AN INVESTIGATION OF ENSEMBLE METHODS TO IMPROVE THE BIAS AND/OR VARIANCE OF OPTION PRICING MODELS BASED ON LÉVY PROCESSES

A thesis submitted to the University of Manchester for the degree of Doctor of Business Administration in the Faculty of Manchester Business School

2014

By
Oliver Steinki
Manchester Business School
Contents

Abstract 11

Declaration 12

Copyright 13

Dedication 15

Acknowledgements 16

1 Introduction 17
   1.1 Motivation .............................................. 17
   1.2 Outline ................................................... 24

2 Option Pricing Models 26
   2.1 Introduction ............................................. 26
   2.2 Black-Scholes Model ...................................... 27
   2.3 Volatility Models ......................................... 28
      2.3.1 Introduction ......................................... 28
      2.3.2 Time Dependent Volatility Models .................... 29
      2.3.3 Local Volatility Models ............................... 30
2.3.3.1 Constant Elasticity of Variance . . . . . . . . . . 31
2.3.3.2 Implied Local Volatility . . . . . . . . . . . . . . 32

2.3.4 Stochastic Volatility Models . . . . . . . . . . . . . . . . 35
  2.3.4.1 Definition . . . . . . . . . . . . . . . . . . . . . . 35
  2.3.4.2 Stochastic Volatility and its Driving Process . . . 36
  2.3.4.3 Stochastic Volatility Models . . . . . . . . . . . . 38

2.4 Lévy Models . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42
  2.4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . 42
  2.4.2 Pure Continuous Lévy Components . . . . . . . . . . . . 44
  2.4.3 Finite Activity Jumps Lévy Components . . . . . . . . . 44
    2.4.3.1 Merton’s Jump Diffusion Model . . . . . . . . . . 44
    2.4.3.2 Kou’s Jump Diffusion Model . . . . . . . . . . . 45
    2.4.3.3 Eraker’s Jump Diffusion Model . . . . . . . . . . 46
  2.4.4 Infinite Activity Jumps Lévy Components . . . . . . . . 47
    2.4.4.1 Generalized Hyperbolic Process . . . . . . . . . . 47
    2.4.4.2 Normal Inverse Gaussian Process . . . . . . . . . 48
    2.4.4.3 Variance Gamma Process . . . . . . . . . . . . . 48
    2.4.4.4 CGMY Process . . . . . . . . . . . . . . . . . . . 50
    2.4.4.5 Log Stable Model . . . . . . . . . . . . . . . . . . 51

2.5 Time-Changed Lévy Models . . . . . . . . . . . . . . . . . . . . . 54
  2.5.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . 54
  2.5.2 General Setup of Time-Changed Lévy Models . . . . . . 56
  2.5.3 Stochastic Exponential of Time-Changed Lévy Models . . 57
  2.5.4 Ordinary Exponential of Time-Changed Lévy Models . . 58
  2.5.5 Cox-Ingersoll-Ross Stochastic Clock . . . . . . . . . . . . 59
  2.5.6 Gamma-Ornstein-Uhlenbeck Stochastic Clock . . . . . . . 60

2.6 Lévy Models with Leverage . . . . . . . . . . . . . . . . . . . . . . 61
3 Model Uncertainty

3.1 Introduction

3.2 Risk versus Uncertainty

3.3 Sources of Model Uncertainty

3.4 Model Uncertainty Measures

3.4.1 Coherent Model Uncertainty Measures

3.4.1.1 Cont’s Axioms

3.4.2 Worst-Case Measures

3.4.3 Convex Model Uncertainty Measures

3.4.3.1 Cont’s Axioms

3.4.4 Risk-Averaging Measures

3.4.5 Risk-Differencing Measures

3.5 Summary

4 Ensemble Methods

4.1 Introduction

4.2 Rationale Behind Ensemble Methods

4.2.1 Ambiguity Decomposition

4.2.2 Bias, Variance and Covariance

4.3 Classifying Ensemble Methods

4.4 Ensemble Generation

4.4.1 Subsampling

4.4.2 Input Features Manipulation

4.4.3 Output Variable Manipulation

4.5 Ensemble Pruning

4.5.1 Partitioning Based Approaches
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2.3 Model Choice</td>
<td>136</td>
</tr>
<tr>
<td>6.2.4 Pricing Technique</td>
<td>138</td>
</tr>
<tr>
<td>6.2.5 Evaluation Measure</td>
<td>142</td>
</tr>
<tr>
<td>6.2.6 Calibration</td>
<td>142</td>
</tr>
<tr>
<td>6.3 Ensemble Generation</td>
<td>145</td>
</tr>
<tr>
<td>6.4 Ensemble Pruning</td>
<td>148</td>
</tr>
<tr>
<td>6.5 Ensemble Integration</td>
<td>150</td>
</tr>
<tr>
<td>6.6 Summary</td>
<td>152</td>
</tr>
<tr>
<td>7 Empirical Results</td>
<td>153</td>
</tr>
<tr>
<td>7.1 In-Sample-Fit</td>
<td>153</td>
</tr>
<tr>
<td>7.1.1 Stand Alone Model Results</td>
<td>153</td>
</tr>
<tr>
<td>7.1.2 Ensemble Generation Results</td>
<td>157</td>
</tr>
<tr>
<td>7.1.3 Ensemble Pruning Results</td>
<td>159</td>
</tr>
<tr>
<td>7.1.4 Ensemble Integration Results</td>
<td>163</td>
</tr>
<tr>
<td>7.1.5 Summary</td>
<td>166</td>
</tr>
<tr>
<td>7.2 Out-of-Sample-Fit</td>
<td>166</td>
</tr>
<tr>
<td>7.2.1 Stand Alone Model Results</td>
<td>166</td>
</tr>
<tr>
<td>7.2.2 Ensemble Generation Results</td>
<td>170</td>
</tr>
<tr>
<td>7.2.3 Ensemble Pruning Results</td>
<td>172</td>
</tr>
<tr>
<td>7.2.4 Ensemble Integration Results</td>
<td>174</td>
</tr>
<tr>
<td>7.2.5 Summary</td>
<td>176</td>
</tr>
<tr>
<td>8 Conclusions</td>
<td>177</td>
</tr>
<tr>
<td>8.1 Summary of Main Findings</td>
<td>177</td>
</tr>
<tr>
<td>8.2 Contribution to Knowledge</td>
<td>179</td>
</tr>
<tr>
<td>8.3 Suggestions for Future Research</td>
<td>180</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Overview of Option Pricing Models .............................. 64
4.1 Overview of Ensemble Generation Methods ....................... 99
4.2 Overview of Ensemble Pruning Methods .......................... 107
4.3 Overview of Ensemble Integration Methods ...................... 113
5.1 Potential Option Pricing Base Models ........................... 121
5.2 Potential Ensemble Generation Methods .......................... 128
5.3 Potential Ensemble Pruning Methods ............................. 130
5.4 Potential Ensemble Integration Methods .......................... 132
6.1 Analyzed Option Pricing Models ................................. 137
6.2 Analyzed Ensemble Generation Methods .......................... 137
6.3 Analyzed Ensemble Pruning Methods ............................. 139
6.4 Analyzed Ensemble Integration Methods .......................... 151
7.1 In-Sample Stand-Alone Option Pricing Model Results ............. 155
7.2 In-Sample Option Pricing Models: Calibrated Parameter Values .... 156
7.3 In-Sample Ensemble Generation Results .......................... 158
7.4 In-Sample Ensemble Pruning Results ............................. 162
7.5 In-Sample Ensemble Integration Results .......................... 165
7.6 Out-of-Sample Stand-Alone Option Pricing Model Results .......... 169
List of Figures

5.1 Option Pricing Ensemble Framework .......................... 116

7.1 In-Sample-Fit of Analyzed Stand Alone Models I ............... 154
7.2 In-Sample-Fit of Analyzed Stand Alone Models II ............... 155
7.3 In-Sample-Fit of Analyzed Ensemble Generation Methods ...... 158
7.4 In-Sample-Fit of Analyzed Ensemble Pruning Methods .......... 161
7.5 In-Sample-Fit of Analyzed Ensemble Integration Methods ...... 164
7.6 Out-of-Sample-Fit of Analyzed Stand Alone Models I .......... 167
7.7 Out-of-Sample-Fit of Analyzed Stand Alone Models II .......... 168
7.8 Out-of-Sample-Fit of Analyzed Ensemble Generation Methods .. 170
7.9 Out-of-Sample-Fit of Analyzed Ensemble Pruning Methods ...... 172
7.10 Out-of-Sample-Fit of Analyzed Ensemble Integration Methods .. 174
Abstract

This thesis introduces a novel theoretical option pricing ensemble framework to improve the bias and variance of option pricing models, especially those based on Lévy Processes. In particular, we present a completely new, yet very general theoretical framework to calibrate and combine several option pricing models using ensemble methods. This framework has four main steps: general option pricing tasks, ensemble generation, ensemble pruning and ensemble integration. The modularity allows for a flexible implementation in terms of asset classes, base models, pricing techniques and ensemble architecture.

Next, we have analyzed the suggested theoretical option pricing ensemble framework extensively from an empirical perspective, using DAX options as an example. We evaluate the performance of various ensemble combination schemes of different option pricing models and demonstrate their ability to improve the option pricing accuracy. We also compare these results to the classical approach in the literature, which is taking the calibrated model with the lowest measure of pricing error on the calibration dataset at time \( t = t_0 \) to price options of the new dataset at time \( t = t_1 \) and find quite encouraging results for the ensemble approach.
Declaration

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

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

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

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

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

To my parents.
Chapter 1

Introduction

1.1 Motivation

This thesis introduces a novel theoretical option pricing ensemble framework to improve the bias and variance of option pricing models, especially those based on Lévy Processes. Following the publication of the seminal papers of Black and Scholes (1973) and Merton (1973), option pricing became a popular branch of financial theory. This was partly driven by the exponential growth in the amount and variety of traded derivatives over the last 40 years. The significant size of the global derivatives market makes correct pricing of vanilla options and more complex instruments such as American, barrier and basket options or credit, rates and volatility derivatives a relevant and potentially profitable task.

Although the Black-Scholes framework remains the benchmark for practitioners and academics alike, it is now widely accepted that it suffers from a number of shortcomings. This led to the development of more sophisticated models trying to better replicate underlying market processes. The empirical literature has

---

1 The notional amount of all outstanding OTC derivatives is estimated to be USD 647 trillion according to the Bank for International Settlements (as of Dec. 2011).
analyzed especially the following three deviations from the Black-Scholes assumptions (Carr and Wu, 2004):

First, the empirical log returns of financial assets do not behave according to a normal distribution. Empirical evidence shows that return distributions exhibit jumps, are skewed to the left, have higher peaks and heavier tails than those of a normal distribution. Second, return volatilities vary stochastically over time. Third, returns and their volatilities are correlated, usually negatively for equities, fixed income securities and commodities.

The derivatives pricing literature has addressed these three main shortcomings over time: In addition to continuous processes, asset price trajectories were allowed to exhibit jumps. The constant volatility assumption was replaced by deterministic functions or by more flexible stochastic volatility processes. More general Lévy processes were introduced to capture the non-normal return behavior of a wide range of financial securities. Applying stochastic time changes to the Lévy processes can generate stochastic volatilities as well as stochastic higher return moments. Furthermore, the leverage effect was incorporated into models, thus connecting the evolution of volatility with the asset price movement.

These modifications have resulted in a vast number of option pricing models. In particular, asset returns have been modeled as a continuous diffusion process with a so-called local volatility (Derman and Kani, 1994; Dupire, 1994; Rubinstein, 1994) or with a stochastic volatility process (Hull and White, 1987; Johnson and Shanno, 1987; Scott, 1987; Stein and Stein, 1991; Heston, 1993). Similar works assume constant elasticity of variance (Cox and Ross, 1976). Alternatively, models with jumps (Ahn and Thompson, 1988; Amin and Ng, 1993; Bates, 1991; Jarrow and Rosenfeld, 1984; Kou, 2002; Merton, 1976) or models with jumps and stochastic volatility (Bates, 1996a, 2000) have been proposed to capture the
empirical return characteristics described above. Recent jump-models based on Lévy processes include the normal-inverse-Gaussian model (Barndorff-Nielsen, 1997), the variance-gamma model (Carr et al., 1998), generalized hyperbolic processes (Eberlein et al., 1998) and the logstable model (Carr and Wu, 2003a). Furthermore, combinations or generalizations of the above have been introduced: affine jump-diffusion models (Duffie et al., 2000) as well as models based on time-changed Lévy processes (Carr et al., 2003; Carr and Wu, 2004).

A number of empirical investigations (Carr et al., 1998; Carr and Wu, 2003a; Schoutens, 2003; Dahlbokum, 2007) attest the more recent models - which adjust for all or parts of the three main weaknesses of the Black-Scholes framework - superior in-sample performance (defined as improved fitting of observed option prices) compared to the Black-Scholes benchmark. However, this improvement comes at the expense of lower analytical tractability and increased model complexity and variance. Furthermore, a common feature of all derivative pricing models is that they must be calibrated to existing market prices of options to avoid arbitrage opportunities before using them. Model calibration is an inverse problem that consists of determining model parameters to reproduce the observed option prices. This calibration process can be challenging due to several reasons (Gupta, 2009):

First, there can be inconsistencies or mispricings in the option prices used for calibration, so that no parameter set could correctly reproduce all the observed option prices. Second, there might not be enough observable market prices to uniquely determine the calibration parameters. Furthermore, the stability of the calibrated models can be quite low - a small change in price of the underlying may lead to significant changes in the calibration parameters of the model (high

\[2\text{For example, the implied volatility smile is an empirical fact that cannot be replicated in the Black-Scholes framework.}\]
Despite these issues, most of the current literature on calibrating derivative pricing models focuses mainly on finding the single best-fit set of parameters that minimizes a certain evaluation measure given a set of market prices and one or several option pricing models, i.e. reducing the in-sample bias of the models to be calibrated. Only recently, attention has been paid to the issue of finding the right and avoiding the wrong parameters/models (model uncertainty), how stable these parameters/models are (robustness/variance minimization) and how one can optimize the pricing task in this context of model uncertainty (Gupta and Reisinger 2011).

Research on model uncertainty and model risk in finance only started to attract academic interest in the mid 1990’s, when mathematicians (Derman 1996; Green and Figlewski 1999; Hull and Suo 2002; Branger and Schlag 2004; Kerkhof and Melenberg 2004; Cont 2006; Gupta 2009; Gupta et al. 2010; Branger et al. 2012) began investigating the issue of model risk/uncertainty and market observers (Elliott 1997) reported on substantial losses due to model uncertainty (Gupta 2009). While initial investigations into misestimation have been done, the literature on misspecification in the context of derivatives pricing seems relatively limited to date.

Huber (1981) defines robustness as “signifying insensitivity to small deviations from the assumptions”. Hence, one needs to define the following two terms: what is considered a small deviation from the assumptions and what measure is used to judge insensitivity? In our context, we would define a small deviation from

\footnote{Defined as parameter set which results in the minimal error measure between model and market prices.}

\footnote{Model uncertainty can be partitioned into two types: misspecification (choosing the wrong model) and misestimation (choosing the wrong parameters).}
the assumptions as the delta in price of the underlying from time $t = t_0$ to time $t = t_1$ and measure insensitivity as continued accurate pricing performance (defined by some measure of pricing error) at time $t = t_1$ with the option pricing model calibrated at time $t = t_0$. In practice, a relevant question for a trader would be how well a calibrated model performs on new data points compared to simpler models such as the Black Scholes benchmark. It is not clear whether the generally more accurate pricing performance of calibrated recent option pricing models will extend to new datasets (due to the generally higher variance of the more recent models) and which of the calibrated models will perform best (defined as having the lowest pricing error) on the new data. One way to approach the issues of model uncertainty and robustness is the use of ensemble methods.

Ensemble methods attempt to maintain or even improve the bias while increasing robustness/reducing variance at the same time.

Ensemble methods make predictions according to a combination of base model predictions and differ in the type of model aggregation scheme utilized. Already in 1818, Laplace claimed that “In combining the results of these two methods, one can obtain a result whose probability law of error will be more rapidly decreasing” (as quoted in Clemen (1989)). Mendes-Moreira et al. (2012) define ensemble methods as follows: ”Ensemble learning is a process that uses a set of models, each of them obtained by applying a learning process to a given problem. This set of models (ensemble) is integrated in some way to obtain the final prediction.”

Ensemble methods are expected to be useful when there is uncertainty in choosing the best model and when it is important to avoid large errors (Armstrong, 2001). Both criteria apply to our context of derivatives pricing. One necessary condition for an ensemble to be more accurate than any of its individual members is to have diverse and accurate base models, accurate being defined as having a low
bias. Diverse base models make different errors on new data points (in our context option prices) and are therefore not perfectly correlated (Dietterich, 2000b). Hence, the strength of an ensemble in our context depends on the strength of the individual option pricing models and a measure of correlation between their pricing errors on new data.

Previous literature has analyzed the success of ensemble methods extensively in diverse application fields such as medicine (Polikar et al., 2008), climate forecasting (Stott and Forest, 2007), image retrieval (Tsoumakas et al., 2005) and astrophysics (Bazell and Aha, 2001). Several empirical studies have shown that ensemble predictions can often be much more accurate than the forecasts of the base learners (Freund and Schapire, 1996; Bauer, 1999; Dietterich, 2000a), reduce variance (Breiman, 1996a; Lam and Suen, 1997) or bias and variance (Breiman, 1998; Kong and Dietterich, 1995) and various authors have suggested justifications for the success of ensemble methods from a theoretical perspective (Kittler et al., 1998; Allwein et al., 2000; Kleinberg, 2000).

Various alternatives to form ensembles have been proposed; see Valentini (2002) for a detailed review. The oldest technique of aggregating different models are Bayesian decision rules (Suen and Lam, 2000), which have only recently been investigated in the context of derivatives pricing (Gupta, 2009; Gupta et al., 2010; Gupta and Reisinger, 2011). Other papers proposed to combine base learners using relatively simple mathematical operations, such as minima, maxima, averages, weighted averages, medians or products of the base models (van Breukelen et al., 1998; Kuncheva, 1997; Schapire, 1990). Other methods attempt to directly influence the accuracy of the ensemble by improving the accuracy of and diversity among the base models, e.g. by changing the input data. Resampling methods (Breiman, 1996a; Freund and Schapire, 1996) and feature selection methods (Ho,
are examples of methods that alter the input data. Mixture of expert methods (Jacobs, 1995) choose different base models for distinctive input regions, whereas output coding methods (Dietterich and Bakiri, 1995) modify the way the different models are aggregated. Test-and-select methods (Roli, 2001; Sharkey et al., 2000) choose the appropriate base learners to include in the final ensemble by judging their performance on a test dataset.

Although the various ensemble methods mentioned above are quite diverse, they have one thing in common: they emphasize the way the models are combined rather than focusing too much on the properties of the base models. However, several authors have shown that the success of ensemble methods is also dependent on the particular choice of base learners, especially their robustness (Breiman, 1998) as well as their individual accuracy (Hansen and Salamon, 1990; Giacinto and Roli, 2001) and diversity (Kuncheva, 2004). These papers motivate the hypothesis that ensembles built upon recently introduced Lévy process based option pricing models (that have generally a lower bias) could perform better, defined as having a lower pricing error on new datasets, than ensembles based on earlier option pricing models.

Model uncertainty and the issue of robustness have only recently gained attention in the derivatives pricing literature (Gupta et al., 2010). To the best of our knowledge, an investigation whether ensemble methods can reduce model uncertainty and improve pricing performance in the context of derivatives pricing models has only been published for Bayesian methods (Gupta, 2009). This is a shortcoming of the current literature and the motivation for the research question addressed in this thesis:

*Can Ensemble Methods Improve the Bias or Variance of Option Pricing Models based on Lévy Processes?*
Hence, this thesis suggests a novel theoretical option pricing ensemble framework to reduce model uncertainty, enhance model robustness and improve pricing accuracy on new datasets, improved pricing accuracy being defined as having a lower measure of pricing error. The merits of this proposed methodology are analyzed empirically by first describing the exact study design and then evaluating the performance of various ensembles of different option pricing models on the DAX index. These results are then compared to the classical approach in the literature, which takes the calibrated model with the lowest measure of pricing error on the calibration dataset at time $t = t_0$ to price options of the new dataset at time $t = t_1$.

1.2 Outline

This thesis aims to demonstrate that the use of ensemble methods can reduce model uncertainty, enhance model robustness and improve pricing performance for derivatives pricing models based on Lévy processes. It is organised as follows:

In the first chapters, the literature is reviewed in three areas relevant to the proposed novel method: the evolution of the derivatives pricing literature, the issue of model uncertainty and the use of ensemble methods to reduce bias and/or variance.

We will start in chapter 2 with the evolution of derivatives pricing models and review most of the above mentioned models, stress their key features and empirical characteristics, outline which shortcomings of the Black-Scholes model they address and how they build upon each other. This detailed introduction is necessary in order to use them as potential base models for the ensemble architecture.\footnote{Ensemble architecture is used as a general term to describe all three steps of the ensemble process: ensemble generation, ensemble pruning and ensemble integration.}
introduce later in this thesis. In chapter 3, we will have a closer look at model un-
certainty. We will differentiate between risk and uncertainty, investigate sources
of model uncertainty and introduce measures to quantify it. Chapter 4 will con-
tain a review of the use of ensemble methods in various areas and look at the
various combination schemes which have been proposed to improve forecasting
accuracy and robustness of ensemble models. To do so, we will split the ensemble
process intro three steps: ensemble generation, ensemble pruning and ensemble
integration following Mendes-Moreira et al. (2012) and analyze each step sepa-
rately. We will also explain the rationale why ensembles are able to outperform
base models.

The main contribution of this thesis is split over chapters 5 to 7 and contains
theoretical as well as empirical elements. First, a novel theoretical option pricing
ensemble framework which uses ensemble methods to increase the performance
of derivatives pricing models is proposed in chapter 5. Second, chapter 6 outlines
details of the empirical study design we have used to analyze the novel option
pricing ensemble framework: the pricing and calibration procedures as well as the
corresponding ensemble methods. Third, chapter 7 presents the empirical results
of the theoretical framework introduced in Chapter 5. We analyze and interpret
the performance of several ensemble combination schemes of various derivatives
pricing models using options on the German DAX index as an example. We an-
alyze these results not only in-sample (as most of the academic literature does),
but also analyze them from a practitioners point of interest, which is taking
the calibrated model with the lowest measure of pricing error on the calibration
dataset at time $t = t_0$ to price options of the new dataset at time $t = t_1$. The
final chapter, Chapter 8, concludes the main findings of the thesis, interprets the
results and recommends possible areas of future research. The references follow.
Chapter 2

Option Pricing Models

2.1 Introduction

This section describes the evolution of the derivatives pricing literature over almost 40 years. We will use the introduced models later in this thesis to build ensembles of option pricing models and analyze their behaviour with respect to in-sample and out-of-sample pricing accuracy, stability and model risk measures as well as hedging performance. We will cover a wide range of different classes of option pricing models. This is motivated by the fact that one of the key factors to build successful ensembles is the diversity of the base models, which are in the context of this thesis option pricing models. More recent option pricing models based on Lévy processes tend to have a better pricing accuracy, which should also help to build successful ensembles.

This chapter is outlined as follows: We start with a short introduction of the famous Black Scholes model (section 2.2), followed by an investigation of volatility models (section 2.3) which were introduced in the 1980’s. After this, we introduce
option pricing models based on Lévy processes (section 2.4). We will not only investigate finite and infinite activity Lévy processes, but also examine time-changed Lévy models (section 2.5) and Lévy models with leverage (section 2.6).

2.2 Black-Scholes Model

The seminal work of Black and Scholes (1973) and Merton (1973) results in a closed form solution for European options, requiring only one variable which is not observable, volatility. The main contribution of the BSM model was the construction of a dynamically replicating portfolio of the payoffs of a European option to value it.

The Black-Scholes portfolio replication argument considers a market with only two securities, i) a riskless bond \( B(t) \) and ii) a stock \( X(t) \) with the following equations:

\[
\begin{align*}
    dB(t) &= r(B)dt \\
    dX(t)/X(t) &= \mu dt + \sigma dW
\end{align*}
\]

with constants \( \mu, \sigma \) and \( r \) representing the risk-free rate. Assuming no arbitrage, Black and Scholes determined the closed form value of a European call on a non dividend paying stock as follows:

\[
C(X(t), t, T, r, \sigma, K) = X(t)\psi(d_1) - Ke^{(-r(T-t))}\psi(d_2) \tag{2.3}
\]

where

\[
d_1 = \frac{ln(X(t)/K) + (r + \frac{1}{2}\sigma^2)(T - t)}{\sigma\sqrt{T-t}}, \tag{2.4}
\]

\[1\] In 1997, the importance of their contribution was highlighted when Myron Scholes and Robert Merton were awarded the Nobel Prize for economics (Fischer Black died in 1995).
\[
d_2 = \frac{\ln(\frac{X(t)}{K}) + (r - \frac{1}{2}\sigma^2)(T - t)}{\sigma\sqrt{T - t}}
\]
\[
= d_1 - \sigma\sqrt{T - t}.
\] (2.5)

In \(C(X(t), t, T, r, \sigma, K)\), \(t\) represents the time when \(C\) is calculated, \(T\) is the expiration date, \(\psi(\cdot)\) is the standard cumulative distribution function and \(K\) represents the strike price. The reader is referred to \cite{Poon2005} for a more detailed derivation of the Black-Scholes formula.

### 2.3 Volatility Models

#### 2.3.1 Introduction

"Suppose we use the standard deviation of possible future returns on a stock as a measure of its volatility. Is it reasonable to take that volatility as a constant over time? I think not.” - Fisher Black, 1976

The question of determining the fair value of an option goes back to 1900, when Louis Bachelier \cite{Bachelier1900} described in his PhD thesis "Theorie de la Speculation" how to use Brownian motion to model the fluctuations of asset prices and tried to value fixed income options. Similar to Black and Scholes \cite{BlackScholes1973}, he assumed the market to be a fair game and used the normal distribution to estimate the odds that prices will move. But Bachelier’s model became suspect once concepts like utility theory and risk aversion became popular in economics. If investor’s are risk averse, shouldn’t that lead to an option price lower than its expected value, similar to stocks? This puzzle was solved by Black and Scholes for a world where stocks drift according to a geometric Brownian motion process and where volatility is constant, as their elegant solution is independent of investor’s risk aversion and return expectation. As the above quotation suggests however, the
assumption of constant volatility was challenged from an early time onwards. Empirical investigations and theoretical considerations (such as the leverage effect) found that constant volatility is not a very realistic assumption. This led to the development of various dynamic volatility models, which can be classified into three categories:

- Time dependent volatility models
- Local volatility models
- Stochastic volatility models

We will turn to all of these categories on the next pages. We will analyze here only those models which assume a Wiener process to capture the drift process of the underlying, often closely following Mitra (2009). Models, which allow the underlying to follow other processes than geometric Brownian motion, will be reviewed from section 2.4 onwards.

### 2.3.2 Time Dependent Volatility Models

One of the earliest empirical observations about volatility was that it varies with an option’s expiry date (Wilmott 1998):

\[
dX/X = \mu dt + \sigma(t)dt
\]  \hspace{1cm} (2.6)

Consequently, Merton (1973) proposed a model where volatility was no longer assumed to be constant but time dependent. The option price is still calculated according to the Black Scholes formula (equation 2.3), except that \( \sigma \) is replaced by \( \sigma_c \) where:

\[
\sigma_c = \sqrt{\frac{1}{T-t} \int_t^T \sigma^2 \tau d\tau}
\]  \hspace{1cm} (2.7)
This results in the following $d_1$ and $d_2$:

\[
d_1 = \frac{\left( \log \left( \frac{X}{K} \right) + \mu (T - t) + \frac{1}{2} \int_t^T \sigma^2(\tau) \, d\tau \right)}{\sqrt{\int_t^T \sigma^2(\tau) \, d\tau}}, \tag{2.8}
\]

\[
d_2 = \frac{\left( \log \left( \frac{X}{K} \right) + \mu (T - t) - \frac{1}{2} \int_t^T \sigma^2(\tau) \, d\tau \right)}{\sqrt{\int_t^T \sigma^2(\tau) \, d\tau}}. \tag{2.9}
\]

Equation 2.7 converts $\sigma(t)$ to its constant volatility equivalent $\sigma_c$ over the time period $t$ to $T$. The distribution of $X(t)$ is the following:

\[
\log \left( \frac{X_T}{X_t} \right) \sim N \left( \left( \mu - \frac{1}{2}\sigma_c^2 \right)(T - t), \sigma_c^2(T - t) \right) \tag{2.10}
\]

The constant volatility $\sigma_c$ changes in value as $t$ and $T$ change, which enables the time dependent volatility to map the empirically observed increase in implied volatility for longer maturities (keeping the strike constant). Although time dependent volatility models were an improvement to the assumption of constant volatility, they were not able to explain all empirical characteristics of volatility, such as the volatility smile or the leverage effect, as volatility did not vary with price.

