



THE UNIVERSITY OF  
**WAIKATO**  
*Te Whare Wānanga o Waikato*

Research Commons

<http://researchcommons.waikato.ac.nz/>

## Research Commons at the University of Waikato

### Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

# **Hokohoko: A Comprehensive Framework for Evaluating Artificial Intelligence-based and Statistical Techniques for Foreign Exchange Speculation**

**A thesis  
submitted in fulfilment  
of the requirements for the Degree  
of  
Master of Science (Research)  
at  
The University of Waikato  
by  
Neil Christopher Bradley**



THE UNIVERSITY OF  
**WAIKATO**  
*Te Whare Wānanga o Waikato*

2020

# Abstract

This thesis investigates the measurement of predictor performance as applied to foreign exchange speculation. It outlines the development of key ideas and techniques over the course of the last 120 years, and examines the datasets and metrics used within a representative sample of the academic corpus. In this examination two problems are identified: first, there is a lack of consistency in the datasets used to test researchers' algorithms; and second, a large variety of metrics are used, most of which are either inappropriate for or inappropriately applied to FOREX speculation. To address these issues, this thesis presents two solutions: a Python library, *Hokohoko*, which provides a consistent dataset and interface for testing FOREX prediction algorithms; and a new metric, Speculative Accuracy, which it argues provides a more appropriate measure of usefulness with regards to speculation. *Hokohoko* is then used to test a series of hypotheses regarding the usefulness of various metrics, alongside Speculative Accuracy.

# Acknowledgements

Completion of a Master's degree requires allocation of resources and dedication of time from both the author and supporters. This thesis was no different, and I am indebted to, and thankful for, the contributions made by other persons whose impact is not necessarily recorded within the thesis' body itself. To that end, I would like to thank Dr. Michael Mayo and Dr. Panos Patros for their supervisory expertise; Dr. Te Taka Keegan for his invaluable assistance in naming *Hokohoko*; Christian Anderson-Scott and the ORCA lab for invigorating discussions over the course of the research; Sarah and Jason Kelly for their assistance with proofreading; my two little boys Thomas and Joshua, who may one day understand why Daddy's office was sometimes out-of-bounds; and last but not least, my wife Sophie for her love and support.

*Dedicated to my family.*

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Foreign Exchange Speculation</b>	<b>3</b>
<b>3</b>	<b>Background: Foreign Exchange Prediction and Modeling</b>	<b>9</b>
3.1	Statistical Properties . . . . .	10
3.2	Efficient Market Hypothesis . . . . .	12
3.3	Auto-Regression . . . . .	14
3.4	Conditional Heteroscedasticity . . . . .	15
3.5	Nonlinearity . . . . .	17
3.6	Chaos Theory . . . . .	19
3.7	Neural Networks . . . . .	23
3.8	Wavelet Decomposition . . . . .	27
3.9	Particle Swarm Optimization . . . . .	28
3.10	Support Vectors . . . . .	29
3.11	Hybrid Predictors . . . . .	31
3.12	Market Efficiency and Chaotic Behaviour . . . . .	32
3.13	Concluding Remarks . . . . .	34
<b>4</b>	<b>Measuring Foreign Exchange Predictions: Common Metrics</b>	<b>37</b>
4.1	Datasets . . . . .	38
4.2	Metrics . . . . .	41
4.2.1	What is Being Measured? . . . . .	43
4.2.2	Metric Family: Statistical . . . . .	45
4.2.3	Metric Family: Directional . . . . .	48
4.2.4	Metric Family: Comparative . . . . .	51
4.2.4.1	Pearson's Correlation Coefficient . . . . .	51
4.2.4.2	Friedman's Test . . . . .	52
4.2.4.3	Theil's U Metric . . . . .	52
4.2.4.4	Box-Pierce and Ljung-Box . . . . .	53
4.2.4.5	Granger-Newbold . . . . .	55
4.2.4.6	Brock-Dechert-Scheinkman Test . . . . .	56

4.2.4.7	Diebold-Mariano Test . . . . .	58
4.2.5	Metric Family: Profit-based . . . . .	59
4.2.5.1	Profits . . . . .	60
4.2.5.2	Sharpe Ratio . . . . .	61
4.2.5.3	Compounded Return . . . . .	62
4.2.5.4	Annualised Return . . . . .	62
4.2.5.5	Drawdown . . . . .	63
4.2.5.6	Stirling Ratio . . . . .	64
4.2.5.7	Profit Factor . . . . .	64
4.2.5.8	Sortino Ratio . . . . .	64
4.2.5.9	Problems with Profit-based Metrics . . . . .	65
4.3	(Not So) Tough Benchmarks . . . . .	66
4.4	Summary . . . . .	67
<b>5</b>	<b>Addressing Incomparable Datasets: Introducing Hokohoko</b>	<b>69</b>
5.1	Data . . . . .	70
5.2	Parallel Execution . . . . .	72
5.3	Simulation . . . . .	78
5.4	Availability . . . . .	80
5.5	Quality Assurance . . . . .	81
5.6	Invocation . . . . .	81
5.7	Parameters . . . . .	82
5.8	Included Predictors . . . . .	83
5.9	Included Assessors . . . . .	85
<b>6</b>	<b>Addressing Inappropriate Metrics: Introducing Speculative Accuracy</b>	<b>86</b>
6.1	The Ideal Metric . . . . .	86
6.2	Speculative Accuracy: A New Metric . . . . .	87
6.3	Examples . . . . .	91
6.4	<i>Hokohoko</i> and Speculative Accuracy? . . . . .	95
<b>7</b>	<b>Experiment Methodology</b>	<b>96</b>
7.1	Predictors . . . . .	97
7.2	Metrics . . . . .	100
7.3	Experiment Aims . . . . .	103
7.4	Experiment: Correlation of Metrics to Profit-per-Position . . .	104
7.5	Experiment: Ranking of Predictors by Metric . . . . .	106
<b>8</b>	<b>Results and Discussion</b>	<b>107</b>
8.1	Results: Correlation of Metrics with Profits . . . . .	108

8.1.1	Performance of Directional Metrics for Random Walk Predictions in <i>Hokohoko</i> . . . . .	108
8.1.2	Performance of CLOSE-based Error Metrics for Random Walk Predictions in <i>Hokohoko</i> . . . . .	111
8.1.3	Performance of HIGH/LOW-based Error Metrics for Random Walk Predictions in <i>Hokohoko</i> . . . . .	113
8.1.4	Performance of Speculative Accuracy Variants for Random Walk in <i>Hokohoko</i> . . . . .	115
8.1.5	Ranking Metrics by Correlation . . . . .	119
8.2	Results: Ranking of Predictors by Metric . . . . .	123
8.3	Hypotheses Evaluation . . . . .	124
8.4	Conclusion . . . . .	126
<b>9</b>	<b>Conclusion</b>	<b>128</b>
	<b>Appendices</b>	<b>131</b>
<b>A</b>	<b>Predictor Code Listing: Example.py</b>	<b>131</b>
<b>B</b>	<b>Assessor Code Listing</b>	<b>136</b>
<b>C</b>	<b>Results Tables for Experiment 2</b>	<b>143</b>
	<b>Bibliography</b>	<b>156</b>

# List of Tables

4.1	Relative Frequency of Metrics . . . . .	42
4.2	Statistical Metrics . . . . .	46
4.3	Directional Correctness . . . . .	49
4.4	Direction Parts . . . . .	49
4.5	Combined Directional Statistics . . . . .	49
4.6	Papers using Theil's U-Metric . . . . .	53
4.7	Diebold-Mariano Examples . . . . .	59
5.1	Hokohoko Dataset Symbols . . . . .	71
5.2	Predictors Included in Hokohoko . . . . .	84
7.1	Metrics used in Experiments . . . . .	101
8.1	Correlation Between Metrics and Realised Profits . . . . .	117
8.2	Metrics by Average Rank . . . . .	118
8.3	Predictor Ranks for GBPUSD . . . . .	121
8.4	Predictor Ranks for EURGBP . . . . .	122



# List of Figures

2.1	The Foreign Exchange Market . . . . .	3
2.2	Buy and Sell Outcomes . . . . .	4
2.3	Orders Stack . . . . .	5
3.1	Gaussian and Paretian Distributions . . . . .	11
3.2	Student- <i>t</i> and Heteroscedastic Distributions . . . . .	15
3.3	Visualization of Lorenz's System . . . . .	20
3.4	A Simple ANN . . . . .	24
3.5	Support Vector Example . . . . .	30
3.6	Predictive Technology Timeline . . . . .	35
4.1	Symbol Distribution within Corpus . . . . .	39
4.2	Visual Representation of GBPUSD Datasets Used . . . . .	39
4.3	Truncation via TAKE_PROFIT and STOP_LOSS . . . . .	43
4.4	Graphical Representations of FOREX Prices . . . . .	44
4.5	Irrelevance of CLOSE Values . . . . .	48
5.1	Overlapping Periods . . . . .	72
5.2	Hokohoko Execution Flowchart . . . . .	73
5.3	Hokohoko's Message Sequence . . . . .	75
6.1	Results Metrics used by Speculative Accuracy . . . . .	89
6.2	Normalisation of Speculative Accuracy . . . . .	94
8.1	Correlation of Directional Metrics . . . . .	108
8.2	Correlation of CLOSE-based Metrics . . . . .	110
8.3	Correlation of HIGH/LOW-based Metrics . . . . .	112
8.4	Correlation of Speculative Accuracy Metrics . . . . .	114



**Format conventions used within this thesis.**

Capitalised	A reference to a class, or named item.
<i>italics</i>	Italics are used for emphasis, or to indicate the introduction of a new concept, or for named software.
<b>bold</b>	Used to highlight important points.
monospace	Indicates code.
LOWERCASE CAPITALS	Items in lowercase capitals are properties related to the foreign exchange market.

# Trading Terms

**ask** The current price that traders are *asking* for their currency. This is the value that a Position is bought for.

**balance** The amount of currency a trading account nominally holds.

**bid** The current price that traders are *offering* to buy the currency. This is the value that a Position is sold for.

**broker** A company that allows smaller traders access to the foreign exchange market.

**candlestick** A type of chart used to show currency movements. It consists of two vertical bars overlaid, one thick and one thin, whose top and bottom indicate different exchange rate values.

**close** The exchange rate for a given Symbol, at the end of a fixed interval, or epoch, of time.

**equity** The amount of currency a trading account holds at a given instant.

**high** The highest exchange rate observed for a given Symbol, within a specific interval of time.

**leverage** A multiplicative factor applied to trades, magnifying profit, losses and trading costs.

**liquidity provider** A large institute that trades on the exchange rate itself, on behalf of businesses, etc. A common example is banks and payment service providers, such as credit card companies.

**low** The lowest exchange rate observed for a given Symbol, within a specific interval of time.

**margin** The ratio of a trading account's equity available compared to equity used.

**open** The exchange rate for a given Symbol, at the start of a fixed interval of time.

**Order** A request to trade, with conditions attached. Some of these conditions include an acceptable exchange rate range to begin trading, the Symbol to trade in, and the volume to trade. It can also include extra information, such as TAKE\_PROFIT and STOP\_LOSS conditions..

**Position** An active trade. The value of a Position changes with exchange rate movements, and the aim of trading is to have Positions that increase in value. BUY Positions increase in value as the exchange rate increases, and SELL Positions increase in value as the exchange rate decreases.

**spread** The difference between ASK and BID rates for a Symbol.

**stop\_loss** An exchange rate programmed into a Position which, if achieved, will cause the Position to close automatically, accepting the losses incurred. The losses are applied to the trading account's balance.

**Symbol** A trading currency pair, made up of a base currency and a target currency. E.g., AUDNZD has the base AUD and the target NZD. When designated 'Symbol' within the text, this refers to the currency pair, and 'SYMBOL' refers to the relevant variable in an Order or Position.

**take\_profit** An exchange rate programmed into a Position which, if achieved, will cause the Position to close automatically, realising the profits made. These profits are applied to the trading account's balance.

**tick** An instantaneous exchange rate movement. Not fixed to any particular time-scale, a tick occurs when the ASK or BID rates of a Symbol change.

**trader** A person or entity who trades on the foreign exchange market.

### Mathematical Symbols

$y_t$	A point $y$ of the exchange rate series $Y$ at time $t$ .
$\delta y_t$	The first-order difference between the point $y_t$ and its predecessor $y_{t-1}$ .
$\hat{y}$	The predicted value.
$\mu$	The mean value.
$n, N$	The number of events within a time series.
$\epsilon_t$	The error value at $t$ .
$E(\hat{y}, y)$	An error function that calculates $\epsilon$ from $\hat{y}$ and $y$ .
$P(x)$	The probability of $x$ within a sample.
$\sigma$	The standard deviation.
$\psi$	A set of variables.
$\xi, \zeta$	Random noise belonging to a distribution.
$\rho$	Pearson's Correlation Coefficient.
$\bar{\rho}$	Indicates the mean correlation for a series of results.

# Chapter 1

## Introduction

Since floating in 1971, the Foreign Exchange (FOREX) market has grown to become the world's largest trading platform. With a daily turnover of USD 6.9 trillion, it is estimated that speculation makes up over 99 percent of trading activities (Borio, Claessens, Mojon, Shin, & Wooldridge, 2020; Global Policy Forum, 2013). To speculate in FOREX, a trader guesses or predicts future exchange rate movements and attempts to profiteer through buying and selling currency pairs (Symbols). Profits made can be magnified through the use of leverage. However, this magnification applies to losses and trading costs also, making accurate prediction an essential element of FOREX speculation.

In response to the need for accurate prediction, many different algorithms have been developed over the years to model and/or predict exchange rate movements. However, within the voluminous body of FOREX literature, scant attention has been paid to the algorithms' relative predictive performance, and even less to the validity of any metrics used. This thesis aims to redress this lack of attention, and thus considers the issue of benchmarking predictors for the purpose of FOREX speculation.

Analysis of a representative subset of the FOREX corpus reveals two main shortcomings within the literature. First, there are as many different datasets used for benchmarking as there are papers, rendering direct comparison between papers impossible. Second, there is a wide variety of



different metrics used, most of which are either inapplicable or inappropriately applied to FOREX speculation, having little correlation with speculative results. These issues combined render the sum total of the speculative results presented within the FOREX corpus largely meaningless.

Based on the hypothesis that a suitable metric will show significant correlation with speculative results, and that a suitable benchmark will enable easily comparable results, this thesis proposes two novel solutions to combat the generally meaningless results pervading the FOREX corpus:

1. An open-source library, *Hokohoko*, which provides an API and dataset for consistent benchmarking of FOREX predictors.
2. A new metric, Speculative Accuracy, which is designed specifically for application to FOREX speculation.

These solutions are subjected to a thorough experimental evaluation, with this thesis proceeding as follows:

Chapter 2 describes the operation of the FOREX market from the view of a trader and outlines the major issues affecting predictability and profitability. Chapter 3 presents a short history of FOREX modeling and prediction, discussing the development of several key ideas and techniques. Chapter 4 analyses the metrics used within a representative subset of the FOREX corpus. Chapter 5 introduces the *Hokohoko* library and Chapter 6 describes the Speculative Accuracy metric. Chapter 7 explains the methodology used to test the suitability of Speculative Accuracy for FOREX speculation, with Chapter 8 presenting the results and discussion thereof. And finally, Chapter 9 concludes with a summary of the work undertaken and suggestions towards future research.

## Chapter 2

# Foreign Exchange Speculation

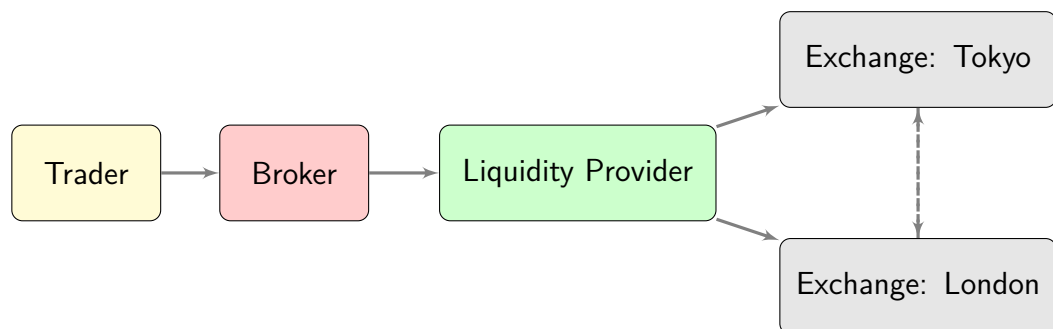


Figure 2.1: A condensed structure of the FOREX market, from a trader's perspective. Each node could be considered representative of many entities, e.g., millions of traders, thousands of brokers and liquidity providers, and a handful of exchanges around the world.

As shown in Fig. 2.1, the FOREX market is made up of multiple layers of participants, each of which has its own role. At the heart of the FOREX market is a network of central currency exchanges, which are responsible for matching trade requests, and are located around the world in different time zones. Individual traders, however, do not access these exchanges directly. Instead, traders hold accounts with *brokers*, who in turn access the market through *liquidity providers*. Liquidity providers are large companies, such as banks and payment service providers, who make the actual trades on the FOREX market—typically in the order of millions of dollars. Brokers are much smaller companies that allow access to the market for clients who do not otherwise possess the required wherewithal for participation. Global trade

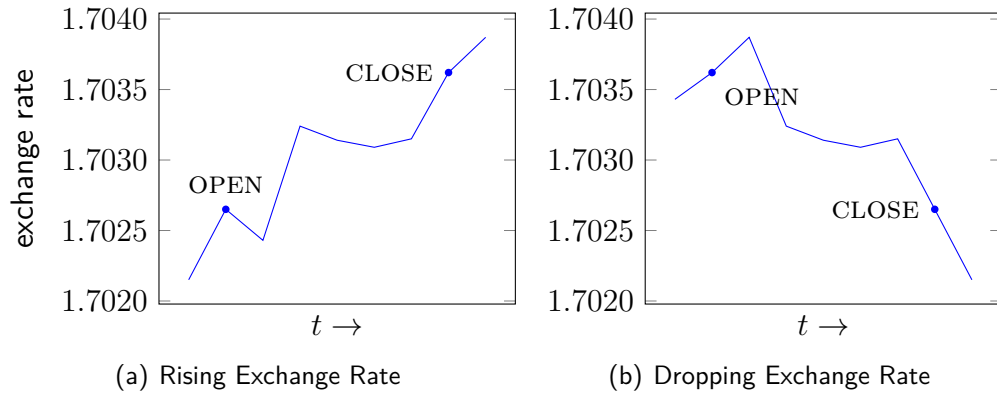


Figure 2.2: Example outcomes for different exchange rate movements. In (a), if the Position opened was a BUY Position, it will have made a profit at CLOSE. On the other hand, if it was a SELL Position, it will close at a loss. In (b), the scenario is reversed: a SELL Position will have made a profit, whereas a BUY will realise a loss.

also happens through the FOREX market, however, these transactions happen directly through the liquidity providers and make up a tiny fraction of FOREX trading.

To speculate in FOREX, a trader guesses or predicts future exchange rate movements, and places BUY or SELL *Orders* through their broker, as they deem appropriate. Orders can be placed with a number of conditions, and these conditions determine how the Order will be processed by the FOREX market. At the most basic level, an Order consists of six values: the SYMBOL to trade, the direction to trade in (either BUY or SELL), the VOLUME to be traded, and optional OPEN, TAKE\_PROFIT and STOP\_LOSS values. If specified, OPEN indicates the exchange rate at which the Order will become an active *Position*, otherwise it will activate immediately at the current rate. TAKE\_PROFIT and STOP\_LOSS specify exchange rates at which the Position will automatically close, otherwise the Position will remain active until closed manually. On close, a Position implicitly becomes an Order in the opposite direction, activating immediately. As shown in Fig. 2.2, for a BUY Position a profit is made when the exchange rate for the Symbol is higher at CLOSE than OPEN, and a loss realised if the exchange rate is lower at CLOSE than OPEN. For a SELL Position the conditions are reversed, with a profit made if CLOSE is lower than OPEN

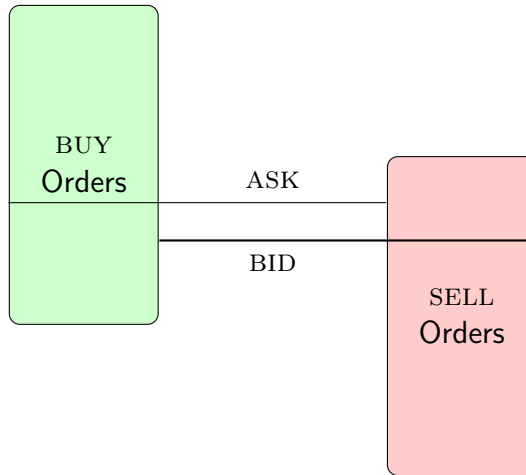


Figure 2.3: The FOREX market as two Orders stacks, for a given Symbol. Whilst the stacks overlap, only the Positions on the 'correct' side of the price are activated, with the others remaining dormant until the exchange rate reaches them.

and a loss realised if CLOSE is higher than OPEN.

Within the FOREX market, Orders are placed into a queue with other Orders of the same direction and rate and fulfilled once they reach the front of the queue (see Fig. 2.3). The exchange rate per-Symbol consists of two values: the ASK and BID rates. The ASK rate is what traders are requesting for their Positions and is the value at which BUY Orders are fulfilled; and the BID rate is what traders are offering to buy Positions for and the value at which SELL Orders are fulfilled. The difference between BID and ASK is known as the *spread*, and a Position needs to move further than the spread in order to make a profit.

Like all speculation, trading in FOREX is risky. If the market were 'fair', it would be possible to make money by the use of analysis and appropriate TAKE\_PROFIT and STOP\_LOSS conditions. However, the market is not fair, and there are extra factors a trader needs to take into account besides just exchange rate movements (Mandelbrot, 1963b; Engel & Hamilton, 1990).

First, placing a trade invokes a series of costs. As well as the spread, there is no guarantee that a Position will open or close at the specified exchange rate. Instead, it will open or close at the nearest available rate, with the difference between the requested and the actual rate called *slippage*. Every day a Position

is held, it accumulates small, interest-based *swap fees*. The brokers themselves also usually charge a small *commission*, typically based on the traded volume. All these costs reduce a trader's profits and therefore need to be considered when placing an Order.

Second, the FOREX market has proven difficult to regulate, and unscrupulous parties have found ways to manipulate the system. Examples of such manipulation include insider trading, price fixing and triggering. Insider trading is when a party has access to information not yet public, such as a pending policy change, or not public at all, such as a liquidity provider's internal Order stack. Price fixing occurs when large trading parties collude, in order to maximise their own profits, at the expense of the rest of the market. And triggering happens when a trading party becomes large enough for its own trading to cause market swings, and they then manipulate the market by issuing false trades and then trading on the reaction. Over the years, a number of parties have been investigated and disciplined for such actions, with jail terms and billions of dollars in fines handed out (BBC, 2015; Chee & Ridley, 2019).

And third, the behaviour of the FOREX market is a complex interaction of the trading forces between principalities, and as such the fiscal policies of different governing bodies can have significant effects on exchange rate movements. An example of this occurred on January 15, 2015, when the Swiss government decided to 'unpeg' the Swiss franc from the Euro. Without warning, the EURCHF exchange rate lost almost twenty percent of its value, causing massive losses for hedge funds around the world ("Swiss Franc", n.d.). Another example is found in the behaviour of the central banks, who have at times intervened in the market in an attempt to maintain the relative value of their currency (Mussa, 1979; LeBaron, 1999; Cheung & Chinn, 2001).

Whilst these issues of costs, manipulation and interference are properties of the FOREX market itself, there is a fourth issue directly under the trader's

control: *leverage*. Leverage is where liquidity providers allow traders to use only a portion of their own equity for a trade, with the remainder made up by the liquidity provider. For example, a liquidity provider may provide a 1:100 leverage rate, which means a trader can place a trade for USD1000, whilst using only USD10 of their own equity. Leverage magnifies all aspects of trading: profits, losses and costs. It thus enables potentially greater profits to be made by the trader. In the process, leverage also magnifies the effect exchange rate movements have on a trader's operating *margin*.

The margin is the ratio between how much equity a trader has risked versus how much equity they hold, and is calculated:

$$\text{margin} = \frac{\text{account equity}}{\text{equity used}} \quad (2.1)$$

Margin is affected by a traders open Positions, with profitable Positions increasing the margin and losing Positions decreasing it. If a trader's margin drops too low the consequences can be dire. First, their liquidity provider may prevent the trader from placing new trades, severely restricting their ability to 'trade out' of their predicament. Then, if the margin reduces further, they may simply start to foreclose the trader's Positions until the trader's margin is above the desired minimum level again, forcing the trader to accept whatever losses this incurs. If a trader's account reaches zero margin, the account is rendered defunct and all trades are then closed. In some scenarios, a trader may even end up with a bill for any excess losses made by the liquidity provider due to their trading. Whilst a trader can close their own trades or inject extra money into their trading account to maintain margin, by far the best approach is to have minimal losing Positions in the first place.

The issues of costs, manipulation, interference and leverage all play an important role in the profitability of FOREX trading. Costs need to be minimised, risks needs to be assessed and mitigated, and margin needs to be

managed in order to maximise profitability. Therefore, it is essential that a trader be able to predict market movements as accurately as possible. Given the importance of accurate prediction, the sheer size of the market, and the massive number of participants, it comes as no surprise that the last century has seen a staggering amount of work published on the topics of modeling and predicting FOREX movements.

# Chapter 3

## Background: Foreign Exchange

### Prediction and Modeling

Since Bachelier's original work on speculative analysis in 1900, *Théorie de la Spéculation*, the world has witnessed the birth of the computer, key fields such as Econophysics and Chaos Theory, and the floating exchange rate (Bachelier, 1900; May, 2006). Thousands of papers have been published by academics, government agencies and private businesses, seeking to understand the FOREX market, its composition, behaviour, movements, volatility and inherent risks. Some are motivated by a quest for understanding, perhaps to influence policy or add to the body of knowledge, whilst others are interested more in applications for investments, such as predicting future price movements and volatility, and both groups are prolific publishers. Since the collection of related work is so vast, this chapter concentrates on the development of a few key ideas and their application to the field of FOREX speculation. These developments are followed in somewhat chronological order, though they were often developed in parallel with each other, and not all contributions have equal merit. Of particular interest will be the development of the economic theory, the Efficient Market Hypothesis and its presence in the FOREX literature.



### 3.1 Statistical Properties

Widely regarded as the first application of mathematics in finance, Bachelier's work was based on the assumption that stock market price changes followed a Gaussian distribution and were therefore predictable through probability theory. Stating:

“The influences which determine the movements of the [financial markets] are innumerable. Events past, present or even anticipated often showing no apparent connection with its fluctuations, yet have repercussions on its course. Beside fluctuations from, as it were, natural causes, artificial causes are also involved. ... The determination of these fluctuations is subject to an infinite number of factors: it is therefore impossible to expect a mathematically exact forecast.” (Bachelier (1900), *trans.* May (2006))

Bachelier went on to state that while Probability Theory will never be able to predict the market exactly, it may well be possible to study the state of the market at any given instant. Bachelier then proposed that, for fixed periods or *epochs* of time  $t$ , the probability of an observed price  $y_t$  can be calculated:

$$P(y_t) = \frac{\sum_{i=1}^t \begin{cases} 1, & \text{if } y_{i-1} \leq y_t \leq y_i \\ 0, & \text{otherwise} \end{cases}}{t} \quad (3.1)$$

and the probability of a price change for  $n$  epochs,  $\delta y_{t,n}$ :

$$P(\delta y_{t,n}) = \frac{\sum_{i=n}^t \begin{cases} 1, & \text{if } y_i - y_{i-n} \geq y_t - y_{t-n} \\ 0, & \text{otherwise} \end{cases}}{t - n} \quad (3.2)$$

Bachelier then proposed that the future price at epoch  $n$  could be predicted, by calculating  $\hat{y}_{t+n} = y_t + \delta y$ , from:

$$P(y_{t+n}) = \max \left( \left[ P(y_t + \delta y) P(\delta y) \right]_{\delta y = \min(\delta y_{i,n}) \text{ for } i=n, \dots, t}^{\max(\delta y_{i,n}) \text{ for } i=n, \dots, t} \right) \quad (3.3)$$

However, Bachelier provided no empirical research to support the theory, and the work was largely ignored for the first half of the twentieth century.

As the twentieth century moved on, it was discovered that Bachelier's assumption regarding the Gaussian distribution of price changes was incorrect. Analysis of the markets revealed the distribution to be *leptokurtic*, or fat-tailed, and of near-infinite variance.

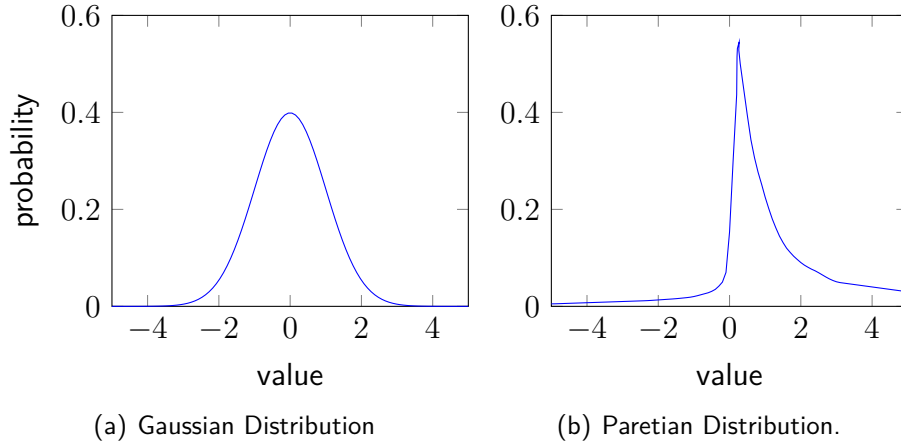


Figure 3.1: Gaussian and Paretian Distributions. (a) shows a Gaussian distribution, with mean  $\mu = 0$  and standard deviation  $\sigma = 1$ , while (b) shows a Paretian distribution, with location  $\delta = 0$ , scale  $\gamma = 1$ , skewness  $\beta = 0.75$  and characteristic exponent  $\alpha = 0.5$ .

Mandelbrot picked up on this in the early 1960's and, considering price changes symptomatic of random variables with infinite variance, suggested that the markets may hold to a Paretian rather than Gaussian distribution (Mandelbrot, 1963a, 1963b). Unlike a Gaussian distribution, which is defined by mean  $\mu$  and standard deviation  $\sigma$ , a stable Paretian distribution has four parameters: location  $\delta$ , scale  $\gamma$ , skewness  $\beta$  and characteristic exponent or *height of tail*  $\alpha$ , as in Fig. 3.1 (Fama, 1965b). This description of the market helped to explain its unpredictability, with causal structural features more likely to be hidden by noise in a Paretian model than a Gaussian model, and Paretian noise appearing to generate patterns that trigger price movements. Mandelbrot also noted that the most important feature of a Paretian distribution is its tail and that Fourier analysis of such systems will always reveal some sort of 'path', even though

there really is not one present. This means that there is no causal link between the detected frequencies and the price movements—they are instead artifacts of sample selection. Mandelbrot went on to state that, if market rates follow a Paretian distribution, then market movements will be more influenced by noise than systemic causes, and thus questioned the use of technical analysis (analysis of past patterns) by speculators. Quoting Keynes,

“are the patterns just historical curve-fitting and description, or are they inductive claims towards the future with reference to the past?” (Keynes, 1936)

Mandelbrot then suggested that using  $\ln(\delta y_t)$  instead of  $\delta y_t$  will transform the distribution from Paretian to Gaussian.

## 3.2 Efficient Market Hypothesis

At the same time, mathematicians were not the only ones pondering the nature of the world’s financial markets. Following empirical analysis of Mandelbrot’s work and building on thirty years of economic theory, Fama summed up the research of both in three heavily influential papers published between 1965 and 1970 (Fama, 1965b, 1965a, 1970), outlining the *Efficient Market Hypothesis* (EMH) and described it thus:

- The market is vast, with a buyer for every seller, and vice-versa.
- The market is made up of many intelligent participants in competition with each other.
- Market knowledge is reflected by participant behaviour, and the current price is an accurate estimation of the true price. However, there is always uncertainty as to the exact price, and so competition between participants causes the value to follow a *Random Walk* (RW), defined as

$$y_t = y_{t-1} + \xi \tag{3.4}$$

where  $y_t$  is the current value,  $y_{t-1}$  is the previous and  $\xi$  is *Independent and Identically Distributed* (IID) random noise. Independence means that there is no serial correlation present. Identically distributed means that it is time-invariant—that is, the shape of the distribution does not change over the series.

- If there is a systematic discrepancy between current and intrinsic value it will be found and nullified, returning the market to random walk.
- New information typically causes an overreaction but will return to random walk. The length and timing of this overreaction are also random.
- Market uncertainty is distributed through transaction costs, keeping the market efficient.

Therefore,

- Past prices hold zero information and thus are unable to assist in the prediction of future movements.
- Price changes are independent and hold to some probability distribution, such as Gaussian, Paretian, etc.
- A simple buy-and-hold trading strategy will be just as good as more sophisticated techniques, especially once development costs are considered.
- Fundamental analysis (seeking to understand price movements through economic factors such as supply-and-demand, interest rates, policy effects, etc.) is still useful; however, the more fundamental analysts there are, the more efficient the market will be.
- Due to the ‘triggering’ nature of the market distribution, it will often experience major shifts, and is thus risky to trade in. Because of the rapid rate of these changes, excessive slippage may prevent loss prevention measures such as STOP\_LOSS from being effective.

Fama continues on to state that if the EMH and Random Walk model are an accurate depiction of reality, then technical analysis is entirely without merit. However, the market is not necessarily unpredictable—it is just that any gains made through correct predictions will be offset by transaction costs.

### 3.3 Auto-Regression

In 1970, Box and Jenkins published the first edition of their now well-known book *Time Series Analysis: Forecasting and Control* (Box & Jenkins, 1970). This book made a significant impact in the field of time series prediction, with its key feature being a clear exposition of the *Auto-Regressive Moving Average* (ARMA) model. This model was Whittle’s extension of Yule’s original Auto-regressive (AR) model and is effectively a linear aggregate of random shocks around a fixed mean (Yule, 1927; Whittle, 1951). The ARMA model is defined as:

$$y_t = c + \epsilon_t + \sum_{i=1}^p \phi_i y_{t-i} - \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (3.5)$$

where  $c$  is a constant,  $\epsilon_t, \epsilon_{t-1}, \dots$  are IID white noise error terms of mean zero, and  $\phi_1, \dots, \phi_p$  and  $\theta_1, \dots, \theta_q$  are parameters such that the model is *stationary*, i.e., its stochastic properties, such as  $\mu$  and  $\sigma$ , do not change over time. The terms  $p$  and  $q$  refer to the number of auto-regressive and moving average terms included in the model respectively. Intuitively,  $p$  is how many lagged terms from  $y_t$  to  $y_{t-p}$  are included in the model, weighted by  $\phi$ ; and  $q$  is how many terms are included in the moving average model, which is a linear regression of observed error terms from  $y_t$  to  $y_{t-q}$ , weighted by  $\theta$ . Box and Jenkins detailed an iterative procedure for choosing the correct parameters  $\phi$  and  $\theta$ . Later in 1976 they added an integrating process, creating the ARIMA( $p,d,q$ ) model, which replaces  $y_i$  in Eq. 3.5 with the  $d$ -th first-order difference of  $y_i - y_{i-1}$  in order to ensure the first-order differential of the model was stationary.

Despite their popularity, these early auto-regressive models were found deficient when applied to foreign exchange movements. In their two papers

comparing the exchange rate movements of the 1970's with an assortment of available models, Meese and Rogoff found that none of the linear AR models could improve upon the Random Walk model (Meese & Rogoff, 1983a, 1983b).

### 3.4 Conditional Heteroscedasticity

Not everyone was convinced by Mandelbrot's Paretian model either. In the early 1970's, Clark proposed that instead of Paretian, the distribution of values may actually be *subordinate*—that is, the values themselves may be re-ordered, or directed by, another stochastic process (Clark, 1973). Clark showed that if the data was reorganised by traded volume then subsamples would be *lognormal* (that is,  $\ln(\delta y)$  normally distributed, where  $\delta y$  is the first-order differences of  $y$ ), with greatly reduced kurtosis, and the resulting distribution would be closer to the data than the equivalent Paretian distribution.

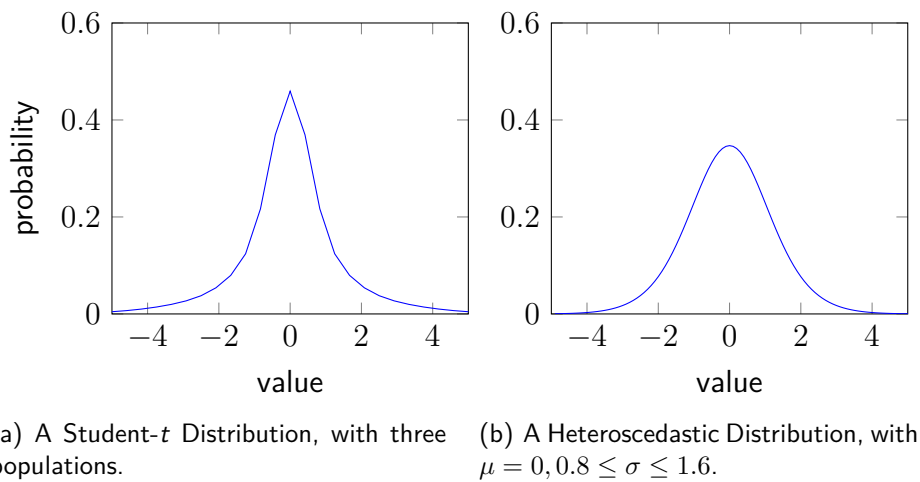


Figure 3.2: Student- $t$  and Heteroscedastic Distributions. The Student- $t$  distribution has three underlying distributions, with the same mean but different variances. The heteroscedastic distribution, however, moves from one variance to another over time, whilst maintaining a constant mean.

Similarly, Blattberg and Gonedes investigated the possibility that the leptokurtic nature of the markets indicated a Student- $t$  distribution (a continuous mixture of normal distributions with different scales), rather than

Gaussian distribution (Blattberg & Gonedes, 1974). A special case of this is *heteroscedasticity*, where the variance of a time-series is not consistent (White, 1980). Fig. 3.2 shows example Student- $t$  and heteroscedastic distributions.

Continuing on with heteroscedasticity, in 1982 Engle developed an *Auto-Regressive Conditional Heteroscedastic* (ARCH) model for time series prediction, defined formally as:

$$\begin{aligned} y_t | \psi_{t-1} &\sim N(x_t \beta, h_t), \\ h_t &= h(\epsilon_{t-1}, \dots, \epsilon_{t-p}, \alpha), \\ \epsilon_t &= y_t - x_t \beta, \end{aligned} \tag{3.6}$$

where  $x_t \beta$  is the mean of  $y_t$ , which is a linear combination of endogenous and exogenous variables included in the set  $\psi_{t-1}$ , with  $\alpha$  and  $\beta$  vectors of unknown parameters (Engle, 1982). Intuitively,  $\psi_{t-1}$  is the set of all data available at time  $t$ , including an auto-regressed variance  $h_t \alpha$  around  $x_t \beta$ . Engle proved that  $\alpha$  and  $\beta$  could be calculated independently, and proposed using an iterative approach of estimating  $\beta$  first with either *Ordinary Least Squares* (OLS) or *Maximum Likelihood Estimate* (MLE), using the resulting residuals to estimate  $\alpha$ , then iteratively refining  $\alpha$  and  $\beta$  by repeating these steps until the model reaches the desired level of conformity. Once the ARCH parameters have been estimated, the one-step-ahead prediction  $\hat{y}_{t+1}$  can be made using  $\psi_t$ .

Engle noted that ARCH was intended to solve the problem of a standard regression with a fixed mean that might potentially be missing some exogenous variables, misspecified or experience structural change over time—such as FOREX prices. However, whilst ARCH is useful for these scenarios, it would be better to fix the problems. Engle showed how the LaGrange Multiplier could be used to test for ARCH effects within a time series.

A few years later, Bollerslev introduced the GARCH model, which is the ARCH model generalized to allow varying mean as well as

variance (Bollerslev, 1986, 1987). Noting that ARCH had an arbitrary declining structure, Bollerslev posited that GARCH was to ARCH as ARMA was to AR, and was able to show that a simple GARCH(1,1) model outperformed an ARCH(8) model. The difference between ARCH and GARCH is a modified function for  $h_t$ ,

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i y_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i} \quad (3.7)$$

where  $p, \alpha_i \geq 0$ , and  $q, \alpha_0, \beta_i > 0$ . Due to the lagged conditional variances that enter into a GARCH( $p, q$ ) function, it could be considered a kind of adaptive learning mechanism; however, a generalized test for GARCH was not considered feasible.

