Bond University
Faculty of Business

Option Pricing Using Artificial Neural Networks: An Australian Perspective

Joachim Tobias Hahn

November 2013

Submitted in total fulfilment of the requirements
of the degree of Doctor of Philosophy

Abstract

The thesis addresses the question of how option pricing can be improved using machine learning techniques. The focus is on the modelling of volatility, the central determinant of an option price, using artificial neural networks. This is done explicitly as a volatility forecast and its accuracy evaluated. In addition, its use in option pricing is tested and compared with a direction option pricing approach.

A review of existing literature demonstrated a lack of clarity with respect to the model development methodology used in the area. This issue is discussed and finally addressed along with a consolidation of the various modelling approaches undertaken previously by researchers in the field. To this end, a consistent process is developed to guide the specific model development.

Previous research has focused on index options, i.e. a single time series and some options related to it. The aim of the research presented here was to extend this to equity options, taking into consideration the particular characteristics of the underlying and the options.

The research focuses on the Australian equity option market before and after the global financial crisis. The results suggest that in the market and over the time frame studied, an explicit volatility model combined with existing deterministic models is preferable.

Beyond the specific results of the study, a detailed discussion of the limitations and methodological issues is presented. These relate not only to the methodology used here but the various choices and tradeoffs faced whenever machine learning techniques are used for volatility or option price modelling. Academic insight as well as practical applications depend critically on the understanding of these choices.

Declaration

This thesis is submitted to Bond University in fulfilment of the requirements of the degree of Doctor of Philosophy. This thesis represents my own original work towards this research degree and contains no material which has been previously submitted for a degree or diploma at this University or any other institution, except where due acknowledgement is made.

Signature Date

Additional Research Outcomes

The following publications and presentations were prepared up to and during the candidature, albeit not directly related to the research presented in this thesis.

Publications

- B. J. Vanstone and T. Hahn (2013). *Momentum Investing and the GFC: The Case of the S&P/ASX 100*. Available at SSRN: `http://ssrn.com/abstract=2312114`. Submitted to the 26th Australasian Finance and Banking Conference 2013
- B. J. Vanstone, T. Hahn, and G. Finnie (Aug. 20, 2012a). "Developing High-Frequency Foreign Exchange Trading Systems". In: *25th Australasian Finance and Banking Conference 2012*. Available at SSRN: `http://ssrn.com/abstract=2132390`
- B. Vanstone, T. Hahn, and G. Finnie (2012b). "Momentum returns to S&P/ASX 100 constituents". In: *JASSA* 3, pp. 12–18
- B. Vanstone, G. Finnie, and T. Hahn (Dec. 2012). "Creating trading systems with fundamental variables and neural networks: The Aby case study". In: *Mathematics and Computers in Simulation* 86. The Seventh International Symposium on Neural Networks. The Conference on Modelling and Optimization of Structures, Processes and Systems, pp. 78–91
- B. Vanstone and T. Hahn (Aug. 23, 2010). *Designing Stockmarket Trading Systems (with and without Soft Computing)*. Harriman House
- B. Vanstone, G. Finnie, and T. Hahn (2009). "Designing short term trading systems with artificial neural networks". In: *Advances in electrical engineering and computational science*. Ed. by S.-I. Ao and L. Gelman. Springer Science+Business Media B.V., pp. 401–409
- B. Vanstone, T. Hahn, and G. Finnie (2009b). "Fundamental investment research – do US results apply to Australian investors?" In: *JASSA* 4, pp. 13–16

- B. J. Vanstone, T. Hahn, and G. Finnie (2009a). "Returns to Selecting Value Stocks in Australia – The Aby Filters". In: *22nd Australasian Finance and Banking Conference 2009*. Available at SSRN: `http://ssrn.com/abstract=1436207`
- B. J. Vanstone and T. Hahn (July 2008). "Creating short-term stockmarket trading strategies using Artificial Neural Networks: A Case Study". In: *2008 International Conference of Computational Intelligence and Intelligent Systems (ICIIS), World Conference on Engineering (WCE 2008), London 2nd – 4th July, 2008*

Acknowledgements

My thanks go to Bond University for giving me the opportunity to undertake this research. I would especially like to thank my supervisors Bruce J. Vanstone and Gavin Finnie for their constructive feedback, support, and encouragement throughout the process.

I would also like to thank my family for their ongoing support and patience.

The research presented in this thesis used data supplied by the Securities Industry Research Centre of Asia-Pacific (SIRCA) including data from Thomson-Reuters and the Australian Securities Exchange (ASX).

Contents

List of Tables

List of Figures

List of Acronyms

AHE	average hedging error
AIC	Akaike information criterion
ANN	artificial neural network
ANOVA	analysis of variance
ASX	Australian Securities Exchange (previously Australian Stock Exchange)
ARCH	autoregressive conditional heteroscedasticity
ATM	at-the-money option
BIC	(Schwarz) Bayesian information criterion
BS	Black-Scholes (option pricing model or formula)
BSM	Black-Scholes-Merton (option pricing model)
CART	classification and regression trees
CEV	constant elasticity of variance model
CME	Chicago Mercentile Exchange
CRR	Cox-Rubinstein-Ross (option pricing model)
CS	Corrado-Su (option pricing model)
DAX	Deutscher Aktienindex
DM	Diebold-Mariano test statistic
DVF	deterministic volatility function
EGARCH	exponential GARCH (see below for GARCH)
EMU	European Monetary Union
ETO	exchange-traded option
GARCH	generalised autoregressive conditional heteroscedasticity
GFC	Global Financial Crisis of 2007–2011
GRNN	generalised regression (artificial) neural network
HHL	Haug-Haug-Lewis option pricing model
HV	historic volatility
ITM	in-the-money option
IV	implied volatility
Libor	British Bankers' Association London Interbank Offered Rate
LIFFE	London International Financial Futures and Options Exchange

MLP multi-layer perceptron
MAE mean absolute error
MAPE mean absolute percentage error
MdAE median absolute error
ME mean error
MPE mean percentage error
MSE mean squared error
MSPE mean squared prediction error
NMSE normalised mean squared error
NRMSE normalised root mean squared error
OLS ordinary least squares regression
OTC over-the-counter
OTM out-of-the-money option
RBF radial basis function
RMSE root (of) mean squared error
SABR stochastic alpha beta rho volatility model
SFE Sydney Futures Exchange
SLP single-layer perceptron
SSE sum of squared errors
SV stochastic volatility
SVM support vector machine
SVR support vector regression
US United States of America
USA United States of America
VIX Chicago Board Options Exchange Volatility Index
WS Wilcoxon signed rank test

List of Symbols

B	price of a bond
c	price of a European-style call option; price of a call option of unspecified type
C	price of an American-style call option
D_t	cash-flow (typically a dividend) received at time t
e	Euler's number ($e \approx 2.718\,28$)
f_t	value of a forward or futures contract at time t
F_t	price of a forward or futures contract at time t
I	income generally or interest specifically
K	strike price of an option or the delivery price of a forward or futures contract
P	price of a put option
q	income yield
r	risk-free rate (return on an idealised risk-free asset)
$\vec{r}$	returns series (optional subscript specifying the beginning and end of the sampling period)
R^2	coefficient of determination
σ	standard deviation
σ^2	variance
S_t	price of a single unit of the underlying security at time t
t	time at observation[1]
T	time to expiry or maturity

[1] t as a subscript may be missing if the context is sufficiently clear, i.e. typically at $t = 0$.

Chapter 1

Introduction

1.1 Background and Motivation

Since the introduction of modern portfolio theory five decades ago, the foundations were laid for the development of increasingly sophisticated financial instruments, markets and valuation models. The insights gained over this period have contributed substantially to the increase in efficiencies in modern economies and the businesses operating in them. They also contributed, however, to a number of challenges for businesses, finance and investment professionals, and individual consumers and individual clients.

The increase in our understanding of finance allowed researchers and practitioners to develop increasingly complex models to value instruments, allocate funds and manage principal-agent problems to varying degrees. The valuation tools are of particular interest for the purpose of this thesis. A major advantage has been the possibility of creating financial securities that are customised to the particular needs of individual clients, whether they are funds, businesses or individuals. They provide significant gains in the efficiency with which funds are allocated.

This development was further supported by an effort towards deregulation, particularly since the 1980s. Deregulating financial markets was a prerequisite to the introduction of new instruments, market mechanisms and distribution of information and products.

The increases in complexity and distribution, however, required significant investments in automation. The number of transactions as well as the speed at which they are settled would not be possible without computer support. The same is true for the valuation of securities. They typically require significant development of models for forecasting and pricing of securities, optimisation of fund allocation as well as accurate and timely performance measurement and reporting.

The success of any financial service provider, business or major individual client critically depends on the systems to perform these functions including development of such computer systems and researchers developing the models on which they are based.

Sophistication, deregulation and automation had another effect on the field of finance: a considerable increase in competition. Banks, who used to be the major providers of financial services together with insurance companies, no longer hold such a significant position in the market. Funds, whether they are mutual funds representing individual savers, businesses, pension funds,[2] as well as hedge funds, large individual investors control significant amounts of capital and exert considerable influence.

All market participants compete for access to and control of a limited number of investment opportunities. In their search for return, or more formally the right risk-return trade-off, they rely on a large number of models to forecast and value individual cash-flows and measure their associated risks. The competitive nature and the significant transparency of financial markets when compared to markets for real assets requires continuous improvement in model accuracy and refinement of both models and processes.

While there are many benefits of such progress, there are, however, several problems. It is commonly argued that the Global Financial Crisis (GFC) of 2007–2011 has its roots in the lack of regulatory supervision as a result of deregulation, inappropriate strategies to address principal-agency problems, especially with respect to the remuneration of finance professionals in several areas of the finance industry. Lastly, an over-reliance on the strictly mathematical models used to value securities without consideration of their limitations both theoretical as well as practical ones is understood to have been a factor as well. This is particularly true with respect to assumptions made on the return process of real estate in the United States of America (USA) and the risk associated with the investment in sovereign debt issued by members of the European Monetary Union (EMU).

Regardless of the merits of deregulation and its effects, the market structure that has emerged resulted, amongst other trends, in the development of a market for derivatives over the past twenty years that is active, broad with respect to geography, industries and firms, and accessible to even individual investors. The

[2] In Australia, these would typically be in the form of Superannuation Funds. Self-managed superannuation funds would be treated like individual investors for the purpose of this discussion.

aforementioned improvements in transactional efficiency and valuation accuracy are of great importance to them.

1.2 Research Goals and Hypotheses

The research presented in this thesis was motivated by a desire to address some of the challenges users of derivative securities face, in particular, challenges relating to the pricing of options. The principal problems faced are those of determining the correct price based on the value of the security, either to be ready to enter a transaction at a price more favourable than the one determined (as a price maker) or to decide whether an advertised price is sufficiently favourable (as a price taker) to enter a transaction. A second issue is the monitoring of the portfolio and the management of open positions, i.e. a decision whether to close out a position at any given point in time.

In both cases a prudent investor would take not only the individual transaction or instrument into consideration but the portfolio as a whole. Since this requires knowledge of all existing positions or their underlying position-generating strategy, such strategies are not typically subject to academic research. It is also difficult to generalise beyond the strategy under consideration limiting research to the processes leading to such decisions and the availability of suitable models to aid decision makers in them.

A final point of mostly practical concern are transaction management and settlement issues. These will not be discussed in any detail here for two reasons in particular: firstly, they are largely determined by the individual situation and internal organisation of the firm as well as the regulatory environment and market structure. Secondly, these areas are for the most part subject to efficiency gains rather than gaining additional insights, model development or any other area of academic research.

Instead, the research focuses on the first and arguably the most important step in the process: the valuation of options for the purpose of pricing them in organised markets. This specifically excludes the valuation for purposes other than market transactions, e.g. valuation for tax purposes or related-party transactions. It also excludes any transaction occurring in the over-the-counter market.

In the process, the focus will be on the valuation of equity options in the Australian market. Australia has one of the most active equity markets and a significant option market given the country's relatively small size. While the exclusion

of the over-the-counter market limits the research to only a fraction of the total market in such instruments, it nevertheless restricts it to the more conservative subset. The organised market is particularly transparent and the (near) absence of counter-party risk limits valuation to the actual cash-flows, their incidence (timing), magnitude (amount) and risk.

The valuation is only one part of the pricing process. In addition to determining the fair value of the instruments, it is also important to consider the market's assessment of the value. In that sense the pricing and valuation may, and typically do, differ.

Another aspect of the research results from a particular feature of option valuation. While the value, and consequently the price, of an option depends on a number of variables, one, the future volatility of the underlying security, is of particular importance. The valuation problem is therefore also a forecasting problem. The two cannot be separated.

The final component determining the research direction are recent developments towards less prescriptive models for valuation purposes. This results from a variety of developments such as the trend towards increasing automation, higher transaction volumes, greater competition and more complex securities. This largely led to a greater focus on data-driven approaches as well as models that can be developed within fairly short time-frames.[3]

It is also likely that the experiences gained during the GFC will lead to even more focus on data-driven modelling. There is greater awareness today than up until 2007 that theoretical return distributions are insufficient to describe security returns. An exacerbating factor is the wider availability of data and the faster and thus easier processing of large data sets in an automated fashion.

As will be discussed in subsequent chapters, the focus of research in this area has been on artificial neural networks. While alternative machine learning techniques are available to researchers and practitioners, artificial neural networks are particularly well-suited to the type of data found in financial markets, such as large data sets of noisy data with non-linear relationships between the variables.

The question resulting from these developments is therefore if machine learning techniques, notably artificial neural networks (ANNs) can be used to enhance the models used for pricing Australian equity options or even replace them entirely.

[3]Short development periods are largely the result of models becoming irrelevant in the face of competition in even decreasing periods themselves. It is therefore necessary to develop models replacing existing ones to respond to the market pressures. While this is largely true for trading systems; these systems typically rely at least in part on valuation models themselves.

The following hypotheses should be seen in the context of this market definition, i.e. they relate to the Australian equity options market, a point that will not be reiterated for the purpose of clarity and brevity.

The problem can be approached in two different ways. The option pricing could be improved by using a better value for the volatility input or the pricing can be substituted with an ANN. The former is expressed in the following working hypothesis:[4]

Hypothesis 1 *ANNs can forecast volatility more accurately than traditional models.*

The hypothesis is based both on prior research as well as the understanding that volatility is a more complex process than many other time series thus lending itself to non-parametric estimation and forecasting techniques.

An important aspect of the hypothesis is the focus on option pricing. As will become clear when discussing the existing volatility forecasting models, there is no such thing as *the best model.* Instead, there is *at least one suitable model for any particular intended use of volatility.* Since the focus of the research is on option pricing, the intended use of the volatility is clear. This also has implications for the performance measure used.

The alternative approach to using machine learning techniques for option pricing is the direct one:

Hypothesis 2 *Option prices based on ANN models outperform those generated using a traditional pricing model with respect to market prices.*

Instead of forecasting volatility explicitly and assuming constancy of the value, the forecast is embedded in the pricing function as it is represented by the ANN. There is some loss of utility in this approach as the volatility forecast is not visible and therefore not accessible. It cannot be observed, reported, such as to a client or regulator, or constrained – though it is unclear if that would of benefit to any particular user– before being passed to the pricing function. The volatility could be extracted from the resulting price, however, using an existing pricing model if any of these points are of concern to the user of such a model.

It can be argued that the latter approach is necessarily better than the former. Given a particular data set, fitting the ANN directly should lead to a result that

[4]Corresponding test hypotheses H_0 will be introduced in Chapter 3.

cannot be worse, at least in-sample, than the volatility forecast followed by the pricing function.

This argument does not necessarily hold out-of-sample as the pricing function may impose constraints on the ultimate price in a way the ANN cannot.

Equally, Hypothesis 1 does not represent a prerequisite to Hypothesis 2. If Hypothesis 1 cannot be supported, i.e. its corresponding H_0 cannot be disproved, it would not necessarily preclude Hypothesis 2 from being true. The pricing model may be misspecified itself and an otherwise good forecast may not lead to a better price if it fails to compensate for, or even exacerbates, the problem.

Finally, a question arises whether the separation of the two problems can be achieved and if this can combine the various benefits, i.e. to have an explicit volatility forecast as well as out-performance due to the use of an ANN for option pricing. This is in contrast to an integrated ANN that does not attempt to model volatility directly. The issue of separability is further discussed in Chapter 3.

The first approach may be preferable to some users of machine learning techniques, particularly in a transition phase. Given their black-box nature, ANNs as well as many other techniques applied in various fields of study have long suffered from a lack of acceptance. Being able to observe at least some of the internal mechanisms, in this case intermediate results, without a loss of performance would be helpful. The fitting of a model so complex is likely to prove difficult.

On the other hand, the argument provided earlier does hold here as well: the pricing ANN should not systematically do worse than the single fitting model as it has access to all components and an additional volatility-forecasting ANN, which can be ignored, in principle. This depends on the nature of the volatility-related input, however, a point elaborated on further below. If a conflict occurs, the best-performing model is likely to be used except for educational uses, given the market pressures.

To answer these questions the thesis is organised as follows: Chapter 2 reviews in detail the literature regarding volatility forecasting, option pricing and machine learning. Some background is given first about the structure of the market and terminology to provide a context for the discussion that follows. A summary is provided detailing not only the major research gaps but also discussing some inherent epistemological limitations of research in this area. The discussion covers the two main areas, the financial background and the machine learning

background separately first with details at the point of convergence discussed in greater detail in section 2.5.

Given the interdisciplinary nature of the research, a brief description of principles, terminology, and notation is given in the literature review in sections 2.1–2.4. The central review of literature is presented in section 2.5 onwards.

Based on the literature review, a methodology for testing the hypotheses is developed and presented in detail in Chapter 3. Of particular importance in that context will be the measurement of performance and the statistics used to compare different models. Equally important, however, is the way the network training is structured. This represents an innovative approach to addressing problems of a similar nature, i.e. the simultaneous as opposed to the step-wise solving of related estimation and forecasting problems.

Chapter 4 presents the general characteristics of the data used for the research, the intermediate results of the various simulation and testing steps and provides a detailed summary of overall results.

The last chapter, Chapter 5, concludes with a summary of the results and offers an explanation of their implications for the individual hypotheses as well as the implications for the research field as a whole. An outline of a research agenda is provided at the end, which is derived from the insights gained through the course of this research.

1.3 Research Contributions

In the process of developing a methodology and analysing the results, several outcomes can be observed that contribute to the existing body of knowledge. These fall into two broad categories. Firstly, a number of methodological contributions are made by the research presented in the following chapters, notably Chapter 3:

- Previous research on machine learning and option pricing is largely limited to individual time series, in particular index value series. In the following chapters, this is extended to panel data, i.e. a data set with a cross-sectional as well as a time series component. There are several implications for the design of the ANN, the preparation of training and testing data, and most importantly for the evaluation of the results.

- Not only was past research focused almost exclusively on simple time series, it was also predominantly driven by a singular objective. Research exists on

volatility forecasting and other research on option pricing but the two are typically dealt with separately. The problem of pricing options, however, requires a comprehensive analysis of at least those two components, which is developed further below.

Secondly, contributions are made through empirical results, in particular the following:

- The question of usefulness of machine learning for the purpose of pricing equity options using actual data is examined whereas past research focuses almost exclusively on index options with equity options or synthetic data sets far less commonly investigated.
- Equally limited is the existing body of knowledge with respect to the transfer of knowledge from pricing options using machine learning techniques from the USA to Australia.
- The question of whether volatility forecasts can be improved through the use of ANNs has not been answered in the past with respect to option pricing. While general forecastibility has been examined and answered, the issue of applicability to option pricing is novel, especially with respect to Australian equity options.

Chapter 5 discusses the contributions in greater detail and with respect to the hypotheses and how they are derived from the empirical results.

Chapter 2

Literature Review

2.1 Overview

2.1.1 The Australian Equity and Option Markets

This chapter provides an overview of previous research in the area with particular focus on machine learning models for derivatives pricing. However, a brief introduction to the pricing of futures and options is given as well as brief discussion of existing volatility models – all without the use of any machine learning methods. The principal motivation is to clarify terminology and discuss the competing, i.e. benchmark models and methods as they are frequently used. In addition, it provides a context for the development of the methodology as it is introduced in this thesis. A summary of the market, its securities, and particular features will be provided.

Those familiar with option pricing may wish to skip sections 2.1 (this section) through 2.3, and those familiar with machine learning section 2.4, respectively.

As noted before the objective of the study is an investigation into the use of machine learning techniques for Australian equity options. The Australian equity market shares many of its characteristics with the equity markets of other large developed countries, though it is certainly smaller and somewhat more concentrated in certain industries such as mining and agriculture traditionally. There is also considerable concentration within the financial services sector[5] and some other regulated industries. The structure and notably the regulatory framework are not substantially different from those of many other countries.

[5]This fact will be of some importance in later chapters when discussing interest rates, the impact of the GFC and the data set being used.

Equity represents fractional ownership of a company and it gives the shareholder a number of rights. In the context of this thesis, the following rights are of particular interest:

Shareholders are entitled to receive their fractional share of the *dividend payouts*. Australia's tax system makes use of franking credits to avoid the issue of double taxation of income at the corporate and investor level. Individual shareholders receive franking credit along with the dividends which is then included in calculating their income tax.

If a company ceases operating, any proceeds are to be distributed to shareholders. As a residual claim to the assets of the firm, the shareholders are only entitled to receive proceeds after claims of other creditors are met. As this implies that the company is not managed as a going concern, it is typically not a central element of economic analysis.

Since shareholders own a fraction of a company they may have an interest to maintain their current level of partial ownership. In the event of new issues or other corporate actions that may dilute their fractional ownership, they have the right to access those issues first. While this right may be important to some investors, it may be waived or is traded separately by many.

Finally, a shareholder is also entitled to vote at the (annual) general meeting, this includes votes on particular topics as well as those affecting the appointment of directors of the company.

All of the above rights will play some role in the following discussion and the methodological design of the study. Only the first, dividend rights, will be modelled directly, however. Special or liquidation dividends are not considered, however, nor is the dilution of fractional ownership beyond the usual adjustments of price series discussed below.

Multiple classes of shares are not a significant part of the investment universe within Australia. A notable exception are Instalment securities of Telstra, a result of the privatisation of the previously government-owned telecommunications company. Equity-linked securities, especially rights offerings and other securities with deferred delivery features are also traded from time to time.

Australia's primary exchange is the Australian Securities Exchange (ASX), which resulted from a merger of the Australian Stock Exchange and the Sydney Futures Exchange and offers a trading environment for equity securities and their derivatives. Regional exchanges do exist but are of limited relevance. Since the middle of 2000, Standard and Poor's provide stock indices covering the Australian

equity market and dividing it in a number of strata. Their S&P/ASX index range is based on criteria such as market capitalisation, liquidity, domicile, as well as additional consideration regarding the stability of the constituent set. The most concentrated set is the S&P/ASX 20, while the S&P/ASX 200 and S&P/ASX 300 are the primary investment and broader market index, respectively. The number in the index name represents the approximate number of companies and thus securities included at any one time.

As mentioned, the ASX is not only a venue for trading equity securities but also for trading their derivatives. These include options with fixed terms, flexible options with negotiated terms, and futures contracts. All such derivatives are subject to central clearing and as is common for exchange-traded derivatives, the counter-party risk is reduced through novation, i.e. by structuring the contracts so that the clearing company is the counter-party to the market participants. Due to their size, contractual arrangements and regulation, such arrangements limit the likelihood and impact of defaults of any one party.

Among such arrangements is a requirement to post and maintain margins. While this feature is common in futures-markets, it is not as prevalent in the case of trading options as the exercise of an option and thus the obligation to deliver the underlying is uncertain, as will be discussed below. Parties to such derivatives contracts are required to provide sufficient funds (collateral) to satisfy the clearing company that they will be able to deliver as and when delivery of the security or cash is due.

2.1.2 Principles of Financial Modelling

Financial models and the time-series analysis underlying many of them is based on three interrelated principles that are of relevance here: the law of one price (the no-arbitrage condition), market efficiency, and the equality of price and value.

The core idea of all financial models is that two securities should have the same price if they represent the same value to their owner. Arbitrage should not be possible, i.e. the trading of the two securities should not yield an economic profit. Indeed, if one security was priced higher than the second, even though they represent the same value, an investor could sell the former, cash the proceeds and reinvest them in the latter and retain the difference. Naturally, such trading activity would increase the supply of the first and the demand for the second security, driving prices to one another until no such difference exists any

longer. The consequence is that trading activity brings about arbitrage and the no-arbitrage condition can only be met if there are investors looking for and acting upon actual arbitrage opportunities.

This leaves open the question of how arbitrage opportunities are found and what they are based on, i.e. how individuals identify situations where arbitrage opportunities exist.

This has natural implications for the characterisation of the whole market. Following Roberts (1967) (as cited in Campbell, Lo, and MacKinlay, 1997), a distinction is made and a hypothesis formulated about actual markets rather than model assumptions, with respect to the extent to which they permit arbitrage and the information set that can, or more specifically cannot, be exploited to generate economic profits. A market is considered to be efficient if no arbitrage opportunities exist.

A distinction is typically made following Fama (1970) and Fama (1991): A market is considered *weak-form efficient* if no economic profits can be made based on past price series; this includes all information related to prices, such as the returns series as well. In a *semistrong-form efficient* market, no economic profits are achievable by acting on publicly available information. Since price series are one such source, semistrong-form efficient markets also cover weak-form efficient ones. This also includes, however, any kind of information about the company's accounts, balance sheet, income statement, statement of cash flows, management or any other communication made by or about the firm in a public forum. The strongest claim would be that markets are *strong-form* efficient. Here, even private information cannot lead to economic profits as it too has been priced in typically through arbitrage by insiders.

While no particular view is taken in this thesis regarding the actual level of efficiency of markets, it is worth noting that the use of econometric models as well as machine learning techniques is typically based on the implicit, or more rarely explicit, assumption that markets are not efficient and that information discovery and resulting arbitrage trade is possible. In a purely efficient market, no such study would be required nor would it be worthwhile as no additional profit can be generated to compensate for the additional cost of conducting the research. A position is unnecessary in part since demonstrating superior performance of an alternative model could in principle be explained by the fact that standard assumptions of the default model are not met and that the market may still price assets rationally and efficiently under modified (and realistic) assumptions.

Equally, a failure to demonstrate a superior model may represent a failure to document that the existing model is inappropriate or that the market may be efficient but does not demonstrate either. For a detailed discussion, the reader is referred to the extensive literature on the epistemological issues of research into market efficiency. The research presented here focuses instead on the merits of the particular modelling techniques outlined below and related methodological concerns.

Finally, the previous concepts are applicable to a single security. Since it represents claims against future cash flows, those claims and the price paid represent the two assets. In equilibrium, i.e. when markets are efficient and no arbitrage opportunities exist, the price of the securities should just equal the value of the security (the claims against cash flows). The identity can be read either way. One can either calculate the value of a security and assert that the amount also equals its fair price. Alternatively, the price can be observed and the value be asserted as equalling that particular amount if one is willing to make the assumption of an equilibrium.

All three concepts but especially the first and third are critically important for the valuation of derivatives and the methodological discussion below. In order to be useful, an additional supporting assumption is made, however: frictionless and unconstrained trading. It is typically assumed that there are no limitations, regulatory or economic in nature, to act on information and to buy or sell securities, or create new securities where necessary. In particular, transactions are assumed to take place in a tax-free environment (at the margin) and that there are no constraints to shorting securities, i.e. to selling without owning them. These assumptions are made in this thesis as well, in particular the absence of transaction cost and taxes, when explicitly modelling cash flow patterns. By design, the ANNs may capture some effects resulting from market deviations from the assumptions implicitly. No information is, however, provided to them for this purpose.

2.2 Options Pricing

2.2.1 Forwards and Futures

In order to discuss the valuation of options, it is necessary to briefly discuss forward and futures contracts. A forward contract is an agreement between two parties to buy (and sell, respectively) an asset at a future point in time T at a

price K agreed upon today, i.e. at $t = 0$. Since the price at that future point in time may, and typically will, be different, the parties benefit or lose out on the transaction depending on whether they are the buyer or the seller, and whether the future price is above or below the delivery price. The parties agree to settle the contract at the future time by paying K to the seller, and delivering the asset(s) to the buyer. Typically a contract covers a larger number of assets of the same type, the quantity is referred to as the contract size.

While a forward contract (in short: a forward) is any contract of this nature, a futures contract (in short: a future) is the standardised version. The specific details of their standardisation are country and asset-specific, although they follow generally the same pattern. The contract size is chosen to be small enough to be useful to a large number of potential market participants, while being large enough to make trading economical. The choice of delivery dates is made with a similar trade-off in mind, however, in the case of physical assets, especially agricultural ones, additional constraints may exist due to the harvesting or breeding cycle. Similarly, commodities forward and futures contracts need to specify the location of delivery if it is a physical delivery, a significant question considering transportation cost related to the securities.

The valuation of a forward is fairly simple given that the contract simply defers delivery and payment for the future (see Hull, 2008, for a detailed discussion on which the following notation is based). The price of a future F_0 today ($t = 0$), based on arbitrage arguments is:

$$F_0 = S_0 e^{rT} \tag{2.1}$$

If the equation were not to hold, a trader could buy (sell) the forward contract and sell (buy) the asset, whose price is S_0 and make a profit from the difference. r is the risk-free rate, the rate at which proceeds from a sale can be reinvested and at which one can borrow funds. This assumes that no further income is associated with holding the asset. This is not true in every case; additional income in form of interest or dividend payments is a possibility among other sources. Depending on whether such income is a discrete payment I or a rate q, the forward price is one of:

$$F_0 = (S_0 - I)e^{rT} \tag{2.2}$$

$$F_0 = S_0 e^{(r-q)T} \tag{2.3}$$

The value of a long forward contract f based on the no-arbitrage assumption is:

$$f = (F_0 - K)e^{-rT} \tag{2.4}$$

since the value of two forwards with delivery price of F_0 and K has to be the discounted price difference at maturity $(F_0 - K)$. From 2.1–2.4 follows that the value of a long forward contract is:

$$f = S_0 - Ke^{-rT} \qquad \text{no income} \tag{2.5}$$

$$f = S_0 - I - Ke^{-rT} \qquad \text{payment of } I \tag{2.6}$$

$$f = S_0 e^{-qT} - Ke^{-rT} \qquad \text{income yield } q \tag{2.7}$$

A comprehensive discussion of pricing options including related literature in the following section is given by Hull (2008).

In addition to its standardisation, the futures contract has a second important feature that applies to Australian options as well. Futures contracts are not between the ultimate buyer and the ultimate seller, instead, they are between each party and the clearing house as mentioned before. In addition to reducing counter-party risk, to the degree that it is eliminated for all practical purposes, this has the effect of daily mark-to-market. The value of the future is determined as per the above discussions and the margin requirements set. Both parties must pledge certain amounts of capital, as set by the clearing house to reduce the default risk to the clearing house. The margin typically consists of two elements, an initial margin at the creation of the position and a maintenance margin that results from changes in the underlying asset price. Additional funds but also a reduction of funds may be needed or possible as a result. It should be noted that margin requirements do not affect the return to an investor since the investor is still entitled to any proceeds from the funds in the margin account. These restrictions do not typically apply to over-the-counter (OTC) derivatives, i.e. forward contracts.

Finally, forwards and futures may not be settled by delivering the actual underlying but rather paying cash. Any such agreement forms part of the contract and is most common for index futures, where no actual delivery is possible given the nature of the underlying asset.

2.2.2 Options Characteristics

As is well known (for more details see, e.g. Hull, 2008, or any other textbook for those unfamiliar with the financial aspect of the thesis), while forward or futures contracts are contractual commitments to settle at maturity, options are substantially more flexible. An option represents the right but not an obligation to the buyer to buy or sell the underlying asset at a future point in time and a predetermined price. To the seller, the contract is an obligation to deliver if the buyer chooses to exercise their right. The underlying asset can be any asset, financial or real. In the latter case, the asset may represent a course of action to be taken, a common occurrence in corporate finance.

If the buyer has the right to buy the underlying asset, the option is a call option. If, on the other hand, the right is one of selling the asset to the writer of the option (the seller of the options contract), the option is a put option.

Further distinctions are made subject to the nature of the underlying contract, the timing of exercise and the way the price is determined. Options cover a large variety of assets but those on futures, especially index and commodities futures, and equity securities are particularly common in context of exchange-traded derivatives. Options without payment in cash but rather a swapping of assets (swaptions), options on interest rates and fixed-income securities are also frequently used.

In regards to the timing of exercise, a common but by no means only distinction is made between European-style and American-style exercise. European-style exercise give the option buyer the right to exercise the option at a single future point in time. On the other hand, American-style exercise offers such a right up until a future point in time. The choice of exercise is largely determined by market conventions but can have significant implications on their valuation. In addition to the two styles discussed here, various alternative exist, with a range of exercise patterns.

The question arises how the price at delivery is determined. The simplest and most common case for exchange-traded options (ETOs) is the use of a fixed, pre-set price, the strike price. A common alternative in certain markets is the use of the average price over the life of the option using various formulae.

Options that do not follow the simple pattern of European- or American-style exercise and a single delivery price are referred to as exotic options and are largely beyond the discussion of this thesis but will play a minor role in discussion of

volatility models. These are almost exclusively traded OTC resulting generally in a lack of data availability, price transparency and potentially significant counterparty risk.

A final note on terminology: an option that has value to its holder is called an in-the-money (ITM) option, in the case of a call (put) this occurs when the current price is above (below) the strike price. An out-of-the-money (OTM) option is the reverse, it represents an option that has currently no value if executed and an at-the-money (ATM) option being the marginal case where strike and current price are equal.

2.2.3 Pricing of European-style Options

Options with European-style exercise are sufficiently simple to allow for a closed-form solution following analytical approaches. While numerical methods are equally applicable, the closed-form solution to these derivatives has become one of the foundations of modern finance.

Black and Scholes (1973) and Merton (1973) derived the pricing model based on the crucial no-arbitrage assumption, which will be referred to from here on as the Black-Scholes-Merton (BSM) model. The following discussion is based on Hull (2008) and Haug (2007). The price at time $t = 0$ of a European call c and a European p is:

$$c = S_0 e^{-qT} \mathrm{N}(d_1) - K e^{-rT} \mathrm{N}(d_2) \tag{2.8}$$

$$p = K e^{-rT} \mathrm{N}(-d_2) - S_0 e^{-qT} \mathrm{N}(-d_1) \tag{2.9}$$

where

$$d_1 = \frac{\ln\left(S_0/K\right) + \left(r - q + \sigma^2/2\right) T}{\sigma\sqrt{T}}$$

$$d_2 = \frac{\ln\left(S_0/K\right) + \left(r - q - \sigma^2/2\right) T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$

using the notation by Hull (2008):

S_0 is the stock price at the time of pricing (“today”),

K the strike price similar to that of a futures contract above,

T the time to expiration,

q the dividend yield,

σ the (annual) volatility, and

r the risk-free rate as before.

$N(\cdot)$ represents the cumulative standardised normal distribution. Note that d_1 and d_2 are shorthand parameters; their subscripts do not refer to future points in time. The additional considerations to make in the case of equity options is that they represent existing stock so no dilution occurs and the absence of dividend payments. Instead of individual payments (Merton, 1973) modified the formula to account for a diffusion process and the result is the use of q as a dividend yield rather than individual payments. He also points out that the model is homogeneous with respect to the underlying price and the strike price, a property exploited frequently in ANN-based pricing.

Alternatively, Hull (2008, p. 298–299) points out that "European options can be analyzed by assuming that the stock price is the sum of two components: a riskless component that corresponds to the known dividends during the life of the option and a risky component" and that "[t]he Black-Scholes formula is […] correct if S_0 is equal to the risky component of the stock price and σ is the volatility of the process followed by the risky component."

It is worth noting that the above pricing formulas are similar to those for futures given their common basis in models based on the no-arbitrage condition. In particular, equation 2.8 is equivalent to equation 2.7 of a long forward both representing a purchase. The difference is in the uncertainty of exercise of the option subject to the price behaviour versus the certainty of delivery in the case of the forward.

The above formulas do not apply, however, to futures options. While Australian options are spot options, i.e. delivery is of the actual stock, other markets use futures to be delivered, which in turn are for the purchase (sale) of the underlying stock. Black (1976) showed that the pricing results in similar formulas, however, in that the dividend yield q is to be substituted with the risk-free rate in equation 2.8.

A final complication arises from the fact that some exchanges use futures-style margining for options in a way similar to that discussed in section 2.2.1. This includes the Sydney Futures Exchange (SFE) (before its merger with the ASX) and currently the ASX (Lajbcygier, 2004, among others by the same author) The margining of the option premium results in a further adjustment of the formula of the model by Black (1976). Under the usual assumptions of those models, the

risk-free rate is effectively set to $r = 0$ (Haug, 2007; Lieu, 1990). The pricing of options with futures-style margining was introduced by Asay (1982). Lieu (1990) shows that the interest rate effectively disappears from the pricing Black-Scholes formula and argues that early exercise is no longer relevant resulting in both exercise types to be priced in the same way (see Lajbcygier et al., 1996, for an example of using the European-style pricing for American-style exercise). Kuo (1993) extends this model to accommodate actual cash flows resulting from the marking-to-market inherent in margining and their financing by market participants. Finally, Chen and Scott (1993) show that the findings by Lieu (1990) apply more generally and confirm that early exercise is not beneficial.

As White (2000b) points out, however not only do the authors ignore marking-to-market (a criticism similar to Kuo's) but also that the models only apply to non-coupon bearing securities. This focus on transaction cost is particularly noteworthy given the suggestion by Dumas, Fleming, and Whaley (1996) that the problems with the Black-Scholes model as it is applied in practice may not be due to a misspecification but a mismatch between supply of and demand for certain options after the market 'crash' of 1987.

Various authors contributed additional adjustments to the BSM model, in particular, to account for a number of deviations from the assumptions made therein. A summary discussion of the most significant adjustments with relevant references is provided by Haug (2007, chapter 6).

2.2.4 Pricing of American-style Options

The pricing of options with American-style exercise generally follows the modelling of those of European-style (Hull, 2008; Haug, 2007). There are two significant differences that influence the choice of pricing model. Firstly, the presence of early exercise renders analytical approaches less useful, limiting them to special circumstances or as only approximate solutions. Secondly, the pricing of calls and puts diverges further beyond the differences in the case of the BSM model or other approaches to European options.

There are four principle ways of pricing options with American-style exercise:

1. closed-form pricing of call options,
2. closed-form approximations to pricing call or put options,
3. tree-based models, and

4. simulation-based approaches.

The approach pursued in this thesis falls in the last category but will be treated separately given the significantly different assumptions and process.

As will be discussed in the following section, early exercise is not optimal for calls on assets that do not pay income. In the context of this thesis, it would not be optimal to exercise a call if the company does not, or rather is not expected to, pay dividends. In such cases, the American call can be treated like a European one and priced accordingly. In any case, the same parameters are used in the pricing models: S_0, K, σ, T, r, and either q or D_t depending on the nature of the dividend model. Continuing the use of the notation by Hull (2008), the American call is referred to as C and the corresponding put as P.

All alternatives to this approach are ultimately approximations due to either the nature of their assumptions or due to the way they are implemented. Closed-form approximations exploit the fact that the early exercise is optimal only just before the dividend payout, in a practical context the ex-dividend date rather than the payment date. They further simplify dividend payments as a yield similar to the yield used in the previous sections on futures and European options.

The model by Barone-Adesi and Whaley (1987) is of practical importance according to Haug (2007) and is based on a quadratic approximation after determining critical values for the underlying prices. An alternative model by Bjerksund and Stensland (2002) is considered somewhat better by Haug (2007) and uses two exercise boundaries over time. The approximation models have the substantial benefit of being very fast when calculating prices and when computing the corresponding derivatives (see the following section). Their major disadvantage is the use of a dividend yield instead of modelling discrete dividends as they are expected to occur.

Tree-based models allow for the explicit modelling of discrete dividends and consequently allow for an accurate pricing of a large variety of options. The structure and evolution of the trees varies by model but is based on the current price of the underlying and moving through time. At each node the price is determined and exercise decisions (if applicable) are made. The values are aggregated back towards the root of the tree to determine the price of the option. The most common model is the binomial method by Cox, Ross, and Rubinstein (1979) and Rendleman and Bartter (1979) and is typically referred to as the Cox-Rubinstein-Ross (CRR) model. Despite it being relatively old, it remains among the most widely used (Haug, 2007). If the steps are chosen sufficiently small and there-

fore sufficiently many, the model converges to the BSM model for the European option.

The concept of a binomial tree, i.e. a tree whose nodes have exactly two child nodes, is also closely related to risk-neutral valuation. It is notable, and entirely a function of the no-arbitrage condition, that the expected return of the underlying asset does not feature in any of the models mentioned so far and does not in any other model. This simplifies modelling and interpretation. Alternatives to the binomial method exist such as the trinomial tree (Boyle, 1986), which introduces a no-change state in addition to the up and down movements in the binomial tree, or implied tree models.

The latter are of great practical importance albeit not for the simple options. They are worth mentioning, however, due to their particular relationship with volatility models. As discussed previously, volatility is one of the parameters of the pricing model and typically the critical one due to the relatively high sensitivity of the option price to changes in volatility but also because it is the one most difficult to estimate (S_0, K and T being known and r being either irrelevant or fairly constant over short periods of time). Implied tree models (Dupire, 1994; Derman and Kani, 1994; Rubinstein, 1994) use the implied volatility observed in ETOs to price exotic OTC options. The structure and pricing mechanisms are similar to the standard binomial and trinomial trees. The benefit is that the volatility and therefore future expectations from a large number of market participants are aggregated and used in markets where no such information exists or is difficult to get. In addition, pricing exotic options and 'vanilla' options (i.e. simple exercise patterns such as the European-style exercise) using the same implied volatilities is in itself a case of arbitrage pricing. The price attached to a particular volatility expectation is the same if the implied tree model is used as in the "vanilla" options thus extending the modelling approach to within-asset class pricing.

While tree-based models are more accurate with respect to dividend payments if executed correctly, they are computationally expensive and the accuracy can be compromised due to numerical issues, i.e. rounding errors and stability, as well as trade-offs between convergence and execution speed.

Lastly, simulation-based models involve the generation of suitable samples from known (joint) distributions of underlying asset prices. Monte-Carlo methods are frequently used when needed but they suffer from even more significant compu-

tational problems than the tree-based models. The use of many machine learning techniques, notably ANNs, also falls in this category.

2.2.5 Option Greeks and Additional Considerations

While the option price is of obvious interest to those seeking to trade options to ensure their assumptions about the future behaviour of the asset are properly reflected, the options' derivatives are of even greater importance. This is particularly true in the case of hedging, i.e. taking positions to reduce the risk to the overall portfolio. The major sensitivities of European-style options used for risk management are (Hull, 2008, p. 360):

$$\begin{aligned} \Delta_c &= \frac{\partial c}{\partial S} & \Delta_p &= \frac{\partial p}{\partial S} \\ \Gamma &= \frac{\partial^2 c}{\partial S^2} & &= \frac{\partial^2 p}{\partial S^2} \\ \mathrm{V} &= \frac{\partial c}{\partial \sigma} & &= \frac{\partial p}{\partial \sigma} \end{aligned}$$

In addition to Δ (Delta) and Γ (Gamma), which measure the sensitivity to the changes in the underlying prices, and 'Vega' (V), which measures the sensitivity to changes in volatility, sensitivity to the passing of time (Θ) and to changes in the interest rate (P) are of some importance. Θ is a natural result of trading, P is generally considered to be somewhat less important. Γ and V are also symmetric, i.e. they are identical for the call and the put. Due to their importance for risk management, it is highly desirable for any pricing model to be able to determine the sensitivities easily as well. They are easy to calculate for closed-form models and can typically be approximated for numerical approaches. As will be discussed further below, the sensitivities can also be used as performance metrics for the pricing function.

A last consideration with respect to the properties of option prices (Hull, 2008) and consequently constraints in the non-parametric fitting (see below) are various relationships between parameters. Firstly, option prices face upper bounds, irrespective of whether they are European- or American-style. Call options cannot be more expensive than the current price of the asset for if they were, it would be cheaper to buy the asset at the current time instead of paying a premium in excess of the asset price and purchase the asset later at an additional (positive) cost.

