Analysis of Financial Time Series using Non-Parametric Bayesian Techniques



Syed Ali Asad Rizvi
Brasenose College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Trinity 2018

Abstract

The overarching aim of this thesis is to show that Gaussian processes and Rényi entropy can be valuable non-parametric tools for forecasting intraday volatility for a wide range of financial time series.

In this thesis empirical volatility forecasting using Gaussian processes (GPs) is presented for stocks, market indices, forex and cryptocurrencies. Key innovations are presented in the application of GPs by using separated negative and positive returns in transformed log space, and the use of Rényi transfer entropy for incorporating information flow from other time series to modify volatility forecasts. Significant performance gains are demonstrated over strong benchmarks for volatility forecasting - the mcsGARCH model that is specially designed for intraday volatility, rolling average, and use of last observed value as next step forecast. A set of robust loss functions are used for assessing performance, and we establish the significance of all results at the 95% confidence interval.

This thesis demonstrates empirically that standalone GPs perform better than GARCH, EGARCH, GJR-GARCH in forecasting intraday financial volatility. This is done by presenting the results from the largest study done in literature to date, that uses GPs for this purpose. Approximately 50K experiments are carried out and over 18 million volatility forecasts are analysed, on 11 years worth of trading data from 50 market symbols sampled at 1-minute frequency to establish the significance of the results.

It is recognized that GARCH, EGARCH, GJR-GARCH have not been designed for intraday volatility forecasting therefore this comparison is not with the best in class. mcsGARCH is chosen as the best in class GARCH model specifically designed for forecasting intraday volatility. After this selection it is empirically established that when mcsGARCH is used for benchmarking the performance of GPs, it is found that plain GPs are substantially worse than mcsGARCH. This negative result helps establish the need for innovation in the application of GPs to volatility forecasting.

An innovation for the GP intraday volatility forecasting model is developed by regressing on the negative and positive returns envelopes separately in log-space. The forecasts generated from this method are found to significantly outperform all benchmarks including mcsGARCH in terms of their loss function performance. This superior performance against mcsGARCH successfully demonstrated that GPs can be a significant tool in the volatility practitioner's forecasting toolkit.

In order to incorporate information from other financial time series into the volatility forecast the concept of Rényi transfer entropy is introduced as a non-parametric measure for calculating the asymmetrical information flows between two financial time series with the ability to emphasize different parts of the event space. The effective Rényi transfer entropy for key market indices is calculated and information flows around key global events are visualized. It is found that very high flows of information occur around global events of financial importance but very little influence is found for politically important events.

A new approach is devised for folding the effective Rényi transfer entropy into the forecast made by envelope based GPs. This new approach of using the ERTE modified envelope GP forecasts is applied for predicting intraday volatility for 6 market indices and 12 forex pairs. The performance improvement from these modified forecasts is benchmarked and found to be significantly better than all benchmarks considered including the standalone envelope GP forecasts. The parameters used in calculating the ERTE forecasts

are found empirically and values which provide the most consistent gains in the training data are used in the comparison against benchmarks.

The approach is extended further and applied to forecast the volatility in cryptocurrency markets by conducting experiments on 6 month long time series from 10 cryptocurrencies sampled at 5-minute frequency. It is found that envelope based GPs significantly outperform against all three benchmarks on this dataset. This is the largest volatility forecasting study done to date by using intraday frequency data from multiple cryptocurrencies.

An innovation is presented for folding the effective Rényi transfer entropy value from multiple time series into one value by extending the non-parametric formulation for ERTE. While this extension results in the need for a larger data set, its application for calculating ERTE remains computationally simple with a high degree of flexibility

The combined time series ERTE value is used for modifying the forecasts from envelope based GPs and applied to the case of using up to 9 secondary cryptocurrency time series for calculating the forecasts for 1 primary cryptocurrency. This technique is trialled for all 10 cryptocurrencies in the data and it is found that although this method still out performs the three benchmarks, its performance against the standalone envelope based GP is not consistent. A potential reason for this is discussed by assessing the relationship between the discretization parameter in ERTE and the requirements it poses for the size of the dataset.

The thesis closes with a discussion of future directions and open questions.

Declaration

The thesis I am submitting is entirely my own work except where otherwise indicated.

Acknowledgements

I thank my supervisor Professor Steve Roberts for his consistent encouragement, sage advice, support and guidance.

I thank my parents Parveen Ali and Abid Rizvi for their kind patience and care.

I thank Rhodes Trust and Oxford-Man Institute of Quantitative Finance for their generous benefaction.

My everlasting gratitude for the many friends who supported me through this, especially Grzegorz and Khadija.

Contents

Preface **9**

1 Introduction **10**

1.1 Objective of this Thesis 10

1.2 Gaps in Existing Literature 11

1.3 Scope of Work 13

1.4 Major Contributions of this Thesis 13

2 Gaussian Process Based Volatility Forecasting **15**

2.1 Introduction 16

2.2 Challenges of Predicting Financial Volatility 17

2.3 Gaussian Processes for Time Series Forecasting 19

2.4 GARCH models for Financial Volatility 21

2.5 Performance and Accuracy Metrics for Forecast Evaluation 22

2.6 Datasets for Understanding Performance of Gaussian Processes 25

2.7 Using Gaussian Processes for Forecasting Financial Volatility 26

2.8 Performance of GPs vs GARCH 30

2.9 Stronger Benchmarks for Volatility Forecast Evaluation 33

2.10 The Multiplicative Component GARCH Model mcsGARCH 33

2.11 Results Against the Stronger Benchmarks 36

2.12 Conclusion 36

3 Envelope Based Gaussian Processes **40**

3.1 Extensions to the Gaussian Process Model 40

3.2 Methodology for Experiments 43

3.3 Discussion of results 44

3.4 Assessing the Distribution of the Observed Returns Relative to the Volatility Forecasts 53

3.5 Computing Time Cost for Assessing GP Performance 53

3.6 Conclusion 54

4 Information Flows between Financial Time Series Using Rényi Entropy **56**

4.1 Introduction 57

4.2 Measuring Information Flows Using Rényi Transfer Entropy 58

4.3 Stock Market Data 63

4.4 Methodology for Experiments 66

4.5 Calculating Monthly Effective Rényi Transfer Entropy 67

4.6 Summary of Highlighted Financial Events 68

4.7 Conclusion 70

5 Forecasting Volatility Using Rényi Information Flows **82**

5.1 Introduction 82

5.2 Combining ERTE with volatility forecasts 83

5.3 Datasets for Market Indices and Forex 85

5.4 Methodology of Experiments 86

5.5 Evolution of ERTE with q 87

5.6 Analysing Gain from ERTE with Discretization Level 96

5.7 Evolution of ERTE with Block Length 96

5.8 Discussion of results 98

5.9 A Better Apple 104

5.10 Conclusion 106

6 Forecasting Volatility in Cryptocurrency markets **107**

6.1 Introduction 108

6.2 Challenges of Working with Cryptocurrency Data 109

6.3 Folding Multiple Time Series into the Volatility Forecast 110

6.4 Volatility Proxy based on OHLC 112

6.5 Cryptocurrency Data 112

6.6 Methodology of Experiments 116

6.7 Discussion of Results 121

6.8 Caveat on Forecasting Performance 125

6.9 Conclusion 125

7 Conclusion **143**

7.1 Evaluation of the Objective 143

7.2 Summary of Key Contributions in the Thesis 144

7.3 Future Directions 146

7.4 Open Questions 147

A Computational Setup **148**

Preface

Imagine a boardgame with 7 billions players, trillions of possible moves at each step, and the possibility of winning life changing money at any given turn. How would you choose your next move?

It's one of life's most difficult questions, one that has ramifications for our social, political, philosophical, and economic existence. Yet this is the very question that we try to answer when we study financial markets. For those who like boardgames, like I do, financial markets are the largest boardgame in play with an endlessly evolving set of rules and the whole global population as its cast of players. This is why financial markets are one of the most interesting studies of collective human behaviour and decision making.

Financial markets are the buzzing machine at the very heart of the human ecosystem. They represent trillions of dollars and countless jobs and we often remain blindly unaware that willingly or unwillingly we all participate in them everyday. Even our smallest decisions such as choosing which song to hear or which sweater to wear, when aggregated together, feed into the vast machinery of the global financial markets. It's a Keynesian beauty contest of the most epic proportions.

When we look at financial markets from a quantitative lens we are attempting to understand, explain, and forecast the quintessentially qualitative decision making of humans at a global scale. The price of Apple Inc. stock on any given day may feel so numeric on the surface, but if we unpack it we will soon see that it represents the desires of millions of Christmas gift givers, billions of creative choices of designers and product developers, and thousands of decisions made by Samsung, Google and Nokia employees.

Let's turn back to the boardgame analogy for a moment. When you are playing a game

with countless players, with the stakes so high, and you can't even see the whole board let alone what each player is doing at any given moment, your decisions will seldom have concrete certainty. Each step you take will carry an inherent level of uncertainty in it. It is this uncertainity about the future, and even the present, that when pooled together across the whole of the financial markets gives rise to the fascinating phenomenon known as volatility. Despite the many quantitative interpretations of financial volatility, at its most fundamentally intuitive level it is the uncertainty in our collective human decisions.

The field of forecasting volatility is an effort to quantify these sentiments about the unknown, the concerns for the future, and the information blind spots that the market players have and then forecast their impact on the state of the game at the next turn.

A quantitative interpretation of volatility is crucial within financial markets for option pricing, risk management and portfolio management. Volatility can now be traded as an instrument itself, so it is no longer merely a variable feeding into other algorithms. Financial market volatility also has a wider impact on financial regulation, monetary policy, and our assesement of a country's political health. It is this global importance of volatility that has fascinated and inspired the work in this thesis.

I owe a wealth of gratitude to my supervisor Prof. Steve Roberts, my family, and my friends - especially Khadija, Grzegorz, Elmarie, Karl, Justin, and Jonny - for motivating me through the final phase and helping in refining this work.

I hope that you will enjoy reading about my research as much as it has been enjoyable for me to develop it.

Syed Ali Asad Rizvi

Chapter 1

Introduction

1.1 Objective of this Thesis

The overarching aim of this thesis is to show that Gaussian processes and Rényi entropy can be valuable non-parametric tools for forecasting intraday volatility for a wide range of financial time series.

First, the objective is to show that Gaussian processes perform consistently better at forecasting intraday volatility than other well established models - namely the GARCH (Generalized Autoregressive Conditional Heteroskidastic) model and some of its variants, especially mcsGARCH (multiplicative component GARCH). Gaussian processes are a highly flexible non-parametric Bayesian technique. They enable the encoding of our prior beliefs about the state of the world and generate probability distributions over forecasts rather than just point estimates. The intention is to show that even when compared to strong contenders like mcsGARCH, Gaussian process based volatility forecasting techniques perform significantly better.

Second, the aim is to present a novel technique for improving volatility forecasting by combining the information flow from other financial time series by using Rényi transfer entropy - a non-parametric measure for studying transfer entropy. Rényi transfer entropy provides an exciting lens to understand the flow of information in financial markets at the global scale, and allows interpretable insights to be drawn about the underlying dynamics

of volatility flow across multiple financial time series. The intention is to demonstrate that Rényi transfer entropy can help in the improvement of volatility forecasts by taking into account the information flowing from multiple financial time series.

Third, the aim is to demonstrate that the higher performance of Gaussian processes is consistent across a long period of time and a breadth of financial time series - including stocks, market indices, forex pairs, as well as cryptocurrencies. The intention is to establish that the performance of Gaussian process based methods make them a valuable tool for predicting the volatility of a diverse set of financial time series.

1.2 Gaps in Existing Literature

In the field of machine learning applied to volatility forecasting there are four key gaps. This thesis is an attempt to address those gaps.

The first gap is the lack of a commonly used data set of financial time series that can be used to rank the performance of different machine learning algorithms against each other. This lack of consistency is challenging for comparing and ranking the prediction performance because different authors end up using very different data sets. Such rankings have been developed in econometrics literature and they help propel the state of the art [43, 32, 18]. A consistent data set would make machine learning based financial volatility research more streamlined and transparent, in the same way that MNIST, CIFAR, or ImageNet have enabled extensive comparison and ranking of machine learning classification techniques.

The second gap is lack of breadth of data that is used to demonstrate performance. In published literature most machine learning algorithms are often only tested on one or two choice financial time series, usually sampled at the daily frequency with a widely varying array of time horizons. This makes it quite difficult to assess if the algorithm is a significant improvement on the state of the art or just have a good 'trading' day. The impact of this inconsistency is that reported performance tables within published papers are hard to compare with each other. An algorithm that does well on one stock market

symbol over a long time period is certainly commendable but there are around 2800 stock symbols in New York Stock Exchange alone, to say nothing of other stock exchanges, forex pairs, market indices, or commodities. Therefore performance of machine learning algorithms needs to be demonstrated on a broader set of financial data sets to establish consistency of performance. Guidelines on using a broad set of data have been laid out in econometrics literature [43, 18, 33] that would prove beneficial if adopted in machine learning.

Third, there is no single set of econometric loss functions that is widely used in the machine learning based financial forecasting literature. Although econometrics research has been done extensively on the impact of loss function choice on forecast assessment [38, 51], this research has not been adopted widely in machine learning. Research published in the field of machine learning based financial forecasting often limits itself to using likelihood based loss functions more common in machine learning, rather than the mean squared error and QLIKE loss functions advocated in econometrics. This lack of consistent loss measures between the two fields make comparison very difficult and keeps the research siloed. A consistent set of econometrics loss functions would allow the direct comparison of developments in the two fields and enable rapid cross pollination of ideas.

Fourth, a particularly overlooked area in volatility forecasting is data sampled at 1-minute frequency. It's not only a gap in machine learning literature but rather in econometrics more widely. The 1-minute frequency has many positive attributes that make it a good candidate for further research. First it generates enough data so as to merit exploration with advanced machine learning techniques that need vast data quantities. Second, unlike second or tick level data it does not suffer from the complex effects of market microstrucutre [2, 1, 37, 48]. Third, at this level most stock market symbols are sufficiently liquid so that the data set does not have to wrangled excessively to manage missing values [1]. Despite these positive features, published research on data at this frequency remains sparse.

1.3 Scope of Work

The work in this thesis focuses on using Gaussian processes as a well established non-parametric Bayesian technique for time series forecasting. Its flexible nature and ability to encode prior information makes it an ideal candidate for exploring financial volatility forecasting.

The dataset used in this thesis spans 50 market symbols including stocks, global market indices, and forex pairs covering a 11 year period sampled at a 1 minute frequency. The data also includes 10 cryptocurrencies covering a 6 month period sampled at 5-minute frequency.

The empirical results are compared to well established models intraday volatility - GARCH and in particular mcsGARCH. Only robust loss functions are used, as established in econometrics literature.

1.4 Major Contributions of this Thesis

The work presented in this thesis builds upon and significantly expands the scale of existing research in the area of non-parametric methods applied to financial volatility forecasting.

This work presents one of the largest academic studies of financial market volatility by carrying out empirical research on an immense dataset of 50 market symbols for an 11 year period sampled at 1-minute frequency, and 10 cryptocurrencies for a period of 6 months sampled at 5-minute frequency. Approximately 50K experiments are carried out and over 18 million volatility forecasts are analysed to establish the significance of the results. All findings are evaluated against well established intraday volatility benchmarks using a robust set of loss functions.

This work also makes two significant contributions to existing literature. First, it introduces the concept of envelope based volatility forecasts by using Gaussian processes separately on the positive and negative envelopes of the returns of a financial time se-

ries. Second, it introduces a method for incorporating the asymmetric flow of information between financial time series into the volatility forecast.

The rest of the thesis is arranged as follows:

Chapter 2 introduces the framework for forecasting volatility using Gaussian processes, along with key definitions, choices of benchmarks, loss fucntions, and dataset descriptions. The results for experiments using Gaussian processes are presented and their performance is analysed.

Chapter 3 presents an extension of the Gaussian process framework by regressing on the negative and positive return envelopes separately in log-space. The impact of this extension is analysed empirically on a large set of stock market symbols.

Chapter 4 introduces the concept of Rényi tranfer entropy and the utilization of it for estimating information flows between financial time series. The information flows between global market indices around global financial events are visualised and discussed.

Chapter 5 presents a novel technique for using in the Rényi transfer entropy for improving the volatility forecast generated by the envelope based Gaussian process. The impact of this technique is empirically tested for 6 global market indices and 11 forex pairs.

Chapter 6 extends the Rényi transfer entropy technique to incorporate the information flow from multiple time series simultaneously. This technique is then tested for forecasting the volatility of 10 cryptocurrencies and the performance gains are analysed.

Chapter 7 concludes the thesis with a summary of the key findings and presents some thoughts on future directions and open questions.

Chapter 2

Gaussian Process Based Volatility Forecasting

Objective

The aim of this chapter is to gain an empirical understanding of the performance of Gaussian processes (GPs) for forecasting volatility for financial market data that has been sampled at 1-minute intervals. The major contribution of this chapter is that it presents the largest study of the performance of GPs for forecasting volatility. The GPs are put to the test in over 48 thousand experimental runs on a data set containing approximately 100 million observations of price data. The data set is diversified by including stocks, market indices, forex, and commodities indices. First the performance of GPs is benchmarked against 3 flavours of the GARCH model. Although these GARCH models had not been developed with intraday frequency in mind, they are the most commonly used yardsticks found in literature. After this the performance of GPs is compared to mcsGARCH which is a model developed specifically for intraday volatility.

This chapter lays out the complete framework used for the experiments. This involves selecting datasets, a proxy for volatility, performance and accuracy metrics, benchmarks to gauge the performance against, as well as an overall methodology and the technical set up required to carry out the experiments.

2.1 Introduction

In financial time series, volatility is a measure of the uncertainty present in the market at a given point in time. Volatility forecasting is important for various types of economic agents, ranging from option traders to price options, to investors looking to forecast exchange rates, and banks looking to hedge their risks. Volatility forecasting is challenging because the phenomenon of conditional variance is unobserved and this makes the evaluation and comparison of various proposed models difficult. A great number of models have been proposed since the introduction of the seminal papers by Engle and Bollerslev [25, 9]. DeGooijer provides a comprehensive overview of the various models that have come to the fore over the last 25 years [18]. Most of the models that are extensively studied and used in practice are parametric in nature and only recently has there been a move towards exploring semi-parametric and non-parametric models. An exploration of various parametric and non-parametric methods for volatility measurement has been summarised in literature [3], as well as the desired properties of good volatility forecasts [4, 55]. An excellent survey of the field of volatility forecasting is presented in Poon (2005) [54].

Financial time series are understood to be heteroskedastic in nature, i.e. the volatility varies with time. The volatility also exhibits a phenomenon known as bunching or volatility clustering, where periods of high volatility are often followed by periods of high volatility and periods of low volatility usually follow periods of low volatility. Negative and positive returns [1] do not affect volatility in the same way [36] It is often seen that large negative returns seem to raise volatility more than positive returns which have the same absolute value [42].

The most widely used parametric model for volatility forecasting is the Generalized Autoregressive Conditional Heteroskedascity (GARCH) model, introduced by Bollerslev in 1986 [9], developed from the ARCH model introduced by Engle in 1982 [25]. GARCH assumes a linear relationship between the current and past values of volatility, and as-

[1] Returns of a financial time series are price movements and are understood to be a measure of the difference between two consecutive values of the time series. Returns are more precisely defined in the next section

sumes that positive and negative returns have the same effect on volatility. Many variations have been introduced to the concept of the GARCH model, most of which try to address one or both of these assumptions, each with its own merits and demerits [32].

This thesis explores a volatility forecasting methods based on Bayesian non-parametrics - Gaussian Processes, and provides a comparison with traditional approaches. The use of Gaussian processes for volatility forecasting has been explored in literature primarily on daily data [23, 53] [53] using a few symbols of the thousands available in the financial markets. A larger study using daily data of 20 forex pairs and 30 stock symbols is presented in Wu et al. (2014) [66], with GARCH, EGARCH (Exponential GARCH) and the so-called GJR-GARCH (Glosten Jagannanthan Runkle GARCH, named after the authors) used as benchmarks. This chapter presents the largest study done using Gaussian processes on minute level price data for forecasting financial volatility. This work extends the state of the art by establishing a baseline for the performance of Gaussian processes benchmarked against well known GARCH models, GARCH, GJR-GARCH and EGARCH. It is discovered that these models are not a like-for-like comparison since they are not developed for intraday volatility. In order to provide a fair comparison mcs-GARCH is introduced as a benchmark.

This chapter begins by introducing the challenges of forecasting financial volatility. This is followed by a short primer on Gaussian processes and GARCH. Then the methodology of experiments and the data under consideration are outlined. The chapter finishes with a discussion of the results.

2.2 Challenges of Predicting Financial Volatility

Volatility can not be observed directly and this necessitates that we use some proxies for volatility in order to carry out any empirical analysis. Some commonly used proxies include squared returns, absolute returns, Open Close High Low, and realized volatility [61] Squared and absolute returns being the most commonly used, followed by realized volatility. The absolute log returns $|r_t|$ are used as the volatility proxy $\widehat{\sigma}$ for the true

volatility $\overset{*}{\sigma}$ throughout this thesis

The choice between squared and absolute returns has been subject to much academic study, but no definitive approach has been established. Zheng (2014) [68] find them both comparable in sensitivity as noisy proxies for volatility. Giles (2008) [29] found that models based on absolute returns result in higher mean squared error than ones based on squared returns. Many studies empirically show that absolute returns outperform square return based volatility measures [27, 24, 46] Ding (1993) [22] suggest that volatility should be measured directly from absolute returns. Davidian & Carroll (1987) [16] show that volatility specification based on absolute returns is more robust against asymmetry and non-normality.

In this thesis, the choice of using absolute returns as the proxy for volatility was further motivated by empirical considerations. In empirical analysis absolute returns provide the best possible range of data without carrying out any scaling. The returns at the minute level are close to zero and using squared returns at the minute level brings the time series even more close to zero and this can create many numerical and computational challenges.

In order to arrive at the returns the price time series needs to be de-trended Financial time series are non-stationary in the mean, and the linear trend in the mean can be removed by taking the first difference of the time series, as given by

$$r_t = (p_t - p_{t-1})/p_{t-1}, \tag{2.1}$$

where p_t and p_{t-1} are the prices at time t and $t-1$ respectively. These are known as arithmetic returns. An alternative to the arithmetic return is to take the geometric returns, also known as the log returns. Log returns r_t are obtained by using:

$$r_t = \log(p_t) - \log(p_{t-1}). \tag{2.2}$$

The absolute value of these log returns is used through out this thesis for processing all price time series, unless explicitly stated otherwise.

A key factor to note at this point is that absolute returns are a downward biased proxy for volatility (Giles 2008), this means that models of volatility based on absolute returns might underestimate the underlying volatility. The exact amount of the downward bias depends on the underlying distribution of the returns. For example, if the underlying distribution was normal then this downward bias is around 0.8 of the true σ^*.

2.3 Gaussian Processes for Time Series Forecasting

A general familiarity with Gaussian processes is assumed and only the information most relevant for this thesis is presented here. For a detailed treatment of Gaussian processes refer to the seminal work by Rasmussen & Williams [64] [56].

Gaussian processes form a class of non-parametric models for classification and non-linear regression. They have become widely popular in the machine learning community where they have been applied to a wide range of problems. Their major strength lies in their flexibility and the ease of their computational implementation.

A Gaussian process is a collection (possibly infinite) of random variables, any finite subset of which have a joint Gaussian distribution. For a function $\mathbf{y} = f(\mathbf{x})$, drawn from a multi-variate Gaussian distribution, where $\mathbf{y} = \{y_1, y_2, ..., y_n\}$ are the values of the dependent variable evaluated at the set of locations $\mathbf{x} = \{x_1, x_2, ..., x_n\}$, we can denote this as

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}; \boldsymbol{\mu}(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x})) \tag{2.3}$$

where $\boldsymbol{\mu}$ is a *mean function*, and $\mathbf{K}(\mathbf{x}, \mathbf{x})$ is the *covariance matrix*, given as

$$\mathbf{K}(\mathbf{x}, \mathbf{x}) = \begin{pmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\ \vdots & \vdots & \vdots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \end{pmatrix} \tag{2.4}$$

Each element of the covariance matrix is given by a function $k(x_i, x_j)$, which is called

the *covariance kernel*. This kernel gives us the covariance between any two sample locations, and the selection of this kernel depends on our prior knowledge and assumptions about the observed data.

In order to evaluate the Gaussian process posterior distribution at a new test point x_* we use the joint distribution of the observed data and the new test point,

$$p\left(\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}(\mathbf{x}) \\ \mu(x_*) \end{bmatrix}, \begin{bmatrix} \mathbf{K}(\mathbf{x},\mathbf{x}) & \mathbf{K}(\mathbf{x},x_*) \\ \mathbf{K}(x_*,\mathbf{x}) & k(x_*,x_*) \end{bmatrix}\right) \tag{2.5}$$

where $\mathbf{K}(\mathbf{x},x_*)$ denotes a column vector comprised of $k(x_1,x_*),...,k(x_n,x_*)$ and $\mathbf{K}(x_*,\mathbf{x})$ is its transpose. By matrix manipulation, we find that the posterior distribution over y_* is Gaussian and its mean and variance are given by

$$m_* = \mu(x_*) + \mathbf{K}(x_*,\mathbf{x})\mathbf{K}(\mathbf{x},\mathbf{x})^{-1}(\mathbf{y} - \boldsymbol{\mu}(\mathbf{x})) \quad \text{and} \tag{2.6}$$

$$\sigma_*^2 = K(x_*,x_*) - \mathbf{K}(x_*,\mathbf{x})\mathbf{K}(\mathbf{x},\mathbf{x})^{-1}\mathbf{K}(\mathbf{x},x_*). \tag{2.7}$$

This can be extended for a set of locations say $\mathbf{x}_*$, to find the posterior distribution of $\mathbf{y}_*$. Using standard results for multivariate Gaussians, the extended equations for the posterior mean and variance are given by

$$p(\mathbf{y}_* \mid \mathbf{y}) = \mathcal{N}(\mathbf{y}_*; \mathbf{m}_*, \mathbf{C}_*) \tag{2.8}$$

where,

$$\mathbf{m}_* = \boldsymbol{\mu}(\mathbf{x}_*) + \mathbf{K}(\mathbf{x}_*,\mathbf{x})\mathbf{K}(\mathbf{x},\mathbf{x})^{-1}(\mathbf{y}(\mathbf{x}) - \boldsymbol{\mu}(\mathbf{x})) \tag{2.9}$$

$$\mathbf{C}_* = \mathbf{K}(\mathbf{x}_*,\mathbf{x}_*) - \mathbf{K}(\mathbf{x}_*,\mathbf{x})\mathbf{K}(\mathbf{x},\mathbf{x})^{-1}\mathbf{K}(\mathbf{x}_*,\mathbf{x})^\top. \tag{2.10}$$

If the observed function values, y_i, have noise associated with them, then we can bring a noise term into the covariance. Since the noise of each sample is expected to be uncorrelated therefore the noise term only adds to the diagonal of $\mathbf{K}$. The covariance for noisy

observations is given as

$$\mathbf{V}(\mathbf{x},\mathbf{x}) = \mathbf{K}(\mathbf{x},\mathbf{x}) + \sigma_n^2\mathbf{I}, \tag{2.11}$$

where $\mathbf{I}$ is the identity matrix and σ_n^2 is a *hyperparameter* for the noise variance. For finding the posterior distributions of new test points for noisy data, we simply replace the $\mathbf{K}(\mathbf{x},\mathbf{x})$ term in the above equations with, $\mathbf{V}(\mathbf{x},\mathbf{x})$ from Eq. 2.11.

Can volatility forecasts be made at the minute level using the ability of Gaussian processes to predict time series? In order to explore this question the financial returns R can be provided as input to a Gaussian process and the mean μ and standard deviation σ from the forecast output can be used as potential tools in predicting the underlying financial volatility of the input series. The next section introduces GARCH models. These are used in order to establish if Gaussian process performance is comparable to existing established volatility forecasting techniques.

2.4 GARCH models for Financial Volatility

This section is a brief overview of GARCH, GJR-GARCH and EGARCH.

The most commonly used model for financial data is GARCH. It assumes that the returns arise from a zero mean, and heteroskedastic Gaussian distribution. Defining a white noise (Wiener) process r_t, as:

$$r_t \sim \mathcal{N}(0,\sigma_t^2) \tag{2.12}$$

The GARCH(p,q) model predicts the underlying variance of a time-series as a linear (auto-regressive) combination of p past observed variances and a moving average of the noise process, through q previous squared returns. This linear combination is uniquely defined by coefficient set $\alpha_{0:q}$ and $\beta_{1:p}$, as:

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j r_{t-j}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \tag{2.13}$$

GARCH, however, suffers from two major drawbacks: first it does not distinguish between the different effects of negative vs positive returns on volatility, second it assumes a linear relationship between the variables. Several mutations of the GARCH have been proposed to overcome these limitations and two of the most commonly used GARCH variants are EGARCH and GJR-GARCH [30, 33, 32]. The exponential GARCH is given by

$$\log(\sigma_t^2) = \alpha_0 + \sum_{j=1}^{q} \alpha_j g(r_{t-j}) + \sum_{i=1}^{p} \beta_i \log(\sigma_{t-i}^2) \tag{2.14}$$

It adds flexibility to the GARCH model by allowing for the asymmetric effect to be captured using $g(x_t)$ where a negative value of θ in

$$g(x_t) = \theta r_t + \lambda |r_t| \tag{2.15}$$

will cause the negative returns to have a more pronounced effect on the volatility. GJR-GARCH, defined by:

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \sum_{j=1}^{q} \alpha_j r_{t-j}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{k=1}^{r} \gamma_k r_{t-k}^2 I_{t-k} \\ I_{t-k} &= \begin{cases} 0, & \text{if } r_{t-k} \geq 0 \\ 1, & \text{if } r_{t-k} < 0 \end{cases} \end{aligned} \tag{2.16}$$

tries to capture the different effects of positive and negative returns by using a leverage term which only activates in the case of the return being negative.

2.5 Performance and Accuracy Metrics for Forecast Evaluation

In order to assess the performance of volatility forecasting techniques in a holistic manner there are two main considerations. The first is the ability of the forecasting technique to

minimise loss on a set of loss functions when compared against an observed proxy for volatility. Second the forecast of volatility should generate a conditional distribution that is in line with our assumptions about the latent volatility distribution.

For our loss function we choose mean squared error (MSE), root mean squared error (RMSE), and QLIKE. The selected loss fucntions are given below, where $\widehat{\sigma}_t$ is the proxy for volatility and h_t is the volatility forecast at time t. The best forecast $\overset{*}{h}_t$ will minimize MSE, RMSE, and QLIKE

$$MSE = \frac{1}{n}\sum_{n}(\widehat{\sigma}_t^2 - h_t^2) \tag{2.17}$$

$$RMSE = \frac{1}{n}\sum_{n}\sqrt{\widehat{\sigma}_t^2 - h_t^2} \tag{2.18}$$

$$QLIKE = \frac{\widehat{\sigma}_t}{h_t} - \log\frac{\widehat{\sigma}_t}{h_t} - 1 \tag{2.19}$$

In a seminal work on robust loss functions, Patton (2011) [51] shows that MSE and QLIKE belong to a class of robust loss functions which rank the accuracy of volatility forecasting models correctly even when a noisy proxy of volatility is used - provided that the proxy is unbiased. In this work squared returns and realized volatility are shown to be unbiased proxies for the underlying squared volatility σ^2. Although absolute returns are a biased proxy for volatility and a robust class of loss functions has not been developed specially for these, in Nagata & Oya (2012) [47] a general class of robust loss functions is developed which can be fine tuned based on the presence of bias in the volatility proxy. They find that MSE and QLIKE can still rank effectively in the presence of proxy bias, although QLIKE is more sensitive to the bias than MSE. RMSE was chosen in addition to MSE and QLIKE to enable comparison with other literature generated in this area.

The significance of the results is highlighted by looking at the notched box plots for the results. In a notched box plot, as shown in Fig. 2.1, the box shows the interquartile range (IQR). The IQR is the 25th to 75th percentile also known as Q1 and Q3. The IQR is where the center 50% of the given data points will fall. The whiskers at each end of the plot add 1.5 times the IQR to the 75th percentile - Q3 and subtract 1.5 times the IQR

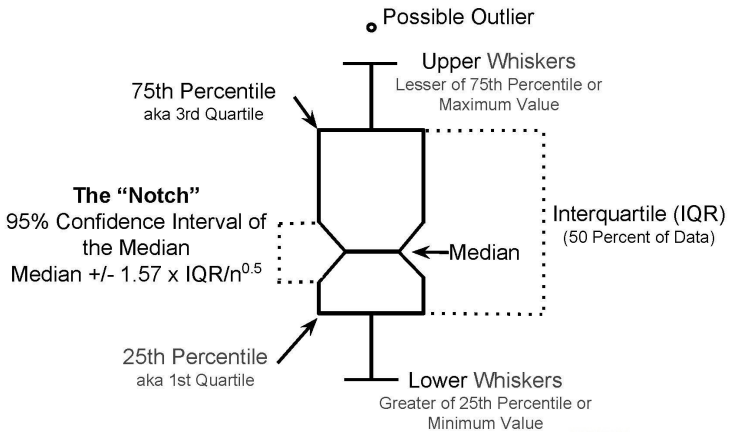


Figure 2.1: Explanation of the notched box plot for testing significance of results [13]

from the 25th percentile - Q1. The whiskers should include 99.3% of the data if from a normal distribution. The line in the centre shows the median of the data The notch displays the a confidence interval around the median. According to Graphical Methods for Data Analysis (Chambers, 1983) if the notches do not overlap with a given value there is 'strong evidence' (95% confidence) that the median is significantly different from that value. For the results under consideration if zero lies outside the notches then the resulting gain is significant at 95%.

Second, in order to assess the distributional characteristics of the forecasts the conditional residuals can be analysed. GARCH and its variants are well established to produce Gaussian residuals. To establish that forecasts produced by a given method are unbiassed, element wise normalization of the actual returns r_t is carried out, using the forecast time series h_t, as given by:

$$e_t = r_t / h_t \tag{2.20}$$

The residuals e_t obtained from the normalization should be close to a white noise process, therefore they should be Gaussian distributed with zero mean and unit standard deviation given by:

$$e_t \sim \mathcal{N}(0,1) \tag{2.21}$$

2.6 Datasets for Understanding Performance of Gaussian Processes

The full data set includes data for 50 stock market symbols, containing 23 individual stocks, 13 foreign exchange rate pairs (forex), 12 stock market indices, and 2 commodity indices, sampled at 1-minute frequency. The time period for the symbols is 10 years and 3 months, from 01 January 2006 to 31 March 2017. This period includes many important events in the recent history of financial markets such as the Great Recession (2007-2012), the Brexit referendum (2016), and the 2016 US presidential election. The complete data set has approximately 100 million price data points. The names of the selected symbols are given in Table 2.1.

The breadth of the dataset helps in understanding the performance of Gaussian Processes in forecasting volatility across a wide range of financial time series. This helps to ensure that any performance boost from Gaussian Processes is consistent across many types of financial series and does not overfit to one particular instance of a single type of asset class. The data is sampled at 1-minute frequency. This provides reasonably high frequency time series without having significant interference from market microstructure.

The 24 stock symbols chosen are the ones that constituted the top one-third of the S&P 500 by weight as of March 2017. The selected symbols were part of the index for the whole period between 2006 to 2017 to ensure consistency in the year to year comparisons. Two key members of the current S&P 500 were excluded due to this filter - Facebook and Google. Facebook was excluded as it was listed on NASDAQ in 2012. Google was excluded because its unconventional stock split in 2014 created new share classes which complicated the relationship of the data to the pre-2014 time series.[2] All the stock symbols are listed on either the New York Stock Exchange (NYSE) or Nasdaq Stock Market (NASDAQ). Their trading day lasts from 9:30 to 16:00 hrs adjusted for daylight savings as needed.

[2]This split was unconventional for that time in that Google created a brand-new class of shares, rather than simply doubling the number of shares outstanding. Due to the this non-standard move Google is excludedfrom the stocks dataset.

The 12 forex pairs included in the data set are the most regularly traded forex pairs. They provide a high coverage of the international forex market. The chosen pairs account for over 70% of all global forex trades with over $3.5 trillion in daily average turnover. The Reuters database provided the 1-minute frequency and the Federal Reserve Bank of St. Louis Economic Database (FRED) provided the daily rates. The data acquired from FRED is the noon buying rates in New York City for cable transfers payable in foreign currencies, and is not seasonally adjusted. Three notable exclusions are Chinese Yuan, Danish Kroner, and Swedish Kronor for which the data obtained from TRTH had data quality issues in the chosen time period, due to which I chose to not include them in the experiments.

The 12 international market indices were selected to cover the major global stock exchanges. There are 3 indices from USA, 5 from Europe, and 4 from Asia and Australia.

The 2 commodity indices include the Philadelphia Gold and Silver index (XAU), SPDR Gold Trust (GLD) which is an ETF with the aim of reflecting the performance of the price of gold bullion.

2.7 Using Gaussian Processes for Forecasting Financial Volatility

In order to understand the baseline performance of Gaussian processes on financial returns data three approaches are explored, one where the unaltered returns (with both positive and negative values, called *signed returns* from here onwards) are used as input and the second approach in which absolute returns are used as input, and third where the squared returns are used as input to Gaussian processes. A moving window approach is used to estimate sequential forecasts, this shows how the Gaussian process will work with live streaming data.

In the first approach, the unaltered returns series $\mathrm{R}_t^{(m)}$ (denoted from now by R_t for simplicity) is given as input to the GP to predict the volatility at time $t+1$, where

Category	Symbols
Stocks	Apple (AAPL), Amazon (AMZN), Bank of America (BAC), Berkshire Hathaway (BRKb), Comcast Corporation (CMCSA), Cisco (CSCO), Chevron (CVX), General Electric (GE), Home Depot (HD), Intel (INTC), Johnson & Johnson (JNJ), JPMorgan Chase (JPM), Coca-Cola (KO), Merck (MRK), Microsoft (MSFT), Pfizer (PFE), Proctor & Gamble (PG), Philip Morris (PM), AT&T(T), Visa (V), Verizon (VZ), Wells Fargo (WFC), Exxon Mobil (XOM)
Forex (against USD)	Australian dollar (AUD), Canadian dollar (CAD), Swiss Franc (CHF), Euro (EUR), British pound (GBP), Japanese yen (JPY), Norwegian krone (NOK), New Zealand dollar (NZD), Polish zloty (PLN), Singapore dollar (SGD), Turkish lira (TRY), South African rand (ZAR)
Market indices	S&P500 (INX), NASDAQ (IXIC), German DAX (GDAXI), All Oridinaries (AORD), BSE Sensex (BSESN), Austrian Traded Index (ATX), Swiss Market Index (SSMI), Hang Seng Index (HSI), Nikkei 225 (N225), Dow Jones Inudstrial Average (DJA), Cotation Assistée en Continu (CAC40), Financial Times Stock Exchange 100 Index (FTSE)
Commodity indices	Philadelphia Gold and Silver index (XAU), SPDR Gold Trust (GLD)

Table 2.1: List of financial market symbols in the dataset

$R_t^{(m)} = \{r_t, r_{t-1}, \cdots, r_{t-m+1}\}$. The standard deviation σ of the GP output is used as the volatility forecast. This volatility forecast is then compared to $|r_{t+1}|$, which is the proxy for volatility at time step $t+1$ and accuracy metrics are calculated.

In the second approach, the absolute returns series $|\mathrm{R}_t|$ is given as input to the GP to predict the volatility at time $t+1$. In this case the mean μ of the GP output is used as the volatility forecast. This volatility forecast is compared to $|r_{t+1}|$, as before.

In the third approach, the squared returns series R_t^2 is given as input to the GP to predict the volatility at time $t+1$. In this case the mean μ of the GP output is used as the volatility forecast. This volatility forecast is compared to r_{t+1}^2.

In order to forecast volatility at time $t+1$ the sequence $R_t^{(m)} = \{r_t, r_{t-1}, \cdots, r_{t-m+1}\}$ is provided as the training data, where m is the size of the training window. The absolute or square of the sequence is taken as required by the approach being tested. The forecast obtained from the GP is used as the is compared to $|r_{t+1}|$ to calculate the loss functions. At the next time step $t+2$, i.e. the next minute of the trading day in this case, the 1-step moved window of the returns sequence $R_{t+1}^{(m)} = \{r_{t+1}, r_t, \cdots, r_{t-m+2}\}$ is provided to the GP. This accuracy of this forecast is compared to $|r_{t+2}|$, and so on.

When applying GPs to financial forecasting, a primary consideration is that selection of an appropriate kernel to best suit the data. This encodes prior knowledge about the function space. Three different kernels were compared in order to identify the best one: the squared exponential (SE) kernel, the Matérn-3/2 kernel ($M_{3/2}$), and the quasi-periodic (QP) kernel outlined in [57]. The kernel that minimizes the MSE is then selected for the rest of the experiments.

The SE kernel was added to the comparison since it is the most commonly used GP kernel where the data follows a smooth function. Even though returns don't show a smoothly varying time series, the SE serves as the baseline for comparing the performance of the other two kernels. The $M_{3/2}$ kernel is compared since it is once differentiable the way a persistent volatility function might be expected to behave. Although the returns themselves might be modelled better by the Matérn-1/2 kernel (Brownian motion or Ornstein-Uhlenbeck) kernel, the underlying volatility has higher persistence and is ex-

pected to follow a once differentiable $M_{3/2}$. The quasi periodic kernel is tested to see if its ability to capture periodicity can be used to fit to the underlying periodic persistence in the volatility.