Engle and Bollerslev worked together for a number of years on the ARCH family, alongside other mathematicians, with a number of variations (such as Exponential-ARCH/GARCH, ARCH/GARCH-M, etc.) proposed and tested. Despite the advances made, ARCH/GARCH models struggled to sufficiently model the leptokurtosis in FOREX movements (Bollerslev, Chou, & Kroner, 1992). Hsieh suggested that ARCH may be too rigid in its approach to heteroscedasticity, and that a more flexible treatment may be required (Hsieh, 1991).

### 3.5 Nonlinearity

Around the same time that conditional heteroscedasticity was being investigated, researchers were also starting to question the use of linear models for FOREX prediction. Tong and Lim suggested that if a nonlinear model was adopted, then one-step-ahead predictions should be possible, though they acknowledged the model should only be considered superior if overall predictions were more accurate than the linear model (Tong & Lim, 1980). Discussing causes of nonlinearity in a time series, they suggested that any models developed would likely need to include exogenous data and would



therefore need to be generalisable to the multi-variate case. Due to the presence of oscillators and fractional cycles, a nonlinear system would not necessarily require continuous input. Expanding Tong's linear *Threshold Auto-Regression* (TAR) model, they then introduced a number of non-linear models: the *Self-Exciting TAR* (SETAR), *Closed-Loop TAR* (TARSC) and *Open-Loop TAR* (TARSO) models.

For all the variations proposed, not all of the included peer-reviews were convinced by Tong and Lim's arguments, with several finding inconclusive evidence of predictive superiority. Other researchers were also unable to improve upon the linear methods in this fashion, with Scheinkman and LeBaron noting that the evidence for non-linearity would appear to contradict the EMH, but no nonlinear method had been shown to outperform it out-of-sample (Scheinkman & LeBaron, 1989). In their analysis of ten exchange rates over fourteen years, Diebold and Nason were unable to find any nonlinearities exploitable for improved prediction (Diebold & Nason, 1989). And Meese and Rose came to the conclusion that:

“We do not deny that non-linear effects are important in understanding even moments of exchange rate processes. However, we do conclude that incorporating nonlinearities into existing structural models of exchange rate determination does not at present appear to be a research strategy which is likely to improve dramatically our ability to understand how exchange rates are determined.” (Meese & Rose, 1991)

Hsieh noted that empirical evidence suggests exchange rates are linearly uncorrelated, but may be non-linearly dependent, and thus not IID. Instead, exchange rate changes are nonlinear stochastic functions of their own past, depending upon third-order moments, which is consistent with conditional heteroscedasticity (Hsieh, 1989a). Hsieh does suggest, however, that the observed nonlinear dependence may be because exchange rate changes are deterministic processes that *look* random (Hsieh, 1989b). In other words, perhaps the nonlinearities are *chaotic* rather than structural.

### 3.6 Chaos Theory

Whilst hints towards it had been present in the works of Cantor, Poincaré and other early mathematicians (Rosser, 2008), it is generally agreed that Chaos Theory was birthed in Lorenz's 1963 paper, *Deterministic Nonperiodic Flow* (Lorenz, 1963). Seeking to adapt a simplification of Saltzman's thermal equations for computation, Lorenz observed that instead of converging to a consistent periodic cycle, the model exhibited aperiodic cycles of unpredictable length. Discussing these results, Lorenz introduced the concept of the unstable *trajectory*, where the path followed by a system through its phase space depends on the system's current state, and was thus able to show that such a system exhibits a *Sensitive Dependence to Initial Conditions* (SDIC). A system is considered *chaotic* if it exhibits both SDIC and *topological transitivity*, which is where trajectories eventually overlap themselves within phase space, increasing topological density (DeVaney, 1992). Lorenz notes that, unless all governing variables are known exactly, long term prediction of a chaotic system is impossible. A visual representation of Lorenz's example, the famous 'butterfly', is shown in Fig. 3.3.

Researchers were not slow to apply Chaos Theory to financial time series. In analysis of the markets, Mandelbrot was able to show that they exhibit *Fractional Brownian Motion* (fBm), evidenced by the fact that Fourier analysis revealed the fundamental frequencies to be proportionate to the sample size (Mandelbrot & Ness, 1968; Mandelbrot, 1972). Mussa picked up on this, stating that while there is evidence that exchange rates may be Paretian, that same evidence could also be used to suggest that the market was anticipating some form of future exogenous input (Mussa, 1979). There is also evidence that the exchange rates followed moving average trends, though at a period of one month or more, ninety percent of the exchange rate movements were unexpected, and no theory so far proposed had been able to sufficiently explain the foreign exchange market. Granger and

**LORENZ ATTRACTOR WITH  
CHAOTIC PROPERTIES**

$$\begin{aligned} \mathbf{x}' &= \sigma(\mathbf{y} - \mathbf{x}) & \sigma &= 10 \\ \mathbf{y}' &= \mathbf{x}(\rho - \mathbf{z}) - \mathbf{y} & \rho &= 28 \\ \mathbf{z}' &= \mathbf{x}\mathbf{y} - \beta\mathbf{z} & \beta &= 8/3 \end{aligned}$$

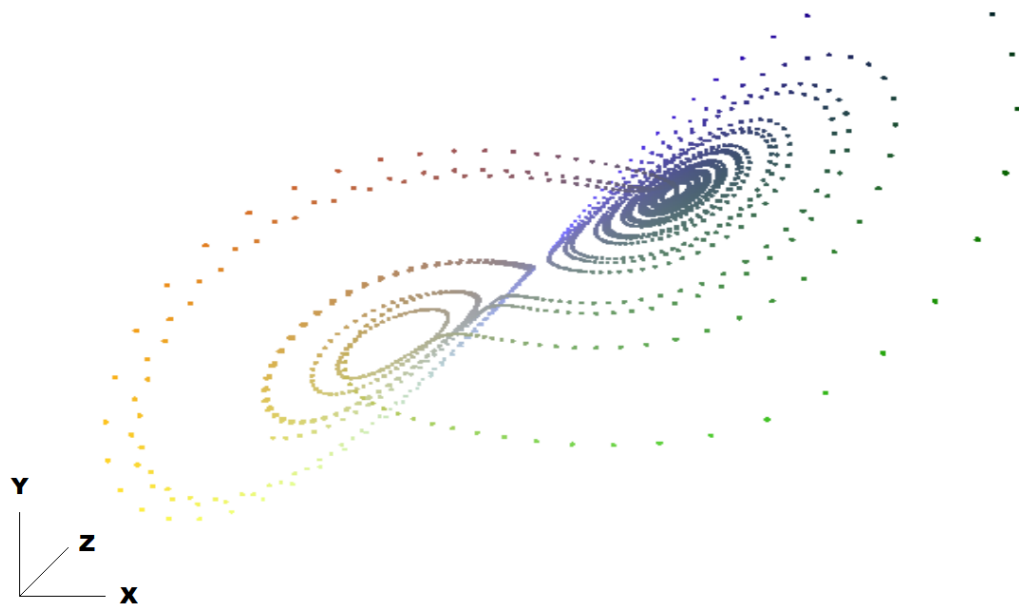


Figure 3.3: Visualization of Lorenz's System

Joyeaux showed how to apply fractional differencing to ARMA models, in order to induce long-term memory (G. Granger & Joyeux, 1980). Parallel and independent to Granger and Joyeaux, Hosking was doing the same for ARIMA models (Hosking, 1981). However, it was not until the work of Packard et al. and Takens that significant applications became possible.

In 1980, Packard et al. showed that time-variant data could be used instead of multiple dimensions to measure the chaotic properties of data (Packard, Crutchfield, Farmer, & Shaw, 1980). This allows the 'reconstruction' of exogenous events through analysis of only a single time series (such as FOREX movements). Their method involved the development of a *return map*, either using time-differenced or derivative values to construct a series of vectors of the desired dimensionality  $N$ , e.g.:

$$\mathbf{Y}_t = \{y_t, y_{t-\tau}, y_{t-2\tau}, \dots, y_{t-N\tau}\}, \text{ OR} \quad (3.8)$$

$$\{y'_t, y''_t, \dots, y^{(N)}_t\}$$

which can be used to calculate the Lyapunov exponent (a measure of

trajectory divergence) of the system, through an iterative process. However, this is only possible if  $\tau \ll I/\Lambda$ , where  $I$  is the degree of accuracy and  $\Lambda$  is the sum of all positive exponents. Takens added that the maximum embedding dimension  $m$  that could be calculated is defined by  $m > 2D_A$ , where  $D_A$  is the dimension of the system's attractor. Thus, if one can calculate the embedding dimension, the maximum number of dimensions required to map the system can be estimated. This allows an estimation of the *Hausdorff Dimension* ( $D_H$ ), as  $D_H \leq D_A$ , and provides a measure of a system's chaos, even if the exogenous variables themselves are not available (Takens, 1981).

Another metric, the *Hurst Exponent* ( $H$ ), can also be calculated from the Hausdorff dimension: for a self-affine system  $H = N + 1 - D_H$ . A Hurst exponent of 0.5 indicates Brownian motion, with lower values indicating increased noise and higher values indicating increased periodicity.

Grassberger and Procaccia noted, however, that Packard et al.'s method of computation is particularly difficult for higher-order chaos ( $D_H > 2$ ) and introduce their own measure, the *Correlation Integral* (CI), defined as:

$$C(\mathbf{Y}, \gamma) \equiv \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{i,j=1}^N \theta(\gamma - \|\mathbf{Y}_i - \mathbf{Y}_j\|) \quad (3.9)$$

where  $\mathbf{Y}$  is a long time series of  $N$  points with fixed time increment  $\tau$ ,  $\gamma$  is a threshold distance and  $\theta$  is the Heaviside function  $H = \frac{d}{dx} \max\{x, 0\}$  (Grassberger & Procaccia, 1983a). They argue that, because the Correlation Integral takes not only the attractor's shape, but also its regional influence into account, which the Hausdorff dimension does not, the Correlation Integral is a more relevant measure. As a bonus, it is significantly easier to calculate.

For small  $\gamma$ ,  $C(\gamma) \approx \gamma^v$ , where  $v$  is the correlation exponent. Generally speaking, if  $v < N$ , the system contains deterministic chaos, whereas if  $v = N$ , either the system is random noise, or more dimensions are required.

If the system contains deterministic chaos, the implication is that with enough historical data points, predictions may be inferred from past behaviour (Grassberger & Procaccia, 1983b).

When considered for application to FOREX prediction, it is necessary to account for the fact that exchange rate movements are extremely noisy, with some currency pairs having a full range of only a few hundred discrete values and thus a low signal-to-noise ratio. Wolf et al. noted that there are two categories of noise within experimental data: statistical and catastrophic (Wolf, Swift, Swinney, & Vastano, 1985). Statistical noise includes problems such as jitter and quantization error. Catastrophic noise is either from too low an embedding dimension or from too little data relative to the complexity of the attractor. In order to minimize statistical noise, it is necessary to ensure successive points are sufficiently greater than the noise level apart (as noted by Packard et al.). Preprocessing the exchange rate data can also help, though this introduces its own rounding errors and its use therefore may be moot. These same issues occur in the processing of the data, where computer rounding errors potentially have a long-term impact (Eckmann, Kamphorst, & Ruelle, 1987).

Another issue is the dearth of available data to analyse. Whilst the attractor does not have to fully evolve within the observed system, a significant number of points are required to detect it. Hsieh noted that physical scientists typically use 100,000 or more data points to detect low dimensional chaotic systems, but financial economists have substantially fewer points available—the largest datasets at the time consisted of approximately 5,000 daily observations. In this scenario, an embedding dimension of 10 only allowed for 500 non-overlapping histories and, due to phase space volume increasing exponentially with dimension, it is difficult to argue that 500 points ‘fill up’ a 10-dimensional space (Hsieh, 1991).

In 1987, Farmer and Sidorowich introduced the *k-Nearest Neighbour* (*k*NN) algorithm to predict chaotic time series, with application to systems that lack

‘first principles’, such as financial time series (Farmer & Siderowich, 1987).

For a given time series  $Y$ , with interval  $\tau$  and  $t$  observations:

1. Find  $k$  states that minimize  $\|y_t - y_i\|, i \in 1, \dots, t - 1$ ; that is, find the  $k$  states closest in phase space to  $y_t$  and assign to a set  $\psi_k$ .
2. Build a map from each identified state,  $y_i \in \psi_k$ , to its successor,  $y_{i+1}$ .
3. Calculate the interpolation of the first-order differences in the mapping,  $P$ .
4. Predict  $y_{t+1} = y_t + P$ .

They note that, while the  $k$ NN may be useful for predicting time series, it could also be used to quickly test for the presence of chaos. Specifically, should the algorithm exhibit any measure of success, that could be taken as a positive indication of low-level chaos. Given the right data structures, this algorithm is fast and highly parallelisable.

By 1989, two distinct groups had emerged in exchange rate prediction studies—those that used stochastic nonlinear time series and those that used deterministic chaotic time series (Diebold & Nason, 1989)—and thus the stage was set for the introduction of *Artificial Neural Networks* (ANNs) into FOREX prediction.

### 3.7 Neural Networks

Despite having been invented almost forty years prior, ANNs had largely been ignored by FOREX researchers. This was, in part, due to Minsky and Papert’s pessimistic appraisal of their performance versus complexity, and the lack of available processing power (Rosenblatt, 1958; Minsky & Papert, 1969; Rumelhart, Hinton, & Williams, 1986). However, with the advent of Chaos Theory and increasing CPU power, researchers again began to look at them as an option, especially since early networks were non-parametric and capable of deriving rules from data without prior analysis.

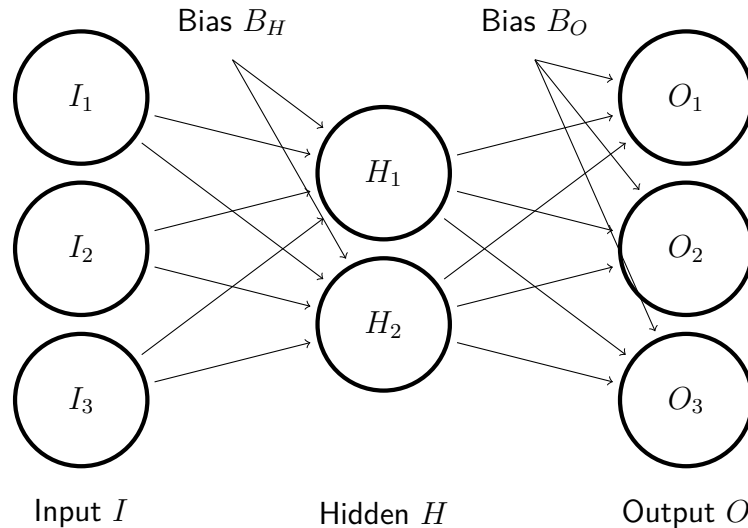


Figure 3.4: A Simple ANN

At the heart of an ANN is the *neuron*, which is effectively a function that transforms a series of input values into an output value. The transform is called an *activation function*, and can include threshold, linear, quasi-linear or squashing, gating or sigma-pi functions (Williams, 1986). A commonly used neuron is the *perceptron*, which calculates the sum of weighted inputs and activates its output if a certain threshold is met (Rosenblatt, 1958).

A basic ANN is made up of one or more neurons, arranged in some way between an input and an output. A common approach, the *Multi-Layered Perceptron* (MLP), is shown in Fig. 3.4. In their 1986 book *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Rumelhart and McClelland et al. presented the then state-of-the-art in ANNs, focusing on *how* neural networks learned to solve problems (Rumelhart & McClelland, 1986). In this book, Rumelhart, Hinton and Williams noted three training methods already proposed: competitive learning, pre-calculated values, or the development of a learning procedure capable of learning the internal state, such as Boltzmann machines (Rumelhart et al., 1986). However, instead of taking these approaches, they chose a new route: the development of a generalized delta rule that applied only to local computations. They called their ANN a *Layered Feed-Forward* network, and it used differentiable nonlinear activation

functions. Because the activation functions were differentiable, the output of the machine could be compared to the expected value and the error back-propagated through the network to update the network weights. The efficacy of the back-propagation is controlled by a parameter known as the *learning rate*, and ANNs may need to be trained repeatedly on the same input/output data in order to ‘learn’ the appropriate weightings.

Mathematically, if an ANN is defined as:

$$\mathbf{\Pi}(a, b, c) = [\mathbf{I}^a, \mathbf{H}^b, \mathbf{O}^c], \{a, b, c \in \mathbb{N}\} \quad (3.10)$$

with  $a$  inputs  $\mathbf{I}$ , a hidden layer  $\mathbf{H}$  with  $b$  neurons, and an output layer  $\mathbf{O}$  with  $b$  inputs and  $c$  outputs, then the forward-propagation algorithms could be defined thus:

$$\begin{aligned} \mathbf{H}_i &= \Gamma \left( \sum_{i=1}^a \mathbf{I}_i \mathbf{w}_i \right), \\ \mathbf{O}_i &= \Gamma \left( \sum_{i=1}^b \mathbf{H}_i \mathbf{w}_i \right) \end{aligned} \quad (3.11)$$

where  $\mathbf{w}_i$  represents each neuron's internal weighting, and  $\Gamma$  the activation function. If the network's output error is then defined as

$$\boldsymbol{\epsilon} = \mathbf{O} - \mathbf{Y} \quad (3.12)$$

the error correction algorithms could be defined thus:

$$\mathbf{w}_i = \mathbf{w}_i + \alpha \Gamma' \left( \sum_{i=1}^c \boldsymbol{\epsilon} \mathbf{w}_{O_i} \right) \quad (3.13)$$

where  $\alpha$  is the learning rate,  $\Gamma'$  the derivative of the activation function  $\Gamma$ , and  $\mathbf{w}_O$  the weights from the output layer.

Whilst immediately obvious that the activation function needs to be differentiable, Rumelhart et al. also recommended that the activation



function be nonlinear and used the logistic activation function

$$\Gamma(\mathbf{L}) = \frac{1}{1 + e^{-\left(\sum_{i=1}^n \mathbf{w}_i \mathbf{L}_i + \theta_i\right)}} \quad (3.14)$$

for layer  $\mathbf{L}$ , where  $\theta$  is a bias akin to a threshold. They recommend the weight  $\mathbf{w}$  should initially be randomized and note that adjusting the learning rate  $\alpha$  requires some experimentation. If the learning rate is too high the system becomes unstable, but if set too low the system will take a long time to converge to a stable solution. In order to combat excessive oscillation, they suggested that the back-propagation algorithm could include a weighted regression of previous errors.

Following this significant contribution, they then pointed towards two future topics of interest, *Recurrent Neural Networks* (rNN) and *Sigma-Pi units*. Giles and Maxwell suggested that ANNs may benefit from the addition of pre-programmed neurons that contain *a priori* information about the problem the network is set to solve (Giles & Maxwell, 1987). Acknowledging the potential of ANNs to solve so-called fuzzy problems, Lippman tested six different ANNs on a set of classification tasks, with seemingly decent results. However, despite no known applications outside of research, early results demonstrated ANNs' performance capability in massively parallel environments (Lippmann, 1987).

Despite their success in other areas, such as speech and handwriting recognition, image classification, etc., ANNs failed to find favour amongst FOREX researchers of the 1990's. Several authors used Chaos theory and data-driven learning algorithms in attempts to predict FOREX movements, however, most concluded that ANNs were inferior to Random Walk. The authors who suggested otherwise were typically found to have invalidated their testing through either sample tampering or snooping (White, 2000). One reason for these failures was that optimizing ANNs required a lot of trial and error (Kaastra & Boyd, 1996). Walczak noted there is a balance, with

too much data being just as detrimental to ANN performance as not enough (Walczak, 2001). And Zhang and Hu were able to demonstrate that ANNs' performance improved with increasing input nodes, however they were still unable to outperform the Random Walk model (Zhang & Hu, 1998).

Another reason for this failure may have been that there was conflicting evidence as to whether the FOREX market was indeed chaotic. Researchers were able to show increasing levels of chaos in simulated markets, as participant ratios changed between fundamentalists, technical analysts and trend followers. However, whilst there was evidence that these ratios were changing in the real FOREX market, a survey of 90's literature covering chaos in the FOREX market shows no concrete evidence of chaos (Jeanne & Masson, 2000; Ausloos, 2000; Farmer & Joshi, 2002). According to Barnett and Serletis:

“A survey of '90s papers reveals conflicting evidence of chaos, some evidence of non-linearity, and [no evidence movements are] IID. However, the tests used were not suitable to the task. The alternative theory, that market movements are caused by 'Random Shocks' seems implausible, yet it is more readily accepted than chaos. The problem is simply that testing for chaos within the wider scope of the economy is beyond our current capabilities, and testing for chaos within the financial time series proven inconclusive.” (Barnett & Serletis, 2000)

Despite the disappointing results, researchers remained optimistic about the future of FOREX prediction, with the emergence of several new approaches, such as *Wavelet Decomposition*, *Particle Swarm Optimisation* (PSO) and *Support Vector Regression* (SVR) being incorporated into future research, along with the rise of hybrid predictors.

### **3.8 Wavelet Decomposition**

Developed independently in the fields of pure mathematics, physics and engineering, the wavelet transform (Eq. 3.15) is a tool that cuts data,

functions or operators into different frequency components, much like the windowed Fourier transform (Daubechies, 1992). Unlike the windowed Fourier transform, however, wavelets have a time-width relative to their frequency spectrum—that is, the higher the frequency, the wider the bandwidth of the wavelet. Wavelets are thus more ‘zoomable’ than Fourier windows and as a result are better able to analyse short-lived phenomena.

$$T_{m,n}^{wav}(f) = a_0^{-m/2} \int dt \mathbf{Y} \psi(a_0^{-m}t - nb_0) \quad (3.15)$$

In application to FOREX prediction, this means short-term events like reactions, spikes and corrections may be better specified as a wavelet input rather than the raw time series or its derivatives. This is particularly applicable if the market contains short self-similar movements, where a group of participants are all using the same algorithm or trading strategy.

Several researchers have used wavelet decomposition either to show the multifractal nature of exchange rates, or as part of their predictor input (Arnéodo, Muzy, & Sornette, 1998; Kantelhardt et al., 2002; Pal, Rao, & Manimaran, 2014; Stošić, Stošić, Stošić, & Stanley, 2015; Shin & Han, 2000; Bekiros & Marcellino, 2013; Bagheri, Peyhani, & Akbari, 2014). Recently, He, Chen and Tso used a modernised version, *Variational Mode Decomposition* (VMD) to outperform an ARMA model (He, Chen, & Tso, 2018).

### 3.9 Particle Swarm Optimization

Introduced in 1995 by Kennedy and Eberhart, PSO was born out of the synthesis of swarm theory, genetic algorithms and evolutionary programming, and was designed to optimize continuous nonlinear functions with special application to neural network weights (Kennedy & Eberhart, 1995). According to them, a particle swarm should be multidimensional and collision-free, with each particle representing a location  $\mathbf{x}$  and velocity  $\mathbf{v}$ .

The swarm trends towards an attractor  $\mathbf{\Lambda}$ , but allows for the introduction of a new attractor  $\mathbf{L}$ , and should exhibit the following properties:

1. The population should be able to carry out simple time and space, or *proximity*, computations.
2. The population should be able to respond to quality factors.
3. The population response should exhibit *diversity*, not committing itself to an excessively narrow outcome.
4. The population should exhibit *stability* and *adaptability*, not being too quick to change yet not averse to it either.

A simple form of PSO is given for each particle,

$$\mathbf{v}_p = \mathbf{v}_q + 2\xi(\mathbf{p}_{best} - \mathbf{x}) + 2\zeta(\mathbf{\Lambda} - \mathbf{x}) \quad (3.16)$$

where  $\mathbf{v}_p$  is the particle's velocity,  $\mathbf{v}_q$  its previous velocity,  $\mathbf{p}_{best}$  the particle's best known location,  $\mathbf{\Lambda}$  the swarm's best known location, and  $\xi$  and  $\zeta$  random values between 0 and 1. For application to neural network optimisations, the swarm has as many dimensions as it has network weights, with the number of particles determined by external computational requirements.

Perhaps due to the significant computational requirements of a large swarm, PSO was not applied to FOREX prediction until recently, with examples found in the works of Sermpinis, Theofilatos, Karathanasopoulos, Georgopoulos, and Dunis (2013), Bagheri et al. (2014), Ravi, Pradeepkumar, and Deb (2017), Pradeepkumar and Ravi (2017) and Hajizadeh, Mahootchi, Esfahanipour, and Kh (2019).

### 3.10 Support Vectors

Also in 1995, Cortes and Vapnik introduced the *Support Vector Network* (SVN) (Cortes & Vapnik, 1995). In an SVN, the input vectors are mapped into a high-dimensional space  $\mathbf{Z}$  through a predefined non-linear mapping or

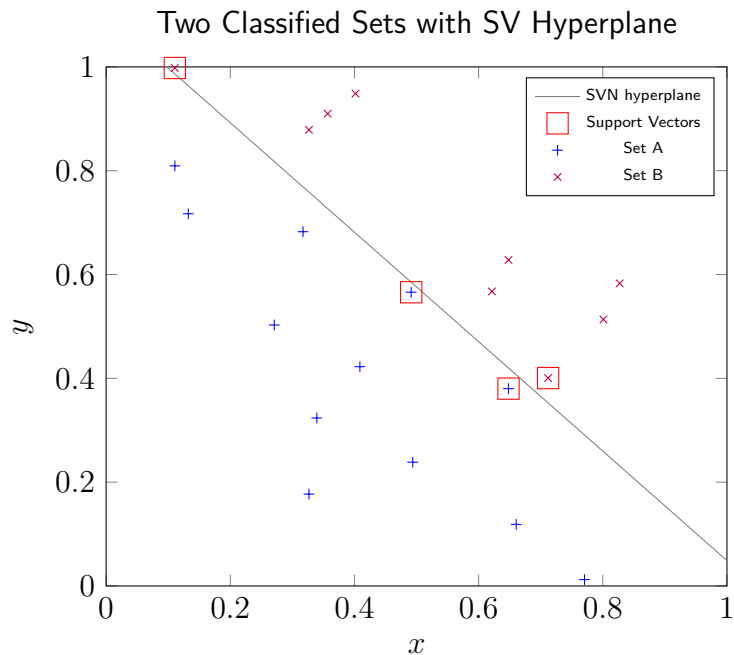


Figure 3.5: Support Vector Example. Whilst there are many data points within the dataset, only the boundary values between the sets are required to calculate the separating hyperplane.

*kernel*. Then, within this space, a linear decision boundary  $\mathbf{x}$  is calculated such that there is maximum separation between classifications. That is, for a set of labelled training patterns

$$(y_1, \mathbf{x}_1), \dots, (y_\ell, \mathbf{x}_\ell), y_i \in \{-1, 1\} \quad (3.17)$$

there exists  $\mathbf{w}$  and  $b$  such that

$$\begin{aligned} \mathbf{w} \cdot \mathbf{x}_i + b &\geq 1 & \text{if } y_i = 1 \\ \mathbf{w} \cdot \mathbf{x}_i + b &\leq -1 & \text{if } y_i = -1 \end{aligned} \quad (3.18)$$

with the optimal hyperplane  $\mathbf{w}_0 + b_0 = 0$  being that which maximises the distance between inputs, whilst minimizing incorrect classification or *loss function*. For the calculation of this boundary, it is not necessary to use the full dataset, but only the subset along the boundaries, or *support vectors* (see Fig. 3.5). However, finding the optimal subset of support vectors and their hyperplane is an NP-hard problem. Aimed at classification, this model was

extended to include *Support Vector Regression* (SVR) the following year, with the family of models called *Support Vector Machines* (SVMs) (Drucker, Burges, Kaufman, Smola, & Vapnik, 1996).

First applied in other financial time series prediction with promising results (e.g., Tay and Cao (2001), Kamruzzaman and Sarker (2004), Kim (2003) and Lu, Lee, and Chiu (2009)), SVMs were found to out-perform neural networks for FOREX prediction and are often used as part of hybrid predictors (Ince & Trafalis, 2006; Yu, Wang, & Lai, 2009; Huang, Chuang, Wu, & Lai, 2010).

### 3.11 Hybrid Predictors

Aside from immediately after the 2008 global financial crisis, every year since the turn of the millennium saw more FOREX predictors published than the previous year. A large number of these predictors are *hybrids* that combine various predictive techniques already discussed in different ways. Giles, Lawrence, and Tsoi (2001) used grammatical inference to develop automata for use as input to their ANN. Nag and Mitra (2002) used evolutionary programming to *evolve* a population of ANNs. Chen and Leung (2004) used *Bayesian Vector Autoregression* (BVAR) for input into their RNN. Yu, Lai, and Wang (2008) wrapped two *Radial Basis Function Neural Networks* (RBF-NN) around a *Generalized Variance* (GVAR) minimization function. Majhi, Panda, and Sahoo (2009) chained, or cascaded, simple single-hidden-node ANNs together, using a shared error function to update all the ANNs simultaneously. Khashei, Bijari, and Ardali (2009) used a fuzzy regression to de-noise the input time series into their neural network, and Ni and Yin (2009) combined *Self-Organising Maps* (SOM), SVR and technical indicators to predict FOREX movements.

Obviously, developing hybrid predictors for FOREX prediction is a verdant area of research, with the aforementioned examples but a handful out of hundreds. Therefore the question has to be asked, why are there so

many different-yet-similar papers being published? The reason for this may lie in the Efficient Market Hypothesis and Chaos Theory.

### 3.12 Market Efficiency and Chaotic Behaviour

According to the Efficient Market Hypothesis, all currently available information is eventually reflected in the current exchange rates. This information includes knowledge of prediction techniques—once a predictor is published or otherwise widely used, it is subsumed into the competitive market, with the intelligent participants anticipating its use and thereby nullifying its effectiveness. At the same time, Chaos Theory suggests that the widespread use of common prediction patterns may cause multifractal behaviour through self-similar feedback loops. As a result, the market will appear predictable for indeterminate periods of time, during which new predictors and methods may be developed that appear to promise excess returns. Again, however, these new methods themselves are subsumed into the market, and the merry-go-round that is the hope for excess profits continues. For those participants that find a particular method profitable for a time, this is always done with the implicit acceptance of the possibility for catastrophic losses in the future. Whilst not all researchers agree with this view, the evidence of historical analysis would seem to support it.

Analysing the exchange rates of the 1920s, 1930s and 1970s, Mussa concluded the the best predictor was the current price, “and that is not a good predictor at all” (Mussa, 1979). In 1986, Sweeney published a popular article, *Beating the Foreign Exchange Market*, making the claim it was easy to make excessive returns through speculation and providing a vague and incomplete algorithm for others to use (Sweeney, 1986). However, that same year saw Frankel and Froot argue that the US dollar was on a ‘bubble path’, with a feedback loop of ever-increasing prices being caused by short-term speculation, such as Sweeney’s (Frankel & Froot, 1986). They noted that

there was a startling difference between short and long-term positions taken by market participants and predicted a significant correction in the near future. The advent of Black Monday on 19 October, 1987 saw this prediction correct, and long-term fundamentalists were thus vindicated in their assessment of short-term speculation.

Since then, positions for and against both the EMH and chaotic market theories have oscillated back and forth in the academic corpus. As has already been noted, most early researchers struggled to beat Random Walk, including Meese and Rogoff (1983a), Alexander and Thomas (1987) and Diebold and Nason (1989). Engel and Hamilton offered hope in the early 90s, when they argued that predictions were possible in the long term (three plus years) (Engel & Hamilton, 1990), and Brock, Lakonishok, and LeBaron (1992) suggested short-term profitability might be possible through the use of simple trading rules—though Levich and Thomas showed that such rules tended to decline in profitability after a while (Levich & Thomas, 1993). Bollerslev et al. (1992) presented the *Information Processing Hypothesis*: that exchange rate changes cascade through others, though Engel and Hamilton did not find multivariate analysis helpful for prediction. Meese and Rose found that accounting for conditional heteroscedasticity brought no notable improvement over Random Walk, and Hsieh found the same with chaos-based auto-regressions (Meese & Rose, 1991; Hsieh, 1991). And in the late 1990s, Allen and Karjalainen also were unable to beat Random Walk with their genetic algorithm (Allen & Karjalainen, 1999). It should be noted that long-term predictability is not at odds with the EMH, as defined by Fama, though many researchers failed to appreciate that.

In 2000, Yao and Tan published their paper, *A Case Study on using Neural Networks to Perform Technical Forecasting of FOREX*, in which they claimed the Random Walk model was outmoded and irrelevant due to a chaotic market implying non-efficiency (Yao & Tan, 2000). At the same time, they also claimed their out-of-sample ANN outperformed an in-sample



ARIMA, on the basis of profits made. Despite a number of critical issues, it has since become the most cited paper in neural networks and foreign exchange, for a rather strange reason: a later researcher misquoted it as beating Random Walk, and the academic community has run with that since.

Since then, several others have also claimed to have beaten Random Walk: Leung, Che, and Daouk (2000), with a *General Regression Neural Network* (GRNN); Clarida, Sarno, Taylor, and Valente (2003), with a linear *Vector Equilibrium Correction Model* (VECM) targeting a modified, *Risk Neutral Random Walk*; Sermpinis et al. (2013), with their adaptive RBF-PSO network; and Shen, Chao, and Zhao (2015), who used a *Deep Belief Network* (DBN). At the same time, others have also failed to beat Random Walk: Kilian and Taylor (2003), with their ESTAR algorithm, and Mendes, Godinho, and Dias (2012), who could only beat it sans trading costs, are notable examples. However, positive claims should be treated with some skepticism: Rossi noted in 2013 that most papers who claimed to beat Random Walk actually beat random *with drift* (RWD), and the simple Random Walk remained the toughest benchmark to beat (Rossi, 2013).

Despite years of intense scrutiny, there is still a notable lack of empirical evidence for improved FOREX prediction. There is also a notable lack of evidence that the predictors published work outside of their research scope. However, there is evidence that managed risk may enable short term profits, at the risk of courting disaster. In the absence of evidence otherwise, it is not unreasonable to conclude that the FOREX market is indeed efficient.

### 3.13 Concluding Remarks

In the 120 years since Bachelier, thousands of papers have been published in the area of financial time-series prediction. This chapter has covered the transition from simple, stochastic models to the state-of-the-art in machine learning, as applied to FOREX speculation—such as autoregression, neural

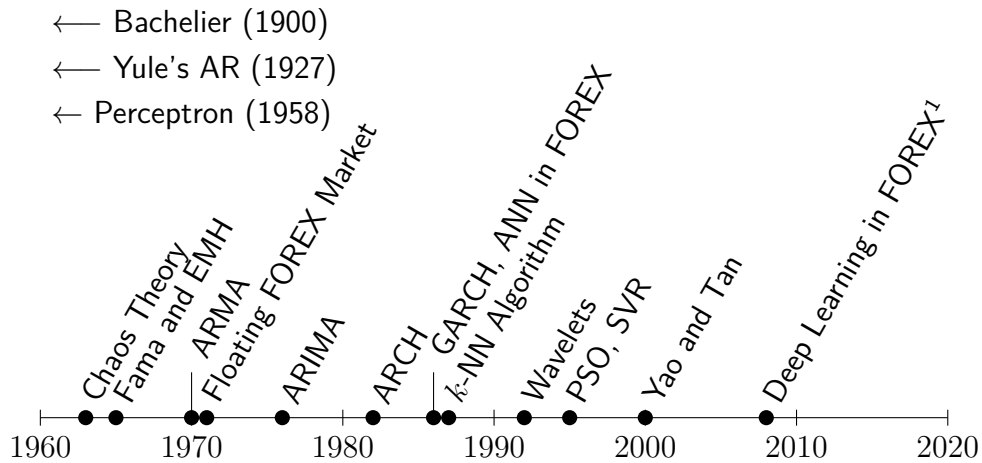


Figure 3.6: A timeline of the development of technologies used in FOREX prediction. <sup>1</sup> Deep Learning's roots can be found in the work of Ivakhnenko (1966), who developed the *Group Method of Data Handling* (GMDH) algorithm to calculate an extremely high-order regression-type polynomial (Farlow, 1981). However, it was not until 2006 that it was applied to FOREX prediction (by Hinton, Osindero, and Teh (2006)); and more recently with a renewed interest from 2015 onwards. However, whilst the published results *look* good, close examination of the published papers for FOREX reveals them to be lacking in academic rigour—the methods used for benchmarking are heavily biased towards the deep-learning algorithms.

networks, support vector machines and hybrid predictors (see Fig. 3.6). Other technologies also exist in the literature, such as genetic algorithms, reinforcement learning, fuzzy logic and deep learning, but have not been explored here due to their relative paucity in the FOREX literature. Research targeted at Stock Market prediction has also been left out-of-scope, except for key background elements, despite the obvious correlation. And whilst not all of the research mentioned was intended for FOREX speculation, it has all been influential in shaping the current state of research.

Given the plethora of predictive technologies available and the continuing impact of Chaos theory and the EMH, the question naturally arises—which predictor is the best? To answer that, another question needs to be asked: how are predictors compared? Is there an objective measure, and if so, is this evidenced in the literature? What does it mean when a researcher claims they have ‘beaten’ another method, such as Random Walk, anyway? These latter

questions are the primary concern of this thesis, and its investigation of them is presented in the next chapter.

# Chapter 4

## Measuring Foreign Exchange

### Predictions: Common Metrics

A targeted search on Google Scholar reveals in excess of 24,000 papers published in the area of financial time-series prediction. In order to obtain a representative subset of this vast array of literature, the following selection procedure was used:

1. The search results were first restricted to peer-reviewed articles available in the *Scopus* database, for the search terms “(forex OR ‘foreign exchange’) AND (‘machine learning’ OR ‘artificial intelligence’ OR ‘time series’)”, with just over 4,000 abstracts, plus references, acquired.
2. All abstracts were read, with false hits removed, leaving approximately 2,500 relevant papers.
3. All references were then collated and counted, in order to capture papers not available in the Scopus database, building a second list of cited references.
4. Papers within each list were given a score based on citations-per-year, with a bias towards more recent papers. Papers without citations were removed from the Scopus list. This process left 944 main articles and 566 referenced works.

5. The 100 highest-scoring and available works from each list were then obtained.

Due to overlap between lists, this processes yielded a total of 177 articles for analysis, of which 59 were background theory and 118 presented new predictors or models. This chapter focuses on the analysis of this representative subset of the FOREX corpus (hereafter referred to as just *the corpus*), and looks at how the researchers involved measured the predictive performance of their proposed model or algorithm.

## 4.1 Datasets

Within the corpus, over forty different Symbols were used for analysis, across different time periods and with varying frequencies. Fig. 4.1 shows the relative frequency per-Symbol, and Fig. 4.2 shows the amount of overlap between time periods used for papers whose dataset included daily GBPUSD values. It is readily apparent from a visual inspection that there is limited correlation between the time periods selected, and numerical analysis confirms this:

Suppose that for any given two time periods,  $A$  and  $B$ , the overlap  $O$  is defined:

$$O = \max\left(0, \frac{\min(A_{end}, B_{end}) - \max(A_{start}, B_{start})}{\max(A_{end}, B_{end}) - \min(A_{start}, B_{start})}\right) \quad (4.1)$$

Analysis of the time periods used for daily GBPUSD predictions (Fig. 4.2) reveals only two papers with datasets overlapping by more than 80%—Pradeepkumar and Ravi (2017) and Ravi et al. (2017)—with the obvious link being the same authors in the same year. Since there is no correlation between the datasets used for daily GBPUSD, the single most common data format, it is not unreasonable to suggest that extrapolation to the wider corpus is likely to find the same situation. As a result, direct comparison between published results, and therefore predictors, would seem to be next to impossible without re-implementation of the proposed algorithms. However, this might not necessarily be the case.