### 2.3.3 Local Volatility Models

Option pricing models which vary only with the level of price are called level dependent volatility models, whereas models which vary with price and time are called local volatility models. Formally, local volatility models are described as follows:

\[
dX/X = \mu \, dt + \sigma(X, t) \, dW \tag{2.11}
\]
The term local is used as volatility is assumed to be known with certainty "locally”, which means when $X$ and $t$ are known. In contrast, volatility is never known with certainty for stochastic volatility models. One advantage of local volatility models over stochastic volatility models is that no second source of randomness is added to the models. Unlike stochastic volatility models with a second source of randomness, local volatility models are complete, which allows theoretically perfect hedging. A second advantage is that local volatility models can be calibrated to fit implied volatility surfaces perfectly (Dupire 1997). In addition, local volatility models are able to account for some theoretical arguments and empirical observations time dependent volatility models fail to incorporate (for example the leverage effect). We will now review some common local volatility models.

2.3.3.1 Constant Elasticity of Variance

Cox and Ross (1976) introduced the constant elasticity of variance model (CEV):

$$dX/X = \mu dt + \sigma(X)dW,$$

$$\sigma(X) = aX^{(n-1)}$$

with $0 \leq n \leq 1$ and $a > 0$ as constants. $n = 0$ results in the model initially proposed by Bachelier (with Brownian motion), while $n = 1$ obtains the geometric Brownian motion from equation 2.2. We can also see that $n$ captures the leverage effect as volatility increases with a lower stock price. In contrast to the local volatility models from Dupire (1994) and Derman and Kani (1994), the CEV model is analytically tractable. Additionally, with the right calibration of $n$ and $a$, the CEV model can fit volatility smiles (Beckers 1980). The CEV has
been modified and extended by various researchers over the years. Examples include the interest rate model of Cox et al. (1985) (see also section 2.5.5). Lo et al. (2000) derive the option pricing formula for the CEV with time dependent parameters (turning a level-dependent model into a local model) and Hsu et al. (2010) determined the probability density function of the CEV model. Schroder (1989) restates the CEV model in terms of a Chi-square distribution, which allows the derivation of closed form solutions. Hagan et al. (2002) extended the CEV model to the stochastic SABR model (SABR stands for stochastic alpha, beta and rho). The SABR model describes the behavior of a forward price $F$ on a given underlying under stochastic volatility and its risk-neutral dynamics are:

$$dF(t) = \sigma(t)F^\beta(t)dW_1^Q,$$

$$d\sigma(t) = \alpha\sigma(t)dW_2^Q,$$

with $\text{corr}(dW_1, dW_2) = \rho \in [-1, 1], \quad 0 \leq n \leq 1, \quad \alpha > 0.$

### 2.3.3.2 Implied Local Volatility

In the early 90’s, Dupire (1994), Derman and Kani (1994) and Rubinstein (1994) developed local volatility concepts where the volatility smile is based on option data. They proved that there exists a unique risk-neutral process which is consistent with option data. Dupire’s approach uses option prices to get the implied distribution of the underlying equity. The price of a European call under risk-neutral valuation is equal to:

$$C = e^{-rT}\int_K^{\infty} (X(T) - K)p(X(T))dx,$$

31
with \( p(X(T)) \) describing the risk-neutral probability density function (pdf) of \( X(T) \). Breeden and Litzenberger (1978) demonstrated with the use of equation 2.16 that the risk-neutral cumulative density function \( F(\cdot) \) at \( K \) is equal to:

\[
\frac{\partial C}{\partial K} = -e^{-rT}F(X(T) \geq K),
\]

(2.17)

and the risk-neutral pdf \( p(X(T) = K) \) is given by:

\[
\frac{\partial^2 C}{\partial K e^{rT}} = p(X(T) = K).
\]

(2.18)

This demonstrates that we can derive the risk-neutral density \( p(X(T)) \) from market prices of options. Dupire then applied \( p(X(T)) \) from Breeden and Litzenberger’s equation 2.18 to the Fokker-Planck equation to obtain the Dupire equation:

\[
\frac{\partial C}{\partial T} = \sigma^2(X,T)\frac{X^2}{2} \frac{\partial^2 C}{\partial X^2} - (r - D)X \frac{\partial C}{\partial X} - DC,
\]

(2.19)

rearranging for \( \sigma(X,T) \) gives:

\[
\sigma(X,T) = \sqrt{\frac{\frac{\partial C}{\partial T} + (r - D)X \frac{\partial C}{\partial X} + DC}{\frac{X^2}{2} \frac{\partial^2 C}{\partial X^2}}}.
\]

(2.20)

Hence, one can extract the local volatility \( \sigma(X,T) \) from market prices of options. However, as we can see in equation 2.20, we need partial differentials with respect to \( T \) and \( K \). This requires a continuous set of options data for all \( K \) and \( T \), which is unrealistic. Furthermore, especially far out-of-the-money and far in-the-money options tend to suffer from low liquidity, which leads to wide bid-ask spreads. Pinder (2003) analyzes the Australian equity options market and finds that bid-ask spreads are influenced by trading volume, time to expiry and volatility. Since option data is discrete, one needs to interpolate to obtain a
continuous data set. However, local volatility calculations react sensitively to the different interpolation methods (Wilmott, 2000).

The approaches of Derman and Kani (1994) and Rubinstein (1994) are quite similar to Dupire’s model, the main difference is that they use the discrete time binomial (or trinomial) tree framework (going back to Cox et al. (1979)) instead of the continuous time framework Dupire applies. The tree fitting approach of Derman/Kani and Rubinstein has the advantage that it does not rely on interpolation methods. A detailed derivation of Dupire’s and Derman and Kani’s approach can be found in Javaheri and Cont (2005). Britten-Jones and Neuberger (2000) further extended the model of Derman and Kani. They model stochastic volatility (not only local volatility as Derman and Kani do) by fitting a tree model which is consistent with market prices of options. Their model is parameterized by up and down movements of option prices; with the probabilities for each branch inferred from empirical option data. Local volatility models imply that option prices contain all relevant information about the volatility of the underlying. However, empirical evidence (e.g. Christensen (1998)) demonstrates the differences between realized and implied volatility. Dumas et al. (1998) show further that the implied volatility surface of S&P 500 options changes over time and conclude that the above mentioned local volatility models have no better pricing performance than the classical Black Scholes Model. As we will see in the next section, stochastic volatility models offer more time-homogeneous volatility smiles.
2.3.4 Stochastic Volatility Models

2.3.4.1 Definition

We have seen above that local volatility models have several advantages compared to time dependent volatility models. However, they still possess several weaknesses. One of them is the perfect correlation (positive or negative) of volatility with stock price, which does not hold empirically. Also, returns empirically exhibit volatility clustering, a pattern that does not necessarily occur under local volatility. These weaknesses motivated the development of stochastic volatility models. Stochastic volatility can be defined as follows, assuming that the underlying $X$ behaves according to the following stochastic differential equation:

$$\frac{dX}{X} = \mu dt + \sigma(\omega)dW_1$$

(2.21)

Volatility is stochastic if $\sigma(\omega)$ is governed by a stochastic process which is driven by another random process than the one that drives the underlying (usually modeled as another Wiener process). The probability space $(\Omega, \mathcal{F}, P)$ is $\Omega = ([0, \infty) : \mathbb{R}^2)$, with the filtration $\{\mathcal{F}_t\}_{t \geq 0}$ representing all information of two Wiener processes $(W_1, W_2)$. As volatility cannot be negative, the process governing $\sigma(\omega)$ must be positive for all values. The two Wiener processes have instantaneous correlation $\rho \in (-1, 1)$ defined as:

$$corr(dW_1(t), dW_2(t)) = \rho dt.$$  

(2.22)

Empirical investigations have shown that $\rho$ tends to be close to 0 in FX markets, but negative in equity markets (due to the leverage effect, see e.g. [Black (1976)]). Although $\rho$ can evolve with time, we assume it is constant for now and will review the leverage effect in more detail in section [2.6]. The main difference to
local volatility is that stochastic volatility is driven by a second random process, whereas local volatility models have only one source of randomness. Stochastic volatility models capture more empirically observed return behaviours than local volatility models do (Lewis 2000). They can also generate return distributions that have fatter left tails and are more peaked than the normal distribution, with \( \rho \) controlling the degree of tail asymmetry (Durham 2007).

Furthermore, Renault and Touzi (1996) demonstrated that volatility independent of the return distribution always produces volatility smiles\(^2\). Also, stochastic volatility is able to model volatility’s empirical dependence on the measured time scale, which should not occur under the two other classes of volatility models. In addition, historic volatility exhibits higher variability than expected from time-dependent or local volatility models, with a significant change in volatility during the 1987 crash (Schwert 1989). However, stochastic volatility models exhibit also some disadvantages compared to local volatility models. First, they introduce a second, non-tradable source of randomness, which means that the market is no longer complete (Wu 2008). This has a negative influence on the practical applications of hedging under stochastic volatility models. In addition, stochastic volatility models tend to be analytically less tractable, often having no closed form solutions and leaving simulation as the only method to calculate option prices.

2.3.4.2 Stochastic Volatility and its Driving Process

Stochastic volatility models can be categorized into different groups according to the process governing the evolution of volatility. Many stochastic volatility models

\(^2\)The reverse does not hold, volatility smiles do not necessary imply stochastic volatility as we have seen above.
assume a mean-reverting stochastic volatility process of the following form:

\[
\sigma = f(Y),
\]

\[
dY = \alpha(m - Y)dt + \beta dW_2
\]  \hspace{1cm} (2.23)

where \( \beta \) is a constant with \( \beta \geq 0 \), \( m \) is the long run value to which \( \sigma \) returns and \( \alpha \) describes the rate of mean reversion. The inclusion of mean reversion in volatility modeling allows the modeling of volatility clustering. Volatility clustering is a characteristic feature of financial time series \( \text{Engle} \ 1982 \) which cannot be captured by local or time dependent volatility models and led to the development of the family of GARCH models \( \text{Bollerslev} \ 1986 \). Equation 2.24 is an Ornstein-Uhlenbeck (OU) process in \( Y \) with the following solution:

\[
Y(t) = m + (Y(0) - m)e^{-\alpha t} + \beta \int_0^t e^{-\alpha t} dW_2,
\]  \hspace{1cm} (2.25)

where the distribution of \( Y(t) \) is described by:

\[
Y(t) \sim N(m + (Y(0) - m)e^{-\alpha t}, \frac{\beta^2}{2\alpha}(1 - e^{-2\alpha t})).
\]  \hspace{1cm} (2.26)

An alternative process to describe mean-reverting volatility would be to use the Cox - Ingersoll - Ross \( \text{Cox and Rubinstein} \ 1985 \) or Feller processes (introduced in \( \text{Feller} \ 1951 \); see \( \text{Fouque et al.} \ 2000 \) for an application) instead of an OU process. The reader is referred to \( \text{Dahlbokum} \ 2007, 2010 \) for a detailed analysis of a wide range of mean-reverting processes. We will revisit these processes in section 2.5 in the context of time-changed Lévy processes.
2.3.4.3 Stochastic Volatility Models

Several researchers have proposed a stochastic evolution of volatility. We will describe below the main models and highlight their differences and similarities.

Johnson and Shanno Model

Johnson and Shanno (1987) were among the first researchers to propose a stochastic volatility model:

\begin{align*}
    dX &= \mu_1 X dt + \sigma X^n dW_1, \, n \geq 0, \quad (2.27) \\
    d\sigma &= \mu_2 \sigma dt + \sigma^k \beta dW_2, \, k \geq 0, \quad (2.28)
\end{align*}

with \( \text{corr}(dW_1(t), dW_2(t)) = \rho dt \). Johnson and Shanno recommend a Monte Carlo method to determine the option price under the stochastic volatility process using risk-neutral valuation. They find that their method generates option prices which are consistent with the empirical observations of volatility smiles and higher implied volatility for longer maturities. However, the Johnson and Shanno model only allows for numerical solutions.

Scott Model

Scott (1987) suggests the use of geometric Brownian motion to model the price path of the underlying stock and a mean-reverting process for the volatility:

\begin{align*}
    \frac{dX}{X} &= \mu dt + \sigma dW_1, \quad (2.29) \\
    d\sigma &= \alpha(m - \sigma) dt + \beta dW_2, \quad (2.30)
\end{align*}

Scott’s proposition of an Ornstein-Uhlenbeck process to model the volatility is based on empirical stock price returns. He also describes a parameter estimation technique based on moments matching. He assumes a correlation of zero between
the two processes to facilitate the computation of option prices. Similar to Johnson and Shanno, his option prices are calculated by Monte Carlo simulations and have only numerical solutions. Furthermore, he observes only a marginal improvement in pricing performance vs. the BSM framework.

Hull-White Model

Hull and White (1987) consider the case where volatility is modelled as follows:

\[
\frac{dX}{X} = \mu dt + \sigma dW_1, \quad (2.31)
\]

\[
\frac{d\sigma^2}{\sigma^2} = \mu_2 dt + \beta dW_2. \quad (2.32)
\]

One of the distinctive features of the Hull-White model is that it provides a closed form solution to European option prices when the correlation between the two processes is zero and for a given risk-neutral measure \( Q \). Let us define:

\[
\tilde{\sigma} = \frac{1}{T-t} \int_t^T \sigma^2(s) ds, \quad (2.33)
\]

where \( \tilde{\sigma} \) is a random variable. They calculate the option price using the standard BS formula under a risk-neutral measure \( Q \) and with volatility \( \sqrt{\tilde{\sigma}^2} \), which is equivalent to:

\[
C(t, X, K, T, \sigma(\omega)) = E^Q[C_{BS}(t, X, K, T, \sqrt{\tilde{\sigma}^2})]. \quad (2.34)
\]

Equation 2.34 provides results consistent with empirically observed currency options with \( \rho \approx 0 \) (Melino and Turnbull 1991). Monte Carlo simulation is used when volatility is correlated with the return of the underlying. Similar to the Scott and the Stein and Stein Model, the assumption of \( \rho \approx 0 \) works reasonably
well for currency markets (e.g. Melino and Turnbull (1991)), but is an unrealistic assumption for equities due to the leverage effect (Black, 1976; Christie, 1982).

**Stein and Stein Model**

Stein and Stein (1991) suggest to model volatility with a mean-reverting process based on tractability and empirical considerations:

\[
\frac{dX}{X} = \mu dt + \sigma dW_1, \tag{2.35}
\]

\[
d\sigma = -\alpha(m - \sigma)dt + \beta \sigma dW_2 \tag{2.36}
\]

with \( \rho = 0 \). Although equation (2.36) allows volatility to be negative, Stein and Stein highlight that only \( \sigma^2 \) is ever used in calculation. In contrast to Johnson and Shanno (1987) and Scott (1987), but similar to Hull and White (1987), the Stein and Stein model has a closed form solution for option prices for a given risk-neutral measure and assuming zero correlation between \( dW_1 \) and \( dW_2 \). They are the authors noticing the possibility to use Fourier transforms in order to solve the Black Scholes partial differential equation. However, the assumption that volatility is uncorrelated with spot returns fails to capture important skewness effects that arise from such correlations.

**Heston Model**

Heston’s model (Heston, 1993) is a significant contribution as it was the first stochastic volatility model that has an analytical solution for European options when \( dW_1 \) and \( dW_2 \) are correlated (depending on several assumptions on the risk-neutral measure \( Q \)).

\[
\frac{dX}{X} = \mu dt + \sigma dW_1, \tag{2.37}
\]
\[ d\sigma^2 = \alpha(m - \sigma^2)dt + \beta \sigma dW_2. \] (2.38)

The modeling of \( \sigma^2 \) is based on the square-root process of Cox and Rubinstein (1985), which was originally used to model the short term behavior of interest rates. The underlying mean-reverting process was originally introduced by Feller (1951). Due to market incompleteness, we have to specify a risk-neutral measure to be able to price an option under risk-neutral valuation. Heston chooses this risk-neutral measure based on economic justifications. He then uses Fourier transforms to find an analytical solution for the option. The analytical solution of Heston’s model has attracted various researchers to further extend the model. Scott (1997) includes stochastic interest rates, whereas Pan (2002) allows for stochastic dividends and Bates (1996b) combines the stochastic volatility process with jumps in the price trajectory of the underlying currency.

The volatility models reviewed so far have a common shortfall: they are all short term volatility models and do not take into account any long term or broader economic factors/trends, which also have an impact on volatility.

While the above considered models have more flexible statistical properties than the original Black-Scholes model, they share with it the property of continuous price trajectories. Bates (1996a) is the only model considered so far that allows for discontinuous price behaviour. However, there is empirical evidence (e.g. Cont (2001)), that empirical stock prices exhibit jumps. We will therefore review now models based on Lévy processes, which can capture the behaviors of return innovations on a full range of financial securities.
2.4 Lévy Models

2.4.1 Introduction

Tankov (2004) lists several factors motivating the inclusion of jumps in models of asset price trajectories:

First, time series of asset prices clearly show the presence of jumps. Second, in models where the price process behaves according to Brownian motion, the probability of a large move over a short time interval is very small, unless one chooses an artificially high value for the volatility parameter. Consequently, these models price short term out-of-the-money (OTM) options much lower than observed market prices of such options. On the other hand, models which include jumps, have a non-negligible probability that - even when the time to maturity is very short - an OTM option will end up in the money. Third, jump models allow risk managers to quantify and account for the risk of significant price movements over short time periods, movements which appear non-existent in a pure diffusion framework. Fourth, models with continuous price behavior generally imply complete markets or markets, which can be made complete by adding one or two additional instruments, like in stochastic volatility models. Since in such markets the payoff structure of options can be replicated (which was the main contribution of Black and Scholes (1973)), options can be considered as redundant assets. However, in real markets, due to the presence of jumps, perfect hedging as assumed by BS is impossible and options allow market participants to hedge risks they cannot hedge with the underlyings alone.

Lévy processes are named after the French mathematician Paul Lévy (1937), who developed the theory of infinite divisibility. They are made of two basic building blocks: a diffusion part and/or a jump part and have been used in mathematical
finance for more than a century. Bachelier (1900) introduced Brownian motion as a stock price model. Osborne (1959) extended this model by proposing the exponential of Brownian motion to model stock prices, based on psychological arguments. Samuelson (1965) suggested the use of the same process in a more systematic manner. The advantage of geometric Brownian motion compared to Brownian motion is that it is nonnegative. Mandelbrot (1963) was one of the first to suggest an exponential non-normal Lévy process to model asset prices. He found that the logarithm of relative price changes of cotton follows a heavy-tailed distribution. Consequently, he suggested to use a symmetric $\alpha$-stable Lévy motion with a tail index $\alpha < 2$ (which is a pure-jump process) instead of Brownian motion. Some years later, Press (1967) recommended an exponential Lévy process model with a non-stable distribution. The resulting log price process is a combination of a Brownian motion and an independent compound Poisson process with normally distributed jumps. Merton (1976) follows Press’ model in adding a compound Poisson process to the Brownian motion assumption of the Black Scholes model to describe the price trajectory of the underlying in an option pricing context. Brownian motion generates normally distributed return innovations. The added compound Poisson process in the Merton model results in a return distribution, which is a mixture of normal distributions weighted by Poisson probabilities. Starting in the 1990’s, several researchers introduced more general Lévy processes, which can generate a wider range of distributional behaviors through various types of jump specifications. A compound Poisson process results in a finite number of jumps within a finite time interval. Such a jump process is suitable to model rare and large events such as market crashes or corporate defaults. However, empirical observations (Cont 2001; Tsay 2005) show that asset prices often display many small jumps. A general Lévy process cannot only model continuous price trajectories via Brownian motion and rare and large
jump events via a compound Poisson process, but it can also generate frequent jumps of different sizes (Wu 2008). The following section will review the different examples of Lévy processes indicated above.

2.4.2 Pure Continuous Lévy Components

The classic BS model from section 2.2 is the only Lévy model without jumps. As a direct result of the Lévy-Khintchine formula (Cont and Tankov 2004), the characteristic function of $X_t$ is given by:

$$\phi(u) = \exp\left[i\mu u - \frac{\sigma^2 u^2}{2}\right], \quad (2.39)$$

with the risk-neutral drift $\mu = -\frac{\sigma^2}{2}$. In the BS world, the martingale measure $Q$ is unique and we don’t have to consider any other pricing rules.

2.4.3 Finite Activity Jumps Lévy Components

2.4.3.1 Merton’s Jump Diffusion Model

The Jump Diffusion Model of Merton (1976) (MJD) is a Lévy process with finite-activity, i.e. with a finite number of jumps during a finite period of time. It adds a compound Poisson jump component with mean arrival rate $\lambda$ to the continuous sample path assumption of the BS model and is one of the first trials to adjust for the leptokurtic probability density function as the jumps fatten the return distribution’s tails. Merton interprets jumps as the arrival of new information with more than a marginal effect on price movements. The jump size is log-normally distributed with mean $\mu_J$ and variance $\delta$, conditional on one jump
occurring. The Lévy density of the compound Poisson process is given by:

\[
\pi(x) = \lambda \frac{1}{\sqrt{2\pi\delta}} \exp\left(-\frac{(x - \mu_J)^2}{2\delta}\right) \tag{2.40}
\]

The characteristic function for the compound Poisson jump is:

\[
\phi_t(u) = E[e^{iuX_t}] = \exp[i\mu ut - \frac{\sigma^2 u^2}{2} - \lambda t(e^{i\mu u} - \frac{1}{2}\lambda^2 u^2)] \tag{2.41}
\]

with the risk-neutral drift \(\mu = -\frac{\sigma^2}{2} - \lambda \xi; \xi = e^{\mu_J + \frac{\lambda^2}{2}} - 1 \) \(\text{Cont and Tankov, 2004}\).

Finite-activity jump processes are used extensively to model large but rare events in the financial literature. In the context of credit-risk, Poisson processes have been used to model the random arrival rate of default events \(\text{Duffie et al., 2003, Duffie and Singleton, 1999, Lando, 1998, Carr and Wu, 2009}\) use a Poisson jump process with a recovery rate of zero to estimate the impact of corporate default on the stock price, which jumps to zero upon arrival of the default event. In a separate paper, \(\text{Carr and Wu, 2007b}\) use a Poisson jump process with a random recovery rate to estimate the impact of sovereign default on the currency of the defaulting country. Upon arrival of the default event, the currency price jumps down by a random amount.

### 2.4.3.2 Kou’s Jump Diffusion Model

\(\text{Kou, 2002}\) extends the Jump Diffusion Model of Merton by proposing a double exponential conditional distribution for the jump size. The characteristic exponent can be written as:

\[
\psi(u) = -\lambda[(\beta_+ - iu)^{-1} - \beta_+^{-1} + (\beta_-iu)^{-1} - \beta_-^{-1}]. \tag{2.42}
\]
Given this specification, the jump arrival rate increases monotonically and has a decreasing jump size. The different exponential coefficients $\beta_+$ and $\beta_-$ result in asymmetry between up and down jumps. Both Kou’s and Merton’s models can generate leptokurtic features (although the kurtosis from Kou’s model is more pronounced), implied volatility smiles, and analytical solutions for call and put options. Contrary to Merton’s model, the double-exponential jump specification of Kou can lead to analytic approximations for finite horizon American options thanks to the memoryless property of exponential density and to analytical solutions for some path-dependent options, such as look-back, barrier and perpetual American options (Kou and Wang, 2004).

2.4.3.3 Eraker’s Jump Diffusion Model

Eraker et al. (2003) and Eraker (2004) introduce a model with discontinuous correlated jumps in stock prices and stock price volatility, and with state dependent arrival intensity. They introduce compound Poisson jumps into the stochastic volatility process, assuming that volatility can only jump upwards and that the jump size is controlled by a one-sided exponential density. The characteristic exponent is given by:

$$\psi(u) = \lambda(1 - \frac{1}{1 - iu\eta}).$$  \hfill (2.43)

They find strong evidence for the simultaneous occurrence of volatility jumps and negative price jumps in the context of no-arbitrage models. Although compound Poisson jumps capture rare and large events such as market crashes and corporate defaults, empirical evidence suggests that asset prices actually display many small jumps (Tsay, 2005). These types of behaviors are better captured by infinite-activity jumps, which generate infinite number of jumps within any finite time interval and to which we will turn in the next section.
2.4.4 Infinite Activity Jumps Lévy Components

2.4.4.1 Generalized Hyperbolic Process

Barndorff-Nielsen and Halgreen (1977) introduced the generalized hyperbolic (GH) distribution as a model for the grain-size distribution of wind-blown sand. In the following years, it was discovered that the hyperbolic shape occurs in a wide range of empirical studies, e.g. in biology, geology or relativity theory (Barndorff-Nielsen and Shephard, 2002). Following a suggestion by Barndorff-Nielsen, Ernst Eberlein and several coworkers began an investigation if hyperbolic laws apply also to finance (Eberlein and Keller, 1995). The characteristic exponent of the GH process is given by:

\[ \psi(u) = -\ln\left(\frac{\sqrt{\alpha^2 - \beta^2}}{\sqrt{\alpha^2 - (\beta + iu)^2}}\right)\lambda K_{\lambda}\left(\delta\sqrt{\alpha^2 - (\beta + iu)^2}\right) \]

(2.44)

where \( \alpha, \beta \) and \( \lambda \) are shape parameters and \( \delta \) is a scale parameter. Eberlein and Keller studied the thicker tailed hyperbolic distribution extensively, including the corresponding option pricing theory and practice (Eberlein et al., 1998; Eberlein, 2000). Eberlein and his coworker’s research was possible as Barndorff-Nielsen and Halgreen (1977) proved the infinite divisibility of the generalized inverse Gaussian distribution. Sorensen and Bibby (2003) describe the applications of generalized hyperbolic processes in finance, while Bingham and Kiesel (2001) analyze the use of hyperbolic processes in finance.

The GH process is a pure jump process which incorporates the Variance Gamma (VG, see section 2.4.4.3) and Normal Inverse Gaussian processes (NIG, see section 2.4.4.2) as special cases. The reader is refered to Schoutens (2003) for discussion of another special case, called the hyperbolic process. Prause and Raible, both PhD students under Eberlein, have plenty of information on generalized...
hyperbolic Lévy Processes in their dissertations (Prause, 1999; Raible, 2000).

2.4.4.2 Normal Inverse Gaussian Process

The normal inverse Gaussian (NIG) process has been introduced by Barndorff-Nielsen (1997). Rydberg (1997a,b) investigates both the calibration and the simulation of NIG processes. Barndorff-Nielsen observed that whenever deviations from the hyperbolic shape occurred, they usually exhibited heavier tails than the hyperbolic distribution. This led him to consider more closely another one of the generalised hyperbolic laws, the normal inverse Gaussian (NIG) distribution. Apart from the paper of Sichel (1973) on the distribution of the size of diamonds, the NIG had not received much attention previously. Barndorff-Nielsen found that it does not only fit a much wider range of financial securities data, but it also has various advantageous mathematical properties the hyperbolic distribution does not possess (Barndorff-Nielsen 1995, 1997, 1998a,b). The Normal Inverse Gaussian (NIG) distribution has a characteristic exponent given by:

\[
\psi(u) = -\lambda(\sqrt{\alpha^2 - \beta^2}) - \sqrt{\alpha^2 - (\beta + iu)^2},
\]

where \(\alpha\) and \(\beta\) are shape parameters \((\alpha > 0, -\alpha < \beta < \alpha)\), and \(\delta > 0\) is a scale parameter. The NIG model is a special case of the GH processes with \(\lambda = -1/2\). It is the result of time-changing a Brownian motion with the inverse Gaussian subordinator (Tankov, 2004).

2.4.4.3 Variance Gamma Process

The class of Variance Gamma (VG) distributions was introduced by Madan and Seneta (1987) as a model for stock returns. They considered the symmetric case
along with Madan and Seneta (1990) and Madan and Milne (1991). Carr et al. (1998) extended the model to a general case with skewness. They are also the first to apply this stochastic process to option pricing and illustrate the superior pricing performance of their model over BS on S&P 500 option data. Carr et al. (2002) further generalize the VG model in their CGMY model, of which the VG is a special case with a power coefficient of 0. The VG process can be constructed in one of the following three ways: The first one is subordinating a BM to a Gamma process in time parameter. The second way is specifying the density function of the process increments to have a VG distribution; the third alternative is specifying the Lévy measure. We will discuss below the first alternative and refer to Binkowski (2008) for a discussion of the two other methods.

In the first approach, the VG process is obtained by sampling Brownian motion with drift at random times, time increments which themselves are determined by another Lévy process. The two additional parameters are the drift of the Brownian motion and the volatility of the time change. These additional parameters provide control over the skewness and kurtosis of the return distribution. In every finite time interval, infinite many jumps can occur. The VG process is described as a subordinated Brownian motion:

$$X_t = G_t + \sigma W(G_t), \quad (2.46)$$

where $G$ is a subordinator, i.e. an increasing Lévy process. The characteristic exponent of the VG model is the following:

$$\phi(u) = (1 - iu\theta\nu + \frac{1}{2}\sigma^2\nu u^2)^{-\frac{1}{\nu}} \quad (2.47)$$
Under the following parameterization, the VG model can be regarded as a special case of the CGMY model (section 2.4.4.4):

\[ C = \lambda, G = \frac{\mu_-}{\nu_-}, M = \frac{\mu_+}{\nu_+}, Y = \alpha = 0. \] (2.48)

### 2.4.4.4 CGMY Process

Carr et al. (2002) developed a continuous time model which allows jumps and diffusions to exhibit either finite or infinite-activity. They define a pure jump process to be one of finite/infinite-activity if the number of price jumps in any time period is finite/infinite. The parameters of their model allow the jump component to have either finite or infinite variation. It therefore synthesizes the features of several continuous time models such as diffusion- (Black and Scholes, 1973; Merton, 1973), pure jump- (Cox and Ross, 1976) and jump-diffusion processes (Merton, 1976). They conclude that the statistical and risk-neutral processes for equities are pure jump processes with infinite-activity and finite variation. The CGMY distribution has the following characteristic exponent:

\[ \psi(u) = CT(-Y)\lbrack M^Y - (M - iu)^Y + G - (G + iu)^Y \rbrack, \] (2.49)

with \( C, G, M > 0 \) and \( Y \in [-1, 2] \).

