $r_t^2$ and estimate the Hidden Markov Model with two volatility states $q_t = 1, 2$. The state dependent emission distributions are assumed to be normally distributed as follows:

$$P(y_t|q_t = i) \simeq \mathcal{N}(\mu_i, \sigma_i) \quad i = 1, 2 \quad (4.12)$$

where $y_t = \{r_t^2\}$ for $1 \leq t \leq T$. The HMM model is estimated via the maximum likelihood method described in section 4.2.2. The estimated parameters are then used in the Viterbi Algorithm (see section 4.2.2) to identify the most likely hidden state sequence.

The estimated HMM parameters (transition matrix $a_{ij}$, emission distribution parameter $\mu_i$ and $\sigma_i$) for the two hidden states are reported in Table 4.3. Next to the estimated parameters are the standard errors, derived from the asymptotic covariance matrix of estimates, which is computed by finite difference approximations. It is clear from Table 4.3 that the volatility level in state 1 is much higher than in state 2. The mean of the state dependent squared return distribution for the 'excited' state ranges between $9.48 \times 10^{-5}$ for GBP to $54.97 \times 10^{-5}$ for BRL. The 'normal' state $r_t^2$ distribution has a range that is lower than for the 'excited' state, in between $0.89 \times 10^{-5}$ for GBP to $2.74 \times 10^{-5}$ for BRL. AUD and BRL have volatility levels distinct from the group, CHF, EUR, CAD and GBP. Of all six currency pairs, BRL is the most persistent and most likely to remain in the same state (high or low volatility) for a long period, whereas AUD tends to stay much longer in the low volatility state than all other other currencies. In terms of the dynamics of the underlying Markov series, the majority of the currencies have a probability $p(q_{t+1} = 1|q_t = 1)$ staying in the excited state lies between $0.318$ (AUD) to $0.369$ (EUR), with the exception of BRL at $0.452$. The probability to remain in the 'normal' state is higher than that for the 'excited' state for all currencies.
\[ \hat{a}_{ij} = p(q_{t+1} = j|q_t = i) \]  
\[ \mathcal{N}_1(\mu_{\sigma^2}) \quad [10^{-5}] \]  
\[ \mathcal{N}_2(\mu_{\sigma^2}) \quad [10^{-5}] \]

<table>
<thead>
<tr>
<th></th>
<th>Transition Matrix</th>
<th>Emission Distribution</th>
</tr>
</thead>
</table>
| AUD      | \[
\begin{pmatrix}
0.318_{(0.024)} & 0.683_{(0.024)} \\
0.146_{(0.008)} & 0.855_{(0.008)}
\end{pmatrix}\] | 31.85_{(2.4)} \quad 2.34_{(0.06)} |
| BRL      | \[
\begin{pmatrix}
0.452_{(0.023)} & 0.549_{(0.023)} \\
0.142_{(0.008)} & 0.859_{(0.008)}
\end{pmatrix}\] | 54.97_{(4.03)} \quad 2.74_{(0.08)} |
| CHF      | \[
\begin{pmatrix}
0.368_{(0.018)} & 0.633_{(0.018)} \\
0.374_{(0.013)} & 0.627_{(0.013)}
\end{pmatrix}\] | 11.2_{(0.39)} \quad 0.99_{(0.03)} |
| CAD      | \[
\begin{pmatrix}
0.341_{(0.021)} & 0.661_{(0.021)} \\
0.21_{(0.01)} & 0.791_{(0.01)}
\end{pmatrix}\] | 11.5_{(0.51)} \quad 0.92_{(0.03)} |
| GBP      | \[
\begin{pmatrix}
0.362_{(0.02)} & 0.639_{(0.02)} \\
0.272_{(0.012)} & 0.729_{(0.012)}
\end{pmatrix}\] | 9.48_{(0.37)} \quad 0.89_{(0.03)} |
| EUR      | \[
\begin{pmatrix}
0.369_{(0.018)} & 0.632_{(0.018)} \\
0.364_{(0.013)} & 0.637_{(0.013)}
\end{pmatrix}\] | 10.29_{(0.33)} \quad 0.91_{(0.03)} |

Table 4.3: Hidden Markov Model Estimates

The estimated transition matrices \( \hat{a}_{ij} = p(q_{t+1} = j|q_t = i) \) for the Markov hidden states are reported with the standard error derived from asymptotic covariance matrix of estimates. In the right column, the estimated state dependent emission distributions for \( r_t^2 \) are reported, with standard error in brackets, for which we assume a normal distribution \( \mathcal{N}_i\), state \( i = 1,2 \). All parameter estimates are statistically significant at the 99% confidence level.

The optimal state sequence associated with a given observation sequence is determined via the Viterbi Algorithm, which is based on dynamic programming methods. The top picture in Figure 4.3 shows the derived state sequence for the EUR time series. The lower picture in Figure 4.3 depicts the associated probability to be in a specific state. The diagram allows for the 'eyeballing' of the goodness of fit for the states. It is clear that there are periods when the identification of the state is less certain than other periods. In the interest of brevity, we omit the diagnostic diagrams for the other currencies.

Figure 4.4 plots the kernel density estimates for \( r_t \) and \( r_t^2 \) distributions of daily
Figure 4.3: State Sequence for the daily EUR/USD exchange rate

The upper graph shows the most likely volatility state sequence as determined by the Viterbi Algorithm for the daily EUR/USD exchange rate. The low volatility state and high volatility state are represented by \{1, 2\}. In the lower graph, the smoothed probabilities of the two states are shown.

EUR/USD. The results are very similar for all the other exchange rates and, hence, they are not presented here. Figure 4.4 shows on the right picture that long tails, both negative and positive, are captured within state 1, the ‘excited’ market state. The smaller positive and negative returns are in state 2, the ‘normal’ market state. Figure 4.4 shows a small overlap of the emission distributions of the two volatility states and a smaller density count for the ‘excited’ state.

Figure 4.4: EUR/USD state dependent return distributions

The left figure shows the state dependent squared returns $r_t^2$ distribution. The right figure shows the return distribution.
In Appendix 4.8 the volatility state sequences for all currencies are shown over
time, with the high volatility states marked in red. The financial crisis of 2008 to
2009 is clearly identifiable as a period when all currency pairs in the sample were
in volatile periods. Other financial markets events are also visible. The crash of
May 2010, the start of the financial crisis in August 2007 and other well known
volatile periods can be observed. The period from August 2006 to May 2007 is an
extended period of tranquillity, in which all currency pairs in the sample had only
a few, short-lived volatile episodes.

4.5 Dependence between Hidden States

In this section, we apply the Shannon entropy to measure the dependencies between
the series. We use the mutual information in section 4.5.1 for measuring the de-
pendence between the states of two series and transfer entropy in section 4.5.2 to
measure the information flow between the states of two series.

We first analyse the entropy and information content for each state series indi-
vidually. Understanding the entropy of the state sequences, especially the entropy
of blocks of the process, are helpful as these form the basic input in the calculation
of mutual information and transfer entropy. They also shed light on how much in-
formation is contained in the past, given the current state of events, can be used
to predict the future of the state sequence. In section 2.3.3, we introduced various
measures to determine the information content of series of random variables, such
as the entropy rate

\[
h_\infty = \lim_{n \to \infty} H(X_0|X_{-1}, \ldots, X_{-(n+1)})
\]

which quantifies the average information needed to predict a future observation given the past. One of the
assumptions of the HMM model is that the hidden states form a Markov process.
Although the hidden state sequence is, generally, not a Markov process, it is inter-
esting to use the entropy measurements of the estimated state sequence to test for
the Markov property.

As the Markov chain is irreducible and aperiodic, the distribution of the \( n \)th

\(^2\)Spreij (2001) proved the conditions under which the hidden state sequence is a Markov process.
random variable tends to a stationary distribution as \( n \to \infty \). In section 2.3.3, we derived the stationary distribution and the corresponding entropy for a Markov process. Given the estimated transition matrices, we can calculate the stationary distribution of the Markov process. In panel A of Table 4.4, the probabilities of the states for the different currency pairs are listed. The relative frequency of the 'excited' state \( \pi_1 \) is lower than that of the 'normal' state \( \pi_2 \) for all currencies. AUD and BRL have the lowest \( \pi_1 \) while EUR has the highest \( \pi_1 \). AUD and BRL also have the highest excess kurtosis in the sample with EUR having the lowest excess kurtosis (see Table 4.2). In our set-up, we stipulated a simple model for squared returns consists of two states with normal distributions. In the case of AUD and BRL, in which the tails are heavier, the 'excited' state’s mean of the emission distribution is drawn more to the tails, capturing the more extreme and less frequent volatility regimes.

The entropy and the entropy rate of the Markov process can be derived from the stationary distribution (see section 2.3.3). Panel B of Table 4.4 lists the entropy of the Markov process \( H(X) \), given the transition probabilities and assuming a stationary distribution. \( \hat{H}(X) \) is the entropy as measured from the state sequence derived through the Viterbi Algorithm. \( H(X) \) is greater than \( \hat{H}(X) \) for all currencies. So the estimated state sequence is more regular and has a lower entropy than its stationary Markovian counterpart. One possible explanation for this is the 'small sample effect', which we mentioned in Chapter 2. The calculated stationary distribution is the theoretical entropy, which does not suffer the bias of the empirical estimate. Another possible explanation for the difference between the two figures is that the estimated state sequence is in fact not a Markov process.

The HMM state series, estimated in the previous section, are assumed to be Markov processes. This implies that the future is independent from the past, given the present. This conditional independence from past histories longer than one day, implies that the entropy of the series can be simplified \( H(X_n|X_{n-1}, \ldots X_1) = H(X_n|X_{n-1}) \). In Panel C of Table 4.4, the entropy rates of the hidden Markov processes \( h_\infty \), assuming stationarity, are listed. The estimated entropy rates \( \hat{h}_\infty = \)
\( \hat{H}(X_0|X_{-1}) \) are slightly lower than the stationary distribution entropy rates, \( H(X) \). This is to be expected given that stationarity is a limiting case for the process for which we only have a finite sample. Also, small sample bias, again, plays a part for the underestimation (Grassberger (2003)). This small sample effect can also be observed when estimating entropies for longer histories. The highest rate associated

| Panel A: Frequencies of Stationary Distribution |
|--------|--------|--------|--------|--------|--------|
| \( \pi_1 \) | 0.177  | 0.206  | 0.241  | 0.372  | 0.366  | 0.299  |
| \( \pi_2 \) | 0.824  | 0.795  | 0.76   | 0.629  | 0.635  | 0.702  |

<table>
<thead>
<tr>
<th>Panel B: Process Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H(X) )</td>
</tr>
<tr>
<td>( \hat{H}(X) )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Entropy Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h_\infty )</td>
</tr>
<tr>
<td>( \hat{H}(X_2</td>
</tr>
</tbody>
</table>

Table 4.4: Entropy rate \( h_\infty \) and sample \( \hat{H}(X_2|X_1) \) in (bit) for the stationary distribution of HMM process.

with CHF is 0.9515 bit, and is close to the maximum of 1. Only EUR has a similar high entropy rate. This relates to the observation that for both currency pairs, the stationary distribution is more equally weighted and hence the states are less predictable.

### 4.5.1 Hidden State Dependence

In Chapter 2, we introduced mutual information \( I(X, Y)_\tau = H(Y) - H(Y|X_\tau) \) as a measure of the dependency between two random variables \( X \) and \( Y \). It measures the average gain in information about \( Y \) from observing \( X_\tau \). We have seen in Chapter 2 that \( I(X, Y) = I(Y, X) \), so \( X \) provides the same amount of information on \( Y \) as \( Y \) does on \( X \). It is in this sense not a directional measure.