Let $d = |x_i - x_j|$, then the three kernels are given as,

$$k_{SE} = \sigma_h^2 \exp\left[-\left(\frac{d}{\sqrt{2}l}\right)^2\right] \tag{2.22}$$

$$k_{\text{Matérn},\nu=3/2} = \left(1 + \frac{\sqrt{3}d}{l}\right)\exp\left[-\left(\frac{\sqrt{3}d}{l}\right)\right] \tag{2.23}$$

$$k_{QP} = \sigma_h^2 \exp\left(-\frac{\sin^2[\pi d/T]}{2w^2} - \frac{d^2}{l^2}\right) + I\sigma_n^2 \tag{2.24}$$

where σ_h is the output scale, l is the input length scale and σ_n^2 is the noise variance. The period is given as T is the period and w is a hyperparameter meant to capture the roughness relative to the period, for a further discussion of this kernel, see [57]. The performance of the kernels is tested on a sample of the data and the performance of the three kernels for the forex is shown in Fig. 2.2. The kernel comparison shows that the Matérn-3/2 gives the best perfomance. It can be seen that generally the Matérn-3/2 gives the best performance out of the three kernels, while the quasi-periodic kernel performs the worst, as can be seen in figure 2.2. All further experiments are conducted using the Matérn 3/2 kernel. The kernel hyperparameters are recalculated at each time step.

For the absolute and squared returns method, the constant setting is used for the mean function, while the zero mean setting is used for the signed returns method. For each run of the experiment a full day's worth of training minutes are forecast step by step. This full set of forecasts and the associated observed returns are used pairwise to calculate the loss function value for that trading day. In order to assess the impact of different lenghts of the moving window, each trading day is tested using 12 different lengths of the moving window m are tried from 30 - 360 minutes, in 30 minute increments.

While testing with longer window lengths Cholesky updating and Cholesky downdating [49] is deployed to update the covariance matrix instead of recomputing it at each

forecasting time step. This ensures that after the initial learning phase has been carried out, all future updates occur at a rapid pace making the technique extremely suitable for online use, with live feed data.

Over the 11 years of data available a subset of 75 days is selected for each symbol, by picking one day randomly from every two months to ensure uniformity of coverage. This helps ensure that the performance of GPs is tested on a large set of the available data. Using this approach more than 48K experiments are carried out in total for the 50 symbols, with 75 trading days per symbol, and 12 window lengths tested per trading day.

The value of the loss functions from the GP forecasts are then compared to the value of the loss functions from the benchmarks. The gain or loss is quantified as a percentage. For example, for MSE this is quanitfied as

$$\%\text{Improvement} = 100 \times (1 - \frac{MSE_{GP}}{MSE_{benchmark}}) \tag{2.25}$$

where a positive value means the GP had a gain in forecasting accuracy compared to the benchmark, whereas a negative value means that the GP performed worse than the benchmark.

All computation with Gaussian processes is implemented using GPflow python library [17]. In the GPflow version used, the hyperparameters are optimized with the default scipy implementation of L-BFGS-B (a Limited-memory implmentation of the Broyden–Fletcher–Goldfarb–Shanno algorithm with bound constraints). The rugarch package [28] is used to implement the GARCH benchmarks.

2.8 Performance of GPs vs GARCH

The performance of GPs is found to be superior to that of GARCH, EGARCH, and GJR-GARCH across all three approaches Tables 2.2, 2.3, and 2.4 summarise the gains in forecasting performance of GPs over the other benchmarks. The gains have been summarised using the symbol categories - stocks, indices, and forex.

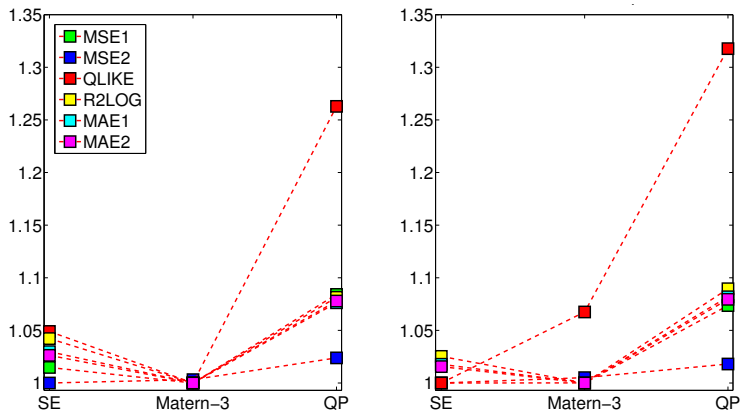


Figure 2.2: Loss function performance of the three kernels for absolute returns (left) and squared returns (right)

Generally it is observed that improvement against vanilla GARCH are the highest, followed by those on GJR-GARCH and EGARCH. Performance improvement is dominated by the gains in stocks. The improvement in indices and forex while less than that of stocks, is still consistent. A comparative visualization of the forecasts by GP vs GARCH is shown in Fig. 2.3.

In order to ensure the significance of the results, the notched box plots are presented in Fig. 2.5. It can be seen that notches of all medians lie away from zero. This establishes that the results are significant with 95% confidence.

The unbiasedness of the residuals obtained from GP regression is shown in figure 2.4. The distributions of the residuals of GARCH, EGARCH and GJR-GARCH are also shown for comparison. It can be observed that the distributions from the GP residuals closely follow a normal distribution establishing that the GP forecasts are unbiased at a comparable level to the GARCH models.

The challenge for these positive results is that GARCH, GJR-GARCH, and EGARCH have not been developed for intraday volatility forecasting. Even though they are the most commonly used benchmarks for comparison in machine learning literature looking at financial data, performance gains against them are not a like-for-like comparison. A more robust model developed for the intraday volatility is mcsGARCH. This stronger contender is introduced in the next section along with two other benchmarks for intraday volatility.

	Improv. over GARCH	Improv. over GJR-GARCH	Improv. over EGARCH
MSE			
Stocks	4.31%	2.89 %	2.12%
Indices	3.08%	1.52%	1.73%
Forex	2.57%	1.03%	1.08%
RMSE			
Stocks	2.92%	1.94%	1.77%
Indices	1.65%	1.53%	1.38%
Forex	1.14%	1.29%	0.96%
QLIKE			
Stocks	3.65%	2.42%	2.96%
Indices	2.51%	2.01%	1.37%
Forex	2.08%	1.96%	0.71%

Table 2.2: Average values for the percentage improvement in MSE, RMSE, and QLIKE of the GPs using signed returns

	Improv. over GARCH	Improv. over GJR-GARCH	Improv. over EGARCH
MSE			
Stocks	3.97%	1.94 %	2.31%
Indices	4.22%	2.03%	1.44%
Forex	3.91%	1.36%	1.20%
RMSE			
Stocks	2.46%	1.27%	1.03%
Indices	2.44%	1.60%	0.78%
Forex	1.02%	1.11%	0.66%
QLIKE			
Stocks	3.13%	2.85%	2.50%
Indices	3.59%	2.11%	1.81%
Forex	2.17%	2.19%	0.99%

Table 2.3: Average values for the percentage improvement in MSE, RMSE, and QLIKE of the GPs using absolute returns

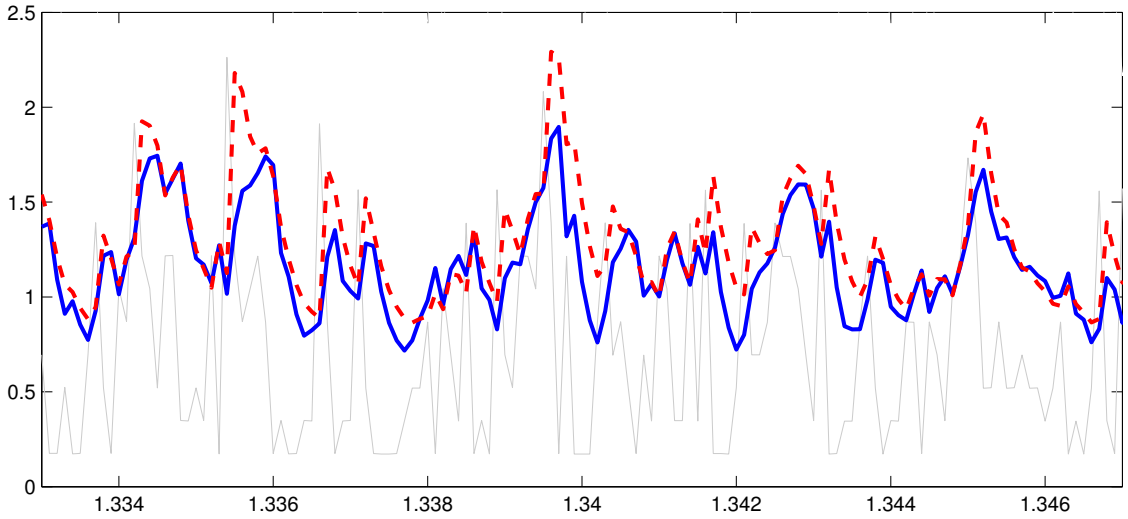


Figure 2.3: GP Volatility forecasts (blue) obtained by regressing over absolute returns (grey), magnitude vs time, GARCH forecasts (red) are shown for comparison.

2.9 Stronger Benchmarks for Volatility Forecast Evaluation

In order to gauge the performance of GPs against like-for-like benchmarks, the forecasts from the Gaussian Process models are evaluated against three main benchmarks - mcsGARCH which is a version of GARCH specially suited to minute level data, a rolling window moving average, and using the last observed return as forecast for the next value. These three are strong contenders for volatility benchmarking at intraday level. Any gains against these would be significant in establishing the performance of forecasting on intraday data.

2.10 The Multiplicative Component GARCH Model mcsGARCH

A newer variant of GARCH(1,1) model is the Multiplicative Component GARCH (mcsGARCH). Engle & Sokalaska (2012) [26] decompose the volatility of high frequency asset returns into multiplicative components: daily, diurnal, stochastic intraday volatility

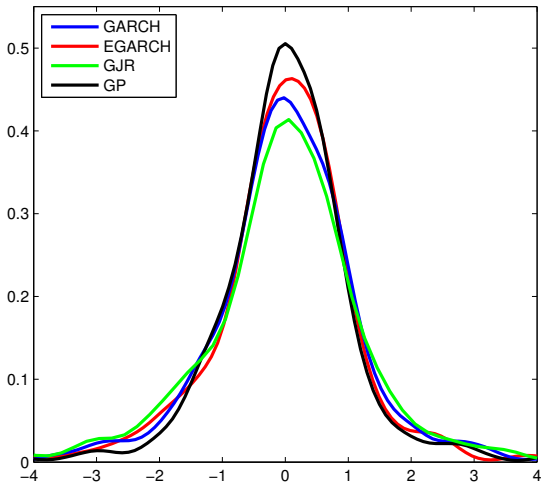


Figure 2.4: Probability distribution of the residuals that are obtained from normalizing the returns by the forecasts produced by GP regression. The residuals produced by GARCH and its variants are provided for comparison to establish unbiasedness of GPs for volatility prediction

components, that can be easily estimated. Andresen and Bollerslev (1997, 1998) proposed similar approach of accounting for intraday volatility using multiplicative components. Since this GARCH variant has been specifically devised for intraday volatility therefore gains against this would be a superior indicator of performance than the benchmarks used before.

The mcsGARCH model defines the intraday return $\mathrm{R_{i,t}}$ process for trading day i as follows

$$\mathrm{R_{i,t}} = \sqrt{h_i s_t q_i}\varepsilon_{i,t}$$

Where s_t is the diurnal variance component for each intraday period, $q_{i,t}$ is the stochastic intraday variance component, and $\varepsilon_{i,t}$ is an error term such that $\varepsilon_{i,t} = a_i z_{i,t}$ where $z_{i,t} \sim \mathcal{N}(0,1)$ and $a_{i,t} = h_i s_t q_{i,t}$

The relationship can be rewritten as

$$s_t = \frac{1}{N}\sum_{t=1}^{N}\frac{R_{i,t}^2}{h_i}$$

	Improv. over GARCH	Improv. over GJR-GARCH	Improv. over EGARCH
MSE			
Stocks	3.01%	2.33%	1.93%
Indices	3.56%	1.81%	2.36%
Forex	2.91%	1.19%	1.10%
RMSE			
Stocks	2.42%	2.05%	1.41%
Indices	2.93%	1.78%	2.22%
Forex	1.99%	1.25%	0.61%
QLIKE			
Stocks	3.55%	2.30%	1.65%
Indices	2.61%	2.10%	1.79%
Forex	2.08%	1.84%	1.16%

Table 2.4: Average values for the percentage improvement in MSE, RMSE, and QLIKE of the GPs using squared returns

The residuals are divided by the diurnal and daily volatility to normalize them

$$\bar{\varepsilon_{i,t}} = \frac{\varepsilon_{i,t}}{h_i s_t} = q_{i,t} z_{i,t}$$

The stochastic component of volatility $q_{i,t}$ is modelled by a simple GARCH(1,1) process as

$$q_{i,t}^2 = \omega + \alpha \bar{\varepsilon_{i,t}}^2 + \beta q_{i,t-1}^2$$

The rugarch package [28] is used to implement the mcsGARCH. The implementation requires daily values of data for estimating the daily volatility component. This is taken from Google Finance API (for stocks and indices) or FRED (for forex) depending on the symbol under consideration. To calculate the diurnal component of volatility, 2 months worth of minute level returns data is provided to the mcsGARCH implementation.

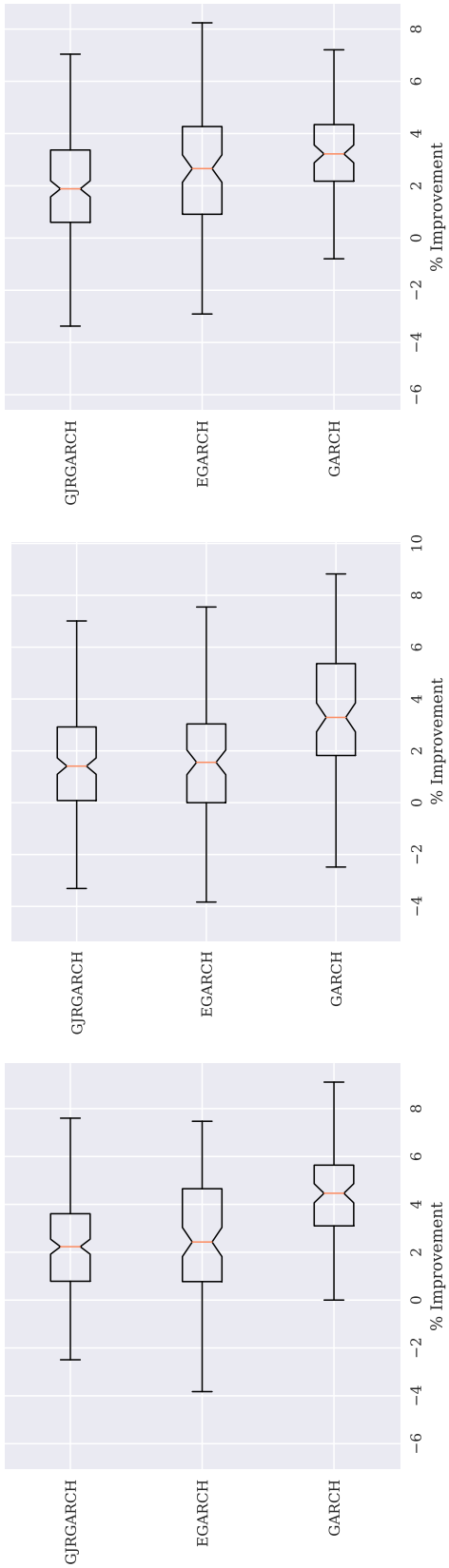
2.11 Results Against the Stronger Benchmarks

These three approaches of using Gaussian processes to predict volatility did not perform well when compared to these stronger benchmarks. The results are summarized in Fig. 2.7 and Fig. 2.6 It can be seen that these approaches perform approximately 100 - 230% worse than these stronger benchmarks on average. Although these results are negative, this is the first time a study of this scale has been carried out for guaging the performance of Gaussian processes against a robust set of volatility benchmarks. This helps in ruling out the naive application of Gaussian processes to signed returns, absolute, or squared returns as a viable method of forecasting financial volatility. These negative results further establish a new benchmark that any GP based forecasting technique must meet in order to show comparable or superior performance.

A new approach based on GPs is developed in the next chapter by building on the negative findings here.

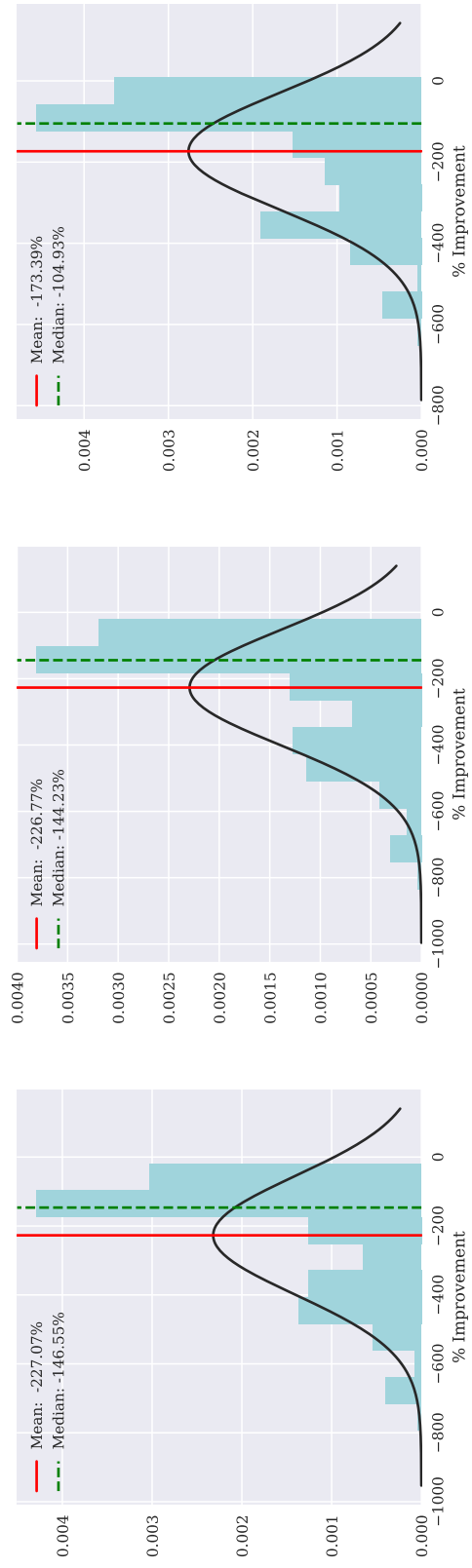
2.12 Conclusion

In this chapter it was established that while Gaussian processes perform well in comparison to GARCH, GJR-GARCH, and EGARCH they perform poorly when benchmarked against more robust volatility models such as mcsGARCH, mcsGARCH, rolling average, and using the last observed value as the next step forecast. In the next chapter we outline a more nuanced approach for using GPs with returns data, which produces significant gains in forecasting performance against all benchmarks.



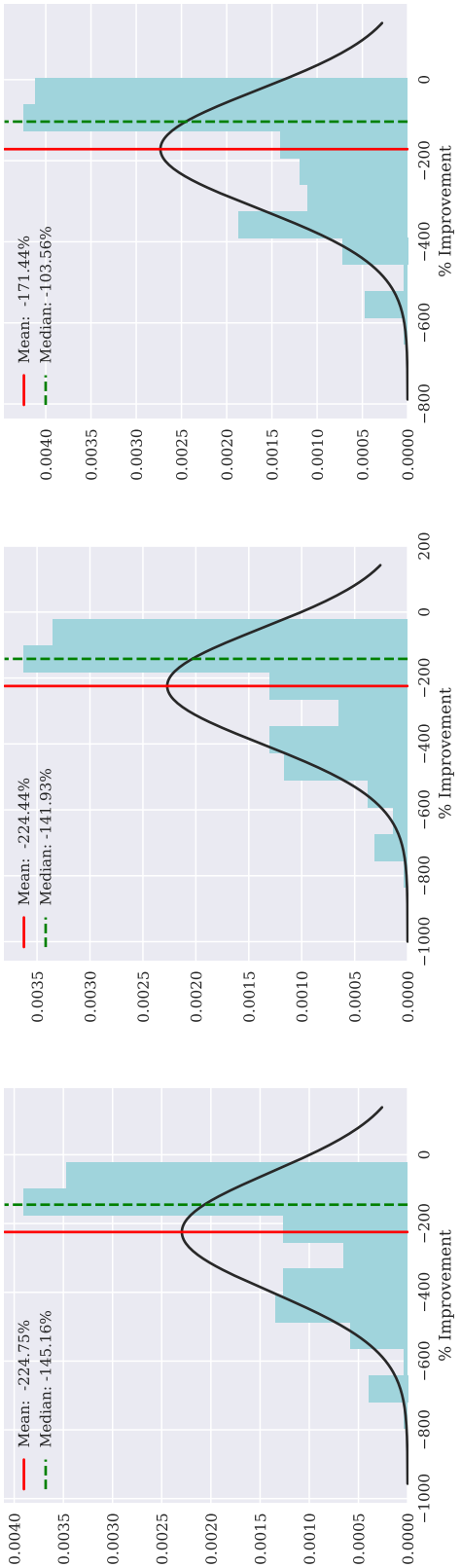
(a) Signed returns (b) Absolute returns (c) Squared returns

Figure 2.5: Improvement in MSE by using Gaussian processes, compared to benchmarks



(a) mcsGARCH (b) Rolling average (c) Previous value

Figure 2.6: Improvement in MSE using envelope method, compared to benchmarks



(a) mcsGARCH (b) Rolling average (c) Previous value

Figure 2.7: Improvement in MSE using envelope method, compared to benchmarks

Chapter 3

Envelope Based Gaussian Processes

Objective

The aim of this chapter is to develop a forecasting approach that consistently performs better than strong competing benchmarks. In this chapter this new technique for forecasting financial volatility is developed by using Gaussian processes. The positive and negative returns of the time series are forecast separately and a transformation of the data space is also used in order to make the application of GPs possible. The aim is to carry out a highly extensive set of experiments using Gaussian processes for financial volatility forecasting on the available data.

3.1 Extensions to the Gaussian Process Model

A key characteristic of volatility is that it is affected asymmetrically by positive and negative returns [54]. Empirical evidence shows a negative relation between market returns and volatility [36]. Negative returns are found to increase volatility more than positive returns Recent literature suggests behavioural explanations [42, 36], leveraging effects and volatility feedback [6, 65] could be the cause of the asymmetric effect. To address this asymmetric response of volatility various volatility forecasting techniques have been developed that treat the positive and negative returns separately. An overview of these

asymmetric volatility forecasting is provided in [6].

Simply using absolute or squared returns does not address this asymmetric affect of returns on volatility. The underlying dynamics of the positive returns can be quite different from each other in terms of distribution and time evolution. These dynamics can be captured separately by separating the returns time series into its positive and negative parts and then using each to predict the volatility at the next step.

A new approach is developed here in order to address the positive and negative returns separately by using Gaussian processes. In order to calculate the volatility forecast at time $t+1$, the return time series R_t is separated into its positive returns R_p and negative returns R_n depending on the sign of the return.

$$R_t = \begin{cases} R_n, & \text{if } r_t < 0 \\ R_p, & \text{if } r_t \geq 0 \end{cases}$$

A sample of separated negative and positive returns can be seen in figure 3.1.

It should be noted that R_p and R_n are essentially the upper and lower envelops of the complete returns time series R as shown in Fig. 3.1 The absolute value of each envelope $|R_p|$ and $|R_n|$ is then input to two separate GPs giving rise to four outputs, the μ_p and σ_p from the positive envelope, and μ_n and σ_n from the negative envelope. The average of $|\mu_n|$ and $|\mu_p|$ is taken as the volatility forecast h_{t+1} for the next time step. This volatility forecast is compared to $|r_{t+1}|$ for calculating the loss functions. In order to simplify comparison only the absolute returns as proxy for volatility approach is retained from the last chapter.

Transforming the Forecast Space

Noting that volatility defined via absolute returns is a strictly positive quantity, therefore GP regression is performed on the log-transformed absolute returns. This has the major advantage of enforcing the positivity constraint on the solutions while retaining a standard GP in the log-space. Predictive measures of uncertainty in the log-space are then readily

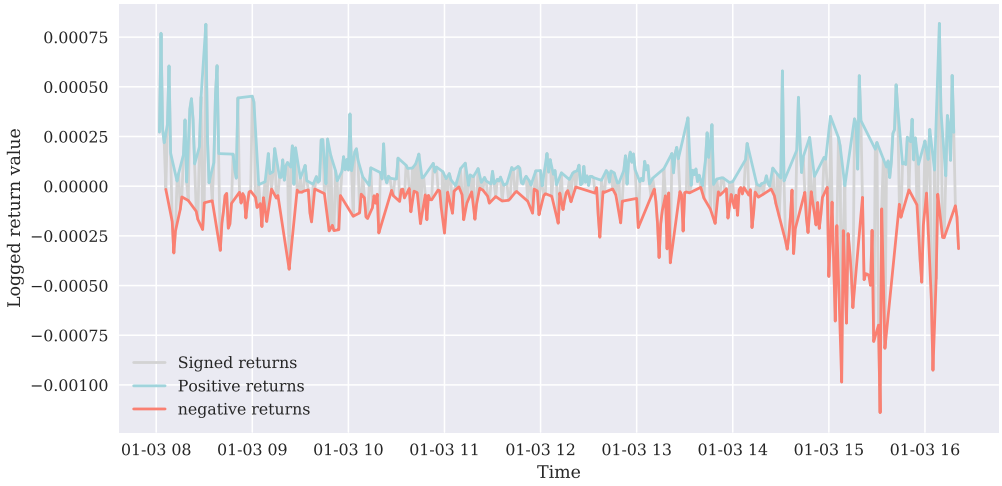


Figure 3.1: Positive (blue) and negative (red) returns separated into two envolopes for DAX on 3rd Jan 2006 from 08:00 - 16:30 hrs UTC.

transformed back, noting that they form asymmetric bounds.

Defining the target variable y as $y = \log(|r_t|)$, and carrying out Gaussian process regression over the set of observed y. It remains a simple task then to transform the next step forecast and credibility intervals of the predictive distribution on y to ones on $|r|$ using exp() to invert the log transform as:

$$\bar{r}_* = \exp(\bar{f}_*) \tag{3.1}$$

and the upper and lower intervals can be recovered as:

$$c_{up} = \exp(\bar{f}_* + 1.96\sigma_*) \tag{3.2}$$

$$c_{low} = \exp(\bar{f}_* - 1.96\sigma_*) \tag{3.3}$$

Here the 95% intervals are chosen and readily obtained from scaling the predictive standard deviation on y by 1.96, and $\sigma_*^2 = \mathbb{V}[f_*]$ from Eq. (2.7).

The effect of this transformation on the negative and positive return envelopes can be seen in the QQ plots in Fig. 3.2. The envelopes of the financial time series display a log-normal distribution as seen in the figure, and the logged envelopes are closer to normal.

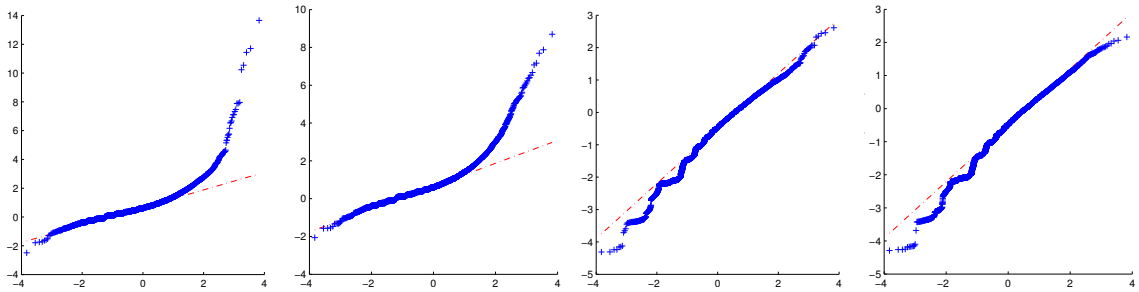


Figure 3.2: QQ plots vs the standard normal, standard normal quantiles on the horizontal axis and quantiles of the input sample on the vertical (a) QQ plot of the maxima (upper) envelope. (b) QQ plot of the minima (lower) envelope. (c) QQ plot of the log of the maxima (upper) envelope. (d) QQ plot of the log of the minima (lower) envelope.

Therefore these logged envelopes are used as input for forecasting with GPs.

3.2 Methodology for Experiments

In this chapter the same 50 symbols outlined in Table 2.1 are used. The same methodology as Sec. 2.7 is used with a moving window of training data to estimate sequential forecasts.

Each window the returns training data is split into two parts - the positive and negative returns. These two training sets are then input to two separate GPs each carrying out its own hyperparameter optimization. To generate the forecast for the next time step the average of the two μ values generated by each GP

It should be noted that in this approach two GPs are used to generate the forecast for each time step, but each GP works on half of the data set. The matern 3/2 kernel is used for each GP and the hyperparameters are recalculated at each time step. The constant mean function setting is used in GPflow.

As before, a full day's worth of training minutes are forecast step by step for each run of the experiment. This full set of forecasts and the associated observed returns are used pairwise to calculate the loss function value for that trading day and the same 12 different lengths of the moving window m are tried from 30 - 360 minutes, in 30 minute increments. The significance of the results is tested by looking at the notched box plots for the results.

In order to assess the distributional characteristics of the forecasts the conditional

residuals are also analysed. Since we defined $r_t = \overset{*}{\sigma}_t \varepsilon$, where $\varepsilon \sim \mathcal{N}(0,1)$, we can assess the probability distribution of the pairwise division of the return series R by the forecast series H. It can be expected that as the forecast accuracy improves the remainders from the pairwise division approach the normal distribution i.e. $\mathrm{P}(\mathrm{R}/\mathrm{H}) \to \mathcal{N}(0,1)$. To test this, the first, second, and third standard deviation of the distribution of remainders from the pairwise division is considered. The ideal forecast series $\overset{*}{\mathrm{H}}$ should give rise to 68%, 95%, and 99.7% for this metric.

3.3 Discussion of results

A set of 48K experiments is carried out using the data sampling specifications laid out in the previous chapter.

This set of experiments represents many empirical firsts. It is the first time that such an extensive set of experiments has been carried out using Gaussian processes for financial volatility forecasting. This also represents the first time such an extensive comparison has been done of mcsGARCH and competing benchmarks on minute level financial data over a period as long as 11 years. This is also the most extensive study of volatility estimation done at the minute level across any other forecasting techniques.

Analysis of the results reveals that the envelope based GP approach for forecasting volatility performs significantly better than all the other benchmarks. The improvement in the loss functions produced by the envelope GP method over the benchmarks is summarised in Table 3.1. The distribution of the percentage improvement for MSE of the envelope approach over the other benchmarks is shown in Fig. 3.8. Across all the experimental results, on average the MSE values from this method are 12.30% better than the mcsGARCH, 12.56% better than using a rolling average, and 26.89% better than using the previous observed value as our volatility forecast.

In order to establish the significance of the performance gains notched box plots are presented in Fig. 3.3. As can be seen in all three notched plots, the results are found to be highly significant at the 95% confidence interval across all comparisons since the notches

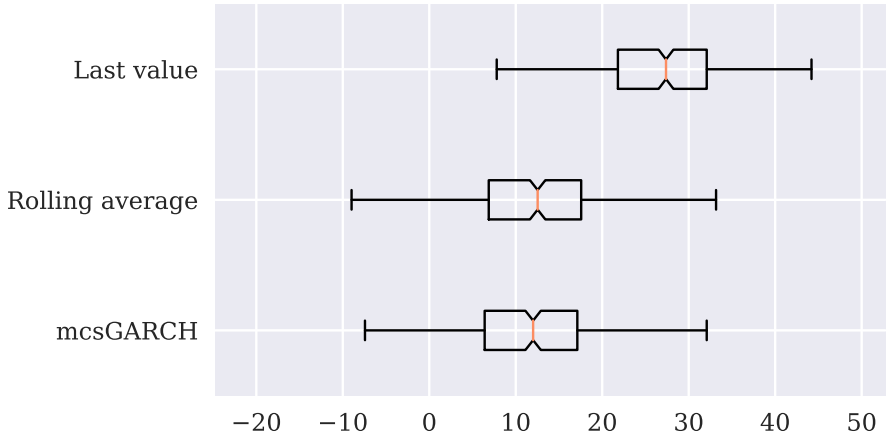


Figure 3.3: Improvement in MSE using envelope method, compared to mcsGARCH, rolling average, and last observed value benchmarks. The results are highly significant since the notches around all the median are in the positive and well beyond zero.

have no overlap with zero in any of these cases. This establishes that the results can be considered significant and that envelope based GPs perform better than mcsGARCH, rolling average, and using last observed value as volatility forecast. It can be seen that the highest gains are in comparison to the last observed value benchmark. The gains over rolling average and mcsGARCH are similar around 12%. This could also indicate that at the intraday level the rolling average for short window lengths is encoding short term returns dynamics in a manner comparable to the mcsGARCH.

In order to gauge the effect of the look back window the results are separately analysed for each window length, as shown in Fig. 3.9. At all lengths there are significant gains over all three benchmarks. A closer inspection reveals that shorter look back windows seem to produce increasingly higher gains in performance for the GPs. This can indicate that the GPs are able to fit to the local dynamics better than the other benchmarks with shorter training windows. This may also indicate that the information gained from the look back windows decays sharply after a few minutes of past data have been included in the training set.

The performance gain is also broken by symbol type as shown in Fig. 3.10. In the comparison with mcsGARCH it can be seen that the performance gains are highest on stocks (14%) followed by market indices (11%), with lower but still significant gains

observed on forex data (7%). This is similar to the gains observed in the comparison against other GARCH benchmarks in the previous chapter where it was found that gains on stock data usually exceed the ones on indices and forex. A deeper analysis of this artefact found that the return envelops of indices and forex do not become normalized in the log space to the same extent as the ones from stocks. This discrepancy can help explain why the gains are not found to be equal across all these different types of symbols.

The symbol wise results for a selection of symbols are shown in Fig. 3.4.

Figure 3.5 shows the 95% variance bounds in normal space, that we obtain when we regress using GPs in log-space, as can be seen that the variance bounds are asymmetric and capture the envelope and thus the variance of the absolute returns. Fig. 3.7 shows a comparison of mcsGARCH and GP predictions.

The highly significant results in this section establish that GPs used on the separated positive and negative returns in log space represent a strong model for forecasting financial volatility across a breadth of assets.

	Improv. over mcsGARCH	Improv. over rolling average	Improv. over previous value
MSE			
Stocks	14.23%	14.85%	28.66%
Indices	10.98%	10.51%	23.19%
Forex	07.00%	06.52%	23.07%
RMSE			
Stocks	06.09%	06.31%	15.72%
Indices	05.31%	05.03%	13.49%
Forex	1.54%	01.33%	11.96%
QLIKE			
Stocks	08.04%	08.42%	17.96%
Indices	05.06%	5.71%	12.13%
Forex	03.68%	03.92%	9.87%

Table 3.1: Average values for the percentage improvement in MSE, RMSE, and QLIKE of the GP envelope method over the benchmarks.

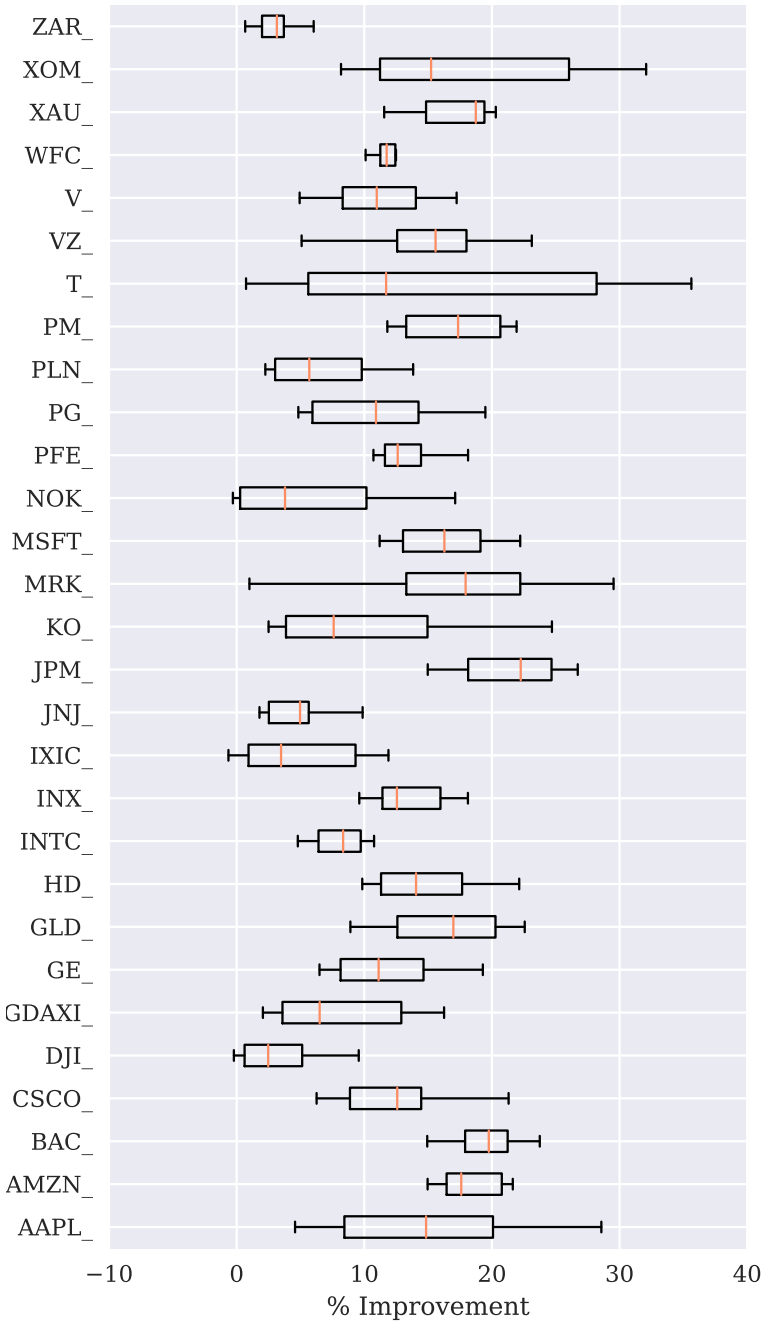


Figure 3.4: Significant improvement in MSE is observed across the data set in comparison against mcsGARCH, as seen here in a sample of selected symbols. Notches not shown to increase clarity.

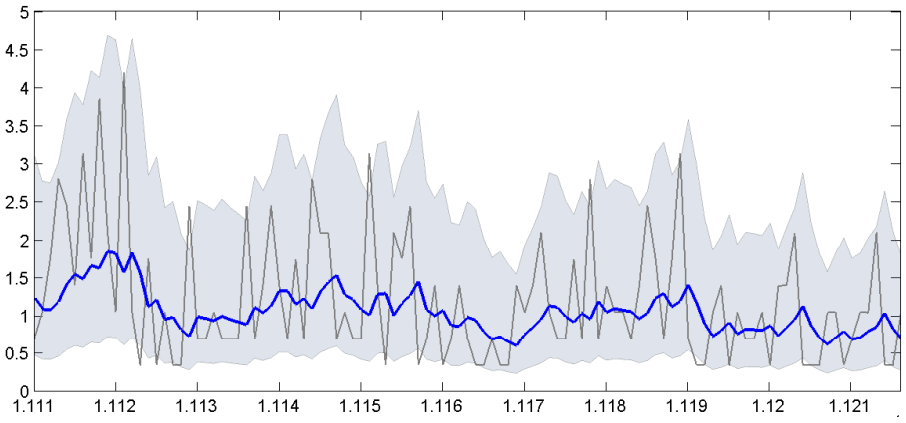


Figure 3.5: Assymetric variance bounds as observed in normal space, magnitude vs time, obtained from regressing in log-space.

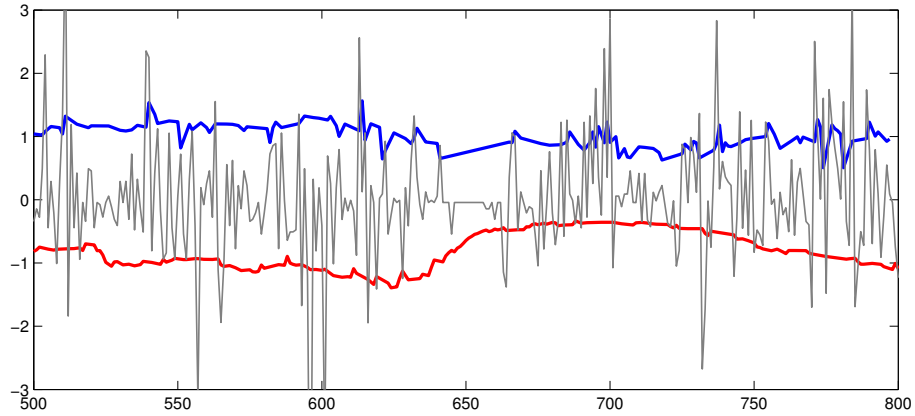


Figure 3.6: Forecasted mean for upper and lower envelopes, magnitude vs time, obtainded by regressing separately on the positive and negative returns.