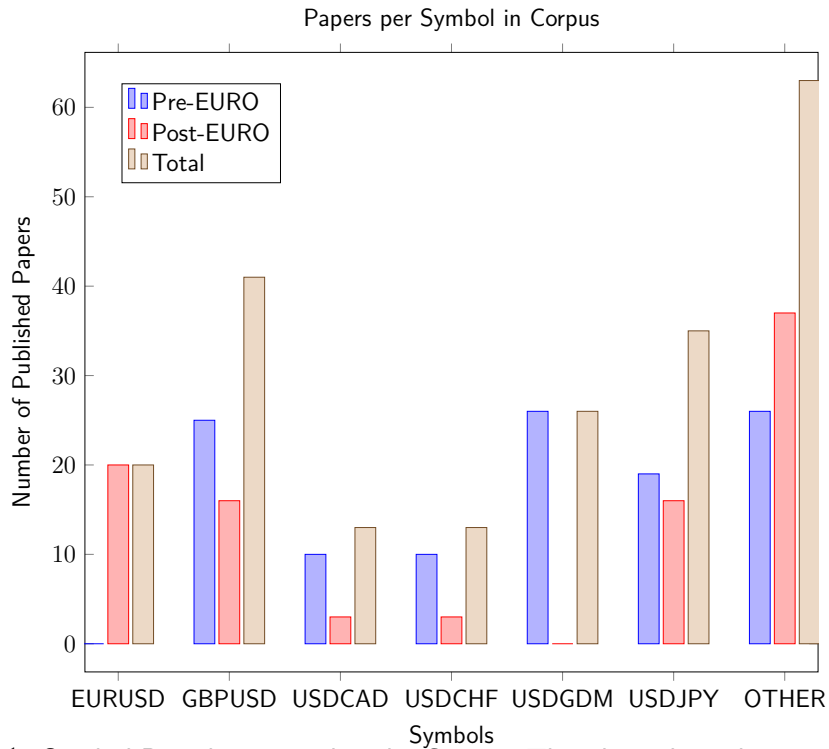


Figure 4.1: Symbol Distribution within the Corpus. This chart shows how many papers within the corpus use each Symbol, and is differentiated into two figures: Symbols used before the advent of the Euro (in blue) and Symbols used after (in red). The total figures per-Symbol are shown in the brown columns. All up, over 40 different Symbols were used, with six main currencies (over 5 papers each), and the remainder grouped in 'OTHER'.

Dates used for Daily GBPUSD Calculations

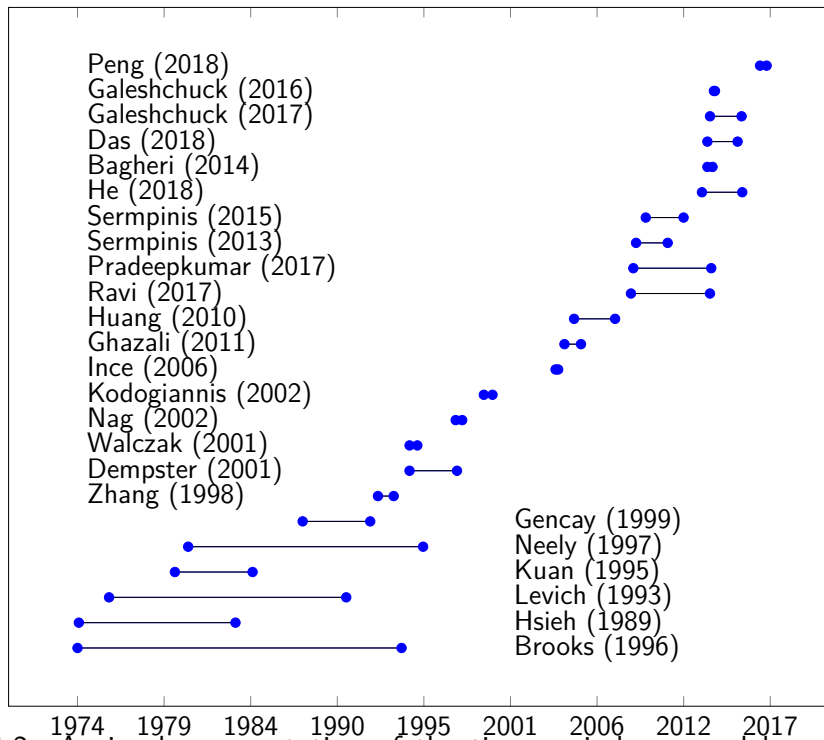


Figure 4.2: A visual representation of the time periods covered by each paper within the corpus, for the GBPUSD symbol.

If the metrics reported by researchers were largely cohesive, and presented across a broad range of scenarios, for the same purposes, comparison might still be possible. Along with defining the EMH, Fama also gave advice as to how speculative performance should be measured:

“For a [predictor] to be judged successful, it must demonstrate consistent ability to outperform the market, over a long time. Otherwise it is no more than a statistic. ... A simple, yet effective test to measure the effectiveness of a [predictor] is this: whenever a position is recommended, also take the position in another similarly risked asset. Over many iterations, it will become clear if the [predictor] is effective” (Fama, 1965b).

In this manner, a predictor’s worth can be tested by comparison with datasets *different* to that which it was developed for. This could be used for two different purposes: first, if the predictor performs consistently across multiple datasets then this would validate the generality of the proposed algorithm; second, if it performs well, out-of-sample, for the given dataset, but poorly in others, this provides a measure of confidence that the algorithm successfully identified unique features specific to the chosen dataset. Also, if the ‘similarly risked’ asset was the same dataset but for a different time period, this would give a measure of consistency, or an assurance of performance.

There are, then, a number of ways in which this could be tested pragmatically:

1. Split the dataset into a number of equally sized, potentially overlapping, smaller time-series, and test for consistency within performance metrics, via either statistical analysis of multiple runs, or cross-validation between the time series.
2. Construct one or more randomized sequences from the first-order differences of the time series (known as bootstrapping), and test for significance (Efron, 1982). This significance could be via a confidence interval,  $t$ -statistic, or some other ranking means.

3. Take the approaches of (1) and (2), but instead of using the same Symbol, test with a similarly distributed Symbol; i.e., if predicting USDCAD, compare with USDCHF (or some other Symbol with similar stochastic properties).

In fact, all these methods have been used within the corpus, although to varying effect. Only a few researchers utilised (1) with any level of statistical significance: Yao and Tan (2000) used 12 offset, overlapping time periods within the same Symbols; Giles et al. (2001) used 30; and Kampouridis and Otero (2017) used 255. A handful of researchers utilised (2), mostly to show via confidence interval that extracted features existed solely in the primary dataset. And while most researchers ran their predictors on more than one Symbol, it was not uncommon to find different periods used for each Symbol, and no study used the other Symbols as a control, like (3) suggests. Bar the previously noted papers, none of the research within the corpus presented more than a snapshot performance of their predictor. There is therefore no evidence provided that the performance of each predictor was anything other than luck, or as Fama called it, ‘a statistic’. Given that nearly all researchers claimed that their predictor performed better than their chosen benchmarks, the question needs to be asked: are the reported results *bona fide*, or are they the result of a cherry-picked dataset?

## 4.2 Metrics

Assuming, however, that the datasets are not selected for favourable performance, and that the sample results provided really are indicative of predictor performance, perhaps it is still possible to compare predictors by the metrics given? Unfortunately, systematic analysis of the corpus indicates that such an approach is also impossible, for two reasons.

First, in the 118 papers surveyed, over forty *different* metrics were used (see Table 4.1). Examination of the equations provided revealed some



<b>Metric</b>	<b>Shorthand</b>	<b>Percentage</b>	<b>Year</b>
Mean Absolute Error	MAE	19.5%	1983
Root Mean Squared Error	RMSE	19.5%	1983
Directional Correctness	DC	16.1%	1995
Mean Squared Error	MSE	15.3%	1989
Mean Absolute Percentage Error	MAPE	14.4%	1996
Absolute Profits	P	8.5%	1997
Diebold-Mariano Test	DM	7.6%	1995
Theil's U	TU	7.6%	1996
Normalised Mean Squared Error	NMSE	6.8%	1996
Sharpe Ratio	ShR	6.8%	1996
Brock-Dechert-Scheinkman Test	BDS	5.8%	1987
Annualised Return	AR	5.1%	1996

*Less than 5%: Box-Pierce Q, Correlation Coefficient,  $\chi^2$ , Compounded Return, Drawdown, Directional Symmetry, Friedman's Test, F-measure, Generalised Least Squares, Granger-Newbold, Hit Rate, Heteroscedastic Mean Absolute Error, Heteroscedastic Root Mean Squared Error, Ljung-Box Test, Mean Error, Mean Squared Error Ratio, Maximum Absolute Error, Ordinary Least Squares, Profit Factor, Pesaran-Timmermann Test,  $R^2$ , Risk Measure, Return on Investment, Signal-to-noise Ratio, Sortino Ratio, Stirling Ratio, Tukey-HSD and Variance.*

Table 4.1: Relative frequency of the metrics used by the corpus by which predictor accuracy was measured. Year is when the metric was first encountered within the sample corpus, not when invented. Due to the sheer number of metrics, Shorthand does not necessarily match what is used in the relevant papers, with priority for conflicting shorthands given to the more popular metrics and alternatives provided for the others.

researchers using the same methods by different names. This became problematic when a researcher published results using a metric for which there are multiple variants, but did not state which variant they used. Whilst it may be possible to convert between some metrics for the purposes of comparison, in general this is not the case, and comparison between papers by metric becomes intuitive rather than observed.

Second, the vast majority of the metrics used were unsuitable for FOREX prediction, being mostly borrowed from other disciplines and then applied inappropriately. It is important to note at this point the difference between *time series* and *FOREX* prediction, *particularly for speculation*, that until now has not been explicated:

Time series prediction is the prediction of the value of a time series at a

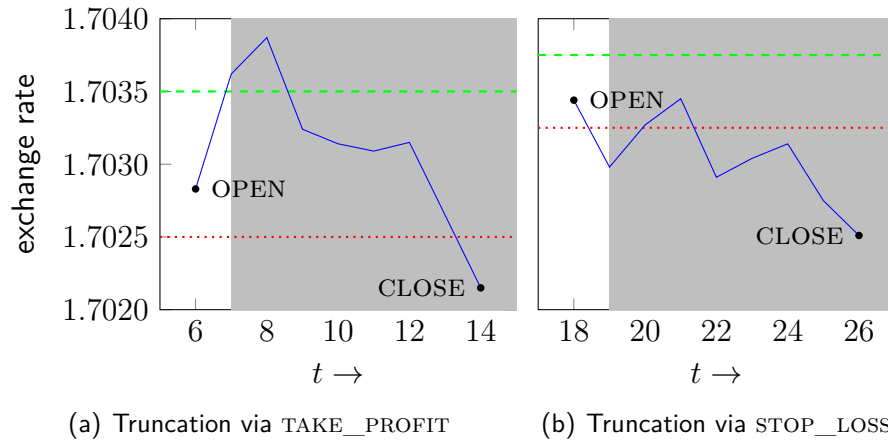


Figure 4.3: Examples of *optional truncation* and their effect on profitability.

In (a), if a BUY Position were opened at  $t = 6$  and closed at  $t = 14$ , it would realise a loss. However, with a TAKE\_PROFIT set, indicated by the green dashed line, the Position would have already closed at  $t = 7$ , for a profit. As a result, the time period is *truncated* (as indicated by the grayed out background) and the remainder of the time period is of no interest for that Position.

In (b), the effect of a STOP\_LOSS is shown. Whilst a BUY Position that opened at  $t = 19$  and closed at  $t = 26$  would realise a loss, the use of a STOP\_LOSS, indicated by the red dotted line, ensures a smaller loss is made when the STOP\_LOSS is triggered at  $t = 19$ . Like the TAKE\_PROFIT in (a), this truncates interest in the remainder of the time period.

specific future point in time. This can include FOREX time series—however for the purposes of speculation, this definition is insufficient. The reason for this deficiency is perhaps not obvious at first, but basically, speculation introduces the concept of *optional truncation* through the application of TAKE\_PROFIT and STOP\_LOSS values (see Fig. 4.3). These truncation values allow for a *range* of values to be considered correct, and none of the synthetic metrics used within the corpus make allowance for this. Whilst it could be argued that profit-based metrics do take these into account, they come with their own set of problems, as will be shown later. In order to better understand what this means, it is necessary to consider *what exactly is being measured?*

### 4.2.1 What is Being Measured?

At the smallest time unit available in FOREX, the *tick*, FOREX time series consist of two values: the BID and ASK prices (Fig. 4.4a). However, ticks follow exchange rate movements, and thus have no consistent timing. It is therefore

common for FOREX movements to be represented as an aggregate time series, commonly referred to as a *candlestick* (Fig. 4.4b). In this format, the FOREX time series now consists of a number of values: OPEN, HIGH, LOW, CLOSE and VOLUME, per-Symbol.

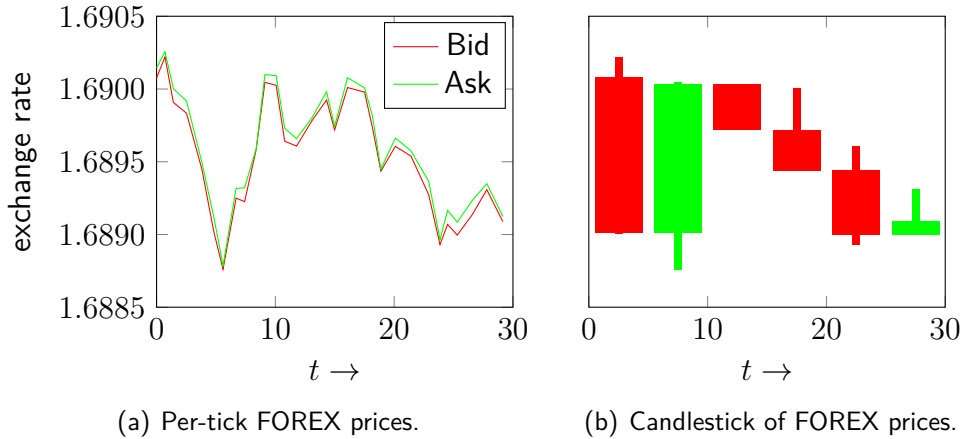


Figure 4.4: Graphical representations of FOREX prices. (a) shows the per-tick data, which happens at irregular intervals, while (b) shows how the tick data is aggregated into regularly-spaced *candlesticks*, for the same data. The thick portion of the candlestick represents the movement between OPEN to CLOSE, and the thin HIGH and LOW. If the price increases from OPEN to CLOSE, the candlestick is green, otherwise it is red. Each successive candlestick starts roughly where the previous candlestick closed.

OPEN is the exchange rate for the symbol at the start of the candlestick, HIGH and LOW are the highest and lowest exchange rates within the candlestick, and CLOSE is the exchange rate at the end. These are typically the ASK prices, and the spread is no longer represented. However a new value, VOLUME, is also available, which is the relative volume of the symbol traded within the candlestick's time period. This time period can be any consistent time period, measured in minutes. Within the corpus, the mode was daily observations, or 1440 minute time periods—though the daily start times reported varied from paper to paper.

Logically then, it could be argued that for the purposes of speculation, any prediction that is realised between the LOW and HIGH values of the predicted period is therefore correct—especially if the prediction is used to inform a TAKE\_PROFIT and thus truncate interest in the remainder of the period. In

the same vein, the prediction of an appropriate `STOP_LOSS` value could be construed as a forecast of the risk required to achieve a profit. However, within the corpus, this distinction is never made. In fact, only three papers referenced candlestick properties, and none made use of them, either as input or for assessment.

Whilst several papers were devoted to long-term modeling, where these values are considered ‘noise’, the vast majority of papers were targeting speculation—especially indicated by those papers that reported predictive performance via profitability. It is therefore surprising to find such a large feature of FOREX speculation missing from the corpus, and it is through this lens that this chapter proceeds to analyse the various metrics used by the corpus.

The metrics used within the corpus fall roughly into four categories: statistical, directional, comparative and profit-based.

#### 4.2.2 Metric Family: Statistical

The statistical family of metrics is the largest and most represented in the corpus, and is listed in Table 4.2. These metrics are typically based on some form of error calculation, i.e.,  $\epsilon = E(y, \hat{y})$ , with the smaller the reported value, the more accurate the predictor considered—for example, *Mean Absolute Error* (MAE) and *Root Mean Squared Error* (RMSE). Due to a number of known issues with these metrics, particularly related to the effect of outliers, it is typical within the corpus to report an assortment of statistical metrics. Nevertheless, for the purpose of FOREX speculation, these metrics have all been applied incorrectly.

The major problem with these metrics is their dependence on a single ‘correct’ value for comparison. But what should that ‘correct’ value be? As has been noted earlier, none of the papers paid any attention to the nature of FOREX speculation, or specifically considered *what* was being predicted. Nowhere is this more evident than in the statistical metrics used, where the

Table 4.2: Statistical metrics used within the surveyed corpus.

Metric	Equation	References
<i>Given</i>		
	$\epsilon_t = y_t - \hat{y}_t, \quad \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i, \quad P(y_i) \text{ is probability of } y_i \text{ in } Y$	
<i>then</i>		
ME	$\frac{1}{N} \sum_{i=1}^N \epsilon_i$	Meese and Rogoff, 1983a.  Meese and Rogoff, 1983a; Zhang and Hu, 1998; Andersen, Bollerslev, and Lange, 1999; Leung, Che, and Daouk, 2000; Tay and Cao, 2001; Kamruzzaman and Sarker, 2004; Clarida, Sarno, Taylor, and Valente, 2003; Ince and Trafalis, 2006; Panda and Narasimhan, 2007; Khashei, Hejazi, and Bijari, 2008; Sarno and Valente, 2009; Huang, Chuang, Wu, and Lai, 2010; Lin, Chiu, and Lin, 2012; Sermpinis, Dunis, Laws, and Stasinakis, 2012; C. Evans, Pappas, and Xhafa, 2013; Sermpinis, Theofilatos, Karathanasopoulos, Georgopoulos, and Dunis, 2013; Rossi, 2013; Shen, Chao, and Zhao, 2015; Peng, Albuquerque, de Sá, Padula, and Montenegro, 2018; Bui, Vu, and Dinh, 2018; Das, Bisoi, and Dash, 2018; Dash, 2018; Hajizadeh, Mahootchi, Esfahanipour, and Kh, 2019.
MAE	$\frac{1}{N} \sum_{i=1}^N  \epsilon_i $	Engel and Hamilton, 1990; Kuan and Liu, 1995; Gençay, 1999; Nag and Mitra, 2002; Ince and Trafalis, 2006; Khashei, Bijari, and Ardali, 2009; Leu, Lee, and Jou, 2009; Huang, Chuang, Wu, and Lai, 2010; Rossi, 2013; Sermpinis, Stasinakis, Theofilatos, and Karathanasopoulos, 2015; Pradeepkumar and Ravi, 2017; Ravi, Pradeepkumar, and Deb, 2017; Bui, Vu, and Dinh, 2018; Dash, 2018; He, Chen, and Tso, 2018; Hann and Steurer, 1996.
MSE	$\frac{1}{N} \sum_{i=1}^N \epsilon_i^2$	Meese and Rogoff, 1983a; Hsieh, 1989b, 1991; Zhang and Hu, 1998; Andersen, Bollerslev, and Lange, 1999; Sarantis, 1999; Leung, Che, and Daouk, 2000; Clarida, Sarno, Taylor, and Valente, 2003; Manzan and Westerhoff, 2007; Panda and Narasimhan, 2007; Sarno and Valente, 2009; Huang, Chuang, Wu, and Lai, 2010; Sermpinis, Dunis, Laws, and Stasinakis, 2012; Lin, Chiu, and Lin, 2012; Sermpinis, Theofilatos, Karathanasopoulos, Georgopoulos, and Dunis, 2013; Rossi, 2013; Rout, Majhi, Majhi, and Panda, 2014; Sermpinis, Stasinakis, Theofilatos, and Karathanasopoulos, 2015; Shen, Chao, and Zhao, 2015; Peng, Albuquerque, de Sá, Padula, and Montenegro, 2018; Das, Bisoi, and Dash, 2018; Dash, 2018; Hajizadeh, Mahootchi, Esfahanipour, and Kh, 2019.
RMSE	$\sqrt{\text{MSE}}$	

Metric	Equation	References
MSPE	$\frac{100}{N} \sum_{i=1}^N \left(\frac{\epsilon_i}{y_i}\right)^2$	Diebold and Nason, 1989; Gençay, 1999; Arifovic and Gençay, 2001.
RMSPE	$\sqrt{\text{MSPE}}$	Kodogiannis and Lolis, 2002
MAPE	$\frac{100}{N} \sum_{i=1}^N \frac{ \epsilon_t }{y_t}$	Diebold and Nason, 1989; Levich and Thomas, 1993; Kaastra and Boyd, 1996; Zhang and Hu, 1998; Kodogiannis and Lolis, 2002; Nag and Mitra, 2002; Panda and Narasimhan, 2007; Majhi, Panda, and Sahoo, 2009; Sermpinis, Dunis, Laws, and Stasinakis, 2012; Lin, Chiu, and Lin, 2012; Sermpinis, Theofilatos, Karathanasopoulos, Georgopoulos, and Dunis, 2013; Rout, Majhi, Majhi, and Panda, 2014; Shen, Chao, and Zhao, 2015; Galeshchuk and Mukherjee, 2017; Das, Bisoi, and Dash, 2018; Dash, 2018; Hajizadeh, Mahootchi, Esfahanipour, and Kh, 2019.
NMSE	$\frac{\sum_{i=1}^N \epsilon_t^2}{\sum_{i=1}^N (y_t - \bar{y}_t)^2}$	Tenti, 1996; Lisi and Schiavo, 1999; Yao and Tan, 2000; Tay and Cao, 2001; Kamruzzaman and Sarker, 2004; Yu, Wang, and Lai, 2005; Yu, Lai, and Wang, 2008; Ghazali, Hussain, and Liatsis, 2011.
R <sup>2</sup>	1 – NMSE	Andersen and Bollerslev, 1998; Andersen, Bollerslev, and Lange, 1999; Nag and Mitra, 2002; M. Evans and Lyons, 2002; Chen and Leung, 2004.
$\chi^2$	$\sum_{i=1}^N \frac{y_i^2}{NP(y_i)} - N$	Nelson, 1991; Meese and Rose, 1991
SNR	$10 \log_{10} \frac{N \max(\mathbf{Y})^2}{\sum_{i=1}^N \epsilon_i^2}$	Ghazali, Hussain, and Liatsis, 2011.

Other statistical metrics used include OLS Engle, 1982; Meese and Rogoff, 1983a; Bollerslev, 1986; Meese and Rose, 1991, GLS Meese and Rogoff, 1983b log-loss, heteroscedastic MAE & RMSE, Andersen, Bollerslev, and Lange, 1999, MaxAE Nag and Mitra, 2002 and Variance Shen, Chao, and Zhao, 2015.

CLOSE value has been used exclusively for comparative purposes. However, the CLOSE value is merely the last observed value of the time period, with  $\text{LOW} \leq \text{CLOSE} \leq \text{HIGH}$ , and has no correlation to the potential profitability, at least in the context of most of the published research (see Fig. 4.5). In these longer time periods, the use of TAKE\_PROFIT and STOP\_LOSS values allow speculators to realise performance not allowed by exclusive use of CLOSE

values by predictors. It is worth noting that, as the interval of interest shortens, the significance, and therefore correlation to profitability, of the CLOSE value increases—however, nearly all researchers use ultra-low frequency intervals. As such, the use of the CLOSE value, within the FOREX corpus related to speculation, is without merit.

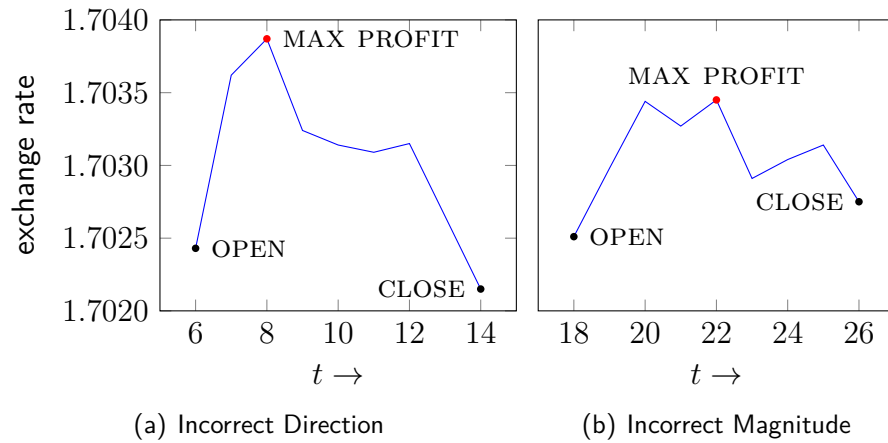


Figure 4.5: Irrelevance of CLOSE values. In (a), CLOSE is less than OPEN, but maximum profitability is at  $t = 8$ , in the opposite direction. In (b), whilst the direction of CLOSE is now correct, it is considerably lower in value than the maximum profitability achieved at  $t = 22$ .

### 4.2.3 Metric Family: Directional

In a similar fashion, this same charge could be leveled against the next family, directional metrics. Whilst there are a number of names given to the various directional metrics, analysis of the equations provided revealed there was, in fact, only a single metric, correct direction. Correct direction is defined:

$$CD(\hat{y}_t) = 1, \text{ if } (\hat{y}_t - y_{t-1})(y_t - y_{t-1}) \geq 0 \quad (4.2)$$

$$0, \text{ otherwise}$$

and has been referred to by several names, as in Table 4.3. It has also been broken into parts, as in Table 4.4, with these parts also sometimes rearranged to form a new metric, as in Table 4.5.

Much like the statistical metrics, however, these metrics all rely on a single

Table 4.3: Names used for Directional Correctness with the corpus.

Name	References
Directional Symmetry	Tay and Cao (2001), Kamruzzaman and Sarker (2004), Leu, Lee, and Jou (2009).
Directional Accuracy	Panda and Narasimhan (2007), Shen, Chao, and Zhao (2015), Nassirtoussi, Aghabozorgi, Wah, and Ngo (2015), Galeshchuk and Mukherjee (2017).
$D_{stat}$	Yu, Wang, and Lai (2005), Yu, Lai, and Wang (2008), Yu, Wang, and Lai (2009), Pradeepkumar and Ravi (2017), Ravi, Pradeepkumar, and Deb (2017).
Hit Rate	Kim (2003), Bagheri, Peyhani, and Akbari (2014), Das, Bisoi, and Dash (2018).
Sign	Kuan and Liu (1995), Diebold and Mariano (1995), Hann and Steurer (1996), Gençay (1999), Arifovic and Gençay (2001), Panda and Narasimhan (2007).

Table 4.4: Parts used from Directional Correctness within the corpus.

Name	References
Direction Change Correct per-Direction	Giles, Lawrence, and Tsoi (2001), Walczak (2001), Ghazali, Hussain, and Liatsis (2011). Kamruzzaman and Sarker (2004), Lin, Chiu, and Lin (2012).

Table 4.5: Combined Directional Statistics

Metric	Equation	References
Pesaran-Timmermann Test	$PT = \frac{P(CC) - P(CD)}{\sigma^2(CC) - \sigma^2(CD)}$	Pesaran and Timmermann (1992), Sermpinis, Theofilatos, Karathanasopoulos, Georgopoulos, and Dunis (2013), Sermpinis, Stasinakis, Theofilatos, and Karathanasopoulos (2015).
F-measure	$FM = \frac{2TP}{2TP + FP + FN}$	Patel, Shah, Thakkar, and Kotecha (2015), Das, Bisoi, and Dash (2018).

TP, TN, FP and FN are the *True Positive*, *True Negative*, *False Positive* and *False Negative* percentages, respectively.  $P(CC)$  is the percentage of correct changes predicted and  $P(CD)$  is the percentage of correct direction predicted.



value and so all were used with CLOSE. Therefore, just as the statistical results reported have little bearing on overall profitability, the same can be said for directional statistics.

Another issue with directional metrics is that FOREX movements within any given time period are likely to be *bi-directional* (that is, they can obtain values both above and below OPEN before attaining their CLOSE value). Analysis of seven years of per-day data across fifty currency pairs (see Section 5.1) shows that almost 98% of daily FOREX rates are bi-directional. Combined with the use of TAKE\_PROFIT values, this means that it is generally possible to make profitable predictions in both directions, and thus limiting prediction results to a single direction is incorrect.

Alongside this is the issue of the inequality operator in Eq. 4.2. Whilst some researchers report *greater-than*, the equation given here is the more commonly listed, and slightly dishonest, *greater-than-or-equal*. The difference is subtle—greater-than-or-equal includes all periods of non-movement as correct, regardless of the researcher’s prediction. This has the effect of inflating the result slightly (e.g., it would report an extra two percent accuracy for the dataset provided by *Hokohoko*, multiplied to four percent when referenced by parts). And this hints to the third issue: researchers reporting favourable statistics.

Reporting by direction gives researchers some nice, large numbers to report. Generally speaking, long-term directional predictions are not particularly difficult—currency pairs are usually either trending or oscillating slowly. In these scenarios, even predicting just a single direction at all times will produce a favourable result—especially if no-change is included as correct—as long as the prediction matches the trend. Given most researchers are reporting results for fairly large time scales and long intervals, directional statistics relate to their ability to predict the direction of heavily filtered, ultralow-frequency data—something that the simple linear-regression indicator already does fairly well. It should be noted that easy-to-predict is a

relative term, and does not necessarily mean easily profitable, as it usually implies increased costs as per the EMH.

Overall, due to directional statistics having no correlation with profitability, and their often conflated values, they are a poor measure of a predictor's performance, if the goal is profitable speculation.

#### 4.2.4 Metric Family: Comparative

Another way that researchers measure the effectiveness of their predictors is by way of comparison with either another predictor, or the time series itself. Because of their major application to other predictor's time series, this thesis calls them the *comparative* family, and defines them as separate from the statistical metrics, though they are closely related. Within the corpus, this family of metrics is made up of Pearson's Correlation Coefficient, Friedman's Test, Theil's U, the Ljung-Box-Pierce Q, Granger-Newbold's test for causality, Brock-Dechert-Scheinkman's test for deterministic chaos, and the Diebold-Mariano test.

##### 4.2.4.1 Pearson's Correlation Coefficient

Developed by Pearson in the late 1800's, the Pearson Correlation Coefficient is a measure of the linear correlation between two time series,  $\mathbf{A}$  and  $\mathbf{B}$ . It is defined as

$$\rho(\mathbf{A}, \mathbf{B}) = \frac{\sum_{i=1}^N ((\mathbf{A}_i - \bar{\mathbf{A}})(\mathbf{B}_i - \bar{\mathbf{B}}))}{\sqrt{\sum_{i=1}^N (\mathbf{A}_i - \bar{\mathbf{A}})^2 \sum_{i=1}^N (\mathbf{B}_i - \bar{\mathbf{B}})^2}} \quad (4.3)$$

and has a range of  $[-1, 1]$ , with 0 indicating no linear correlation, and 1,  $-1$  positive and negative correlation, respectively. It was used in the corpus by Panda and Narasimhan (2007), C. Evans, Pappas, and Xhafa (2013) and Shen et al. (2015) to measure the correlation between different predictors' outputs against future data, with the predictor having the highest score

declared the best. Reported results ranged from 0.66 to 0.88, indicating a measure of correlation, with the control Random Walk ranging from 0.27 to 0.65.

#### 4.2.4.2 Friedman's Test

Introduced in 1937, Friedman's test is intended for use to determine which out of multiple factors has the most effect on a dependent variable. It can be applied to the comparison of predictors by giving each predictor a rank, per-time period  $t$ , and calculated:

$$F_R = \frac{12}{NK(K+1)} \sum_{j=1}^K \left( \sum_{i=1}^N \mathbf{R}_{ij} \right)^2 - 3N(K+1) \quad (4.4)$$

where  $K$  is the number of ranks (or predictors),  $N$  the length of the time series, and  $\mathbf{R}$  a vector of all ranks by predictor and time (Friedman, 1937). Friedman's test was used by Das, Bisoi, and Dash (2018), both on its own and as input into *Tukey's Honestly Significant Difference* (HSD) test, for statistical differencing between means. Their application ranked  $|\epsilon_{ij}|$ , and that subjects this technique to the same critique as the other metrics: it is dependent on a single 'best' value, with CLOSE used.

#### 4.2.4.3 Theil's U Metric

Dating from 1958 and 1966 respectively, there are two versions of Theil's U metric. The first,  $U1$ , is a measure of predictive *accuracy*:

$$U1 = \frac{\sqrt{\sum_{i=1}^N (\mathbf{A}_i - \mathbf{B}_i)^2}}{\sqrt{\sum_{i=1}^N \mathbf{A}_i^2}} \quad (4.5)$$

Table 4.6: Papers using Theil's U-Metric

Version	Papers
U1	Leung, Che, and Daouk (2000), Dash (2018).
U2	Sermpinis, Dunis, Laws, and Stasinakis (2012), Sermpinis, Laws, Karathanasopoulos, and Dunis (2012), Sermpinis, Theofilatos, Karathanasopoulos, Georgopoulos, and Dunis (2013), Pradeepkumar and Ravi (2017), Ravi, Pradeepkumar, and Deb (2017).
<i>unknown</i>	Hann and Steurer (1996).

and the second,  $U_2$ , is a measure of the *quality* of predictions:

$$U_2 = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (\mathbf{A}_i - \mathbf{B}_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N \mathbf{A}_i^2} + \sqrt{\frac{1}{N} \sum_{i=1}^N \mathbf{B}_i^2}} \quad (4.6)$$

with  $\mathbf{A}$  and  $\mathbf{B}$  representing the time series being compared (Theil, Cramer, Moerman, & Russchen, 1958; Theil, 1966). For both these metrics, the closer the value is to 0, the closer the two time series match.  $U_2$  is a correction of  $U_1$ , which gives an upper bound of 1, whereas  $U_1$  is unbounded positively (though such an error would be catastrophic and is highly unlikely). These metrics were typically applied against the researchers' definition of Random Walk, with examples from the corpus given in Table 4.6.

#### 4.2.4.4 Box-Pierce and Ljung-Box

In 1970, Box and Pierce noted a need for fitness tests in the newly developed ARMA and ARIMA models (Box & Pierce, 1970). To address this issue they posited that, if a model were an accurate estimate of a time series, the residuals (or errors) would consist of purely white noise. However, if the fit was lacking, there may be further complexities present, signalled by the presence of autocorrelations within the residuals. They therefore developed a

test for the presence of autocorrelations within residuals, defined:

$$Q_{BP} = N \sum_{k=1}^h \left( \frac{\sum_{i=1}^N \epsilon_i \epsilon_{i-k}}{\sum_{i=1}^N \epsilon_i^2} \right)^2 \quad (4.7)$$

where  $h$  is the number of lags being tested. Theoretically, the smaller  $Q_{BP}$ , the better the fit of the model, though they recommend caution in interpretation. Indeed, over the next few years, it was found through application that the test sometimes yielded suspiciously low values. To remedy this, Ljung and Box published a modification in 1978, with a changed variance estimate, yielding:

$$Q_{LBP} = N(N+2) \sum_{k=1}^h \frac{\left( \frac{\sum_{i=1}^N \epsilon_i \epsilon_{i-k}}{\sum_{i=1}^N \epsilon_i^2} \right)^2}{N-k} \quad (4.8)$$

which is sometimes referred to as the *Ljung-Box-Jenkins* test (Ljung & Box, 1978). Bollerslev et al. (1992) used it to show a simple GARCH(1, 1) model outperformed a more complex ARCH(12) model; Levich and Thomas (1993) applied the tests directly to FOREX data, as a test for autocorrelation; and Bera and Higgins (1993) used it to determine the AR component for their ARCH and GARCH models. Diebold (1986, 1988) found that, in the presence of conditional heteroscedasticity, these metrics are largely invalidated, and recommended using ARCH-corrected standard errors, though only Bera and Higgins (1993) make note of doing so.

Silvapulle and Evans (1998) tested several variants and, finding Diebold's error-prone in the presence of unexplained disturbances, recommended the use of modifications suggested by Wooldridge (1991): changing the underlying error function to utilise a conditional mean. Like all of the error-based metrics so far, the Ljung-Box-Pierce Q assumes a single correct value per prediction.

#### 4.2.4.5 Granger-Newbold

Based on early definitions of causality by Granger in 1969, the Granger-Newbold test is designed to test for causality between two related variables with respect to a third (C. Granger, 1969; G. Granger & Joyeux, 1980). Given two error-series  $e_1$  and  $e_2$  related to the same predicted series, if

$$\begin{aligned}x_t &= e_{1t} + e_{2t} \\z_t &= e_{1t} - e_{2t}\end{aligned}\tag{4.9}$$

then regressing

$$x_t = \beta z_t + \epsilon\tag{4.10}$$

where  $\beta$  is a coefficient and  $\epsilon_t$  bias, provides a metric by which to compare the two time series errors (Mariano, 2000). Assuming that the error loss is quadratic, the forecast errors have zero mean and follow a Gaussian distribution, and are serially uncorrelated: if  $\beta$  is 0 then the predictions are close to one another. If  $\beta$  is significantly different from 0, then they are different. This needs to be tempered by analysis of  $\epsilon$ , however. If  $\epsilon$  is non-zero, then either  $x$  or  $z$  is biased, and the test assumptions are invalidated (Hann & Steurer, 1996).

Several variants of this test for causality have been published by Granger and others, in an effort to address its limited application to single-step-ahead predictions and reliance on quadratic error-loss functions. Granger noted that it can be argued that variance is not the proper criterion to use as a measure of closeness of a predictor to the true value—but it is easy to use (C. Granger, 1969). It is often hard to detect causality, as data that appears instantly caused may in fact be undersampled, and purely deterministic series cannot be said to have any causal influences other than themselves. Therefore, this method is not really suited for chaotic time series.

Within the corpus, Meese and Rogoff (1983a) used it as a metric to compare different exchange rate models of the 1970s; Hann and Steurer (1996) used

it to test for significant difference between the predictions of two models at specific time periods; and Lisi and Schiavo (1999) used it to test the statistical significance of their predictions.

Another, less-used variant is of the form

$$\begin{aligned} \text{MGN}(\mathbf{x}, \mathbf{z}) &= \frac{\hat{\rho}(\mathbf{x}, \mathbf{z})}{\sqrt{\frac{1 - (\hat{\rho}(\mathbf{x}, \mathbf{z}))^2}{N - 1}}} \\ \hat{\rho}(\mathbf{x}, \mathbf{z}) &= \frac{\mathbf{x}^\top \mathbf{z}}{\sqrt{(\mathbf{x}^\top \mathbf{x})(\mathbf{z}^\top \mathbf{z})}} \end{aligned} \quad (4.11)$$

Like all the statistical metrics, this method relies on an error function dependent on a single ‘correct’ result, and within the corpus its use is restricted to CLOSE values.

#### 4.2.4.6 Brock-Dechert-Scheinkman Test

In 1987 Brock, Dechert and Scheinkman proposed a non-linear test analogous to the linear Ljung-Box-Pierce Q test, stating that linear time series methods, such as spectral analysis and autocovariance functions, may not be able to observationally distinguish between deterministic and random systems (Brock, Dechert, & Scheinkman, 1987). They suggested that, if the correlation dimension of a time series is low, and the estimated largest Lyapunov exponent is positive, the residuals from a misspecified model for a chaotic time series will have the same correlation dimension and Lyapunov exponent as the time series itself. Therefore, a test for chaos, applied to the residuals, should be useful to verify the accuracy of the model. They proposed that, analogous to LBP-Q, if a model fits the data then the resulting residuals will be white noise only, and presented a test for nonlinear residuals based on the earlier work of Wolf et al., Grassberger et al. and Takens:

$$\text{BDS}(\mathbf{Y}_m, \gamma) = \frac{\sqrt{N}(C(\mathbf{Y}_m, \gamma) - P(C(\mathbf{Y}_1, \gamma))^m)}{\sigma(\sqrt{N}(C(\mathbf{Y}_m, \gamma) - P(C(\mathbf{Y}_1, \gamma))^m))} \quad (4.12)$$

$\mathbf{Y}_m$  is the  $m$ -embedded time series drawn from  $\mathbf{Y}$ ,  $N$  the length of  $\mathbf{Y}_m$ ,  $\gamma$  a specified error,  $C$  the correlation integral (see Eq. 3.9),  $\sigma$  the standard deviation function and  $P(C(\mathbf{Y}_1, \gamma))$  the probability of the 1-dimensional error.

They also noted, however, that if the noise in the time series is large enough relative to the variation of  $\mathbf{Y}_t$  across the attractor  $\mathbf{\Lambda}$ , then the structure will be obscured, and the Lyapunov exponents potentially undefined. Therefore, they proposed a smoothing process first be applied to reduce noise, which theoretically should not have any effect on the detection of low-dimensional chaos. This test can then be applied as a goodness-of-fit test to any model built on IID errors (Brock et al., 1987; Brock, Scheinkman, Dechert, & LeBaron, n.d.).