We estimate mutual information for lags \( \tau = 0, 1, -1 \), to analyse volatility co-movement (\( \tau = 0 \)), and volatility spillovers (\( \tau = \pm 1 \)). The estimates of finite samples are biased Grassberger (2003). These biases are related to sample size, the size of the state space and the entropy rate of the underlying process. Here, we
Panel A: $\tau = 0$

<table>
<thead>
<tr>
<th></th>
<th>AUD$_{t+\tau}$</th>
<th>BRL$_{t+\tau}$</th>
<th>EUR$_{t+\tau}$</th>
<th>GBP$_{t+\tau}$</th>
<th>CAD$_{t+\tau}$</th>
<th>CHF$_{t+\tau}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUD$_t$</td>
<td>16.611 (1.442)</td>
<td>44.18 (1.512)</td>
<td>43.649 (1.523)</td>
<td>63.445 (1.537)</td>
<td>32.566 (1.532)</td>
<td></td>
</tr>
<tr>
<td>BRL$_t$</td>
<td>6.082 (1.484)</td>
<td>4.979 (1.502)</td>
<td>10.154 (1.506)</td>
<td>1.785 (1.555)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EUR$_t$</td>
<td>74.079 (1.474)</td>
<td>16.987 (1.468)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBP$_t$</td>
<td>25.378 (1.536)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAD$_t$</td>
<td>12.19 (1.475)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHF$_t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel B: $\tau = -1$

<table>
<thead>
<tr>
<th></th>
<th>AUD$_{t+\tau}$</th>
<th>BRL$_{t+\tau}$</th>
<th>EUR$_{t+\tau}$</th>
<th>GBP$_{t+\tau}$</th>
<th>CAD$_{t+\tau}$</th>
<th>CHF$_{t+\tau}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUD$_t$</td>
<td>4.578 (1.447)</td>
<td>3.805 (1.526)</td>
<td>5.673 (1.517)</td>
<td>5.118 (1.542)</td>
<td>0.437 (1.526)</td>
<td></td>
</tr>
<tr>
<td>BRL$_t$</td>
<td>0.035 (1.526)</td>
<td>1.663 (1.509)</td>
<td>2.589 (1.512)</td>
<td>0.001 (1.548)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EUR$_t$</td>
<td>1.127 (1.47)</td>
<td>6.417 (1.53)</td>
<td></td>
<td></td>
<td>0.048 (1.505)</td>
<td></td>
</tr>
<tr>
<td>GBP$_t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.162 (1.509)</td>
<td></td>
</tr>
<tr>
<td>CAD$_t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHF$_t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel C: $\tau = +1$

<table>
<thead>
<tr>
<th></th>
<th>AUD$_{t+\tau}$</th>
<th>BRL$_{t+\tau}$</th>
<th>EUR$_{t+\tau}$</th>
<th>GBP$_{t+\tau}$</th>
<th>CAD$_{t+\tau}$</th>
<th>CHF$_{t+\tau}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUD$_t$</td>
<td>3.947 (1.568)</td>
<td>0.839 (1.529)</td>
<td>7.525 (1.517)</td>
<td>7.305 (1.521)</td>
<td>1.027 (1.526)</td>
<td></td>
</tr>
<tr>
<td>BRL$_t$</td>
<td>0.136 (1.512)</td>
<td>0.664 (1.509)</td>
<td>3.872 (1.489)</td>
<td>0.191 (1.526)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EUR$_t$</td>
<td>0.292 (1.47)</td>
<td>0.173 (1.488)</td>
<td></td>
<td></td>
<td>0.11 (1.522)</td>
<td></td>
</tr>
<tr>
<td>GBP$_t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.197 (1.5)</td>
<td>0.07 (1.495)</td>
</tr>
<tr>
<td>CAD$_t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.552 (1.509)</td>
</tr>
<tr>
<td>CHF$_t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.5: Volatility States Co-Movement and Interdependence

This table shows the estimates for Mutual Information $\hat{I}(X,Y)_{{\tau}} [10^{-3}]$ (in bits) for the hidden state sequences with different time lags. Volatility co-movement ($\tau = 0$) is shown in panel A and volatility states interdependence relations ($\tau = \pm 1$) are shown in panels B and C. The error estimates are shown in brackets. These are produced by generating reshuffled series of one of the series. The 99% quantile is reported $q_{0.99}(\hat{I}(X_{sh},Y))$ [10$^{-3}$]. In panel A, all measured mutual information $\hat{I}(X,Y)$ are significant. The significant measurements in panels B and C are shown in bold.

follow the approach in Chapter 4 to quantifying this bias by calculating the mutual information with one process being reshuffled, and thereby destroying all of its time series patterns. With sufficient repetitions of this reshuffling process, we generate a sample for which the mutual information includes the noise due to small sample. As an estimate of the noise, we take the 99% quantile of the measured mutual information of the reshuffled series $q_{0.99}(\hat{I}(X_{sh},Y))$ and report them in brackets in the table. Panels A, B and C in Table 4.5, show the results for the different lags $\tau = 0, -1, +1$.

We have seen in section 2.3 that mutual information measures the divergence of two random variables from being independent. In this particular setting, with binary valued series, mutual information is a summation of four terms, repre-
senting the possible combinations of values in the two random variables $X, Y$:

$$I(X; Y)_\tau = - \sum_{x,y \in \{1,2\}} p(x_\tau, y) \cdot \log_2 \frac{p(x_\tau, y)}{p(x_\tau)p(y)}.$$ 

Any statistical significant measurement of $\hat{I}(X, Y)_\tau > 0$, implies that the random variables are more often in the same state, high volatility or low volatility, and at the specified time (lagged by $\tau$ in the case of $X$), than if those variables are independent.

For all currency pairs, the mutual information peaks at lag $\tau = 0$, where all estimated values are above the 99% quantile of 'noise'. The strongest observed relationship is noted between EUR and CHF, with a mutual information of $305.02 \times 10^{-3}$ bits. This is approximately five times stronger than that for any of the other currencies. The next strongest co-movement relationship is that between GBP and EUR with $74.079 \times 10^{-3}$ bits. The connection between GBP and CHF with $\hat{I} = 60.996 \times 10^{-3}$ bits completes the triangle strong link of European currencies in the sample. The other strong co-movement relationship outside this triangle of European group is that between AUD and CAD with $63.445 \times 10^{-3}$ bits measured.

For all currency pairs, mutual information dropped sharply when $\tau \neq 0$. This means that a change in volatility regime in one series has the strongest influence on the volatility regime of another series on the same day. Panels B and C show that at $\tau \pm 1$, only 13 of the 30 pairs have mutual information statistically significant at the 99% confidence level. The strongest two-way lead-lag connections are AUD-GBP, AUD-CAD, AUD-BRL, CAD-GBP and, to a lesser extent, CAD-BRL. The weaker one-way connections include AUD$_0$-EUR$_{-1}$, BRL$_0$-GBP$_{-1}$ and EUR$_0$-CAD$_{-1}$.

### 4.5.2 Hidden State Information Flow

While mutual information quantifies the deviation from $X$ and $Y$ being independent, transfer entropy quantifies the deviation from $X$ being determined by its own history only (via conditional probabilities). Unlike mutual information, transfer entropy $T_{Y \rightarrow X}(m, l)$, introduced in Chapter 2 and defined below, is not symmetric and takes into account only the statistical dependencies originating from $Y$ and not those from...
To measure any dependence of volatility states estimated by the HMM-model, the transfer entropy between the hidden state sequence of two series is estimated. The transfer entropy $T_{Y \rightarrow X}$ from $Y$ to $X$ is the degree to which $Y$ makes less uncertain the future of $X$ beyond the degree that $X$ is already giving information on resolving the future of $X$. In this context, a detected information flow from $Y$ to $X$ indicates that past volatility states in $Y$ (up to specified length $l$) contain information, that makes the next day’s volatility state in $X$ more predictable. We will in the present case use the most likely hidden states determined by the Viterbi algorithm.

We use the two-letter alphabet for the volatility state and the Grassberger estimator $\hat{H}$ in equation (3.15) for the estimation of transfer entropy. To calculate the estimation error, we follow Marschinski and Kantz (2002) (as in Chapter 3) and estimate the mean and standard error by the bootstrapping method, involving reshuffling the predictor series, destroying all its time series pattern. This process is repeated enough times, after which a mean, $\mu(\hat{T}_{sh})$, and a standard deviation, $\sigma$ is calculated. The effective transfer entropy (ET) (see Chapter 2 for details) is formed by deducting the mean of the shuffling simulations from the estimated transfer entropy, $ET_{Y \rightarrow X}(m, l) \equiv T_{Y \rightarrow X}(m, l) - \mu(\hat{T}_{sh})$. For $ET_{Y \rightarrow X}(m, l)$, $m$ refers to the amount of own past history of $X$ included in the prediction, whereas $l$ refer to the amount of past history of $Y$ used in predicting $X$. Also, as in Chapter 3, we use the REA ($relative \ explanation \ added$) measure from Marschinski and Kantz (2002) to guage the relative information gain due to the transfer entropy. REA defined below relates the information gain from TE to the entropy of the series:

$$REA(m, l) = \frac{ET_{Y \rightarrow X}(m, l)}{H(X_0|X_{-1}, \ldots X_{-(m)})}$$

Table 4.6 shows that the largest information flow, as measured by the effective transfer entropy $ET$, is that between BRL and CAD with $\hat{ET}_{BRL\rightarrow CAD}(4, 4) =$
21.521 [10^{-3}]. This information flow is also the strongest in terms of REA. It adds
REA = 2.963 [%] to the information already contained in the history of volatil-
ity states of CAD itself. This compares well with the REA = 1.32% reported in
Marschinski and Kantz (2002) for an information flow from the Dow Jones to the
DAX equity index based on one-minute intraday returns over the period from May

The currency pairs BRL, CAD with AUD form a triple group producing the
strongest information flow compared with the whole sample. A direct linkage be-
tween the currency pairs can be clearly seen for AUD and CAD, which have a
higher correlation \( \hat{\rho} = 0.623 \) (Table 4.1) and for which we measured a high mutual
information \( \hat{I}(\text{AUD}, \text{CAD})_{r=0} = 63.445 \) (Table 4.5). Transfer Entropy picks up a
strong information flow between BRL and CAD, with BRL \( \rightarrow \) CAD being stronger
than CAD \( \rightarrow \) BRL with \( \hat{ET}_{\text{CAD} \rightarrow \text{BRL}}(4, 4) = 15.666 \ [10^{-3}] \). So, although the co-
movement between the volatility states of BRL and CAD measured with Mutual
Information is weak, we can detect with transfer entropy a strong flow, pointing to
BRL volatility influencing CAD volatility.

The strong information flow in Europe is among CHF, EUR and GBP with
\( \hat{ET}_{\text{CHF} \rightarrow \text{EUR}}(4, 4) = 19.038 \ [10^{-3}] \) and \( \hat{ET}_{\text{EUR} \rightarrow \text{GBP}}(4, 4) = 15.889 \ [10^{-3}] \). Inagaki
(2006) examines the currency pairs GBP and EUR for volatility spillover in the
1999-2004 period. Consistent with our results, the author found support for a uni-
directional volatility spillover from EUR to GBP, with the EUR having a one-sided
impact on GBP.

In the previous chapter, we investigated the relationship between daily and hourly
draws in GBP and EUR. The transfer entropy measured between volatility states
\( \hat{ET}_{\text{EUR} \rightarrow \text{GBP}}(4, 4) = 15.889 \ [10^{-3}] \) is comparable to our results from chapter 3. For
draws on daily returns we estimated \( \hat{ET}_{\text{EUR} \rightarrow \text{GBP}}(2, 4) = 11.451 \ [10^{-3}] \), with a
shorter history of GBP of two days. This corresponds to findings in Rebonato and
Chen (2009), which points to a link between large draws and excited volatility states.
Rebonato and Gaspari (2006) conjectured that there are at least two regimes (‘nor-
mal’ and ‘excited’) in the US$ interest rate market. Following this insight, Rebonato
and Chen (2009) used a hidden Markov model for interest rates. The authors show that large draws are primarily from excited states, especially if the draws have been short. The excited states exhibit positive auto-correlation (where consecutive bursts exist), and in the ‘normal’ and ‘quiet’ states, returns have negative auto-correlation, with high fractions of reversals in the normal state. In Appendix 4.7 we compare the duration of the series, staying in the same volatility states, and draws in our data, to highlight the connection between large draws and high volatility states.
Table 4.6: Volatility States Information Flows
This table presents the results on information flows between the hidden volatility state series. All combinations of currency pairs are listed by the 'source' process. For all currency pairs, the combination of blocks \((m, l)\) from one to four days past history is selected with the highest \(\hat{E}T\). For each flow \(Y \rightarrow X\), \(m\) blocks of past states in \(X\) are taken into account and \(l\) past states in \(Y\). Transfer Entropy \(\hat{T}\), effective TE \(\hat{ET}\) and the relative explanation added \(REA\) is shown.

<table>
<thead>
<tr>
<th>(Y \rightarrow X)</th>
<th>Block Entropy</th>
<th>Transfer Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{E}T_{Y\rightarrow X}(m, l))</td>
<td>(m)</td>
<td>(T) (bit)</td>
</tr>
<tr>
<td>Panel A: AUD (\rightarrow) X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUD (\rightarrow) BRL</td>
<td>1 4</td>
<td>10.192</td>
</tr>
<tr>
<td>AUD (\rightarrow) CAD</td>
<td>4 4</td>
<td>38.986</td>
</tr>
<tr>
<td>AUD (\rightarrow) CHF</td>
<td>1 2</td>
<td>4.354</td>
</tr>
<tr>
<td>AUD (\rightarrow) EUR</td>
<td>1 4</td>
<td>6.701</td>
</tr>
<tr>
<td>AUD (\rightarrow) GBP</td>
<td>1 4</td>
<td>15.203</td>
</tr>
<tr>
<td>Panel B: BRL (\rightarrow) X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRL (\rightarrow) AUD</td>
<td>4 4</td>
<td>27.868</td>
</tr>
<tr>
<td>BRL (\rightarrow) CHF</td>
<td>4 4</td>
<td>42.537</td>
</tr>
<tr>
<td>BRL (\rightarrow) EUR</td>
<td>1 4</td>
<td>2.314</td>
</tr>
<tr>
<td>BRL (\rightarrow) GBP</td>
<td>4 4</td>
<td>25.327</td>
</tr>
<tr>
<td>Panel C: EUR (\rightarrow) X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EUR (\rightarrow) AUD</td>
<td>3 4</td>
<td>16.152</td>
</tr>
<tr>
<td>EUR (\rightarrow) BRL</td>
<td>1 3</td>
<td>2.468</td>
</tr>
<tr>
<td>EUR (\rightarrow) CAD</td>
<td>1 4</td>
<td>6.311</td>
</tr>
<tr>
<td>EUR (\rightarrow) CHF</td>
<td>4 4</td>
<td>21.894</td>
</tr>
<tr>
<td>EUR (\rightarrow) GBP</td>
<td>4 4</td>
<td>32.756</td>
</tr>
<tr>
<td>Panel D: CHF (\rightarrow) X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHF (\rightarrow) AUD</td>
<td>3 2</td>
<td>1.201</td>
</tr>
<tr>
<td>CHF (\rightarrow) BRL</td>
<td>2 4</td>
<td>7.522</td>
</tr>
<tr>
<td>CHF (\rightarrow) CAD</td>
<td>4 4</td>
<td>25.756</td>
</tr>
<tr>
<td>CHF (\rightarrow) EUR</td>
<td>4 4</td>
<td>31.099</td>
</tr>
<tr>
<td>CHF (\rightarrow) GBP</td>
<td>4 4</td>
<td>27.589</td>
</tr>
<tr>
<td>Panel E: CAD (\rightarrow) X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAD (\rightarrow) AUD</td>
<td>4 4</td>
<td>35.408</td>
</tr>
<tr>
<td>CAD (\rightarrow) BRL</td>
<td>4 4</td>
<td>38.384</td>
</tr>
<tr>
<td>CAD (\rightarrow) CHF</td>
<td>1 4</td>
<td>2.674</td>
</tr>
<tr>
<td>CAD (\rightarrow) EUR</td>
<td>4 4</td>
<td>31.312</td>
</tr>
<tr>
<td>CAD (\rightarrow) GBP</td>
<td>1 4</td>
<td>9.957</td>
</tr>
<tr>
<td>Panel F: GBP (\rightarrow) X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBP (\rightarrow) AUD</td>
<td>1 4</td>
<td>11.395</td>
</tr>
<tr>
<td>GBP (\rightarrow) BRL</td>
<td>4 4</td>
<td>29.678</td>
</tr>
<tr>
<td>GBP (\rightarrow) CAD</td>
<td>1 4</td>
<td>11.946</td>
</tr>
<tr>
<td>GBP (\rightarrow) CHF</td>
<td>1 4</td>
<td>4.863</td>
</tr>
<tr>
<td>GBP (\rightarrow) EUR</td>
<td>1 2</td>
<td>1.586</td>
</tr>
</tbody>
</table>
4.6 Conclusion