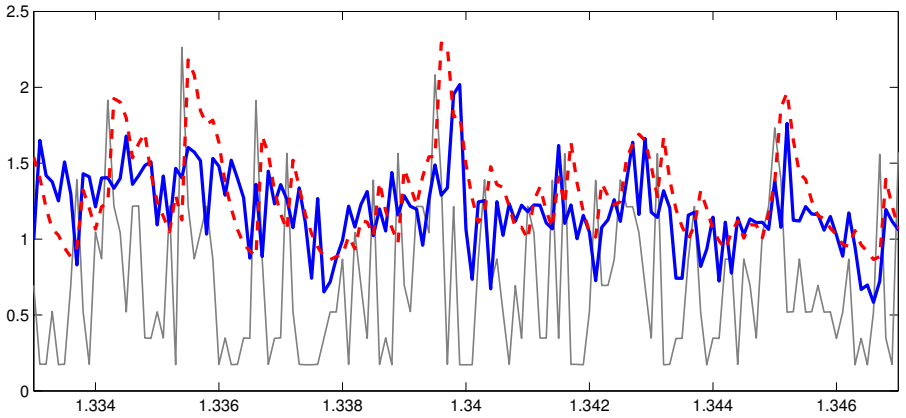
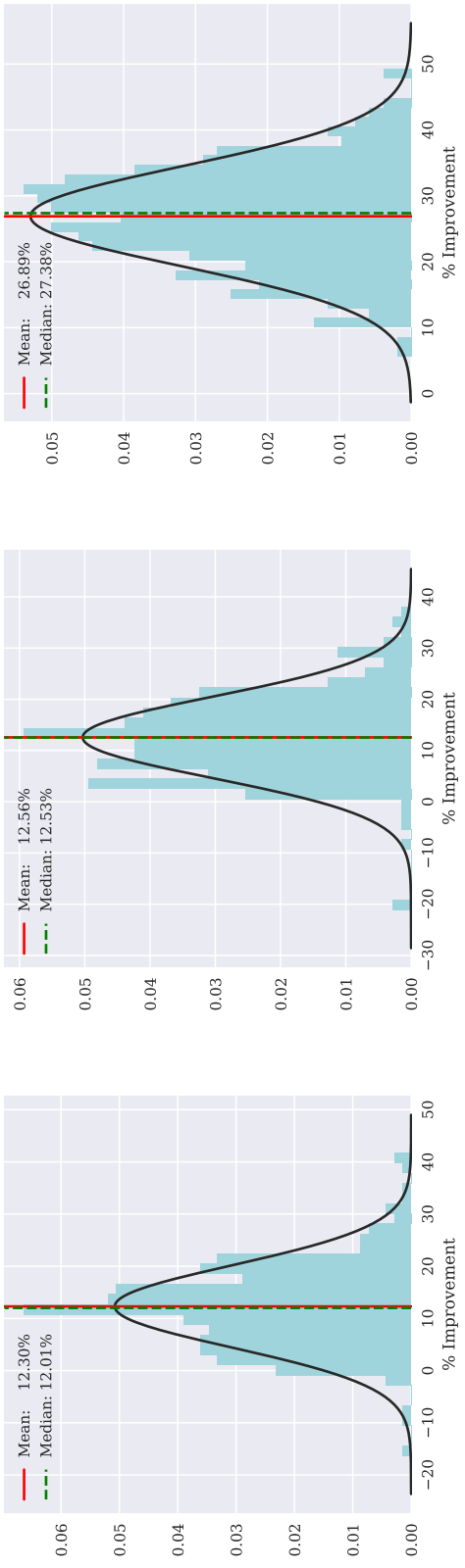
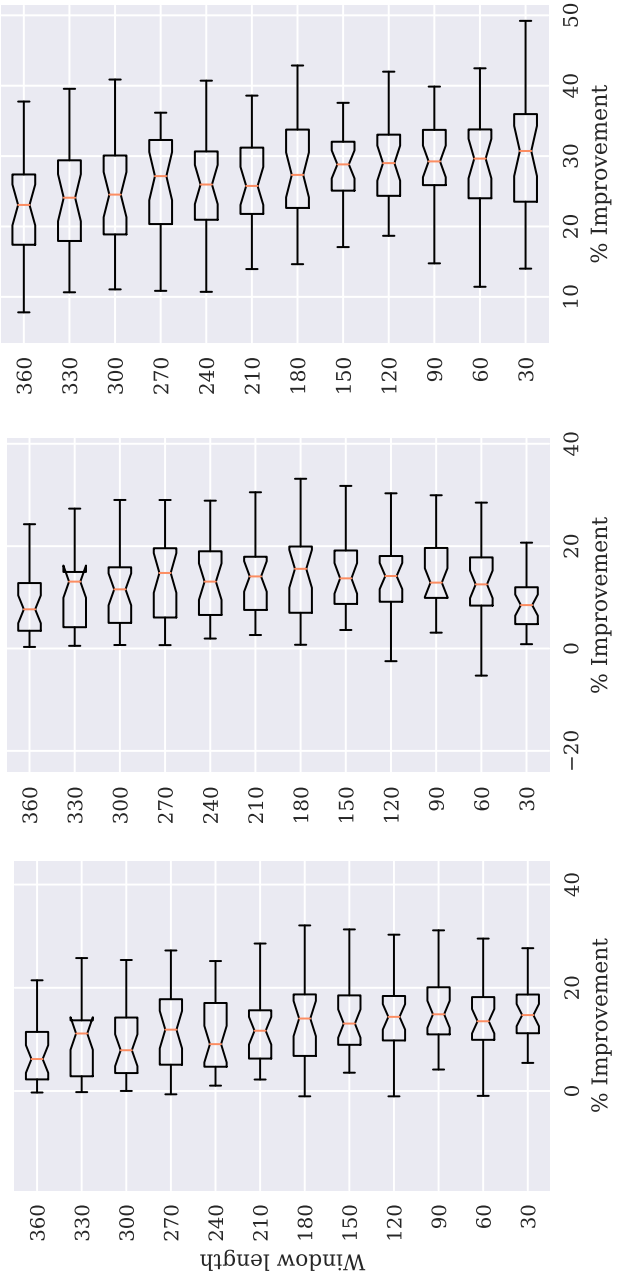


Figure 3.7: GP Volatility forecasts obtained by combining the regressions from the maxima envelope of the positive returns and the minima envelope of the negative returns, magnitude vs time, GARCH forecasts (red), and absolute returns (grey) are shown for comparison.



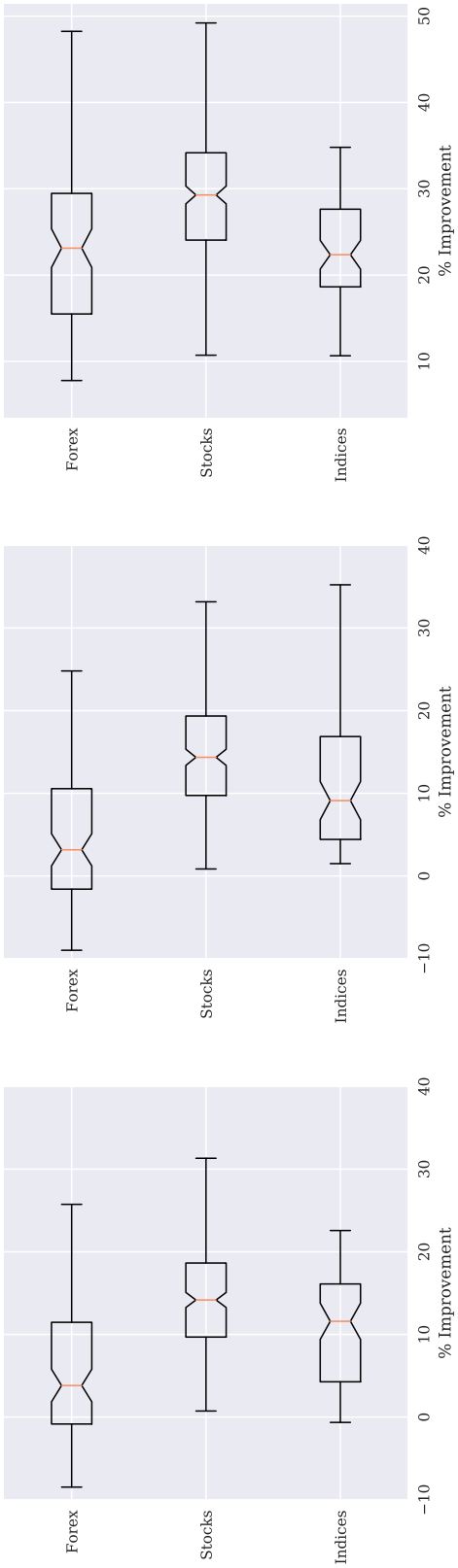
(a) mcsGARCH (b) Rolling average (c) Previous value

Figure 3.8: Improvement in MSE using envelope method, compared to benchmarks.



(a) mcsGARCH (b) Rolling average (c) Previous value

Figure 3.9: Improvement in MSE using envelope method, compared to benchmarks.



(a) mcsGARCH (b) Rolling average (c) Previous value

Figure 3.10: Results broken down by type of symbol for the improvement in MSE using envelope method, compared to benchmarks.

3.4 Assessing the Distribution of the Observed Returns Relative to the Volatility Forecasts

In order to understand the distributional dynamics of the GP forecasts a further analysis is carried out by using the notion that $r_t = \sigma_t \varepsilon$ where $\varepsilon \sim \mathcal{N}(0,1)$. From this it may be surmised that the returns time series normalized by the true volatility should follow a Gaussian distribution. This does not place any constraints on the distribution of the observed returns nor on the distribution of the true volatility - only that the residuals follow the normal distribution.

In order to test this empirically, the observed return at each time step is divided by the volatility forecast for that time step. The percentage of the observed returns that falls within the first, second, and third multiple of the forecast is calculated. For the true volatility sequence this would be close to 68%, 95% and 99.7% as per the normal distribution.

The results of this assessment are shown in Fig. 3.11. The first, second, and third multiples are empirically found to be at 67.1%, 91.7%, and 97.7%. While these values are not a complete match for what would be expected from completely normal residuals, the strong clustering around these values means that the forecasts from the GP are able to model the latent volatility sequence with high fidelity.

3.5 Computing Time Cost for Assessing GP Performance

Gaussian process based forecasts took 81% less time than the mcsGARCH based forecasts. This means the GP based scheme was able to be executed more than 5 times faster than the mcsGARCH.

The faster execution time combined with the significant performance gains in forecasting establishes that there are substantial quantitative benefits of using a GP based approach. The average time for computing the forecast at each time step was 1.73 seconds using the GPs on envelopes of the returns series for any given training window of up

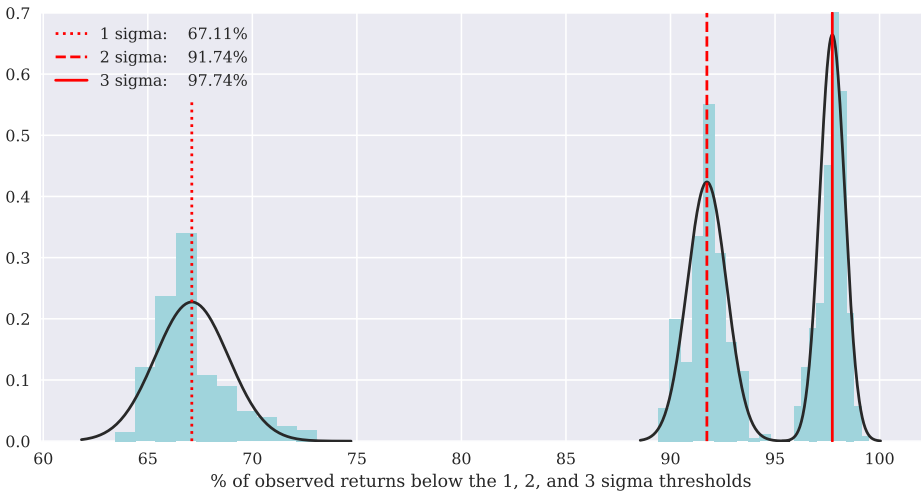


Figure 3.11: The distribution of absolute returns falling under 1, 2, and 3 multiples of the GP volatility forecasts

to 360 minutes. A whole trading day with 390 trading minutes can be forecast in approximately 11 minutes. The full set of GP experiments took approximately 11K hours (1.25 years) of clock time for the 50 asset symbols. All experiments were parallelised for fast execution. The average time for mcsGARCH forecasts at each time step was found to be 9.31 seconds.

These time are given here for completeness since it should be possible to optimize these execution time much further by using further parallelization, better hardware capability, and optimized code in programming languages such as C++.

3.6 Conclusion

This chapter showed that the envelope based approach of forecasting volatility with Gaussian processes performs significantly better than all three benchmarks - mcsGARCH, rolling average, and using the last observed value. The proposed approach performed 12% better on average than the mcsGARCH with the MSE loss function. The performance gains are highest on stocks (14%) followed by market indices (11%), with lower but still significant gains observed on forex data (7%).

The highly significant results in this section firmly establish that the log space enve-

lope based approach with GPs represents a strong model for forecasting financial volatility across a breadth of assets including stocks, market indices, and forex. This is the first time that such an extensive set of experiments has been carried out using Gaussian processes for financial volatility forecasting. This also represents the first time such an extensive comparison has been done of competing benchmarks on minute level financial data over a period as long as 11 years.

In the next chapter the concept of Rényi entropy is explored to assess how information flows between financial markets. This flow of information is later used to create an approach for further enhancing the volatility forecasts.

Chapter 4

Information Flows between Financial Time Series Using Rényi Entropy

Objective

The aim of this chapter is to contribute in various ways to the vast and growing literature on equity market linkages. First, a model-free non-parametric entropy measure Rényi transfer entropy is used to establish a framework for calculating information flows between two financial time series. This model-free methodology allows for great flexibility in measuring information flows. Second, extending the empirical results available on Rényi transfer entropy using a very large period of 1-minute frequency intraday data across a range of global indices. Third, performing an in-depth analysis of the key events that have impacted information flows globally and tracking the varying strength of information transmission broken down to a monthly analysis.

The results presented in this chapter on monthly information flows are the largest study done using Rényi Transfer Entropy for the global market indices in terms of time period covered and the granularity of data used.

4.1 Introduction

Equity market linkages have been explored in several recent research papers. Econophysics literature is increasingly exploring the dependencies of financial time series by means of transfer entropy. The analysis in this chapter uses the concept of Rényi transfer entropy, which is a flexible non-parametric method that accounts for linear as well as non-linear dependencies [59]. Using Rényi transfer entropy as an approach to measure financial time series linkages exploits the empirical distribution of the data using a completely model-free measure which is based on information theory. In order to apply this method, no specific time series model is required to estimate a dependence measure unlike other approaches such as correlation or Granger causality [19].

The concept of Rényi transfer entropy is similar to the more common Shannon transfer entropy [39]. Both of these are non-parametric measures based on the Kullback–Leibler distance between probability distributions [40]. Rényi transfer entropy has an additional parameter that allows focusing on specific parts of a distribution, such as centre or tail observations. When dealing with financial return data this is an especially interesting feature that allows exploration of the effect of tail events which are assumed to be more informative than observations located in the centre of the distribution.

As far back as 2005, before the financial crisis took its full effect, Bonfiglioli & Favero (2005) [10] examined the long-term and short-term interdependencies between the US and the German stock market and found evidence for contagion in the short-term dynamics, but no long-term interdependencies. The role of the US market as a leader among the world's major stock markets was also studied by Bessler & Yang (2003) [8]. Marschinski & Kantz (2002) [44] and Kwon & Yang (2008) [41] applied the Shannon transfer entropy to quantify the information flows between stock market indices. Jizba et al. (2012) [40] quantified the information flows in similar data using the Rényi transfer entropy. Dimpfl & Peter (2013) [19] use the Shannon transfer entropy to analyze information transfer between credit derivative and bond markets. Dimpfl & Peter (2014) [20] use Rényi transfer entropy and effective Rényi transfer entropy to explore the transatlantic information flows

between stock market indices before, during, and after the 2008 financial crisis. Chunxia et al. (2016) [15] studied the contagion among American industries based on transfer entropy. He & Shang (2017) [35] looked at a comparison between different approaches to calculating the transfer entropy with application to the Chinese stock market. Teng & Shang (2017) [63] developed an approach to merge the information flows between two systems happening at multiple time scales. Dimpfl & Peter (2018) [21] used transfer entropy to analyse the transmission of volatility between the oil, stock, gold, and currency markets. Behrendt & Zimmermann (2018) [5] used Shannon transfer entropy to infer collective investor behaviour from large-scale Wikipedia search data for individual stocks. An excellent overview of work in this field is presented in Bossomaier et al. (2016) [11].

The research presented in this chapter finds significant bi-directional information flows at a one-minute frequency between major European stock market indices and the S&P 500. This flow spikes heavily around events of financial significance. Politically important events are found to be less important in their impact on financial information flows. It is also established that tail events contribute substantially larger portion to the flow of information compared to the central events. This supports the notion that tail observations of the empirical return distribution are more informative than observations in the centre

The rest of this chapter is arranged as follows: Shannon and Rényi transfer entropies are introduced followed by an introduction to the market indices data being used. Next the experimental methodology is presented followed by a discussion of results at the end.

4.2 Measuring Information Flows Using Rényi Transfer Entropy

First, Shannon entropy is introduced in order to illustrate the concept of Rényi entropy transfer. Shannon entropy measures the uncertainty associated with draws from a given probability distribution. In 1928, Hartley [34] defined that when we observe one specific state x of probability $p(x)$, the amount of information gained is $-\log_2 p(j)$. In 1948,

Shannon [60] introduced its application to quantify the information content of a message in the context of information theory. Consider a discrete random variable X with probability distribution $p(x)$ where x denotes the possible states or outcomes of X.

The measurement units are bits when the base 2 logarithm is used. Based on this definition, the Shannon entropy of $p(j)$ is given by

$$H_X = -\sum_x p(x) \log_2 p(x)$$

This measure of uncertainty is maximized for an equally distributed variable. As the probabilities start to move away from the uniform distribution the Shannon entropy starts to get lower and our underlying uncertainty reduces.

Shannon entropy by itself for a given time series is a univariate measure and the concept of mutual information can be used to extend it to the bivariate case. Mutual information is a symmetric measure for the difference between two probability distributions - the Kullback-Leibler distance. Any form of statistical dependency is accounted for by the mutual information. If we assume two discrete random processes X and Y with marginal probability distribution $p(x)$ and $p(y)$ and join probability distribution $p_{XY}(x,y)$, then the Shannon mutual information between them is given by

$$M_{XY} = \sum_{x,y} p_{XY}(x,y) \log_2 \frac{p_{XY}(x,y)}{p(x)p(y)}$$

where all possible values of x and y contribute to the summation. The Shannon mutual information measures the reduction in uncertainty from observing both processes compared to the case of independence of X and Y. If both process were independent then the marginal distribution $p_{XY}(x,y) = p(x)p(y)$ then the mutual information would be zero. Mutual information quantifies the benefit gained from knowing one process when we are forecasting the outcomes of another process. Shannon mutual entropy is a symmetric measure and can not differentiate if one of the underlying process contributes more information to our understanding than the other.

In the context of financial markets, measures that can capture the asymmetric and dy-

namic nature of information flow in financial time series are needed. In 2000, Schreiber [59] considered transition probabilities to introduce a dynamic structure to mutual information. Consider a stationary Markov process X of order k. The probability to observe X in state x at time $t+1$ conditional on the previous k observations is $p(x_{t+1} \mid x_t, \cdots, x_{t-k+1}) = p(x_{t+1} \mid x_t, \cdots, x_{t-k})$. For measuring the information flow from process Y to X, Schireiber [59] proposes quantifying the deviation from the generalized Markov property $p(x_{t+1} \mid x_t^{(k)}) = p(x_{t+1} \mid x_t^{(k)}, y_t^{(l)})$ based on the Kullback-Leibler distance and derives the Shannon transfer entropy as

$$\mathrm{TE}_{Y\to X}(k,l) = \sum_{x,y} p(x_{t+1}, x_t^{(k)}, y_t^{(l)}) \log \frac{p(x_{t+1} \mid x_t^{(k)}, y_t^{(l)})}{p(x_{t+1} \mid x_t^{(k)})} \tag{4.1}$$

Transfer entropy is an asymmetric measure because the fraction being logged in Eq. 4.1 will be different depending on whether we are measuring the information from from Y to X or X to Y. The Shannon transfer entropy quantifies the additional information that we gain about the future value of X by by observing the past values of Y, given the history of X is already known. The information flow from X to Y can be evaluated by calculating $TE_{X\to Y}(l,k)$

In 2012, Jizba et al. [40] introduced the concept of evaluating transfer entropy by using the Rényi entropy rather than Shannon entropy. Rényi introduced the concept of Rényi entropy in 1970 [39]. It depends on a weighting parameter q which can be used to empahsize different parts of the probability distribution of X

$$\mathrm{H}_X^q = \frac{1}{1-q} \log \sum_x p_q(x)$$

where $q > 0$. As $q \to 1$, the Rényi entropy converges to the Shannon entropy. Rényi entropy enables a more differentiated analysis of the underlying distributions by varying the values of q. The parameter q is an emphasis parameter, and helps in emphasising different parts of the underlying probability distribution. The tail events can be highlighted for $0 < q < 1$, while the more probable central events are given more weight when $q > 1$. Fig. 4.1 shows how the original probabilities are transformed for different values of q.

This is very interesting from the perspective of financial time series where fat tails play a key role in our analysis.

Jizba et al [40] derived the Rényi transfer entropy to measure the information flow from Y to X as

$$\mathrm{RTE}^{q}_{Y\to X}(k,l) = \frac{1}{1-q}\log\frac{\sum_x \phi_q(x_t^{(k)})p^q(x_{t+1}\mid x_t^{(k)})}{\sum_{x,y}\phi_q(x_t^{(k)},y_t^{(l)})p^q(x_{t+1}\mid x_t^{(k)},y_t^{(l)})} \tag{4.2}$$

with $q>0$ and the concept of escort distribution [1] $\phi_q(x) = p^q(x)/\sum_x p^q(x)$ is used to normalize the weighted distribution. As before, $\mathrm{RT}_{X\to Y}(l,k)$ to measure the information flow in the inverse direction can be defined analogously.

There are two important things to note about the Rényi transfer entropy. First that a value of $\mathrm{RTE}^{q}_{Y\to X}(k,l)=0$ does not mean that processes are independent since the q parameter is being used and the zero value might change with change in q. Second, Rényi transfer entropy can also be negative in value. In the case of Shannon based measures, information about past events of Y can not increase the uncertainty of future values of X, and the worst case might simply be that it adds no further information leaving the measure unchanged. The Rényi transfer entropy of a time series with itself is always zero.

In the case of Rényi based measures, negative estimates occur when observing past values of Y might increase the probability of a future event of X. This means that observing X implies greater exposure to the risk of X than would have been expected by observing the values of X by itself. Since extreme tail events are assumed to be more informative than median observations [58], so Rényi transfer entropy is a very nuanced tool for analysing information flows between financial time series.

It is important to note that X and Y are discrete random variables. The continuous values of financial time series need to be converted into S discrete symbols. A compromise needs to be made between the right value for S, k, and l. Marschinski et al [44] outlined some of the potential pitfalls of using too small or too large values for each of these parameters. The amount of data required grows like S^{k+l}.

[1] Details on the escort distribution can be explored in Beck and Schögl (1993, chap. 9).

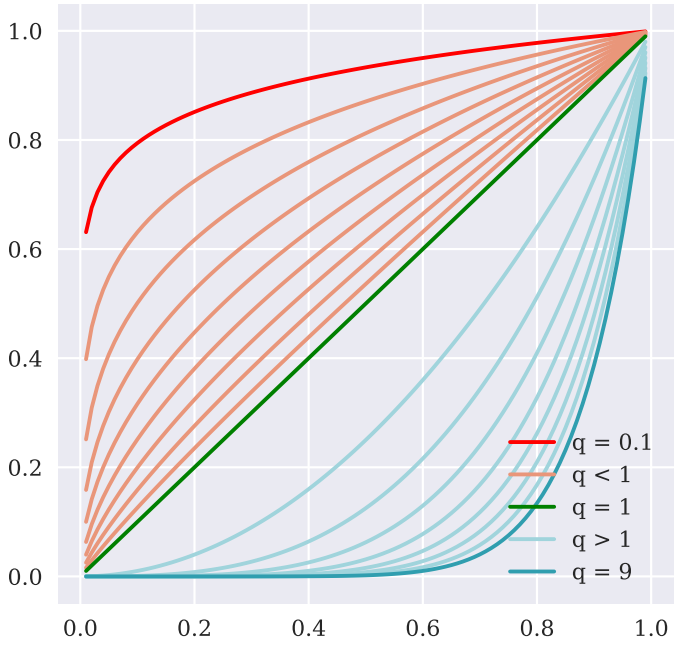


Figure 4.1: Values from the original distribution are transformed by the q parameter. Values with lower probability are assigned increasing weight as q goes to zero while values of $q > 1$ enhance the probability contribution of more commonly occurring events.

In order to minimize mis-estimation of the transfer entropy due to finite sample effects Marschinski et al [44] introduced the concept of effective transfer entropy, which Jiba et al [40] extended to effective Rényi transfer entropy. The effective Rényi transfer entropy from Y to X is given by

$$\mathrm{ERTE}^{q}_{Y\to X}(k,l) = \mathrm{RTE}^{q}_{Y\to X}(k,l) - \mathrm{RTE}^{q}_{Y_{shuffled}\to X}(k,l) \tag{4.3}$$

where $Y_{shuffled}$ is the shuffled series obtained by randomly rearranging the observed values of Y in order to destroy any structure in the underlying data. Any non-zero value of $\mathrm{RTE}^{q}_{Y_{shuffled}\to X}(k,l)$ is considered an artefact of finite sample size and therefore removed from the calculation of transfer entropy.

Rényi transfer entropy and effective Rényi transfer entropy is used to explore how information is flowing between global financial markets by looking specifically at the time series of stock market indices.

4.3 Stock Market Data

In order to analyse the effectiveness of this approach at a granular level, the data set is comprised of 6 indices - the primary series is S&P500, and the 5 large global indices whose opening hours overlap with S&P500 are the secondary series i.e. FTSE (UK), CAC40 (France), DAX (Germany), SSMI (Switzerland), ATX (Austria). The indices used are described in table 4.1 The time range of the data is from 1 January, 2006 to 31 March 2017. All data are obtained from Thomson Reuters Tick History (TRTH) API.

This builds on the seminal research done by Jizba et al in their 2012 work on effective Rényi transfer entropy. Jizba et al [40] looked at stock exchange indices for the period between January 1998 - 31 December 2009. The data used in this chapter covers the period from 1 January, 2006 to 31 March 2017.

Only valid values are used in the analysis where both indices under consideration are trading actively. Periods without trading activity (weekends, night time, holidays) in one or both stock exchanges are excluded. The log returns from the index values are used and the opening and closing auctions are discarded. This means that generally there is an overlap of 120-180 index observations each day between the between the primary series and any of the secondary series. All time series under consideration were synced to UTC to simplify time stamp matching between the indices.

The summary statistics of the return series used for the minute level analysis are presented in Table 4.2. These statistics are for the restricted duration of the overlapping trading hours. The data shows a more pronounced value of skew and kurtosis than the full returns series because the data set is restricted to the first two overlapping trading hours of NYSE.

Indices	Description	Country
S&P500	Standard and Poor 500 index represents the 500 stocks actively traded in the US based on market capitalization	USA
DAX (GDAXI)	Dax Indices stock of 30 major German companies	Germany
CAC40	Cotation Assistée en Continu is a benchmark French stock market index that represents a capitalization-weighted measure of the 40 most significant values among the 100 highest market caps on the Euronext Paris	France
FTSE	Financial Times Stock Exchange 100 Index is a share index of the 100 companies with the highest market capitalisation listed on the London Stock Exchange	United Kingdom
ATX	The Austrian Traded Index of the Wiener Börse consists of 20 stocks	Austria
SSMI	The Swiss Market Index is a capitalization-weighted index of the 20 largest and most liquid stocks representing about 85% of the free-float market capitalization of the Swiss equity market	Swiss

Table 4.1: Data from 6 global stock market indices was used to explore Effective Rényi Transfer Entropy (ERTE). Time period was between 2006-2017 for all the indices.

Index	Count	Mean	Std	Min	25%	75%	Max	Skew	Kurtuosis
S&P500	-	0.0000	0.0005	-0.0139	-0.0002	0.0002	0.0111	-0.1639	27.9864
GDAXI (DAX)	336,592	0.0000	0.0005	-0.0097	-0.0002	0.0002	0.0121	-0.1690	14.1836
CAC40	118,874	0.0000	0.0007	-0.0115	-0.0003	0.0003	0.0113	-0.0608	9.7442
FTSE	118,874	0.0000	0.0006	-0.0151	-0.0002	0.0002	0.0121	0.6393	39.4606
ATX	118,874	0.0000	0.0005	-0.0109	-0.0002	0.0002	0.0081	-0.1470	13.5321
SSMI	118,874	0.0000	0.0006	-0.0132	-0.0002	0.0002	0.0188	0.1174	20.5367

Table 4.2: Data from 12 global stock market indices was used to explore Effective Rényi Transfer Entropy (ERTE). Time period was between 2006-2017 for all the indices. Count shows the overlapping trading minutes with S&P500.

4.4 Methodology for Experiments

For calculating the ERTE and RTE two financial time series are selected. The main series of interest is referred to as the primary series R_p and the series providing flow of information is termed the secondary series R_s

In order to calculate the relative frequencies for the PDF needed in Eq. 4.2 the returns of both series are divided into discrete bins based on their amplitude and each bin is assigned a symbol.

Three bins are used to divide the return data along the 5% and the 95% quantiles of the full data set, denoted as $q^r_{[0.05]}$ and $q^r_{[0.95]}$ respectively. The symbolic encoding S_t replaces each value of the return time series by the respective symbol for upper quantile (U), middle (M), or lower quantile (L)

$$S_t = \begin{cases} L, & \text{if } r_t \leq q_{[0.05]^r} \\ U, & \text{if } r_t \geq q_{[0.95]^r} \\ M, & \text{otherwise} \end{cases}$$

General consensus in the literature exists [20, 19, 40] that this choice of quantiles is reasonable for identifying extreme tail events. Dimpfl & Peter (2014) [20] found that by moving further into the tails, for example taking the 1% and 99% quantiles or moving inwards towards the centre dilutes the amount of information flow detectable.

Given look back periods k and l, k copies of the primary series each shifted by one step is joined with l copies of the secondary series each of which is also one step shifted. This gives us a sequence where each member is a of the form

$$i = (r_{p,t}, r_{p,t-1}, \cdots, r_{p,t-k+1}, r_{s,t}, \cdots, r_{s,t-l-1})$$

where r_p and r_s are observations from the primary and secondary series. The occurrence of each combination of discrete symbols is counted and normalized to arrive at the relative frequency $P(i)$. In order to avoid spurious data effects, the shifting of the series is

reinitialized for each trading day so that returns that are temporally distant do not end up close to each other. All components of 4.2 are calculated similarly.

The look back period and the discretization possible depends on the number of data points available, with a larger the number of data points allowing a higher value of each of these parameters. Smaller data samples will limit the ability to sufficiently observe all possible symbol sequences.

A month long moving window of trading days is used for parsing through the data. From each window only the overlapping time stamps are taken for contributing towards the transfer entropy calculation. The RTE and $ERTE$ are calculated for a range of values of q to analyse how different parts of the underlying distributions contribute to the information flow between the two indices under consideration.

The transfer entropy series is overlayed with key events that impacted financial markets during the time period under consideration. This provides context for what real world events are being picked up in the analysis.

4.5 Calculating Monthly Effective Rényi Transfer Entropy

The results in this section are the largest study done using Rényi Transfer Entropy for the global market indices in terms of time period covered and the granularity of data used. The insights for the period from 2011 - 2017 are novel and show that significant information flows occur around events of global financial importance. Another key finding in this section is that key political events seem to have a less potent effect on market information flows than key financial events.

For calculating the monthly level information flows, the daily minute level data from the time series is used. S&P500 and 5 European indices (DAX, FTSE, CAC40, ATX, SSMI) which overlap in time are used. Data from the whole subset is partitioned into monthly buckets, and these buckets are used to calculate the pairwise effective transfer entropy to understand the evolution of information flows. Only the hours in which two indices overlapped are used. A sample of overlapping returns is shown in Fig. 4.2

To illustrate, consider the example of DAX and S&P500. There is an overlap of 2 hours on average on any given trading day. This gives 2700 return data points in an average month of 22 trading days. Since this is a much smaller data sample effective renyi transfer entropy as advocated by Jizba et al (2012) [40], rather than the transfer entropy. For calculating the ERTE a symbol encoding of 6 equal quantiles is used, and $l = k = 1$ as in Marschinski & Kartz (2002) [44], and Dimpfl & Peter (2014) [20]. The values of q in $0 < q < 1$ and $1 < q < 2$ are used as before.

The pairwise flow of information between S&P500 and the European indices is shown in Figs 4.3 - 4.11.

There is a pronounced flow of information from S&P500 to all other market indices. This effect is magnified as the central events are given more focus by increasing the value of q. It should also be noted that the amount of RTE in bits is higher for lower values of q compared to higher values. This further strengthens the belief that rare events contribute a higher component of information across markets.

The DAX seems to be a significant information partner for the S&P500 as it can be seen from the Fig. 4.3 that the flow of information between DAX and S&P500 is quite balanced at all levels of q, except in the two year period between 2011 and 2013 where significantly higher level of information flows from DAX to S&P500 than the other way around.

Some key events like the the Lehman Brothers bankruptcy, 2010 Flash Crash, and a change in the credit rating of USA all show spikes of information that persist over many months. Slightly counter intuitively it is found that major political events such as the 2016 Presidential election and the Brexit referendum are found to cause no significant change in the flow of information

4.6 Summary of Highlighted Financial Events

The key global financial events highlighted in the figures are summarised below. Some key political events are also included to ground the time line. The **Lehman Brothers**



Figure 4.2: Positive (blue) and negatve (red) returns separated into two envolopes for DAX on 3rd Jan 2006 from 08:00 - 16:30 hrs UTC.

bankruptcy on September 15, 2008, remains the largest bankruptcy filing in U.S. history, with Lehman holding over US$600 trillion in assets. This bankruptcy triggered a one-day drop of 4.5% in the Dow Jones Industrial Average, the largest decline recorded to that point since the September 11, 2001 attacks. This is a key event in the precipitation of the Great recession globally.

The **May 6, 2010, Flash Crash** was a United States trillion-dollar stock market crash, which lasted for approximately 36 minutes. Stock indexes, such as the S&P 500, Dow Jones Industrial Average and Nasdaq Composite, collapsed and rebounded. The Dow Jones Industrial Average had its second biggest intraday point drop plunging about 9% within minutes, only to recover a large part of the loss. Due to the volatile prices of stocks, stock index futures, options and exchange-traded fund (ETFs) trading volume spiked. A CFTC 2014 report described it as one of the most turbulent periods in the history of the financial market.

The **U.S. credit rating was lowered** by several credit rating agencies around the world, including Standard & Poor's (S&P) which reduced the country's rating from AAA (outstanding) to AA+ (excellent) on August 5, 2011

The **Swiss Franc was unpegged from Euro** on January 15th, when the Swiss Na-

tional Bank (SNB) unexpectedly announced that the Swiss Franc won't be held at a fixed exchange rate with the Euro any longer. The Franc rose rapidly and the Euro fell in value. A number of hedge funds across the world made big losses and the Swiss stock market collapsed. The exchange rate peg was introduced in 2011 while the financial markets around the world were in turmoil.

The **federal funds interest rate was increased** on December 16, 2015. This was a widely expected move, with the rate banks charge each other for overnight loans being changed from the 0-.25% range to the .25-.50% range. U.S. stock markets rallied on the news. In the lead up to this change oil price had reached its lowest since December 2008 along with other indicators in broader markets that had raised investor concern. This increase was announced as the Federal Open Market Committee believed that economic conditions had sufficiently improved since the 2008 financial crash.

The **Brexit referendum** took place on June 23, 2016 in the United Kingdom (UK) and Gibraltar to gauge support for the country either remaining a member of, or leaving, the European Union (EU) under the given legal provisions. The referendum resulted in a simple majority of 51.9% being in favour of leaving the EU

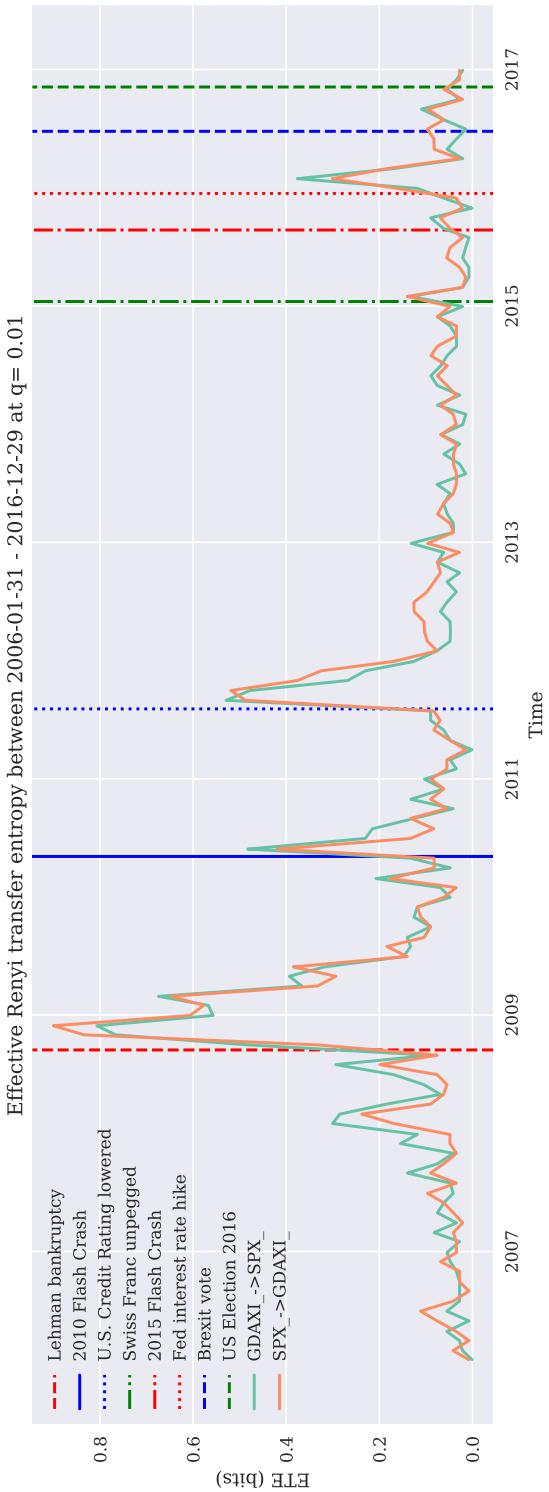
The **US Presidential election 2016** was the 58th American presidential election, held on Tuesday, November 8, 2016. In a surprise victory, the Republican candidate Donald Trump defeated former Secretary of State Hillary Clinton. Trump took office as the 45th President on January 20, 2017.

4.7 Conclusion

In this chapter it was established that significant information flows exist in the global markets as measured using Rényi Transfer Entropy between the returns time series for 6 global market indices. The evolution of information flows with time was analysed by calculating the monthly RTE between S&P500 and 5 European markets during their overlapping active trading hours. The results presented on monthly information flows are the largest study done to date using Rényi Transfer Entropy for the global market indices

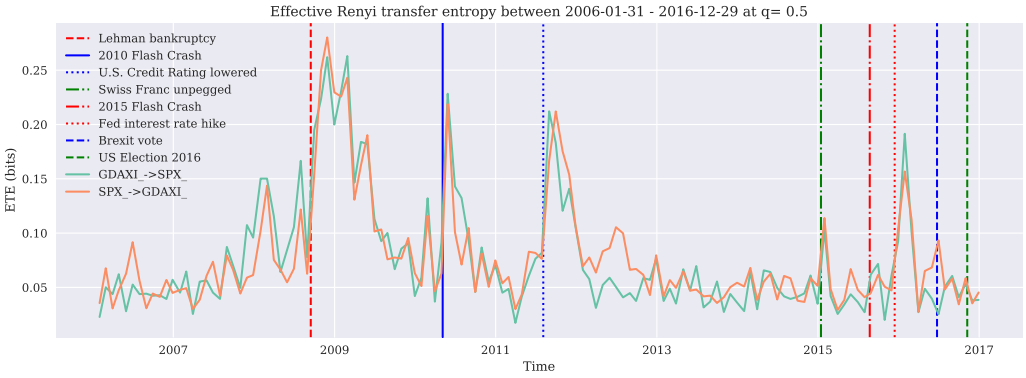
in terms of time period covered and the granularity of data used.

The novel insights for the period from 2011 - 2017 demonstrate that significant information flows occur around events of global financial importance. It is also found that key political events seem to have a less potent effect on market information flows than key financial events.

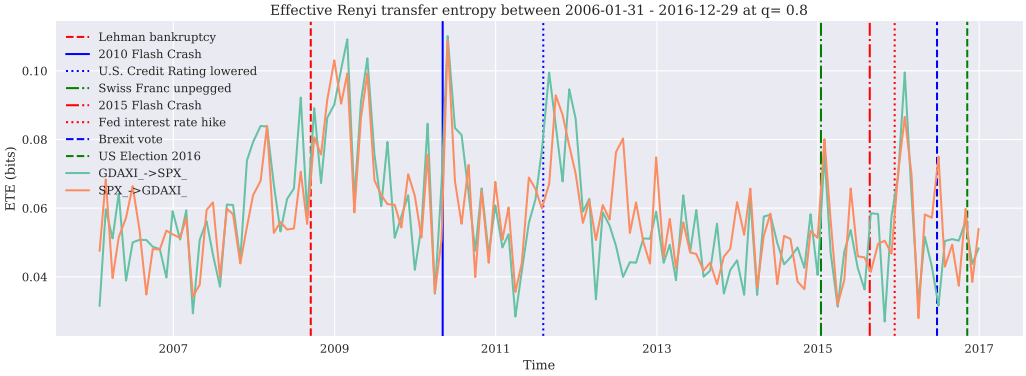


(a) $q = 0.01$

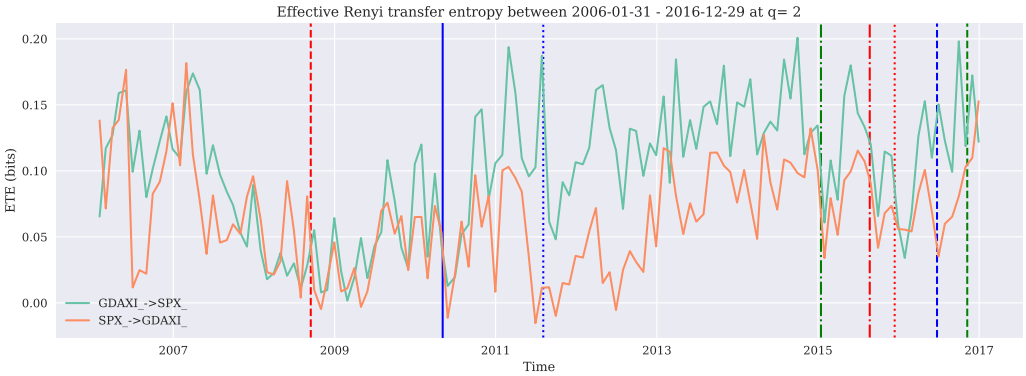
Figure 4.3: Effective Rényi transfer entropy between DAX and S&P500 for various values of q.