Since publication, the BDS test has become the most popular test for low-dimensional chaos, with Hsieh (1989b) using it to attempt to distinguish between different types of nonlinearity; Scheinkman and LeBaron (1989) using it to compare nonlinear models with ARCH models; Bollerslev et al. (1992) using it to compare ARCH and GARCH models; and Hann and Steurer (1996) using it to compare nonlinear and linear models. In particular, Hann and Steurer (1996) supposed that, if the residuals are nonlinear, then the use of a neural network for prediction may be beneficial. However, they found no evidence of nonlinearity, and thus concluded that the use of nonlinear methods will provide no extra performance. Barnett and Serletis (2000) also noted that the BDS test is actually a measure of *whiteness*, and all other interpretations are therefore indirect and insufficient in-and-of-themselves to draw any conclusions regarding chaos, linearity and nonlinearity.

Despite its relative popularity, the BDS test has been shown to be weak against smaller sample sizes, especially those found in financial time series, with the LaGrange multiplier providing more consistent performance (Brock et al., 1992; Lee, White, & Granger, 1993). However, whilst often used to optimise models, the LaGrange Multiplier has not been used as a performance



metric within the corpus, whereas the BDS has.

#### 4.2.4.7 Diebold-Mariano Test

In 1995, Diebold and Mariano noted that there was generally a rather casual attitude towards the application of metrics within the economic corpus (Diebold & Mariano, 1995). They noted that point estimates are usually sampled, without any attempt to assess their uncertainty, and that the economic loss associated with a forecast may be poorly related to the usual statistical metrics. They therefore proposed a test for the null hypothesis that there is no difference between two competing forecast time series.

Given an error function  $E(y, \hat{y})$ , the DM statistic between two series  $\hat{\mathbf{w}}, \hat{\mathbf{x}}$ , with respect to  $\mathbf{y}$  is defined:

$$\begin{aligned}
 d_t &= E(\mathbf{y}_t, \hat{w}_t) - E(\mathbf{y}_t, \hat{x}_t) \\
 \bar{d} &= \frac{1}{N} \sum_{i=1}^N d_t \\
 \hat{\omega}_{\bar{d}} &= \gamma_0 + 2 \sum_{i=1}^N \gamma_i, \quad \gamma_i = \text{cov}(d_i, d_{N-i}) \\
 \text{DM} &= \frac{\bar{d}}{\sqrt{\frac{\hat{\omega}_{\bar{d}}}{N}}}
 \end{aligned} \tag{4.13}$$

A strongly negative score indicates the first predictor superior to the second, a strongly positive score vice-versa, and a close to zero score identical. This test has significant advantages over the other proposed metrics, as it is applicable to a wide range of loss functions, with the loss functions not needing to be quadratic, symmetric or even continuous. The error distributions can be non-zero-mean, non-Gaussian and contemporaneously or serially correlated.

Within the corpus, the DM test is used in two different ways: as a comparison between two competing time series predictions; or to generate a statistical measure within a bootstrap environment. Examples of the former

Table 4.7: Works using Diebold-Mariano for bootstrap analysis of predictor performance.

---

Gençay (1999)
Clarida, Sarno, Taylor, and Valente (2003)
Kilian and Taylor (2003)
Manzan and Westerhoff (2007)
Sarno and Valente (2009)
Sermpinis, Dunis, Laws, and Stasinakis (2012)
Sermpinis, Laws, Karathanasopoulos, and Dunis (2012)
Bekiros and Marcellino (2013)
Rossi (2013)
Sermpinis, Theofilatos, Karathanasopoulos, Georgopoulos, and Dunis (2013)
Sermpinis, Stasinakis, Theofilatos, and Karathanasopoulos (2015)
Pradeepkumar and Ravi (2017)
Ravi, Pradeepkumar, and Deb (2017)
Peng, Albuquirqui, de Sá, Padula, and Montenegro (2018)

---

can be found in the work of Jongen, Verschoor, Wolff, and Zwinkels (2012), Das et al. (2018) and Hajizadeh et al. (2019), who all used Absolute Error as their error function. The latter can be seen in the papers listed in Table 4.7.

However, the choice of error function was largely the same across these papers, with most using MAE, MSE, or RMSE—only Pradeepkumar and Ravi (2017) extended the selection to include Directional Correctness and Theil’s U2. So whilst the use of the DM test may provide a measure of statistical significance, the trend within the literature of relying on error functions that depend solely on CLOSE continued, and the overall results are thus fairly useless as a measure of speculative performance. As Rossi (2013) noted, “the choice of metric matters, and a lot!”.

#### 4.2.5 Metric Family: Profit-based

There is, however, a family of metrics that do not depend on an error function. Reporting results by profitability comes rather late in the corpus, as most researchers prior to 1990 were mathematicians and economists, nearly all of whom rejected profits as a measure *a priori*. Computer scientists, however, have shown no such reticence; and as the balance of

researchers in FOREX prediction has shifted away from mathematicians and economists towards computer scientists, the propensity for profit-based reporting that proliferates ‘popular’ literature has filtered into the academic corpus. Much like the other metric families, a number of different profit-based metrics have been used over the years. Within the selected corpus, researchers have used raw profits, the Sharpe ratio, compounded returns, the Stirling ratio, risk measure, annualised return, drawdown, profit factor, and the Sortino ratio to measure the effectiveness of their predictor/algorithm.

#### 4.2.5.1 Profits

Whilst the first metric referenced in the corpus, raw profits were not generally reported as a metric until the turn of the millennium. The main reason for this was the implicit understanding that profits were dependent on more than just the predictor. A well-known effect of the EMH in the markets is that, in a rising market, everyone does well (Sharpe, 1966). Therefore external causes unrelated to the predictor may help its performance, depending on the Symbol and time period used. Profits gained may also be a measure of the *trading strategy* applied—how much risk is used, leverage required, margin maintained, costs managed, etc. The former issue could be countered by using multiple time-periods for testing, as suggested by Fama, however the latter cannot. The best that can be done in this situation is the publication of all account settings used, which allows transparency but not comparability. Nevertheless, with the advent of easy-to-use, freely available trading software with back-testing algorithms (such as the *MetaTrader* and *cTrader* platforms), it has become significantly easier for researchers to report the raw profits directly from their trading software.

Profits are sometimes also reported as *Return On Investment* (ROI). This

is the percentage increase in equity, defined:

$$\text{ROI} = \frac{\text{equity}_{\text{final}}}{\text{equity}_{\text{initial}}} \quad (4.14)$$

Examples of reporting raw profits can be found in Yao and Tan (2000), Walczak (2001), Dempster and Jones (2001), Dempster, Payne, Romahi, and Thompson (2001), Park and Irwin (2007), Hirabayashi, Aranha, and Iba (2009) and Ozturk, Toroslu, and Fidan (2016). Examples of ROI can be found in Dymova, Sevastjanov, and Kaczmarek (2016) and de Almeida, Neves, and Horta (2018).

#### 4.2.5.2 Sharpe Ratio

In 1966, following the work of Fama and Markowitz, Sharpe was interested in the problem of predicting mutual fund performance. In order to do this, a way of measuring their performance and relating profits to risk was required. Sharpe therefore expanded upon the  $T$ -metric proposed earlier by Treynor (1965), and developed what is now known as the *Sharpe ratio*, defined:

$$R_{\text{Sharpe}} = \frac{A - B}{\sqrt{(A - B)^2}} \quad (4.15)$$

where  $A$  is the actual return, and  $B$  the risk-free return—such as investing in fixed-term bonds, etc. (Sharpe, 1966, 1994). Whilst the original examples were based on the yearly returns made by mutual funds, Sharpe’s formula has been used by researchers in other econometric fields, such as FOREX speculation. Most researchers, however, have not used it as a comparison, instead setting  $B$  to 0 or no investment, with only  $C$ . Evans et al. (2013) using it to make comparison between an alternative investment. This application results in a modified equation:

$$R_{\text{Sharpe}} = \frac{A}{\sigma(A)} \quad (4.16)$$

which is found in the work of Hann and Steurer (1996), Neely, Weller, and Dittmar (1997), LeBaron (1999), Dunis, Laws, and Sermpinis (2011) and Korczak and Hernes (2017).

#### 4.2.5.3 Compounded Return

Another method of measuring profits is the *compounded return*, calculated iteratively as:

$$\begin{aligned} \text{CR}(0) &= 1 \\ \text{CR}(y, i) &= \text{CR}(y, i - 1) \frac{y_i - y_{i-1}}{y_i} - c_{y_i} \\ \text{profit}_{\text{CR}} &= k \text{CR}(y, N) \end{aligned} \quad (4.17)$$

where  $\text{CR}(y, N)$  is the compounded return,  $k$  leverage and  $c_{y_i}$  the trading costs associated with the trade  $y_i$ . This formula has the advantage that it can be calculated quickly, and is often simplified further by the removal of trading costs. In the corpus, Allen and Karjalainen (1999) generated their compounded returns with fees. However, Ausloos (2000) and Jordà and Taylor (2011) did not.

#### 4.2.5.4 Annualised Return

Similar to the compounded returned is the *Annualised Return*, or the percentage of profits returned on a per-annum basis. This is calculated relative to the period and frequency of prediction:

$$\text{AR} = \text{equity}_{\text{initial}} \left( \frac{\text{equity}_{\text{final}}}{\text{equity}_{\text{initial}}} \right)^{\frac{f}{N}} \quad (4.18)$$

where  $f$  is the amount of  $N$  events in a year, e.g. 365 for daily, and  $N$  the observed events. Within the corpus, only one paper that reported annualised return, C. Evans et al. (2013), used an out-of-sample period shorter than one year. The others, Hann and Steurer (1996), Neely et al. (1997), Yu, Wang, and Lai (2005), Ghazali, Hussain, and Liatsis (2011) and Sermpinis, Dunis, Laws, and Stasinakis (2012), Sermpinis, Stasinakis, Theofilatos, and

Karathanasopoulos (2015), all used periods substantially longer.

#### 4.2.5.5 Drawdown

Alongside profits, several researchers also reported some variety of risk measurement. Given its bearing on the validity of reported results, *Maximum Drawdown* is often used. This is an important statistic for FOREX speculators, because this metric relates to the viability of an account.

Closely related to Eq. 2.1, where margin is calculated with respect to the account equity, drawdown is defined:

$$\text{drawdown} = \text{balance} - \text{equity} \quad (4.19)$$

Recall that, if the account’s margin reduces to a certain level, a trader’s ability to control their account is reduced—sometimes to the point of forced closure of trades, which is almost always calamitous for the account. Therefore, it is of utmost importance to maximise margin by minimizing drawdown, and thus, knowing the drawdown of an algorithm, regardless of its temporal locality, can be useful for estimating the minimum buy-in for a predictor. On the other hand, without knowing *when* the conditions under which maximum drawdown occurred, it might not be that helpful.

In the corpus, Ghazali et al. (2011) showed that their Dynamic Ridge Polynomial Neural Network had less drawdown than their competitors, as did Bakhach, Tsang, and Jalalian (2016) with their Backlash Agent. Sermpinis, Dunis, et al. (2012) showed that better predictors tended to have lower drawdowns; and de Almeida et al. (2018) considered themselves forced to use drawdown as a metric, noting that this was due to a lack of transparency in Sermpinis, Dunis, et al., whom they wanted to compare their results with.

#### 4.2.5.6 Stirling Ratio

Another application of the drawdown and profits is found in the *Stirling ratio*, defined as

$$R_{\text{Stirling}} = \frac{\text{profit}}{\text{maximum drawdown}} \quad (4.20)$$

with the intention being to report a comparable value. A number of definitions of the Stirling ratio have been given, but this general formula is the most used in the corpus, being used by Hryshko and Downs (2004) and Mendes et al. (2012). The bottom value, maximum drawdown, is sometimes modified to allow for an acceptable level of risk, as in Dempster and Jones (2001), and an annualized version of it also exists (“Sterling ratio”, n.d.).

#### 4.2.5.7 Profit Factor

Another ratio, proposed by Bakhach, Tsang, Ng, and Chinthalapati (2016) is the *Profit Factor*, which is calculated as

$$\text{PF}(r) = \frac{\sum_{i=1}^N r_i, \text{ where } r_i > 0}{\sum_{i=1}^N r_i, \text{ where } r_i < 0} \quad (4.21)$$

with  $r$  being the series of returns—or in words, gross profits divided by gross losses. A value of 1 signifies breaking even, with less being a losing strategy, and more a profitable strategy. Their idea was that this metric would measure the amount of profit per unit of risk.

#### 4.2.5.8 Sortino Ratio

Also referenced by Dempster and Jones (2001), and Bakhach, Tsang, Ng, and Chinthalapati (2016) was the *Sortino Ratio*. This is a measure of risk similar to the Sharpe ratio, but only calculated against returns that are less than a targeted rate of return,  $r_T$ , and therefore dependent on the *downside risk*. It

is defined

$$R_{\text{Sortino}} = \frac{\sum_{i=1}^N r_i - r_T}{\sqrt{\frac{\sum_{i=1}^N \theta(r_i - r_T)(r_i - r_T)^2}{N}}} \quad (4.22)$$

where  $\theta$  is the Heaviside function. Sortino and Forsey (1996), however, recommended the use of a continuous function instead, to allow for a *range* of possibly correct values. This recommendation was not followed in the corpus.

#### 4.2.5.9 Problems with Profit-based Metrics

Given the variety of profit-related metrics, it is difficult to compare their respective predictors. As well as being entirely related to the dataset the results are drawn from, profits are also affected by more than just the predictor accuracy. Trading strategy plays a big part in the performance of a predictor, and whilst it is true that a better predictor should perform better with the same strategy, the reality is that there is no set trading strategy used within the corpus. In fact, it is usual to find a trading strategy defined alongside a predictor, and the separation of the predictor's performance from the trading strategies is impossible. There is also the issue of trading costs, with some researchers reporting results with costs and others without. And there is no evidence any of the researchers took exogenous exchange rates into account—that is, in a multi-Symbol scenario, adjusting the values back to a base currency—either for prediction or assessment. Therefore, despite the ease with which profit-based metrics may be perceived, there is no comparability between them, and their use as a metric is more hype than substance.



### 4.3 (Not So) Tough Benchmarks

Given Rossi's assertion that the Random Walk is still the toughest benchmark to beat, it is not surprising that it is often given as one of the benchmarks in the corpus (27 percent of papers use it). Giles et al. (2001) noted that,

“If equal performance is obtained on the random walk data as compared to the real data then no significant predictability has been found.”

There is, however, some confusion in the corpus as to its specific application: some used no price change to represent Random Walk, whilst others used the same first order change. As has already been noted, researchers have reported mixed results across the board for Random Walk, and so its validity as a benchmark remains intact.

There are other issues besides mere choice of benchmark in the corpus as well. The 2000 paper from Yao and Tan provides several examples of these:

- Their data sets were tiny. Although they used twelve overlapping time periods, each out-of-sample sequence contained only twenty-six points. They also used weekly samples, whereas even then the mode was daily, and much higher frequency data was available. (In their defense, they did acknowledge this).
- Despite most prior literature showing ANNs significantly inferior to Random Walk, they did not make this comparison. They claimed this was because the market was chaotic, but their Hurst interpretation was a little overzealous, with values in  $[0.532, 0.555]$  for their given Symbols. Whilst the case could be made for extremely weak chaos, it would be fragile indeed, given how close these values are to Brownian motion.
- Another issue was the ‘ARIMA’ methods they claimed to use. They were, in fact, ARMAs, and unoptimised—they tested only ARMA(1,1) and ARMA(2,2).

- They presented their results as profits only, however, as has been noted already, these depend on more than just their predictor’s accuracy.
- They only presented the results of their best optimised predictor. This is a form of data-snooping, as in reality the optimal settings would not be known in advance. By the criteria set by White, this renders their results invalid (White, 2000).

Surprisingly, despite the frequency with which this work is cited, no one seems to have picked up on these issues. As was noted earlier in Chapter 2, this paper has often been referenced through another, and it seems likely from the evidence that most citing authors haven’t actually read the paper. It could also be argued that Yao and Tan appeared to say what researchers wanted to hear—“The Random Walk has been beaten by ANNs”—and this then explains its heavy citation. Nevertheless, after Yao and Tan there was an explosion of research into using ANNs for FOREX prediction. However, many papers still suffer the same issues.

## 4.4 Summary

This chapter has looked at the metrics used by a representative sample of papers in the area of FOREX prediction, with a particular emphasis on prediction for speculation. In the 118 papers surveyed, over forty different Symbols were used, with nearly every paper using different datasets, along with different time frames and sample frequencies. There were also over forty different metrics used, and no consistent or standard benchmark. Analysis of these metrics showed little comparability between them, and so the conclusion can be drawn:

**Of the surveyed papers, nearly all recently published papers are incomparable with one another, rendering their *effective* contribution to the application of machine learning to foreign exchange speculation questionable.**

To repeat Rossi's assertion, "the main conclusion that emerges from the literature is that the choice of evaluation method matters, and a lot!". Looking at the metrics in depth, there was an obvious dependence on an underlying error function. The most common error functions used were Absolute Error  $|\epsilon|$  and Squared Error  $\epsilon^2$ , but both these and all other error functions were shown unsuitable for FOREX speculation, due to their dependence on a single 'correct' value. At the same time, profit-based metrics were also investigated, however, they too were found to be incomparable due to the interference of external factors, such as trading strategy and costs. This leads to a second conclusion:

**Of the surveyed papers, those that calculated statistical metrics exclusively focused on close values, despite the limited applicability to speculation. Those that reported profit-based metrics were inconsistent, incomparable and heavily biased. On the whole, all these papers' reported results regarding the prediction of foreign exchange movements for speculation are effectively meaningless.**

It is thus the aim of this thesis to address these conclusions, and so two potential remedies are presented in the next two chapters.

# Chapter 5

## Addressing Incomparable

## Datasets: Introducing Hokohoko

In order to address the problem of incomparable datasets and time periods, this thesis presents a newly-developed Python 3 library, *Hokohoko*<sup>1</sup>. *Hokohoko* is intended for use by researchers in FOREX speculation, providing a consistent framework to test prediction algorithms, backed by a large dataset of historical FOREX rates. As detailed throughout this chapter, *Hokohoko*:

- Features an algorithm that breaks the dataset up into multiple overlapping sets, and can execute predictors on them in parallel.
- Is easy-to-use and cross-platform.
- Is implemented in Python 3, allowing integration with common machine learning libraries such as *Keras*, *TensorFlow* and *PyTorch*.
- Is intended for release into PyPI, with minimal dependencies (*NumPy* and *SciPy*).
- Can be invoked from the command line or imported into a project.
- Offers an API similar to real-world trading software.

---

<sup>1</sup>*Hokohoko* means “to trade, barter, exchange, sell, buy, export, alternate” in Te Reo Māori. The name was determined in conjunction with Associate Professor Te Taka Keegan, Associate Dean Māori of Wananga Putaiao, Te Wananga o Waikato (Division of Health, Engineering, Computing and Science, University of Waikato).

- Offers a number of baseline predictors to compare with, such as ARMA, ARCH and Random Walk.
- Provides standardised parameters drawn from the FOREX corpus.
- Includes a number of metrics, including this thesis' second proposal, the Speculative Accuracy metric (see Chapter 6).
- Includes a simulator function to allow comparison of metrics with profitability, with the correlation providing a measure of *usefulness* for speculation.
- Includes a fully-deterministic Random Number Generator (RNG) to provide repeatable results.
- Is open-source, with the source code hosted on GitHub and documentation online. Documentation is co-located in code, and extractable via *Sphinx*.
- Is unit tested, with test definitions co-located in code via annotations, and a utility included to generate new test cases from new annotations.

This chapter explores a number of key issues related to *Hokohoko* and its implementation.

## 5.1 Data

*Hokohoko* comes with a data file containing seven years of historical per-minute values for fifty Symbols, from 7 July 2012 to 7 July 2019 (see Table 5.1), sourced from the *cTrader* platform. Extra care was taken in the collection of this data to ensure contiguity, with any Symbols not available from the source for the full time-frame removed from the dataset. The data was acquired as per-minute, per-Symbol *Bars*:

$$\text{Bar}[\text{SYMBOL}] = \begin{bmatrix} \text{TIMESTAMP} \\ \text{OPEN} \\ \text{HIGH} \\ \text{LOW} \\ \text{CLOSE} \\ \text{VOLUME} \end{bmatrix} \quad (5.1)$$

Within the data file, each row represents a single minute, and is made up of all the per-Symbol Bars for that minute. Duplicate minutes, where every value except timestamp was identical, were deleted in order to remove weekend and holiday data from the dataset. The data was also examined for corruption from the source, with one solitary impossible (negative) value being replaced by an interpolated one. As a result, the dataset contains just over 2.6 million minutes of contiguous FOREX history, with approximately 130 million individual candlesticks. The resulting data file is 2.6GB, which is too large to be included in the library, and so is made available as a separate download. Instructions for download are included in the online documentation.

Base	Target
AUD	CAD, CHF, JPY, NZD, USD
CAD	CHF, JPY
CHF	JPY
EUR	AUD, CAD, CHF, GBP, HKD, JPY, MXN, NOK, PLN, SEK, TRY, USD, ZAR
GBP	AUD, CAD, CHF, JPY, NOK, NZD, SGD, USP
NZD	CAD, CHF, JPY, SGD, USD
SGD	JPY
USD	CAD, CHF, CZK, DKK, HKD, JPY, MXN, NOK, PLN, SEK, SGD, TRY, ZAR
XAG	USD
XAU	USD

Table 5.1: Symbols available in *Hokohoko*'s dataset, from 7 July 2012 to 7 July 2019. Symbol names are the concatenation of Base and Target, e.g. AUDNZD is the Australian Dollar-New Zealand Dollar exchange rate.

Due to this separation from the library, it is therefore possible to release updated datasets, with the requirement for researchers to state which dataset

they used. For example, a second dataset termed ‘Dataset2’ could potentially be released containing data up to 7 July 2020, or extra Symbols, etc.

## 5.2 Parallel Execution

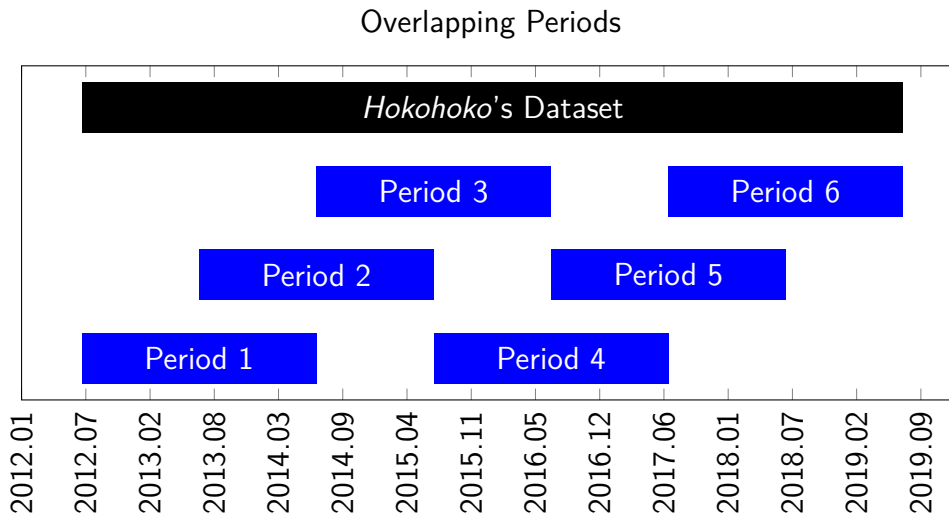
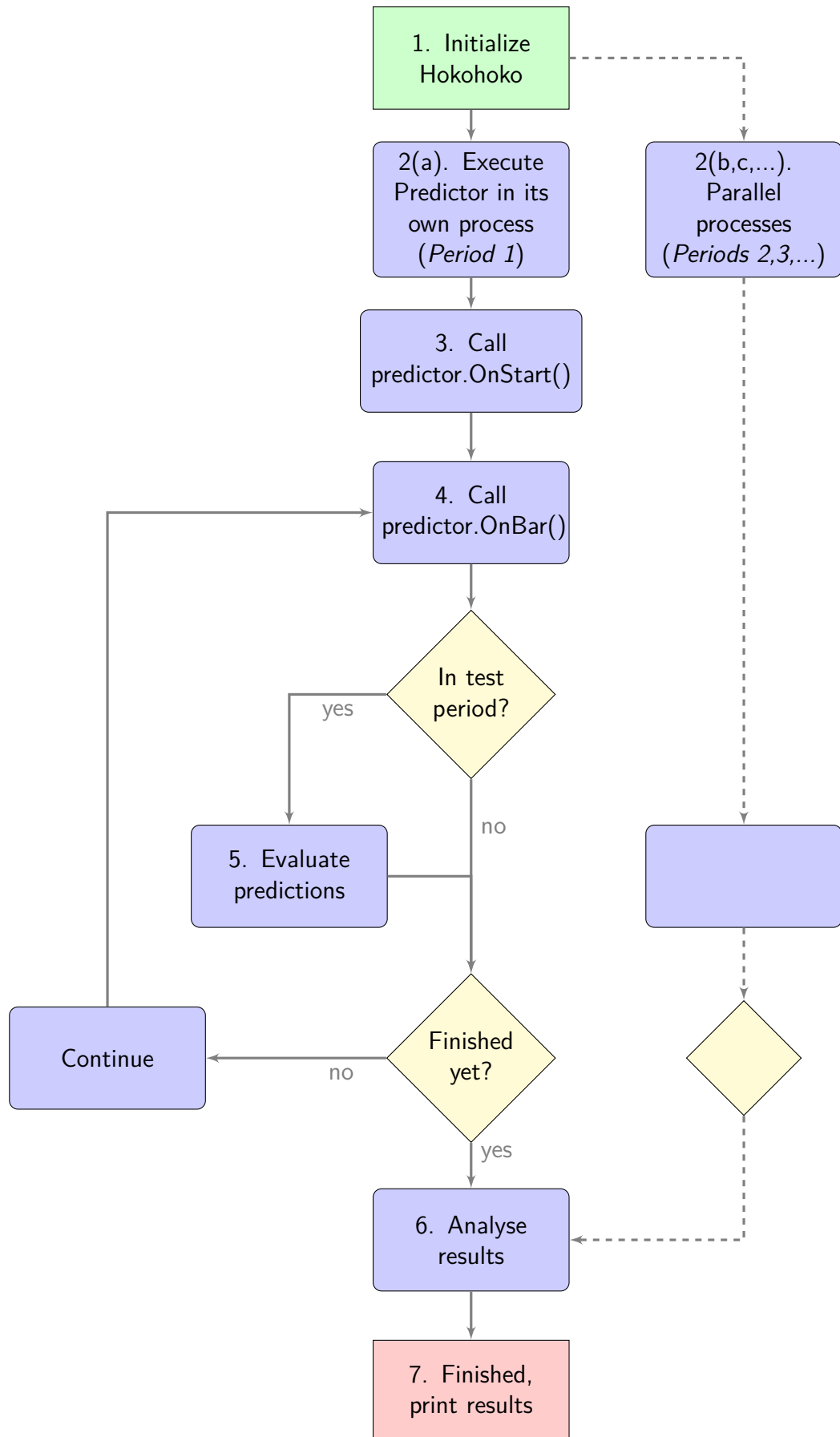


Figure 5.1: An illustration of six overlapping two-year Periods, drawn from *Hokohoko's* dataset. *Hokohoko's* default setting is 256 Periods.

Given the size of the dataset in *Hokohoko*, it is relatively easy to generate many overlapping multivariate time series, allowing for parallel processing of different *Periods*, as in Fig. 5.1. By default, *Hokohoko* splits the data up into a number of these Periods, and executes separate instances of the desired predictor on each Period. Each Period is provided with either the full array, or a user-specified subset, of the available Symbols. While running, each parallel execution is in its own isolated memory space, and the predictor is triggered by *Hokohoko* at specific intervals, simulating the `OnBar()` function available in popular trading platforms. There are a number of configurable parameters (see Section 5.7) that control this behaviour, but it is intended that, for the purposes of benchmarking, the default values are used. A high-level overview of operations is given in Fig. 5.2.

Figure 5.2: Flowchart of *Hokohoko* Execution.



## Initialization

*Hokohoko* relies on three entities to function: a *Data* class, *Predictor* class, and one or more *Assessor* classes. On initialization, it first loads the *Data* class and its associated data file in order to find out what Symbols are available. If a subset of Symbols is requested, *Hokohoko* calculates which Symbols are needed to perform simulation of the subset. It also calculates the start and end dates, from the dataset, for each Period to use. Periods are equal-sized but offset, so the same periods do not exist at different levels of parallelism—that is to say, the Periods in a two-Period run are not a subset of those used in a four-Period run. Finally, it checks that it can spawn the requested Assessor(s), which wait(s) to process results. The full message sequence for a run can be seen in Fig. 5.3.

## Spawn Process

Once the settings for all the Periods have been calculated, *Hokohoko* creates a process pool and queues a new instance of the Predictor for each Period. Whilst expensive on RAM, this is necessary to avoid Python’s Global Interpreter Lock (GIL). For CPU-bound processes, such as *Hokohoko*, the GIL effectively reduces multi-threaded performance to that of a single thread, and so it is necessary to use separate Python processes. Upon process spawn, each Period then loads its specific data from the data source, initialises the local predictor and calls the Predictor’s `OnStart()` method.

## OnStart()

The `OnStart()` method allows the Predictor to configure some initial values by passing in the last CLOSE values prior to the Period’s start. If the Predictor requires some form of persistent state, such as network weights, etc., these should be *declared* in the Predictor’s `__init__()` method and *initialized* here. *Hokohoko*’s deterministic Random Number Generator (RNG) is configured

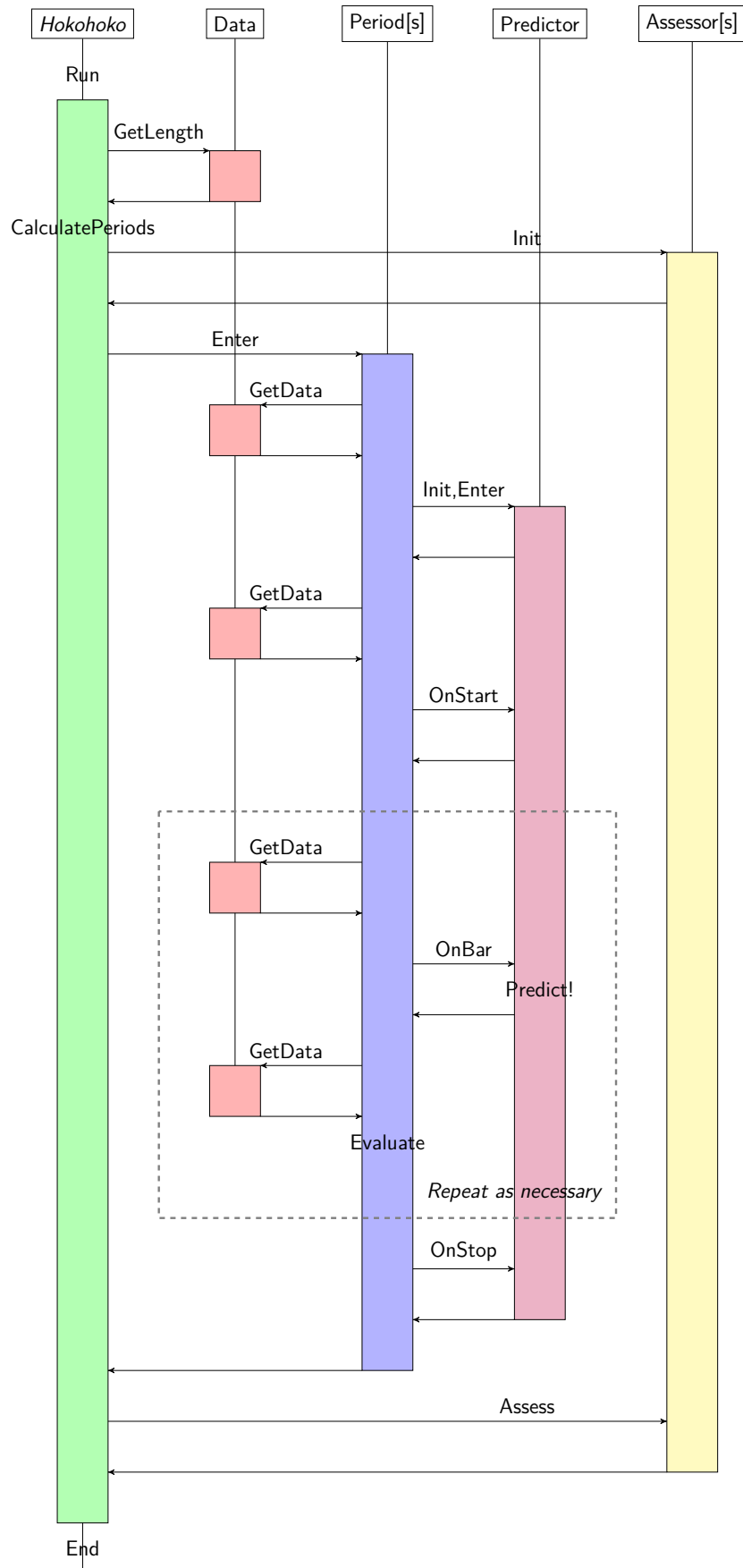


Figure 5.3: Message sequences during *Hokohoko*'s execution. Each vertical line represents a major class within *Hokohoko*.

prior to `OnStart()` being called.

## OnBar()

Once the Predictor has been started, *Hokohoko* then begins to provide it with a stream of data. Iterating through the per-minute data within the dataset, it triggers the Predictor's `OnBar()` method at set intervals, passing in the latest Bar (an array of per-Symbol candlesticks for the previous interval). This is calculated as:

$$\text{Bar}[\text{SYMBOL\_ID}] = \begin{bmatrix} \mathbf{m}_{1,\text{OPEN}} \\ \max([\mathbf{m}_{i,\text{HIGH}}]_{i=1}^N) \\ \min([\mathbf{m}_{i,\text{LOW}}]_{i=1}^N) \\ \mathbf{m}_{N,\text{CLOSE}} \\ \sum_{i=1}^N \mathbf{m}_{i,\text{VOLUME}} \end{bmatrix} \quad (5.2)$$

with  $N$  being the number of minutes in the interval. Upon returning from `OnBar()`, *Hokohoko* then evaluates the predictions made by the Predictor.

Predictions are made within `OnBar` by generating `Order` objects, which can be placed into the internal Orders queue via `self.place_order()` or `self.place_orders()`. An `Order` is defined as:

$$\text{Order} = \begin{bmatrix} \text{SYMBOL\_ID} \\ \text{DIRECTION} \\ \text{OPEN\_BID} \\ \text{TAKE\_PROFIT} \\ \text{STOP\_LOSS} \end{bmatrix} \quad (5.3)$$

with `OPEN_BID`, `TAKE_PROFIT` and `STOP_LOSS` being optional. *Hokohoko* uses `TAKE_PROFIT` as the prediction, and `STOP_LOSS` as risk management. Generally, `OPEN_BID` should be left as `None`, but ambitious researchers may try to min-max their predictions by ascribing an opening value. When the `OPEN_BID` is set to `None`, a `Position` is opened immediately, at the start of the Bar. Otherwise, with a prescribed `OPEN_BID`, an `Order` will be placed,

which will not activate until the exchange rate achieves or transitions past the `OPEN_BID`. Orders that are not activated are discarded at the end of the Bar.

In the same token, `TAKE_PROFIT` and `STOP_LOSS` can also be set to `None`, with the prediction then being restricted to just the direction change over the next interval. Trade direction can be any value of `BUY`, `SELL`, `DONT_BUY` or `DONT_SELL`—with the latter provided to indicate intentionally not trading.

## Evaluation

In *Hokohoko*'s framework, predictors are not aware of future data, so they can only make predictions on a case-by-case basis. They are provided no indication as to whether they are in training or testing, and therefore should return a prediction for each and every Bar. However, only predictions made after the internal test-point will be evaluated.

After the Predictor returns control to *Hokohoko*, at the end of `OnBar()`, *Hokohoko* checks to see if it is in the testing period. If it is, *Hokohoko* checks that every requested Symbol has both `BUY` and `SELL` directions specified, replacing any missing directions with `DONT_BUY` or `DONT_SELL` as appropriate, before iterating through the next Bar.

During the iteration process, *Hokohoko* runs the simulator algorithm on any Orders present. Orders that have specified `None` for their `OPEN_BID` are immediately converted into *Positions*, otherwise they are left in the Orders queue until the `OPEN_BID` is met. Similarly, Positions are closed either when their `TAKE_PROFIT` or `STOP_LOSS` is met, or the Bar ends. *Hokohoko* then returns to `OnBar()` (Section 5.2) until the Period concludes. Whilst *Hokohoko* does not explicitly inform the Predictor on the outcomes of any predictions made, the history is available through `self.account`—otherwise it is up to the Predictor to keep track of its own predictions.

## Analysis

Once all parallel Periods have finished processing, the results are passed into the waiting Assessor(s), as an array of *Accounts*. A *Hokohoko Account* keeps a record of each Period's per-minute balance and equity, as well as a history of all the predictions made and their outcomes. Predictions are stored in the history as *Positions*, with each Position having the structure:

```
Position {
    Order    order;          // The Position's creating Order.
    Bar      future;        // A Bar containing the future Bar.
    Status   status;        // A Status of what happened.
    long     open_time;     // The time the Position opened.
    long     close_time;    // The time the Position closed.
    float    open_rate;     // The exchange rate at opening.
    float    close_rate;    // The exchange rate at closing.
    float    held_value;    // Internal, based on exogeneous rate.
    float    initial_value; // Internal, " " " "
    float    final_value;   // Internal, " " " "
```

Depending on the selected assessor, the results will be analysed per-Period, per-Symbol and overall. `FUTURE` is the Bar over which the Position was evaluated. `Status` can be any of the values `PENDING`, `OPEN`, `CLOSED`, `CLOSED_TAKE_PROFIT` or `CLOSED_STOP_LOSS`. A `PENDING` Position was an Order that never activated; an `OPEN` Position was an Order that activated but did not close; a `CLOSED` Position closed at the end of the future Bar; and `CLOSED_TAKE_PROFIT` and `CLOSED_STOP_LOSS` are Positions that closed during the future Bar, for the stated reason.

## 5.3 Simulation

In order to provide fair comparisons, *Hokohoko's* simulator operates as follows:

- All Periods start with a balance of USD0.
- The Account is at a 1:1 leverage.
- Orders can be placed in any Symbol the predictor registered (the default is all available Symbols).
- All placed Orders are for USD1000. This is the reason why Symbols that cannot be converted are not able to be simulated—every Symbol included in *Hokohoko*'s dataset can be converted to USD, but if the account's base currency is changed, that might not be possible.
- Position values are calculated every minute, with respect to both the currency rate and exogenous rate if required. The current equity is:

$$\text{equity} = \text{balance} + \sum_{i=1}^N (P_{i,\text{held}} - P_{i,\text{initial}}) \quad (5.4)$$

where  $P$  represents all currently open Positions, and  $P_{i,\text{held}} - P_{i,\text{initial}}$  is the  $i$ -th Position's profitability for the moment. Basically, equity is the balance plus the sum of all open Positions' current profits, which may be positive or negative.

- Whenever a Position is closed, its profit is added to the balance.
- The simulator is cost-less: there are no trading costs, such as commission, swap fees or spread.
- The simulator includes slippage, as a function of the dataset—some successive Bars' OPEN differs from their predecessor's CLOSE value; if a Position that is calculated through an exogenous exchange rate closes to a TAKE\_PROFIT or STOP\_LOSS condition, whilst the primary exchange rate is known exactly, the exogenous exchange rate is not, and so the exogenous rate is calculated as an average of the candlestick values for the minute in which the Position closed.

Because the Period accounts all start at zero, margin level is not tracked, and the simulator will not close a malfunctioning account (that is, an account that runs out of margin/equity). This has advantages over using back-testing from trading software in that it is not required to keep the account at a minimum level for operation. It also highlights again the issue with most profit-based metrics, as it is plainly obvious in this scenario that post-selection of a favourable starting balance enables a range of possible values per-metric.

The other important issue to note here is that there are no compounding returns in this model. All the returns are *additive*, which is a) necessary for multi-currency and b) more akin to reality. The Predictor also has no say over other account conditions, such as margins, leverage or costs. As such, predictive performance is divorced from trading strategy.

## 5.4 Availability

*Hokohoko* has been designed from the outset with the intention of open-sourcing. The source code is in a GitHub repository<sup>2</sup> and includes unit testing and scripts to generate documentation and distribution files.

The documentation is written in the source code's docstrings, and a script is included that extracts the documentation for external compilation with *Sphinx*. It is intended that, once released, the documentation will be available at [hokohoko.readthedocs.io](https://hokohoko.readthedocs.io)<sup>3</sup>.