The same parameterization as seen in section 2.4.4.3 is also found in the literature.

The above family of distributions is also called the KoBoL family by some authors (following Novikov 1994, Koponen 1995, Boyarchenko and Levendorskii 1999, 2000, 2002). Initially, the term truncated Lévy process was used based on the physics literature on turbulence. Truncated Lévy flights were first mentioned by Mantegna and Stanley (1994), who pioneered also their use in finance (Mantegna
The power coefficient \( Y = \alpha \) controls the arrival frequency of small jumps and therefore the jump type. With the power coefficient equal to \(-1\), the Lévy density becomes the double-exponential specification of Kou (see section 2.4.3.2). The CGMY model generates finite-activity jumps as long as \( \alpha < 0 \). When \( \alpha \in [0, 1] \), the model generates jumps with infinite-activity and finite variation; \( \alpha = 0 \) results in the VG process (section 2.4.4.3). If \( \alpha \in [1, 2] \), the jump process exhibits infinite variation.

2.4.4.5 Log Stable Model

The logstable model (LS) of Carr and Wu (2003a) is a model with fat tails on all time horizons. They found the central limit theorem violated in the volatility smile pattern of S&P 500 index options over different maturity horizons. If the central limit theorem would hold, the implied volatility smile should flatten out as maturity increases. The above mentioned models of Merton and Kou, the VG and the CGMY, the GH and the NIG distributions all model the log return as a Lévy process with stationary independent increments. For all of these models, the central limit theorem states that the absolute value of skewness declines like the reciprocal of the square-root of maturity, while the kurtosis diminishes with the reciprocal of maturity (Madan and Konikov, 2000). This results in the implied volatility smile flattening out relatively quickly as time to maturity increases. Heston (1993) (section 2.3.4.3) incorporates a persistent stochastic volatility process, which slows down the speed of convergence, but doesn’t stop it as long as the volatility process is stationary. Carr and Wu deliberately violate the central limit theorem conditions by introducing return distributions with infinite moments of any order equal to or greater than two. Nevertheless, their model guarantees the finiteness of all moments of the index level itself, which
ensures the existence of an equivalent martingale measure as well as the finiteness of option prices at all maturities. The LS model is based on an $\alpha$-stable processes with maximum negative skewness to combine infinite return moments with finite price movements. The transition probability distribution $p_t(X_t = x)$ of an $\alpha$-stable process typically has a power-law decay as $x \to \infty$ when $\alpha < 2$ (i.e. for non-normal leptokurtic distributions); consequently $E[exp(X_t)]$ does not exist and call values would be infinite. Carr and Wu’s key innovation is setting the skewness parameter $\beta = -1$ (what they call maximum negative skewness), results in a more rapid decay, $E[exp(X_t)] < \infty$, which leads to finite call values and the following characteristic function (Lewis, 2001):

\[
\phi_t(u) = \exp[iu\mu t - (iu\sigma)^\alpha t \sec(\pi \alpha / 2)].
\] (2.50)

The LS model is a special case of the CGMY process when $G = M = 0$ and $C = 0$ when $x > 0$. Carr and Wu (2003a) compare the option pricing performance of their model vs. the performance of the MJD and VG model and find that the LS model outperforms both other models.

More general, several empirical studies assign infinite-activity jumps superior performance in describing the statistical behavior of equities and equity indices. Carr et al. (2002) find that estimates for the power coefficient $\alpha$ are mostly greater than zero for equity markets. Empirical studies of options also find that using infinite-activity jumps is superior to finite-activity jumps. Huang and Wu (2004) investigate various time-changes (see section 2.5) to generate stochastic volatilities on the three jump specifications reviewed above (no jumps, finite- and infinite-activity jumps) and find that infinite-activity jumps perform best in pricing options.

Earlier studies using compound Poisson jumps to capture rare and large price
movements made it imperative to add also a diffusion component to fill the gaps in between the arrival of the jumps. However, if the starting point is an infinite-activity jump process that can generate an infinite number of small and large movements within a finite interval, it is not clear whether the diffusion component is still needed. Carr et al. (2002) as well as Carr and Wu (2003b) find that it is not necessary, whereas Huang and Wu (2004) conclude that a diffusion component is beneficial in their time-changed Lévy process setting to generate correlations (see also section 2.6) with the diffusive activity rate process. However, the diffusion return component might not be necessary anymore if the activity rate also follows a pure jump process and correlations are implemented through parallel jumps in returns and the activity rate.

Carr and Wu (2003a) investigate the short-maturity behavior of ATM and OTM options to identify if the asset price process of the underlying exhibits jumps and/or diffusion components. They discover that the short-maturity behavior of OTM options is dominated by jump components, whereas the short-maturity behavior of ATM options is dominated by a diffusion component. However, they also find that an infinite-variation jump component could generate similar short-maturity behavior for ATM options, which makes the identification of the diffusion component more difficult if an infinite-variation jump component is present as well.

Ait-Sahalia (2004) shows that, even in the presence of infinite-variation jumps, the diffusion variance can be identified from discretely sampled time-series data (using maximum likelihood methods), as long as the jump component doesn’t have a power coefficient to close to 2. In a separate paper, he and Jacod Ait-Sahalia and Jacod (2008) illustrate that it is possible to distinguish the continuous part of a Lévy process from its jumps part, and even different types of jumps from one another.

52
The empirical success of pure jump Lévy processes does not hold if one looks at the variation of option prices across different maturities. Konikov and Madan (2002) find that homogenous Lévy processes impose strict conditions on the term structure of the risk-neutral variance, kurtosis and skewness. Specifically, these Lévy processes assume a constant variance over the term, skewness which is inversely proportional to the square root of the term and kurtosis which is inversely proportional to the term. Consequently, the literature developed new methods which can incorporate a changing behavior in moments across maturities. We will turn to these methods in the next section.

2.5 Time-Changed Lévy Models

2.5.1 Introduction

We have seen several examples of Lévy processes in the preceeding section. The stochastic property of asset return volatility is a well-known empirical fact (Engle, 2004). Recent investigations (David and Veronesi, 1999; Johnson, 2002; Carr and Wu, 2007a) show that higher returns moments vary also over time. To capture the evidence of stochastic volatility and higher moments, one can apply stochastic time changes to the Lévy processes, which amounts to stochastically altering the clock on which the Lévy processes run. Intuitively, one can see the original clock as calendar time and the new random clock as business time. A more active trading day with high volatility implies a faster business clock. Bochner (1949, 1955) was the first one to bring up the idea of time deformation and subordination, Clark (1973) introduced the idea in the context of asset pricing and Ané and Geman (2000) recently extended and generalized Clark’s result using high frequency data.
Time-changed Lévy (TCL) processes were introduced by Carr et al. (2003) and Carr and Wu (2004). The class of TCL processes covers a wide range of models in the literature and is yet relatively parsimonious. They unify two large strands of the option pricing literature. The first strand uses compound Poisson processes to generate jumps (following Merton (1976), and employs a mean-reverting square-root process to generate stochastic volatility (following Heston (1993)). Examples of this strand of research include Bates (1996b), Bates (2000), Bakshi et al. (1997), Andersen et al. (2002) and Pan (2002). All these examples can be seen as applications of the affine jump diffusion framework of Duffie et al. (2000), where potentially correlated jumps are allowed to occur at both mean and volatility levels. Although the affine framework represents an important theoretical advance, the exclusive use of compound Poisson processes to model jumps is a limitation as they can only generate a finite number of jumps within a finite time period.

This limitation and the empirical fact that asset prices actually display many small jumps on a finite time scale led to the development of a second strand in the option pricing literature, which considers more general jump structures. Examples of this strand, which permits an infinite number of jumps to occur within any finite time interval, include the Normal Inverse Gaussian model of Barndorff-Nielsen (1998b) (section 2.4.4.2), the generalized hyperbolic class of Eberlein, Keller and Prause (section 2.4.4.1), the VG model of Carr, Chang and Madan (section 2.4.4.3), the CGMY process as generalization of VG by Carr, Geman, Madan and Yor (section 2.4.4.4) as well as the finite moment log-stable model of Carr and Wu (section 2.4.4.5).

The empirical investigations of these authors generally support the use of infinite-activity Lévy processes as a way to model returns in a relatively parsimonious way. Barndorff-Nielsen and Shephard (2001a) and Carr et al. (2003) further extended these models to allow for stochastic volatility. However, these models assume that
volatility changes are non-correlated with asset returns, and the leverage effect is only considered under special cases.

TCL processes can combine some of the best features from the two literature streams described above as they generalize both strands simultaneously. In particular, they relax the affine requirement and jump specifications of Duffie, Ken and Singleton and generalize the stochastic volatility/higher return moments of Lévy models by allowing correlations between volatility and asset returns. Hence, the TCL framework allows the combination of complex jump structures with stochastic volatility/higher return moments and the leverage effect. Furthermore, to capture the correlation between returns and their volatilities, Lévy process innovations can be correlated with innovations in the random clock on which it is run (see section 2.6). When this correlation is negative, the clock tends to run faster when the asset price falls, thus capturing the leverage effect first discussed by Black (1976).

2.5.2 General Setup of Time-Changed Lévy Models

The price process $S_t, t > 0$ of the underlying asset is modeled as an exponential Lévy process $X_t, t > 0$:

$$S_t = S_0 exp(rt + X_{Y_t})$$

running on the stochastic time $Y_t$, with $r$ denoting the risk free rate. The time change from $t$ to $Y$ results in stochastic volatility of the jump process $X_t$. As time is a strictly increasing process, $Y_t$ is set up as the following integral, called the instantaneous rate of time change:

$$Y_t = \int_0^t y_s ds$$
Due to the non-negativity and mean-reverting properties of \( y_t \), interest rate models are the natural choice to model the stochastic time change; two of them will be reviewed in sections 2.5.5 and 2.5.6. Time-changed Lévy Processes \( X_{Y_t} \) have two different sources of randomness. Next to the stochastic nature of \( X_t \), the development of time \( Y_t \) is also uncertain. Therefore, the characteristic function of \( X_t Y_t \) involves expectations over two different sources of randomness:

\[
\phi_{X_t Y_t} = E[\exp(i\nu X_t Y_t)] = E[E[\exp(i\nu X_t Y_t)|Y_t]].
\]  

(2.53)

If the processes \( X_t \) and \( Y_t \) are independent from each other, the inner integration can be carried out as follows:

\[
\phi_{X_t Y_t} = E[\exp(Y_t \psi_X(\nu))] = \psi_Y(-i\psi_X(\nu)),
\]  

(2.54)

where \( \psi_X \) is the characteristic exponent of the Lévy process and \( \phi_{Y_t} \) is the characteristic function of the time change \( Y_t \). The assumption of independence between \( X_t \) and \( Y_t \) excludes models with leverage effect. We will return to this issue in section 2.6 and illustrate how correlations between the Lévy process and the time-change can be incorporated. We will now investigate two popular approaches of constructing a risk-neutral version of the simple uncorrelated setup based on [Dahlbokum 2007, 2010].

2.5.3 Stochastic Exponential of Time-Changed Lévy Models

[Carr et al. 2003] provide the mathematical background of stochastically exponentiating a time-changed process \( X_{Y_t} \). [Cont and Tankov 2004] provide an
alternative access which will be outlined below; it basically states that time-
changing a martingale will result in another martingale. Therefore, we have to
take the risk-neutral version of the Lévy processes presented in section 2.5 and
perform the desired time change. Doing so results in a drift-adjusted process on
the changed time $Y_t$:

$$S_t = S_0 \exp[rt + X_{Y_t} - Y_t \psi_X(-i)].$$  \hspace{1cm} (2.55)

Defining $X_{Y_t} = X_{t} - Y_t \psi_X(-i)$, with superscript $S$ marking a martingale via
stochastic exponentiation, the characteristic function of the risk-neutral process
becomes the following:

$$\phi_{X_{Y_t}^S} (\nu) = E[e^{i\nu X_{Y_t}^S}] = \phi_{Y_t}(-i\psi_X(\nu) - \nu\psi_X(-i)).$$  \hspace{1cm} (2.56)

The reader is referred to Carr and Wu (2004) for details of the martingale con-
struction.

2.5.4 Ordinary Exponential of Time-Changed Lévy Models

The transfer to the risk-neutral world in this approach is achieved by mean-
correcting the basic setup of equation 2.51:

$$S_t = S_0 \frac{(exp[r t + X_{Y_t}])}{E[exp(X_{Y_t})]}.$$  \hspace{1cm} (2.57)
Defining \( X^O_t = \frac{X_Y_t}{E[exp(X_Y_t)]} \), with superscript \( O \) marking a martingale via ordinary exponentiation, and using:

\[
E[exp(X_Y_t)] = \phi_Y_t(-i\psi_X(-i)),
\]

the characteristic function of the arbitrage-free process becomes:

\[
\phi_t(X^O_Y(\nu)) = E[e^{i\nu X^O_Y_t}] = \frac{\phi_Y_t(-i\psi_X(\nu))}{\phi_Y_t(-i\psi_X(-i))}.
\]

We refer to Schoutens et al. (2004) and Carr et al. (2003) for further details of the approach. As we have seen in section 2.5, the stochastic time \( Y_t \) is not modelled directly but via its rate of change \( y_t \) as specified in equation 2.52. We will now turn to two different processes for \( y_t \) demonstrated by Schoutens et al. (2004).

### 2.5.5 Cox-Ingersoll-Ross Stochastic Clock

The intuition behind TCL arises from the Brownian scaling property. The scaling property relates changes in scale with changes in time, therefore random changes in volatility can also be captured via random changes in time. The instantaneous rate of time change has to be positive if the new ”business” clock is increasing and has to be mean-reverting if the random time changes are to persist. The square root process of Cox et al. (1985) (hereafter CIR) is the classical example of a mean-reverting process and is therefore often used to model the stochastic time change. The CIR process is given by the following stochastic differential equation:

\[
dy_t = \nu(\eta - y_t)dt + \xi \sqrt{y_t}dW_t,
\]

We refer to Schoutens et al. (2004) and Carr et al. (2003) for further details of the approach.
where $W_t, t > 0$ describes a Brownian motion. $\eta$ is the long term level, $\nu$ controls the strength of the mean-reversion and $\xi$ describes the volatility of the time-change. Schoutens et al. (2004) demonstrate that the starting value $y_0$ can be normalized to 1, leaving the process with three free parameters. The characteristic function is given by:

$$\phi = E[\exp(i\nu Y_t | y_0)] = \frac{\left(\exp(\frac{\nu^2 \eta t}{2})\exp\left(\frac{-2y_0 i\nu}{\nu + \omega \cosh\left(\frac{\omega t}{2}\right)}\right)\right)}{\cosh\left(\frac{\omega t}{2}\right) + \nu \sinh\left(\frac{\omega t}{2}\right)^{\frac{2\nu \eta}{\omega^2}}}, \quad (2.61)$$

with $\omega = \sqrt{\nu^2 - 2\xi^2 i\nu}$ (Schoutens et al., 2004).

### 2.5.6 Gamma-Ornstein-Uhlenbeck Stochastic Clock

Barndorff-Nielsen and Shephard (2001b) coined the term ”background driving Lévy Process" (BDLP), which is another mean-reverting processes, precisely a Ornstein-Uhlenbeck process governed by a one sided pure jump Lévy process:

$$dy_t = -\nu y_t dt + dZ_{\nu t}, \nu > 0 \quad (2.62)$$

with $Z_t, t > 0$ being a Lévy subordinator. In case of the Gamma Ornstein Uhlenbeck ($\Gamma - OU$) process, $Z_t$ is a compound Poisson process with intensity $\zeta$ and an exponential distribution with mean $\frac{1}{\beta}$ describing the jumps. The characteristic function of the ($\Gamma - OU$) process (given $y_0$) is (Schoutens et al., 2004):

$$\phi = \exp\left[\frac{iv\nu y_0}{\nu}(1 - e^{-\nu t}) + \frac{\nu \zeta}{iv} \left(\frac{\beta}{\nu} - \frac{\beta}{\nu(1 - e^{-\nu t})} - iv t\right)\right]. \quad (2.63)$$

The model setup we discussed above assumes no correlation between the the Lévy process (asset prices) and the time change (volatility). This constraint must be relaxed to incorporate the leverage effect. This involves finding a way to correlate
processes with jump and diffusion parts. Carr et al. (2003) as well as Carr and Wu (2004) present several approaches how one can achieve this; we will turn to them in the next section.

2.6 Lévy Models with Leverage

The term leverage effect has become generic in describing the negative correlation between equity returns and their volatilities. However, various other explanations have also been proposed in the literature (Black, 1976; Kahnemann and Tversky, 1979; Cao et al., 2008). The leverage effect in option pricing models based on time-changed Lévy processes is implemented via negative correlations between increments in the the Lévy process and increments in the time change process. Recall that every pure jump component is orthogonal to a purely continuous component Carr and Wu (2004). Consequently, when we have a purely continuous Lévy process, nonzero correlation can only be achieved by a continuous component in the activity rate process. Similarly, when we have a pure jump Lévy process, nonzero correlation can only be induced by a jump component in the activity rate process. In addition, if the Lévy process has a pure jump Lévy process with finite-activity, nonzero correlation can only be achieved by using a finite-activity jump component in the activity rate process; the analogue holds for pure jump Lévy processes with infinite-activity. We will outline below the case where price and volatility process have a continuous diffusion part and refer to Carr and Wu (2004) for examples of implementing leverage via finite- and infinite-activity jumps. Let us consider Heston’s model (Heston, 1993) as an example where the leverage effect is implemented by correlating the involved Brownian motions. The
CIR-time-changed diffusion is given by:

\[ X_{Y_t} = \sigma W_{Y_t}, \quad (2.64) \]

\[ dy_t = \nu(\eta - y_t) dt + \xi \sqrt{y_t} dV_t, \quad (2.65) \]

\[ E[dW_t dV_t] = \rho dt. \quad (2.66) \]

As indicated above, this approach does not work for pure jump processes or for other methods of time-changes, such as the \( \Gamma - OU \) type. \[ \text{Carr and Wu (2004)} \] derive risk-neutral versions of models with diffusion-based leverage effects via the stochastic exponential (see section 2.5.3). They introduce a leverage-neutral complex-valued measure \( \tilde{Q} \), under which equation \( 2.54 \) represents the characteristic function. For our example of CIR-time-changed Lévy processes, they obtain the following characteristic function:

\[ \phi_{X_{Y_t}}(\nu) = \phi_{Y_t}(-i\psi_X(\nu)) = \exp\{-b(t)y_0 - c(t)\} \quad (2.67) \]

with

\[ b(t) = \frac{2\psi_X(\nu)(1 - e^{-\varpi t})}{2\varpi - (\varpi - \nu^*)(1 - e^{-\varpi t})}, \quad (2.68) \]

\[ c(t) = \frac{\eta \nu}{\xi^2} [2\ln(1 - \frac{\varpi - \nu^*}{2\varpi}(1 - e^{\varpi t})) + (\varpi - \nu^*)t] \quad (2.69) \]

and

\[ \varpi = \sqrt{(\nu^*)^2 + 2\xi^2 \psi_X(\nu)}, \quad \nu^* = \nu - i\nu \rho \sigma \xi, \quad y_0 = 1. \quad (2.70) \]

\[ \text{Schoutens et al. (2004)} \] demonstrated how to normalize the current rate of time change \( y_0 \) to 1. The parameters \( \nu, \eta \) and \( \xi \) belong to the CIR modeled time-change, while \( \psi_X(\nu) \) describes the characteristic exponent of the Lévy process. The leverage effect is limited between \(-1 \leq \rho \leq 1\).
This concludes the discussion of the building blocks of Lévy processes to model financial security returns. Based on these building blocks, one can create various time-changed Lévy processes by choosing among the Lévy Processes (section 2.4) and combining them with time-changes (section 2.5) to model nonnormal return distributions and stochastic volatilities. These models can be constructed via the ordinary (section 2.5.4) or stochastic (section 2.5.3) exponential. Using the stochastic exponential further allows us to equip the model with the leverage effect.

2.7 Summary

This section has reviewed the evolution of the derivatives pricing literature over time. It started with Bachelier at the beginning of the twentieth century and closed with CGMY at the beginning of the twentyfirst century. It equips us with a selection of base models we can use for the empirical test (chapter 7) of the ensemble methods later in this thesis. In conclusion, our review shows that the development of option pricing models has progressed in a logical order; new models mainly focused on addressing key shortcomings of previous models. The late 80's/early 90's have seen an overhaul of the constant volatility assumption (resulting in stochastic volatility applied directly to the BS model). Important models of this generation have been reviewed in section 2.3.4.3. Starting from the 90’s, Lévy processes (generalizations of Brownian motion, section 2.4) gained significantly in popularity. Table 2.1 summarizes the reviewed option pricing models.

As most introduced option pricing methods focus on European options, we will conduct the empirical test on DAX index options, as they are also of European type.
<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>Diffusion</th>
<th>Jumps</th>
<th>Stoch. Vol.</th>
<th>Leverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black Scholes Merton</td>
<td>Black and Scholes (1973) &amp; Merton (1973)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEV</td>
<td>Cox and Ross (1976)</td>
<td>✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dupire</td>
<td>Dupire (1994)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Derman</td>
<td>Derman and Kani (1994)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rubinstein</td>
<td>Rubinstein (1994)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson &amp; Shanno</td>
<td>Johnson and Shanno (1987)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scott</td>
<td>Scott (1987)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hull-White</td>
<td>Hull and White (1987)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stein and Stein</td>
<td>Stein and Stein (1991)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heston</td>
<td>Heston (1993)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Merton JD</td>
<td>Merton (1976)</td>
<td>✓</td>
<td></td>
<td>Poisson</td>
<td>✓</td>
</tr>
<tr>
<td>Kou</td>
<td>Kou (2002)</td>
<td>✓</td>
<td></td>
<td>Double Exponential</td>
<td></td>
</tr>
<tr>
<td>NIG</td>
<td>Barndorff-Nielsen (1997) &amp; Rydberg (1997a,b)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VG</td>
<td>Madan and Seneta (1987) &amp; Carr et al. (1998)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGMY</td>
<td>Carr et al. (2002)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS</td>
<td>Carr and Wu (2003a)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIG(_{CIR})</td>
<td>Barndorff-Nielsen (1997) &amp; Rydberg (1997a,b) &amp; Cox et al. (1985)</td>
<td>✓</td>
<td></td>
<td>CIR</td>
<td>✓</td>
</tr>
<tr>
<td>NIG(_{OU})</td>
<td>Barndorff-Nielsen (1997) &amp; Rydberg (1997a,b) &amp; Barndorff-Nielsen and Shephard (2001b)</td>
<td>✓</td>
<td></td>
<td>OU</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 2.1: Overview of Option Pricing Models
As we have seen, Lévy processes cannot only generate stochastic volatility (by applying stochastic time changes - an indirect way compared to the one above, see section 2.5), they can also generate stochastic higher return moments as empirical investigations of return distributions of financial assets have shown that one needs more than two parameters (not only mean and variance, but also skewness and kurtosis) to describe these distributions properly. If we further include the leverage effect in our model (section 2.6), one can theoretically adjust for all three main shortcomings of the Black-Scholes assumptions outlined in the introduction of this thesis. However, the tradeoff associated with improved fitting of observed option prices comes at the expense of analytical tractability and increased model complexity and variance.

Furthermore, a common feature of all derivative pricing models is that they must be calibrated to existing market prices of options to avoid arbitrage opportunities before using them. Model calibration is an inverse problem that consists of determining model parameters to reproduce the observed vanilla option prices. This calibration process can be challenging due to several reasons (Gupta, 2009):

First, there can be inconsistencies or mispricings in the option prices used for calibration, so that no parameter set could correctly reproduce all the observed option prices. Second, there might not be enough observable market prices to uniquely determine the calibration parameters. Furthermore, the stability of the calibrated models can be quite low - a small change in price of the underlying may lead to significant changes in the calibration parameters of the model (high variance).

Despite these issues, most of the current literature on calibrating derivative pricing models focuses mainly on finding the single best-fit parameter set that minimizes a certain measure of pricing error given a set of market prices and one or several option pricing models, i.e. reducing the bias of the model to be calibrated.
Only recently, attention has been paid to the issue of finding the right and avoiding the wrong parameters/models (model uncertainty), how stable these parameterized models are (robustness/variance minimization) and how one can optimize the pricing task in this context of model uncertainty (Gupta and Reisinger, 2011). The next section will therefore review the current literature on model uncertainty with a special focus on derivatives pricing.
Chapter 3

Model Uncertainty

3.1 Introduction

Model uncertainty in the context of this thesis deals with the problem of not knowing the best option pricing model to use. Following the taxonomy of Kerkhof and Melenberg (2004), we define model misestimation as choosing the wrong model parameters and model misspecification as selecting the wrong model type. We outlined in the previous section that the evolution of the derivatives pricing literature led to more complex, but also more realistic financial models. However, before theoretical models can be used in practice and calibrated to existing market prices to avoid arbitrage opportunities, one faces the problem of model uncertainty. In increasing order of complexity, one has to deal with the following issues (Branger and Schlag, 2004):

- Parameter uncertainty: One knows the true model and has to find the correct model parameters, hence to calibrate the model to market prices of options

- Specification uncertainty: One knows the true model class (e.g. stochastic
volatility models, section 2.3.3), but not which model is the best one to choose from this class. The issue becomes even more complex if one is uncertain what the true model class is (e.g. whether one has to include jumps, stochastic volatility or both, see Branger et al. (2012)).

As the pricing of derivatives is very sensitive to changes in the chosen option pricing model or parameter set, the issue of model uncertainty is very relevant. This applies especially to practitioners as the wrong model selection might result in significant financial losses.

The issue of model uncertainty in the context of derivatives pricing has only recently attracted academic interest and we follow closely the work of Gupta (Gupta, 2009; Gupta et al., 2010; Gupta and Reisinger, 2010, 2011, 2012).

As mentioned, the model selection problem is an inverse one. We have a model or set of models to price a financial claim; we then take some market prices and try to find the model or model parameter set \( \theta \) which minimizes the difference between model and market prices. It is an inverse problem as we know how to get from the option pricing model to the prices but not how to get from the market prices to the corresponding model. Several papers have examined this inverse problem (Dupire, 1994, 1997; Rubinstein, 1994; Jackwerth and Rubinstein, 1996).

As it is not clear whether perfect replication of market prices is possible in practice, most approaches focus on minimizing the calibration error. There exists a wealth of literature on the issue of model calibration. Taking local volatility models (section 2.3.3) as an example, Osher and Lagnado (1997), Jackson et al. (1999), Coleman et al. (2001), as well as Chiarella et al. (2000) investigated minimization techniques and regularization functions to find the best-fit local volatility surface.

However, most of the current literature on derivatives with respect to pricing,
risk management and hedging focuses mainly on finding the best-fit in-sample parameter set for one or several option pricing models. This basically ignores model uncertainty as the models used are considered given. Little attention is paid to measuring how robust the parameter set is or whether other models or parameter sets can reproduce market prices equally/almost as well, but are more stable and hence have better performance on out-of-sample data sets. This is a shortcoming of the current literature, as it lacks a substantial investigation of robustness, uniqueness and model uncertainty and focusses mainly on minimizing the in-sample pricing error.

This section of the literature review is organized as follows. We will start with a distinction between risk and uncertainty and turn then to the literature on model uncertainty. We will investigate various ways how model uncertainty may manifest itself and what the potential consequences are. After this, we will review several risk measures which will be used to judge the performance of our proposed ensemble approach in chapter 7.

3.2 Risk versus Uncertainty

The problem of using the wrong model is often quoted as model risk in the literature, although we think that model uncertainty is the more accurate terminology following the arguments of Cont (2006) and Gupta et al. (2010). The distinction between risk and uncertainty was first discussed by Knight (1921).

Option pricing theory usually starts with an underlying asset $S$ that evolves on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\Omega$ representing the set of future scenarios, $\mathcal{F}$ a filtration of $\Omega$ and $\mathbb{P}$ representing a probability measure on $\mathcal{F}$ (see also section 2.3.4). Hence, risk is defined as not knowing which future scenario $\omega \in \Omega$ will
occur, whereas uncertainty is defined as not knowing the probability measure \( P \) on \( \mathcal{F} \) (Gupta et al., 2010). Epstein (1999) defines ambiguity as the situation where we have to choose among several possible specifications \( P_1, P_2, \ldots, P_N \) for probabilities of future outcomes.