In this chapter, we follow Hamilton (1989) to model volatility regimes for a set of currencies quoted against the US dollar. A two state hidden Markov model was defined, with which we modelled high and normal volatility states. Following Chapter 3, we used mutual information and transfer entropy to evaluate various volatility regime relationships between the currency pairs. We found volatility regime co-movement relationships among the European currencies (EUR/USD, GBP/USD, CHF/USD). CAD/USD and AUD/USD exhibit a volatility co-movement relationship and, to a lesser extent, interdependency with a time lag. We also find evidence for information flows from EUR/USD to GBP/USD, which indicates a causal volatility spillover relationship. This finding is in line with findings in Inagaki (2006), reported a unidirectional volatility spillover from EUR to GBP in the period 1999-2004. The work showcases also the usefulness of the concepts of mutual information and transfer entropy when applying them to analyse volatility co-movements and spillover relationships among volatility states.

4.7 Appendix - State Duration

In this appendix, we define drawdowns (drawups) and draws, as in Chapter 3. All draw statistics for the sample are summarized in Table 4.7. Table 4.7 shows $E[D], \sigma[D], E[U]$ and $\sigma[U]$, which are the mean and standard deviations of the drawdowns and drawups. BRL stands out with a higher mean and volatility than the others and an asymmetry between drawdowns and drawups. $E[d], \sigma[d], E[u], \sigma[u]$ gives the average daily price drop within a drawdown (drawup) and the standard deviation. $E[D]/E[d] = E[l_d]$ is the ratio of the average drawdown magnitude to the average daily price drop in a drawdown. In the case of $iid$, Gaussian increments, this equals 2, a result which is valid even in the case where the time series lacks homoskedasticity. A value below 2 represents a thin tail distribution, and a value above 2 represents thick tail distribution. Table 4.2 shows $E[l_u] > E[l_d]$ for all cur-
Panel A: Draws

\[
\begin{align*}
E[D]/E[d] &= E[l_d] 
E[U]/E[u] &= E[l_u] 
E[D] &= 0.012 \quad 0.016 \quad 0.008 \quad 0.01 \quad 0.01 \\
\sigma(D) &= 0.014 \quad 0.023 \quad 0.009 \quad 0.009 \quad 0.01 \\
E[U] &= 0.013 \quad 0.016 \quad 0.009 \quad 0.01 \quad 0.01 \\
\sigma(U) &= 0.013 \quad 0.017 \quad 0.009 \quad 0.01 \quad 0.01 \\
\end{align*}
\]

Panel B: Draw Distribution

\[
\begin{align*}
\hat{z}(D) &= 0.993 \quad 0.859 \quad 1.026 \quad 1.163 \quad 1.072 \quad 1.03 \\
\sigma_{\hat{z}}(D) &= 0.023 \quad 0.02 \quad 0.023 \quad 0.025 \quad 0.023 \quad 0.024 \\
\hat{\chi}(D) [10^{-3}] &= 11.6 \quad 14.4 \quad 8.1 \quad 10.1 \quad 9.7 \quad 8.9 \\
\sigma_{\hat{\chi}}(D) [10^{-3}] &= 0.4 \quad 0.5 \quad 0.3 \quad 0.3 \quad 0.3 \quad 0.3 \\
\hat{z}(U) &= 1.104 \quad 1.013 \quad 1.087 \quad 1.088 \quad 1.132 \quad 1.096 \\
\sigma_{\hat{z}}(U) &= 0.025 \quad 0.023 \quad 0.023 \quad 0.025 \quad 0.027 \quad 0.024 \\
\hat{\chi}(U) [10^{-3}] &= 12.7 \quad 15.4 \quad 8.8 \quad 10.4 \quad 10.1 \quad 9 \\
\sigma_{\hat{\chi}}(U) [10^{-3}] &= 0.4 \quad 0.5 \quad 0.3 \quad 0.3 \quad 0.3 \quad 0.3 \\
\end{align*}
\]

<table>
<thead>
<tr>
<th>Currency Pair</th>
<th>AUD</th>
<th>BRL</th>
<th>CAD</th>
<th>CHF</th>
<th>EUR</th>
<th>GBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E[D]/E[d] = E[l_d])</td>
<td>1.827</td>
<td>2.036</td>
<td>1.874</td>
<td>1.84</td>
<td>1.877</td>
<td>1.958</td>
</tr>
<tr>
<td>(E[U]/E[u] = E[l_u])</td>
<td>2.12</td>
<td>2.113</td>
<td>1.984</td>
<td>1.894</td>
<td>1.932</td>
<td>2.008</td>
</tr>
<tr>
<td>(E[D])</td>
<td>0.012</td>
<td>0.016</td>
<td>0.008</td>
<td>0.01</td>
<td>0.01</td>
<td>0.009</td>
</tr>
<tr>
<td>(\sigma(D))</td>
<td>0.014</td>
<td>0.023</td>
<td>0.009</td>
<td>0.009</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>(E[U])</td>
<td>0.013</td>
<td>0.016</td>
<td>0.009</td>
<td>0.01</td>
<td>0.01</td>
<td>0.009</td>
</tr>
<tr>
<td>(\sigma(U))</td>
<td>0.013</td>
<td>0.017</td>
<td>0.009</td>
<td>0.01</td>
<td>0.01</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Table 4.7: Summary Statistics
Statistics of drawdowns and drawups for all sample currencies.

Except for AUD and BRL, many currency pairs have \(E[l_d]\) lower than 2. Interestingly for interest rates, this measure is consistently lower than the iid Gaussian value of 2, as shown in Rebonato and Gaspari (2006).

Johansen and Sornette (2001) show that draws follow a modified exponential distribution

\[ P_{se}(D) = P(D = 0)e^{-\left(\frac{|D|}{\varphi}\right)^z} \]

where \(z > 1\) represents lower kurtosis and \(z < 1\) higher kurtosis. Table 4.7 show that most currencies have a \(z\) close to 1, with the exception of BRL/USD drawdowns (\(\hat{z} = 0.859\)) indicating a high kurtosis for drawdowns.

Given the transition probabilities of the hidden state sequence, we can calculate the probability of a sequence of same states with a given length to occur. The probability of a sequence of length \(n\) to stay in state \(i\), \(q_k = i, \ldots , q_{k+n-1} = i\) is \(a_{ii}^{n-1} \cdot (1 - a_{ii})\). The expected duration of a sequence in state \(i\) is then \(\sum_{n \geq 0} n \cdot a_{ii}^{n-1} \cdot (1 - a_{ii}) = 1/(1 - a_{ii})\).

For a transition probability of \(\frac{1}{2}\), which is the example of a fair coin, we have, on average, two consecutive heads or tails being tossed. For the six currency pairs, the
expected times of remaining in a certain state are listed in the Table 4.8.

<table>
<thead>
<tr>
<th></th>
<th>AUDUSD</th>
<th>BRLUSD</th>
<th>CADUSD</th>
<th>CHFUSD</th>
<th>GBPUSD</th>
<th>EURUSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>State 1</td>
<td>1.466</td>
<td>1.824</td>
<td>1.517</td>
<td>1.582</td>
<td>1.567</td>
<td>1.584</td>
</tr>
<tr>
<td>State 2</td>
<td>6.859</td>
<td>7.048</td>
<td>4.781</td>
<td>2.681</td>
<td>3.689</td>
<td>2.754</td>
</tr>
</tbody>
</table>

Table 4.8: Expected Time State Duration

The distribution of state sequences staying in a given state is depicted in Figure 4.5 for state 1 and in Figure 4.6 for state 2. The sequences remaining in one state follow the theoretical predictions. For longer periods there are marked deviations to the theoretical predictions for the underlying Markov sequences. The duration

Figure 4.5: Duration of sequence remaining in the high volatility state 1

Figure 4.6: Duration of sequence remaining in the low volatility state 2

of same state sequences relates to the question of whether the HMM model is able to capture large drawdowns and large drawups over different lengths within such a sequences.
## 4.8 Appendix - Currency State Series

<table>
<thead>
<tr>
<th>Month</th>
<th>Jan-06</th>
<th>Feb-06</th>
<th>Mar-06</th>
<th>Apr-06</th>
<th>May-06</th>
<th>Jun-06</th>
<th>Jul-06</th>
<th>Aug-06</th>
<th>Sep-06</th>
<th>Oct-06</th>
<th>Nov-06</th>
<th>Dec-06</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAD/USD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBP/USD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHF/USD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The table above contains currency state series data for CAD/USD, GBP/USD, and CHF/USD from January to December of 2006.
Chapter 5

Volatility Timing of Hedge Funds and Entropy

Abstract

This study uses the approximate entropy of S&P realised volatility to detect changes in volatility regime in order to assess the market volatility timing and investing strategies of US equity focus hedge funds. A one-factor model is used, conditioned on the entropy of market volatility, to measure the dynamic of hedge funds equity exposure. On a cross section of 1,903 hedge funds, we find that, over the period from 2000 to 2014, these hedge funds adjust their exposures dynamically in response to changes in volatility regime. This adds to the literature on the volatility timing behaviour of hedge fund managers, but using entropy as a model independent measure of volatility regime change.
5.1 Introduction

Hedge fund managers are in general free to change trading strategies, to allocate capital to asset classes and to choose leverage. Hedge fund managers often reallocate their capital dynamically in response to a change in market environment. To pursue their investment goals, hedge fund managers scan the markets for signals and information that can give support to their analyses regarding the current and the future states of the market. From the information gathered, the fund manager forms an expectation of the future states of the market and develops a trading strategy to make profitable use of this knowledge. Absolute and relative price levels, market volatility and trends are information that play an important role in asset allocation and portfolio selection.

A high degree of irregularity and unpredictability in market prices and volatility inhibits the managers’ ability to predict future development of the markets. Fund managers are more likely to choose lower levels of equity exposure when the market is in a state of heightened irregularity where no clear view can be formed on expected returns or volatility. In this case, ceteris paribus, the fund managers may respond by actively reducing the capital allocation of investments in risky assets or delaying further investments in equity markets. Conversely, fund managers will seek more (long or short) exposure to the market in an environment with higher regularity and predictability.

We extend the market timing model of Ferson and Schadt (1996), which assumes funds’ equity risk exposure is a linear function of the market risk factors. Various modifications to this baseline model are tested using the approximate entropy (ApEn) of market volatility series as a state variable of the market condition. ApEn captures serial correlation and it is sensitive to changes in market volatility. A drop in entropy could signal volatility regime change. A switch to high volatility state will prompt fund asset allocation response and severely curtail the manager’s ability to future market movements. In response to a change in volatility regime, the manager may increase or reduce equity risk exposure by adjusting the funds’ market
beta. Volatility is an important determinant for the level of market beta a fund is choosing. For instance, for funds wishing to keep their funds volatility stable over time targeting at a specific Sharpe ratio, the managers are likely to reduce the fund’s beta when a switch to higher volatility regime is detected.

Hedge fund managers make good candidates for the study. They are lightly regulated investment managers who have a lot more flexibility in implementing an investment strategy than mutual funds managers. Brunnermeier (2004) found hedge fund managers are among the most sophisticated investors, who dynamically adjusting their portfolio to changing markets. We use a cross section of hedge funds from Lipper Tass over the sample period from 2000 to 2014, who have a geographic focus on the US equity markets. As a volatility variable, we use the realized volatility based on five-minute price ticks and aggregated to the daily figure, as well as the VIX index for market implied volatility.