Equally, put options cannot be more expensive than the strike price. The argument is the same, only a loss could be achieved by buying the put and exercising the option. This implies that if a situation of this kind were to arise, simple arbitrage could yield a profit by selling the option and taking the opposite position in the stock today (the direction would depend on whether it is a call or a put). This is excluded based on the no-arbitrage assumption or would disappear from the market place as a result of arbitrage activity:

$$c \leq S_0, \quad C \leq S_0, \quad p \leq K, \quad \text{and} \quad P \leq K \tag{2.10}$$

An additional boundary exists for European puts; these cannot be worth more than the discounted strike price as this is the only time the option can be exercised. Given that the American put can be exercised at any point until then, the argument does not apply to an American put:

$$p \leq Ke^{-rT} \tag{2.11}$$

Lower boundaries also exist. While Hull (2008) uses an arbitrage argument, there is an alternative point to make: consider a long (short) future, an agreement to buy (sell) an asset at a future point in time for price K given the current price of S_0. Equation 2.5 showed that the price of such a futures contract is $f_l = S_0 - Ke^{-rT}$ ($f_s = -f_l$, respectively). Since an otherwise equal option can result in the same purchase (sale) but with the additional value to the option holder of being able to chose whether to exercise or not, the option cannot be cheaper than the corresponding futures value. The option also cannot be cheaper than being free. In the absence of dividends (note that equation 2.5 was used, the no-income case) the lower boundaries are thus:

$$c \geq \max\left(S_0 - Ke^{-rT}, 0\right) \qquad \text{and} \qquad p \geq \max\left(Ke^{-rT} - S_0, 0\right) \tag{2.12}$$

and since the American option can be exercised immediately, it also follows that

$$P \geq K - S_0 \tag{2.13}$$

If additional flexibility (optionality) gives rise to a lower boundary, it also follows that American options cannot be cheaper than their corresponding European

ones:

$$C \geq c \qquad \text{and} \qquad P \geq p \tag{2.14}$$

Finally an important relationship exists between a European call and a European put with identical features, namely (Hull, 2008):

$$c - p = S_0 - Ke^{-rT} \tag{2.15}$$

allowing for the creation of synthetic positions in one through a combination of the three remaining assets with Ke^{-rT} representing a cash position. In the case of American options and in the presence of dividends the relationships are as follows:

$$c - p + D = S_0 - Ke^{-rT} \tag{2.16}$$

$$S_0 - D - K \leq C - P \leq S_0 - Ke^{-rT} \tag{2.17}$$

A consequence of the various relationships is that early exercise is never optimal for an American call except just before the ex-dividend date and it is optimal for an American put whenever the current stock price is sufficiently low.

Finally, as time approaches expiry, the option price of ITM call options should converge towards the nominal difference between the strike price and the current price and an OTM call option should approach a value of $c = 0$.

2.3 Volatility Models

2.3.1 Overview of Modelling Approaches

As mentioned previously, the option contract specifies the underlying, the time to maturity, and the delivery price – in addition to the lot size and various other details. The valuation requires knowledge of three other variables, however: the current price of the underlying, the risk-free rate, and the volatility of returns. While the former two components are observable, volatility is the one element of the valuation model that is not determined nor is it easily observed. Its estimation is consequently the focus of both academic research and practitioners' modelling efforts.

Volatility modelling is complicated by the fact that volatility – in the way it is defined in these models, i.e. as standard deviation σ – is used in various areas

of finance. It is therefore impossible to give a comprehensive overview of models, techniques, and approaches. In addition to its use in derivatives pricing, volatility is central to most asset pricing models where it is the, or at least one of various, measures of risk. Equilibrium returns are, consequently, subject to the riskiness of the asset. The estimation of future risk is an integral part of valuing financial assets. This extends to the modelling of covariance for portfolio management as well as the measuring – though not the forecasting – of volatility for performance measurement and attribution purposes.

Limiting the volatility modelling to derivatives pricing still leaves a considerable amount of models. However, they can be grouped by their basic assumptions and level of sophistication, which leads to the following categories:

- historical volatility,
- stochastic volatility,
- volatility term-structure and volatility surface models (including local volatility), and
- non-parametric volatility, volatility term-structure and surface models.

These will be summarised in this and the following sections excluding, however, non-parametric models, which will be discussed in section 2.5.2.

2.3.2 Historical Volatility

Historical volatility makes few assumptions regarding the source or changes in volatility levels. It is limited to assuming that past volatility is the best estimate of future volatility and constitutes a naïve approach. It implies that volatility is constant as well, at least as far as the forecasting horizon is concerned.

Despite its limited set of assumptions, historical volatility models still require a number of parameters:

- What is the time frame over which to measure (past) volatility?
- What is the frequency at which to sample, e.g. daily, weekly, monthly?
- Which time series should be used and how is volatility to be calculated, i.e. what formula to use?
- What adjustments need to be to made?

The first two questions are somewhat related given that a sufficient sample size is required to achieve meaningful estimates, i.e. to minimise the impact of noise. Treating the two independently, however, the question of the time frame is common to most approaches and a clear answer is not given in the literature. The choice appears to be a function of the specific and implicit assumptions made by researchers of what constitutes a sufficient data set.

The common goal is to choose a time frame that reflects the current or future volatility level and not to include observations so old as to be no longer relevant to the security, such as would result from corporate restructurings leading to changes in the risk-level, for example. This question is significant for option pricing as options available for trading vary in time to maturity. If a functional dependency between the observation time frame and the forecasting horizon is assumed or suspected, a single time frame is clearly not feasible.

Similar to choices of the time frame, the choice of frequency appears to depend on the personal preferences of researchers in addition to the objectives of any particular study. A clear preference for any one choice does not emerge from the literature. This is in part due to the conflicting evidence from theoretical research on the one hand and practitioners' experiences. Poon and Granger (2003) comment on this pointing out that "[i]n general, volatility forecast accuracy improves as data sampling frequency increases relative to forecast horizon [(Andersen, Bollerslev, and Lange, 1999)]."

It is noteworthy, as discussed by the same authors, that Figlewski (1997) found that long-term forecasts benefit from aggregation, i.e. from lower sampling frequencies. This in turn is contrary to the theoretical discussion by Drost and Nijman (1993), who show that aggregation should preserve the features of the volatility process. Poon and Granger (2003) stress, however, that "it is well known that this is not the case in practice [...and that t]his further complicates any attempt to generalize volatility patterns and forecasting results." In light of these findings and the conclusions by these widely-cited authors, the question arises if any results from previous literature are applicable to a specific problem or if they are rather of interest with respect to the process they follow. The estimation and evaluation methodologies rather than the resulting findings are the likely contribution of volatility modelling research.

The above discussions regarding aggregation, sampling frequency and time frame are of some importance to the historical volatility models as well as to the stochastic volatility models. They both typically assume a single observation,

i.e. a single price or return, on which to base the volatility estimate. This single variable approach is only one of the choices, however. It is commonly applied to the closing price of a day, week, or month. The volatility is calculated as the standard deviation of (log-)returns over the sampling period.

Alternatively, the trading ranges can be used, a technique used more frequently prior to the availability of high-frequency data. Parkinson (1980) proposed a volatility measure based on the high and low observation instead of the closing price, assuming prior aggregation. Garman and Klass (1980) on the other hand propose a combination of the trading range and the close where the range is used as the current observation and the change in closing prices as the reference point. All these suffer from various problems, however, and none appear to be used frequently. For a discussion and related literature see Haug (2007).

2.3.3 Stochastic Volatility and the ARCH-Family of Models

Despite the various choices and adjustments in the process of developing a historical volatility model, they are fairly simple, parsimonious, and well-behaved. Over time, researchers realised, however, that they cannot reflect various features that are frequently (over time and across asset classes) present in observed volatility: heteroscedasticity, volatility clustering, and surprisingly long memory of volatility shocks.

In response to this and the more general view by finance professionals that volatility represents an asset in its own right, led to the development of more complex models that address both the forecasting power as well as the statistical properties of the model and its estimation procedures. The two notable classes are stochastic volatility (SV) models and autoregressive conditional heteroscedasticity (ARCH) models (Engle, 1982), in particular. While neither is particularly new, they are still actively researched and it is beyond the scope of this review to discuss them in any detail. The purpose is to introduce them only to the degree necessary for a reference model implementation.

Relaxing the assumption of constant volatility allows for the modelling of any number of models depending on the nature of the stochastic process underlying volatility. The models assume a constant mean and an error component, which follows a pre-specified stochastic process. This allows for the modelling of the asset returns. The difference is in the structure of the error term. It can be of a particular functional form (typically as differential equations) to account

for the various stylised facts, e.g. Heston (1993), Cox (1975) for the constant elasticity of variance (CEV) model, or Hagan et al. (2002) for the stochastic alpha beta rho (SABR) model. More commonly, researchers follow the traditional approach of a combination of moving averages and autoregressive components. The ARCH model only uses the latter, the generalised autoregressive conditional heteroscedasticity (GARCH) uses both components (Bollerslev, 1986), here using the notation by Poon and Granger (2003):

$$r_t = \mu + \epsilon_t \tag{2.18}$$

$$\epsilon_t = \sqrt{h_t} s_t \tag{2.19}$$

$$h_t = \omega + \sum_{k=1}^{q} \alpha_k \epsilon_{t-k}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{2.20}$$

ARCH(q) only uses the first two terms of h_t in equation 2.20 while GARCH(p, q) follows the full specification as above. Further ARCH-family models exist to reflect particular features such as asymmetric responses to shocks.

The comprehensive review by Poon and Granger (2003), which is also discussed by Poon and Granger (2007), revealed that implied volatility is usually better than models from the ARCH-family (referred to there as GARCH) but the difference is not as clear in the case of historical volatility compared with implied volatility and even less so when comparing historical and ARCH-family models concluding "as a rule of thumb, historical volatility methods work equally well compared with more sophisticated ARCH class and SV models" (Poon and Granger, 2003).

The authors stress that it is critical to understand the objective of the volatility model and choose selection criteria accordingly. They also emphasised that different models may work better or worse depending on the asset being studied. Finally, the review reveals that the GARCH(1, 1) model is among the most common in the ARCH family (where parameters are given at all), a point also made across the wider literature.

2.3.4 Volatility Adjustments

Regardless of the choice of parameters, a number of adjustments may have to be made to the volatility estimate to arrive at a suitable parameter for option pricing. The first and most significant adjustment, numerically, is the adjustment for the time frame. Option pricing formulas use the annual volatility by convention. It is therefore necessary to state volatility in annual terms irrespective of the chosen

sampling frequency. The adjustment can be stated in terms of the number of sampling periods in a typical year or in terms of the length each sample covers:

$$\sigma = \sigma_s \sqrt{N} = \frac{\sigma_s}{\sqrt{\tau}} \tag{2.21}$$

where N is the number of trading days (weeks, months, periods) per year when sampling at daily (weekly, monthly, any particular) frequency and τ the length of a sample as a fractional year $\tau = \frac{1}{N}$ (Hull, 2008; Haug, 2007).

In addition to frequency adjustments, additional corrections may have to be made to arrive at an accurate volatility estimate. These additional adjustments are model-specific in that they are meant to correct for particular assumptions of the option pricing model rather than for the volatility as such. It is therefore important to be clear about the purpose of the volatility model. This is true in more general terms as Poon and Granger (2007) explain in detail.

The issue of dividends in option pricing has been discussed before. As is pointed out above, one possibility is to separate the problem into two component, a riskless and a risky one. The volatility needs to be adjusted in this case to correct for the additional cash flow. In the case of such a presence of dividends, the common practice is that the BSM model requires an adjustment to the volatility estimate by a factor of

$$\frac{S_0}{S_0 - D} \tag{2.22}$$

where D represents the present value of dividends (Hull, 2008, p. 298, footnote 12) and S_0 the price of the underlying. This is the necessary adjustment in the presence of cash dividends.

The adjustment is needed as the dividend represents an additional benefit to the owner of the stock (see the discussion at the beginning of the chapter). The treatment of stock dividends is different as they do not represent additional value but instead add to the number of shares held but not (if done proportionately) to the fractional share of ownership. Depending on the nature of the issue and their source, different adjustments may have to be made to the option price. These are typically excluded from discussions of option pricing and indeed from their implementation (including various models in MATLAB). This literature review like the thesis as a whole do not address the question of stock dividends – or warrants for the same reason.

The asymmetry in dealing with dividends is only one example. As discussed in detail in Australian Securities Exchange (2010) and Australian Securities Exchange (2012), various events are treated quite differently. While ordinary dividends do not result in contractual adjustments, special dividends as well as a large variety of corporate action events result in adjustment of the contract and are consequently irrelevant for pricing purposes. These do not add to the uncertainty and consequently to the value of options.

Haug (2007) compares a variety of adjustment methods (for the purpose of dividend payments) for both European and American options and concludes that the above volatility adjustment performs poorly. The author shows that none of the computationally simple methods performs well across a variety of parameters. In particular the timing of dividend payments throughout the life of the option is frequently a source for mispricings. Only the computationally more expensive methods perform well[6] but the author also points out that the above adjustment is frequently used by practitioners.

Additional adjustments, especially those for different number of days (of the underlying and the interest payments) are discussed by Haug (2007). Gatheral (2006) points out a related issue of preventing arbitrage in such cases. Neither issue is of significance here but these imply some limitations of the methodology introduced in the next chapter.

The statements regarding time frame, frequency and adjustments apply not only to historical but also stochastic volatility models as well. In particular, they too require a conversion to annual equivalents of the volatility measure.

2.3.5 Volatility Term Structure and Surface Models

Assuming, for the following discussion, that a volatility model has been found, i.e. a method for forecasting volatility over at least a short forward period is possible with sufficient expected success, the question arises how to apply the volatility model to option pricing. Two issues need to be addressed: what adjustment is needed for the various levels of time to expiry of a set of options, and secondly, how is the estimate to be adjusted for varying strike prices. The problems can be discussed independently or simultaneously.

[6] A particular model proposed by the same author and others is also included but it too requires a numerical evaluation in at least some cases despite not being as expensive as traditional models.

The time-varying nature of volatility was one of the motivating factors for the development of ARCH-models. It is for the same reason that a single estimate cannot be applied to options with varying time to expiry (assuming a constant strike price). The volatility to be used for pricing needs to match that of the relevant period. One possibility is to develop a sufficient number of forecasting models for the various possible forward periods but this is likely infeasible. In the case of ARCH models, Poon and Granger (2003) suggests an iterative procedure to arrive at additional forecasts.

In general terms, a way of converting between different forward periods would be desirable. The relationship between the time to expiry and the volatility is referred to as the volatility term structure. Recognising this, Haug and Haug (1996) (as cited in Haug, 2007) show that the implied forward volatility reflects the informational content of the term structure embedded in the base volatilities. The corresponding model according to Haug (2007) is:

$$\sigma_F = \sqrt{\frac{\sigma_2^2 T_2 - \sigma_1^2 T_1}{T_2 - T_1}} \tag{2.23}$$

and there is a lower boundary for

$$\sigma_2 = \sigma_1 \sqrt{\frac{T_1}{T_2}} \tag{2.24}$$

under the usual assumptions and suitable conditions, in particular that $T_2 > T_1$. Since it is only the lower boundary it does not necessarily represent the best estimate but a useful starting point if only one volatility is given but another needed (over a longer time frame but fully covering the first).

Another option is to use the local volatility. Gatheral (2006) shows that the implied volatility can then[7] be found based on stochastic volatility models and notes

> "So sometimes it's possible to fit the term structure of the at-the-money volatility with a stochastic volatility model, but it's never possible to fit the term structure of the volatility skew for short expirations. That's one reason why practitioners prefer local volatility models: a stochastic volatility model with time-homogeneous parameters

[7] The author uses a single example to show that it is possible to infer the Black-Scholes implied volatility from local volatilities using the Heston model.

> cannot fit market prices! Perhaps an extended stochastic volatility model with correlated jumps in stock price and volatility [...] might fit better. But how would traders choose their input parameters?" (Gatheral, 2006, p. 39-40)

A similar problem exists in the case of varying strike prices. Here too, a functional relationship is needed between a single volatility estimate and the volatilities to be used for pricing. The latter need to reflect the misspecification of the pricing model and are thus model-specific. In the case of the Black-Scholes model, non-normality is one such significant deviation and its implications are discussed by Hull (2008). This gives rise to the well-known volatility smile. Explicit modelling appears relatively rarely in the literature; Gatheral (2004) proposes one such model with four parameters in addition to a volatility forecast for a given time to expiry.

The solution discussed by Hull (2008) is to use an adjustment table. One would start with a single forecast and find the appropriate volatility for a given time to expiry and strike level. Formally, one needs:

$$\sigma_t(K,T) = f(\sigma_t, K, T) \tag{2.25}$$

The table of adjustment factors in Hull (2008) is one such way and may be more broadly applied once it is stated in terms of the discounted strike or forward price. In either case a general relationship is assumed to hold over time.

One possible solution is to infer the functional form from other instruments. This is exactly the previously discussed implied tree method. Here a local volatility is inferred, local as it is specific to a particular state or node in the tree $\sigma_L(S,t)$. Instead of inferring volatility from single estimate to the volatility surface, the surface is inferred from existing prices in related (vanilla) options and applied to the instruments to be priced (exotic options). This process is discussed in great detail by Gatheral (2006). The approach requires an independent security, however, as the implied volatility from one security cannot be used for pricing of the same security[8]. The local volatility approach and the related models are thus not particularly useful for the original research questions. Nonetheless, they illustrate

[8]It is possible to use the implied volatility but the result would be the same price as observed in the previous transaction adjusted only by Δ and the result of passage of time. If every market participant acted this way, no additional information would be reflected in the prices. Additional information needs to be added through the trading activity itself or through parameter updates by some participants, however.

the usefulness of such an approach for specific securities. It also demonstrates the fairly large number of parameters needed even in the case of analytical models.

Given that the Heston model parameters, in particular, move slowly over time (Gatheral, 2006), the model may in principle be useful even when the condition of independent securities is not met. A related approach is the combination of modelling the volatility smile and fitting it across expiration times using splines as discussed by Gatheral (2006).

Finally, some researchers attempted to model volatility through functional forms of varying complexity, in particular with attempts at introducing interaction terms. While no standard model has emerged so far, a frequently cited set of functions is the one by Dumas, Fleming, and Whaley (1996), and Peña, Rubio, and Serna (1999). The author introduced a number of functions including quadratic terms to model the curvature of the smile.

What is common to all these models is that the number of parameters is fairly large and the fitting procedure is not trivial. Gatheral's concern about the ability to find a fitted model is of particular significance as it highlights the importance of the problem as well as the practical limitations. In an environment such as this, non-parametric models may be more suitable as neither the functional form nor the parameters need to be determined. Instead, both elements are inferred from the data. This problem like the earlier ones may thus benefit from machine learning techniques.

2.4 Machine Learning

Machine learning techniques have become increasingly important tools in both research and development work in academia and the wider industry. Its foundations can be found in research into artificial intelligence, i.e. the replication and ultimately extension of intelligence with all its aspects at least similar to the ones of humans. The goal, which is still the long-term vision of many researchers in the area is to replicate at least the behaviour of human beings using computers. The goal quickly proved overly ambitious given hardware and theoretical limitations.

In recent years, machine learning has become a sub-discipline aimed at supporting decision makers or even replacing them entirely in specific industries and certain roles. To this end, knowledge representation and inference models have been developed in addition to processes and methodologies for their development and embedding in the business process and the organisation as a whole.

The focus of the thesis is exclusively on how option pricing decisions can be aided using such techniques, which modelling approaches are suitable, and how such models are built and fitted to improve results.

Common to all machine learning techniques is their empirical nature. Rather than making assumptions as is frequently done in academic research, and inferring particular models from them, machine learning focuses on observed data in various forms and the models are designed so that they maximise the use of existing evidence for future decisions.

The models can be classified (e.g. Vanstone and Hahn, 2010) by operating principle, e.g. expert systems, case-based reasoning, swarm intelligence models, ANNs, or meta-learners such as genetic algorithms. Alternatively, they can be organised by problem type, e.g. classification and time-series analysis. The former requires assigning distinct labels presented to the model, the latter are a special case of regression problems, not unlike those studied in statistics. Both problem types are of some significance in option pricing. One possible application of classification is the decision of whether to exercise an (American) option early. This would require a predictive classification model. Time-series analysis is the central topic of this thesis and will be discussed in more detail in this and the following sections.

A large number of learning models exist for regression problems (see, for example, Hastie, Tibshirani, and Friedman, 2009). Most, though not all, are also applicable to the special case of time-series analysis. Among the most frequently used in financial research are ANNs. This class of techniques has a number of benefits, while not unique are fairly rare, and will be discussed at the end of this section.

As the name implies Artificial Neural Networks are an attempt at replicating a mechanism observed in nature. They were inspired by the structure and functioning of the brain, including that of humans. In a substantially simplified view, the brain resembles a very large network with their nodes (neurons, i.e. cells) interconnected with each other through links or edges (synapses in the biological context). Through chemical and electrical processing, signals are passed from one node to another. The signalling results and is in turn dependent on the excitement level. Through the complex interaction of these signals, thought and behaviour emerge in the individual.

Replicating a human brain, apart from ethical and other philosophical issues, is currently not possible even at a basic technological level due to its size and

complexity, i.e. the number of nodes and the high degree of interconnectedness. Simpler networks can be built, however, and several models have been proposed in the literature. By far the most common, and in many ways simplest, is the single-layer perceptron (SLP).

It consists of an input layer, a hidden layer, and an output layer. The independent variables represent the input vector and each value is provided to the network at one of the input layer nodes, each node consequently represents an explanatory variable. The desired output is determined as the value of the output layer nodes. In the case of regression, only one such node exists, typically. The hidden layer represents internal states. As the name implies, the values are not typically available for inspection nor, as will be discussed further, can they be interpreted in any meaningful way.

As an extension multiple hidden layers are possible, the general case of the multi-layer perceptron (MLP). Regardless of the total number of layers, each node is connected to all nodes in the previous and all in the following layer. This implies the absence of loops and is not true for all other network types.

The perceptron is a supervised learning technique. In order to build a model for the set problem, a number of samples need to be given to the network in learning mode. The inputs and outputs are provided and the network is trained. During training the network is evaluated. The initial values are used and weights, which are attached to the edges using randomised initial values, are applied to them. Furthermore, an activation function, which is often of the sigmoidal type, is applied to the combined value. This in turn is passed to the next layer.

In the output layer, the identity function is typically used, i.e. the weighted average represents the output value. During learning, the resulting value is compared to the provided output, the target value, and an error is calculated. This error is then used to modify the weights in the opposite direction throughout the network. This process is repeated for each sample, and possibly for several epochs. The resulting weights reflect a non-linear combination of the inputs of arbitrary complexity. This is only limited by the architecture of the network (for a formal discussion see Hastie, Tibshirani, and Friedman, 2009).

Networks of this particular kind are also called feedforward-backpropagation networks since evaluation is done in only one direction (forward) and weights are modified by a traversing the network in reverse direction (backpropagation). Alternative algorithms with better convergence characteristics exist as well.

Alternative architectures allow for loops in the network design or apply less strict structural limitations on the architecture. A common deviation is to substitute the sigmoidal transfer function with a radial basis.

ANNs have a number of benefits but users also face a number of problems in developing models and applying them successfully in a practice. Their principal advantage is that they are 'universal approximators.' The models can fit any continuous function, including those with non-linear features. This finding was proven by Cybenko (1989) (see also Irie, 1988; White, 1988; Funahashi, 1989). The author limited the discussion to particular features, a limitation that Hornik, Stinchcombe, and White (1989) showed is not necessary. Instead, the universal approximator property is shown to result from the general architecture and training process. These proofs are highly significant as they show that any sufficiently large network will approximate a given function arbitrarily well. The limitations stem only from architecture choices and data quality. Furthermore, Hornik, Stinchcombe, and White (1990) showed that this does not only apply to the function itself but also to its derivatives. Given the practical importance of the option greeks, this insight explains some of the interest in the use of ANNs in financial research.

The approximation properties make ANNs excellent predictors given sufficient high-quality data. Despite these benefits, a number of disadvantages exist that are a consequence of the great flexibility of the technique. Firstly, the models derived from the data have no explanatory power. It is not possible to attribute the prediction quality to any particular attribute or set of attributes. In fact, it is not even clear if any particular attribute contributes anything to the final outcome, i.e. to the classification or regression output. Unlike standard regression models, which allow for statistical tests of the significance of individual parameters (variable weights), no such mechanism exists in the case of neural networks. The resulting model is therefore a black-box. The inability to understand the inner workings has traditionally contributed to significant scepticism and resulted in their use largely in environments of high competition and an absence of competing models. Over time, some methods have been developed for the removal of irrelevant inputs and increasing explanatory power, largely through rule extraction (e.g. Baesens et al., 2003, for financial applications).

Secondly, the proofs by the various authors only demonstrated that the network can approximate any particular function but no standard architecture or process for choosing an architecture exists. The choice of input variables, the number and

size of hidden layers, the choice of learning parameters, pre- and post-processing steps are still left to individual users with little guidance other than by reviewing successful (rarely failed) implementations in related areas of research or use. This is a particular problem for the choice of architecture and the learning process. A similar problem exists in regards to the chosen input variables but since this is domain-specific, the selection of variables needs to be made for any approach and is thus not specific to ANNs.

Significant problems arise especially from the sensitivity of the model weights to the initial weights. The nonconvexity of the error function results in multiple minima and thus in the risk of getting trapped in one of them during learning. One suggestion (e.g. Hastie, Tibshirani, and Friedman, 2009) is to train networks multiple times with different initial weights. This is a common approach in such situations. Vanstone (2005) does not solve the problem explicitly but combines the issue with the question of the size of the hidden layer. The author starts at $\lceil \sqrt{n} \rceil$ hidden nodes where n is the number of input variables, repeating the process with increasing hidden layer size and stopping when the utility fails to increase further. It is important to note that the author splits the error measure in two components, one for training the network and one for model selection. The stopping is subject to the latter constraint, which in the case of that particular research is the benefit of the network for trading.

The alternative and textbook approach is to be "guided by background knowledge and experimentation" (Hastie, Tibshirani, and Friedman, 2009) although the authors provide some basic advice of using between 5 and 100 nodes. Based on the work by Kolmogorov (1957, among others), which proved the existence of a universal approximator, a maximum of $2n + 1$ nodes are needed for the hidden layer (Vanstone, 2005).

2.5 Financial Applications of Machine Learning

2.5.1 Option Pricing

Overview

Following the significant developments in ANN research, various researchers started applying the technique to financial problems including option pricing. The extensive but not exhaustive review focuses on the series of innovations while

also reporting variable and parameter choices. A unified notation is used in the presentation where possible, which is based on the previous sections rather than using the symbols as they appear in the various publications. Dates are given in international format and network specifications are given as x–y–z with x input nodes, y hidden nodes, and z output nodes. Training methods, parameters, and even input variables are frequently missing in the literature but are included where available .

The review is organised as follows: Early studies are presented first in largely chronological order as these set the standard for the modelling and statistical evaluation of option pricing models outside the classical financial literature. This is followed by a discussion of various modelling approaches and a separate subsection for exotic options. Finally, Australian studies and research focused primarily on methodological advances is summarised. Research on option pricing with a significant focus on the underlying volatility measure is discussed in the section on volatility forecasting after this section.

Research into machine learning techniques and option pricing and to an even greater extent into volatility forecasting does not appear to follow a clear path. Consequently, most articles are combinations of methodological changes, insights into particular markets, and technological outcomes. They are organised here by their major contribution from the perspective of this thesis.

Early International Evidence

Among the earliest uses of ANNs are those by Malliaris and Salchenberger (1993a) (see also Malliaris and Salchenberger, 1993b). They find that neural networks can indeed be used for option pricing. They test this on S&P 100 index options, using daily observations between 1990–01–01 and 1990–06–30, the 3-month Treasury bill for the risk-free rate, and the ATM implied volatility. In addition to the usual five parameters of the pricing model, the lagged index value and the lagged option price are used. Networks with varying sizes are trained for five points in time with a minimum history of 30 days and a testing time frame of two weeks. Their rationale is that this allows the capturing of the volatility dynamics. Their results suggest that there are pricing biases for both approaches but that the ANN produces generally lower pricing errors. They also note that the results are sensitive to choices of model architecture and parameters having used a learning rate of 0.9 and momentum of 0.6 with a sigmoidal activation function. The au-

thors note that combinations and in particular the use of the ANN to a pricing function in a next step may be beneficial. An important point made is that they "would not expect to achieve results that are significantly different than those of Black-Scholes if many traders are using the Black-Scholes model and the market prices reflect their strategies."

Hutchinson, Lo, and Poggio (1994) apply three nonparametric pricing models, projection pursuit, the radial-basis function network, and the MLP network, to option pricing and hedging (see also Hutchinson, 1994). The MLP training is done in on-line, rather than batch mode, and gradient descent. As discussed before, price-homogeneity is an important property and by exploiting this property and assuming constant volatility and the risk-free rate, the authors reduce the functional form to only two parameters S_t/K and T to estimate c/K. Using R^2, the average hedging error (AHE), i.e. the present value of the absolute deviations of a replicating portfolio, and a combined measure of mean and variance analogous to the definition of the mean squared error of a predictor in statistics ν, they find that the techniques work well for synthetic data, where the MLP is chosen to have four hidden nodes. Furthermore, they tested the same approach using real data, S&P 500 index call options between 1987–01 and 1991-12, fitting the models on each of the first nine of ten sub-periods, and testing them in the subsequent one. Volatility is based on the 60-day historical estimate and the 3-month Treasury bill is used again for the risk-free rate. Here too, they find that the models outperform the Black-Scholes (BS) formula for pricing and hedging. An important concern identified is that additional predictors and statistical tests are needed for future research.

An early deviation from using call options was the working paper by Kelly (1994), who investigated the pricing of American put options on four major (by volume) US companies' common stock. The study period was fairly short with only 1369 observations between 1993–10–01 and 1994–04–13. Using a one-year historical volatility estimate (but reporting similar results for alternative timeframes) and the 3-month Treasury bill for the risk-free rate, the ANN explains 99.6 % of the variability and thus much higher than the competing CRR model. The author also demonstrates that the ANN can be used for hedging.

Barucci, Cherubini, and Landi (1996) expand on their earlier work (Barucci, Cherubini, and Landi, 1995) on the use of ANNs to approximate partial differential equations under the no-arbitrage conditions using the Galerkin technique (see original articles for details). The demonstrate how to allow for stochastic

volatility and in particular for the modelling of the volatility smile. They also note, based on analytical considerations, that "the approximating solution […] can be looked at as a NN with one hidden layer augmented by a linear term, i.e. a direct input output connection." This is similar to what was observed earlier empirically by Malliaris and Salchenberger (1993b). Their approximation is $c(S,T) = S - \sum_{i=0}^{N+1} w_i(T)\Phi_i(S)$, the latter term representing the weighted trial functions introduced by the authors. The article further demonstrates the approach using an example "inspired to the S&P 500 options market" but the authors stress that no attempt at learning the critical parameters for stochastic volatility from market data had been made.

Herrmann and Narr (1997) further explore the non-parametric pricing models in the German market and using intraday data. They use put and call options on the Deutscher Aktienindex (DAX) (European-style German stock index options traded at the Deutsche Termin Börse) between 1995–01–01 and 1995–12–31 (with certain observations excluded, see the paper for details). The Frankfurt Interbank Offer rates were used for the risk-free rate and the implied volatility index for the DAX (VDAX). Like Hutchinson, Lo, and Poggio (1994), the authors used a synthetic data set but only to determine the suitability and MLP network complexity required. They confirm prior research that ANNs are able to price options better (more accurately) than the BS formula and are able to implicitly model the derivatives with some notable deviations from the closed form with regard to P and V, they speculate that this may be due to correlations between the variables.

Qi and Maddala (1996) extend earlier work on pricing S&P 500 index options by including open interest as an input variable. They use only a short period of time (1994–12–01 to 1995–01–19) of daily data, the three month treasury bill as a proxy for the risk-free rate and the volatility over the past 106 days. Instead of a simple split, they apply five-fold cross-validation. They train feedforward networks with a single hidden layer using five input variables (excluding volatility but including open interest for network training) by first experimenting with various network sizes concluding five nodes to be appropriate. The data is first normalised, the learning rate set to 0.05 and momentum to 0.95. Learning is stopped after 15 000 epochs with additional constraints to weight updates. They report small improvements over the BS formula. By analysing the actual weights in the network, the authors argue that the network reflects the typical economic relationships and that open interest is an "important factor" in pricing. They also

note that R^2 is a better performance measure as it is not affected by the variance of the target variable.

Lachtermacher and Rodrigues Gaspar (1996) studied the uses of ANNs in the Brazilian market. Sampling TELEBRÁS data between 1994–10–04 and 1994–11–30, the underlying price data, volatility of the stock returns over the 22 previous periods. The "futures market interest rate" was used as a proxy for the risk-free rate. Like Qi and Maddala (1996), the authors used cross-validation for training a network with two hidden nodes in a single layer, using a learning rate of 0.1 and momentum of 0.9. The size of the hidden layer was increased twice by one node each time to determine the appropriate network size and find that three hidden nodes results in the best overall fit.

White (2000b) (also White, 2000a, from where some results are reported and which is thus not discussed separately here) models the option pricing function using Genetic Adaptive Neural Networks and adopts a methodology to allow for learning simulated as well as observed data. The author focuses on a particular feature of the LIFFE, which similar to the SFE and now the ASX uses futures-style margining for options. The options are Eurodollar futures options between 1994–01–04 and 1994–07–29. The network uses the future rate, strike, maturity and volatility (60-day historical estimate) as inputs and predicts either the call or put price in any network. Rather than using traditional back-propagation, a genetically-optimised network is trained. Based on the mean squared error (MSE), the best network (for real data) is one with 18 hidden nodes after significant fitting of the data (no improvement in sum of squared errors and additional 30 000 generations). The performance was measured against market rates and compared with the adjusted model by Chen and Scott (1993). The author points out that the use of historical volatility represents a "naïve" model comparison but also points to evidence by White, Hatfield, and Dorsey (1998) that still shows significantly improved performance. Like other studies, this one attempts to overcome some of the disadvantages of the black-box nature of ANNs, here by studying the sensitivity to changes in variables (essentially testing for the sign and of the greeks and the curvature in some cases). Finally, the author tests if additional inputs, moneyness and the 3-month Eurodollar interest rate, can contribute to pricing performance, which can be supported.[9] Regarding the use of the risk-free rate the author concludes that "[a]t the time of this writing, very few studies have

[9]The author retains the network complexity, i.e. the number of hidden nodes, despite increasing the potential model complexity for comparison reasons.

addressed this type of option [,with future-style margining, and that t]he studies that have pertained to this issue disagree as to whether or not a risk-free rate should be included in the option pricing model." The author points out that the agreement is that if payment is required by both, the rate should be dropped.

Another study on LIFFE-traded options is presented by Raberto et al. (2000). They introduce as input $\frac{|S-K|}{T}$ to model the smile in addition to the usual $\frac{S}{K}$ and T showing graphically a good match to observed prices after training the estimator on synthetic data.

With assumptions similar to those in the original article by Hutchinson, Lo, and Poggio (1994), Yao, Li, and Tan (2000) limit input variables to moneyness, time-to-expiry and the risk-free rate. Thus volatility is strictly implied and assumed constant for the training, validation, and test set. Various network sizes are tested starting at half the number of inputs with a step-wise increment of 1. Using Nikkei 225 index call options trading at the Singapore International Monetary Exchange between 1995–01–04 and 1995–12–29, they find as did previous researchers that the ANNs can outperform the BS formula in particular for non-ATM options. They also find that time-indexing improves performance (with respect to normalised mean squared error (NMSE)) but suggest that volatility does not need to be modelled separately.

Healy et al. (2002) study LIFFE FTSE 100 index futures options between 1992 and 1997. They deliberately do not assume homogeneity and find that the bid-ask spread as a proxy for transaction cost and open interest have explanatory power while volume does not. In particular, the bid-ask spread has greater explanatory power than the risk-free rate, using the 90-day Libor rate. They use implied volatility and eleven hidden nodes for the network but despite the use of market rates and volatility conclude that performance is not constant over time thus raising the question of how long a model derived from market prices is useful.

Adding to the international evidence, Amilon (2003) tests the usefulness of ANNs in the Swedish stock index option market, focusing on calls between 1997-06 and 1999-03 excluding April and May each year due to concerns regarding dividend-adjustments made to the index during that period. The risk-free rate is the 90-day local treasury bill. In addition to the usual parameters, time-to-expiry is measured in trading and calendar dates separately. As mentioned above, various adjustments to closed-form pricing models exist, including corrections for differing day-counts between the risky asset and the risk-free one. The adjustment is, however, made to the rate itself rather than an additional parameter

directly. Homogeneity is assumed and the bid and ask prices are estimated separately. Furthermore, five lagged index values are used as inputs instead of only the current one as well as two historical volatility estimates based on 10 and 30 past returns respectively. The hyperbolic tangent is used for the hidden layer activation functions, and the logistic function for the output layer. Another unusual choice is regarding the partitioning. For each year the first four months are used for training, the following two for validation and one more month for testing, this is repeated by moving forward by a month but keeping the starting point of the training period the same thus expanding the training set. This ultimately yields non-overlapping testing sets, however.

Carverhill and Cheuk (2003) studies S&P 500 index futures options (calls and puts). Departing from prior studies, homogeneity is assumed but the ratio reversed (K/S), the risk-free rate interpolated using market data (Libor and Eurodollar futures) and several networks trained using weighted average past data with sampling at weekly frequency to avoid day-of-week effects. The modelling is approached by separating the pricing from the estimation of the derivatives, the greeks. Three hidden nodes are used in either case but the network estimating the greeks, has two outputs, Δ and V. The results show improvements with respect to the volatility in hedging, i.e. better hedging, than the reference model (CRR).

Modelling Stochastic Processes and Risk-Neutral Pricing

Given the strong foundation of derivatives pricing in stochastic calculus, several authors have attempted to gain insights by combining these areas.

According to Schittenkopf and Dorffner (2001), for example, the use of mixture density networks, which approximate the conditional density function as a mixture of Gaussians is a viable alternative to the standard pricing function.[10] Their parameters are estimated using MLPs with a single input, the time to expiry T. The functional form used in the transformation of MLP to density ensures both positive priors and non-negative variances. This constraint is particularly useful given that nonparametric approaches can often result in any output, even those without meaning in the domain being studied. The authors test two network specifications of varying complexity using FTSE 100 data (from the United Kingdom). They use intraday transactional observations from 1993–01–04 to 1997–10–22 but restrict it to near-the-money options with ordinary volatility levels not expiring

[10] This approach is suitable in the risk-neutral valuation framework.

in the near future. Their results show that the networks are better than both the traditional Black-Scholes formula as well as one adjusted for higher moments (especially skewness and kurtosis). This holds for pricing and hedging errors.

A similar approach is taken by Healy et al. (2007). Studying LIFFE FTSE 100 put options (European and American), the authors infer the risk-neutral densities. Using a 5–11–1 architecture option prices were estimated permitting differentiation to arrive at the densities. The results are found to be unbiased with respect to the price at expiration but biased with respect to realised volatility.

Extended ANN Models and Support Vector Machines

Zapart (2002) approaches the problem not by estimating the risk-neutral densities but rather the evolution of prices. Using wavelets and a neural network, the author models local volatility over time, which is then used in a binomial tree. Studying the options on three United States of America (US) companies, hedging risk is determined and shown to be lower for the combined wavelet-ANN model (five hidden nodes) than the BS formula between 2001–08 and 2002–01. Zapart (2003a) extends this data set to 2002–07 and 53 companies again in the US demonstrating how such a network can be used for trading, in this particular case to creating delta-hedged portfolios with a single threshold parameter. Finally, Zapart (2003b) extends the original approach (Zapart, 2002) by using genetic algorithms to find a suitable network, a particularly time-consuming task considering that each time series is to be modelled by a separate network (not unlike GARCH volatility modelling).

Montagna et al. (2003) use a synthetic data set of European and American options. The authors study a path-integral algorithm for pricing first and apply radial basis function (RBF) networks to the prices derived using this algorithm. The authors find that both approaches can be very useful for pricing options and that the ANN results depend "as expected" on the choice of the spread (training) parameter and the number of observations used for training. They generally find a very low deviation of the ANN-based results from the reference model. Morelli et al. (2004) extended this research focusing on the differences between the RBF and MLP models with respect to pricing and the estimation of option greeks. They find that the RBF approach is much faster and thus more suitable for "preliminary checks" while the MLP performs better in the long run but is more

expensive to develop (with respect to training time). As in the previous article, a synthetic data set was used and the error terms used for comparison.

Bennell and Sutcliffe (2004) follow the standard approach of focusing on FTSE 100 European call options traded at the LIFFE between 1998–01–01 and 1999–03–31. They compared a MLP-based model and the reference BSM formula. Unique to the research is the use of the dividend yield, for which they retrieved the historical yield from DataStream as a proxy for the forward-looking annualised dividend yield. The authors used the usual option pricing inputs as well as open interest and daily volume, both of which were eliminated as candidate inputs in the process, and the homogeneity hint, whose impact was studied in various ways. The authors also experimented with a number of different network sizes. The data was partitioned such that the first quarter of 1999 constituted the test set, the earlier data used for training with one third used for cross-validation. An extensive set of performance statistics was used comprising R^2, mean error (ME), mean absolute error (MAE), mean percentage error (MPE), and MSE. Similar to previous research (including that discussed in the subsequent section), the authors find mixed results. The BSM model is generally better while the MLP performs comparably in the restricted region, i.e. when limiting the set to near-ATM options expiring within just under a year. On balance, the authors conclude that the ANN solution is superior to the BSM one considering that the restricted set does not impose a significant constraint with respect to the number of observations excluded.

While most authors consider improvements of pricing accuracy the goal, Choi et al. (2004) focus on the actual learning algorithm. The authors compare five different learning algorithms for ANNs to their own algorithm, which approaches training as a minimisation problem using a regularisation parameter ("Globally Regularized Neural Network"). Using the Korean Stock Exchange's KOSPI European-style index options for the year 2000, the authors find large improvements with respect to generalisation (i.e. lower test errors) and faster training.

Further research into various training methods and models was published by Dindar and Marwala (2004), and Dindar (2004). The authors compared MLP and RBF networks as well as a number of committees of networks and particle-swarm optimisation. Using data from the "South African Foreign Exchange"[11] covering

[11]The references to this exchange may be wrong. While it could not be verified fully, it is likely that the South African Futures Exchange was meant. This is supported by subsequent research with a common co-author as well as the links provided in the references section of the thesis and the articles.

the years 2001–2003. The data consisted of the usual stock and option data as well as the "market to market," high and low prices. The authors applied cross-validation though it is not clear if a separate out-of-sample set was used or what option terms were, and whether the volatility parameter was implied volatility. Focusing on the technical aspects the authors found that a circular committee approach leads to the best performance with respect to the error metric while the best RBF and MLP networks performed relatively well, too. Importantly, they note that the performance of the MLPs tends to be more stable across models. Given the difficulties of finding a well-performing network, this insight is particularly relevant to ANN-related research in option pricing as local minima and highly sensitive results are generally of little value.

Using the same data set and limiting investigation to the futures and options of the All Share Index Pires (2005) (see also Pires and Marwala, 2004; Pires and Marwala, 2005) compared Baysian MLPs and support vector machines (SVMs), and comparing the Baysian approach to the maximum likelihood method. Based on the ME and the maximum error, the author concludes that the SVMs outperform the MLPs regardless of the method and the differences can be very large.

Promising results were also reported by Hamid and Habib (2005) using S&P 500 index options between March 1983 and June 1995. Since the underlying were index futures, the Black (1976) model was used as a reference point and to extract implied volatilities. The authors compare prices and implied volatilities for the nearest call option at specified expiry dates with respect to the reference model and report t-test, MAE, and root (of) mean squared error (RMSE) for the different models. Unlike other studies, they find that ANN has some difficulties at very short maturities. As they do not provide the futures volatility as an input they speculate that the data may be too noisy to infer the function in that region for this data set. Note that many studies that do exclude data because of fitting problems, do so at the very long maturities and typically find a good fit for shorter periods.

Another comparison of models is given by Liang, Zhang, and Yang (2006). The authors limit their investigation to four stocks of the Hong Kong market during March through July 2005. The authors propose a hybrid model, which does not use the usual option pricing inputs but rather four different pricing functions (binomial model, the BS formula, the finite difference, and the Monte-Carlo model). The actual input is the difference between those functions and the mean pricing result on a particular day. The resulting hybrid model is therefore an ensemble

(committee) estimator of the constituent models. The number of hidden neurons of the resulting neural network has some impact on the results. The sliding window, which determines which observations are used for training for the purpose of pricing for a particular day is considered an important parameter choice. The model is thus not fixed for the period but is modified over it. This is in contrast to many, if not all, previous research, which limits training to a single period or relatively few periods. Due to the large number of combinations of parameters, no details are provided here. Liang, Zhang, and Li (2009) extend this work using related but different data to forecast option prices. To this end they add a sliding window of past prices as input and modify the BS, finite difference and Monte-Carlo methods, which are used in combination with a linear neural network. They report lower forecast errors for the MLP relative to the linear neural networks. Among the outcomes is that the length of the sliding window has a large impact on the linear model but less so on the non-linear ones.

Liang et al. (2009) further extend the earlier work using a larger data set from 2006–01–01 to 2007–12–31. As in the earlier study (Liang, Zhang, and Yang, 2006), a hybrid model is developed based on the three statistical methods. In addition to the combination of the three methods, an MLP, and an SVM (for regression rather than classification) are applied. Again several sliding window lengths are used. The authors conclude that forecasting can be improved and that the order (from poorest to best) model is: linear neural network, MLP, and SVM (with differences being relatively small). The relative (to the previous day), mispricing error is used as the metric.

A similar rolling method is applied by Mitra (2006), confirming conclusions reached in other markets, including several Asian markets, for the Indian stock market. The study covers the call options on the Nifty (National Stock Exchange) between 2004–05–28 and 2005–06–30 and uses a 60-day historical volatility as a proxy for future volatility. The results using a network with two hidden nodes show a reduction of the error term by about 50 % in-sample and out-of-sample and across error metrics (sum of squared errors (SSE), RMSE, total absolute error, MAE).