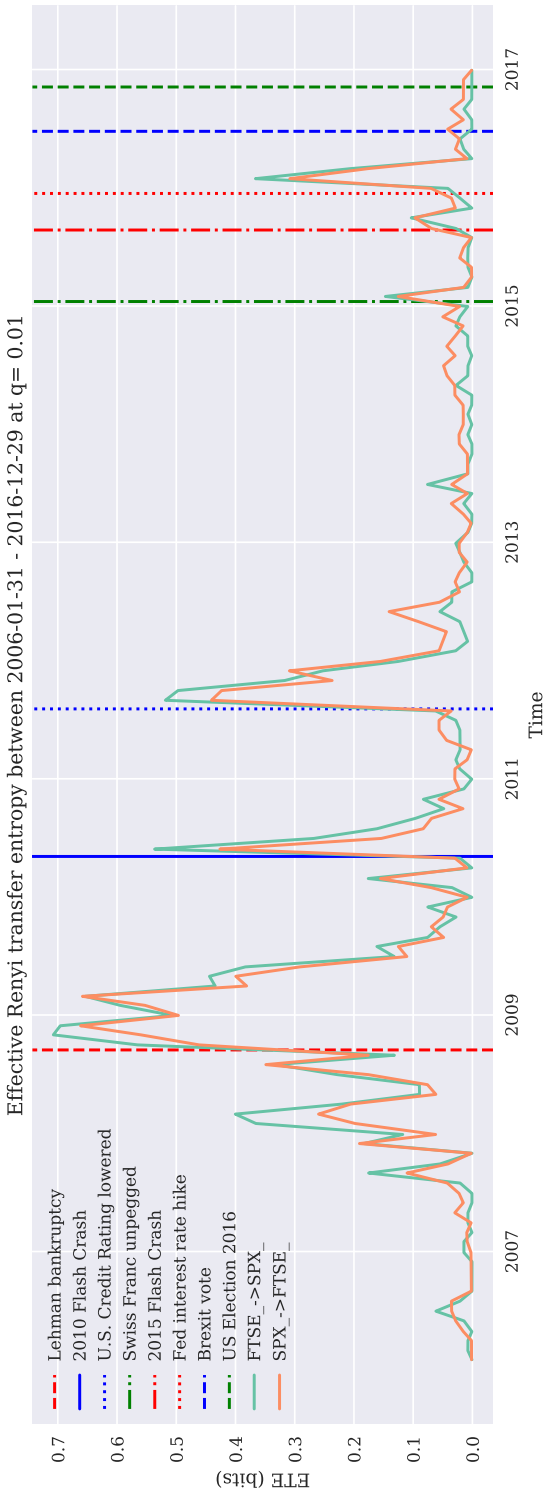
(c) $q = 0.5$



(d) $q = 0.8$

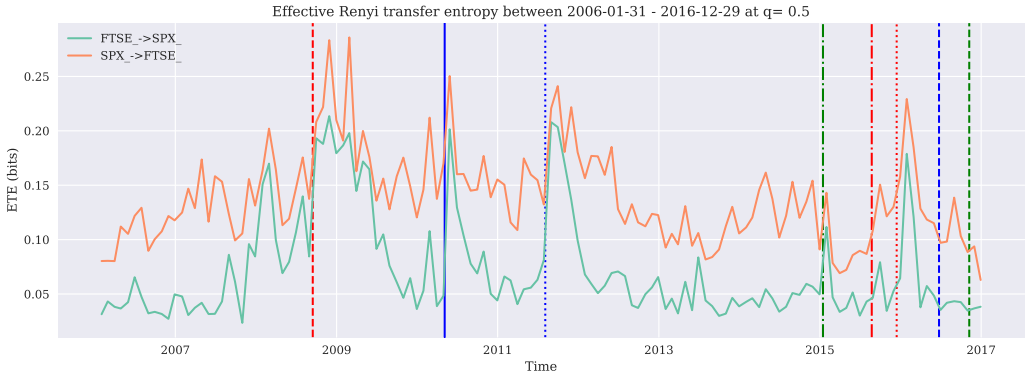


(e) $q = 2.0$

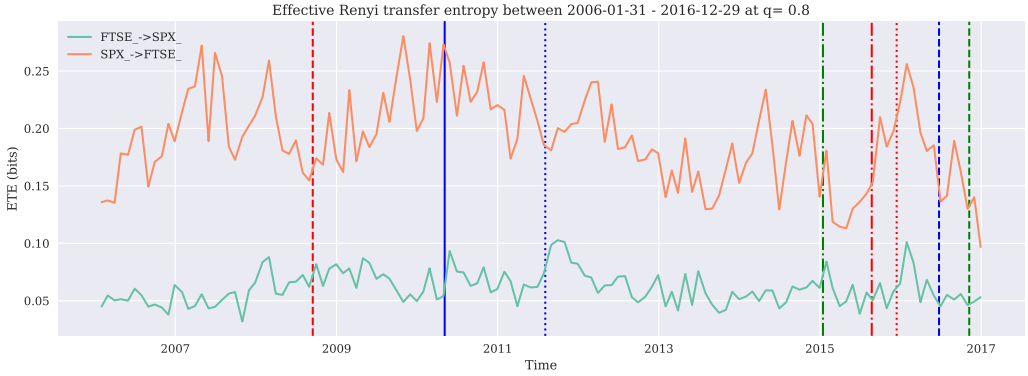


(a) $q = 0.01$

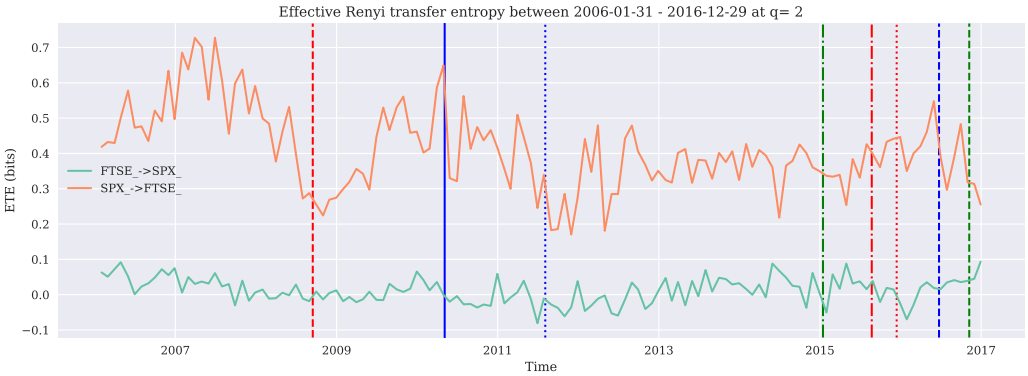
Figure 4.5: Effective Rényi transfer entropy between FTSE and S&P500 for various values of q.



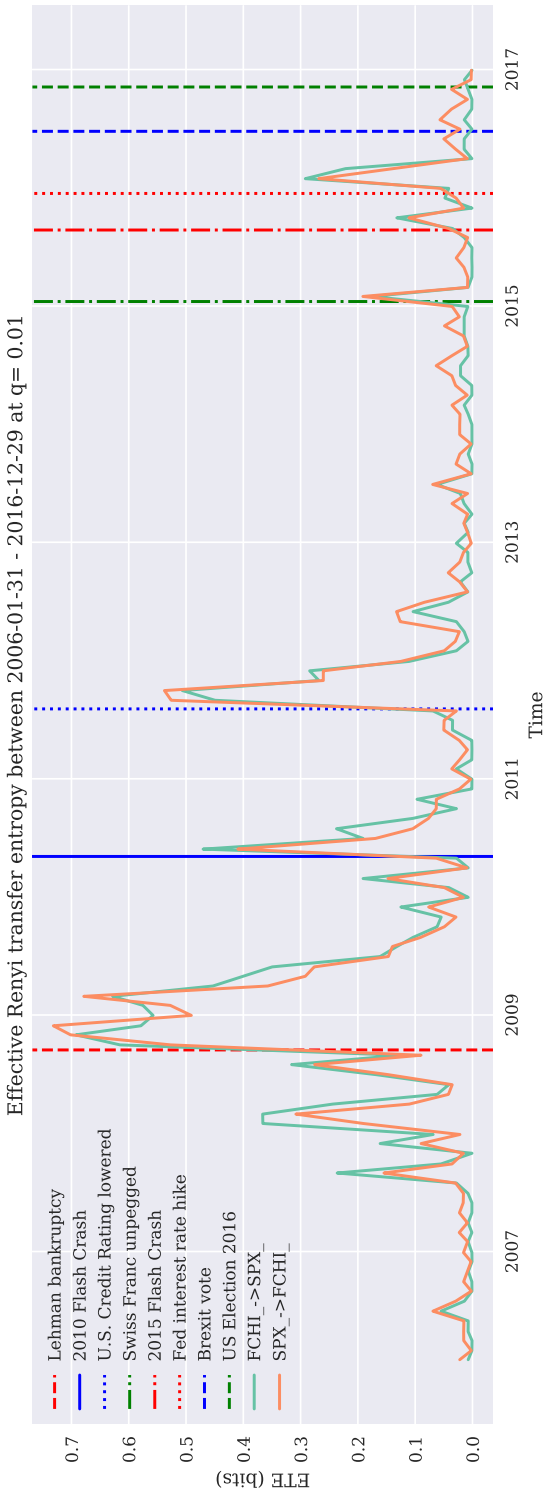
(c) $q = 0.5$



(d) $q = 0.8$

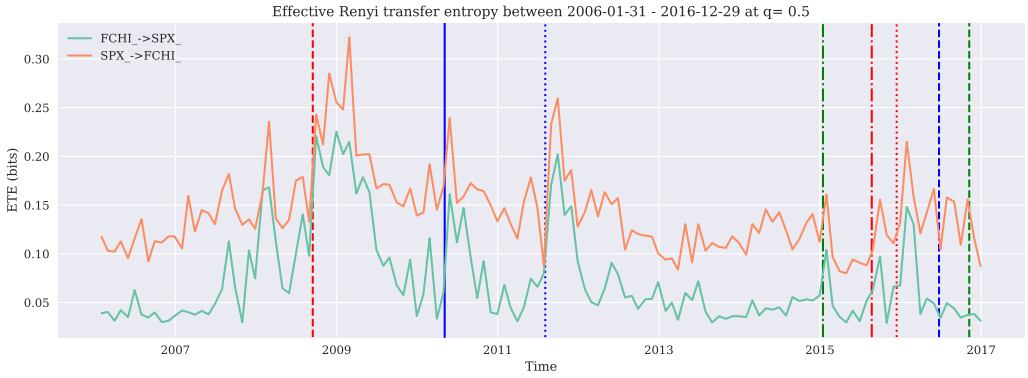


(e) $q = 2.0$

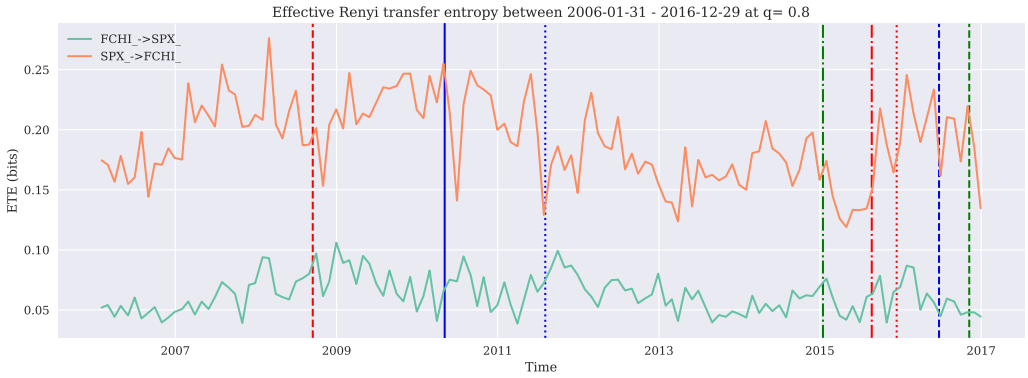


(a) $q = 0.01$

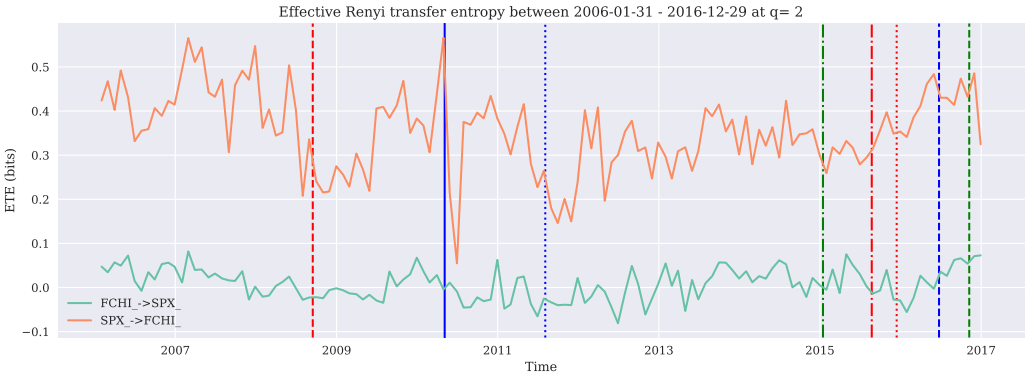
Figure 4.7: Effective Rényi transfer entropy between CAC40 (FCHI) and S&P500 for various values of q.



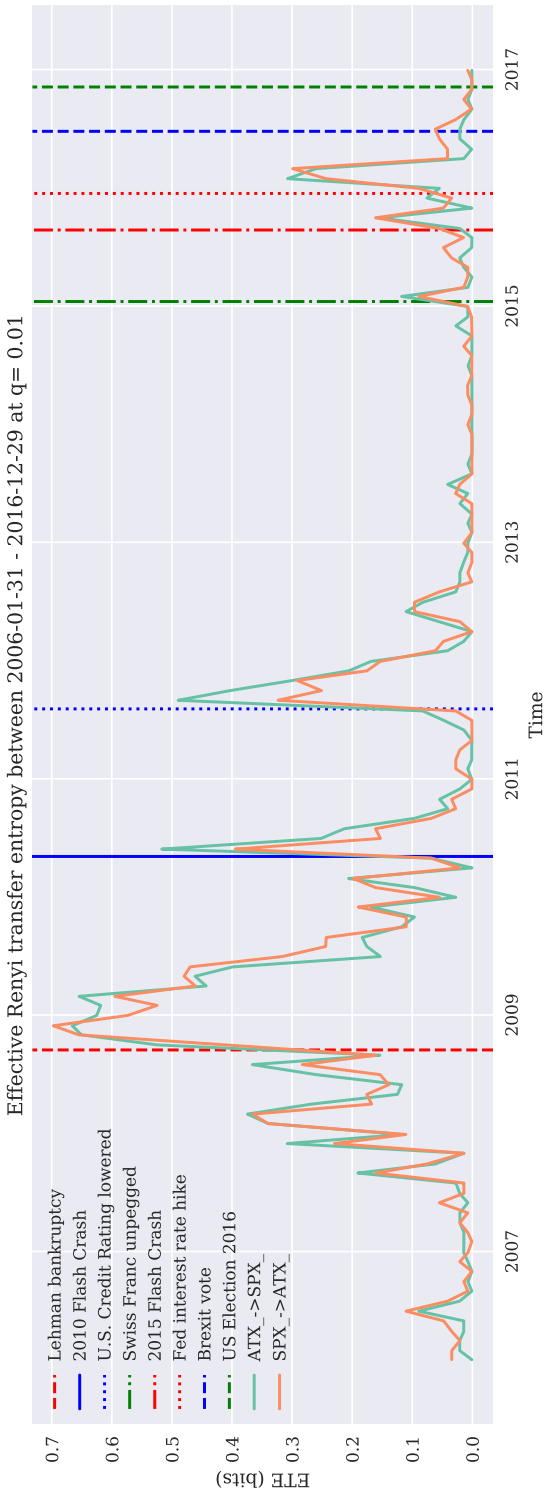
(c) $q = 0.5$



(d) $q = 0.8$

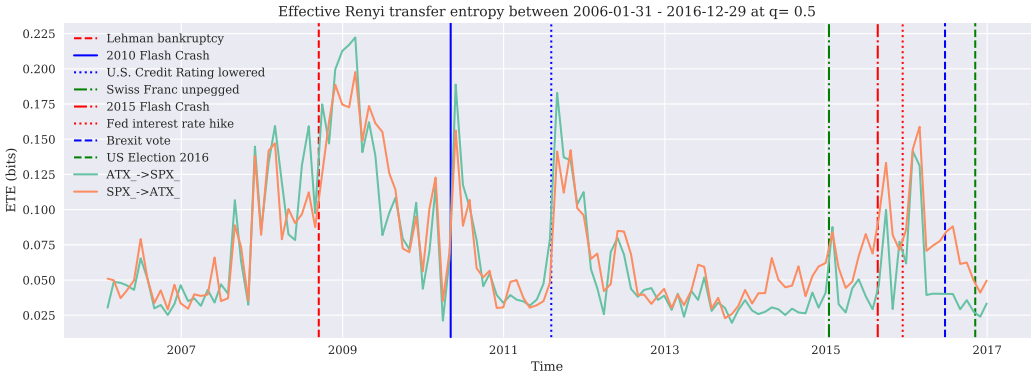


(e) $q = 2.0$

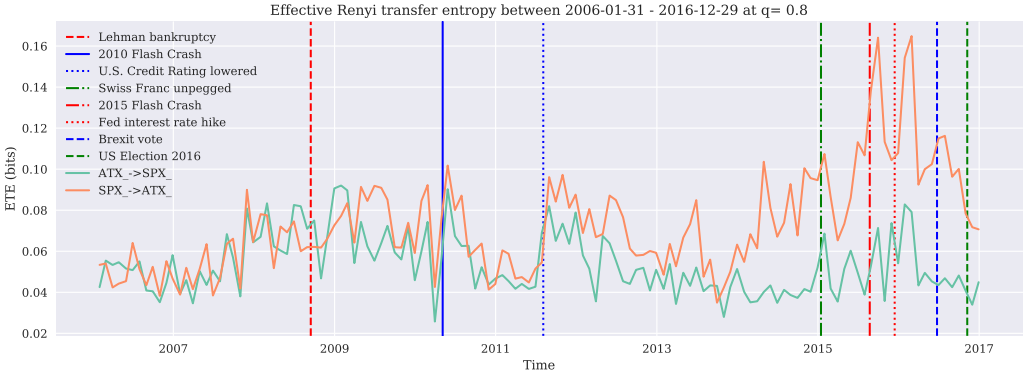


(a) $q = 0.01$

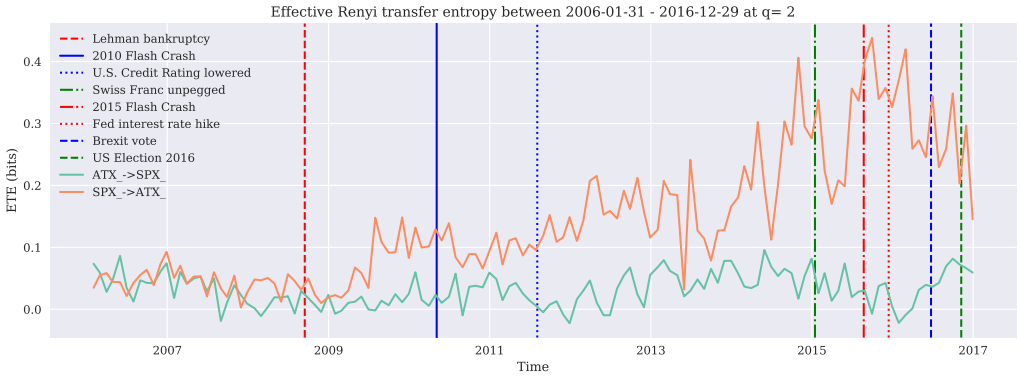
Figure 4.9: Effective Rényi transfer entropy between ATX and S&P500 for various values of q.



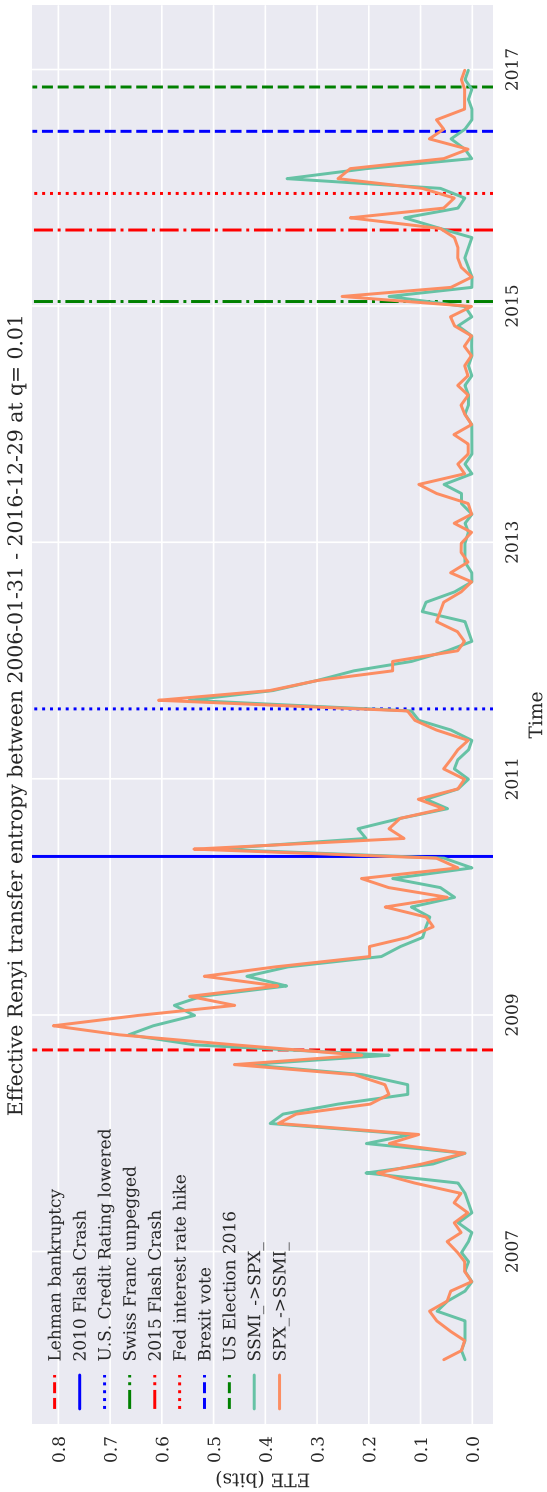
(c) $q = 0.5$



(d) $q = 0.8$

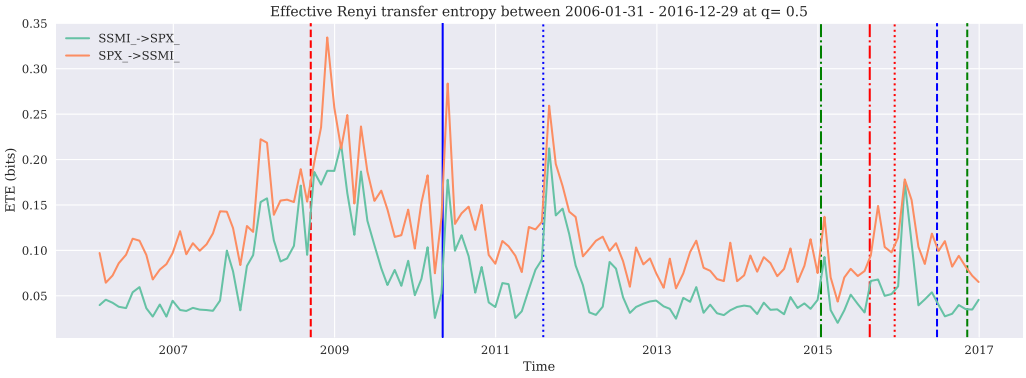


(e) $q = 2.0$

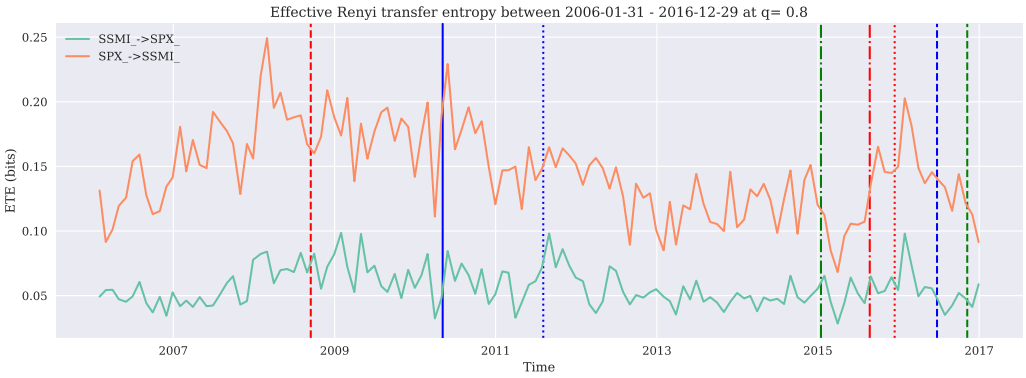


(a) $q = 0.01$

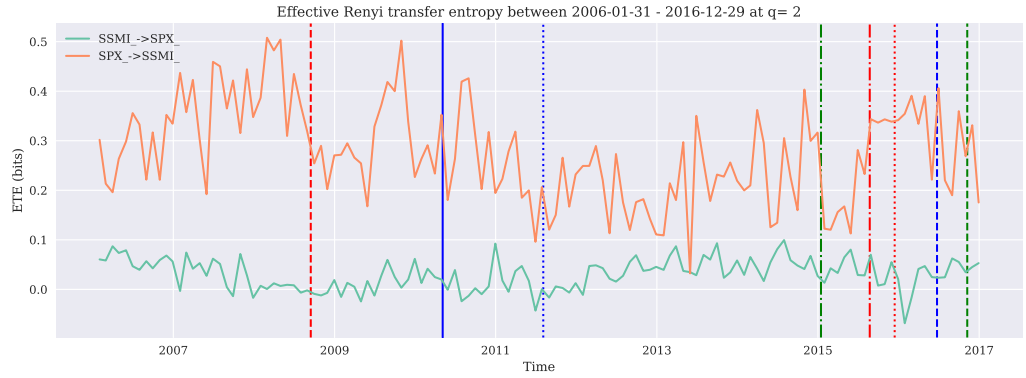
Figure 4.11: Effective Rényi transfer entropy between SSMI and S&P500 for various values of q.



(c) $q = 0.5$



(d) $q = 0.8$



(e) $q = 2.0$

Chapter 5

Forecasting Volatility Using Rényi Information Flows

Objective

The aim of this chapter to present a novel method for improving volatility forecasts by uniting the information flow from effective Rényi transfer entropy with the envelope based Gaussian process method. This chapter presents a large empirical study of the various ERTE parameters and their impact on improving loss performance of the envelope GP method for forecasting financial volatility of forex and market indices. The aim is to empirically find the values of emphasis parameter q, the block length k, and the discretization level S that produce the most consistent substantial gains in performance over the standalone envelope GP forecasts and the mcsGARCH benchmark.

5.1 Introduction

For the return time series of a given symbol, the envelope Gaussian process method introduced in Chapter 3 performs significantly well compared to strong benchmarks such as mcsGARCH. In the previous Chapter, effective Rényi transfer entropy was introduced and it was found that it can help in understanding the information flow between financial

markets. Recent literature has discussed the potential application of using entropy based measures for measuring financial volatility [7, 62, 67]. Taking inspiration from this body of literature, a new approach for forecasting volatility is developed here by folding the effective Rényi transfer entropy into the forecasts from the envelope GPs developed earlier. This technique is then tested empirically across a large dataset of market indices and forex data. The parameters for ERTE calculation are empirically determined in order to achieve consistent performance gains over benchmarks.

The rest of the chapter is arranged as follows: the method for combining ERTE with GP forecasts is introduced, followed by the methodology of experiments. The empirical selection of the emphasis parameter q, discretization parameter S, and look back window k and l is discussed. This is followed by a discussion of results against benchmarks. A further comparison is carried out by using ERTE in tandem with mcsGARCH forecasts to establish the utility of the ERTE folding method.

5.2 Combining ERTE with volatility forecasts

In order to generate the volatility forecast for a return time series R_t the GP forecast can be combined with the information flow from the ERTE by bridging the two using the entropy of a normal distribution.

First the volatility forecast for $t+1$ is calculated by using the envelope GP. This forecast is denoted by $\sigma_{GP,t+1}$. Since the volatility distribution at time $t+1$ is not accessible empirically any further assumptions might constrain it unnecessarily. When provided with only a single moment namely with a specified variance σ, the normal distribution has maximum entropy among all real-valued distributions supported on the number line for that particular moment. Therefore, the assumption of normality imposes the minimal prior structural constraint beyond this moment.

The normal distribution denoted by $\mathcal{N}(\mu, \sigma^2)$, has a density function

$$p(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

If $\mathcal{N}_t$ is a normal distribution with a time evolving standard deviation σ_t. The entropy of the normal distribution at time t is given as

$$H_t = \frac{1}{2}\log(2\pi e\sigma_t^2) \tag{5.1}$$

If the GP forecast value of variance $\sigma_{GP,t+1}$ is used, then the entropy of the forecast for time $t+1$ can be written as

$$H_{GP,t+1} = \frac{1}{2}\log(2\pi e\sigma_{GP,t+1}^2) \tag{5.2}$$

Given the returns series R_t, the effective Rényi transfer entropy $ERTE_{t+1}$ between two time series can be found using Eq. 4.3. The choice of secondary time series for calculating the ERTE is dependant on any prior information regarding the interplay of the two series. The ERTE can be considered an addition of entropy to the forecast entropy $H_{GP,t+1}$, such that the combined entropy for $t+1$ is

$$H_{c,t+1} = H_{GP,t+1} + ERTE_{t+1} \tag{5.3}$$

By using Eq. 5.1 again Eq. 5.3 can be written as

$$\frac{1}{2}\log(2\pi e\sigma_{c,t+1}^2) = \frac{1}{2}\log(2\pi e\sigma_{GP,t+1}^2) + ERTE$$

upon simplification this reduces to

$$\sigma_{c,t+1} = \exp(ERTE) \times \sigma_{GP,t+1} \tag{5.4}$$

This means that the ratio of the ERTE adjusted volatility forecast and the GP generated forecast is given by the exponentaiting the value of the ERTE. Using 5.4 the new forecast can be generated by using the GP forecast and the ERTE.

The percentage change in value of the GP forecast based on ERTE is shown in Fig.

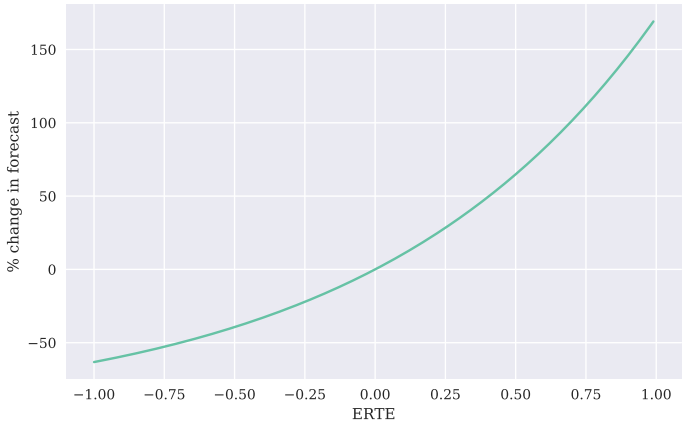


Figure 5.1: The percentage change in the volatility forecast given the effective Rényi transfer entropy. Positive values of ERTE increase the value of our forecast demonstrating a decrease in certainty about the next value to be observed the and negative values decrease the forecast increasing our confidence about the value of the next observation.

5.1. The effect can be seen to be assymetric in nature. For example, it can be seen that a Rényi transfer entropy value of 0.75 bits will give rise to a modification of the volatility forecast by approximately 100%, while a -0.75% entropy transfer results in a downward modification of approximately -50%.

From the discussion in the previous chapter, it is understood that ERTE depends on three key factors: the value of the emphasis paramter q, the discretization S of the returns, and the length of the look back windows k and l for the primary and secondary series.

In the next few sections the impact of each of these on ERTE is explored in order to arrive at the parametrization that provides the highest gains in MSE improvement.

5.3 Datasets for Market Indices and Forex

The data set of global market indices and forex is used in this chapter in order to empirically measure the impact of combining ERTE with envelope GP forecasts. The data consists of the 6 market indices used in the last chapter as given in Table 4.2, and the 11 Forex pairs as given in Table 2.1. This is data is sampled at 1-minute frequency and covers the period from January 01, 2006 to March 31, 2017.

5.4 Methodology of Experiments

In order to calculate the combined forecast from the envelope GPs and the ERTE, each of these is first calculated separately. Given a returns sequence R_t for forecasting the volatility at time $t+1$ the forecast from envelope based GPs is generated in the same manner as Sec. 3.2. The largest window length for the training period for the GP is kept fixed at 360. This is done in order to simply the amount of experiments.

In order to calculate the ERTE two time series are chosen, the primary time series is the one that was used for calculating the GP forecast in the last step, and a secondary series which will provide the information flow to our primary series. The overlapping trading minutes for the two time series are used in the ERTE calculation. Two trading days worth of data is used for calculating the ERTE value. For market indices this is around 240 trading minutes, while for forex pairs this is around 2880 trading minutes. As before the starting and finishing auctions are discarded for indices.

In order to evaluate the ERTE for the chosen time series, three parameter values need to be chosen. First for the emphasis parameter q, second for the look back windows k and l for the primary and secondary series respectively, and third for the discretization parameter S. Since ERTE values can change significantly based on the choice of these parameter values a batch of experiments is carried out using a range of values in order to select the find the values that provide the most consistent performance gains. For these experiments different values of the parameters are tested step by step and the ERTE is calculated from each set of parameter values.

The values of ERTE are combined with the GP forecast in order to generate the modified forecast using the method outlined in the Sec. 5.2. This modified forecast is evaluated against the original GP based forecast for gains in MSE using Eq. 2.25, repeated here for convenience:

$$\%\text{Improvement} = 100 \times (1 - \frac{MSE_{GP}}{MSE_{benchmark}})$$

This is carried out for all the overlapping trading minutes possible for the two time series on a given trading day. This is around 120 for the market indices and 1440 for the

Forex pairs. This is repeated across 75 randomly sampled trading days over the 2006-2017 period. For a given set of overlapping indices this yields around 9000 forecast samples per symbol.

This process is repeated for all the forex symbols and market indices. For these experiments the ERTE for indices is calculated from S&P500 to DAX, and from all 5 European indices to S&P500, for forex pairs this is calculated from all forex pairs to GBP. This leads to experimental results for approximately 1300 trading days or 1.24 million individual returns forecasts. Parameter values are chosen by analysing this large set of results. Preference is given to parameter values which provide the most consistent performance across the experiments for a majority of symbols rather than over-fitting the parameter values to each symbol independently.

The data from 2006-07 is used for empirically finding the ERTE parameter values, these found parameter values are fixed and then benchmarking is done on data from 2008-2017. This separation of training data ensures that no unintentional future information is leaking into the testing results. The experimental results from the selected set of parameters is benchmarked against mcsGARCH, rolling average, and the last observed value. Significance of the results is visualized using notched box plots as before.

5.5 Evolution of ERTE with q

The parameter q is the most significant parameter influencing the ERTE observed between two time series. The function for Rényi transfer entropy $ERTE(q)$ for $q < 1$ varies depending on the time series pair under consideration and the time period for which they are being considered.

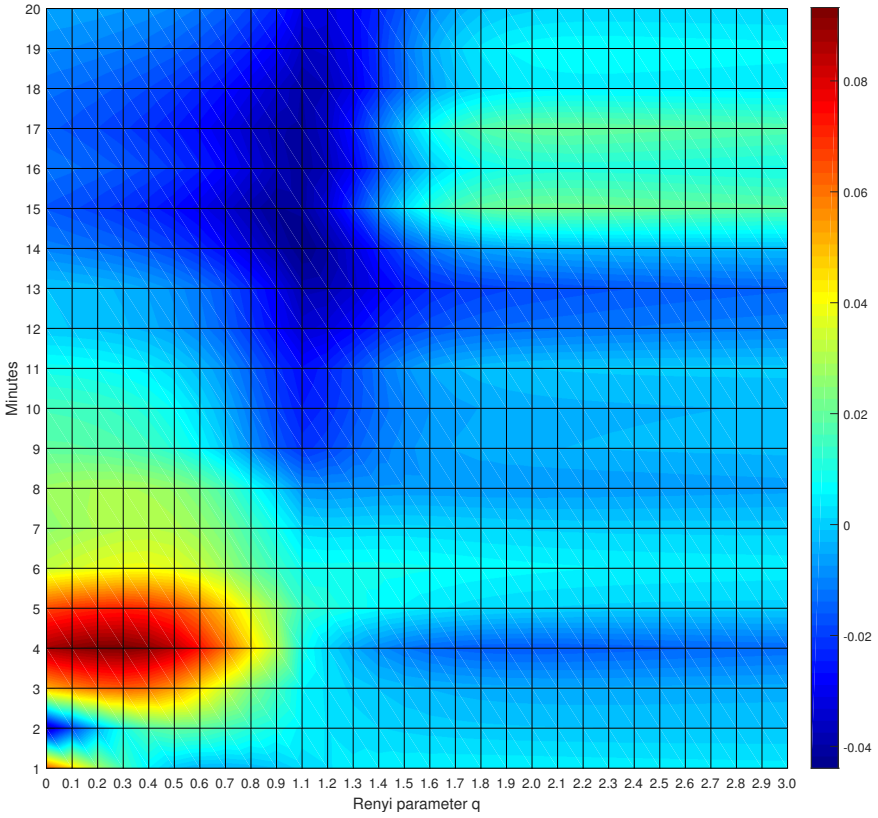
In order to illustrate this relationship two sets of forex pairs are chosen - Euro & GBP and JPY & GBP, and their ERTE is visualised for a range of values for q and the time period of the look back windows k and l. This is presented as heat maps shown in Fig. 5.2 and 5.3. From the heat maps it can be seen that the flows in each direction are substantially different and change significantly based on the choice for q. This is observed

both between the Euro and GBP, as well as JPY and GBP.

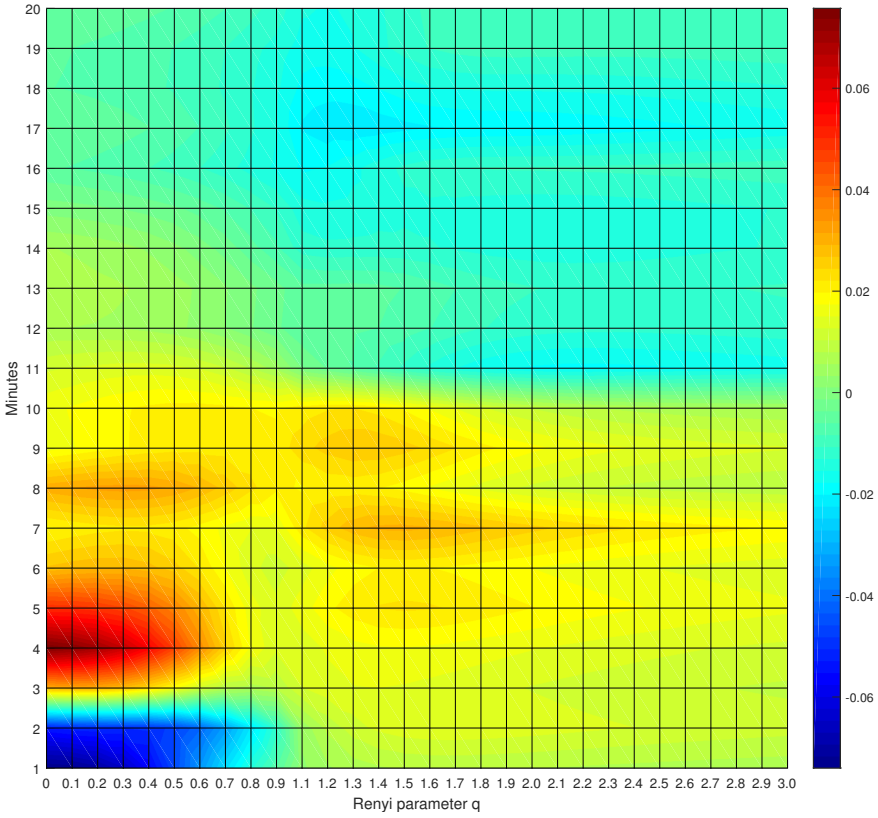
In order to further establish this concept, the minute by minute dynamically evolving ERTE between DAX and S&P500, and between FTSE and S&P500 is shown in Figs. 5.4 and 5.6 over a three week period, calculated at 1-minute frequency using a rolling window of two trading days. It can be seen that the ERTE changed substantially over the course of these few days and the magnitude changes with different values of q, with higher values of ERTE recorded for lower values of q.

A set of experiments is carried out to choose the parameter value of q that provides the best improvement in forecast performance. A range of values of q are tested out with the market indices and the forex data. The other two parameters are kept fixed at $S = 6$ and $k = l = 1$. The results of this The results from these experiments are shown in Table 5.1. The last column shows the result when the ERTE for all four values is averaged and used for the forecast modification. It can be seen from the results that while no single q values performs consistently better but the ERTE averaged from multiple q values produces the best performance gains consistently across all 6 indices and forex pairs.

Therefore instead of choosing one single value for the emphasis parameter q, the technique for combined ERTE from four q values is selected.