It is intended that *Hokohoko* will also be available in PyPI, the Python Package Index. The scripts to do this have also been included in the source and have been tested on [test.pypi.org](https://test.pypi.org). As has already been mentioned, the data file for *Hokohoko* is too large to be included in the distribution, so it is hosted alongside the documentation. Instructions on how to obtain the data file can be found in the documentation.

<sup>2</sup><https://github.com/nc-bradley/Hokohoko>

<sup>3</sup>In the meantime, it is available at <https://bebecom.co.nz/hokohoko/>.

*Hokohoko* has been tested on Python 3.6-3.8, on Debian, Ubuntu and Windows 10.

## 5.5 Quality Assurance

*Hokohoko* went through a number of iterations during development. There were three main goals kept in sight over these iterations: the software must be *maintainable*; it must be *tested*; and it must be *documented*. Given it is written in Python, every attempt was made to keep it as *Pythonic* as possible.

Maintainability meant keeping the code simple and easy to read, as often as possible. An annotation class was added, which enabled the definition of unit tests within the code body, and Sphinx documentation styles were applied to the docstrings. This design pattern is used so as to facilitate easy understanding of a module or function—the docstrings describe the class or function’s purpose and the unit tests prescribe the desired behaviour. In this manner, all relevant information as to a class or function’s design is co-located with its code.

## 5.6 Invocation

*Hokohoko* can be imported as a module:

```
import hokohoko

from hokohoko.entities import Config, Predictor

class MyPredictor(Predictor):
    ...

config = Config(
    predictor="MyPredictor None",
    assessors=["hokohoko.assessors.Accuracy"]
)
```



```
hokohoko.run(config)
```

Or invoked via the command-line:

```
python -m hokohoko.Hokohoko \
  -P "hokohoko.predictors.SameAsLast --direction EXACT_SAME" \
  -A "hokohoko.assessors.Accuracy ---show-results True"
```

## 5.7 Parameters

*Hokohoko's* behaviour can be controlled through a number of parameters:

- predictor\_class:** A string containing the fully-qualified Predictor name and its arguments. Argument syntax is defined by the Predictor.
- assessors:** A list of strings containing the fully-qualified Assessor names and their arguments, one string per Assessor.
- data\_class:** The Data class to be loaded. This will not typically need to be changed.
- data\_parameters:** The parameters for the data class. If using *Hokohoko's* data file, these will not need to be changed.
- data\_subset:** A comma-separated list of Symbols to restrict analysis to. This can be helpful for reducing development time by loading only a single Symbol instead of all those contained in the data file.
- period\_count:** How many Periods to load. This should only be adjusted for developmental purposes—a benchmark should leave this on the default setting.
- process\_count:** How many processes are allowed to be run in parallel at the same time. This should be adjusted depending on the user's desired CPU/RAM load.

<b>past__minutes:</b>	This is the frequency at which <code>OnBar()</code> will be called. Default is 1440, or one day.
<b>hold__minutes:</b>	How many minutes the evaluation function is allowed to look ahead to check predictions. Ideally, this will match <code>past__minutes</code> , but flexibility is allowed.
<b>load__limit:</b>	Sets the maximum amount of minutes to load from the data file. This can be useful during development to speed up load times, otherwise this should be left at <code>None</code> .
<b>training__minutes:</b>	How many minutes of data is used for the training set. This defaults to 18 months and should only be adjusted for development purposes.
<b>test__minutes:</b>	How many minutes of data is used for the test set. This defaults to 6 months and should only be adjusted for development purposes.
<b>debug:</b>	Sets <code>self.debug</code> within the Predictor and also provides extra information during processing.
<b>profiling:</b>	Enables Python's profiling tool and logs the performance metrics per-Period, which can be useful for finding bottlenecks.

*Hokohoko's* documentation provides more information about these parameters, and the default values can be found in `hokohoko/defaults.py`.

## 5.8 Included Predictors

*Hokohoko* currently includes interpretations of some landmark predictors: a simplified form of Bachelier's Gaussian-based predictor; a predictor based on Yule's AR model; a simple 6-3-1 neural network similar to Yao and Tan's; as well as some simpler applications, such as `DoNothing`, `SameAsLast`, and `Random`. Provision has also been made for the inclusion of Box and Jenkin's

ARMA and ARIMA models, Engle’s ARCH and Bollerslev’s GARCH; however, their implementation is as-of-yet incomplete. Where possible, every effort has been made to use external libraries and thus not reinvent the wheel, however these remain for future work, as noted in Table 5.2.

Table 5.2: Predictors Included in Hokohoko

Predictor	Description
DoNothing	Always says DONT_BUY, DONT_SELL.
RandomWalk	Predicts the same movement as the previous Bar.
Static	Makes static predictions, e.g. TAKE_PROFIT = Bar[CLOSE] + 0.01.
Random	Opens a single Position in a random direction and a DONT_X in the other. Due to the fully deterministic RNG of <i>Hokohoko</i> , this will produce identical results with identical configuration.
SameAsLast	Predicts the same movement (OPEN, HIGH, LOW and CLOSE) as the previous Bar. Also includes options to invert, offset or invert and offset the Bar values.
Bachelier1900	This is a simplified version of Bachelier’s predictor. It keeps track of $P(y_t)$ and $P(y_t - y_{t-1})$ and calculates the maximum likely value as its prediction.
Yule1927	A traditional AR model.
YaoTan2000	A simple FF-NN using the same 6-3-1 structure as reported by Yao and Tan (2000).
Planned	It is intended that additional predictors will be added to <i>Hokohoko</i> in the future. Already planned for inclusion include implementations of ARMA-, ARIMA-, ARCH- and GARCH-based predictors.

As well as providing the above predictors, *Hokohoko* provides an easily

extensible framework by which researchers can ‘plug-in’ their own predictors. The base `hokohoko.entities.Predictor` class can be inherited from, and the following methods overridden:

```

__enter__()
__exit__(exc_type: Any, exc_val: Any, exc_tb: Any)
on_start(bars: Iterable[hokohoko.entities.Bar])
on_bar(bars: Iterable[hokohoko.entities.Bar])
on_stop()

```

Optionally, `__init__()` can also be overridden to set up frameworks for persistent state. As designed, state will not be shared between Periods without the researcher explicitly putting in external structures to do so, so testing the veracity of a researcher’s predictor should be fairly easy.

The Orders structure and deterministic RNG can be accessed through `self`, as `Predictor` inherits from `Random`. It reseeds itself every `Bar` to maintain a consistent RNG per-`Bar` per-`Period`.

## 5.9 Included Assessors

In its current state, *Hokohoko* includes two Assessors. The first is `AccountViewer`, which provides insight into the behaviour of the predictor during its runs—particularly its profitability, via a minute-by-minute history of balance and equity. The second is this thesis’ proposed metric `Speculative Accuracy`, which will be covered next.

# Chapter 6

## Addressing Inappropriate Metrics: Introducing Speculative Accuracy

In Chapter 4, it was noted that most of the statistical and comparative metrics used by researchers within the corpus failed to take into account *what* they were predicting. In particular, for application to FOREX speculation, the values being predicted are a bi-directional range, moving both up and down within the same Bar, yet every researcher based their metric on a single value, CLOSE. This reliance on a single value exacerbated the problem of incomparable metrics, as the CLOSE value was stated to have little relevance to the actual profitability of a predictor for the longer prediction periods used by researchers, and therefore has limited *usefulness*. Consideration of these two problems and their effects gives rise to the question: what constitutes an ideal metric, for the purposes of FOREX speculation?

### 6.1 The Ideal Metric

In order to provide results that are comparable and meaningful, the ideal metric should:

1. Require standardised conditions for researchers to test their algorithms against.

2. Report a single, easily comparable result, with minimal interpretation needed.
3. Be agnostic towards account conditions and trading strategies.
4. Take into account all available information appropriately, reflecting the fact that FOREX movements are represented by a bi-directional range.
5. Correlate to actual performance—that is, it must be *useful*. In application to FOREX speculation, this means a higher score should equate to greater profitability.
6. Set a high, preferably unobtainable, standard.
7. Be easy to obtain and use.

If (1), (2) and (3) are met, then the result can be used to compare predictor performance in otherwise unrelated papers, without the need for researchers to acquire and/or implement other researchers' predictors. If (4) and (5) are met, then practitioners can be confident of the applicability of the metric to real trading. If (6) is met, then there remains within the metric future scope for improvement, until such time as no improvement can be made. And if (7) is met, it should encourage uptake within the research community, so that the community may all benefit from (1)-(6). In order to meet these requirements, *Hokohoko* introduces a new metric, Speculative Accuracy.

## 6.2 Speculative Accuracy: A New Metric

If, for a given prediction, the result  $R$  is:

$$R = \left[ \begin{array}{l} \text{MAX\_PROFIT} \in \mathbb{R}_{\geq 0}, \\ \text{MAX\_LOSS} \in \mathbb{R}_{\leq 0}, \\ \text{ACTUAL\_PROFIT} \in \mathbb{R}, \\ \text{TAKEN} \in \{0, 1\} \end{array} \right] \quad (6.1)$$

then a new metric, Speculative Accuracy, can be defined as follows:

$$SA(R) = \begin{cases} \frac{ACTUAL\_PROFIT}{MAX\_PROFIT}, & \text{if } MAX\_PROFIT > 0, \\ \text{else: } 1.0, & \text{if } MAX\_PROFIT = 0 \wedge TAKEN = 0, \\ \text{else: } \frac{-ACTUAL\_PROFIT}{MAX\_LOSS}, & \text{if } MAX\_LOSS < 0, \\ \text{else: } 0.0 & \end{cases} \quad (6.2)$$

The logic behind this metric is that, for any given future period of time, the conditions of a taken Position may be met, and thus the prediction is correct. However, it is only maximally correct if the predicted TAKE\_PROFIT value matches the maximum observed HIGH/LOW, as appropriate for the prediction's direction for the forecast time period. Therefore, the predictor's performance is defined as the ratio of actual to maximum results. Under these conditions it is possible for a STOP\_LOSS condition to prevent the realisation of profit, but this will reduce the magnitude of any potential loss and raise the average score. The STOP\_LOSS is the maximum drawdown, or risk, forecast as required to allow the realisation of the prediction (or TAKE\_PROFIT). Thus, obtaining a reasonable score requires a balancing act between correctly predicting the TAKE\_PROFIT and STOP\_LOSS values in order to both maximise profit and minimise risk. Examples of this are shown in Fig. 6.1.

It is noted, however, that Eq. 6.2 is

- Unbounded to the negative, with incorrect predictions heavily penalised. That is, either the prediction was incorrect; or a STOP\_LOSS was activated before the prediction could be realised, indicating insufficient drawdown allowance.
- Highly optimistic, such that making a loss for a tiny potential profit is heavily penalised.
- Heavily skewed for correct DONT\_BUY and DONT\_SELL predictions. This is not a major issue, as these predictions are only correct for just over 2% of the provided dataset, and the result is logically correct.

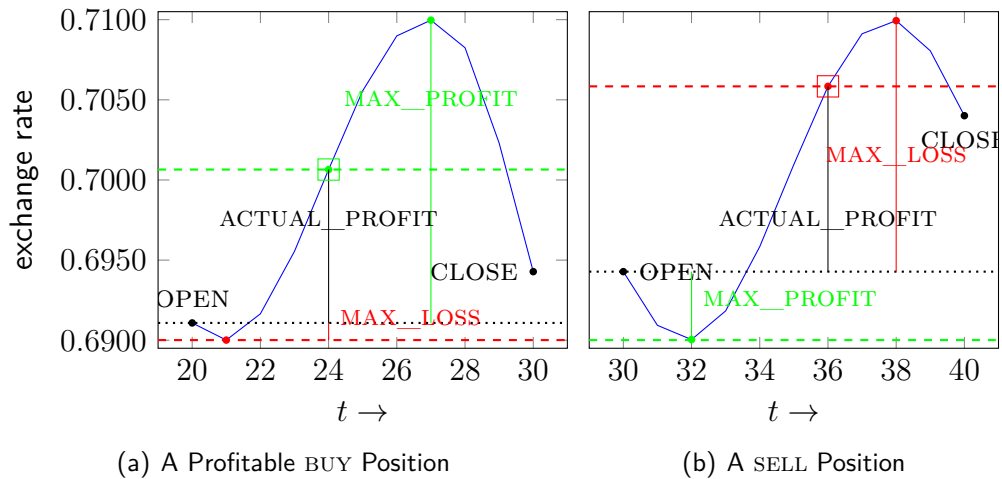


Figure 6.1: Results metrics used by Speculative Accuracy. In these figures, OPEN and CLOSE relate to the opening and closing of a Bar, or candlestick. (a) shows a BUY Position that closed at a profit, whereas (b) shows a SELL Position that closed at a loss. Both examples show how the values MAX\_PROFIT, MAX\_LOSS and ACTUAL\_PROFIT are calculated. In (a), MAX\_PROFIT equates to the difference between HIGH and OPEN, and MAX\_LOSS the difference between LOW and OPEN, for a BUY Position. These are reversed for SELL Positions, as in (b). In these examples, TAKEN = 1, as the Positions were BUY and SELL, respectively. Whilst each of (a) and (b) only correlate to a single Position in these examples, in *Hokohoko* both directions would be calculated for each Bar.

- Not the absolute maximum profit possible, allowing scores above 100%. In order to achieve a score above 100%, the Predictor needs to specify an OPEN\_BID in the opposite direction to the prediction.

For example, say a Predictor specified an OPEN\_BID 0.01 below the current exchange rate for a BUY Order, and then a TAKE\_PROFIT for 0.01 above the current exchange rate. If the exchange rate dropped low enough to activate the Position *and* then rose enough to close it at the TAKE\_PROFIT, *within the Bar*, it would realise a movement of 0.02. Suppose the difference between HIGH and OPEN (or the MAX\_PROFIT) for this Bar was only 0.015 pips, however. In this scenario, the Predictor would receive a Speculative Accuracy score of 133% ( $\frac{0.02}{0.015}$ ).

This is a very risky and difficult prediction to make. But, if a researcher's Predictor can predict the dips and rises that allow for profits greater than MAX\_PROFIT, a higher score can be justified.



Under normal circumstances, where a Predictor uses an OPEN\_BID of **None**, and even in the presence of slippage, this is not possible—as MAX\_PROFIT is then calculated from the same OPEN value as the Position.

Of these, only the heavy bias for a loss from a tiny potential profit is a real issue. This can be alleviated, however, if a normalised version of the first clause were considered; such that, if a per-Period-per-Symbol *weight*  $W$  is defined:

$$W_R = \frac{\text{MAX\_PROFIT}_R}{\max([\mathbf{R}_{P,S}.\text{MAX\_PROFIT}]_1^N)} \quad (6.3)$$

where  $\mathbf{R}_{P,S}$  is the set of results for a specific Period  $P$  and Symbol  $S$ , and  $N$  the number of results within the Period. Then a new metric, *Weighted Speculative Accuracy*, is defined thus:

$$\text{WSA}(R) = \begin{cases} W \frac{\text{ACTUAL\_PROFIT}}{\text{MAX\_PROFIT}}, & \text{if } \text{MAX\_PROFIT} > 0, \\ \text{else: } 1.0, & \text{if } \text{MAX\_PROFIT} = 0 \wedge \text{TAKEN} = 0, \\ \text{else: } \frac{-\text{ACTUAL\_PROFIT}}{\text{MAX\_LOSS}}, & \text{if } \text{MAX\_LOSS} < 0, \\ \text{else: } 0.0 & \end{cases} \quad (6.4)$$

This weighting has the effect of normalising all Periods, because

$$\max(\text{MAX\_LOSS}) = -\max(\text{MAX\_PROFIT})$$

for any given  $\mathbf{R}_{P,S}$ , and reduces the impact of both massively negative and highly positive scores associated with predictions in low profitability intervals. The weighting is only applied to the first equation, when MAX\_PROFIT is positive, as the other equations are already within the range  $[-1, 1]$ .

Other normalisation methods are possible also, such as hyperbolic tangent or symmetric log transformations.

## 6.3 Examples

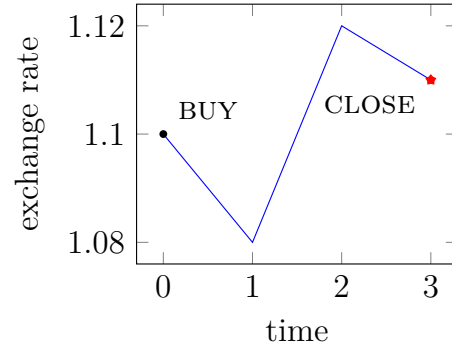
### Example 1: Profit possible, profit made

If:

$$R \begin{bmatrix} \text{MAX\_PROFIT} \\ \text{MAX\_LOSS} \\ \text{ACTUAL\_PROFIT} \\ \text{TAKEN} \end{bmatrix} = \begin{bmatrix} 0.02 \\ -0.02 \\ 0.01 \\ 1.00 \end{bmatrix}$$

then:

$$SA(R) = 0.5$$



In this case, the result is 50% because, whilst correct, the prediction only resulted in half the possible profit. This score has a range of  $[0, \infty]$ . However, getting a score above 1 is difficult, as noted earlier.

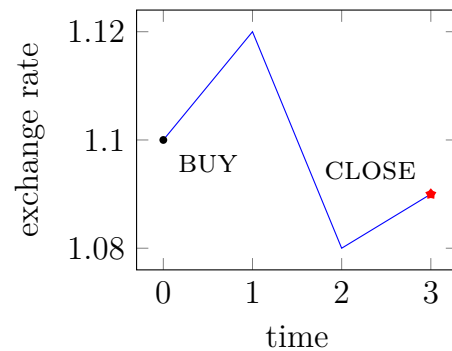
### Example 2: Profit possible, loss made

If:

$$R \begin{bmatrix} \text{MAX\_PROFIT} \\ \text{MAX\_LOSS} \\ \text{ACTUAL\_PROFIT} \\ \text{TAKEN} \end{bmatrix} = \begin{bmatrix} 0.02 \\ -0.02 \\ -0.01 \\ 1.00 \end{bmatrix}$$

then:

$$SA(R) = -0.5$$



In this case, the result is -50%, because the loss made equalled half the potential profit. There are three possible causes for this: either there was no TAKE\_PROFIT; or it was set too high; or the STOP\_LOSS triggered before the TAKE\_PROFIT, and was thus too low. This score has a range of  $[-\infty, 0]$ .

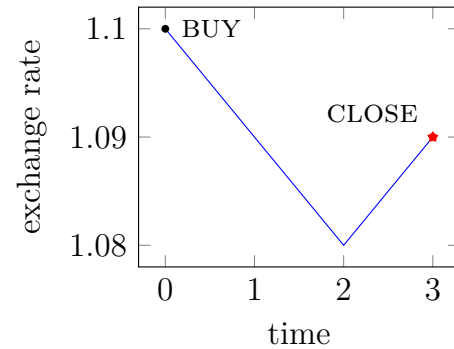
### Example 3: No profit possible, loss made

If:

$$R \begin{bmatrix} \text{MAX\_PROFIT} \\ \text{MAX\_LOSS} \\ \text{ACTUAL\_PROFIT} \\ \text{TAKEN} \end{bmatrix} = \begin{bmatrix} 0.00 \\ -0.02 \\ -0.01 \\ 1.00 \end{bmatrix}$$

then:

$$SA(R) = -0.5$$



In this case, no profit was possible but a position was taken nonetheless. The score is negative, to reflect the fact that it was a loss, but the result is only -50%, as the loss made was half of the potential loss. In this scenario, there is no way of linking the loss to potential profitability, and so the score has a range of  $[-1, 0]$ .

It is notable that this is the same score as in Example 2, though for a different reason. In Example 2, the score of -50% indicated that the prediction was half the size of the correct value, in the wrong direction. In this example, the score is half the size of the maximum possible loss, as profit was impossible for the direction used. The first case is a penalizing score for getting an incorrect prediction, the second is a reward for not losing as much as might have been lost.

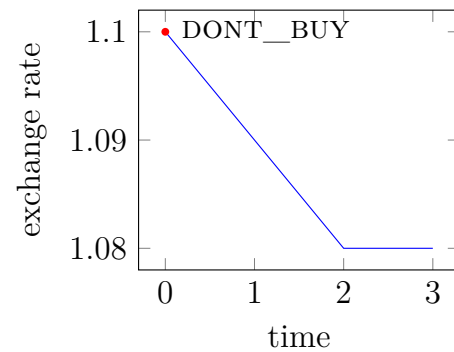
### Example 4: No profit possible, position not taken

If:

$$R \begin{bmatrix} \text{MAX\_PROFIT} \\ \text{MAX\_LOSS} \\ \text{ACTUAL\_PROFIT} \\ \text{TAKEN} \end{bmatrix} = \begin{bmatrix} 0.00 \\ -0.02 \\ 0.00 \\ 0.00 \end{bmatrix}$$

then:

$$SA(R) = 1.0$$



In this case, no profit was possible and no position taken. Because this is the logically correct result, it is given a score of 100%. This score is always given when no action (DONT\_BUY or DONT\_SELL) was taken, and no profit was possible (MAX\_PROFIT = 0).

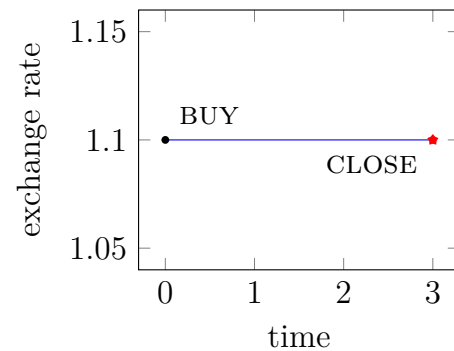
### Example 5: No exchange rate movement, position taken

If:

$$R \begin{bmatrix} \text{MAX\_PROFIT} \\ \text{MAX\_LOSS} \\ \text{ACTUAL\_PROFIT} \\ \text{TAKEN} \end{bmatrix} = \begin{bmatrix} 0.00 \\ 0.00 \\ 0.00 \\ 1.00 \end{bmatrix}$$

then:

$$SA(R) = 0.0$$



In this case, a prediction was made that resulted in neither a profit nor loss, but was not the correct DONT\_BUY or DONT\_SELL. The logically correct result here is given as 0%, which is better than making a loss, but worse than making a profit. In real-world trading, such a score is highly unlikely due to the presence of trading costs.

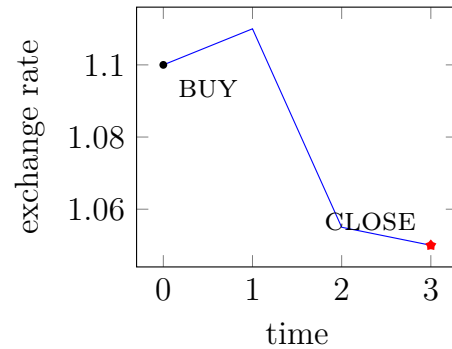
### Example 6: A heavily biased loss

If:

$$R \begin{bmatrix} \text{MAX\_PROFIT} \\ \text{MAX\_LOSS} \\ \text{ACTUAL\_PROFIT} \\ \text{TAKEN} \end{bmatrix} = \begin{bmatrix} 0.001 \\ -0.050 \\ -0.045 \\ 1.000 \end{bmatrix}$$

then:

$$SA(R) = -45$$



Due to the heavy loss made during the given period, with a tiny profit possible, the score obtained is a hefty -4500%. However, suppose that the highest possible profit for the Period and Symbol was 0.1. Then, according to Eq. 6.3  $W = \frac{0.001}{0.1}$ , and using Eq. 6.4 gives a score of  $SA(R) = -45\%$ . This is still not a good score, however it no longer skews the result distribution quite so much as the unweighted score does. Fig. 6.2 shows the effect of normalisation on the score distribution.

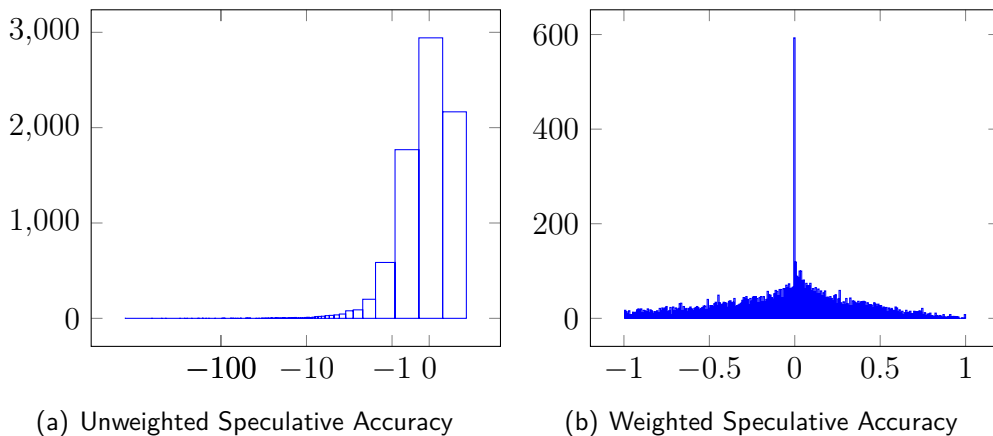


Figure 6.2: The effect of normalisation on randomly distributed speculative accuracy results. (a) is the distribution of raw Speculative Accuracy scores. Note the long tail to the negative—these are caused by significant losses in the presence of minute possible profits. (b) is the same scores, normalised by the maximum possible profit. The data was 8192 MAX\_PROFIT and MAX\_LOSS values with ACTUAL\_PROFIT between them, randomly generated with a 3% chance of not taking a Position.

## 6.4 *Hokohoko* and Speculative Accuracy?

Combined, *Hokohoko* and Speculative Accuracy meet the requirements for an ideal benchmark, as listed in Section 6.1. *Hokohoko* meets requirements (1), (3) and (7) by being a standardised benchmark, providing a fixed trading strategy and being readily available. Speculative Accuracy meets requirements (2), (4), (5) and (6) by reporting a single comparable result that takes into account the bi-directional future of FOREX speculation, setting a high standard of comparison with near-maximum profitability, and correlating to actual performance.

This metric can be used either on its own, or as an error function in a comparative metric, such as Diebold-Mariano's. In *Hokohoko*'s current implementation, the mean is reported per-Period, per-Symbol and overall. The raw results are also made available by passing "`--show-results True`" to the assessor, so that researchers may perform their own statistical analysis.

The remainder of this thesis investigates the *usefulness*, or correlation with profitability, of Speculative Accuracy and other common metrics, using *Hokohoko*.

# Chapter 7

## Experiment Methodology

This thesis has made a number of claims regarding the metrics used to measure the efficacy of various predictors for the purpose of FOREX speculation:

1. Due to the exclusive use of CLOSE values in the underlying error functions, the statistical metrics used produce results that have no correlation to profitability.
2. Directional metrics based on CLOSE also have little correlation to profitability.
3. Papers that target speculation, as indicated by reporting of profitability, are incomparable due to inconsistencies in account settings, time periods and dataset selection.

In response to these issues, this thesis has offered two solutions:

1. The library *Hokohoko* has been developed to provide a consistent framework for benchmarking FOREX predictions, including an account settings- and trading strategy-agnostic simulator.
2. A new metric, *Speculative Accuracy*, has been developed and is argued to be more *useful* for FOREX prediction. For the purposes of benchmarking for speculation, *usefulness* is defined as correlating to profitability—that is, a useful benchmark will be able to provide a

relative measure of the profitability of a prediction algorithm when applied to FOREX speculation.

It is hypothesized that Speculative Accuracy provides a better measure of a predictor's performance, in the context of FOREX speculation, than those metrics commonly used in the FOREX corpus.

This chapter thus details experiments run to test this hypothesis, using *Hokohoko*, with the *usefulness* of Speculative Accuracy and various other metrics investigated. It is structured as follows: first, the Predictors utilised are explained in detail, followed by the metrics tested. Then the aims of the experiments are laid out, followed by the experiment methodologies. The results are presented and discussed in the next chapter.

## 7.1 Predictors

A number of simple predictors are included in *Hokohoko* (see Table 5.2), and these were used for the experiments.

### Random Walk

For the purposes of these experiments, the Random Walk prediction is defined:

$$\begin{aligned}
 \hat{y} &= y_t + \delta y_t \\
 &= y_t + (y_t - y_{t-1}) \\
 &= 2y_t - y_{t-1}
 \end{aligned}
 \tag{7.1}$$

or, the first order difference of the next Bar will be the same as the current Bar. If  $\hat{y} > y_{t,\text{CLOSE}}$ , then a BUY Order is generated; if  $\hat{y} < y_{t,\text{CLOSE}}$ , then a SELL Order is generated; otherwise no Orders are placed. In placing the Order, the TAKE\_PROFIT is set to  $\hat{y}$ , with no STOP\_LOSS specified.



## Random

The Random predictor uses *Hokohoko*'s inbuilt deterministic RNG to generate a value  $\hat{y}$  in  $[0,1]$ . If  $\hat{y} \geq 0.51$ , it generates a BUY Order; if  $\hat{y} \leq 0.49$ , it generates a SELL Order. Unlike the other predictors in *Hokohoko*, it does not specify either TAKE\_PROFIT or STOP\_LOSS values. Consequently, it makes a profit if the guessed direction is correct, and a loss otherwise. The interval between (0.49, 0.51) allows the predictor a 2 percent chance of making no prediction, matching the observed non-movement rate in *Hokohoko*'s dataset.

## Autoregression

*Hokohoko*'s AR predictor keeps track of the Bars' HIGH and LOW values in two windowed queues. It uses `statsmodels.tsa.ar_model.AR` to fit the data to AR models, and then uses these models to make 1-step-ahead predictions for each of HIGH and LOW,  $\hat{y}_{\text{HIGH}}$  and  $\hat{y}_{\text{LOW}}$ , respectively. Based on the predicted values, it then chooses actions to take: if  $\hat{y}_{\text{HIGH}} > y_t$ , it places a BUY Order with TAKE\_PROFIT set to  $\hat{y}_{\text{HIGH}}$ ; if  $\hat{y}_{\text{LOW}} < y_t$ , it places a SELL Order with TAKE\_PROFIT set to  $\hat{y}_{\text{LOW}}$ . In both scenarios, STOP\_LOSS is set to None.

## Static

The Static Predictor is not really a predictor, but a trading strategy. All it does is place an Order each way, BUY and SELL, with a TAKE\_PROFIT set at  $\pm 0.01$  as appropriate. In the interests of keeping the comparison with the other Predictors fair, STOP\_LOSS was left as None.

## Artificial Neural Network

The ANN used in these experiments is a 6-3-1 Feed-Forward Neural Network loosely based on Yao and Tan (2000) and Nielsen (2015). Like Yao et al., the network was given moving averages of the past values, of lengths 5, 10, 20, 60, and 120 each, plus the last observed value, as input. As the predictor

progressed through its run, the already observed data was split into a training and a validation set, with a 7:2 split. In this application, the selected activation function was the hyperbolic tangent  $\tanh(\epsilon)$ , with the derivative  $1 - \tanh(\epsilon)^2$  used for back-propagation. The ANN used Root Mean Squared Error to test for convergence and up to 1000 training runs per-Bar. Early exit was allowed if the standard deviation of the RMSEs for a moving window dropped below a certain threshold. The weights and biases were initialized from a Gaussian distribution with  $\mu = 0$  and  $\sigma = 1$ . The ANN retrained every Bar, using the previous weights and biases as the starting point for training. Within the validation sets, the ANN achieved an average Correct Direction score of 67%, and a RMSE of 0.00053. No efforts were made to optimise the hyper-parameters, other than testing for convergence.

## SameAsLast

*Hokohoko's* **SameAsLast** predictor is a variation on Random Walk which, based on a configuration flag, places BUY and SELL Orders as appropriate. The flag, `direction`, can have one of four values: `EXACT_SAME`, `EXACT_OPPOSITE`, `OFFSET_SAME` and `OFFSET_OPPOSITE`. The first, `EXACT_SAME` specifies that the next Bar will be identical to the last. For a prediction, this means that the HIGH and LOW values are predicted to be the same. The `OPPOSITE` flag signifies the predictions should be inverted around the OPEN value. The `OFFSET` flag instructs the shifting of the HIGH and LOW relative values from the OPEN value to the CLOSE value, with `OFFSET_OPPOSITE` indicating both a shift and inversion. Based on the predicted HIGH and LOW, **SameAsLast** opens BUY and SELL Orders as appropriate, or `DONT_BUY` and `DONT_SELL` otherwise. Whilst ordinarily a `STOP_LOSS` value would be set, for the experiments here this was set to `None`.

## Bachelier

Based on the earliest known paper on mathematical analysis of market movements, the **Bachelier** Predictor included in *Hokohoko* keeps track of two sets of probabilities: the probability of a given  $y$ ,  $P_y$ , and the first order differences  $\delta y$ ,  $P_{\delta y}$ . In order to make a prediction, it searches through both probabilities, seeking to maximise  $P(\hat{y}) = P_y(y_t + \delta y)P_{\delta y}(\delta y)$ , where  $\hat{y} = y_t + \delta y$ . If there are multiple  $\hat{y}$  with equal probability, it averages the values. Once it has identified the exchange rate with maximum probability, this value is set as the `TAKE_PROFIT`, and a `BUY` or `SELL` Order is placed depending on the relative position of the `TAKE_PROFIT` with the `CLOSE` value of the previous Bar.

## 7.2 Metrics

Twenty metrics were tested in the experiments, including Correct Direction, F-measure, Error-based variants and 6 variants of Speculative Accuracy.

### Profits-per-Position

In order to assess the applicability of a metric, the experiments compared each metric's score with the *Profit-per-Position* ( $P^3$ ). Three variants of  $P^3$  were used: *Realised Profits*, which is the profits simulated by *Hokohoko* using GBPUSD; *Maximum Profits*, which were the highest possible profit for the particular Bar for GBPUSD; and *EURGBP*, which were the realised profits in the presence of an exogenous exchange rate, in this case back to USD from EUR. Despite this thesis' critique of profits as a metric, it is appropriate to use them in this scenario, for the following reasons:

1. The primary goal of these experiments is to measure the *usefulness* of a given metric. Usefulness is determined by a metric's ability to differentiate between different predictors' profitability.

2. Many of the arguments against the use of profits do not apply in this scenario. *Hokohoko* provides consistent trading account settings for all the experiments, and a consistent trading strategy.
3. *Hokohoko* provides the ability to test across multiple time periods, so there is no data selection-bias.

Table 7.1 details the definition of each metric, within the experiments.

Table 7.1: Metrics used in Experiments

Metric	Description
Observed Direction	<p>This is the actual direction the exchange rate moved.</p> <p>It is defined:</p> $OD = \begin{cases} 1, & \text{if CLOSE} > \text{OPEN} \\ 0, & \text{if CLOSE} = \text{OPEN} \\ -1, & \text{if CLOSE} < \text{OPEN} \end{cases}$
Correct Direction	<p>Measures if the predicted direction was correct.</p> $CD = \begin{cases} 1, & \text{if CLOSE} > \text{OPEN} \wedge \text{DIRECTION} = \text{BUY} \\ 1, & \text{if CLOSE} < \text{OPEN} \wedge \text{DIRECTION} = \text{SELL} \\ 0, & \text{otherwise} \end{cases}$
F-measure	<p>Unlike all the other metrics, this one requires full runs. Therefore it was calculated each Period and compared with the sum of all the profits for the Period. Its definition can be found in Table 4.5.</p>
$\epsilon_{\text{CLOSE}}$	<p>The standard error function <math>\hat{y} - y</math>. In this case, the error is based on the CLOSE rate of the future Bar.</p>
$ \epsilon_{\text{CLOSE}} $	<p>Absolute Error for the CLOSE rate.</p>
$\epsilon_{\text{CLOSE}}^2$	<p>Squared Error for the CLOSE rate.</p>

Metric	Description
$\frac{\epsilon_{\text{CLOSE}}}{\delta y}$	This is the Percentage Error of the standard error for CLOSE rate. In this case, the percentage is calculated relative to the first-order movement. The reason for this is that, due to the relatively tiny movements exchange rates make to their value, basing a percentage on absolute values is effectively just rescaled (and slightly distorted) standard errors.
$\frac{ \epsilon_{\text{CLOSE}} }{\delta y}$	Absolute Percentage Error for the CLOSE rate.
$\left(\frac{\epsilon_{\text{CLOSE}}}{\delta y}\right)^2$	Squared Percentage Error for the CLOSE rate.
$\epsilon_{\text{H/L}}$	This is the standard error function, except based on HIGH for BUY Positions, and LOW for SELL Positions.
$ \epsilon_{\text{H/L}} $	Absolute Error for the appropriate CLOSE or HIGH rate.
$\epsilon_{\text{H/L}}^2$	Squared Error for the appropriate CLOSE or HIGH rate.
$\frac{\epsilon_{\text{H/L}}}{\delta y}$	Percentage Error based on the appropriate HIGH or LOW rate.
$\frac{ \epsilon_{\text{H/L}} }{\delta y}$	Absolute Percentage Error for the appropriate HIGH or LOW rate.
$\left(\frac{\epsilon_{\text{H/L}}}{\delta y}\right)^2$	Square Percentage Error for the appropriate HIGH or LOW rate.
SA	Speculative Accuracy, as defined in Eq. 6.2.
SA <sub>-1</sub>	This is a variant of Speculative Accuracy, where taken Positions with no profit possible are given a score of -1.

Metric	Description
$\text{symlog}_{10}(\text{SA})$	<p>In order to address the long tail of Speculative Accuracy, this is a variant that uses a base-10 symmetric log outside of <math>[-1,1]</math> to condense the SA score:</p> $\text{symlog}_{10}(\text{SA}) = \begin{cases} -\log_{10}(-\text{SA}) & \text{if } \text{SA} < -1 \\ \text{SA} & \text{if } -1 \leq \text{SA} \leq 1 \\ \log_{10}(\text{SA}) & \text{if } \text{SA} > 1 \end{cases} \quad (7.2)$
$\text{symlog}_e(\text{SA})$	As $\text{symlog}_{10}(\text{SA})$ , but with base $e$ .
$\frac{\tanh(\text{SA})}{\tanh(1)}$	This is a variant that normalises Speculative Accuracy using the tanh function. In order to normalise such that $(-1,1)$ remains $(-1,1)$ , the result is divided by $\tanh(1)$ , with a new range of $(-1.31,1.31)$ possible.
WSA	Weighted Speculative Accuracy, as defined in Eq. 6.4.

## 7.3 Experiment Aims

The primary aim of these experiments is to measure the *usefulness* of various metrics for application to FOREX speculation, via exploration of the following questions:

1. Is there any correlation between the observed direction of an exchange rate movement and realised profit for a given Bar, where realised profit is the equity change in the Position over the course of the Bar.
2. Is there any correlation between the observed direction of an exchange rate movement and maximum profitability for a given Bar, for the trading direction specified by the Order.

3. Is there any correlation between the metrics listed in Table 7.1 and realised profit.
4. Is there any correlation between the metrics listed in Table 7.1 and maximum profitability.

A smaller, secondary aim is to investigate the influence of metric choice on predictor ranking—that is, ordering predictors by performance.

## 7.4 Experiment: Correlation of Metrics to Profit-per-Position

In order to test the usefulness of the metrics, a new Assessor containing all of the metrics was written (see Appendix B) and run against the `RandomWalk` predictor:

```
python -m hokohoko.Hokohoko
    -P "hokohoko.assessors.RandomWalk"
    -A "hokohoko.assessors.Experiments"
    -n 512
    [-S "GBPUSD"]    # Run 1, single Symbol with Account currency.
    [-S "EURGBP"]   # Run 2, single Symbol with exogenous currency.
```

Two Symbols were used because *Hokohoko*'s simulator converts the traded currency back to its internal Account's base currency, USD. This conversion invokes an exogenous exchange rate, which none of the tested metrics take into account. The experiments with EURGBP therefore reveal the comparative strength of metric correlation for application to multi-currency accounts. This experiment was run with 512 overlapping Periods and the results were collated on a per-Position basis. With each run consisting of a test period of 183 days, this gives up to 187,392 Positions to test. Because several of the predictors only make predictions in a single direction, `DONT_BUY` and `DONT_SELL` Positions were ignored in the metric testing.

The results are presented graphically in Chapter 8 as score versus profit for each of the conditions being tested, with the overall result for each test being Pearson's Correlation Coefficient, as calculated by `numpy.corrcoeff`.