This research is extended and confirmed by Mitra (2012) using data from 2008–07–01 to 2011–06–30. A paired t-test comparing the BS model and the ANN model shows them to be significantly different at the 1 %-level.

The dynamic complexity of the option pricing problem is well understood and the sliding window models are only one way of addressing them. Teddy, Lai, and

Quek (2006) (for more detail see also Teddy, Lai, and Quek, 2008) proposes an alternative model inspired – as the ANN itself – by the human brain. The proposed model uses the weighted Gaussian neighbourhood output for activation. The new model fails to outperform a 3–8–1 MLP model using the RMSE of the mispricing of British Pound–US Dollar foreign exchange futures call options traded on the Chicago Mercentile Exchange (CME). The authors stress, however, that the new model allows for the extraction of discrete pricing rules, which they claim is very difficult for traditional MLPs. Furthermore, they show how the model (without comparison to alternatives) can produce risk-free portfolios from trading system using mispriced options.

Kakati (2008) makes a similar point regarding an alternative model that has at least some explanatory power. In this article an adaptive neuro-fuzzy system (commonly abbreviated ANFIS) is proposed and compared to the BS model. In line with the existing body of knowledge, the model outperforms the traditional pricing model but also a simple ANN. This is demonstrated using 40 American-style call options on seven Indian stocks. Volatility is estimated using GARCH models, 60-day historical volatility as well as implied volatility. The author also adopts the homogeneity hint.

The results of the study are interesting in several ways. Firstly, the ANN is only rarely the best model, which is surprising given past evidence, and even the BS model outperforms when used with implied volatility. Given how implied volatility is derived, this may not be surprising. However, the ANFIS approach generally outperforms the competing models under all volatility measures.

Quek, Pasquier, and Kumar (2008) also consider an alternative network structure, RBF using Monte-Carlo Evaluation Selection to determine the input set. In addition to the usual steps, the authors also use the predicted values to compute a theoretical portfolio returns series from a trading model (delta hedging). Of the input variables tested, open, close, high, low and the previous two open prices are reported to be relevant. Two instruments are investigated, gold and the British Pound-US Dollar futures and options, covering the period of 2000–2002 (gold) and 2002–10 to 2003–06 (currency, respectively). Unusually for option pricing, the authors develop the networks to make a directional and price prediction forecast using an MLP, an Elman recurrent network, a special fuzzy-neural network, and the innovation model. The latter is a modified recurrent network with a modified learning algorithm. The Elman network performed better than the MLP; the fuzzy-neural network had generally poor prediction ability. The methods were

chosen due to the limited amount of data and the authors conclude that their proposed model outperforms the MLP and the Elman network leading to an accuracy of up to 90 %. This work is further extended by Tung and Quek (2011), who introduce an evolving fuzzy rule set and related trading system for trading volatility based on the novel approach.

Hybrid Models

Andreou, Charalambous, and Martzoukos (2002) compare standard ANNs to hybrid versions, which estimate not the option price but the difference between observed price and the BSM model each under the assumption of homogeneity with respect to the strike price. They determine the optimal network configurations and input parameters by simulation of rolling (and overlapping) four-month training and six-month testing periods. They find that the non-parametric models outperform the BSM model for S&P 500 European call options between 1998–05 and 2000-12 with respect to the median absolute error (MdAE). Unlike other researchers, they use interpolated risk-free rate, an adjusted moneyness measure, and three measures of volatility, 30- and 60-day historical volatility as well as the Chicago Board Options Exchange Volatility Index (VIX) for each model.

Extending this research Andreou, Charalambous, and Martzoukos (2006) estimate the networks using the Huber function, training and testing a large number of networks over a wide range of parameters of the Huber function. They find that for the period between 1998–04 and 2001–08, the use of the Huber function instead of least squares during training is beneficial. In addition to modifying the training process, various alternative volatility models are considered, principally focusing on implied volatility. The least-squares-based networks underperform the parametric ones, however, unless the Huber-function is used. The hybrid networks outperform even under the common least-squares procedure.

In later research, the authors return to the least-squares estimation procedure, however (Andreou, Charalambous, and Martzoukos, 2008). They extend the data set back to 1998–01 and create ten rolling sets of data for training, validation, and testing. Using the 60-day and implied volatility index, and dividend-adjusted moneyness, they compare the hybrid networks with the traditional BS and the model by Corrado and Su (1996). Andreou, Charalambous, and Martzoukos (2008) find that hybrid BS-ANN performs best when contract-specific param-

eters are given. Finally, they conclude that hedging profits depend on the level of transaction-cost and suggest that some inefficiencies may still exist in the market.

Andreou, Charalambous, and Martzoukos (2010) extend prior work on hybrid networks and the use of the deterministic volatility function (DVF) for reference models to price European call options on the S&P 500 index between 2002–01 and 2004–08. Splitting the data into twelve months training, two for validation, and one month for testing, with rolling partitions but non-overlapping test sets, allows for continuous re-estimation of parameters (both the parametric reference models and the non-parametric innovation models). The ANNs are further modified by adding a parametric pricing function as a fourth layer to a standard single-hidden-layer perceptron so that "the network structure embeds knowledge from the parametric model during estimation (thus resulting in a semi-parametric option pricing method)." The resulting models, which use the dividend yield and in the case of ANNs a dividend-adjusted moneyness measure, are shown to be comparable to models of stochastic volatility with jumps with respect to the chosen error measure (notably RMSE but also reporting MAE and MdAE). They do not generally and unconditionally outperform those but do outperform the conventional pricing formulas (BS and CS). The authors re-iterate the importance of choosing functions for hedging based on their hedging results and note merely their pricing performance, i.e. the performance metric needs to be aligned with the intended use of the model. The specific enhancements within the model are with respect to volatility, skewness, and kurtosis.

Departing from prior research into ANN use, Andreou, Charalambous, and Martzoukos (2009) use SVM regression for option pricing and find positive results in the case of hybrid models, which use not average volatility but volatility from a DVF. The data used consists of the typical observation with an interpolated risk-free rate and daily option prices (using bid-ask midpoints) between 2003–02 to 2004–08. The use of a DVF, however, is important since it allows for contract-specific volatility estimates.

Charalambous and Martzoukos (2005) study the applicability of ANNs to two option pricing problems using a hybrid approach similar to the one proposed in the methodology section of this thesis. The authors use synthetic data sets for financial and real options and build a simple 20-hidden-nodes ANN, as well as a hybrid network using the numerical option pricing model as the baseline model while one fits the difference. Studying the MSE, MAE, and the maximum absolute error of the networks as well as the numerical pricing model as a reference

point, they conclude that the hybrid approach is superior, especially for the very large network and is slightly better even for a very small 2-hidden-nodes network compared to the reference model.

A similar approach is taken by Blynski and Faseruk (2006). Again the hybrid model as the difference between the target value and the baseline model is developed. The authors use a wider range of performance metrics, however, R^2, ME, MAE, mean absolute percentage error (MAPE), NMSE, and MSE. They split the Chicago Board of Exchange OEX index call options data between 1986 and June 1993 into 60 % training, 20 % validation, and 20 % testing data and removed non-representative data. As previous authors, including the research discussed in the next section, they find some support for the use of ANNs. A two-stage hybrid model, which estimates implied volatility and applies Lajbcygier's hybrid approach discussed below, fails to improve the outcome. At least for this market, the researchers conclude that the market appears to be efficient with respect to pricing such options. In particular they find that using the implied volatility, the BS model is fairly good by comparison.[12]

A similar model is developed by Amornwattana, Enke, and Dagli (2007). Two ANNs are trained one estimating the implied volatility, the second pricing the difference between the observed price and the reference BS model. For comparison, a model using historical volatility instead of the first network is used, which has a window size of 90 days. The model is tested on five stocks that are members of the DJIA from 2002–07–01 to 2002–10–15, where the October data was used for out-of-sample evaluation. Unusual for option pricing evaluation, the authors calculated various non-parametric tests and scores using the MAE and MSE series: the Wilcoxon test, median, Van der Waerden and Savage scores. These demonstrate that the lower errors (compared to the reference model using BS and historic volatility (HV), and the network learning the difference between the observation and the BS value) are also statistically significant in many cases.

Saxena (2008) follows the standard methodology for hybrid models, using the ANN to learn the additional price relative to the baseline BS model. They test this on the S&P CNX Nifty index (India) options between 2005–11–01 and 2007–01–25, using a one year historical volatility and exploiting price-homogeneity. Following the splitting of data into 40 % for training, 30 % for validation, and the remainder for testing, the usual statistics R^2, ME, MAE, MPE,

[12] The authors do not discuss methodological issues arising from the use of implied volatility derived from the same model to which it is applied in a later step, however.

and MSE are calculated and the hybrid model outperforms the baseline model with respect to each of them.

A different approach to hybrid models is the one proposed by Ko (2009). Instead of combining the machine learning technique with an underlying pricing function, a (neural) regression model is developed, which combines various input networks in a single functional form similar to committee models. The input variables are the BS inputs and the experiments were conducted using index future options from the Taiwan Futures Exchange. The options are European-style and data was collected for the period 2005–01–03 to 2006–12–31, of which 80 % was used for training, the remainder for testing. A learning rate of 0.1 and two hidden layers with six neurons each, was used for each network. The study concludes that the model is superior to the BS model with respect to the average absolute delta-hedging error.

Not all evidence is supportive of the use of ANN. Gradojevic and Kukolj (2011), for example, concludes that a parametric model based on fuzzy rule-based system is no worse than a non-parametric feedforward network using European-style S&P 500 index call options for the period of 1987 to 1993 (by expiry date). No model can be said to be better than the other using the Diebold-Mariano test statistic. This is despite using a network with a hint similar to prior research, which typically found hints to improve results and lead to improved performance.

Applications to Exotic Options Pricing and Non-Equity Securities

Extending the use of ANNs to more complex options, Lu and Ohta (2003a) and Lu and Ohta (2003b) generate a synthetic data set based on a number of assumptions informed by NYSE options. The authors study the applicability of machine learning to the pricing of complex power and rainbow options and modify the standard approach of training the networks using the contract inputs and volatility parameters by supplementing the input data set with a pricing hint based on digital contracts, i.e. pricing based on the binomial model. They show using Monte-Carlo simulations that the hinting improves pricing performance measured as the RMSE and stress that the approach can be generalised to a variety of complex options.

Using LIFFE data as well, Xu et al. (2004) generate a synthetic data set of barrier option data points based on the European-style actual observations. The authors build two different models, who differ in the exclusion vs. inclusion of the

trading date as an input, which did not appear to lead to better results. They find based on the R^2 and paired t-tests that ANNs can be a valuable tool for option pricing and that no additional steps had to be taken to address the barrier feature of the options.

Another application outside the equity (index) options is presented by Leung, Chen, and Mancha (2009). The authors compare MLP and general regression neural network as well as a number of projection models and the BS formula. These are applied to foreign exchange futurees options for the British Pound, the Canadian Dollar and the Japanese Yen as traded on the CME covering the period of 1990–01 to 2002–12 (in sample) and 2003–01 to 2006–12 (out-of-sample). Subject to transaction cost, the general regression model is generally best. This is also supported by pairwise comparisons of trading returns using the bootstrap method. Furthermore, an ensemble (i.e. a committee) model combining the two machine learning techniques shows a lower coefficient of variation of returns than each model individually, which is not surprising givne past statistical literature. This is, however, economically important as variability is risk in this context.

Chen and Sutcliffe (2012) (see also Chen and Sutcliffe, 2011) investigate the use of hybrid ANNs and compare it to the model by Black (1976). The data set covers NYSE LIFFE tick data for short sterling (British Pound) call and put options and the relevant futures for the period of 2005–01–04 and 2006–12–29. The authors test a simple model pricing the option and one pricing the difference between the reference model and the observation separately. In addition, two different strategies of calculating the hedging requirements are proposed, calculated based on the priced option, or as the target variable of either of the networks. Paired t-tests are used to compare the pricing differences. Based on the MSE, MAE, ME of pricing and hedging, the authors conclude that ANNs can be used successfully for interest rate options with the hybrid model strictly better, and the simple network better or no worse than the traditional model.

Evidence for Australian Index Futures Options

Lajbcygier et al. (1996) focused on the use of ANNs for pricing American-style index futures options (SPI futures options) in the Australian market. Using end-of-day option prices between January 1993 and December 1994, the latest prior intraday underlying price, the 90-day bank bill as a proxy for the risk-free rate and a volatility estimate similar to Hutchinson, Lo, and Poggio (1994). In re-

gards to the latter, the authors point out that the original article (Hutchinson, Lo, and Poggio, 1994) contained an error, which, according to the authors, was acknowledged in private communication but was not present in the actual implementation. The resulting networks assume price-homogeneity and are evaluated using R^2, normalised root mean squared error (NRMSE), and MAPE. Similar to Hutchinson, Lo, and Poggio (1994), the authors (dealing with only one underlying) use two configurations, one with a reduced set of input variables (moneyness and time to expiry), the other with the full specification of all four parameters. The data set was further split into a training and test set (20 %), with the test set formed randomly from all observations. The authors conclude that the neural networks can be beneficial in at least some cases, notably in the reduced region (10 % around ATM and for expiry no longer than one fifth of a year) but they do not show substantial improvements over the benchmark methods, the Black-Scholes and the Barone-Adesi/Whaley formulas. They also note the surprising finding that the Black-Scholes model fits the Australian data extremely well when compared to the US case (Hutchinson, Lo, and Poggio, 1994) especially considering that it is not directly applicable to American-style options.[13]

Lajbcygier and Connor (1997a) concentrate on small data set of intraday SPI options in 1993 (see also Lajbcygier and Connor, 1997b). Using the first six months for training with again 20 % of data reserved for cross-validation, the authors train a 3–15–1 network similar to the networks before with two notable innovations: Following their prior research (Lajbcygier and Flitman, 1996) they train not a direct pricing network but hybrid model, which generates the price as the difference between the modified BS formula and a trained ANN model. In doing so they implement the suggestion by Malliaris and Salchenberger (1993a) and a functional form common in forecasting problems (e.g. volatility forecasting as pointed out by Poon and Granger, 2003) and not dissimilar from the one used by Barucci, Cherubini, and Landi (1996). Furthermore, they use a weighted implied volatility (IV) scheme to derive a volatility estimate acknowledging problems with this approach in general terms (see the article for details or Poon and Granger, 2003, for a more general discussion). Finally, they use bootstrapping to infer confidence intervals, and bootstrapping and bagging to address model bias.[14] In doing so they address earlier concerns by researchers given that the error terms

[13] Note that early exercise is not meaningful for non-dividend paying securities in any case as per the earlier discussion.

[14] A detailed discussion of the theoretical foundations and implementation notes regarding the two techniques can be found in Hastie, Tibshirani, and Friedman (2009).

of such pricing models often do not meet the assumptions of standard statistical tests. The objective is to infer points of mispricing and thus profitable trades. They conclude that bootstrap bias reduction is superior to bagging in the case of option pricing and that the error appear greater near the boundaries, notably at the very short maturities.

In a later paper (Lajbcygier, 2003a), the author re-examines the same data set as in Lajbcygier et al. (1996) adding statistical tests to determine which models are better than others using dependent t-tests on the test set, which again is a random 20 % of the data in both the reduced and the full region. The independent t-tests are used to compare the distribution (not the pair-wise comparison) of the absolute errors between the reduced and full region in the test set to determine any differences between them. Their statistical findings support the earlier evidence that ANN-based models lead to significantly improved pricing in the economically important (restricted) region. In the full region only the largest network appears competitive. Furthermore, the networks are not substantially different from one another. The authors highlight again the surprisingly good fit of the Black-Scholes formula compared to the Barone-Adesi/Whaley model suggesting that the argument by Lieu (1990) holds in Australia.

Finally, Lajbcygier (2004) attempts to address pricing biases found in earlier research (Lajbcygier and Connor, 1997a) near the boundary (see also Lajbcygier, 2003b). Some of those were discussed earlier in section 2.2.5. In particular, Lajbcygier (2004) modifies the learning algorithm such that the hybrid model (Lajbcygier and Connor, 1997a) yields meaningful results as one approaches expiry $T = 0$, moneyness is close to 0, or volatility (again using the weighted implied standard deviation) approaches 0. In all three cases, the neural network should return 0 (recall that it represents a penalty over the closed-form modified Black model). For the three years studied, 1993–1995, the authors find statistically significant and much improved behaviour of the pricing function near the boundaries as a result from the constraints. The author suggests further research studying the use of bootstrap methods in combination with constraints, the combination with alternative models, such as SV, and exogenous variables.

Methodological Advances

Anders, Korn, and Schmitt (1998) demonstrate how the network architecture can be derived based on a two-step process of slowly increasing the number of

nodes and then decreasing the connections based on statistical tests. They use the DAX index call options (European-style) between of 1994 with all transactions between 11:00 and 11:30 using minutely data. Furthermore, unusual observations (violating boundary conditions, or situations near the boundary) were excluded. The risk-free rate is derived by interpolation of interbank lending rates at various maturities. Two volatility estimates, 30-day HV, and IV (using the VDAX), are investigated. The resulting networks (without closed-form model as a hint), are not fully connected and have three hidden nodes. Once the BS result is provided to the output node, the network collapses further leaving only moneyness and time-to-expiry as additional variables.[15] The authors conclude that the statistical techniques used for configuration can help with the pricing accuracy preserving the usual relationships between variables as suggested by theoretical considerations (option greeks). Notably, the index level has predictive power when combined with historical volatility.

A related problem is the dynamic structure of the network. Instead of determining the network structure at the time of development, the structure is fixed but the parameters are allowed to change over time (non-stationarity). Ormoneit (1999) shows that Kalman filters can be used to influence the weight updates in ANNs and applies this to option pricing. The use of Kalman filters is not new; Niranjan (1996), for example, demonstrated how they can be used to model volatility and the risk-free rate. In contrast, Ormoneit allows for continuous updates of network weights but imposes a penalty (regularisation) using Kalman filters. The strategy is tested using DAX index call options between 1997–03 and 1997–12, all of which expiring at the end of the period. A constant risk-free rate of $ln(1.05)$ is assumed and various fixed, previous implied, or historical volatility of the futures contract are used in combination with the reference model, the Black-Scholes formula. They find that the networks perform very well with respect to the hedging error but not as well for pricing. The problem appeared to be near expiry and the author attributes it to limited complexity of the model to accommodate the boundary condition. Like Lajbcygier (2004), Ormoneit modifies the network but in this case by switching to an activation function and related evaluation along the network path that are consistent with the risk-neutral pricing approach and reflect the boundary condition. This results in improved pricing and hedging.

[15] Note that this is identical to the standard specification of the volatility surface.

In another study regarding network design and its relationship to the option pricing problem Galindo-Flores (1999) builds 13 920 candidates to determine the characteristics of several machine learning techniques, classification and regression trees (CART) decision trees, feedforward ANN, k-nearest neighbour, and ordinary least squares (OLS) regression, and their parameter choices. The author suggests that it is difficult to decide a priori, which technique should be used and how the models are to be constructed (i.e. parameter choices). In particular, the neural networks used for option pricing use three input and one output node (as in Hutchinson, Lo, and Poggio, 1994) with six and 18 hidden nodes in two models generated for comparison. Focusing on structural risk (Vapnik, 1995) and using a synthetic pricing data set under the usual assumptions of constant interest rate and volatility, it is found that ANNs offer the best performance especially when the sample size is larger, OLS being preferable when the sample size is restricted. The ANN-specific finding is that the modified Newton method in combination with the larger network (3–18–1), and a large number of iterations (450) is preferred. The author cautions against generalisation, however, as the data set is simplified, 'less nonlinear.'

Garcia and Gençay (2000) show how to exploit one of the assumptions normally made in option pricing, and in pricing securities generally. As discussed before, the pricing function is considered homogeneous with respect to the price and strike of the underlying.[16] Using European-style S&P 500 index call options (and initially a synthetic data set) from 1987–01 to 1994–10, they demonstrate that splitting the pricing function (homogeneity hint) into two components one driven by moneyness $\frac{S}{K}$ and one by time-to-maturity, leads to better out-of-sample pricing. They train networks of varying complexity (up to 10 hidden nodes) in the first half of each year, select the network based on the third quarter, and test the results in the last quarter. Diebold-Mariano (DM) statistic and the mean squared prediction error (MSPE) are used for pricing evaluation. They also point out that the hedging performance needs to be considered in the network selection process rather than being used simply after the best network with respect to pricing be chosen.

An alternative methodological improvement is suggested by Gençay and Qi (2001). The authors test the effect of regularisation techniques, in particular Bayesian regularisation, early stopping, and bagging in the context of non-

[16]Note that at least one study (Anders, Korn, and Schmitt, 1998) indicates that the price level of the underlying (an index) has some limited predictive power.

parametric option pricing. They use S&P 500 call options between 1988–01 and 1993–12. In addition to the pricing data, the three-month T-bill is used as a proxy for the risk-free rate, and a three month moving sample of volatility is used. The results are compared to a neural network without special considerations during training. Each year is split into training set (first half), validation (third quarter), and test set (fourth quarter). The authors stress the shortcomings of splitting the data set in this way. To test whether improvements were made, the mean squared prediction error and the average hedging error are presented along with a measure combining mean and variance $\nu = \sqrt{\mu^2 + \sigma^2}$; the authors also apply the DM test statistic and the Wilcoxon signed rank (WS) test.

Gençay and Salih (2003) further investigates these results using S&P 500 index options between 1988–01 and 1993–10, adding also the Bayesian information criterion (BIC) as an alternative to the regularisation techniques and studying in particular, the relationships between input variables and mispricing between ANN and BS models. They find that much of the bias found in the BS model is eliminated by the use of ANN. They follow the partitioning scheme of the previous article discussed and use historical volatility matching time-to-expiry but not less than 22 observations.

Dugas et al. (2001) Dugas et al. (also 2009) further show that in a process similar to the two previous studies, alternative approximators can be used, which benefit from the derivatives. While the approach is different from the ones seen before (Lajbcygier, 2004; Choi et al., 2004; Ormoneit, 1999), it relies on a different functional form and the neural network is used as a benchmark. They apply this to S&P 500 European index call options between 1988 and 1993. They find that the use of such knowledge is beneficial for forecasting, especially generalising, in reducing the MSE.

Focusing on synthetic European call options data in the presence of dividends (as a continuous yield), Le Roux and Du Toit (2001) confirm that the pricing function can be learned but were unable to find an optimal network architecture despite testing a fairly large range of designs, one or two hidden layers, and for the single hidden layer between six and 24 hidden nodes.

An alternative to the use of Kalman filters above is given by Ghosn and Bengio (2002). Instead of building a single network and allowing for parameter drift in some way, a number of models are built simultaneously with somewhat different parameters. Those parameters are however constrained and those constraints form the "domain-specific bias." Using European call options related to

the S&P 500 index between 1987 and 1993, the multi-task method (i.e. the innovation model) almost always outperforms the traditional single network. The models assume homogeneity and vary in terms of the time frame used for training. Unrelated to the main question, the author points out, however, that very long timeframes (1250 days) lead to poor performance. Instead a two-input network is used subsequently, which is the solution Hutchinson, Lo, and Poggio (1994) arrived at by assuming constant volatility over the whole period.

An alternative to the bootstrapping for inference (Lajbcygier and Connor, 1997a) is the three-phase process introduced by Healy et al. (2003) as well as Healy et al. (2004). The authors demonstrate how, by training an additional network on what is the validation data set for the primary one to be analysed to infer confidence limits. They demonstrate its use on LIFFE FTSE 100 options pricing using ANNs.

Huang and Wu (2006b), and Huang and Wu (2006a) (see also Huang, 2008) compare a number of hybrid models but rather than combining machine learning techniques and the reference model only, they combine filters and an additional model to the problem, i.e. Monte-Carlo filters and Unscented Kalman filters, respectively. Using the same data set, call and put options from the Taiwan Futures Exchange between 2004–09–16 and 2005–06–14. they find that the hybrid model using Monte-Carlo filters (and Unscented Kalman filters, respectively) and SVMs outperform competing hybrid models, including a combination of an ANN with the respective filter. Very little information is given on the fitting procedures, however. The authors focused on minimising the RMSE.

Traditional models also underperformed relative to a 'hyperparameterised' model proposed by Jung, Kim, and Lee (2006). The model selects parameters using a neural network kernel for pricing and implied volatility as the relevant input parameter in addition to the usual set. The risk-free rate was not used as it changed little during the study period. Testing the model on the Korea Stock Exchange KOSPI 200 index for the year 2000, the model results in lower errors than the competing MLPs with various learning algorithms as well as the BS model and an RBF network. This was true in the training and test subset.

Another way of deriving the network configuration is through the use of evolutionary algorithms. Wang (2006) tests this approach, where the network size and connectedness is determined by an evolutionary algorithm followed by training. The resulting network is tested in Taiwan using two warrants with the same underlying security and a two-month sample period of 2000–02–10 to 2000–04-06

using daily data. Using a delta-hedging strategy, the proposed model nearly doubles the profit from the trading compared to the BSM model.

Lee et al. (2007) also report positive results in regards to using particle swarm optimisation (similar to Dindar and Marwala, 2004; Dindar, 2004). Unlike Wang (2006), it is not neural networks but the model itself that is determined. The authors find it preferable to genetic algorithms Wang (conceptually similar to 2006) when estimating implied volatility of Korean KOSPI 200 call options between 2005–02–27 and 2005–03–10.

One of the few to address the specific question of network configuration without introducing a new training algorithm or optimisation technique are Thomaidis, Tzastoudis, and Dounias (2007). For the specific purpose of option pricing, the authors suggest various processes a researcher can follow. In particular they suggest that the process should be informed by statistical tests or metrics. The process is principally based on Lagrange Multiplier tests and information criteria such as Akaike information criterion (AIC) for the simple-to-complex version. Neurons are added until their marginal contribution falls below a particular (probability or information criterion) threshold.

Alternatively, they suggest starting with a large model and tentatively removing neurons, testing for its marginal contribution. In essence the process is similar to step-wise regression with the exception that the tests are intended to allow for non-linearity and the steps taken are specific to the architecture of neural networks.

The authors test their methodologies using European-style S&P 500 equity option contracts for the period of 2002–05–17 to 2002–07–29. They find that the simple-to-complex models perform better in terms of fit and generalisation.

Quek, Pasquier, and Kumar (2008) provides some additional guidance on the topic. Citing prior research the authors note that some heuristics do not work and raises the issue, which may affect the prior article's methodology, i.e. that pruning may work only if all variables are normalised. This may not apply in all cases and care needs to be taken as variables are redundant at 0 only if that is the case.

Despite considerable research in this area, it is noteworthy that the two points of interest identified by Hutchinson, Lo, and Poggio (1994), additional pricing information (variables) and performance metrics, little progress has been made in both areas. Instead, the rather technical details of the model development process and data-specific concerns have remained the primary focus of attention.

It may well be that the former cannot be addressed until a better understanding of the latter has been achieved.

2.5.2 Volatility Modelling

Overview

As discussed previously, volatility modelling is a central part of many finance-related activities. Even a limited discussion of competing models would require more space that is reasonable for the development of the methodology in the next chapter. Instead of providing a detailed discussion, the focus is on three aspects, which will be reviewed in sequence:

1. option pricing with special consideration for the volatility input,
2. ANN-based forecasts, and
3. hybrid models.

Option Pricing Models with Special Consideration for the Volatility Input

Meissner and Kawano (2001) combined the pricing networks with a (modified) GARCH(1, 1) volatility estimate assuming homogeneity. Importantly, they test whether a single network for various underlying securities, and individual networks (one per underlying) outperform the reference model (BS). They confirm both using options on 10 US stocks from 1999–05–01 to 2000–01–31, excluding various observations based on a number of criteria. The results show, among others, that the MLP and GRNN outperform the reference model. As other studies, they too report large differences in architectures, ranging from 8 to 18 hidden nodes for per-underlying networks (in one case with two hidden layers).

Pande and Sahu (2006) approaches the volatility estimation problem embedded in ANN-based option pricing differently. Rather than limiting the choice to historical volatility or using implied volatility, a principal components analysis is conducted. A number of potential candidate models are provided; the results of this step are not reported. The output vector is then an input to the option pricing problem using fairly small networks of three to five hidden nodes. These are compared using correlation, ME, and MSE. Using data provided by BSE India on Satyam Stock, two thirds of which were used for training, the authors report results consistent with findings in other markets, i.e. that the BS formula is fairly

good, and here better, at pricing ITM options, and ANNs are better in cases of OTM options.

Tzastoudis, Thomaidis, and Dounias (2006) improve the fit through two approaches, firstly, they simplify the functional form by converting part of the problem to the forward price (note that this conversion was discussed in chapter 2 in the context of future pricing and the dividend adjustment sometimes made). Secondly, the study infers the volatility surface in the form of a price multiplier as a function of moneyness and time-to-expiry on two distinct days and uses a 45-day historical volatility as a proxy. The data set consisting of European-style options on the S&P 500 index on 2002–05–08 (in-sample) and 2002–07–19 (out-of-sample, respectively). The resulting networks, which are modest in size having between one and five hidden nodes, captured some of the prominent features of the options. However, an additional weighting scheme for the MSE gives greater weight to options near ATM levels. Interestingly, the best networks out-of-sample were those with very few hidden nodes. The resulting hybrid models outperform the BS model as had been reported previously.

As discussed before, Lee et al. (2007) applies particle swarm optimisation to the implied volatility estimate using Korean data.

Tseng et al. (2008) compare two different GARCH models in combination with artificial neural networks, the standard EGARCH and a modified Grey-EGARCH. Applied to Taiwan Futures Exchange index options (2005–01–03 to 2006–12–29), of which 70 % was used for training (the authors term it a "forecasting model"). The results are mixed, depending on the error measure though they favour the proposed Grey-EGARCH model.

Similar results on the same data set are reported by Wang (2009a) with three competing underlying volatility models: GARCH, GJR-GARCH, and Grey-GJR-GARCH. Here too, the quality varies across error measures but the GARCH-based model generally underperforms. The authors conclude on balance that the Grey-GJR-GARCH volatility leads to better results in the context of ANNs.

A similar study was conducted by Wang (2011), who compared a novel hybrid stochastic volatility with jump and support vector regression model on the one hand, with several competing models. These included support vector regression with stochastic volatility, with GARCH-based volatility estimates, the Garman-Kohlhagen pricing model, and an ANN . The options being studied were currency options on Australian Dollar, Euro, Japanese Yen and British Pound (all vs. the

US Dollar) during 2009. The author conclude that the proposed model outperforms the others but also that the ANN is preferred over the traditional model.

The review of volatility forecasting literature using machine learning techniques is less exhaustive given the large number of publications. Instead, the review aims to cover a wide range of different approaches with respect to variables, modelling parameters, and design choices. As in the previous section, the notation is again largely standardised.

Gonzalez Miranda and Burgess (1995) study European Ibex 35 options forecasting hourly changes using an integrated modelling approach as argued in this thesis before, expected the ANN to outperform on the full set given the nature of the learning process without this holding true necessarily for the reduced set, an argument similar to the one in the introductory section of this thesis. They also find that momentum is reflected and that changes in the implied volatility appear to be a function of the strike state without, however, providing additional details.

Carelli, Silani, and Stella (2000) approaches the option pricing problem in a way very similar to the this thesis focussing on network selection, however. The authors introduce a complex procedure to guide in the (feedforward) network design. Using USD/DEM call and put options, they model the volatility surface as a local volatility surface. They conclude that the process is helpful especially in the presence of limited data and, as the discussion of local volatility above showed, that the pricing assumptions made by market participants could be used for pricing related instruments, the no-arbitrage argument.

Wang et al. (2012) (see also Wang, 2009b) offer one of the most comprehensive studies of volatility modelling for option pricing albeit with the aim of forecasting a price. They study historical volatility (30 trading days), implied volatility, a deterministic volatility function using a quadratic form similar to Dumas, Fleming, and Whaley (1996), and Peña, Rubio, and Serna (1999) (this is a function of moneyness only instead of the volatility surface), GARCH and GM-GARCH models (Grey-Model GARCH). Based on these models backpropagation networks are fitted to the intraday call and put option prices for TXO (TAIEX options). In addition to the various volatility models, an adjustment is made by using the index future instead of the spot value due to the difference in dividend handling. The time period included the full years 2008–2009 thus specifically including the GFC. Networks are trained for various sub-periods and the total period though testing errors across various time periods are not reported, only the author's

own performance ratio metric (for RMSE, MAE, and MAPE) by sub-period. The authors conclude that the choice of activation has no significant impact on performance, more neurons (tested up to 4) are preferable over fewer. They also add that there was a drop in performance from 2008 to 2009 suggesting that characteristics of the time-series changed over the course of the GFC. There is insufficient data presented in the article, however, to understand the impact of the GFC on generalisation and more broadly model fit across a range of models. The discrete volatility function is second best, implied volatility is preferred. As suggested in the econometrics literature, the GARCH-style models fail to outperform historical volatility in this study as well. These results are not consistent with those by Wang (2009b), who compares a number of volatility models using 2003–2004 data concluding that the best models are of of the GARCH-family. The methodology is otherwise similar if less complex.

ANN-based Volatility Forecasting

Among the earlier studies, Malliaris and Salchenberger (1996), and Donaldson and Kamstra (1997) stand out. Malliaris and Salchenberger (1996) uses a similar approach as in the case of option pricing and studies the implied volatility of S&P 100 index futures options for 1992. The explanatory variables are thus the lagged volatility, current volatility, closing price, time to expiry, additional volatility at different forward times, put open interest, as well as combined prices of options and market prices. The latter implicitly modelling the relationships between option prices and the prices of the underlying securities discussed above.

Donaldson and Kamstra (1997) compare various models as well as an ANN using S&P 500, Toronto Stock Exchange Composite Index, Japan's NIKKEI, and London's FTSE index between 1969–01–01 and 1990–12–31. The authors report that neural networks with lagged unexpected returns as inputs can be useful in forecasting volatility. Importantly, they find that there "may be important differences between the processes driving returns volatility in the four countries [they] study." This is a conclusion Lajbcygier also reached with respect to ANN-based option pricing compared with the US.

Hu and Tsoukalas (1999) compare various ARCH-based models and simple statistical models to an ANN combining the former as well as an averaging model, i.e. creating an ensemble learner in a way. They find that it does not improve results generally though it does perform better with respect to mean absolute

error (instead of the square residuals) and "behaves well in the crisis period [of 1993]."

Consistent with previous results and using option straddles for evaluation purposes, Dunis and Huang (2002) find that recurrent neural networks outperform traditional models as well as model combinations at forecasting volatility. The study covers British Pound and Japanese Yen exchange rates (each against the US Dollar) between 1993–12 and 2000–05. The input vector covers "exchange rate volatilities [...], the evolution of important stock and commodity prices, and, [...], the evolution of the yield curve. In addition to the usual RMSE, MAE, Theil's U, and the percentage of correct directional forecast are evaluated for forecasting accuracy evaluation in addition to the Wald test statistic. The trading strategy using straddles assumed fixed holding periods and included transaction cost.

Using lagged returns of US Dollar/Deutsche Mark (DM) exchange rates between 1980–01–01 and 2000–01–01, Gavrishchaka and Ganguli (2003) studies the use of SVMs. The authors conclude that the SVM can be used to model the long memory of the volatility processes in particular, and outperforms traditional GARCH models.

Dash, Hanumara, and Kajiji (2003) use a larger data set including also the US Dollar rates against the Japanese Yen and the Swiss Franc and hourly quotes during 1999 (and a limited set for Deutsche Mark quotes). The inputs to the networks include currency data, yield curve data, and GARCH data, which also serves as a reference model. The actual subset used for any model is found using genetic optimisation or a special algorithm. The results are largely consistent with prior literature, i.e. That improvements above the GARCH model can be made using specific networks. The actual performance, and here the inclusion of specific inputs, is to some degree specific to particular securities.

Positive results are reported by Hamid and Iqbal (2004) with respect to forecasting S&P 500 index future volatility (several contracts) between 1984–02–01 and 1994–01–31 when compared to the implied volatility of the Barone-Adesi and Whaley model. The study implements multiple forecasting horizons and uses non-overlapping data. The explanatory variables are two index values, lagged futures on the index and seven commodities, an exchange rate (Japanese Yen), and several points on the yield curve. The selection is based on correlation and the "relative contribution coefficient." In addition to the usual MAE and RMSE statistics, the Mann-Whitney test is used for comparison.

Kim, Lee, and Lee (2006) use an alternative method for training RBF networks. Two networks are trained first to model the local volatility, then to minimise the pricing error of the first stage network. Lower pricing and hedging errors are reported for 1995 S&P 500 index European call options.

While not using ANNs, Audrino and Colangelo (2010) model the implied volatility surface using a semi-parametric technique, regression tress. They use Option Metrics' Ivy database with S&P 500 index options between 1996–01–04 and 2003–08–29. Given that a surface is evaluated, they suggest using not only the daily and overall SSE but also a weighted metric, "the daily and the overall averaged empirical criteria." They also experiment with a number of predictor variables, notably the yield curve, and option-price related data based on a discrete-time reference pricing model. For the specific technique applied, moneyness and time to expiry are the most important factors, with leading and lagging factors playing a smaller role. The authors conclude that the approach leads to the best model of volatility surface dynamics and that it may even be used when there are structural breaks.

According to Mantri, Gahan, and Nayak (2010), ANNs are no different in their ability to forecast stock market volatility compared to GARCH, EGARCH, IGARCH, and GJR-GARCH models. Using the open, high, low, close values (or only the close) of two Indian index series (BSE and NIFTY), the ANN forecasts are less volatile but using analysis of variance (ANOVA) to compare the annual observations, no difference could be found. The data covered the years 1995 to 2008 and only the annual observations were used for the statistical analysis.

A combination of SVM and GARCH is introduced by Chen, Härdle, and Jeong (2010). This approach is found in later discussions on the topic. An SVM is fitted to the modified returns series and conditional standard deviation in an iterative way. This is conceptually similar to traditional recurrent networks or time-delayed feedforward networks. Using MAE, directional accuracy measures, and the Diebold-Mariano test for differences in MAE on daily British Pound (against the US Dollar) exchange rates, NYSE composite index forecasts (both 2004 to 2007), and a synthetic data set, the authors conclude that MLP with an additional feedback connection and recurrent SVM GARCH models outperform alternative ones in one-period ahead forecasts.

Ou and Wang (2010a) compares standard GARCH, EGARCH and GJR models to least-squares SVMs. They term this a hybrid approach though it is structured differently from what is considered 'hybrid' here. The authors estimate the

GARCH parameters using SVMs. Apart from finding that these combined models perform better generally, the authors stress that they are more robust to the changes brought about by the GFC. The data covers the stock market index of each of Singapore, the Philippines and Kuala Lumpur comparing the years 2007 and 2008 directly.

Ou and Wang (2010b) and Hossain and Nasser (2011) compare such support vector machines and related models for forecasting volatility. Studying the Shanghai Composite Index between 2001 and 2006, and the Bombay Stock Exchange (2006–10–05 to 2010–11-01) and NIKKEI 225 (2001–01–04 to 2010–11–01) respectively, the authors compare the standard GARCH model to support vector machines (for regression) and relevance vector machines (a probabilistic version of SVM). Importantly, for the methodology developed below, the models are structured similar to the GARCH-model. They use past observations of returns and residuals of the GARCH-form model to predict the next observation. Ou and Wang (2010b) finds that the relevance vector machine performs best and the standard GARCH model worst, the SVM models are relatively good as well. Hossain and Nasser (2011) add that the ARMA-GARCH model is superior to GARCH but more importantly that the vector machine models are the only ones meeting robustness criteria. This is surprising given the strong theoretical foundation of GARCH.

Ahn et al. (2012) employ a strategy of repeated training and testing using overlapping periods, and option implied volatility and greeks, to forecast future implied volatility for the KOSPI 200 (Korean index) ATM options successfully.

Hybrid Models for Volatility Forecasting

As mentioned before, one of the difficulties of using ANNs is the identification of input variables, not least because many statistical techniques, which are used for this purpose, assume linearity, which is precisely the limitation an ANN is meant to overcome. Hyup Roh (2007) makes an important contribution. In addition to comparing a pure ANN model and three hybrid techniques, the contribution of various input variables to the forecast is presented. The study focuses on the KOSPI 200 index volatility and it's 22-day forward estimate. The factors, whose contribution are determined are the square return, return, volume, index level (current and previous), two government interest rates (at 3 month and 1 year maturities), open interest, premium average, contract volume, the previous day's

squared volatility or similar estimates depending on the model, i.e. the last term varies with the model that is combined to form the hybrid network. The resulting hybrid model (a combination of ANN and exponential GARCH (EGARCH)) is found to have greater predictive power with respect to volatility and directional forecast.

Aragones, Blanco, and Estevez (2007) simplify the inputs to an RBF network, and consequently to an equivalent MLP, to the implied volatility and 11-day momentum of the IBEX–30 index and its futures for a number of sub-periods. To determine the quality of the forecasts, the authors used linear regression of predicted volatility and observed volatility. They find that the RBF model adds value above the implied volatility estimate and that this is true even in the presence of external shocks.

A different form of hybrid models, at least in the context of volatility modelling and option pricing is that proposed by Chang (2006), and Chang and Tsai (2008). In both cases two competing models are built, one of which is a non-parametric one, and their results are combined using a weighting learned from data using a machine learning technique. The models are a fuzzy neural system and an non-linear GARCH model combined using support vector regression (SVR), and a combination of an SVR and the Grey Model and GARCH modelling using an ANN, respectively. Using index data of four major equity markets and data from London International Financial Futures and Options Exchange (LIFFE), they find that the use of machine learning leads to better predictions while the combined model performs especially well.

Andreou, Charalambous, and Martzoukos (2010) study the fitting of deterministic volatility regression functions using neural networks. They are function modelling the volatility surface. Similar to Dumas, Fleming, and Whaley (1996), or Peña, Rubio, and Serna (1999), a number of functions are suggested, the most complex one being quadratic with an interaction term. An ordinary ANN is used with an additional layer containing the parametric pricing function. This allows for the training of the network to configure the parametric model, i.e. to supply the parameters accordingly. Using S&P 500 index call options for 2002–01 to 2004–08, the authors conclude that the proposed model are preferable. They also note that if the model is chosen according to its hedging performance, it performs better at this function than one chosen according to pricing accuracy. This supports earlier results and implies that the decision criterion is critical when

choosing amongst competing models. Of the parametric models the stochastic volatility with jump model is best.

Hybrid models can also be formed directly from a time-series, an approach adopted by Tung and Quek (2011). They train a fuzzy-neural network based on implied and historical volatility. It is a time-delayed network with the last ten observations used to forecast the next observation in the same series. The authors also develop a methodology for creating a trading system based on such volatility forecasts using the a particular options strategy, a straddle, and technical indicators. They test it successfully in the Hong Kong market but unfortunately only with data between 2002 and 2006 thus not including the GFC.

Finally, Hajizadeh et al. (2012) combines the insights of previous research by several authors and compares several GARCH (including GJR-GARCH) and EGARCH models as well as two hybrid ones. S&P 500 data between 1998–01–02 and 2009–08–31 was used with a forecast of 10 and 15 days forward. The hybrid ANN models combine the best-performing underlying model, EGARCH(3, 3), with a number of explanatory variables identified using a correlation analysis. They were the index price, the NASDAQ price, the Dow Jones price, 1-day lagged volatility, 3- and 6- month daily treasury yield, index squared return, volatility using the preferred model and the traded volume. Of the two hybrid models, the first is a simple hybrid network similar to those seen before, the second one includes added time series, which were created based on the statistical properties of the actual series. A split ratio for network modelling of 70 %–20 %–10 %. Both hybrid models and especially the second one outperform the statistical models on all error measures ME, RMSE, MAE, and MAPE.

2.6 Open Research Problems

The above literature review has demonstrated a persistent research interest in the use of machine learning techniques in finance and in derivatives pricing and volatility modelling in particular. This is not surprising given the competitive nature of the industry and the difficulties decision makers face on an ongoing basis. The volatility forecasting literature is even more extensive then what is discussed in the previous sections.

A significant number of questions have been left unanswered so far, or addressed only superficially:

- No prior research has been conducted on Australian equity options even though the literature regarding Australian index options is quite extensive. The question arises if the limited evidence regarding equity options, especially in the presence of dividends is transferable to:
  - the Australian market and
  - the post-GFC period beyond some preliminary findings resulting from overlapping periods, and no conclusions regarding the applicability post-GFC have been reached. Only some of the later publications such as Wang et al. (2012) address the issue of the GFC though not specifically its impact.

- Furthermore, little systematic analysis has been conducted regarding the interaction of option pricing and volatility modelling when either or both use machine learning. This is especially true for the modelling of the volatility surface or even just the volatility term structure.

- Despite (or possibly because of) the general difficulties of attribution and the lack of explanatory power of ANNs, relatively little prior research has been conducted with regard to the benefits of a volatility model as opposed to a direct pricing mechanism. Some exceptions to this were discussed in section 2.5.2.

The research throughout this thesis aims to address these shortcomings of the prior literature by focusing on the specific hypotheses stated in chapter 1.