Although the differentiation between risk and uncertainty might seem negligible, traders and other market participants will have different aversions against not knowing the future states or not knowing the probabilities of possible future states. For example, Ellsberg (1961) studied the difference in aversion against uncertainty and risk as defined above and demonstrated that uncertainty aversion can help to justify some anomalies in classical decision theory. More recent work related to uncertainty aversion in capital markets include Cao et al. (2005), Chen and Epstein (2003) and Gagliardini et al. (2009).

The problem of model uncertainty is also related to the issue of market incompleteness. However, an incomplete market does not necessarily imply model uncertainty and vice versa (Branger et al., 2012). A market is considered incomplete, if there are payoff profiles that cannot be replicated by self-financing dynamic strategies (Branger et al., 2004). Under market incompleteness, the probability measure is assumed to be known, but we do not have a unique martingale measure as the number of risk factors is too high compared to the number of traded assets. Under model uncertainty, the probability measure \( P \) is unknown and the market can be either complete or incomplete.

The literature has proposed two different paradigms to deal with model uncertainty (Cont, 2006). The first one, which we will investigate further in this thesis, is the aggregation of several candidate models, e.g. through averaging. This approach has been used extensively in the statistical literature. Several empirical

\footnote{He defines ambiguity analog to our definition of uncertainty above.}
studies have shown that ensemble predictions can often be much more accurate than the forecasts of the base learners (Freund and Schapire 1996; Bauer 1999; Dietterich 2000a) and various authors have suggested justifications for the success of ensemble methods from a theoretical perspective (Kittler et al. 1998; Allwein et al. 2000; Kleinberg 2000). This approach will be reviewed in more detail in section 4.

The second paradigm is based on worst case or maxmin approaches and has been introduced by Gilboa and Schmeidler (1989) and analyzed in the context of asset pricing among others by Epstein and Wang (1994) and Routledge and Zin (2009). Coherent (Artzner et al. 1999; Frittelli and Gianin 2002; Cont 2006) and convex (Föllmer and Schied 2002; Frittelli and Gianin 2002; Cont 2006) measures of risk as well as extensions of it (Acerbi and Scandolo 2008; Branger et al. 2012; Elliott et al. 2007; Frittelli and Gianin 2002; Gupta 2009; Gupta et al. 2010) are related to the worst case approach and will be presented in section 3.4.

### 3.3 Sources of Model Uncertainty

The financial crisis of 2008 has shown that wrong model assumptions can have extremely unwelcome financial consequences. An investigation of model risk only started to attract academic interest in the mid 90’s, when mathematicians such as Derman (1996), Green and Figlewski (1999), Li (1999), Britten-Jones and Neuberger (2000), Hull and Suo (2002), Kerkhof and Melenberg (2004) as well as Cont (2006), Cont and Tankov (2006) and Gupta (2009), Gupta et al. (2010), Gupta and Reisinger (2011) began investigating the issue of model risk and model uncertainty and market observers like Elliott (1997) reported on substantial losses due to model risk.
Derman [1996] distinguishes between seven different sources of model risk:

The first one is the potential inapplicability of modelling: One might not be able to predict financial asset movements with complex mathematical models as brinkmanship or psychology might be more important factors concerning the movement of stocks. Derman states that a clear understanding of the driving factors is needed before mathematics can be used to describe the underlying process.

A second source of risk is the choice of an incorrect model by either forgetting some factors completely or modelling them incorrectly, e.g. as stochastic instead of deterministic or vice versa. Also, the inclusion of market frictions such as transaction costs might make a model no longer applicable. Furthermore, relationships between variables might be wrongly modeled, e.g. correlations between interest and prepayment rates in the context of CDO modelling. Moreover, a model might work under stable market conditions, but fail during a financial crisis. In addition, the model for a financial asset, e.g. a stock, might be inappropriate to model a similar asset in another market with diverging volatility and/or different yield curve characteristics. Derman mentions as a third possible risk that the model might be correct, but the analytical solution found might be incorrect. As a fourth risk he states that even if the correct analytical solution is found, the numerical approximation might not be accurate enough. A related risk considers the issue of software or hardware bugs used to implement the numerical schemes.

Another important risk is the inappropriate use of the correct model, e.g. that a model based on Monte Carlo simulations is used with too few simulations. Derman highlights as the final source of risk the sensitivity of the model to input data. He stresses the importance to check whether small changes in the underlying assumptions of the model or observed market data cause disproportionate changes in option prices, the Greeks or any other model inference. If market conditions change suddenly or the input data is actually very noisy, then it is
imperative that the model is robust enough to account for this uncertainty.

Green and Figlewski (1999) differentiate between three different sources of model risk: model misspecification, unobservability of input parameters and market violation of assumptions. The notion of model misspecification is similar to what Derman (1996) calls incorrect model choice. Examples for unobservability of input parameters (also called model misestimation) would be the inability to detect the jump density in a general Lévy process. Figlewski and Green analyze empirically the market and model risk exposures for someone trading European style vanilla options. They consider the example of arbitrage-free fair option values as an example for market violation of assumptions. They state that most financial models find the fair price by constructing a replication strategy of the pay-offs, however transaction costs often make this replication very expensive and therefore unrealistic in real markets. Their empirical contribution reviews volatility forecasting methods on four underlyings: Deutschmark Exchange Rate, 10 Year US Treasury Yields, 3 months USD LIBOR and the S & P 500 index. They find that unhedged positions can lead to very large risk exposures which are difficulty to diversify. They conclude that the risk of misestimating volatility in trading and hedging derivatives positions can be very large and hence very costly. They recommend that agents use a volatility markup (and hence a higher option price) to reduce this risk.

Hull and Suo (2002) investigate model risk from the perspective of model misspecification and look at the pricing errors arising from a continually recalibrated implied volatility model. They find that the model works reasonably well to price and hedge European and compound options, but can have significant model uncertainty for barrier options. They explain that for exotic options, the performance of the model depends on the degree of path dependence to calculate the
The more path dependent an option is, the more its pricing is exposed to model risk. Hence, Hull and Suo conclude that the failure to correctly price barrier options is to be expected from the local volatility model as they have a much higher degree of path dependence than European or compound options. 

Britten-Jones and Neuberger (2000) also analyze model risk in the context of an implied volatility model and provide another example of specification uncertainty. They demonstrate that several different option pricing models can be calibrated to fit market prices of vanilla options quite well - hence the prevalence of specification uncertainty.

Li (1999) studies the implications of parameter uncertainty on hedging when exact calibration is not feasible. Similar to the option “Greeks”, he calculates the sensitivity of the option portfolio to model parameters and deduces appropriate hedging ratios. He recommends two hedging strategies: the least-squares hedge and the absolute-value hedge, the former is additive and the latter one self-recoverable. Li (1999) stresses the importance of an “instrumentally” hedged portfolio. However, in practice, this idea is difficult to implement as the parameter sensitivities are difficult to estimate and additional transaction costs might reduce the value added by his approach.

Monoyios (2007) is an example of hedging optimization in the context of parameter uncertainty in an incomplete market. He demonstrates that misestimation of the drift parameter can have benign effects if the drift parameter is overestimated but enormous if the drift parameter is underestimated.

Kerkhof and Melenberg (2004) try to quantify model risk by determining how much regulatory capital should be set aside to account for model uncertainty. They find that model risk due to misspecification significantly exceeds model risk
due to estimation error. Using a simulation study, they analyze a wide range of popular risk measures such as Value at Risk and Expected Shortfall and determine the multiplication factor to account for model uncertainty under the Basel Accord backtesting requirements. We will analyze their introduced risk measure further in section 3.4.5. Although Kerkhof and Melenberg (2004) differentiate between parameter (misestimation risk) and model (misspecification risk) uncertainty, we will outline in the next section, following Cont (2006) and Gupta (2009), why this distinction is not necessary in this thesis.

Cont (2006) introduces a quantitative framework to measure model uncertainty in the context of derivative pricing. He develops two different methods: one based on a coherent risk measure compatible with market prices of derivatives, while the second method is based on a convex risk measure. Furthermore, Cont suggests several measures to manage model risk. We will further investigate his contributions in section 3.4.

Gupta (2009) investigates in his doctoral thesis a Bayesian approach to model uncertainty and shows - using the example of a local volatility model - how to price and hedge in a more robust way under the Bayesian framework. Furthermore, he extends the risk measures of Artzner et al. (1999), Föllmer and Schied (2002), Frittelli and Gianin (2002) and Cont (2006). Together with his PhD supervisor Christoph Reisinger, he published several interesting papers summarizing and extending the work of his doctoral thesis (Gupta et al., 2010, Gupta and Reisinger, 2010, 2011, 2012) which we used as a guideline and very valuable source for this section of the thesis.
### 3.4 Model Uncertainty Measures

In the previous section, we reviewed the literature on model uncertainty and identified several occurrences of parameter and/or specification uncertainty. In this section, we will investigate methods to assign a quantitative measure to this uncertainty. We will consider parameter and specification uncertainty under a unified framework and refer to both of them as model uncertainty as a meta-model could be built in which different parametrizations of models correspond to different models (Gupta and Reisinger, 2011).

Market risk measures such as Value at Risk (Duffie and Pan, 1997) are used to estimate the amount of capital required to hold a risky position. This implies that one would have a realistic model for the considered underlying $S$. Hence, one could calculate a risk measure $\rho(X)$ for a claim $X$ written on $S$ given the specific model $\theta$. Other important risk measures have been introduced by Artzner et al. (1999), Föllmer and Schied (2002) and Frittelli and Gianin (2002) in asset pricing and by Cont (2006) and Gupta (2009) in the context of derivatives pricing.

Market risk measures (and most models introduced in section 2), consider the variable $X$ to be random through its dependence on a state of the future $\omega \in \Omega$, i.e. $X = X(\omega)$. If the right model including its parameters were known, the law of $X$ would be fully determined. However, under model uncertainty, the model $\theta \in \Theta$ is unknown. Hence, we denote the claim by $X(\omega, \theta)$ and the market risk measure under model $\theta$ by $\rho^{P_\theta}$ (Gupta et al., 2010).

Introducing measures of parameter or specification uncertainty is important for several reasons (Gupta, 2009). On the one hand, it provides us with a quantitative and hence comparable result of several competing models and/or model calibrations. Second, it is also useful for similar tasks such as risk management.
and hedging as we can not only determine a measure of pricing performance but also establish potential pricing errors and hedging losses as well as other related financial quantities.

We outline below five different model uncertainty measures based on [Gupta et al. (2010)].

### 3.4.1 Coherent Model Uncertainty Measures

[Artzner et al. (1999)] introduced multiple desirable properties of risk measures and defined them as coherent if they fulfill several criteria. It was further developed by [Delbaen (2002)].

To introduce coherent market measures, we assume the following: We have a set $\Omega$ of scenarios $\omega$, $\mathcal{X}$ is a set of valued functions $X$ on $\Omega$ and $\mathbb{P}$ is a probability measure on $\Omega$. Then, a market risk measure is defined as a mapping from $\mathcal{X}$ into $\mathbb{R}$ and a market risk measure $\rho$ is defined as coherent if it fullfills the following four axioms (Gupta, 2009):

**Translational Invariance:** $\forall X \in \mathcal{X}$ and $\forall a \in \mathbb{R}$, $\rho(X + a) = \rho(X) - a$

**Monotonicity:** $\forall X, Y \in \mathcal{X}$ with $X \leq Y$, we have $\rho(X) \geq \rho(Y)$

**Subadditivity:** $\forall X, Y \in \mathcal{X}$, $\rho(X + Y) \leq \rho(X) + \rho(Y)$

**Positive Homogeneity:** $\forall \lambda \geq 0$ and $\forall X \in \mathcal{X}$, $\rho(\lambda X) = \lambda \rho(X)$

**Representation Theorem** Following [Artzner et al. (1999)], a market risk measure $\rho : \mathcal{X} \to \mathbb{R}$ is considered coherent iff there exists a family of risk measures $P_{\rho}$ of probability measures $\mathbb{P}$ on the set $\Omega$ of scenarios $\omega$ such that:

$$\rho(X) = \sup_{P \in P_{\rho}} \{\mathbb{E}^{P}[X]\}$$
Examples of coherent risk measures are (Gupta, 2009):

i) worst-case \( \rho_0(X) = \sup_{\omega \in \Omega} \{-X(\omega)\} \)

ii) average value \( \rho_1(X) = \mathbb{E}^\mathbb{P}[-X(\omega)] \)

iii) expected shortfall \( \rho_\beta(X) = \frac{1}{\beta} \int_{\Omega} -X(\omega) 1_{X \leq q_X(\beta)} \mathbb{P}[d\omega] \) for some \( \beta \in (0, 1) \)

and where the quantiles \( q \) are given by \( \beta = \mathbb{P}[X \leq q_X(\beta)] \)

Based on the definitions above, one can demonstrate that \( \rho_0 = \lim_{\beta \to 0} \rho_\beta \) and \( \rho_1 = \lim_{\beta \to 1} \rho_\beta \). \( \beta \) is considered a risk aversion parameter: the smaller its value, the higher the risk aversion. Typical values for \( \beta \) are \( \beta = 0.01 \) or \( \beta = 0.05 \) (Gupta, 2009).

3.4.1.1 Cont’s Axioms

Cont (2006) analyzes ways to measure the uncertainty of choosing the wrong option pricing model. He introduces in his widely cited paper a quantitative framework to measure model uncertainty and defines several properties a model uncertainty measure should fulfill, extending the work of Artzner et al. (1999). Assume we have various claims \( C_i \) inside a bid-ask range \([V^{(i)}\text{bid}}, V^{(i)}\text{ask}]\) for \( i \in I \) and varous models \( \theta \) belonging to the set of models \( \Theta \). As usual, we have the risk neutral probability measure \( \mathbb{Q}_\theta \) for the asset price process of \( S \) dependent on the specific model \( \theta \). We further assume that:

\[
\forall \theta \in \Theta, \quad \mathbb{E}^{\mathbb{Q}_\theta}[C_i] \in [V^{(i)}\text{bid}}, V^{(i)}\text{ask}] \quad \forall i \in I,
\]

which means that all models \( \theta \in \Theta \) are able to replicate option prices within their bid-ask spreads (Gupta et al., 2010). We define \( \mathcal{X} = \{X : \forall \theta \in \Theta, \mathbb{E}^{\mathbb{Q}_\theta}[|X|] < \infty\} \) as the set of all contingent claims that have well-defined prices in all considered
models. Furthermore, \( \Phi \) describes the set of all trading strategies \( \phi \) which result in \( E_{Q^\theta}^\Theta \left[ \int_0^T \phi_u dS_u \right] = 0 \) for all models \( \theta \in \Theta \). For the sake of simplicity, Cont (2006) assumes a risk free rate of zero. He then defines a function \( \mu : \mathcal{X} \rightarrow [0, \infty) \) to be a coherent model uncertainty measure if it fulfills the following four axioms next to equation (3.1):

**Cont.1** All benchmark options have a lower model uncertainty than the bid-ask spread:
\[
\forall i \in I, \quad \mu(C_i) \leq |V^{(i)\text{ask}} - V^{(i)\text{bid}}|
\]

**Cont.2** As hedging is model dependent, model dependent dynamic hedging does not reduce model uncertainty:
\[
\forall \phi \in \Phi, \quad \mu \left( X + \int_0^T \phi_t dS_t \right) = \mu(X)
\]
However, if the option can be fully replicated in a model-free way, no model uncertainty exists:
\[
\text{if } \exists x \in \mathbb{R}, \phi \in \Phi \text{ s. t. } \forall \theta \in \Theta, \quad X = x + \int_0^T \phi_t dS_t \theta \text{ a. s. } \mu(X) = 0
\]

**Cont.3** Diversification cannot increase model uncertainty of a portfolio of claims:
\[
\forall X_1, X_2 \in \mathcal{X}, \forall \lambda \in [0, 1], \quad \mu(\lambda X_1 + (1 - \lambda)X_2) \leq \lambda \mu(X_1) + (1 - \lambda)\mu(X_2)
\]

**Cont.4** Static hedging reduces model uncertainty. The hedging of a claim with traded options is less than or equal to the sum of the model uncertainty of that claim and the corresponding uncertainty about the cost of the replication strategy:
\[
\forall X \in \mathcal{X}, \forall a \in \mathbb{R}^d, \quad \mu \left( X + \sum_{i=1}^d a_i C_i \right) \leq \mu(X) + \sum_{i=1}^d |a_i||V^{(i)\text{ask}} - V^{(i)\text{bid}}|
\]
3.4.2 Worst-Case Measures

Cont (2006) introduces the following equation:

$$
\mu_0(X) = \sup_{\theta \in \Theta} \{E_Q^\theta[X]\} - \inf_{\theta \in \Theta} \{E_Q^\theta[X]\} \tag{3.2}
$$

as a coherent measure of model uncertainty, fulfilling equation (3.1) and the axioms outlined in the previous section. The proposed risk measure 3.2 calculates the difference between the model with the highest and the one with the lowest option price calculated by all models \( \theta \) of the set \( \Theta \). It is a worst-case measure as it finds the biggest spread among all prices \( E_\theta[X] \) for the option \( X \).

3.4.3 Convex Model Uncertainty Measures

Cont (2006) further extends coherent model uncertainty measures to convex model uncertainty measures, based on the work of Frittelli and Gianin (2002) and Föllmer and Schied (2002). Gupta (2009) shows how to generalize Cont (2006) to the case of bid-ask spreads. Using the same setup as in section 3.4.1, we have a set \( \Omega \) of scenarios \( \omega \), \( X \) is defined as a set of valued functions \( X \) on \( \Omega \) and \( P \) is a probability measure on \( \Omega \). Then, a market risk measure is defined as a mapping from \( X \) into \( \mathbb{R} \) and a market risk measure \( \rho \) is defined as convex if it fulfills the following three axioms (Gupta, 2009):

**Translational Invariance:** \( \forall X \in X \) and \( \forall a \in \mathbb{R} \), \( \rho(X + a) = \rho(X) - a \)

**Monotonicity:** \( \forall X, Y \in X \) with \( X \leq Y \), we have \( \rho(X) \geq \rho(Y) \)

**Subadditivity:** \( \forall X, Y \in X \), \( \rho(X + Y) \leq \rho(X) + \rho(Y) \)

**Representation Theorem** Following Frittelli and Gianin (2002), a market risk measure \( \rho : X \to \mathbb{R} \) is considered a finite valued and lower semicontinuous
convex market risk measure iff there exists a convex penalty functional $\alpha$ and non empty convex set $P_\rho$ of probability measures $\mathbb{P}$ on the set $\Omega$ of scenarios $\omega$ such that:

$$\rho(X) = \sup_{\mathbb{P} \in P_\rho} \left\{ \mathbb{E}^{\mathbb{P}}[ -X ] - \alpha(\mathbb{P}) \right\}$$

Gupta (2009) points out that the convex market risk measure converges to a coherent one if $\alpha = 0$ or $\alpha = \infty$.

3.4.3.1 Cont’s Axioms

The idea behind convex measures is the addition of a penalty term which behaves according to a functional $\alpha$. Different from equation 3.1, where we assume all models can reproduce market prices perfectly inside the bid-ask spreads and hence do not have to introduce a penalty term, we now consider all models $\theta \in \Theta$ but penalize them according to $\alpha$ if they do not perfectly replicate observed market prices $V$. To do so, Cont (2006) weakens assumption 3.1 and replaces it with the following assumption:

$$\exists \theta \in \Theta, \quad \mathbb{E}^{Q_\theta}[C_i] \in [V^{(i)\text{bid}}, V^{(i)\text{ask}}] \quad \forall i \in I, \quad (3.3)$$

This means that instead of all models $\theta$ replicating option prices within their bid-ask spreads as in equation 3.1, equation 3.3 only requires that there exists at least one model which does so. Cont (2006) then defines a function $\mu^* : \mathcal{X} \to [0, \infty)$ by:

$$\mu^*(X) = \sup_{\theta \in \Theta} \left\{ \mathbb{E}^{Q_\theta}[X] - \alpha(\theta) \right\} - \inf_{\theta \in \Theta} \left\{ \mathbb{E}^{Q_\theta}[X] + \alpha(\theta) \right\} \quad (3.4)$$
for $\alpha$ satisfying

$$\alpha(\theta) \geq \max \{V^{(i)ask} - E^Q[C_i], E^Q[C_i] - V^{(i)bid}, 0\} \tag{3.5}$$

for all $i \in I$ and models $\theta \in \Theta$ (Gupta 2009). $\mu^*$ is then defined as a convex model uncertainty measure if it fulfills the following three axioms next to equation 3.3:

**Cont.1** All benchmark options have a lower model uncertainty than the bid-ask spread:

$$\forall i \in I, \quad \mu(C_i) \leq |V^{(i)ask} - V^{(i)bid}|$$

**Cont.2** As hedging is model dependent, model dependent dynamic hedging does not reduce model uncertainty:

$$\forall \phi \in \Phi, \quad \mu \left( X + \int_0^T \phi_t dS_t \right) = \mu(X)$$

However, if the option can be fully replicated in a model-free way, no model uncertainty exists:

$$\text{if } \exists x \in \mathbb{R}, \phi \in \Phi \text{ s. t. } \forall \theta \in \Theta, \quad X = x + \int_0^T \phi_t dS_t \theta \text{ a. s. } \mu(X) = 0.$$  

**Cont.3** Diversification cannot increase model uncertainty of a portfolio of claims:

$$\forall X_1, X_2 \in \mathcal{X}, \forall \lambda \in [0, 1], \quad \mu(\lambda X_1 + (1 - \lambda)X_2) \leq \lambda \mu(X_1) + (1 - \lambda)\mu(X_2)$$

**Cont.4’** Static hedging via long positions in a claim reduces model uncertainty:

$$\forall X \in \mathcal{X}, \forall a \in [0, 1]^{d+1} \text{ s. t. } \sum_{i=0}^d a_i = 1, \quad \mu \left( a_0 X + \sum_{i=1}^d a_i C_i \right) \leq \mu(X) + \sum_{i=1}^d a_i |V^{(i)ask} - V^{(i)bid}|$$

The last axiom Cont.4 had to be modified to Cont.4’ to limit $a$ to the set $\{a \in [0, 1]^{d+1} : a_0 + \sum_i I a_i = 1\}$ as a convex risk measure cannot extrapolate risk in a proportional way (Gupta 2009).

Cont. (2006) suggests the following calibration error measure $\mu^*_0$:

$$\mu^*(X) = \sup_{\theta \in \Theta} \{E^Q[X] - \alpha_0(\theta)\} - \inf_{\theta \in \Theta} \{E^Q[X] + \alpha_0(\theta)\}$$

81
The convex penalty function $\alpha_0$ is defined by:

$$
\alpha_0(\theta) = ||E^\theta Q[C] \triangleright V||,
$$

with $|| \cdot ||$ being a vector norm on $\mathbb{R}^{|I|}$ and

$$(E^\theta Q[C] \triangleright V)_i = \max\{V^{(i)ask} - E^\theta Q[C_i], E^\theta Q[C_i] - V^{(i)bid}\}.$$ 

This assumes that there exists a model $\theta \in \Theta$ which reproduces all options to within their bid-ask spreads so that $\alpha_0(\theta) = 0$.

### 3.4.4 Risk-Averaging Measures

Branger and Schlag (2004) analyze a worst-case and a Bayesian market risk framework which is close to the one used by Gupta et al. (2010) and Gupta and Reisinger (2011). Given a set $\Theta$ of candidate models denoted by $\theta$, the probability of model $\theta$ is described by $p(\theta)$. $P^\theta$ is then defined as the probability measure for the set of future outcomes under the model $\theta$ and $\rho^{P^\theta}(X)$ measures the market risk of the contract $X$ under the probability measure $P^\theta$. For the Bayesian approach, Branger and Schlag (2004) introduce two methods to combine model and market risk: The first one is called model integration, and it consists of capturing market and model risk in one aggregate model. They call the second method risk integration, which they define as the expectation of some function $\phi$ over all risk measures of the candidate models.

More precisely, model integration according to Branger and Schlag (2004) calculates the distribution of the payoff $X$ at expiry and aggregates all these probability distributions into one big distribution with the following market risk measure:

$$
\rho(X) = \rho^{P}(X), \quad \text{with} \quad P = \sum_{\theta \in \Theta} p(\theta) P^\theta
$$

(3.6)
As one can see, the only thing relevant for the model integration approach are the probabilities of the the states and on the candidate models. However, model and market risk are treated alike in this approach, although most market participants might find it more useful to treat them separately.

The second method introduced by Branger and Schlag (2004), risk integration, defines the risk measure as the expectation of the risk aversion function $\phi$ over the risk measures in the candidate models:

$$
\rho(X) = \sum_{\theta \in \Theta} p(\theta) \phi(\rho^\theta(X)).
$$

(3.7)

The risk aversion function $\phi$ takes different forms dependent on the risk aversion of the decision maker. Assuming a linear function means that the agent is model risk neutral, a convex function corresponds to a risk averse decision maker and a concave function would imply a model risk preferring agent. Branger and Schlag (2004) suggest $\phi(X) = x^n$ with $n > 1$ as convex risk aversion functions.

### 3.4.5 Risk-Differencing Measures

Kerkhof et al. (2002) propose a general framework for the quantification of model risk to determine how much regulatory capital should be set aside extending the work of Artzner et al. (1999). They calculate a worst case measure that should be set aside to respect the capital adequacy framework set out by the Basel Committee by introducing several risk measurement methods. More precisely, Kerkhof et al. (2002) define model uncertainty $\mu$ as the spread between the worst-case market risk measure $\rho$ and a benchmark market risk measure belonging to
a benchmark model $\alpha \in \Theta$:

$$
\mu(X) = \sup_{\theta \in \Theta} \rho^{P_{\theta}}(X) - \rho^{P_{\alpha}}(X),
$$

(3.8)

with each model $\theta$ in $\Theta$ corresponding to a different probability measure $P_{\theta}$ and hence resulting in a different market risk measure $\rho^{P_{\theta}}(X)$ for the claim $X$.

Kerkhof et al. (2002) propose and analyze their worst-case approach for two examples of risk measures $\rho$: Value at Risk and tail conditional expectation. Not surprisingly, they find that one should set aside capital requirements not only to account for market risk, but also for model uncertainty. They further decompose model uncertainty in parameter and specification uncertainty and conclude that specification uncertainty is often much higher than estimation uncertainty.

### 3.5 Summary

This section has clarified the distinction between risk and uncertainty, reviewed the literature on model uncertainty and introduced five model uncertainty measures following Gupta et al. (2010). The review of the literature on model uncertainty has highlighted the lack of literature depth on model/parameter robustness and uniqueness. While initial investigations into misestimation have been done, the literature on misspecification in the context of derivatives pricing seems relatively limited to date.

As we have outlined, the literature has proposed two different paradigms to deal with model uncertainty (Cont 2006). The first one, which we will investigate further in this thesis, is the aggregation of several candidate models through ensemble methods. This approach has been used extensively in the statistical
literature. Several empirical studies have shown that ensemble predictions can often be much more accurate than the forecasts of the base learners (Freund and Schapire 1996; Bauer 1999; Dietterich 2000a) and various authors have suggested justifications for the success of ensemble methods from a theoretical perspective (Kittler et al. 1998; Allwein et al. 2000; Kleinberg 2000). This approach will be reviewed in more detail in the next section.

The second paradigm is based on worst case or maxmin approaches and has been introduced by Gilboa and Schmeidler (1989) and analyzed in the context of asset pricing among others by Epstein and Wang (1994) and Routledge and Zin (2009). We have analyzed these approaches in the context of model uncertainty measures. As we have shown, model uncertainty measures can be broken down into two main categories: Coherent (Artzner et al. 1999; Frittelli and Gianin 2002; Cont 2006) and convex (Föllmer and Schied 2002; Frittelli and Gianin 2002; Cont 2006) measures.
Chapter 4

Ensemble Methods

4.1 Introduction

The previous two sections reviewed the literature on option pricing models and model uncertainty. Although the option pricing literature introduced many advanced and generally more accurate models than the Black-Scholes benchmark, model uncertainty remains. The more recent models based on Lévy processes can significantly reduce - but not completely eliminate - in-sample pricing errors as they overcome the main shortcomings of the Black-Scholes framework (Carr et al. 1998, Carr and Wu 2003a). However, due to their higher variance, we do not know if their improved pricing performance will extend to out-of-sample data. More general, if we could build the “perfect” derivatives pricing model, which would give us correct prices for all maturities and strikes for in-sample and out-of-sample data, model uncertainty would not exist and we would not need to investigate the use of ensemble methods. The main principle behind ensemble methods is that perfect models do not exist in reality (Brown 2010).

The idea of combining several opinions, forecasts or models to reduce uncertainty
is not new and has already been formalized approximately 200 years ago by Laplace (1818), who demonstrated that a suitable combination of two probabilistic methods will have superior performance than either one of the two. Plenty of similar terms next to ensemble methods have been introduced to describe the combination of several models into one, e.g. aggregation, combination, committee or fusion (Dindar and Marwala 2004; Drucker et al. 1994; Jain et al. 2000; Kittler et al. 1998; Lam and Suen 1995; Polikar et al. 2008). Despite the recent success of ensemble methods, a unified theory about their success does not yet exist (Re and Valentini 2012).