## Hypotheses

This thesis has claimed that there is little correlation between direction or CLOSE values and maximum profitability, particularly at longer intervals. Therefore, the expected results for the daily Bars provided by *Hokohoko* are as follows:

1. It is expected that there will be limited correlation between observed direction and maximum profitability.
2. It is expected that there will be positive correlation between Correct Direction and reported profits, as a positive profit value likely originates from a correct direction prediction. In the presence of exogenous Symbols, as provided by *Hokohoko*, this may be obfuscated somewhat. However, positive correlation is still expected.
3. It is expected there will be little to no correlation between Correct Direction and maximum profitability.
4. Being based on Correct Direction, it is expected that there will be observable correlation between F-measure and realised profit, but little to none between F-measure and maximum profitability.
5. There will be little to no correlation between CLOSE-based error metrics and realised profit.
6. There will be little to no correlation between CLOSE-based error metrics and maximum profit.
7. There will be little to no correlation between HIGH/LOW-based error metrics and realised profit.



8. There will be positive correlation between HIGH/LOW-based error metrics and maximum profit.
9. As it was designed for this task, it is expected that all variants of Speculative Accuracy will show strong correlation with realised profits.
10. It is also expected that all variants will show strong correlation with maximum profits.
11. Weighted Speculative Accuracy is expected to have the highest correlation, as the weighting makes its results linear, per-Period.

## 7.5 Experiment: Ranking of Predictors by Metric

In this experiment, *Hokohoko* was run with the Predictors in Section 7.1, with the scores collated per-Predictor-per-metric and ranked. Rankings were calculated by mean  $\mu$  for the metrics that had positive correlation (such as Directional Correctness, F-measure and Speculative Accuracy), and absolute mean  $|\mu|$  for the error-based metrics.

### Expected Results

For the metrics that have no correlation to profits, it is expected that the observed rankings will show no clear pattern, in themselves or against simulated profits. For the metrics modified to use HIGH and LOW, it is expected that some consistency in rankings will be observable. For Speculative Accuracy, it is expected that the rankings will most closely match simulated profits, and therefore, usefulness.

This chapter has outlined the experiments used to test the various metrics and Predictors included in *Hokohoko*. The results of these experiments are presented and discussed in the next chapter.

# Chapter 8

## Results and Discussion

This chapter contains the results and discussion for the experiments outlined in the previous chapter:

For Experiment 1, the results are presented in graphical form, with each graph consisting of 2,500 randomly selected Position/profit samples from the generated results. From the full results, comprising approximately 93,000 Position/profit samples, the correlations between each metric's score, realised profit and maximum profit was calculated and are listed in Table 8.1. The metrics are then ranked by absolute average correlation  $|\bar{\rho}|$  in Table 8.2, then discussed.

For Experiment 2, the rankings of each tested Predictor versus metric scores are presented in Tables 8.3 and 8.4, then discussed.

Finally, this chapter ends with a discussion of the observed results, with a particular interest in validating or refuting the various hypotheses tested.

## 8.1 Results: Correlation of Metrics with Profits

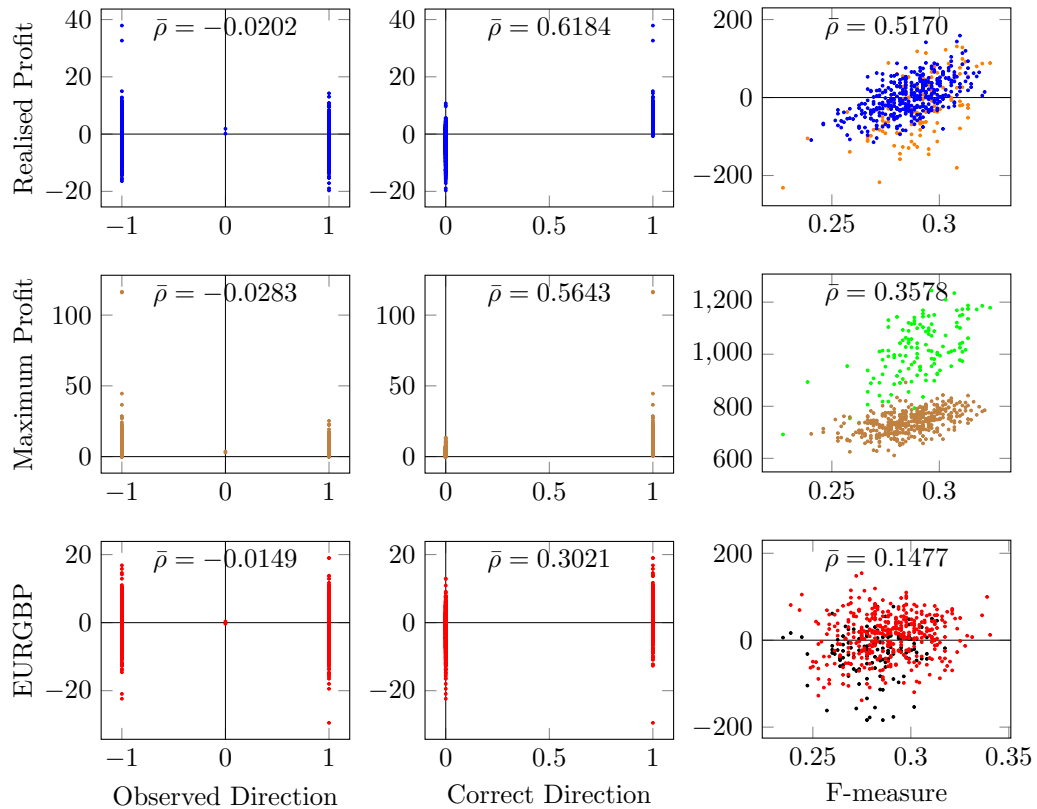


Figure 8.1: Correlation of direction-based metrics, with respects to realised and maximum profit for GBPUSD, and realised profit for EURGBP. For Observed and Correct Direction, realised and maximum profits are per-Position; for F-measure, they are the sum per-Period. The orange, green and black sets in the F-measure figures are Periods 142-269, which span the 2016 Brexit referendum.

### 8.1.1 Performance of Directional Metrics for Random Walk Predictions in *Hokohoko*

It was hypothesised that there would be little to no correlation between the observed direction of exchange rate movements and profitability, and this is confirmed in Fig. 8.1, with both Symbols' correlation coefficients close to 0 for both realised and maximum profits. It was also hypothesised that there would be a clear correlation between Correct Direction and realised profits, with weaker correlation for the exogenous case, and this also was confirmed. However, there is only a small difference between  $\bar{\rho}$  for realised and maximum profits, suggesting the hypothesis of little to no correlation

between Correct Direction and maximum profitability incorrect—at least for the daily intervals used in this experiment. Even so, this might not be entirely true, as examination of the F-measure correlation reveals:

Due to its reliance on Correct Direction, correlation is also observed in the F-measure metric for realised profits and GBPUSD. And while the reported  $\bar{\rho}$  for maximum profits appears weaker, it is apparent that there are two distinct clusters of profitability, which have been highlighted. The cluster with higher maximum profits consists of Periods 142-269, which spanned the 2016 Brexit referendum, and its accompanying market volatility is reflected in increased potential profits. This is only noticeable in the F-measure results because they consist of summation per-Symbol-per-Period, whereas the other metrics' profits are per-individual prediction. With the Brexit cluster accounting for 25% of the tested Periods, this shows how F-measure, and thus Correct Direction, lacks correlation with maximum profits. As the variance in potential profitability increases, the correlation with Correct Direction and F-measure decreases. Therefore, the hypothesis of little to no correlation between Correct Direction and maximum profits is confirmed, in the general case.

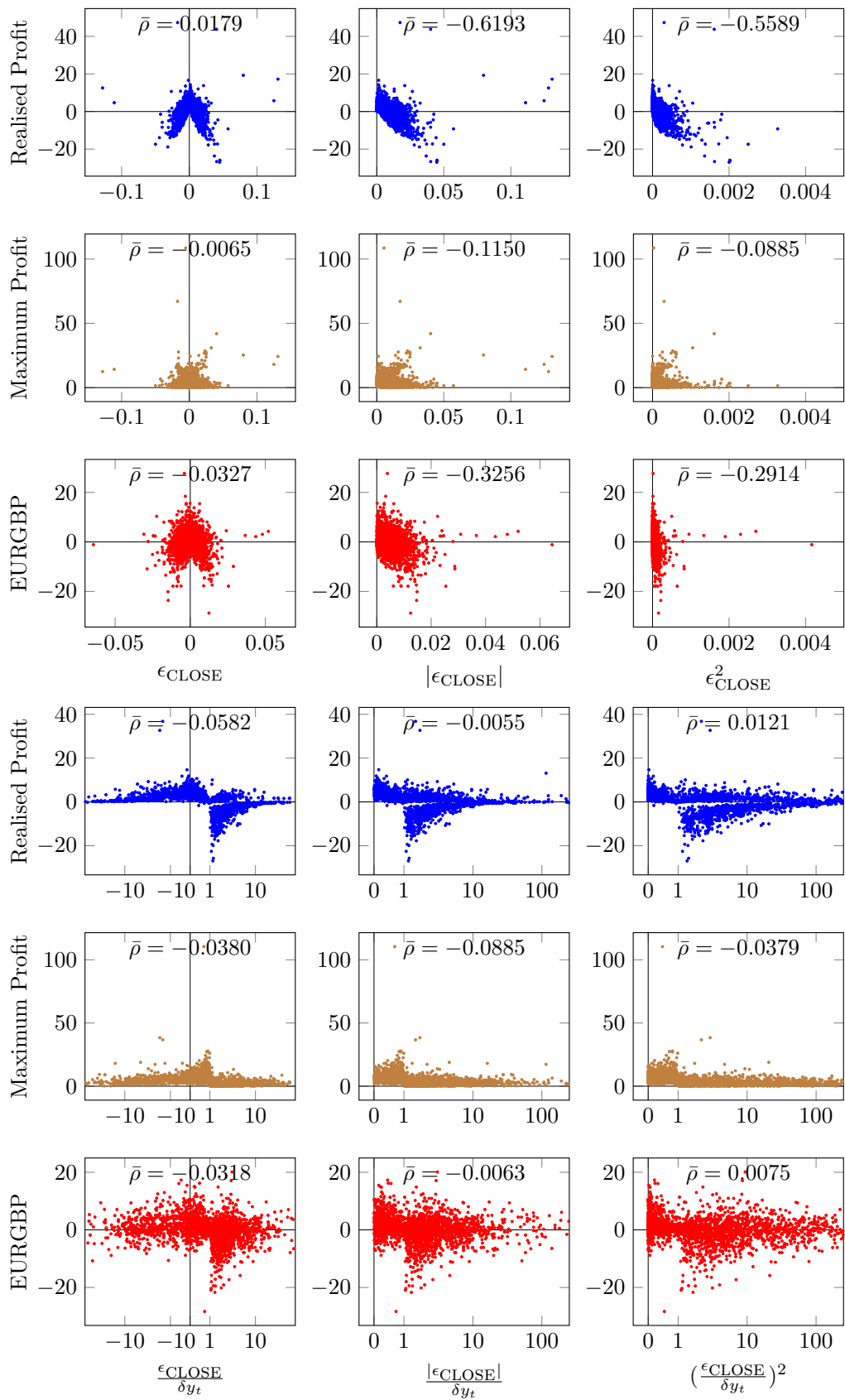


Figure 8.2: Correlation of CLOBASE-based metrics, with respects to realised and maximum profit for GBPUSD, and realised profit for EURGBP.

### 8.1.2 Performance of close-based Error Metrics for Random Walk Predictions in *Hokohoko*

In Fig. 8.2, it can be seen that  $\epsilon_{\text{CLOSE}}$  trends towards profitability as it approaches 0, from both directions, for both realised and maximum profits. This is not reflected in the correlation coefficient, however, with  $\bar{\rho} < 0.02$ , as the distribution is the sum of two trends. By rectifying  $\epsilon_{\text{CLOSE}}$ ,  $|\epsilon_{\text{CLOSE}}|$  and  $\epsilon_{\text{CLOSE}}^2$  merge these two trends together, and it is thus apparent that all three metrics are correlated to the realised profits. However, they show little correlation with maximum profits.

For the percentage error metrics, the situation is slightly more complex. Whilst there is correlation present, it is split into parts across the spectrum of possible scores, with significant overlap between profits and losses above  $\frac{\epsilon_{\text{CLOSE}}}{\delta y} = 1$ . This overlap reflects the fact that trades which conclude below CLOSE may still be profitable. In this case, rectification does not help, with both absolute and squared methods increasing the overlap to almost the entire spectrum of results and annihilating any correlation that might have been present.

It was hypothesised that there would be little to no correlation between CLOSE-based error metrics and either realised or maximum profits. That has been confirmed for the percentage-based error functions, as well as  $\epsilon_{\text{CLOSE}}$ . However, both  $|\epsilon_{\text{CLOSE}}|$  and  $\epsilon_{\text{CLOSE}}^2$  exhibited correlation with realised profits, suggesting that this hypothesis is only partially correct.

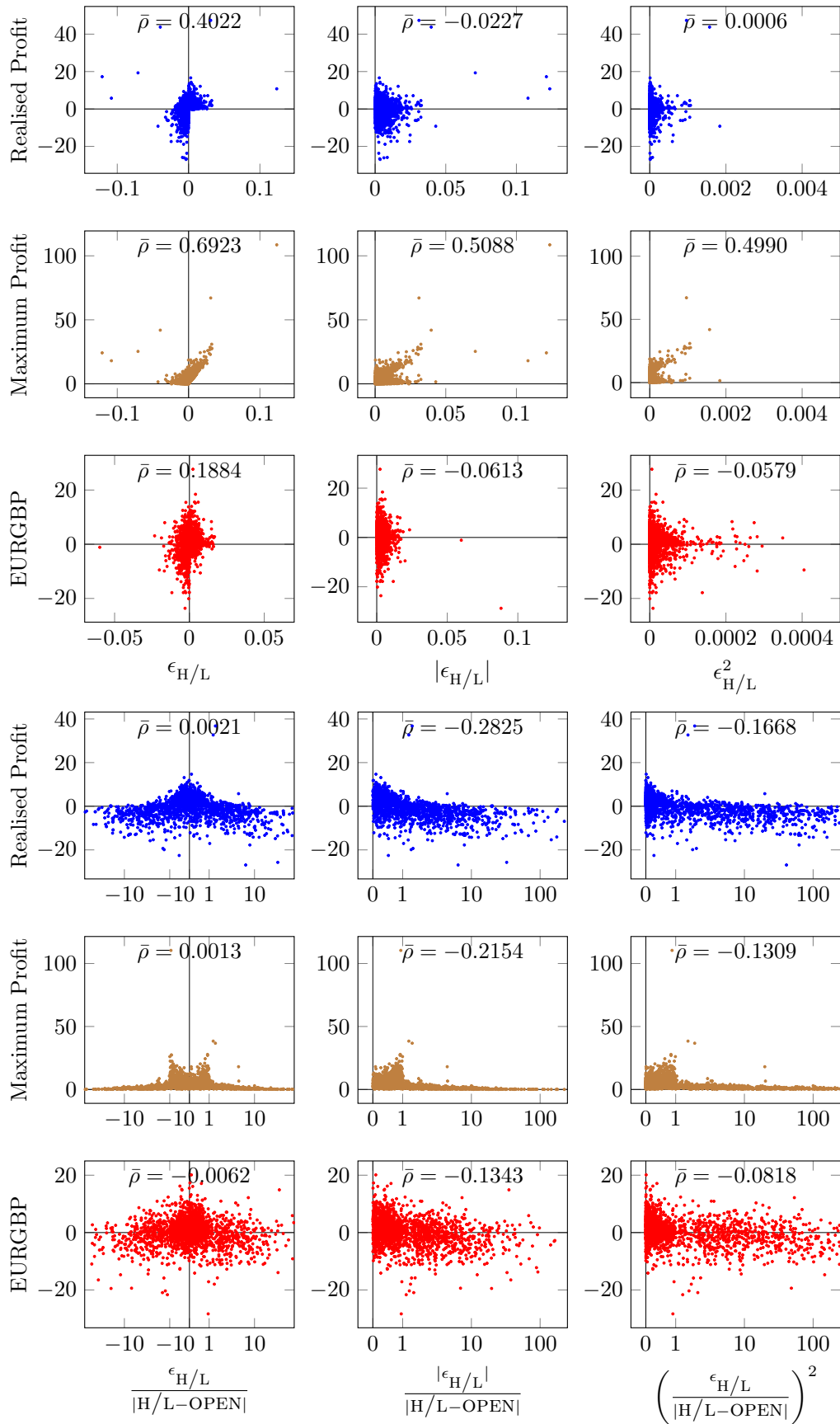


Figure 8.3: Correlation of HIGH/LOW-based metrics, with respects to realised and maximum profit for GBPUSD, and realised profit for EURGBP.

### 8.1.3 Performance of high/low-based Error Metrics for Random Walk Predictions in *Hokohoko*

While Fig. 8.3 shows clear correlation between  $\epsilon_{H/L}$ ,  $|\epsilon_{H/L}|$  and  $\epsilon_{H/L}^2$ , with respect to maximum profits, this level of correlation does not extend to the realised profits. Only  $\epsilon_{H/L}$  shows any level of correlation, and that is only because of the high correlation above 0. However,  $\epsilon_{H/L}$  is always profitable above 0, as this occurs when TAKE\_PROFIT is set too high. In trading, such positions would not do as well as the prediction would suggest, as they would be left to conclude at either CLOSE or a value set by a trading strategy, such as a trailing stop loss. Therefore, the apparent correlation between  $\epsilon_{H/L}$  is deceptive, as there is no *useful* correlation.

Much like  $\epsilon_{CLOSE}$ , the percentage metric  $\frac{\epsilon_{H/L}}{|H/L-OPEN|}$  consists of two trends converging at 0. As such, rectification via either absolute or square functions reveal that weak correlation is present; however, it is reduced due to the fact that losses are fairly evenly distributed across scores, for both realised and maximum profits, again hindering the usefulness of the apparent correlation.

It was hypothesised that there would be little to no correlation between HIGH/LOW-based error metrics and realised profits, but some correlation would exist with maximum profit, and the evidence would indicate this to be true.



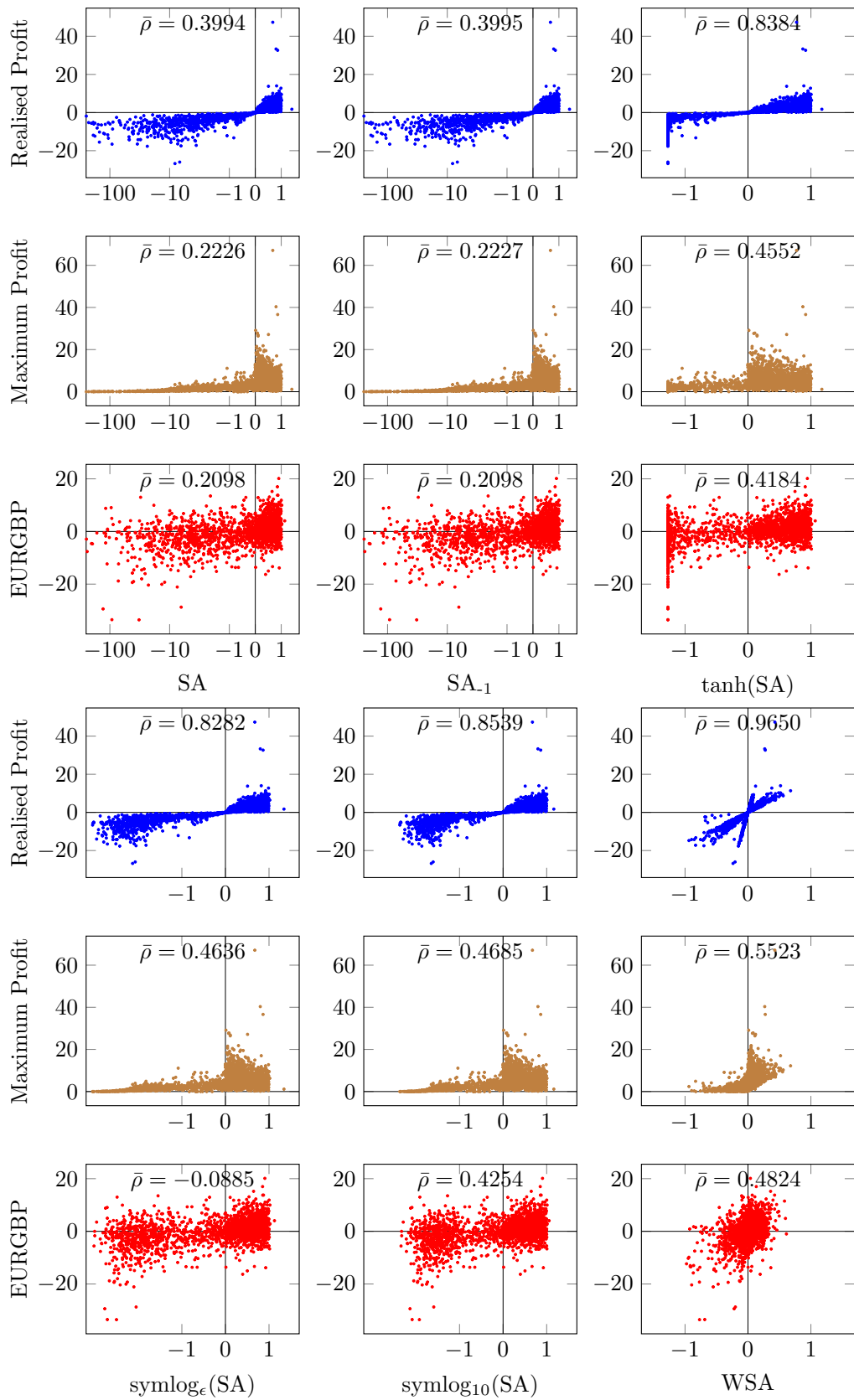


Figure 8.4: Correlation of various Speculative Accuracy metrics, with respects to realised and maximum profit for GBPUSD, and realised profit for EURGBP.

### 8.1.4 Performance of Speculative Accuracy Variants for Random Walk in *Hokohoko*.

In Fig. 8.4, the two basic forms of Speculative Accuracy exhibit weak correlation with both realised and maximum profits, whilst the four normalised variants exhibit strong correlation. The difference is due to the excessively long tail to the negative for the basic versions. The normalised versions show increasing correlation the closer to linear they become, based on the aggressiveness of the normalisation function. Weighted Speculative Accuracy is shown to be exceptionally strongly correlated, and therefore warrants further examination:

In actuality, WSA is strongly correlated because WSA *is* basically the realised profits themselves. Simplification of the mathematics, per-prediction, yields:

$$\text{WSA} = \frac{\text{max\_profit}}{\max([max\_profit_i]_{i=1}^N)} \frac{\text{actual\_profit}}{\text{max\_profit}}$$

Essentially, WSA becomes the realised profit scaled by the maximum profit seen for the Period and Symbol, with slight variance caused by slippage. As a result, the maximum score of WSA is not 1, as intended, but rather the average of all `max_profits` for the Period and Symbol. In order to rectify this, it would perhaps be better to scale by the mean of `max_profits` instead, giving a perfect average score of 1. Regardless, it is incorrect to claim WSA a metric, as presented here, when it is really just normalised profits.

Speculative Accuracy and its variants were expected to show strong correlation with realised profits, and significant correlation with maximum profit. This expectation was partially confirmed, with the three normalised Speculative Accuracy metrics showing the strongest correlation with realised profits. These variants also exhibited correlation with maximum profits. However, all were outperformed by Correct Direction,  $\epsilon_{H/L}$ ,  $|\epsilon_{H/L}|$  and  $\epsilon_{H/L}^2$ —although Correct Direction could be removed from consideration, as per discussion in 8.1.1. Whilst Weighted Speculative Accuracy was indeed

shown to have the strongest correlation, as expected, its design has been shown to be fundamentally flawed. The simpler versions of Speculative Accuracy also exhibited correlation with realised profits, however, they also were outperformed by the other metrics.

## Correlation Between Metrics and Realised Profits

Metric	Realised Profits		Maximum Profits		EURGBP	
	$\bar{\rho}$	$\sigma$	$\bar{\rho}$	$\sigma$	$\bar{\rho}$	$\sigma$
Observed	-0.0202	0.0771	-0.0283	0.0701	-0.0149	0.1320
Direction						
Correct Direction	0.6184	0.0495	0.5643	0.0948	0.3021	0.0654
F-measure	0.5170		0.3578		0.1477	
$\epsilon_{\text{CLOSE}}$	0.0179	0.1511	-0.0065	0.1009	-0.0327	0.1633
$ \epsilon_{\text{CLOSE}} $	-0.6193	0.1591	-0.1150	0.1523	-0.3256	0.0885
$\epsilon_{\text{CLOSE}}^2$	-0.5589	0.2288	-0.0885	0.1613	-0.2914	0.1290
$\frac{\epsilon_{\text{CLOSE}}}{\delta y_t}$	-0.0582	0.0526	-0.0380	0.0551	-0.0318	0.0714
$\frac{ \epsilon_{\text{CLOSE}} }{\delta y_t}$	-0.0055	0.0361	-0.0885	0.0588	-0.0063	0.0634
$\left(\frac{\epsilon_{\text{CLOSE}}}{\delta y_t}\right)^2$	0.0121	0.0293	-0.0379	0.0439	0.0075	0.0648
$\epsilon_{\text{H/L}}$	0.4022	0.1015	0.6923	0.1116	0.1884	0.1344
$ \epsilon_{\text{H/L}} $	-0.0227	0.1029	0.5088	0.1560	-0.0613	0.1143
$\epsilon_{\text{H/L}}^2$	0.0006	0.1102	0.4990	0.1809	-0.0579	0.1235
$\frac{\epsilon_{\text{H/L}}}{ H/L - \text{OPEN} }$	0.0021	0.1041	0.0013	0.0732	-0.0062	0.1077
$\frac{ \epsilon_{\text{H/L}} }{ H/L - \text{OPEN} }$	-0.2825	0.1084	-0.2154	0.0767	-0.1343	0.0922
$\left(\frac{\epsilon_{\text{H/L}}}{ H/L - \text{OPEN} }\right)^2$	-0.1668	0.0872	-0.1309	0.0489	-0.0818	0.0905
SA	0.3994	0.1334	0.2226	0.0773	0.2098	0.1175
SA <sub>-1</sub>	0.3995	0.1334	0.2227	0.0774	0.2098	0.1175
$\text{symlog}_{10}(\text{SA})$	0.8539	0.0581	0.4685	0.0931	0.4254	0.0707
$\text{symlog}_e(\text{SA})$	0.8282	0.0592	0.4636	0.0887	0.4128	0.0743
$\frac{\tanh(\text{SA})}{\tanh(1)}$	0.8384	0.0599	0.4552	0.0947	0.4184	0.0666
Norm. Profits	0.9650	0.0792	0.5523	0.1108	0.4824	0.0932

Table 8.1: Correlation between metrics and realised profits.  $\bar{\rho}$  is the average of the Pearson Correlation Coefficient for the 512 parallel Periods run by *Hokohoko*, with  $\sigma$  being the standard deviation. At the bottom of the table is Normalised Profits, replacing WSA. F-measure does not have standard deviations as  $\|\rho\| = 1$  for it.

### Ranking of Metrics by Correlation

	GBPUSD		EURGBP	Average Rank
	Real. Profit	Max. Profit	Real. Profit	
Normalised Profits	1	3	1	1.7
$\text{symlog}_{10}(\text{SA})$	2	6	2	3.3
Correct Direction	6	2	6	4.7
$\frac{\tanh(\text{SA})}{\tanh(1)}$	3	8	3	4.7
$\text{symlog}_e(\text{SA})$	4	7	4	5.0
$\epsilon_{\text{H/L}}$	9	1	10	6.7
$ \epsilon_{\text{CLOSE}} $	5	14	5	8.0
F-measure	8	9	11	9.3
$\text{SA}_{-1}$	10	10	8	9.3
$\epsilon_{\text{CLOSE}}^2$	7	15	7	9.7
SA	11	11	8	10.0
$ \epsilon_{\text{H/L}} $	15	4	14	11.0
$\frac{ \epsilon_{\text{H/L}} }{ \text{H/L} - \text{OPEN} }$	12	12	12	12.0
$\left(\frac{\epsilon_{\text{H/L}}}{ \text{H/L} - \text{OPEN} }\right)^2$	13	13	13	13.0
$\epsilon_{\text{H/L}}^2$	21	5	15	13.7
$\frac{\epsilon_{\text{CLOSE}}}{\delta y_t}$	14	17	17	16.0
Observed Direction	16	19	18	17.7
$\epsilon_{\text{CLOSE}}$	17	20	16	17.7
$\frac{ \epsilon_{\text{CLOSE}} }{\delta y_t}$	19	15	20	18.0
$\left(\frac{\epsilon_{\text{CLOSE}}}{\delta y_t}\right)^2$	18	18	19	18.3
$\frac{\epsilon_{\text{H/L}}}{ \text{H/L} - \text{OPEN} }$	20	21	21	20.7

Table 8.2: Metrics ordered by average rank, in application to Random Walk and realised and maximum profits. Ranks were calculated using  $|\bar{\rho}|$  from Table 8.1.

### 8.1.5 Ranking Metrics by Correlation

Table 8.1 shows the mean and standard deviations of all the tested metrics' correlation with realised and maximum profits and Table 8.2 shows the metrics ranked in average order of correlation across the three scenarios: GBPUSD realised profits, GBPUSD maximum profits and EURGBP realised profits. The rankings are based on the predictor's relative ranking within each of the three scenarios. In these tables, Normalised Profits refers to the results previously attributed to WSA.

While there is generally a reduction in the mean correlation between metric and realised profit between GBPUSD and EURGBP (48.8% for reductions), the ordering of the metrics' performance is fairly stable. If results below 50% are considered more random than structural, only eight metrics show promising performance in the GBPUSD Periods—three variants of Speculative Accuracy,  $|\epsilon_{\text{CLOSE}}|$ ,  $\epsilon_{\text{CLOSE}}^2$ , Correct Direction and F-measure, in addition to the renamed Normalised Profits. Ignoring Normalised Profits for the time being,  $\text{symlog}_{10}(\text{SA})$ ,  $\text{symlog}_e(\text{SA})$  and  $\tanh(\text{SA})$  all outperform the next best, Correct Direction and  $|\epsilon_{\text{CLOSE}}|$  by over 40%, with Correct Direction and  $|\epsilon_{\text{CLOSE}}|$  having the same correlation, but in different directions. Therefore, if *usefulness* for speculation is considered by correlation with profitability, these metrics all show usefulness, with the Speculative Accuracy variants more useful than the others.

By the same logic, several of the metrics show no measure of usefulness at all. These are all the metrics that exhibit symmetry around either one or both axes, such as  $\epsilon_{\text{CLOSE}}$ , its percentage variants and all of the HIGH/LOW-based error metrics. This strongly indicates that these metrics should not be used for assessing predictor performance with regard to speculation.

Interestingly, when compared with maximum profit, Table 8.1 shows some significant changes to the ranking of metrics. Correct Direction remains largely unchanged; however, the CLOSE-based error metrics cease to possess any information at all, whilst the HIGH/LOW error metrics come to the fore.

This is not a surprise, though  $|\epsilon_{H/L}|$  and  $\epsilon_{H/L}^2$  are on the cusp of being informative, with the three Speculative Accuracy metrics not quite close enough for consideration.

It should be noted that there are three possible applications for these metrics, which is not necessarily obvious, but helps to explain the range of correlation values in the results. These applications are benchmarking, optimisation and model-checking. The distinction between these applications can be seen in the rankings held by different metrics. First, the metrics that exhibit strong correlation to observed results, such as the Speculative Accuracy family, are most suited for performance assessment. Second, the metrics that exhibit strong correlation to the desired results (in this case, maximum profitability), such as  $\epsilon_{H/L}$ , are most suited to optimisation. And third, the metrics that exhibit clear symmetry, such as  $\epsilon_{CLOSE}$ , are most suited to model-checking. Whilst metrics can fit more than one category, it is important to verify their *usefulness* for the given task.

At this stage, it is appropriate to revisit Normalised Profits. This thesis has argued at length against using profits as a metric, largely based on the way they are misused within the academic corpus. However, when using *Hokohoko* the arguments about trading strategies, account settings and balance tampering no longer apply, as *Hokohoko* provides a consistent testing environment. In this regard, the best metric for measuring the performance of FOREX predictions, if the aim is speculation, would be Normalised Profits—if consistency of simulation regimes is assured, such as through *Hokohoko*—closely followed by the normalised Speculative Accuracy metrics.

### Ranking of Predictors by Metric for GBPUSD

	RW	Random	AR	Static	ANN	SAL_ES	SAL_EO	SAL_OS	SAL_OO	Bachelier	SE w/ RP
Correct Direction	2	4	9	8	10	6	3	7	5	1	2.6
F-measure	2	6	9	8	10	5	3	7	4	1	2.9
$\epsilon_{\text{CLOSE}}$	1	5	2	6	3	8	10	4	9	7	4.8
$ \epsilon_{\text{CLOSE}} $	5	2	8	9	3	4	10	7	6	1	3.2
$\epsilon_{\text{CLOSE}}^2$	4	1	8	9	3	5	10	6	7	2	3.3
$\frac{\epsilon_{\text{CLOSE}}}{\delta y_t}$	10	1	6	7	5	4	2	9	3	8	3.6
$\frac{ \epsilon_{\text{CLOSE}} }{\delta y_t}$	7	1	8	9	3	4	10	6	5	2	3.2
$\left(\frac{\epsilon_{\text{CLOSE}}}{\delta y_t}\right)^2$	7	1	8	9	3	4	10	5	6	2	3.3
$\epsilon_{\text{H/L}}$	4	10	5	6	8	2	9	3	1	7	4.3
$ \epsilon_{\text{H/L}} $	3	9	7	8	1	2	10	5	4	6	4.8
$\epsilon_{\text{H/L}}^2$	4	5	9	2	1	6	10	7	8	3	3.6
$\frac{\epsilon_{\text{H/L}}}{ \text{H/L} - \text{OPEN} }$	1	9	3	6	5	7	10	2	8	4	4.8
$\frac{ \epsilon_{\text{H/L}} }{ \text{H/L} - \text{OPEN} }$	3	8	7	9	2	6	10	4	5	1	4.4
$\left(\frac{\epsilon_{\text{H/L}}}{ \text{H/L} - \text{OPEN} }\right)^2$	3	9	7	8	2	6	10	5	4	1	4.3
SA	2	8	6	9	7	5	10	4	3	1	3.8
SA <sub>-1</sub>	2	8	6	9	7	5	10	4	3	1	3.8
symlog <sub>10</sub> (SA)	3	10	7	9	2	6	8	5	4	1	4.5
symlog <sub>e</sub> (SA)	3	10	7	9	2	6	8	4	5	1	4.6
$\frac{\tanh(\text{SA})}{\tanh(1)}$	6	10	7	9	2	5	8	3	4	1	4.6
Norm. Profits	5	2	9	4	8	7	10	3	6	1	2.2
<b>Avg. Ranking</b>	<b>3.9</b>	<b>6.0</b>	<b>6.9</b>	<b>7.7</b>	<b>4.4</b>	<b>5.2</b>	<b>8.6</b>	<b>5.0</b>	<b>5.0</b>	<b>2.6</b>	
Realised Profit	6	1	9	3	10	5	7	8	4	2	

Table 8.3: Ranking different Predictors by metric. The rankings are calculated by  $\bar{\rho}$  for linear metrics, and  $|\bar{\rho}|$  for error-based metrics.



### Ranking of Predictors by Metric for EURGBP

	RW	Random	AR	Static	ANN	SAL_ES	SAL_EO	SAL_OS	SAL_OO	Bachelier	SE w/ RP
Correct Direction	1	5	2	8	3	6	9	4	7	10	2.8
F-measure	1	6	2	7	8	4	9	3	5	10	2.2
$\epsilon_{\text{CLOSE}}$	1	5	4	6	10	3	9	2	8	7	3.1
$ \epsilon_{\text{CLOSE}} $	4	2	8	10	3	5	9	6	7	1	4.4
$\epsilon_{\text{CLOSE}}^2$	4	2	8	10	3	5	9	6	7	1	4.4
$\frac{\epsilon_{\text{CLOSE}}}{\delta y_t}$	10	1	8	6	2	5	3	9	4	7	5.2
$\frac{ \epsilon_{\text{CLOSE}} }{\delta y_t}$	4	1	8	10	3	5	9	6	7	2	4.4
$\left(\frac{\epsilon_{\text{CLOSE}}}{\delta y_t}\right)^2$	4	1	8	10	3	5	9	7	6	2	4.5
$\epsilon_{\text{H/L}}$	4	9	5	10	6	3	8	2	1	7	1.6
$ \epsilon_{\text{H/L}} $	2	8	6	10	1	3	9	5	4	7	2.7
$\epsilon_{\text{H/L}}^2$	2	4	8	9	5	3	10	6	7	1	4.1
$\frac{\epsilon_{\text{H/L}}}{ \text{H/L} - \text{OPEN} }$	4	6	5	10	9	1	8	3	7	2	3.2
$\frac{ \epsilon_{\text{H/L}} }{ \text{H/L} - \text{OPEN} }$	6	8	7	10	2	3	9	5	4	1	3.3
$\left(\frac{\epsilon_{\text{H/L}}}{ \text{H/L} - \text{OPEN} }\right)^2$	5	7	6	9	2	3	10	4	8	1	3.5
SA	3	10	7	9	2	6	8	4	5	1	3.1
SA <sub>-1</sub>	3	10	7	9	2	6	8	4	5	1	3.1
$\text{symlog}_{10}(\text{SA})$	4	10	7	9	2	6	8	3	5	1	3.1
$\text{symlog}_e(\text{SA})$	4	10	7	9	2	6	8	3	5	1	3.1
$\frac{\tanh(\text{SA})}{\tanh(1)}$	4	10	7	9	2	6	8	3	5	1	3.1
Norm. Profits	4	8	1	7	3	6	9	2	5	10	1.8
<b>Avg. Ranking</b>	<b>3.7</b>	<b>6.2</b>	<b>6.1</b>	<b>8.9</b>	<b>3.7</b>	<b>4.5</b>	<b>8.5</b>	<b>4.4</b>	<b>5.6</b>	<b>3.7</b>	
Realised Profit	4	10	1	9	6	5	8	2	3	7	

Table 8.4: Ranking different Predictors by metric. The rankings are calculated by  $\bar{\rho}$  for linear metrics, and  $|\bar{\rho}|$  for error-based metrics.

## 8.2 Results: Ranking of Predictors by Metric

In Tables 8.3 and 8.4, each tested Predictor is ranked, per-metric, against all the other Predictors. Using the metrics already tested in Experiment 1, most of the tested predictors failed to significantly outperform either Random Walk or each other. However, there were two predictors that did manage to outperform Random Walk—Bachelier’s for GBPUSD, and the 6-3-1 Artificial Neural Network with Moving Average inputs. These results are not immediately obvious, however, with some of the other predictors’ relative rankings masking their performance.

First, `SameAsLast`<sub>EXACT\_OPPOSITE</sub> performs poorly nearly universally, except for  $\frac{\epsilon_{\text{CLOSE}}}{\delta y_t}$ . Second, the Random predictor also has skewed results in the CLOSE-based errors, as it is heavily skewed for correct directions (due to no TAKE\_PROFIT). Removing these two predictors from the statistics reveals Bachelier’s to be better than Random Walk. In the GBPUSD scenario, the Bachelier predictor outperformed nearly all the other predictors, including Random Walk. This superior performance is also evidenced in its average realised profit also, with the second-highest average profits. In the EURGBP scenario, both Bachelier and the ANN outperformed Random Walk, with the Bachelier predictor slightly ahead of the ANN.

Despite the apparently ‘good’ performances of some of the predictors, this does not mean that the predictors were profitable. In fact, in GBPUSD only the Bachelier and Random predictors recorded a profit, while half the predictors were profitable in EURGBP. However, in all cases, the average profits were dwarfed by the standard deviation, with the Z-scores for break-even in the range  $[-0.017, 0.012]$  indicating no real profitability present (full results are in Appendix C). Crucially, if trading costs had been included, only the AR in EURGBP would have broken even, making 1.5c per-USD1000 trade—roughly a 0.4% return per annum on daily trades. According to the EMH, this is the expected result—if a predictor is *a priori* unable to beat the Random Walk predictor then that implies its predictive

performance should be the same as Random Walk, i.e. 50%, as a lower performance level would imply improved performance by taking an opposite prediction. This leads to the conclusion that random performances imply random rankings, and this makes it impossible to demonstrate the efficacy of metrics using real exchange rate data.

Besides this impossibility, there is another issue with this experiment. It is possible that the lack of clarity in rankings could be attributed to the predictors' poor performance—the predictors tested were either toys, or out-dated. If a Predictor cannot out-predict Random Walk, then its results are effectively random. Without better predictors, the only way to measure the efficacy of metrics is by synthetic measures, as in Experiment 1.