In addition to these questions, others will be left for future research, further discussed in 5.3, but they impact the methodology design to some degree and are thus stated explicitly:

- Despite extensive research over the past few decades, surprisingly little progress has been made with respect to the design of ANNs, their architecture and their learning process (except with respect to regularisation, bagging and boosting) which are now fairly well-understood. It is therefore difficult to replicate designs or form a justifiable view of whether a particular design is good especially when it fails to perform as expected. In the absence of clear general principles, the failure may be due to the user or a result of the data; it is difficult to form an opinion in this case.

- A reference model for the volatility surface is also missing from the literature unless one is willing to use the local volatility models. This is particularly true for vanilla options where it is not possible to find a source from which to infer the volatilities.

Chapter 3

Methodology

3.1 Model Development and Experimental Design

3.1.1 Overview

Answering the questions identified in the previous chapter requires a dual approach to the option pricing problem. Firstly, an appropriate volatility forecasting model needs to be found with the quality metric subject to various assumptions. Secondly, the pricing model problem needs investigation with respect to the reliability of the pricing mechanism. Identifying the two components individually does not necessarily lead to an improved model to support pricing decisions. As discussed in the previous chapter, such a model may simply lead to better forecasts with similar or even inferior results with respect to pricing.

Thirdly, due to the interaction effects evidenced by the very existence of volatility surface models, an evaluation of the combined effect is needed. This allows for the identification of the likely source of any improvement, whether the volatility forecasts need improvement, the pricing mechanism, or if the problems are interdependent to the degree that they are inherently inseparable. This is conceptually the same consideration as Hull's integrated Black-Scholes and GARCH formula.

The problem is further complicated by the variety of markets and market conditions under which prior research has been conducted and the question naturally arises to what degree such research and the resulting design decisions are transferable to the market of interest here. Some of those assumptions and decisions have been discussed previously; in this section a detailed description of the methodological issues and design choices is given.

The remainder of this section discusses the model design and variable choices, how they were derived and the way of interaction between the various models. A

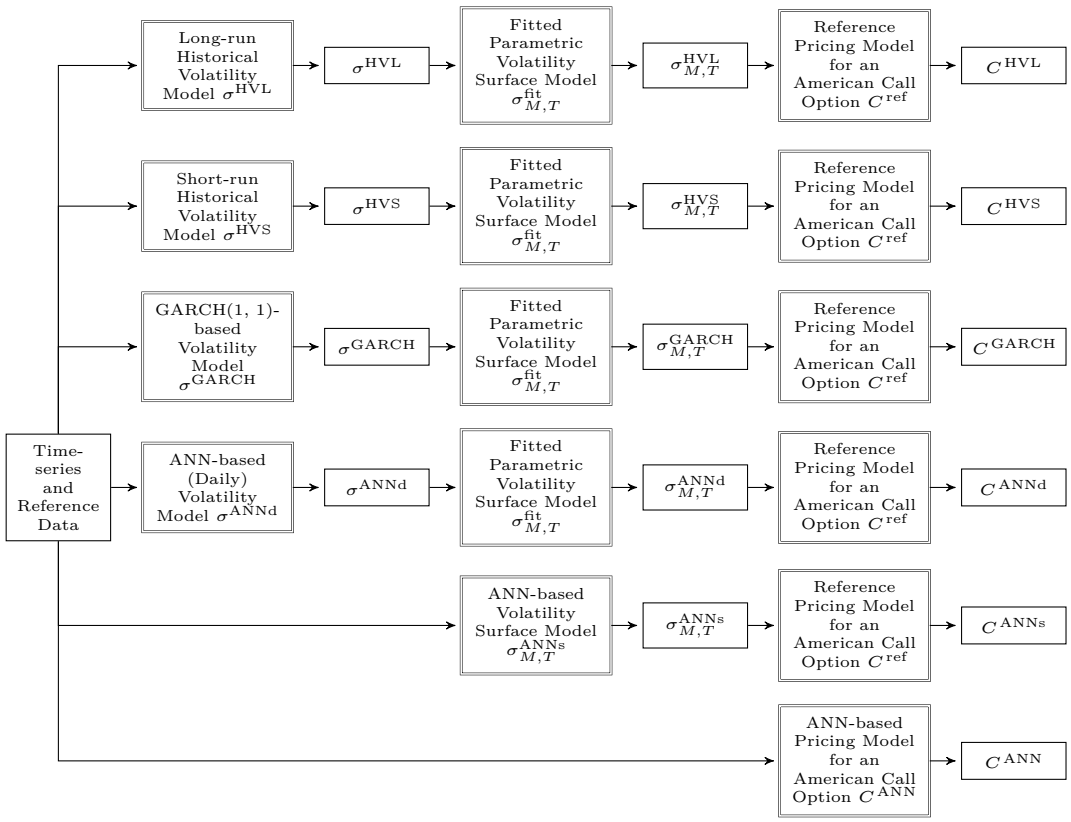

Figure 3.1: Overview of models used including implied volatility models, which are used for evaluation only. σ, $\sigma_{M,T}$, and C superscripts indicate the significant model difference, i.e. the first model that is different in the process.
Market data used as input variables is discussed in greater detail in subsequent sections of this chapter.

brief description of the data collection process follows along with a more detailed explanation of the implementation, especially with respect to the architecture and learning process of the ANNs. An overview of evaluation metrics is given and the chapter concludes with some critical remarks and epistemological questions that remain.

Figure 3.1 gives an overview of the methodology outlining the essential models, explanatory and explained variables for each model. The ultimate goal is to identify whether the benefits of using ANNs exist in the current Australian equity options market insofar as they are applied to pricing. This notably excludes the analysis of decision support for whether and when to exercise options, specific trading, investment, hedging, or market making strategies, etc. The focus is principally on the *general* applicability of the technique to pricing and therefore on

their average performance with respect to some metric. Of particular and specific interest is if and how ANNs can improve the option pricing process, whether it is through a better volatility forecast as the most significant input, through a better pricing mechanism when presented with a volatility forecast or if the performance improvement is due to the combined effects of forecasting and pricing.

3.1.2 Volatility Forecast (Hypothesis 1)

The objective of the volatility forecast, the first step in the research process, is to correctly forecast the realised volatility between the current point in time and the time of expiry of the option. Evidently, only one such value is ultimately correct. Equally, only one number can be the expected value, or the best estimate for the volatility. Consequently, there should be only a single volatility forecast for any stock and time to expiry. Deriving the best estimate is a classic forecasting problem and as the literature review showed, is typically derived from past observations of volatility. It is principally possible to consider additional attributes to aid the forecasting. In particular exogenous shocks should be expected and may be modelled. Little evidence for the benefit or even a consensus on a basic set of such attributes could be found. By far the most prevalent models are those limiting their analysis to the past return series alone.

The baseline ANN is one using only historical volatility of varying past periods to allow for long memory effects as well as short-term deviations of volatility from its long-term mean. This provides sufficient flexibility to the learning network to respond by adjusting the result in favour of long or short-term volatility depending on the circumstances.

For clarity, the following notation will be used to refer to any model:

$$f_\theta^{\text{name}} : \vec{x} \mapsto y \quad \text{or} \quad f_\theta^{\text{name}}(\vec{x}) = y \tag{3.1}$$

where 'name' is the model name, y refers to the desired output, θ to the model specification, and $\vec{x}$ to the input vector. Three particular types of model classes are introduced and using the above notation, results in them to be one of:

$$\begin{aligned} \sigma(...) &\quad \text{for volatility forecasts} \\ \sigma(M = \tfrac{S_0}{K}, T, ...) &\quad \text{for volatility surfaces} \\ C(...) &\quad \text{for (call) option prices} \end{aligned}$$

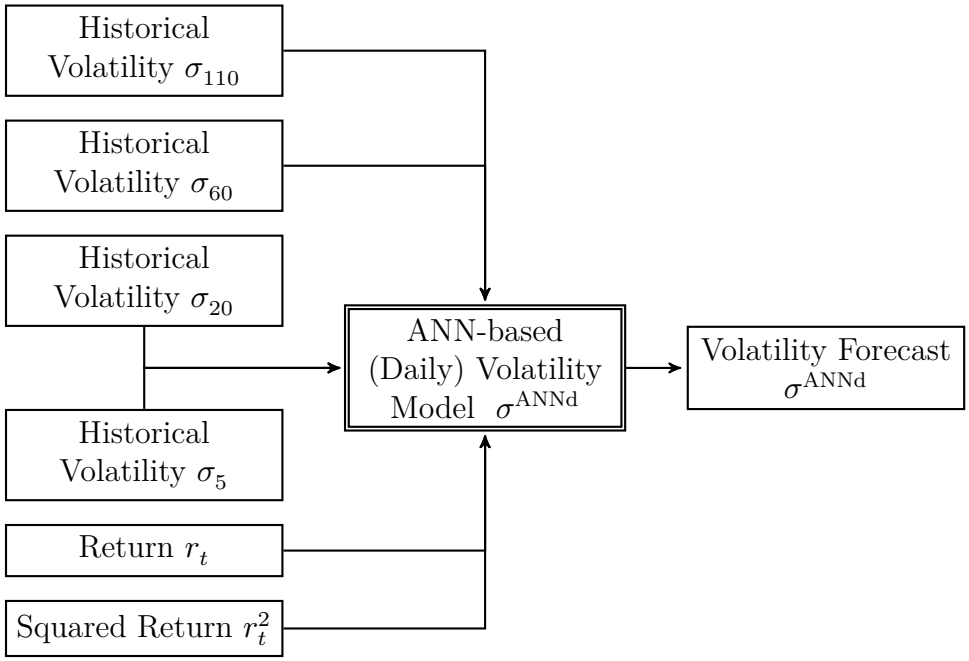

Figure 3.2: ANN-based (Daily) Volatility Model

The baseline volatility network (see Figure 3.2) is thus:

$$\sigma^{\text{ANNd}} : (\sigma_{110}, \sigma_{60}, \sigma_{20}, \sigma_5, r_t, r_t^2) \mapsto \sigma\sqrt{N} \tag{3.2}$$

the network provides the annualised volatility (see equation 2.21) forecast based on the past n days of historical volatility, the last return observation, and its squared value. The inclusion of the previous five day volatility is largely due to what appears to be common practice by professionals to use a single week's worth of observations. As Haug (2007) points out, this is a very unreliable measure, the number of observations is far too low to yield a meaningful result. A transition to intraday data is an alternative, whether such a move would be beneficial is unclear, however. Here the goal is to build a simple reference ANN-based model without any particular attempt at optimising. The high noise ratio in the last component typically results in the output to be independent from that particular input variable.

As is apparent from the review by Poon and Granger (2003), the evaluation is based on a variety of proxies for the volatility including realised volatility over a range of forward periods. The argument made with respect to the wide confidence interval in the σ_5 input, also applies to the target value. This is both true during

the learning where a value is presented as well as to the error calculated during model evaluation.

A four-week time frame is used for the forward period. There is no particular reason for the choice, as there are typically no particular reasons for such choices in past literature. Four weeks is sufficiently long to allow for a reasonable amount of data – albeit not as the conventional 30 observations used in statistics – but short enough to test the response of the model to temporary fluctuations. The average time to maturity of options considered, starting either at the time of creation or at the time of first sale, or indeed at any other point in time of significance (by volume, open interest, etc.) are all alternative choices. In the absence of additional criteria or more specific objectives of the decision maker, the shortest period is used for consistency with the reference models. Furthermore, time-weighted volatility or volatility weighted by option "greeks" could be used but they too are most useful if a specific objective is pursued.

As with traditional econometric models, the forecast at t is based on the previous data only and is thus applicable to any pricing and evaluation at t and not $t + 1$. This means that no additional delay needs to be used for a valid model. This is also true for the networks trained while the actual implementation differs slightly. Due to the way time is represented, the implementation uses all data up to and including the current day, i.e. including the closing price of the current day. When synchronising time series across models, the one day (specifically one business day) lagged volatility measure is used for pricing options. This ensures that volatility estimates are known at the beginning of the trading day and based on the previous close and earlier information.

The literature review demonstrated that past studies used a number of additional variables for the input. Given the focus on index volatility, these often referred to related indices. No such information is included here. There are a number of methodological reasons for their omission. Firstly, determining a related time series is less clear for an equity security. While a related (benchmark) index could be chosen, it is not clear why any or a particular index should be a suitable input.

Secondly, unless the same inputs are presented to the competing models, it is not clear if the different model form or the additional inputs are the driver for varying performance. Such comparisons would have to include ARCH-style models with exogenous variables and regression models using historical volatility along with the enhanced ANNs. This is a question left for future research (see

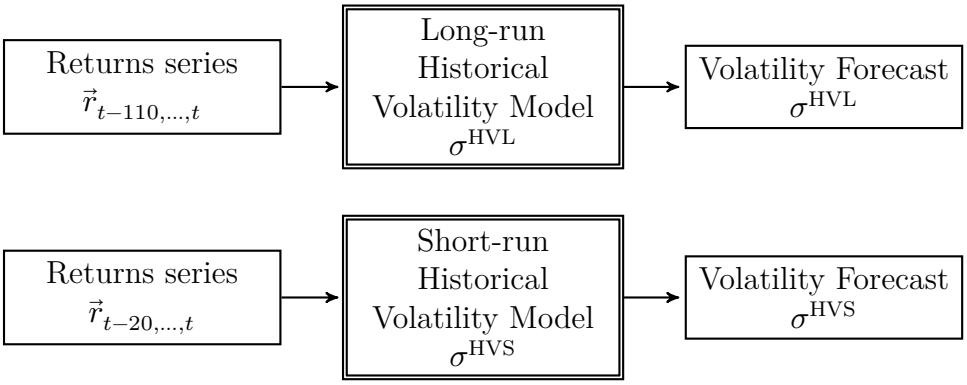

Figure 3.3: Historical Volatility Models

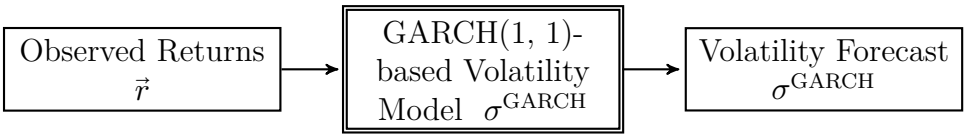

Figure 3.4: GARCH-based Volatility Model

section 5.3). The use of the risk-free rate, i.e. a government bond yield as a proxy, is particularly difficult. It is certainly justifiable on the grounds of being a proxy for investor's risk-preferences.

Thirdly, using it in the context of volatility forecasting implies its use in the subsequent pricing step, where it is – under the assumptions made here – incorrect due to the margining regime.

The choice of forward period is only relevant with respect to the evaluation of volatility models on their own. Once a volatility estimate is needed for option pricing, the period would have to match that of the time to expiry. As explained before, an individual model for each such period may have to be built, or the one period model applied iteratively as was suggested for the GARCH model. This is feasible in the case of GARCH models as these are combined forecasts of return and volatility. This approach fails in the case of direct volatility forecasts using an ANN. This requires particular attention when choosing the time frame for which to measure realised volatility and when comparing different models for selection purposes.

The models are built for panel data, i.e. training includes data over the in-sample period and across several securities. This is benchmarked against standard econometric models, specifically historical long- and short-term historical

volatility σ^{HVL} and σ^{HVS} (see Figure 3.3, both corresponding to the long and near-short-term specification of the ANN-based model), and the GARCH(1, 1) model σ^{GARCH}(see Figure 3.4). Those were chosen as representative of frequently used models in research and practice based on the review literature. They, and the GARCH model in particular, do not necessarily represent good models in absolute terms but sufficiently well-understood models to serve as benchmarks. They are also used in prior ANN volatility forecasting research.

Note that the historical and GARCH(1, 1) models return volatility for the specific sampling frequency σ_s requiring the usual adjustment (see equation 2.21) to arrive at annualised volatility σ. This is done implicitly in the case of the neural networks. For this reason, the networks are also presented with annualised volatility but not annualised returns.

Hypothesis 1 can then be answered by comparing the competing models:

- ANN-based (Daily) Volatility Model σ^{ANNd},
- Long-run Historical Volatility Model σ^{HVL},
- Short-run Historical Volatility Model σ^{HVS}, and
- GARCH(1, 1)-based Volatility Model σ^{GARCH}

by comparison to the realised volatility over the fixed time frame as specified in the ANN-based models (including applying the GARCH model repeatedly to arrive at a comparable time frame). Days of historical volatility refers to the number of deviations from the mean evaluated and the number of underlying returns is thus greater by one.

3.1.3 Option Pricing (Hypothesis 2)

The resulting volatility models allow for a comparison of simple forecast but not for the forecast for purposes of option pricing due to the fact that the volatility surface is not flat (a plane parallel to that given by the strike and time-to-expiry axes). Instead the surface needs to be derived based on the fixed-period volatility forecasts at time t.

The process is more complex in the case of the GARCH model and the simple volatility forecasts. Here a transform is needed and a function similar to those by Dumas, Fleming, and Whaley (1996), Peña, Rubio, and Serna (1999), or Andreou, Charalambous, and Martzoukos (2010) are used. The models are labelled

$\sigma^{\text{fit}}_{M,T}(M, T, \cdot)$, where the last parameter represents the volatility forecast as per any of the competing models above:

$$\beta_0 + \beta_1 M + \beta_2 M^2 + \beta_3 MT + \beta_4 T^2 + \beta_5 T \tag{3.3}$$

The parameters are found by OLS regression. For practical purposes, the volatility will be set to 0 (similar to the lower bound in the literature mentioned above) where the regression leads to negative values. Unlike those authors, the question that arises here is how to integrate the reference volatility forecast. Apart from adding it as another explanatory variable and fitting more parameters, including those of additional interaction terms, two options are available, which do not require a larger set of parameters:

- the surface is considered the additional volatility, relative to the reference forecast and thus β_0 in the regression formula is substituted with $\sigma + \beta_0$;
- the surface is the relevant scaling function, i.e. points on the surface represent factors, which are to be applied to the reference volatility forecast.

The former resembles the – usually successful – hybrid learning approaches, which use a reference model as a base and learn the difference. The latter is similar to some adjustments that are made by practitioners using tables to modify point forecasts for use in parametric models. Furthermore, this preserves the shape of the surface when there are changes to its level, the latter is chosen:

$$\sigma^{\text{fit}}_{M,T}(M, T, \sigma) : \beta_0 + \beta_1 M + \beta_2 M^2 + \beta_3 MT + \beta_4 T^2 + \beta_5 T = \frac{\sigma_{M,T}}{\sigma} \tag{3.4}$$

There is no valid reason for limiting the ANN to forecasting volatility for a single time to expiry or level of moneyness, however. Given the flexibility they offer, it is equally valid, and potentially beneficial to derive the volatility surface directly from the ordinary volatility forecasting inputs and the additional parameters of moneyness and time-to-expiry. The resulting model specification is:

$$\sigma^{\text{ANNs}}_{M,T} : (M, T, \sigma_{110}, \sigma_{60}, \sigma_{20}, \sigma_5, r_t, r_t^2) \mapsto \sigma_{M,T} \tag{3.5}$$

shown in Figure 3.5

This is exactly σ^{ANNd} with the two additional parameters. The reason for using $M = S/K$ lies in the assumed homogeneity as discussed by Garcia and Gençay (1998) as well as Garcia and Gençay (2000) in regards to option pricing and is

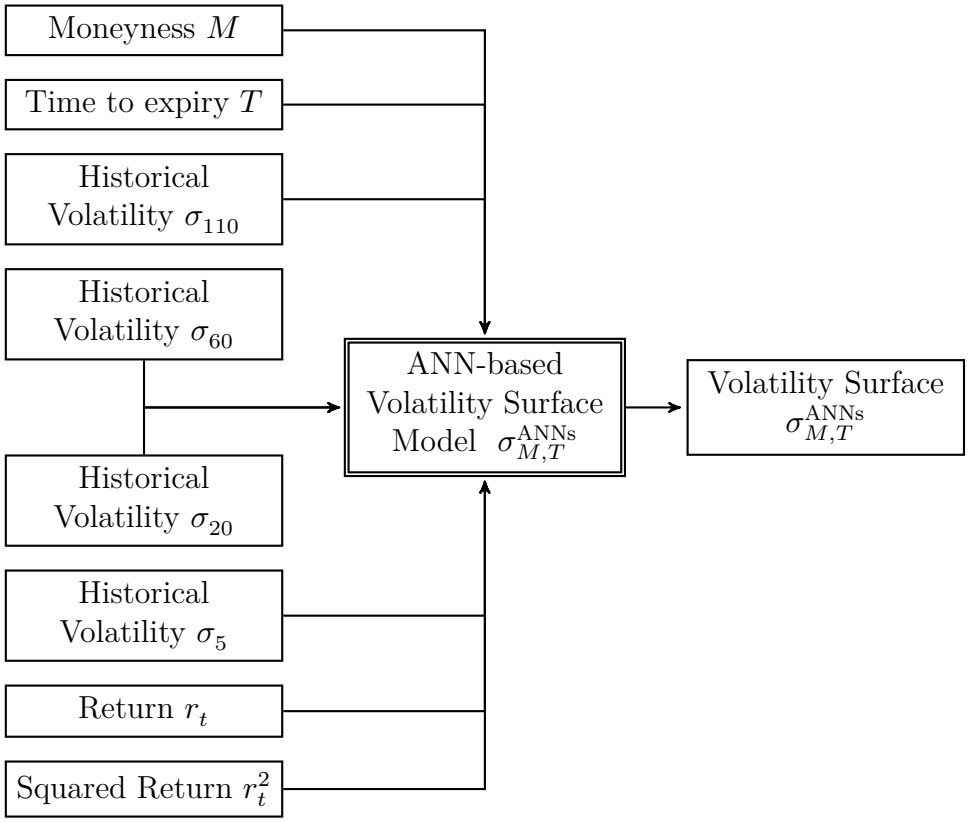

Figure 3.5: ANN-based Volatility Surface Model

chosen here too for consistency reasons. This allows for a direct comparison of the resulting option prices to those of the direct pricing network.

Hypothesis 2 could then be answered by comparison of these models in regards to their fit to the implied volatility surface or at the money but with varying time-to-expiry, to realised volatility.

The selection criteria and the underlying error measures and statistical inference methods used are discussed later in this chapter. In regards to the option pricing model, it shall suffice to note here that the appropriate choice between the two sets depends on the objective of the user of such a model. If the goal is trading highly-liquid options, the former will be sufficient as this limits the modelling to ATM options, or those relatively close to it, in most cases. If the question is of a more general nature, as is the case in this thesis, the second approach is needed as it covers ATM, ITM, as well as OTM options equally. The comparison of implied volatility surfaces, rather than observations on them, is thus the valid choice for this thesis.

While an evaluation of implied volatility surfaces is a possibility, Hypothesis 2 can also be answered by determining which model results in a better option price. Since the hypothesis only refers to superiority of the ANN-based model over traditional models, not all models previously discussed need to be evaluated here. Instead it is sufficient to evaluate the ANN models and the reference model using the best volatility forecast, the surface derived by the fitting procedure above and applied to the option price. This has the additional benefit of limiting the impact of sparse data on the surface as more liquid regions also have more observations attached to them. A surface-to-surface comparison may cause problems of both theoretical and practical nature.

Regardless of which network is the best to be chosen for this purpose, a single reference model for the option pricing needs to be chosen. Several possibilities exist:

- The (modified) Black-Scholes formula could be used. Despite the options of interest being equity options with American-style exercise, the formula provides a simple reference point that is likely close to the actual option values. As pointed out previously, the European-style option price is the lower boundary of the American-style option with otherwise identical features. This fact together with the absence of dividends, which would justify early exercise for calls, are the main reasons for this particular choice by Lajbcygier (2003b), for example. The argument is weaker for pure equity options due to the presence of dividends.

- The CRR model is feasible and a common choice in cases of American-style exercise. It has the previously discussed benefit of pricing the dividend stream correctly as well. It is computationally expensive however, and it's iterative nature requires additional choices with respect to the number of steps. The additional accuracy of the model as such is easily undermined by a lack of convergence due to too few steps. The trade-off between accuracy and speed, which *is* a consideration in a practical context, is not easily resolved.

- The Haug-Haug-Lewis (HHL) model (Haug, Haug, and Lewis, 2003), which is also applicable in the presence of discrete dividends, is even more costly with respect to processing time.

- If the discrete dividends are considered proportional to the price, i.e. they can be expressed as a yield but paid at discrete intervals, the method by Villiger (2006) can be used to approximate the price of an American call.

- Finally, approximation models such as that by Barone-Adesi and Whaley (1987) is used frequently in practice according to Haug (2007), it allows for a cost-of-carry to be specified as a continuous yield to allow for the inclusion of dividend benefits. The same author also notes that the Bjerksund and Stensland (2002) model is an even better approximation, it is a suitable alternative to Barone-Adesi/Whaley one with the same simplifying assumption.

With the consideration of dividends being one of the contributions of the research, the first of the options is not a suitable choice. Among the remaining options, the choice depends partially on what assumptions and simplifications should be made. While the CRR approach is the most realistic, it is only beneficial if the payments are known. While this can be assumed for research purposes, especially in countries and for industries that offer comparatively stable dividend streams, the model by Villiger (2006) may be preferable if dividend payments are not known or preference is to be given to the assumption of a constant yield. On balance, and in large part due to the comment on its frequent practical use[17], the Bjerksund and Stensland model is used for this thesis. This requires a conversion of the dividend payments to a cost-of-carry rate.

The notation is modified to reflect the choice resulting in C^{ref} referring to the Bjerksund and Stensland approximation for the price of an American call option with the usual parameters including the dividend yield.

Finally, Figure 3.6 shows the additional model required for Hypothesis 2. It is designed to answer the question whether there is a benefit to splitting the problem into two parts, one for volatility modelling (as above) and the other for option pricing, or whether the two are best integrated given a parsimonious specification:

$$C^{\text{ANN}} : (\sigma_{110}, \sigma_{60}, \sigma_{20}, \sigma_5, r_t, r_t^2, M, T, q) \mapsto C. \tag{3.6}$$

The innovation is the model replacing C^{ref} (in Figure 3.1). It represents an option pricing model as previous models. Instead of supplying a volatility forecast,

[17] The additional benefit and direct consequence of its popularity is that the model is already implemented in MATLAB.

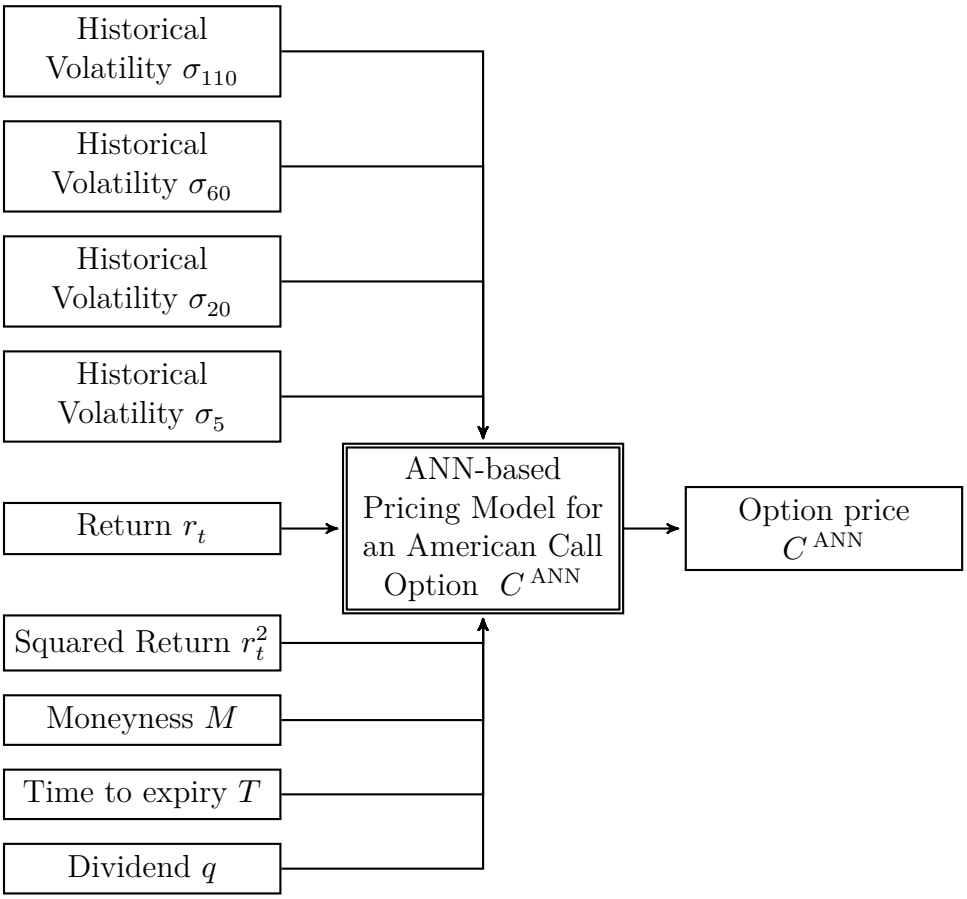

Figure 3.6: ANN-based Pricing Model for an American Call Option

the inputs used for the volatility forecast in $\sigma_{M,T}^{\text{ANNs}}$ are supplied directly to the pricing function. Here another benefit of the use of the various approximation models, including the one used for the reference model, becomes apparent. Just as in the case of the homogeneity hint used even for volatility forecasting, here the use of the dividend yield allows for a direct comparison of models. Were the CRR model be chosen for the option pricing, any over- or under-performance could not be attributed easily to the pricing function as such. Rather, it may be the result of more realistic or more commonly used representations of the dividend stream that are driving the performance. By using the yield in both, the issue is avoided altogether although the same would be true if the discrete dividend yield were used.

Supplying the dividends directly, in the form of discrete payments, is not easily achieved as the representation of this information is not clearly defined. A dividend stream consists of two vectors, one of dates, one of payment amounts. For each sample, these vectors may be of varying length. Such models would require additional assumptions and conventions regarding the minimum and maximum number of payments to be modelled and permitted in the network as well as a decision of how fewer than the maximum number of dividends are to be represented to the network.

The models designs, as they are chosen avoid these issues entirely and permit direct comparisons and evaluation of models.

All pricing models presented so far differ in the way they represent volatility, as a single input or as a combination of constituent attributes. Volatility is only one factor, however. The typical definition of such formulas includes, at a minimum, the traditional 5-tuple: current price, strike price, the risk-free rate, time to expiry, and volatility. Thus the remaining four inputs need to be supplied as well. As Garcia and Gençay (1998) and Garcia and Gençay (2000) point out, however, it is beneficial to represent the first two parameters as a ratio, rather than as individual attributes. Instead of supplying current and strike price as individual values, moneyness is calculated and used in the formulas. Conceptually, this means scaling the option not to one unit of ownership but to one unit of money at the time of exercise. This also means that the option price needs to be divided by the strike price K. No further adjustments are needed. In particular, since volatility represents volatility of a rate of return and return is independent of the level, no adjustment is required.

The latter points to an important assumption in the process presented by Garcia and Gençay (1998) and Garcia and Gençay (2000). It assumes homogeneity with respect to the price. The stock price level in this case is assumed to be unrelated to the return and pricing process. This is a natural assumption considering the principles of financial modelling covered in the previous chapter. In particular, the number of shares outstanding is arbitrary and so is the fractional value. It is the combination of units held and the price of a single unit that matters. There is one important consideration, however. Many market participants believe that there are price levels of some significance, these are treated as support and resistance lines or in other more complex ways. If their views were true or if they believed they are, there may be, if briefly, a relationship between the price and the return or at least it would offer an opportunity for such a relationship to emerge. In accepting the homogeneity hint as proposed by the authors and used by researchers subsequently, one rejects the existence or significance of that particular effect.

3.2 Data Scope and Sources

All data related to options with the exception of that set covering the underlying prices and index membership for the data filter (see below) has been retrieved from the Thomson Reuters TickHistory Database provided by *Securities Industry Research Centre of Asia-Pacific (SIRCA)* (2010–2012). The data was requested in two stages in 2010 and 2012 for an extended set. All equity and equity-linked options data was retrieved including reference data, which covers symbology and reference data changes. The latter refers to the changes in contract details as they occur. Individual transactions were requested and the request verified. A number of securities could not be retrieved though this is almost exclusively the result of failing to retrieve test data which need to be excluded from the data set in any case. This test subset does not represent actual market data and does consequently not qualify for inclusion. The few remaining failed series relate to instruments that would not have qualified for inclusion under the rules (see below for the definition of the candidate set).

The data covers individual transactions and in the case of the first data set also quotes. The quotes were discarded along with reports of implied volatility. The implied volatility reported by the database refers to regulatory implied volatility, which according to Australian Securities Exchange (2010) is published in

response to price changes for the purpose of informing participants about margin requirements and not for pricing purposes.

The latest reference data and transaction information was retained and converted from text format to binary representation as explained below.

Equity data was retrieved from two sources but both through *Securities Industry Research Centre of Asia-Pacific (SIRCA)* (2010–2012). Current prices were retrieved at 5-minute intervals from Thomson Reuters TickHistory including additional data relating to the securities and their symbology. Although the database also provides the dividend and corporate action data, it was decided to use the CRD database from *Securities Industry Research Centre of Asia-Pacific (SIRCA)* (2010–2012) and use the pre-calculated dilution factors to adjust returns. The use of the existing formula and implementation allow in principle for replication and it is assumed that the research character of the database implies additional prior scrutiny of the adjustment methodology as well as the underlying data.[18]

The equity data was restricted to the members of the Standard and Poor's (S&P) ASX 20 index membership on June 30, 2007 excluding the special Telstra equity security but including the standard Telstra shares (TLS). The data covers the sampling period of July 2000 to June 2011. During processing, the data was synchronised, i.e. equity and options data aligned. Since the equity data constitutes the most constrained set, it also defines the overall data set resulting in samples only over that period and only for members of the index on the given date.

Unlike some research conducted in the past, there has been no attempt at fitting models repeatedly. Consequently, no rolling-window subsets of data were created and no update to the candidate set, i.e. the investment universe to which the models are applied, needed to be made. It should be noted that the indices published for Australia are designed such as to have few constituent changes among other criteria (Standard and Poor's, 2011).

[18]In research unrelated to this dissertation, the author's supervisor and the author investigated pricing differences resulting from varying adjustments or base data across databases to decide on a particular source of data for research. Pricing differences did exist but they were usually quite small and did not raise concerns in that context.

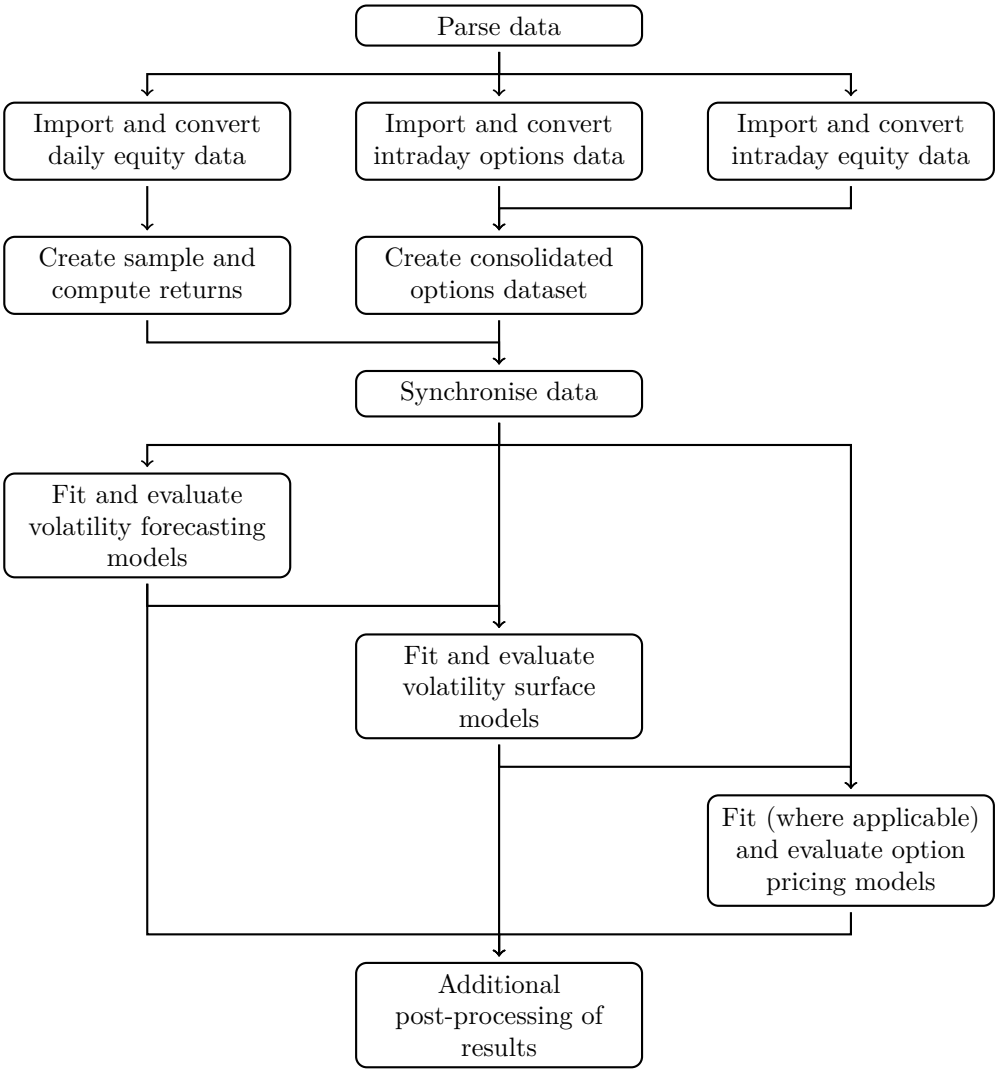

Figure 3.7: Schematic Overview of the Model Development and Processing Workflow

3.3 Model Fitting and Testing

3.3.1 Simulation Implementation

The simulations were largely implemented in MATLAB; in particular, the training and evaluation of neural networks was done using the Neural Network toolbox.

The data was imported from the text files retrieved from the data providers and converted to binary format to avoid multiple parsing and cumulative rounding errors. Preprocessing focused on the removal of duplicate observations.

The following steps are necessary to process data, implement, and run or simulate the models (see Figure 3.7 for a schematic illustration):[19]

1. Parsing and importing of data, including removal of unnecessary and duplicate entries;
2. Creation of a complete equity subset including returns;
3. Consolidation of option reference data, such as strike price, time to expiry, etc. and transaction data to yield a complete option pricing data set as a set of observation vectors;
4. Synchronisation of daily and intra-day data through the use of lookup indexes;
5. For each hypothesis:
   a) Splitting of the data set into in-sample and out-of-sample subsets so all fitting (statistical models) and training (ANNs) can be done within the in-sample set and all evaluation in the out-of-sample set;
   b) Fitting and applying of the volatility forecast models (Hypothesis 1), the volatility surface (Hypotheses 1 and 2), and option pricing models; networks are trained using the in-sample data set only (Hypotheses 1 and 2);
   c) Evaluation of the resulting models, applying the volatility forecasting models, the deterministic volatility (adjustment) function (surface model), or networks using out-of-sample data;
6. Computed summary statistics generating and formatting tables and plots (see Chapter 4).

The first two steps are largely self-explanatory and mainly aimed at enabling further steps and improving processing time in later stages. It is critical that during synchronisation no bias is introduced. Time-series offsets need to be chosen such that at any point in time a decision is made (or is simulated to be made), it is based only on available information, this implies strict time precedence.

[19]The steps do not represent a strict sequence due to varying dependencies of the selected models and some steps were done out of the specified order. For example, comparisons between in-sample and out-of-sample results can be made without reference to competing models and were thus partially done at the time the models were fitted.

There is one exception in the case of option pricing and specific to derivatives markets. Given that theoretical pricing models are based on arbitrage arguments, the process is conceptually symmetrical. The no-arbitrage situation can be created by the equity price moving in response to the option price or vice versa. Given the high liquidity of the equity market and the more restrictive regulatory environment in derivatives trading, the standard time-ordering was used regardless. The price of the underlying was observed and the price of the option based on it. The reverse was not investigated or simulated. Under the assumption that equity markets drive derivatives markets, no look-ahead bias results from this treatment. [20]

The data was then split into in-sample and out-of sample data. The same cut-off point was used for the split as for the index membership date. This treatment is equivalent to assuming that the decision to use the ANNs was made on June 30, 2007 and that all simulation as well as the investment universe they are based on, were run at that point in time. The date is in all likelihood the most conservative of cut-off points as it splits the data set not only into in-sample versus out-of-sample but also marks the end of a particularly good and stable economic period and the beginning of the GFC.

Using the in-sample data set only, the GARCH models were fitted using standard parameters and a specification of GARCH(1, 1). The GARCH volatility model is then synchronised again with the main data set. The resulting set is a full in-sample set.

An adjustment is needed with respect to volatility. As pointed out in the literature review, the volatility in the pricing model needs to be stated in terms of annual volatility requiring the usual adjustment. This can be done during volatility modelling or during pricing. As a matter of convention and convenience, the adjustment is made before reporting volatility forecasts, i.e. volatility forecasts represent annualised p-period forward volatility in all tables and figures.

The simulation, time series creation and synchronisation steps were then repeated using the fixed model parameters for all models (including GARCH, IV, volatility surface, ANN). The results were stored for later retrieval, statistics calculated (see 3.4 for evaluation methodology and Chapter 4 for results) and tables as well as figures created.

[20] This approach is not sufficient if it is assumed that there is an interaction, i.e. a mutual influence, and it is not heavily biased in favour of the equity market.

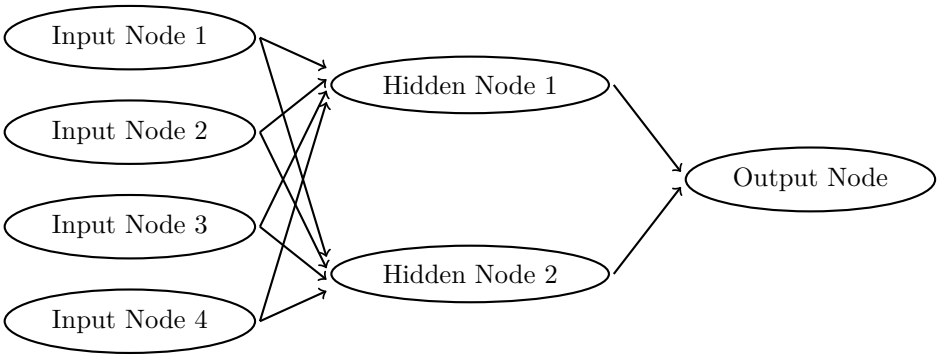

Figure 3.8: Schematic ANN Architecture

3.3.2 ANN Training

The training phase of the networks is particularly difficult given the large number of open questions that remain in the area, especially with respect to architecture, learning algorithms and their parameters, and finally the training process. As was alluded to before, this research follows Vanstone (2005) as well as Vanstone and Hahn (2010). Due to different objectives and software, it needs to be adjusted.

The original process requires the definition of input parameters, preparation of data (the software used in the research performs standardisation of the input range automatically), and initial network design choices. The network starts with $\lceil \sqrt{n} \rceil$ hidden nodes in a single hidden layer for n input variables (Figure 3.8 provides an example of the starting network architecture for a $n = 4$ network). A network is trained and the in-sample results stored along with the network parameters. The process is repeated until the in-sample results fail to improve. Each learning cycle is limited in length by a maximum number of epochs, i.e. data presentations to the network, without improvement. The authors also provide some guidance regarding the training parameters, which determine the speed with which weights are updated and learning takes place but also potentially limits the search for improved performance by especially fast conversion to a local minimum.

The problem of local minima is particularly acute for ANNs. Not only is it likely to find a local minimum instead of the global one but the lack of explanatory power limits the understanding of the resulting model. Decision makers cannot, therefore, detect such problems even when they otherwise have considerable domain knowledge and intuition for the problem. The suggestion is to train many networks and choose the best with respect to some metric.

The problem is avoided to some degree by following the process discussed before. The approach already provides for iterations and thus for opportunities to escape local minima. This comes at a price of choosing a less parsimonious model than needed. An additional iteration may not improve the results because of the additional hidden node but rather because of the different starting point. This network would be chosen even though the same or similar result could have been found with a smaller number of hidden nodes. The process does, however, provide an opportunity to arrive at fairly parsimonious models by its design compared with some other alternatives.

The Levenberg-Marquardt back-propagation algorithm was used in conjuction with the MSE performance metric. The hidden layer used the sigmoidal and the output layer the linear transfer function.

The methodology by Thomaidis, Tzastoudis, and Dounias (2007) discussed before was considered as well but ultimately rejected for the following reasons. Firstly, there appears to be little research following this specific process. Secondly, it is not clear where to start. While the different steps are outlined, it is not clear if starting with one neuron or variable instead of another one leads to different results and how such problems can be detected or corrected for.

Thirdly, the process does not address the issue of local minima and thus additional training is required in any case. Finally, it is so far unclear if the methodology is sufficient to account for non-linearities both in the process and in the statistics and metrics used. It should be noted that the use of well-defined statistics and a clear process is extremely attractive from both a research and a practitioner's perspective and should be investigated further.

3.3.3 ANN Model Selection

The iterative approach requires two steps in addition to the network design and learning parameter choices. Firstly, the question arises which network to choose for any particular estimation or forecasting problem. Secondly, it requires a determination of whether the particular model chosen is sufficiently better – with respect to the same or a different metric – than an existing competing model. While the latter is discussed in greater detail in the next section, the former is largely straight-forward.