This section will start with an analysis of the generalization error to understand why ensembles are successful. We will then investigate different methods to build ensembles based on Mendes-Moreira et al. (2012). In particular, we will review the three steps of the ensemble process: ensemble generation, ensemble pruning and ensemble integration.

4.2 Rationale Behind Ensemble Methods

Re and Valentini (2012) present three different theories why ensembles are successful. The first one describes ensemble methods in the context of large margin classifiers following Mason et al. (2000). They demonstrate that ensembles enlarge the margins and improve the generalization capabilities of output coding techniques (Allwein et al. 2000) and boosting based ensemble methods (Schapire et al. 1998).

The second theory mentionned by Re and Valentini (2012) is based on the classical bias-variance decomposition of Geman et al. (1992), which we will review in more detail in section 4.2.2. This decomposition demonstrates that ensembles
can reduce variance (Breiman, 1996a; Lam and Suen, 1997) or bias and variance (Breiman, 1998; Kong and Dietterich, 1995). Domingos (1995) showed how to unify these two theories, as they are basically two ways of looking at the same issue.

The third theory why ensembles work is based on Kleinberg (1996) and Kleinberg (2000). His stochastic modeling method combines many weak learners to form a strong ensemble and has the advantage that it has relatively low variance and continued low bias when applied to out-of-sample data.

Dietterich (2000b) lists three fundamental reasons why ensembles are successful in machine learning applications. The first one is statistical. Models can be seen as searching a hypothesis space $\mathcal{H}$ to identify the best hypothesis. However, the statistical problem arises as we often have only limited datasets in practice. Hence, we can find many different hypotheses in $\mathcal{H}$ which fit reasonable well and we do not know which one of them has the best generalization performance. This makes it difficult to choose among them. Therefore, the use of ensemble methods can help to avoid this issue by averaging over several models to get a good approximation of the unknown true hypothesis.

The second reason is computational. Many models work by performing some form of local searches such as gradient descent to minimize error functions that could get stuck in local optima. An ensemble constructed by starting the local search from many different points may provide a better approximation to the true unknown function.

The third argument provided by Dietterich is representational. In many situations, the unknown function we are looking for is not included in $\mathcal{H}$. However, a combination of several hypotheses drawn from $\mathcal{H}$ can enlarge the space of representable functions, which could then also include the unknown true function.
Dietterich (2000b) also stresses that one key to successful ensemble methods is to construct individual learning algorithms with prediction accuracy above 50% whose errors are at least somewhat uncorrelated. Hence, using a diverse set of accurate models is important for the success of ensembles and motivation for the introduction of a wide range of diverse option pricing models in section 2. The fact that error correlations between base models contribute to ensemble performance has been known for a while (Brown and Wyatt, 2005). Perrone (1993) proved in his doctoral thesis that we obtain a \( \frac{1}{N} \), \( N = |\Theta| \) variance reduction if the correlation between the base models is zero. We will investigate further the concept of error diversity in the following two sections.

### 4.2.1 Ambiguity Decomposition

Krogh and Vedelsby (1995) demonstrated that the average squared error of an ensemble for any estimated data point is always lower than or equal to the average squared error of the single model predictions. Although they describe the ambiguity decomposition for an ensemble of \( N \) Neural Networks, we will outline below the example for option pricing models. Assuming that the ensemble price is a weighted price of the base model prices (\( \hat{V}_{\text{ens}} = \sum_{i=1}^{N} w_i \hat{V}_i \) and \( \sum_{i=1}^{N} w_i = 1 \)), they show that the squared error of a linearly combined ensemble of \( N = |\Theta| \) base models \( \theta \) can be broken down into two components:

\[
(\hat{V}_{\text{ens}} - V)^2 = \sum_{i=1}^{N} [w_i (\hat{V}_i - V)^2] - \sum_{i=1}^{N} [w_i (\hat{V}_i - \hat{V}_{\text{ens}})^2]
\]

(4.1)

with \( V \) describing the observed plain vanilla option market price, \( \hat{V}_{\text{ens}} \) the ensemble estimate and \( \hat{V}_i \) the estimate of base model \( \theta_i \).

\footnote{The extension to bid-ask spreads is left out for the time being. We refer the interested reader to Gupta (2009), whose approach could also be applied to our context.}
The reader is referred to the original for further details of the proof and to Brown (2004) for two examples that prove the ambiguity decomposition in more detail. We will review different methods to weight the base models in section 4.3. Following Brown et al. (2005), we will outline below an alternative proof similar to the bias-variance decomposition originally introduced by Geman et al. (1992):

\[
\sum_i w_i (\hat{V}_i - V)^2 = \sum_i w_i (\hat{V}_i - \hat{V}_{ens} + \hat{V}_{ens} - V)^2 \\
= \sum_i w_i [(\hat{V}_i - \hat{V}_{ens})^2 + (\hat{V}_{ens} - V)^2 + 2(\hat{V}_i - \hat{V}_{ens})(\hat{V}_{ens} - V)] \\
= \sum_i w_i (\hat{V}_i - \hat{V}_{ens})^2 + (\hat{V}_{ens} - V)^2 \\
(V_{ens} - V)^2 = \sum_i w_i (\hat{V}_i - V)^2 - \sum_i w_i (\hat{V}_i - \hat{V}_{ens})^2
\]

(4.2)

The first term on the right hand side of the equation describes the average squared error of the individual base models. The second term, called ambiguity, quantifies the error correlation between the base models as it measures the difference between the forecasts of the base models and the ensemble. As this second term is always positive, ambiguity decomposition basically tells us that the average squared error of the ensemble estimate will always be lower than or equal to the average squared error of the base models.

The larger the second term, the higher will be the diversity effect for the ensemble. However, as the difference between the base model forecasts increases, the first term increases as well, hence reducing the accuracy. This tradeoff is called diversity - accuracy tradeoff as one needs to find the right balance between the two terms to minimize the ensemble error. Several models we will review in section 4.3 will explicitly take this tradeoff into account.

As one can see from the above, it is possible for a single base model to have
a lower error than the ensemble error, but the difficulty lies in identifying this model beforehand. As we do not know how to identify the best base model in advance, we compare the ensemble to picking a random base model.

The ambiguity decomposition outlined above is a property of ensembles trained on a single dataset. However, it doesn’t consider the generalization to out-of-sample data, the distribution over possible training sets or all potential weight initialisations (Brown et al., 2005). Hence, we turn now to the bias-variance-covariance decomposition of Ueda and Nakano (1996) as it takes into account the expected error on new data points such as out-of-sample option prices.

### 4.2.2 Bias, Variance and Covariance

Geman et al. (1992) introduced the bias-variance decomposition for a neural network. They demonstrate that the bias-variance decomposition for squared error functions breaks the generalisation error of an estimator into two parts: bias and variance. However, these two work diametrically: A reduction of variance will lead to an increase in bias and vice versa leading to the following decomposition:

$$
\begin{align*}
\mathbb{E}[(\hat{V} - V)^2] &= \mathbb{E}[(\hat{V} - \mathbb{E}[\hat{V}])^2] + (\mathbb{E}[\hat{V}] - V)^2 \\
MSE(\hat{V}) &= var(\hat{V}) + bias(\hat{V})^2
\end{align*}
$$

(4.3)

with the same notation as before. If the estimator is a convex combined ensemble, the variance component breaks down further, leading to the bias-variance-covariance decomposition of Ueda and Nakano (1996). Following their work, we define the three elements of the decomposition below, starting with bias, the
average bias of the base models:

\[
\overline{\text{bias}} = \frac{1}{N} \sum_{i=1}^{N} (\mathbb{E}[\hat{V}_i] - V).
\] (4.4)

The next element is the averaged variance of the base models denoted by \( \overline{\text{var}} \):

\[
\overline{\text{var}} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}[(\hat{V}_i - \mathbb{E}[\hat{V}_i])^2]
\] (4.5)

The last element is the averaged covariance of the base models:

\[
\overline{\text{covar}} = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i, j=1}^{N} \mathbb{E}[(\hat{V}_i - \mathbb{E}[\hat{V}_i])(\hat{V}_j - \mathbb{E}[\hat{V}_j])]
\] (4.6)

We can now combine these three elements to the bias-variance-covariance decomposition of the mean-squared-error (Ueda and Nakano, 1996):

\[
\mathbb{E}[(\hat{V}_{\text{ens}} - V)^2] = \overline{\text{bias}}^2 + \frac{1}{N} \overline{\text{var}} + (1 - \frac{1}{N}) \overline{\text{covar}}
\] (4.7)

The covariance term shows us that the mean squared error of an ensemble of option pricing models is significantly influenced by the amount of pricing error correlation between the base models. Hence, we would like to decrease this covariance term without influencing the bias or variance and will have a closer look at methods such as negative correlation learning ensemble methods in section 4.5.

### 4.3 Classifying Ensemble Methods

Following the significant amount of research on ensemble methods in recent years, a wide range of ensemble methods and algorithms have been developed with promising results in classification and regression tasks (e.g. Breiman (2001) or
Liu and Maheu (2009). However, most work on ensemble methods focuses on classification tasks, although in our context of derivatives pricing we face a regression problem. As classification techniques are often not applicable for regression tasks, we will only review the literature on ensemble methods for regression, closely following the survey of Mendes-Moreira et al. (2012). More precisely, we partition the ensemble process into three steps: ensemble generation, ensemble pruning and ensemble integration. In our context, the step of ensemble generation (section 4.4) includes the calibration of several candidate option pricing models to existing market prices. The second step, ensemble pruning (section 4.5), will remove some of the models calibrated in step 1, e.g. due to unsatisfactory pricing performance or similar criteria. The third and last step of the process, ensemble integration (section 4.6), defines the aggregation scheme of the models which have not been eliminated in step 2. At the end of step 3, we will obtain the ensemble prediction (denoted by $\hat{V}_{\text{ens}}$) of option prices. Further details how ensemble methods can be used to reduce the bias and/or variance of option pricing models based on Lévy process will be outlined in chapters 5 and 6.

4.4 Ensemble Generation

As described above, the first step of the ensemble process is called ensemble generation, starting with a set of candidate models:

$$\Theta_0 = \{\theta_i, \quad i = 1, ..., N_0\} \quad (4.8)$$

An ensemble is called homogeneous if all base models $\theta$ belong to the same class of models (e.g. stochastic volatility models, section 2.3.4.3). If the base models are more diverse, the ensemble is called heterogeneous (Mendes-Moreira et al. 2012).
This second approach, used in this thesis, is expected to obtain a more diverse ensemble (Webb and Zheng, 2004) with generally better performance (Wichard et al., 2003). As we have seen in section 4.2, diversity is next to accuracy one of the key success factors of ensembles (Perrone and Cooper, 1993). However, we do not have control about the diversity of the base models in the ensemble generation phase. For example, our option pricing models used as base models could have very correlated pricing errors. Nevertheless, by calibrating a larger number of models from different classes of option pricing models, we increase the likelihood of having an accurate and diverse subset of base models, although at the expense of computational requirements. This increased probability is motivation for the introduction of a diverse range of option pricing models in section 2. The task of choosing the appropriate subset of base models will be dealt with during the ensemble pruning phase in section 4.5.

Ensemble generation methods can be classified according to how they attempt to generate different base models: either through manipulating the data or through manipulating the modeling process (Mendes-Moreira et al., 2012). Data manipulation can be further broken down into subsampling from the training data and manipulating input features or output variables. The modeling process manipulation can also be subdivided further: It can be achieved by using different parameter sets or manipulating the induction algorithm or the resulting model. As we do not want to change the Lévy Process base models, we do not analyze model manipulation in this literature review and refer the interested reader to Mendes-Moreira et al. (2012) for details of this approach.
4.4.1 Subsampling

Subsampling from the training set is reported to be successful for unstable learning algorithms such as decision trees or neural nets (Breiman, 1996b; Dietterich, 1997). (Kim et al., 2003) also reports their successful application to support vector machines.

Bagging (Breiman, 1996a) stands for bootstrap aggregation and is a well-known method of this approach. It generates multiple versions of one base model by training it on several subsamples and uses these models to build an ensemble. The aggregation scheme averages over the subsamples when predicting a numerical outcome and does a plurality vote in classification tasks. The subsamples of the training set are formed by making bootstrap replicates of the learning data and uses these as new learning sets. Empirical tests on real and simulated data sets attest bagging substantial gains in accuracy and explain the theoretical justification behind it (Breiman, 1996a; Domingos, 1997).

Extensions of bagging, which focus mainly on reducing the computational cost by reducing the number of generated subsamples of the training data include Buehlmann (2010) and Buja and Stuetzle (2006).

Boosting (Schapire, 1990) is another well-known method of this ensemble approach. Freund and Schapire (1996) introduced AdaBoost (adaptive boosting), probably the most popular ensemble technique based on boosting. The main idea behind all boosting techniques is that the combination of many weak learners (defined as having a forecasting accuracy slightly above 0.5) can build a strong ensemble. The main difference to bagging is that subsamples are initially chosen at random (like in bagging), but later an increased weight is allocated to subsamples with less accurate predictions, hence they have a higher probability of being selected. As boosting was originally developed for classification tasks, several
authors suggested extensions of the algorithm to apply it to regression problems. 

Freund and Schapire (1997) were the first to introduce an adaptation of the AdaBoost algorithm for regression problems. Similar to Avnimelech and Intrator (1999) and Shrestha and Solomatine (2006), they modified the error function used by AdaBoost to be applicable to regression tasks (Mendes-Moreira et al., 2012). The forward stage-wise additive model, introduced in Friedman (2001), is another example of subsampling. It shares with boosting the intuition to focus the learning process on the examples which have been misclassified. However, instead of assigning higher weights to examples with low prediction accuracy as in boosting, it focuses on minimizing the residuals of the predictions made at the previous iteration. Friedman (2002) further extended this approach to avoid overfitting and is known as stochastic gradient descent boosting.

4.4.2 Input Features Manipulation

Manipulating the input features involves changing the representation of examples to create several different training sets. Two different variants of this approach exist: The first one, called feature selection, creates the new training sets as subsets of the original training data. The second one, called feature transformation, applies some transformation to the elements of the original training set to create new training sets. Feature selection methods include the random subspace method of Ho (1998) and improvements of it by Schclar and Rokach (2009). Opitz (1999) proposes a quite ambitious example of feature selection methods as his algorithm tries to minimize the error of the base models and maximize the ambiguity term of the ensemble at the same time. However, he claims to have superior results than bagging or AdaBoost methods.

A relatively simple feature transformation example was introduced by Frank and
Pfahringer (2006) under the name of input smearing. The basic idea behind this algorithm is adding Gaussian noise to the data points of the training data. An earlier example of introducing Gaussian noise to transform the training data has been presented in Raviv and Intrator (1996).

Rodríguez et al. (2006) provide an example how to combine both approaches of feature manipulation. Their approach, called rotation forest, splits the training set in disjoint subsets to increase the likelihood of improved diversity.

### 4.4.3 Output Variable Manipulation

Manipulating the output data received relatively little attention in the academic literature so far, especially in the context of regression. However, we can highlight the work of Breiman (2000), whose basic idea of adding Gaussian noise to the output variable (called output smearing) is similar to the one from Raviv and Intrator (1996) and Frank and Pfahringer (2006) for transforming the input features. Breiman (2000) also analyzed his approach for classification tasks with the same idea and calls it output flipping. An empirical comparison with bagging shows the slightly improved generalization error of output smearing and output flipping, but mixed results relative to boosting. However, it is not clear how such techniques should be applied in the context of derivatives pricing where the output variables (listed vanilla option prices) are given. Table 4.1 summarizes the different ensemble generation methods introduced in this section.\(^2\)

\(^2\)Table adapted from Mendes-Moreira et al. (2012).
<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>Approach</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bagging</td>
<td>Breiman (1996a)</td>
<td>Subsampling</td>
<td>Constructs multiple data sets by randomly selecting subsets</td>
</tr>
<tr>
<td>Forward Stage-Wise Additive Model</td>
<td>Friedman (2001)</td>
<td>Subsampling</td>
<td>Modifies Adaboost by minimizing residuals of previous round predictions</td>
</tr>
<tr>
<td>Multiple Additive Regression Trees</td>
<td>Friedman (2002)</td>
<td>Subsampling</td>
<td>Modifies Adaboost by minimizing stochastic residuals of previous round predictions</td>
</tr>
<tr>
<td>AdaBoost.RT</td>
<td>Shrestha and Solomatine (2006)</td>
<td>Subsampling</td>
<td>Modifies AdaBoost classification algorithm by discretizing error into a binary one</td>
</tr>
<tr>
<td>Random Subspace</td>
<td>Ho (1998)</td>
<td>Input Feature Manipulation</td>
<td>Constructs subsample datasets by randomly selecting them</td>
</tr>
<tr>
<td>Ensemble Feature Selection</td>
<td>Opitz (1999)</td>
<td>Input Feature Manipulation</td>
<td>Aims to minimize the error of the base models and maximize the ambiguity term of the ensemble at the same time</td>
</tr>
<tr>
<td>Bootstrap Ensemble with Noise</td>
<td>Raviv and Intrator (1996)</td>
<td>Input Feature Manipulation</td>
<td>Adds Gaussian noise to data points of the training data</td>
</tr>
<tr>
<td>Input Smearing</td>
<td>Frank and Pfahringer (2006)</td>
<td>Input Feature Manipulation</td>
<td>Adds Gaussian noise to data points of the training data</td>
</tr>
<tr>
<td>Rotation Forests</td>
<td>Rodríguez et al. (2006)</td>
<td>Input Feature Manipulation</td>
<td>Creates disjoint subsets to increase the likelihood of improved diversity</td>
</tr>
<tr>
<td>Output Smearing</td>
<td>Breiman (2000)</td>
<td>Output Feature Manipulation</td>
<td>Adds Gaussian noise to target output variable</td>
</tr>
</tbody>
</table>

Table 4.1: Overview of Ensemble Generation Methods
4.5 Ensemble Pruning

Continuing to follow closely Mendes-Moreira et al. (2012) and Mendes-Moreira et al. (2007), we investigate in this section different ensemble pruning methods which eliminate some of the models generated in the ensemble generation step. The methods introduced for ensemble generation create a diverse set of ensembles, however they do not ensure that we use only as many models as needed to ensure the best accuracy.

Ensemble pruning describes the process of choosing the appropriate subset $\Theta_1$ of the initial candidate pool of models $\Theta_0$:

$$\Theta_1 \subseteq \Theta_0$$

(4.9)

Ensemble pruning methods try to improve ensemble accuracy and/or to reduce computational cost. They can be divided into partitioning-based and search-based methods. Partitioning based approaches split the candidate models into subgroups based on certain predefined criteria. Search based approaches try to find a subset with improved ensemble accuracy by either adding or removing models from the initial candidate pool. Furthermore, the different pruning approaches could be classified according to their stopping criterion: Direct ensemble pruning methods are approaches, where the number of models used is determined ex ante, whereas evaluation ensemble methods determine the number of models used according to the ensemble accuracy (Mendes-Moreira et al., 2012).

4.5.1 Partitioning Based Approaches

The main idea behind partitioning is that many models of certain subcategories are similar and therefore redundant. In our context, this would mean that we
might choose only the best model(s) of the different types of option pricing model categories introduced in chapter 2.

To achieve the best possible ensemble of option pricing models, we need to combine a set of accurate, yet diverse models. In partitioning based methods, this is achieved by taking the most accurate models of various subgroups, which are supposed to be diverse, and then combining them.

One can either have a pre-defined categorization of models (e.g. as we have shown in chapter 2) or classify models in subcategories according to the similarity of their predictions. Lazarevic and Obradovic (2001) group the candidate models into subcategories based on the similarity of their prediction vectors. However, for their approach, one has to specify the number of different subcategories in advance. Coelho and von Zuben (2006) use the adaptive radius immune algorithm to cluster similar base models. Their approach has the advantage that the number of subcategories is chosen automatically and hence does not have to be specified in advance.

### 4.5.2 Search Based Approaches

Search based approaches can be split into three different categories following the terminology of the feature selection literature (Aha and Bankert 1996; Molina et al. 2002): exponential, randomized and sequential approaches.

#### 4.5.2.1 Exponential Pruning Algorithms

These methods search all possible subsets $\Theta_1$ of the candidate pool $\Theta_0$ calibrated during the ensemble generation step. Given $N$ different models, we have $2^N - 1$ different possibilities for the subset $\Theta_1$. Perrone and Cooper (1993) recommend this ensemble pruning method for small values of $N$. However, the exponential
search approach is computationally expensive and recommended only for \( N \leq 30 \) (Martínez-Munoz and Suárez, 2006).

### 4.5.2.2 Randomized Pruning Algorithms

Ruta and Gabrys (2001) investigated three evolutionary optimization methods for the classifier selection task, but their results have limited generalization validity due to the small pool of base models. Partridge and Yates (1996) also analyze randomized pruning algorithms, but find that the results are not promising. Zhou et al. (2002) claim that using many base models could be better than using all base models and develop the GASEN (Genetic Algorithm based Selective Ensemble) approach upon this idea. This methods starts by assigning random weights to the base models and Zhou et al. (2002) report empirical results superior than bagging and boosting.

### 4.5.2.3 Sequential Pruning Algorithms

Sequential pruning algorithms continuously change the subset \( \Theta_1 \) until they have found the ideal one. There are four different approaches to do so:

- **Forward Subset Selection:** This method starts with an empty subset \( \Theta_1 \) and adds successively models in each iteration to the ensemble.

- **Backward Subset Selection:** This methods starts with the subset \( \Theta_1 = \Theta_0 \), hence including all candidate models in the ensemble at the beginning. Successively, models are eliminated from the ensemble in each iteration.

- **Forward-Backward Subset Selection:** This method can have backward and forward steps in the subset selection process.
• **Ranking-Based Subset Selection:** This method ranks the candidate models according to a certain criteria such as pricing error and includes only the top \( n \) models in the ensemble.

**Forward Subset Selection**

Forward subset selection methods build ensembles by starting with an empty subset \( \Theta_1 \) and iteratively adding models to minimize the expected pricing error of the ensemble in our context of derivatives pricing. Coelho and von Zuben (2006) introduce two forward subset selection methods: forward without exploration and forward with exploration. The first one ranks all candidate models according to their performance (pricing accuracy in our context). It starts with the best one, adds the second best model etc. until the performance of the ensemble is not improved by adding the next best ensemble. Forward subset selection also starts with the best model, but then tests all potential models for addition to the ensemble and adds the one with the best effect on the overall ensemble accuracy. Both approaches continue this process until the performance of the ensemble is not improved anymore by adding new models. These methods have been introduced by Perrone and Cooper (1993), who used them in the ensemble pruning step.

Partridge and Yates (1996) present a method similar to forward with exploration introduced above. Their inclusion criteria is a measure of diversity instead of accuracy and the number of included models is fixed in advance.

Martinez-Munoz and Suárez (2006) also present an algorithm similar to forward with exploration. At each iteration step, one adds the model with the highest reduction of the ensemble generalization error to the ensemble.
Backward Subset Selection
Backward subset selection approaches include in the initial step all models $\Theta_0$ calibrated during the ensemble generation step in the subset $\Theta_1$. Models are then iteratively removed to improve the overall ensemble performance. [Coelho and von Zuben (2006)] introduce two backward subset selection methods: pruning without exploration and pruning with exploration. In the first one, all candidate option pricing models would be ranked according to accuracy and the worst one is removed from the model. This step is repeated until removal of the worst base model does no longer improve ensemble accuracy. However, this approach does not take into account error correlation of the different models. Backward subset selection works analogue to forward subset selection. At each iteration, not the worst model (according to some measure of pricing accuracy) is removed, but the one which results in the best improvement of ensemble performance. This process stops when the removal of an option pricing model does not improve ensemble performance anymore.

Forward-Backward Subset Selection
Forward-backward subset selection methods combine features of the two above introduced approaches. [Moreira (2006)] introduce a method that starts from a predetermined number of $n$ models. At each iteration, it conducts a forward, as well as a backward step as described above. The forward step includes the candidate model which leads to the highest ensemble improvement. The backward step then removes the model which has the smallest effect on improving ensemble accuracy. This process continues until the forward and backward step select the same model.
Margineantu and Dietterich (1997) is another example of a forward-backward subset selection algorithm and initially similar to the forward subset selection approach of Coelho and von Zuben (2006) described above. However, starting from the iteration where the $n$th model $\theta$ is added to the ensemble, the algorithm tests whether the replacement of any of the already selected models by a new candidate model could improve overall ensemble performance. It stops as soon as it reaches a pre-defined number of iterations.

**Ranking Based Subset Selection**

Ranking based subset selection methods order all candidate models $\Theta_0$ according to a performance metric such as pricing accuracy in our context. They build then an ensemble based on the top $n$ models. $n$ is either specified in advance or can be dynamically adjusted based on a certain criteria such as minimum pricing error on the validation data set etc.

Partridge and Yates (1996) analyze such a method where the performance metric is model accuracy and hence the $n$ most accurate models are chosen to build the ensemble. However, results are not very good as this approach does not guarantee diversity. Rooney et al. (2004) introduce a method called dynamic integration that aims to mitigate this problem by using a performance metric that balances the accuracy-diversity tradeoff.

Perrone and Cooper (1993) propose an algorithm which removes similar models from the candidate pool $\Theta_0$ and then uses a ranking based approach to combine the remaining models. The correlation matrix of the model predictions is used to define similarity and models with a correlation above a certain threshold are removed from the candidate pool.

The literature on ensemble pruning methods - especially in a regression context
- is quite limited so far. In general, ensemble pruning is usually applied to and analyzed for ensemble generation methods which produce a large number of base models. Hernandez-Lobato and Martinez-Munoz (2006) attest ensemble pruning methods an improvement in ensemble performance, especially for ensemble generation methods such as bagging, which do not typically include a pruning step.

However, no detailed comparisons between different ensemble pruning methods are available in the literature. Some authors of the above introduced ensemble pruning methods compare their suggested methods to previous ones (e.g. Partridge and Yates (1996), Coelho and von Zuben (2006) or Ruta and Gabrys (2001)), but due to the limited scope and bias of their comparisons, these results cannot be generalized.

Table 4.2 summarizes the analyzed ensemble pruning methods.3 We will now turn to the last step of the ensemble process, ensemble integration.

---

3Table adapted from Mendes-Moreira et al. (2012).
<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>Approach</th>
<th>Stoping Criterion</th>
<th>Evaluation Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustering by Deterministic Annealing</td>
<td>Lazarevic and Obradovic [2001]</td>
<td>Partitioning-Based</td>
<td>Pre-defined number of clusters with best model from each cluster</td>
<td>Partitioning according to weighted distance, selection based on base model accuracy</td>
</tr>
<tr>
<td>Adaptive Radius Immune Algorithm</td>
<td>Coelho and von Zuben [2006]</td>
<td>Partitioning-Based</td>
<td>ARIA partitions similar base models into groups, best model of each cluster used in ensemble</td>
<td>Partitioning according to ARIA, selection based on base model accuracy</td>
</tr>
<tr>
<td>Correlation between Errors</td>
<td>Perrone and Cooper [1993]</td>
<td>Search-Based: Exponential</td>
<td>Pre-defined number of models</td>
<td>Mean pairwise correlation between errors</td>
</tr>
<tr>
<td>GASEN</td>
<td>Zhou et al. (2002)</td>
<td>Search-Based: Randomized</td>
<td>Pre-defined threshold for the fitness</td>
<td>Fitness ($\frac{1}{\text{error}}$) of the base models</td>
</tr>
<tr>
<td>Forward without Exploration</td>
<td>Coelho and von Zuben [2006]</td>
<td>Search-Based: Forward</td>
<td>Ensemble accuracy starts to decrease</td>
<td>Ranking of base model’s accuracy</td>
</tr>
<tr>
<td>Forward with Exploration</td>
<td>Coelho and von Zuben [2006]</td>
<td>Search-Based: Forward</td>
<td>Ensemble accuracy starts to decrease</td>
<td>Ensemble accuracy</td>
</tr>
<tr>
<td>Picking Maximum Diversity</td>
<td>Partridge and Yates [1996]</td>
<td>Search-Based: Forward</td>
<td>Pre-defined number of models is reached</td>
<td>Base model diversity</td>
</tr>
<tr>
<td>Ordering Bagging Ensembles</td>
<td>Martinez-Munoz and Suárez [2006]</td>
<td>Search-Based: Forward</td>
<td>Pre-defined number of models is reached</td>
<td>Base model with best reduction of ensemble generalization error</td>
</tr>
<tr>
<td>Backward without Exploration</td>
<td>Coelho and von Zuben [2006]</td>
<td>Search-Based: Backward</td>
<td>Ensemble accuracy starts to decrease</td>
<td>Ranking of base model’s accuracy</td>
</tr>
<tr>
<td>Backward with Exploration</td>
<td>Coelho and von Zuben [2006]</td>
<td>Search-Based: Backward</td>
<td>Ensemble accuracy starts to decrease</td>
<td>Ensemble accuracy</td>
</tr>
<tr>
<td>Greedy Search</td>
<td>Moreira [2006]</td>
<td>Search-Based: Forward-Backward</td>
<td>Last added model = last removed model</td>
<td>Ensemble accuracy</td>
</tr>
<tr>
<td>Reduce-Error Pruning with Backfitting</td>
<td>Margineantu and Dietterich [1997]</td>
<td>Search-Based: Forward-Backward</td>
<td>Pre-defined number of iterations</td>
<td>Ensemble accuracy</td>
</tr>
</tbody>
</table>

Table 4.2: Overview of Ensemble Pruning Methods
4.6 Ensemble Integration

Following the ensemble generation and ensemble pruning step, we will now investigate the last step of the ensemble process: ensemble integration. The ensemble integration step describes how the remaining calibrated models are combined to a single ensemble forecast. This can be formalized in our context as:

$$\hat{V}_{ens} = \sum_{i=1}^{N} w_{\theta_i} \ast \hat{V}_{\theta_i}$$ (4.10)

with $\hat{V}_{ens}$ describing the ensemble price forecast, $w_{\theta_i}$ the weight of model $\theta_i$ and $\hat{V}_{\theta_i}$ the individual model option price prediction.