The current analysis adds to the market volatility timing research on hedge funds. Research on fund managers’ volatility timing ability has been thus far concentrated on mutual funds, whereas our focus is on hedge funds. We will demonstrate the use of ApEn in detecting changes in the time series structure of volatility. Simulations based on popular econometric models of volatility clearly show the usefulness of ApEn measure in distinguishing different volatility patterns and in detecting volatility regime changes. On a sample of 1,903 hedge funds, we find that managers adjust the funds’ market beta as volatility is settling into a new regime. Dead funds show no such market beta adjustment skill, differing from live funds where such an adjustment ability was noted. The results are consistent with the findings in Busse (1999) on mutual funds. The results also show the usefulness of entropy as a model independent measure of changes in volatility regime and patterns.

There is an extensive literature on market timing models and conditional performance evaluation of actively-managed investment vehicles. Treynor and Mazuy (1966) were among the first to study whether investment manager were able to profitably time changes in market conditions. The authors formulated a model to test the timing ability, extending the CAPM by a quadratic term to capture the
nonlinear exposure to the market, depending on an expected high or low market return. In the sample of 57 mutual funds, the hypothesis of market timing ability was accepted for just one fund. Henriksson and Merton (1981) formulated a market timing model where the manager forecasts market excess return over the risk-free rate. Depending on the forecast, the manager adjusts the portfolio’s beta. This is modelled as an option payoff on the market return, with the strike price equal to the risk-free rate. On the basis of this model, the authors developed parametric and non-parametric tests for an investment manager’s forecasting skill. Ferson and Schadt (1996) formulated a model of market timing where the risk exposure evolves linearly with the market observables. In their analysis, various variables that are public information are used as market conditions. Using a monthly data set of mutual funds, the authors found evidence for several variables of a change in risk exposure in response to public information on the economy. Furthermore, they were able to show that an unconditional version of the Treynor and Mazuy (1966) model and the Henriksson and Merton (1981) models are misspecified when applied to naïve strategies, in which conditional market timing models are an improvement.

In measuring performance, researchers have attempted to distinguish market timing ability from stock selection skill. These skills are not easy to separate. Korajczyk and Jagannathan (1986) found that if managers trade options or levered securities, evidence for market skill is spurious. The authors construct portfolios that ”show artificial timing ability when no true timing ability exists”. To test for market timing ability, Jiang, Yao, and Yu (2007) used mutual funds holdings information to avoid spurious findings. On the basis of the holdings based test, the authors were able to find positive market timing ability in US equity mutual funds. Using daily information, Bollen and Busse (2001) showed that mutual funds may possess market timing ability. The higher frequency information adds to the power of the tests used. Unfortunately, higher frequency information on a fund’s performance is unavailable for hedge funds. Using modified versions of various market timing models with daily conditioning variables, Patton and Ramadorai (2013) found evidence of varying hedge fund risk exposures across and within months. Market timing ability
has also been investigated in relation to fund managers investing in bond markets, as in the study by Chen, Ferson, and Peters (2010).

Timing market volatility has also been actively researched. Busse (1999) uses daily returns of mutual funds to investigate a fund’s ability to time market volatility. His findings suggest that funds decrease market exposure when market volatility is high. The value of this timing ability is in higher Sharpe ratios for successful volatility timing managers. The author further shows that surviving funds show sensitivity to market volatility, whereas non-surviving funds do not show this sensitivity. The economic value of volatility timing for investors has been investigated by Fleming, Kirby, and Ostdiek (2001). Using conditional analysis on short horizon asset allocation strategies, the authors found volatility timing strategies outperform unconditional static strategies that have the same target return and volatility.

Various information theoretic quantities have been used to measure and analyse the irregularity, patterns and serial structure of time series. Pincus (1991) used approximate entropy (ApEn) to investigate and classify systems in terms of their complexity. ApEn is a family of measures that is able to differentiate varying degrees of serial correlation within a time series. For specific stochastic processes, such as stationary processes, analytical formulas for the calculation of ApEn are shown (see Pincus (1991)). In a series of papers, including Pincus (1995), Pincus and Huang (1992) and Pincus (1992), various statistical qualities of ApEn have been investigated. So far there have been few applications of ApEn to address economics research questions. Pincus (2008) showcases the usefulness of ApEn as a model independent measure of serial structure and irregularity in financial and economic time series. Duan and Stanley (2010) investigate empirical changes in patterns of volatility and the degree of predictability using daily data of equity, foreign exchange, commodity and interest rates. The authors found that ApEn is useful for measuring predictability of a time series and found that longer sample period increases the degree of predictability. The investigated time series exhibited a entropy value different from that of a random walk, from which the authors concluded that the financial time series is to some degree predictable.
5.2 Entropy of Volatility

Various information measures have been introduced to enhance our understanding of stochastic time series. In section 2.4, we introduced approximate entropy, $\text{ApEn}(m, r, N)(u)$, as a measure for the regularity or dependence structure of a time series. Let $u$ be a sequence of numbers $u(1), \ldots, u(N)$ of length $N$. Given a non-negative number $m$ with $m \leq N$, we form $m$-block of subsequences $x(i) \equiv (u(i), u(i+1), \ldots, u(i+m-1))$. The distance between two blocks is measured by $d(x(i), x(j)) \equiv \max_{k=1,2,\ldots,m}|u(i+k-1) - u(j+k-1)|$. For a given block $x(i)$ we count the fraction of block $x(j)$ that have a distance of less than $r$ to the given block $C^m(r)$, to quantify the regularity of a particular pattern. With $\Phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log C^m(r)$ the approximate entropy $\text{ApEn}(m, r, N)$ is defined as follows.

$$\text{ApEn}(m, r, N)(u) = \Phi^m(r) - \Phi^{m+1}(r), \quad m \geq 0 \quad (5.1)$$

$$\text{ApEn}(0, r, N)(u) = -\Phi^1(r) \quad (5.2)$$

We proved the following formula in section 2.4, which supports the intuition on $\text{ApEn}$.

$$-\text{ApEn}(m, r, N)(u) \approx \frac{1}{N-m} \sum_{i=1}^{N-m} \log \left( \frac{C^{m+1}(r)}{C^m(r)} \right) \quad (5.3)$$

Equation (5.3) shows that $\text{ApEn}$ measures the logarithmic likelihood that sequences of patterns that are close for $m$ observations will remain close for the next $m$ observations. $\text{ApEn}$ reflects persistence, serial correlation and regularity. Lower $\text{ApEn}$ values correspond to higher persistence, stronger dependence and predictability, and vice versa for higher $\text{ApEn}$ values.

Here, we are concerned with the dynamics of volatility. The following sub-sections will showcase the behaviour of approximate entropy of volatility using models that are well known for capturing the salient features of financial markets volatility. Although it is possible to deduct analytically tractable expressions for the approximate entropy of specific processes (see equation (2.34)), this is not, in general, the case...
for the random processes we are concerned with in this analysis. Therefore, we will use simulations to examine how the entropy of volatility changes with respect to changing volatility level and persistence.

In line with Pincus and Kalman (2004), the threshold \( r \) is set equal to \( 0.2\hat{\sigma} \) with \( \hat{\sigma} \) being the sample standard deviation of the series being investigated. If the observed series is volatility, then \( \hat{\sigma} \) is the volatility of volatility. The chosen threshold value strikes a balance between statistical validity and the amount of data available. A larger \( r \) achieves a smaller standard error and a lower bias, but at the expense of losing much information on the data generating process. By normalising \( r \) using the estimated volatility \( \hat{\sigma} \) of the time series, ApEn becomes a scale and translation invariant measure. This makes ApEn a clean measure for irregularity complementing a measure for variability, such as standard deviation. Constructed in this way, one can compare irregularity in data sets that have very different volatilities.\(^1\)

5.2.1 GARCH(1,1)

Since the introduction of the General Auto Regressive Conditional Heteroscedasticity GARCH\((p, q)\) model by Baillie and Bollerslev (1989), it has become one of the most widely used volatility models for financial time series. The GARCH model is easy to estimate and it captures the salient features of volatility, namely clustering and mean reversion, as well as fat tailed asset returns distribution. The most popular choice of the GARCH\((p, q)\) specification has been the GARCH(1,1) model, defined as follows:

\[
\sigma_t^2 = \alpha_0 + \alpha_1 \cdot a_{t-1}^2 + \beta_1 \cdot \sigma_{t-1}^2
\]

where \( a_t = r_t - \mu_t \) and \( \sigma_t \cdot \epsilon_t \) is the stock return innovations at time \( t \) for return series \( r_t \) with mean \( \mu_t \).\(^2\)

To determine the ‘regularity’ of volatility from a GARCH(1,1) process, one can

---

\(^1\)Again, it is worth pointing out that the time series in this chapter to which we apply ApEn to are various volatility measures. Since we are interested in the regularity of volatility; all the ‘volatility’ mentioned here is in fact ‘volatility of volatility’.  
\(^2\)The reader is reminded that that \( r_t \) with a time subscript denotes returns, whereas \( r \) without a subscript denotes the threshold of the approximate entropy measure.
calculate ApEn($m = 1, r, N$) of the squared innovations $\{a_t^2\}$. Assuming $|a_{t+j-1}^2 - a_{t-1}^2| < r$, we have from equation (5.4),

$$|\sigma_{t+j}^2 - \sigma_t^2| = |\alpha_0 + \alpha_1 a_{t+j-1}^2 + \beta_1 \sigma_{t+j-1}^2 - (\alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2)|$$

$$= |\alpha_1 (a_{t+j-1}^2 - a_{t-1}^2) + \beta_1 (\sigma_{t+j-1}^2 - \sigma_{t-1}^2)|$$

$$\leq \alpha_1 |a_{t+j-1}^2 - a_{t-1}^2| + \beta_1 |\sigma_{t+j-1}^2 - \sigma_{t-1}^2|$$

$$\leq \alpha_1 r + \beta_1 |\sigma_{t+j-1}^2 - \sigma_{t-1}^2|$$

Since $a_{t+j}^2 = \sigma_{t+j}^2 \epsilon_{t+j}^2$,

$$|E(a_{t+j}^2) - E(a_t^2)| = |\sigma_{t+j}^2 - \sigma_t^2|$$

(5.5)

$$\leq \alpha_1 r + \beta_1 |\sigma_{t+j-1}^2 - \sigma_{t-1}^2|$$

(5.6)

If $|\sigma_{t+j-1}^2 - \sigma_{t-1}^2| < r$, then $|E(a_{t+j}^2) - E(a_t^2)| \leq \alpha_1 r + \beta_1 r < r$ as $\alpha_1 + \beta_1 < 1$. Equation (5.6) provides the conditions for a high likelihood that a succeeding pair of squared innovations to be close, when the preceding squared innovations of a GARCH process are close $|a_{t+j-1}^2 - a_{t-1}^2| < r$. If the likelihood is high, then the entropy of $\{a_t^2\}$ is low, which means that the conditional volatility shows more regularity, that is, a higher degree of serial correlation.

To study how the parameters, $\alpha_1$ and $\beta_1$ impact on ApEn of squared returns, we ran Monte Carlo simulations on GARCH(1,1) processes with various $\alpha_1$ and $\beta_1$ parameters to produce a series with 400 time steps. The mean and standard error of the simulated ApEn($m = 1, r = 0.2\sigma, N = 400$) were recorded. The ApEn values are shown in Figure 5.1 in a contour plot for all parameter combinations. Similar levels of ApEn are connected with lines. The upper right-hand corner of the plot is empty as only parameter combinations with $\alpha_1 + \beta_1 < 1$ are considered.

The ApEn range from a low of 0.29 bit at $(\alpha_1, \beta_1) = (0.9, 0.05)$ to a high of 1.25 bit at $(\alpha_1, \beta_1) = (0.05, 0.7)$. The recorded maximum, $1.26 \pm 0.08$ bit, matches that of a Brownian motion using the same ApEn parameters ($m = 1, r = 0.2\sigma, N = 400$). The lower (higher) ApEn levels are associated with higher (lower) $\alpha_1$, the auto
ApEn($m = 1, r = 0.2\sigma, N = 400$) measured for $r_t^2$, where returns $\{r_t\}$ are generated from GARCH(1,1) process with normal distributed shocks, and varying combination of $\alpha_1$ (x-axis), $\beta_1$ (y-axis) with constraint $\alpha_1 + \beta_1 < 1$. Same levels of ApEn are connected by a contour line and graphed in different colours, where white corresponds to the highest ApEn values measured.

regressive component of the squared shocks. Furthermore, $\alpha_0$, which reflects the overall unconditional variance of the GARCH process in equation (5.6), has no impact on ApEn.