Once a set of networks has been trained and a stopping criterion (for the in-between models stopping, not the within-model training stopping) applied, the

resulting model needs to be constructed. If multiple networks are trained for any particular design, this choice is a dual one of choosing between network designs and between within-design networks. Since the simple process by Vanstone (2005) is used, this complication does not apply.

The same process also requires the choosing of the best network from the various designs. Assuming the error metric is appropriate for the application domain, the best network provides the best fit to the data. It is also possible, however, that such an approach leads to overfitting and that the error metric used for stopping and evaluation is not consistent with the ultimate use of the model. The metric for design choice may, therefore, be different from the one used for training. Following Vanstone (2005) and other researchers as discussed in the literature review, the network design with the best performance is used and no additional distinction is made with respect to error metrics.

These issues are common and not unique to ANNs but are potentially more significant given the universal approximation ability. In principle, the best-performing network may also be the one most sensitive to structural changes and may perform particularly poorly out-of-sample. It is not a foregone conclusion to use the best network if such an effect is anticipated. Further, the problem is not eliminated through regularisation or similar approaches as these are still within the given data set. It may thus be beneficial to determine, where possible, such sensitivity and use it in determining the best of the chosen networks, best not with respect to an error metric but best with respect to the wider set of performance constraints.

Even in the presence of such analysis, it is not clear that a single network even should be sought or whether a combination of networks along the development path may be superior. Combining forecasts is a common approach in statistics and econometrics and Poon and Granger (2003) point out one application of a learner regression in a purely econometric framework.

No attempts have been made in regards to the analysis of sensitivity or combining networks in option pricing previously and the number as well as the interdependency of choices require a separate analysis of the issue. Instead of introducing several innovations at the same time, only a single network is chosen, the best with respect to the same error metric as is used for training using the process by Tan. However, the total number of parameters will be reported along with the error metrics to enable at least a basic discussion of the complexity of the networks, which is the principle cause of overfitting. This allows for the determination of

measures of parsimony and aids in the second step, the question of whether to choose the ANN or a standard model as discussed next.

3.4 Model Evaluation and Comparisons

Measuring forecasting and pricing errors is central to model development and evaluation. They are used for the following purposes:

- The regression error of the network needs to be determined for each sample presented to the network during training in order to determine how the weights need to be modified in order to minimise the average error (see 3.3.2).
- If several networks are trained simultaneously either to determine a suitable network architecture or to avoid the issue of local minima, an error measure is needed to choose one from amongst those models (see 3.3.3).
- A model needs to be chosen among a set of competing (benchmark) models or a combination of such models. This can be based fully or partly on the forecasting or estimation error.

Common to all is the need to determine what metric to use and the data it is applied to, i.e. which error of what. Table 3.1 shows the standard definitions of common error formulas as found in the literature (see Chapter 2).

Due to the frequent use in (linear) regression and generally favourable statistical and numerical properties, it is not surprising that the MSE is frequently used

Acronym	**Name**	**Formula**
MSE	mean squared error	$\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2$
RMSE	root (of) mean squared error	$\sqrt{\text{MSE}}$
MAE	mean absolute error	$\frac{1}{n}\sum_{i=1}^{n}\lvert x_i - y_i\rvert$
MAPE	mean absolute percentage error	$\frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{x_i}{y_i} - 1\right\rvert$

Table 3.1: Definitions of Error Metrics of Forecast or Estimate x_i and Observation y_i

during ANN training and as discussed above for network selection as well. It does have two main problems, however, it penalises large errors significantly (due to the squaring) and it is difficult to interpret with a squared unit. This is similar to a problem found in expressing volatility as σ^2 and for the same reason, the positive root is often reported instead. While MSE is used for network development purposes, the results reported in the following chapter will therefore represent the RMSE instead.

The use of the absolute measures and in particular of MAPE is frequently found as it is not sensitive to the level of the estimate nor does it focus as much on large errors, which may be due to outliers. Neither issue is likely to exist in the methodological framework used in this thesis: The use of the homogeneity hint combined with the data preparation steps are expected – and indeed are designed – to result in a data set that is not sensitive to price levels. In the case of volatility forecasting, the principal variables are rates and while large differences may exist within as well as between time series, they are not expected to be so frequent and so significant to cause problems. It can also be argued that the greater penalty for large errors is beneficial in the case of option pricing. This is a security class typically used for hedging, large unexpected events are thus of particular concern.

All average error metrics are reported, principally to allow for comparisons with prior studies but only the MSE is used for decision making in the presence of competing models. The model with the lowest error is the preferred one.

Even with the measure decided, the question remains the error of what is to be calculated and used. In the case of the first and last hypotheses, no distinction need be made. The variable being forecast (estimated) is also the variable of interest, the error is thus between the forecast (estimation, respectively) and the actual observation.

When evaluating the remaining hypothesis, the question is subtler. The question is if the volatility forecast results in a better option price. The question thus is whether the goal should be to minimise the forecasting error (all forecasting errors), to minimise the pricing error resulting from the forecasts (all pricing errors), or if the forecasting error is to be minimised during training but the model chosen based on pricing errors (mixed metrics).

The first allows for a consistent treatment of error metrics between the first and the second hypothesis while the second allows for a consistent treatment between the second and the third (with respect to error measures). The third choice appears particularly suited to accommodate the competing goals. While super-

ficially advantageous, the alternative is not particularly convincing, however. If the goal is to ultimately minimise the pricing error, then the forecasting error should be used for training purposes. Unless the two metrics are such that they always result in the same outcome, in which case the decision maker would be indifferent between them, choosing the mixed approach will yield pricing errors not smaller than the all-pricing-error approach (ignoring for the purpose of the argument the issue of local minima).

Therefore, the question of what to measure the error of needs to be answered by focusing on the specific research goal. The motivation of the research is to arrive at an appropriate decision making process for option pricing[21]. This supports the use of all-pricing errors in principle but not from a purely procedural perspective.

As pointed out in the discussion of time series alignment and look-ahead bias, the view taken in this thesis is that the underlying security prices are observed and the option prices built upon them on the basis of a volatility forecast. The question to answer is whether the volatility forecasting can be improved to result in better pricing. It is thus necessary to compare the volatility forecasting with a volatility surface model and both compete for use in pricing. Since both are volatility forecasts they need to be evaluated on this basis using the same error measures.

It should also be noted that the pricing error measure is not rejected here on the basis of inconsistency between the benchmark model and the ANN model. The benchmark model is based on implied volatility and thus implicitly on option prices. It would be valid to compare it to an ANN that is trained to minimise pricing error resulting from volatility forecasts.

All error metrics discussed are only useful for comparing relative benefits of models. Two additional questions need to be answered, whether the differences are statistically significant and what their benefits are relative to the problem or to the competing model.

Since the objective in all cases is to determine if the mean error is significantly different from another set, the ANOVA framework is used. Based on the variance of the sample, it allows for the determination of the probability of observing difference in means by chance only. Where applicable ANOVA is used simulta-

[21] This point implies that either approach is valid depending on one's perspective. If the focus is on outcome rather than process, the use of all-pricing-errors is likely beneficial as it offers the opportunity to learn the 'right' forecasting model strictly for pricing. The view taken here is that the volatility forecast drives the pricing and the regression is thus biased towards the beginning of the process rather than its end.

neously for multiple models, rather than pairwise, to reduce the compounding of errors.

ANOVA results are reported but not used for model choice. Instead they are used to draw conclusions at the end not guide in the development.

3.5 Theoretical and Practical Limitations

The methodology introduced in this chapter is designed to answer the particular research questions raised in this thesis. Several limitations apply, however, both with respect to answering these questions and to applying it to broader research questions.

Regarding the specific research questions expressed in the hypotheses of Chapter 1, the methodology is limited to a step-wise model development process. This sequential approach simplifies development and decision making ensuring a reasonable basis for the final choices and resulting beliefs but does not necessarily account for all the complexity inherent in the problem.

In particular, when applying the result of one hypothesis as an input to the next, the non-linearity of the problem may result in the best choice of an input model to be a suboptimal choices of the integrated model. This may be the result of varying sensitivity of model outputs to parameters as in the discussion of error measures. An integrated approach would be preferable but is likely to require an investigation of all – or at least many – combinations of models for input parameters. This is not unique to the presented methodology but a general problem in option pricing as well as other areas of financial research and partly motivation for this research.

Not only would such research be considerably more complex and time-consuming, it is also more likely to be biased, in particular by means of data mining, and requires more sophisticated statistical tests to avoid inference errors. The present methodology is not free of those but being limited to a single uni-directional development process using decision criteria as they have been used in the prior literature, the current process is more robust.

Another limitation results from the large number of choices of variables, sampling method, parameters, and benchmark models. While each has been defended based on their use in previous literature or common practice and discussed in the context of the methodology development, generalisation beyond the existing data set is not fully defensible. This is especially true with respect to volatil-

ity modelling and when attempting to extrapolate to other markets, whether by geography, liquidity, underlying assets, or other criteria.

This limits the research outcomes strictly to the securities, and a regulatory and economic environment similar to the one studied. The use of out-of-sample data allows for generalisation over time to some degree. It is the process, however, that is central to the research, which may also prove useful for practitioners attempting to replicate and extend this methodology to suit their particular needs.

More significant limitations exist when such attempts beyond the data set and research goals of this thesis are made. No attempts have been made, in particular, to improve hedging performance or to simply evaluate it. Some error metrics exist for this purpose (see 2.5.1) in part as the general ability of ANNs has been demonstrated. If hedging is central to the research or trading activity, the methodology needs to be adjusted accordingly. The research outcomes reported in the next chapter do not necessarily apply to such questions.

Secondly, if alternative benchmark models exist that are considered superior for a particular application, these need to be added at the appropriate stage. Benchmark models were chosen largely based on prior literature and consistency with the stated research goals (thus excluding, notably, local volatility models). They do not necessarily represent the best available model for the problem but a common and well-understood alternative.

Finally, the process cannot help to explain why ANNs typically improve performance. The methodology is only designed such as to provide a first step towards answering which area of modelling, i.e. volatility forecasting, volatility surface modelling, or option pricing, improve results but not why or how such improvements are achieved.

Chapter 4

Analysis of Data

4.1 Overview

In this chapter the intermediate statistics of the simulations and their results are reported. The presentation is structured along the development process of the various models, i.e. from left to right in Figure 3.1.

The chapter is organised as follows: general data characteristics are presented, including a number of pre-tests required for analysis. This is followed by training and testing results of the various networks and error terms of the parametric models used for comparison for volatility forecasts, volatility surface models, and option pricing models.

For each stage, the data characteristics and fitting of data is reported. This is followed by a comparison of models and a comparison of performance between in-sample data and the out-of sample set.

When individual securities are reported, their latest symbol in the sampling period is used.

4.2 General Characteristics of the Data Set

The data set consists of broadly two subsets as explained above: equity data, and option data. Each also contains reference data such as the names of securities, adjustments to be made and the contract information for options, i.e. strike price, expiry date.

These data sets were imported and synchronised with adjustments for changes of the symbol and other related data. Further subsampling was applied at the appropriate stages and is detailed below.

Security	$\boldsymbol{\sigma}$	**ARCH-test**		**Q-test**	
		p	**statistic**	**p**	**statistic**
AMP	0.0202	0.6531	0.2020	0.0009[b]	10.9630
ANZ	0.0121	0.8725	0.0257	0.6858	0.1637
BHP	0.0176	0.0007[b]	11.3656	0.6787	0.1716
BXB	0.0185	0.5725	0.3185	0.0653	3.3973
CBA	0.0101	0.0000[b]	59.2177	0.0531	3.7411
CGJ	0.0148	0.0411[a]	4.1720	0.9334	0.0070
FGL	0.0120	0.0241[a]	5.0861	0.8393	0.0411
MQG	0.0151	0.0000[b]	39.6275	0.0003[b]	12.8271
NAB	0.0114	0.0117[a]	6.3540	0.9928	0.0001
QBE	0.0259	0.0000[b]	229.9609	0.0000[b]	44.2125
RIN	0.0172	0.0724	3.2279	0.3508	0.8706
RIO	0.0160	0.0001[b]	15.6079	0.3135	1.0158
SGB	0.0100	0.1367	2.2147	0.9969	0.0000
SUN	0.0128	0.0000[b]	116.8904	0.1441	2.1334
TLS	0.0117	0.0000[b]	120.7925	0.3325	0.9390
WBC	0.0106	0.0000[b]	28.5261	0.2444	1.3552
WDC	0.1836	0.5216	0.4108	0.9804	0.0006
WES	0.0138	0.0000[b]	18.1731	0.4774	0.5049
WOW	0.0116	0.0000[b]	44.2569	0.9264	0.0085
WPL	0.0157	0.0000[b]	72.1385	0.4209	0.6478

Table 4.1: Descriptive and Test Statistics for Underlying Equity Securities for the In-sample Period. Significance at the 5% (1%) level is indicated by [a] ([b]).

The data covers the period 2000–01–01 to 2011-06-30 and is split into three parts. The first year, 2000, is used to compute cumulative dividends and the resulting yield. It is also used to compute initial historical volatility if applicable. Furthermore, the first year's 252 trading days were used as the basis for the annualisation in all subsequent years. The following period up to and including 2007–06–30 was used to fit models or train the networks, respectively. The remaining data, including the GFC period was used for out-of-sample testing.

For the purpose of volatility forecasting, the problem consists of a set of time series. Each needs to be analysed separately for the historical volatility and GARCH model. The latter is only useful in the presence of anomalies. In particular, Engle's ARCH test and the Ljung-Box-Q test are applied to the series to test for heteroscedasticity and residual auto-correlation, respectively. The results can be found in Table 4.1.

The null hypothesis of no heteroscedasticity, is frequently rejected, as is the null hypothesis that no autocorrelation exists at lag 1. As can be expected, the confidence in the various tests varies but it is equally obvious that the series do not meet the conditions normally defined for time series. While it would be possible to fit a separate model (of separate classes) to each series, any summary conclusion requires the testing of defined models. Therefore, no varying lags were used and standard GARCH(1, 1) models fitted to the in-sample data and subsequently applied to out-of sample data.

Due to the nature of the option pricing data set, descriptive statistics are not meaningful. The data set is a subset of all qualifying options, i.e. across strikes and maturities for qualifying underlying securities. Only call options were considered and only reported trades, rather than including bids and asks. The following discussion of specific networks also covers the definition of 'qualifying' as there are minor variations across the models.

4.3 Volatility Forecast Evaluation

Volatility forecasts form the largest subset of models by type. A model is fitted to each security's time series with the exception of the ANN, which is fitted to the panel data, i.e. over time and across series. There were 31 109 observations in the in-sample set for the ANN, which counts individual returns rather than aligned and padded series, and 17 282 observations in the out-of-sample set.

The two historical volatility models, whose parameters were chosen a priori, were applied to the data set without any modifications or limitations other than that the first observation was chosen such that it started one window length (look-back period) from the first observation. Missing data (due to trading inactivity) was ignored and the volatility estimate is thus based on the available data during the period and as such with fewer observations than would be expected had there been no trading inactivity. Given the high liquidity of the chosen index's constituents, this is not a major concern. A summary of the error terms of the volatility models can be found in Tables 4.2, 4.3, 4.4, and 4.5.

It is evident that the forecast errors for these models differ greatly. The reasons for these differences are not the subject of this research but it should be noted that a number of securities are not part of the data set throughout the whole sample period due to changes to their trading status. In particular, observations are missing at the beginning of the period for RIN and WDC, and at its end for

Security	**MSE**	**RMSE**	**MAE**	**MAPE**
AMP	0.046 409	0.215 428	0.114 905	0.455 103
ANZ	0.012 810	0.113 179	0.061 759	0.376 628
BHP	0.007 399	0.086 016	0.066 517	0.253 556
BXB	0.023 849	0.154 430	0.091 919	0.375 067
CBA	0.002 307	0.048 032	0.035 048	0.246 015
CGJ	0.012 849	0.113 353	0.079 456	0.404 967
FGL	0.003 621	0.060 176	0.045 630	0.263 175
MQG	0.007 904	0.088 905	0.067 332	0.325 238
NAB	0.006 350	0.079 685	0.052 554	0.336 557
QBE	0.165 122	0.406 352	0.145 280	0.471 545
RIN	0.019 066	0.138 079	0.090 678	0.520 687
RIO	0.004 590	0.067 748	0.053 082	0.230 020
SGB	0.002 089	0.045 701	0.034 016	0.238 276
SUN	0.005 909	0.076 871	0.057 495	0.345 141
TLS	0.004 352	0.065 973	0.049 328	0.305 120
WBC	0.002 230	0.047 222	0.037 559	0.245 667
WDC	0.001 773	0.042 113	0.030 013	0.185 251
WES	0.006 570	0.081 055	0.055 343	0.288 805
WOW	0.002 811	0.053 016	0.041 396	0.252 252
WPL	0.004 911	0.070 079	0.052 440	0.238 587

Table 4.2: σ^{HVL} Error Measures (In-sample)

Security	**MSE**	**RMSE**	**MAE**	**MAPE**
AMP	0.017 287	0.131 481	0.098 264	0.320 481
ANZ	0.016 250	0.127 477	0.095 893	0.292 509
BHP	0.028 462	0.168 707	0.113 394	0.301 902
BXB	0.018 219	0.134 977	0.100 205	0.300 722
CBA	0.015 399	0.124 094	0.092 821	0.331 675
CGJ	0.012 082	0.109 918	0.091 609	
FGL	0.007 372	0.085 863	0.065 477	0.311 015
MQG	0.065 423	0.255 780	0.163 911	0.347 660
NAB	0.020 252	0.142 309	0.101 347	0.320 014
QBE	0.014 649	0.121 035	0.093 862	0.320 653
RIN	0.005 167	0.071 883	0.063 062	
RIO	0.075 203	0.274 232	0.178 697	0.369 869
SGB	0.034 529	0.185 821	0.142 876	
SUN	0.029 430	0.171 552	0.119 433	0.307 283
TLS	0.007 512	0.086 673	0.066 530	0.310 338
WBC	0.014 313	0.119 636	0.090 693	0.310 063
WDC	0.020 157	0.141 974	0.097 207	0.294 923
WES	0.024 495	0.156 507	0.104 294	0.321 403
WOW	0.006 798	0.082 448	0.057 904	0.272 287
WPL	0.019 830	0.140 818	0.103 527	0.341 464

Table 4.3: σ^{HVL} Error Measures (Out-of-sample)

Security	**MSE**	**RMSE**	**MAE**	**MAPE**
AMP	0.060 641	0.246 254	0.116 348	0.421 352
ANZ	0.015 555	0.124 721	0.061 345	0.332 640
BHP	0.007 463	0.086 391	0.064 406	0.237 159
BXB	0.036 076	0.189 937	0.105 387	0.411 807
CBA	0.003 009	0.054 851	0.040 917	0.272 438
CGJ	0.017 641	0.132 821	0.086 911	0.426 005
FGL	0.004 301	0.065 585	0.050 564	0.282 137
MQG	0.011 690	0.108 120	0.075 176	0.330 861
NAB	0.006 734	0.082 063	0.054 081	0.322 541
QBE	0.158 775	0.398 466	0.115 437	0.305 474
RIN	0.022 132	0.148 768	0.084 845	0.419 757
RIO	0.005 577	0.074 678	0.058 513	0.248 220
SGB	0.002 718	0.052 130	0.038 383	0.255 050
SUN	0.007 177	0.084 720	0.062 310	0.348 837
TLS	0.005 899	0.076 805	0.054 105	0.315 450
WBC	0.002 874	0.053 611	0.042 134	0.264 211
WDC	0.002 229	0.047 217	0.037 430	0.247 467
WES	0.008 405	0.091 681	0.061 617	0.299 844
WOW	0.003 690	0.060 741	0.045 854	0.273 301
WPL	0.007 892	0.088 838	0.064 695	0.277 516

Table 4.4: σ^{HVS} Error Measures (In-sample)

Security	**MSE**	**RMSE**	**MAE**	**MAPE**
AMP	0.017 971	0.134 055	0.094 851	0.298 839
ANZ	0.017 461	0.132 140	0.093 274	0.286 768
BHP	0.019 501	0.139 646	0.092 628	0.255 326
BXB	0.024 831	0.157 579	0.120 295	0.349 621
CBA	0.012 682	0.112 616	0.078 896	0.271 314
CGJ	0.019 639	0.140 141	0.124 504	
FGL	0.009 419	0.097 052	0.074 733	0.325 052
MQG	0.064 407	0.253 785	0.156 490	0.311 511
NAB	0.020 292	0.142 451	0.100 066	0.304 185
QBE	0.017 347	0.131 707	0.100 336	0.349 427
RIN	0.001 619	0.040 233	0.032 287	
RIO	0.057 615	0.240 030	0.151 237	0.323 903
SGB	0.049 761	0.223 071	0.173 599	
SUN	0.031 511	0.177 514	0.113 052	0.289 859
TLS	0.010 917	0.104 482	0.075 755	0.331 569
WBC	0.011 509	0.107 280	0.077 395	0.257 963
WDC	0.019 442	0.139 433	0.096 610	0.281 898
WES	0.018 734	0.136 870	0.091 483	0.287 988
WOW	0.007 185	0.084 765	0.056 871	0.256 846
WPL	0.018 095	0.134 519	0.100 830	0.327 101

Table 4.5: σ^{HVS} Error Measures (Out-of-sample)

Security	**MSE**	**RMSE**	**MAE**	**MAPE**
AMP	0.070 312	0.265 164	0.158 077	0.720 824
ANZ	0.010 646	0.103 177	0.073 451	0.510 458
BHP	0.005 261	0.072 530	0.054 715	0.216 818
BXB	0.045 220	0.212 650	0.139 452	0.657 664
CBA	0.002 148	0.046 343	0.035 746	0.256 122
CGJ	0.013 810	0.117 516	0.094 558	0.569 583
FGL	0.003 141	0.056 045	0.044 376	0.264 481
MQG	0.007 560	0.086 949	0.069 187	0.350 923
NAB	0.004 935	0.070 253	0.048 952	0.345 262
QBE	0.145 906	0.381 976	0.141 370	0.506 385
RIN	0.013 715	0.117 109	0.078 333	0.445 397
RIO	0.003 988	0.063 152	0.049 457	0.221 775
SGB	0.001 847	0.042 977	0.031 779	0.225 496
SUN	0.005 317	0.072 915	0.054 429	0.339 646
TLS	0.004 404	0.066 360	0.048 802	0.334 940
WBC	0.002 022	0.044 967	0.035 201	0.235 933
WDC	0.083 837	0.289 547	0.284 264	1.995 850
WES	0.006 251	0.079 063	0.054 902	0.289 230
WOW	0.002 650	0.051 477	0.038 837	0.247 094
WPL	0.004 847	0.069 620	0.053 842	0.258 854

Table 4.6: σ^{GARCH} Error Measures (In-sample)

CGJ, RIN, and SGB. For these securities, changes affecting the whole market do not affect their average forecasting errors if the security was not yet or is no longer traded at that time. The identification of regime changes and the resulting forecastibility is beyond the scope of this research. Therefore, no exclusions or adjustments were made in this regard.

Following the application of the historical volatility model, the GARCH models were fitted. Any missing values were removed in this instance as well. Since historical volatility models and GARCH models treat missing values somewhat differently, percentage errors are undefined (due to a target value of 0) for different observations. The resulting models (see Appendix A for details on the model specifications) were applied to the time series in-sample and out-of sample. Tables 4.6 and 4.7 show the in-sample (out-of-sample) results for individual securities.

Security	MSE	RMSE	MAE	MAPE
AMP	0.024 155	0.155 419	0.128 849	0.500 790
ANZ	0.038 243	0.195 557	0.132 509	0.332 137
BHP	0.017 921	0.133 869	0.090 801	0.251 694
BXB	0.029 714	0.172 378	0.131 208	0.431 612
CBA	0.012 902	0.113 589	0.075 391	0.229 388
CGJ	0.009 699	0.098 486	0.083 717	
FGL	0.007 164	0.084 643	0.065 112	0.284 687
MQG	0.089 619	0.299 365	0.190 709	0.328 158
NAB	0.023 614	0.153 667	0.100 572	0.265 519
QBE	0.020 247	0.142 291	0.110 576	0.420 060
RIN	0.031 456	0.177 358	0.173 921	
RIO	0.049 783	0.223 121	0.136 985	0.270 578
SGB	0.045 998	0.214 471	0.160 264	
SUN	0.036 029	0.189 812	0.117 715	0.259 397
TLS	0.008 304	0.091 125	0.065 045	0.268 010
WBC	0.015 846	0.125 882	0.087 729	0.248 772
WDC	0.076 008	0.275 696	0.231 361	0.989 084
WES	0.016 655	0.129 055	0.087 327	0.277 711
WOW	0.006 099	0.078 093	0.051 504	0.250 133
WPL	0.028 195	0.167 913	0.116 362	0.350 788

Table 4.7: σ^{GARCH} Error Measures (Out-of-sample)

The network training followed the process described in the previous chapter. Prior to training, all variables are modified to fit into a fixed interval. Figures 4.1 and 4.2 show the original values and their ranges, outliers are compressed at the extremes.[22] These plots, like all other box plots in this chapter, show the central half of the data inside each box. It is unsurprising that the value range of volatility observations is different for the out-of-sample data. As would be expected, the GFC led to an increase of volatility. The general characteristics of the inputs are, however, similar.

The validation set was used to choose the best configuration. The best network configuration for volatility forecasting was the one with four nodes in the hidden layer, whose training record is shown in Figure 4.3. Figure 4.4 shows the error terms of the various models in-sample. This includes not only the validation but also the training set, i.e. the complete in-sample data set. While the mean error showed in the plot was lower, this was the result of the network's tendency to overfit and the error in the validation set was slightly higher than the next smaller

[22]Plots in this chapter typically treat the lower and upper 2.5 % value ranges as extreme values, which are compressed (with markers) or removed (without markers). Plots showing the full data set are so marked.

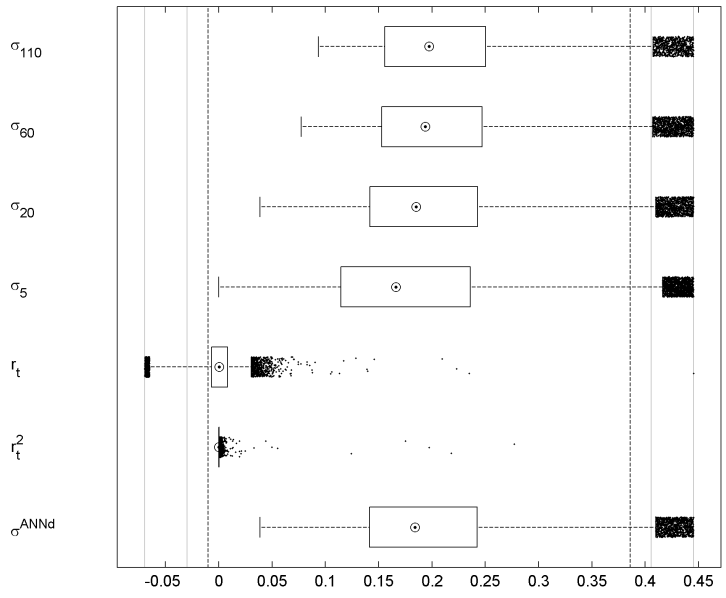

Figure 4.1: Characteristics of Variable Value Ranges (In-sample) for σ^{ANNd}

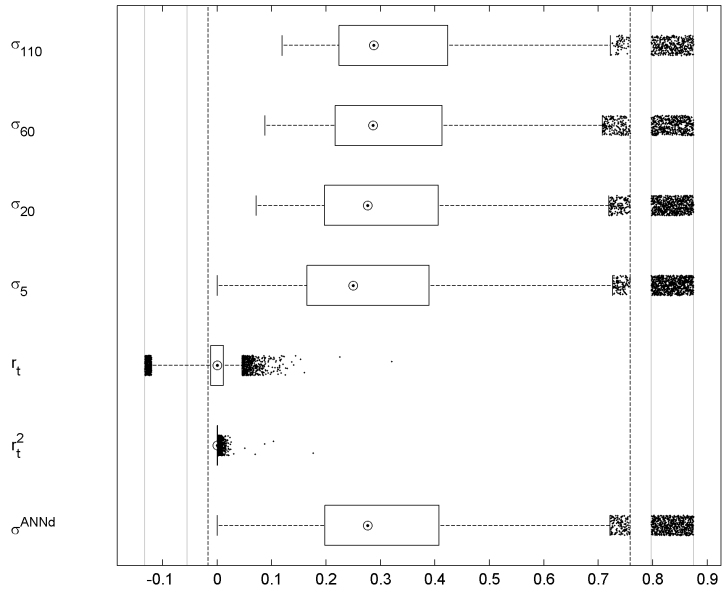

Figure 4.2: Characteristics of Variable Value Ranges (Out-of-sample) for σ^{ANNd}

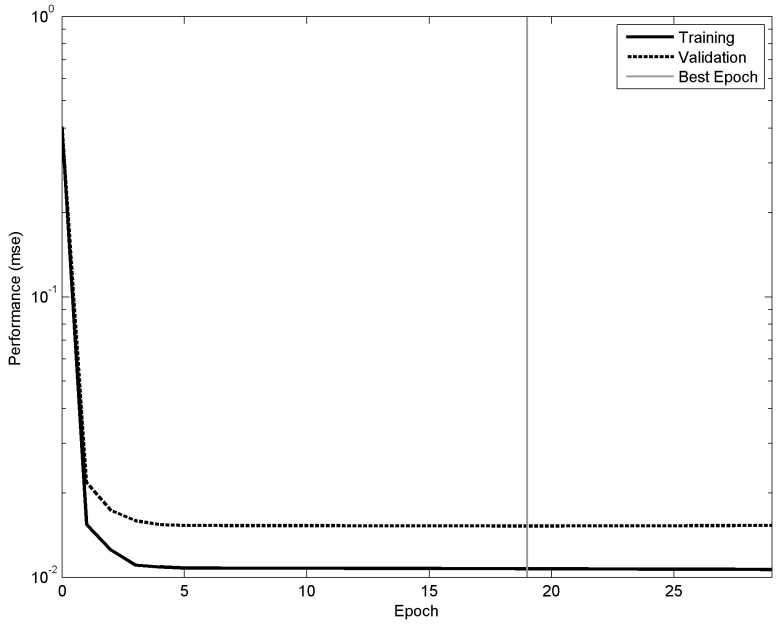

Figure 4.3: σ^{ANNd} Training Record

design. As a result the network with four nodes was chosen. However, it is evident that the differences between the network designs are very small.

An interesting observation can be made when examining the relationship between target values and network outputs. Figure 4.5 shows the two variables and a regression line is included. Figure 4.6 shows only the central 95 % of the value range (see Figures 4.7 and 4.8 show the same for the out-of-sample set).

In both instances, the network appears to average the forecast, overestimating at low target values and underestimating at high values. Whether this is due to a larger than expected number of outliers, whether this is a feature of the data set, or whether it is the lack of additional explanatory variables is not the focus of this research. It should be noted, however, that this behaviour may not be desirable for all applications. The distribution of observations in Figures 4.7 and 4.8 is particularly interesting as it not only shows this effect but it also appears to show a distictive shape. It may thus be possible to improve performance by transforming data or changing the modelling parameters. Since the use of the out-of-sample set for modelling decisions would change its character as an *out-of-sample* set, this line of research was not pursued further. It may be of general interest for future research, however.

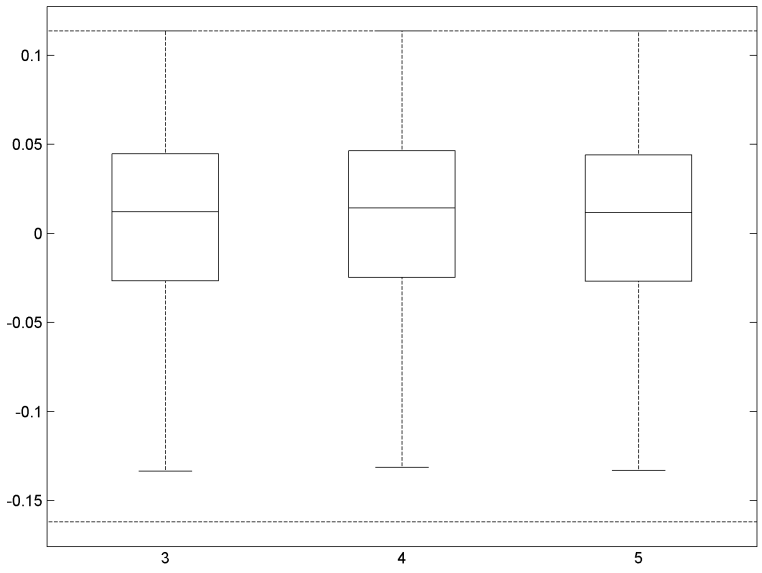

Figure 4.4: Error Distribution of Trained Network Architectures for σ^{ANNd} (In-sample)

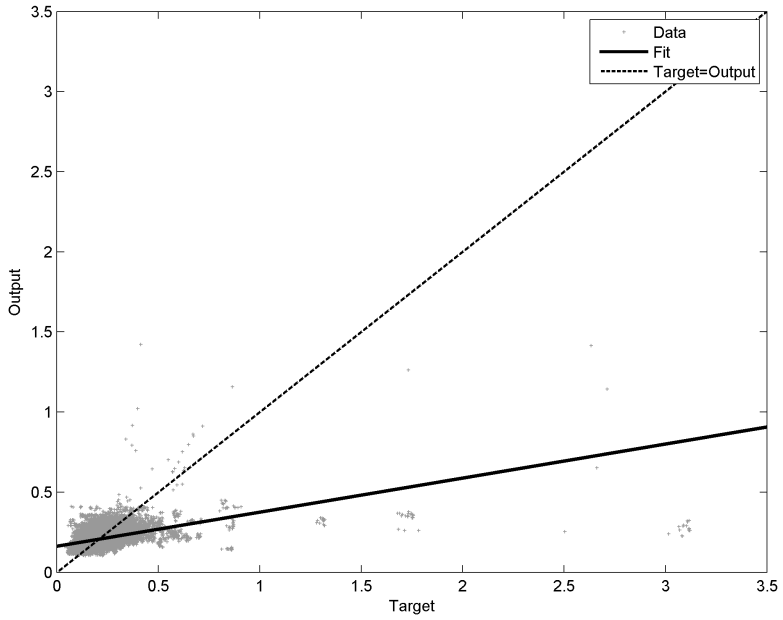

Figure 4.5: Target and Output Values for the σ^{ANNd} In-sample Data

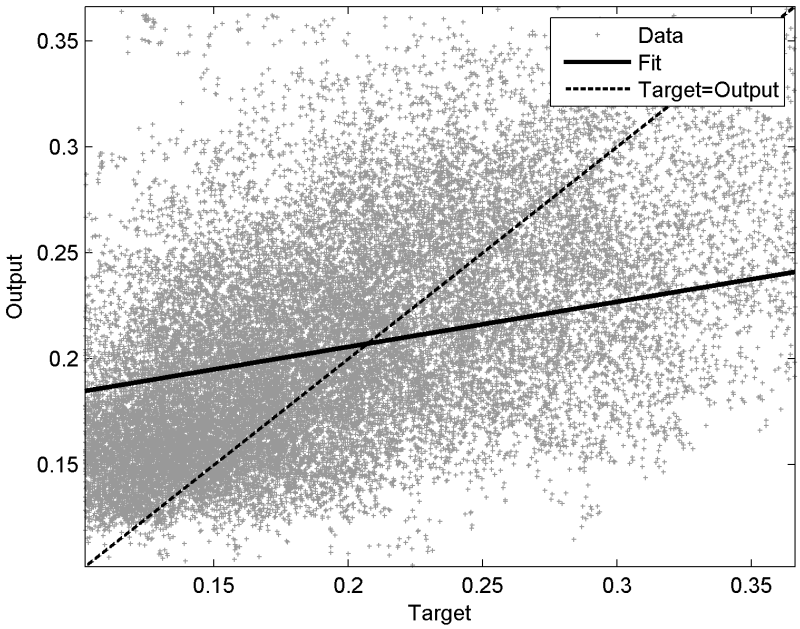

Figure 4.6: Target and Output Values for the σ^{ANNd} In-sample Data (Without Outliers)

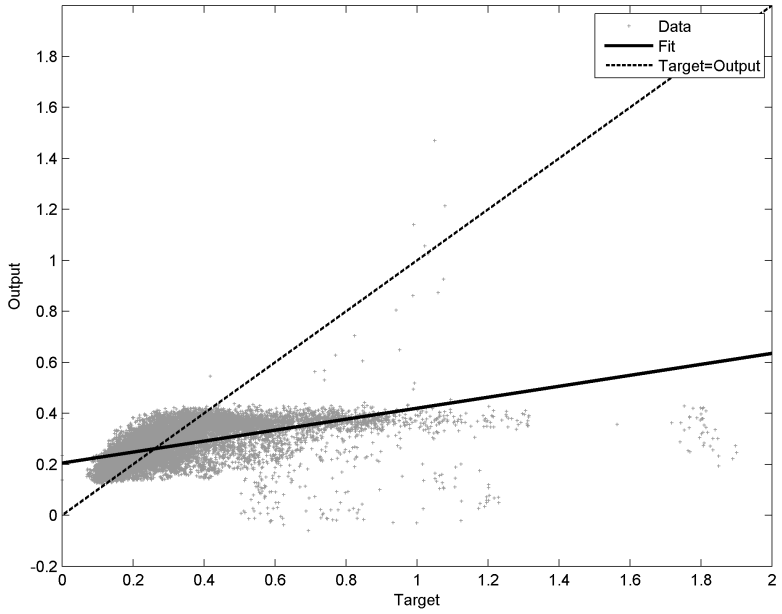

Figure 4.7: Target and Output Values for the σ^{ANNd} Out-of-sample Data

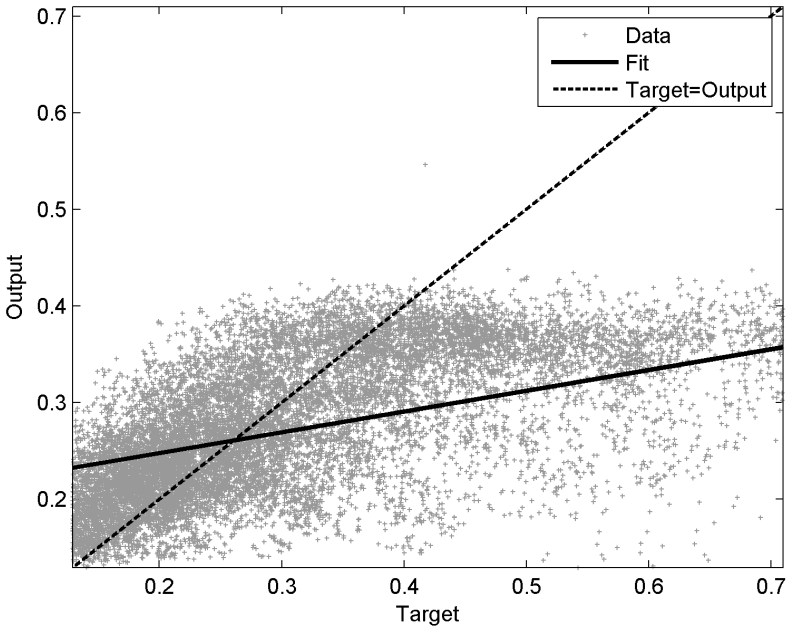

Figure 4.8: Target and Output Values for the σ^{ANNd} Out-of-sample Data (Without Outliers)

The first hypothesis suggest that ANNs can forecast volatility more accurately. This appears to be the case for the combined data set (all securities) in the in-sample period as shown in Table 4.8. The ANN shows lower errors regardless of the measure. However, the results of the out-of-sample period (Table 4.9) are not consistent with this view. The network either overfits the data or a structural shift caused a change in model ranks. Two other aspects are noteworthy. Firstly, the GARCH model is among the worst of those tested regardless of the error measure and period, which is broadly consistent with past literature. Secondly, the out-of-sample period favours the model with the shorter lookback period, which could indicate a more unstable environment.

σ **Model**	**MSE**	**RMSE**	**MAE**	**MAPE**
HVL	0.017 596	0.132 649	0.063 544	0.317 710
HVS	0.019 975	0.141 334	0.066 497	0.314 571
GARCH	0.020 287	0.142 433	0.073 716	0.405 985
ANNd	0.011 853	0.108 870	0.053 966	0.275 371

Table 4.8: Comparison of Error Measures across Volatility Forecasting Models (In-sample)

σ Model	MSE	RMSE	MAE	MAPE
HVL	0.023 738	0.154 070	0.103 287	
HVS	0.022 817	0.151 054	0.100 134	
GARCH	0.029 671	0.172 252	0.113 770	
ANNd	0.030 579	0.174 867	0.100 451	

Table 4.9: Comparison of Error Measures across Volatility Forecasting Models (Out-of-sample)

Source	SS	df	MS	F	p
Groups	22.73	3	7.5761	442.5410	4.9288×10^{-286}
Error	2138.94	124 941	0.0171		
Total	2161.67	124 944			

Table 4.10: ANOVA of Volatility Forecasting Errors (In-sample)

To determine the statistical significance of these results, Table 4.10, reports the ANOVA results. The difference between at least one pair is significant and Table 4.11 shows the pair-wise model comparisons indicating significant rows. The table shows the lower end of the confidence interval, the mean of group differences, and the upper end of the confidence interval. The null hypothesis of no difference, i.e. a difference of 0, cannot be rejected when zero is inside the confidence interval. The tests show that only the combination of σ^{HVS} and the ANN are not significantly different at the 5 % level. The differences are very small, however, from a practical perspective, which is illustrated in Figure 4.9. The large sample size allows for the findings of relatively minor effects. It is important to note, in particular when ranking the models, that the above tables use the mean squared error while the ANOVA tables and box plots use the observed forecasting (and later pricing) errors, their means, and the difference between their means. The MSE is useful as it does not allow for one error to reduce the impact of another in the opposite direction and is thus used for network training and model estimation. That is also the reason it was used as the performance measure during development. However, the question of the following analysis is what the characteristics of an average model prediction are.

The differences are larger showing significance between all groups in the out-of-sample period. Tables 4.12 and 4.13 report the equivalent statistics for this subset, and a visual comparison in Figure 4.10.

Model	**Model**	**Low**	**Mean**	**High**	**Comment**
HVL	HVS	0.0075	0.0102	0.0129	sign.
HVL	GARCH	−0.0261	−0.0234	−0.0207	sign.
HVL	ANNd	0.0062	0.0088	0.0115	sign.
HVS	GARCH	−0.0363	−0.0336	−0.0309	sign.
HVS	ANNd	−0.0041	−0.0014	0.0013	
GARCH	ANNd	0.0295	0.0322	0.0349	sign.

Table 4.11: Multiple Comparison of Volatility Forecasting Models (In-sample)

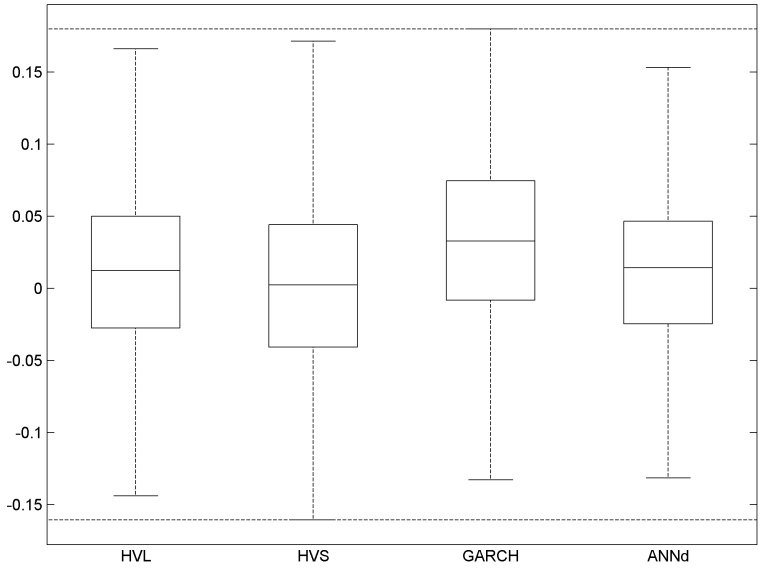

Figure 4.9: Comparison of Volatility Forecasting Errors (In-sample)

Source	**SS**	**df**	**MS**	**F**	**p**
Groups	42.37	3	14.1234	546.6610	0
Error	1788.89	69 241	0.0258		
Total	1831.26	69 244			

Table 4.12: ANOVA of Volatility Forecasting Errors (Out-of-sample)

Model	Model	Low	Mean	High	Comment
HVL	HVS	0.0071	0.0116	0.0160	sign.
HVL	GARCH	0.0261	0.0305	0.0350	sign.
HVL	ANNd	0.0609	0.0653	0.0698	sign.
HVS	GARCH	0.0145	0.0189	0.0234	sign.
HVS	ANNd	0.0493	0.0538	0.0582	sign.
GARCH	ANNd	0.0304	0.0348	0.0393	sign.