### 8.3 Hypotheses Evaluation

So far, this chapter has presented the results of the experiments run to test the *usefulness* of various metrics, including Speculative Accuracy, for FOREX speculation using *Hokohoko*. These experiments covered four scenarios: the correlation of various metrics in relation to realised and maximum profits for GBPUSD, and realised profits for EURGBP; and the relative performance of different predictors under various metrics. Alongside the experiments, several hypotheses were proposed, which need verification:

For Experiment 1:

1. It was expected that there would be limited correlation between observed direction, and maximum profitability. This was confirmed, with no evidence of correlation with either realised or maximum profits measured.
2. It was expected that there would be a positive correlation between Correct Direction and realised profits. This also was confirmed, with Correct Direction ranking on average second-equal for correlation,

across all three tested scenarios.

3. It was expected that interference from an exogenous Symbol would obscure, but not remove, this correlation. This too was confirmed, with the correlation approximately halved in the EURGBP tests.
4. It was expected that there would be little correlation between Correct Direction and maximum profit. However, Correct Direction was only beaten by  $\epsilon_{H/L}$ , in Table 8.1.
5. It was expected that there would be positive correlation between F-measure and realised profits, but not maximum profit. This was confirmed.
6. It was expected that there would be little to no correlation between CLOSE-based metrics and either realised or maximum profits. In fact, this was only mostly true, with Absolute CLOSE Error and Squared CLOSE Error showing correlation to realised profits, in the GBPUSD case. However, it was confirmed in all other instances of the CLOSE-based Error metrics.
7. It was expected that there would be little to no correlation between HIGH/LOW-based metrics and realised profit, but some correlation would exist with maximum profit. This was confirmed, with only the non-percentage HIGH/LOW-Error metrics showing correlation with maximum profits.
8. Speculative Accuracy and its variants were expected to show strong correlation with realised profits, and significant correlation with maximum profit. This was partially confirmed, with the three normalised versions of Speculative Accuracy showing the strongest correlation with realised profits. However, whilst they showed significant correlation with maximum profits, other metrics were shown to have higher correlation, particularly  $\epsilon_{H/L}$ .

For Experiment 2:

1. It was expected that the metrics with low correlation to profits would have no clear pattern. This was indeed the case, however, it cannot be attributed to a property of the metrics themselves. Rather, the lack of a clear pattern was inherent in the data itself, and so the veracity of this expectation remains unknown.
2. It was expected that the HIGH/LOW-based metrics would exhibit some consistency in rankings. This also was not the case, again due to the inability of predictors to consistently beat Random Walk.
3. It was expected that Speculative Accuracy would most closely match realised profit. Again, this was not verified, due to the inability for reported profits to differentiate between predictors.

## 8.4 Conclusion

The experiments conducted in this thesis were intended to serve three purposes: First, they were intended to verify certain hypotheses regarding the usefulness of different metrics. Second, they were intended to show how various metrics could be applied, through *Hokohoko*, to test the efficacy of various FOREX predictors. And third, they were intended to show the use of *Hokohoko* by application.

Of the hypotheses tested, most proved likely to be true—but not all. Notably, Correct Direction was shown to have higher correlation with maximum profits than expected, for the daily interval. Similarly, the rectified  $\epsilon_{\text{CLOSE}}$  metrics,  $|\epsilon_{\text{CLOSE}}|$  and  $\epsilon_{\text{CLOSE}}^2$  were shown to have higher correlation with realised profits than expected. These are probably related, and warrant further exploration. Across the board, the percentage-type metrics were shown to have no correlation with either realised or maximum profits. And the normalised Speculative Accuracy metrics ( $\text{symlog}_{10}(\text{SA})$ ),

$\text{symlog}_e(\text{SA})$  and  $\text{tanh}(\text{SA})$ ) were shown to have very strong correlation with realised profits but less with maximum profits.

In addition, the proposal to weight Speculative Accuracy was shown to be flawed, with weighting based on maximum observed profits per-Symbol-per-Period effectively transforming the metric to rescaled profits. For the purposes of speculation, these ‘Normalised Profits’ were shown to be most useful, however, the argument is tautological. Nevertheless, it must be acknowledged that, if a predictor is to be ranked by its profitability, it makes sense to rescale profits between Periods and Symbols, such as dividing realised profits by the average profits for its Period and Symbol.

Attempts were also made to rank a variety of Predictors, available out-of-the-box in *Hokohoko*, by various metrics. Had this experiment been successful, such ranking ability would imply the negation of the EMH. However, while some predictors appeared to rank consistently above others, the margins of error were too high for any meaningful conclusions to be drawn. As a result, this experiment was unable to either rank predictors or disprove the EMH, which is not surprising given the dominance of the EMH over the last fifty years.

Searching by ‘FOREX’ or ‘foreign exchange’ in PyPI returned 86 results, most of which were wrappers to trading APIs, and none were deemed suitable for the task of investigating metrics. Therefore, *Hokohoko* was written and used for the experiments, providing a consistent interface and consistent test conditions, under which the hypotheses were tested. These experiments thus also showed the application of the *Hokohoko* framework, and how it will enable future challenges to the EMH to be quickly and extensively tested.

# Chapter 9

## Conclusion

This thesis has investigated the measurement of predictor performance as applied to foreign exchange speculation. It outlined the development of key ideas and techniques over the course of the last 120 years, and examined the datasets and metrics used within a representative sample of the academic corpus. In this examination two problems were identified: first, there was a lack of consistency in the datasets used to test researchers' algorithms; and second, a large variety of metrics were used, most of which were either inappropriate for or inappropriately applied to FOREX speculation. This thesis then presented two solutions to these issues: a Python library, *Hokohoko*, which provided a consistent dataset and interface for testing FOREX prediction algorithms; and a new metric, Speculative Accuracy, which it argued provides a more appropriate measure of usefulness with regards to speculation. *Hokohoko* was then used to test a series of hypotheses regarding the usefulness of various metrics, alongside Speculative Accuracy. From these experiments a number of conclusions can be drawn:

- It was shown that there was no correlation between the observed direction of foreign exchange rate movements and the maximum profitability for the given epoch.
- None of the tested predictors were able to conclusively 'beat' Random Walk. This suggests that the FOREX market is indeed efficient, and

therefore correct prediction of future FOREX market movements is difficult.

- For the purposes of assessing prediction performance for speculation, Normalised Profits provides the best measure of *usefulness*, with normalised Speculative Accuracy metrics, such as  $\tanh(\text{SA})$ , coming a close second.
- Correct Direction was shown to have less correlation with realised profits than Speculative Accuracy. However, its correlation was higher than expected, which tempers this thesis' critique of papers using it.
- $\epsilon_{\text{HIGH/LOW}}$  was shown to have the highest correlation with maximum profits, suggesting it would be the best metric by which to train predictors.
- $\epsilon_{\text{CLOSE}}$ -based metrics were shown to have less correlation with realised profits than either the normalised Speculative Accuracy metrics or Correct Direction. They were also shown to have less correlation with maximum profitability than  $\epsilon_{\text{HIGH/LOW}}$ . This strongly suggests that they should not be used in application to FOREX speculation.
- None of the percentage-type metrics were shown to have any significant correlation with either realised or maximum profitability. This strongly suggests that they also should not be used for FOREX speculation.

Alongside these conclusions, this thesis also points towards a number of future research opportunities:

- The analysis of the academic research was based on a subset of the FOREX corpus, focusing on the most cited works. It is possible that some significant contributions were missed, therefore future research could expand upon this selection, particularly by increasing the number of more recent works included.



- This thesis used 512 overlapping Periods in the generation of its results. It would be interesting for future research to investigate how the number of Periods used affects the stability of the observed results.
- This thesis used two Symbols, GBPUSD and EURGBP, in drawing its conclusions. *Hokohoko* provides the ability to test predictors across 50 different Symbols, and so this research could be expanded upon by using all available Symbols.
- For the experiments, TAKE\_\_PROFIT was set, but STOP\_\_LOSS was left as None. Given the inability of all the tested predictors to outperform Random Walk, this suggests that risk management may be more important in FOREX speculation than predictor accuracy. Therefore, there is scope for future research to investigate the relationship between STOP\_\_LOSSES, TAKE\_\_PROFITS and realised or maximum profits.
- The experimental results in this thesis indicate that there is greater than hypothesised correlation between Correct Direction and  $\epsilon_{\text{CLOSE}}$  with regards to maximum profitability at the fixed interval of 1,440 minutes. Future research could investigate the effect of differing interval lengths on observed correlations.
- A number of comparative metrics were described in Chapter 4, *Measuring Foreign Exchange Predictions: Common Metrics*. Future work could expand upon this thesis by also testing those metrics' correlation with realised and maximum profits.
- Another future project could be the addition of bootstrap testing to the *Hokohoko* framework.
- Finally, *Hokohoko* was designed with the intention of being made open source, and thus being available for other researchers to test their own predictors. Therefore, future research could include a comparative study which tests the most recently published predictors using *Hokohoko*.

# Appendix A

## Predictor Code Listing:

### Example.py

```
"""
```

```
=====
```

```
Example.py
```

```
=====
```

```
This file is to show how a researcher may use the Hokohoko /  
library. It assumes there is a file data.npz in the current  
directory. Details on how to obtain the data file can be found in  
Hokohoko's documentation.
```

```
Author: Neil Bradley
```

```
License: MIT License
```

```
"""
```

```
import multiprocessing as mp
```

```
import sys
```

```
from typing import List, Optional
```

```
from hokohoko import Hokohoko
```

```
from hokohoko.entities
```

```
    import Bar, Config, Direction, Order, Predictor
```

```

class Example(Predictor):
    """
    This is an example Predictor. It makes a single random
    prediction per Bar. It demonstrates:

    1. How to inherit Predictor and override the appropriate
        methods.
    2. How to use self.random() to access fully-deterministic RNG.
    3. How to access self.Account.
    """
    def __init__(
        self,
        lock: mp.Lock,
        parameters: Optional[str] = None,
        debug: Optional[bool] = False
    ) -> None:
        """
        Initialises the Predictor, saving the given parameters for
        use in on_start. It provides the following instance
        objects for use:

        :param lock:          An lock which is shared between all
                               concurrent Predictor processes.
                               Intended use is for shared access to
                               external resources, etc. Stored in
                               self.lock, and provided for custom
                               Predictors which might need access to
                               shared external resources.

        :type lock:           multiprocessing.Lock

        :param parameters:    User-customisable parameters, which
                               get stored in self.parameters.

        :type parameters:     str, optional

        :param debug:         Pass-through debug flag.

        :type debug:          bool, optional

```

*Note that it is not strictly necessary to override, this is just to show how to add your own data structures in.*

```

"""
super().__init__(lock, parameters, debug)
self.exclamation = "I inherited Predictor!"

def __enter__(self) -> Predictor:
    """
    Simple demonstration of overriding __enter__.

    :returns: As per Python `with` specification, self.
    :rtype: hokohoko.entities.Predictor

    """
    print(self.exclamation)
    return self

def __exit__(self, exc_type, exc_val, exc_tb):
    """
    This has to be overridden, but does nothing in this
    example (there are no resources to be released).

    """
    pass

def on_start(self, bars: List[Bar]) -> None:
    """
    Has to be overridden, the base class raises a
    NotImplementedError otherwise.

    :param bars: The list of Bars containing the opening
                  values for each currency pair. If set,
                  only the requested subset is provided.
    :type bars: List[hokohoko.entities.Bar]

    """

```

```
print(f"This is an example Bar: {bars[0]}")
```

```
def on_bar(self, bars: List[Bar]) -> None:
    """
    This implementation picks a random direction for each
    symbol. Predictors provide access to a deterministic RNG
    source (themselves), so all random calls should be
    self.random(), self.randint(), etc.

    :param bars:    The latest bar in the data. This is an
                    array of the selected currencies.
    :type bars:     List[hokohoko.entities.Bar]
    """

    orders = []

    for b in bars:
        if self.random() > 0.5:
            direction = Direction.BUY
            take_profit = None
            stop_loss = b.close - 0.001
        else:
            direction = Direction.SELL
            take_profit = None
            stop_loss = b.close + 0.001

        orders.append(Order(
            symbol_id=b.symbol_id,
            direction=direction,
            open_bid=None,
            take_profit=None,
            stop_loss=None
        ))

    self.place_orders(orders)
```

```
if __name__ == "__main__":
    conf = Config(
        predictor_class="Example",
        assessors=[
            "hokohoko.assessors.Accuracy --show-results True",
            "hokohoko.assessors.AccountViewer"
        ],
        data_parameters="data.npz",
        data_subset="EURGBP",
        period_count=32,
        process_count=8
    )
    Hokohoko.run(conf)

    sys.exit()
```

# Appendix B

## Assessor Code Listing

```
from collections import defaultdict
from typing import Iterable

import numpy as np

from hokohoko.entities import Assessor, Direction, Status
from hokohoko.assessors import Accuracy
from hokohoko.utils import convert_id_to_symbol

class Experiment5(Assessor):
    """
    This assessor was used for Experiment 1 in Neil Bradley's
    thesis. The aim of this experiment is to determine
    the correlation between Directional metrics and Profits.
    """
    def analyse(self, period_results: Iterable) -> None:

        table = {}

        print("# Period, Symbol, CD, "
              "c_e, c_ae, c_e2, c_pe, c_ape, c_spe, "
              "hl_e, hl_ae, hl_e2, hl_pe, hl_ape, hl_spe, "
              "sa0, sa1, sa_10, sa_e, sa_tanh, wsa, "
              "r_profit, m_profit")
```

```

)
for pr in period_results:
    period, account = pr.get()

    # 1. Find the weights per-Symbol within the run. (For WSA)
    maxes = defaultdict(float)
    for h_id, history in account.history.items():
        max_profit, max_loss, actual_profit = Accuracy.calculate_rate_changes(history)
        maxes[history.order.symbol_id] = max(maxes[history.order.symbol_id], max_profit)

    # 2. Now cycle through, printing out the stats.
    for h_id, history in account.history.items():

        key = (period, history.order.symbol_id)
        if key not in table.keys():
            table[key] = {}
            for s in [
                "CD", "tp", "tn", "fp", "fn",
                "c_e", "c_ae", "c_e2", "c_pe", "c_ape", "c_spe",
                "hl_e", "hl_ae", "hl_e2", "hl_pe", "hl_ape", "hl_spe",
                "sa0", "sa1", "sa_10", "sa_e", "sa_tanh", "sa_w",
                "r_profit", "m_profit"
            ]:
                table[key][s] = []

    # Skip non-predictions.
    if history.order.direction in (Direction.DONT_BUY, Direction.DONT_SELL):
        continue

```



```

max_profit, max_loss, actual_profit = Accuracy.
    calculate_rate_changes(history)

# Correct Direction
if (
    history.future.close > history.future.open and
        history.order.direction == Direction.BUY
    or
    history.future.close < history.future.open and
        history.order.direction == Direction.SELL
):
    correct_direction = 1
else:
    correct_direction = 0

# FM
if correct_direction == 1:
    tp = 1
    tn = 2
    fp = 0
    fn = 0
else:
    tp = 0
    tn = 1
    fp = 1
    fn = 1

# CLOSE
if history.order.take_profit is None or history.
    order.direction in (Direction.DONT_BUY,
        Direction.DONT_SELL):
    c_e = history.future.close - history.future.
        open
else:
    c_e = history.future.close - history.order.
        take_profit

```

```

c_pe = c_e / (history.future.close - history.
    future.open)
c_ae = abs(c_e)
c_e2 = c_e * c_e
c_ape = abs(c_pe)
c_spe = c_pe * c_pe

# HIGH/LOW
if history.order.direction == Direction.BUY:
    if history.order.take_profit is None:
        hl_e = history.future.high - history.
            future.close
    else:
        hl_e = history.future.high - history.order
            .take_profit
    hl_pe = hl_e / (history.future.high - history.
        future.open)
elif history.order.direction == Direction.SELL:
    if history.order.take_profit is None:
        hl_e = history.future.close - history.
            future.low
    else:
        hl_e = history.order.take_profit - history
            .future.low
    hl_pe = hl_e / (history.future.low - history.
        future.open)
elif history.order.direction == Direction.DONT_BUY
:
    hl_e = history.future.high - history.future.
        open
    hl_pe = hl_e / (history.future.high - history.
        future.open)
else:
    hl_e = history.future.open - history.future.
        low
    hl_pe = hl_e / (history.future.low - history.

```

```

        future.open)
hl_ae = abs(hl_e)
hl_e2 = hl_e * hl_e
hl_ape = abs(hl_pe)
hl_spe = hl_pe * hl_pe

# SA Variants
max_profit, max_loss, actual_profit = Accuracy.
    calculate_rate_changes(history)
sa0 = Accuracy.calculate_accuracy(max_profit,
    max_loss, actual_profit, history.status)
sa1 = Accuracy.calculate_accuracy(max_profit,
    max_loss, actual_profit, history.status, 1)
sa_10 = Accuracy.calculate_accuracy(max_profit,
    max_loss, actual_profit, history.status, 2)
sa_e = Accuracy.calculate_accuracy(max_profit,
    max_loss, actual_profit, history.status, 3)
sa_tanh = Accuracy.calculate_accuracy(max_profit,
    max_loss, actual_profit, history.status, 4)
sa_w = Accuracy.calculate_accuracy(
    max_profit, max_loss, actual_profit, history.
        status, 5,
        maxes[history.order.symbol_id] if maxes[
            history.order.symbol_id] != 0 else 1
    )

# Record Values
table[key]["CD"].append(correct_direction)
table[key]["tp"].append(tp)
table[key]["tn"].append(tn)
table[key]["fp"].append(fp)
table[key]["fn"].append(fn)
table[key]["c_e"].append(c_e)
table[key]["c_ae"].append(c_ae)
table[key]["c_e2"].append(c_e2)
table[key]["c_pe"].append(c_pe)

```

```

table[key]["c_ape"].append(c_ape)
table[key]["c_spe"].append(c_spe)
table[key]["hl_e"].append(hl_e)
table[key]["hl_ae"].append(hl_ae)
table[key]["hl_e2"].append(hl_e2)
table[key]["hl_pe"].append(hl_pe)
table[key]["hl_ape"].append(hl_ape)
table[key]["hl_spe"].append(hl_spe)
table[key]["sa0"].append(sa0)
table[key]["sa1"].append(sa1)
table[key]["sa_10"].append(sa_10)
table[key]["sa_e"].append(sa_e)
table[key]["sa_tanh"].append(sa_tanh)
table[key]["sa_w"].append(sa_w)
table[key]["r_profit"].append(history.final_value
    - history.initial_value)
table[key]["m_profit"].append(max_profit / history
    .future.open)

print(
    f'{period},{convert_id_to_symbol(history.order
        .symbol_id)},'
    f'{correct_direction},'
    f'{c_e:.4g},{c_ae:.4g},{c_e2:.4g},{c_pe:.4g},{
        c_ape:.4g},{c_spe:.4g},'
    f'{hl_e:.4g},{hl_ae:.4g},{hl_e2:.4g},{hl_pe:.4
        g},{hl_ape:.4g},{hl_spe:.4g},'
    f'{sa0:.4g},'
    f'{sa1:.4g},'
    f'{sa_10:.4g},'
    f'{sa_e:.4g},'
    f'{sa_tanh:.4g},'
    f'{sa_w:.4g},'
    f'{table[key]["r_profit"][-1]:.4g},'
    f'{table[key]["m_profit"][-1]*1000:.4g}'
)

```

```

# 3. Now to put together the other statistics.
r_periods = {}

print("##### FM Stats:")
print("# Period,FM,r_profits,m_profits")
for key in table.keys():
    line = table[key]
    if key[0] not in r_periods:
        r_periods[key[0]] = {
            "FM": [],
            "r_profits": [],
            "m_profits": []
        }

fm = 2 * np.mean(line["tp"]) / (2 * np.mean(line["tp"
    ]) + np.mean(line["fp"]) + np.mean(line["fn"]))

# Here, we need the data to graph FM against.
print(
    f'{key[0]},'
    f'{fm:.4g},'
    f'{np.sum(line["r_profit"]):.4g},'
    f'{np.sum(line["m_profit"])*1000:.4g}'
)

```

# Appendix C

## Results Tables for Experiment 2

Table C.1: Correct Direction Rankings

GBPUSD	Mean	Median	SD	Rank(Mean)	Rank(Median)
RW	0.5029	1.0000	0.5000	2	1
Random	0.5007	1.0000	0.5000	4	1
Yule1927	0.4990	0.0000	0.5000	9	7
Static	0.5000	0.0000	0.5000	8	7
YaoTan2000	0.4924	0.0000	0.4999	10	7
SAL_ES	0.5003	1.0000	0.5000	6	1
SAL_EO	0.5016	1.0000	0.5000	3	1
SAL_OS	0.5000	0.0000	0.5000	7	7
SAL_OO	0.5004	1.0000	0.5000	5	1
Bachelier1900	0.5172	1.0000	0.4997	1	1
EURGBP	Mean	Median	SD	Rank(Mean)	Rank(Median)
RW	0.5031	1.0000	0.5000	1	1
Random	0.4998	0.0000	0.5000	5	4
Yule1927	0.5022	1.0000	0.5000	2	1
Static	0.4993	0.0000	0.5000	8	4
YaoTan2000	0.5007	1.0000	0.5000	3	1
SAL_ES	0.4998	0.0000	0.5000	6	4
SAL_EO	0.4980	0.0000	0.5000	9	4
SAL_OS	0.4999	0.0000	0.5000	4	4
SAL_OO	0.4997	0.0000	0.5000	7	4
Bachelier1900	0.4859	0.0000	0.4998	10	4

Table C.2: F-measure Rankings

GBPUSD	Mean	Median	SD	Rank(Mean)	Rank(Median)
RW	0.5740	0.5741	0.0293	2	2
Random	0.5717	0.5741	0.0382	6	2
Yule1927	0.5704	0.5707	0.0077	9	9
Static	0.5714	0.5714	0.0013	8	6
YaoTan2000	0.5636	0.5634	0.0317	10	10
SAL_ES	0.5717	0.5714	0.0018	5	6
SAL_EO	0.5729	0.5735	0.0173	3	4
SAL_OS	0.5714	0.5714	0.0020	7	6
SAL_OO	0.5718	0.5728	0.0022	4	5
Bachelier1900	0.5888	0.5948	0.0493	1	1
EURGBP	Mean	Median	SD	Rank(Mean)	Rank(Median)
RW	0.5740	0.5741	0.0358	1	1
Random	0.5709	0.5714	0.0366	6	3
Yule1927	0.5736	0.5729	0.0075	2	2
Static	0.5707	0.5714	0.0014	7	3
YaoTan2000	0.5701	0.5664	0.0513	8	9
SAL_ES	0.5712	0.5714	0.0019	4	3
SAL_EO	0.5694	0.5694	0.0205	9	8
SAL_OS	0.5714	0.5714	0.0020	3	3
SAL_OO	0.5711	0.5714	0.0023	5	3
Bachelier1900	0.5557	0.5553	0.0461	10	10

Table C.3: Error<sub>CLOSE</sub> Rankings

GBPUSD	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	0.0000	0.0000	0.0119	0.0000	0.0000	1	1
Random	-0.0002	-0.0002	0.0085	0.0002	0.0002	5	7
Yule1927	0.0000	0.0001	0.0131	0.0000	0.0001	2	5
Static	-0.0002	-0.0001	0.0131	0.0002	0.0001	6	3
YaoTan2000	0.0000	0.0001	0.0094	0.0000	0.0001	3	2
SAL_ES	-0.0003	-0.0002	0.0123	0.0003	0.0002	8	9
SAL_EO	-0.0007	-0.0004	0.0198	0.0007	0.0004	10	10
SAL_OS	-0.0001	-0.0001	0.0124	0.0001	0.0001	4	4
SAL_OO	-0.0004	-0.0002	0.0125	0.0004	0.0002	9	8
Bachelier1900	-0.0003	-0.0002	0.0087	0.0003	0.0002	7	6
EURGBP	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	0.0000	0.0001	0.0064	0.0000	0.0001	1	6
Random	0.0001	0.0000	0.0046	0.0001	0.0000	5	4
Yule1927	-0.0001	0.0000	0.0072	0.0001	0.0000	4	3
Static	0.0001	0.0003	0.0110	0.0001	0.0003	6	9
YaoTan2000	-0.0006	-0.0007	0.0050	0.0006	0.0007	10	10
SAL_ES	0.0001	0.0000	0.0067	0.0001	0.0000	3	2
SAL_EO	0.0004	0.0002	0.0109	0.0004	0.0002	9	8
SAL_OS	0.0000	0.0000	0.0068	0.0000	0.0000	2	5
SAL_OO	0.0002	0.0001	0.0069	0.0002	0.0001	8	7
Bachelier1900	0.0002	0.0000	0.0044	0.0002	0.0000	7	1

Table C.4: Absolute Error<sub>CLOSE</sub> Rankings

GBPUSD	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	0.0086	0.0068	0.0082	0.0086	0.0068	5	7
Random	0.0059	0.0046	0.0061	0.0059	0.0046	2	2
Yule1927	0.0091	0.0071	0.0094	0.0091	0.0071	8	8
Static	0.0109	0.0101	0.0074	0.0109	0.0101	9	9
YaoTan2000	0.0066	0.0051	0.0067	0.0066	0.0051	3	3
SAL_ES	0.0086	0.0067	0.0088	0.0086	0.0067	4	4
SAL_EO	0.0137	0.0106	0.0144	0.0137	0.0106	10	10
SAL_OS	0.0087	0.0067	0.0089	0.0087	0.0067	7	6
SAL_OO	0.0086	0.0067	0.0091	0.0086	0.0067	6	5
Bachelier1900	0.0059	0.0045	0.0064	0.0059	0.0045	1	1
EURGBP	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	0.0048	0.0037	0.0043	0.0048	0.0037	4	4
Random	0.0033	0.0026	0.0031	0.0033	0.0026	2	1
Yule1927	0.0051	0.0040	0.0050	0.0051	0.0040	8	8
Static	0.0101	0.0100	0.0043	0.0101	0.0100	10	10
YaoTan2000	0.0037	0.0029	0.0034	0.0037	0.0029	3	3
SAL_ES	0.0049	0.0038	0.0046	0.0049	0.0038	5	5
SAL_EO	0.0077	0.0061	0.0077	0.0077	0.0061	9	9
SAL_OS	0.0049	0.0038	0.0047	0.0049	0.0038	6	6
SAL_OO	0.0049	0.0039	0.0049	0.0049	0.0039	7	7
Bachelier1900	0.0033	0.0026	0.0030	0.0033	0.0026	1	1

Table C.5: Squared Error<sub>CLOSE</sub> Rankings

GBPUSD	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	1.00E-04	4.58E-05	7.00E-04	1.00E-04	4.58E-05	4	7
Random	7.26E-05	2.09E-05	4.52E-04	7.26E-05	2.09E-05	1	2
Yule1927	1.72E-04	4.99E-05	9.33E-04	1.72E-04	4.99E-05	8	8
Static	1.73E-04	1.03E-04	4.90E-04	1.73E-04	1.03E-04	9	9
YaoTan2000	8.76E-05	2.61E-05	4.72E-04	8.76E-05	2.61E-05	3	3
SAL_ES	1.50E-04	4.44E-05	8.45E-04	1.50E-04	4.44E-05	5	4
SAL_EO	3.94E-04	1.13E-04	2.71E-03	3.94E-04	1.13E-04	10	10
SAL_OS	1.53E-04	4.53E-05	7.85E-04	1.53E-04	4.53E-05	6	6
SAL_OO	1.58E-04	4.49E-05	1.04E-03	1.58E-04	4.49E-05	7	5
Bachelier1900	7.52E-05	2.01E-05	4.90E-04	7.52E-05	2.01E-05	2	1
EURGBP	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	4.09E-05	1.36E-05	1.07E-04	4.09E-05	1.36E-05	4	4
Random	2.10E-05	6.76E-06	7.51E-05	2.10E-05	6.76E-06	2	1
Yule1927	5.16E-05	1.64E-05	2.33E-04	5.16E-05	1.64E-05	8	8
Static	1.21E-04	1.00E-04	1.19E-04	1.21E-04	1.00E-04	10	10
YaoTan2000	2.50E-05	8.24E-06	8.00E-05	2.50E-05	8.24E-06	3	3
SAL_ES	4.52E-05	1.44E-05	1.84E-04	4.52E-05	1.44E-05	5	5
SAL_EO	1.18E-04	3.72E-05	5.67E-04	1.18E-04	3.72E-05	9	9
SAL_OS	4.63E-05	1.45E-05	2.09E-04	4.63E-05	1.45E-05	6	6
SAL_OO	4.82E-05	1.50E-05	2.66E-04	4.82E-05	1.50E-05	7	7
Bachelier1900	1.98E-05	6.76E-06	5.27E-05	1.98E-05	6.76E-06	1	1



Table C.6: Percentage Error<sub>CLOSE</sub> Rankings

GBPUSD	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	1.2540	0.9886	39.3365	1.2540	0.9886	10	3
Random	-0.4993	0.0000	0.5000	0.4993	0.0000	1	1
Yule1927	0.9371	1.0030	41.0665	0.9371	1.0030	6	9
Static	0.9959	0.8935	48.5536	0.9959	0.8935	7	2
YaoTan2000	0.9227	1.0090	15.6134	0.9227	1.0090	5	10
SAL_ES	0.8774	0.9944	37.2635	0.8774	0.9944	4	5
SAL_EO	0.5881	0.9891	78.5276	0.5881	0.9891	2	4
SAL_OS	1.1392	0.9986	38.2080	1.1392	0.9986	9	8
SAL_OO	0.8492	0.9974	38.3648	0.8492	0.9974	3	7
Bachelier1900	0.9962	0.9966	2.7324	0.9962	0.9966	8	6
EURGBP	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	1.0571	0.9885	25.5249	1.0571	0.9885	10	3
Random	-0.5002	-1.0000	0.5000	0.5002	1.0000	1	8
Yule1927	1.0246	0.9812	29.2827	1.0246	0.9812	8	2
Static	0.9999	0.8027	58.1124	0.9999	0.8027	6	1
YaoTan2000	0.8671	0.9974	11.8158	0.8671	0.9974	2	5
SAL_ES	0.9746	0.9956	25.8203	0.9746	0.9956	5	4
SAL_EO	0.9063	1.0090	52.7497	0.9063	1.0090	3	10
SAL_OS	1.0317	0.9975	27.0312	1.0317	0.9975	9	6
SAL_OO	0.9658	0.9999	27.0207	0.9658	0.9999	4	7
Bachelier1900	1.0008	1.0020	2.8463	1.0008	1.0020	7	9

Table C.7: Absolute Percentage Error<sub>CLOSE</sub> Rankings

GBPUSD	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	5.6489	1.4210	38.9489	5.6489	1.4210	7	5
Random	0.4993	0.0000	0.5000	0.4993	0.0000	1	1
Yule1927	6.0440	1.5600	40.6301	6.0440	1.5600	8	8
Static	9.0508	2.4740	47.7130	9.0508	2.4740	9	10
YaoTan2000	2.7042	1.1020	15.4051	2.7042	1.1020	3	3
SAL_ES	5.5065	1.4510	36.8648	5.5065	1.4510	4	7
SAL_EO	10.7590	2.2280	77.7893	10.7590	2.2280	10	9
SAL_OS	5.5315	1.4130	37.8227	5.5315	1.4130	6	4
SAL_OO	5.5306	1.4240	37.9736	5.5306	1.4240	5	6
Bachelier1900	1.2341	1.0000	2.6335	1.2341	1.0000	2	2
EURGBP	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	4.8015	1.3960	25.0915	4.8015	1.3960	4	4
Random	0.5002	1.0000	0.5000	0.5002	1.0000	1	1
Yule1927	5.5419	1.5980	28.7717	5.5419	1.5980	8	8
Static	13.9706	4.0460	56.4169	13.9706	4.0460	10	10
YaoTan2000	2.4328	1.0910	11.5951	2.4328	1.0910	3	3
SAL_ES	4.9361	1.4750	25.3629	4.9361	1.4750	5	7
SAL_EO	9.3999	2.2760	51.9133	9.3999	2.2760	9	9
SAL_OS	4.9481	1.4290	26.5945	4.9481	1.4290	6	5
SAL_OO	4.9551	1.4500	26.5800	4.9551	1.4500	7	6
Bachelier1900	1.2563	1.0060	2.7431	1.2563	1.0060	2	2

Table C.8: Squared Percentage Error<sub>CLOSE</sub> Rankings

GBPUSD	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	1,548.9600	2.0190	50,816.7300	1,548.9600	2.0190	7	5
Random	0.4993	0.0000	0.5000	0.4993	0.0000	1	1
Yule1927	1,687.3475	2.4320	60,312.2251	1,687.3475	2.4320	8	8
Static	2,358.0039	6.1200	40,765.1809	2,358.0039	6.1200	9	10
YaoTan2000	244.6292	1.2130	7,777.1655	244.6292	1.2130	3	3
SAL_ES	1,389.2622	2.1050	45,910.9165	1,389.2622	2.1050	4	7
SAL_EO	6,166.6787	4.9660	206,230.4594	6,166.6787	4.9660	10	9
SAL_OS	1,461.1516	1.9970	50,645.5558	1,461.1516	1.9970	5	4
SAL_OO	1,472.5458	2.0260	51,009.1962	1,472.5458	2.0260	6	6
Bachelier1900	8.4583	1.0000	210.1639	8.4583	1.0000	2	2
EURGBP	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	652.6592	1.9480	15,978.0661	652.6592	1.9480	4	4
Random	0.5002	1.0000	0.5000	0.5002	1.0000	1	1
Yule1927	858.5050	2.5530	22,649.4242	858.5050	2.5530	8	8
Static	3,377.1715	16.3700	46,276.8506	3,377.1715	16.3700	10	10
YaoTan2000	140.3675	1.1900	4,009.3983	140.3675	1.1900	3	3
SAL_ES	667.6850	2.1750	16,832.2668	667.6850	2.1750	5	7
SAL_EO	2,783.4095	5.1800	70,787.3584	2,783.4095	5.1800	9	9
SAL_OS	731.7236	2.0410	19,831.1263	731.7236	2.0410	7	5
SAL_OO	731.0164	2.1040	19,778.7327	731.0164	2.1040	6	6
Bachelier1900	9.1025	1.0110	245.5842	9.1025	1.0110	2	2

Table C.9: Error<sub>HIGH/LOW</sub> Rankings

GBPUSD	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	0.0001	0.0000	0.0083	0.0001	0.0000	4	2
Random	0.0060	0.0046	0.0062	0.0060	0.0046	10	10
Yule1927	-0.0008	-0.0007	0.0092	0.0008	0.0007	5	5
Static	0.0034	0.0023	0.0071	0.0034	0.0023	6	6
YaoTan2000	0.0054	0.0040	0.0073	0.0054	0.0040	8	7
SAL_ES	0.0000	-0.0002	0.0088	0.0000	0.0002	2	4
SAL_EO	-0.0060	-0.0042	0.0142	0.0060	0.0042	9	9
SAL_OS	0.0000	0.0000	0.0091	0.0000	0.0000	3	1
SAL_OO	0.0000	0.0000	0.0092	0.0000	0.0000	1	3
Bachelier1900	0.0054	0.0041	0.0061	0.0054	0.0041	7	8
EURGBP	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	0.0001	0.0001	0.0045	0.0001	0.0001	4	3
Random	0.0035	0.0027	0.0035	0.0035	0.0027	9	9
Yule1927	-0.0004	-0.0004	0.0051	0.0004	0.0004	5	5
Static	-0.0065	-0.0073	0.0038	0.0065	0.0073	10	10
YaoTan2000	0.0022	0.0015	0.0046	0.0022	0.0015	6	6
SAL_ES	0.0000	-0.0001	0.0049	0.0000	0.0001	3	4
SAL_EO	-0.0033	-0.0024	0.0076	0.0033	0.0024	8	8
SAL_OS	0.0000	0.0000	0.0051	0.0000	0.0000	2	1
SAL_OO	0.0000	0.0000	0.0051	0.0000	0.0000	1	2
Bachelier1900	0.0030	0.0023	0.0030	0.0030	0.0023	7	7

Table C.10: Absolute Error<sub>HIGH/LOW</sub> Rankings

GBPUSD	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	0.0053	0.0039	0.0064	0.0053	0.0039	3	5
Random	0.0060	0.0046	0.0062	0.0060	0.0046	9	8
Yule1927	0.0055	0.0039	0.0074	0.0055	0.0039	7	6
Static	0.0059	0.0059	0.0052	0.0059	0.0059	8	9
YaoTan2000	0.0045	0.0030	0.0064	0.0045	0.0030	1	1
SAL_ES	0.0053	0.0037	0.0070	0.0053	0.0037	2	2
SAL_EO	0.0093	0.0064	0.0123	0.0093	0.0064	10	10
SAL_OS	0.0054	0.0038	0.0073	0.0054	0.0038	5	3
SAL_OO	0.0054	0.0038	0.0074	0.0054	0.0038	4	3
Bachelier1900	0.0055	0.0041	0.0060	0.0055	0.0041	6	7
EURGBP	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	0.0030	0.0021	0.0035	0.0030	0.0021	2	3
Random	0.0035	0.0027	0.0035	0.0035	0.0027	8	8
Yule1927	0.0031	0.0022	0.0041	0.0031	0.0022	6	5
Static	0.0069	0.0074	0.0031	0.0069	0.0074	10	10
YaoTan2000	0.0028	0.0018	0.0042	0.0028	0.0018	1	1
SAL_ES	0.0030	0.0021	0.0039	0.0030	0.0021	3	2
SAL_EO	0.0052	0.0036	0.0065	0.0052	0.0036	9	9
SAL_OS	0.0031	0.0022	0.0040	0.0031	0.0022	5	4
SAL_OO	0.0030	0.0022	0.0041	0.0030	0.0022	4	6
Bachelier1900	0.0031	0.0023	0.0029	0.0031	0.0023	7	7

Table C.11: Squared Error<sub>HIGH/LOW</sub> Rankings

GBPUSD	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	6.89E-05	1.50E-05	5.64E-04	6.89E-05	1.50E-05	4	5
Random	7.49E-05	2.15E-05	4.95E-04	7.49E-05	2.15E-05	5	8
Yule1927	8.59E-05	1.53E-05	7.52E-04	8.59E-05	1.53E-05	9	6
Static	6.19E-05	3.53E-05	6.01E-04	6.19E-05	3.53E-05	2	9
YaoTan2000	6.19E-05	8.88E-06	7.21E-04	6.19E-05	8.88E-06	1	1
SAL_ES	7.70E-05	1.39E-05	7.14E-04	7.70E-05	1.39E-05	6	2
SAL_EO	2.38E-04	4.07E-05	2.02E-03	2.38E-04	4.07E-05	10	10
SAL_OS	8.21E-05	1.48E-05	7.92E-04	8.21E-05	1.48E-05	7	3
SAL_OO	8.39E-05	1.48E-05	8.36E-04	8.39E-05	1.48E-05	8	3
Bachelier1900	6.56E-05	1.71E-05	5.22E-04	6.56E-05	1.71E-05	3	7
EURGBP	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	2.07E-05	4.49E-06	1.82E-04	2.07E-05	4.49E-06	2	3
Random	2.40E-05	7.24E-06	1.46E-04	2.40E-05	7.24E-06	4	8
Yule1927	2.67E-05	4.74E-06	2.52E-04	2.67E-05	4.74E-06	8	5
Static	5.68E-05	5.43E-05	1.65E-04	5.68E-05	5.43E-05	9	10
YaoTan2000	2.56E-05	3.20E-06	2.95E-04	2.56E-05	3.20E-06	5	1
SAL_ES	2.37E-05	4.33E-06	2.23E-04	2.37E-05	4.33E-06	3	2
SAL_EO	6.95E-05	1.31E-05	4.54E-04	6.95E-05	1.31E-05	10	9
SAL_OS	2.56E-05	4.62E-06	2.51E-04	2.56E-05	4.62E-06	6	4
SAL_OO	2.64E-05	4.75E-06	2.71E-04	2.64E-05	4.75E-06	7	6
Bachelier1900	1.79E-05	5.43E-06	3.80E-05	1.79E-05	5.43E-06	1	7