### 5.2.2 Changing Volatility Level

We now turn our attention to the ability of ApEn to detect a change in the level of volatility in a time series. By this we mean a jump up or down in the level of volatility in the data generating process, similar to the empirically observed alternating high and low volatility clustering. We perform a simple numerical experiment which gives us insight into the dynamics of entropy when the level of volatility is shifting. A time series of 1000 time steps is simulated based on equation (5.7) below:

$$
  r_t = \begin{cases} 
    r_t \in \mathcal{N}(0, \sigma_1^2) & t \leq 500 \\
    r_t \in \mathcal{N}(0, \sigma_2^2) & t > 500 
  \end{cases}
$$

(5.7)
The data generating process has a predetermined jump from a lower to a higher volatility $\sigma_1 < \sigma_2$ at the middle of the time series. ApEn is measured on the squared returns over the rolling time window: $\text{ApEn}(m = 1, r = 0.2\sigma, N = 100)$. As shown in the top panel (i) of Figure 5.2, for $t \leq 500$, ApEn over the rolling window is stable at a level of $1.22 \pm 0.009$ bit. This is also the case for $t > 601$, when the entire rolling window is in the high volatility state. For $500 < t \leq 600$, when the rolling time window includes the point where the switch in volatility occurs, the ApEn falls to a low of 0.22 bit at $t = 520$. It is clear that the threshold $r = 0.2\sigma$ used in the calculation of ApEn is sensitive to the sample standard deviation of the series (volatility of squared return in this case), and shifted upwards as the volatility (of volatility) increases. With a higher threshold, the likelihood of the observations (squared returns in this case) to stay close becomes higher for the lower volatility part of the return series, lowering the entropy. The volatility of squared returns is measured by the kurtosis, the fourth moment of the data generating process. In top panel (i) of Figure 5.2, the (standardised) kurtosis is plotted over time. It follows closely the shape of entropy in response to the step change in volatility: a sharp drop when crossing $t = 500$, followed by a slow increase. This is not surprising as, apart from the step change in volatility, the data generating process is a random walk with iid white noise. The process is not meant to exhibit any persistence.

A similar picture emerges when the switch is from a high volatility state to a low volatility state as shown in the bottom panel, panel (ii), of Figure 5.2. Unlike the previous case, the increase in ApEn and kurtosis is characterised by a slower change, followed by a sharper reversion to the mean level.

As mentioned above, the change in ApEn during a volatility regime change can be partly attributed to the change in the higher moments of the distribution in the rolling window. To understand the extent to which the serial correlation within the two regimes in this scenario plays a role, we set up an additional simulation, in which we take samples from a mixture of normal distributions. Specifically, we run numerical simulations, where we draw from a mixture normal distribution $w_1 \cdot \mathcal{N}(0, \sigma_1^2) + w_2 \cdot \mathcal{N}(0, \sigma_2^2)$, with varying weights $w_1 + w_2 = 1$. The normal distributions
have the same specification as in the previous experiment and the approximate entropy parameters are the same: $m = 1, N = 100, r = 0.2\sigma$. According to the different weights, a sample path is produced by (uniformly) randomly selecting one of the normal distributions to draw from, thus producing a time series of squared returns which are drawn from the same distributions in the same weighted manner as the volatility regime experiment specified in equation (5.7).

Figure 5.3 shows both the result from this numerical experiment and the result from the volatility regime change experiment. The rolling time window starts and ends with a sample of 100 data points drawn from either one of the normal distributions. In between, the rolling window moves from the high volatility regime to
Figure 5.3: ApEn: Regime change versus mixed normal distribution.
Mean of ApEn\((m = 1, N = 100, r = 0.2\hat{\sigma})\) from monte carlo simulations of a mixed normal distribution \(w_1 \cdot \mathcal{N}(0, \sigma_1^2) + w_2 \cdot \mathcal{N}(0, \sigma_2^2)\) (line 'Mixed') and for the rolling window of a regime change (line 'Dynamic') as in Figure 5.2 (ii). The average of the estimated ApEn over 1000 simulations is shown, with the grey band around the mean ApEn of the mixed distribution showing the standard error. The mixing weights range from \(w_2 = 1\), so drawing only from the high volatility distribution, to \(w_1 = 1\) (all low volatility draws) on the right end. The rolling window of the dynamic regime change simulation runs accordingly from a rolling window with only high volatility squared returns, to one with only low volatility squared returns.

The graph depicting the mixture distribution ('Mixed' in Figure 5.3) starts at the left, with weights from \(w_1 = 0\), \(w_2 = 1\) (so selecting only from the higher volatility normal distribution) and ends with a mixture of \(w_1 = 1\), \(w_2 = 0\) at the right end of the figure. One can see that the ApEn paths follow the same shape, but with the regime change time series 'Dynamic' following a more pronounced path. The lowest entropy in the mixture simulation is \(\text{ApEn} = 0.882577 \pm 0.1634\) bit, measured at \((w_1 = 0.67, w_2 = 0.33)\). The main difference between the two time series is the serial correlation structure in the volatility regime change example, which is absent in the simulations from a static mixed distribution. ApEn measures the log likelihood of two squared returns being close, conditional on the preceding two squared returns are also close. Drawn from a random choice of two distributions in the mixed distribution, the likelihood of staying close is only dependent on the shape of the distribution and not from the previous
draw. The case of a volatility change has a serial structure when the change is within
the rolling window.

5.2.3 Markov Switching Volatility

Hamilton (1989) introduced a popular stochastic volatility model that allows the
variance of returns to switch between discrete levels and the switching is deter-
mined by a hidden Markov chain. This specification retains from the GARCH(p,q,)
model the important feature of volatility clustering but offers the added flexibility of
different volatility persistence levels at different states. For a model with two states,
denoted as \( h_n = 1, 2 \), and normally distributed white noise in each state, the return
process can be written as

\[
P(r_t | h_t = i) \sim \mathcal{N}(\mu_i, \sigma_i) \quad i = 1, 2
\]

The switching between the two states is driven by a first order Markov pro-
cess in which the transition probability between the states, \( P(h_{t+1} | h_t) \), is given
by the following matrix, the elements of which are the transition probability
\( T_{i,j} = P(h_{t+1} = i | h_t = j) \),

\[
T = \begin{pmatrix}
P(1|1) & P(2|1) \\
P(1|2) & P(2|2)
\end{pmatrix} = \begin{pmatrix}
p & 1 - p \\
1 - q & q
\end{pmatrix}
\]

(5.9)

We simulate and estimate the impact of transition probabilities on the ApEn
measure. The time series of returns are drawn from a normal distribution in the nu-
umerical experiment, with the state dependent volatility controlled by the transition
matrix. In general, if the finite-state Markov chain is irreducible and aperiodic, the
stationary distribution is unique. Given any starting distribution, the distribution
of \( h_t \) tends to the stationary distribution as \( t \to \infty \) which can be easily calculated
given the transition matrix. We simulated a sample length of 400 data points using
various specifications for the transition matrix in equation (5.9) and non-zero transi-
tion probabilities. State 1 is the low volatility state with $\sigma_1 = 0.1$ and state 2 is the high volatility state with $\sigma_2 = 0.5$. The estimated $\text{ApEn}(m = 1, r = 0.2\hat{\sigma}, N = 400)$ for the squared returns, presented in Figure 5.4, range from a low of $0.8075 \pm 0.0933$ bit at $(p, q) = (0.9, 0.8)$ to a high of $1.2047 \pm 0.0883$ bit at $(p, q) = (0.1, 0.9)$.

Figure 5.4: $\text{ApEn}$: Markov switching process
Numerical simulation of Markov process with combinations of transition probabilities $(p, q)$ in equation (5.9). For each combination, the mean $\text{ApEn}(1, r = 0.2\hat{\sigma}, N = 400)$ of the numerical simulations is reported.

The maximum $\text{ApEn}$ of a Markov switching process is associated with a low probability of switching from the high to the low volatility state ($P(1|2) = 1 - q = 0.1$) and a high probability of the reverse transition ($P(2|1) = 1 - p = 0.9$). For larger sample sizes, where the distribution of states tends to the equilibrium, such a combination of transition probabilities is associated with a process that is often in high volatility state $\frac{P(2|1)}{P(1|2) + P(2|1)} = 0.9$ and seldom in the low volatility state $\frac{P(1|2)}{P(1|2) + P(2|1)} = 0.1$. Similarly, a low $\text{ApEn}$ is to be expected for high probability of low volatility state. This is consistent with the pattern of $\text{ApEn}$ response to a change in volatility shown above in Figure 5.2.
5.3 Data

Monthly hedge fund data, such as return performances, investment style and assets under management (AuM) have been extracted from the Lipper Tass database for the January 2000 to December 2014 period. The Lipper database constitutes one the most extensive hedge fund data sources and has been widely used in academic research. Since 1994, the database also retains data on funds that went out of business, which we also include in our analysis. The information provided by any hedge fund to Lipper, or any other database, is self-reported, without any legal or regulatory obligation involved. The motivation for self-reporting is purely to get added recognition and legitimacy. As such, it gives rise to various biases in the data (Fung and Hsieh (2000)). Hence, to correct for the well-documented survivorship and back fill biases, we select all funds with a minimum of 24 months reported performances and delete the first 12 months of fund performances at the time when the fund started self-reporting.

Only hedge funds are selected and not mutual funds. We further narrowed the selection to hedge funds that have the United States as the self-reported geographical area of focus. Volatility timing is most prevalent in equity markets. Consequently, we exclude all fixed income, credit or currency focus funds, which account for 334 of the 2,704 US focus hedge funds. Given the equity focus, the set of factors that are of potential relevance in explaining fund performance is reduced. Our final data set, after adjusting for the aforementioned biases and restrictions, contains 1,903 hedge funds.

The top diagram in Figure 5.5 shows the number of live funds in the sample over time. A peak is reached in 2005 followed by a significant drop due to consolidation in the industry and the effect of the financial crisis of 2007, when a large number of funds went out of business. The bottom diagram of Figure 5.5 shows the total amount of AuM in million USD managed by the funds included in the sample. Again, the financial crisis is clearly identifiable, marked with a substantial drop in AuM occurring in 2007 due to the double effects of capital outflows and losses on
investments. Since then, there has been a continuous rise until the second half of 2014. Increased financial market volatility, which began in late Q3 2014 and was further heightened in Q4, resulted in redemptions by hedge fund investors. This capital outflow is noticeable by the drop in AuM and number of live funds at the end of 2014, as shown in Figure 5.5.

Figure 5.5: Hedge Funds: AuM, live funds

Hedge funds from January 2000 to December 2014
(i) The number of live funds

(ii) Assets under management (AuM, $ million)

In our sample, 753 funds have been liquidated at some point during the sample period, while 1,150 funds remain going concerns. The most popular subcategory, ‘Long-Short Equity’ accounts for 801 funds (including funds of hedge funds) in the sample. The summary statistics of the monthly returns are reported in panel A of Table 5.1. The average monthly return for all funds is 0.455% and the ‘Long-Short Equity’ subcategory has a slightly higher mean return of 0.578%. The live funds have a higher average monthly return than that of the dead funds, but they also

---

have a much higher standard deviation and a much longer left tail.

<table>
<thead>
<tr>
<th>Panel A: Fund Returns</th>
<th>N</th>
<th>Mean</th>
<th>Std</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Funds</td>
<td>1903</td>
<td>0.455</td>
<td>8.762</td>
<td>-4.33</td>
<td>-1.18</td>
<td>0.5</td>
<td>2.1</td>
<td>4.68</td>
</tr>
<tr>
<td>Long-Short Equity</td>
<td>801</td>
<td>0.578</td>
<td>5.256</td>
<td>-4.18</td>
<td>-1.26</td>
<td>0.59</td>
<td>2.46</td>
<td>5.15</td>
</tr>
<tr>
<td>Live</td>
<td>1150</td>
<td>0.512</td>
<td>10.045</td>
<td>-4.53</td>
<td>-1.28</td>
<td>0.53</td>
<td>2.22</td>
<td>4.9</td>
</tr>
<tr>
<td>Dead</td>
<td>753</td>
<td>0.336</td>
<td>5.16</td>
<td>-3.88</td>
<td>-1.01</td>
<td>0.47</td>
<td>1.87</td>
<td>4.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Volatility Levels</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>VIX- RV</td>
<td>5.385</td>
<td>6.033</td>
<td>1.21</td>
<td>3.816</td>
<td>6.233</td>
<td>8.737</td>
<td>11.496</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Entropy (N=60 days)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ApEn(m = 1, r = 0.2σ) (VIX)</td>
<td>0.988</td>
<td>0.238</td>
<td>0.66</td>
<td>0.829</td>
<td>1.009</td>
<td>1.163</td>
<td>1.286</td>
<td></td>
</tr>
<tr>
<td>ApEn(m = 1, r = 0.2σ) (√RV)</td>
<td>1.154</td>
<td>0.284</td>
<td>0.749</td>
<td>0.99</td>
<td>1.215</td>
<td>1.376</td>
<td>1.453</td>
<td></td>
</tr>
<tr>
<td>ApEn(m = 1, r = 0.2σ) (VIX - √RV)</td>
<td>1.422</td>
<td>0.118</td>
<td>1.288</td>
<td>1.367</td>
<td>1.441</td>
<td>1.495</td>
<td>1.542</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1: Summary statistics
This table presents summary statistics for the 1,903 hedge funds in our sample. Panel A summarizes returns for 'All Funds', 'Long-Short Equity' subcategory, and the 'Live' and the 'Dead' funds subsets. We report the number of funds N, mean, standard deviation and the quantiles from 10% to 90%. Panel B lists the statistics on the volatility measures used in percentage points and annualized. RV is the realised variance of S&P 500, the square root of which is the realised volatility. VIX is the CBOE volatility index for S&P 500. Panel C shows the summary statistics for the estimated approximate entropy of the respective volatility time series calculated using a rolling window of 60 days.