Table 4.13: Multiple Comparison of Volatility Forecasting Models (Out-of-sample)

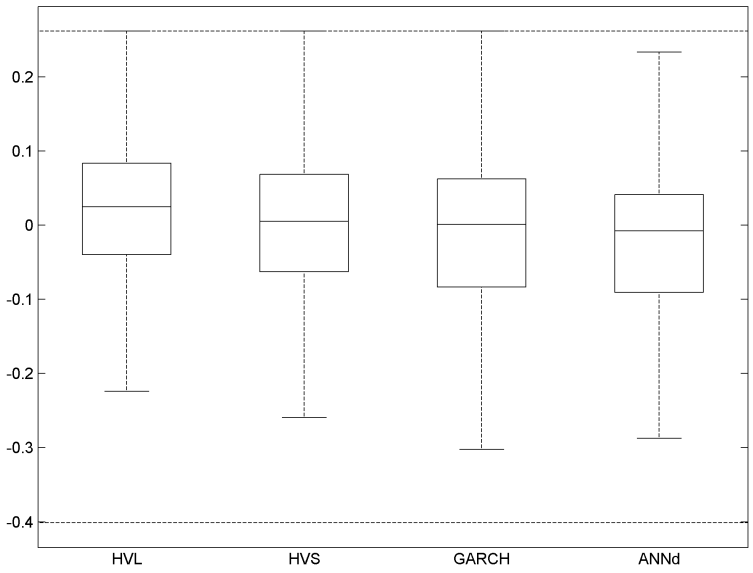

Figure 4.10: Comparison of Volatility Forecasting Errors (Out-of-sample)

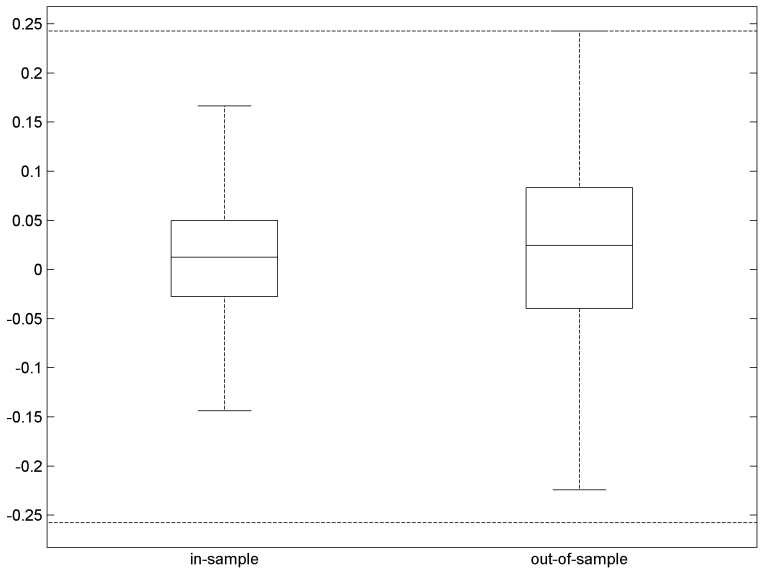

Figure 4.11: Comparison of σ^{HVL} Errors Applied to In-sample and Out-of-sample Data

Based on the in-sample results, one would have used the ANN and in this regard the hypothesis was correct. It does not hold, however, when tested in a new set and in changed circumstances. This raises the question of stability of the models generally. When comparing the in-sample and out-of-sample errors for each model, the user of a model would prefer to see no significant difference between them, i.e. the model applies equally well in either period.

The more complex models, σ^{GARCH} (Table 4.16 and Figure 4.13) and σ^{ANNd} (Table 4.17 and Figure 4.14), show a significant difference between in-sample and out-of sample errors. The same is not true for the simple models σ^{HVL} (Table 4.14 and Figure 4.11) and σ^{HVS} (Table 4.15 and Figure 4.12), which suggests that their forecasting characteristics have not changed significantly in the transition from the in-sample to the out-of-sample set. All models show a much wider range of forecasting errors in the out-of-sample set, which was expected considering the period it covered as well as the nature of any out-of-sample testing.

Source	SS	df	MS	F	p
Groups	0.00	1	0.0012	0.0630	0.8019
Error	953.63	48 465	0.0197		
Total	953.63	48 466			

Table 4.14: ANOVA of σ^{HVL} In-sample and Out-of-sample Errors

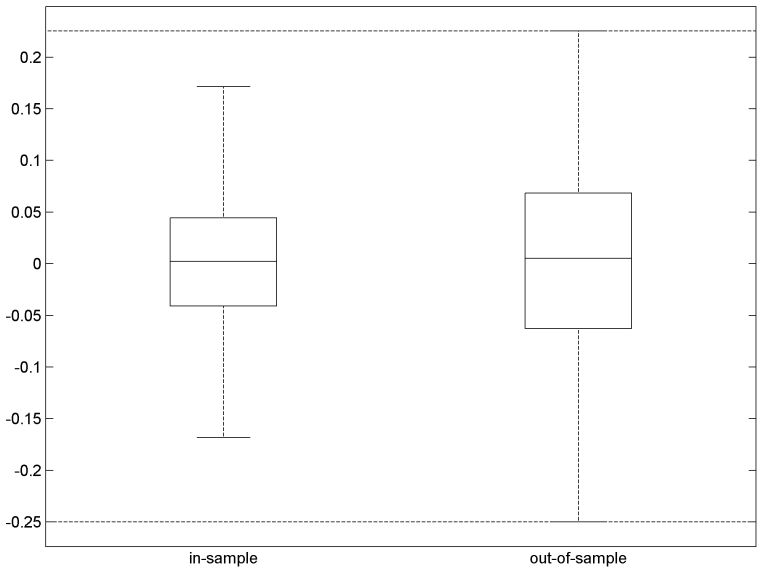

Figure 4.12: Comparison of σ^{HVS} Errors Applied to In-sample and Out-of-sample Data

Source	SS	df	MS	F	p
Groups	0.01	1	0.0111	0.5313	0.4661
Error	1020.95	48 645	0.0210		
Total	1020.96	48 646			

Table 4.15: ANOVA of σ^{HVS} In-sample and Out-of-sample Errors

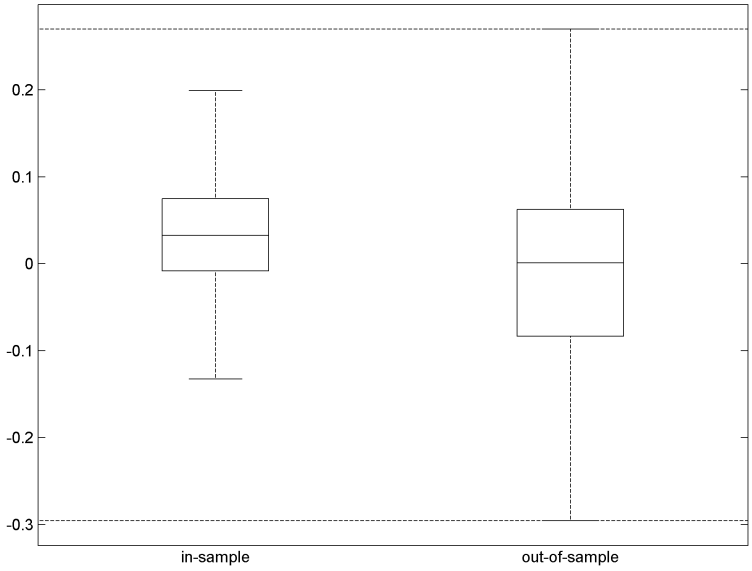

Figure 4.13: Comparison of σ^{GARCH} Errors Applied to In-sample and Out-of-sample Data

Source	**SS**	**df**	**MS**	**F**	**p**
Groups	32.03	1	32.0283	1408.1100	8.0446×10^{-304}
Error	1107.33	48 683	0.0227		
Total	1139.36	48 684			

Table 4.16: ANOVA of σ^{GARCH} In-sample and Out-of-sample Errors

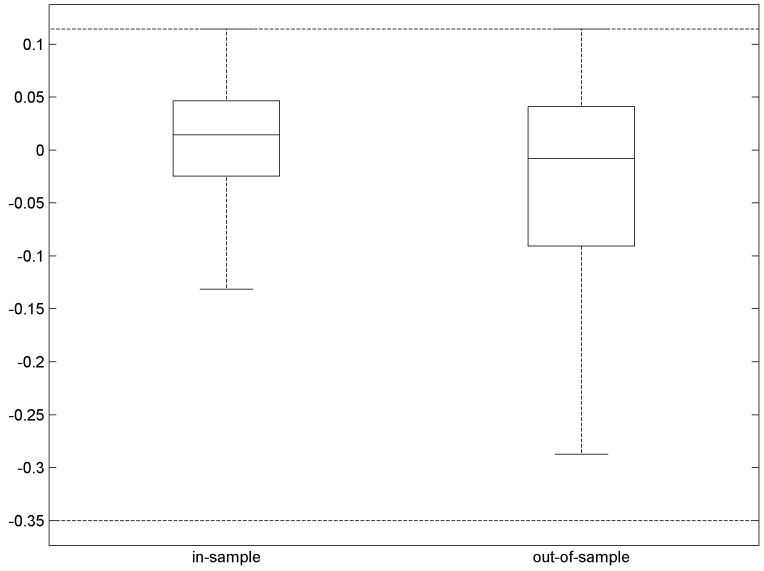

Figure 4.14: Comparison of σ^{ANNd} Errors Applied to In-sample and Out-of-sample Data

Source	**SS**	**df**	**MS**	**F**	**p**
Groups	35.04	1	35.0449	2004.6300	0
Error	845.94	48 389	0.0175		
Total	880.98	48 390			

Table 4.17: ANOVA of σ^{ANNd} In-sample and Out-of-sample Errors

4.4 Volatility Surface Fitting

The next two steps are to address the second hypothesis. For this reason, this section will report results only and conclusions will be discussed in the following section in the context of pricing options.

Additional data was synchronised with the option prices, including the underlying security reference, the strike price, and the expiry date. These were taken from relevant fields, data update events and the description field. The approximate expiry date was also taken by inference from the trading symbol to determine if any of the update events is applicable. In particular, any data was not applied more than 5 years forward as trading symbols can be re-used over longer periods. Furthermore, the dividend yield was calculated as the continuous yield of the previous year's sum of interim and final cash dividends relative to the current price of the underlying. Special dividends were not considered. This is updated each period with only the information a market participant would have received at the time taken into consideration.

In order to fit a volatility surface, the quadratic model is fitted using robust linear regression. This requires the implied volatility as used by the option pricing (reference) model in the next step. A small subsample of similar order of magnitude as the forecasting model was thus taken from all qualifying options with complete information (expiry date, strike price, underlying security). Qualifying options were those, whose underlying security is one of the securities listed previously.

The subsample comprised 51 814 in-sample and 46 768 out-of-sample records initially (see Tables 4.17 and 4.44 for a comparison of the total number of observations used for volatility forecasting and option pricing). While representing only a fraction of all trades, the long sampling period and the inclusion of several underlying securities results in a sizeable data set. More importantly, however, the use of individual trades during the day does not impose a 'special' meaning onto the final trade, and is not biased where such a significance (of the last trade of a day) may exist. Instead, it represents any transaction during the trading day regardless of its timing.

Finally, two types of exclusion critera are applied. Firstly, if the implied volatility could not be calculated, the observation was removed from these models. Secondly, a small number of securities with M greater than 5 or those with an option price larger than the strike were removed. The latter should not find a

	Value	SE	t	p
(Intercept)	3.0965	0.0274	113.01	0
M	−5.6255	0.0523	−107.52	0
T	0.9824	0.0298	32.94	2.1668×10^{-235}
MT	−0.7727	0.0282	−27.43	2.2345×10^{-164}
M^2	3.6312	0.0288	126.29	0
T^2	−0.0282	0.0029	−9.64	5.6076×10^{-22}

Table 4.18: Volatility Surface Model Parameters for $\sigma^{\text{HVL}}_{M,T}$

	Value	SE	t	p
(Intercept)	0.4480	0.0370	12.10	1.1325×10^{-33}
M	−0.6262	0.0707	−8.86	8.2597×10^{-19}
T	1.2770	0.0403	31.69	3.0362×10^{-218}
MT	−1.0653	0.0381	−27.99	4.9112×10^{-171}
M^2	1.3446	0.0388	34.62	2.0880×10^{-259}
T^2	−0.0329	0.0040	−8.33	8.1324×10^{-17}

Table 4.19: Volatility Surface Model Parameters for $\sigma^{\text{HVS}}_{M,T}$

	Value	SE	t	p
(Intercept)	5.1333	0.0247	207.94	0
M	−9.5323	0.0471	−202.22	0
T	0.5940	0.0269	22.11	8.9661×10^{-108}
MT	−0.5157	0.0254	−20.32	2.2516×10^{-91}
M^2	5.4243	0.0259	209.39	0
T^2	0.0018	0.0026	0.70	0.4846

Table 4.20: Volatility Surface Model Parameters for $\sigma^{\text{GARCH}}_{M,T}$

buyer since it would be cheaper to buy the underlying security rather than buying the option for the same price and having to pay for the underlying again later. This removed 20 in-sample observations, and 45 out-of-sample, respectively from each model. After the removal of observations missing implied volatility, 48 846 in-sample records were applied and the models tested on 41 897 out-of-sample records. ANNs models further remove incomplete observations, which result from the lagged historical volatility time series.

Once the data was prepared, the model parameters were estimated. Tables 4.18 to 4.21 show the model parameters and probability values.

	Value	**SE**	**t**	**p**
(Intercept)	5.4782	0.0215	255.39	0
M	−10.3374	0.0410	−252.38	0
T	0.7543	0.0233	32.30	1.4992×10^{-226}
MT	−0.6066	0.0221	−27.50	2.8361×10^{-165}
M^2	5.9548	0.0225	264.54	0
T^2	−0.0147	0.0023	−6.41	1.4416×10^{-10}

Table 4.21: Volatility Surface Model Parameters for $\sigma_{M,T}^{\text{ANNd}}$

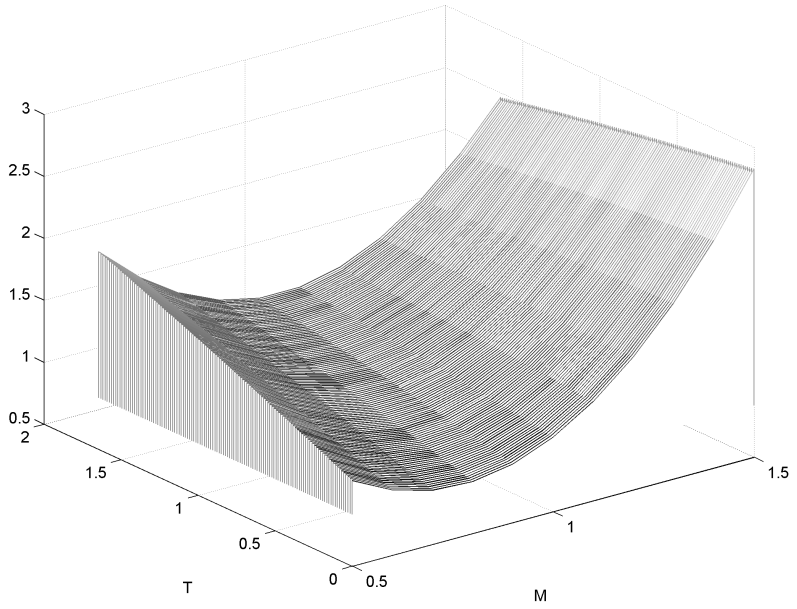

Figure 4.15: Sample $\sigma_{M,T}^{\text{HVL}}$Volatility Surface (Multiplier)

Figures 4.15 to 4.18 show the fitted surface models. The range is chosen near price-strike equality and for a relatively short time frame. The surface represents correction factors to the underlying volatility forecast. Considering that same implied volatility values were used in fitting all models, they show significant differences in shape. It is important to note that the value of each of these surfaces has to be considered in the context of the underlying volatility model. For this reason, the evaluation includes volatility modelling errors, rather than errors in factor levels.

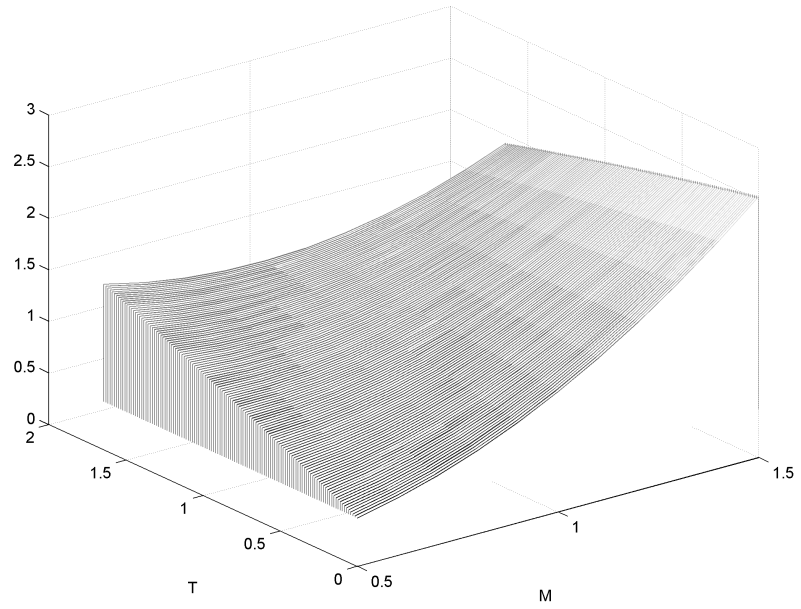

Figure 4.16: Sample $\sigma_{M,T}^{\text{HVS}}$Volatility Surface (Multiplier)

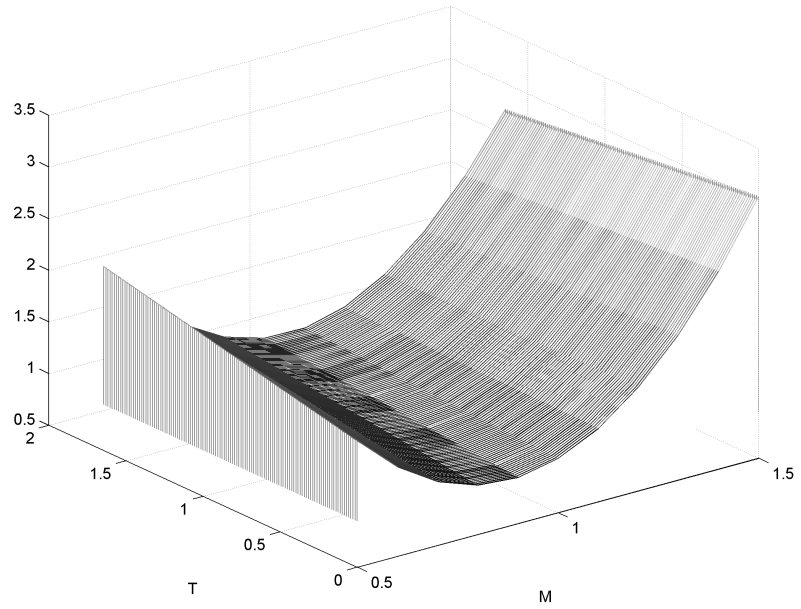

Figure 4.17: Sample $\sigma_{M,T}^{\text{GARCH}}$Volatility Surface (Multiplier)

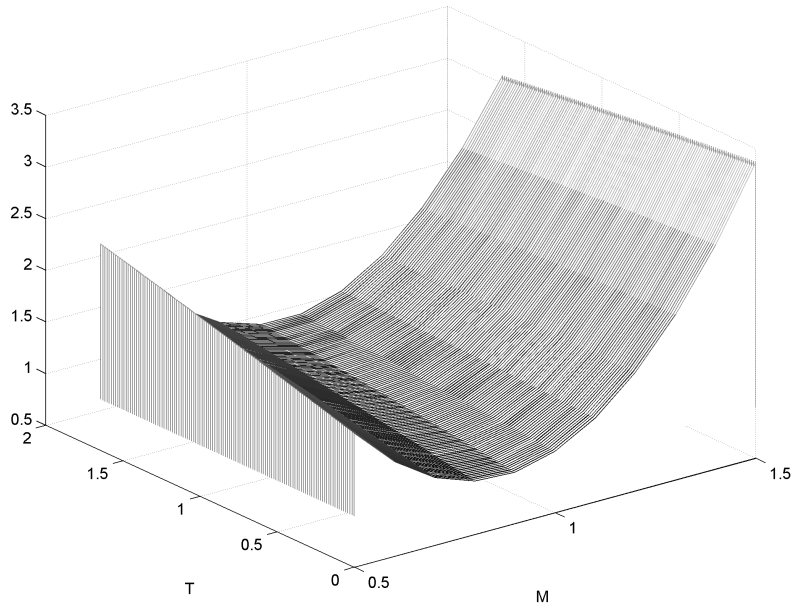

Figure 4.18: Sample $\sigma_{M,T}^{\text{ANNd}}$Volatility Surface (Multiplier)

Using the same implied volatility estimates as the target values, an ANN was trained. The variable value ranges are shown in Figures 4.19 and 4.20. They show similar characteristics of the input values except for the generally higher volatility levels in the out-of-sample set (see above), and a number of outliers in the lower range of the moneyness values. The network training terminated very early (see Figure 4.22) when the network with four hidden nodes failed to improve results. Thus the smallest network with three nodes was chosen. The training record is shown in Figure 4.21, which shows the expected decrease in the error terms.

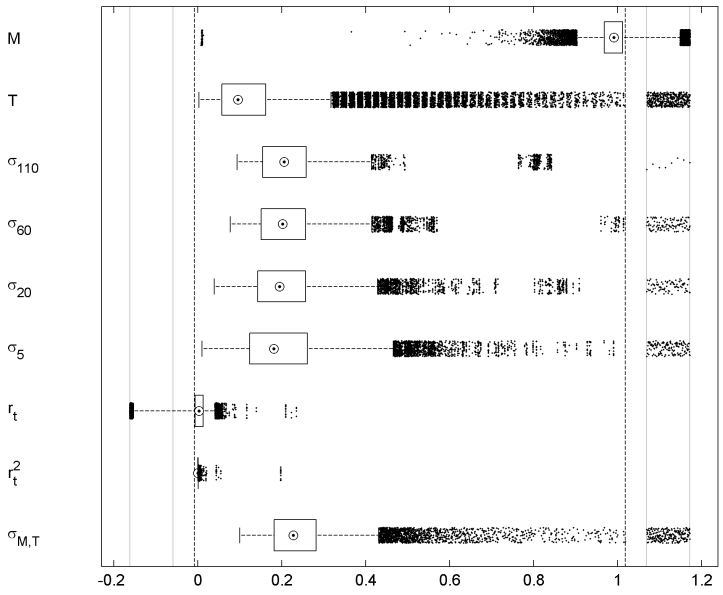

Figure 4.19: Characteristics of Variable Value Ranges (In-sample) for $\sigma_{M,T}^{\text{ANNs}}$

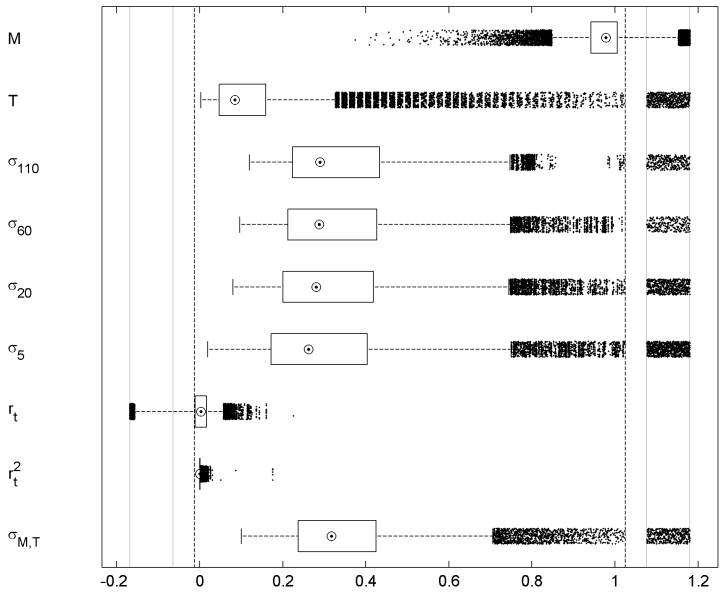

Figure 4.20: Characteristics of Variable Value Ranges (Out-of-sample) for $\sigma_{M,T}^{\text{ANNs}}$

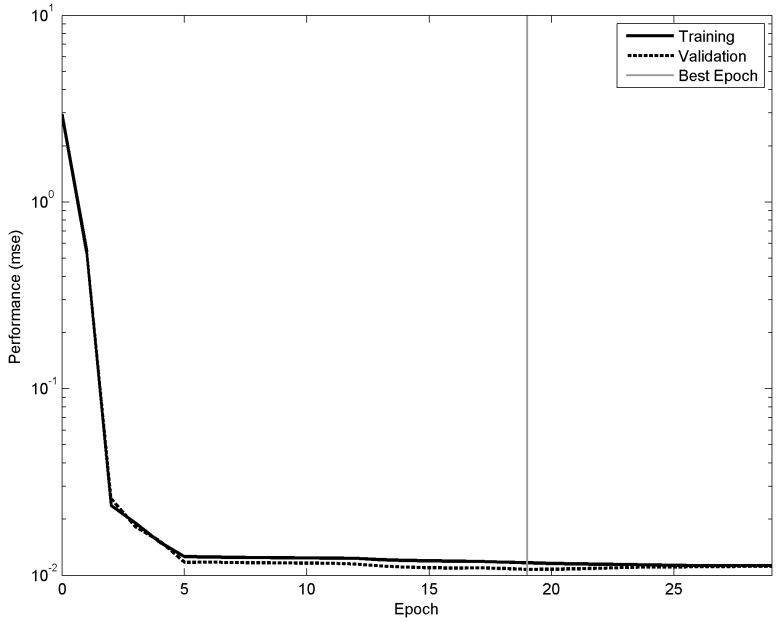

Figure 4.21: $\sigma_{M,T}^{\text{ANNs}}$ Training Record

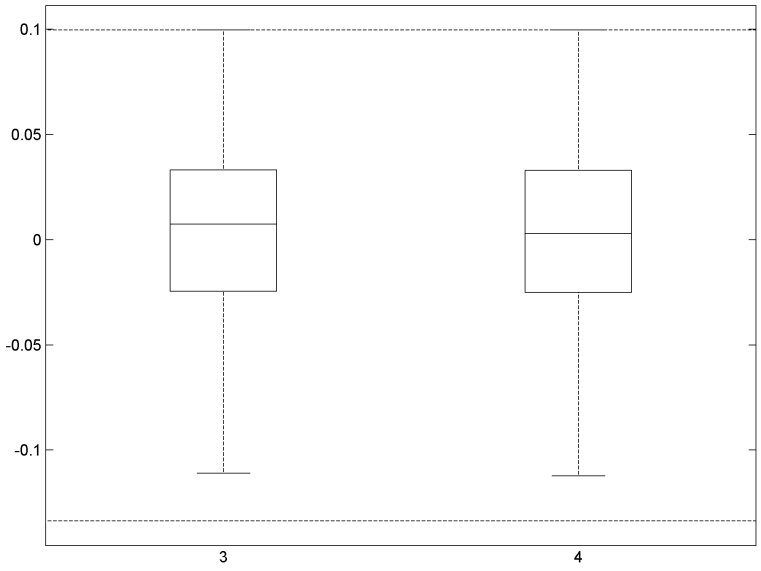

Figure 4.22: Error Distribution of Trained Network Architectures for $\sigma_{M,T}^{\text{ANNs}}$ (In-sample)

The plots of actual compared to output values show a similar pattern as before. Focusing on the central data range Figures 4.23 and 4.24 already show a greater difference and tendency towards an average value. This further demonstrates the need for additional research; the same argument against basing modelling decisions on out-of-sample results applies, however.

Similar to volatility forecasting, the volatility surface models need to be compared to each other and with respect to their ability to generalise beyond the training set.

It is evident that the network models are preferable compared with the alternative models (see Tables 4.22 and tab:r:d:err:models:o:s), not only with respect to their MSE and other metrics but also with regard to the location and distribution of the prediction error (see Tables 4.24, 4.25, 4.26, and 4.27). More important than the fact that the volatility forecasting network rather than the surface model appears strongest in the out-of-sample period is the fact that the volatility forecasting model combined with a regression model offers multipliers such that the errors of implied volatility are in a narrower range (see Figure 4.25). This remains true even in the out-of-sample data set (see Figure 4.26.

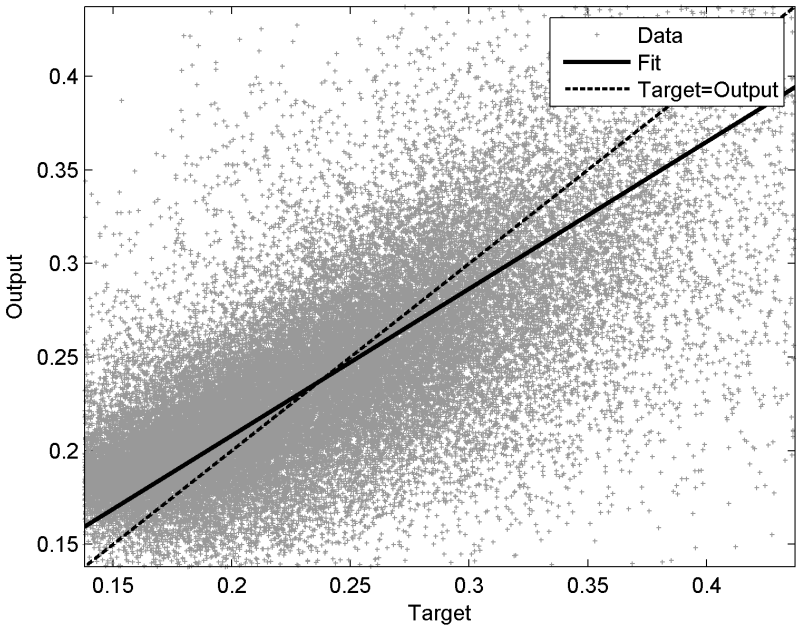

Figure 4.23: Target and Output Values for the $\sigma_{M,T}^{\text{ANNs}}$ In-sample Data (Without Outliers)

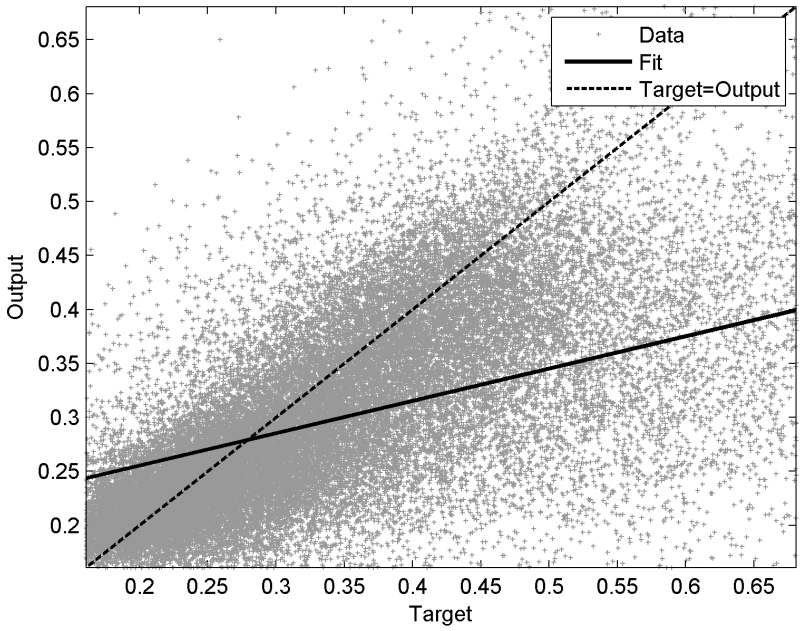

Figure 4.24: Target and Output Values for the $\sigma_{M,T}^{\text{ANNs}}$ Out-of-sample Data (Without Outliers)

$\boldsymbol{\sigma}_{\mathbf{M,T}}$ **Model**	**MSE**	**RMSE**	**MAE**	**MAPE**
HVL	0.048 439	0.220 089	0.057 709	0.189 387
HVS	0.062 097	0.249 192	0.071 288	0.241 269
GARCH	0.051 221	0.226 320	0.059 271	0.196 969
ANNd	0.040 245	0.200 612	0.047 450	0.151 263
ANNs	0.011 468	0.107 086	0.045 948	0.178 562

Table 4.22: Comparison of Error Measures across Volatility Surface Models (In-sample)

$\boldsymbol{\sigma}_{\mathbf{M,T}}$ **Model**	**MSE**	**RMSE**	**MAE**	**MAPE**
HVL	0.030 358	0.174 235	0.092 886	0.239 364
HVS	0.036 675	0.191 508	0.101 733	0.263 480
GARCH	0.032 894	0.181 366	0.093 804	0.231 255
ANNd	0.029 049	0.170 437	0.077 960	0.178 584
ANNs	0.029 989	0.173 174	0.087 149	0.206 772

Table 4.23: Comparison of Error Measures across Volatility Surface Models (Out-of-sample)

Source	SS	df	MS	F	p
Groups	4.93	4	1.2333	28.9261	4.4733×10^{-24}
Error	10 407.30	244 104	0.0426		
Total	10 412.20	244 108			

Table 4.24: ANOVA of Volatility Surface Errors (In-sample)

Model	Model	Low	Mean	High	Comment
HVL	HVS	−0.0101	−0.0065	−0.0029	sign.
HVL	GARCH	−0.0066	−0.0030	0.0006	
HVL	ANNd	0.0027	0.0063	0.0099	sign.
HVL	ANNs	−0.0084	−0.0047	−0.0011	sign.
HVS	GARCH	−0.0001	0.0035	0.0071	
HVS	ANNd	0.0092	0.0128	0.0164	sign.
HVS	ANNs	−0.0019	0.0018	0.0054	
GARCH	ANNd	0.0057	0.0093	0.0129	sign.
GARCH	ANNs	−0.0054	−0.0017	0.0019	
ANNd	ANNs	−0.0147	−0.0111	−0.0075	sign.

Table 4.25: Multiple Comparison of Volatility Surface Models (In-sample)

Source	SS	df	MS	F	p
Groups	287.81	4	71.9512	2387.0200	0
Error	6314.28	209 480	0.0301		
Total	6602.09	209 484			

Table 4.26: ANOVA of Volatility Surface Errors (Out-of-sample)

Model	Model	Low	Mean	High	Comment
HVL	HVS	−0.0056	−0.0023	0.0009	
HVL	GARCH	0.0635	0.0667	0.0700	sign.
HVL	ANNd	0.0691	0.0724	0.0757	sign.
HVL	ANNs	0.0790	0.0823	0.0855	sign.
HVS	GARCH	0.0658	0.0691	0.0723	sign.
HVS	ANNd	0.0715	0.0747	0.0780	sign.
HVS	ANNs	0.0813	0.0846	0.0879	sign.
GARCH	ANNd	0.0024	0.0057	0.0090	sign.
GARCH	ANNs	0.0123	0.0155	0.0188	sign.
ANNd	ANNs	0.0066	0.0098	0.0131	sign.

Table 4.27: Multiple Comparison of Volatility Surface Models (Out-of-sample)

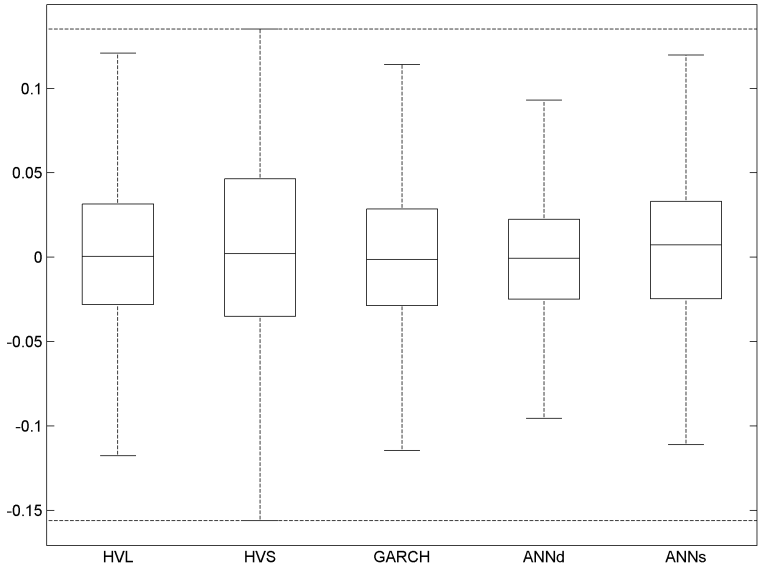

Figure 4.25: Comparison of Volatility Surface Errors (In-sample)

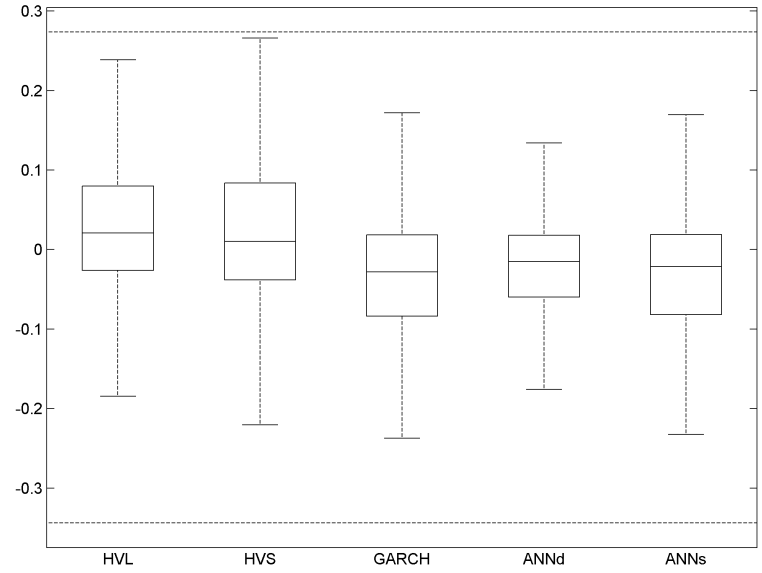

Figure 4.26: Comparison of Volatility Surface Errors (Out-of-sample)

As in the previous section, volatility surface models perform worse and significantly differently out-of-sample. Evidence can be found in Figures 4.27 to 4.31, as well as Tables 4.28 to 4.32. The difference between the two sample periods and the sample size are together large enough to reject the null hypothesis that they are the same across all models.

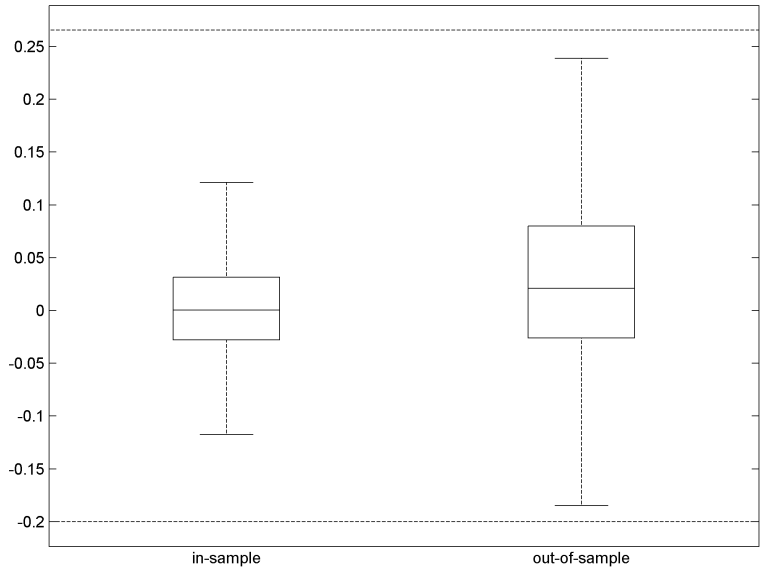

Figure 4.27: Comparison of $\sigma_{M,T}^{\text{HVL}}$ Errors Applied to In-sample and Out-of-sample Data

Source	**SS**	**df**	**MS**	**F**	**p**
Groups	28.22	1	28.2238	710.7330	5.5602×10^{-156}
Error	3601.85	90 702	0.0397		
Total	3630.08	90 703			

Table 4.28: ANOVA of $\sigma_{M,T}^{\text{HVL}}$ In-sample and Out-of-sample Errors

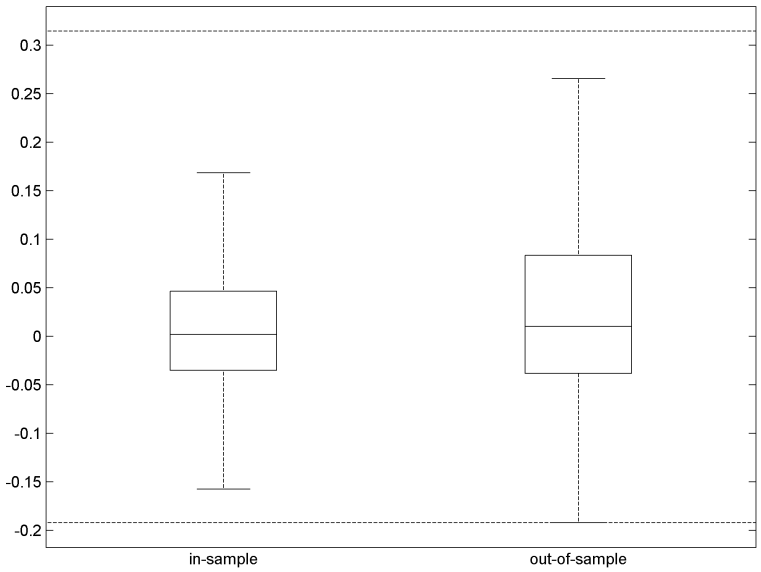

Figure 4.28: Comparison of $\sigma_{M,T}^{\mathrm{HVS}}$ Errors Applied to In-sample and Out-of-sample Data

Source	**SS**	**df**	**MS**	**F**	**p**
Groups	21.96	1	21.9621	439.6230	2.2289×10^{-97}
Error	4532.93	90 737	0.0500		
Total	4554.89	90 738			

Table 4.29: ANOVA of $\sigma_{M,T}^{\mathrm{HVS}}$ In-sample and Out-of-sample Errors

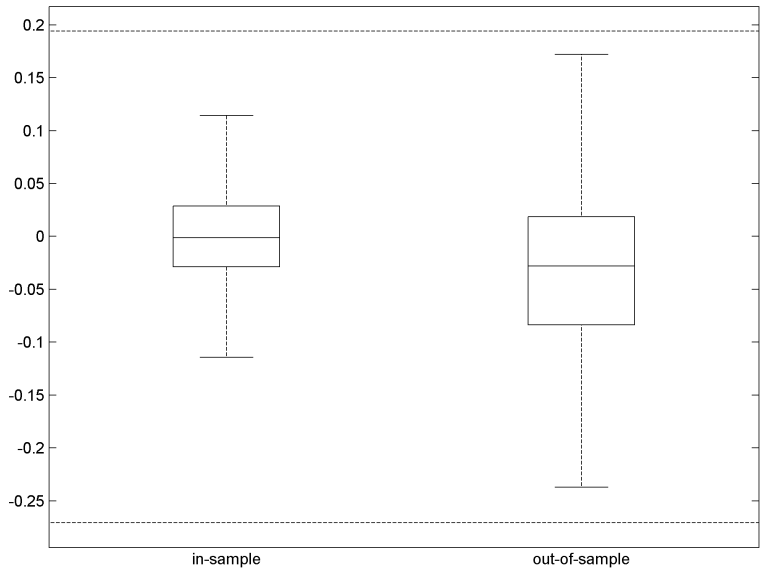

Figure 4.29: Comparison of $\sigma_{M,T}^{\text{GARCH}}$ Errors Applied to In-sample and Out-of-sample Data

Source	**SS**	**df**	**MS**	**F**	**p**
Groups	26.61	1	26.6070	633.1610	3.0924×10^{-139}
Error	3813.16	90 741	0.0420		
Total	3839.77	90 742			

Table 4.30: ANOVA of $\sigma_{M,T}^{\text{GARCH}}$ In-sample and Out-of-sample Errors

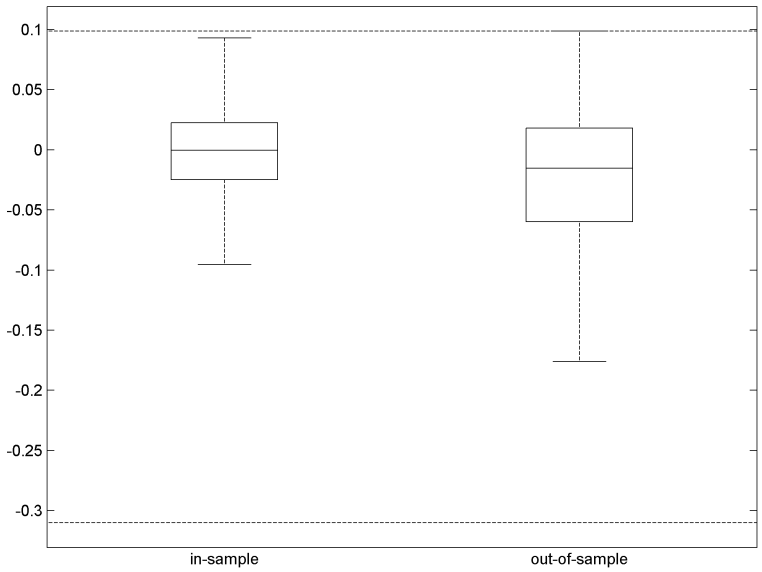

Figure 4.30: Comparison of $\sigma_{M,T}^{\mathrm{ANNd}}$ Errors Applied to In-sample and Out-of-sample Data

Source	**SS**	**df**	**MS**	**F**	**p**
Groups	21.27	1	21.2677	625.2450	1.5860×10^{-137}
Error	3085.23	90 702	0.0340		
Total	3106.50	90 703			

Table 4.31: ANOVA of $\sigma_{M,T}^{\mathrm{ANNd}}$ In-sample and Out-of-sample Errors

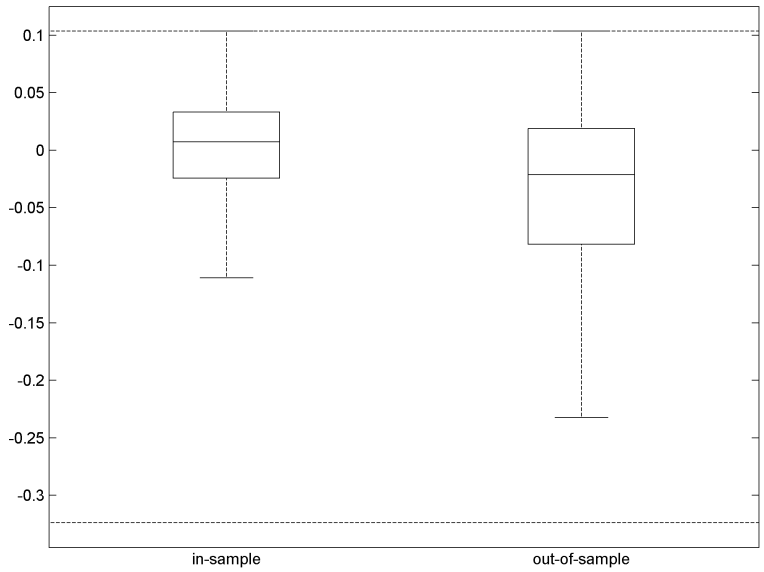

Figure 4.31: Comparison of $\sigma_{M,T}^{\text{ANNs}}$ Errors Applied to In-sample and Out-of-sample Data

Source	**SS**	**df**	**MS**	**F**	**p**
Groups	60.08	1	60.0767	3227.3600	0
Error	1688.40	90 702	0.0186		
Total	1748.48	90 703			

Table 4.32: ANOVA of $\sigma_{M,T}^{\text{ANNs}}$ In-sample and Out-of-sample Errors

4.5 Option Pricing Evaluation

In the previous step, models from the first stage were applied to yield an enhanced model for the volatility surface model, i.e. a correction or modifier, which is applied to the underlying volatility forecast. The evaluation of option pricing models follows the same process. The underlying models were used to price options. Observed option prices serve as the target values.