Table C.12: Percentage Error<sub>HIGH/LOW</sub> Rankings

GBPUSD	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	0.0172	0.0164	33.6777	0.0172	0.0164	1	5
Random	0.2821	0.0000	48.7025	0.2821	0.0000	9	1
Yule1927	-0.0280	0.0308	42.2894	0.0280	0.0308	3	8
Static	-0.0739	-0.0089	47.4691	0.0739	0.0089	6	3
YaoTan2000	0.0727	-0.0249	15.8859	0.0727	0.0249	5	6
SAL_ES	-0.2030	-0.0253	36.5699	0.2030	0.0253	7	7
SAL_EO	-0.7300	-0.0352	70.7985	0.7300	0.0352	10	9
SAL_OS	0.0189	0.0063	35.7991	0.0189	0.0063	2	2
SAL_OO	-0.2641	-0.0141	35.1154	0.2641	0.0141	8	4
Bachelier1900	0.0514	0.5136	4.8078	0.0514	0.5136	4	10
EURGBP	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	-0.0637	0.0278	26.3006	0.0637	0.0278	4	6
Random	-0.1995	0.0000	29.2183	0.1995	0.0000	6	1
Yule1927	-0.1211	-0.0244	29.1325	0.1211	0.0244	5	4
Static	0.4907	0.0483	52.7903	0.4907	0.0483	10	8
YaoTan2000	-0.3095	0.3533	7.7128	0.3095	0.3533	9	10
SAL_ES	-0.0227	0.0029	25.4163	0.0227	0.0029	1	2
SAL_EO	0.2905	0.0275	57.5226	0.2905	0.0275	8	5
SAL_OS	-0.0611	0.0070	25.5001	0.0611	0.0070	3	3
SAL_OO	0.2380	0.0356	30.0826	0.2380	0.0356	7	7
Bachelier1900	0.0549	0.2000	2.3533	0.0549	0.2000	2	9

Table C.13: Absolute Percentage Error<sub>HIGH/LOW</sub> Rankings

GBPUSD	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	4.1623	0.7547	33.4195	4.1623	0.7547	3	6
Random	6.5221	0.9905	48.2646	6.5221	0.9905	8	8
Yule1927	5.0022	0.7402	41.9925	5.0022	0.7402	7	2
Static	7.5076	1.1670	46.8717	7.5076	1.1670	9	10
YaoTan2000	2.1156	0.7426	15.7445	2.1156	0.7426	2	3
SAL_ES	4.3967	0.7214	36.3052	4.3967	0.7214	6	1
SAL_EO	8.9642	1.0310	70.2325	8.9642	1.0310	10	9
SAL_OS	4.3401	0.7427	35.5351	4.3401	0.7427	4	4
SAL_OO	4.3701	0.7525	34.8434	4.3701	0.7525	5	5
Bachelier1900	1.0643	0.9371	4.6888	1.0643	0.9371	1	7
EURGBP	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	3.7068	0.7429	26.0382	3.7068	0.7429	6	4
Random	5.3057	0.9932	28.7332	5.3057	0.9932	8	8
Yule1927	3.8548	0.7267	28.8766	3.8548	0.7267	7	2
Static	11.2824	2.7540	51.5729	11.2824	2.7540	10	10
YaoTan2000	1.5178	0.7496	7.5683	1.5178	0.7496	2	6
SAL_ES	3.3990	0.7192	25.1880	3.3990	0.7192	3	1
SAL_EO	6.9816	1.0020	57.0981	6.9816	1.0020	9	9
SAL_OS	3.6787	0.7337	25.2334	3.6787	0.7337	5	3
SAL_OO	3.5581	0.7457	29.8724	3.5581	0.7457	4	5
Bachelier1900	1.0245	0.9397	2.1193	1.0245	0.9397	1	7

Table C.14: Squared Percentage Error<sub>HIGH/LOW</sub> Rankings

GBPUSD	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	1,134.2079	0.5696	50,463.1681	1,134.2079	0.5696	3	6
Random	2,371.9171	0.9812	87,332.3657	2,371.9171	0.9812	9	8
Yule1927	1,788.4704	0.5478	64,009.4300	1,788.4704	0.5478	7	2
Static	2,253.2070	1.3610	40,900.9738	2,253.2070	1.3610	8	10
YaoTan2000	252.3706	0.5514	7,161.3698	252.3706	0.5514	2	3
SAL_ES	1,337.4032	0.5205	43,297.2318	1,337.4032	0.5205	6	1
SAL_EO	5,013.1543	1.0630	147,999.1779	5,013.1543	1.0630	10	9
SAL_OS	1,281.5799	0.5517	57,543.2245	1,281.5799	0.5517	5	4
SAL_OO	1,233.1040	0.5663	38,775.0841	1,233.1040	0.5663	4	5
Bachelier1900	23.1185	0.8782	1,386.1664	23.1185	0.8782	1	7
EURGBP	Mean	Median	SD	Mean	Median	Rank( Mean )	Rank( Median )
RW	691.7366	0.5519	24,420.6707	691.7366	0.5519	5	4
Random	853.7514	0.9865	21,837.8995	853.7514	0.9865	7	8
Yule1927	848.7073	0.5280	79,348.5200	848.7073	0.5280	6	2
Static	2,786.7089	7.5830	42,497.7066	2,786.7089	7.5830	9	10
YaoTan2000	59.5818	0.5619	1,505.1834	59.5818	0.5619	2	6
SAL_ES	645.9464	0.5172	72,231.2094	645.9464	0.5172	3	1
SAL_EO	3,309.0520	1.0040	400,183.1173	3,309.0520	1.0040	10	9
SAL_OS	650.2529	0.5383	24,155.2589	650.2529	0.5383	4	3
SAL_OO	905.0018	0.5560	127,367.6878	905.0018	0.5560	8	5
Bachelier1900	5.5403	0.8830	107.8698	5.5403	0.8830	1	7

Table C.15: Speculative Accuracy Results

GBPUSD	Mean	Median	SD	Rank(Mean)	Rank(Median)
RW	-4.9823	0.1808	46.7639	2	6
Random	-5.4203	0.0041	47.2621	8	10
Yule1927	-5.3350	0.1963	49.2195	6	2
Static	-5.4924	0.0372	48.0993	9	9
YaoTan2000	-5.4096	0.1834	54.3293	7	5
SAL_ES	-5.2780	0.2127	48.3546	5	1
SAL_EO	-5.6870	0.1466	50.0570	10	7
SAL_OS	-5.2604	0.1951	48.4494	4	3
SAL_OO	-5.2342	0.1875	48.3087	3	4
Bachelier1900	-3.4028	0.0561	38.0326	1	8
EURGBP	Mean	Median	SD	Rank(Mean)	Rank(Median)
RW	-3.8681	0.1923	27.4897	3	4
Random	-4.2246	0.0026	28.0373	10	10
Yule1927	-3.9649	0.2120	29.5461	7	1
Static	-4.1669	0.0040	28.7982	9	9
YaoTan2000	-3.4694	0.1828	30.5779	2	6
SAL_ES	-3.9533	0.2047	29.1606	6	2
SAL_EO	-4.1208	0.1402	29.6785	8	7
SAL_OS	-3.8825	0.2029	29.1087	4	3
SAL_OO	-3.9007	0.1860	28.9130	5	5
Bachelier1900	-2.8164	0.0537	49.5507	1	8

Table C.16: Speculative Accuracy<sub>-1</sub> Results

GBPUSD	Mean	Median	SD	Rank(Mean)	Rank(Median)
RW	-4.9836	0.1808	46.7638	2	6
Random	-5.4214	0.0041	47.2620	8	10
Yule1927	-5.3361	0.1963	49.2194	6	2
Static	-5.4936	0.0372	48.0991	9	9
YaoTan2000	-5.4107	0.1834	54.3292	7	5
SAL_ES	-5.2792	0.2127	48.3545	5	1
SAL_EO	-5.6882	0.1466	50.0569	10	7
SAL_OS	-5.2616	0.1951	48.4493	4	3
SAL_OO	-5.2354	0.1875	48.3086	3	4
Bachelier1900	-3.4035	0.0561	38.0326	1	8
EURGBP	Mean	Median	SD	Rank(Mean)	Rank(Median)
RW	-3.8692	0.1923	27.4895	3	4
Random	-4.2256	0.0026	28.0372	10	10
Yule1927	-3.9658	0.2120	29.5460	7	1
Static	-4.1679	0.0040	28.7981	9	9
YaoTan2000	-3.4704	0.1828	30.5778	2	6
SAL_ES	-3.9543	0.2047	29.1605	6	2
SAL_EO	-4.1218	0.1402	29.6784	8	7
SAL_OS	-3.8835	0.2029	29.1086	4	3
SAL_OO	-3.9017	0.1860	28.9129	5	5
Bachelier1900	-2.8177	0.0537	49.5506	1	8

Table C.17: Symlog<sub>10</sub>(Speculative Accuracy) Results

GBPUSD	Mean	Median	SD	Rank(Mean)	Rank(Median)
RW	-0.2137	0.1808	1.0218	3	6
Random	-0.3349	0.0041	1.0912	10	10
Yule1927	-0.2417	0.1963	1.0527	7	2
Static	-0.3253	0.0372	1.0962	9	9
YaoTan2000	-0.1222	0.1834	0.9508	2	5
SAL_ES	-0.2198	0.2127	1.0420	6	1
SAL_EO	-0.2761	0.1466	1.0695	8	7
SAL_OS	-0.2144	0.1951	1.0328	5	3
SAL_OO	-0.2143	0.1875	1.0329	4	4
Bachelier1900	-0.0011	0.0561	0.6018	1	8
EURGBP	Mean	Median	SD	Rank(Mean)	Rank(Median)
RW	-0.1933	0.1923	1.0148	4	4
Random	-0.3291	0.0026	1.0744	10	10
Yule1927	-0.2131	0.2120	1.0363	7	1
Static	-0.3263	0.0040	1.0762	9	9
YaoTan2000	-0.0876	0.1828	0.8992	2	6
SAL_ES	-0.2094	0.2047	1.0251	6	2
SAL_EO	-0.2674	0.1402	1.0459	8	7
SAL_OS	-0.1859	0.2029	1.0165	3	3
SAL_OO	-0.2033	0.1860	1.0166	5	5
Bachelier1900	0.0078	0.0537	0.5685	1	8

Table C.18:  $\text{Symlog}_e$ (Speculative Accuracy) Results

GBPUSD	Mean	Median	SD	Rank(Mean)	Rank(Median)
RW	-0.4719	0.1808	1.5399	3	6
Random	-0.6346	0.0041	1.6077	10	10
Yule1927	-0.5167	0.1963	1.5772	7	2
Static	-0.6252	0.0372	1.6152	9	9
YaoTan2000	-0.3467	0.1834	1.4780	2	5
SAL_ES	-0.4888	0.2127	1.5692	6	1
SAL_EO	-0.5619	0.1466	1.5989	8	7
SAL_OS	-0.4774	0.1951	1.5574	4	3
SAL_OO	-0.4777	0.1875	1.5571	5	4
Bachelier1900	-0.0921	0.0561	1.0052	1	8
EURGBP	Mean	Median	SD	Rank(Mean)	Rank(Median)
RW	-0.4441	0.1923	1.5158	4	4
Random	-0.6188	0.0026	1.5720	10	10
Yule1927	-0.4726	0.2120	1.5352	7	1
Static	-0.6154	0.0040	1.5718	9	9
YaoTan2000	-0.2858	0.1828	1.3812	2	6
SAL_ES	-0.4666	0.2047	1.5278	6	2
SAL_EO	-0.5382	0.1402	1.5462	8	7
SAL_OS	-0.4345	0.2029	1.5155	3	3
SAL_OO	-0.4552	0.1860	1.5161	5	5
Bachelier1900	-0.0719	0.0537	0.9387	1	8



Table C.19: tanh(Speculative Accuracy) Results

GBPUSD	Mean	Median	SD	Rank(Mean)	Rank(Median)
RW	-0.0484	0.2348	0.8239	6	6
Random	-0.1689	0.0053	0.9068	10	10
Yule1927	-0.0705	0.2544	0.8558	7	2
Static	-0.1572	0.0488	0.9107	9	9
YaoTan2000	0.0396	0.2381	0.7292	2	5
SAL_ES	-0.0470	0.2752	0.8410	5	1
SAL_EO	-0.1050	0.1911	0.8722	8	7
SAL_OS	-0.0458	0.2530	0.8315	3	3
SAL_OO	-0.0458	0.2433	0.8320	4	4
Bachelier1900	0.0811	0.0736	0.4062	1	8
EURGBP	Mean	Median	SD	Rank(Mean)	Rank(Median)
RW	-0.0326	0.2495	0.8261	4	4
Random	-0.1704	0.0034	0.9006	10	10
Yule1927	-0.0510	0.2743	0.8537	7	1
Static	-0.1682	0.0053	0.9035	9	9
YaoTan2000	0.0599	0.2373	0.7031	2	6
SAL_ES	-0.0463	0.2651	0.8383	6	2
SAL_EO	-0.1074	0.1829	0.8659	8	7
SAL_OS	-0.0258	0.2628	0.8297	3	3
SAL_OO	-0.0435	0.2414	0.8304	5	5
Bachelier1900	0.0821	0.0705	0.3941	1	8

Table C.20: Weighted Speculative Accuracy

GBPUSD	Mean	Median	SD	Rank(Mean)	Rank(Median)
RW	-0.0020	0.0191	0.1680	5	6
Random	-0.0015	0.0003	0.2160	2	10
Yule1927	-0.0027	0.0208	0.1767	9	2
Static	-0.0019	0.0027	0.2047	4	9
YaoTan2000	-0.0021	0.0204	0.1355	8	3
SAL_ES	-0.0020	0.0225	0.1710	7	1
SAL_EO	-0.0031	0.0145	0.1923	10	7
SAL_OS	-0.0016	0.0197	0.1705	3	5
SAL_OO	-0.0020	0.0198	0.1706	6	4
Bachelier1900	0.0001	0.0059	0.0745	1	8
EURGBP	Mean	Median	SD	Rank(Mean)	Rank(Median)
RW	0.0013	0.0208	0.1690	4	6
Random	-0.0007	0.0001	0.2136	8	10
Yule1927	0.0028	0.0235	0.1741	1	2
Static	-0.0006	0.0003	0.2123	7	9
YaoTan2000	0.0016	0.0218	0.1289	3	4
SAL_ES	-0.0002	0.0235	0.1689	6	1
SAL_EO	-0.0010	0.0143	0.1863	9	7
SAL_OS	0.0026	0.0227	0.1675	2	3
SAL_OO	0.0004	0.0212	0.1674	5	5
Bachelier1900	-0.0011	0.0057	0.0742	10	8

# Bibliography

- Alexander, D., & Thomas, L. (1987). Monetary/asset models of exchange rate determination: How well have they performed in the 1980's? *International Journal of Forecasting*, 3, 53–64.
- Allen, F., & Karjalainen, R. (1999). Using genetic algorithms to find technical trading rules. *Journal of Financial Economics*, 51, 245–271.
- Andersen, T., & Bollerslev, T. (1998). Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. *The Journal of Finance*, 53(1), 219–265.
- Andersen, T., Bollerslev, T., & Lange, S. (1999). Forecasting financial market volatility: Sample frequency vis-à-vis forecast horizon. *Journal of Empirical Finance*, 6, 457–477.
- Arifovic, J., & Gençay, R. (2001). Using genetic algorithms to select architecture of a feedforward artificial neural network. *Physica A*, 289, 574–594.
- Arnéodo, A., Muzy, J.-F., & Sornette, D. (1998). “direct” causal cascade in the stock market. *The European Physical Journal B*, 2, 277–282.
- Ausloos, M. (2000). Statistical physics in foreign exchange currency and stock markets. *Physica A*, (285), 48–65.
- Bachelier, L. (1900). Théorie de la spéculation. *Annales Scientifiques de l'École Normale Supérieure*, 3(17), 21–86.
- Bagheri, A., Peyhani, H., & Akbari, M. (2014). Financial forecasting using anfis networks with quantum-behaved particle swarm optimization. *Expert Systems with Applications*, 41, 6235–6250.

- Bakhach, A., Tsang, E., & Jalalian, H. (2016). Forecasting directional changes in the fx markets. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)* (Web, 1–8). IEEE.
- Bakhach, A., Tsang, E., Ng, W., & Chinthapati, R. (2016). Backlash agent: A trading strategy based on directional change. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)* (Web, 1–9). IEEE.
- Barnett, W., & Serletis, A. (2000). Martingales, non-linearity, and chaos. *Journal of Economic Dynamics & Control*, *24*, 703–724.
- BBC. (2015). Record fines for currency market fix. Accessed 02-03-2020. Retrieved from <https://www.bbc.com/news/business-32817114>
- Bekiros, S., & Marcellino, M. (2013). The multiscale causal dynamics of foreign exchange markets. *Journal of International Money and Finance*, *33*, 282–305.
- Bera, A., & Higgins, M. (1993). Arch models, properties, estimation and testing. *Journal of Economic Surveys*, *7*(4), 305–366.
- Blattberg, R., & Gonedes, N. (1974). A comparison of the stable paretian and student distribution as statistical model for prices. *Journal of Business*, *47*, 244–280.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, *31*, 307–327.
- Bollerslev, T. (1987). A conditionally heteroscedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, *69*(3), 542–547.
- Bollerslev, T., Chou, R., & Kroner, K. (1992). Arch modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, *52*, 5–59.
- Borio, C., Claessens, S., Mojon, B., Shin, H., & Wooldridge, P. (Eds.). (2020). *International banking and financial market developments, march 2020*.
- Box, G., & Jenkins, J. (1970). *Time series analysis: Forecasting and control*. Holden-Day.

- Box, G., & Pierce, D. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332), 1509–1526.
- Brock, W., Dechert, W., & Scheinkman, J. (1987). *A test for independence based on the correlation dimension*. University of Wisconsin at Madison.
- Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock prices. *The Journal of Finance*, 47(5), 1731–1764.
- Brock, W., Scheinkman, J., Dechert, W., & LeBaron, B. (n.d.). A test for independence based on the correlation dimension. *Econometric Reviews*, 15(3), 197–235.
- Bui, L., Vu, V., & Dinh, T. (2018). A novel evolutionary multi-objective ensemble learning approach to forecasting exchange rates. *Data & Knowledge Engineering*, 114, 40–66.
- Chee, F., & Ridley, K. (2019). Eu fines barclays, citi, jp morgan, mufg and rbs \$1.2 billion for fx rigging. Retrieved from <https://www.reuters.com/article/us-eu-antitrust-banks-idUSKCN1SM0XS>
- Chen, A.-S., & Leung, M. (2004). Regression neural network for error correction in foreign exchange forecasting and trading. *Computers & Operations Research*, 31, 1049–1068.
- Cheung, Y.-W., & Chinn, M. (2001). Currency traders and exchange rate dynamics: A survey of the us market. *Journal of International Money and Finance*, 20, 439–471.
- Clarida, R., Sarno, L., Taylor, M., & Valente, G. (2003). The out-of-sample success of term structure models as exchange rate predictors: A step beyond. *Journal of International Economics*, 60, 61–83.
- Clark, P. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1), 135–155.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273–297.

- Das, P., Bisoi, R., & Dash, P. (2018). Data decomposition based fast reduced kernel extreme learning machine for currency exchange rate forecasting and trend analysis. *Expert Systems with Applications*, *96*, 427–449.
- Dash, R. (2018). Performance analysis of a higher order neural network with an improved shuffled frog leaping algorithm for currency exchange rate prediction. *Applied Soft Computing*, *67*, 215–231.
- Daubechies, I. (1992). *Ten lectures on wavelets*. Springer-Verlag.
- de Almeida, B., Neves, R., & Horta, N. (2018). Combining support vector machine with genetic algorithms to optimize investments in forex market with high leverage. *Applied Soft Computing*, *64*, 596–613.
- Dempster, M., & Jones, C. (2001). A real-time adaptive trading system using genetic programming. *Quantitative Finance*, *1*, 397–413.
- Dempster, M., Payne, T., Romahi, Y., & Thompson, G. (2001). Computational learning techniques for intraday fx trading using popular technical indicators. *IEEE Transactions on Neural Networks*, *12*(4), 744–754.
- DeVaney, R. (1992). *A first course in chaotic dynamical systems: Theory and experiment*. Addison-Wesley.
- Diebold, F. (1986). Testing for serial correlation in the presence of heteroskedasticity. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 323–328.
- Diebold, F. (1988). *Empirical modeling of exchange rate dynamics*. Springer Verlag, New York.
- Diebold, F., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.
- Diebold, F., & Nason, J. (1989). Nonparametric exchange rate prediction. *Journal of International Economics*, *28*, 512–332.
- Drucker, H., Burges, C., Kaufman, L., Smola, A., & Vapnik, V. (1996). Support vector regression machines. *Proceedings of the 9<sup>th</sup> International Conference on Neural Information Processing Systems*, 155–161.

- Dunis, C., Laws, J., & Sermpinis, G. (2011). Higher order and recurrent neural architectures for trading the eur/usd exchange rate. *Quantitative Finance*, *11*(4), 615–629.
- Dymova, L., Sevastjanov, P., & Kaczmarek, K. (2016). A forex trading expert system based on a new approach to the rule-base evidential reasoning. *Expert Systems with Applications*, *51*, 1–13.
- Eckmann, J., Kamphorst, S., & Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhysics Letters*, *4*(9), 973–977.
- Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans*. Society for Industrial and Applied Mathematics.
- Engel, C., & Hamilton, J. (1990). Long swings in the dollar: Are they in the data and do markets know it? *The American Economic Review*, *80*(4), 689–713.
- Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, *50*(4), 987–1007.
- Evans, C., Pappas, K., & Xhafa, F. (2013). Utilizing artificial neural networks and genetic algorithms to build and algo-trading model for intra-day foreign exchange speculation. *Mathematical and Computer Modelling*, *58*, 1249–1266.
- Evans, M., & Lyons, R. (2002). Order flow and exchange rate dynamics. *Journal of Political Economy*, *110*(1), 170–180.
- Fama, E. (1965a). Random walks in stock market prices. *Financial Analysts Journal*, *21*(5), 55–59.
- Fama, E. (1965b). The behaviour of stock market prices. *The Journal of Business*, *38*, 34–105.
- Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.
- Farlow, S. (1981). The gmdh algorithm of ivakhnenko. *The American Statistician*, *35*(4), 210–215.
- Farmer, J., & Joshi, S. (2002). The price dynamics of common trading strategies. *Journal of Economic Behaviour & Organisation*, *49*, 149–171.

- Farmer, J., & Siderowich, J. (1987). Predicting chaotic time series. *Physical Review Letters*, 59(8), 845–848.
- Frankel, F., & Froot, K. (1986). Understanding the us dollar in the eighties: The expectations of chartists and fundamentalists. *Economic Record, Special Edition*, 24–38.
- Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200), 675–701.
- Galeshchuk, S., & Mukherjee, S. (2017). Deep networks for predicting direction of change in foreign exchange rates. *Intelligent Systems in Accounting, Finance and Management*, 24, 100–110.
- Gençay, R. (1999). Linear, non-linear and essential foreign exchange rate prediction with simple technical trading rules. *Journal of International Economics*, 4, 91–107.
- Ghazali, R., Hussain, A., & Liatsis, P. (2011). Dynamic ridge polynomial neural network: Forecasting the univariate non-stationary and stationary trading signals. *Expert Systems with Applications*, 38, 3765–3776.
- Giles, C., Lawrence, S., & Tsoi, A. (2001). Noisy time series prediction using recurrent neural networks and grammatical inference. *Machine Learning*, 44, 161–183.
- Giles, C., & Maxwell, T. (1987). Learning, invariance and generalisation in high-order neural networks. *Applied Optics*, 26(23), 4972–4978.
- Global Policy Forum. (2013). Currency transaction taxes. Retrieved from <https://www.globalpolicy.org/global-taxes/currency-transaction-taxes.html>
- Granger, C. (1969). Investigating causal relations by econometric models and cross spectral methods. *Econometrica*, 37(3), 429–439.
- Granger, G., & Joyeux, R. (1980). An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1, 15–29.



- Grassberger, P., & Procaccia, I. (1983a). Characterization of strange attractors. *Physical Review Letters*, *50*(5), 189–208.
- Grassberger, P., & Procaccia, I. (1983b). Measuring the strangeness of strange attractors. *Physica 9D*, *9*, 189–208.
- Grassberger, P., & Procaccia, I. (1983c). Measuring the strangeness of strange attractors. *Physica D*, *9*, 189–208.
- Hajizadeh, E., Mahootchi, M., Esfahanipour, A., & Kh, M. (2019). A new nn-pso hybrid model for forecasting euro/dollar exchange rate volatility. *Neural Computing and Applications*, *31*, 2063–2071.
- Hann, T., & Steurer, E. (1996). Much ado about nothing? exchange rate forecasting: Neural networks vs. linear models using monthly and weekly data. *Neurocomputing*, *10*, 323–339.
- He, K., Chen, Y., & Tso, G. (2018). Forecasting exchange rate using variational mode decomposition and entropy theory. *Physica A*, *510*, 15–25.
- Hinton, G., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*, 1527–1554.
- Hirabayashi, A., Aranha, C., & Iba, H. (2009). Optimization of the trading rule in foreign exchange using genetic algorithm. *Proceedings of the 2009 IASTED International Conference on Advances in Computer Science and Engineering*, 1529–1536.
- Hosking, J. (1981). Fractional differencing. *Biometrika*, *68*(1), 165–176.
- Hryshko, A., & Downs, T. (2004). System for foreign exchange trading using genetic algorithms and reinforcement learning. *International Journal of Systems Science*, *35*, 763–764.
- Hsieh, A. (1989a). Modeling heteroscedasticity in daily foreign-exchange rates. *Journal of Business & Economic Statistics*, *7*(3), 307–317.
- Hsieh, A. (1989b). Testing for nonlinear dependence in daily foreign exchange rates. *The Journal of Business*, *62*(3), 339–368.
- Hsieh, A. (1991). Chaos and nonlinear dynamics: Application to financial markets. *The Journal of Finance*, *46*(5), 1839–1877.

- Huang, S.-C., Chuang, P.-J., Wu, C.-F., & Lai, H.-J. (2010). Chaos-based support vector regressions for exchange rate forecasting. *Expert Systems with Applications*, 37, 8590–8598.
- Ince, H., & Trafalis, T. (2006). A hybrid model for exchange rate prediction. *Decision Support Systems*, 42, 1054–1062.
- Ivakhnenko, A. (1966). Group method of data handling — a rival of the method of stochastic approximation. *Technological Forecasting and Social Change*, 12, 213–227.
- Jeanne, O., & Masson, P. (2000). Currency crises, sunspots and markov-switching regimes. *Journal of International Economics*, 50, 327–350.
- Jongen, R., Verschoor, W., Wolff, C., & Zwinkels, R. (2012). Explaining dispersion in foreign exchange expectations: A heterogeneous agent approach. *Journal of Economic Dynamics and Control*, 36, 719–735.
- Jordà, Ò., & Taylor, A. (2011). Performance evaluation of zero net-investment strategies. *NBER Working Paper No. 17150*, 1–40.
- Kaasra, I., & Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 10, 215–236.
- Kampouridis, M., & Otero, F. (2017). Evolving trading strategies using directional changes. *Expert Systems with Applications*, 73, 145–160.
- Kamruzzaman, J., & Sarker, R. (2004). Ann-based forecasting of foreign currency exchange rates. *Neural Information Processing – Letters and Reviews*, 3(2), 49–58.
- Kantelhardt, J., Zschiegner, S., Koscielny-Bunde, E., Havlin, S., Bunde, A., & Stanley, H. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A*, 316, 87–114.
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 – International Conference on Neural Networks*, 4, 1942–1948.
- Keynes, J. (1936). *The general theory of employment, interest and money*. MacMillan.

- Khashei, M., Bijari, M., & Ardali, G. (2009). Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks. *Neurocomputing*, *72*, 956–967.
- Khashei, M., Hejazi, S., & Bijari, M. (2008). A new hybrid artificial neural networks and fuzzy regression model for time series forecasting. *Fuzzy Sets and Systems*, *159*, 769–786.
- Kilian, L., & Taylor, M. (2003). Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics*, *60*, 85–107.
- Kim, K.-J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, *55*, 307–319.
- Kodogiannis, V., & Lolis, A. (2002). Forecasting financial time series using neural network and fuzzy system-based techniques. *Neural Computing & Applications*, *11*, 90–102.
- Korczak, J., & Hernes, M. (2017). Deep learning for financial time series forecasting in a-trader system. In *2017 proceedings of the federated conference on computer science and information systems* (pp. 905–912).
- Kuan, C.-M., & Liu, T. (1995). Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics*, *10*(4), 347–364.
- LeBaron, B. (1999). Technical trading profitability and foreign exchange intervention. *Journal of International Economics*, *49*, 125–143.
- Lee, T., White, H., & Granger, C. (1993). Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. *Journal of Econometrics*, *56*, 269–290.
- Leu, Y., Lee, C.-P., & Jou, Y.-Z. (2009). A distance-based fuzzy time series model for exchange rates forecasting. *Expert Systems with Applications*, *36*, 8107–8114.
- Leung, M., Che, A.-S., & Daouk, H. (2000). Forecasting exchange rates using general regression neural networks. *Computers & Operations Research*, *27*, 1093–1110.

- Levich, R., & Thomas, L. (1993). The significance of technical trading-rule profits in the foreign exchange market: A bootstrap approach. *Journal of International Money and Finance*, 12, 451–474.
- Lin, C.-S., Chiu, S.-H., & Lin, T.-Y. (2012). Empirical mode decomposition-based least squares support vector regression for foreign exchange rate forecasting. *Economic Modelling*, 29, 2583–2590.
- Lippmann, R. (1987). An introduction to computing with neural nets. *IEEE Journal of Acoustics, Speech and Signal Processing*, 4–22.
- Lisi, F., & Schiavo, R. (1999). A comparison between neural networks and chaotic models for exchange rate prediction. *Computational Statistics & Data Analysis*, 30, 87–102.
- Ljung, G., & Box, G. (1978). On a measure of lack of fit in time series models. *Biometrika*, 66(2), 297–303.
- Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, 20, 130–141.
- Lu, C.-J., Lee, T.-S., & Chiu, C.-C. (2009). Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems*, 47, 115–125.
- Majhi, R., Panda, G., & Sahoo, G. (2009). Efficient prediction of exchange rates with low complexity artificial neural network models. *Expert Systems with Applications*, 36, 181–189.
- Mandelbrot, B. (1963a). New methods in statistical economics. *Journal of Political Economy*, 71, 421–440.
- Mandelbrot, B. (1963b). The variation of certain speculative prices. *The Journal of Business*, 36, 394–419.
- Mandelbrot, B. (1972). Statistical methodology for nonparametric cycles: From covariance to  $r/s$  analysis. *Annals of Economic and Social Measurement*, 1(3), 259–290.
- Mandelbrot, B., & Ness, J. (1968). Fractional brownian motions, fractional noises and applications. *SIAM Review*, 10(4), 422–437.

- Manzan, S., & Westerhoff, F. (2007). Heterogeneous expectations, exchange rate dynamics and predictability. *Journal of Economic Behavior & Organization*, 64, 111–128.
- Mariano, R. (2000). *Testing forecast accuracy*.
- May, D. (2006). Theory of speculation. Translation of L. Bachelier's 1900 PhD.
- Meese, R., & Rogoff, K. (1983a). Empirical exchange rates of the seventies: Do they fit out-of-sample? *Journal of International Economics*, 14, 3–24.
- Meese, R., & Rogoff, K. (1983b). The out-of-sample failure of empirical exchange rate models: Sampling error or misspecification? In J. Frenkel (Ed.), *Exchange rate and international macroeconomics* (pp. 67–112). University of Chicago Press.
- Meese, R., & Rose, A. (1991). An empirical assessment of non-linearities in models of exchange rate determination. *The Review of Economic Studies*, 58(3), 603–619.
- Mendes, L., Godinho, P., & Dias, J. (2012). A forex trading system based on a genetic algorithm. *Journal of Heuristics*, 18, 627–656.
- Minsky, M., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. MIT Press.
- Mussa, M. (1979). Empirical regularities in the behaviour of exchange rates and theories of the foreign exchange market. *Carnegie-Rochester Conference Series on Public Policy*, 11, 9–57.
- Nag, A., & Mitra, A. (2002). Forecasting daily foreign exchange rates using genetically optimized neural networks. *Journal of Forecasting*, 21, 501–511.
- Nassirtoussi, A., Aghabozorgi, S., Wah, T., & Ngo, D. (2015). Text mining of news-headlines for forex market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment. *Expert Systems with Applications*, 42, 306–324.
- Neely, C., Weller, P., & Dittmar, R. (1997). Is technical analysis in the foreign exchange market profitable, a genetic programming approach. *The Journal of Financial and Quantitative Analysis*, 32(4), 405–426.

- Nelson, D. (1991). Conditional heteroscedasticity in asset returns: A new approach. *Econometrica*, *59*(2), 347–370.
- Ni, H., & Yin, H. (2009). A self-organising mixture autoregressive network for fx time series modelling and prediction. *Neurocomputing*, *72*, 3529–3537.
- Nielsen, M. (2015). *Neural networks and deep learning*. Determination Press.
- Ozturk, M., Toroslu, I., & Fidan, G. (2016). Heuristic based trading system on forex data using technical indicator rules. *Applied Soft Computing*, *43*, 170–186.
- Packard, N., Crutchfield, J., Farmer, J., & Shaw, R. (1980). Geometry from a time series. *Physical Review Letters*, *45*(9), 712–716.
- Pal, M., Rao, P., & Manimaran, P. (2014). Multifractal detrended cross-correlation analysis on gold, crude oil and foreign exchange rate time series. *Physica A*, *416*, 452–460.
- Panda, C., & Narasimhan, V. (2007). Forecasting exchange rate better with artificial neural network. *Journal of Policy Modeling*, *29*, 227–236.
- Park, C., & Irwin, S. (2007). What do we know about the profitability of technical analysis? *Journal of Economic Surveys*, *21*(4), 786–826.
- Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Prediction stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, *42*, 259–268.
- Peng, Y., Albuquirqui, P., de Sá, J., Padula, A., & Montenegro, M. (2018). The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with support vector regression. *Expert Systems with Applications*, *97*, 177–192.
- Pesaran, M., & Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, *10*(4), 461–465.
- Pradeepkumar, D., & Ravi, V. (2017). Forecasting financial time series volatility using particle swarm optimization trained quantile regression neural network. *Applied Soft Computing*, *58*, 35–52.

- Ravi, V., Pradeepkumar, D., & Deb, K. (2017). Financial time series prediction using hybrids of chaos theory, multi-layer perceptron and multi-objective evolutionary algorithms. *Swarm and Evolutionary Computation*, 36, 136–149.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408.
- Rosser, J. (2008). Chaos theory before lorenze. *Nonlinear Dynamics, Psychology, and Life Sciences Special Issue: The Impact of Edward Lorenz*, 13(3), 257–269.
- Rossi, B. (2013). Exchange rate predictability. *Journal of Economic Literature*, 51(4), 1063–1119.
- Rout, M., Majhi, B., Majhi, R., & Panda, G. (2014). Forecasting of currency exchange rates using an adaptive arma model with differential evolution based training. *Journal of King Saud University – Computer and Information Sciences*, 26, 7–18.
- Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (pp. 318–362). The MIT Press.
- Rumelhart, D., & McClelland, J. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. The MIT Press.
- Sarantis, N. (1999). Modeling non-linearities in real effective exchange rates. *Journal of International Money and Finance*, 18, 27–45.
- Sarno, L., & Valente, G. (2009). Exchange rates and fundamentals: Footloose or evolving relationship? *Journal of the European Economic Association*, 7(4), 786–830.
- Scheinkman, J., & LeBaron, B. (1989). Nonlinear dynamics and stock returns. *The Journal of Business*, 62(3), 311–337.

- Sermpinis, G., Dunis, C., Laws, J., & Stasinakis, C. (2012). Forecasting and trading the eur/usd exchange rate with stochastic neural network combination and time-varying leverage. *Decision Support Systems, 54*, 316–329.
- Sermpinis, G., Laws, J., Karathanasopoulos, A., & Dunis, C. (2012). Forecasting and trading the eur/usd exchange rate with gene expression and psi sigma neural networks. *Expert Systems with Applications, 39*, 8865–8877.
- Sermpinis, G., Stasinakis, C., Theofilatos, K., & Karathanasopoulos, A. (2015). Modeling, forecasting and trading the eur exchange rates with hybrid rolling genetic algorithms – support vector regression forecast combinations. *European Journal of Operational Research, 247*, 831–846.
- Sermpinis, G., Theofilatos, K., Karathanasopoulos, A., Georgopoulos, E., & Dunis, C. (2013). Forecasting foreign exchange rates with adaptive neural networks using radial-basis functions and particle swarm optimization. *European Journal of Operational Research, 225*, 528–540.
- Sharpe, W. (1966). Mutual fund performance. *The Journal of Business, 39*(1), 119–138.
- Sharpe, W. (1994). The sharpe ratio. *The Journal of Portfolio Management, 21*(1), 49–58.
- Shen, F., Chao, J., & Zhao, J. (2015). Forecasting exchange rate using deep belief networks and conjugate gradient method. *Neurocomputing, 167*, 243–253.
- Shin, T., & Han, I. (2000). Optimal signal multi-resolution by genetic algorithms to support artificial neural networks for exchange rate forecasting. *Expert Systems with Applications, 18*, 257–269.
- Silvapulle, P., & Evans, M. (1998). Testing for serial correlation in the presence of dynamic heteroscedasticity. *Econometric Reviews, 17*(1), 35–55.
- Sortino, F., & Forsey, H. (1996). On the use and misuse of downside risk. *Journal of Portfolio Management, 22*, 35–42.
- Sterling ratio. (n.d.). Accessed 10-03-2020. Wikipedia. Retrieved from [https://en.wikipedia.org/wiki/Sterling\\_ratio](https://en.wikipedia.org/wiki/Sterling_ratio)



- Stošić, D., Stošić, D., Stošić, T., & Stanley, H. (2015). Multifractal analysis of managed and independent float exchange rates. *Physica A*, 428, 13–18.
- Sweeney, R. (1986). Beating the foreign exchange market. *The Journal of Finance*, 41(1), 163–182.
- Swiss Franc. (n.d.). Accessed 26-03-2020. Wikipedia. Retrieved from [https://en.wikipedia.org/wiki/Swiss\\_franc#End\\_of\\_capping](https://en.wikipedia.org/wiki/Swiss_franc#End_of_capping)
- Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand & L.-S. Young (Eds.), *Dynamical systems and turbulence* (pp. 366–381). Springer Berlin Heidelberg.
- Tay, F., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega: The International Journal of Management Science*, 29, 309–317.
- Tenti, P. (1996). Forecasting foreign exchange rates using recurrent neural networks. *Applied Artificial Intelligence*, 10, 567–581.
- Theil, H. (1966). *Applied economic forecasting*. Rand McNally.
- Theil, H., Cramer, J., Moerman, H., & Russchen, A. (1958). Economic forecasts and policy. *Contributions to Economic Analysis*, 31(35).
- Tong, H., & Lim, K. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society, Series B (Methodological)*, 42(3), 245–292.
- Treynor, J. (1965). How to rate management of investment funds. *Harvard Business Review*, 43, 63–75.
- Walczak, S. (2001). An empirical analysis of data requirements for financial forecasting with neural networks. *Journal of Management Information Systems*, 17(4), 203–222.
- White, H. (1980). A heteroskedasticity-constant covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838.
- White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5), 1097–1126.

- Whittle, P. (1951). *Hypothesis testing in time series analysis*. Uppsala, Almqvist & Wiksells Boktryckeri.
- Williams, R. (1986). The logic of activation functions. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (pp. 423–443). The MIT Press.
- Wolf, A., Swift, J., Swinney, H., & Vastano, J. (1985). Determining lyapunov exponents from a time series. *Physica D*, *16*, 285–317.
- Wooldridge, J. (1991). On the application of robust, regression-based diagnostics to models of conditional means and conditional variances. *Journal of Econometrics*, *47*, 5–46.
- Yao, J., & Tan, C. (2000). A case study on using neural networks to perform technical forecasting of forex. *Neurocomputing*, *34*, 79–98.
- Yu, L., Lai, K., & Wang, S. (2008). Multistage rbf neural network ensemble learning for exchange rates forecasting. *Neurocomputing*, *71*, 3295–3302.
- Yu, L., Wang, S., & Lai, K. (2005). A novel nonlinear ensemble forecasting model incorporating glar and ann for foreign exchange rates. *Computers & Operations Research*, *32*, 2523–2541.
- Yu, L., Wang, S., & Lai, K. (2009). A neural-network-based nonlinear metamodeling approach to financial time series forecasting. *Applied Soft Computing*, *9*, 563–574.
- Yule, G. (1927). On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *226*, 267–298.
- Zhang, G., & Hu, M. (1998). Neural network forecasting of the british pound/us dollar exchange rate. *Omega: The International Journal of Management Science*, *26*(4), 495–506.