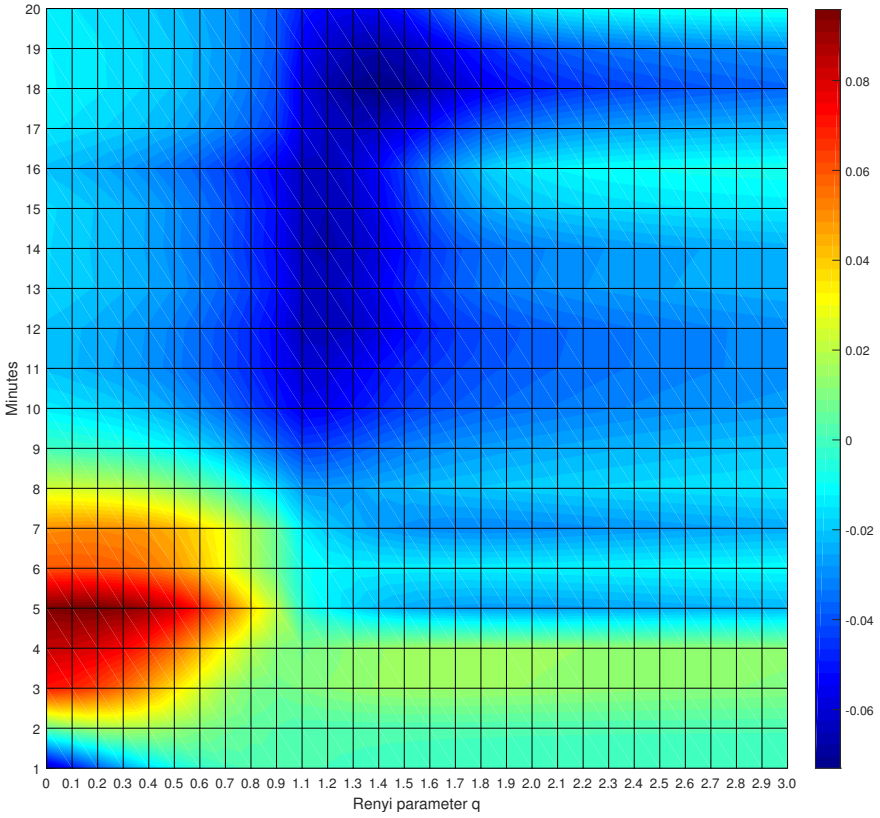


(a) $ERTE_{GBP \to EUR}$

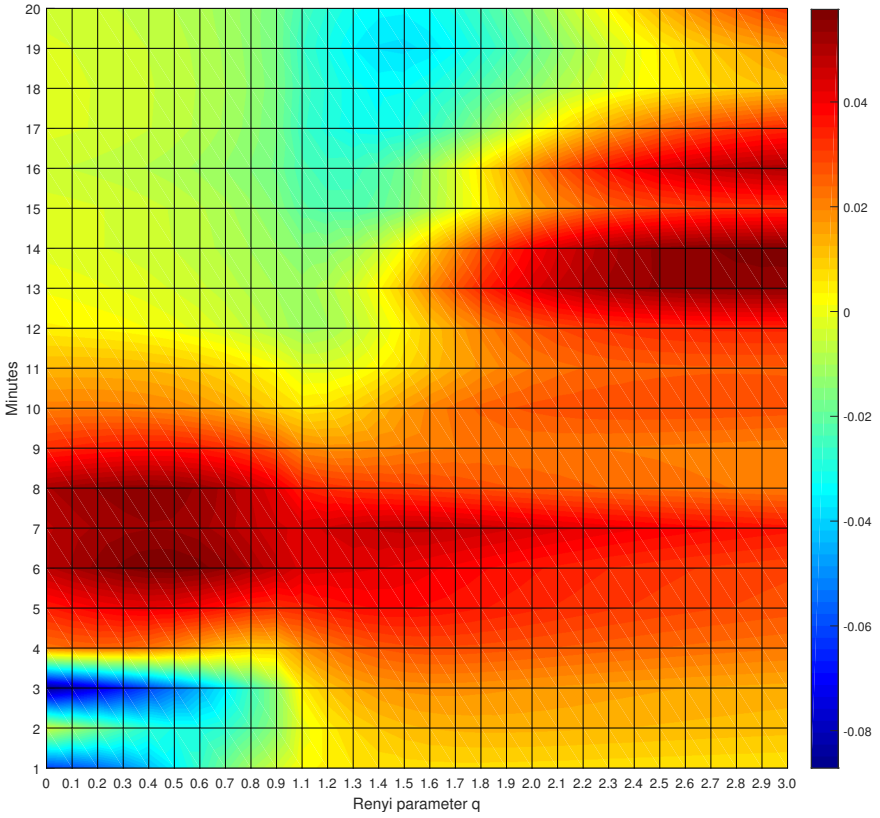


(b) $ERTE_{EUR \to GBP}$

Figure 5.2: The effective Rényi transfer entropy between GBP and EUR with $0 < q < 3$ on the horizontal axis and $k = l = [1, 20]$ minutes on the vertical axis. Positive values indicate an influx of information which results in decrease in variance.



(a) $ERTE_{GBP \to JPY}$

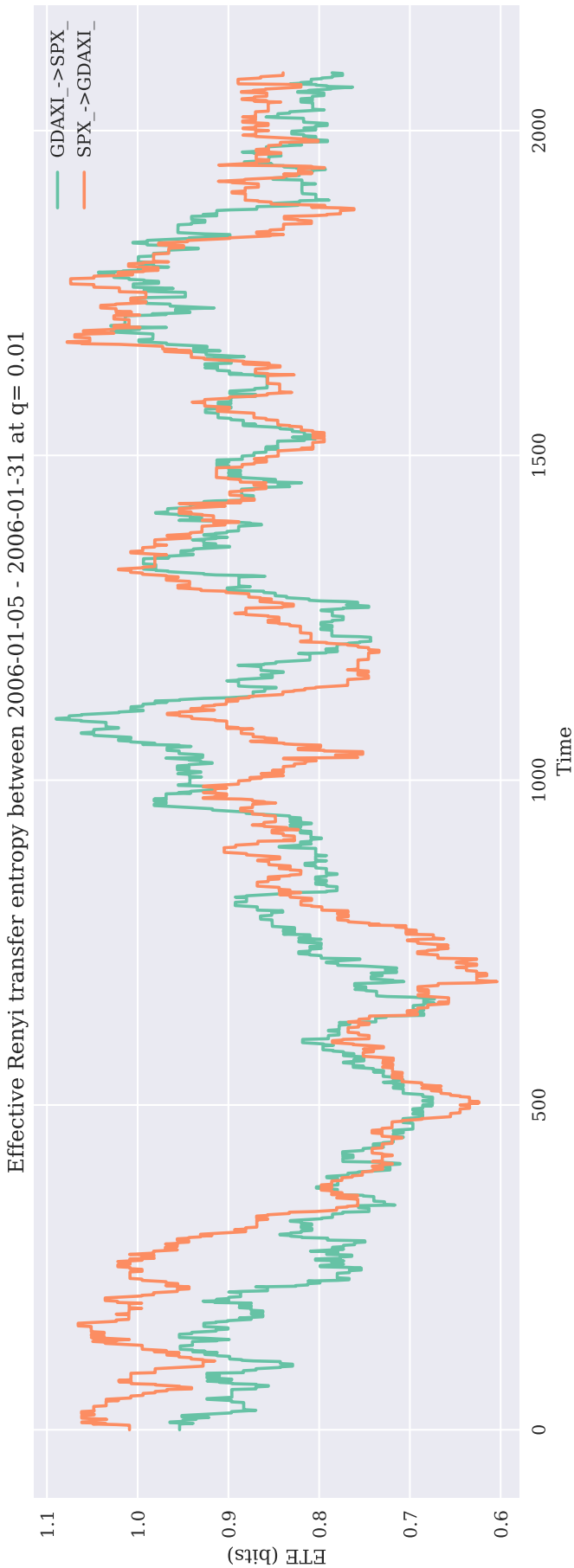


(b) $ERTE_{JPY \to GBP}$

Figure 5.3: The effective Rényi transfer entropy between GBP and JPY with $0 < q < 3$ on the horizontal axis and $k = l = [1, 20]$ minutes on the vertical axis. Positive values indicate an influx of information which results in decrease in variance.

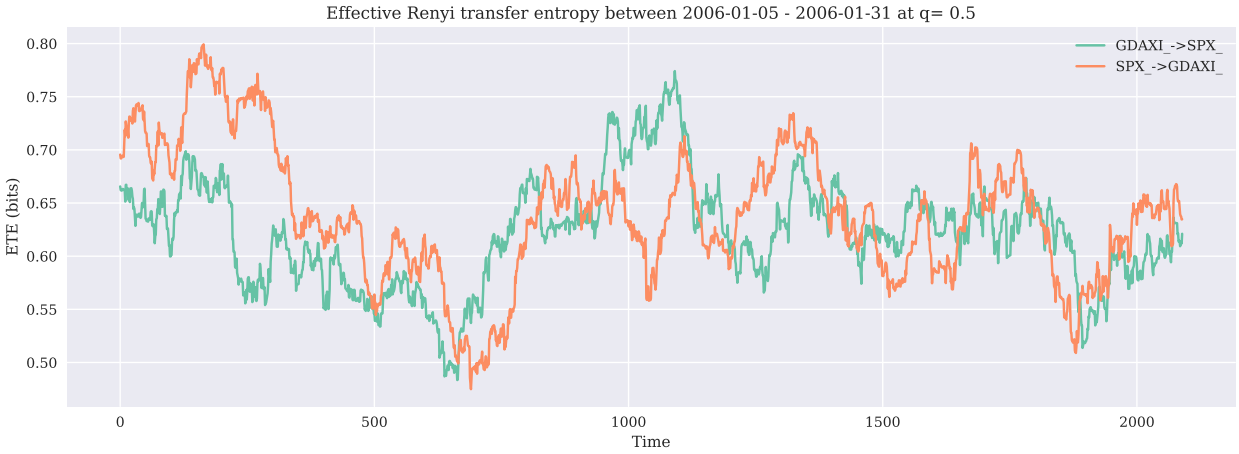
ERTE pair	$q = 0.01$	$q = 0.5$	$q = 0.8$	$q = 2.0$	$q =$averaged
DAX $\rightarrow$ S&P500	4.81%	5.52%	2.49%	2.78%	4.90%
FTSE $\rightarrow$ S&P500	2.90%	2.82%	3.47%	3.92%	1.58%
CAC40 $\rightarrow$ S&P500	3.03%	0.61%	4.16%	4.43%	5.70%
ATX $\rightarrow$ S&P500	2.07%	4.01%	2.93%	2.26%	3.38%
SSMI $\rightarrow$ S&P500	1.95%	4.65%	0.35%	1.27%	2.36%
AUD $\rightarrow$ GBP	2.35%	4.29%	4.17%	2.69%	2.96%
CAD $\rightarrow$ GBP	3.63%	5.31%	1.00%	4.25%	4.90%
CHF $\rightarrow$ GBP	5.01%	4.84%	3.36%	-3.00%	3.98%
EUR $\rightarrow$ GBP	7.36%	2.65%	4.60%	3.15%	7.21%
JPY $\rightarrow$ GBP	4.75%	-5.88%	3.01%	6.43%	4.46%
NOK $\rightarrow$ GBP	4.03%	8.07%	-1.54%	4.67%	3.54%
NZD $\rightarrow$ GBP	2.24%	3.27%	-4.87%	8.25%	3.36%
PLN $\rightarrow$ GBP	2.35%	3.56%	3.42%	4.77%	3.82%
SGD $\rightarrow$ GBP	2.90%	-2.53%	4.89%	3.72%	4.00%
TRY $\rightarrow$ GBP	1.23%	3.01%	5.73%	4.01%	2.86%
ZAR $\rightarrow$ GBP	1.26%	6.44%	-5.81%	-8.32%	1.88%

Table 5.1: Improvement in MSE by using ERTE enhanced envelope GP forecasts compared to envelope GP forecasts alone, for various values of q. Other parameters fixed with $S = 6$, $k = l = 1$

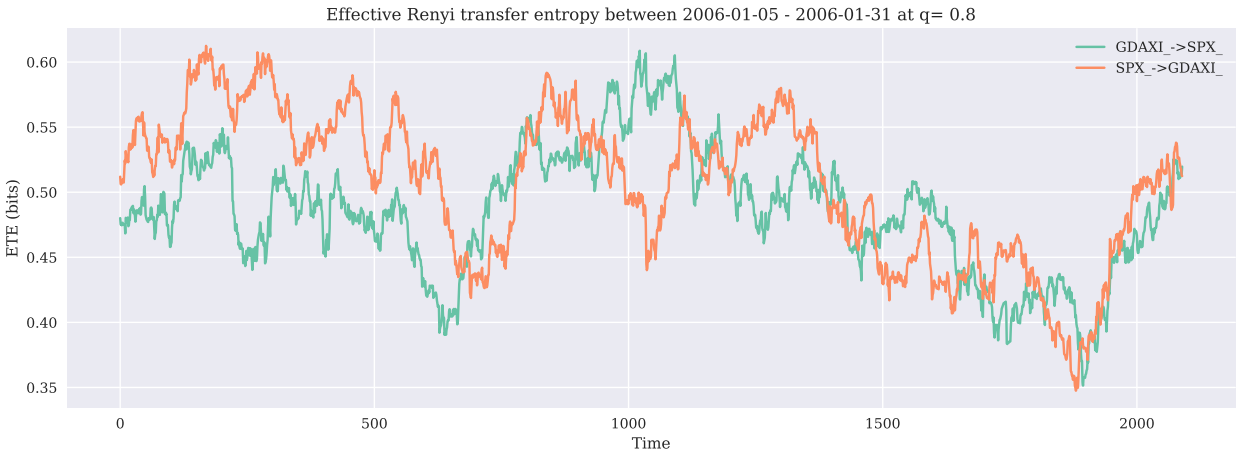


(a) $q = 0.01$

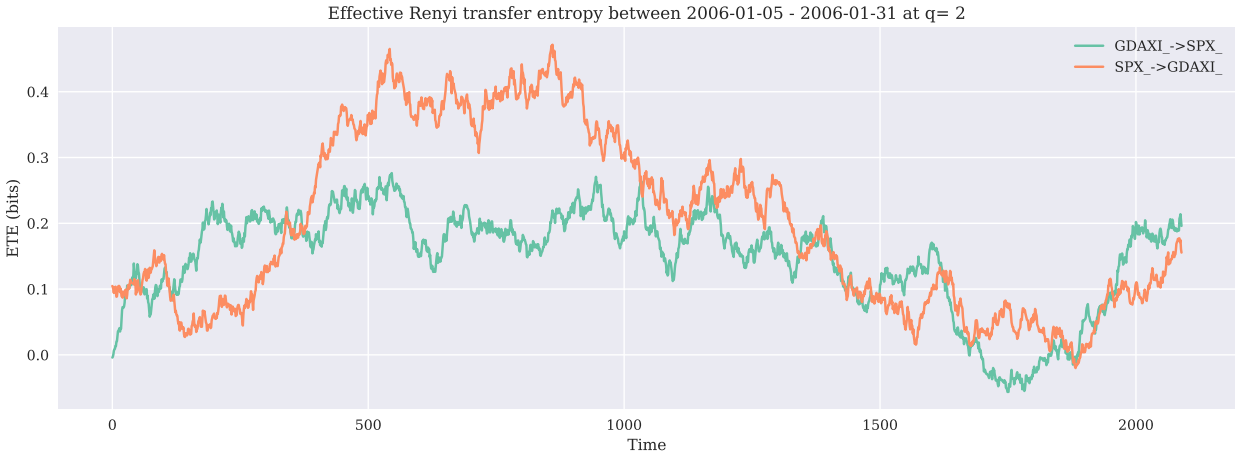
Figure 5.4: Minute level variation in effective Rényi transfer entropy between DAX and S&P500 for various values of q.



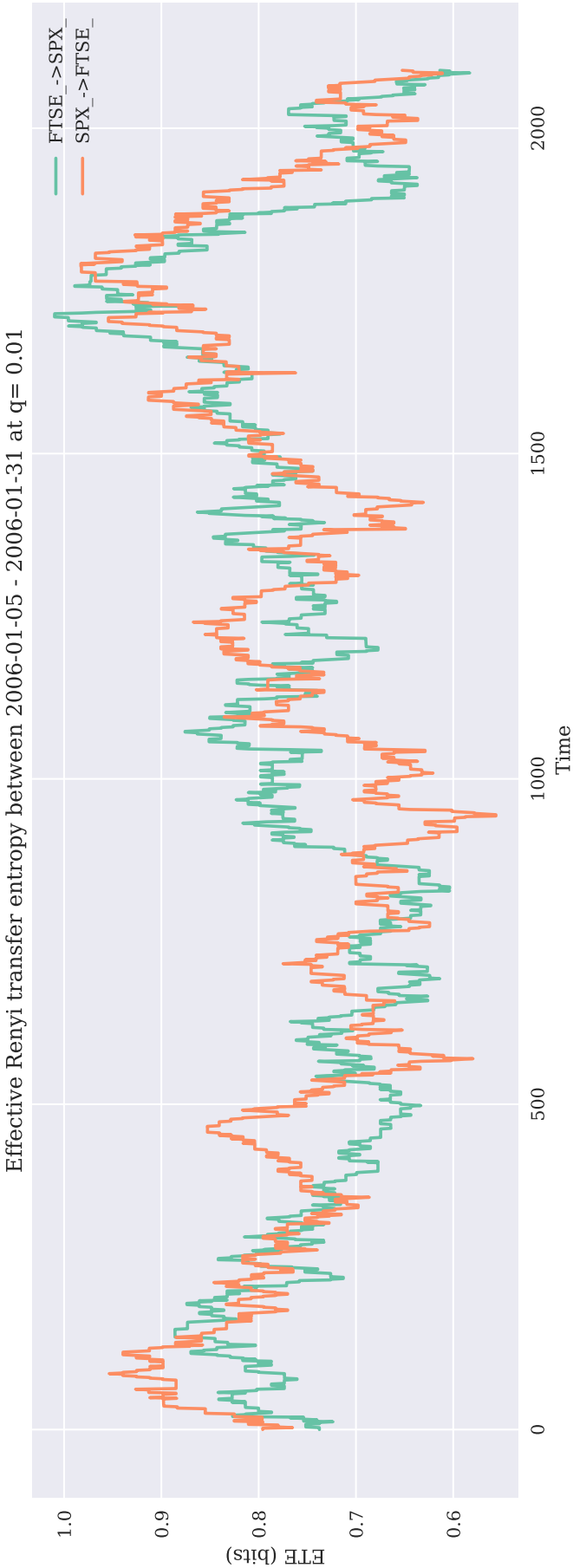
(c) $q = 0.5$



(d) $q = 0.8$

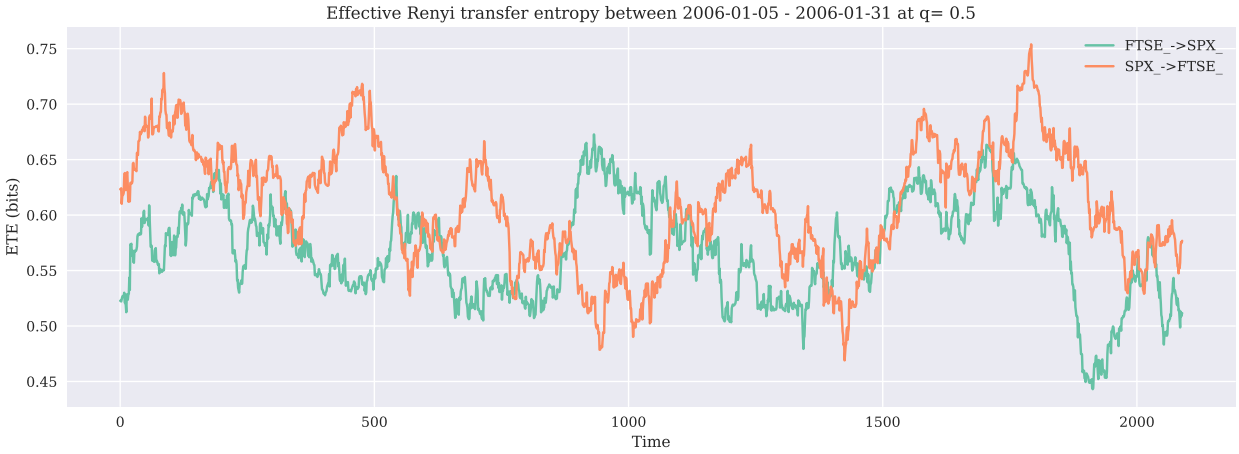


(e) $q = 2.0$

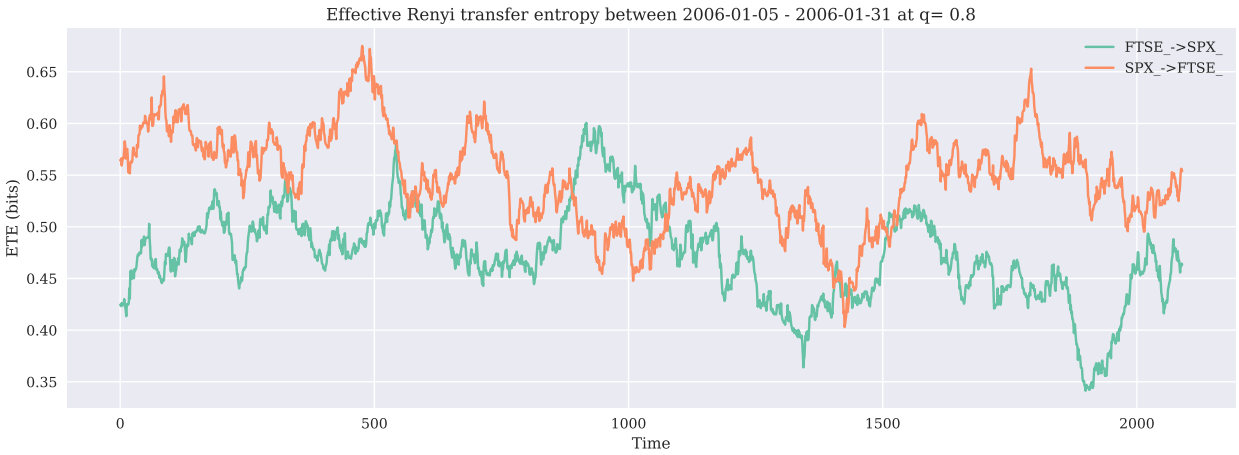


(a) $q = 0.01$

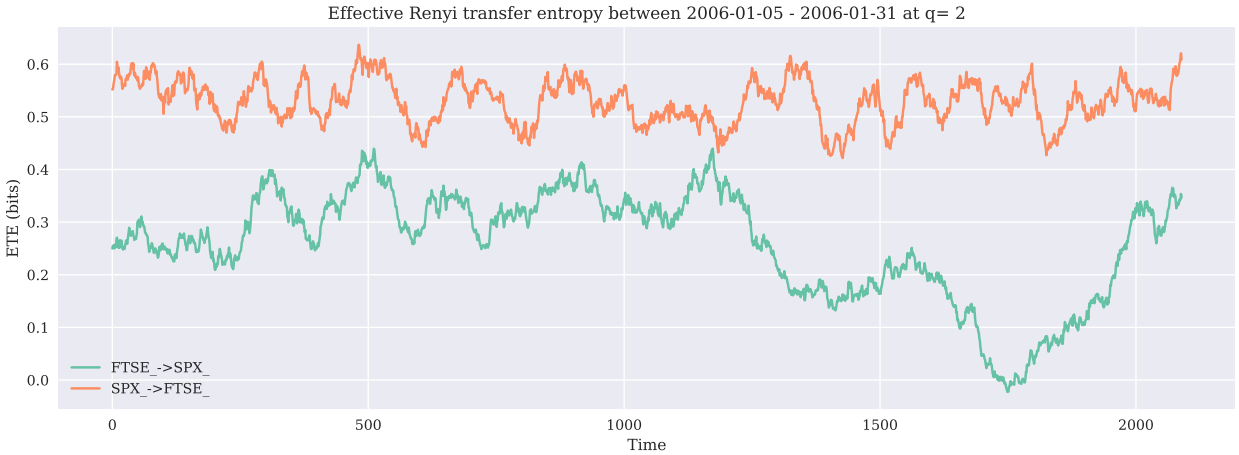
Figure 5.6: Minute level variation in effective Rényi transfer entropy between FTSE and S&P500 for various values of q.



(c) $q = 0.5$



(d) $q = 0.8$



(e) $q = 2.0$

5.6 Analysing Gain from ERTE with Discretization Level

It is well documented in literature that the discretization level S of the returns series has significant impact on the measured value of effective Rényi transfer entropy [45, 40, 19, 20]. The effect of S on ERTE is empirically studied here in order to improve the gains possible by using ERTE with GPs.

The experiments are carried out for five values of S from 2 to 6. When $S = 2$ only the signs of the returns are being used to exchange the information. With $S > 2$, information about the magnitude of the returns also contributes to the ERTE. Marchinski et al. (2002) [44] found that the value of ERTE increased up to $S = 5$ and decayed rapidly thereafter. In the experiments carried out here it is found that the same effect occurs when $S > 6$ for the data under consideration. For the experiments the other two parameters are fixed at $q = 0.01$ and $k = l = 1$.

The results from the experiments are given in Table 5.2. It can be seen that a steady increase in performance is observed as S increases up to 6. When $S > 6$ is selected finite sample effects become highly magnified and ERTE values become highly noisy and don't compute consistently.

The discretization parameter S is selected to be 6 from this set of experiments

5.7 Evolution of ERTE with Block Length

The importance of the block length k and l for ERTE was seen earlier in this chapter in the heatmaps in Figs. 5.2 and 5.3. The value of k and l possible depends on the size of data available The data requirement grows as a multiple of S^{k+l} [44]. Jizba et al. (2012) [40] established that given a data set l should be fixed to 1 and the value of k should be maximized. Dimpfl & Peter (2014) [20] used $k = l = 1$ for calculating monthly ERTE. In this section the value of l for the secondary series is fixed to 1 and the effect of changing k is empirically tested for 5 values of k from 1 to 5. The values of other two parameters are fixed at $q = 0.01$ and $S = 6$ for the experiments.

ERTE pair	$S=2$	$S=3$	$S=4$	$S=5$	$S=6$
DAX $\to$ S&P500	1.88%	1.70%	2.42%	3.12%	4.81%
FTSE $\to$ S&P500	-0.01%	0.20%	1.68%	1.45%	2.90%
CAC40 $\to$ S&P500	1.04%	1.56%	0.29%	3.41%	3.03%
ATX $\to$ S&P500	-0.32%	-0.77%	1.03%	1.24%	2.07%
SSMI $\to$ S&P500	-2.65%	0.81%	0.29%	2.36%	1.95%
AUD $\to$ GBP	-0.10%	2.05%	1.14%	3.39%	2.35%
CAD $\to$ GBP	0.35%	-0.26%	3.74%	2.94%	3.63%
CHF $\to$ GBP	1.23%	2.52%	-0.18%	4.23%	5.01%
EUR $\to$ GBP	2.83%	2.28%	4.73%	6.72%	7.36%
JPY $\to$ GBP	-0.78%	2.65%	1.09%	5.63%	4.75%
NOK $\to$ GBP	0.66%	1.48%	2.50%	3.85%	4.03%
NZD $\to$ GBP	0.73%	3.31%	3.50%	2.11%	2.24%
PLN $\to$ GBP	-3.61%	1.48%	3.34%	3.10%	2.35%
SGD $\to$ GBP	1.23%	-2.88%	0.11%	1.19%	2.90%
TRY $\to$ GBP	1.23%	-0.29%	2.74%	-0.19%	1.23%
ZAR $\to$ GBP	-0.06%	-2.51%	1.03%	0.50%	1.26%

Table 5.2: Improvement in MSE by using ERTE enhanced envelope GP forecasts compared to envelope GP forecasts alone, for various values of S. Other parameters fixed with $q = 0.01$, $k = l = 1$.

The results from the set of experiments are summarized in Table 5.3. It can be seen that for minute level forecasting, the gains in ERTE performance decay rapidly for $k > 1$, with the ERTE assisted forecasts performing worse than standalone GP forecasts in some cases for values of $k > 2$. From this set of results the value of the block length parameter k is fixed to 1 for subsequent experiments.

ERTE pair	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
DAX $\rightarrow$ S&P500	4.81%	3.83%	0.81%	0.99%	0.93%
FTSE $\rightarrow$ S&P500	2.90%	1.98%	0.78%	-2.3%	0.37%
CAC40 $\rightarrow$ S&P500	3.03%	1.47%	1.05%	0.96%	-2.49%
ATX $\rightarrow$ S&P500	2.07%	2.28%	2.46%	-3.97%	-0.75%
SSMI $\rightarrow$ S&P500	1.95%	2.14%	-0.71%	1.70%	0.14%
AUD $\rightarrow$ GBP	2.35%	3.26%	0.81%	-2.67%	-1.13%
CAD $\rightarrow$ GBP	3.63%	2.03%	3.96%	-0.19%	-0.44%
CHF $\rightarrow$ GBP	5.01%	2.59%	-0.07%	3.51%	1.23%
EUR $\rightarrow$ GBP	7.36%	3.65%	0.48%	-0.01%	-2.53%
JPY $\rightarrow$ GBP	4.75%	3.84%	1.09%	1.61%	0.61%
NOK $\rightarrow$ GBP	4.03%	3.81%	5.58%	0.44%	-1.55%
NZD $\rightarrow$ GBP	2.24%	3.02%	-3.47%	0.85%	-1.16%
PLN $\rightarrow$ GBP	2.35%	2.76%	2.10%	2.69%	1.90%
SGD $\rightarrow$ GBP	2.90%	2.21%	4.36%	2.03%	-2.01%
TRY $\rightarrow$ GBP	1.23%	1.26%	-3.47%	-0.377%	0.44%
ZAR $\rightarrow$ GBP	1.26%	1.01%	-0.37%	-1.33%	-1.21%

Table 5.3: Improvement in MSE by using ERTE enhanced envelope GP forecasts compared to envelope GP forecasts alone, for various values of k. Other parameters fixed with $q = 0.01$, $S = 6$, $l = 1$.

5.8 Discussion of results

It is found from the experiments that the ERTE modified forecasts perform significantly better than all other benchmarks under consideration, performing 14.3% better than the

mcsGARC benchmark for market indices and 11.7% for forex pairs. It also outperforms the standalone envelope based GP forecasts by 3.88% across the 18 symbols under consideration, performing 4.06% better on the market indices and 3.79% better on the forex data in the MSE improvement. The significance of the results is shown in Fig. 5.8 as notched boxed plots.

For the calculating improvement in S&P500 volatility forecasts DAX is used as the second series for calculating $ERTE_{DAX \to S\&P500}$. For the improvement in the other 5 market indices S&P500 is used as secondary series for calculating the the $ERTE_{S\&P500 \to Index}$ for each index under consideration. A sample visualization of the forecasts is provided in Fig. 5.9. It can be seen that the transformed forecast benefits from the underlying forecast from GP and folds in further information about potential changes in the time series through the ERTE. As an aside, the downward diurnal trend in absolute returns can also be observed in this figure since the returns are from the beginning of the trading day for S&P500.

Similarly for the forex symbols, the secondary series for all symbols is the GBP in calculating the $ERTE_{GBP \to FX}$ for each forex symbol. For GBP the Euro is used as the secondary series to calculate $ERTE_{EUR \to GBP}$.

These results demonstrate that ERTE can be a very important addition to supplement the forecasting performance of existing techniques. It is a flexible and non-parametric way of folding information from other time series in to the primary time series of interest. The data does not have to be limited to other financial time series. Varied data such as Google search trends and Wikipedia access results have also been subjected to ERTE analysis (Zimmerman 2018)

	Improv. over envelope GP	Improv. over mcsGARCH	Improv. over rolling average	Improv. over previous value
MSE				
S&P500	4.90%	16.06%	16.26%	30.93%
DAX	2.01%	12.91%	11.19%	25.78%
FTSE	4.54%	14.33%	15.47%	27.57%
CAC40	3.16%	11.29%	12.53%	26.81%
ATX	5.97%	17.43%	16.26%	28.45%
SSMI	3.81%	13.93%	14.37%	28.33%
RMSE				
S&P500	1.13%	5.25%	5.51%	17.71%
DAX	3.34%	4.72%	5.28%	14.38%
FTSE	1.01%	6.23%	8.24%	11.88%
CAC40	3.12%	7.11%	6.27%	9.95%
ATX	1.55%	11.10%	8.33%	16.07%
SSMI	2.10%	6.62%	7.54%	11.06%
QLIKE				
S&P500	1.54%	4.95%	6.05%	15.51%
DAX	4.00%	4.98%	4.14%	13.63%
FTSE	1.22%	6.84%	7.92%	19.47%
CAC40	1.51%	9.08%	6.95%	17.49%
ATX	2.26%	8.38%	7.73%	15.05%
SSMI	2.43%	5.53%	6.05%	15.45%

Table 5.4: Averaged values for the percentage improvement in MSE, RMSE, and QLIKE of using the Rényi transfer entropy with the GP envelope method over the benchmarks.

	Improv. over envelope GP	Improv. over mcsGARCH	Improv. over rolling average	Improv. over previous value
MSE				
AUD	3.71%	10.73%	11.67%	22.39%
CAD	2.47%	11.97%	11.37%	35.93%
CHF	1.28%	9.66%	10.08%	22.86%
EUR	3.47%	11.45%	13.44%	30.17%
GBP	7.21%	13.93%	15.10%	31.11%
JPY	4.36%	12.21%	10.36%	26.56%
NOK	3.05%	13.17%	14.05%	27.34%
NZD	2.81%	13.53%	12.88%	32.32%
PLN	6.68%	10.95%	13.55%	28.46%
SGD	4.98%	9.54%	9.12%	33.29%
TRY	3.10%	14.03%	15.89%	23.32%
ZAR	2.39%	9.01%	11.75%	23.55%

Table 5.5: Averaged values for the percentage improvement in MSE by using the Rényi transfer entropy with the GP envelope method over the benchmarks.

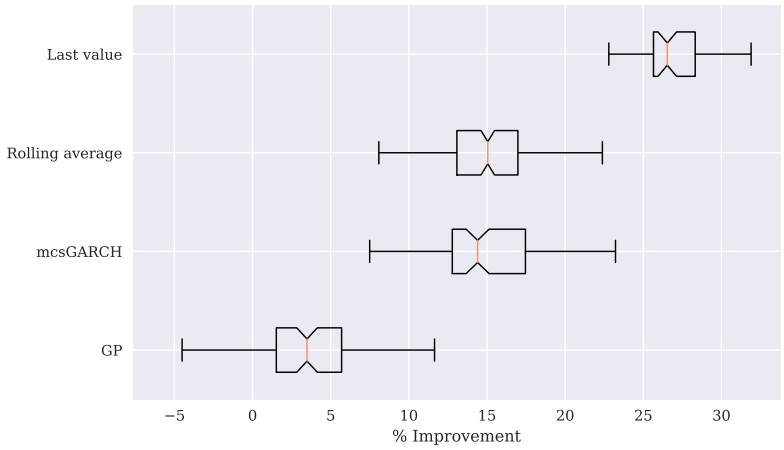


Figure 5.8: Improvement in MSE using ERTE combined with the envelope method, compared to standalone envelope GP, mcsGARCH, rolling average, and last observed value benchmarks. The results are all highly significant since the notches around all the median are in the positive and well beyond zero.

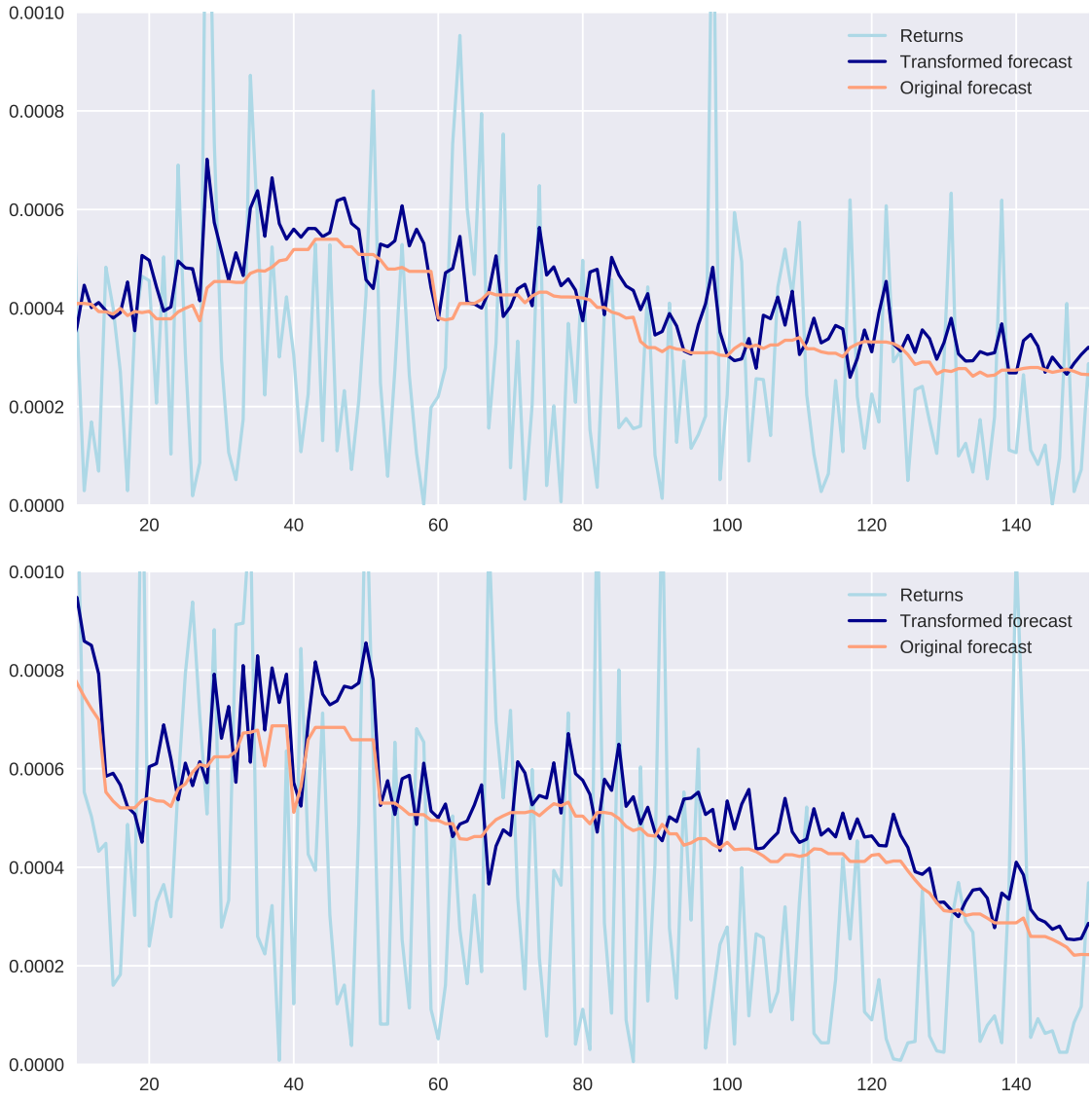


Figure 5.9: ERTE based trasnformed forecasts for S&P500 shown in dark blue, compared to the original forecasts in red, using $ERTE_{\text{DAX}\to\text{S\&P500}}$ to modify the envelope GP forecasts. $q = combined$, $S = 6$, $k = l = 1$.

5.9 A Better Apple

Although the results presented in this chapter are highly significant, it might be construed that it's not a pure apples-to-apples comparison since the other benchmarks don't have access to data from the secondary time series.

In order to make the comparison fair, a benchmark that makes use of information from two time series was searched for. While many econometric models exist for making use of two time series, none were found that were specifically crafted for intraday volatility prediction and had been tested in literature for 1-minute frequency data. In order to overcome this challenge, a better apple is designed by using the mcsGARCH forecasts and modifying them using the ERTE. These modified forecasts are compared against the mcsGARCH standalone forecasts, the envelope GP forecasts, and the modified envelope GP forecasts.

The results from this comparison are presented in Table. 5.6. It is found that while the modified forecasts from mcsGARCH don't outperform the modified GP forecasts, they perform better than mcsGARCH across all symbols and in a few instances perform better even than the standalone GP forecasts. This indicates that the ERTE modification of forecasts from another established volatility prediction model can serve as valuable tool in the volatility practitioner's tool kit

ERTE pair	Improvement against		
	env. GP + ERTE	env. GP	mcsGARCH
S&P500	−24.66%	−5.17%	4.64%
DAX	−19.26%	−5.91%	3.87%
FTSE	−18.96%	−1.43%	2.59%
CAC40	−19.85%	−2.87%	6.51%
ATX	−23.56%	−7.10%	2.03%
SSMI	−27.61%	2.33%	9.84%
AUD	−22.04%	0.85%	8.95%
CAD	−17.56%	−4.70%	3.49%
CHF	−13.89%	2.59%	2.89%
EUR	−19.55%	1.38%	5.87%
GBP	−23.59%	−1.08%	6.37%
JPY	−17.26%	−1.77%	2.95%
NOK	−19.15%	−2.61%	3.78%
NZD	−20.01%	−5.58%	2.53%
PLN	−14.93%	−3.44%	5.43%
SGD	−16.45%	1.07%	6.29%
TRY	−21.31%	−2.71%	4.29%
ZAR	−12.49%	−3.95%	2.09%

Table 5.6: Improvement in MSE by using ERTE enhanced mcsGARCH forecasts compared to other benchmarks. Negative values indicate that the mcsGARCH performed worse than the benchmark. Parameters fixed at $q = \textit{combined}$, $S = 6$, $k = l = 1$.

5.10 Conclusion

In this chapter a novel approach for volatility forecasting using Rényi transfer entropy was developed. The effective Rényi transfer entropy provides a valuable model free-approach for quantifying information flows among forex pairs and market indices. Envelope based GP forecasts are updated by using the Rényi information flows calculated between pairs of financial time series. This approach is applied to minute-level data set of 6 market indices and 12 currency pairs. The parameter values for emphasis parameter q, discretization parameter S, and lookback windows k and l are chosen empirically for performance gains.

The experiments establish that the ERTE modified GP forecasts perform better than all other benchmarks under consideration. They perform 14.3% better than the mcsGARCH benchmark for market indices and 11.7% for forex pairs. They also outperform the standalone envelope based GP forecasts by 3.88% across the 18 symbols under consideration, performing 4.06% better on the market indices and 3.79% better on the forex data in the MSE improvement. A further comparison is carried out by using ERTE to modify mcsGARCH forecasts and it is found that while these do not outperform ERTE modified GP forecasts, they do show performance gains over the standalone mcsGARCH forecasts.