Merz (1998) splits ensemble integration methods into constant and nonconstant weighting schemes in his doctoral thesis. In the first approach, the weights $w$ are the same for all models, whereas with the second method the weights vary according to which model $\theta$ is used for the option price prediction. He also highlights the issue of multicollinearity, which occurs when the predictions of the base models are highly correlated. Suggestions to avoid or reduce the existence of multicollinearity include several methods applied during the ensemble generation or ensemble pruning step to guarantee an accurate, yet diverse (and hence not perfectly correlated) set of base models.

We will outline below constant and non-constant weighting functions for the ensemble integration step, based on the work of Merz (1998), Mendes-Moreira et al. (2007) and Mendes-Moreira et al. (2012).

4.6.1 Constant Weighting Functions

Constant weighting integration methods use the same set of weights, independent of which data is estimated (e.g. it does not differentiate weights for ITM, ATM
and OTM options). Perrone and Cooper (1993) were among the first to classify ensemble integration methods and the Basic Ensemble Method (BEM) is nothing else but the simple average of the base models predictions, hence all $w = \frac{1}{N}$:

$$\hat{V}_{BEM} = \frac{1}{N} \sum_{i=1}^{N} \hat{V}_{\theta_i}$$

(4.11)

This method implies that the base model errors ($V - \hat{V}_i$) are not correlated and have mean zero. In the same paper, Perrone and Cooper (1993) also introduced a more complex ensemble integration method, called Generalized Ensemble Method (GEM), to address the issue of mutual independence and zero mean assumed by BEM. In GEM, the weights $w$ are dependent on the model errors (the higher the errors of model $\theta$ on the validation data set, the lower its weight). Furthermore, GEM takes into account the correlation between the errors of the different base models (Mendes-Moreira et al., 2012). However, contrary to BEM, GEM suffers the issue of multicollinearity. The simple median is similar to BEM and also doesn’t suffer multicollinearity. Buehlmann (2010) demonstrates its use as integration method for bagging.

Breiman (1996c) introduced the concept of stacked regression, based on the stacking framework previously suggested for classification tasks by Wolpert (1992). The general idea behind stacking is the combination of a set of calibrated base models which takes into account the bias of these models on the specific training data. Assuming we have a calibration data set with $M$ option prices, the algorithm searches the weights $w_i$ that minimize the squared error:

$$\sum_{j=1}^{M} \left[ V - \sum_{i=1}^{N} w_i \times \hat{V}_i \right]^2$$

(4.12)

This approach tends to overfit as the same training data is used to calibrate...
the models and determine their weight, but Breiman suggests the use of cross validation to ameliorate this risk. However, multicollinearity is another possible issue (Breiman, 1996c). An investigation into ridge regression to solve this issue did not give satisfactory results. Another observation of Breiman is that many weights $w_i$ are estimated at or close to zero, hence underlining the importance of the ensemble pruning step.

Wang et al. (2003) propose a method that used the weights inversely proportional to the expected error of $\hat{V}_i$.

Dynamic Weighting methods, introduced by Puuronen et al. (1999), overweight base models with good performance on the training data relative to weaker base models to build the ensemble based on a weighted average. Rooney et al. (2004) analyzed the dynamic weighting methods and assigns them similar performance as stacked regression based on an empirical analysis of a range of standardized datasets.

Search methods, similar to those analyzed in section 4.5.2, have also been investigated to estimate the weights $w$. For example, Ortiz-Boyer et al. (2005), attest an evolutionary search algorithm superior performance than GEM and BEM integration methods. Caruana and Niculescu-Mizil (2004) analyze an algorithm similar to BEM which is also search based, but with the difference that the integration step is embedded in the pruning step and models can be selected multiple times. The final integration method is a simple average as in BEM, but with the difference that models could be used several times. Merz (1998) conducts in his doctoral thesis an empirical comparison and analytical explanation of several weighting functions for classification and regression.
tasks. He finds that his developed algorithm singular value decomposition (SVD) helps to avoid multicollinearity without discarding any models with good prediction accuracy. SVD does so by taking into account the correlation between different model forecasts and retaining only uncorrelated models. These models are then combined using a simple weighted average as in equation 4.11.

4.6.2 Non-Constant Weighting Functions

Non-Constant Weighting functions use different weighting functions according to which data is predicted, e.g. if the option is ITM, ATM or OTM or depending on the time to maturity. Mendes-Moreira et al. (2007) classified non-constant weighting schemes as static (weighting is determined during the calibration process) or dynamic (weighting is determined at the time of prediction). Static methods can either predetermine models to use for certain areas, e.g. models $\theta_1$, $\theta_2$, $\theta_3$ for ATM option prices.

Another approach of static models defines certain areas of expertise for each model and uses only this model for the forecasting. Kuncheva (2002) refers to this method as static selection. Todorovski and Džeroski (2003) apply the area of expertise concept in a classification context, however their approach could also be applied to regression tasks.

Dynamic weighting methods choose the appropriate base models to use at the time of the prediction, hence it chooses the expected best base models on the fly (Mendes-Moreira et al., 2012). Contrary to areas of expertise methods introduced earlier, these areas are not previously defined, but based on the performance of the base models on similar data of the training set. In our context, this might be the option pricing models which have the lowest measure of pricing error on similar options in terms of moneyness and maturity. Merz (1996) analyzes this...
dynamic approach in detail. If only the best base model on similar data is used, this is usually known as dynamic selection or adaptive selection (Woods et al., 1997; Kuncheva, 2002). If an ensemble is formed based on the best $n$ models on similar data, the approach is called dynamic fusion (Woods et al., 1997). Table 4.3 summarizes the different ensemble integration methods.\footnote{Table adapted from Mendes-Moreira et al. (2012).}
<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>Integration Approach</th>
<th>Dynamic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Ensemble Method</td>
<td>Perrone and Cooper (1993)</td>
<td>Constant: Simple Average</td>
<td></td>
<td>Simple average of base model predictions</td>
</tr>
<tr>
<td>Bragging</td>
<td>Buehlmann (2010)</td>
<td>Constant: Median</td>
<td></td>
<td>Median of base model predictions</td>
</tr>
<tr>
<td>Dynamic Fusion</td>
<td>Woods et al. (1997)</td>
<td>Constant: Selection</td>
<td>✓</td>
<td>The n best models form the ensemble</td>
</tr>
<tr>
<td>Dynamic Selection</td>
<td>Rooney et al. (2004)</td>
<td>Constant: Selection</td>
<td>✓</td>
<td>Model with least errors on k-nearest neighbors set selected</td>
</tr>
<tr>
<td>Generalized Ensemble Method</td>
<td>Perrone and Cooper (1993)</td>
<td>Constant: Weighted Average</td>
<td></td>
<td>Model weights are inversely proportional to the error in the validation set and sum to 1</td>
</tr>
<tr>
<td>Linear Regression Model</td>
<td>Breiman (1996c)</td>
<td>Constant: Weighted Average</td>
<td></td>
<td>Model weights are inversely proportional to the error in the validation set</td>
</tr>
<tr>
<td>Accuracy Weighted Ensembles</td>
<td>Wang et al. (2003)</td>
<td>Constant: Weighted Average</td>
<td></td>
<td>Model weights are inversely proportional to the expected prediction error</td>
</tr>
<tr>
<td>Forward Selection with Replacement</td>
<td>Caruana and Niculescu-Mizil (2004)</td>
<td>Constant: Weighted Average</td>
<td></td>
<td>It selects those models (potentially multiple times) that improve ensemble accuracy and then averages over all selected models</td>
</tr>
<tr>
<td>Singular Value Decomposition</td>
<td>Merz (1998)</td>
<td>Constant: Weighted Average</td>
<td></td>
<td>SVD considers correlation of model predictions and builds weighted average of uncorrelated models</td>
</tr>
<tr>
<td>Dynamic Integration Algorithm</td>
<td>Puuronen et al. (1999)</td>
<td>Constant: Weighted Average</td>
<td>✓</td>
<td>Model weights based on test data prediction accuracy</td>
</tr>
<tr>
<td>Dynamic Weighting</td>
<td>Rooney et al. (2004)</td>
<td>Constant: Weighted Average</td>
<td>✓</td>
<td>Model weights are inversely proportional to the error in the k-nearest neighbor set</td>
</tr>
<tr>
<td>Dynamic Weighting with Selection</td>
<td>Rooney et al. (2004)</td>
<td>Constant: Selection and Weighted Average</td>
<td>✓</td>
<td>Like dynamic weighting, additionally removes models with an error above a predefined threshold</td>
</tr>
<tr>
<td>Static Selection</td>
<td>Kuncheva (2002)</td>
<td>Non-Constant: Selection</td>
<td></td>
<td>Tree helps to select the appropriate model</td>
</tr>
<tr>
<td>Meta Decision Trees</td>
<td>Todorovski and Džeroski (2003)</td>
<td>Non-Constant: Selection</td>
<td></td>
<td>Like dynamic weighting, additionally removes models with an error above a predefined threshold</td>
</tr>
</tbody>
</table>

Table 4.3: Overview of Ensemble Integration Methods
4.7 Summary

This section has reviewed the literature on ensemble methods. We started with an explanation of the rationale behind ensemble methods and why they are successful. We then detailed the three steps of constructing an ensemble: ensemble generation, ensemble pruning and ensemble integration. For each of the three steps we have reviewed popular methods closely following the extensive reviews by Mendes-Moreira et al. (2012) and Mendes-Moreira et al. (2007). We are now equipped to introduce a completely novel option pricing ensemble framework over the course of three chapters, using the models introduced in chapter 2 as base models. We will start with an outline of the theoretical framework in chapter 5, then show how to test this framework empirically in chapter 6 and finally present the encouraging results of the empirical analysis in chapter 7.
Chapter 5

Option Pricing Ensemble Framework

5.1 Introduction

Chapters 2, 3 and 4 have reviewed the literature in three different areas. While the option pricing literature advanced significantly over the last 40 years, the literature on model uncertainty and ensemble methods is still relatively limited so far. To the best of our knowledge, no attempts exist yet of combining these different areas, i.e. using ensemble methods to reduce the issues of model uncertainty for option pricing tasks.

Such a new toolkit, which could be useful for practitioners as it might reduce bias and/or variance of the option pricing forecast, will be proposed in this section and is the key theoretical contribution of this thesis. In particular, we will outline a very general theoretical framework to calibrate and combine several option pricing models using ensemble methods. Its modularity is displayed in figure 5.1 and allows for a flexible implementation in terms of asset classes, base models, pricing techniques and ensemble architecture. A practical implementation of this
Figure 5.1: Option Pricing Ensemble Framework
framework will be described in chapter 6 and tested empirically in chapter 7. The empirical analysis has been conducted in MATLAB, where we used the general option pricing framework of Kienitz and Wetterau (2012) as guideline for the general option pricing tasks. We have chosen MATLAB as implementation tool as it is an easy platform to develop models. It offers a lot of pre-built, reliable and fast optimization routines used by thousands of people in academia and the financial industry. This use of standardized mathematical software, where mathematical functions like optimisations, such as Fourier transform, random number generators are prebuilt, reduces the coding effort required to implement the proposed option pricing ensemble framework.

We have split the framework in four main steps and will describe them including their substeps in further detail below. All steps with Roman and Arabic numbers have to be executed in consecutive order (e.g. starting with I.1). Letters represent different alternatives for a step described by Arabic or Roman numbers, e.g. one can choose whether to use step II A or II B etc.

5.2 General Option Pricing Tasks

As displayed in figure 5.1, the general option pricing tasks consists of six consecutive steps:

1. Market Data
2. Filtering and Weighting
3. Model Choice
4. Pricing Technique
5. Evaluation Measure
6. Calibration

We will outline details and different alternatives for each step in the following sections.

5.2.1 Market Data

As a first step, one has to download the appropriate market information from data providers such as Reuters, Bloomberg, etc. In general, the following data points are necessary for each option which should be priced:

- Current time $t$
- Expiry date $T$ of the option
- Strike $K$ of the option
- Price level $S$ of the underlying at time $t$
- Option market price $V$ ($C$ for a call, $P$ for a put)
- Risk-free interest rate $r$ applicable for the time period $T - t$

5.2.2 Filtering and Weighting

To avoid option pricing errors, most practitioners filter out extreme $|K/S|$ levels, options with limited trading volume or wide bid-ask spreads. Furthermore, some approaches price only calls or puts and infer the prices of the other option type via put-call-parity. Also, several researchers focus only on pricing ATM or OTM options. Sometimes, different weightings are also applied, so that a higher weight is given to ATM options. However, in our framework we would suggest to include
a weighting scheme in the ensemble architecture and only filter out extreme or illiquid options. Details of the filtering rules applied in the empirical section of this thesis are outlined in section 6.2.2.

5.2.3 Model Choice

Chapter 2 has reviewed the evolution of the derivatives pricing literature extensively and hence introduced a wide range of potential base models for our ensemble approach. Starting from the benchmark models of Black and Scholes (1973) and Merton (1973), asset returns have been modeled as a continuous diffusion process with a so-called local volatility (Derman and Kani, 1994; Dupire, 1994; Rubinstein, 1994) or with a stochastic volatility process (Heston, 1993; Hull and White, 1987; Johnson and Shanno, 1987; Scott, 1977; Stein and Stein, 1991). Similar works assume constant elasticity of variance (Cox and Ross, 1976). Alternatively, models with jumps (Ahn and Thompson, 1988; Amin and Ng, 1993; Bates, 1991; Jarrow and Rosenfeld, 1984; Kou, 2002; Merton, 1976) or models with jumps and stochastic volatility (Bates, 1996a, 2000) have been proposed to capture the empirical return characteristics described above. Recent jump-models based on Lévy processes include the normal-inverse-Gaussian model (Barndorff-Nielsen, 1997), the variance-gamma model (Carr et al., 1998), generalized hyperbolic processes (Eberlein et al., 1998) and the logstable model (Carr and Wu, 2003a). Furthermore, combinations or generalizations of the above have been introduced: affine jump-diffusion models (Duffie et al., 2000) as well as models based on time-changed Lévy processes (Carr et al., 2003; Carr and Wu, 2004). Table 5.1 summarizes the option pricing models introduced in chapter 2 and represents a pool of different option pricing models one could use in an option pricing ensemble framework. This list is not complete, but includes the main
option pricing models of the different classes of option pricing models according to our view.
<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>Diffusion</th>
<th>Jumps</th>
<th>Stoch. Vol.</th>
<th>Leverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black Scholes Merton</td>
<td>Black and Scholes (1973) &amp; Merton (1973)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEV</td>
<td>Cox and Ross (1976)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dupire</td>
<td>Dupire (1994)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Derman</td>
<td>Derman and Kani (1994)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rubinstein</td>
<td>Rubinstein (1994)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson &amp; Shanno</td>
<td>Johnson and Shanno (1987)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scott</td>
<td>Scott (1987)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hull-White</td>
<td>Hull and White (1987)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stein and Stein</td>
<td>Stein and Stein (1991)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heston</td>
<td>Heston (1993)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Merton JD</td>
<td>Merton (1976)</td>
<td>✓</td>
<td></td>
<td>Poisson</td>
<td>✓</td>
</tr>
<tr>
<td>Kou</td>
<td>Kou (2002)</td>
<td>✓</td>
<td>Double Exponential</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIG</td>
<td>Barndorff-Nielsen (1997) &amp; Rydberg (1997a,b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VG</td>
<td>Madan and Seneta (1987) &amp; Carr et al. (1998)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGMY</td>
<td>Carr et al. (2002)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS</td>
<td>Carr and Wu (2003a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIG_{CIR}</td>
<td>Barndorff-Nielsen (1997) &amp; Rydberg (1997a,b) &amp; Cox et al. (1985)</td>
<td></td>
<td></td>
<td>CIR</td>
<td>✓</td>
</tr>
<tr>
<td>NIG_{OU}</td>
<td>Barndorff-Nielsen (1997) &amp; Rydberg (1997a,b) &amp; Barndorff-Nielsen and Shephard (2001b)</td>
<td></td>
<td></td>
<td>OU</td>
<td>✓</td>
</tr>
<tr>
<td>VG_{CIR}</td>
<td>Madan and Seneta (1987) &amp; Carr et al. (1998) &amp; Cox et al. (1985)</td>
<td></td>
<td></td>
<td>CIR</td>
<td>✓</td>
</tr>
<tr>
<td>VG_{OU}</td>
<td>Madan and Seneta (1987) &amp; Carr et al. (1998) &amp; Barndorff-Nielsen and Shephard (2001b)</td>
<td></td>
<td></td>
<td>OU</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 5.1: Potential Option Pricing Base Models
Opposite to the classical option pricing approach of just selecting the best option pricing model and motivated by the investigations into model uncertainty (chapter 3), our ensemble approach implies the use of multiple models to build an ensemble. Potential ways how to choose from and to combine the above mentioned (or an even more extensive list of) option pricing models will be presented in the sections on ensemble architecture later in this chapter.

5.2.4 Pricing Technique

This thesis deals with building ensembles to improve the pricing of options under Lévy processes, for which several methods exist. One could derive the partial differential equation and solve it numerically. However, this method has proven quite complex given the jump part. A second option is the use of Monte Carlo methods. Using this method, one needs to simulate a relevant number of paths of the risk-neutral process, average the final value and discount it back to the pricing date to recover the option price. A disadvantage of this method is the high computational intensity. A third method takes advantage of the fact that - although the probability density function of a Lévy process is often not known - its characteristic function always is. This is the main motivation behind Fourier based option pricing methods for exponential Lévy models.

Stein and Stein (1991) were the first to propose Fourier inversion methods in the field of finance by introducing a stochastic volatility model which uses the Fourier transform to find the distribution of the underlying. Heston (1993) used the characteristic function to obtain an analytical option pricing formula for European options, where the volatility of the underlying changes over time. Since then, Fourier transform methods have been researched actively in the financial literature. Bakshi and Madan (2000) gave an economic foundation
for characteristic functions, generalized Heston’s and Stein and Stein’s approach and developed several valuation formulae for a wide array of contingent claims. Duffie et al. (2000) came to the conclusion that Fourier methods are applicable for a wide range of stochastic processes and coined the term of exponential affine jump diffusion processes.

Carr and Madan (1999) introduced a numerically very efficient method to obtain the values of European call and put options by using Fourier transforms. Lee (2004) extended Carr and Madan’s approach, generalized it to other payoff profiles and combined it with other known Fourier pricing elements. Carr and Wu (2004) also build upon Carr and Madan (1999), broaden their methodology to general claims and apply them to time-changed Lévy processes (TCL), quadratic activity rate models (Leippold and Wu, 2002) and the class of generalized affine models (Filipović, 2001).

Lewis (2001) suggests a modular pricing framework based on the characteristic function for a given price process and a Fourier transformed payoff function which determines the contract type. This framework allows him to separate the underlying stochastic process from the option payoff by using the Plancherel - Parseval Theorem (Schmelzle, 2010).

This is in contrast to the approach of Carr and Madan (1999), where the whole option price (including the particular payoff function) is Fourier transformed.

Kienitz and Wetterau (2012) compare various pricing techniques based on Fourier transforms and direct integration of the risk neutral density. In particular, Kienitz and Wetterau (2012) present and compare the following pricing techniques in terms of stability and speed:

- Carr-Madan method (Carr and Madan, 1999)
- Lewis method (Lewis, 2001)
• Attari method (Attari, 2004)
• Tankov method (Tankov, 2006)
• Convolution method (Lord et al., 2008)
• Cosine method (Fang and Oosterlee, 2008)
• Fourier Space Time Stepping method (Jaimungal and Surkov, 2009)

A detailed analysis of all mentioned techniques is beyond the scope of this thesis. We refer to the original sources or to Kienitz and Wetterau (2012) for a detailed comparison of the different methods and will outline in chapter 6 only the Carr Madan method (Carr and Madan, 1999), as it is the one applied in the empirical section of this thesis.

5.2.5 Evaluation Measure

The pricing performance of the different option pricing models is judged based on a certain evaluation measure, which represents the difference between market and model prices and can hence be seen as a measure of fit. It is this evaluation measure that the calibration algorithm tries to minimize by changing the parameter set. We will introduce six evaluation measures below: average percentage error (APE), average absolute error (AAE), average relative percentage error (ARPE), average absolute relative percentage error (AARPE), mean squared relative error (MSRE) and root mean squared error (RMSE) following Schoutens (2003) and Huang and Wu (2004). Alternative evaluation measures and a discussion of these can be found in Huang and Wu (2004) and Schoutens (2003).

• APE: \[ \frac{1}{\text{mean option price}} \sum_{1}^{M} \frac{V - \hat{V}}{M} \]
As before, \( M \) represents the number of options priced, \( V \) the option market price and \( \hat{V} \) the option model price.

5.2.6 Calibration

In the previous sections, we have outlined how to calculate option prices for given parameters specifying jump and volatility components. For empirical investigations, we have to specify these parameters. There are two ways to do so: We could either analyze time series data of the underlying to specify these parameters or use the market prices of liquid options to derive these parameters. In our empirical section (chapters 6 and 7), we calibrate the parameters based on market prices of listed European options on the DAX index.

In general, the calibration process is an inverse problem: we have a parametrised model to price a given financial instrument; we then take some market prices and try to find the set of model parameters which minimizes the difference between model and market prices. It is an inverse problem as we know how to get from the option pricing model to the prices but not how to get from the market prices to the corresponding model. Gupta (2009) investigates the calibration problem/parameter uncertainty as well as a Bayesian approach to it in detail in his doctoral thesis.
As we are searching for a parameter vector that fits the model prices to markets prices, the calibration procedure can be seen as an optimization problem. Kienitz and Wetterau (2012) summarize various optimization methods to find the parameter set with the minimal difference between market and model prices. They describe the following six calibration methods:

- Downhill Simplex method (Nelder and Mead, 1965)
- Damped Gauss-Newton method (Levenberg, 1944; Marquardt, 1963)
- L-BFGS Quasi-Newton method (Nocedal, 1980)
- Sequential Quadratic Programming (Geiger and Kanzow, 2002)
- Differential Evolution (Storn and Price, 1997)
- Simulated Annealing (Metropolis et al., 1953)

A detailed analysis of these techniques is beyond the scope of this thesis. We refer to the original sources or to Kienitz and Wetterau (2012) for a very good comparison of the different methods and will outline in chapter 6 only the least squared errors method following Cont and Tankov (2004) and Lindstrom (2008), as it is the most widely used one in the financial literature (Bates, 1996b; Cont and Tankov, 2004; Hull and Suo, 2002; Schoutens et al., 2004; Lindstrom, 2008) and therefore the one applied in the empirical section of this thesis.

5.3 Ensemble Generation

As seen in chapter 4, the ensemble process can be partitioned into three steps: ensemble generation, ensemble pruning and ensemble integration. In our context, the step of ensemble generation (section 4.4) includes the calibration of several
candidate option pricing models to existing market prices. The second step, ensemble pruning (section 4.5), will remove some of the models calibrated in step 1, e.g. due to unsatisfactory pricing performance or similar criteria. The third and last step of the process, ensemble integration (section 4.6), defines the aggregation scheme of the models which have not been eliminated in step 2. At the end of step 3, we will obtain the ensemble prediction (denoted by $\hat{V}_{\text{ens}}$) of option prices. Table 5.2 summarizes the ensemble generation methods reviewed in section 4.4 and represents a pool of different ensemble generation methods one could use in an option pricing ensemble framework. This list is not complete, but includes the mayor ensemble generation methods according to our view.
<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>Approach</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bagging</td>
<td>Breiman (1996a)</td>
<td>Subsampling</td>
<td>Constructs multiple data sets by randomly selecting subsets</td>
</tr>
<tr>
<td>Forward Stage-Wise Additive Model</td>
<td>Friedman (2001)</td>
<td>Subsampling</td>
<td>Modifies Adaboost by minimizing residuals of previous round predictions</td>
</tr>
<tr>
<td>Multiple Additive Regression Trees</td>
<td>Friedman (2002)</td>
<td>Subsampling</td>
<td>Modifies Adaboost by minimizing stochastic residuals of previous round predictions</td>
</tr>
<tr>
<td>AdaBoost.RT</td>
<td>Shrestha and Solomatine (2006)</td>
<td>Subsampling</td>
<td>Modifies AdaBoost classification algorithm by discretizing error into a binary one</td>
</tr>
<tr>
<td>Random Subspace</td>
<td>Ho (1998)</td>
<td>Input Feature Manipulation</td>
<td>Constructs subsample datasets by randomly selecting them</td>
</tr>
<tr>
<td>Ensemble Feature Selection</td>
<td>Opitz (1999)</td>
<td>Input Feature Manipulation</td>
<td>Aims to minimize the error of the base models and maximize the ambiguity term of the ensemble at the same time</td>
</tr>
<tr>
<td>Bootstrap Ensemble with Noise</td>
<td>Raviv and Intrator (1996)</td>
<td>Input Feature Manipulation</td>
<td>Adds Gaussian noise to data points of the training data</td>
</tr>
<tr>
<td>Input Smearing</td>
<td>Frank and Pfahringer (2006)</td>
<td>Input Feature Manipulation</td>
<td>Adds Gaussian noise to data points of the training data</td>
</tr>
<tr>
<td>Rotation Forests</td>
<td>Rodríguez et al. (2006)</td>
<td>Input Feature Manipulation</td>
<td>Creates disjoint subsets to increase the likelihood of improved diversity</td>
</tr>
<tr>
<td>Output Smearing</td>
<td>Breiman (2000)</td>
<td>Output Feature Manipulation</td>
<td>Adds Gaussian noise to target output variable</td>
</tr>
</tbody>
</table>

Table 5.2: Potential Ensemble Generation Methods
5.4 Ensemble Pruning

Ensemble pruning (section 4.5) describes the process of removing some of the models calibrated in the ensemble generation step, e.g. due to unsatisfactory pricing performance or similar criteria. It aims to improve prediction accuracy or reduce computational cost. Table 5.3 summarizes the ensemble pruning methods reviewed in section 4.5 and represents a list of potential ensemble pruning methods one could use in an option pricing ensemble framework. This list is not complete, but includes the mayor ensemble pruning methods according to our view.
<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>Approach</th>
<th>Stopping Criterion</th>
<th>Evaluation Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustering by Deterministic Annealing</td>
<td>Lazarevic and Obadovic (2001)</td>
<td>Partitioning-Based</td>
<td>Pre-defined number of clusters with best model from each cluster</td>
<td>Partitioning according to weighted distance, selection based on base model accuracy</td>
</tr>
<tr>
<td>Adaptive Radius Immune Algorithm</td>
<td>Coelho and von Zuben (2006)</td>
<td>Partitioning-Based</td>
<td>ARIA partitions similar base models into groups, best model of each cluster used in ensemble</td>
<td>Partitioning according to ARIA, selection based on base model accuracy</td>
</tr>
<tr>
<td>Correlation between Errors</td>
<td>Perrone and Cooper (1993)</td>
<td>Search-Based: Exponential</td>
<td>Pre-defined number of models</td>
<td>Mean pairwise correlation between errors</td>
</tr>
<tr>
<td>GASEN</td>
<td>Zhou et al. (2002)</td>
<td>Search-Based: Randomized</td>
<td>Pre-defined threshold for the fitness</td>
<td>Fitness ($\frac{1}{\text{error}}$) of the base models</td>
</tr>
<tr>
<td>Forward without Exploration</td>
<td>Coelho and von Zuben (2006)</td>
<td>Search-Based: Forward</td>
<td>Ensemble accuracy starts to decrease</td>
<td>Ranking of base model’s accuracy</td>
</tr>
<tr>
<td>Forward with Exploration</td>
<td>Coelho and von Zuben (2006)</td>
<td>Search-Based: Forward</td>
<td>Ensemble accuracy starts to decrease</td>
<td>Ensemble accuracy</td>
</tr>
<tr>
<td>Picking Maximum Diversity</td>
<td>Partridge and Yates (1996)</td>
<td>Search-Based: Forward</td>
<td>Pre-defined number of models is reached</td>
<td>Base model diversity</td>
</tr>
<tr>
<td>Ordering Bagging Ensembles</td>
<td>Martinez-Munoz and Suarez (2006)</td>
<td>Search-Based: Forward</td>
<td>Pre-defined number of models is reached</td>
<td>Base model with best reduction of ensemble generalization error</td>
</tr>
<tr>
<td>Backward without Exploration</td>
<td>Coelho and von Zuben (2006)</td>
<td>Search-Based: Backward</td>
<td>Ensemble accuracy starts to decrease</td>
<td>Ranking of base model’s accuracy</td>
</tr>
<tr>
<td>Backward with Exploration</td>
<td>Coelho and von Zuben (2006)</td>
<td>Search-Based: Backward</td>
<td>Ensemble accuracy starts to decrease</td>
<td>Ensemble accuracy</td>
</tr>
<tr>
<td>Greedy Search</td>
<td>Moreira (2006)</td>
<td>Search-Based: Forward-Backward</td>
<td>Last added model = last removed model</td>
<td>Ensemble accuracy</td>
</tr>
<tr>
<td>Reduce-Error Pruning with Backfitting</td>
<td>Margineantu and Dietterich (1997)</td>
<td>Search-Based: Forward-Backward</td>
<td>Pre-defined number of iterations</td>
<td>Ensemble accuracy</td>
</tr>
</tbody>
</table>