In this chapter, our focus is on the volatility of the US equity markets, for which we select the S&P 500 index as a representative. As a measure of actual volatility, we use the daily realized variance downloaded from the Oxford-Man Institute’s Realized Library. Furthermore, the daily VIX index is downloaded from the CBOE (Chicago Board of Options Exchange) web site. VIX is a composite measure of market implied volatility, backed out from a series of options written on the S&P 500. All volatility time series are annualized. Panel B of Table 5.1 provides the summary statistics for the various volatility measures. It is clear that the VIX level is higher than the realised measure on average. Both have similar standard deviations and are right

\[
RV_t = \sum_{0 \leq t_j-t \leq 1} x_{j,t}^2, \quad x_{j,t} = X_{t+t_j,t} - X_{t+t_j-1,t}
\]

where \( X \) is the logarithmic of stock price, and \( t_{j,t} \) is the normalised time of trade on \( t \)-th day. See Shephard and Sheppard (2010) for details.

\[ \text{The realised volatility is annualized by multiplying } \sqrt{252}, \text{ the square root of average trading days in a year and the result is expressed in percentage points. The VIX values reported are already annualised.} \]
skewed, as evidenced by their means being greater than their respective medians.

5.4 Entropy of Realized and Implied Volatility

We estimate the ApEn of various volatility measures (i.e. VIX, $\sqrt{RV}$ and their difference) for the S&P 500. As discussed in the previous section, ApEn suffers from a larger bias with smaller samples. To balance the need for more data points in order to produce an accurate entropy measure with the need to track the dynamics of volatility as it changes, we decided, after some trial and error, to use a rolling window of 60 days. In all cases, we set the threshold $r = 0.2\sigma$ and $m = 1$ day. Panel C of Table 5.1 provides the summary statistics for the three sets of ApEn estimates. Of the three volatility measures, VIX has the lowest mean ApEn and the spread (i.e. the difference between VIX and $\sqrt{RV}$) has the highest average ApEn. In particular, the mean ApEn of the spread is high relative to its standard deviation, which may suggest a lesser degree of regularity for the spread.

5.4.1 Realised Volatility

Figure 5.6 plots the estimated ApEn(RV) alongside the realized volatility time series itself over the January 2000 to February 2015 period. The entropy is calculated on the realized volatility level, i.e. it measures the conditional logarithmic likelihood that realized volatility levels that are close will remain close on the following day. A low ApEn signals higher regularity or a change in the volatility regime. From the left diagram of Figure 5.6, we can see that the ApEn(RV) reached the minimum of 0.18 bit on October 9, 2008, and a maximum of 1.7 bit on July 25, 2003.

To further investigate the above finding, Figure 5.7 plots the RV over the 90 days period from May 1 to July 25, 2003. On the left diagram of Figure 5.7, the RV resembles a random walk, which is less predictable, and more irregular. The right diagram of Figure 5.7 shows the lack of a relationship between RV on consecutive days. These patterns are associated with a high entropy regime that fluctuates
Figure 5.6: Realized Variance

The left diagram shows the entropy of realized variance $\text{ApEn}(m = 1, r = 0.29, N = 60)(\text{RV})$, measured daily over a rolling window of 60 days. The right diagram shows the realized variance time series, which is truncated at $3[10^{-3}]$, excluding the largest spike on October 10, 2010 where $\text{RV} = 7.7[10^{-3}]$.

between 1 and 1.7 bit for the July 2003 to February 2007 period. Volatility during this period is consistently lower, with a few short-lived periods of slightly higher volatility.

Figure 5.8 plots the RV over the 60 days period from July 11 to October 9, 2008, on the left, which is marked by a low volatility initial phase as it entered into a high volatility state at the beginning of September 2008. The dramatic increase in volatility, followed by the volatility staying persistently high for a period of time is similar to that simulated in section 5.2.2. The drop in entropy to 0.86 bit on the February 27, 2007 marks the beginning of a change in the volatility pattern, with significant shifts in volatility regimes. On this day, US equity markets dropped steeply in response to the concerns regarding domestic and international economic growth. The right-hand diagram of Figure 5.8 shows there is a great deal of dependence between consecutive days RV, which drove the ApEn values down.
Figure 5.7: RV: Maximum entropy episode  
Left figure shows the realized variance over the 90 days from 1 May to 25 July 2003. The entropy $\text{ApEn}(m = 1, r = 0.2\hat{\sigma})(\text{RV}_{\text{SP500}}) = 1.708231$ measured for this period has been the highest in the sample. The right picture shows the sequence of pairs of consecutive realized variances $(x = RV_{t-1}, y = RV_t)$ as dots, coloured according to their contribution to the overall entropy of the sample.

5.4.2 VIX

The VIX index is a composite measure of the implied volatility of out-of-the-money put and call options on the S&P 500. It is a forward-looking volatility measure, reflecting the market’s expectation of volatility up to option maturity. Figure 5.9 plots the ApEn of VIX over the 2000 to 2015 period alongside the time series of VIX level of the same period. Entropy $\text{ApEn}(\text{VIX})$ ranges between a maximum of 1.52 bit on April 2, 2009 to a minimum of 0.18 bit on October 20, 2008. Compared with the ApEn of RV in figure 5.6, the ApEn of VIX is consistently lower with little change. This suggests that VIX is more persistent.

5.4.3 Spread

We also analyse the spread between implied volatility and realized volatility. To make the two figures comparable, we annualise the realized volatility. The entropy is, as before, estimated on a rolling window of 60 days. In Figure 5.10, the left diagram shows the estimated approximate entropy and the right picture, the corresponding
Figure 5.8: RV: Minimum entropy episode

The left diagram shows ApEn \((m = 1, r = 0.2\hat{\sigma}, N = 60)\) (RV) from July 11 to October 9, 2008. The lowest ApEn is 0.18. The right diagram shows the sequence of pairs of consecutive realized variances \((x = RV_{t-1}, y = RV_t)\) as dots, coloured according to their contribution to the overall entropy of the sample.

The time series of the spread. The entropy hovers around the median of 1.4393 bit most of the time. Two periods stand out, however. In 2008, the entropy falls to 0.79 bit on October 24, 2008. Realized volatility in the markets spike around that time to 139 percentage points annualized. The VIX index spiked a few days later, leading to wild swings in the spread, and in turn this leads to lower entropy levels. Later, the entropy falls to a low of 0.78 bit on May 25, 2010. Again, this started off as a low volatility regime for both RV and VIX. Then realized volatility first spiked, followed by the VIX index. As a whole, there is a greater degree of irregularity in the spread series and this is confirmed by the high ApEn values.
Figure 5.9: VIX
The left diagram shows $\text{ApEn}(m = 1, r = 0.2\sigma, N = 60)(\text{VIX})$, measured daily over a rolling window of 60 days. The right diagram shows the VIX time series with daily close values.

Figure 5.10: Volatility Spread $(\sqrt{\text{RV}} - \text{VIX})$
The left diagram shows $\text{ApEn}(m = 1, r = 0.2\sigma, N = 60)(\sqrt{\text{RV}} - \text{VIX})$, measured daily over a rolling window of 60 days. The right diagram shows the time series of the spread.
5.5 Empirical Results

5.5.1 Timing Model

Linear factor models such as the Arbitrage Pricing Theory (see Cochrane (2005)), or the three-factor Fama-French model (Fama and French (1996)) are at the heart of the asset pricing literature. Given hedge fund returns, \( r_t \), the standard functional form of the factor regression, used, for example, by Fung and Hsieh (2004), is of the form set out below:

\[
r_t = \alpha + \beta^T \cdot r_{M,t} + \epsilon_t \tag{5.11}
\]

\[
r_t = \alpha + \sum_{i=1}^{N} \beta_i \cdot F_{i,t} + \epsilon_t \tag{5.12}
\]

where \( r_{M,t} \) is the factor returns and \( \beta \) is the factor loading.

This factor model has been extended in various directions to capture the different characteristics of the investment strategies that hedge funds employ. One strand focusses on a manager’s ability to forecast market conditions and on the basis of his forecasts, make adjustments to the market exposure of the fund. Timing models, such as those of Treynor and Mazuy (1966) and Ferson and Schadt (1996) among others, have been used to test a fund managers’ ability to time markets.

Following Ferson and Schadt (1996), we use a model with conditional variables that takes into account varying market conditions. The fund managers seeks exposure to the market by adjusting the market beta of his portfolio, to generate portfolio returns as follows:

\[
r_{it} = \alpha_i + \beta_{it} \cdot r_{M,t} + \epsilon_{it} \tag{5.13}
\]

where \( r_{it} \) is the individual monthly excess return of fund \( i \) at time \( t \), \( r_{M,t} \) is the excess return of the market over the same period, and \( \epsilon_{it} \) is the residual error term. The returns are calculated to be in excess of the risk free rate, which we proxy with the returns on the one month Treasury bill. At time \( t - 1 \), the fund manager uses the information on the market conditions available, \( Z_{t-1} \), to forecast the market in the following period at \( t \). Depending on the forecast the fund manager will adjust
the fund’s market exposure. The time evolution of the market exposure $\beta_{it}$ is shown below:

$$\beta_{it} = \beta_i + \gamma_i \cdot Z_{t-1}$$ (5.14)

It is a Taylor expansion to the first linear approximation of $\beta_{it}$ in an additional variable $Z_{t-1}$. Equations (5.13) and (5.14) combined to give the following factor model:

$$r_{it} = \alpha_i + \beta_{it} \cdot r_{M,t} + \gamma_i \cdot r_{M,t} \cdot Z_{t-1} + \epsilon_{it}$$ (5.15)

The return on fund $i$ is determined by some factor $r_{M,t}$ with the factor loading depending on information $Z_{t-1}$. Using the previous period instead of the current period enables us to interpret the change in factor loading as an adjustment of the fund exposure in reaction to a change in the market conditions prediction.

Several studies examined the fund manager’s ability to forecast market returns or volatility. In this chapter, the forecasted market condition is the market volatility regularity. In particular, we focus on the regime change in volatility as measured by the approximate entropy of volatility of S&P 500. In sections 2.4 and 5.2 we introduced ApEn as a model independent measure of the serial structure of a time series. Applied to volatility (see section 5.4), ApEn is sensitive to changes in the level of volatility and its persistence. We use the ApEn of volatility as the state variable of market condition, $Z$, in model (5.15).

### 5.5.2 Panel Regression

Ferson and Schadt (1996) use publicly available information as conditioning factors. The selected factors, representing public information, have been shown to be useful in predicting security returns and risks over time. These include the lagged one-month Treasury bill yield, the dividend yield of the value-weighted CRSP equity index combining the New York Stock Exchange (NYSE) and the American Stock Exchange (AMEX), a measure of the credit spread of corporate bonds.

In this chapter, we used entropy measures of volatility as conditioning variable,
Z, to test if funds adjust their exposure in response to a change in the regularity of volatility or a volatility regime change. In particular we estimated the following models:

\[ r_{it} = \alpha_i + \beta_i \cdot r_{M,t} + \epsilon_{it} \]  
(5.16)

\[ r_{it} = \alpha_i + \beta_i \cdot r_{M,t} + \gamma_i \cdot r_{M,t} \cdot Z_{RV_{t-1}} + \epsilon_{it} \]  
(5.17)

\[ r_{it} = \alpha_i + \beta_i \cdot r_{M,t} + \delta_i \cdot r_{M,t} \cdot Z_{ApEn(RV)_{t-1}} + \epsilon_{it} \]  
(5.18)

\[ r_{it} = \alpha_i + \beta_i \cdot r_{M,t} + \nu_i \cdot r_{M,t} \cdot Z_{ApEn(RV)_{t-1}} \cdot Z_{RV_{t-1}} + \epsilon_{it} \]  
(5.19)

The conditioning variables \( Z_t \) are derived from realized volatility \( RV_t \) at \( t \) and the entropy thereof. The ApEn calculated at time \( t \) uses daily RV from the preceding 60 days period. Other volatilities, such as the implied volatility, will be discussed later in this section. We follow the custom in the timing literature (Ferson and Schadt (1996)) to demean and standardize the conditioning series, ApEn and RV, to make the inference independent of the scale of the variables.

Equation (5.18) is similar in structure to the Ferson Schadt model in (5.17), with the only difference being the choice of the conditioning variable. In equation (5.17), the conditioning variable is the volatility level, whereas in equation (5.18), it is the regularity of volatility with a low ApEn value associated with a higher regularity, indicating a possible change in volatility regime.

The regression method used is the pooled ordinary least squared regression. To test for the appropriate model specification, we ran a series of statistical tests for each regression equation. The Hausman test was used to decide between fixed or random effects, with the null hypothesis being that the preferred model is random effects against the fixed effects as the alternative. We also test if a pooled OLS model would be a more appropriate choice by performing a Lagrange multiplier test of individual fixed or time effects. The test results suggest that there is no substantial inter-fund variation. Accordingly, we use the pooled OLS model in all regressions.

Table 5.2 gives the empirical results for the models in (5.16) to (5.19). The \( F \)
Table 5.2: Panel Regression Conditional CAPM

Pooled OLS model estimation over the entire fund sample of various volatility timing models as defined in equations (5.16) to (5.19). The estimates for the various predictors along with the standard error are shown. The statistical significance of the various estimates is shown according to the associated p-value of a two tailed hypothesis test that the value is zero. In the lower part, statistics such as the $R^2$, adjusted $R^2$ and the F-statistics on the estimated model are shown.