This chapter empirically establishes that ERTE modified envelope based GPs are a strong method for volatility forecasting and the ERTE modification of forecasts from another established volatility prediction model can prove to be a valuable tool for the volatility practitioner.

Chapter 6

Forecasting Volatility in Cryptocurrency markets

Objective

The aim of this chapter is to apply the volatility forecasting techniques developed so far to one of the most volatile investment vehicles of our times - cryptocurrencies. This chapter demonstrates that Gaussian processes perform significantly better than competing benchmarks for forecasting volatility on cryptocurrency data. This chapter presents the largest study of its kind in literature, carried out on a set of as many as 10 crypto currency pairs, sampled at a 5-minute frequency. It also provides the first presentation of applying Gaussian processes to a large set of cryptcurrencies to forecast volatility. This chapter also extends the previous work to incorporate information from multiple time series into the volatility forecast. An innovation is devised for generating the effective Rényi transfer entropy from multiple secondary series towards one primary series.

6.1 Introduction

> "Dad, I want **1** bitcoin for my birthday", a boy asked his bitcoin-investing dad. His dad replied,
>
> "What? You want $15,554? $14,354 is a lot of money! What do you need $16,782 for anyway?"

Cryptocurrencies are infamous for their volatility. They have been considered one of the larger bubbles of recent times with high amount of speculation, a large asymmetry in the knowledge of market participants, and new crypto offerings entering the market everyday.

Cryptocurrencies are digital or virtual currencies that maintain one form or another of a secured unalterable historical ledger of trades and balances. Bitcoin was the first decentralized cryptocurrency powered by a public ledger called a blockchain that records all historical transactions. At the time of writing there are around 2017 cryptocurrencies available to trade with a total current market cap of approximately $215 billion. There are an estimated 14,000 exchanges where these cryptocurrenices can be traded.

Cryptocurrencies can be catergorized into coins and tokens due to the difference in their structure. Coins are separate currencies with their own separate blockchain mechanism, while tokens operate on top of another blockchain. Examples of tokens include OmiseGo, Augur, and Tether. Some coins are based on the bitcoin blockchain protocol e.g. Litecoin and BitcoinCash, while others have come up with their own independent and alternative blockchain protocol e.g. Ethereum, Monero, and Neo. This creates some interesting linkages between the price fluctuations of cryptocurrencies, with token and coin prices being impacted by changes in the price of the parent blockchain protocol. Osterrieder et al. (2017) [50] present a statistical analysis of the top traded cryptocurrencies.

The known work on cryptocurrencies mostly focuses on the study of Bitcoin and/or is found to use daily data. Chu et al. (2017) [14] evaluate the use of of 12 GARCH models to forecast daily volatility for Bitcoin. Pichl (2017) [52] show that HARmodels are more robuts in modelling Bitcoin volatility than GARCH models. Bouri et al. (2017) [12] use asymmetric GARCH models to investigate the relationship between daily price returns

and volatility changes in the Bitcoin market around the price crash of 2013. The work in literature most similar to the one presented in this chapter is Guo et al. (2018) [31] where they evaluate multiple machine learning methods including an instance of GPs for forecasting volatility using hourly data for Bitcoin. Some key research choices made by them are differnt than the frame work used in this thesis. For example, they use non-robust loss functions in their evaluation, use arithmetic rather than log returns in their framework, and benchmark the performance against GARCH and EGARCH models.

This chapter adds to this growing body of literature and provides one of the largest set of experiments carried out on forecasting volatility in cryptocurrency markets at 5-minute frequency for a data set consisting of 10 cryptocoins. This chapter also presents a novel method for estimating effective Rényi transfer entropy from multiple secondary series towards one primary series.

The rest of this chapter is arranged as follows: first the challenges of gathering cryptocurrency data are briefly outlined, then the novel measure for calculating ERTE from multiple time series is described. Then the choice of volatility proxy is described, followed by a brief description of the dataset and the methodology of experiments. Finally, the results of the experiments are presented with some commentary on potential caveats in evaluating the forecasting performance.

6.2 Challenges of Working with Cryptocurrency Data

Unlike stocks, market indices, and forex, the price for cryptocurrencies does not originate from a single data source and this makes accurate price discovery quite challenging. Most commercially available datasources such as Coinmarketcap aggregate the date from multiple crypto trading exchanges to produce the quoted price. This means that the aggregation algorithm, along with varying liquidity accross different exchanges can give rise to multiple versions of the ground truth.

For the data presented in this chapter a custom API was used for getting Open-Low-High-Close (OHLC) data at a 5-minute frequency from three different crypto trading

exchanges: Bittrex, Bitfinex, and Kraken. All three are leading exchanges for cryptocurrencies and provide sufficient liquidity for price accuracy.

Crypto currencies don't behave like traditional markets in many respects. Since the landscape of crypto exachanges is still quite fragmented many arbitrage opportunities exist in these markets. A large volume of cryptcurrencies also trade in dark pools and these volumes fail to be reflected in the price dynamics observed in price time series.

6.3 Folding Multiple Time Series into the Volatility Forecast

This section presents an important innovation for the non-parametric method for calculating the information flow from multiple time series. In order to incorporate the effective Rényi transfer entropy into the forecast, previously one pair was selected and the ERTE from it was incorporated into the forecast generated from the envelope based Gaussian process. In this section this approach is explored further to expand its application to multiple time series. The key innovation presented here is that the non-parametric formulation for ERTE is extended to enable it to use data from multiple secondary time series and generate a combined ERTE. This combined ERTE is to be used later with the volatility forecast generated from envelope GPs for the primary series.

In order to calculate the modified forecast σ_{t+1} for time $t+1$ for primary series $R_{p,t}$ it is possible to calculate the effective Rényi tranfer entropy from multiple secondary series $R_{s_1,t}, R_{s_2,t}, \cdot, R_{s_m,t}$. Each ERTE generated will produce a different value of the modified forecast. One simple method might be to combine multiple ERTEs into one averaged value, but a key conceptual challenge with simple averaging is that it does not take into account the space in which effective Rényi transfer entropies might operate. If this space is a vector space then a simple average will fail to represent the combined impact of multiple ERTEs on the primary series.

For a more nuanced empirical approach the equation for the Rényi transfer entropy is

revisited here. The information flow from Y to X as

$$\mathrm{RTE}^{q}_{Y\to X}(k,l) = \frac{1}{1-q}\log\frac{\sum_x \phi_q(x_t^{(k)})p^q(x_{t+1}\mid x_t^{(k)})}{\sum_{x,y}\phi_q(x_t^{(k)},y_t^{(l)})p^q(x_{t+1}\mid x_t^{(k)},y_t^{(l)})} \tag{6.1}$$

with $q>0$ and escort distribution $\phi_q(x)=p^q(x)/\sum_x p^q(x)$ is used to normalize the weighted distribution.

In the simple case of one primary series and one secondary series it is taken to be the case that each observation $x_t^{(k)}$ of the primary series is of the form $(r_{p,t}, r_{p,t-1}, \cdots, r_{p,t-k+1})$ and each observation $y_t^{(l)}$ of the secondary series is of the form $(r_{s,t}, r_{s,t-1}, \cdots, r_{s,t-l-1})$ where r_p and r_s are observations from the primary and secondary series R_p and R_s. But if R_s might be comprised from m secondary series $R_{s_1}, R_{s_1}, \cdots, R_{s_m}$ such that each observation each observation of $y_t^{(l)}$ is of the form,

$$(r_{s_1,t}, r_{s_1,t-1}, \cdots, r_{s_1,t-l-1}, \quad r_{s_2,t}, r_{s_2,t-1}, \cdots, r_{s_2,t-l-1}, \quad \cdots \quad r_{s_m,t}, r_{s_m,t-1}, \cdots, r_{s_m,t-l-1})$$

The PDF calculated from y_t defined this way can be used with Eq. 4.2, and will be able to produce the ERTE from multiple secondary time series towards one primary series. A key point to note here is that this formulation is highly flexible yet still computationally simple to implement. A differnt value of l may be chosen for each time series enabling different look back windows for each time series incorporated into the calculation.

The challenge presented by this formulation is that the data size requirement has grown from S^{k+l} to $S^{k+(l\times m)}$, which means a substantially larger data set is needed to calculate meaningful values of ERTE. Since m is dependant on the number of secondary series that need to be utilized, therefore the size of the look back windows k and l and discretization level S must be adjusted. Previously it was discussed that $l=1$ is a standard practice advocated in literature (Jizba et al 2012). Further it was empirically found that $k=1$ was able to produce the highest gains in performance over benchmarks. Therefore, in order to use the 9 secondary time series for one primary time series, using $k=l=1$, and using discretization $S=2$, the data size requirement is 1024 data points at minimum. This is equal to 3 days worth of 5-minute frequency observations.

In order to calculate the evolving transfer entropy from these multiple series and min-

imize spurious finite sample effects the rolling window over 7 days is used to calculate the combined ERTE.

6.4 Volatility Proxy based on OHLC

The Open-High-Low-Close (OHLC) values for each 5-minute interval are available in the dataset therefore an OHLC based proxy for volatility can be utilized. The returns data is found to have zero drift, therefore the Garman-Klass estimate for volatility is chosen. For open O_t, high H_t, low L_t, and close C_t the Garman-Klass volatility σ_{GK} is given as,

$$\sigma_{GK} = \sqrt{\frac{0.5}{n}\sum_{t=1}^{n}(u_t - d_t)^2 - \frac{0.39}{n}\sum_{t=1}^{n}c_t^2} \tag{6.2}$$

where $u_t = \ln\frac{H_t}{O_t}$, $c = \ln\frac{C_t}{O_t}$, and $d = \ln\frac{L_t}{O_t}$. For estimating the volatility σ_{GK} at each 5-min interval the value of n is fixed as 1 and the OHLC values from the data are used with Eq. 6.2.

6.5 Cryptocurrency Data

The data for 10 crypto currency pairs is used in the experiments presented in this chapter. The coins are a sample of the top 25 with respect to market cap and represent over 75% of the total market cap of cryptocurrency markets. The list of the coins along with their description is given in Table 6.1. The data covers the 6 month period from Nov 01, 2017 to May 01, 2018 and is sampled at 5-minute frequency. The Open-High-Low-Close (OHLC) data for each price observation are available. There are a total of approximately 430K price observations in the full dataset. This represents one of the largest cryptcurrency datasets to date being used in literature to study volatility in cryptocurrency markets.

The summary statistics of the log returns for all the cryptocurrencies are given in Table 6.2. It can be seen that the top coins show a marked positive skew, far more than is usually observed for stocks, market indices, or forex. This indicates that for investors there are

frequent small losses but also a chance of a few extreme gains.

Name (Symbol)	Description	Market Cap ($ Million)
Bitcoin (BTC)	The first decentralized ledger currency. Cryptocurrency with the highest market capitalization	165,947
Ethereum (ETH)	A platform for programmable contracts and money. Supports Turing-complete smart contracts.	79,955
Ripple (XRP)	An enterprise payment settlement network that is designed for peer to peer debt transfer. Not based on bitcoin	34,965
Litecoin (LTC)	A faster enbaler of the Bitcoin concept. The first cryptocurrency to use Scrypt as a hashing algorithm	9,894
Neo (NEO)	Chinese based cryptocurrency to enable the development of digital assets and smart contracts	5,511
Dash (DASH)	A bitcoin-based currency featuring instant transactions, decentralized governance and budgeting, and private transactions	4,030
Monero (XMR)	Privacy-centric coin using the CryptoNote protocol with improvements for scalability and decentralization	3,824
Ethereum Classic (ETC)	An alternative version of Ethereum whose blockchain does not include the DAO Hard-fork. Supports Turing-complete smart contracts.	2,464
OmiseGO (OMG)	A token for enabling an open payment platform and decentralised exchange issued on Ethereum. It provides banking, remittance, and exchange capabilities. First Ethereum project to exceed US$1 billion valuation	1,750
Zcash (ZEC)	The first open, permissionless financial system employing zero-knowledge security	1,165

Table 6.1: Data from 10 crypto currencies is analysed for the 6 months between Nov 01, 2017 and May 01, 2018. Market cap as on May 6th, 2018

Name	Count	Mean	Std	Min	25%	75%	Max	Skew	Kurtuosis
Bitcoin (BTC)	53,142	0.0000	0.0043	−0.0258	−0.0015	0.0014	0.1110	1.1367	24.3642
Ethereum (ETH)	52,611	0.0000	0.0046	−0.0609	−0.0016	0.0016	0.1226	1.0751	29.0850
Ripple (XRP)	51,782	0.0000	0.0069	−0.0918	−0.0024	0.0020	0.1337	1.4205	25.2119
Litecoin (LTC)	50,944	0.0000	0.0060	−0.1655	−0.0024	0.0021	0.1932	0.9705	45.2421
Neo (NEO)	50,739	0.0000	0.0071	−0.0641	−0.0026	0.0026	0.1650	−0.8565	15.4215
Dash (DASH)	39,422	0.0000	0.0074	−0.0878	−0.0026	0.0026	0.2082	0.8277	23.3832
Monero (XMR)	42,720	0.0000	0.0073	−0.0571	−0.0029	0.0029	0.1042	0.4230	7.6221
Ethereum Classic (ETC)	49,085	0.0000	0.0069	−0.073	−0.0029	0.0028	0.1269	0.5041	10.4402
OmiseGO (OMG)	46,592	0.0000	0.0082	−0.1269	−0.0029	0.0029	0.2102	0.7166	19.4655
Zcash (ZEC)	44,314	0.0000	0.0069	-0.0623	-0.0028	0.0027	0.0936	0.3023	6.5742

Table 6.2: Summary statistics for the returns of the 10 crypto currencies in the dataset, for the 6 months between Nov 01, 2017 and May 01, 2018

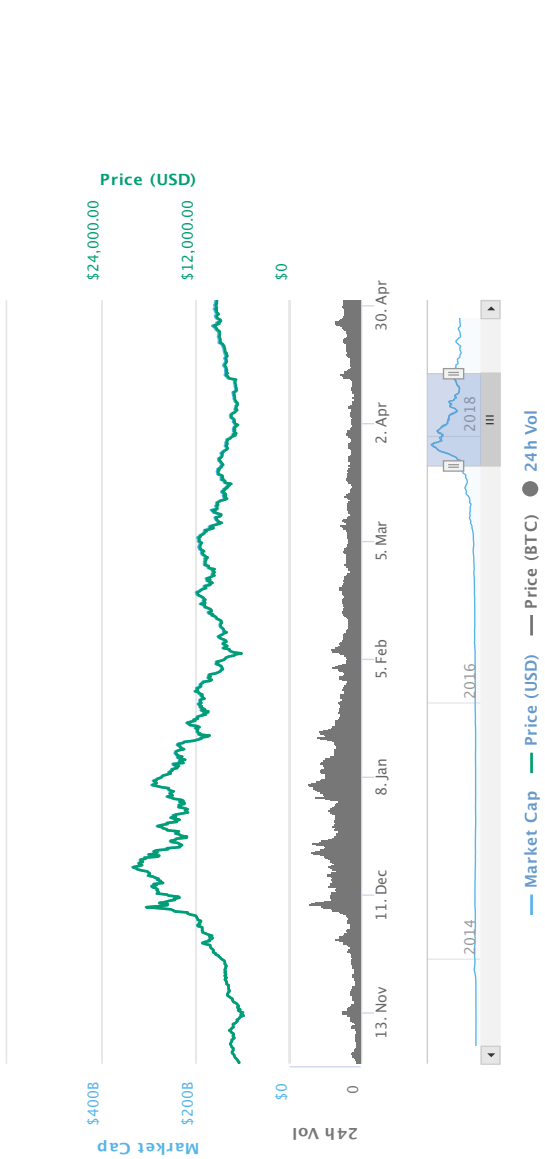
6.6 Methodology of Experiments

In order to estimate the volatility forecasts for a given primary series from the cryptocurrencies, a two step approach is used.

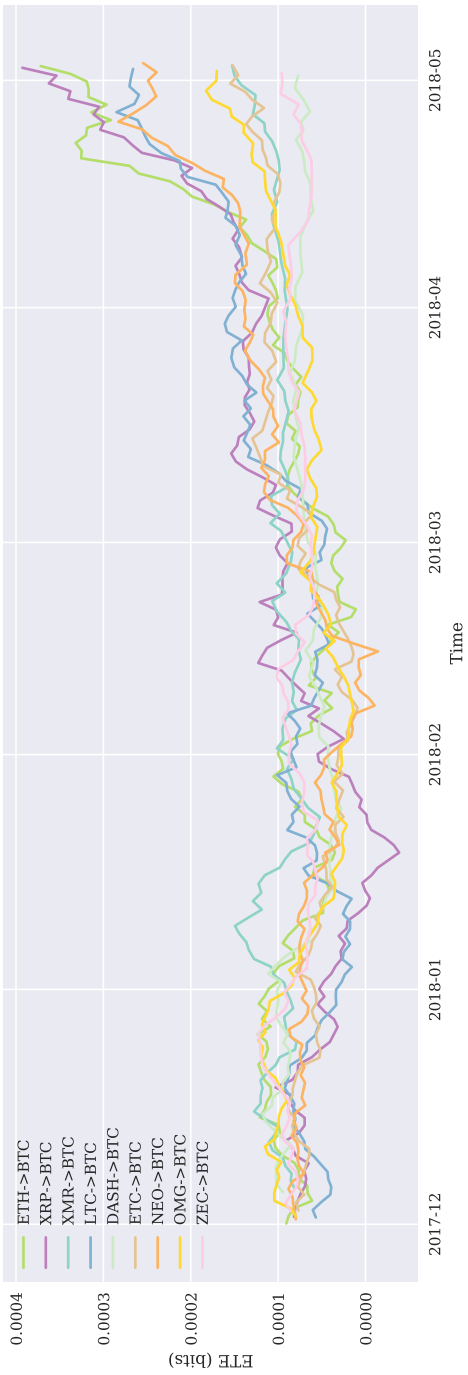
First, a forecast is generated using the envelope GP method developed in Chapter 3. A rolling training window of 1 trading day or 288 data points is provided to the GP for producing the next time step volatility forecast using the envelope method. Rolling window forecasts are generated step by step for one complete trading day and the hyperparameters are calculated at each step. The loss functions are calculated for the forecast series against the Garman-Klass volatility proxy using Eq. 6.2. The loss function performance of the GP method is compared to the three benchmarks - mcsGARCH, rolling window, and using the last observed value as next step forecast.

Second, the ERTE is generated from the 9 secondary series to the primary series using the method from Sec. 6.3. This ERTE is used to modify the forecasts from the GP method generated in the last step. The loss function values are calculated as before and the performance is benchmarked. The 9 ERTE series from each cryptocurrency to BTC, and ETH is shown in Figs. 6.1 and 6.3 for various values of q, along with the price and market cap for the cryptocurrency. Similar charts for all the other currencies showing the pairwise ERTE for the whole data set are shown in Figs. 6.7 - 6.21 at the end of the chapter.

The testing period is started from Jan 01, 2018 since mcsGARCH requires 2 months worth of data for calculating the diurnal volatility. This leaves 125 days for testing performance. Using the rolling window approach forecasts for all 125 trading days are carried out. There are 288 trading points in each trading day due to the 5 minute frequency therefore 36K forecasts are generated per cyrptocurrency or 360K overall. The large number of experimental results provides significant confidence in the findings.

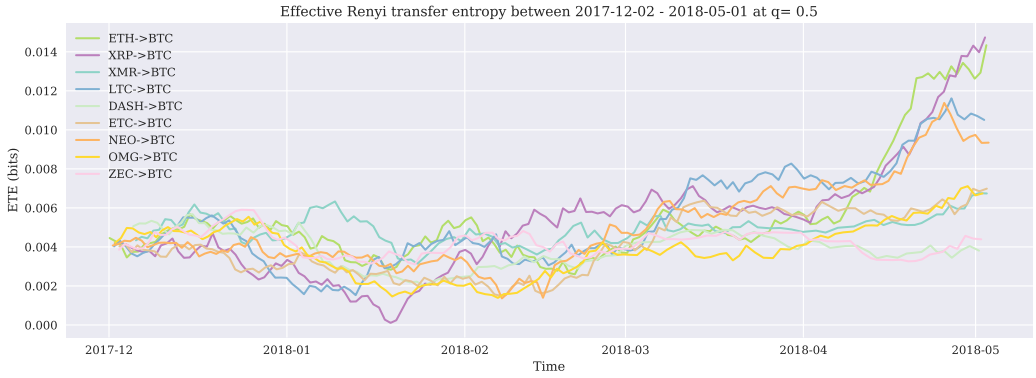


(a) Price and market cap

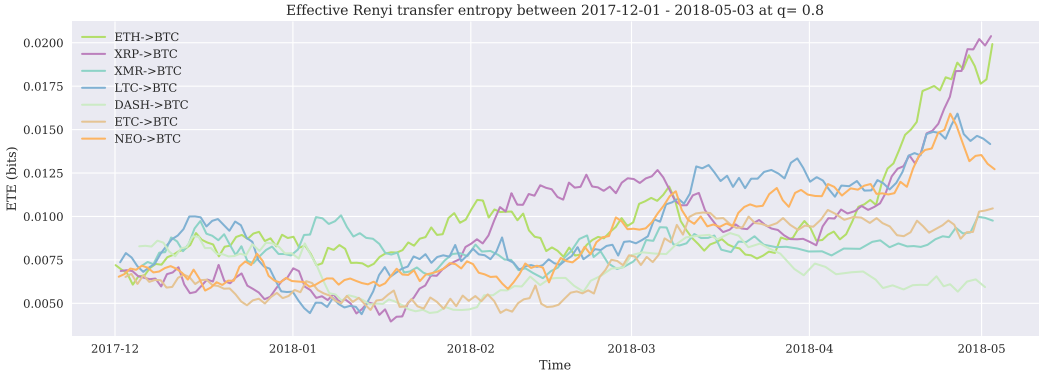


(b) $q = 0.01$

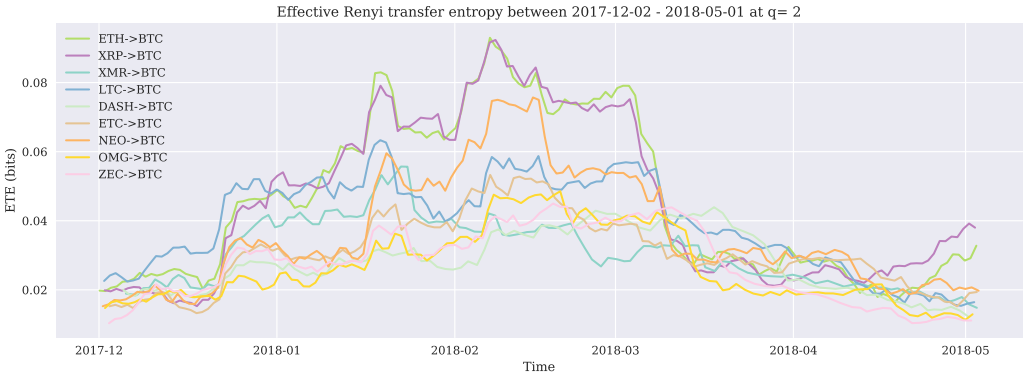
Figure 6.1: Effective Rényi transfer entropy from other cryptocurrencies towards BTC for various values of q



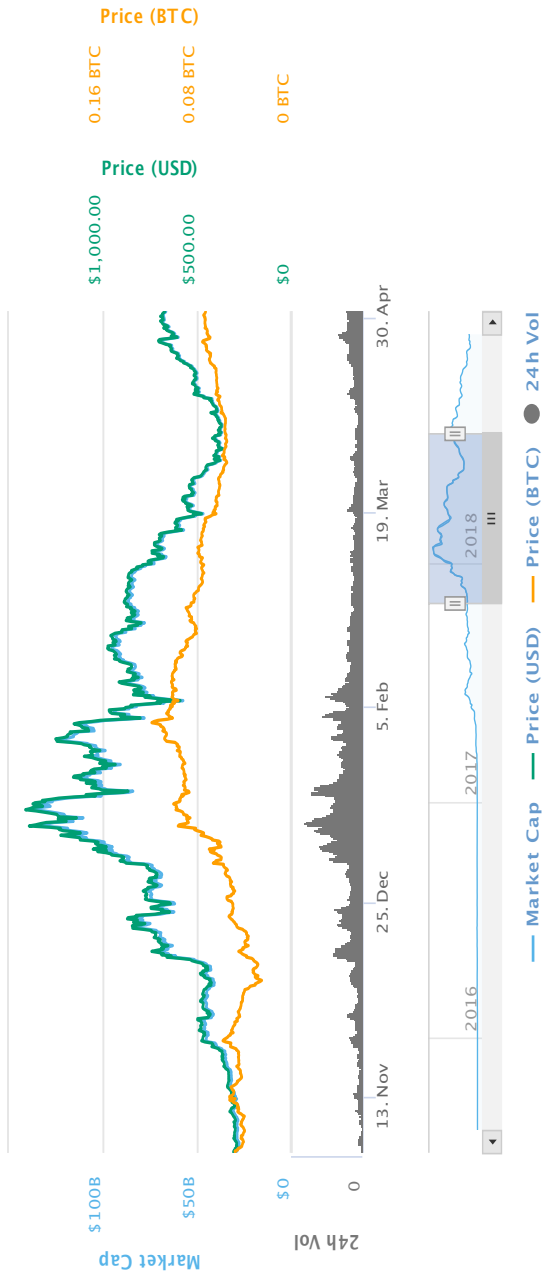
(c) $q = 0.5$



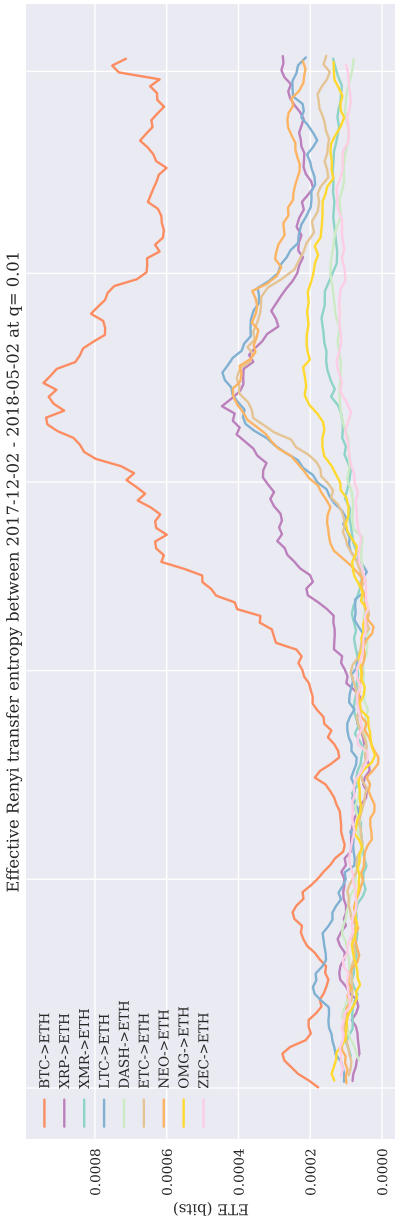
(d) $q = 0.8$



(e) $q = 2.0$

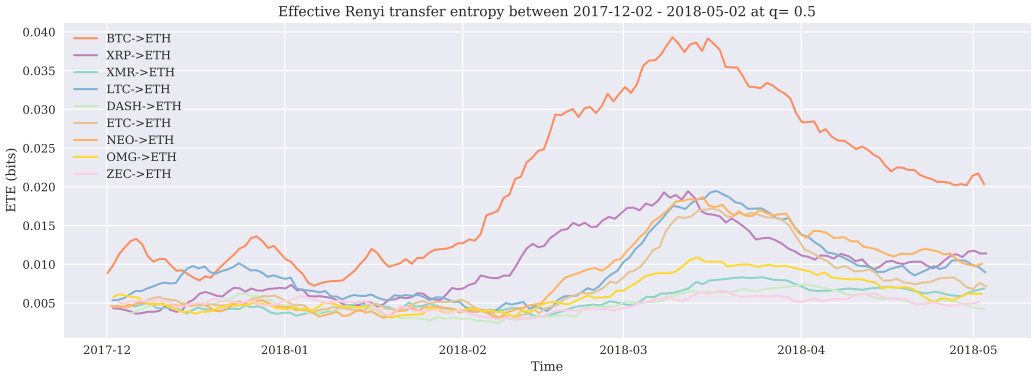


(a) Price and market cap

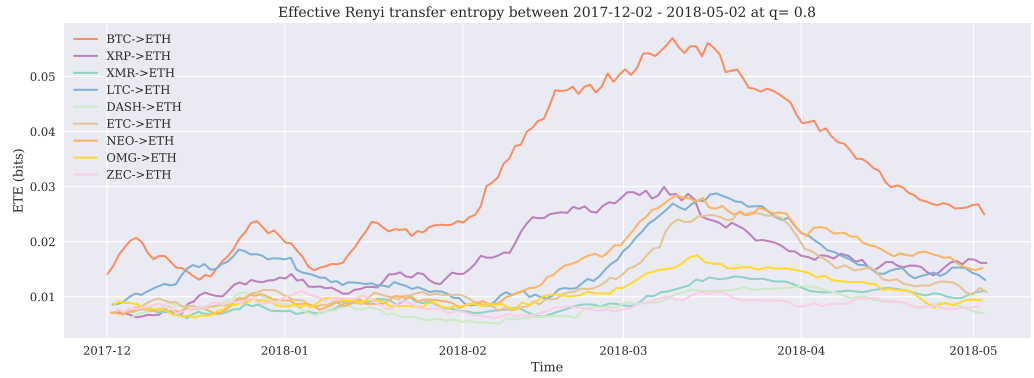


(b) $q = 0.01$

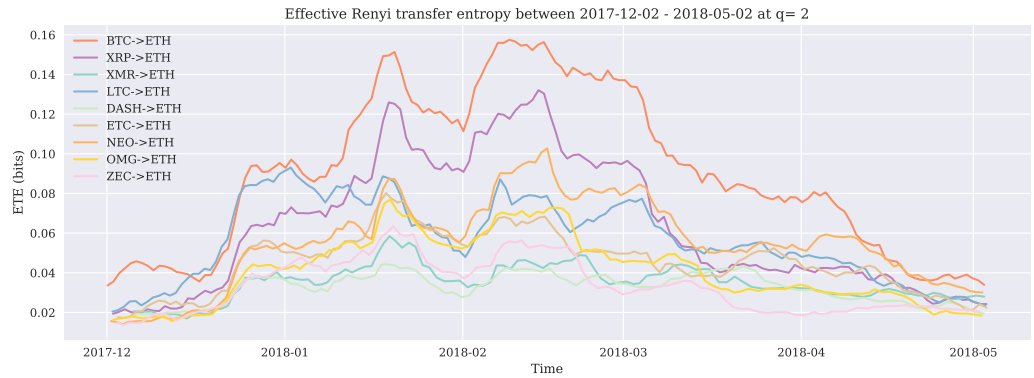
Figure 6.3: Effective Rényi transfer entropy from other cryptocurrencies towards ETH for various values of q



(d) $q = 0.5$



(e) $q = 0.8$



(f) $q = 2.0$

6.7 Discussion of Results

Analysis of the experiments reveals that that envelope GPs consistently perform better than all other benchmarks across the 10 cryptocurrencies. They are found to perform 4.3% better than mcsGARCH, 6.9% better than rolling average, and 14.5% better than using the last observed value, on average over the 10 cryptocurrencies for MCS loss function. Similar results are found for RMSE and QLIKE, across all three benchmarks. The results of the experiments are presented in Table 6.3. The results are found to be highly significant as can be seen in the notched box plot presented in Fig. 6.5.

It can be seen that the highest gains in performance for MSE loss function are found for Ethereum (ETH), Ethereum Classic (ETC), LiteCoin (LTC), and Ripple (XRP). The lowest positive gain is realised for Neo (NEO).

Next the results for the ERTE modified GP forecasts are analysed. Here it is also found that using the ERTE modified forecasts provides significant gains in performance over the three benchmarks, and a 1% average improvemnt over the standalone GP benchmark. Although in some instances it performs worse than the standalone envelope GPs on average for Ripple (XRP), DASH, and Ethereum Classic (ETC). A potential explanation for this is discussed in the next section. The results for the ERTE modified GP forecasts are presented in Table 6.4. The significance of the results is presented in the notched box plots in Fig. 6.6. It can be seen that the results are highly significant as the notches are all well away from zero for the 95% confidence level.

	Improv. over mcsGARCH	Improv. over rolling average	Improv. over previous value
MSE			
BTC	2.72%	4.79%	21.45%
ETH	8.08%	8.93%	15.71%
XRP	5.38%	6.65%	14.65%
LTC	6.76%	9.34%	10.28%
NEO	1.13%	5.77%	13.74%
DASH	2.48%	10.13%	15.96%
XMR	1.06%	3.30%	13.52%
ETC	9.98%	8.72%	16.30%
OMG	2.46%	5.93%	8.78%
ZEC	3.33%	6.41%	15.03%
RMSE			
BTC	1.82%	2.98%	11.72%
ETH	2.56%	3.41%	8.41%
XRP	4.49%	3.37%	7.03%
LTC	3.36%	3.25%	8.08%
NEO	3.74%	2.81%	5.53%
DASH	4.97%	3.64%	8.72%
XMR	2.18%	3.59%	6.41%
ETC	2.87%	3.44%	7.39%
OMG	2.43%	2.67%	4.42%
ZEC	1.99%	2.09%	8.00%
QLIKE			
BTC	1.82%	3.56%	7.03%
ETH	2.76%	2.48%	9.09%
XRP	3.92%	3.38%	4.32%
LTC	5.63%	2.72%	6.77%
NEO	2.42%	4.16%	4.68%
DASH	0.97%	3.17%	7.63%
XMR	0.73%	2.09%	5.65%
ETC	1.86%	2.80%	8.48%
OMG	1.00%	2.72%	6.44%
ZEC	4.14%	3.61%	9.85%

Table 6.3: Averaged values for the percentage improvement in MSE, RMSE, and QLIKE of using the the GP envelope method over the benchmarks

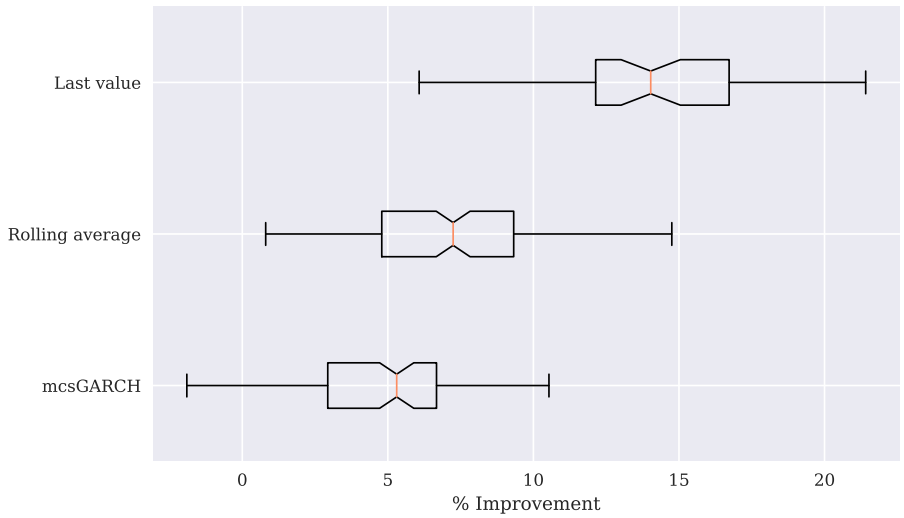


Figure 6.5: Improvement in MSE using envelope GP method, compared to mcsGARCH, rolling average, and last observed value benchmarks. The results are highly significant since the notches around all the median are positive and well beyond zero

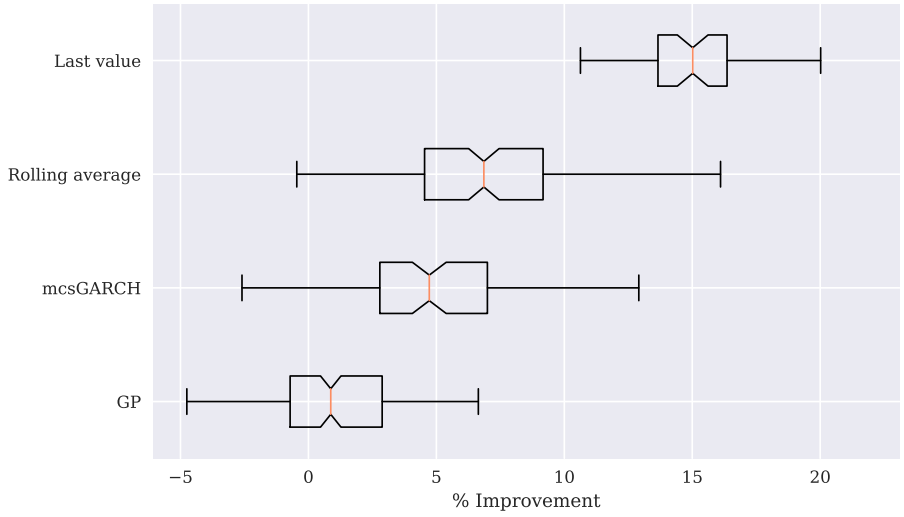


Figure 6.6: Improvement in MSE using ERTE modified envelope GP method, compared to stand alone envelope GP, mcsGARCH, rolling average, and last observed value benchmarks. The results are highly significant since the notches around all the median are positive and beyond zero

	Improv. over envelope GP	Improv. over mcsGARCH	Improv. over rolling average	Improv. over previous value
MSE				
BTC	2.71%	7.73%	6.49%	12.48%
ETH	0.52%	2.93%	4.32%	18.82%
XRP	−0.48%	3.54%	6.03%	16.14%
LTC	1.51%	7.01%	8.34%	17.11%
NEO	2.78%	4.54%	5.21%	15.94%
DASH	−0.83%	4.88%	9.78%	13.23%
XMR	1.90%	5.25%	6.98%	16.03%
ETC	−1.18%	1.39%	6.71%	20.18%
OMG	2.47%	7.84%	4.44%	14.54%
ZEC	1.33%	6.86%	10.05%	11.28%

Table 6.4: Averaged values for the percentage improvement in MSE by using the GP envelope method over the benchmarks

6.8 Caveat on Forecasting Performance

In the previous section an improvement in performance is found when ERTE is combined with envelope GPs. Although the significance of the results is established from the notched box plots, it can be seen in the table that in some instances the gains can be negative i.e. performance gets worse by including the ERTE in those instances. This can also be seen in Fig. 6.6 that although the notches of the median are away from zero, some mass of the box is below zero. The key reason for this is the discretization parameter S that was fixed to 2 in the experiments.

It was seen in Sec. 5.6 that the performance improves as S is increased, but this increase in S requires and substantial increase in the available data. In the case of the cryptocurrency data if S was set to 3, then $S^{k+(l\times m)}$ with $k = l = 1$ and $m = 9$ will evaluate to 59,000, meaning more than 6 months worth of data is needed to evaluate a single value of ERTE from the 9 series. This is more than the data available in the whole dataset, therefore S has been limited to 2 but if data at a higher frequency was available, say 1-minute, then only data from approx. 40 days will be needed to calculate one ERTE value. This is still substantial so another method might be to explore the ERTE with fewer time series, since at 1 minute frequency level up to 5 secondary time series can be accommodated for calclating the ERTE by using only a single trading day worth of data when $S = 3$. This points to the need for higher frequency data for cryptocurrencies to be available in a standardized form so more nuanced techniques can be applied with ease.

6.9 Conclusion

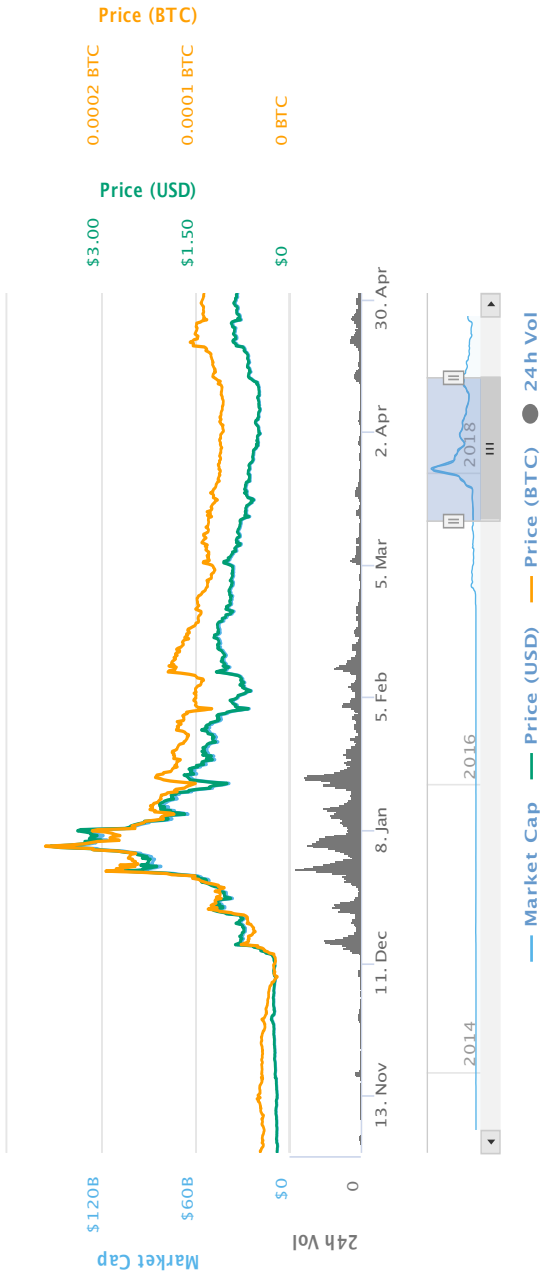
This chapter presented a significant set of experiments carried out on one of the largest datasets in literature, using a set of 10 cryptocurrencies sampled at 5-minute frequency over 6 months.