Table 5.3: Potential Ensemble Pruning Methods
5.5 Ensemble Integration

The third step of the ensemble process, ensemble integration (section 4.6), defines the aggregation scheme of the models which have not been eliminated in step 2. At the end of step 3, we will obtain the ensemble prediction (denoted by $\hat{V}_{\text{ens}}$) of option prices. Table 5.4 summarizes the ensemble integration methods reviewed in section 4.5 and suggests a number of potential ensemble integration methods one could use in an option pricing ensemble framework. This list is not complete, but includes the mayor ensemble integration methods according to our view.
<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>Integration Approach</th>
<th>Dynamic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Ensemble Method</td>
<td>Perrone and Cooper (1993)</td>
<td>Constant: Simple Average</td>
<td></td>
<td>Simple average of base model predictions</td>
</tr>
<tr>
<td>Bragging</td>
<td>Buehlmann (2010)</td>
<td>Constant: Median</td>
<td></td>
<td>Median of base model predictions</td>
</tr>
<tr>
<td>Dynamic Fusion</td>
<td>Woods et al. (1997)</td>
<td>Constant: Selection</td>
<td>✓</td>
<td>The $n$ best models form the ensemble</td>
</tr>
<tr>
<td>Dynamic Selection</td>
<td>Rooney et al. (2004)</td>
<td>Constant: Selection</td>
<td>✓</td>
<td>Model with least errors on k-nearest neighbors set selected</td>
</tr>
<tr>
<td>Generalized Ensemble Method</td>
<td>Perrone and Cooper (1993)</td>
<td>Constant: Weighted Average</td>
<td></td>
<td>Model weights are inversely proportional to the error in the validation set and sum to 1</td>
</tr>
<tr>
<td>Linear Regression Model</td>
<td>Breiman (1996c)</td>
<td>Constant: Weighted Average</td>
<td></td>
<td>Model weights are inversely proportional to the error in the validation set</td>
</tr>
<tr>
<td>Accuracy Weighted Ensembles</td>
<td>Wang et al. (2003)</td>
<td>Constant: Weighted Average</td>
<td></td>
<td>Model weights are inversely proportional to the expected prediction error</td>
</tr>
<tr>
<td>Forward Selection with Replacement</td>
<td>Caruana and Niculescu-Mizil (2004)</td>
<td>Constant: Weighted Average</td>
<td></td>
<td>It selects those models (potentially multiple times) that improve ensemble accuracy and then averages over all selected models</td>
</tr>
<tr>
<td>Singular Value Decomposition</td>
<td>Merz (1998)</td>
<td>Constant: Weighted Average</td>
<td></td>
<td>SVD considers correlation of model predictions and builds weighted average of uncorrelated models</td>
</tr>
<tr>
<td>Dynamic Integration Algorithm</td>
<td>Puuronen et al. (1999)</td>
<td>Constant: Weighted Average</td>
<td>✓</td>
<td>Model weights based on test data prediction accuracy</td>
</tr>
<tr>
<td>Dynamic Weighting</td>
<td>Rooney et al. (2004)</td>
<td>Constant: Weighted Average</td>
<td>✓</td>
<td>Model weights are inversely proportional to the error in the k-nearest neighbor set</td>
</tr>
<tr>
<td>Dynamic Weighting with Selection</td>
<td>Rooney et al. (2004)</td>
<td>Constant: Selection and Weighted Average</td>
<td>✓</td>
<td>Like dynamic weighting, additionally removes models with an error above a pre-defined threshold</td>
</tr>
<tr>
<td>Static Selection</td>
<td>Kuncheva (2002)</td>
<td>Non-Constant: Selection</td>
<td></td>
<td>Tree helps to select the appropriate model</td>
</tr>
<tr>
<td>Meta Decision Trees</td>
<td>Todorovski and Džeroski (2003)</td>
<td>Non-Constant: Selection</td>
<td></td>
<td>Like dynamic weighting, additionally removes models with an error above a pre-defined threshold</td>
</tr>
</tbody>
</table>

Table 5.4: Potential Ensemble Integration Methods
5.6 Summary

This section has introduced a very flexible option pricing ensemble framework to calibrate and combine several option pricing models using ensemble methods. The general framework is displayed in figure 5.1. Its modularity allows for a flexible implementation in terms of asset classes, base models, pricing techniques and ensemble architecture. We will now turn to a practical implementation of this framework, by first describing the exact study design we have used in chapter 6 and then presenting the results of this empirical investigation in chapter 7.
Chapter 6

Empirical Study Design

6.1 Introduction

This chapter will outline the study design we have chosen to conduct an empirical analysis of a practical implementation of the option pricing ensemble framework introduced in chapter 5 and summarized in figure 5.1. We will follow the same structure as in the previous chapter and explain for each of the four steps, which technique / method we have applied. We start with a description of the general option pricing tasks and turn then to the ensemble architecture. An analysis of every possible combination of the general option pricing tasks and ensemble architecture elements introduced in previous chapters would outsize this thesis. We will therefore focus only on several examples to indicate the potential of the suggested option pricing ensemble framework, emphasizing the ensemble pruning and ensemble integration step. We leave an investigation of other combination possibilities to future research. The empirical results of this investigation are then presented in chapter 7.
6.2 General Option Pricing Tasks

6.2.1 Market Data

The analysis of the pricing performance of several ensembles of option pricing models has been conducted with DAX index options. As most option pricing models have initially been developed for equity rather than forex or commodities, we have chosen an equity index to test the proposed option pricing ensemble framework. One advantage of index options - unlike most stock options - is that they are European style as the methods we developed in previous chapters apply to European options. Furthermore, the DAX is one of the most liquid index futures available, hence the pricing should be relatively efficient. Additionally, as the DAX is a performance index, we do not have to adjust for dividends.

The German DAX index represents the performance of the 30 largest German companies in terms of free-float adjusted market capitalization according to the Deutsche Boerse definition. The index is based on Xetra prices, the electronic trading system of Deutsche Boerse. On every trading day, the calculation of the index starts at 9.00am and ends at 5.30pm. The DAX index is the underlying of DAX options traded on the EUREX, the biggest European derivatives exchange. As DAX options are European style, they can only be exercised on the settlement day, which is usually the third Friday of the month.

Our final data set consists of 1400 DAX index call option settlement prices as well as corresponding DAX prices and EURIBOR rates (reflecting the time to maturity of the options) between Jan 3 and 14, 2011. All data has been downloaded from Bloomberg.
6.2.2 Filtering and Weighting

As far OTM and far ITM as well as very long dated options are usually quite illiquid, we have excluded the following options from our initial data set. After applying these filters, the data set is reduced to 750 options.

- Options with maturities $(T < 30, T > 730)$
- Options with extreme strikes $(|S/K - 1| > 0.35)$
- Options with bid-ask spreads greater than 15% as it is usually a good proxy for illiquidity
- Options without trading volume have also been excluded as those prices may be distorted

The Fourier method gives unsatisfactory results when maturity tends to zero and when the options are deep-in-the-money or deep-out-of-the-money (Carr and Madan, 1999). The filters introduced above should help to avoid these issues. Although other researchers (Dahlbokum, 2007; Huang and Wu, 2004) focus only on OTM options due to higher liquidity, we have used ITM, ATM and OTM call options as we applied the above mentioned filters. As indicated in chapter 5, we refrain from applying different weightings to different options at this step and consider such possibilities only in the design of the ensemble architecture.

6.2.3 Model Choice

As we use an ensemble of different base models, we highlight in blue the names of all candidate models we took into account for the ensemble architecture construction:
<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>Diffusion</th>
<th>Jumps</th>
<th>Stoch. Vol.</th>
<th>Leverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black Scholes Merton</td>
<td>Black and Scholes (1973) &amp; Merton (1973)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEV</td>
<td>Cox and Ross (1976)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dupire</td>
<td>Dupire (1994)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Derman</td>
<td>Derman and Kani (1994)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rubinstein</td>
<td>Rubinstein (1994)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson &amp; Shanno</td>
<td>Johnson and Shanno (1987)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scott</td>
<td>Scott (1987)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hull-White</td>
<td>Hull and White (1987)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stein and Stein</td>
<td>Stein and Stein (1991)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heston</td>
<td>Heston (1993)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Merton JD</td>
<td>Merton (1976)</td>
<td>✓</td>
<td></td>
<td>Poisson</td>
<td>✓</td>
</tr>
<tr>
<td>Kou</td>
<td>Kou (2002)</td>
<td>✓</td>
<td></td>
<td>Double Exponential</td>
<td></td>
</tr>
<tr>
<td>NIG</td>
<td>Barndorff-Nielsen (1997) &amp; Rydberg (1997a,b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VG</td>
<td>Madan and Seneta (1987) &amp; Carr et al. (1998)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGMY</td>
<td>Carr et al. (2002)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS</td>
<td>Carr and Wu (2003a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIG_{CIR}</td>
<td>Barndorff-Nielsen (1997) &amp; Rydberg (1997a,b) &amp; Cox et al. (1985)</td>
<td></td>
<td></td>
<td>CIR</td>
<td>✓</td>
</tr>
<tr>
<td>NIG_{OU}</td>
<td>Barndorff-Nielsen (1997) &amp; Rydberg (1997a,b) &amp; Barndorff-Nielsen and Shephard (2001b)</td>
<td></td>
<td></td>
<td>OU</td>
<td>✓</td>
</tr>
<tr>
<td>VG_{CIR}</td>
<td>Madan and Seneta (1987) &amp; Carr et al. (1998) &amp; Cox et al. (1985)</td>
<td></td>
<td></td>
<td>CIR</td>
<td>✓</td>
</tr>
<tr>
<td>VG_{OU}</td>
<td>Madan and Seneta (1987) &amp; Carr et al. (1998) &amp; Barndorff-Nielsen and Shephard (2001b)</td>
<td></td>
<td></td>
<td>OU</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 6.1: Analyzed Option Pricing Models
6.2.4 Pricing Technique

We have listed examples of various option pricing techniques in section 5.2.4. For the Lévy Process models mentioned above, we use the Carr Madan method (Carr and Madan, 1999) in the empirical section of this thesis as this is the most widely used on in the option pricing literature (Kienitz and Wetterau, 2012). Hence, we outline below a short overview (following Carr and Madan (1999) and Dahlbokum (2010)) of the underlying theory and show how to use characteristic functions and Fourier transforms to price options. The idea behind Fourier based option pricing is to express the Fourier transform of the option price through the underlying’s characteristic function (Dahlbokum, 2010). We refer to Matsuda (2004) for an introduction to option pricing via Fourrier transforms and to Carr et al. (2003) for an extensive list of references to other articles using Fourier transforms.

Carr and Madan (1999) define \( k \) as the log of the strike price \( K \) and the risk neutral density of the log price \( s_T \) as \( q_T(s) \). The value of a European call option (such as DAX options analyzed in chapter 7) maturing at time \( T \) as a function of \( k \) is given by:

\[
C_T(k) = \int_{k}^{\infty} e^{-rT}(e^s - e^k)q_T(s)ds
\]  

(6.1)

Since \( C_T(k) \) tends to \( S_0 \) as \( k \) tends to \(-\infty\), the call pricing function is not square integrable in the log strike direction. Hence, one has to use a suitable transformation. Carr and Madan (1999) propose the following modified call price function:

\[
e_T(k) = e^{\alpha k}C_T(k),
\]  

(6.2)

with \( C(k) \) being the price of the call, \( k \) being the log-strike and \( \alpha \) being the dampening factor. The modified call price \( e_T(k) \) is square integrable for a range
of \( \alpha > 0 \) values and \( \forall k \). Dahlbokum (2007) investigates the use of several alternatives to the Madan and Carr approach, which are computationally more efficient, but as the Carr and Madan approach remains the benchmark in academia and industry and we are not primarily concerned with computational performance, we will use the Carr and Madan approach. The Fourier transform and the inverse Fourier transform of the modified call price are defined as:

\[
F_{c_T}(\nu) = \psi_T(\nu) = \int_{-\infty}^{\infty} e^{i\nu k} c_T(k) dk
\]  

(6.3)

\[
c_T(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\nu k} F_{c_T}(\nu) d\nu = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\nu k} \psi_T(\nu)
\]  

(6.4)

Therefore:

\[
C_T(k) = e^{-\alpha k} c_T(k) = e^{-\alpha k} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\nu k} \psi_T(\nu)
\]  

(6.5)

Since:

\[
\int_{-\infty}^{\infty} e^{-i\nu k} \psi_T(\nu) = \int_{0}^{\infty} e^{-i\nu k} \psi_T(\nu) + \int_{-\infty}^{0} e^{-i\nu k} \psi_T(\nu)
\]  

(6.6)

\[
= \int_{0}^{\infty} e^{-i\nu k} \psi_T(\nu) + \int_{0}^{\infty} e^{-i\nu k} \psi_T(-\nu)
\]

As we have \( \psi_T(\nu) = \psi_T(-\nu) \) when \( \psi_T(\nu) \) is symmetric, it follows that:

\[
\int_{-\infty}^{\infty} e^{-i\nu k} \psi_T(\nu) = 2Re \int_{0}^{\infty} e^{-i\nu k} \psi_T(\nu)
\]  

(6.7)

where \( Re(.) \) stands for the real part. As \( C_T(k) \) is real, we have:

\[
C_T(k) = \frac{e^{-\alpha k}}{\pi} \int_{0}^{\infty} e^{-i\nu k} \psi_T(\nu)
\]  

(6.8)

Now all we have to do is to retrieve \( \psi_T(\nu) \) in analytical form, enter it in equation (6.8) and we get back the option price. Carr and Madan (1999) obtain the
following analytical form for $\psi_T(\nu)$:

$$
\psi_T(\nu) = \frac{e^{-rT} \phi_T(\nu - (\alpha + 1)i)}{\alpha^2 + \alpha - \nu^2 + i(2\alpha + 1)\nu}
$$

(6.9)

We will outline below how to derive this result based on DeVille (2008) and Matsuda (2004). Combining equations (6.1), (6.2) and (6.3):

$$
\psi_T(\nu) = \int_{-\infty}^{\infty} e^{ik} \int_{-\infty}^{s} e^{ak} e^{-rT}(e^s - e^k)q_T(s) ds dk = \int_{-\infty}^{\infty} e^{-rT} q_T(s) \left[ \int_{-\infty}^{s} e^{s+\alpha+i\nu}k dk - \int_{-\infty}^{s} e^{(1+\alpha+i\nu)k} dk \right] ds.
$$

(6.10)

As the following equation holds:

$$
\int_{a}^{b} f'(x) e^{f(x)} dx = e^{f(x)}
$$

(6.11)

we can rewrite the following two terms:

$$
\int_{-\infty}^{s} e^{s+\alpha+i\nu}k dk = \left[ \frac{e^{s+\alpha+i\nu}k}{\alpha + i\nu} \right]_{-\infty}^{s} = \frac{e^{s+\alpha+i\nu}s}{\alpha + i\nu} - \lim_{k \to -\infty} \left( \frac{e^{s+\alpha+i\nu}k}{\alpha + i\nu} \right) = \frac{e^{(1+\alpha+i\nu)s}}{\alpha + i\nu}
$$

$$
\int_{-\infty}^{s} e^{(1+\alpha+i\nu)k} dk = \left[ \frac{e^{1+\alpha+i\nu}k}{\alpha + 1 + i\nu} \right]_{-\infty}^{s} = \frac{e^{1+\alpha+i\nu}s}{\alpha + 1 + i\nu} - \lim_{k \to -\infty} \left( \frac{e^{1+\alpha+i\nu}k}{\alpha + 1 + i\nu} \right) = \frac{e^{(1+\alpha+i\nu)s}}{\alpha + 1 + i\nu}
$$
and obtain:

\[
\psi_T(\nu) = \int_{-\infty}^{\infty} e^{-rT} q_T(s) \left[ \frac{e^{(1+\alpha+i\nu)s}}{\alpha + i\nu} - \frac{e^{(1+\alpha+i\nu)s}}{\alpha + 1 + i\nu} \right] ds \\
= \int_{-\infty}^{\infty} e^{-rT} q_T(s) \left[ \frac{(1 + \alpha + i\nu)e^{(1+\alpha+i\nu)s} - (\alpha + i\nu)e^{(1+\alpha+i\nu)s}}{e^{(1+\alpha+i\nu)s}(\alpha + 1 + i\nu)} \right] ds \\
= \int_{-\infty}^{\infty} e^{-rT} q_T(s) \left[ \frac{e^{(1+\alpha+i\nu)s}}{\alpha^2 + \alpha + \nu^2 + i(2\alpha + 1)\nu} \right] ds \\
= \frac{e^{-rT}}{\alpha^2 + \alpha + \nu^2 + i(2\alpha + 1)\nu} \int_{-\infty}^{\infty} q_T(s)e^{(1+\alpha+i\nu)s} ds.
\]

Furthermore, we can see that the following relation must hold:

\[
(1 + \alpha + i\nu)s = \left( \frac{\alpha i}{i} + \frac{i}{i} + i\nu \right)s \\
= i \left( \frac{\alpha}{i} + \frac{1}{i} + \nu \right)s \\
= i (-i\alpha - i + \nu)s \\
= i (\nu - (\alpha + 1)i)s.
\]

As we have:

\[
\phi_T(\nu) = \int_{-\infty}^{\infty} q_T(s)e^{i\nu s} ds
\]

and

\[
\psi_T(\nu) = \frac{e^{-rT}}{\alpha^2 + \alpha + \nu^2 + i(2\alpha + 1)\nu} \int_{-\infty}^{\infty} q_T(s)e^{(1+\alpha+i\nu)s} ds
\]

we can obtain the following analytical expression given by [Carr and Madan (1999)]:

\[
\psi_T(\nu) = \frac{e^{-rT}\psi_T(\nu - (\alpha + 1)i)}{\alpha^2 + \alpha + \nu^2 + i(2\alpha + 1)\nu}.
\]

This concludes the proof. After demonstrating how one can price options using the Fourier transform, we will turn next to the evaluation measure according to which we will calibrate the option pricing models.
6.2.5 Evaluation Measure

We have outlined several evaluation measures in chapter 5. In the empirical section, the pricing performance of the different models is judged based on the average absolute relative percentage error over the analyzed time period. Hence, the calibration process aims to minimize the following formula (Schoutens, 2003):

- **AARPE:** \( \frac{1}{M} \sum_{1}^{M} |\frac{V - \hat{V}}{V}| \)

We will now turn to the calibration technique used to minimize the ARPE.

6.2.6 Calibration

Least squared errors methods are the dominating calibration method in the financial literature (Bates, 1996b; Cont and Tankov, 2004; Hull and Suo, 2002; Schoutens et al., 2004; Lindstrom, 2008). In this section, we will follow closely Cont and Tankov (2004) and Lindstrom (2008) to explain the calibration issue in more detail.

The concrete calibration procedure is relatively simple. One starts with collecting \( M \) options and their prices on the respective underlying at the same time. These options have different maturities and strike levels. We define \( V_j \) as the \( j \)th option’s market price and \( \hat{V}_{\theta_i,j} \) represents the corresponding model price of model \( \theta_i \). In total we have \( N \) different models. We then search for each model the parameter set \( \theta_i \) which minimizes the following expression:

\[
\sum_{i=1}^{N} \sum_{j=1}^{M} \left( V_j - \hat{V}_{\theta_i,j} \right)^2
\]  
(6.17)

This is equivalent to minimizing the sum of the squared residuals (difference...
between market and model prices) by changing the parameters of the specific model. To find these parameter sets, we use gradient decent as an optimization algorithm. To determine a local minimum using this algorithm, one takes steps proportional to the minus gradient of the respective function at the current point. Gradient descent is based on the observation that if the real-valued function \( f(x) \) is defined and differentiable in the neighborhood of a given point \( x_0 \), then \( f(x) \) will decrease the fastest when one goes from \( x_0 \) towards the direction of the minus gradient of \( f \) at \( x_0 \):

\[
x_1 = x_0 - \epsilon \nabla f(x_0)
\]  

by choosing \( \epsilon > 0 \), \( f(x_0) \geq f(x_1) \). One starts with an initial guess of \( x_0 \) for a local minimum of \( f \) and then generates the sequence \( x_0, x_1, x_2, \ldots \) via the following equation:

\[
x_{n+1} = x_n - \epsilon \nabla f(x_n), \quad n = 0, 1, 2, \ldots
\]  

A local minimum is reached when the gradient tends to zero. In our analysis, we have fixed the value of the step size \( \epsilon \), but other algorithms change the step size deterministically or stochastically after each step (Lindstrom 2008). In our context of option pricing, the function to minimize is the difference between market and model prices. \( x_0 \) represents a first guess of a vector of model parameters which could do so. Since the analytic form of the function \( f \) is not known, we approximate its gradient by using the finite difference gradient, resulting in the following approximation for the gradient:

\[
\frac{\partial f(x)}{\partial x} = \frac{f(x + h) - f(x)}{h}
\]  

with \( h > 0 \) determining the step size of the finite-difference gradient procedure. Hence, we have to choose two step sizes in a gradient descent algorithm, \( h \) (the
finite-difference procedure step) and $\epsilon$ (the step size of the gradient descent procedure).

Although this process seems straightforward, there are several issues when calibrating option pricing models to market prices (see also chapter 3). One of them is the lack of a single quoted market price. Instead, we can only observe bid and ask prices; a common approach to mitigate this issue (used also in our analysis) is approximating the market price by the mid price of these two quotes. However, this adds to the risk of overfitting as any parameter vector yielding option values within the bid-ask spread would be acceptable with respect to the no-arbitrage conditions. This risk of overfitting is even more pronounced for complex models where many parameters have to be calibrated.

Furthermore, most option pricing models rely on a parametric model for the underlying. Such a process, even if quite advanced, will only be a proxy for the real process driving the underlying and hence unable to capture all finer details of the price process.

A similar issue is related to the evolving nature of market properties over time, e.g. due to macro-economic events or changing investor preferences, which result in time varying parameters. The required adaptive calibration is easier to apply for simpler models, what makes them more popular among market participants. Furthermore, one can apply data mining techniques if the main goal is the interpolation or prediction of future prices. However, if the main purpose is hedging, risk management or exotic option pricing, good parameter estimates are needed. [Cont and Tankov (2004)] highlight the issue that least squared error methods tend to provide good in sample predictions, but are usually not robust to outliers giving bad parameter estimates. Another common issue with least squared errors
calibration methods is multimodal loss functions, which make numerical optimization difficult. Although several algorithms with additional penalties have been suggested to reduce this problem and usually improve numerical stability, they generally do not improve the quality of the parameter estimates (Dahlbokum, 2010).

As we have shown above, it is not clear whether perfect replication of market prices is possible in practice. This is why most approaches focus on minimizing the evaluation measure. We will now turn to the ensemble architecture used in the empirical analysis before applying the described calibration and Fourier transform methods to the option pricing models outlined previously as ensemble base models.

6.3 Ensemble Generation

Ensemble generation methods can be classified according to how they attempt to generate different base models: either through manipulating the data or through manipulating the modeling process (Mendes-Moreira et al., 2012). Data manipulation methods have been analyzed in section 4.4 and can be broken down into subsampling, input feature or output feature manipulation. As we do not want to change the option pricing base models, we have not analyzed model manipulation in section 4.4. Furthermore, our main intention in this thesis is to demonstrate the general applicability of ensemble methods to improve the bias or variance of recent option pricing models on a given set of training data (option market prices). As we do not aim to extend the more recent models nor manipulate the data (as a practitioner is interested in real option prices, not manipulated ones), we have analysed only the probably most popular ensemble generation method.
and focus instead more on the ensemble pruning and integration steps. However, we encourage an investigation of the other described ensemble generation methods in future research publications. We highlight in blue the ensemble generation method we tested empirically in this thesis.
<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>Approach</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bagging</td>
<td>Breiman (1996a)</td>
<td>Subsampling</td>
<td>Constructs multiple data sets by randomly selecting subsets</td>
</tr>
<tr>
<td>Forward Stage-Wise Additive Model</td>
<td>Friedman (2001)</td>
<td>Subsampling</td>
<td>Modifies Adaboost by minimizing residuals of previous round predictions</td>
</tr>
<tr>
<td>Multiple Additive Regression Trees</td>
<td>Friedman (2002)</td>
<td>Subsampling</td>
<td>Modifies Adaboost by minimizing stochastic residuals of previous round predictions</td>
</tr>
<tr>
<td>AdaBoost.RT</td>
<td>Shrestha and Solomatine (2006)</td>
<td>Subsampling</td>
<td>Modifies AdaBoost classification algorithm by discretizing error into a binary one</td>
</tr>
<tr>
<td>Random Subspace</td>
<td>Ho (1998)</td>
<td>Input Feature Manipulation</td>
<td>Constructs subsample datasets by randomly selecting them</td>
</tr>
<tr>
<td>Ensemble Feature Selection</td>
<td>Opitz (1999)</td>
<td>Input Feature Manipulation</td>
<td>Aims to minimize the error of the base models and maximize the ambiguity term of the ensemble at the same time</td>
</tr>
<tr>
<td>Bootstrap Ensemble with Noise</td>
<td>Raviv and Intrator (1996)</td>
<td>Input Feature Manipulation</td>
<td>Adds Gaussian noise to data points of the training data</td>
</tr>
<tr>
<td>Input Smearing</td>
<td>Frank and Pfahringer (2006)</td>
<td>Input Feature Manipulation</td>
<td>Adds Gaussian noise to data points of the training data</td>
</tr>
<tr>
<td>Rotation Forests</td>
<td>Rodríguez et al. (2006)</td>
<td>Input Feature Manipulation</td>
<td>Creates disjoint subsets to increase the likelihood of improved diversity</td>
</tr>
<tr>
<td>Output Smearing</td>
<td>Breiman (2000)</td>
<td>Output Feature Manipulation</td>
<td>Adds Gaussian noise to target output variable</td>
</tr>
</tbody>
</table>

Table 6.2: Analyzed Ensemble Generation Methods
6.4 Ensemble Pruning

Ensemble pruning (section 4.5) describes the process of removing some of the models calibrated in the ensemble generation step. It aims to improve prediction accuracy or reduce computational cost. As we investigate different ensemble architectures to price options, we highlight in blue the names of all ensemble pruning methods we analysed empirically in this thesis:
<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>Approach</th>
<th>Stoping Criterion</th>
<th>Evaluation Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustering by Deterministic Annealing</td>
<td>Lazarevic and Obradovic [2001]</td>
<td>Partitioning-Based</td>
<td>Pre-defined number of clusters with best model from each cluster</td>
<td>Partitioning according to weighted distance, selection based on base model accuracy</td>
</tr>
<tr>
<td>Adaptive Radius Immune Algorithm</td>
<td>Coelho and von Zuben [2006]</td>
<td>Partitioning-Based</td>
<td>ARIA partitions similar base models into groups, best model of each cluster used in ensemble</td>
<td>Partitioning according to ARIA, selection based on base model accuracy</td>
</tr>
<tr>
<td>Correlation between Errors</td>
<td>Perrone and Cooper [1993]</td>
<td>Search-Based: Exponential</td>
<td>Pre-defined number of models</td>
<td>Mean pairwise correlation between errors</td>
</tr>
<tr>
<td>GASEN</td>
<td>Zhou et al. [2002]</td>
<td>Search-Based: Randomized</td>
<td>Pre-defined threshold for the fitness</td>
<td>Fitness ($\frac{1}{\text{error}}$) of the base models</td>
</tr>
<tr>
<td>Forward without Exploration</td>
<td>Coelho and von Zuben [2006]</td>
<td>Search-Based: Forward</td>
<td>Ensemble accuracy starts to decrease</td>
<td>Ranking of base model’s accuracy</td>
</tr>
<tr>
<td>Forward with Exploration</td>
<td>Coelho and von Zuben [2006]</td>
<td>Search-Based: Forward</td>
<td>Ensemble accuracy starts to decrease</td>
<td>Ensemble accuracy</td>
</tr>
<tr>
<td>Picking Maximum Diversity</td>
<td>Partridge and Yates [1996]</td>
<td>Search-Based: Forward</td>
<td>Pre-defined number of models is reached</td>
<td>Base model diversity</td>
</tr>
<tr>
<td>Ordering Bagging Ensembles</td>
<td>Martinez-Munoz and Suárez [2006]</td>
<td>Search-Based: Forward</td>
<td>Pre-defined number of models is reached</td>
<td>Base model with best reduction of ensemble generalization error</td>
</tr>
<tr>
<td>Backward without Exploration</td>
<td>Coelho and von Zuben [2006]</td>
<td>Search-Based: Backward</td>
<td>Ensemble accuracy starts to decrease</td>
<td>Ranking of base model’s accuracy</td>
</tr>
<tr>
<td>Backward with Exploration</td>
<td>Coelho and von Zuben [2006]</td>
<td>Search-Based: Backward</td>
<td>Ensemble accuracy starts to decrease</td>
<td>Ensemble accuracy</td>
</tr>
<tr>
<td>Greedy Search</td>
<td>Moreira [2006]</td>
<td>Search-Based: Forward-Backward</td>
<td>Last added model = last removed model</td>
<td>Ensemble accuracy</td>
</tr>
<tr>
<td>Reduce-Error Pruning with Backfitting</td>
<td>Margineantu and Dietterich [1997]</td>
<td>Search-Based: Forward-Backward</td>
<td>Pre-defined number of iterations</td>
<td>Ensemble accuracy</td>
</tr>
</tbody>
</table>