The data set is the same as in the previous step with two small changes. Firstly, the pricing function (Bjerksund and Stensland, 2002) requires strictly positive dividend yield. Where the yield was 0, it was replaced with the smallest positive floating point value, that could be represented by hardware architecture. Secondly, just as some implied volatilities could not be found, some options could not be priced. These observations were ignored by the statistical functions during evaluation. This did not affect the full ANN model.

In addition to these consideration, a transformation to the output of the ANN was necessary. All models except for the full machine learning model result in a price as their final output. Consistent with prior literature, the network is, however, trained to learn the option price per unit of money not unit of stock. This was achieved by dividing by the strike price. In order to allow for comparisons, the process was reversed before statistics were calculated. The network output was thus multiplied by the applicable strike price and the resulting option price was used for evaluation.

The additional model in this step is the full ANN. It is based on all previous inputs; no additions were made beyond those in moving from the underlying (volatility forecasting) to the option (volatility surface). The training process is the same as in the previous stages of modelling. The network went through two states during training as is visible in the training record 4.34. Note that the last of the parameters in Figures 4.32 and 4.33 represent the price of the option divided by the strike and that this process was reversed for the purpose of the following analysis. Four network architectures were trained and evaluated using the validation set and the network with six hidden nodes failed to improve performance so the network with five hidden nodes was chosen (see Figure 4.35 for a comparison of error terms).

Comparing the results shows that the resulting neural network does not perform well. The existing networks, on the other hand combine very well with the

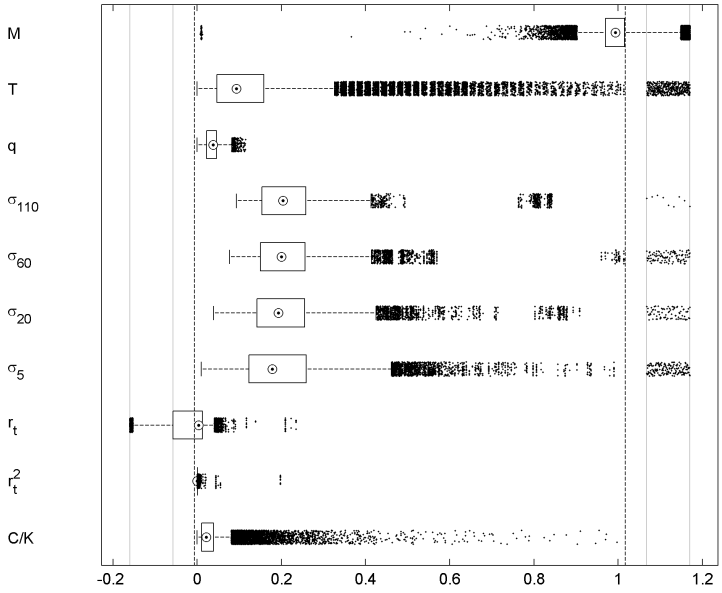

Figure 4.32: Characteristics of Variable Value Ranges (In-sample) for C^{ANN}

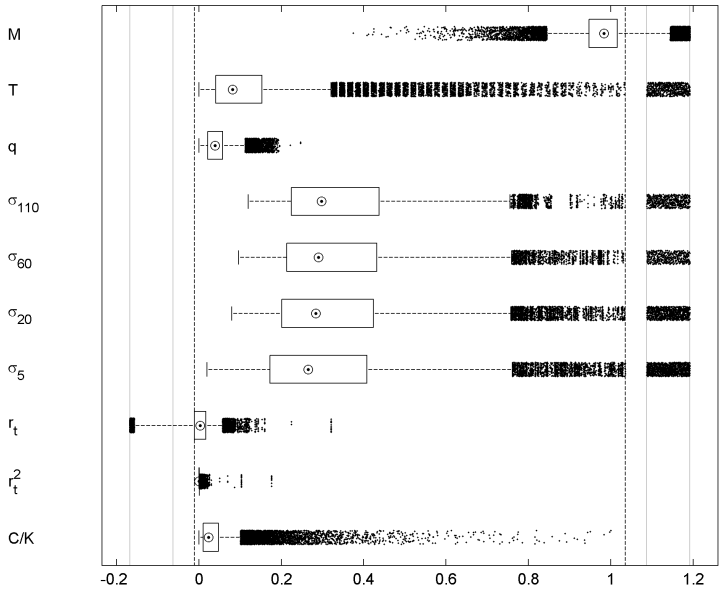

Figure 4.33: Characteristics of Variable Value Ranges (Out-of-sample) for C^{ANN}

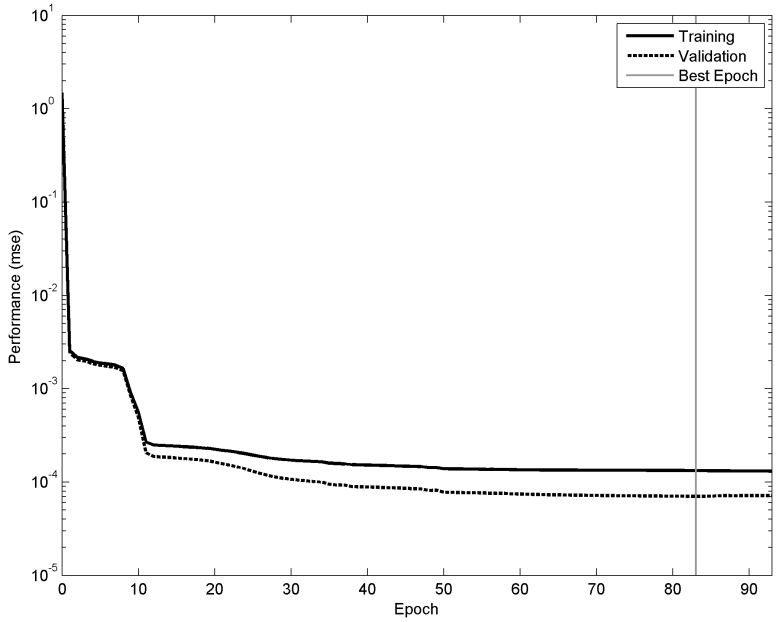

Figure 4.34: C^{ANN} Training Record

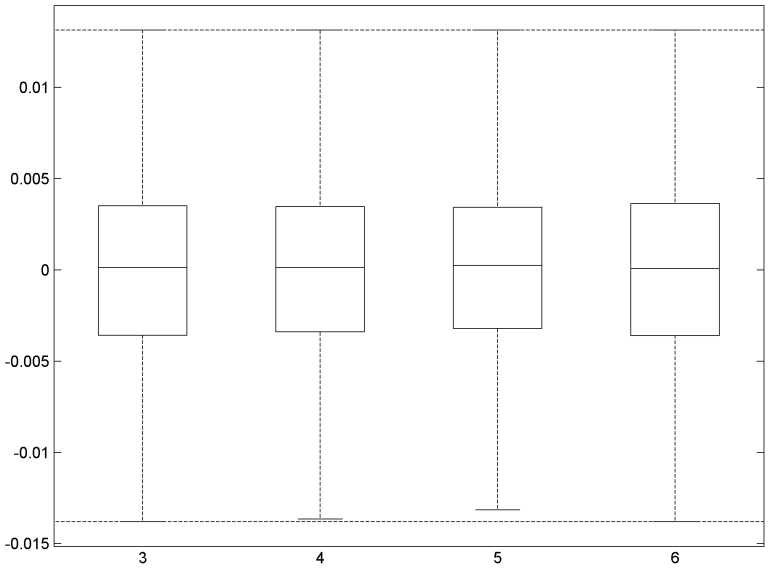

Figure 4.35: Error Distribution of Trained Network Architectures for C^{ANN} (In-sample)

C Model	MSE	RMSE	MAE	MAPE
HVL	0.055 505	0.235 594	0.109 127	0.252 030
HVS	0.101 411	0.318 450	0.150 484	0.332 511
GARCH	0.072 472	0.269 206	0.111 781	0.274 537
ANNd	0.056 563	0.237 830	0.089 010	0.200 126
ANNs	0.086 734	0.294 507	0.115 556	0.245 223
ANN	3.400 510	1.844 050	0.169 881	1.346 040

Table 4.33: Comparison of Error Measures across Option Pricing Models (In-sample)

C Model	MSE	RMSE	MAE	MAPE
HVL	0.744 880	0.863 064	0.319 320	0.429 064
HVS	0.976 326	0.988 092	0.357 043	0.438 967
GARCH	0.660 226	0.812 543	0.274 976	0.404 852
ANNd	0.369 346	0.607 739	0.215 487	0.283 973
ANNs	0.450 993	0.671 560	0.263 247	0.309 993
ANN	24.383 300	4.937 940	1.296 070	4.789 660

Table 4.34: Comparison of Error Measures across Option Pricing Models (Out-of-sample)

reference pricing model and the volatility surface model. Tables 4.33 and 4.34 summarise the error statistics.

Figure 4.38 reveals that it is a certain type of option that particularly affects the results. These options largely characterised by their low option value relative to their strike (the plots show the data after multiplying by the strike again) and they are typically overpriced by the network; the effect is not visible in the core set (see Figure 4.39) suggesting it is an outlier effect. This only applies to the out-of-sample data (see Figure 4.38), however, and does not explain the poor results more generally (see Figures 4.36 and 4.37 for comparison).

The remaining networks, which use the same input data do not show poor performance. Rather, they perform very well in both subsets. It is noteworthy, however, that all models perform much worse out-of-sample.

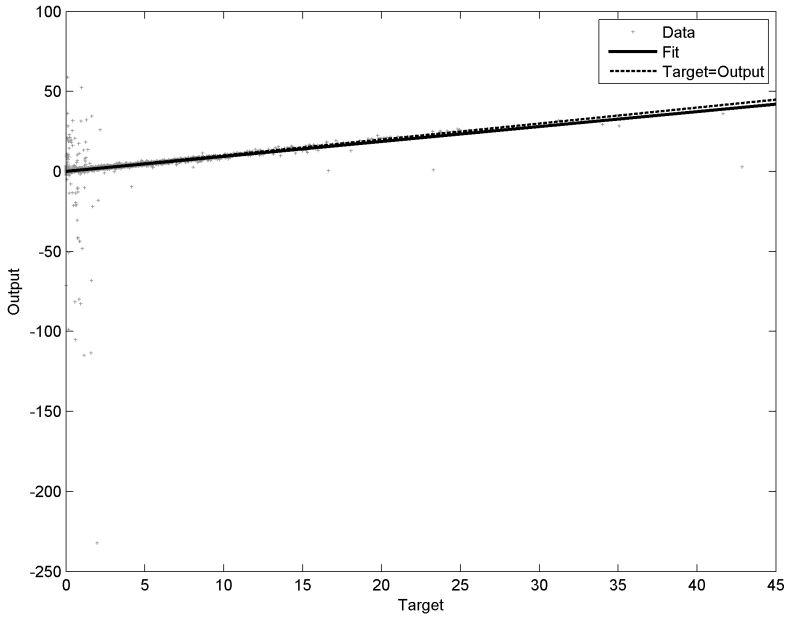

Figure 4.36: Target and Output Values for the C^{ANN} In-sample Data

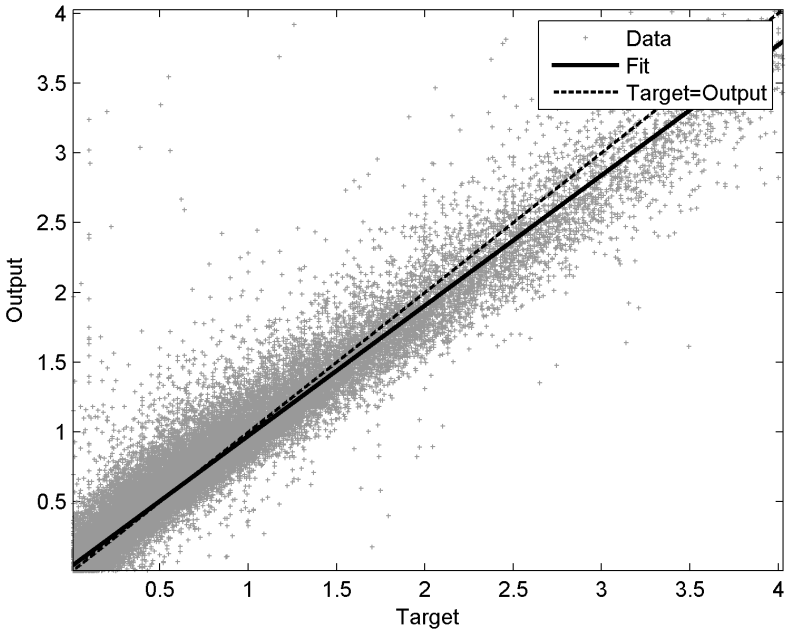

Figure 4.37: Target and Output Values for the C^{ANN} In-sample Data (Without Outliers)

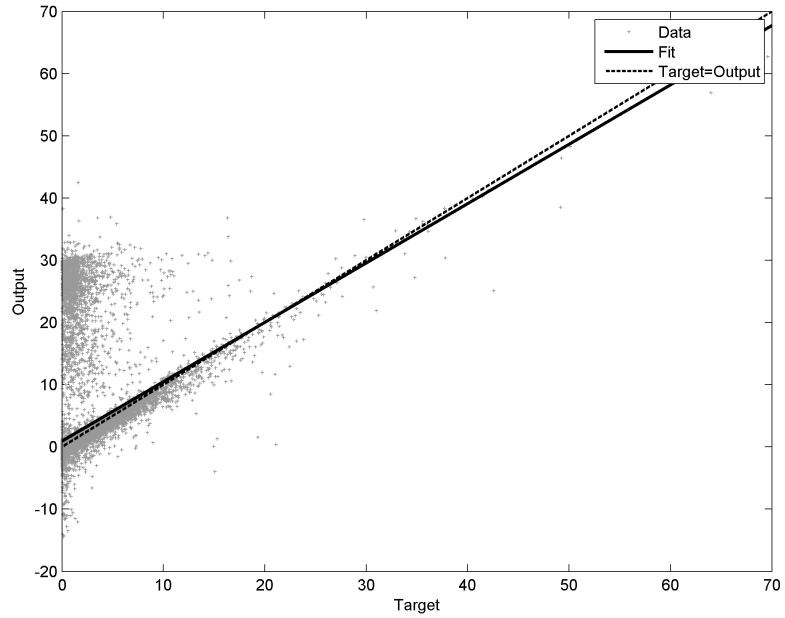

Figure 4.38: Target and Output Values for the C^{ANN} Out-of-sample Data

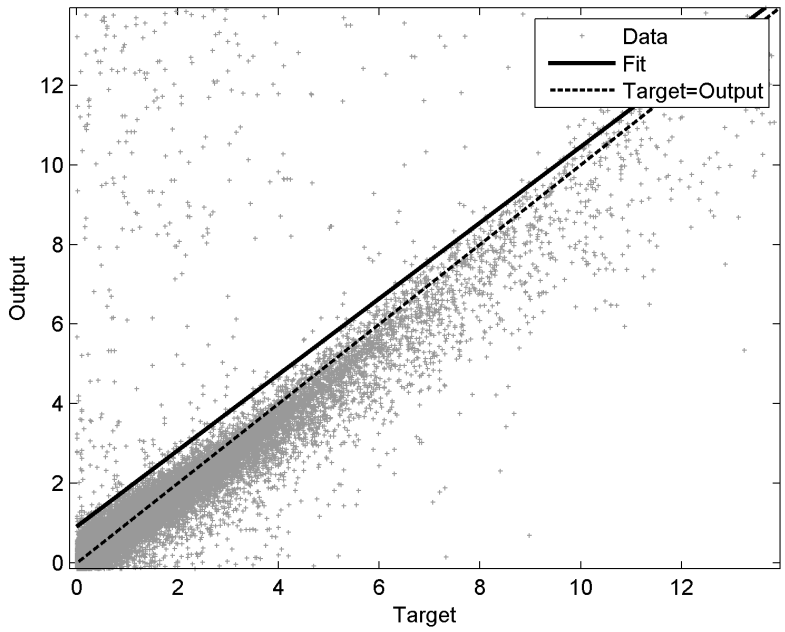

Figure 4.39: Target and Output Values for the C^{ANN} Out-of-sample Data (Without Outliers)

Source	**SS**	**df**	**MS**	**F**	**p**
Groups	108.02	5	21.6035	32.4450	3.3945×10^{-33}
Error	193 677	290 873	0.6658		
Total	193 785	290 878			

Table 4.35: ANOVA of Option Pricing Errors (In-sample)

Model	**Model**	**Low**	**Mean**	**High**	**Comment**
HVL	HVS	−0.0421	−0.0271	−0.0120	sign.
HVL	GARCH	−0.0061	0.0089	0.0240	
HVL	ANNd	−0.0027	0.0123	0.0273	
HVL	ANNs	0.0068	0.0218	0.0369	sign.
HVL	ANN	0.0197	0.0344	0.0492	sign.
HVS	GARCH	0.0210	0.0360	0.0511	sign.
HVS	ANNd	0.0243	0.0394	0.0544	sign.
HVS	ANNs	0.0339	0.0489	0.0640	sign.
HVS	ANN	0.0468	0.0615	0.0763	sign.
GARCH	ANNd	−0.0117	0.0033	0.0184	
GARCH	ANNs	−0.0021	0.0129	0.0279	
GARCH	ANN	0.0107	0.0255	0.0402	sign.
ANNd	ANNs	−0.0055	0.0096	0.0246	
ANNd	ANN	0.0074	0.0222	0.0369	sign.
ANNs	ANN	−0.0022	0.0126	0.0273	

Table 4.36: Multiple Comparison of Option Pricing Models (In-sample)

Source	**SS**	**df**	**MS**	**F**	**p**
Groups	30 478.10	5	6095.61	1258.1000	0
Error	1.23×10^6	254 700	4.85		
Total	1.26×10^6	254 705			

Table 4.37: ANOVA of Option Pricing Errors (Out-of-sample)

Model	Model	Low	Mean	High	Comment
HVL	HVS	−0.0541	−0.0106	0.0329	
HVL	GARCH	0.2058	0.2493	0.2927	sign.
HVL	ANNd	0.2218	0.2653	0.3088	sign.
HVL	ANNs	0.3009	0.3444	0.3878	sign.
HVL	ANN	−0.6984	−0.6561	−0.6138	sign.
HVS	GARCH	0.2163	0.2598	0.3033	sign.
HVS	ANNd	0.2324	0.2759	0.3194	sign.
HVS	ANNs	0.3114	0.3549	0.3984	sign.
HVS	ANN	−0.6878	−0.6456	−0.6033	sign.
GARCH	ANNd	−0.0275	0.0160	0.0595	
GARCH	ANNs	0.0516	0.0951	0.1386	sign.
GARCH	ANN	−0.9477	−0.9054	−0.8631	sign.
ANNd	ANNs	0.0356	0.0791	0.1226	sign.
ANNd	ANN	−0.9637	−0.9214	−0.8791	sign.
ANNs	ANN	−1.0428	−1.0005	−0.9582	sign.

Table 4.38: Multiple Comparison of Option Pricing Models (Out-of-sample)

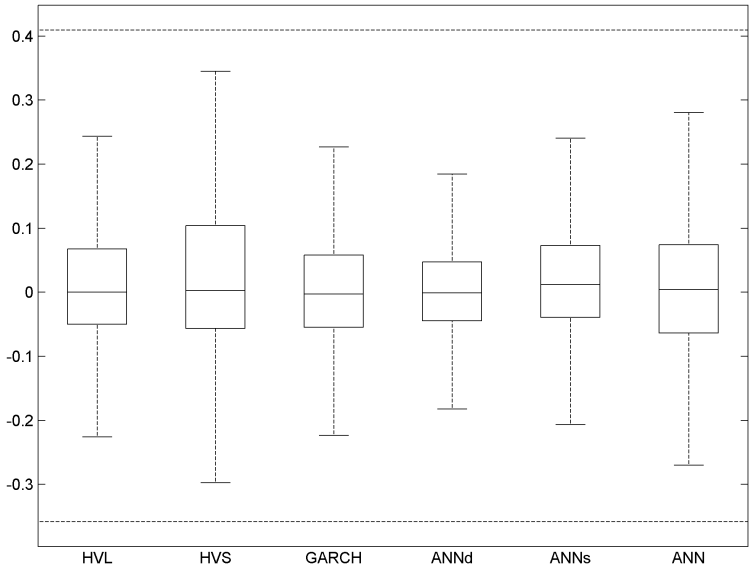

Figure 4.40: Comparison of Option Pricing Errors (In-sample)

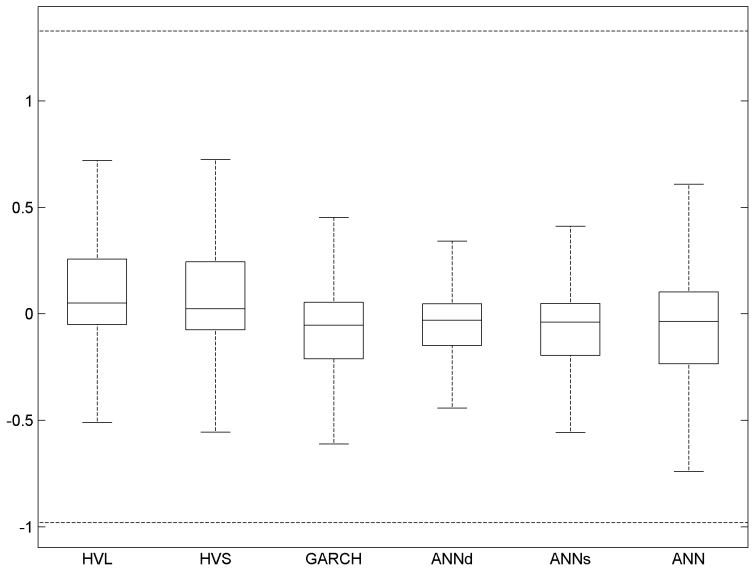

Figure 4.41: Comparison of Option Pricing Errors (Out-of-sample)

As in the previous stages, here too the comparison of errors reveals significant difference, again in part due to the large number of observations. Comparisons across models (see Figures 4.40 and 4.41) and across subsets (see Figures 4.42 to 4.47) show that models have significant differences and behave differently in the later period (out-of-sample). Of the models evaluated, the C^{ANNd} model performs best and shows the lowest inter-quartile range.

Unlike in the previous steps, the pair-wise comparison results are more mixed. While the C^{ANNd} and C^{ANNs} models are better, their means are not significantly different from that of the C^{GARCH} model in-sample (see Table 4.36) and the best performing model C^{ANNd} is not significantly different from it in the out-of-sample period (see Table 4.38) despite the lower error forecasting errors (see Table 4.34). Both outperform the remaining choices, however, and are significantly different. This and the narrower range of errors (ignoring outliers), suggests C^{ANNd} is the preferred model followed by C^{ANNs}.

Similar to the previous forecasting and surface models, the difference between the in-sample and out-of-sample period are quite large and significant. Tables 4.39 to 4.44 show very high levels of significance, i.e. very low p-values. This is also evident in the corresponding box plots in Figures 4.42 to 4.47. Any performance one would have expected based on the in-sample period, would not have been

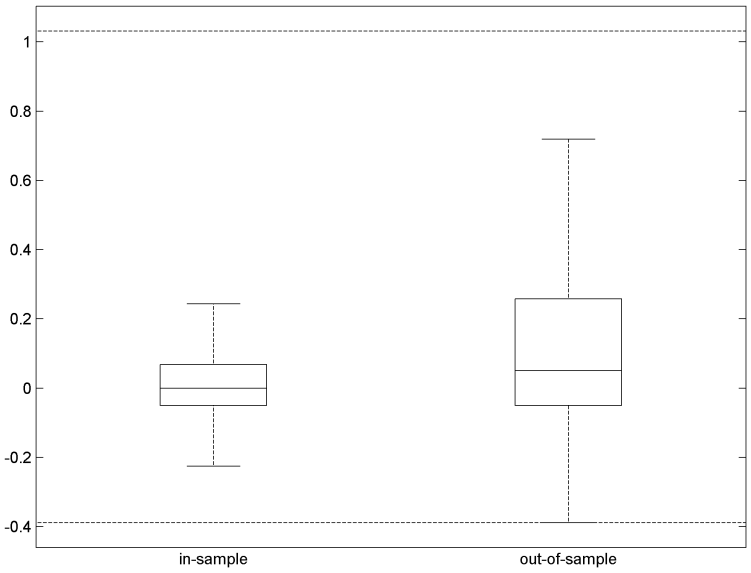

Figure 4.42: Comparison of C^{HVL} Errors Applied to In-sample and Out-of-sample Data

Source	**SS**	**df**	**MS**	**F**	**p**
Groups	689.55	1	689.5450	1923.7900	0
Error	32 054.40	89 430	0.3584		
Total	32 744	89 431			

Table 4.39: ANOVA of C^{HVL} In-sample and Out-of-sample Errors

achieved out-of-sample. As the range of input variables discussed above showed, volatility was different (larger) in the out-of-sample period compared to the in-sample set.

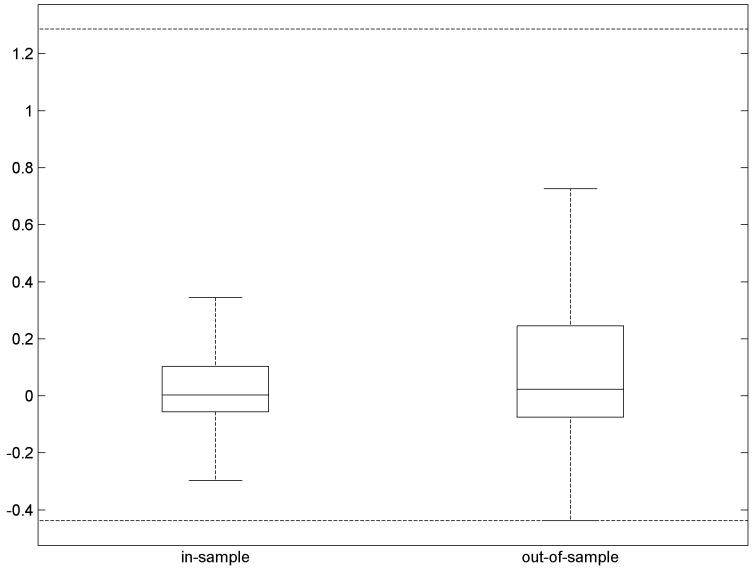

Figure 4.43: Comparison of C^{HVS} Errors Applied to In-sample and Out-of-sample Data

Source	**SS**	**df**	**MS**	**F**	**p**
Groups	566.24	1	566.2370	1160.8900	8.1147×10^{-253}
Error	43 619.80	89 429	0.4878		
Total	44 186.10	89 430			

Table 4.40: ANOVA of C^{HVS} In-sample and Out-of-sample Errors

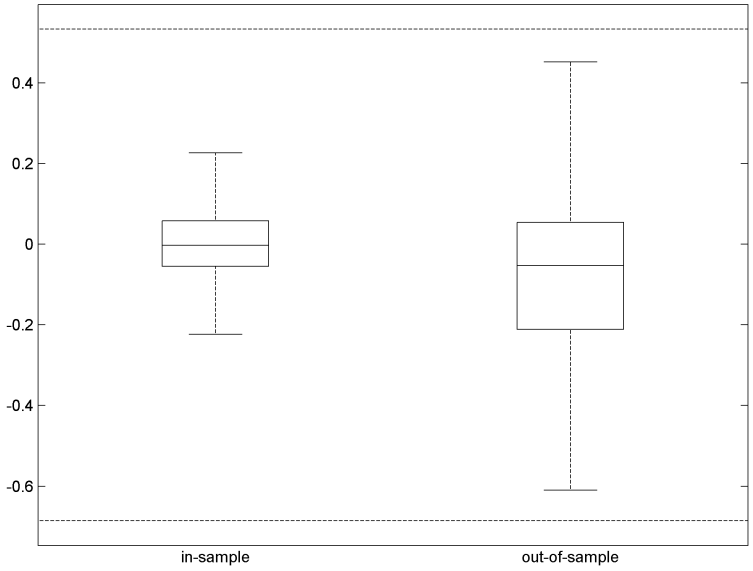

Figure 4.44: Comparison of $C^{\,\mathrm{GARCH}}$ Errors Applied to In-sample and Out-of-sample Data

Source	**SS**	**df**	**MS**	**F**	**p**
Groups	91.90	1	91.9005	266.7760	6.9889×10^{-60}
Error	30 807.30	89 430	0.3445		
Total	30 899.20	89 431			

Table 4.41: ANOVA of $C^{\,\mathrm{GARCH}}$ In-sample and Out-of-sample Errors

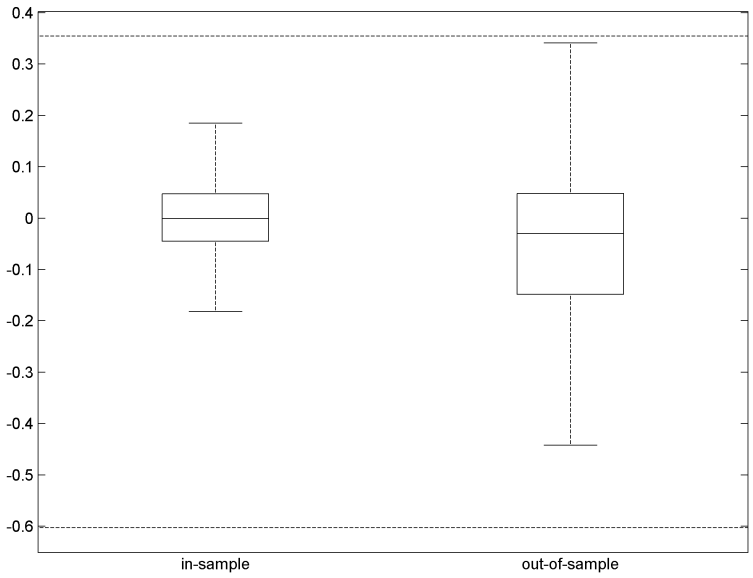

Figure 4.45: Comparison of $C^{\,\mathrm{ANNd}}$ Errors Applied to In-sample and Out-of-sample Data

Source	**SS**	**df**	**MS**	**F**	**p**
Groups	131.70	1	131.6970	659.6120	6.1125×10^{-145}
Error	17 846.40	89 385	0.1997		
Total	17 978.10	89 386			

Table 4.42: ANOVA of $C^{\,\mathrm{ANNd}}$ In-sample and Out-of-sample Errors

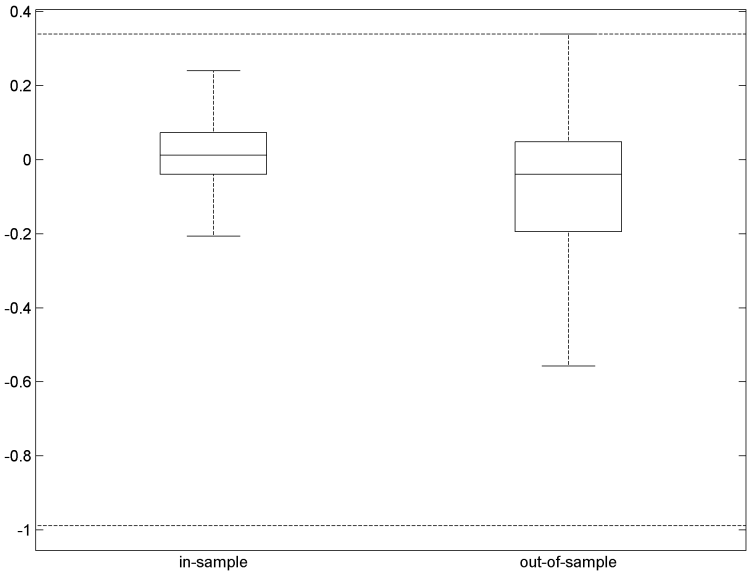

Figure 4.46: Comparison of C^{ANNs} Errors Applied to In-sample and Out-of-sample Data

Source	**SS**	**df**	**MS**	**F**	**p**
Groups	477.27	1	477.2740	1942.0600	0
Error	21 976.50	89 424	0.2458		
Total	22 453.80	89 425			

Table 4.43: ANOVA of C^{ANNs} In-sample and Out-of-sample Errors

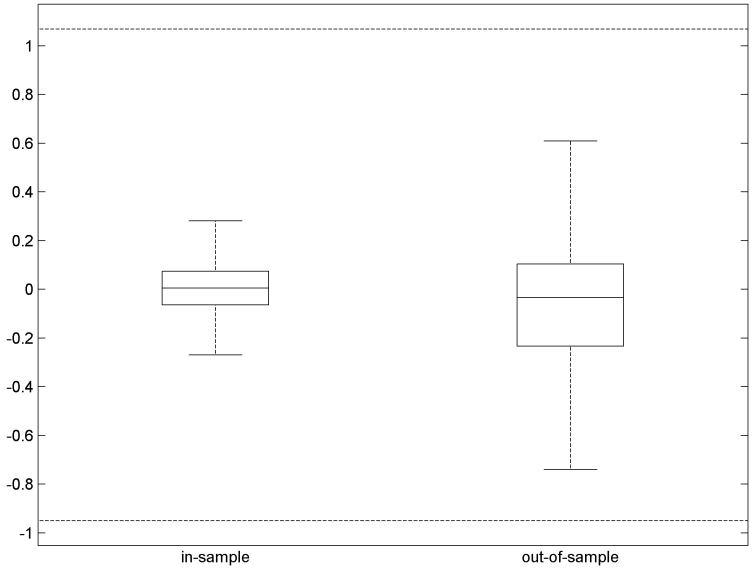

Figure 4.47: Comparison of C^{ANN} Errors Applied to In-sample and Out-of-sample Data

Source	**SS**	**df**	**MS**	**F**	**p**
Groups	18 441.20	1	18 441.20	1417.1700	6.1278×10^{-308}
Error	1.28×10^6	98 475	13.01		
Total	1.30×10^6	98 476			

Table 4.44: ANOVA of C^{ANN} In-sample and Out-of-sample Errors

4.6 Summary and Implications

The results presented in this chapter demonstrate the applicability of ANNs to both volatility modelling and option pricing. The greatest benefits are achieved by focusing on a combination of machine learning techniques and formal modelling. The importance of a formal model to supplement the data-driven approach may be domain-specific, however. Past research has already shown that some markets may benefit more from the addition of ANNs than others.

Interestingly, the models do not show consistency in performance when combined with other techniques. This lack of consistency and the poor fit of the C^{ANN} model may also be a function of the absence of further explanatory variables. In addition to the historical volatility estimates, past return observations were used as they were shown to be beneficial. Only in the case of index options does prior research offer guidance for additional variables. Those could be used with econometric models too, however.

The fact that the best model was an ANN but that it was not trained for the specific target and yet outperforms the models trained for that one purpose is puzzling and leads to a number of research questions for the future. Such research would be particularly valuable since the models were specifically structured such that they share a common variable set, enlarged only when changing the objective. The underperforming network had as much information available as the well-performing one. This raises the question whether structural knowledge can be learned by relatively simple networks.

It is also possible that the volatility forecasting ANN was simply more appropriate for the post-GFC period. The shorter period historical volatility model also performed well out-of-sample.

While possible future directions for researchers are summarised in the next chapter, the research has a number of implications for practitioners:

- The use of machine learning techniques should be considered. While there is no guarantee that this will produce benefits to a business, it is quite possible that such benefits exist in outcome or process.

- Generalisation is particularly problematic even for relatively constrained, theory-driven models. Unexpected events can have a significant impact on the outcomes and may lead to models with limited or no value for practical purposes. Care should thus be taken in the development and application of

any model by monitoring the continuing applicability of a model to market behaviour.

- Modelling knowledge is difficult to transfer from one domain to the other and even from one market to another.

Chapter 5

Conclusion

5.1 Summary of Results and Implications for the Hypotheses

The results as presented in the previous chapter allow for a number of summary conclusions. In regards to the working hypotheses proposed in Chapter 3, a number of direction conclusions can be reached.

Hypothesis 1 *ANNs can forecast volatility more accurately than traditional models.*

The evidence presented in the previous chapter offers some support for this hypothesis. It is certainly possible for volatility forecasting to benefit from non-parametric models. The size of the contribution will depend on the nature of the time series. A judgement needs to be made by the user of the model as to whether the benefits of fitting justify the loss in explanatory power of the model compared to the close-form solutions available through the models of the ARCH-family.

It should be noted that this conclusion is reached in the context of the modelling limitations discussed previously. In particular, additional explanatory variables may help as may rolling-fit (see below for a more detailed discussion).

Hypothesis 2 *Option prices based on ANN models outperform those generated using a traditional pricing model with respect to market prices.*

In regards to the option pricing, the results are clearer. Option pricing does benefit from the use of non-parametric cases even in the case of Australian equity

options. Given the relatively modest and conditional benefits demonstrated by Lajbcygier for index options, this is an important finding as it broadens the investment universe to which machine learning appears to be applicable.

While support for the use can be found in principle, the means by which this is achieved, remain an open question and this clearly requires more research.

However, the networks performed well out-of-sample, which was a particularly difficult period.

5.2 Summary of Contributions

These results demonstrate a certain level of usefulness of ANNs for the purpose of volatility forecasting in general, an aspect that is well understood but also for the specific purpose of valuing and pricing options, one of the contributions of this thesis. At least with respect to the data set used here, the additional effort of training a number of ANNs and calculating options prices based on the ANN output is beneficial.

Although additional research is required as to the possibility of generalising beyond the specific data set, it is a valuable research outcome as the data covered one of the most difficult times for markets in general and covers a market that is structurally transparent, liquid and regulated enough to provide a valid benchmark for many other options markets, while also being on the conservative side of the market efficiency argument. Particular features, especially the mandatory and regulated pension (superannuation) system and the collateral requirements could have a negative impact on the transferability of results to other markets, however.

Another contribution is the observation that the single-step appears to be inferior. This would indicate that any benefits are derived primarily from the volatility forecast rather than the alternative pricing function. This question was the main reason for developing a three-level methodology, to locate the area where progress can be made.

It was also observed that the techniques that have proved to be useful for index option valuation in the past are also helpful when pricing equity options. This outcome was by no means obvious as pointed out before. The fact that there is both a change in the type of security that is used as the underlying as well as a change in option structure, i.e. a change in the exercise right, required an empirical analysis to determine whether results of past research apply here too.

The final empirical contribution relates to the applicability of the comparatively large body of knowledge relating to the USA to the Australian market. As shown above, the results support the hypothesis that machine learning can improve volatility forecasting and option pricing in the Australian equity options market. The results for the US market are consequently less likely to result from particular market microstructure, regulatory or other country-specific aspects.

While each of the above is a valuable contribution in its own right, the value of the empirical results as a whole is particularly important. The results show that machine learning, while shown to work for index options in several markets, can prove to be a valuable tool in option pricing. Care must be taken in the cases and circumstances discussed before and some limitations exist for the application of the methodology used in this research thesis.

In addition to contributions to empirical research, there are also two aspects relating to purely theoretical aspects. Firstly, a viable methodology for extending single time series index option fitting and testing is introduced throughout this thesis and its effectiveness demonstrated for Australian equity options. The panel data approach for options is still somewhat limited and future research will be discussed in the following section. However, the methodology can be used successfully to determine the quality of competing black-box models in the presence of interrelated securities.

The second theoretical contribution is closely related but distinct. In addition to the use of panel data, another novel extension to the testing methodology is introduced. It is well understood within the econometrics field that the quality of the volatility model is related to its objective, i.e. the best model is only the best for a particular purpose in the presence of uncertainty. No clear distinction is made in the literature in this regard. The particular way of formulating the two hypotheses in this thesis combined with the strict methodology developed in it, allows for the fitting and selecting of an appropriate volatility and option pricing function, in combination or independently, with the well-defined and common objective of improving option pricing. This allowed for the specific conclusions to be reached but also for the question to be asked why the combination of volatility forecasting ANN and standard models proves so useful. Without a clear separation of the two aspects, this would not have been possible.

5.3 Future Research

The research outcomes and the process leading to them raise a number of questions, many of which are potentially of great interest to either academics or practitioners. While some have been alluded to in Chapters 2 and 3, others will be introduced in this section.

This thesis followed White (2000b) in regards to the choice of the risk-free rate in the presence of margin requirements. As the author pointed out then, the choice of adjustment depends on the assumptions one makes and to the best of the thesis' author's knowledge, no settlement on the issue has been reached in the literature. This is complicated further by the varied rules across markets and over time governing margin requirements and their implementation. Further research is required both on a theoretical level, i.e. which adjustments to make under what circumstances, and also empirically, i.e. in what way they affect option pricing in practice. This may affect not only the closed-form solutions and approximation methods but also how pricing models are created using machine learning techniques.

While the issue ultimately did not have impact the choices of the research (assuming a rate of 0), one of the operationally trivial but theoretically important questions is the question of what to use as a *proxy for the risk-free rate.* Only for the USA is there a somewhat stronger consensus. While there are some overviews of the choices that are made, it would be of significant value to a large number of researchers in academia and in commercial research organisations and departments, to have access to an extensive review of such choices. The choice obviously depends on the country – or possibly in the case of the EMU on a broader political grouping – but also on the type of research (e.g. security valuation, market efficiency), the research objective (e.g. academic or commercial, theory development or empirical discovery), the time frame and possibly others. A review of past research should be supplemented by a summary of the proxies used at major organisations and by data providers and trading platforms.

A similar situation exists with respect to *sampling frequencies of time series.* Various researchers hold individual beliefs about which sampling frequency is sufficiently low to avoid market microstructure effects. While it is likely that there is no clear point but a continuum, occasional comments in various publications suggest that at least some critical values are observable. Clearly when trading is not continuous as in the foreign exchange markets, one natural window size is the

length of the trading day and possibly of one week. The question remains whether there are equally useful if not equally clear window sizes at higher frequencies.

The broader question therefore is whether and where market microstructure effects end, what sampling frequencies are typically used in which areas of research within finance and how these can be accounted and corrected for. While such research exists for individual asset classes and individual markets, a broader theoretical framework would be desirable. The same applies to the required sample sizes when a continuous time series is not strictly needed. While it is generally preferable to use more rather than less data, there are at least practical limits to this. In addition, inference is difficult for extremely large data sets, where any difference (random or not) appears significant at reasonable significance levels.