In the analysis of experiments it was found that for cryptocurrencies envelope GPs outperform the other benchmarks in volatility forecasting. A novel method of evaluating

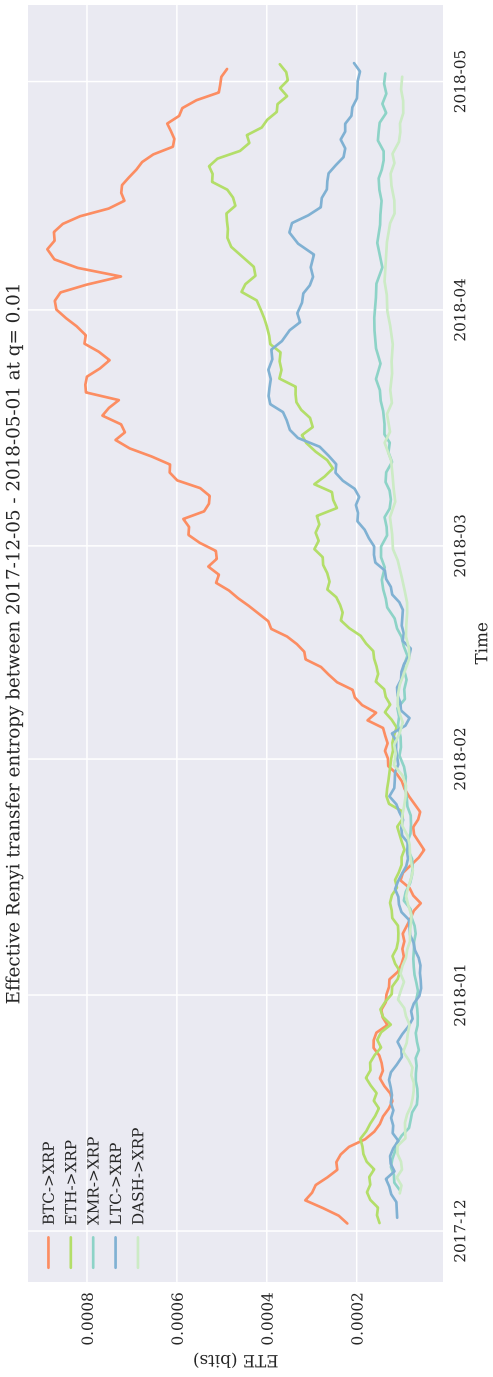
the effective Rényi transfer entropy from multiple time series is also described in this chapter. This ERTE was used for modifying the forecasts of envelope GPs and it is found to improve the performance over compared benchmarks.

This chapter also discussed some of the caveats in interpreting forecasting performance for ERTE based envelope GPs for the cases where the discretization level has to be kept low due to data size limitations.

Figures for the pairwise entropy transfer between all cryptocurrencies are presented in the following pages.

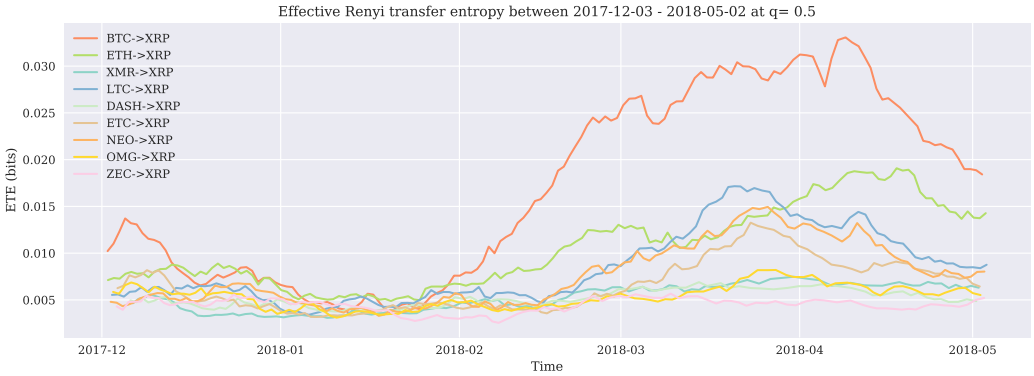


(a) Price and market cap

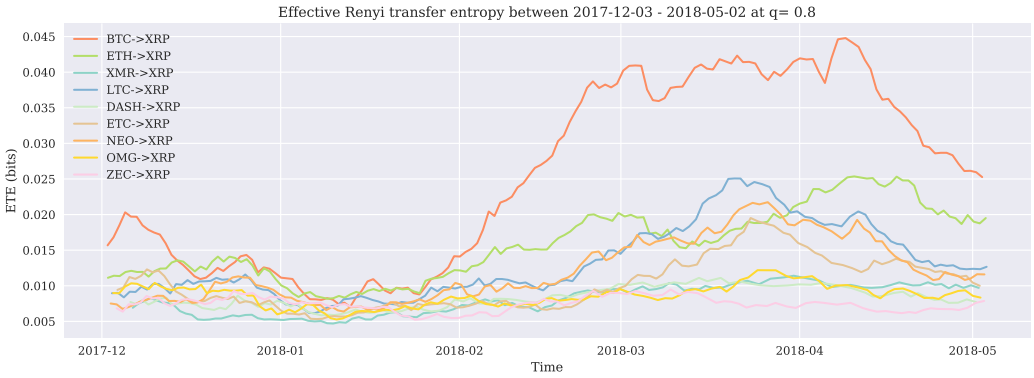


(b) $q = 0.01$

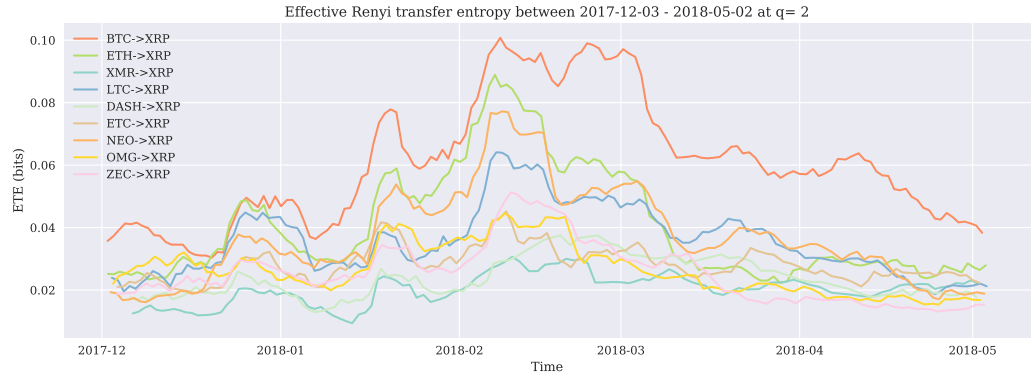
Figure 6.7: Effective Rényi transfer entropy from other cryptocurrencies towards XRP for various values of q



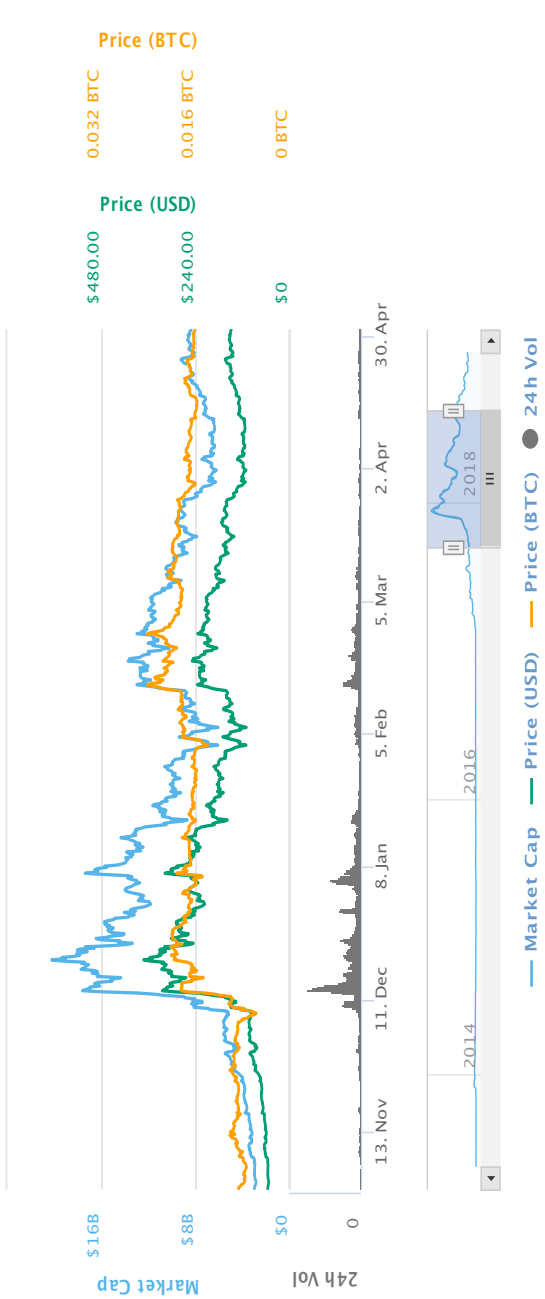
(d) $q = 0.5$



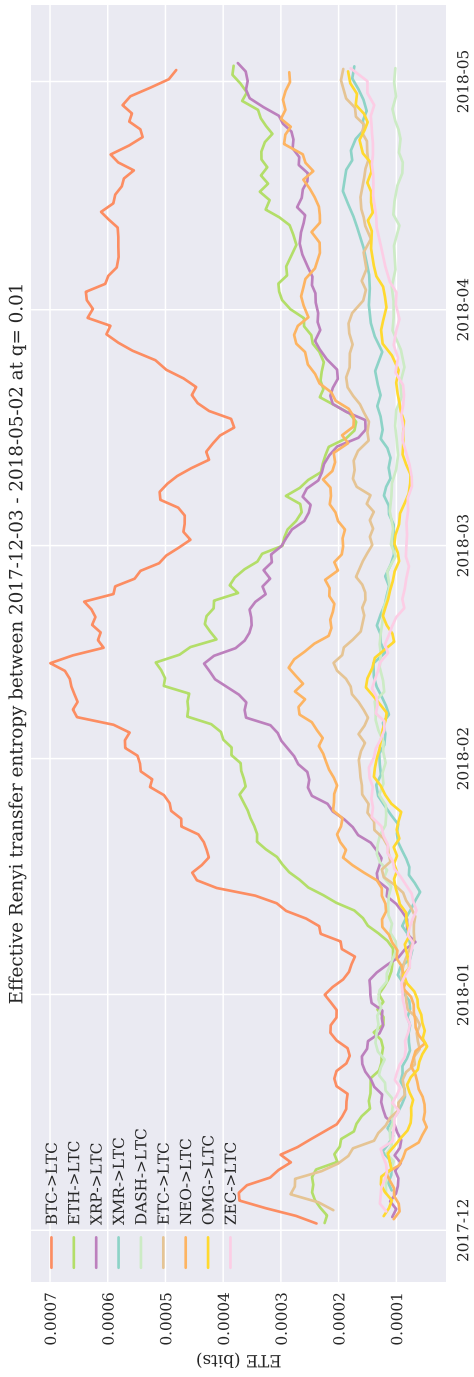
(e) $q = 0.8$



(f) $q = 2.0$

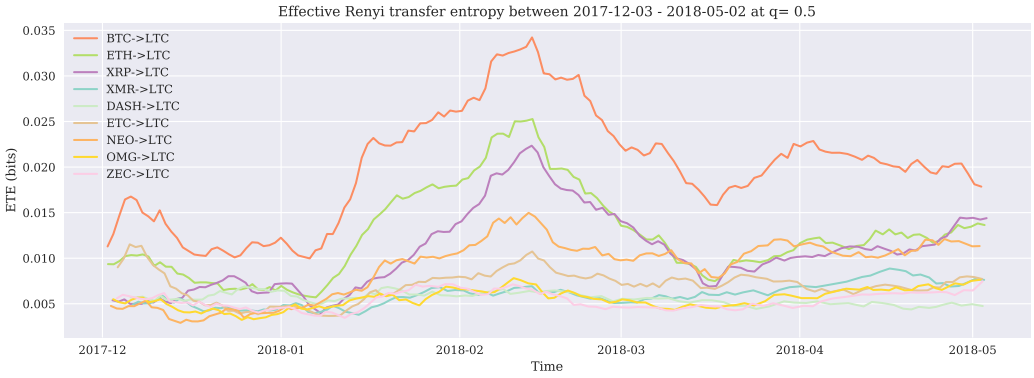


(a) Price and market cap

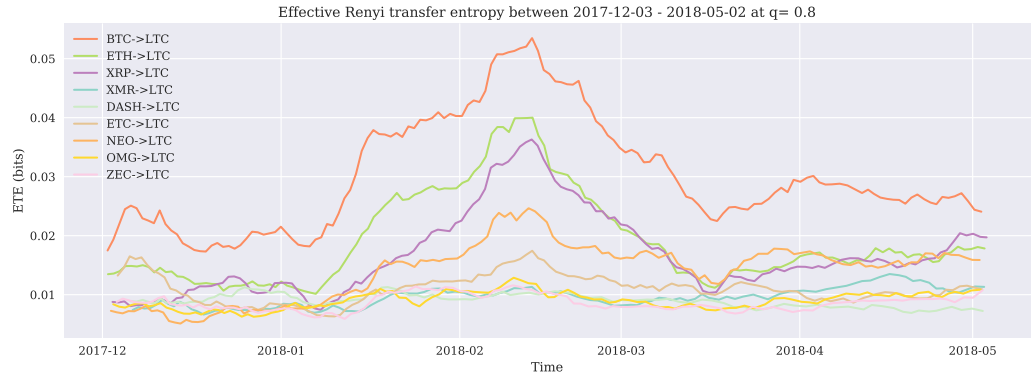


(b) $q = 0.01$

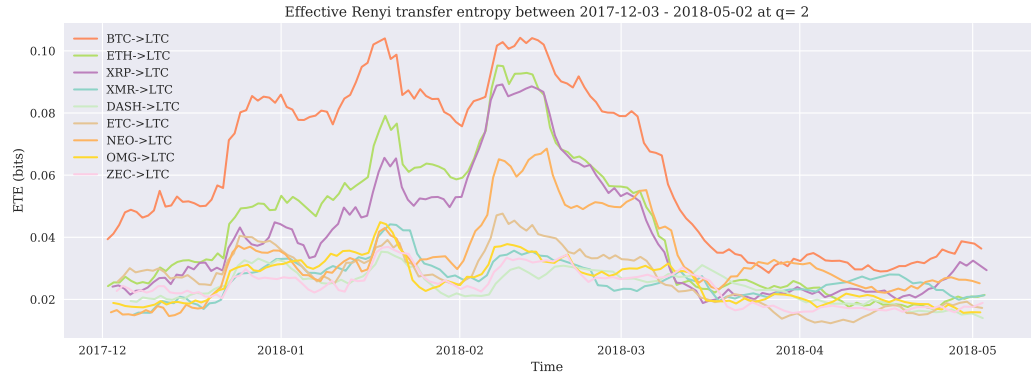
Figure 6.9: Effective Rényi transfer entropy from other cryptocurrencies towards LTC for various values of q



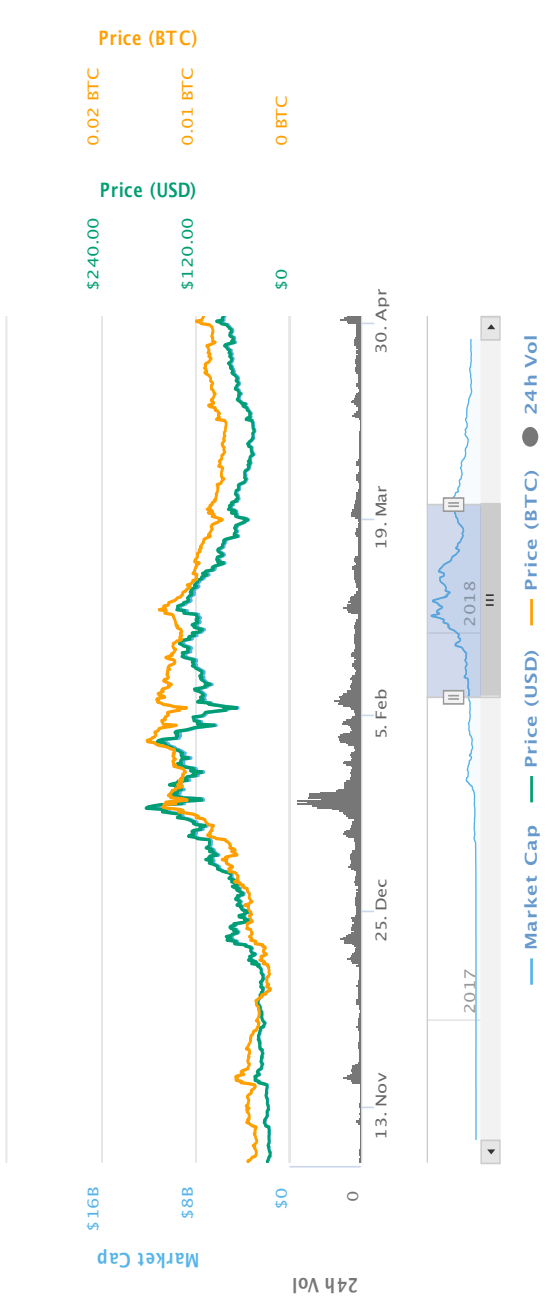
(d) $q = 0.5$



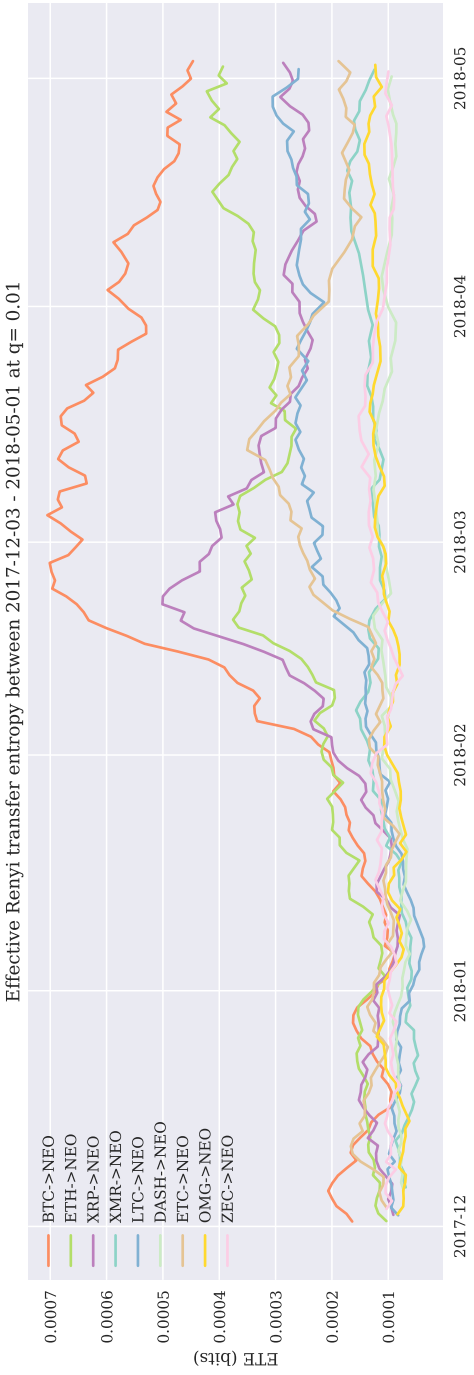
(e) $q = 0.8$



(f) $q = 2.0$

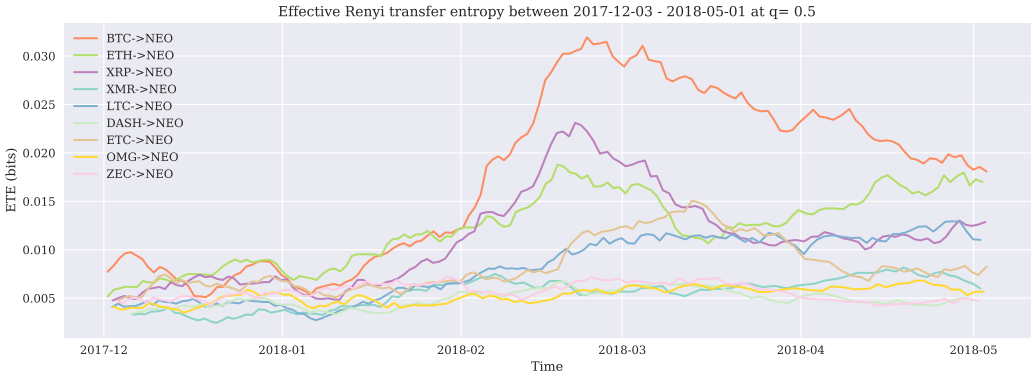


(a) Price and market cap

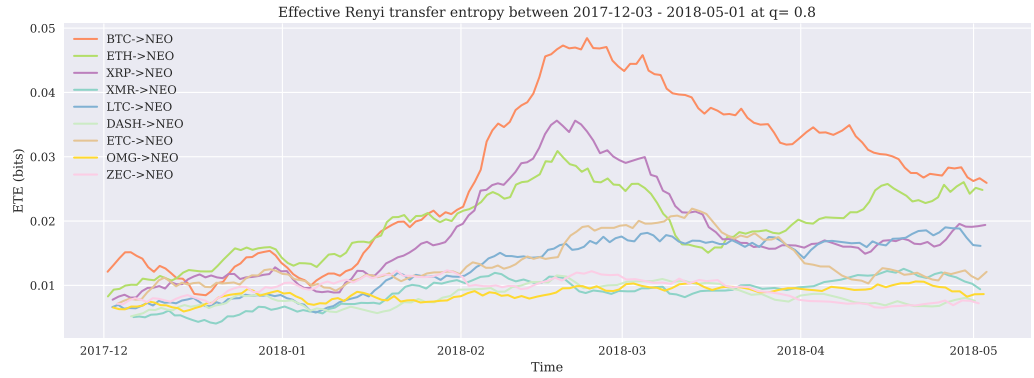


(b) $q = 0.01$

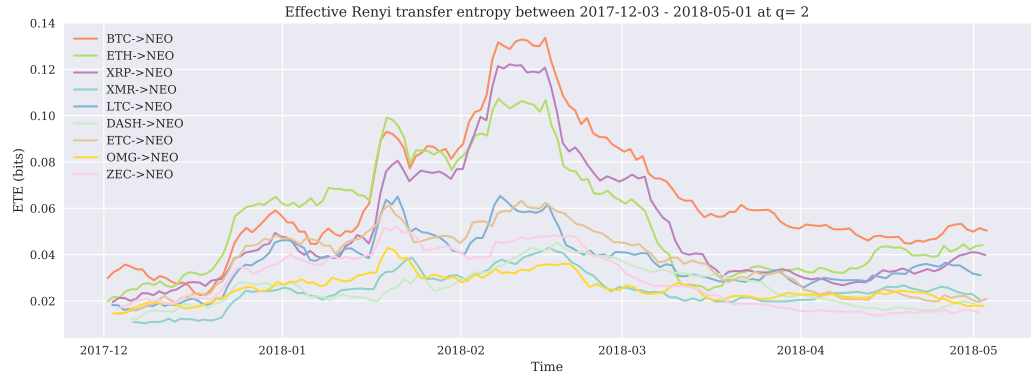
Figure 6.11: Effective Rényi transfer entropy from other cryptocurrencies towards NEO for various values of q



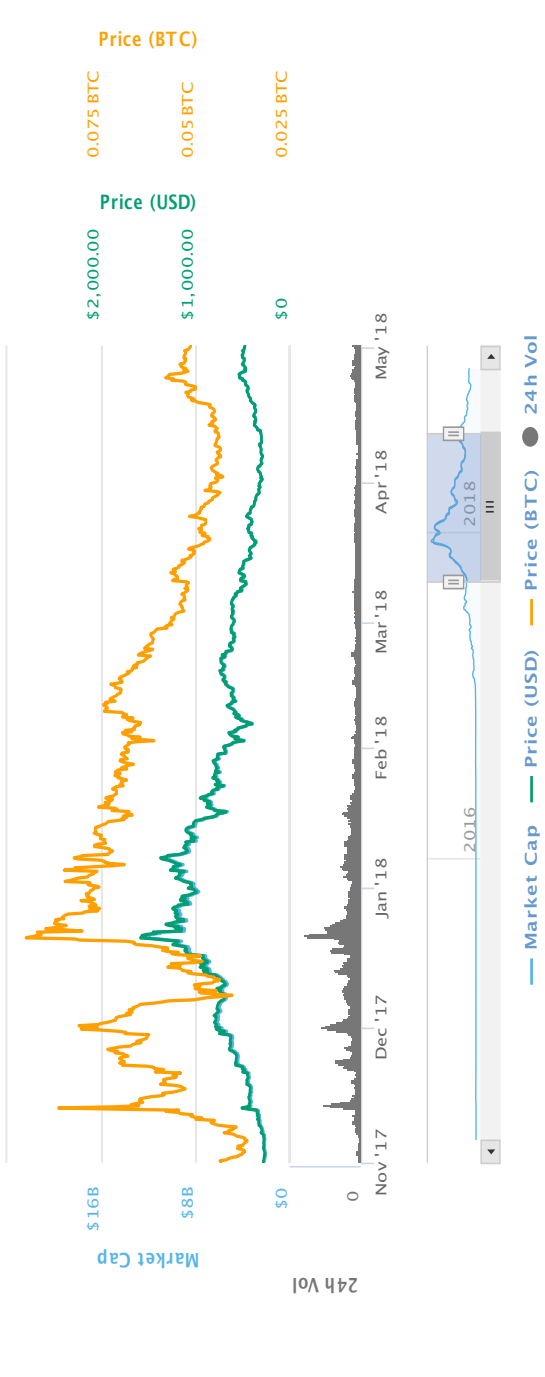
(d) $q = 0.5$



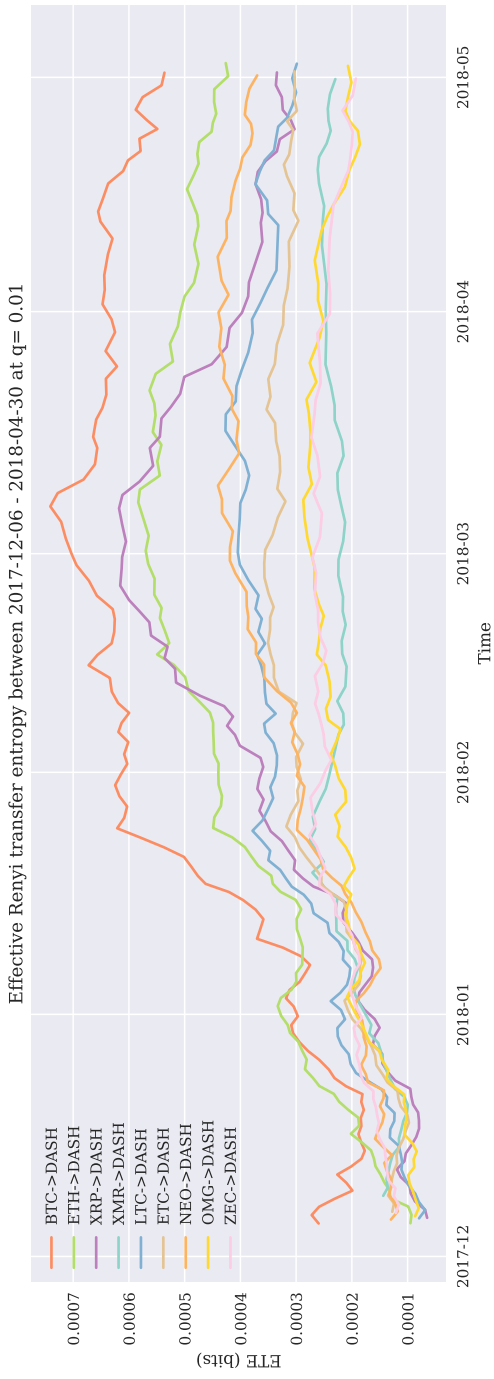
(e) $q = 0.8$



(f) $q = 2.0$

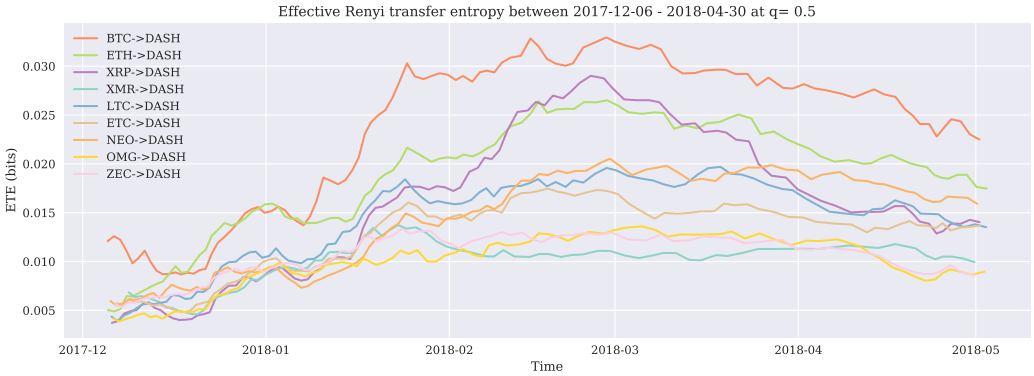


(a) Price and market cap

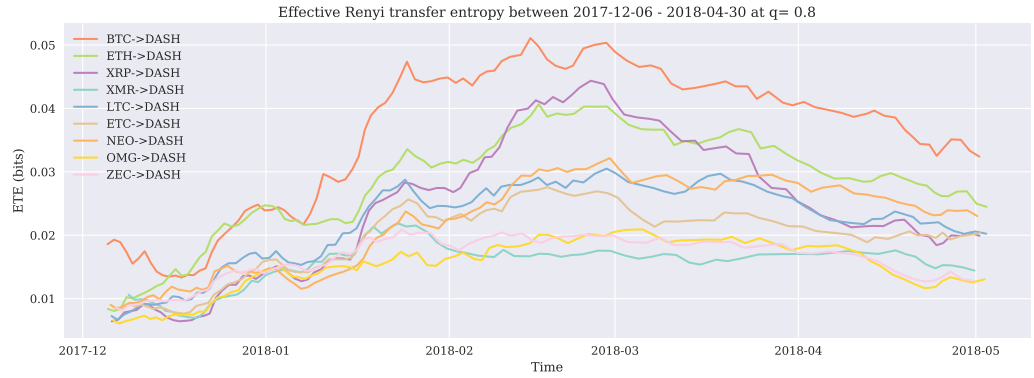


(b) $q = 0.01$

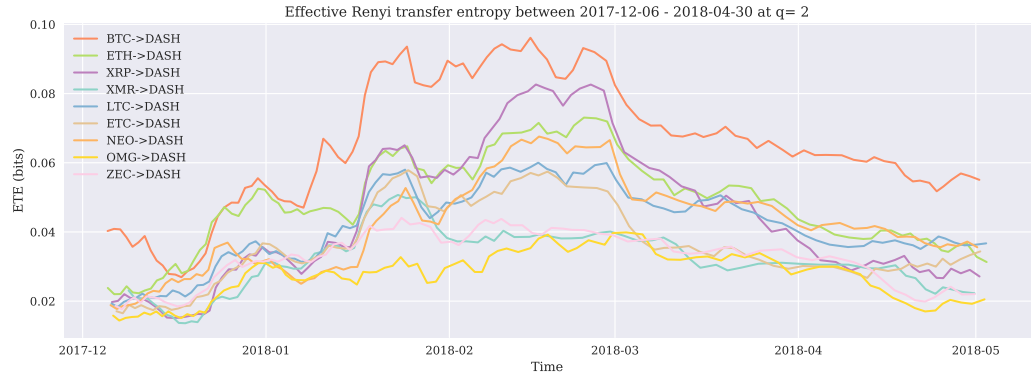
Figure 6.13: Effective Rényi transfer entropy from other cryptocurrencies towards DASH for various values of q



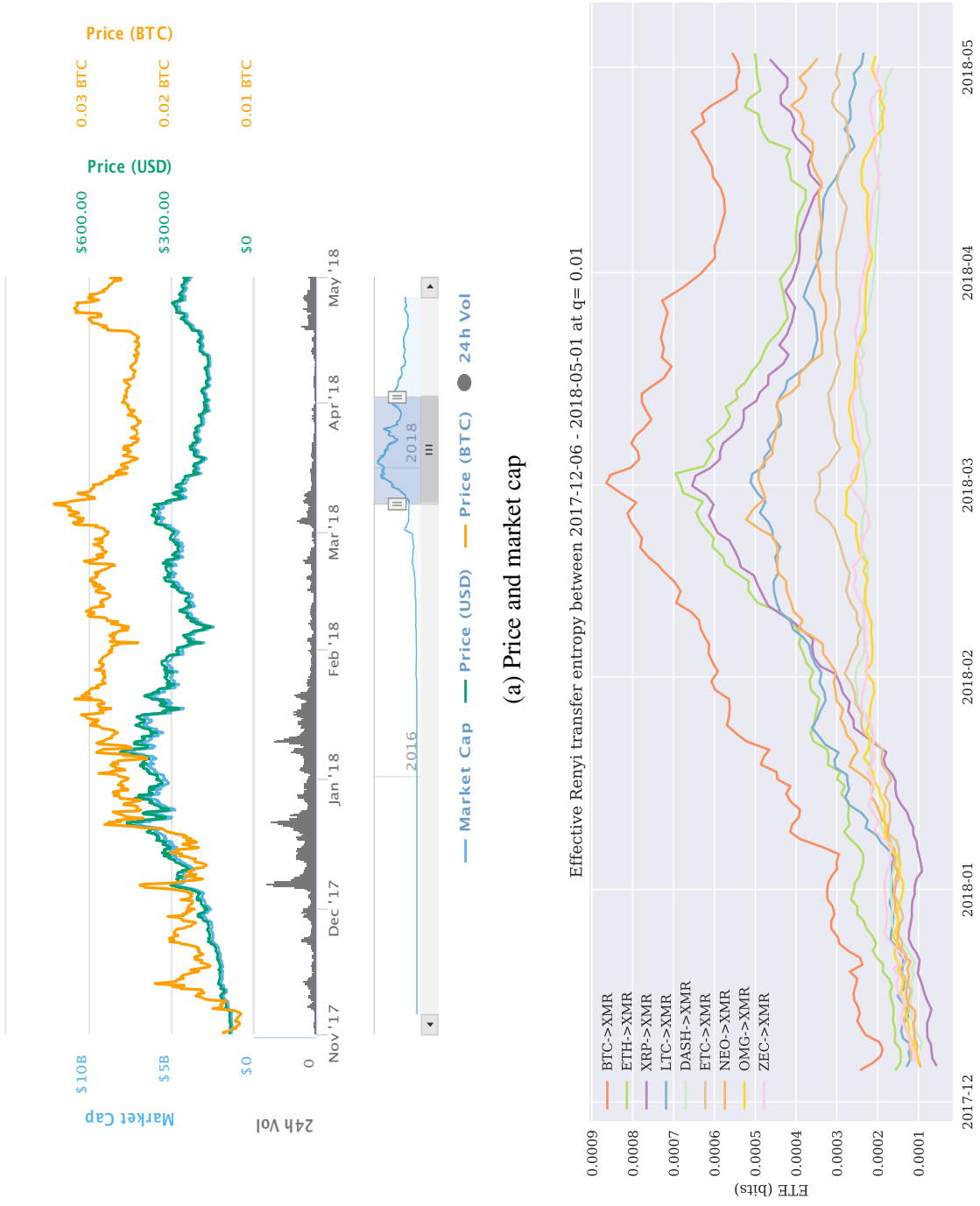
(d) $q = 0.5$



(e) $q = 0.8$



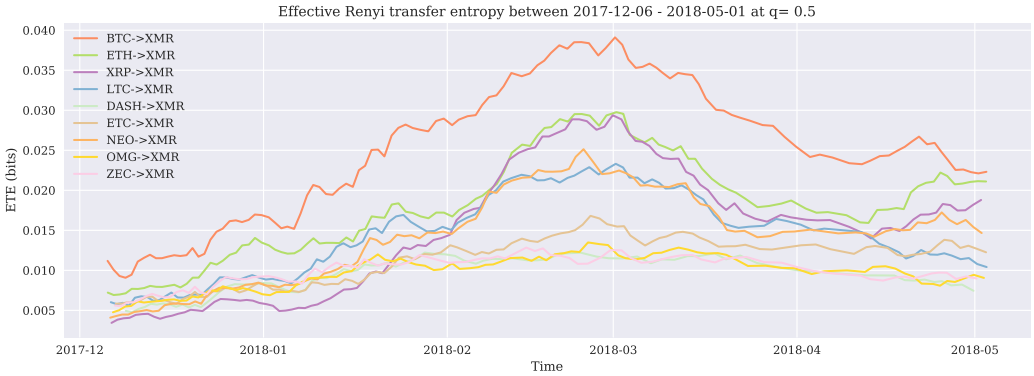
(f) $q = 2.0$



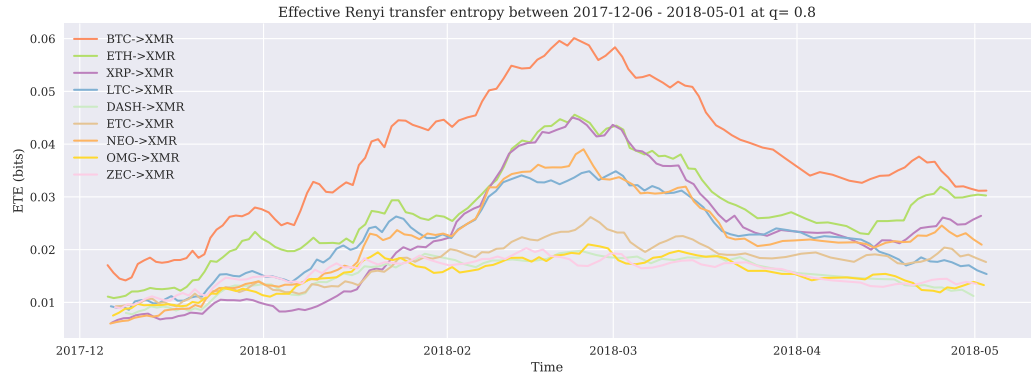
(a) Price and market cap

(b) $q = 0.01$

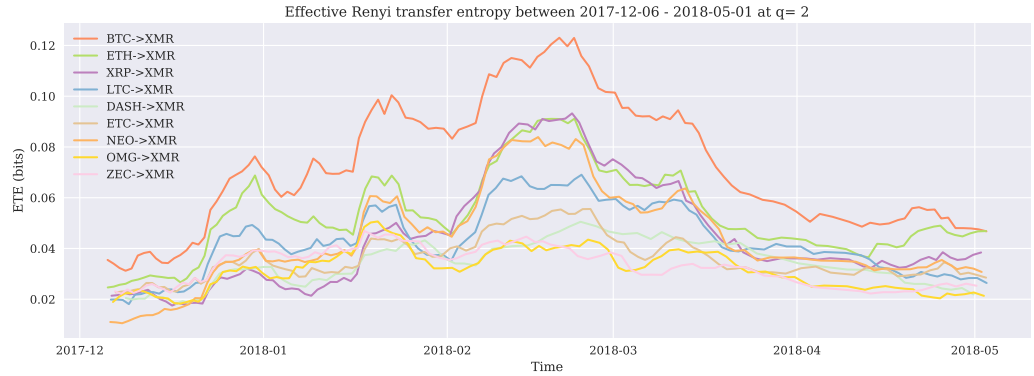
Figure 6.15: Effective Rényi transfer entropy from other cryptocurrencies towards XMR for various values of q