<table>
<thead>
<tr>
<th>Dependent variable: $r_{it}$</th>
<th>$\beta_{it} \cdot r_{Mt}$</th>
<th>$\gamma_{i} \cdot r_{Mt} \cdot Z_{RV_{i-1}}$</th>
<th>$\delta_{i} \cdot r_{Mt} \cdot Z_{ApEn(RV)_{i-1}}$</th>
<th>$\nu_{i} \cdot r_{Mt} \cdot Z_{ApEn(RV)<em>{i-1}} \cdot Z</em>{RV_{i-1}}$</th>
<th>$\alpha_{i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(5.16)</td>
<td>0.187***</td>
<td>-0.031***</td>
<td>0.035***</td>
<td>-0.001***</td>
<td>0.178***</td>
</tr>
<tr>
<td>(5.17)</td>
<td>0.202***</td>
<td>-0.028***</td>
<td>0.032***</td>
<td>-0.002***</td>
<td>0.162***</td>
</tr>
<tr>
<td>(5.18)</td>
<td>0.200***</td>
<td>-0.023*</td>
<td>0.042***</td>
<td>0.004</td>
<td>0.142***</td>
</tr>
<tr>
<td>(5.19)</td>
<td>0.222***</td>
<td>0.0005</td>
<td>-0.0002</td>
<td>0.0002</td>
<td>0.178***</td>
</tr>
<tr>
<td>(5.17 &amp; 5.18)</td>
<td>0.214***</td>
<td>-0.002***</td>
<td>0.0005</td>
<td>0.001</td>
<td>0.126***</td>
</tr>
<tr>
<td>(5.17 &amp; 5.19)</td>
<td>0.247***</td>
<td>0.0004</td>
<td>0.0002</td>
<td>0.001</td>
<td>0.129***</td>
</tr>
<tr>
<td>(5.18 &amp; 5.19)</td>
<td>0.208***</td>
<td>0.001</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.166***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.009</td>
<td>0.009</td>
<td>0.009</td>
<td>0.009</td>
<td>0.010</td>
<td>0.010</td>
<td>0.009</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.009</td>
<td>0.009</td>
<td>0.009</td>
<td>0.009</td>
<td>0.010</td>
<td>0.010</td>
<td>0.009</td>
</tr>
<tr>
<td>F Statistic</td>
<td>1,065.425***</td>
<td>553.230***</td>
<td>551.065***</td>
<td>551.635***</td>
<td>384.235***</td>
<td>393.833***</td>
<td>368.975***</td>
</tr>
<tr>
<td></td>
<td>(df = 1; 117270)</td>
<td>(df = 2; 116710)</td>
<td>(df = 2; 116710)</td>
<td>(df = 2; 116710)</td>
<td>(df = 3; 116709)</td>
<td>(df = 3; 116709)</td>
<td>(df = 3; 116709)</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01
test (at the bottom of Table 5.2) rejects the null hypothesis for all regressions with combinations of conditional variables that the additional variables do not matter. The results for the CAPM base model, (5.16), as reported in the first column of Table 5.2, show a positive cross-sectional equity market exposure over time in the sample. Factor loadings for the predictor variable \( \hat{\beta}_{it} = 0.187(0.006) \) and \( \hat{\alpha}_i = 0.178(0.026) \) are positive and statistically significant. Given the overall equity focus of hedge fund strategies, this is as expected. The low value of the unconditional beta is explained by the fact that only a subset of the funds in the sample pursue an investment strategy which exclusively invests in equity markets. It is also plausible that the hedge funds are active in areas other than the US equity market. Following the unconditional regression, the CAPM model conditional on realized volatility is estimated. The results are shown in the second column of Table 5.2. The estimate for the market condition variable \( r_{M,t} \cdot Z_{RVt} \) is negative \( \gamma_i = -0.031(0.004) \), confirming the results in Ferson and Schadt (1996) that fund betas respond negatively to a rise in market volatility.

We now turn to the discussion on the regression results using conditioning information from the entropy of volatility. The conditioning variable \( Z_{ApEn(RV)_{t-1}} \) in equation (5.18) is the standardized time series of ApEn(RV) calculated based on the realized volatility measured over a 60-day period. The results reported in the third column of Table 5.2, show similar factor loadings as the Ferson and Schadt (1996) model in (5.17). All coefficients for the predictor variables test statistically significantly differently from zero. The unconditional variables \( \hat{\beta}_{it} \) and \( \hat{\alpha}_i \) have lower factor loadings, whereas the conditional variable \( Z_{ApEn(RV)_{t-1}} \) has a higher factor loading \( \delta_i = 0.035(0.005) \), which is bigger than \( \gamma_i \), in absolute terms. The difference in sign is that a downward move in ApEn suggests a higher regularity and a move towards a change in the volatility regime, as was shown in the numerical experiments in section 5.2. A positive loading means the fund’s beta reacts negatively (positively) to a decrease (increase) in the entropy of market volatility in the previous period.

However, a decrease in entropy could be due either to a market volatility regime shift or less persistence in the volatility time series. To be able to distinguish between
these two effects, we introduce a combined conditional variable. The combined factors in $\nu_i \cdot r_{M,t} \cdot Z_{ApEn(RV)_{t-1}} \cdot Z_{RV_{t-1}}$ measure a funds beta adjustment when both entropy and volatility level change. The column marked (5.17) & (5.18) shows the results. The estimates for $\hat{\beta}_i$ and $\hat{\delta}_i$ are similar compared to the simple entropy conditioning model in (5.18). The estimated combined conditional variable $\hat{\nu}_i = -0.002(0.0002)$ is statistically significant. It measures a fund beta response to a combined change in volatility and entropy of volatility. The negative estimate $\hat{\nu}_i$ is consistent with a positive beta response to a drop in volatility entropy and a negative response to an increase in volatility level, and vice versa. This suggests that, on average, fund managers respond to the market volatility settling into a new regime. Depending on whether the new regime has a low volatility or a high volatility regime, the fund manager will make a positive or negative market beta adjustment accordingly. In the last column, we report the results on the model with all combinations of conditional variables. All conditioning variables are statistically significant. The results show a more pronounced result due to the incorporation of all the individual market conditions; most of the factor loadings increased in absolute term.

So, the refined model below is consistent with the empirical findings on the volatility timing ability of fund managers, and that in a situation when volatility settles into a new regime, the funds beta adjustment is the most pronounced;

$$r_{it} = \alpha_i + \beta_{it} \cdot r_{M,t} + \delta_i \cdot r_{M,t} \cdot Z_{ApEn(RV)_{t-1}}$$
$$+ \nu_i \cdot r_{M,t} \cdot Z_{ApEn(RV)_{t-1}} \cdot Z_{RV_{t-1}} + \epsilon_{it}$$ (5.20)

The basic and refined models in equations (5.17) to (5.19) with combinations, add an explanatory power as measured by $R^2$ and the adjusted $R^2$, which are listed in Table 5.2. The low $R^2$ could be due to some omitted variables and might be improved in a non-pooled regression. Both volatility timing models improved the fit marginally compared to the unconditional CAPM in terms of $R^2$. 

122
5.5.3 Strategy and Survivorship

We now turn to the question of whether the managers, volatility timing behaviour in model (5.20) is different when pursuing a specific strategy. The 'Long-Short Equity' investment strategy provides a good test in this respect. It is a strategy where the equity-related factors should be the most appropriate, and it is the most popular investment strategy represented in the fund sample with a sufficient number of funds for valid statistical inference (see Table 5.1 for the summary statistics for the sample funds). The 'Long-Short Equity' returns come from directional as well as spread bets on the equity market. Exposure to the market can be either in long or short equity positions. Table 5.3 show the results for the entropy volatility timing model. The funds market exposure $\hat{\beta}_m = 0.554(0.006)$ is more than double of the sample containing all funds. This market beta is consistent with findings from Fung and Hsieh (2011), who found, for a sample of long/short equity between 1994-2006, that the beta value was between 0.3 to 0.6. The sensitivity to a change in market volatility entropy $\hat{\nu}_i$ is smaller than in the overall sample. With a change in entropy in the market volatility, the fund manager is adjusting less compared with the overall sample. Overall, the results suggest that the 'Long-Short Equity' fund manager responds less forcefully to a change in volatility condition.

We have re-estimated the models, also distinguishing funds which, by 2014, went out of business and those that are still active. The results are listed in Table 5.3 and show significantly different behaviours among the two subsets. Managers of live funds do seem to adjust the portfolio beta in response to an expected change in market volatility more drastically ($\hat{\delta}_i = 0.058$) than dead funds ($\hat{\delta}_i = 0.007$). This finding is consistent with the findings in Busse (1999) in the context of mutual funds, where the author shows that the market beta of surviving funds is sensitive to market volatility, which is not the case for non-surviving funds.
Table 5.3: Panel Regression Conditional CAPM: Live, Dead and LS Equity Funds

Pooled OLS model estimation of one-factor conditional model. Subsets of the fund sample used are: 'Long Short (LS) Equity', 'live' and 'dead' funds. The estimates for the various predictors along with the standard error are shown. The statistical significance of the various estimates is shown according to the associated p-value of a two tailed hypothesis test that the value is zero. In the lower part, statistics such as the $R^2$, adjusted $R^2$ and the F-statistics on the estimated model are shown.

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>LS Equity</th>
<th>Live</th>
<th>Dead</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{it} \cdot r_{M,t}$</td>
<td>0.247***</td>
<td>0.554***</td>
<td>0.224***</td>
<td>0.292***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.006)</td>
<td>(0.011)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>$\delta_{it} \cdot r_{M,t} \cdot Z_{ApEn(RV)}_{t-1}$</td>
<td>0.042***</td>
<td>0.020***</td>
<td>0.058***</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.007)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>$\nu_{it} \cdot r_{M,t} \cdot Z_{ApEn(RV)}<em>{t-1} \cdot Z</em>{RV_{t-1}}$</td>
<td>-0.002***</td>
<td>-0.002***</td>
<td>-0.001***</td>
<td>-0.002***</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0001)</td>
<td>(0.0003)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>$\alpha_{it}$</td>
<td>0.129***</td>
<td>0.182***</td>
<td>0.183***</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.020)</td>
<td>(0.037)</td>
<td>(0.025)</td>
</tr>
</tbody>
</table>

Observations: 116,713   49,601   79,781   36,932
$R^2$: 0.010   0.217   0.006   0.054
Adjusted $R^2$: 0.010   0.217   0.006   0.054
F Statistic: 393.833***   4,585.394***   156.242***   706.728***
(df = 3; 116709) (df = 3; 49597) (df = 3; 79777) (df = 3; 36928)

Note: *p<0.1; **p<0.05; ***p<0.01

5.5.4 Conditional Fama French Model

In previous section 5.5.3, we investigated fund manager behaviour when the regularity of market volatility was changing with the conditional volatility model. We used a conditional CAPM model to find evidence of market beta adjustments when entropy is changing within the sample. Although the market beta of a fund is arguably the most important factor for an equity-focused fund, there are other risk factors that a fund is exposed to. The Fama French three-factor model is a widely used model in the literature (Fung and Hsieh (2011), and Patton and Ramadorai (2013)) to identify common risk factors in the risk exposure of equity focussed funds. It has a factor SMB, which denotes the excess returns of small cap stocks over large cap stocks. HML denotes the excess returns of a portfolio of high book-to-market stocks over returns from a portfolio with low book-to-market stocks. The third factor in the model is the market excess returns over a risk free asset.

In the spirit of the conditional multi-factor models that were employed by Ferson
and Schadt (1996), we extend in the current section the conditional CAPM model to a conditional three-factor model, incorporating the Fama and French (1996) factors (SMB, HML) as additional factors into our model. The extended conditional model is used to show the robustness of the results from the previous section. Our Fama French volatility entropy timing models have the following specifications.

$$r_{it} = \alpha_i + \beta_{it} \cdot r_{M,t} + \beta_{it}^{SMB} \cdot r_{SMB,t} + \beta_{it}^{HML} \cdot r_{HML,t} + \gamma_i \cdot r_{M,t} \cdot Z_{ApEn(RV),t-1} + \epsilon_{it} \quad (5.21)$$

In Table 5.4, the results from the panel regression with this model are shown. Column (1) of the table shows the results from the conditional CAPM of equation (5.18) and column (2) the from the Fama French model. A comparison of the adjusted and unadjusted $R^2$ shows that the Fama French model adds explanatory power, compared to the one factor conditional entropy model. Column (3) shows that a simplified version of the model in equation (5.21), accounting only for an entropy change in the overall market, improves further the $R^2$. One notices furthermore, that the entropy factor loading $\gamma_i \cdot r_{M,t} \cdot Z_{ApEn(RV),t-1}$ is lower for the conditional three factor model (3) than that for the one-factor model (1), showing that the individual risk factors SMB and HML can account for a part of the factor loading. The resulting $\hat{\gamma}_i = 0.008(0.004)$ is according to the t-test statistically significant, providing evidence that the results from the one-factor model are robust also to an enhanced risk factor model. The results from the fully specified model, as in equation (5.22), are shown in column (4) of Table 5.4. Taking into consideration also the changes to the risk factors SMB and HML in reaction to a change in entropy, the sensitivity to the overall market volatility, $\hat{\gamma}_i = 0.004(0.004)$ is no longer statistically significant. However, $\hat{\gamma}_i^{SMB}$ and $\hat{\gamma}_i^{HML}$ are statistically significant suggesting that the Long Short equity managers in the sample change the risk factor exposures to SMB and HML in response to an expected volatility regime change, with only marginal
adjustments being made in the overall market beta of the portfolio. This sheds also some light on our finding from section 5.5.3. There, we found that Long Short Equity funds responded less strongly to a change in market volatility regularity than the average fund in the sample. The factor loadings suggest that the adjustments in the portfolio in reaction to a change in market conditions are not picked up by the market beta, but instead by other risk factor exposure changes.