Despite extensive research in machine learning, there still appears to be no consensus about the various choices and strategies used to select size and architecture, parameters, learning and evaluation functions, and the *meta-learning strategy for ANNs*. A decision support system, or a process that results in such a system if these are domain-specific, requires additional research. This is particularly true for the meta-learning strategy, i.e. how to select a starting architecture, how to modify it and when to stop. Individual approaches exist such as the one employed in this thesis but also some for self-configuring networks. It is possible that a different network architecture and a different treatment of data results in a pricing network outperforming the volatility forecasting network. A systematic approach to modelling problems in this context is needed to address any issues that arise but also to rule out the possibility of incorrect network architecture in any particular case.

Additional research appears to be needed in this area despite those, i.e. research beyond an extensive review of the literature. Even if self-configuring networks are an option for a particular problem, the question remains whether any of those approaches can be generalised to option pricing and if so, how the most suitable method is to be determined in the context of option pricing and the presence of time-varying parameters.

Furthermore, it is not clear what the limits of machine learning are. Can all structural and functional relationships inherently be modelled and what are the requirements for the process and model.

In addition to these questions, which are intended to broaden the perspective of the researcher, extending it beyond the question being researched, there are several problems and research questions that relate more narrowly to the re-

search outcomes of this thesis. As a natural extension of the research conducted and presented here, it may prove helpful to apply the same principles to additional markets and instruments. The question also arises what to attribute the outperformance to.

From a purely theoretical perspective, there is little room for improvements through the use of machine learning techniques. It should be limited to forecasting volatility and compensating for model misspecifications. The question is thus *what market or return process characteristics offer opportunities for alternative pricing mechanisms* in particular. For example, would such techniques apply to emerging markets as well? The question is significant not merely because of improved pricing functions. Rather, the factors that determine the potential for improvement may also be proxies or even causes for market-related risks for academics or opportunities for excess return generation for practitioners.

On a related point, to the extent that improvements to the pricing functions can be made, it is worth testing if they could be exploited in the context of trading systems and whether they are *economically viable strategies* in the medium or long term. This would require a significantly different methodological approach as the focus is not on hedging, i.e. the risk perspective, but on return generation.

Another question relates to the use of *input variables*. The thesis is intentionally limited to those variables that are known and understood to play a role in option pricing as well as historical measures of volatility of the underlying security. Additional variables may exist that would improve the option pricing. Even those taking a very narrow view of the option pricing mechanism, may find it useful to consider additional input variables, not as input to the option pricing itself but to the volatility forecast that it implies. Factors relating to the expected risk of individual firms, industries or the market as a whole may be worth looking into. The resulting models could be benchmarked against econometric ones from the family of ARCH models with exogenous variables. One candidate variable is the VIX, though it is not clear which markets it can be used for.

This point is of special importance given the methodological limitations of the volatility forecasting models discussed before. Past research often include additional time series as explanatory variables. Given their focus on index volatility, the choice of what to include is somewhat easier to make. However, it also affects the reference model for a valid comparison. As discussed, it would not be possible to conclude that the neural network or any other model is better than an alternative one unless both had access to the same information set, i.e. all else

was equal. ARCH-style models with exogenous variables are certainly one way to make such a comparison in addition to simple regression models using historical volatility and a number of predictors. Equally, an extension of past volatility research into autoregressive neural networks into option pricing could lead to interesting results.

Finally, it would be of considerable interest – though somewhat related to the question about meta-learning strategies – if other machine learning techniques such as SVM regression outperform ANNs in this particular domain and whether they are easier to fit to existing volatility and option data.

Bibliography

Ahn, J. J., D. H. Kim, K. J. Oh, and T. Y. Kim (2012). "Applying option Greeks to directional forecasting of implied volatility in the options market: An intelligent approach". In: *Expert Systems With Applications* 39.10, p. 9315.

Amilon, H. (July 2003). "A Neural Network Versus Black-Scholes: A Comparison of Pricing and Hedging Performances". In: *Journal of Forecasting* 22.4, pp. 317–335.

Amornwattana, S., D. Enke, and C. H. Dagli (Oct. 2007). "A hybrid option pricing model using a neural network for estimating volatility". In: *International Journal of General Systems* 36.5, pp. 558–573.

Anders, U., O. Korn, and C. Schmitt (1998). "Improving the Pricing of Options: A Neural Network Approach". In: *Journal of Forecasting* 17.5-6, pp. 369–388.

Andersen, T. G., T. Bollerslev, and S. Lange (1999). "Forecasting Financial Market Volatility: Sample Frequency vis-à-vis Forecast Horizon". In: *Journal of Empirical Finance* 6.5, pp. 457–477.

Andreou, P. C., C. Charalambous, and S. H. Martzoukos (2008). "Pricing and trading European options by combining artificial neural networks and parametric models with implied parameters". In: *European Journal of Operational Research* 185.3, pp. 1415–1433.

Andreou, P. C., C. Charalambous, and S. H. Martzoukos (2009). "European Option Pricing by Using the Support Vector Regression Approach". In: *Artificial Neural Networks – ICANN 2009: 19th International Conference, Limassol, Cyprus, September 14–17, 2009, Proceedings*. Ed. by C. Alippi, M. M. Polycarpou, and C. Panayiotou. Berlin: Springer-Verlag, pp. 874–883.

Andreou, P. C., C. Charalambous, and S. H. Martzoukos (2010). "Generalized parameter functions for option pricing". In: *Journal of Banking and Finance* 34.3, pp. 633–646.

Andreou, P., C. Charalambous, and S. Martzoukos (2002). "Critical assessment of option pricing methods using artificial neural networks". In: *Artificial Neural*

Networks – ICANN 2002. Ed. by J. Dorronsoro. Vol. 2415. Lecture Notes in Computer Science. Berlin: Springer-Verlag, pp. 1131–1136.

Andreou, P., C. Charalambous, and S. Martzoukos (2006). "Robust Artificial Neural Networks for Pricing of European Options". In: *Computational Economics* 27.2, pp. 329–351.

Aragones, J. R., C. Blanco, and P. G. Estevez (2007). "Neural network volatility forecasts". In: *Intelligent Systems in Accounting, Finance and Management* 15.3-4, pp. 107–121.

Asay, M. R. (1982). "A Note on the Design of Commodity Option Contracts". In: *Journal of Futures Markets* 52, pp. 1–7.

Australian Securities Exchange (June 14, 2012). *Explanatory Note for ASX Option Adjustments*. Explanatory Note.

Australian Securities Exchange (Nov. 2010). *Understanding Options Trading*. Booklet. Exchange Centre, 20 Bridge Street, Sydney NSW 2000.

Audrino, F. and D. Colangelo (2010). "Semi-parametric forecasts of the implied volatility surface using regression trees". In: *Statistics and Computing* 20.4, pp. 421–434.

Baesens, B., R. Setiono, C. Mues, and J. Vanthienen (Mar. 2003). "Using Neural Network Rule Extraction and Decision Tables for Credit-Risk Evaluation". In: *Management Science* 49.3, pp. 312–329.

Barone-Adesi, G. and R. E. Whaley (1987). "Efficient Analytic Approximation of American Option Values". In: *Journal of Finance* 42.2, pp. 301–320.

Barucci, E., U. Cherubini, and L. Landi (1995). *No-Arbitrage Asset Pricing with Neural Networks under Stochastic Volatility*. Tech. rep. Italy: Univ. of Florence.

Barucci, E., U. Cherubini, and L. Landi (1996). "No-Arbitrage Asset Pricing with Neural Networks under Stochastic Volatility". In: *Neural Networks in Financial Engineering: Proceedings of the Third International Conference on Neural Networks in the Capital Markets, London, England, 11-13 October 1995*. Singapore: World Scientific, pp. 3–16.

Bennell, J. and C. Sutcliffe (2004). "Black-Scholes Versus Artificial Neural Networks in Pricing FTSE 100 Options". In: *Intelligent Systems in Accounting, Finance and Management* 12, pp. 243–260.

Bjerksund, P. and G. Stensland (2002). *Closed-Form Valuation of American Options*. Working Paper. NHH.

Black, F. (1976). "The Pricing of Commodity Contracts". In: *Journal of Financial Economics* 3, pp. 167–179.

Black, F. and M. Scholes (1973). "The Pricing of Options and Corporate Liabilities". In: *Journal of Political Economy* 81, pp. 637–654.

Blynski, L. and A. Faseruk (2006). "Comparison of the effectiveness of option price forecasting: Black-Scholes vs. simple and hybrid neural networks". In: *Journal of Financial Management & Analysis* 19.2, p. 46.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroscedasticity". In: *Journal of Econometrics* 31, pp. 307–328.

Boyle, P. P. (1986). "Option Valuation Using a Three Jump Process". In: *International Options Journal* 3, pp. 7–12.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997). *The econometrics of financial markets*. Princeton, New Jersey: Princeton University Press.

Carelli, A., S. Silani, and F. Stella (2000). "Profiling neural networks for option pricing". In: *International Journal of Theoretical and Applied Finance* 3.2, pp. 183–204.

Carverhill, A. P. and T. H. Cheuk (2003). *Alternative Neural Network Approach for Option Pricing and Hedging*. Working paper. Available at SSRN: http://ssrn.com/abstract=480562.

Chang, B. R. (2006). "Applying nonlinear generalized autoregressive conditional heteroscedasticity to compensate ANFIS outputs tuned by adaptive support vector regression". In: *Fuzzy Sets and Systems* 157.13, pp. 1832–1850.

Chang, B. R. and H. F. Tsai (2008). "Forecast approach using neural network adaptation to support vector regression grey model and generalized autoregressive conditional heteroscedasticity". In: *Expert Systems with Applications* 34.2, pp. 925–934.

Charalambous, C. and S. Martzoukos (2005). "Hybrid artificial neural networks for efficient valuation of real options and financial derivatives". In: *Computational Management Science* 2.2, pp. 155–161.

Chen, F. and C. Sutcliffe (Sept. 2011). *Pricing and hedging short Sterling options using Neural Networks*. Tech. rep. DP2011–18.

Chen, F. and C. Sutcliffe (2012). "Pricing and hedging short Sterling options using Neural Networks". In: *Intelligent Systems in Accounting, Finance and Management* 19.2.

Chen, R.-R. and L. Scott (1993). "Pricing Interest Rate Futures Options with Futures-Style Margining". In: *Journal of Futures Markets* 13.1, pp. 15–22.

Chen, S., W. K. Härdle, and K. Jeong (2010). "Forecasting Volatility with Support Vector Machine-Based GARCH Model". In: *Journal of Forecasting* 29, pp. 406–433.

Choi, H., H. Lee, G. Han, and J. Lee (2004). "Efficient option pricing via a globally regularized neural network". In: *Advances in Neural Networks - ISNN 2004, Part 2*. Ed. by F. Yin, J. Wang, and C. Guo. Vol. 3174. Lecture Notes in Computer Science. Berlin: Springer-Verlag, pp. 988–993.

Corrado, C. J. and T. Su (1996). "Skewness and kurtosis in S&P 500 index returns implied by option prices". In: *Journal of Financial Research* 19.2, pp. 175–192.

Cox, J. (1975). *Notes on Option Pricing IL Constant Elasticity of Variance Diffusions*. Working Paper. Stanford University.

Cox, J. C., S. A. Ross, and M. Rubinstein (1979). "Option Pricing: A Simplified Approach". In: *Journal of Financial Economics* 7, pp. 229–263.

Cybenko, G. (1989). "Approximation by superposition of a sigmoidal Function". In: *Mathematics of Control, Signals, and Systems* 2, pp. 303–314.

Dash, G. H., C. R. Hanumara, and N. Kajiji (2003). "Neural network architectures for efficient modeling of FX futures options volatility". In: *Operational Research* 3.1, pp. 3–23.

Derman, E. and I. Kani (1994). "Riding on a Smile". In: *Risk Magazine* 7.2.

Dindar, Z. and T. Marwala (2004). "Option pricing using a committee of neural networks and optimized networks". In: *2004 IEEE International Conference on Systems, Man and Cybernetics*. Vol. 1, pp. 434–438.

Dindar, Z. A. (2004). "Artificial Neural Networks Applied to Option Pricing". MA thesis. Johannesburg: Faculty of Engineering and the Built Environment, University of the Witwatersrand.

Donaldson, R. G. and M. Kamstra (1997). "An artificial neural network-GARCH model for international stock return volatility". In: *Journal of Empirical Finance* 4.1, pp. 17–46.

Drost, F. C. and T. E. Nijman (1993). "Temporal Aggregation of GARCH Processes". In: *Econometrica* 61.4, pp. 909–937.

Dugas, C., Y. Bengio, F. Bélisle, C. Nadeau, and R. Garcia (2001). "Incorporating Second-Order Functional Knowledge for Better Option Pricing". In: *Advances in Neural Information Processing Systems 13*. Ed. by T. Leen, T. Dietterich, and V. Tresp. MIT Press.

Dugas, C., Y. Bengio, F. Bèlisle, C. Nadeau, and R. Garcia (2009). "Incorporating Functional Knowledge in Neural Networks". In: *Journal of Machine Learning Research* 10, pp. 1239–1262.

Dumas, B., J. Fleming, and R. E. Whaley (Mar. 1996). *Implied Volatility Functions: Empirical Tests*. Working Paper 5500. National Bureau of Economic Research.

Dunis, C. L. and X. Huang (2002). "Forecasting and trading currency volatility: an application of recurrent neural regression and model combination". In: *Journal of Forecasting* 21.5, pp. 317–354.

Dupire, B. (1994). "Pricing with a Smile". In: *Risk Magazine* 7.1.

Engle, R. F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation". In: *Econometrica* 50.4, pp. 987–1007.

Fama, E. F. (May 1970). "Efficient Capital Markets: A Review of Theory and Empricial Work". In: *Journal of Finance* 25.2, pp. 383–417.

Fama, E. F. (Dec. 1991). "Efficient Capital Markets: II". In: *Journal of Finance* 46.5, pp. 1575–1617.

Figlewski, S. (1997). "Forecasting Volatility". In: *Financial Markets, Institutions and Instruments* 6.2. Available at SSRN: http://ssrn.com/abstract=8312.

Funahashi, K. (1989). "On the Approximate Realization of Continuous Mappings by Neural Networks". In: *Neural Networks* 2, pp. 183–192.

Galindo-Flores, J. (Sept. 1999). "A framework for comparative analysis of statistical and machine learning methods: an application to the Black–Scholes option pricing model". In: *Computational Finance 1999*. Ed. by A. S. Weigend, A. W. Lo, and B. LeBaron. Cambridge, MA, USA: MIT Press, pp. 635–660.

Garcia, R. and R. Gençay (1998). "Option pricing with neural networks and a homogeneity hint". In: *Decision Technologies for Computational Finance: Proceedings of the Fifth International Conference Computational Finance*. Ed. by A.-P. N. Refenes, A. N. Burgess, and J. E. Moody, pp. 195–205.

Garcia, R. and R. Gençay (2000). "Pricing and hedging derivative securities with neural networks and a homogeneity hint". In: *Journal of Econometrics* 94.1-2, pp. 93–115.

Garman, M. B. and M. J. Klass (1980). "On the Estimation of Security Price Volatilities from Historical Data". In: *Journal of Business* 53.1, pp. 67–78.

Gatheral, J. (2004). *A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives*. Global Derivatives & Risk Management. Madrid 2004. Presentation slides.

Gatheral, J. (2006). *The volatility surface: a practitioner's guide*. Hoboken, New Jersey: John Wiley & Sons.

Gavrishchaka, V. V. and S. B. Ganguli (2003). "Volatility forecasting from multiscale and high-dimensional market data". In: *Neurocomputing* 55.1-2, pp. 285–305.

Gençay, R. and M. Qi (July 2001). "Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging". In: *IEEE Transactions on Neural Networks* 12.4, pp. 726–734.

Gençay, R. and A. Salih (2003). "Degree of Mispricing with the Black-Scholes Model and Nonparametric Cures". In: *Annals of Economics and Finance* 4, pp. 73–102.

Ghosn, J. and Y. Bengio (2002). *Multi-task learning for option pricing with modular neural networks*. Working Paper 2002s-53. Montréal: CIRANO.

Gonzalez Miranda, F. and A. Burgess (1995). "Intraday Volatility Forecasting for Option Pricing Using a Neural Network Approach". In: *Conference on Computational Intelligence for Financial Engineering (CIFEr), April 9–11, 1995, New York City*, p. 31.

Gradojevic, N. and D. Kukolj (2011). "Parametric option pricing: A divide-and-conquer approach". In: *Physica D: Nonlinear Phenomena* 240.19, pp. 1528–1535.

Hagan, P. S., D. Kumar, A. S. Lesniewski, and D. E. Woodward (2002). "Managing smile risk". In: *Wilmott Magazine*, pp. 84–108.

Hajizadeh, E., A. Seifi, M. F. Zarandi, and I. Turksen (2012). "A hybrid modeling approach for forecasting the volatility of S&P 500 index return". In: *Expert Systems with Applications* 39, pp. 431–436.

Hamid, S. and A. Habib (2005). *Can Neural Networks Learn the Black-Scholes Model?: A simplified approach*. Working Paper.

Hamid, S. A. and Z. Iqbal (2004). "Using neural networks for forecasting volatility of S&P 500 Index futures prices". In: *Journal of Business Research* 57.10, pp. 1116–1125.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning*. 2nd ed. Springer.

Haug, E. G. and J. Haug (1996). *Implied Forward Volatility*. Third Nordic Symposium on Contingent Claims Analysis in Finance, Iceland.

Haug, E. G., J. Haug, and A. Lewis (Sept. 2003). "Back to Basics A New Approach to the Discrete Dividend Problem". In: *Wilmott Magazine*.

Haug, E. G. (2007). *The Complete Guide to Option Pricing Formulas*. New York: McGraw-Hill.

Healy, J. V., M. Dixon, B. J. Read, and F. F. Cai (2004). "Confidence limits for data mining models of options prices". In: *Physica A: Statistical Mechanics and its Applications* 344.1, pp. 162–167.

Healy, J., M. Dixon, B. Read, and F. F. Cai (2002). "A data-centric approach to understanding the pricing of financial options". In: *The European Physical Journal B* 27.2, pp. 219–227.

Healy, J. V., M. Dixon, B. J. Read, and F. F. Cai (2007). "Non-parametric extraction of implied asset price distributions". In: *Physica A: Statistical Mechanics and its Applications* 382.1, pp. 121–128.

Healy, J., M. Dixon, B. Read, and F. Cai (2003). "Confidence in data mining model predictions: a financial engineering application". In: *Industrial Electronics Society, 2003. IECON'03. The 29th Annual Conference of the IEEE*. Vol. 2.

Herrmann, R. and A. Narr (1997). *Neural Networks and the Valuation of Derivatives–Some Insights into the implied Pricing Mechanism of German Stock Index Options*. Discussion Paper. Germany: University of Karlsruhe.

Heston, S. (1993). "A closed-form solution for options with stochastic volatility, with application to bond and currency options". In: *Review of Financial Studies* 6, pp. 327–343.

Hornik, K., M. Stinchcombe, and H. White (1989). "Multilayer feedforward networks are universal approximators". In: *Neural Networks* 2, pp. 359–366.

Hornik, K., M. Stinchcombe, and H. White (1990). "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks". In: *Neural Networks* 3, pp. 551–560.

Hossain, A. and M. Nasser (2011). "Recurrent Support and Relevance Vector Machines Based Model with Application to Forecasting Volatility of Financial Returns". In: *Journal of Intelligent Learning Systems and Applications* 3, pp. 230–241.

Hu, M. Y. and C. Tsoukalas (1999). "Combining conditional volatility forecasts using neural networks: an application to the EMS exchange rates". In: *Journal of International Financial Markets, Institutions and Money* 9.4, pp. 407–422.

Huang, S. and T. Wu (2006a). "A Hybrid Unscented Kalman Filter and Support Vector Machine Model in Option Price Forecasting". In: *Advances in Natural Computation*. Ed. by L. Jiao, L. Wang, X.-b. Gao, J. Liu, and F. Wu. Vol. 4221. Lecture Notes in Computer Science. Berlin Heidelberg: Springer, p. 303.

Huang, S. and T. Wu (2006b). "Combining Monte Carlo Filters with Support Vector Machines for Option Price Forecasting". In: *Rough Sets and Current Trends in Computing*. Ed. by S. Greco, Y. Hata, S. Hirano, M. Inuiguchi, S. Miyamoto, H. Nguyen, and R. Słowiński. Vol. 4259. Lecture Notes in Computer Science. Berlin Heidelberg: Springer, p. 607.

Huang, S.-C. (2008). "Online option price forecasting by using unscented Kalman filters and support vector machines". In: *Expert Systems with Applications* 34.4, pp. 2819–2825.

Hull, J. C. (2008). *Options, Futures, and other Derivatives*. 7th ed. Upper Saddle River, New Jersey: Pearson Education.

Hutchinson, J. M. (1994). "A Radial Basis Function Approach to Financial Time Series Analysis". PhD thesis. Massachusetts Institute of Technology.

Hutchinson, J. M., A. W. Lo, and T. Poggio (1994). "A Nonparametric Approach to Pricing and Hedging Derivative Securities via Learning Networks". In: *Journal of Finance* 49.3, pp. 851–889.

Hyup Roh, T. (2007). "Forecasting the volatility of stock price index". In: *Expert Systems With Applications* 33.4, pp. 916–922.

Irie, B. (1988). "Capabilities of three-layered perceptrons". In: *IEEE International Conference on Neural Networks, 24–27 July 1988*. Vol. 1, pp. 641–648.

Jung, K.-H., H.-C. Kim, and J. Lee (2006). "A Novel Learning Network for Option Pricing with Confidence Interval Information". In: *Advances in Neural Networks - ISNN 2006*. Ed. by J. Wang, Z. Yi, J. Zurada, B.-L. Lu, and H. Yin. Vol. 3973. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 491–497.

Kakati, M. (2008). "Option Pricing Using Adaptive Neuro-Fuzzy System (ANFIS)". In: *ICFAI Journal of Derivatives Markets* 5.2, pp. 53–62.

Kelly, D. L. (1994). *Valuing and Hedging American Put Options Using Neural Networks*. Working Paper. Carnie Mellon University and University of California Santa Barbara CA.

Kim, B., D. Lee, and J. Lee (2006). "Local Volatility Function Approximation Using Reconstructed Radial Basis Function Networks". In: *Advances in Neural Networks: Third International Symposium on Neural Networks, Chengdu, China, May 28 – June 1, 2006, Proceedings, Part III*. Ed. by J. Wang, Z. Yi, J. M. Zurada, B.-L. Lu, and H. Yin. Vol. 3973. Lecture Notes in Computer Science. Springer, p. 524.

Ko, P.-C. (2009). "Option valuation based on the neural regression model". In: *Expert Systems with Applications* 36.1, pp. 464–471.

Kolmogorov, A. N. (1957). "On the Representation of Continuous Functions of Several Variables by Superposition of Continuous Functions of one Variable and Addition". In: *Doklady Akademii Nauk USSR* 114, pp. 679–681.

Kuo, C. K. (1993). "The Valuation of Futures-Style Options". In: *The Review of Futures Markets* 10.3.

Lachtermacher, G. and L. A. Rodrigues Gaspar (1996). "Neural Networks in Derivative Securities Pricing Forecasting in Brazilian Capital Markets". In: *Neural Networks in Financial Engineering: Proceedings of the Third International Conference on Neural Networks in the Capital Markets, London, England, 11-13 October 1995*. Singapore: World Scientific, pp. 92–97.

Lajbcygier, P. (2003a). "Comparing Conventional and Artificial Neural Network Models for the Pricing of Options". In: *Neural Networks in Business: Techniques and Applications*. Ed. by K. Smith and J. Gupta. Idea Group Publishing, pp. 220–235.

Lajbcygier, P. (2003b). "Option pricing with the product constrained hybrid neural network". In: *Artificial Neural Networks and Neural Information Processing – ICANN/ICONIP 2003, Joint International Conference ICANN/ICONIP 2003, Istanbul, Turkey, June 26–29, 2003, Proceedings*. Vol. 2714. Berlin: Springer-Verlag, pp. 615–621.

Lajbcygier, P. and A. Flitman (1996). "Comparison of Non-parametric Regression Techniques for the Pricing of Options Using an Implied Volatility". In: *Decision Technologies for Financial Engineering: Proceedings of the Fourth International Conference on Neural Networks in Capital Markets*. Ed. by A. S. Weigend, Y. Abu-Mostafa, and A.-P. N. Refenes. New York: World Scientific.

Lajbcygier, P. (2004). "Improving Option Pricing With the Product Constrained Hybrid Neural Network". In: *IEEE Transactions on Neural Networks* 15.2, pp. 465–476.

Lajbcygier, P. R. and J. T. Connor (1997a). "Improved option pricing using artificial neural networks and bootstrap methods". In: *International Journal of Neural Systems* 8.4, pp. 457–471.

Lajbcygier, P. R. and J. T. Connor (1997b). "Improved option pricing using bootstrap methods". In: *Proceedings of the 1997 IEEE International Conference on Neural Networks*. Vol. 4, pp. 2193–2197.

Lajbcygier, P., C. Boek, M. Palaniswami, and A. Flitman (1996). "Neural Network Pricing of All Ordinaries SPI Options on Futures". In: *Neural Networks in Financial Engineering: Proceedings of the Third International Conference on Neural Networks in the Capital Markets, London, England, 11-13 October 1995*. Singapore: World Scientific, pp. 64–77.

Le Roux, L. and G. Du Toit (2001). "Emulating the Black-Scholes model with a neural networks". In: *Southern African Business Review* 5.1.

Lee, S., J. Lee, D. Shim, and M. Jeon (2007). "Binary Particle Swarm Optimization for Black-Scholes Option Pricing". In: *Knowledge-Based Intelligent Information and Engineering Systems*. Ed. by B. Apolloni, R. Howlett, and L. Jain. Vol. 4692. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 85–92.

Leung, M. T., A.-S. Chen, and R. Mancha (2009). "Making trading decisions for financial-engineered derivatives: a novel ensemble of neural networks using information content". In: *Intelligent Systems in Accounting, Finance & Management* 16.4, pp. 257–277.

Liang, X., H. Zhang, and J. Yang (2006). "Pricing options in Hong Kong market based on neural networks". In: *Proceedings of the 13th International Conference on Neural Information Processing 2006, Part 3*. Vol. 4234. Berlin: Springer-Verlag, pp. 410–419.

Liang, X., H. Zhang, and X. Li (2009). "A Simple Method of Forecasting Option Prices Based on Neural Networks". In: *Next-Generation Applied Intelligence*. Ed. by B.-C. Chien, T.-P. Hong, S.-M. Chen, and M. Ali. Vol. 5579. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 586–593.

Liang, X., H. Zhang, J. Xiao, and Y. Chen (2009). "Improving option price forecasts with neural networks and support vector regressions". In: *Neurocomputing* 72.13, pp. 3055–3065.

Lieu, D. (1990). "Option Pricing With Futures-style Margining". In: *Journal of Futures Markets* 10.4, pp. 327–338.

Lu, J. and H. Ohta (2003a). "A data and digital-contracts driven method for pricing complex derivatives". In: *Quantitative Finance* 3.3, pp. 212–219.

Lu, J. and H. Ohta (2003b). "Digital Contracts-Driven Method for Pricing Complex Derivatives". In: *The Journal of the Operational Research Society* 54.9, pp. 1002–1010.

Malliaris, M. and L. Salchenberger (1993a). "A neural network model for estimating option prices". In: *Applied Intelligence* 3.3, pp. 193–206.

Malliaris, M. and L. Salchenberger (1993b). "Beating the best: A neural network challenges the Black-Scholes formula". In: *Proceedings of the Ninth Conference on Artificial Intelligence for Applications*, pp. 445–449.

Malliaris, M. and L. Salchenberger (1996). "Using neural networks to forecast the S&P 100 implied volatility". In: *Neurocomputing* 10.2, pp. 183–195.

Mantri, J. K., P. Gahan, and B. B. Nayak (2010). "Artificial Neural Networks – An Application to Stock Market Volatility". In: *International Journal of Engineering Science and Technology* 2.5, pp. 1451–1460.

Meissner, G. and N. Kawano (2001). "Capturing the volatility smile of options on high-tech stocks–a combined GARCH-Neural Network approach". In: *Journal of Economics and Finance* 25.3, p. 276.

Merton, R. C. (1973). "Theory of Rational Option Pricing". In: *Bell Journal of Economics and Management Science* 4, pp. 141–183.

Mitra, S. K. (Feb. 2012). "An Option Pricing Model That Combines Neural Network Approach and Black Scholes Formula". In: *Global Journal of Computer Science and Technology* 12 (4).

Mitra, S. (2006). *Improving Accuracy of Option Price Estimation using Artificial Neural Networks*. Working Paper. Institute of Rural Management, India.

Montagna, G., M. Morelli, O. Nicrosini, P. Amato, and M. Farina (2003). "Pricing derivatives by path integral and neural networks". In: *Physica A: Statistical Mechanics and its Applications* 324.1-2, pp. 189–195.

Morelli, M. J., G. Montagna, O. Nicrosini, M. Treccani, M. Farina, and P. Amato (2004). "Pricing financial derivatives with neural networks". In: *Physica A: Statistical Mechanics and its Applications* 338.1-2, pp. 160–165.

Niranjan, M. (1996). "Sequential tracking in pricing financial options using model based and neural network approaches". In: *Advances in Neural Information*

Processing Systems. Ed. by M. Mozer and T. Petsche. Vol. 9. MIT Press, pp. 960–966.

Ormoneit, D. (1999). "A regularization approach to continuous learning with an application to financial derivatives pricing". In: *Neural Networks* 12.10, pp. 1405–1412.

Ou, P. and H. Wang (Feb. 2010a). "Financial Volatility Forecasting by Least Square Support Vector Machine Based on GARCH, EGARCH and GJR Models: Evidence from ASEAN Stock Markets". In: *International Journal of Economics and Finance* 2.1.

Ou, P. and H. Wang (May 2010b). "Predict GARCH Based Volatility of Shanghai Composite Index by Recurrent Relevant Vector Machines and Recurrent Least Square Support Vector Machines". In: *Journal of Mathematics Research* 2.2, pp. 11–19.

Pande, A. and R. Sahu (2006). "A new approach to volatility estimation and option price prediction for dividend paying stocks". In: *WEHIA 2006 – 1st International Conference on Economic Sciences with Heterogeneous Interacting Agents; 15–17 June 2006, University of Bologna, Italy*.

Parkinson, M. (1980). "The Extreme Value Method for Estimating the Variance of the Rate of Return". In: *Journal of Business* 53.1, pp. 61–65.

Peña, I., G. Rubio, and G. Serna (1999). "Why do we smile? On the determinants of the implied volatility function". In: *Journal of Banking & Finance* 23, pp. 1151–1179.

Pires, M. (2005). "American Option Pricing Using Computational Intelligence Methods". MA thesis. University of Witwatersrand.

Pires, M. and T. Marwala (2004). "American option pricing using multi-layer perceptron and support vector machine". In: *Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics*. Vol. 2.

Pires, M. and T. Marwala (2005). "American option pricing using Bayesian multilayer perceptrons and Bayesian support vector machines". In: *Proceedings of the 3rd IEEE International Conference on Computational Cybernetics, 2005*, pp. 219–224.

Poon, S.-H. and C. Granger (2007). "Practical Issues in Volatility Forecasting". In: *Financial Analysts Journal* 61.1, pp. 45–56.

Poon, S.-H. and C. W. J. Granger (2003). "Forecasting Volatility in Financial Markets: A Review". In: *Journal of Economic Literature* 41, pp. 478–539.

Qi, M. and G. Maddala (1996). "Option pricing using artificial neural networks: The case of S&P 500 index call options". In: *Neural Networks in Financial Engineering: Proceedings of the Third International Conference on Neural Networks in the Capital Markets, London, England, 11-13 October 1995*. Singapore: World Scientific, pp. 78–91.

Quek, C., M. Pasquier, and N. Kumar (2008). "A novel recurrent neural network-based prediction system for option trading and hedging". In: *Applied Intelligence* 29.2, pp. 138–151.

Raberto, M., G. Cuniberti, E. Scalas, M. Riani, F. Mainardi, and G. Servizi (2000). "Learning short-option valuation in the presence of rare events". In: *International Journal of Theoretical and Applied Finance* 3.3, pp. 563–564.

Rendleman, R. J. and B. J. Bartter (1979). "Two-State Option Pricing". In: *Journal of Finance* 34, pp. 1093–1110.

Roberts, H. (May 1967). *Statistical versus Clinical Prediction of the Stock Market*. Manuscript. Center for Research in Security Prices, University of Chicago.

Rubinstein, M. (1994). "Implied Binomial Trees". In: *Journal of Finance* 49, pp. 771–818.

Saxena, A. (2008). *Valuation of S&P CNX options: comparison of Black-Scholes and hybrid ANN model*. Paper 162–2008. Presentation at the SAS Global Forum 2008.

Schittenkopf, C. and G. Dorffner (2001). "Risk-neutral Density Extraction from Option Prices: Improved Pricing with Mixture Density Networks". In: *IEEE Transactions on Neural Networks* 12.4, pp. 716–725.

Securities Industry Research Centre of Asia-Pacific (SIRCA) (2010–2012).

Standard and Poor's (July 2011). *S&P/ASX Australian Indices Methodology*.

Teddy, S., E. Lai, and C. Quek (2006). "A Brain-Inspired Cerebellar Associative Memory Approach to Option Pricing and Arbitrage Trading". In: *Neural Information Processing: 13th International Conference, ICONIP 2006, Hong Kong, China, October 3-6, 2006. Proceedings, Part III*. Ed. by I. King, J. Wang, L.-W. Chan, and D. Wang. Vol. 4234. Lecture Notes in Computer Science. Springer, p. 370.

Teddy, S., E. Lai, and C. Quek (2008). "A cerebellar associative memory approach to option pricing and arbitrage trading". In: *Neurocomputing* 71.16-18, pp. 3303–3315.

Thomaidis, N. S., V. S. Tzastoudis, and G. D. Dounias (2007). "A comparison of neural network model selection strategies for the pricing of S&P 500 stock

index options". In: *International Journal on Artificial Intelligence Tools* 16.6, p. 1093.

Tseng, C.-H., S.-T. Cheng, Y.-H. Wang, and J.-T. Peng (2008). "Artificial neural network model of the hybrid EGARCH volatility of the Taiwan stock index option prices". In: *Physica A: Statistical Mechanics and its Applications* 387.13, pp. 3192–3200.

Tung, W. L. and C. Quek (2011). "Financial volatility trading using a self-organising neural-fuzzy semantic network and option straddle-based approach". In: *Expert Systems With Applications* 38.5, pp. 4668–4688.

Tzastoudis, V. S., N. S. Thomaidis, and G. D. Dounias (2006). "Improving neural network based option price forecasting". In: *Advances in Artificial Intelligence, Proceedings*. Ed. by G. Antoniou, G. Potamias, C. Spyropoulos, and D. Plexousakis. Vol. 3955. Lecture Notes in Computer Science. Berlin: Springer-Verlag, pp. 378–388.

Vanstone, B. J. and T. Hahn (July 2008). "Creating short-term stockmarket trading strategies using Artificial Neural Networks: A Case Study". In: *2008 International Conference of Computational Intelligence and Intelligent Systems (ICIIS), World Conference on Engineering (WCE 2008), London 2nd – 4th July, 2008*.

Vanstone, B. J., T. Hahn, and G. Finnie (2009a). "Returns to Selecting Value Stocks in Australia – The Aby Filters". In: *22nd Australasian Finance and Banking Conference 2009*. Available at SSRN: `http://ssrn.com/abstract=1436207`.

Vanstone, B. J. (2005). "Trading in the Australian Stockmarket using Artificial Neural Networks". PhD thesis. School of Information Technology, Bond University.

Vanstone, B. J. and T. Hahn (2013). *Momentum Investing and the GFC: The Case of the S&P/ASX 100*. Available at SSRN: `http://ssrn.com/abstract=2312114`. Submitted to the 26th Australasian Finance and Banking Conference 2013.

Vanstone, B. J., T. Hahn, and G. Finnie (Aug. 20, 2012a). "Developing High-Frequency Foreign Exchange Trading Systems". In: *25th Australasian Finance and Banking Conference 2012*. Available at SSRN: `http://ssrn.com/abstract=2132390`.

Vanstone, B., G. Finnie, and T. Hahn (2009). "Designing short term trading systems with artificial neural networks". In: *Advances in electrical engineering*

and computational science. Ed. by S.-I. Ao and L. Gelman. Springer Science+Business Media B.V., pp. 401–409.

Vanstone, B., G. Finnie, and T. Hahn (Dec. 2012). "Creating trading systems with fundamental variables and neural networks: The Aby case study". In: *Mathematics and Computers in Simulation* 86. The Seventh International Symposium on Neural Networks. The Conference on Modelling and Optimization of Structures, Processes and Systems, pp. 78–91.

Vanstone, B. and T. Hahn (Aug. 23, 2010). *Designing Stockmarket Trading Systems (with and without Soft Computing)*. Harriman House.

Vanstone, B., T. Hahn, and G. Finnie (2009b). "Fundamental investment research – do US results apply to Australian investors?" In: *JASSA* 4, pp. 13–16.

Vanstone, B., T. Hahn, and G. Finnie (2012b). "Momentum returns to S&P/ASX 100 constituents". In: *JASSA* 3, pp. 12–18.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York, NY, USA: Springer-Verlag.

Villiger, R. (Mar. 2006). "Valuation of American Call Options". In: *Wilmott Magazine*.

Wang, C.-P., S.-H. Lin, H.-H. Huang, and P.-C. Wu (2012). "Using neural network for forecasting TXO price under different volatility models". In: *Expert Systems with Applications* 39.5, pp. 5025–5032.

Wang, Y.-H. (2009a). "Nonlinear neural network forecasting model for stock index option price: Hybrid GJR-GARCH approach". In: *Expert Systems with Applications* 36.1, pp. 564–570.

Wang, Y.-H. (2009b). "Using neural network to forecast stock index option price: a new hybrid GARCH approach". In: *Quality & Quantity* 43.5, pp. 833–843.

Wang, H.-W. (2006). "Dual Derivatives Spreading and Hedging With Evolutionary Data Mining". In: *Journal of American Academy of Business* 9.2, p. 45.

Wang, P. (2011). "Pricing currency options with support vector regression and stochastic volatility model with jumps". In: *Expert Systems With Applications* 38.1, pp. 1–7.

White, A. J. (2000a). "A genetic adaptive neural network approach to pricing options: a simulation analysis". In: *Journal of Computational Intelligence in Finance* 6.2.

White, A. J. (2000b). *Pricing options with futures-style margining: a genetic adaptive neural network approach*. New York: Garland Publishing.

White, A. J., G. Hatfield, and R. Dorsey (1998). *A Genetic Adaptive Neural Network Approach to Pricing Options With Futures Style Margining*. Working Paper.

White, H. (1988). "Economic Prediction using Neural Networks: The case of IBM Daily Stock Returns". In: *Second Annual IEEE Conference on Neural Networks*, pp. 451–458.

Xu, L., M. Dixon, B. A. Eales, F. F. Cai, B. J. Read, and J. V. Healy (2004). "Barrier option pricing: modelling with neural nets". In: *Physica A: Statistical Mechanics and its Applications* 344.1, pp. 289–293.

Yao, J., Y. Li, and C. L. Tan (2000). "Option price forecasting using neural networks". In: *Omega* 28.4, pp. 455–466.

Zapart, C. (2002). "Stochastic volatility options pricing with wavelets and artificial neural networks". In: *Quantitative Finance* 2.6, pp. 487–495.

Zapart, C. A. (2003a). "Statistical arbitrage trading with wavelets and artificial neural networks". In: *Proceedings of the 2003 IEEE International Conference on Computational Intelligence for Financial Engineering*, pp. 429–435.

Zapart, C. (2003b). "Beyond Black-Scholes: A Neural Networks-Based Approach to Options Pricing". In: *International Journal of Theoretical and Applied Finance* 6.5, pp. 469–490.

Appendix A

GARCH Model Specifications

The following tables summaries the model specification of the fitted GARCH(1, 1) models including the conditional mean constant (C), the conditional variance (K), and the GARCH and ARCH terms. In addition to the parameter values, the standard errors (SE) and t-statistics (t) are reported.

AMP	**Value**	**SE**	**t**
C	0.0002	0.0003	0.5892
K	1.9209×10^{-5}	2.2871×10^{-6}	8.3985
GARCH	0.7359	0.0188	39.0778
ARCH	0.2641	0.0194	13.6398

ANZ	**Value**	**SE**	**t**
C	0.0016	0.0003	5.4501
K	7.3118×10^{-5}	5.8229×10^{-6}	12.5570
GARCH	0.2634	0.0455	5.7859
ARCH	0.3397	0.0550	6.1720

BHP	**Value**	**SE**	**t**
C	0.0012	0.0004	2.9871
K	8.7124×10^{-6}	3.2465×10^{-6}	2.6836
GARCH	0.9135	0.0168	54.4378
ARCH	0.0596	0.0091	6.5501

BXB	**Value**	**SE**	**t**
C	0.0007	0.0003	2.0608
K	1.7473×10^{-5}	2.1097×10^{-6}	8.2819
GARCH	0.7678	0.0182	42.3044
ARCH	0.2322	0.0204	11.3957

CBA	**Value**	**SE**	**t**
C	0.0007	0.0002	3.1895
K	3.1700×10^{-6}	8.5430×10^{-7}	3.7107
GARCH	0.8858	0.0186	47.7470
ARCH	0.0827	0.0129	6.4331

CGJ	**Value**	**SE**	**t**
C	0.0004	0.0002	1.6147
K	0.0002	6.5848×10^{-6}	23.3010
GARCH	0	0.0356	0
ARCH	0.4170	0.0159	26.2239

FGL	**Value**	**SE**	**t**
C	0.0003	0.0003	1.2352
K	2.7529×10^{-6}	7.0557×10^{-7}	3.9016
GARCH	0.9292	0.0113	82.2604
ARCH	0.0524	0.0078	6.7116

MQG	**Value**	**SE**	**t**
C	0.0013	0.0004	3.5899
K	2.6140×10^{-5}	4.1957×10^{-6}	6.2302
GARCH	0.7694	0.0322	23.8641
ARCH	0.1195	0.0180	6.6561

NAB	**Value**	**SE**	**t**
C	0.0006	0.0003	2.2476
K	8.2778×10^{-6}	1.3237×10^{-6}	6.2537
GARCH	0.8287	0.0215	38.6281
ARCH	0.1115	0.0099	11.2991

QBE	Value	SE	t
C	0.0016	0.0003	5.1791
K	1.2398×10^{-5}	3.0076×10^{-6}	4.1223
GARCH	0.7713	0.0206	37.4108
ARCH	0.2287	0.0133	17.2011

RIN	Value	SE	t
C	0.0008	0.0005	1.8602
K	0.0002	8.6192×10^{-6}	21.9865
GARCH	0	0.0223	0
ARCH	0.3545	0.0210	16.9169

RIO	Value	SE	t
C	0.0009	0.0004	2.4686
K	5.2402×10^{-6}	2.1147×10^{-6}	2.4780
GARCH	0.9097	0.0177	51.3468
ARCH	0.0706	0.0136	5.2021

SGB	Value	SE	t
C	0.0008	0.0002	3.4032
K	2.0097×10^{-6}	5.9153×10^{-7}	3.3975
GARCH	0.9440	0.0091	103.1870
ARCH	0.0360	0.0068	5.2612

SUN	Value	SE	t
C	0.0008	0.0003	2.8540
K	1.0107×10^{-5}	1.9521×10^{-6}	5.1773
GARCH	0.8210	0.0179	45.8350
ARCH	0.1220	0.0121	10.0795

TLS	Value	SE	t
C	0.0001	0.0002	0.5348
K	1.4205×10^{-5}	2.4751×10^{-6}	5.7393
GARCH	0.7383	0.0249	29.6654
ARCH	0.1681	0.0136	12.3975

WBC	**Value**	**SE**	**t**
C	0.0008	0.0002	3.2745
K	5.8446×10^{-6}	1.5840×10^{-6}	3.6898
GARCH	0.8571	0.0242	35.4337
ARCH	0.0912	0.0157	5.7924

WDC	**Value**	**SE**	**t**
C	0.0006	0.0003	1.8913
K	7.0120×10^{-5}	4.0116×10^{-6}	17.4792
GARCH	3.0280×10^{-6}	0.0049	0.0006
ARCH	1.0000	0.0036	274.2910

WES	**Value**	**SE**	**t**
C	0.0008	0.0003	2.5874
K	1.1107×10^{-6}	3.8227×10^{-7}	2.9055
GARCH	0.9576	0.0054	176.3950
ARCH	0.0386	0.0051	7.5275

WOW	**Value**	**SE**	**t**
C	0.0007	0.0003	2.8515
K	2.9186×10^{-6}	7.4501×10^{-7}	3.9176
GARCH	0.9250	0.0128	72.2025
ARCH	0.0540	0.0094	5.7268

WPL	**Value**	**SE**	**t**
C	0.0008	0.0004	2.1814
K	3.0376×10^{-5}	8.0633×10^{-6}	3.7671
GARCH	0.8094	0.0471	17.1671
ARCH	0.0675	0.0164	4.1068