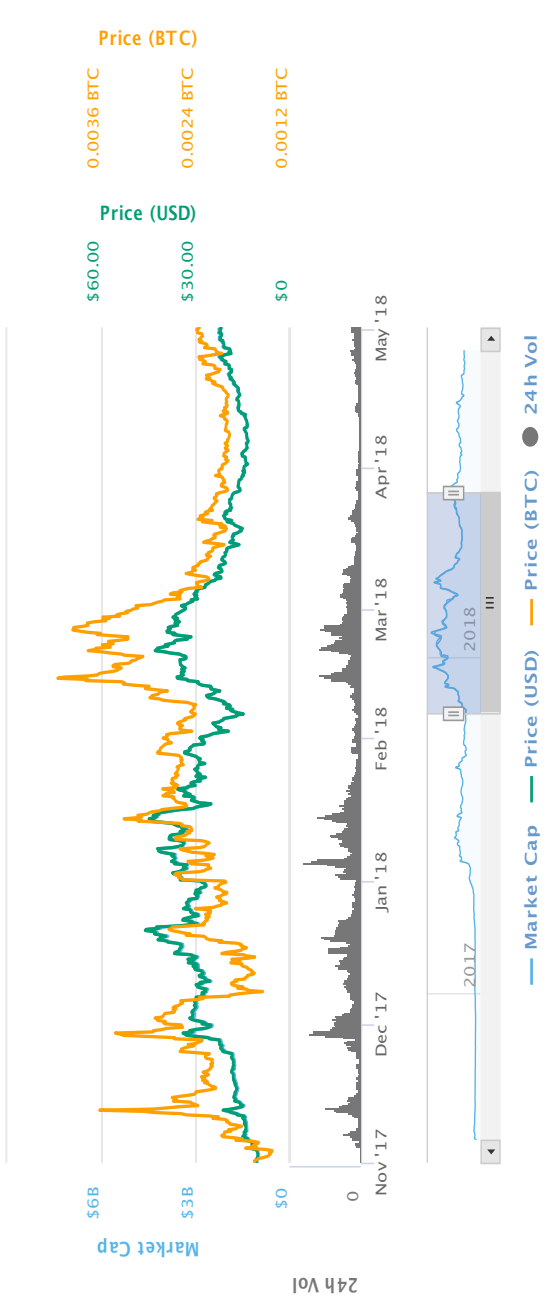
(d) $q = 0.5$



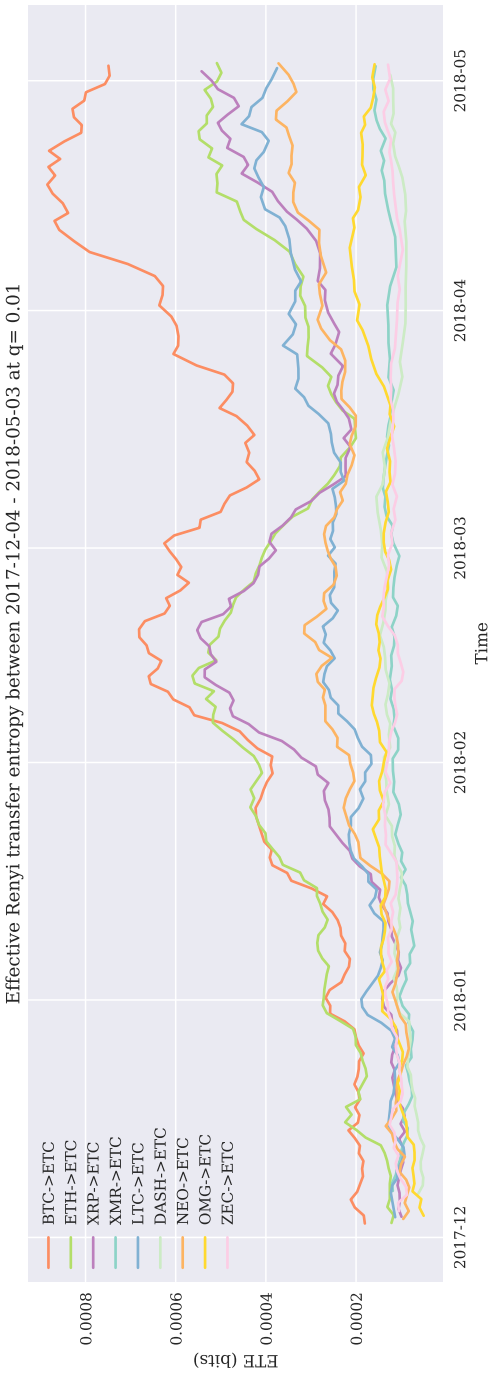
(e) $q = 0.8$



(f) $q = 2.0$

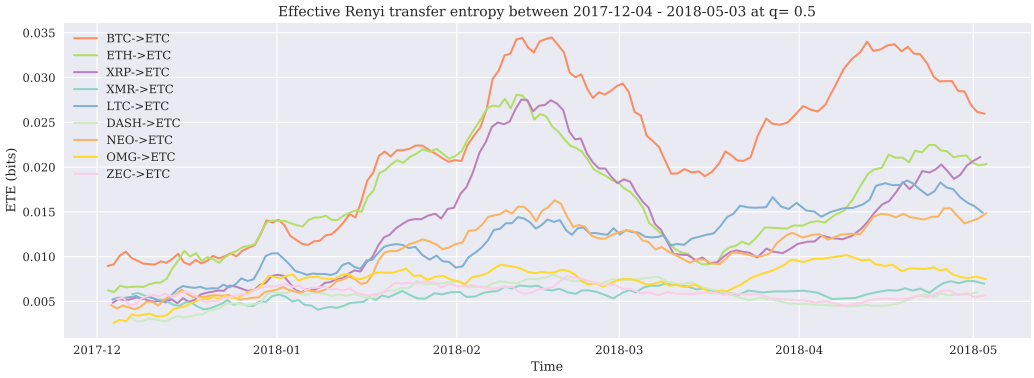


(a) Price and market cap

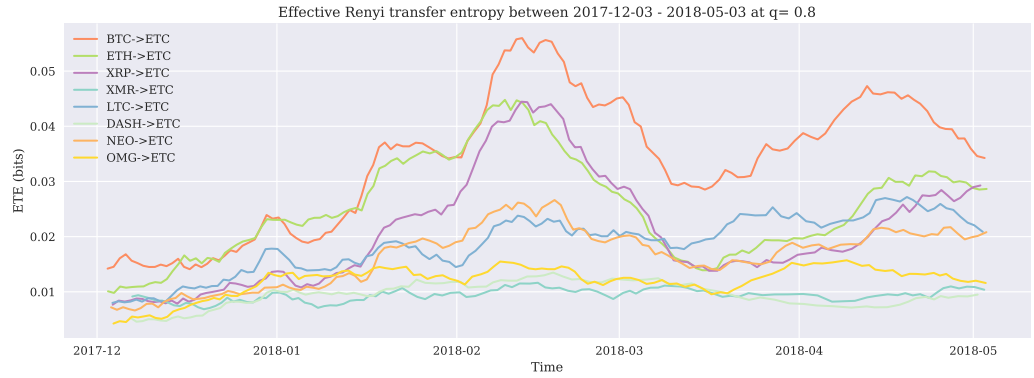


(b) $q = 0.01$

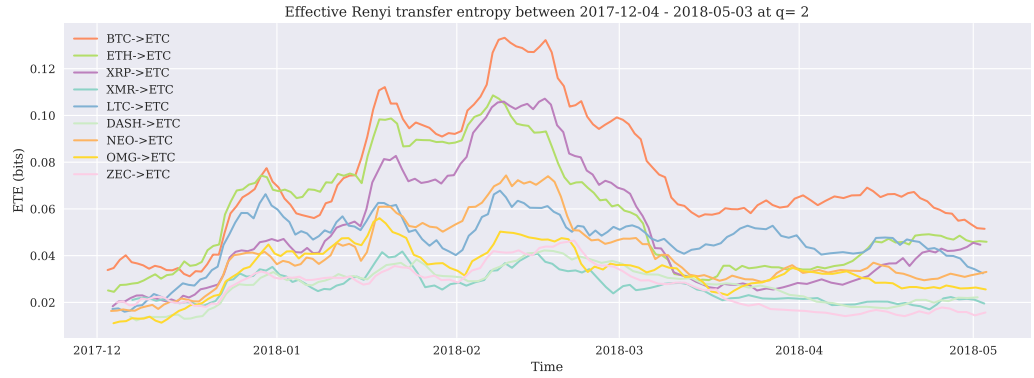
Figure 6.17: Effective Rényi transfer entropy from other cryptocurrencies towards ETC for various values of q



(d) $q = 0.5$



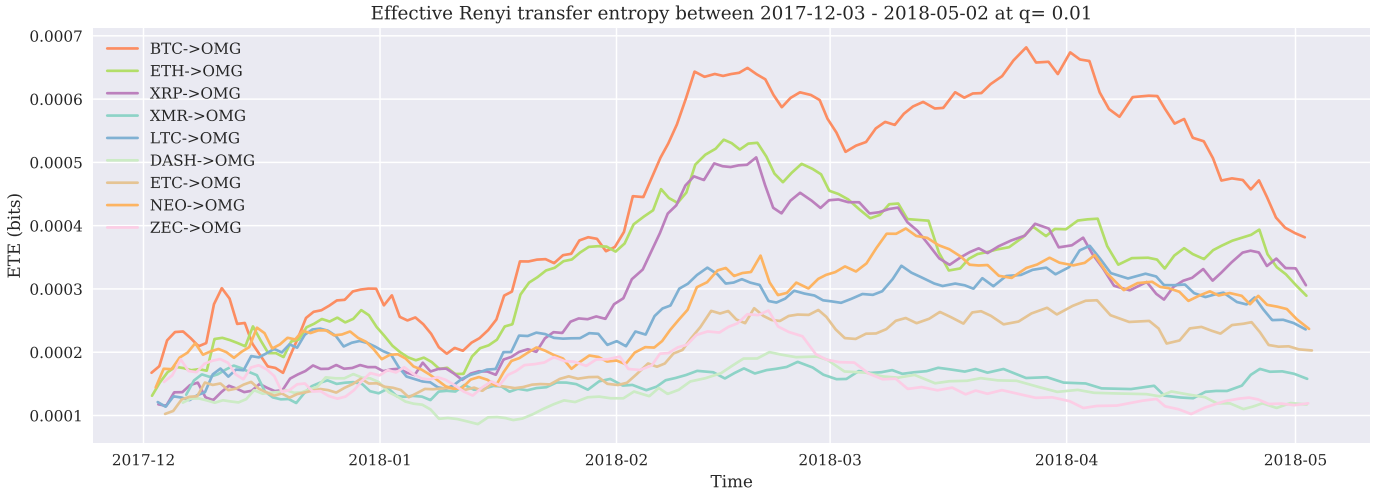
(e) $q = 0.8$



(f) $q = 2.0$

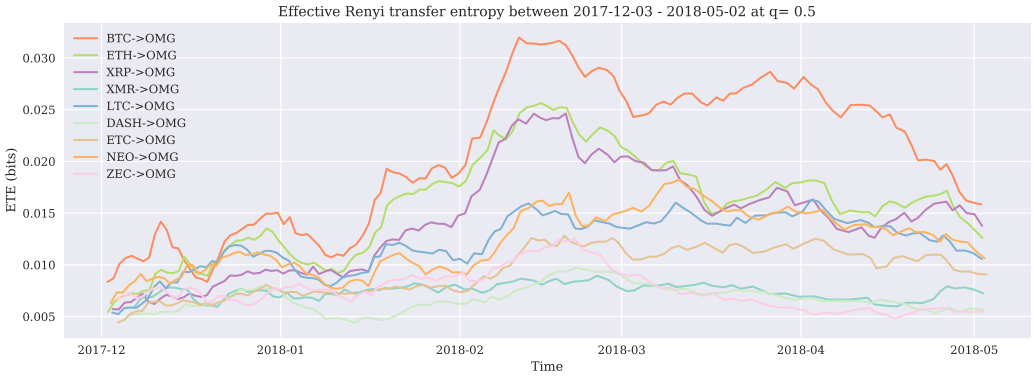


(a) Price and market cap

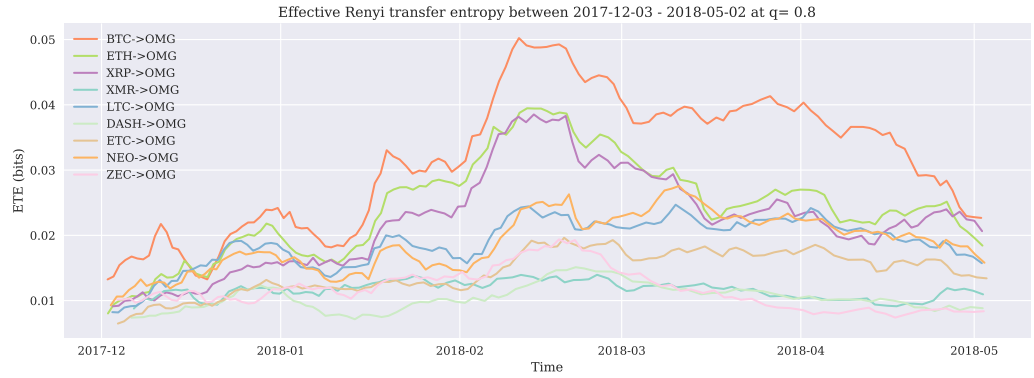


(b) $q = 0.01$

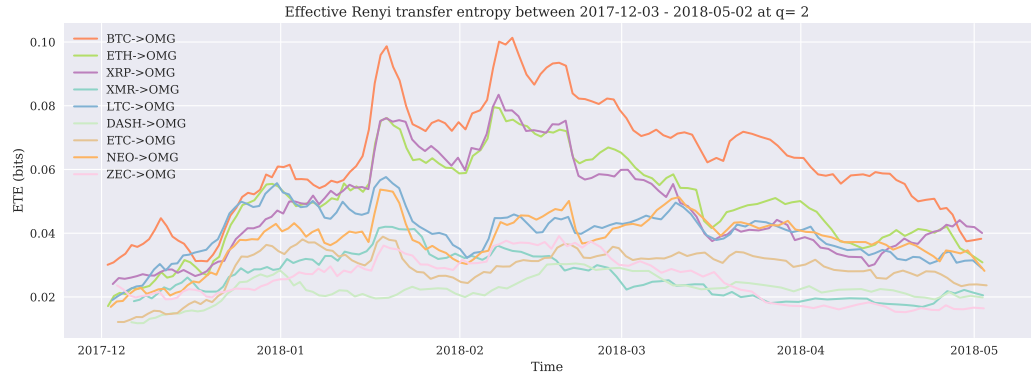
Figure 6.19: Effective Rényi transfer entropy from other cryptocurrencies towards OMG for various values of q



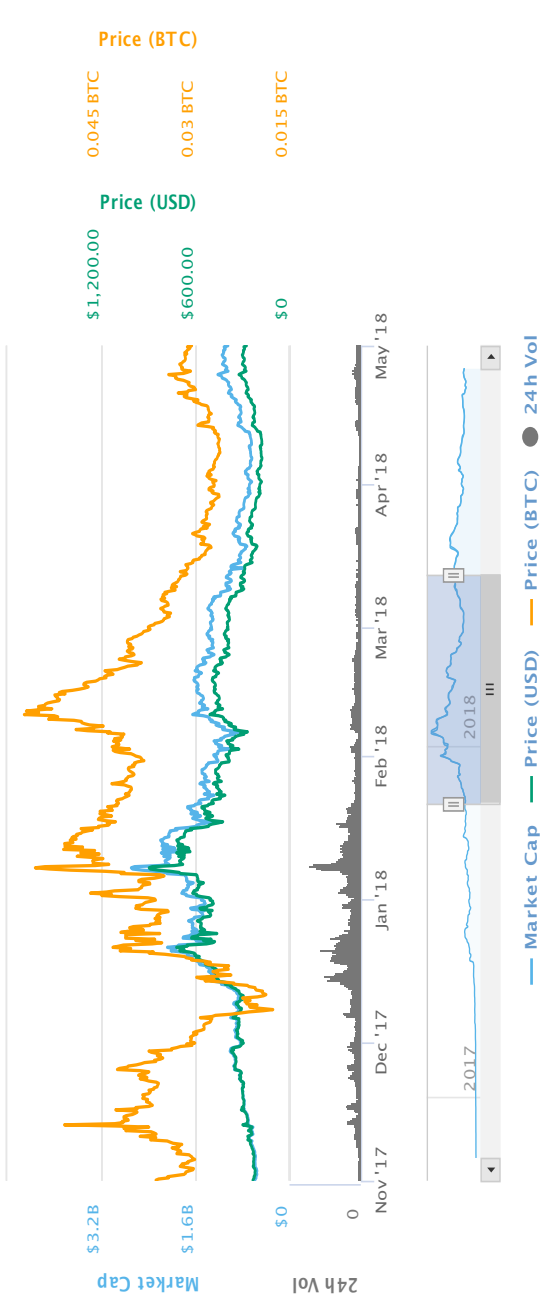
(d) $q = 0.5$



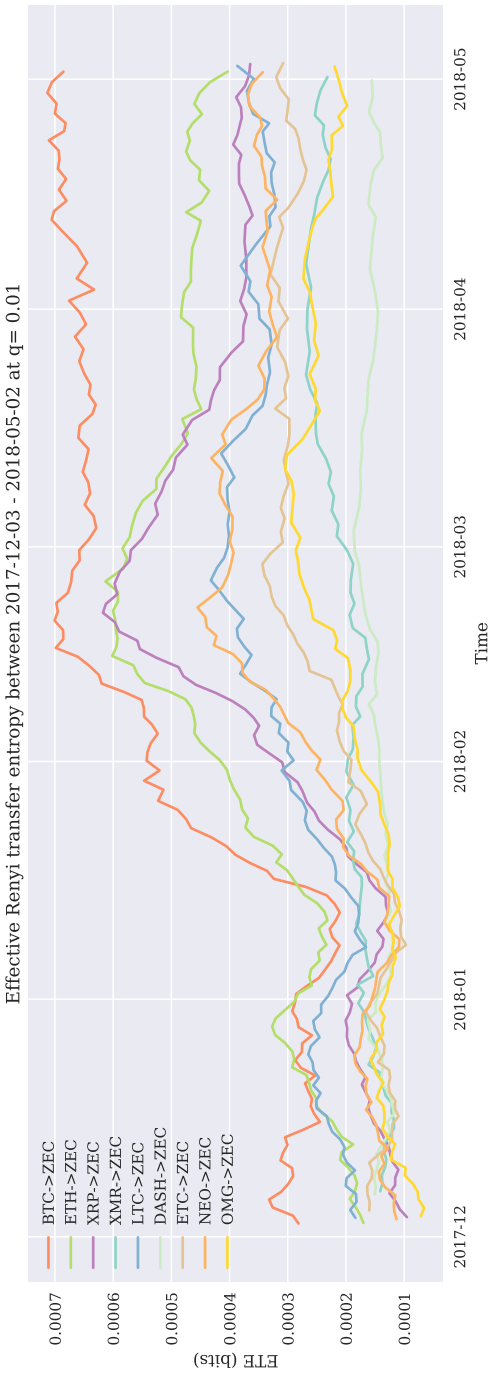
(e) $q = 0.8$



(f) $q = 2.0$

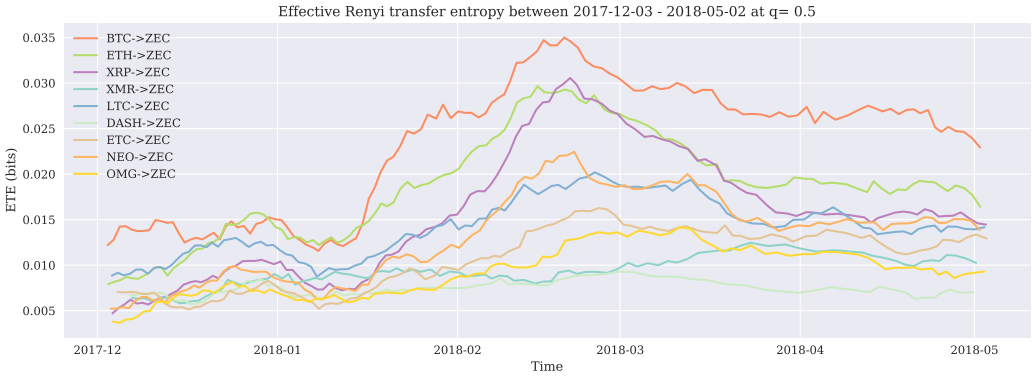


(a) Price and market cap

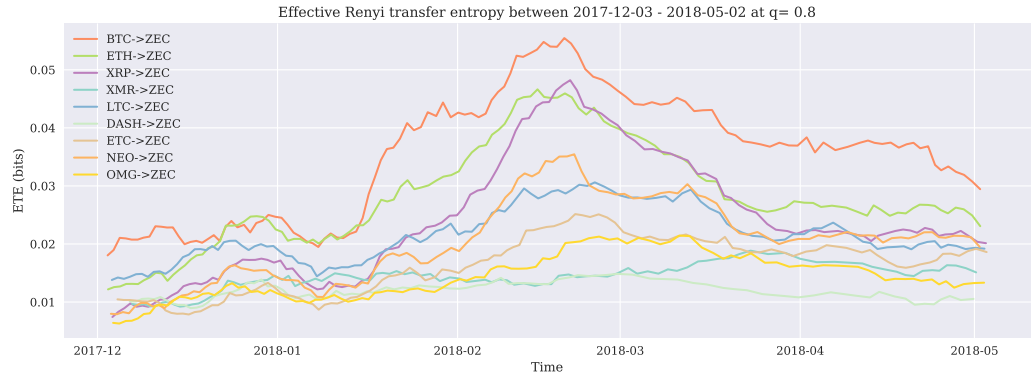


(b) $q = 0.01$

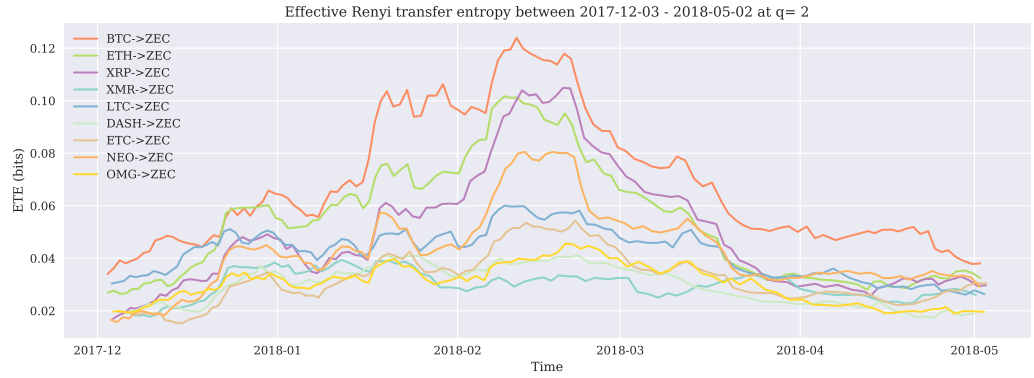
Figure 6.21: Effective Rényi transfer entropy from other cryptocurrencies towards ZEC for various values of q



(d) $q = 0.5$



(e) $q = 0.8$



(f) $q = 2.0$

Chapter 7

Conclusion

7.1 Evaluation of the Objective

The overarching aim of this thesis was to show that Gaussian processes and Rényi entropy can be valuable non-parametric tools for forecasting intraday volatility for a wide range of financial time series. This aim has been met by demonstrating the improvement in performance achieved by using Gaussian processes and Rényi transfer entropy across a large set of experiments encompassing data from many parts of the financial markets.

Another objective was to show that Gaussian processes perform consistently better at forecasting intraday volatility than other well established models - namely GARCH and some of its variants, especially mcsGARCH. This objective was met by demonstrating that Gaussian process based volatility forecasting techniques perform significantly better than these benchmarks, on a set of robust loss functions.

The aim was also to present a novel technique for improving volatility forecasting by combining the information flow from other financial time series by using effective Rényi transfer entropy - a non-parametric measure for studying transfer entropy. This technique was demonstrated by showing that effective Rényi entropy modified volatility forecasts performed significantly better than forecasts from standalone GPs and all other benchmarks.

The final aim was to establish that the higher performance of Gaussian processes is

consistent across a long period of time and a breadth of financial time series This was done by using a dataset of 50 symbols from financial markets over an 11 year period at 1-minute frequency and 10 cyroptocurrencies over a 6 month period at 5-minute frequency This demonstrated that the performance of Gaussian process based methods can be utilized for predicting the volatility of a diverse set of financial time series.

7.2 Summary of Key Contributions in the Thesis

In this thesis empirical volatility forecasting using Gaussian processes (GPs) was presented for stocks, market indices, forex and cryptocurrencies. Key innovations were presented in the application of GPs by using separated negative and positive returns in transformed log space, and the use of Rényi transfer entropy for incorporating information flow from other time series to modify volatility forecasts. Significant performance gains were demonstrated over strong benchmarks for volatility forecasting - the mcsGARCH model that is specially designed for intraday volatility, rolling average, and use of last observed value as next step forecast. A set of robust loss functions was used for assessing performance, and we established the significance of all results at the 95% confidence interval.

This thesis achieved the aim of demonstrating empirically that standalone GPs perform better than GARCH, EGARCH, GJR-GARCH in forecasting intraday financial volatility. This was done by presenting the results from the largest study done in literature to date, that uses GPs for this purpose. Approximately 50K experiments were carried out and over 18 million volatility forecasts were analysed, on 11 years worth of trading data from 50 market symbols sampled at 1-minute frequency to establish the significance of the results.

It was recognized that GARCH, EGARCH, GJR-GARCH have not been designed for intraday volatility forecasting therefore this comparison is not with the best in class. mcsGARCH is chosen as the best in class GARCH model specifically designed for forecasting intraday volatility. After this selection it was empirically established that when

mcsGARCH is used for benchmarking the performance of GPs, it is found that plain GPs are substantially worse than mcsGARCH. This negative result helped establish a need for innovation in the application of GPs to volatility forecasting.

An innovation for the GP intraday volatility forecasting model was developed by regressing on the negative and positive returns envelopes separately in log-space. The forecasts generated from this method were found to significantly outperform all benchmarks including mcsGARCH in terms of their loss function performance. This superior performance against mcsGARCH successfully demonstrated that GPs can be a significant tool in the volatility practitioner's forecasting toolkit.

In order to incorporate information from other financial time series into the volatility forecast the concept of Rényi transfer entropy was introduced as a non-parametric measure for calculating the asymmetrical information flows between two financial time series with the ability to emphasize different parts of the event space. The effective Rényi transfer entropy for key market indices was calculated and information flows around key global events were visualized. It was found that very high flows of information occur around global events of financial importance but very little influence was found for politically important events.

A new approach was devised for folding the effective Rényi transfer entropy into the forecast made by envelope based GPs. This new approach of using the ERTE modified envelope GP forecasts was applied for predicting intraday volatility for 6 market indices and 12 forex pairs. The performance improvement from these modified forecasts was benchmarked and found to be significantly better than all benchmarks considered including the standalone envelope GP forecasts. The parameters used in calculating the ERTE forecasts were found empirically and values which provided the most consistent gains in the training data were used in the comparison against benchmarks.

The approach was extended further and applied to forecast the volatility in cryptocurrency markets by conducting experiments on 6 month long time series from 10 cryptocurrencies sampled at 5-minute frequency. It was found that envelope based GPs significantly outperform against all three benchmarks on this dataset. This was the largest

volatility forecasting study done to date by using intraday frequency data from multiple cryptocurrencies.

An innovation was presented for folding the effective Rényi transfer entropy value from multiple time series into one value by extending the non-parametric formulation for ERTE. While this extension results in the need for a larger data set, its application for calculating ERTE remains computationally simple with a high degree of flexibility

The combined time series ERTE value was used for modifying the forecasts from envelope based GPs and applied to the case of using up to 9 secondary cryptocurrency time series for calculating the forecasts for 1 primary cryptocurrency. This technique was trialled for all 10 cryptocurrencies in the data and it was found that although this method still out performs the three benchmarks, its performance against the standalone envelope based GP was not consistent. A potential reason for this was discussed by assessing the relationship between the discretization parameter in ERTE and the requirements it poses for the size of the dataset.

7.3 Future Directions

Volatility forecasting and machine learning are two incredibly active research fields. The state of the art in machine learning based financial forecasting can improve at an accelerated pace by promoting more ways of bridging the gaps between the two fields.

The first step would be to promote and democratize the use a commonly used data set of high frequency financial time series that can be used to rank the performance of different machine learning algorithms against each other. Currently high frequency data is very expensive to access and only a few researchers globally have the funding or industry partnerships to use this data. If a large data set of high frequency financial time series were made publicly available it would enable researchers globally to participate in conducting research. A consistent data set with a breadth of symbols would make machine learning based financial volatility research more streamlined and transparent.

Second, the use of robust econometric loss functions should be promoted in machine

learning literature as this will make comparison of results easier across the two fields. This will allow developments in machine learning to be easily translated to the language that econometrics practitioners trust and enable rapid cross pollination of ideas.

Third, use of comparisons with strong benchmarks should be promoted in machine learning and econometrics literature. New literature coming out on intraday volatility in both fields still often chooses to benchmark performance against the GARCH (1,1). Many improvements have been made to GARCH - which was originally introduced in 1982. New sports cars would not choose to benchmark their performance against a 1982 sports car. Machine learning algorithms should not do this either and stronger, more recent benchmarks should be utilized in performance comparisons.

7.4 Open Questions

Exciting new developments are happening in machine learning each day. The use of asymmetric transfer entropies to study financial markets is also expanding rapidly. This gives rise to many possibilities for exploration. The questions I find most exciting are as follows:

- What does the transfer of entropy between separated negative and positive returns look like? Do negative returns in one financial time series contribute information differently than positive returns?
- What is the impact on forecasting performance if continuous Rényi transfer entropy is used instead of the more established discrete version?
- What is the impact of testing different likelihoods with Gaussian processes, for example student-t?
- What would be the impact of folding in information flows from volume of trades across different financial time series?

The future of volatility research looks exciting and full of possibility

Appendix A

Computational Setup

Database setup

I used a PostgreSQL 9.5.2 database hosted over a stand alone server, similar in capacity to an AWS db.t2.micro instance with SSD Storage of 200 GB, and CPU capacity equivalent to a 1.0–1.2 GHz 2007 Opteron or 2007 Xeon processor. This hosted the complete minute level data for all stock symbols as well as the results generated from all experiment runs.

Computing setup

The python interface code was run from a computing cluster with multiple cores (max 30) of Intel(R) Xeon(R) CPU E5-2667 v2 @ 3.30GHz

Programming setup

All code was created on Python 2.7.13, via Anaconda 4.3.1 (64-bit) running GCC 4.4.7 20120313 (Red Hat 4.4.7-1) on linux2.

Explicitly called packages and their versions are given below:

- pandas version 0.19.2
- numpy version 1.13.3

- scipy version 0.19.0
- GPflow version 0.3.6
- tensorflow version 1.0.1
- SQLalchmey version 1.1.5
- pandas_market_calendars version 0.12
- pytz version 2017.3

Bibliography

[1] Aït-Sahalia, Y., Mykland, P.A., Zhang, L.: How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise. The Review of Financial Studies 18(2), 351–416 (2005), http://dx.doi.org/10.1093/rfs/hhi016

[2] Aït-Sahalia, Y., Yu, J.: High frequency market microstructure noise estimates and liquidity measures. Annals of Applied Statistics 3(1), 422–457 (2009)

[3] Andersen, T.G., Bollerslev, T., Diebold, F.X.: Parametric and Nonparametric Volatility Measurement. In: Hansen, L., Ait-Sahalia, Y. (eds.) Handbook of Financial Econometrics. Amsterdam: North-Holland (2005)

[4] Armstrong, J.: Standards and practices for forecasting. In: Armstrong, J.S. (ed.) Principles of Forecasting: A Handbook for Researchers and Practitioners. Norwell, MA: Kluwer Academic Publishers (2001), http://link.springer.com/chapter/10.1007/978-0-306-47630-3_31

[5] Behrendt, S., Zimmermann, D.J., Paper, W.: Wikipedia search momentum and stock returns. https://dx.doi.org/10.2139/ssrn.3220053 pp. 1–42 (2018)

[6] Bekaert, G., Wu, G.: Asymmetric Volatility and Risk in Equity Markets. Review of Financial Studies 13(1), 1–42 (2000)

[7] Bentes, S.R., Menezes, R.: Entropy: A new measure of stock market volatility? Journal of Physics: Conference Series 394(1) (2012)

[8] Bessler, D., Yang, J.: The structure of interdependence in international stock markets. Journal of International Money and Finance 22(2), 261–287 (2003), https://econpapers.repec.org/RePEc:eee:jimfin:v:22:y:2003:i:2:p:261-287

[9] Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. Journal of econometrics 31, 307–327 (1986), http://www.sciencedirect.com/science/article/pii/0304407686900631

[10] Bonfiglioli, A.: Equities and Inequality. Ssrn pp. 1–47 (2005)

[11] Bossomaier, T., Barnett, L., Harré, M., Lizier, J.T.: An Introduction to Transfer Entropy (2016), http://link.springer.com/10.1007/978-3-319-43222-9

[12] Bouri, E., Azzi, G., Dyhrberg, A.H.: On the return-volatility relationship in the bitcoin market around the price crash of 2013. Economics 11, 1–17 (2017)

[13] Chambers, J.: Graphical Methods for Data Analysis. Chapman and Hall/CRC, New York, NY (1983)

[14] Chu, J., Chan, S., Nadarajah, S., Osterrieder, J.: GARCH Modelling of Cryptocurrencies. Journal of Risk and Financial Management 10(4), 17 (2017), http://www.mdpi.com/1911-8074/10/4/17

[15] Chunxia, Y., Xueshuai, Z., Luoluo, J., Sen, H., He, L.: Study on the contagion among American industries. Physica A: Statistical Mechanics and its Applications 444(C), 601–612 (2016), https://econpapers.repec.org/RePEc:eee:phsmap:v:444:y:2016:i:c:p:601-612

[16] Davidian, M., Carroll, R.J.: Variance Function Estimation. Journal of the American Statistical Association 82(400), 1079–1091 (1987), https://www.tandfonline.com/doi/abs/10.1080/01621459.1987.10478543

[17] De G. Matthews, A.G., Van Der Wilk, M., Nickson, T., Fujii, K., Boukouvalas, A., León-Villagrá, P., Ghahramani, Z., Hensman, J.: GPflow: A Gaussian Pro-

cess Library Using Tensorflow. J. Mach. Learn. Res. 18(1), 1299–1304 (jan 2017), http://dl.acm.org/citation.cfm?id=3122009.3122049

[18] De Gooijer, J.G., Hyndman, R.J.: 25 Years of Time Series Forecasting. International Journal of Forecasting 22(3), 443–473 (jan 2006), http://linkinghub.elsevier.com/retrieve/pii/S0169207006000021

[19] Dimpfl, T., Peter, F.J.: Using transfer entropy to measure information flows between financial markets. Studies in Nonlinear Dynamics and Econometrics 17(1), 85–102 (2013)

[20] Dimpfl, T., Peter, F.J.: The impact of the financial crisis on transatlantic information flows: An intraday analysis. Journal of International Financial Markets, Institutions and Money 31, 1–13 (jul 2014), http://linkinghub.elsevier.com/retrieve/pii/S1042443114000304

[21] Dimpfl, T., Peter, F.J.: Analyzing volatility transmission using group transfer entropy. Energy Economics 75, 368–376 (2018), https://doi.org/10.1016/j.eneco.2018.08.008

[22] Ding, Z., Granger, C.W., Engle, R.F.: A long memory property of stock market returns and a new model. Journal of Empirical Finance 1(1), 83–106 (1993)

[23] Dorion, C., Chapados, N.: Volatility Forecasting and Explanatory Variables: A Tractable Bayesian Approach to Stochastic Volatility. Available at SSRN 1747945 (2013), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1747945

[24] Ederington, L.H., Guan, W.: Longer-term time-series volatility forecasts. Journal of Financial and Quantitative Analysis 45(4), 1055–1076 (2010)

[25] Engle, R.: Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the Econometric Society 50(4), 987–1007 (1982), http://www.jstor.org/stable/10.2307/1912773

[26] Engle, R.F., Sokalska, M.E.: Forecasting intraday volatility in the US equity market. Multiplicative component GARCH. Journal of Financial Econometrics 10(1), 54–83 (2012)

[27] Forsberg, L., Ghysels, E.: Why Do Absolute Returns Predict Volatility So Well? Journal of Financial Econometrics 5(1), 31–67 (2007), http://dx.doi.org/10.1093/jjfinec/nbl010

[28] Ghalanos, A.: rugarch: Univariate GARCH models pp. 1–36 (2018)

[29] Giles, D.E.: Some properties of absolute returns as a proxy for volatility. Applied Financial Economics Letters 4(5), 347–350 (2008)

[30] Glosten, L.R., Jagannanthan, R., Runkle, D.E.: On The Relationship Between The Expected Value and The Volatility of The Nominal Excess Returns on Stocks. Journal of Finance 48(5), 1779–1801 (1993)

[31] Guo, T., Antulov-Fantulin, N.: An experimental study of Bitcoin fluctuation using machine learning methods. In: Association for Computing Machinery Conference (Conference'18) (2018), http://arxiv.org/abs/1802.04065

[32] Hansen, P.R., Lunde, A.: Consistent ranking of volatility models. Journal of Econometrics 131(1-2), 97–121 (mar 2006), http://linkinghub.elsevier.com/retrieve/pii/S0304407605000072

[33] Hansen, P., Lunde, A., Nason, J.: Choosing the best volatility models: The model confidence set approach*. Oxford Bulletin of Economics ... (401) (2003), http://onlinelibrary.wiley.com/doi/10.1046/j.0305-9049.2003.00086.x/full

[34] Hartley, R.: Transmission of Information. Bell System Technical Journal (September), 535–563 (1928)

[35] He, J., Shang, P.: Comparison of transfer entropy methods for financial time series. Physica A: Statistical Mechanics and its Applications 482, 772–785 (2017), http://dx.doi.org/10.1016/j.physa.2017.04.089

[36] Hibbert, A.M., Daigler, R.T., Dupoyet, B.: A behavioral explanation for the negative asymmetric return-volatility relation. Journal of Banking and Finance 32(10), 2254–2266 (2008)

[37] Huang, R.D., Stoll, H.R.: Market Microstructure and Stock Return Predictions. The Review of Financial Studies 7(1), 179–213 (1994), http://dx.doi.org/10.1093/rfs/7.1.179

[38] Hyndman, R.J., Koehler, A.B.: Another look at measures of forecast accuracy. International Journal of Forecasting 22(4), 679–688 (oct 2006), http://linkinghub.elsevier.com/retrieve/pii/S0169207006000239

[39] Jizba, P., Arimitsu, T.: The world according to Rényi: thermodynamics of multifractal systems. Annals of Physics 312(1), 17–59 (2004), http://linkinghub.elsevier.com/retrieve/pii/S0003491604000132

[40] Jizba, P., Kleinert, H., Shefaat, M.: Rnyi's information transfer between financial time series. Physica A: Statistical Mechanics and its Applications 391(10), 2971–2989 (2012)

[41] Kwon, O., Yang, J.S.: Information flow between composite stock index and individual stocks. Physica A: Statistical Mechanics and its Applications 387(12), 2851–2856 (may 2008), http://linkinghub.elsevier.com/retrieve/pii/S0378437108000186

[42] Low, C.: The Fear and Exuberance from Implied Volatility of S & P 100 Index Options. The Journal of Business 77(3), 527–546 (2004)

[43] Makridakis, S., Hibon, M.: The M3-Competition: results, conclusions and implications. International Journal of Forecasting 16(4), 451–476 (oct 2000), http://linkinghub.elsevier.com/retrieve/pii/S0169207000000571

[44] Marschinski, R., Kantz, H.: Analysing the information flow between financial time series. The European Physical Journal B 30(2), 275–281 (nov 2002), http://www.springerlink.com/index/10.1140/epjb/e2002-00379-2

[45] Marschinski, R., Kantz, H.: Analysing the information flow between financial time series - An improved estimator for transfer entropy. Eur. Phys. J. B 30(2), 275–281 (2002), https://doi.org/10.1140/epjb/e2002-00379-2

[46] Mckenzie, M.D.: The impact of exchange rate volatility on Australian trade flows. Journal of International Financial Markets, Institutions and Money 8(1), 21–38 (1998)

[47] Nagata, S., Oya, K.: Volatility Forecast Comparison with Biased Proxy and Related Statistic. In: 7th CSDA International Conference on Computational and Financial Econometrics (2013)

[48] O'Hara, M.: Making market microstructure matter. Financial Management 28(2), 83–90 (1999), http://www.jstor.org/stable/3666197

[49] Osborne, M.a., Roberts, S.J., Rogers, a., Ramchurn, S.D., Jennings, N.R.: Towards Real-Time Information Processing of Sensor Network Data Using Computationally Efficient Multi-output Gaussian Processes. 2008 International Conference on Information Processing in Sensor Networks (ipsn 2008) pp. 109–120 (apr 2008), http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4505467

[50] Osterrieder, J., Chan, S., Chu, J., Nadarajah, S.: A Statistical Analysis of Cryptocurrencies. Ssrn (February), 1–23 (2017)

[51] Patton, A.J.: Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160(1), 246–256 (2011), http://dx.doi.org/10.1016/j.jeconom.2010.03.034

[52] Pichl, L., Kaizoji, T.: Volatility Analysis of Bitcoin Price Time Series (2017)

[53] Platanios, E., Chatzis, S.: Nonparametric Mixtures of Multi-Output Heteroscedastic Gaussian Processes for Volatility Modeling pp. 1–4 (2013), http://www.academia.edu/download/30490812/Nonparametric_-Mixtures_of_Multi-Output_Heteroscedastic_Gaussian_Processes_-for_Volatility_Modeling_-_Paper.pdf

[54] Poon, S.H.: A Practical Guide to Forecasting Financial Market Volatility. Wiley, New York (2005)

[55] Poon, S.H., Granger, C.: Practical Issues in Forecasting Volatility Ser-Huang Poon and Clive Granger. Financial Analysts Journal 61(1), 45–56 (2005)

[56] Rasmussen, C.E., Williams, C.K.I.: Gaussian processes for machine learning., vol. 14. the MIT Press (apr 2006), http://www.ncbi.nlm.nih.gov/pubmed/15112367

[57] Roberts, S., Osborne, M., Ebden, M., Reece, S., Gibson, N., Aigrain, S.: Gaussian processes for time-series modelling. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences 371 (feb 2013), http://www.ncbi.nlm.nih.gov/pubmed/23277607

[58] Rocco, M.: Extreme Value Theory for Finance: A Survey. SSRN Electronic Journal (99), 1–42 (2012), http://www.ssrn.com/abstract=1998740

[59] Schreiber, T.: Measuring information transfer. Physical review letters pp. 461–464 (2000), http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.85.461

[60] Shannon, C.: A Mathematical Theory of Communication. Bell System Technology 27(I), 379–656 (1948)

[61] Sheppard, K., Liu, L., Patton, A.: Does Anything Beat 5-Minute RV? A Comparison of Realized Measures Across Multiple Asset Classes. Economics Work-

ing Papers Series 645, University of Oxford, Department of Economics. (2013), http://ideas.repec.org/p/oxf/wpaper/645.html

[62] Sheraz, M., Dedu, S., Preda, V.: Entropy Measures for Assessing Volatile Markets. Procedia Economics and Finance 22(November 2014), 655–662 (2015), http://linkinghub.elsevier.com/retrieve/pii/S2212567115002798

[63] Teng, Y., Shang, P.: Transfer entropy coefficient: Quantifying level of information flow between financial time series. Physica A: Statistical Mechanics and its Applications 469, 60–70 (2017), http://dx.doi.org/10.1016/j.physa.2016.11.061

[64] Williams, C., Rasmussen, C.: Gaussian processes for regression. In: Touretzky, D.S., Mozer, M.C., Hasselmo, M.E. (eds.) Advances in Neural Information Processing Systems 8. pp. 1–7. MIT Press (1996), http://eprints.aston.ac.uk/651/

[65] Wu, G.: The Determinants of. The review of Financial Studies 14(2), 837–859 (2001)

[66] Wu, Y., Hernández-Lobato, J.M., Ghahramani, Z.: Gaussian Process Volatility Model. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 27, pp. 1044–1052. Curran Associates, Inc. (2014), http://papers.nips.cc/paper/5376-gaussian-process-volatility-model.pdf

[67] Xiao, X., Zhou, C.: Entropy-Based Implied Volatility and its Information Content. Ssrn pp. 1–46 (2015)

[68] Zheng, Z., Qiao, Z., Takaishi, T., Stanley, H.E., Li, B.: Realized Volatility and Absolute Return Volatility: A Comparison Indicating Market Risk. PLOS ONE 9(7), 1–10 (2014), https://doi.org/10.1371/journal.pone.0102940