Table 5.4: Panel Regression Conditional 3F model: LS Equity funds.

Pooled OLS model estimation of conditional 1-factor and 3-factor models for Long Short (LS) Equity funds as defined in equations (5.18) and (5.22). The estimates for the various predictors along with the standard error are shown. The statistical significance of the various estimates is shown according to the associated p-value of a two tailed hypothesis test that the value is zero. In the lower part, statistics such as the $R^2$, adjusted $R^2$ and the F-statistics on the estimated model are shown.

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{it} \cdot r_{Mt}$</td>
<td>0.507***</td>
<td>0.460***</td>
<td>0.463***</td>
<td>0.465***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>$\gamma_i \cdot r_{Mt} \cdot Z_{ApEn(RV)_{t-1}}$</td>
<td>0.013***</td>
<td>0.008**</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td></td>
</tr>
<tr>
<td>$\beta_{it}^{SMB} \cdot r_{SMB,t}$</td>
<td>0.185***</td>
<td>0.185***</td>
<td>0.182***</td>
<td>0.023***</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>$\gamma_i^{SMB} \cdot r_{SMB,t} \cdot Z_{ApEn(RV)_{t-1}}$</td>
<td>0.003</td>
<td>0.004</td>
<td>0.014**</td>
<td>0.034***</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>$\beta_{it}^{HML} \cdot r_{HML,t}$</td>
<td>0.191***</td>
<td>0.137***</td>
<td>0.127***</td>
<td>0.097***</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.020)</td>
<td>(0.020)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>$\gamma_i^{HML} \cdot r_{HML,t} \cdot Z_{ApEn(RV)_{t-1}}$</td>
<td>0.215</td>
<td>0.223</td>
<td>0.225</td>
<td>0.226</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.023)</td>
</tr>
</tbody>
</table>

Observations: 49,601
R$^2$: 0.215
Adjusted R$^2$: 0.226
F Statistic: 6,793.545*** (df = 2; 49598)

Note: *p<0.1; **p<0.05; ***p<0.01

5.5.5 VIX and Volatility Spread

So far we have discussed the results of estimating a conditional CAPM model using ApEn estimated from realized volatility. A different perspective is offered when turning to measures of implied volatility and the spread between realized and implied
volatility. As we have seen, the entropy of implied volatility is substantially lower (see Table 5.1) than that of the realized volatility. This reflects the fact that the expectation on future volatility, which implied volatility is expressing, is moving slower than the actual market volatility, showing higher persistence, resulting in a lower entropy value. For the spread between implied and realized volatility, the entropy results in Table 5.1 suggests substantially higher levels of irregularity and less persistence.

The conditional CAPM model in (5.20) was re-estimated over the entire sample of funds with these other volatility measures, along with the Ferson-Schadt model of equation (5.15). In Table 5.5 the results are shown. The models display results for the unconditional beta $\beta_{it}$ in the range of 0.202-0.218. The next three lines in the table display the conditional market beta $\hat{\beta}_i$ associated with the respective volatility measures as in the Ferson-Schadt model. The beta sensitivity is strongest to a change in implied volatility, and least strong to a change in the volatility spread.

The reaction to a change in the volatility spread $\gamma_i \cdot r_{MA,t} \cdot Z_{(RV-VIX)_{t-1}}$ is lower $\hat{\gamma}_i = -0.027(0.005)$. The result is consistent with the fact that systematic volatility strategies that are betting on changes in the spread are a minority in the sample (‘CTA’, ‘Managed Futures’ and ’Option Arbitrage’ pursuing funds make less than 11% of overall funds). Although the average hedge fund will exploit market conditions of ‘cheap’ volatility relative to higher expected future volatility in buying protection for example, the result is not likely to be substantial for the overall market beta of a hedge fund.

A slightly different aspect of the story is told by looking at the results for $\delta_i \cdot r_{MA,t} \cdot Z_{ApEn(x)_{t-1}}$, where the beta adjustment to a change in volatility is measured. The strongest reaction from funds is seen when the entropy is calculated based on the volatility spread: $\delta_i = 0.101(0.007)$ for $Z_{ApEn((RV-VIX)_{t-1})}$. If we recall, a downward (upward) movement of entropy signals that the volatility (spread) is entering a new regime (settling into a new regime). As Bollerslev and Todorov (2011) have shown, compensation for rare events account for a large part of the average volatility spread. Any shock to the variance risk premium (volatility spread), for example due
to heightened levels of implied volatility without changes in realised volatility (or
delayed response), conveys uncertainty for the further development of the market
(rare events), implied volatility or realized volatility, and possibly signals a change
in regime. The funds in the sample respond to this signal by reducing market
exposure in a manner stronger than to the entropy of RV or VIX. Furthermore,
although the funds are most sensitive to a change in implied volatility, as measured
by \( \gamma_i \cdot r_{M,t} \cdot Z_{VIX,t} \) \( (\hat{\gamma}_i = -0.037) \), they are most sensitive to a possible regime
change in variance risk premium, as measured by its entropy, \( Z_{ApEn((RV-VIX)_{t-1})} \).
Table 5.5: Panel Regression: RV, VIX, $\sqrt{RV} - VIX$.

Pooled OLS model estimation over the entire fund sample of the volatility timing models as defined in the equations (5.16). The conditional market variables depend on realized volatility $Z_{RV_{t-1}}$, implied volatility $Z_{VIX_{t-1}}$, and the spread between implied and realized volatility $Z_{(\sqrt{RV_{t-1}} - VIX_{t-1})}$. The estimates for the various predictors along with the standard error are shown. The statistical significance of the various estimates is shown according to the associated p-value of a two tailed hypothesis test that the value is zero. In the second lower part, statistics such as the $R^2$, adjusted $R^2$ and the F-statistics on the estimated model are shown.

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>rtn</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>$\beta_{it} \cdot r_{M,t}$</td>
<td>0.202***</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
</tr>
<tr>
<td>$\gamma_{it} \cdot r_{M,t} \cdot Z_{RV_{t-1}}$</td>
<td>-0.031***</td>
</tr>
<tr>
<td>$\gamma_{it} \cdot r_{M,t} \cdot Z_{VIX_{t-1}}$</td>
<td>-0.037***</td>
</tr>
<tr>
<td>$\gamma_{it} \cdot r_{M,t} \cdot Z_{(\sqrt{RV} - VIX)_{t-1}}$</td>
<td>-0.027***</td>
</tr>
<tr>
<td>$\delta_{it} \cdot r_{M,t} \cdot Z_{ApEn(x)_{t-1}}$</td>
<td>0.042***</td>
</tr>
<tr>
<td>$x$: RV$<em>{t-1}, VIX</em>{t-1}, (\sqrt{RV} - VIX)_{t-1}$</td>
<td>(0.005)</td>
</tr>
<tr>
<td>$\nu_{it} \cdot r_{M,t} \cdot Z_{ApEn(RV)<em>{t-1}} \cdot Z</em>{RV_{t-1}}$</td>
<td>-0.002***</td>
</tr>
<tr>
<td>$\nu_{it} \cdot r_{M,t} \cdot Z_{ApEn(VIX)<em>{t-1}} \cdot Z</em>{VIX_{t-1}}$</td>
<td>-0.018***</td>
</tr>
<tr>
<td>$\nu_{it} \cdot r_{M,t} \cdot Z_{ApEn(\sqrt{RV} - VIX)<em>{t-1}} \cdot Z</em>{(\sqrt{RV} - VIX)_{t-1}}$</td>
<td>-0.023***</td>
</tr>
<tr>
<td>$\alpha_{it}$</td>
<td>0.162***</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
</tr>
</tbody>
</table>

| Observations       | 116,713 | 117,272 | 117,272 | 116,713 | 117,272 | 117,272 |
| R$^2$              | 0.009   | 0.010   | 0.009   | 0.010   | 0.010   | 0.011   |
| Adjusted R$^2$     | 0.009   | 0.010   | 0.009   | 0.010   | 0.010   | 0.011   |
| F Statistic        | 553.230*** | 565.588*** | 549.599*** | 393.833*** | 386.244*** | 427.753*** |
|                   | (df = 2; 116710) | (df = 2; 117269) | (df = 2; 117269) | (df = 3; 116709) | (df = 3; 117268) | (df = 3; 117268) |

Note: *p<0.1; **p<0.05; ***p<0.01
5.6 Conclusion

In this chapter, we use for the first time approximate entropy (ApEn) in a conditional CAPM model to identify volatility timing behaviour in hedge funds. We show the use of ApEn in measuring changes in the serial structure of volatility. Various numerical simulations showcase the pattern of ApEn in economic scenarios such as a change of volatility regime. We derive conditions for low levels of entropy analytically in popular volatility models and processes, demonstrating the usefulness of ApEn in distinguishing different volatility patterns. We find that entropy is more sensitive to volatility shocks than to serial correlation. This way ApEn is a useful and model free measure of a change in volatility regime.

On a sample of hedge funds, we found that manager adjust the funds’ market beta when volatility is settling into a new regime. Dead funds show no such market beta adjustment in response to a change in volatility regularity, differing from live funds where such an adjustment is much stronger. The results are consistent with the findings in Busse (1999) on mutual funds. For ‘Long-Short Equity’ funds, managers adjust their risk exposure to SMB and HML, but not the market beta, when there is a change in volatility regime. The results we obtained with the conditional CAPM are robust when extending the model to a conditional Fama-French model. The results are robust when the entropy is calculated using VIX and spread (defined as the difference between VIX and realised volatility), though the qualitative interpretation of the results is slightly different. Our research also showcase the usefulness of entropy as a model independent measure of volatility regime change.
Chapter 6

Conclusion

In this thesis, three studies have been conducted, which have the application of information theoretic tools as a common thread. The first study formulates a model to analyse, for the first time, draws in a bivariate setting. Encoding the draw information into a time series of discrete values, we used entropy on the derived series to analyse the correlation and cross-correlations of drawdowns and drawups. We investigated the daily and hourly exchange rates of EUR/USD and GBP/USD from 2001 to 2012. For the hourly and the daily data, we found evidence of dependence among the largest draws. We were also able to identify information flows between drawup/drawdown of the two exchange rates, both for the hourly and daily data. Both daily and hourly series show clear evidence of information flowing from EUR/USD to GBP/USD and, in a slightly more pronounced way in the reverse direction. Robustness tests show that the information measured is not due to noise, from which we conclude that there is a measurable information transfer between the large draws of the two exchange rates, which is potentially useful for forecasting and risk management.

In the second study, we defined a two-state hidden Markov model to identify volatility regimes in a set of currencies quoted against the US dollar. The estimated state process is a process taking discrete values, which lends itself to analysis with information theoretic tools. Following the procedures in the previous study, we used the entropy framework with mutual information and transfer entropy to examine in-
interdependence and volatility spillover among the currencies in the sample in a novel way. We found evidence of various volatility regime relationships between the currency pairs in the sample. Among the European currencies (EUR/USD, GBP/USD, CHF/USD), volatility regime co-movements were identified. We also found evidence of a co-movement relationship between CAD/USD and AUD/USD volatility states. We were able to find evidence of information flows between the volatility states of currency pairs, most notably from EUR/USD to GBP/USD, which indicates a causal volatility spillover relationship, confirming findings in the literature (Inagaki (2006)), but using a different model and set-up. The work showcased the usefulness of the concepts of mutual information and transfer entropy when applying them to volatility spillover studies.

For the last study, we extended the market timing model of Ferson and Schadt (1996), using the approximate entropy (ApEn) of market volatility series as a state variable. We demonstrated the sensitivity of ApEn to volatility regime changes and serial correlation patterns, and used it in the conditional factor model. The analysis adds to the market volatility timing research for hedge funds using ApEn as a model independent measure for volatility regime changes. On a sample of hedge funds, we found that the managers adjust a fund’s market beta when volatility is settling into a new regime. Dead funds show no such market beta adjustment skill, differing from live funds, where such an adjustment can be identified.
Bibliography


Darbellay, G. and D. Wuertz (2000). The entropy as a tool for analysing statistical
dependences in financial time series. In *Physica A-Statistical Mechanics and
Signaux, CH-1015 Lausanne, Switzerland.

Dimpfl, T. and L. Huergo (2011). Using transfer entropy to measure information
flows from and to the CDS market.

Dionisio, A., R. Menezes, and D. A. Mendes (2004, July). Mutual information:
a measure of dependency for nonlinear time series. *Physica A: Statistical Me-
chanics and its Applications* 344, 326–329.

Duan, W.-Q. and H. E. Stanley (2010, May). Volatility, irregularity, and pre-
dictable degree of accumulative return series. *Physical Review E. Statistical,

Edwards, S. and R. Susmel (2000, December). Volatility dependence and conta-
gion in emerging equity markets. *Journal of Development Economics* 66(2),
505–532.

Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market. *Econo-
metrica* 58(3), 525–542.


425–461.

Fleming, J., C. Kirby, and B. Ostdiek (2001, February). The Economic Value of

Funds and Commodity Funds: Natural vs. Spurious Biases. *The Journal of
Financial and Quantitative Analysis* 35(3), 291.


Fung, W. and D. Hsieh (2011, August). The risk in hedge fund strategies: The-
ory and evidence from long/short equity hedge funds. *Journal of Empirical
Finance* 18(4), 23–23.


*arXiv.org*, 7138.