Table 6.3: Analyzed Ensemble Pruning Methods
6.5 Ensemble Integration

The third step of the ensemble process, ensemble integration (section 4.6), defines the aggregation scheme of the models which have not been eliminated in step 2. At the end of step 3, we will obtain the ensemble prediction (denoted by $\hat{V}_{\text{ens}}$) of option prices. As we investigate different ensemble architectures to price options, we highlight in blue the names of all ensemble integration methods we investigated empirically in this thesis:
<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>Integration Approach</th>
<th>Dynamic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Ensemble Method</td>
<td>Perrone and Cooper (1993)</td>
<td>Constant: Simple Average</td>
<td></td>
<td>Simple average of base model predictions</td>
</tr>
<tr>
<td>Bragging</td>
<td>Buehlmann (2010)</td>
<td>Constant: Median</td>
<td></td>
<td>Median of base model predictions</td>
</tr>
<tr>
<td>Dynamic Fusion</td>
<td>Woods et al. (1997)</td>
<td>Constant: Selection ✓</td>
<td></td>
<td>The n best models form the ensemble</td>
</tr>
<tr>
<td>Dynamic Selection</td>
<td>Rooney et al. (2004)</td>
<td>Constant: Selection ✓</td>
<td></td>
<td>Model with least errors on k-nearest neighbors set selected</td>
</tr>
<tr>
<td>Generalized Ensemble Method</td>
<td>Perrone and Cooper (1993)</td>
<td>Constant: Weighted Average</td>
<td></td>
<td>Model weights are inversely proportional to the error in the validation set</td>
</tr>
<tr>
<td>Linear Regression Model</td>
<td>Breiman (1996c)</td>
<td>Constant: Weighted Average</td>
<td></td>
<td>Model weights are inversely proportional to the error in the validation set</td>
</tr>
<tr>
<td>Accuracy Weighted Ensembles</td>
<td>Wang et al. (2003)</td>
<td>Constant: Weighted Average</td>
<td></td>
<td>Model weights are inversely proportional to the expected prediction error</td>
</tr>
<tr>
<td>Forward Selection with Replacement</td>
<td>Caruana and Niculescu-Mizil (2004)</td>
<td>Constant: Weighted Average</td>
<td></td>
<td>It selects those models (potentially multiple times) that improve ensemble accuracy and then averages over all selected models</td>
</tr>
<tr>
<td>Singular Value Decomposition</td>
<td>Merz (1998)</td>
<td>Constant: Weighted Average</td>
<td></td>
<td>SVD considers correlation of model predictions and builds weighted average of uncorrelated models</td>
</tr>
<tr>
<td>Dynamic Integration Algorithm</td>
<td>Puuronen et al. (1999)</td>
<td>Constant: Weighted Average ✓</td>
<td></td>
<td>Model weights based on test data prediction accuracy</td>
</tr>
<tr>
<td>Dynamic Weighting</td>
<td>Rooney et al. (2004)</td>
<td>Constant: Weighted Average ✓</td>
<td></td>
<td>Model weights are inversely proportional to the error in the k-nearest neighbor set</td>
</tr>
<tr>
<td>Dynamic Weighting with Selection</td>
<td>Rooney et al. (2004)</td>
<td>Constant: Selection and Weighted Average ✓</td>
<td></td>
<td>Like dynamic weighting, additionally removes models with an error above a predefined threshold</td>
</tr>
<tr>
<td>Static Selection</td>
<td>Kuncheva (2002)</td>
<td>Non-Constant: Selection</td>
<td></td>
<td>Tree helps to select the appropriate model</td>
</tr>
<tr>
<td>Meta Decision Trees</td>
<td>Todorovski and Džeroski (2003)</td>
<td>Non-Constant: Selection</td>
<td></td>
<td>Like dynamic weighting, additionally removes models with an error above a pre-defined threshold</td>
</tr>
</tbody>
</table>

Table 6.4: Analyzed Ensemble Integration Methods
6.6 Summary

This chapter described the study design we have chosen to conduct an empirical analysis of the option pricing ensemble framework introduced in chapter 5. We have outlined which methods we have used for the general option pricing tasks and described our design of the ensemble architecture, where we put our emphasis on the ensemble pruning and integration step. This is motivated by the aim to demonstrate the general applicability of ensemble methods to the context of option pricing rather than improving the base models. We will now turn to the presentation of the empirical results.
Chapter 7

Empirical Results

This section will display the empirical results based on the study design outlined in chapter 6. We will start with in-sample-fit results for the analyzed stand alone models before turning to the results of the analyzed ensemble generation, ensemble pruning and ensemble integration methods. After this, we will review out-of-sample fit, again beginning with the results of the analyzed stand alone models before evaluating the three steps of the ensemble process.

7.1 In-Sample-Fit

7.1.1 Stand Alone Model Results

Based on the general option pricing tasks outlined in section 6.2 we have calibrated all analyzed base models and show their in-sample pricing performance below. In all graphs in this thesis, circles (’o’) represent market prices and stars (’*’) represent model or ensemble prices.
Figure 7.1: In-Sample-Fit of Analyzed Stand Alone Models
Table 7.1: In-Sample Stand-Alone Option Pricing Model Results
<table>
<thead>
<tr>
<th>Name</th>
<th>Parameter 1</th>
<th>Parameter 2</th>
<th>Parameter 3</th>
<th>Parameter 4</th>
<th>Parameter 5</th>
<th>Parameter 6</th>
<th>Parameter 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black Scholes Merton</td>
<td>$\sigma$: 0.1858</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heston</td>
<td>vInst: -1.4179</td>
<td>vLong: 3.5403</td>
<td>$\kappa$: 2.1014</td>
<td>$\omega$: 11.1607</td>
<td>$\rho$: -3.0929</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Merton JD</td>
<td>$\sigma$: 0.1569</td>
<td>$\lambda$: 0.07491</td>
<td>jumpa: -1.1985</td>
<td>jumpb: 0.2354</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kou</td>
<td>$\sigma$: 0.1583</td>
<td>$\lambda$: 0.0648</td>
<td>prob: 0.0398</td>
<td>$\eta_1$: 0.2168</td>
<td>$\eta_2$: 0.3434</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GH</td>
<td>$\alpha$: -13.7367</td>
<td>$\beta$: -7.9863</td>
<td>$\delta$: -0.0354</td>
<td>$\nu$: 1.1512</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIG</td>
<td>$\alpha$: 17.0443</td>
<td>$\beta$: -11.3358</td>
<td>$\delta$: 0.4133</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VG</td>
<td>$\sigma$: 0.18407</td>
<td>$\theta$: -0.2812</td>
<td>$\nu$: 0.3298</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGMY</td>
<td>C: 1.0000</td>
<td>G: 0.7566</td>
<td>M: 1.0006</td>
<td>Y: 0.2546</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VG_{CIR}</td>
<td>C: 2.8165</td>
<td>G: 12.9119</td>
<td>M: 3.7839</td>
<td>$\kappa$: 1.5054</td>
<td>$\eta$: -0.6571</td>
<td>$\lambda$: 0.6337</td>
<td>$y_0$: 1.6601</td>
</tr>
<tr>
<td>VG_{OU}</td>
<td>C: 10.6943</td>
<td>G: 3.6770</td>
<td>M: 7.8421</td>
<td>$\lambda$: 7.4657</td>
<td>$\alpha$: -1.2290</td>
<td>$b$: -24.1830</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.2: In-Sample Option Pricing Models: Calibrated Parameter Values
Although we can compare the pricing performance of the different models by looking at the AARPE in table ??, it is not necessary that the same ranking will extend to out-of-sample data as we will see in section 7.2. Without claiming generality, we found that all Lévy Process models were able to outperform the Black Scholes benchmark on the analyzed dataset. The detailed calibration results can be found in table 7.2. However, the most advanced time-changed Lévy Process models were not as strong as expected. Instead, the two models from the finite activity Lévy Process category performed best. These findings are in line with empirical results of Dahlbokum (2010), but not with Carr and Wu (2004). As expected, we can support the claim that more flexible distributions of the Lévy Process family are more suitable than the normal distribution to describe returns of financial securities.

7.1.2 Ensemble Generation Results

This thesis aims to show that an ensemble of many different models can help to reduce bias and/or variance relative to single model forecasts. As described in section 6.3, we have therefore only analyzed subsampling out of the many different ensemble generation methods available as these methods generally rely on alterations of the model, which is not our intention. Nevertheless, we use the example of bagging with the classical Black Scholes model to demonstrate that an ensemble based on multiple Black Scholes models generated on different subsets of the original option pricing data, can still outperform the best fit parameter set model. To do so, we have generated 10 different subsets (each subset covering 80% of the original dataset), calibrated the ideal Black Scholes parameter \( \sigma \) on each of them and combine the resulting prices of the different subsets via simple averaging. As we can see, this approach slightly outperforms the single best fit
Name | IS - AARPE | Approach | Description
--- | --- | --- | ---
Bagging | 16.0897% | Subsampling | Constructs 10 different data subsets by randomly selecting subsets and takes best parameter set of each subset
Black Scholes | 16.1575% | IS Best fit parameter results

Table 7.3: In-Sample Ensemble Generation Results

parameter calibration of the Black Scholes model.

Figure 7.3: In-Sample-Fit of Analyzed Ensemble Generation Methods
7.1.3 Ensemble Pruning Results

As outlined in section 6.4, we analyzed five different ensemble pruning methods:

- **GASEN**: This method combines all models which are better than the specified fitness criterion.

- **Forward without Exploration (FwoE)**: This method starts with an empty subset $\Theta_1$ and adds successively models in each iteration to the ensemble. It starts with the two models with the best in-sample performance and then keeps adding models according to their in-sample performance ranking until the ensemble performance starts to decrease.

- **Forward with Exploration (FwE)**: This method starts with an empty subset $\Theta_1$ and adds successively models in each iteration to the ensemble. It starts with the model with the best in-sample performance and then explores the addition of any of the remaining models. It will add the model which leads to the biggest improvement of ensemble performance and continues to add models to the subset until ensemble accuracy starts to decrease.

- **Backward without Exploration (BwoE)**: This method starts with a subset $\Theta_1$ that is equal to all candidate models and removes successively models in each iteration from the ensemble. It starts removing the model with the worst in-sample performance and then keeps removing based on the ranking of in-sample performance until the ensemble performance starts to decrease.

- **Backward with Exploration (BwE)**: This method starts with a subset $\Theta_1$ that is equal to all candidate models and removes successively models
in each iteration from the ensemble. It starts removing the model which results in the best improvement of ensemble performance and then keeps removing models as long as ensemble performance continues to improve.

We display the results below and explain the applied stopping criterion for our data set. To ensure comparability of the different methods, we have applied the basic ensemble method to integrate all models and apply the criteria that an ensemble has to consist of at least two models. BwE and FwE methods are the best performing in-sample ensemble pruning methods and result in the same AARPE as they both end up combining the same candidate models. All methods are able to significantly outperform the average AARPE of all analyzed models. GASEN, FwE and BwE are even able to outperform the best performing standalone model. Although the results are already encouraging, further improvements could be achieved by combining different methods from the three steps of the ensemble process.
<table>
<thead>
<tr>
<th>DAX level K / S</th>
<th>0.8</th>
<th>0.85</th>
<th>0.9</th>
<th>0.95</th>
<th>1</th>
<th>1.05</th>
<th>1.1</th>
<th>1.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>option price in % of S</td>
<td>5.0%</td>
<td>10.0%</td>
<td>15.0%</td>
<td>20.0%</td>
<td>25.0%</td>
<td>30.0%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 7.4: In-Sample-Fit of Analyzed Ensemble Pruning Methods
<table>
<thead>
<tr>
<th>Name</th>
<th>IS - AARPE</th>
<th>Stoping Criterion</th>
<th>Applied Stoping Criterion Comment</th>
<th>Evaluation Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>GASEN</td>
<td>2.6561%</td>
<td>Pre-defined threshold for the fitness</td>
<td>IS - Model Fitness &gt;20</td>
<td>Fitness ($\frac{1}{\text{error}}$) of the base models</td>
</tr>
<tr>
<td>FwoE</td>
<td>2.7672%</td>
<td>Ensemble accuracy starts to decrease</td>
<td>Ensemble accuracy decreases with addition of third model</td>
<td>Ranking of base model’s accuracy</td>
</tr>
<tr>
<td>FwE</td>
<td>2.5724%</td>
<td>Ensemble accuracy starts to decrease</td>
<td>Ensemble accuracy highest with use of Heston, Merton JD, Kou and $V_{GOU}$ models</td>
<td>Ensemble accuracy</td>
</tr>
<tr>
<td>BwoE</td>
<td>3.3003%</td>
<td>Ensemble accuracy starts to decrease</td>
<td>Ensemble accuracy highest after removing Black Scholes and CGMY model</td>
<td>Ranking of base model’s accuracy</td>
</tr>
<tr>
<td>BwE</td>
<td>2.5724%</td>
<td>Ensemble accuracy starts to decrease</td>
<td>Ensemble accuracy highest with use of Heston, Merton JD, Kou and $V_{GOU}$ models</td>
<td>Ensemble accuracy</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>2.7738%</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.4: In-Sample Ensemble Pruning Results
7.1.4 Ensemble Integration Results

In this section, we present the results from the third step of the ensemble process, the ensemble integration step. As outlined in section 6.5, we analyzed four different ensemble pruning methods:

- **Basic Ensemble Method (BEM)**: This method combines all models by calculating the simple average of all candidate models.

- **Bragging**: This method combines all models by calculating the median of all candidate models.

- **Dynamic Fusion (DF)**: This method takes the mean of the best $n$ models to form the ensemble.

- **Generalized Ensemble Method (GEM)**: This method takes a weighted average of all models. Model weights are determined based on the error on the validation set.

As before and to ensure comparability of the different methods, we have applied only the specified integration method and not applied any ensemble generation or pruning method. All methods are able to significantly outperform the average AARPE of all analyzed models. Again, although the results are already encouraging, further improvements could be achieved by combining different methods from the three steps of the ensemble process, which is strongly suggested for future research.
Figure 7.5: In-Sample-Fit of Analyzed Ensemble Integration Methods
<table>
<thead>
<tr>
<th>Name</th>
<th>IS - AARPE</th>
<th>Integration Approach</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEM</td>
<td>4.1211%</td>
<td>Constant: Simple Average</td>
<td>Simple average of base model predictions</td>
</tr>
<tr>
<td>Bragging</td>
<td>4.2763%</td>
<td>Constant: Median</td>
<td>Median of base model predictions</td>
</tr>
<tr>
<td>DF</td>
<td>2.7879%</td>
<td>Constant: Selection</td>
<td>The top $n = 50%$ models form the ensemble</td>
</tr>
<tr>
<td>GEM</td>
<td>3.3522%</td>
<td>Constant: Weighted Average</td>
<td>Model weights are inversely proportional to the error in the validation set and sum to 1</td>
</tr>
<tr>
<td>Average</td>
<td>3.6344%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.5: In-Sample Ensemble Integration Results
7.1.5 Summary

This section has reviewed the in-sample performance of the stand-alone option pricing models as well as the three steps of the ensemble process. We have seen that all methods can outperform the average AARPE of all analyzed models significantly. We will turn now to an investigation of out-of-sample performance to compare these encouraging results to the classical approach in the literature, which is taking the calibrated model with the lowest measure of pricing error on the calibration dataset at time \( t = t_0 \) to price options of the new dataset at time \( t = t_1 \).

7.2 Out-of-Sample-Fit

7.2.1 Stand Alone Model Results

Based on the general option pricing tasks outlined in section 6.2, we have calibrated all analyzed models on in-sample data and show their out-of-sample pricing performance below. We use data from the next close of trading as out-of-sample data and can see a significant decrease (on average by more than 33%) in pricing performance except for the Black Sholes model. We can also see that the pricing performance of the models becomes more similar, with the best in-sample models having the biggest decrease in pricing performance on out-of-sample data. Although the ranking remains similar at the top and at the bottom, it is not the same as in-sample and another indicator that the models with the best in-sample performance do not necessarily need to have the best out-of-sample performance.
Figure 7.6: Out-of-Sample-Fit of Analyzed Stand Alone Models I
Figure 7.7: Out-of-Sample-Fit of Analyzed Stand Alone Models II
<table>
<thead>
<tr>
<th>Name</th>
<th># of parameters</th>
<th>OS - AARPE</th>
<th>OS - Rank</th>
<th>IS - AARPE</th>
<th>IS - Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black Scholes Merton</td>
<td>1</td>
<td>14.2306%</td>
<td>10</td>
<td>16.1575%</td>
<td>10</td>
</tr>
<tr>
<td>Heston</td>
<td>5</td>
<td>9.1348%</td>
<td>9</td>
<td>4.7524%</td>
<td>3</td>
</tr>
<tr>
<td>Merton JD</td>
<td>4</td>
<td>7.6397%</td>
<td>1</td>
<td>2.7071%</td>
<td>1</td>
</tr>
<tr>
<td>Kou</td>
<td>5</td>
<td>7.6808%</td>
<td>1</td>
<td>2.8450%</td>
<td>2</td>
</tr>
<tr>
<td>GH</td>
<td>4</td>
<td>8.0404%</td>
<td>5</td>
<td>6.1380%</td>
<td>8</td>
</tr>
<tr>
<td>NIG</td>
<td>3</td>
<td>7.8603%</td>
<td>4</td>
<td>5.2523%</td>
<td>6</td>
</tr>
<tr>
<td>VG</td>
<td>3</td>
<td>8.7156%</td>
<td>7</td>
<td>4.9642%</td>
<td>4</td>
</tr>
<tr>
<td>CGMY</td>
<td>4</td>
<td>7.8551%</td>
<td>3</td>
<td>6.4725%</td>
<td>9</td>
</tr>
<tr>
<td>VG\textsubscript{CIR}</td>
<td>7</td>
<td>8.7152%</td>
<td>6</td>
<td>5.4498%</td>
<td>7</td>
</tr>
<tr>
<td>VG\textsubscript{OU}</td>
<td>6</td>
<td>8.8199%</td>
<td>8</td>
<td>5.0419%</td>
<td>5</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td><strong>8.8692%</strong></td>
<td></td>
<td><strong>5.9793%</strong></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.6: Out-of-Sample Stand-Alone Option Pricing Model Results
7.2.2 Ensemble Generation Results

We display out-of-sample results for bagging, the only ensemble generation method we analyzed empirically. For the Black Scholes model, we see an improvement in pricing performance on out-of-sample data. However, bagging is still able to outperform the best fit in-sample parameter Black Scholes model and even improves its outperformance in relative terms.

Figure 7.8: Out-of-Sample-Fit of Analyzed Ensemble Generation Methods
<table>
<thead>
<tr>
<th>Name</th>
<th>OS - AARPE</th>
<th>IS - AARPE</th>
<th>Approach</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bagging</td>
<td>13.9092%</td>
<td>16.0897%</td>
<td>Subsampling</td>
<td>Constructs 10 different data sets by randomly selecting subsets and takes best parameter set of each subset.</td>
</tr>
<tr>
<td>Black Scholes</td>
<td>14.2306%</td>
<td>16.1575%</td>
<td></td>
<td>IS Best fit parameter results</td>
</tr>
</tbody>
</table>

Table 7.7: Out-of-Sample Ensemble Generation Results
7.2.3 Ensemble Pruning Results

We present below the out-of-sample performance of the five analyzed ensemble pruning methods described in section 7.1.3. All five ensemble pruning methods are able to outperform the out-of-sample average AARPE of the ten analyzed option pricing models, with BwoE even able to outperform the strongest out-of-sample stand-alone model.

Figure 7.9: Out-of-Sample-Fit of Analyzed Ensemble Pruning Methods
<table>
<thead>
<tr>
<th>Name</th>
<th>OS - AARPE</th>
<th>IS - AARPE</th>
<th>Stoping Criterion</th>
<th>Applied Stoping Criterion Comment</th>
<th>Evaluation Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>GASEN</td>
<td>7.6877%</td>
<td>2.6561%</td>
<td>Pre-defined threshold for the fitness</td>
<td>IS - Model Fitness &gt; 20</td>
<td>Fitness ($\frac{1}{\text{error}}$) of the base models</td>
</tr>
<tr>
<td>FwoE</td>
<td>7.6565%</td>
<td>2.7672%</td>
<td>Ensemble accuracy starts to decrease</td>
<td>IS Ensemble accuracy decreases with addition of third model</td>
<td>Ranking of base model’s accuracy</td>
</tr>
<tr>
<td>FwE</td>
<td>7.7086%</td>
<td>2.5724%</td>
<td>Ensemble accuracy starts to decrease</td>
<td>IS Ensemble accuracy highest with use of Heston, Merton JD, Kou and VG$_{OU}$ models</td>
<td>Ensemble accuracy</td>
</tr>
<tr>
<td>BwoE</td>
<td>7.6135%</td>
<td>3.3003%</td>
<td>Ensemble accuracy starts to decrease</td>
<td>IS Ensemble accuracy highest after removing Black Scholes and CGMY model</td>
<td>Ranking of base model’s accuracy</td>
</tr>
<tr>
<td>BwE</td>
<td>7.7086%</td>
<td>2.5724%</td>
<td>Ensemble accuracy starts to decrease</td>
<td>IS Ensemble accuracy highest with use of Heston, Merton JD, Kou and VG$_{OU}$ models</td>
<td>Ensemble accuracy</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>7.6750%</strong></td>
<td><strong>2.9287%</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.8: Out-of-Sample Ensemble Pruning Results
7.2.4 Ensemble Integration Results

We display the out-of-sample results for the four analyzed ensemble integration methods below. Not only are all four methods able to outperform the out-of-sample average AARPE of the ten analyzed stand-alone models, BEM, Bragging and GEM are also able to outperform the best performing stand-alone model.

Figure 7.10: Out-of-Sample-Fit of Analyzed Ensemble Integration Methods
<table>
<thead>
<tr>
<th>Name</th>
<th>IS - AARPE</th>
<th>Integration Approach</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEM</td>
<td>7.0847%</td>
<td>Constant: Simple Average</td>
<td>Simple average of base model predictions</td>
</tr>
<tr>
<td>Bragging</td>
<td>7.6239%</td>
<td>Constant: Median</td>
<td>Median of base model predictions</td>
</tr>
<tr>
<td>DF</td>
<td>7.7132%</td>
<td>Constant: Selection</td>
<td>The top $n = 50%$ models form the ensemble</td>
</tr>
<tr>
<td>GEM</td>
<td>7.1956%</td>
<td>Constant: Weighted Average</td>
<td>Model weights are inversely proportional to the error in the validation set and sum to 1</td>
</tr>
<tr>
<td>Average</td>
<td>7.4044%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.9: Out-of-Sample Ensemble Integration Results
7.2.5 Summary

This section has reviewed the out-of-sample performance of the stand-alone option pricing models as well as the three steps of the ensemble process. We have seen that all methods of the three steps lead to better results than the out-of-sample average AARPE of the ten analyzed stand-alone models, some ensemble methods are even able to outperform the best performing stand-alone model.
Chapter 8

Conclusions

8.1 Summary of Main Findings

This thesis aimed to demonstrate that the use of ensemble methods can reduce model uncertainty, enhance model robustness and improve pricing performance for derivatives pricing models based on Lévy processes. This proof of concept has been achieved. In order to do so, the thesis started with a review of the evolution of the derivatives pricing literature over time. We have seen that newer models are able to adjust for the three main shortcomings of the Black-Scholes assumptions outlined in the introduction. However, the tradeoff associated with improved fitting of observed option prices comes at the expense of analytical tractability, increased complexity and higher variance. We have then turned to an investigation of the calibration problem and model uncertainty. The current literature focuses mainly on finding the single best-fit set of parameters on in-sample data and pays little attention to measuring how robust the estimate is or whether other set of parameters can reproduce market prices equally/sufficiently well. This is a major shortcoming in the current literature and motivation for this doctoral research. Although initial investigations on parameter uncertainty have
been conducted (Gupta, 2009; Gupta et al., 2010; Gupta and Reisinger, 2011), the issue of specification uncertainty has not been analyzed extensively in the derivatives pricing literature so far.

Consequently, we analyzed the issue of parameter and specification uncertainty further and found that ensemble concepts are one potential way to deal with parameter and specification uncertainty. The idea of ensemble methods is to build a predictive model by combining multiple models, thus keeping the low bias of complex models, but reducing their variance. Ensemble methods have been used in many other disciplines as diverse as climate modelling, spam filtering, handwriting recognition and medical tests to improve prediction performance, but have received only very little attention in the derivatives pricing literature so far. In order to answer my research question: "Can Ensemble Methods Improve the Bias or Variance of Option Pricing Models based on Lévy Processes?", we have then continued with a review of ensemble methods as we examine the use of ensembles of advanced derivatives pricing models to reduce parameter and specification uncertainty and improve predictive performance.

Following the literature review of the derivatives pricing literature (chapter 2), model uncertainty (chapter 3) and ensemble methods (chapter 4), we have then constructed and tested a completely novel option pricing ensemble framework over the course of three chapters:

Chapter 5 outlined the theoretical option pricing ensemble framework which is used to help find a positive answer to the research question.

In order to test this framework empirically, chapter 6 described the specific study design we have chosen for our empirical analysis of the introduced framework. The encouraging empirical results of this study design have then been presented
in chapter\textsuperscript{7}. We demonstrated that ensemble methods can be used to improve the bias and variance of advanced option pricing models and give therefore a positive answer to the research question. We have seen that ensemble methods are not only able to outperform the average error of the analyzed stand-alone option pricing models, some of them are even able to outperform the single best performing model on in-sample as well as out-of-sample data. We can therefore give a positive answer to the research question and encourage the use of ensemble methods in the context of derivatives pricing.

8.2 Contribution to Knowledge

As the pricing of derivatives is very sensitive to changes in the chosen option pricing model or parameter set, the issue of model uncertainty is very relevant. This applies especially to practitioners as the wrong model selection might result in significant financial losses and a more accurate pricing performance than the market can lead to financial gains. We have introduced a novel theoretical option pricing ensemble framework to reduce this uncertainty and tested it empirically. This framework has been able to improve bias and variance in our empirical analysis. The contribution to knowledge of this thesis comes therefore along three dimensions:

- **Methodological**: We have introduced a novel theoretical framework how to price options. Although the ensemble concept is more demanding in terms of computational requirements, it is able to significantly outperform the current industry standard of using the single option pricing model which has the best in-sample performance. Although the idea is very simple, it is yet a powerful approach, as it is able to outperform the average model, as
one doesn’t know a priori which model will perform best.

- **Theoretical:** We have demonstrated that ensemble methods can be successfully used in the context of option pricing due to the ambiguity decomposition (section 4.2.1). The use of the introduced novel option pricing ensemble framework is therefore not only relevant from an academic perspective, but can also have a significant financial impact for practitioners. For example, I have started my own alternative asset management company based on the ideas from my thesis.

- **Empirical:** We have conducted a very extensive empirical analysis of multiple advanced option pricing models as well as ensemble methods. Just the calibration alone of such a wide range of option pricing models is very rare in the literature, considering that the calibration of some stand-alone models took up to 48 hours per model to run, which is not only due to the limited computational power of the used computer. The empirical results have shown that all investigated ensemble methods have been able to outperform the average AARPE of the ten analyzed stand-alone models on in-sample and out-of-sample data. Furthermore, several ensemble methods have been able to outperform the best performing stand-alone model on in-sample as well as out-of-sample data.

## 8.3 Suggestions for Future Research

Although this thesis has introduced a novel option pricing ensemble framework, there are several suggestions for future research:

First, the successful application of the analyzed ensemble methods should
be extended to multiple other underlyings and different time periods. Second, not all introduced option pricing models and ensemble methods could be tested empirically in this thesis. Therefore, future research should also analyze the empirical results of these option pricing models and ensemble methods. Furthermore, we have analyzed each step of the ensemble process separately. An interesting future investigation would combine successful methods of the ensemble generation, ensemble pruning and ensemble pruning steps and explore whether pricing performance could be improved even further through such a combination. Furthermore, practical issues such as the implications on hedging when using such ensemble methods should also be analyzed in future research.
Bibliography


Knight, F. (1921). *Risk, Uncertainty and Profit*.


Laplace, P. S. (1818). Deuxieme supplement a la theorie analytique des probabilites.


